

Ten complete mitochondrial genomes of Gymnocharacini (Stethaprioninae, Characiformes): evolutionary relationships and a repetitive element in the Control Region (D-loop)

Rubens Pasa^{1,3}, Fabiano Bezerra Menegídio², Igor Henrique Rodrigues-Oliveira¹, Iuri Batista da Silva¹, Matheus Lewi Cruz Bonaccorsi de Campos¹, Dinaíza Abadia Rocha-Reis¹, John Seymour (Pat) Heslop-Harrison⁴, Trude Schwarzacher⁴, Karine Frehner Kavalco^{1,3}

Corresponding author: Rubens Pasa^{1,3}

E-mail: rpasa@ufv.br **Phone:** +55 (34) 99773-8285

Postal Address: Federal University of Viçosa, Route MG 230, km 7, LAE Building, room 101. Zip code 38810-000, Rio Paranaíba, MG, Brazil.

¹ Laboratory of Ecological and Evolutionary Genetics, Federal University of Viçosa, campus Rio Paranaíba, Rio Paranaíba, Brazil.

² Integrated Nucleus of Biotechnology, University of Mogi das Cruzes, Brazil.

³ Laboratory of Bioinformatics and Genomics, Federal University of Viçosa, campus Rio Paranaíba, Brazil.

⁴ Department of Genetics and Genome Biology, University of Leicester, United Kingdom.

Declaration of Interest: none

Abstract

We are presenting the complete mitogenomes of eight fish species/cytotypes from Neotropical region belonging to the *Astyanax* and *Psalidodon* genus: *A. aeneus*, *A. altiparanae*, *P. fasciatus* (from two locations - Upper Paraná and São Francisco river basins), *A. lacustris*, *P. rivularis* (two cytotypes) and *P. rioparanaibano*. We perform the whole-genome sequencing for six of these species in a Novaseq 6000 - by Illumina, meanwhile two genomes were assembled from raw data available in databases. Plus, we reassembled and annotated the mitochondrial genomes for *A. mexicanus* and *P. paranae*, both already described and with raw data available online. All the genomes presented the same organization, with 13 protein-coding genes, 22 tRNA genes and two rRNA genes. Aiming to contribute to the understanding of the several cryptic species complexes and phylogeny of the genus, we perform Bayesian analysis using the 13 protein-coding genes from these species, plus *Deuterodon giton* and using a *Brycon* species as outgroup.

Key-words: *Astyanax* spp., bayesian, phylogeny, mtDNA, whole-genome sequencing

1. Introduction

Stethaprioninae is a subfamily of characiform fish that comprises small animals popularly known as tetras. *Astyanax* was the first genus in the number of species within the subfamily, being the most diverse in the Neotropical region [1]. Historically, the genus has a wide geographical distribution, through the south USA to the north of Argentina [2], but some *Astyanax* species share several common features that difficult their identification and recognition [3] leading to efforts to identify diagnostic characteristics or molecular signatures for the group. Recently, the genus *Astyanax* underwent an initial taxonomic review and several of its species have been reclassified in new or disused (old) genera [4]. Some of the species formerly identified as belonging to *Astyanax* were reassigned as *Psalidodon* or *Deuterodon*, for example. Terán et al [4] moved some species of *Astyanax* to others

six genera, including the species complexes *A. fasciatus* [5] and *A. scabripinnis* [6] (relocated to *Psalidodon*) and the *Astyanax* species from Coastal river basins of Brazil, like as *A. giton* (relocated to the genus *Deuterodon*), while the species of the complex *A. bimaculatus* [7] and the North American species remain in the *Astyanax* genus [4]. Especially for the tetras from South America, there is a good background of information from chromosomal and phylogenetic data that indicates trends in the evolution of this group [8].

Despite the great diversity and the taxonomical issues, at least one species of the genus (*A. mexicanus*) have been used as a model to understand the development of eyes and the evolution of complex traits [9] [10] [11]. This species was named *A. fasciatus* in the past, with epigeal and hypogeal (troglomorphic) morphotypes. However, currently, *P. fasciatus* is a designation restrict to tetras from Brazilian rivers and probably corresponds to a group of cryptic species with the largest diversity in diploid numbers in *Psalidodon* genus, with $2n=45, 46, 47, 48$ and 49 chromosomes, plus the presence of B chromosomes and heterochromatin polymorphisms in some populations [12] [13] [14], within low molecular diversification [15]. Morphologically, the populations of the São Francisco river basin are different from the specimens from upper Paraná River and Paraíba do Sul [16] and the original morphological description is so broad that it certainly covers other species outside the complex [17]. In this case, *P. fasciatus* should be only the specimens from the São Francisco river basin (original basin of the type species), and the others may be cryptic species of the complex or even other species [17]. On the other side, despite a clear genetic structure between populations from Upper Paraná and São Francisco river basins, morphometric traits seem to be homoplasy [18].

