Genetica (2017) 145:51–66
DOI 10.1007/s10709-016-9945-7

ORIGINAL PAPER

Complete mitochondrial genome from South American catfish *Pseudoplatystoma reticulatum* (Eigenmann & Eigenmann) and its impact in Siluriformes phylogenetic tree

Luciana Cristine Vasques Villela[1,2] · Anderson Luis Alves[1] · Eduardo Sousa Varela[1] · Michel Eduardo Beleza Yamagishi[3] · Poliana Fernanda Giachetto[3] · Naiara Milagres Augusto da Silva[4] · Josi Margarete Ponzetto[5] · Samuel Rezende Paiva[6,7] · Alexandre Rodrigues Caetano[2,4,7]

Received: 10 September 2016 / Accepted: 22 December 2016 / Published online: 3 February 2017

Abstract The cachara (*Pseudoplatystoma reticulatum*) is a Neotropical freshwater catfish from family Pimelodidae (Siluriformes) native to Brazil. The species is of relative economic importance for local aquaculture production and basic biological information is under development to help boost efforts to domesticate and raise the species in commercial systems. The complete cachara mitochondrial genome was obtained by assembling Illumina RNA-seq data from pooled samples. The full mitogenome was found to be 16,576 bp in length, showing the same basic structure, order, and genetic organization observed in other Pimelodidae, with 13 protein-coding genes, 2 rNA genes, 22 trNAs, and a control region. Observed base composition was 24.63% T, 28.47% C, 31.45% A, and 15.44% G. With the exception of *NAD6* and eight tRNAs, all of the observed mitochondrial genes were found to be coded on the H strand. A total of 107 SNPs were identified in *P. reticulatum* mtDNA, 67 of which were located in coding regions. Of these SNPs, 10 result in amino acid changes. Analysis of the obtained sequence with 94 publicly available full Siluriformes mitogenomes resulted in a phylogenetic tree that generally agreed with available phylogenetic proposals for the order. The first report of the complete *Pseudoplatystoma reticulatum* mitochondrial genome sequence revealed general gene organization, structure, content, and order similar to most vertebrates. Specific sequence and content features were observed and may have functional attributes which are now available for further investigation.

Keywords Mitogenome · Siluriformes · RNA-Seq · Phylogenetic analysis · Catfish · SNPs

Electronic supplementary material The online version of this article (doi:10.1007/s10709-016-9945-7) contains supplementary material, which is available to authorized users.

✉ Alexandre Rodrigues Caetano
alexandre.caetano@embrapa.br

[1] Embrapa Pesca e Aquicultura, CNPASA, Palmas, TO, Brazil

[2] Programa de Pós-Graduação em Ciências Animais, Faculdade de Agronomia e Veterinária, Universidade de Brasília (UnB), Brasília, DF, Brazil

[3] Embrapa Informática Agropecuária, CNPTIA, Campinas, SP, Brazil

[4] Embrapa Recursos Genéticos e Biotecnologia, CENARGEN, Brasília, DF, Brazil

[5] Programa de Pós-Graduação em Genética Evolutiva e Biologia Molecular, Centro de Ciências Biológicas e da Saúde, Universidade Federal de São Carlos (UFSCar), São Carlos, SP, Brazil

[6] Embrapa Labex EUA, Secretaria de Relações Internacionais, Brasília, DF, Brazil

[7] CNPq Fellow, Brasília, DF, Brazil

Introduction

Several species of Neotropical freshwater catfish have high commercial value and are of growing importance for production in aquaculture systems. Among these is the cachara (*Pseudoplatystoma reticulatum*, Eigenmann & Eigenmann, 1889). The biology of this species is largely unknown, despite its relative economic importance (Bignotto et al. 2009), and generation of basic biological information and tools for developing essential knowledge for rearing the species will greatly facilitate the expansion of commercial aquaculture operations.

Species of the genus *Pseudoplatystoma* (family Pimelodidae, order Siluriformes) are migratory and are distributed throughout the main South American drainage basins. *P.*

reticulatum has been found exclusively in the Paraná-Paraguay Basin. However, after a recent taxonomic review, its occurrence in the Amazon Basin has been under discussion (Buitrago-Suárez and Burr 2007; Torrico et al. 2009). Publicly available genetic information for the cachara is mostly restricted to the characterization of genetic diversity of wild populations with microsatellites (Prado et al. 2014) and partial mitochondrial and nuclear gene sequences (Bignotto et al. 2009; Torrico et al. 2009; Carvalho et al. 2011, 2012; Carvalho-Costa et al. 2011; Prado et al. 2011; Hashimoto et al. 2013; Pereira et al. 2013).

The generation of extensive genomic information for the species provided by recent developments in the field of genomics, and in particular high-throughput sequencing (NGS), will help the development of solid tools for use in phylogenetics, phylogeography, and population structure and diversity studies of wild populations. Moreover, the incorporation of genomic tools in breeding programs under development for the species will help accelerate genetic gains in traits associated with productivity, quality and profitability (Yáñez et al. 2015).

Mitochondrial DNA (mtDNA) sequence data has been widely used in studies on phylogenetics, evolution, phylogeography, and population and conservation genetics of different fish species (Avise 1994; Sosa et al. 2012; Borba et al. 2013a, b; Hahn et al. 2013). Most studies have been performed with sequencing data from fragments of a combination of mitochondrial genes. More recently, as whole mtDNA sequences become ever more available, methods to analyze complete mtDNA sequences have been proposed and are likely to yield better-supported results (Yu et al. 2007; Powell et al. 2013; Alam et al. 2014).

Thousands of species have already had their mitogenome completely sequenced and made available in public databases (Hahn et al. 2013). Out of 95 Siluriformes mitogenomes published as of October 2015, only four belong to species from the family Pimelodidae. The current study is the first to report the complete *P. reticulatum* mitochondrial genome sequence, in addition to performing the first phylogenetic analysis with previously published complete mitochondrial sequences from other Siluriformes.

Materials and methods

Sampling, sample processing, and NGS sequencing

Seven distinct tissues were sampled (white and red muscle, gills, liver, kidney, pituitary gland, and gonad) from 12 *P. reticulatum* captured at different locations within the Paraguay River Basin (Fig. S1, Supplementary material). Each individual was tested for species purity with a SNaPshot© assay (Villela et al. 2017a) developed based on previously published species-specific markers (Prado et al. 2011; Hashimoto et al. 2013).

Tissues were processed for total RNA extraction using Trizol® (Ambion), following manufacturer's instructions. Total RNA samples were combined in equimolar amounts into eight different pools: one total RNA pool for each of the seven tissues sampled and a pool containing total RNA from all sampled tissues.

cDNA libraries were constructed for each of the seven tissue-specific RNA pools using TruSeq Stranded RNA Sample Preparation Kits (Illumina, San Diego, CA, USA) following manufacturer's protocols. Average size of cDNA fragments was 250 bp, ranging from 100 to 580 bp. The seven tissue-specific libraries were sequenced on two lanes with an Illumina HiSeq 2000 platform with a protocol to produce 100 bp paired-end reads (SBS Sequencing Kit, v3). A cDNA library was also constructed with an RNA pool from all seven tissues and sequenced in a single run with an Illumina MiSeq platform, with a protocol to produce 300 bp paired-end reads (SBS Sequencing Kit, v3). Sequencing and library construction were performed at the W. M. Keck Center for Comparative and Functional Genomics at the University of Illinois, Urbana-Champaign (of *Pseudoplatystoma reticulatum* RNA samples were sent to the United States under exportation license No. 15BR016615/DF).

RNA-Seq and assembly of *P. reticulatum* mitochondrial genome

All obtained sequences were processed with the program Trimmomatic (Bolger et al. 2014) to identify and remove adapters and determine quality scores and trim sequences, removing end bases with quality scores <25. For the remaining positions, a sliding window with the same minimum quality requirements was applied. The MiSeq paired-end reads >300 bp were merged to obtain super-reads with the FLASh program (Magoc and Salzberg 2011).

De novo transcriptome assembly was performed using Trinity (Haas et al. 2013). Three assembled transcripts derived from the mtDNA were identified within the assembled transcriptome using BLAST searches with the *P. corruscans* mtDNA. These three transcripts were further processed, along with super-reads generated with MiSeq sequences, using the SSAKE program (Warren et al. 2007) to generate a final consensus mtDNA sequence.

Quality-trimmed reads were mapped to the *P. reticulatum* complete mitogenome using BWA software (Li and Durbin 2010) and SNP (*Single Nucleotide Polymorphism*) detection was performed using Freebayes (Garrison and Marth 2012). SNPs with minor allele frequencies (MAF) >0.09 in positions with Read Depth (RD) >522 were considered for further analysis and annotation.

