CYNTHIA APARECIDA VALIATI BARRETO

SEQUENCIAMENTO, MONTAGEM E ANOTAÇÃO DOS GENOMAS MITOCONDRIAIS DE *Astyanax giton* EIGENMANN, 1908 E *Oligosarcus argenteus* GÜNTHER, 1864

Dissertação apresentada à Universidade Federal de Viçosa, como parte das exigências do Programa de Pós-Graduação em Biologia Animal, para obtenção do título de *Magister Scientiae*.

VIÇOSA
MINAS GERAIS – BRASIL
2017

Ficha catalográfica preparada pela Biblioteca Central da Universidade Federal de Viçosa - Câmpus Viçosa

T
B273s
2017

Barreto, Cynthia Aparecida Valiati, 1994-
Sequenciamento, montagem e anotação dos genomas mitocondriais de *Astyanax giton* Eigenmann, 1908 e *Oligosarcus argenteus* Günther, 1864 / Cynthia Aparecida Valiati Barreto. – Viçosa, MG, 2017.
x, 30f. : il. (algumas color.) ; 29 cm.

Orientador: Jorge Abdala Dergam dos Santos.
Dissertação (mestrado) - Universidade Federal de Viçosa.
Inclui bibliografia.

1. Genoma mitocondrial. 2. Filogenia. 3. Characiformes. I. Universidade Federal de Viçosa. Departamento de Biologia Animal. Programa de Pós-graduação em Biologia Animal. II. Título.

CDD 22 ed. 597.48

CYNTHIA APARECIDA VALIATI BARRETO

SEQUENCIAMENTO, MONTAGEM E ANOTAÇÃO DOS GENOMAS MITOCONDRIAIS DE *Astyanax giton* EIGENMANN, 1908 E *Oligosarcus argenteus* GÜNTHER, 1864

Dissertação apresentada à Universidade Federal de Viçosa, como parte das exigências do Programa de Pós-Graduação em Biologia Animal, para obtenção do título de *Magister Scientiae*.

APROVADA: 11 de outubro de 2017

Evanguedes Kalapothakis

Pedro Marcus Pereira Vidigal
(Coorientador)

Jorge Abdala Dergam dos Santos
(Orientador)

AGRADECIMENTOS

Depois de quase sete anos, me despeço de Viçosa e da UFV. Gostaria de agradecer primeiramente a Deus, meu equilíbrio espiritual. Quero agradecer também a minha família, que é minha base, sem ela eu não seria nada. A minha mãe Angelita e meu pai Sergio, que estiveram comigo em todos os momentos, sejam eles felizes ou tristes, me apoiando, aconselhando e fazendo o possível para me darem do melhor. A meu irmão Eduardo, que quando cheguei era um menino e agora retorno para a casa e vejo que tenho um irmão que já é quase um homem, um profissional que a cada dia me surpreende mais com seu belo trabalho. À minhas avós e avôs, que mesmo distante sempre rezaram por mim e me desejaram o melhor. A meus primos e ouros familiares, que sempre torceram pelo meu sucesso e a Tereza pelas orações.

Ao professor Jorge, que me recebeu de portas abertas no mestrado e com quem aprendi muito. Aos meus coorientadores, Pedro do NuBioMol, que me instruiu durante o mestrado e a pessoa que eu devo essa conquista, sem seus ensinamentos não teria chegado até aqui e a Manu, que foi meu suporte quando mudei de projeto e me encaminhou para que eu conseguisse realizar minha dissertação.

Ao professor Evanguedes por ter aceito o convite para participar da minha banca de defesa.

A Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES), pela bolsa concedida e a UFV, pela oportunidade de fazer o mestrado. Sou grata ao Núcleo de Análise de Biomoléculas da Universidade Federal de Viçosa, por fornecer as instalações para a condução dos experimentos.

A república “Casa da mãe Joana”, que foi meu lar durante seis anos de Viçosa, onde compartilhei muitos momentos bons e que carregarei para o resto da minha vida. E também a meu novo lar, aqui no apto 1002 da ladeira, onde conheci pessoas maravilhosas. Obrigada Barbara e Marlene.

Aos meus amigos de Viçosa, muitos estão ao meu lado desde 2011, me apoiando e me fazendo companhia, minha família fora de casa. Muito obrigada a Sofia, Lando, Fabinho, Michelle, Flávio, Isabel e todos que fazem dos meus dias mais felizes. Gostaria de agradecer também ao Fefe, que mesmo morando em JF, se mostrou um

amigo de verdade, onde nós compartilhamos nossas alegrais e frustrações. E também a Cintia, que sempre esteve comigo me apoiando.

Gostaria de agradecer ao Beagle, nesse tempo que fiquei lá, muitas pessoas passaram pelo laboratório e com cada uma aprendi coisas importantes. Bruno, Cassia, Daniel, Daniela, Fred nada, Jonas, Nati, Simone e Mariane. Um agradecimento especial ao Cristhian, que além de namorado, foi um ótimo amigo, que esteve sempre disposto a me ajudar e me fazer crescer. A Kessinha, amiga companheira. A Ariana, com quem compartilhei grande parte do mestrado, onde passamos por muitas coisas que nos ajudaram a crescer e a Sanely, um anjo que passou por lá, para alegrar o lab.

SUMÁRIO

LISTA DE FIGURAS

ARTIGO I

LISTA DE TABELAS

ARTIGO I

ARTIGO II

RESUMO

BARRETO, Cynthia Aparecida Valiati, M.Sc., Universidade Federal de Viçosa, outubro de 2017. **Sequenciamento, montagem e anotação dos genomas mitocondriais de *Astyanax giton* Eigenmann, 1908 e *Oligosarcus argenteus* Günther, 1864.** Orientador: Jorge Abdala Dergam dos Santos. Coorientadores: Manuela Maria Calvacante Granja e Pedro Marcus Pereira Vidigal**.**

Characidae é a maior e mais complexa família dentro de Characiformes e atualmente compreende 1123 espécies válidas. Entre seus representantes, 88 gêneros, incluindo *Astyanax* e *Oligosarcus*, são reconhecidos como *incertae sedis*. A espécie *Astyanax giton* distribui-se na bacia do rio Paraíba do Sul e rios costeiros do Espírito Santo e Rio de Janeiro. *Oligosarcus argenteus* distribui-se nos riachos formadores da drenagem do rio Doce, do rio das Velhas e do rio Paraopeba, estes dois na bacia do rio São Francisco. Até o momento, os genomas mitocondriais dessas espécies ainda não foram descritos. O estudo desses genomas permite avaliar o grau de informação dos genes mitocondriais na resolução filogenética, sendo úteis para a reconstrução da história evolutiva das espécies. Portanto, este trabalho tem como objetivo apresentar o primeiro estudo dos DNA mitocondriais de *A. giton* e *O. argenteus,* além da realização de análises dos polimorfismos, estimativa das taxas de substituição e filogenéticas do DNA mitocondrial de *O. argenteus* e de outros Characiformes. Os resultados mostram que os genomas mitocondriais de *A. giton* e *O. argenteus* possuem características típicas de vertebrados, com 37 genes anotados, sendo 13 sequencias codificadoras de DNA (CDS), dois RNA ribossomais (rRNA), 22 RNA de transferência (tRNA) e uma região de controle (D-loop). Os genomas mitocondriais disponíveis de Characiformes inclusos neste estudo, apresentam uma estrutura similar com o de *O. argenteus* e média geral de 85% de similaridade entre as sequências de aminoácidos das proteínas. Todas as CDS estão sob efeito da seleção purificadora, particularmente o gene COI que está sob a pressão mais seletiva entre todos os genes mitocondriais. O monofiletismo de Characiformes e sua separação em duas subordens (Characoidei + Citharinoidei) é suportado, além das relações Lebiasinidae + (((Chalceidae + ((Bryconidae + (Acestrorhynchidae + Characidae))) terem sido propostas. A topologia indicada pela árvore filogenética

obtida por inferência bayesiana coloca Cynodontidae e Hemiodontidae como grupos irmãos e representa uma nova hipótese das relações evolutivas desses táxons.

ABSTRACT

BARRETO, Cynthia Aparecida Valiati, M.Sc., Universidade Federal de Viçosa, October, 2017. **Sequencing, assembling and annotation of the mitochondrial genomes of *Astyanax giton* Eigenmann, 1908 and *Oligosarcus argenteus* Günther, 1864.** Adviser: Jorge Abdala Dergam dos Santos. Co-advisers: Manuela Maria Calvacante Granja and Pedro Marcus Pereira Vidigal**.**

Characidae is the largest and most complex family within Characiformes and currently comprises 1123 valid species. Among its representatives, 88 genera, including *Astyanax* and *Oligosarcus*, are recognized as *incertae sedis*. The species *Astyanax giton* is distributed among the Paraíba do Sul River Basin and coastal rivers of Espírito Santo and Rio de Janeiro, and *Oligosarcus argenteus* in the streams forming the Doce River, the Velhas River and the Paraopeba River, the two latter are tributaries of the São Francisco Basin. Until the present date, there are no studies describing the mitochondrial genomes of these species. The study of these genomes allow evaluating amount of information of mitochondrial genes in the phylogenetic resolution, being useful for the reconstruction of the evolutionary history of the species. So, the aim at this work is to present the first study of the mitochondrial DNA of *A. giton* and *O. argenteus*, besides the analysis of the polymorphisms, estimation of the substitution and phylogenetic rates of the mitochondrial DNA of *O. argenteus* and other Characiformes. The results show that the mitochondrial genomes of *A. giton* and *O. argenteus* have typical characteristics of vertebrates with 37 annotated genes, 13 codins DNA sequences (CDS), two ribosomal RNA (rRNA), transfer RNA (tRNA) and one control region (D-loop). The available mitochondrial genomes of Characiformes included in this study have a similar structure of *O. argenteus* and an average of 85% similarity between the amino acid sequences of the protein coding genes. All CDS are on the effect of purifying selection, particularly the *COI* gene which is under the most selective pressure among all mitochondrial genes. The results support the monophyly of Characiformes and their separation into two suborders (Characoidei + Citharinoidei), as well as the Lebiasinidae + (((Chalceidae + ((Bryconidae + (Acestrorhynchidae + Characidae))) relationships. The topology indicated by the phylogenetic tree obtained by bayesian inference

places Cynodontidae and Hemiodontidae as sister groups and represents a novel hypothesis of the evolutionary of these taxa.