The *A. bimaculatus* complex is currently represented by the homonymous species, along with others such as *A. altiparanae* (considered as a junior synonym of *A. lacustris* by some authors), *A. lacustris*, *A. assuncionensis* and *A. abramis*. With a preference for calm waters, these species habits mainly in the upper Paraná, Paraguay, Iguassu and São Francisco river basins [19]. Contrary to what is observed for in *P. fasciatus* and *P. scabripinnis*, *A. bimaculatus* shows a constant diploidy number in different populations, $2n = 50$ chromosomes, considered a symplesiomorphic character in Gymnocharacini [20] [21] [22]. The diversity in the

group refers to differences in its karyotypic formula, fundamental number and general symmetry of the karyotypes [20] [22]. These cytogenetic data, associated with the molecular ones, suggest relatively recent divergence, as well as the monophyletic status of this branch [8].

The third species complex explored in this work, *P. scabripinnis*, was proposed by Moreira-Filho & Bertollo [6] based on morphological and chromosomal characteristics of specimens collected in the Paraná and São Francisco river basins. In a review of *P. scabripinnis* group, Bertaco & Lucena [23] pointed out the existence of 15 species, including *P. paranae* and *P. rivularis*. The species of this complex are known for their wide karyotypic diversity, with diploid numbers ranging from $2n=46$ to 50 chromosomes [6] [24] [25]. In recent studies using molecular phylogeography and geometric morphometry, Rocha et al. [26] reinforced the validity of *P. rivularis* and *P. paranae*, sister species of the complex inhabiting the São Francisco and Paraná river basins, respectively. However, among the populations from Upper Rio Paranaíba, the existence of a new species of the complex was observed due to morphometric and mtDNA data [27]. This new species, called *Psalidodon rioparanaibano*, was collected only in a small tributary of the Paranaíba river, surrounded by populations of *P. paranae*. Moreover, within *P. paranae* and *P. rivularis* groups, karyotypic diversity is also present [28] (Rodrigues-Oliveira et al., in preparation), indicating that even though delimited by individual lineages, these groups still constitute compilations of cryptic species.

Hundreds of papers describing mitogenomes have been published in the last few years, and despite being the most sequenced genome nowadays [29], to this date, just the mitogenomes of *A. mexicanus* [30], *P. paranae* [31], and *Deuterodon giton* [32], *P. fasciatus* and *A. altiparanae* [33] are published. In an attempt to fill this gap, we present now the complete sequences of the mitochondrial genome of ten species/cytotypes of *Astyanax/Psalidodon* genus, discussing the evolutionary relationships among these species and the presence of a repeat in the control region.

2. Material and Methods

We analyzed specimens belonging to the *Astyanax/Psalidodon* genus. We collected the samples in different locations through the major Brazilian rivers and their vouchers are deposited in the ichthyological collection of the Laboratory of Ecological and Evolutionary Genetics at the Federal University of Viçosa, campus Rio Paranaíba, Brazil (Table 1). After sampling, we brought the living specimens to the laboratory, euthanized them according to the ethical standards of CONCEA - National Council for Control of Animal Experimentation of Brazil and CEUA/UFV - Animal Use Ethics Committee/Federal University of Viçosa (760/2018). We performed the sampling with licenses provided by SISBIO/ICMBIO - Biodiversity Authorization and Information System (1938128) and SISGEN - National System for the Management of Genetic Heritage and Associated Traditional Knowledge (A9FE946).

We extracted the total genomic DNA of six specimens (*Psalidodon rivularis* with 46 chromosomes, *Psalidodon rivularis* with 50 chromosomes, *Psalidodon fasciatus* from São Francisco river basin, *Psalidodon rioparanaibano*, *Astyanax lacustris*, and *Astyanax altiparanae*) from the liver and heart tissue according to the instructions of the Invitrogen's PureLink DNA extraction and purification kit. After quality checking using fluorometer Qubit (Thermo Fisher Scientific), the Whole Genome Sequencing was performed in a Novaseq 6000 (Illumina, San Diego, CA) at Novogene company, UK.