D-Loop sequencing

Part of the *P. reticulatum* D-loop (control) region sequence was obtained by sequencing a 481 bp PCR fragment amplified from each of the same 12 individual used for RNA sampling. Amplicons were generated with primers CTC TTCCTGCACCCTACCAA and ATTGAGGGCATTCTC ACAGG with the following conditions: 30 ng of genomic DNA, 1 pmol of each primer, 5.0 µL of Qiagen Multiplex PCR plus solution, 1.0 µL of QSolution, in a final volume of 10 µL. Thermalcycling conditions were: 95 °C (15 min), followed by 35 cycles of 94 °C (30 s), 50 °C (1 min) and 72 °C (1 min), with a final step of 72 °C (30 min). PCR products purified with ExoSAP-IT were sequenced in both directions with an ABI Prism 3100 automated sequencer with BigDye Terminator chemistry, following manufacturer's instructions. Obtained sequences were assembled using Phred and Phrap (Ewing and Green 1998; Ewing et al. 1998). Individual sequences were aligned using ClustalW to identify polymorphisms and obtain a consensus sequence that was integrated with the mtDNA sequence derived from the transcriptome.

mtDNA sequence composition and annotation

The *P. reticulatum* mtDNA protein-coding genes were first annotated manually using the obtained sequence in comparisons with the *P. magdaleniatum* (NC_026526.1) and *P. corruscans* (NC_026846.1) mitochondrial genomes, using BLASTx at the NCBI site. tRNA genes and their secondary structures were identified with tRNAScan-SE (v1.21) (Lowe and Eddy 1997) using default parameters. *P. reticulatum* rRNA and $tRNA^{Ser(GCT)}$ genes were compared with *P. magdaleniatum* and *P. corruscans* mitogenomes. In addition, automated annotations of the *P. reticulatum* mitogenome were performed with MitoAnnotator (Iwasaki et al. 2013), DOGMA (Wyman et al. 2004) and MITOS (Bernt et al. 2013).

Mitochondrial genome organization

Nucleotide composition of mitogenome and protein-coding genes were determined using BioEdit (v. 7.2.5). Codon usage and RSCU values were calculated with MEGA 6.0 software (Tamura et al. 2013). Codon usage bias (CUB) and its statistical significance were characterized by Codon Deviation Coefficients (CDC). CDC, GC and purine (AG) contents were estimated in 13 protein-coding genes with the Composition Analysis Toolkit -CAT (http://cbrc.kaust.edu.sa/CAT), using the algorithm developed by Zhang et al. (2012). Strand asymmetry (AT and GC skew) was calculated according to Perna and Kocher (1995). L-strand replication origin (O_L) and H-strand replication origin (O_H) observed in the non-coding control region (D-Loop) of the *P. reticulatum* mitogenome were identified through alignments with previously annotated sequences from *P. reticulatum* and *P. magdaleniatum*. Regulatory elements were identified through alignments with D-Loop sequences from *Tropheus moorii* (Fischer et al. 2013), *Petrochromis trewavasae* (Fischer et al. 2013), *Macropodus opercularis* (Mu et al. 2015), *Oreochromis niloticus* (He et al. 2011), *Oreochromis aureus* (He et al. 2011) and *Danio rerio* (Broughton et al. 2001).

Sequence comparisons with mitogenomes from pimelodidae catfish species

The *P. reticulatum* mitogenome was compared with mitogenomes from four other catfish species from the family Pimelodidae: *P. magdaleniatum* (GenBank acc. #: NC_026526.1), *P. corruscans* (NC_026846.1), *Pimelodus pictus* (NC_015797.1), and *Sorubim cuspicaudus* (NC_026211.1), through the alignment of the sequences usingBioEdit (v. 7.2.5) and MEGA 6.0 (Tamura et al. 2013).

A 460 bp fragment of cytochrome b gene (CYTB) from 55 previously published reference sequences from eight different *Pseudoplatystoma* species (Carvalho-Costa et al. 2011; Lundberg et al. 2011; Torrico et al. 2009) was analyzed along with the obtained *P. reticulatum* mitochondrial sequence. The optimal nucleotide substitution model was identified with JModeltest 2 (Darriba et al. 2012). BEAST v.1.8.0 (Drummond et al. 2012) was used to run the Yule speciation model, considering a strict molecular clock, through 100 million MCMC iterations. Obtained results were visualized with FigTree v. 1.4.2 (tree.bio.ed.ac.uk/software/figtree/).

Phylogenetic analysis of siluriformes with complete mitogenome sequences

The full *P. reticulatum* mitochondrial obtained sequence was analyzed with all 94 publicly available Siluriformes complete mitogenomes, including samples from 29 of 36 described families (Ferraris 2007). The optimal evolutionary model selected was GTR + G + I: General Time Reversible with non-uniformity of evolutionary rates among sites using a discrete Gamma distribution (+G) with 5 rate categories and by assuming that a certain fraction of sites are evolutionarily invariable (+I). The phylogenetic tree was estimated by maximum likelihood bootstrap analysis (10,000 replications) performed with MEGA v6 (Tamura et al. 2013).

Results and discussion

RNA-Seq and assembly of *P. reticulatum* mitochondrial genome

Reciprocal interspecific hybrids between *P. reticulatum* and *P. corruscans* (pintado) are commonly produced for commercial rearing in aquaculture operations (Porto-Foresti et al. 2008). Hybrids are fertile (Prado et al. 2012) and can naturally backcross to purebred individuals in the wild or in captive conditions. Considering that visual evaluation is not fully precise for identifying hybrids, DNA testing of each individual was used to confirm that sampled individuals were all pure *P. reticulatum*.

De novo assembly of the transcriptome generated 312,766 contigs. Three of these contings, 9069, 3798, and 1692 bp in length, were observed to be highly homologous to *P. corruscans* mtDNA sequences (NC_026846.1). Further assembly of these fragments with merged cDNA sequences produced a single 16,281 bp sequence. Because sequencing was performed using cDNAs, the complete D-loop region could not be recovered. Gap filling with Sanger sequencing of PCR products amplified across the region revealed that only 295 bp of the *P. reticulatum* mtDNA had not been sequenced and assembled. The complete *P. reticulatum* mtDNA sequence was observed to have 16,576 bp in length, representing the first report of the complete mitogenome for the species (Genbank acc. # KU291530).

Sequencing of pooled samples allowed for the identification of mitogenome polymorphisms. A total of 77 SNPs with MAF>0.09 and RD>522, composed of 71 transitions (92.2%) and 6 transversions (7.8%), were observed (Table 1). Eight SNPs were observed in the 12 and 16 S ribosomal genes, two were observed in tRNA genes and the remaining SNPs were observed in protein coding genes, while 30 SNPs were observed in non-coding regions. A total of 13% (n = 10) of SNPs observed in coding regions result in amino acid changes (Table S1, Supplementary material). MAF and RD cutoff values were used to eliminate SNPs resulting from sequencing artifacts, observed especially at 3´ ends of reads, from the final annotated sequence (data not shown).

Complete mitochondrial genomes have been typically obtained through sequencing of PCR-amplified overlapping fragments (e.g., He et al. 2011; Hrbek and Farias 2008; Mu et al. 2014; Wang et al. 2014). The rapid progress observed in different NGS technologies has greatly increased the number of complete mtDNA sequences made publicly available for different fish species, including tropical siluriformes, such as the trans-andean shovelnose catfish (*Sorubim cuspicaudus*, Restrepo-Escobar et al. 2014) and *Pseudoplatystoma magdaleniatum* (Rangel-Medrano et al. 2015). However, only two previous reports of studies using cDNA sequences to generate complete mitogenome *de novo* assemblies were observed (Heidtmann 2014; Mu et al. 2015). The analyzed NGS sequences were originally generated for characterizing the *P. reticulatum* transcriptome (Villela et al. 2017b), and although the applied strategy to assemble a complete mitogenome for the species required supplemental sequencing to unveil the untrasncribed portion of the mtDNA control region, the final result was highly satisfactory when compared to other published complete mitogenome sequences.

mtDNA sequence composition and annotation

Manual and automated annotation of the complete *P. reticulatum* mtDNA obtained with DOGMA (Wyman et al. 2004), MITOS (Bernt et al. 2013) and the MitoAnnotator database (Iwasaki et al. 2013) are detailed in Table S2 (Supplementary material). Comparisons of the obtained results revealed differences in rRNA gene sizes of up to 23 bp (e.g. *12* and *16 S*) and of up to 99 bp in protein-coding genes sizes. Consequently, the initial and/or final positions of these genes also differed depending on the applied annotation method (Table S2, Supplementary material).

There were no differences in start/end positions for tRNA genes annotated manually (with tRNAScan) or automatically with MITOS, DOGMA and MitoAnnotator (Table S2, Supplementary material). The $tRNA^{Ser(GCT)}$ gene was not identified in the manual annotation, and $tRNA^{Phe(GAA)}$ and $tRNA^{Met(TGG)}$ genes were only identified with DOGMA. The $tRNA^{Ser(TGA)}$ and $tRNA^{Lys(TTT)}$ genes were mislabeled as $tRNA^{Phe(AAA)}$ and $tRNA^{Asp(AGG)}$ with DOGMA, respectively, but were correctly identified by other methods. Start and end positions for protein-coding genes were equally identified with MitoAnnotator and DOGMA, while observed differences resulted from the omission of Stop codon bases from the final annotation provided by MitoAnnotator. Considering all obtained results, the annotation information obtained with MitoAnnotator was chosen for submission to Genkank (Table 2; Fig. 1). This annotation was compared to publicly available mtDNA annotations from four different

Table 1 Substitution types of 77 SNPs observed in the *P. reticulatum* mitogenome

	Transition		Transversion			
	G↔A	C↔T	A↔T	C↔G	A↔C	T↔G
N (%)	37 (48.1%)	34 (44.2%)	2 (2.6%)	1 (1.3%)	2 (2.6%)	1 (1.3%)
Total (%)	71 (92.2%)		6 (7.8%)			

Table 2 Complete automated annotation of the *P. reticulatum* mitochondrial genome (16,576 bp) generated with MitoAnnotator (Iwasaki et al. 2013)