INTRODUÇÃO GERAL

A ordem Characiformes é composta por 18 famílias e 270 gêneros de peixes, sendo a família Characidae a maior e mais complexa (REIS et al., 2003). Atualmente são reconhecidos dentro de Characidae 1.123 espécies e 145 gêneros, que ocorrem no sudoeste do Texas, México, América Central e do Sul (FROESE; PAULY, 2017). Esta família apresenta uma alta diversidade taxonômica, com uma ampla gama de tamanhos corporais e nichos ecológicos. Dentro de Characidae, 88 gêneros são reconhecidos como *incertae sedis*, ou seja, sem atribuição de subfamília (LIMA et al., 2003). Javonillo et al. (2010) propuseram uma hipótese filogenética para Characidae analisando 98 táxons usando marcadores moleculares e a família foi separada em três clados principais (Clados A, B e C). O clado A é grupo irmão do clado B e o clado C é o mais externo dos três. No clado C, foram incluídos representantes de alguns gêneros *incertae sedis,* entre eles *Astyanax* e *Oligosarcus.*

Os peixes do gênero *Oligosarcus* são popularmente conhecidos como peixes-cachorro e compreendem 21 espécies descritas (MENEZES; RIBEIRO, 2015; ALMIRÓN *et al., 2015; e* MENEZES; RIBEIRO, 2010), que são encontradas na América do Sul abaixo de 14º de latitude sul e distribuídas em muitas bacias hidrográficas brasileiras, além de rios na Bolívia, Argentina e Uruguai (MENEZES, 1987; MENEZES, 1988). A espécie *Oligosarcus argenteus* foi descrita por Günther (1864) com o Brasil como localidade tipo e se distribui por mais de uma bacia hidrográfica, sendo encontrados nos riachos formadores da drenagem do rio Doce, rio das Velhas e rio Paraopeba (FROESE; PAULY, 2017).

Já os peixes do gênero *Astyanax* apresentam 146 espécies válidas que são amplamente distribuídas desde a região neotropical do sul dos Estados Unidos até o norte da Argentina (FROESE; PAULY, 2017). São conhecidos popularmente como lambaris ou piabas e estão entre os mais abundantes e menos conhecidos táxons de peixes de água doce da América do Sul (SANTOS; NOVAES, 2008). Populações de *Astyanax giton* se distribuem na bacia do rio Paraíba do Sul e rios costeiros do Espírito Santo e Rio de Janeiro, Brasil. Atualmente, um trabalho com caracteres morfológicos e moleculares incluiu esta espécie no clado Probolodini o qual é composto por espécies dos gêneros *Deuterodon*, *Probolodus*, *Myxiops*,

Hyphessobrycon luetkenii, espécies de *Astyanax* da região costeira do Brasil e parte das espécies de *Jupiaba* (SILVA, 2017).

Com o desenvolvimento das técnicas de sequenciamento de nova geração (NGS), houve uma expansão de informações sobre o genoma de diversos organismos, inclusive sobre o DNA mitocondrial de peixes, havendo um acréscimo no número de trabalhos sobre esse tema nos últimos anos. Estes trabalhos tem permitido avaliar a atuação de genes mitocondriais na resolução filogenética, revelando-se como um recurso importante para filogeografia, biologia evolutiva, conservação e estudos de genética em peixes neotropicais (GUTIERREZ et al., 2015; CARVALHO et al., 2016).

O genoma mitocondrial típico de vertebrados, é uma molécula compacta de aproximadamente 16 kb que contém 37 genes, sendo 13 codificadores de proteínas, 22 RNA de transferência e dois genes ribossomais, além de uma região de controle (D-loop) (BOORE, 1999). De origem materna, os mitogenomas são muito utilizados em análises comparativas e filogenéticas devido a algumas de suas características, como abundância nos animais, pequeno tamanho e altas taxas de substituição nucleotídica, quando comparados a marcadores moleculares nucleares (BOORE, 1999; SATOH et al., 2016; TANGPHATSORNRUANG et al., 2016; WANG et al., 2016).

Até o momento, existem poucos genomas mitocondriais completos disponíveis no GenBank de Characiformes, de um total de 2.141 espécies válidas (FROESE; PAULY, 2017). Tendo em vista a escassez de informações sobre os mtDNA dessa Ordem, este trabalho tem o objetivo de pela primeira vez sequenciar, montar e realizar a anotação completa dos genomas mitocondriais das espécies *Astyanax giton* e *Oligosarcus argenteus,* além da realização de análises dos polimorfismos, estimativa das taxas de substituição e filogenéticas do DNA mitocondrial de *O. argenteus* e de outros Characiformes.

REFERÊNCIAS

ALMIRÓN, A.; CASCIOTTA, J.; PIÁLEK, L.; DOUBNEVORÁ, K.; RICAN, O. *Oligosarcus amome* (Ostariophysi: Characidae), a new species from the río Uruguay Basin, Misiones, Argentina. **Zootaxa,** v. 3915, n 4, *2015.*

BOORE, J. L. Animal mitochondrial genomes. **Nucleic Acids Research** , v. 27, n. 8, p. 1767–1780, 1999.

CARVALHO, D. C.; PERINI, R.; BASTOS, A. S.; et al. The complete mitochondrial genome of the threatened Neotropical catfish *Lophiosilurus alexandri* (Silurifomes : Pseudopimelodidae) and phylogenomic analysis indicate monophyly of Pimelodoidea. **Genetics and Molecular Biology**, p. 2013–2016, 2016.

FROESE, R.; PAULY, D. Editors 2011. FishBase. World Wide Web electronic publication. 2011. Disponível em: <www.fishbase.org>. Acessado em 14/08/2017

GUTIERREZ, V.; REGO, N.; NAYA, H.; GARCIA, G. First complete mitochondrial genome of the South American annual fish *Austrolebias charrua* (Cyprinodontiformes: Rivulidae): peculiar features among cyprinodontiforms mitogenomes. **BMC Genomics**, v. 16, n. 1, p. 879, 2015. BMC Genomics. Disponível em: <http://www.ncbi.nlm.nih.gov/pubmed/26511223>.

LIMA, F. C. T.; MALABARBA, L. R.; BUBKUP, P. A.; SILVA, J. F.; VARI, R. P.; HAROLD, A.; BENINE, R.; OYAKAWA, O. T.; PAVANELLI C. S.; MENEZES N. A.; LUCENA, C. A. S.; MALABARBA, M. C. S. L.; LUCENA, Z. M. S.; REIS, R. E.; LANGEANI, F.; CASSATI, L.; BERTAÇO, V. A.; MOREIRA, C.; LUCINDA P. H. F. (2003) Genera *Incertae Sedis* in Characidae. In: Reis, R. E.; Kullander, S. O.; Ferraris Jr., C. J. (Eds). p. 106-169. **Check list of the freshwater fishes of South and Central America.** Porto Alegre, Edipucrs, 729p.

REIS, R. E.; KULLANDER, S. O.; FERRARIS JR., C. J. 2003. **Checklist of the Freshwater 132 fishes of Central and South America**. Edipucrs, PUC, Porto Alegre, RS. 729p.

MENEZES, N. A. Três espécies novas de *Oligosarcus* Günther, 1864 e redefinição taxonômica das demais espécies do gênero (Osteichthyes, Teleostei, Characidae). **Boletim de Zoologia**, São Paulo, v. 11, p. 1–39, 1987.

MENEZES, N. A. Implications of the distribution patterns of the species of *Oligosarcus* (Teleostei; Characidae) from central and southern South America. *In*: W.R. Heyers & P.E. Vanzolini (Eds.). p. 295-304. **Proceedings**

of a Workshop on Neotropical Distribution Patterns. Rio de Janeiro, Academia Brasileira de Ciências, 488p, 1988.

MENEZES, N. A.; RIBEIRO, A. C. *Oligosarcus jacuiensis* (Characiformes: Characidae), a new species from the Uruguay and Jacuí River basins, southern Brazil. **Neotropical Ichthyology**, v. 8, n. 3, p. 649–653, 2010.

MENEZES, N. A.; RIBEIRO, A. C. A new species of the lowland *Oligosarcus* Günther species group (Teleostei: Ostariophysi: Characidae). **Neotropical Ichthyology**, v. 13, n. 3, p. 541–546, 2015.