For broader comparisons, we also assembled the mitogenome of two species with raw reads available on ENA (European Nucleotide Archive): *P. fasciatus* from Upper Parana river basin (SRR8476332) and *A. aeneus* from Mexico (SRR1927238). Aiming to validate our methodology we reassembled the mitogenomes of *P. paranae* (SRR5461470) and *A. mexicanus* (SRR2040423). In the Bayesian analysis, we also included the mitochondrial complete sequence of *Deuterodon giton* (NC_044970.1), and *Brycon orbignianus* (KY825192.1) as outgroup.

We assembled the mitogenome from raw reads on Novoplasty v3.7 [34] in a parallel cluster computer (64 Gb RAM) using the mitogenome of *Psalidodon paranae* available on GenBank (SRR5461470) as seed (supplement). We annotated the obtained sequences on MitoAnnotator [35] at MitoFish (<http://mitofish.aori.u-tokyo.ac.jp>). We uploaded the 2x150 raw reads to the Galaxy [36] public platform at usegalaxy.eu, where accessed the quality of raw reads (using FastQC, Babraham Bioinformatics) and filtered with Fastp tool [37].

We perform comparative genomics analysis by BLAST comparison of all *Astyanax/Psalidodon* mitochondrial genomes against a reference (*Psalidodon paranae*) generated by Blast Ring Image Generator (BRIG) [38]. To access the repetitive region we analyze the mitochondrial sequences with Tandem Repeats Finder [39], and aligned on ClustalW [40] to found the repeat motif.

We aligned Fasta sequences with ClustalW [40] and calculated the p-distance with MEGA X software [41]. We used the 13 protein-coding genes (PCGs) in Bayesian phylogenetic inference with MrBayes 3.2.7 [42] after calculation of best evolutionary models for each segment with Partition Finder 2.1.1 [43]. Bayesian analyses were performed using four independent chains with 10-million-generations and posteriorly was verifying the effective sample size (ESS) and strand convergence in Tracer 1.7 software [44]. The first 25% of the generations were discarded as burn-in. For the Maximum Likelihood analysis, we have used the concatenated 13 PCGs after the test of the best model in MEGA X software [41].

3. Results

Our results have shown that all mitogenomes content and gene order were identical (Figure 1), with 13 protein-coding genes (PCGs), 22 tRNA genes and two rRNA genes with the order expected according to already described Characiformes mitogenomes, including to other *Astyanax/Psalidodon/Deuterodon* species as *A. mexicanus* [30], *P. paranae* [31] and *D. giton* [32]. All PCGs, except the *Nd6* gene, are on the heavy chain. All but eight tRNAs are on the heavy chain as well.

The statistics of the sequences are summarized in table 2. The length of mitochondrial sequence range from 16,626 bp in *Psalidodon fasciatus* from São Francisco basin, to 16,812 bp in *Psalidodon rivularis* 2n=50. The average length of D-loop was 1,061 bp. We found a repeat of 35 bp in all D-loops, except in *Deuterodon giton* (Table 3), and with alignment, we got a repeated motif (TATGTATTAGTACATATTATGCATAATTATACATA) slightly variable in some species.

The genetic distance among species (Table 4) is reflected in both the Maximum Likelihood and the Bayesian tree (Figure 2) that shows strong construction with high bootstrap value and posterior probabilities, respectively.

4. Discussion

The topology of the tree is congruent with those inferred for Rossini *et al.* [45] and Pazza *et al.* [8] except for the North American clade, which appear as a sister group of *A. altiparanae* and *A. lacustris* here. Besides, our study reinforces the taxonomic review by Terán *et al.* [4].

In this study, we observed that the populations of *P. fasciatus* belonging to the Paraná and São Francisco river basins had the lowest genetic distances (0.5%, Figure 2). The genetic distance found is similar to that obtained by Pazza *et al.* [18] for the mitochondrial Cytochrome Oxidase I gene (COI; 0.6%), and substantially lower than that of the mitochondrial ATPase 6/8 gene (1.9%). Thus, it is possible that among the analyzed species complexes, *P. fasciatus* is the one with the most recent adaptive radiation history. Combined with the genetic structure found between different basins [18], the high karyotype diversity present in the group [12] [13] [14] [15] reinforces a tendency to maintain different Evolutionarily Significant Units (ESUs) in each population [46].