Gene	Strand[a]	Position (bp)		Length (bp)[c,d]	Intergenic nucleotides[e]	Codon		Anti-codon
		Initial[b]	Final[b]			Start	Stop[f]	
tRNA^{Phe}	H	1	69	69	0			GAA
rRNA *12 S*	H	70	1030 (5)	961(5)	0			
tRNA^{Val}	H	1031	1102	72	0			TAC
rRNA *16 S*	H	1103	2782	1680	0			
tRNA$^{Leu\ (TAA)}$	H	2783	2857	75	1			TAA
NAD1	H	2859	3830	972	1	ATG	TAA	
tRNA^{Ile}	H	3832	3903	72	−1			GAT
tRNA^{Gln}	L	3903	3973	71	−1			TTG
tRNA^{Met}	H	3973	4042	70	0			CAT
NAD2	H	4043	5087 (96)	1045 (948)	0	ATG	T	
tRNA^{Trp}	H	5088	5158	71	18			TCA
tRNA^{Ala}	L	5177	5245	69	1			TGC
tRNA^{Asn}	L	5247	5319	73	0			GTT
O^{L}	–	5320	5348	29	0			
tRNA^{Cys}	L	5349	5415	67	1			GCA
tRNA^{Tyr}	L	5417	5486	70	1			GTA
COX1	H	5488	7038	1551	0	GTG	TAA	
tRNA$^{Ser\ (TGA)}$	L	7039	7109	71	4			TGA
tRNA^{Asp}	H	7114	7186	73	14			GTC
COX2	H	7201	7891	691	0	ATG	T	
tRNA^{Lys}	H	7892	7965	74	1			TTT
ATP8	H	7967	8134	168	−10	ATG	TAA	
ATP6	H	8125	8807	683	0	ATG	TA	
COX3	H	8808	9591	784	0	ATG	T	
tRNA^{Gly}	H	9592	9664	73	0			TCC
NAD3	H	9665 (−51)	10,013	349 (51)	0	ATG	T	
tRNA^{Arg}	H	10,014	10,084	71	0			TCG
NAD4L	H	10,085	10,381 (18)	297 (18)	−7	ATG	TAA	
NAD4	H	10,375	11,752	1378	0	ATG	T	
tRNA^{His}	H	11,753	11,822	70	0			GTG
tRNA^{Ser}	H	11,823	11,889	67	5			GCT
tRNA^{Leu}	H	11,895	11,967	73	0			TAG
NAD5	H	11,968 (−54)	13,794 (42)	1827 (96)	−4	ATG	TAA	
NAD6	L	13,791 (−3)	14,306 (−3)	516	0	ATG	TAA	
tRNA^{Glu}	L	14,307	14,375	69	1			TTC
CYTB	H	14,377	15,514	1138	0	ATG	T	
tRNA^{Thr}	H	15,515	15,586	72	−2			TGT
tRNA^{Pro}	L	15,585	15,654	70	0			TGG
D-*Loop*	-	15,655	16,576	922				

[a]Strand: H = heavy (+); L = light (−)

[b]Numbers in parentheses indicate differences in initial or final positions relative to the positions found by manual annotation

[c]Size of protein-coding genes including stop codons

[d]Numbers in parentheses indicate differences in gene size relative to the sizes obtained by manual annotation

[e]Positive numbers indicate nucleotides separating two adjacent genes. Negative numbers indicate nucleotide overlap

[f]T and TA indicate an incomplete stop codon

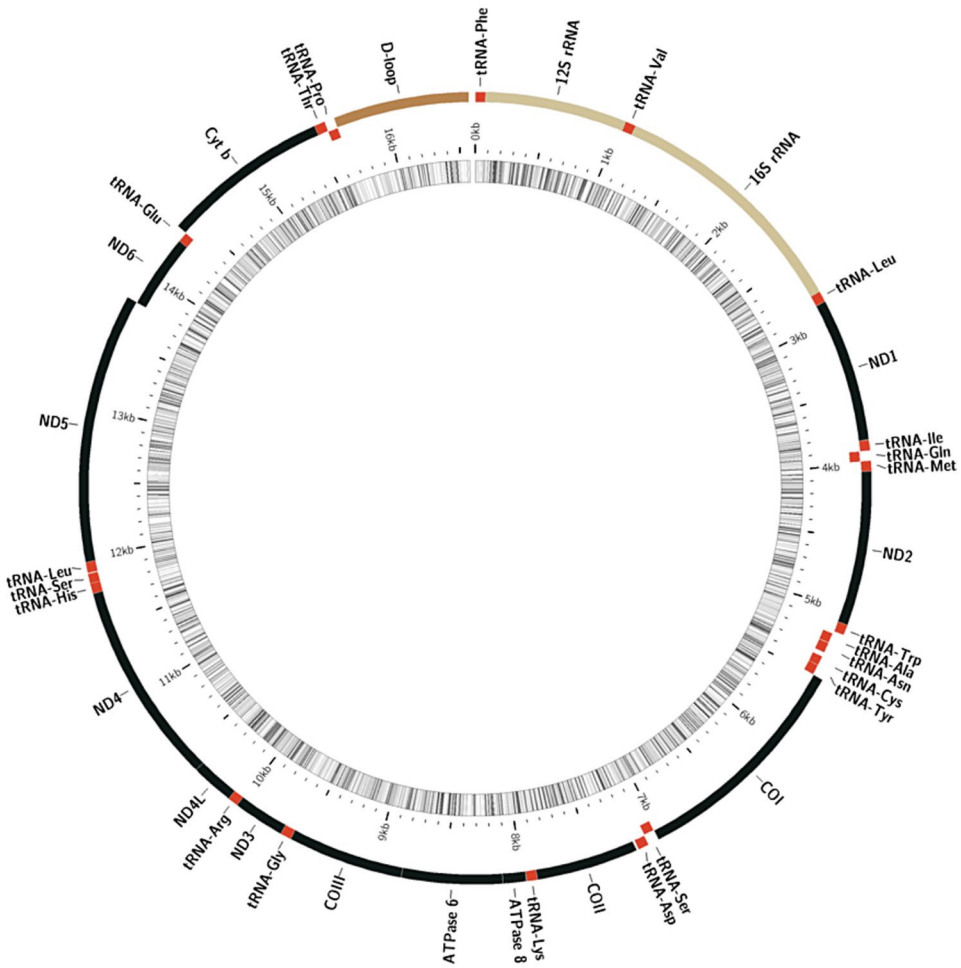

Fig. 1 *P. reticulatum* mtDNA map generated with MitoAnnotator. Protein-coding genes, tRNAs, rRNAs, and non-coding regions are shown in different colors. Genes located within the *outer circle* are coded on the H-strand whereas the remaining genes are coded on the L-strand

Pimelodidae species. Results can be observed in Table S3 (Supplementary material).

Mitochondrial genome composition and organization

The *P. reticulatum* mitogenome observed structure is highly similar to other vertebrate mitogenomes. All 37 mitochondrial structural genes (13 respiratory chain complex protein-coding genes, 2 rRNAs, and 22 tRNAs) in addition to the control region were observed in the assembled mtDNA sequence (Table 2; Fig. 1). The same Heavy (H) and Light (L) strand coding pattern previously reported for most vertebrates was also observed in the *P. reticulatum* mitogenome. Only NADH dehydrogenase subunit 6 (*NAD6*) and the 8 tRNA genes ($tRNA^{Gln(TTG)}$, $tRNA^{Ala(TGC)}$, $tRNA^{Asn(GTT)}$, $tRNA^{Cys(GCA)}$, $tRNA^{Tyr(GTA)}$, $tRNA^{Ser(TGA)}$, $tRNA^{Glu(TTC)}$, and $tRNA^{Pro(TGG)}$) were observed to be coded on the L-strand, as the remaining genes are coded on the H-strand.

An atypical start codon (GTG) was only observed on *COX1* while incomplete stop codons were observed in several protein-coding genes. The presence of incomplete stop codons is a common feature of mitochondrial genes of several vertebrates, including numerous species of fish (Kartavtsev et al. 2007; Alam et al. 2014; Jiang et al. 2014; Mu et al. 2015). Seven genes were observed to have incomplete stop codons: *NAD2, COX2, COX3, NAD3, NAD4* and *CYTB* presented T as stop codon and *ATP6* presented TA. All remaining genes were found to have the same (TAA) stop codon, which is generated via post-transcriptional polyadenylation of the mRNA 3′ end (Ojala et al. 1981).

Intergenic overlaps of protein-coding regions are common within vertebrate mitogenomes and have been reported for several fish species (Broughton et al. 2001; Kim and Lee 2004; Guo et al. 2008; Hrbek and Farias 2008; Prosdocimi et al. 2011; Fischer et al. 2013; Jiang et al. 2014; Mu et al. 2015). Overlap lengths between *ATP8* and *ATP6* (10 nucleotides), *NAD4L* and *NAD4* (seven nucleotides), and *NAD5* and *NAD6* (four nucleotides) observed in *P. reticulatum* were within expected sizes. Intergenic spacer regions were found to be 48 bp in 11 regions (Table 2).