SANTOS, A. C. A.; NOVAES, J. L. C. Population structure of two *Astyanax* Baird & Girard, 1854 (Teleostei, Characidae) species from Upper Paraguaçu River. 755 **Brazilian Archives of Biology and Technology**, p. 105–112, 2008.

SATOH, T. P.; MIYA, M.; MABUCHI, K.; NISHIDA, M. Structure and variation of the mitochondrial genome of fishes. **BMC Genomics**, p. 1–20, 2016. BMC Genomics. Disponível em: <http://dx.doi.org/10.1186/s12864-016-3054-y>.

SILVA, P.C. Sistemática integrativa – diversidade e relações de *Deuterodon* e gêneros afins. 26/04/2017. 199 f. Tese (Doutorado em Biologia Animal). Universidade Federal do Rio Grande do Sul, Rio Grande do Sul. 2017.

TANGPHATSORNRUANG, S.; RUANG-AREERATE, P.; SANGSRAKRU, D.; RUJIRAWAT, T. Comparative mitochondrial genome analysis of *Pythium insidiosum* and related oomycete species provides new insights into genetic variation and phylogenetic relationships. **Gene**, v. 575, n. 1, p. 34–41, 2016. Elsevier B.V. Disponível em: <http://dx.doi.org/10.1016/j.gene.2015.08.036>.

WANG, K.; LI, X.; DING, S.; et al. The complete mitochondrial genome of the *Atylotus miser* (Diptera : Tabanomorpha : Tabanidae), with mitochondrial genome phylogeny of lower Brachycera (Orthorrhapha). **Gene**, v. 586, n. 1, p. 184–196, 2016. Elsevier B.V. Disponível em: <http://dx.doi.org/10.1016/j.gene.2016.04.013>.

ARTIGO I.

Normas da revista *Gene*. ISSN: 0378-1119

Complete mitochondrial genome of *Oligosarcus argenteus* (Ostariophysi, Characidae): comparative and phylogenetic analysis within Characiformes

Cynthia Aparecida Valiati Barreto[a], Manuela Maria Cavalcante Granja[a], Pedro Marcus Pereira Vidigal[b], Cristhian Camilo Conde-Saldaña[a], Jorge Abdala Dergam[a]

[a] Laboratory of Molecular Systematics - Beagle, Department of Animal Biology, Universidade Federal de Viçosa, Viçosa, Minas Gerais, Brazil
[b] Núcleo de Análise de Biomoléculas (NuBioMol), Center of Biological Sciences (CCB), Universidade Federal de Viçosa, Minas Gerais, Brazil

Abstract

The mitogenome of *Oligosarcus argenteus* was sequenced and annotated. Following the general characteristics of vertebrates, this species has 16,711 bp, 13 coding DNA sequences (CDS), two ribosomal RNA genes (*rRNA*), 22 transference RNA genes (*tRNA*) and a control region (D-loop) of 1,036 bp. Within the order Characiformes, the analysis of non-synonimous/synonimous (dN/dS) substitutions of the 13 CDS revealed relatively lower values for the *COI* gene, suggesting strong purifying selection acting on Characiformes. The phylogenetic analysis of the CDS suggests a close relatedness between *Oligosarcus* and *Astyanax*. The taxonomic position of some taxa within the Characiformes remained unclear in the phylogenetic tree, and the position of Cynodontidae and Hemiodontidae as sister groups represents a novel hypothesis for these taxa.

Keywords: Characiformes, mitogenome, phylogenetic analysis, dN/dS.

1. Introduction

Characiformes fish are widely distributed in South and Central America, southern part of North America and Africa (Reis et al., 2003). These species predominate in the New World, with only 11% of African representatives (Eschmeyer and Fong, 2017). The understanding of the relationships between Characiformes has been the objective of numerous studies using morphological and molecular characters, however the relative position of some groups remain unresolved (Calcagnotto et al., 2005; Vari, 1979).

The Characidae family is the most diverse among the Characiformes and comprises 1,123 valid species (Eschmeyer and Fong, 2017). Included in this family is the genus *Oligosarcus*, which is popularly known as dogfish or bocarras and currently comprise 21 valid species (Almirón et al., 2015). This genus is found in South America below 14º south latitude and distributed in hydrographic basins of Brazil, Bolivia, Argentina and Uruguay (Menezes, 1987; Menezes, 1988). The genus *Oligosarcus* was previously classified in the subfamily Acestrorhynchinae, but currently this genus is considered *incertae sedis* and is not included in any suprageneric category (Lima et al., 2003), but is included in Clade C of Javonillo et al. (2010), which contains representatives of large *incertae sedis* genera.

Recently, the mitochondrial genomes have been used in phylogenetic analyses, being useful for the reconstruction of the evolutionary history of the species (Gutierrez et al., 2015). However, the complete sequences of mitochondrial Characiformes genomes, that are available in the NCBI Organelle Genome Resources database, are limited, evidencing the scarcity of information of this order that consist in 2,141 valid species (Eschmeyer and Fong, 2017).

In this study, we present the complete sequence and annotation of the *Oligosarcus argenteus* mitochondrial genome and compare its sequence with the genomes of Characiformes. In addition, the evolutionary relationships among *O. argenteus* and Characiformes species were analyzed based on the coding DNA sequences (CDS) present in their mitochondrial genomes.

2. Materials and methods

2.1. Sample collection and DNA extraction

Total genomic DNA from *Oligosarcus argenteus* (voucher number: CT4477) was extracted from muscle tissue using the proteinase K and phenol method (Sambrook et al., 2001). The specimen was collected at the head of the Latão Stream, in the city of Coimbra (20º49'66" S, 42º49'58" W) in the state of Minas Gerais, in southeastern Brazil. Samples of muscular tissues were collected at the Beagle Molecular Systematics Laboratory and the specimen was fixed in 10% formalin and deposited in the scientific collection of the Museum of Zoology João Moojen, both at the Federal University of Viçosa, Minas Gerais, Brazil.

2.2. DNA sequencing, assembly and annotation

The genomic library was sequenced with 2 X 300bp *paired-end* reads using the Illumina MiSeq (Illumina Inc., San Diego, CA). FASTQC version 0.11.5 was used to evaluate the quality of the sequencing and the *reads* were trimmed (Q20 score) and filtered by size (75 nt) using Trimmomatic version 0.33 (Bolger et al., 2014). The genome of *O. argenteus* was assembled using *de novo assembly* algorithm of the CLC Genomics Workbench version 6.5.1 (CLC bio). The mitochondrial genome sequence was selected and annotated using MitoAnnotator (Iwasaki et al., 2013), which contains a specific pipeline to analyze fish mitogenomes.

2.3. Comparative analysis

The mitogenome sequences of 32 fish-species belonging to the order Characiformes (Table 1) were downloaded from the NCBI Organelle Genome Resources database (https://www.ncbi.nlm.nih.gov/genome/organelle/). The sequences were aligned with the *O. argenteus* mitogenome using the *progressive Mauve* algorithm (Darling et al., 2010) and the *locally collinear blocks* (LCBs) were analyzed using *Mauve* version 2.4.0.

Table 1. GenBank accession codes to the taxa used in this study and their respective mitochondrial genome length, and GC content.

Family	Species	GenBank accession number	Total length (bp)	GC(%)	References
Acestrorhynchidae	*Acestrorhynchus sp.*	AP011981	16,758	42.72	Nakatani et al. (2011)
Parodontidae	*Apareiodon affinis*	AP011998	16,679	43.37	Nakatani et al. (2011)
Characidae	*Astyanax paranae*	KX609386	16,707	42.88	Silva et al. (2016)
Bryconidae	*Brycon henni*	KP027535	16,885	44.57	Landínez-García et al. (2014)
Bryconidae	*Brycon orbignyanus*	KM245044	16,800	42.79	Siqueira et al. (2014)
Chalceidae	*Chalceus macrolepidotus*	AB054130	16,850	44.18	Saitoh et al. (2003)
Chilodontidae	*Chilodus punctatus*	AP011984	16,869	40.16	Nakatani et al. (2011)
Citharinidae	*Citharinus congicus*	AP011985	16,453	46.28	Nakatani et al. (2011)
Curimatidae	*Curimata mivartii*	KP025764	16,705	43.22	Landínez-García et al. (2016a)
Distichodontidae	*Distichodus sexfasciatus*	AB070242	16,555	42.94	Nakatani et al. (2011)
Characidae	*Grundulus bogotensis*	KM677190	17,123	39.84	Isaza et al. (2014)
Characidae	*Hasemania nana*	AB861475	16,581	41.73	Xu et al. (2014)
Characidae	*Hemigrammus bleheri*	LC074360	17,021	41.62	Li et al. (2015)
Hemiodontidae	*Hemiodopsis gracilis*	AP011990	16,731	45.13	Nakatani et al. (2011)
Hepsetidae	*Hepsetus odoe*	AP011991	16,803	46.91	Nakatani et al. (2011)
Cynodontidae	*Hydrolycus scomberoides*	AP011989	16,548	47.64	Nakatani et al. (2011)
Distichodontidae	*Ichthyborus sp.*	AP011993	16,524	44.35	Nakatani et al. (2011)
Prochilodontidae	*Ichthyoelephas longirostris*	KP025763	16,840	45.45	Landínez-García et al. (2016b)
Lebiasinidae	*Lebiasina astrigata*	AP011995	16,899	42.58	Nakatani et al. (2011)
Anostomidae	*Megaleporinus piavussu*	KM886569	16,682	42.99	Yazbeck et al. (2014)
Alestidae	*Micralestes sp.*	AP011996	16,660	42.89	Nakatani et al. (2011)
Serrasalmidae	*Myleus sp.*	AP011997	16,664	47.89	Nakatani et al. (2011)
Characidae	*Oligosarcus argenteus*	MF805814	16,711	42.40	This study
Characidae	*Paracheirodon innesi*	KP013092	16,759	41.60	Yan et al. (2016)
Alestidae	*Phenacogrammus interruptus*	AB054129	16,652	44.63	Saitoh et al. (2003)
Serrasalmidae	*Piaractus brachypomus*	KJ993871	16,722	45.29	Chen et al. (2014)
Serrasalmidae	*Piaractus mesopotamicus*	KM245046	16,722	45.29	Pimentel et al. (2014)
Prochilodontidae	*Prochilodus argenteus*	KR014816	16,697	44.45	Chagas et al. (2016)
Prochilodontidae	*Prochilodus costatus*	KR014817	16,699	44.51	Chagas et al. (2015)
Prochilodontidae	*Prochilodus lineatus*	KM245045	16,699	44.48	do Carmo et al. (2014)
Serrasalmidae	*Pygocentrus nattereri*	AP012000	16,706	47.55	Nakatani et al. (2011)
Bryconidae	*Salminus brasiliensis*	KM245047	17,721	44.26	Brandão-Dias et al. (2014)