In a review based on meristic characters, Lucena and Soares [47] consider *A. bimaculatus* and *A. altiparanae* to be a junior synonym of *A. lacustris*. Despite the monophyly of these two species (Figure 2), previously observed within the *A.*

bimaculatus complex [8], the genetic distance between them (1.8%, Table 4) suggests that both units correspond to independent species isolated in different river basins. Additionally, studies of geometric morphometry indicate that the morphological similarity in this group has been shaped by parallel evolution [48], which would explain the difficulty in establishing autapomorphies for each species.

Instead, *A. altiparanae* and *A. lacustris*, which occur widely in different hydrographic basins, the two analyzed cytotypes of *P. rivularis* coexist in a small geographical area (Abaeté sub-basin river). In this scenario, the genetic distance found (1.76%, Table 4) added to the karyotypic differences ($2n = 46$ and $2n = 50$), are strong indications that both cytotypes correspond to cryptic species with distribution limited to the Upper São Francisco river basin. It is possible that speciation and the evolution in *P. rivularis* was, or is being driven mainly by chromosomal diversification, with the subsequent genetic divergence, as observed in *P. fasciatus* [15].

Our study also reaffirmed the validity of *P. rioparanaibano* as an evolutionary significant unity (ESU) distinguishable from *P. paranae* [27], although both are found in the Paranaíba river basin. Besides the great genetic distance observed between *P. paranae* and *P. rioparanaibano* (2.3%, Table 4), it was possible to observe that both cytotypes of *P. rivularis*, inhabitants of the São Francisco River, are closer to *P. paranae* than *P. rioparanaibano* (Figure 2).

Deepening the knowledge of the mitogenome control region, called D-loop, can play a fundamental role in understanding the evolutionary history of the *Astyanax* and *Psalidodon* genera. In this work, we observed that the size variation between different *Astyanax* / *Psalidodon* mitogenomes occurs mainly due to the extension of the D-loop. Neglecting this region in the reconstruction of mitogenomes can result in a valuable loss of information, since in addition to the variation in size, we found a repetitive sequence of 35bp in 9 of the 10 mitogenomes studied (table 3).

D. giton was the only species that did not present the repetitive sequence in the D-loop. Plus to occupying a sister group position in the recuperated phylogeny (Figure 2) when describing the mitogenome of the species, Barreto et al [32]

observed in their phylogeny of Characidae that the species *Grundulus bogotensis* was closer to the *Astyanax/Psalidodon* group than *D. giton*. Therefore, we not only confirm *D. giton* as a species outside the genus *Astyanax* as suggested by the taxonomic review [4], as the repetitive sequence found in the D-loop may correspond to a synapomorphy absent in the *Deuterodon* group. To confirm this it would be necessary to reassemble the genome of *D. giton* under our methodology and complementary studies need to be done on species of the *Deuterodon* group to clarify this issue.

The methodology used in the reconstruction of the mitochondrial genome proved to be satisfactory and able to access the length of this type of genome, plus the composition and nature of the D-loop, solving possible gaps in previous methodologies [31] [32] [33]. Besides, the study of the complete mitochondrial genome proves to be a tool with the potential to solve taxonomic problems and to help to understand the evolutionary relationships in species complexes, such as *A. bimaculatus*, *P. fasciatus* and *P. scabripinnis*.

Declaration of competing interest:

The authors state that there are no conflicts to declare.

Acknowledgments

Authors thank Henrique Peluzio for his support on the UFV cluster.

The Galaxy server that was used for some calculations is in part funded by Collaborative Research Centre 992 Medical Epigenetics (DFG grant SFB 992/1 2012) and German Federal Ministry of Education and Research (BMBF grants 031 A538A/A538C RBC, 031L0101B/031L0101C de.NBI-epi, 031L0106 de.STAIR (de.NBI)).