Nucleotide composition of the complete *P. reticulatum* mitochondrial genome (Table 3) was as follows: A: 31.45%; C: 28.47%; T: 24.63%; and G: 15.44%. GC content (43.92%) was lower than the AT content (56.08%), confirming the existing bias against G in the nucleotide

Table 3 Base composition of *P. reticulatum* mitogenome protein-coding genes, rRNAs, tRNAs, and the control region

	Length (bp)	% Nucleotide composition				% GC Content	% AT Content	AT Skew	GC Skew
		T (U)	C	A	G				
Complete mitogenome	16,576	24.63	28.47	31.45	15.44	43.92	56.08	0.12	−0.30
Protein-coding genes*	11,399	25.36 (40.70)	30.12 (9.30)	30.26 (13.76)	14.26 (36.24)	44.38 (45.54)	55.62 (54.46)	−0.36 (0.59)	0.09 (−0.49)
1st codon position	3804	20.59 (33.14)	27.69 (8.72)	26.97 (12.79)	24.75 (45.35)	52.44 (54.07)	47.56 (45.93)	−0.06 (0.68)	0.13 (−0.44)
2nd codon position	3798	40.52 (45.93)	27.44 (16.28)	19.15 (12.21)	12.88 (25.58)	40.32 (41.86)	59.68 (58.14)	−0.36 (0.22)	−0.36 (−0.58)
3rd codon position	3797	14.96 (43.02)	35.24 (2.91)	44.68 (16.28)	5.12 (37.79)	40.36 (40.70)	59.64 (59.30)	−0.75 (0.86)	0.50 (−0.45)
tRNAs	1562	27.21	21.45	28.94	22.41	43.85	56.15	0.02 (0.02)	0.03 (0.03)
rRNAs	2641	20.94	25.03	19.99	20.94	45.02	54.98	−0.09 (−0.09)	−0.02 (−0.02)
Control Region	922	33.51	21.69	31.45	13.23	34.92	64.97	−0.24 (−0.24)	−0.03 (−0.03)

*Based on the 12 protein-coding genes located on the H-strand. Base composition for *NAD6* (L-strand) is shown in parentheses

composition of most vertebrate mitogenomes (Broughton et al. 2001; Guo et al. 2008; Fischer et al. 2013). Heavy (H) and Light (L) strand compositional asymmetries have been reported for vertebrate mitochondrial DNA, where the major coding strand (H) has been observed to be relatively rich in A and C in comparison to the L strand (Perna and Kocher 1995; Min and Hickey 2007). Similar trends of strand compositional bias were observed in *P. reticulatum* mtDNA (Table 3), where the AT and GC skews (Perna and Kocher 1995) were 0.12 and −0.30; representing an excess of A over T and a shortage of G over C in the H strand, respectively.

The observed nucleotide composition in *P. reticulatum* mitochondrial protein-coding genes (Table 3) was A: 30.26%, C: 30.12%, T: 25.36%, G: 14.26%. AT content was greater than GC content. Average AT and GC skew values of the 12 protein-coding genes were 0.09 and −0.36, while for *NAD6* (L-encoded) values were −0.49 and 0.59, respectively. Protein-coding genes also showed overall anti-G bias (14.26% of G). In the third codon position, the anti-G bias was larger (5.12% of sites were G), which has been similarly reported in other vertebrate species (Clayton 1991; Meyer 1993; Fonseca et al. 2008). Anti-G biases in third codon positions of mitochondrial protein-coding genes have been reported for *Danio rerio* (Broughton et al. 2001), where 7% of sites were G; *Arapaima gigas* (Hrbek and Farias 2008), with anti-G bias of 3.8% G in the third position of the 12 heavy-strand encoded genes; *Macropodus opercularis* (Mu et al. 2015), which presented 14.8% G in the third codon position; and *Oreochromis niloticus* and *O. aureus* (He et al. 2011).

Min and Hickey (2007) showed the impact nucleotide skews have on amino acid composition of mitochondrial proteins. Mitogenomes with negative GC and positive AT skews encode proteins highly enriched in amino acids encoded by CA-rich codons. The most frequently used amino acids observed in *P. reticulatum* mitochondrial genome are leucine, threonine, proline and isoleucine (Table 4; Fig. 2). Threonine and proline are amino acids encoded by CA-rich codons and account for 16.57% of encoded amino acids.

Preferred codons for each amino acid in protein-coding gene sequences were identified with the highest estimated RSCU values and were matched to all 22 identified tRNAs (Table 4), with the exception of methionine, alanine and threonine. The CUA codon (leucine, RSCU = 2.39) was the most frequently used, followed by AAC (lysine, RSCU = 1.42), AAU (asparagine, RSCU = 1.17) and AUC (isoleucine, RSCU = 1.01). GCG codon (alanine, RSCU = 0.16) was the least frequently used (Table 4).

Considering only amino acids with fourfold degenerate third positions (valine, serine, proline, threonine, alanine, arginine and glycine), codons complementary to the tRNAs ending in A were the most frequently observed in the *P. reticulatum* mitogenome. Among the twofold amino acids, C was the most frequently used nucleotide in the third codon position (Table 4). These results were consistent with the anti-G bias identified in the *P. reticulatum* mitogenome, since the G nucleotide was the least frequent in the

Table 4 Pattern of codon usage in the *P. reticulatum* mitogenome

Amino acid	Codon	Number[a]	Frequency (%)	RSCU[b]	Amino acid	Codon	Number[a]	Frequency (%)	RSCU
Phenylalanine (Phe/F)	UUU	81	1.47	0.76	Tyrosine (Tyr/Y)	UAU	102	1.85	0.94
	UUC*	132	2.39	1.24		UAC*	114	2.06	1.06
Leucine (Leu/L)	UUA*	113	2.05	1	Stop	UAA	86	1.56	1.54
	UUG	39	0.71	0.35		UAG	52	0.94	0.93
	CUU	107	1.94	0.95	Histidine (His/H)	CAU	67	1.21	0.64
	CUC	93	1.68	0.83		CAC*	143	2.59	1.36
	CUA*	269	4.87	2.39	Glutamine (Gln/Q)	CAA*	131	2.37	1.43
	CUG	54	0.98	0.48		CAG	52	0.94	0.57
Isoleucine (Ile/I)	AUU	156	2.82	0.99	Asparagine (Asn/N)	AAU	113	2.05	0.83
	AUC*	158	2.86	1.01		AAC*	160	2.90	1.17
Methionine (Met/M)	AUA	141	2.55	1.39	Lysine (Lys/K)	AAA*	179	3.24	1.42
	AUG*	62	1.12	0.61		AAG	74	1.34	0.58
Valine (Val/V)	GUU	46	0.83	0.83	Aspartic Acid (Asp/D)	GAU	44	0.80	0.68
	GUC	45	0.81	0.81		GAC*	86	1.56	1.32
	GUA*	99	1.79	1.78	Glutamic Acid (Glu/E)	GAA*	94	1.70	1.31
	GUG	32	0.58	0.58		GAG	50	0.91	0.69
Serine (Ser/S)	UCU	67	1.21	0.9	Cysteine (Cys/C)	UGU	34	0.62	0.89
	UCC	96	1.74	1.29		UGC*	42	0.76	1.11
	UCA*	106	1.92	1.42	Tryptophan (Trp/W)	UGA*	96	1.74	1.44
	UCG	21	0.38	0.28		UGG	37	0.67	0.56
Proline (Pro/P)	CCU	122	2.21	1.11	Arginine (Arg/R)	CGU	23	0.42	0.66
	CCC	141	2.55	1.28		CGC	41	0.74	1.18
	CCA*	142	2.57	1.29		CGA*	60	1.09	1.73
	CCG	34	0.62	0.31		CGG	15	0.27	0.43
Threonine (Thr/T)	ACU	132	2.39	1.11	Serine (Ser/S)	AGU	45	0.81	0.6
	ACC	153	2.77	1.29		AGC*	113	2.05	1.51
	ACA*	150	2.72	1.26	Stop	AGA	38	0.69	0.68
	ACG	41	0.74	0.34		AGG	48	0.87	0.86
Alanine (Ala/A)	GCU	58	1.05	0.76	Glycine (Gly/G)	GGU	40	0.72	0.64
	GCC	123	2.23	1.62		GGC	76	1.38	1.22
	GCA*	111	2.01	1.46		GGA*	86	1.56	1.38
	GCG	12	0.22	0.16		GGG	47	0.85	0.76

RSCU relative synonymous codon usage

*Codons that are complementary to the tRNA genes

[a]Number of codons

third codon position. This pattern was similarly reported for *Danio rerio* (Broughton et al. 2001) and is observed in most vertebrates.

Codon usage bias (CUB) has been widely observed in different organisms (Ikemura 1985; Bulmer 1991; Hershberg and Petrov 2008) and is stronger in highly expressed genes used for production of large amounts of protein (Bulmer 1991; Hershberg and Petrov 2008; Zhang et al. 2012; Ma et al. 2014). Codon composition has been reported to have an important role in the regulation of gene expression (Paul et al. 2014), as genes that use codons recognized by more abundant tRNAs may be translated more efficiently (Bulmer 1991; Zhang et al. 2012; Ma et al. 2014). Therefore, accurate quantification of CUBs in protein-coding genes may be important for unrevealing mechanisms involved in gene evolution, function and expression regulation. Several methods have been proposed for estimating CUB. Codon deviation coefficient (CDC) estimates account for both GC and purine (AG) contents, addressing heterogeneous background nucleotide composition (BNC) in all three codon positions, and is considered to be a better estimate to support hypothesis of evolutionary pressures on gene function, independently of underlying reference data sets (Zhang et al. 2012; Paul et al. 2014). Obtained CDC

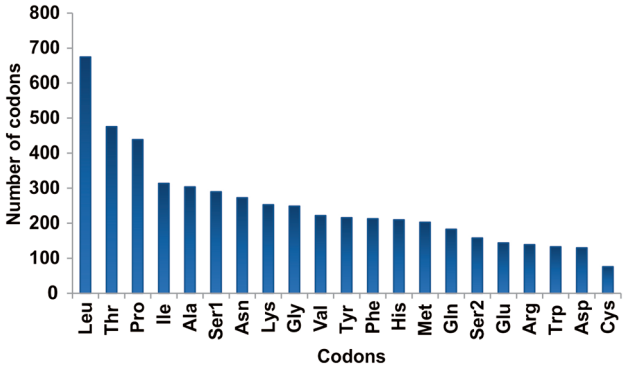

Fig. 2 *P. reticulatum* predicted mitochondrial amino acid usage pattern

estimates ranged from 0.18 to 0.37. NAD4L was the gene with the highest significant CDC value (0.37, $P=0.00$), which supports its function as a highly expressed respiratory chain Complex I gene (Ma et al. 2014). Conversely, COX2, which is part of the cytochrome c oxidase complex also with a key role in the mitochondrial function, was the gene with the lowest significant CDC value (0.18, $P=0.00$). In addition, estimated CUB for ATP8 (0.31), the shortest gene analyzed (153 pb), was not significant, which is consistent with simulations reported by Zhang et al. (2012), showing that short sequences tend to generate unreliable CUB estimates (Table 5).