2.4. Polymorphism analysis and estimation of substitution rates

The CDS of mitogenomes were analyzed to evaluate conservation levels and to estimate the nucleotide substitution rates. The CDS and their respective proteins were aligned using the ClustalW algorithm (Thompson et al., 1994) in Mega version 7.0.14 (Kumar et al., 2016). The *stop* codons were excluded from the alignments obtained and the regions containing ambiguities were manually checked. The number of substitutions and sequence identity were calculated from the alignments using *pairwise distances* matrices.

Selective pressures acting on CDS codons were analyzed using the Data-Monkey server (http://www.datamonkey.org) (Delport et al., 2010) and the overall values of dN / dS were calculated following the method Single-likelihood ancestor counting (SLAC).

2.5. Phylogenetic analysis

Phylogenetic trees were obtained using the Maximum Likelihood (ML) and Bayesian inference (BI) methods. In these analyses, the catfish *Pimelodus pictus* (mitogenome access code in GenBank: AP012019) was selected as outgroup to root the phylogenetic trees. The sequences of the 13 CDS present in the mitogenomes of the 33 species were concatenated with the Sequence Matrix version 1.8 (Vaidya et al., 2011) and used for phylogenetic reconstruction.

The molecular evolution model was selected based on Akaike Information Criterion (AIC) in Partition Finder version 1.1.1 (Lanfear et al., 2012) and each codon position of the coding genes was treated as a separate partition, allowing different mutation rates among the codon positions. The model GTR + I + G (Tavaré et al., 1986) was chosen for 24 of the 39 partitions and the model SYM + I + G (Zharkikh, 1994) for the rest.

Maximum likelihood analysis (Felsenstein, 1981) was performed using Iq-tree version 1.5.4 (Nguyen et al., 2015) and clade consistency was verified from 10,000 pseudoreplicates using the ultrafast bootstrap method (Minh et al. Al., 2013).

Bayesian inference analysis was performed using the MrBayes version 3.2.6 (Ronquist et al., 2012) and the Markov chain Monte Carlo method (MCMC), from four independent chains with 10,000,000 generations, one phylogenetic tree was sampled every 10,000 generations. The first 25% of the generations were discarded with burn-in and the remaining trees were used to generate the statistics and topologies of the consensus phylogenetic tree. The trees and the posterior probability and bootstrap values were inspected using the FigTree version 1.4.0 (Rambaut and Drummond, 2012).

3. Results and discussion

3.1. Structure and organization of O. argenteus mitogenome

The mitogenome of *O. argenteus* was sequenced with a 322-fold coverage and has a sequence containing 16,711 nucleotides, with GC content of 42.4% (Figure 1). The sequences of the mitogenomes of the Characiformes species analyzed in this study have an average size of 16,763 bp (ranging from 16,453 bp to 17,721 bp) and an average GC content of 43.96% (ranging from 39.84% to 47.89%). When compared to these sequences, the mitogenome of *O. argenteus* is larger than the genomes of *Hydrolycus scomberoides* (16,548 bp) and *Curimata mivartii* (16,705 bp), and smaller than the genomes of *Salminus brasiliensis* (17,721 bp) and *Brycon henni* (16,885 bp). These size and GC content variation of the mitogenomes is related to insertions or deletions of nucleotides in intergenic regions or in the D-loop (Mcknight and Shafser, 1997; Rand, 1993).

The mitogenome of *O. argenteus* has a control region (D-loop) of 1,036 nt and a gene density of 2.21 genes / Kb. This genome has 14 intergenic regions (excluding the control region) with sizes ranging from 1 nt to 30 bp and adding up to 104 bp, corresponding to 0.62% of the genome. Thirty-seven genes were predicted and functionally annotated (Table 2), 13 coding DNA sequences (CDS), two ribosomal RNA genes (*rRNA*) and 22 transfer RNA (*tRNA*) genes. Most of the genes are located on the heavy strand, with exception for the *ND6* gene and eight *tRNAs* (*tRNAGln, tRNAAla, tRNAAsn, tRNACys, tRNATyr, tRNASer, tRNAGlu* and *tRNAPro*) that are located on the light strand (Table 2). The tRNAs encoded by the genes have sizes ranging from 68 to 75 nt, and only the serine and leucine tRNAs are duplicated in the genome.

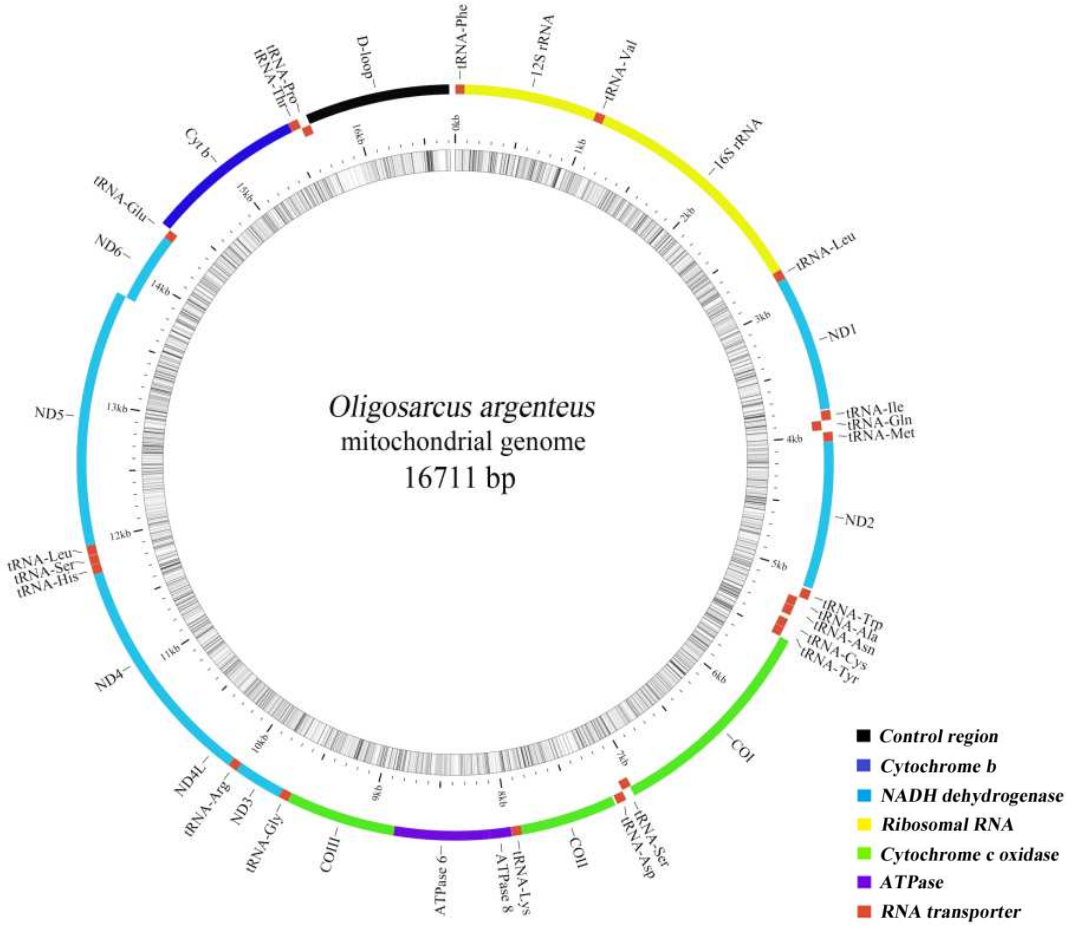

Fig. 1. Map of the mitochondrial genome of *Oligosarcus argenteus*. The genes outside the circle are transcribed heavy strand, while the genes inside are transcribed light strand. The inner ring shadow indicates the GC content of the genome.