References

- [1] M.L. Orsi, E.D. Carvalho, F. Foresti, Biologia populacional de *Astyanax altiparanae* Garutti & Brutti & Britski (Teleostei, Characidae) do médio Rio Paranapanema, Paraná. Brasil. Rev. Bras. Zool. 21.2 (2004) 207-218. <https://doi.org/10.1590/S0101-81752004000200008>.
- [2] C.P Ornelas-Garcia, O. Dominguez-Dominguez, I. Doadrio, Evolutionary history of the fish genus *Astyanax* Baird & Girard (1854) (Actinopterygii, Characidae) in Mesoamerica reveals multiple morphological homoplasies. BMC Evol. Biol. 8.1 (2008) 340. <https://doi.org/10.1186/1471-2148-8-340>.
- [3] S.H. Weitzman, L. Malabarba, Perspectives about the phylogeny and classification of the Characidae (Teleostei: Characiformes), in: L.R. Malabarba, R.E. Reis, R.P. Vari, Z.M.S. Lucena, C.A.S. Lucena (Eds.), Phylogeny and classification of Neotropical fishes, Edipucrs, Porto Alegre, 1998, pp. 161-170.
- [4] G.E. Terán, M.F. Benitez, J.M. Mirande, Opening the Trojan horse: phylogeny of *Astyanax*, two new genera and resurrection of *Psalidodon* (Teleostei: Characidae). Zool. J. Linnean Soc. zlaa019 (2020). <https://doi.org/10.1093/zoolinnean/zlaa019>.
- [5] A.J. Justi, Caracterização cariotípica de populações de *Astyanax fasciatus* (Cuvier, 1819) Pisces, Characidae, em três bacias hidrográficas, Master's Thesis, Universidade Federal de São Carlos, São Carlos (Unpublished results).
- [6] O. Moreira-Filho, L.A.C. Bertollo, *Astyanax scabripinnis* (Pisces, Characidae): a species complex. Brazil. J. Genet. 14 (1991) 331-357.
- [7] V. Garutti, Revisão taxonômica dos *Astyanax* (Pisces, Characidae), com mancha umeral ovalada e mancha no pedúnculo caudal, estendendo-se à extremidade dos raios caudais medianos, das bacias do Paraná, São Francisco e Amazônica, Livre-Docência Thesis, Universidade Estadual Paulista Júlio de Mesquita Filho, São José do Rio Preto (Unpublished results).
- [8] R. Pazza, J.A. Dergam, K.F. Kavalco, Trends in Karyotype Evolution in *Astyanax* (Teleostei, Characiformes, Characidae): Insights From Molecular Data. Front. Genet. 9 (2018) 131. <https://doi.org/10.3389/fgene.2018.00131>.

- [9] W.R. Jeffery, Cavefish as a model system in evolutionary developmental biology. *Dev. Biol.* 231.1 (2001) 1-12. <https://doi.org/10.1006/dbio.2000.0121>.
- [10] W.R. Jeffery, Emerging model systems in evo-devo: cavefish and microevolution of development. *Evol. Dev.* 10.3 (2008) 265-272. <https://doi.org/10.1111/j.1525-142X.2008.00235.x>.
- [11] R. Borowsky, *Astyanax mexicanus*, the blind Mexican cave fish: a model for studies in development and morphology. *Cold Spring Harb. Protoc.* 2008.11 (2008) pdb-emo107. <https://doi.org/10.1101/pdb.emo107>.
- [12] R. Pazza, K.F. Kavalco, L.A.C. Bertollo, Chromosome polymorphism in *Astyanax fasciatus* (Teleostei, Characidae). 1. Karyotype analysis, Ag-NORs and mapping of the 18S and 5S ribosomal genes in sympatric karyotypes and their possible hybrid forms. *Cytogenet. Genome Res.* 112.3-4 (2006) 313-319. <https://doi.org/10.1159/000089886>.
- [13] M. Ferreira-Neto, R.F. Artoni, M.R. Vicari, O. Moreira-Filho, J.P. Camacho, M. Bakkali, C. de Oliveira, F. Foresti, Three sympatric karyomorphs in the fish *Astyanax fasciatus* (Teleostei, Characidae) do not seem to hybridize in natural populations. *Comp. Cytogenet.* 6.1 (2012) 29–40. <https://doi.org/10.3897/CompCytogen.v6i1.2151>.
- [14] K.F. Kavalco, R. Pazza, K. de Oliveira Brandão, L.F. de Almeida-Toledo, Biogeographic patterns in the chromosomal distribution of a satellite DNA in the banded tetra *Astyanax fasciatus* (Teleostei: Characiformes). *Org. Divers. Evol.* 13 (2013) 67–76. <https://doi.org/10.1007/s13127-012-0100-8>.
- [15] K.F. Kavalco, R. Pazza, K.D.O. Brandao, C. Garcia, L.A.C. Bertollo, L.F. de Almeida-Toledo, Chromosomal diversification higher than molecular variation in *Astyanax aff. fasciatus* (Teleostei, Characidae). *Zebrafish.* 13.4 (2016) 345-353. <https://doi.org/10.1089/zeb.2016.1272>.