Table 5 Estimates of codon usage bias (CUB) for 13 protein-coding genes of *P. reticulatum* mitogenome

ID	Length	CDC	P value	GC	AG	GC_1	AG_1	GC_2	AG_2	GC_3	AG_3
NAD1	945	0.27	0.00	0.45	0.44	0.55	0.52	0.40	0.27	0.40	0.53
NAD2	1011	0.34	0.00	0.46	0.45	0.55	0.55	0.45	0.25	0.39	0.54
COX1	1500	0.20	0.00	0.45	0.45	0.55	0.57	0.39	0.31	0.41	0.48
COX2	675	0.18	0.00	0.43	0.47	0.59	0.56	0.32	0.37	0.38	0.48
ATP8	153	0.31	0.40	0.41	0.46	0.43	0.59	0.35	0.31	0.43	0.47
ATP6	669	0.26	0.00	0.43	0.43	0.56	0.53	0.38	0.25	0.36	0.52
COX3	750	0.26	0.00	0.47	0.43	0.58	0.50	0.40	0.35	0.42	0.46
NAD3	336	0.28	0.00	0.45	0.39	0.58	0.45	0.35	0.27	0.43	0.46
NAD4L	291	0.37	0.00	0.52	0.41	0.61	0.46	0.48	0.26	0.47	0.49
NAD4	1338	0.21	0.00	0.43	0.45	0.51	0.52	0.41	0.30	0.38	0.53
NAD5	1788	0.26	0.00	0.43	0.44	0.49	0.57	0.38	0.31	0.43	0.45
NAD6	507	0.36	0.00	0.46	0.50	0.55	0.59	0.41	0.37	0.41	0.53
CYTB	1101	0.31	0.00	0.46	0.42	0.55	0.50	0.35	0.31	0.47	0.44

CDC codon deviation coefficients

Fig. 3 Variable regions and conserved blocks in the non-conding region (D-Loop) of *P. reticulatum* mitochondrial genome

Table 6 Similarity of *P. reticulatum* control region regulatory elements with the others species

Regulatory elements	*Tropheus moorii* (NC_018815.1)	*Petrochromis trewavasae* (NC_018814.1)	*Macropodus opercularis* (KM588227)	*Oreochromis niloticus* (GU370126)	*O. aureus* (GU370125)	*Danio rerio* (KM244705.1)
TAS	0.57	0.52	0.67	0.63	0.54	0.60
CSB-F	0.72	0.72	–	0.72	0.72	–
CSB-E	0.71	0.71	0.75	0.71	0.71	–
CSB-1	0.71	0.44	0.55	0.68	0.68	0.60
CSB-2	0.88	0.53	0.94	0.88	0.88	0.88
CSB-3	0.59	0.70	0.56	0.59	0.59	1.00

Table 7 Size and structure of the complete mitogenomes of *P. reticulatum* and four previously published catfish species from family Pimelodidae

Species	NCBI GenBank ID	Author	Length (bp)	Sequence identity (%)	Protein-coding genes* (bp)	rRNA genes (bp)	tRNA genes (bp)	D-loop (bp)
P. reticulatum		This study	16,576	–	11,399	2641	1562	922
P. corruscans	NC_026846.1	Perini et al. (unpublished data)	16,123	95	11,400	2193	1561	912
P. magdaleniatum	NC_026526.1	Rangel-Medrano et al. (2015)	16,568	93	11,401	2647	1559	912
Pimelodus pictus	NC_015797.1	Nakatani et al. (2011)	16,575	87	11,402	2588	1614	923
Sorubim cuspicaudus	NC_026211.1	Restrepo-Escobar et al. (2014)	16,544	88	11,403	2630	1558	909

*Stop codon included in the sequences

rRNA and tRNA genes

All *P. reticulatum* mitochondrial tRNA genes were observed to have anticodons matching the vertebrate mitochondrial genetic code. The two possible codons for serine (TGA and GCT) and leucine (TAA and TAG) were observed, following the pattern reported for most vertebrates. Sequence overlaps were observed for $tRNA^{Ile}$ and $tRNA^{Gln}$ (1 bp), $tRNA^{Gln}$ and $tRNA^{Met}$ (1 bp), and $tRNA^{Thr}$ and $tRNA^{Pro}$ (2 bp). With the exception of $tRNA^{Ser(GCT)}$, all of the tRNA genes were predicted to fold into classic cloverleaf secondary structures (Table S4, Supplementary material). Amino acid arm and anticodon loop are commonly 7 bp in length, while TΨC and anticodon arms are 5 bp in length and the DHU arm is 4 bp in length (Broughton et al. 2001; Kim and Lee 2004; Hrbek and Farias 2008). However, variations in the size, position, nucleotide composition, and several non-complementary pairings were identified in most of predicted *P. reticulatum* tRNA secondary structures (Table S4, Supplementary material).

Non-coding sequences

Two major non-coding regions were observed. A short region with a length of 29 bp (5′-CTT TCC CCG CCG CCT TAA AAA GGC GGG GA-3′), which corresponds to the L-strand replication origin (O_L), was observed between $tRNA^{Asn}$ and $tRNA^{Cys}$. In addition, one large non-coding region (922 bp), corresponding to the control region (D-Loop) was observed between $tRNA^{Pro}$ and $tRNA^{Phe}$. Base composition analysis revealed the control region is richer in AT (64.97%, Table 3) than the remaining of the mitogenome (56.08%), as previously reported (Mu et al. 2014, 2015).

Important regulatory elements for the replication and expression of the mitogenome were found in three domains of the D-Loop (Fig. 3), showing high identity in most sequence comparisons with other species (Table 6). The conserved GGGGG-box, which functions as the origin of the H- strand replication (O_H), was observed in the central domain of the D-Loop region. The highly conserved central domains (CSB-E and

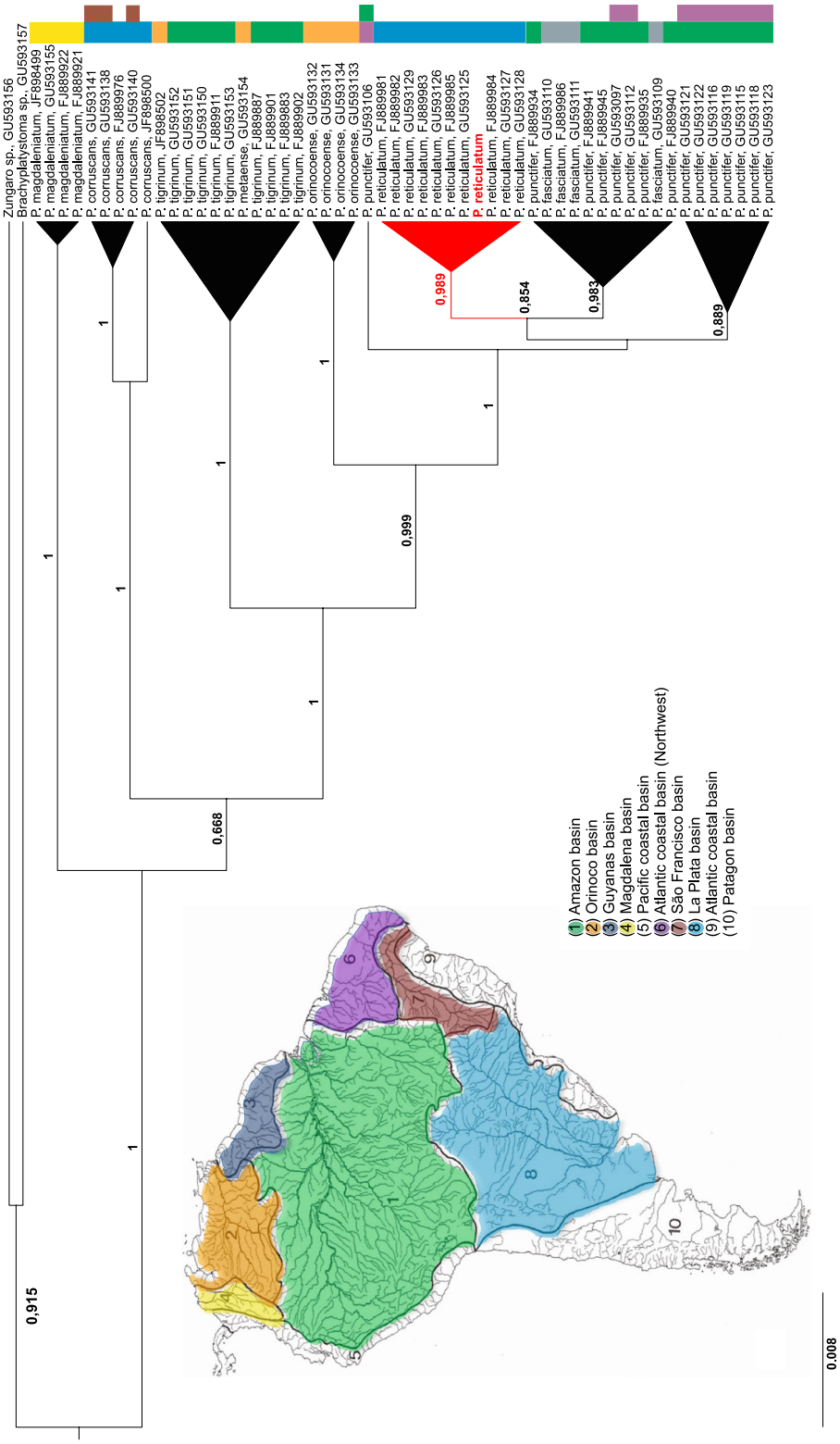

Fig. 4 Phylogenetic relationship of *Pseudoplatystoma* based on bayesian analysis of a fragment from Cytochrome b (CYTB) using a Yule model in BEAST. Branch length fitted a strict molecular clock. *P reticulatum* highlighted in *red* indicates the generated sequence. *Zungaro* and *Brachyplatystoma* sequences were used as outgroup. *Numbers* above the branches indicate posterior probabilities. *Colored bars* in the genealogy represent South American river basins marked on the map