The mitochondrial genomes of vertebrates are highly conserved and syntenic, and the same genic sequential arrangement can be found from lampreys to mammalian species (Hwang et al., 2013, Wang et al., 2014). This synteny was also observed in the Characiformes analyzed in this study, and the comparative analysis with Mauve did not detected any evidence of gene rearrangements in the mitochondrial genomes. Gene rearrangements occur in mitochondrial genomes of some fish orders such as Myctophiformes, Stomiiformes and Batrachoidiformes, which may lead to duplication of genes or displacement of a DNA fragment to another position in the genome (Boore 1999, Satoh et al., 2016).

Table 2. Characteristics of mitogenome of *Oligosarcus argenteus*. * Numbers correspond to nucleotides separating different genes, negative numbers indicate gene overlapping. (H = heavy strand, L = light strand).

	Location		Size		Codon		Intergenic	
Gene	**Start**	**End**	**Nucleotide (nt)**	**Amino acid**	**Start**	**Stop**	**nucleotide (bp)***	**Strand**
tRNA-Phe	1	68	68					H
12S rRNA	69	1019	951				0	H
tRNA-Val	1020	1091	72				0	H
16S rRNA	1092	2758	1667				0	H
tRNA-Leu(1)	2759	2833	75				0	H
ND1	2834	3805	972	323	ATG	TAA	0	H
tRNA-Ile	3815	3886	72				9	H
tRNA-Gln	3885	3955	71				0	L
tRNA-Met	3970	4039	70				14	H
ND2	4041	5096	1056	351	ATG	TAA	1	H
tRNA-Trp	5112	5182	71				15	H
tRNA-Ala	5190	5258	69				7	L
tRNA-Asn	5260	5332	72				1	L
tRNA-Cys	5363	5430	68				30	L
tRNA-Tyr	5430	5500	71				-1	L
COI	5502	7061	1560	519	GTG	AGG	1	H
tRNA-Ser(1)	7049	7120	72				-13	L
tRNA-Asp	7125	7196	72				4	H
COII	7211	7901	691	230	ATG	T--	14	H
tRNA-Lys	7902	7974	73				0	H
ATPase 8	7976	8143	168	55	ATG	TAA	1	H
ATPase 6	8134	8815	682	227	GTG	T--	-10	H
COIII	8816	9599	784	261	ATG	T--	0	H
tRNA-Gly	9600	9672	73				0	H
ND3	9673	10021	349	116	ATG	T--	0	H
tRNA-Arg	10022	10090	69				0	H
ND4L	10091	10387	297	98	ATG	TAA	0	H
ND4	10381	11761	1381	460	ATG	T--	-7	H
tRNA-His	11762	11830	69				0	H
tRNA-Ser(2)	11831	11898	68				0	H
tRNA-Leu(2)	11900	11972	73				1	H
ND5	11973	13811	1839	612	ATG	TAA	0	H
ND6	13808	14323	516	171	ATG	TAG	-4	L
tRNA-Glu	14324	14391	68				0	L
Cytb	14397	15533	1137	378	ATG	TAA	5	H
tRNA-Thr	15535	15606	72				1	H
tRNA-Pro	15605	15674	70				-2	L
Control region	15675	16711	1036				0	-

3.2. Characteristics of coding DNA sequences

The 13 CDS of *O. argenteus* mitochondrial genome have 11,432 nt in length, corresponding to 68.41% of the mitochondrial genome. Eleven genes start with the ATG codon, while the *COI* and *ATPase6* genes start with GTG. In fish species, the GTG primer codon is commonly found for the *COI* gene and rarely found in other genes (Satoh et al., 2016). Among the Characiformes, the species *Astyanax paranae* and *Ichthyoelephas longirostris* also present GTG as the initiator codon for the ATPase6 gene (Landínez-García et al., 2014; Silva et al., 2016). The *stop* codon TAA for *O. argenteus* is found in the *ND1, ND2, ATPase8, ND4L, ND5* and *Cytb* genes, whereas the TAG codon is present in *ND6*. The AGG codon acts as a *stop* codon for the *COI* and the other genes present an incomplete *stop* codon (T), which is also common in some fish species mitogenomes (do Carmo et al., 2014; Siqueira et al., 2014; Xu et al., 2015).

The amino acid sequences of proteins of the Characiform mitogenomes analyzed in this study show an average identity of 85%, with a mean of 38.72 substitutions in each gene (Table 3). The most conserved and most variable amino acid sequences was *COI* (519 nt) and ATPase8 (55 nt), with 95.9% and 70.3% of identity, respectively. The *ND5* gene was the most affected, with 110.73 substitutions, followed by *ND2*, with 97.16 substitutions. Such a high variation for *ND5* is related to the effect of the Bryconidae family standard, which showed 262 substitutions for this gene. Without the Bryconidae, the number of substitutions drops to 78.25. For the families Characidae, Prochilodontidae and Serrasalminae, the *ND2* gene predominates with a higher average of substitution.

3.3. Non-synonymous and synonymous substitutions

The relationship between non-synonymous and synonyms substitutions (dN / dS) is widely used to estimate selection pressures in protein coding regions (Kryazhimskiy and Plotkin, 2008). The ratio dN / dS = 1 indicates neutrality, dN / dS > 1 positive selection and dN / dS < 1 negative or purifying selection (Nei, 2005). To evaluate the differences in the evolutionary rates of the 13 CDS of the 32 Characiformes species, the overall values of dN / dS were analyzed. The dN / dS

values of all the protein coding genes were less than 1, indicating the action of the purifying (negative) selection eliminating the deleterious alleles (Table 3). The lowest dN / dS values were observed for the *COI* gene, followed by *COIII* and *CYTB*. These results are similar to those observed in the species of Perciformes of the Labridae family, evidencing a greater purifying selection on these genes (Zhu et al., 2017). On the other hand, the highest dN / dS values were observed in the *ATPase8* gene, whereas the *ND5* gene had the highest dN / dS values in Perciformes (Zhu et al., 2017).

Table 3. Values for the number of substitutions and identity for the amino acid sequences, and dN/dS for the genes of the 13 CDS of the Characiform mitogenomes.

Gene	Substitutions	Identity	dN/dS
ATPase6	30.341	86.6%	0.074
ATPase8	15.746	70.3%	0.183
COI	21.28	95.9%	0.022
COII	19.988	91.2%	0.052
COIII	16.669	93.6%	0.049
CYTB	35.224	90.5%	0.050
ND1	32.363	89.7%	0.058
ND2	97.161	71.7%	0.136
ND3	14.583	87.1%	0.092
ND4	62.212	86.5%	0.075
ND4L	13.405	86.3%	0.077
ND5	110.73	81.5%	0.122
ND6	33.645	79.6%	0.144

DNA barcoding is a molecular taxonomic identification tool, that uses the mitochondrial gene cytochrome oxidase subunit I, or *COI*. This gene was proposed by Hebert et al. (2003), as it has a robust universal primer, capable of amplifying the gene in several different species and presents a strong phylogenetic signal. This gene also has lower amino acid substitution rates when compared to other CDS (Lynch and Jarrell, 1993), and it may be possible to assign a higher taxonomic group to an unidentified organism (Hebert et al., 2003). Fish studies were able to separate the species with *COI*, in addition to the observation of some phylogenetic characters (Ward et al., 2005). This present work corroborates what has already been observed by Lynch and Jarrell (1993), being the *COI* gene with lower substitution rate among the CDS and strong selective pressure acting on this gene, which makes it a suitable

and robust genetic marker for Characiformes.

3.6. Phylogenetic analysis

The topology of phylogenetic trees obtained by Maximum Likelihood (ML) and Bayesian inference (BI) presented some conflicts, which will be discussed later. On the other hand, the topologies obtained with the two methods were identical in the following aspects: Characiformes monophyletism was supported with high values of posterior probability (PP=1), as well as the separation of the clades of the suborders Citharinoidei and Characoidei, the first one composed exclusively of African species. In the first group, the posterior probability (PP=1) and bootstrap values (BV=100%) corroborated the close relationship of the African families Citharinidae and Distichodontidae. Both taxa are also the sister group of the other Characiformes, based on osteological characters (Vari, 1979). Vari's hypothesis was later confirmed by analyses using molecular markers (Calcagnotto et al., 2005; Oliveira et al., 2011).

Within the Characoidei, the African families Hepsetidae and Alestidae are the sister group of the South American families. The monophyly of the families with two or more individuals analyzed in this study (Prochilodontidae, Serrasalminae, Alestidae, Characidae, Bryconidae and Distichodontidae) was supported with high values of PP and BV.

The phylogenetic reconstruction of the relationships of Anastomoidea clade (Anostomidae, Chilodontidae, Curimatidae and Prochilodontidae) with other families has varied among trees based on molecular (Oliveira et al., 2011; Calcagnotto et al., 2005) and morphological (Buckup 1998) data. In this study, Parodontidae is more closely related to Anastomoidea, and these form a sister group of a clade composed of Hemiodontidae, Cynodontidae and Serrasalmidae. However, in the study by Oliveira et al. (2011), Anastomoidea appears as sister group of a clade formed by Serrasalmidae plus Hemiodontidae, and Cynodontidae as the sister group of the tree. These divergences suggest that further studies are necessary to elucidate the relationships of Anastomoidea with other families, within Characiformes.