CSB-F) were observed to be flanked by Termination Associated Sequences (TAS) and the conserved sequence blocks (CSB-1, CSB-2 and CSB-3), at the 5′ and 3′ ends, respectively. TAS has been reported to be associated with termination of newly synthesized H-strands during replication (Sbisà et al. 1997; Broughton et al. 2001). CSB-1, CSB-2 and CSB-3 contain the origin of the H-strand replication (O_H) and two promoters for transcription of

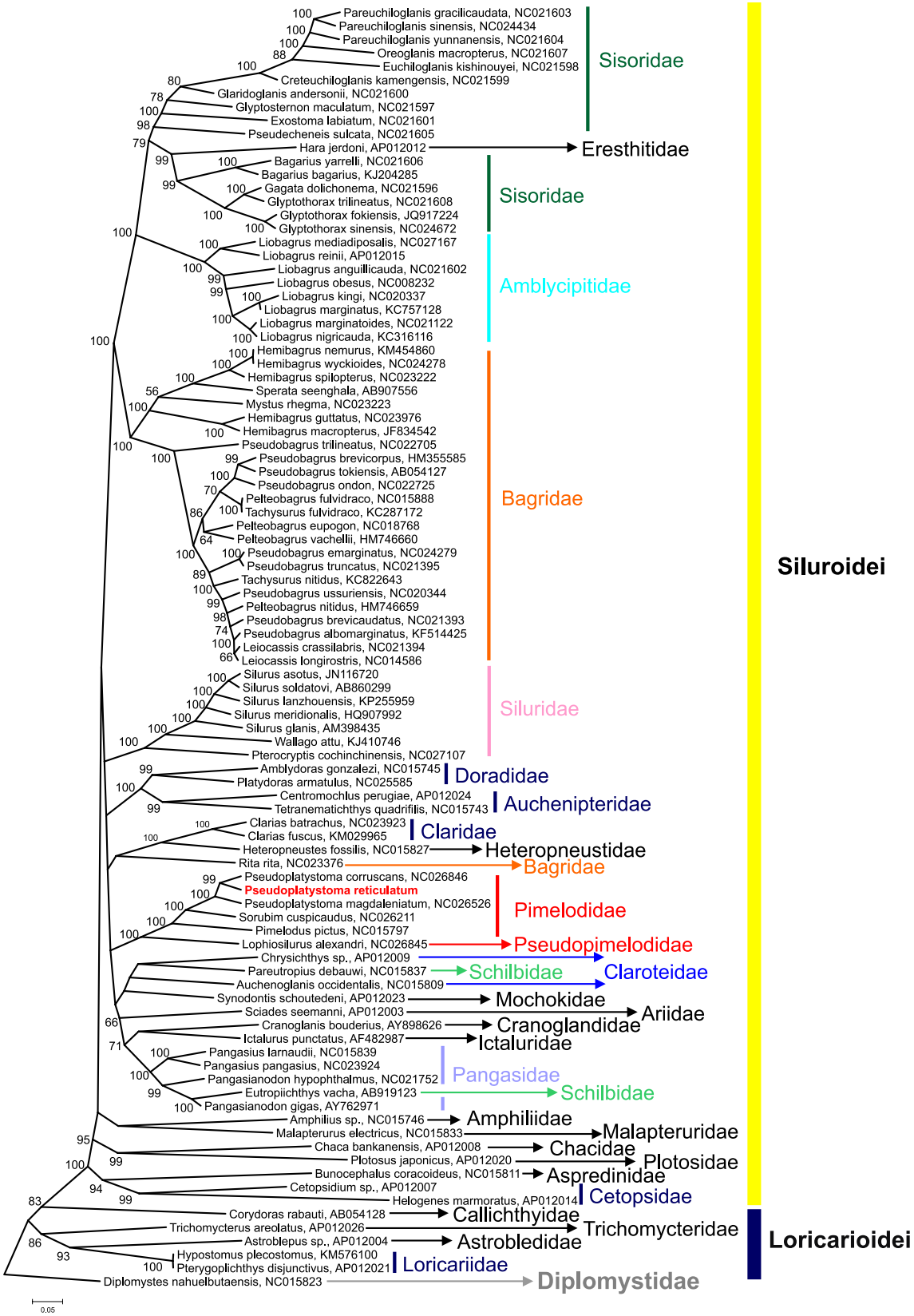

Fig. 5 Phylogeny of Siluriformes families based on the *P. reticulatum* obtained sequence (in *red*) and 94 publicly available full mitogenome sequences. The tree was obtained by Maximum likelihood analysis and branch support was estimated after 10,000 bootstrap runs (values above 50% are shown). Main lineages from Diplomystidae, Loricarioidei and Siluroidei are highlighted

the Light (L-) and H-strands (LSP and HSP, respectively) (Walberg and Clayton 1981; Clayton 1991; Saccone et al. 1991; Taberlet 1996; Jemt et al. 2015).

Sequence comparisons with mitogenomes from Pimelodidae Catfish species

The structure, content, and order of the mitochondrial structural genes were observed to be similar in all five mitogenomes analyzed (Table S3, Supplementary material). Overall sequence identities between *P. reticulatum* and *P. corruscans, P. magdaleniatum, Sorubim cuspicaudus* and *Pimelodus pictus* were 95, 93, 88 and 87%, respectively (Table 7). Sequence length varied from 16,544 bp to 16,576 bp, with the exception of *P. corruscans* (NC_026846.1, 16,123 bp), which has a smaller *12 S* rRNA (515 bp), when compared to the average of other species (959 bp). An alternative *P. corruscans 12 S* published sequence (Lundberg et al. 2011) is 944 bp in length, indicating therefore that the complete *P.corruscans* mitogenome sequence (NC_026846.1) is likely to be misassembled.

Analysis of a cytochrome b fragment (CYTB) derived from the full *P. reticulatum* mitochondrial consensus sequence, along with previously published CYTB sequences from eight different *Pseudoplatystoma* species, placed the samples within the generated *P. reticulatum* group (Fig. 4), confirming that all sampled specimens were from this species. *P. reticulatum* is part of the *P. fasciatum* species complex, which has been recently separated into eight species based on their morphological characteristics and geographic distribution (Buitrago-Suárez and Burr 2007). Available molecular data based on nuclear e partial mtDNA sequences (Torrico et al. 2009; Carvalho-Costa et al. 2011) do not fully corroborate the proposed classification, but are not enough to fully support alternative phylogenetic proposals for the *Pseudoplatystoma* clade. Analysis of the complete mitochondrial genome of all species in the clade will help stablish the most solid phylogenetic proposal for the group.

Phylogenetic analysis of siluriformes with complete mitogenome sequences

A total of 14,652 bp of the complete *P. reticulatum* mitogenome sequence, analyzed along with 94 publicly available Siluriformes complete mitogenomes from 29 families, revealed 8,573 variable sites (1,288 singletons and 7,285 parsimony informative). Thirteen of all analyzed families had two or more represented species, and 10 of these were placed in monophyletic groups (Fig. 5). The Bagridae group was observed to be monophyletic with the exception of *Rita rita*, which was closer to Claridae and Heteropneustidae, with low statistical support. Schilbidae appears to be monophyletic only when species of African (e.g. Claroteidae, without statistical support) and Asian (Panagasidae, with high statistical support) origin were considered. In the Claroteidae family, a single species of each of the two main subfamilies (Claroteinae and Auchenoglanidinae) was sampled and a high similarity with the African group Schilbidae was observed.

When the Diplomytidae family was considered as the basal group, the expected division between the two major groups Loricarioidei and Siluroidei was observed, confirming previous phylogeny proposals based on sequence analysis from nuclear gene fragments (Sullivan et al. 2006) and morphological data (Diogo 2007). Within Loricarioidei, Callichthydae was the most differentiated group rather than Trichomycteridae, based on only one species of each family. The group termed Big Asia, containing Bagridae (except *Rita rita*), Sisoridae, Eresthidae and Ampblyciptidae (Sullivan et al. 2006), was observed to have the largest statistical support in the phylogenetic tree among the main groups, which can be better supported by the larger number of representative species analyzed within each family.