In this study, Serrasalmidae was found as a sister group of the clade formed by Cynodontidae and Hemiodontidae. However, the phylogenetic position of Serrasalmidae has been controversial. In the hypothesis generated by Orti and Meyer (1997), based on two mitochondrial markers, the phylogenetic position of Serrasalmidae was unclear. Serrasalmidae has received weak support as a sister group of Anastomoidea plus Hemiodontidae and Parodontidae (Calcagnotto et al., 2005), and as sister group of Hemiodontidae in the most recent phylogenetic reconstruction of the order Characiformes (Oliveira et al. 2011). The present study supported Serrasalmidae as a monophyletic group, however, future studies are needed to clarify their relationships with related families. On the other hand, the topology of Cynodontidae and Hemiodontidae as sister groups (PP=1; BV=45%), represents a novel hypothesis of the relations between these taxa.

Several hypotheses have been raised to try to explain the positioning of Hepsetidae within Characiformes. In a consensus cladogram, Buckup (1998) proposed that Hepsetidae and Ctenoluciidae formed a sister group, and they were related to Lesbiasinidae and Erythrinidae, but this relationship was unresolved. Subsequently, Calcagnotto et al. (2005) hypothesized that Hepsetidae, Ctenolucidae and Lebiasinidae were more closely related, and these families formed a sister group of Alestidae. The results of this study support the hypothesis of Oliveira et al. (2011), suggesting a close relationship between Alestidae and Hepsetidae (PP=100%; BV=95%).

The Lebiasinidae, Chalceidae, Bryconidae, Acestrorhynchidae and Characidae families formed a monophyletic group (PP=1; BV=76%). Within this group, Lebiasinidae was considered as the sister group of all remaining families and Bryconidae appeared as the sister group of Acestrorhynchidae and Characidae (PP=100%; BV=98%). A similar pattern was found by Oliveira et al. (2011), however in their study Acestrorhynchidae appears as the sister group of Characidae and Byconidae. The hypotheses of Oliveira et al. (2011) and Calcagnotto et al. (2005) are consistent. However, in Calcagnotto et al. (2005), Lebiasinidae forms a clade with Ctenoluciidae and Hepsetidae.

The species *Oligosarcus argenteus* and *Astyanax paranae* formed a monophyletic group (PP=100%; BV=100%) and their close relationship had previously been described (Ortí and Meyer, 1997; Javonillo et al., 2010; Oliveira et al., 2011). They are included in the Clade C of Javonillo et al. (2010), placed together with the subfamily Stethaprioninae and other currently *incertae sedis* genera.

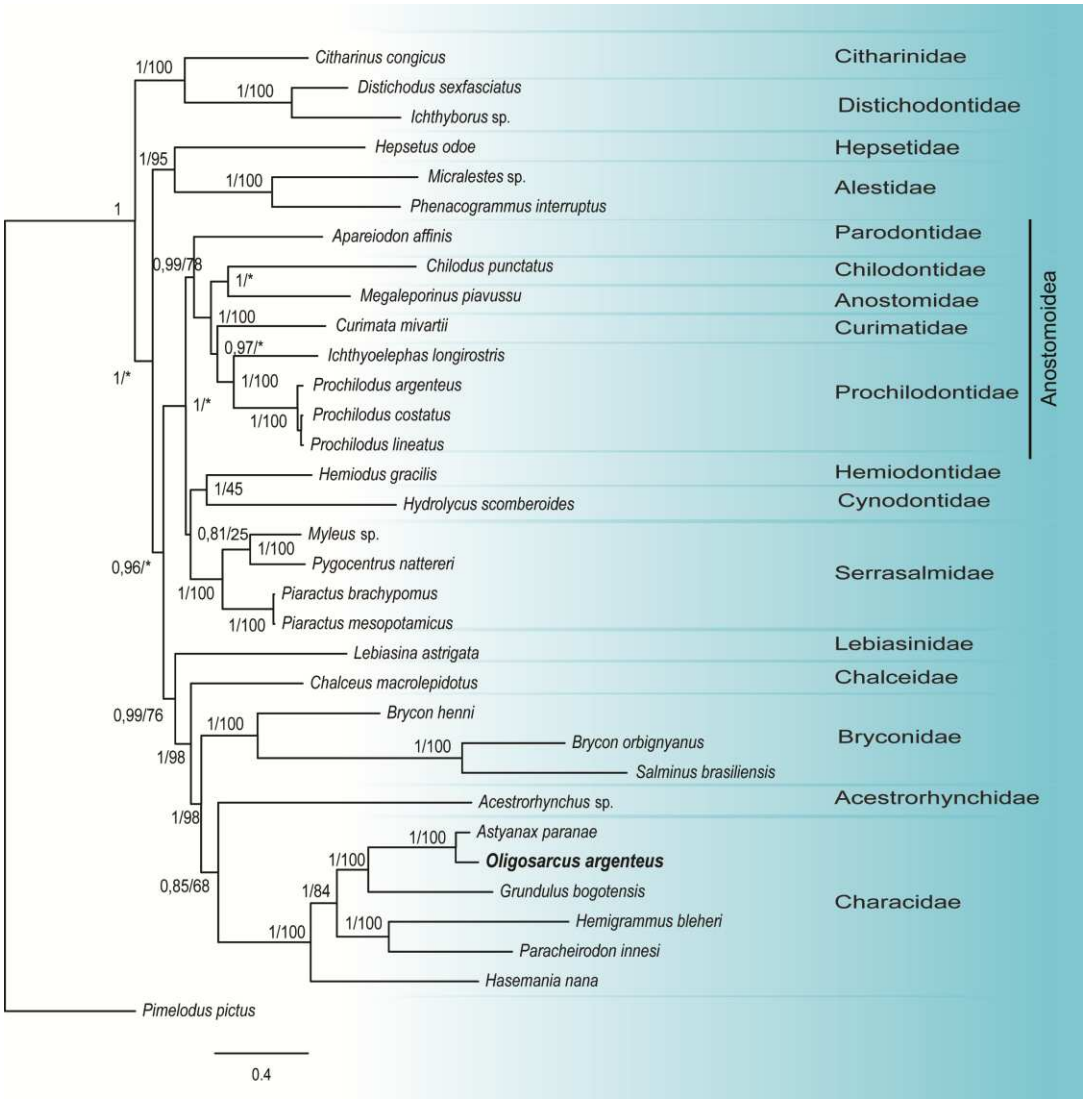

Fig. 2. Bayesian phylogenetic tree using the 13 coding DNA sequences (CDS) of mtDNA for Characiformes. Numbers at each node represents the posterior probability (PP) obtained in Bayesian analysis (BI), and percentage of bootstrap values (BV) obtained by Maximum Likelihood (ML). Asterisks represent nodes that were not obtained using ML analyses.

4. Conclusions

The complete mitochondrial genome of *Oligosarcus argenteus* presented in this study adds new information to the set of available genomes of the Characiformes species, contributing to the studies of evolutionary biology of this order. The mitochondrial genome of *O. argenteus* has typical vertebrate characteristics of 16,711 bp and contains 13 genes encoding proteins, two ribosomal RNAs, 22 RNA transfer and a control region of 1,036 nt. The available mitochondrial genomes of Characiformes included in this study have a similar structure with *O. argenteus* and an average of 85% similarity among the amino acid sequences of mitochondrial proteins. All CDS are under the effect of purifying selection, particularly the *COI* gene, which had the lowest values of dN / dS. The results support the monophyly of Characiformes and their separation into two suborders (Characoidei + Citharinoidei), as well as the Lebiasinidae + (((Chalceidae + ((Bryconidae + (Acestrorhynchidae + Characidae))) relationships. The topology indicated by the BI tree of Cynodontidae and Hemiodontidae as sister groups represents a novel hypothesis of the relations of these taxa.

5. References

Almirón, A., Casciotta, J., Piálek, L., Doubnerová, K. and Ríčan, O., 2015. *Oligosarcus amome* (Ostariophysi: Characidae), a new species from the río Uruguay Basin, Misiones, Argentina. Zootaxa, 3915.

Bolger, A.M., Lohse, M., Usadel, B., 2014. Genome analysis Trimmomatic : a flexible trimmer for Illumina sequence data. Bioinformatics 1–7. doi:10.1093/bioinformatics/btu170

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

Brandão-Dias, P.F.P., Carmo, A.O. do, Martins, A.P.V., Pimenta, R.J.G., Alves, C.B.M., Kalapothakis, E., 2014. Complete mitochondrial genome of *Salminus brasiliensis* (Characiformes, Characidae). Mitochondrial DNA 1736, 1–2. doi:10.3109/19401736.2014.958676

Calcagnotto, D., Schaefer, S.A., Desalle, R., 2005. Relationships among characiform fishes inferred from analysis of nuclear and mitochondrial gene sequences. Mol. Phylogenet. Evol. 36, 135–153. doi:10.1016/j.ympev.2005.01.004

Chen, H., Li, S., Xie, Z., Zhang, Y., Zhu, C., Deng, S., Li, G., Huang, H., 2016. The complete mitochondrial genome of the *Piaractus brachypomus* (Characiformes: Characidae). Mitochondrial DNA 27, 1289–1290. doi:10.3109/19401736.2014.945560

Darling, A.E., Mau, B., Perna, N.T., 2010. Progressivemauve: Multiple genome alignment with gene gain, loss and rearrangement. PLoS One 5. doi:10.1371/journal.pone.0011147

Delport, W., Poon, A.F.Y., Frost, S.D.W., Pond, S.L.K., 2010. Datamonkey 2010 : a suite of phylogenetic analysis tools for evolutionary biology. Bioinformatics 26, 2455–2457. doi:10.1093/bioinformatics/btq429

do Carmo, A.O., Brandão Dias, P.F.P., Martins, A.P.V., Bedore, A.G., Kalapothakis, E., 2014. Complete mitochondrial genome sequence of *Prochilodus lineatus* (Characiformes, Prochilodontidae). Mitochondrial DNA 1736, 1–2. doi:10.3109/19401736.2014.971300

Gutierrez, V., Rego, N., Naya, H., Garcia, G., 2015. First complete mitochondrial genome of the South American annual fish Austrolebias charrua (Cyprinodontiformes: Rivulidae): peculiar features among cyprinodontiforms mitogenomes. BMC Genomics 16, 879. doi:10.1186/s12864-015-2090-3

Hwang, D., Byeon, H.W.A.K.U.N., Lee, J., 2013. Complete mitochondrial genome of the sand lamprey , Lampetra. Mitochondrial DNA 1–2. doi:10.3109/19401736.2013.773989

Hebert, P.D.N., Cywinska, A., Ball, S.L., Jeremy, R., 2003. Biological identifications through DNA barcodes. The Royal Society. 313–321. doi:10.1098/rspb.2002.2218

Isaza, J.P., Alzate, J.F., Maldonado-Ocampo, J. a, 2014. Complete mitochondrial genome sequence of *Grundulus bogotensis* (Humboldt, 1821). Mitochondrial DNA 1736, 1–3. doi:10.3109/19401736.2014.982563

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. doi:10.1093/molbev/mst141

Javonillo, R., Malabarba, L.R., Weitzman, S.H., Burns, J.R., 2010. Molecular Phylogenetics and evolution relationships among major lineages of characid fishes (Teleostei : Ostariophysi : Characiformes), based on molecular sequence data. Mol. Phylogenet. Evol. 54, 498–511. doi:10.1016/j.ympev.2009.08.026

Kryazhimskiy, S., Plotkin, J.B., 2008. The population genetics of dN / dS. PloS Genet. 4. doi:10.1371/journal.pgen.1000304

Kumar, S., Stecher, G., Tamura, K., 2016. MEGA7 : Molecular evolutionary genetics analysis version 7.0 for bigger datasets. Mol. Biol. Evol. 33, 1870–1874. doi:10.1093/molbev/msw054

Landínez-García, R.M., Alzate, J.F., Márquez, E.J., 2016a. Mitochondrial genome of the neotropical detritivore fish *Curimata mivartii* Steindachner 1878 (Characiformes, Curimatidae). Mitochondrial DNA Part A 27, 2196–2197. doi:10.3109/19401736.2014.982623

Landínez-García, R.M., Alzate, J.F., Márquez, E.J., 2016b. Mitochondrial genome of the neotropical trans-Andean fish *Ichthyoelephas longirostris*, Steindachner 1879 (Characiformes, Prochilodontidae). Mitochondrial DNA Part A 27, 2194–2195. doi:10.3109/19401736.2014.982622

Landínez-García, R.M., Alzate, J.F., Márquez, E.J., 2014. Complete mitogenome of the neotropical fish *Brycon henni*, Eigenmann 1913 (Characiformes, Bryconidae). Mitochondrial DNA Part A 27, 2259–2260. doi:10.3109/19401736.2014.984170

Li, C., Sun, Z., Feng, S., Jiang, J., Wu, H., Zhang, Z., Cai, C., Wang, Y., 2015. The complete mitochondrial genome of *Hemigrammus bleberi*. Mitochondrial DNA 1736, 1–2. doi:10.3109/19401736.2015.1089565

Lynch, M., Jarrell, P.E., 1993. A Method for Calibrating Molecular Clocks and Its Application to Animal Mitochondrial DNA. Genetics 135, 1197–1208.

Mcknight, M.L., Shafser, H.B., 1997. Large, rapidly evolving intergenic spacers in the mitochondrial the Salamander Family Ambystomatidae (Amphibia : Caudata) DNA of. Mol. Biol. Evol. 14, 1167–1176.

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. doi:10.1186/1471-2148-11-177

Oliveira, C., Avelino, G.S., Abe, K.T., Mariguela, T.C., Benine, R.C., Ortí, G., Vari,

R.P., Corrêa, R.M., 2011. Phylogenetic relationships within the speciose family Characidae (Teleostei : Ostariophysi : Characiformes) based on multilocus analysis and extensive ingroup sampling. BMC Evol. Biol. 11, 1–25.

Pimentel, J. da S.M., Carmo, A.O. do, Maciel, D. de C.L., Siqueira, F. de F., Kalapothakis, E., 2014. Complete mitochondrial genome sequence of Piaractus mesopotamicus (Holmberg, 1887). Mitochondrial DNA 1736, 1–2. doi:10.3109/19401736.2014.971297

Rand, D.M., 1993. Endotherms, ectotherms, and mitochondrial genome-size variation. J. Mol. Evol. 37, 281–295.

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. doi:10.1007/s00239-002-2417-y

Sambrook, J., Russell, D., 2001. Molecular cloning: A laboratory manual, 3rd Ed. Cold Spring harbor Laboratory Press, Cold Spring Harbor, NY, USA, 682p.

Satoh, T.P., Miya, M., Mabuchi, K., Nishida, M., 2016. Structure and variation of the mitochondrial genome of fishes. BMC Genomics 1–20. doi:10.1186/s12864-016-3054-y

Silva, D.M.Z.D.A., Utsunomia, R., Ruiz-ruano, F.J., Foresti, F., Silva, D.M.Z.D.A., Utsunomia, R., Ruiz-ruano, F.J., 2016. The complete mitochondrial genome sequence of *Astyanax paranae* (Teleostei : characiformes). Mitochondrial DNA Part B Resour. doi:10.1080/23802359.2016.1222251

Siqueira, D.F., Oliveira, A., Martins, S., 2014. Complete mitochondrial genome sequence of *Brycon orbignyanus*. Mitochondrial DNA 1736, 1–2. doi:10.3109/19401736.2014.971298

Wang, W., Liu, J., Wang, H., Yang, M., Liu, Q., Ding, M., Wang, W., 2014. The complete mitochondrial genome of white-tufted-ear marmoset , *Callithrix jacchus* (Primates : Callitrichinae). Mitochondrial DNA 1736, 1–2. doi:10.3109/19401736.2014.971287

Ward, R.D., Zemlak, T.S., Innes, B.H., Last, P.R., Hebert, P.D.N., 2005. DNA barcoding Australia's fish species. Philos. Transacions R. Soc. 1–11. doi:10.1098/rstb.2005.1716

Xu, R., Zhao, Z., Xu, P., Sun, X., 2015. The complete mitochondrial genome of the silvertip tetra , *Hasemania nana* (Characiformes : Characidae). Mitochondrial DNA 1736. doi:10.3109/19401736.2013.861445

Yan, A., Liu, F., Jiang, H., Feng, C., Tang, D., Yan, A., Liu, F., Jiang, H., Feng, C., Tang, D., 2016. The complete mitochondrial genome of *Paracheirodon innesi*. Mitochondrial DNA 1736. doi:10.3109/19401736.2015.112683

Zhu, K., Wu, N., Sun, X., Guo, H., Zhang, N., Jiang, S., Zhang, D., 2017. Characterization of complete mitochondrial genome of five tripe wrasse (*Thalassoma quinquevittatum* , Lay & Bennett , 1839) and phylogenetic analysis. Mitochondrial DNA 598, 71–78. doi:10.1016/j.gene.2016.10.042

ARTIGO II.

Normas da revista *Mitochondrial DNA Part B*. ISSN: 2380-2359

Complete mitochondrial genome sequence of Neotropical fish *Astyanax giton* Eigenmann 1908 (Ostariophysi, Characidae)

Cynthia Aparecida Valiati Barreto[a], Manuela Maria Cavalcante Granja[a], Pedro Marcus Pereira Vidigal[b], Jorge Abdala Dergam[a]

[a] Laboratory of Molecular Systematics - Beagle, Department of Animal Biology, Universidade Federal de Viçosa, Viçosa, Minas Gerais, Brazil

[b] Núcleo de Análise de Biomoléculas (NuBioMol), Center of Biological Sciences (CCB), Universidade Federal de Viçosa, Minas Gerais, Brazil

Abstract

We present, for the first time, the complete mitochondrial genome of *Astyanax giton*, which together with other species, are popularly known as tetras. The mitogenome's length is 16,643 bp, containing 13 protein-encoding genes (CDS), two ribosomal RNA (rRNA), 22 RNA transfer (tRNA) and one control region (D-loop). The order of the genes is typical of vertebrates and the nucleotide content was 29.7% for A, 25.10% for C, 15.6% for G and 29.5% for T. As for other vertebrates, all genes are encoded on the heavy strand except for *ND6* and eight *tRNA* genes.