Close analysis of the groups from South America showed that families contained in Pimeloidoidea (Pimelodidae and Pseudopimelodidae) were highly supported for their respective monophylies, while genus *Pseudoplatystoma* was also monophyletic within Pimelodidae. Therefore, most of the observed results are in agreement with other studies (Sullivan et al. 2006; Jondeung et al. 2007) and reinforce the importance of periodical revisions of molecular phylogenies as additional data are generated and made publicly available. Further analysis considering multiple molecular clocks (Drummond and Suchard 2010) and especially including full mtDNA sequences from additional *Pseudoplatystoma* sp. will contribute to refine the proposed Siluriformes phylogeny.

Conclusions

The first report of the complete *Pseudoplatystoma reticulatum* mitochondrial genome sequence revealed general gene organization, structure, content, and order similar to most vertebrates and other published Pimelodidae mitogenomes. Specific sequence and content differences were observed and may have functional attributes, for further investigation. Additional phylogenetic analysis considering multiple molecular clocks should be considered for improving the current phylogenetic proposal for the *Pseudoplatystoma* clade.

Acknowledgements The authors are grateful for the financial support from the Brazilian National Council for Scientific and

Technological Development (CNPQ), Grant No. 479665/2011-7, and from The Brazilian Agricultural Research Corporation (EMBRAPA), Grant No. 01.11.07.002.00.00. J.M.P. received a graduate scholarship from FAPESP (Grant No. 2012/03553-8). We would like to thank EMBRAPA Multiuser Bioinformatics Lab (Laboratório Multiusuário de Bioinformática da Embrapa) for providing additional computational infrastructure; Daniel Chaves Webber for elaborating the map of cachara collection points (Fig. S1, Supplementary material); and Jefferson Cristiano Christofoletti for formatting the images in Figs. 4 and 5.

Conflict of interest The authors declare that they have no conflict of interest.

Ethical approval All procedures performed in studies involving animals were in accordance with the ethical standards of the institution or practice at which the studies were conducted.

References

Alam MT, Petit RA III, Read TD, Dove ADM (2014) The complete mitochondrial genome sequence of the world's largest fish, the whale shark (*Rhincodon typus*), and its comparison with those of related shark species. Gene 539:44–49

Avise JC (1994) Animal Mitochondrial DNA. In: Avise JC (ed) Molecular markers, natural history and evolution. Chapman & Hall, New York, pp 60–68

Bernt M, Donath A, Jühling F, Externbrink F, Florentz C, Fritzsch G, Pütz J, Middendorf M, Stadler PF (2013) MITOS: Improved de novo metazoan mitochondrial genome annotation. Mol Phylogenet Evol 69:313–319

Bignotto TS, Prioli AJ, Prioli SMAP, Maniglia TC, Boni TA, Lucio LC, Gomes VN, Prioli RA, Oliveira AV, Júlio-Junior HF, Prioli LM (2009) Genetic divergence between *Pseudoplatystoma corruscans* and *Pseudoplatystoma reticulatum* (Siluriformes: Pimelodidae) in the Paraná River Basin. Braz J Biol 69:681–689

Bolger AM, Lohse M, Usadel B (2014) Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics. doi:10.1093/bioinformatics/btu170

Borba RS de, Silva EL da, Ponzetto JM, Pozzobon APB, Centofante L, Alves AL, Parise-Maltempi PP (2013a) Genetic structure of the ornamental tetra fish species*Piabucus melanostomus*Holmberg, 1891 (CHARACIDAE, IGUANODECTINAE) in the Brazilian Pantanal wetlands inferred by mitochondrial DNA sequences. Biota Neotrop 13:42–46

Borba RS de, Zawadzki CH, Oliveira C, Perdices A, Parise-Maltempi PP, Alves AL (2013b) Phylogeography of *Hypostomus strigaticeps* (Siluriformes: Loricariidae) inferred by mitochondrial DNA reveals its distribution in the upper Paraná River basin. Neotrop Ichthyol 11:111–116

Broughton RE, Milam JE, Roe BA (2001) The complete sequence of the zebrafish (*Danio rerio*) mitochondrial genome and evolutionary patterns in vertebrate mitochondrial DNA. Genome Res 11:1958–1967

Buitrago-Suárez UA, Burr BM (2007) Taxonomy of the catfish genus*Pseudoplatystoma* Bleeker (Siluriformes: Pimelodidae) with recognition of eight species. Zootaxa 1512:1–38

Bulmer M (1991) The selection-mutation-drift theory of synonymous codon usage. Genetics 129:897–907

Carvalho DC, Neto DAP, Brasil BSAF, Oliveira DAA (2011) DNA barcoding unveils a high rate of mislabeling in a commercial freshwater catfish from Brazil. Mitochondr DNA 22:97–105

Carvalho DC, Oliveira DAA, Beheregaray LB, Torres RA (2012) Hidden genetic diversity and distinct evolutionarily significant units in an commercially important Neotropical apex predator, the catfish *Pseudoplatystoma corruscans*. Conserv Genet 13:1671–1675

Carvalho-Costa LF, Piorski NM, Willis SC, Galetti PM Jr, Ortí G (2011) Molecular systematics of the neotropical shovelnose catfish genus *Pseudoplatystoma* Bleeker 1862 based on nuclear and mtDNA markers. Mol Phylogenet Evol 59:177–194

Clayton DA (1991) Replication and transcription of vertebrate mitochondrial DNA. Annu Rev Cell Biol 7:453–478

Darriba D, Taboada GL, Doallo R, Posada D (2012) jModelTest 2: more models, new heuristics and parallel computing. Nat Methods 9:772

Diogo R (2007) Homoplasies, consistency index and the complexity of morphological evolution: catfishes as a case study for general discussions on phylogeny and macroevolution. Int. J Morphol 25:831–837

Prado FD do, Hashimoto DT, Mendonça FF, Senhorini JA, Foresti F, Porto-Foresti F (2011) Molecular identification of hybrids between Neotropical catfish species *Pseudoplatystoma corruscans* and *Pseudoplatystoma reticulatum*. Aquac Res 42:1890–1894

do Prado FD, Hashimoto DT, Senhorini JA, Foresti F, Porto-Foresti F (2012) Detection of hybrids and genetic introgression in wild stocks of two catfish species (Siluriformes: Pimelodidae): The impact of hatcheries in Brazil. Fish Res 125–126:300–305

Drummond AJ, Suchard MA (2010) Bayesian random local clocks, or one rate to rule them all. BMC Biol 8:114

Drummond AJ, Suchard MA, Xie D, Rambaut A (2012) Bayesian phylogenetics with BEAUti and the BEAST 1.7. Mol Biol Evol 29:1969–1973

Ewing B, Green P (1998) Base-calling of automated sequencer traces using *Phred*. II. Error probabilities. Genome Res 8:186–194

Ewing B, Hillier L, Wendl MC, Green P (1998) Base-calling of automated sequencer traces using *Phred*. I. Accuracy assessment. Genome Res 8:175–185

Ferraris CJ Jr (2007) Checklist of catfishes, recent and fossil (Osteichthyes: Siluriformes), and catalogue of siluriform primary types. Zootaxa 1418:1–628

Fischer C, Koblmüller S, Gülly C, Schlötterer C, Sturmbauer C, Thallinger GG (2013) Complete mitochondrial DNA sequences of the threadfin cichlid (*Petrochromis trewavasae*) and the bluthead cichlid (*Tropheus moorii*) and patterns of mitochondrial genome evolution in cichlid fishes. Plos One. doi:10.1371/journal.pone.0067048

Fonseca MM, Posada D, Harris DJ (2008) Inverted replication of vertebrate mitochondria. Mol Biol Evol 25:805–808

Garrison E, Marth G (2012) Haplotype-based variant detection from short-read sequencing. http://arxiv.org/pdf/1207.3907v2.pdf. Accessed 16 May 2016

Guo Y, Wang Z, Liu C, Liu Y (2008) Sequencing and analysis of the complete mitochondrial DNA of Russell's snapper (*L. russellii*). Prog Nat Sci 18:1233–1238

Haas BJ, Papanicolaou A, Yassour M, Grabherr M, Blood PD, Bowden J, Couger MB, Eccles D, Li B, Lieber M, MacManes MD, Ott M, Orvis J, Pochet N, Strozzi F, Weeks N, Westerman R, William T, Dewey CN, Henschel R, LeDuc RD, Friedman N, Regev A (2013) *De novo* transcript sequence reconstruction from RNA-seq using the Trinity platform for reference generation and analysis. Nat Protoc 8:1494–1512

Hahn C, Bachmann L, Chevreux B (2013) Reconstructing mitochondrial genomes directly from genomic next-generation sequencing reads—a baiting and iterative mapping approach. Nucleic Acids Res. doi:10.1093/nar/gkt371

Hashimoto DT, Prado FD do, Senhorini JA, Foresti F, Porto-Foresti F (2013) Detection of post-F1 fish hybrids in broodstock using molecular markers: approaches for genetic management in aquaculture. Aquac Res 44:876–884

He A, Luo Y, Yang H, Liu L, Li S, Wang C (2011) Complete mitochondrial DNA sequences of the Nile tilapia (*Oreochromis niloticus*) and Blue tilapia (*Oreochromis aureus*): genome characterization and phylogeny applications. Mol Biol Rep 38:2015–2021

Heidtmann LM (2014) Caracterização do genoma mitocondrial de onça-pintada (*Panthera onca*) e elucidação da filogenia mitogenômica do gênero pantera. Dissertação, Pontifícia Universidade Católica do Rio Grande do Sul

Hershberg R, Petrov DA (2008) Selection on codon bias. Annu Rev Genet 42:287–299

Hrbek T, Farias IP (2008) The complete mitochondrial genome of the pirarucu (*Arapaima gigas*, Arapaimidae, Osteoglossiformes). Genet Mol Biol 31:293–302