Keywords: freshwater fish, mitogenome, next generation sequencing

In the Characidae family, the genus *Astyanax* is a taxon with little known phylogenetic kinships and are popularly known as tetras. It currently comprises 146 valid species with a wide distribution on the Neotropical region, occurring from the southern United States to northern Argentina (Froese & Pauly, 2016). The taxonomy of the *Astyanax* genus is still unsettled, although morphological (Mirande, 2009) and molecular data (Javonillo et al., 2010) indicate that this taxon is paraphyletic. Until the present date, only two mitogenomes of the genus *Astyanax* have been described, which underlines the relevance of the description of the complete mitogenome of *Astyanax giton*.

The specimen *A. giton* was collected in the Doce River Basin, at the headwaters of the Latão Stream, in the city of Coimbra (20º49'66" S, 42º49'58" W) in the state of Minas Gerais, in southeastern Brazil. Samples of muscular tissues were collected at the Beagle Molecular Systematics Laboratory and the specimen was fixed in 10% formalin and deposited in the scientific collection of the Museum of Zoology João Moojen, both at the Federal University of Viçosa, Minas Gerais, Brazil. (voucher number: CT2205). The genomic library was sequenced with 2 X 300 bp *paired-end* reads using the Illumina MiSeq (Illumina Inc., San Diego, CA). The quality of the sequencing was evaluated using FASTQC v.0.11.5 and the *reads* were trimmed (Q20 score) and filtered by size (75 nt) using Trimmomatic v0.33 (Bolger et al., 2014). The mitogenome of *A. giton* was assembled using a *de novo assembly* in CLC Genomics Workbench v6.5.1 (CLC bio). Among the sequences obtained, a contig of ~ 16.6 Kb size significantly aligned showed ~ 77% identity with the *Danio rerio* mitogenome (GenBank accession number NC_002333) using BLASTN in the CLC Genomics Workbench. This *contig* was selected and the genes were predicted and annotated using MitoAnnotator (Iwasaki et al., 2013).

The mitogenome of *A. giton* is typical of vertebrates (Boore, 1999) and its length is 16,643 bp (GenBank Access MF805815), with a 121-fold coverage. When compared to the mitogenomas of *Astyanax paranae* (KX609386) and *Astyanax mexicanus* (AP011982), the mitogenoma of *A. giton* shows 93% identity with *A. paranae* and 87% with *A. mexicanus*. The GC content is 40% and the total base composition is A: 29.7%, C: 25.10%, G: 15.6% and T: 29.5%. Thirty-seven genes were functionally annotated, 13 coding DNA sequences (CDS), two ribosomal RNA

(*rRNA*), 22 transfer RNA (*tRNA)* and a control region (D-loop) of 997 bp (Table 1). Most of the genes are located on the heavy strand (H), except for eight *tRNAs* (*tRNAGln, tRNAAla, tRNAAsn, tRNACys, tRNATyr, tRNASer, tRNAGlu* and *tRNAPro*) and a CDS (*ND6*). *Astyanax giton* has 14 intergenic regions in its mitogenome (excluding the D-loop), which add up to 94 bp in length. Gene overlapping occurs at six sites (*tRNAIli* / *tRNAGln, COI* / *tRNASer, ATPase6* / *ATPase8, ND4* / *ND4L, ND5* / *ND6* and *tRNAThr* / *tRNAPro*), ranging from 1 to 13 bp overlapping.

The 13 CDS of *A. giton* represents 68.6% of the total mitogenome. The *COI* and *ND3* genes start with the GTG codon and the others with the usual ATG codon. Three of the 13 CDS contains the stop codon TAA (*ND1*, *ND4L* and *ND6*); five the stop codon incomplete T- (*COII*, *ATPase6, COIII*, *ND3* and *ND4*); three stop codon TAG (*ND2*, *ATPase8* and *CYTB*) and two the stop codon AGG (*COI* and *ND5*).

Table 1. Annotation of the mitochondrial genome of *Astyanax giton*. * Numbers correspond to nucleotides separating different genes, negative numbers indicate gene overlapping. (H = heavy strand, L = light strand).

	Location		Size (nt)	Amino acid	Codon		Intergenic nucleotide (bp)*	Strand
Gene	**Start**	**End**			**Start**	**Stop**		
tRNA-Phe	1	68	68					H
12S rRNA	69	1022	954				0	H
tRNA-Val	1023	1094	72				0	H
16S rRNA	1095	2766	1672				0	H
tRNA-Leu(1)	2767	2841	75				0	H
ND1	2842	3813	972	323	ATG	TAA	0	H
tRNA-Ile	3822	3893	72				8	H
tRNA-Gln	3892	3962	71				-2	L
tRNA-Met	3975	4044	70				12	H
ND2	4047	5099	1053	350	ATG	TAG	2	H
tRNA-Trp	5107	5177	71				7	H
tRNA-Ala	5188	5256	69				9	L
tRNA-Asn	5258	5330	73				1	L
tRNA-Cys	5361	5427	67				30	L
tRNA-Tyr	5428	5498	71				0	L
COI	5500	7059	1560	519	GTG	AGG	1	H
tRNA-Ser(1)	7047	7118	72				-13	L
tRNA-Asp	7120	7189	70				1	H
COII	7203	7893	691	230	ATG	T--	13	H
tRNA-Lys	7894	7966	73				0	H
ATPase 8	7968	8135	168	55	ATG	TAG	1	H
ATPase 6	8126	8807	682	227	ATG	T--	-10	H
COIII	8808	9591	784	261	ATG	T--	0	H
tRNA-Gly	9592	9664	73				0	H
ND3	9665	10013	349	116	GTG	T--	0	H
tRNA-Arg	10014	10083	70				0	H
ND4L	10084	10380	297	98	ATG	TAA	0	H
ND4	10374	11754	1381	460	ATG	T--	-7	H
tRNA-His	11755	11823	69				0	H
tRNA-Ser(2)	11824	11891	68				0	H
tRNA-Leu(2)	11893	11965	73				1	H
ND5	11966	13801	1836	611	ATG	AGG	0	H
ND6	13796	14311	516	171	ATG	TAA	-6	L
tRNA-Glu	14312	14379	68				0	L
Cytb	14385	15521	1137	378	ATG	TAG	5	H
tRNA-Thr	15525	15597	73				3	H
tRNA-Pro	15597	15666	70				-1	L
Control region	15667	16643	977				0	-

References

Bolger AM, Lohse M, Usadel, B. 2014. Trimmomatic: A flexible trimmer for Illumina Sequence Data. Bioinformatics. 1–7.

Boore JL. 1999. Animal mitochondrial genomes. Nucleic Acids Res. 27:1767–80.

Froese R, Pauly D. 2011. FishBase: World Wide Web electronic publication; [cited 2017 April 19], Available from:www.fishbase.org.

Iwasaki W, Fukunaga T, Isagozawa R, Yamada K, Maeda Y, Satoh TP, et al. 2013. MitoFish and MitoAnnotator: a mitochondrial genome database of fish with an accurate and automatic annotation pipeline. Mol Biol Evol 30:2531–2540.

Javonillo R, Malabarba LR, Weitzman SH, Burns JR. 2010. Relationships 651 among major lineages of characid fishes (Teleostei: Ostariophysi: Characiformes), 652 based on molecular sequence data. Mol Phylogenet and Evol. 54:653 498–511.

Lima FCT, Malabarba LR, Buckup PA, da Silva JFP, Vari RP, Harold A, Benine R, et al. 2003. Characidae, genera incertae sedis. In: Reis RE, Kullander SP, Ferraris CJ, editors. Check List of the freshwater fishes of South and Central America. Porto Alegre: Edipucrs. p 106–69.

Mirande JM. 2009. Weighted parsimony phylogeny of the family Characidae. 692 (Teleostei: Characiformes). Cladistics. 25:574–613.

CONCLUSÕES GERAIS

Este trabalho representa o primeiro sequenciamento, montagem e anotação completa dos genomas mitocondriais das espécies *Astyanax giton* e *Oligosarcus argenteus*. Ele acrescenta novas informações ao conjunto de genomas disponíveis das espécies da ordem Characiformes, contribuindo para os estudos de biologia comparativa e evolutiva. Os genomas mitocondriais de ambos as espécies são típicos de vertebrados, com os tamanhos ~16 Kb e 37 genes.

A análise comparativa das sequências de aminoácidos das sequencias codificadoras de DNA (CDS) apresentaram uma média geral de 85% de identidade entre *O. argenteus* e os Characiformes inclusos neste estudo. Todas as CDS estão sobre efeito da seleção purificadora, sendo o gene *COI* sobre maior pressão seletiva e o gene *ATPase8* sobre menor pressão seletiva. Os resultados da analise filogenética suportam o monofiletismo de Characiformes e sua separação em duas subordens (Characoidei + Citharinoidei), além das relações Lebiasinidae + (((Chalceidae + ((Bryconidae + (Acestrorhynchidae + Characidae))) terem sido propostas. É evidenciada a necessidade de futuros estudos para esclarecer o posicionamento de Anastomoidea e Serrasalmidae. A topologia indicada pela inferência bayesiana (IB) de Cynodontidae e Hemiodontidae como grupos irmãos representa uma nova hipótese das relações desses táxons.