Ikemura T (1985) Codon usage and tRNA content in unicellular and multicellular organisms. Mol Biol Evol 2:13–34

Iwasaki W, Fukunaga T, Isagozawa R, Yamada K, Maeda Y, Satoh TP, 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

Jemt E, Persson O, Shi Y, Mehmedovic M, Uhler JP, López MD, Freyer C, Gustafsson CM, Samuelsson T, Falkenberg M (2015) Regulation of DNA replication at the end of the mitochondrial D-loop involves the helicase TWINKLE and a conserved sequence element. Nucleic Acids Res. doi:10.1093/nar/gkv804

Jiang M, Yang C, Wen H (2014) The complete mitochondrial genome of *Aspiorhynchus laticeps*and its phylogenetic analysis. Meta. Gene 2:218–225

Jondeung A, Sangthong P, Zardoya R (2007) The complete mitochondrial DNA sequence of the Mekong giant catfish (*Pangasianodon gigas*), and the phylogenetic relationships among Siluriformes. Gene 387:49–57

Kartavtsev YP, Jung SO, Lee YM, Byeon HK, Lee JS (2007) Complete mitochondrial genome of the bullhead torrent catfish, *Liobagrus obesus* (Siluriformes, Amblycipididae): Genome description and phylogenetic considerations inferred from the Cyt b and 16 S rRNA genes. Gene 396:13–27

Kim C II, Lee JS (2004) The complete mitochondrial genome of the rockfish *Sebastes schlegeli* (Scorpaeniformes, Scorpaenidae). Mol Cells 17:322–328

Li H, Durbin R (2010) Fast and accurate long-read alignment with Burrows-Wheeler transform. Bioinformatics 26:589–595

Lowe TM, Eddy SR (1997) TRNAscan-SE: a program for improved detection of transfer RNA genes in genomic sequence. Nucleic Acids Res 25:955–964

Lundberg JG, Sullivan JP, Hardman M (2011) Phylogenetics of the South American catfish family Pimelodidae (Teleostei: Siluriformes) using nuclear and mitochondrial gene sequences. P Acad Nat Sci Phila 161:153–189

Ma L, Cui P, Zhu J, Zhang Z, Zhang Z (2014) Translational selection in human: more pronounced in housekeeping genes. Biol Direct 9:17

Magoc T, Salzberg SL (2011) FLASH: fast length adjustment of short reads to improve genome assemblies. Bioinformatics 27:2957–2963

Meyer A (1993) Evolution of mitochondrial DNA in fishes. In: Hochachka PW, Mommsen TP (eds) Biochemistry and molecular biology of fishes, Elsevier, pp 1–38

Min XJ, Hickey DA (2007) DNA asymmetric strand bias affects the amino acid composition of mitochondrial proteins. DNA Res 14:201–206

Mu X, Liu Y, Wang X, Liu C, Song H, Hu Y, Luo J (2014) Characterization of the mitochondrial genome and phylogeny of the black arowana (*Osteoglossum ferreirai*). Biologia 69:1222–1230

Mu X, Liu Y, Lai M, Song H, Wang X, Hu Y, Luo J (2015) Characterization of the *Macropodus opercularis* complete mitochondrial genome and family Channidae taxonomy using Illumina-based *de novo* transcriptome sequencing. Gene 559:189–195

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

Ojala D, Montoya J, Attardi G (1981) tRNA punctuation model of RNA processing in human mitochondria. Nature 290:470–474

Paul P, Mazumder TH, Chakraborty S (2014) *In silico* comparison of nucleotide composition and codon usage bias between the essential and non-essential genes of *Staphylococcus aureus* NCTC 8325. Int J Curr Microbiol App Sci 3:8–15

Pereira LHG, Hanner R, Foresti F, Oliveira C (2013) Can DNA barcoding accurately discriminate megadiverse Neotropical freshwater fish fauna? BMC Genet 14:20

Perna NT, Kocher TD (1995) Patterns of nucleotide composition at fourfold degenerate sites of animal mitochondrial genomes. J Mol Evol 41:353–358

Porto-Foresti F, Hashimoto DT, Alves AL, Almeida RBC, Senhorini JÁ, Bortolozzi J, Foresti F (2008) Cytogenetic markers as diagnoses in the identification of the hybrid between Piauçu (*Leporinus microcephalus*) and Piapara (*Leporinus elongatus*). Genet Mol Biol 31:195–202

Powell AFLA, Barker FK, Lanyon SM (2013) Empirical evaluation of partitioning schemes for phylogenetic analyses of mitogenomic data: An avian case study. Mol Phylogenet Evol 66:69–79

Prado FD, Pardo BG, Guerra-Varela J, Senhorini JA, Martínez P, Foresti F, Porto-Foresti F (2014) Development and characterization of 16 microsatellites for the Neotropical catfish *Pseudoplatystoma reticulatum* and cross species analysis. Conservation Genet Resour 6:679–681

Prosdocimi F, Carvalho DC de, Almeida RN de, Beheregaray LB (2011) The complete mitochondrial genome of two recently derived species of the fish genus Nannoperca (Perciformes, Percichthyidae). Mol Biol Rep. doi:10.1007/s11033-011-1034-5

Rangel-Medrano JD, Alzate JF, Márquez EJ (2015) Complete mitochondrial genome of the Neotropical catfish*Pseudoplatystoma magdaleniatum* (Siluriformes, Pimelodidae). Mitochondr DNA. doi:10.3109/19401736.2014.1003830

Restrepo-Escobar N, Alzate JF, Márquez EJ (2014) Mitochondrial genome of the Trans-Andean shovelnose catfish*Sorubim cuspicaudus*(Siluriformes, Pimelodidae). Mitochondr DNA. doi :10.3109/19401736.2014.989506

Saccone C, Pesole G, Sbisà E (1991) The main regulatory region of mammalian mitochondrial DNA: structure–function model and evolutionary pattern. J Mol Evol 33:83–91

Sbisà E, Tanzariello F, Reyes A, Pesole G, Saccone C (1997) Mammalian mitochondrial D-loop region structural analysis: identification of new conserved sequences and their functional and evolutionary implications. Gene 205:125–140

Sosa MX, Sivakumar IKA, Maragh S, Veeramachaneni V, Hariharan R, Parulekar M, Fredrikson KM, Harkins TT, Lin J, Feldman AB, Tata P, Ehret GB, Chakravarti A (2012) Next-generation sequencing of human mitochondrial reference genomes uncovers high heteroplasmy frequency. Plos Comput Biol. doi:10.1371/journal.pcbi.1002737

Sullivan JP, Lundberg JG, Hardman M (2006) A phylogenetic analysis of the major groups of catfishes (Teleostei: Siluriformes) using *rag 1* e *rag 2* nuclear gene sequences. Mol Phylogenet Evol 41:636–662

Taberlet P (1996) The use of mitochondrial DNA control region sequencing in conservation genetics. In: Smith TB, Wayne RK (eds) Molecular genetic approaches in conservation. Oxford University Press, New York, pp 125–142

Tamura K, Stecher G, Peterson D, Filipski A, Kumar S (2013) MEGA6: Molecular evolutionary genetics analysis version 6.0. Mol Biol Evol 30:2725–2729

Torrico JP, Hubert N, Desmarais E, Duponchelle F, Nuñez Rodriguez J, Montoya-Burgos J, Garcia Davila C, Carvajal-Vallejos FM, Grajales AA, Bonhomme F, Renno J-F (2009) Molecular phylogeny of the genus *Pseudoplatystoma* (Bleeker, 1862): biogeographic and evolutionary implications. Mol Phylogenet Evol 51:588–594

Villela LCV, Alves AL, Varela ES, da Silva NMA, Caetano AR (2017a) Development of a minisequencing assay to identify cachara (*Pseudoplatystoma reticulatum*) and pintado (*Pseudoplatystoma corruscans*) F1 hybrids (In preparation)

Villela LCV, Alves AL, Varela ES, Yamagishi MEB, Giachetto PF, da Silva NMA, Paiva SR, Caetano AR (2017b) Sequencing and characterization of the cachara (*Pseudoplatystoma reticulatum*) transcriptome (In preparation)

Walberg MW, Clayton DA (1981) Sequence and properties of the human KB cell and mouse L cell D-Loop regions of mitochondrial DNA. Nucleic Acids Res 9:5411–5421

Wang SY, Shi W, Miao XG, Kong XY (2014). Complete mitochondrial genome sequences of three rhombosoleid fishes and comparative analyses with other flatfishes (Pleuronectiformes). Zool Stud 53:80

Warren RL, Sutton GG, Jones SJM, Holt RA (2007) Assembling millions of short DNA sequences using SSAKE. Bioinformatics 23:500–501

Wyman SK, Jansen RK, Boore JL (2004) Automatic annotation of organellar genomes with DOGMA. Bioinformatics 20:3252–3255

Yáñez JM, Newman S, Houston RD (2015) Genomics in Aquaculture to better understand species biology and accelerate genetic progress. Front Genet. doi:10.3389/fgene.2015.00128

Yu L, Li Y-W, Ryder OA, Zhang Y-P (2007) Analysis of complete mitochondrial genome sequences increases phylogenetic resolution of bears (Ursidae), a mammalian family that experienced rapid speciation. BMC Evol Biol. doi:10.1186/1471-2148-7-198

Zhang Z, Li J, Cui P, Ding F, Li A, Townsend JP, Yu J (2012) Codon deviation coefficient: a novel measure for estimating codon usage bias and its statistical significance. BMC Bioinf 13:43