- [16] F.A.G. Melo, Revisão taxonômica das espécies do gênero *Astyanax* Baird e Girard 1854 (Teleostei, Characiformes, Characidae) da região da Serra dos Órgãos. Arq. Mus. Nac. Rio de Janeiro. 59 (2001)1–46.
- [17] F.A.G. Melo, P.A. Buckup, *Astyanax henseli*, a new name for *Tetragonopterus aeneus* Hensel (1870), from Southern Brazil (Teleostei, Characiformes). Neotrop. Ichthyol. 4 (2006) 45–52. <http://dx.doi.org/10.1590/S1679-62252006000100003>.
- [18] R. Pazza, L.M. Cruvinel, K.F. Kavalco, Parallel evolution evidenced by molecular data in the banded-tetra (*Astyanax fasciatus*). Biochem. Syst. 70 (2017) 141-146. <https://doi.org/10.1016/j.bse.2016.10.024>.
- [19] M. de S. Domingues, M.R. Vicari, V., Abilhoa, J.P. Wamser, M.M. Cestari, L.A.C. Bertollo, M.C. de Almeida, R.F. Artoni, Cytogenetic and comparative morphology of two allopatric populations of *Astyanax altiparanae* Garutti & Britski, 2000 (Teleostei: Characidae) from upper rio Paraná basin. Neotrop. Ichthyol. 5.1 (2007) 37-44. <https://dx.doi.org/10.1590/S1679-62252007000100005>.
- [20] K.F. Kavalco, R. Pazza, K.D.O. Brandao, C. Garcia, L.F.D. Almeida-Toledo, Comparative cytogenetics and molecular phylogeography in the group *Astyanax altiparanae*–*Astyanax* aff. *bimaculatus* (Teleostei, Characidae). Cytogenet. Genome Res. 134.2 (2011) 108-119. <https://doi.org/10.1159/000325539>.
- [21] E.R. Martinez, A.L. Alves, S.M. Silveira, F. Foresti, C. Oliveira, Cytogenetic analysis in the incertae sedis species *Astyanax altiparanae* Garutti and Britzki, 2000 and *Hyphessobrycon eques* Steindachner, 1882 (Characiformes, Characidae) from the upper Paraná river basin. Comp. Cytogenet. 6.1 (2012) 41–51. <https://doi.org/10.3897/CompCytogen.v6i1.1873>.
- [22] C.A. Fernandes, R.H. da Rocha, D. Bailly, Z.D.R. Guterres, D.S. Alves, I.C. Martins-Santos, Cytogenetic and Morphometric Analysis in the Species *Astyanax altiparanae* Garutti & Britski, 2000 (Teleostei, Characidae) from the Iguatemi River Basin, Brazil. Folia Biol. 62.3 (2014) 193-201. https://doi.org/10.3409/fb62_3.193.

[23] V.A. Bertaco, C.A. Lucena, Two new species of *Astyanax* (Ostariophysi: Characiformes: Characidae) from eastern Brazil, with a synopsis of the *Astyanax scabripinnis* species complex. Neotrop Ichthyol. 4.1 (2006) 53-60.

<https://doi.org/10.1590/S1679-62252006000100004>.

[24] C.A. Fernandes, I.C. Martins-Santos, Sympatric occurrence of three cytotypes and four morphological types of B chromosomes of *Astyanax scabripinnis* (Pisces, Characiformes) in the River Ivaí Basin, state of Paraná, Brazil. Genetica. 124 (2005) 301-306. <https://doi.org/10.1007/s10709-005-4751-7>.

[25] J.P. Klassmann, I.C.M. dos Santos, Karyotype Diversity in *Astyanax scabripinnis* (Pisces, Characiformes) from the São Francisco River Basin, Brazil. Cytologia. 82.2 (2017) 199-203. <https://doi.org/10.1508/cytologia.82.199>.

[26] R.R. Rocha, R. de M. Alves, R. Pasa, K.F. Kavalco, Morphological and genetic structure of two equivalent *Astyanax* species (Characiformes: Characidae) in the region of Paranaíba Arc. Sci. World. J. 2019. (2019).

<https://doi.org/10.1155/2019/6507954>.

[27] R. de M. Alves, I.H. Rodrigues-Oliveira, R. Pasa, K.F. Kavalco, A new *Astyanax* species related to the *Astyanax paranae* complex (Characiformes, Characidae) from Upper Paranaíba river basin, Brazil, supported by genetic and morphometric data. Biodivers. J. (in press).

[28] E.L. Maistro, C. Oliveira, F. Foresti, Comparative cytogenetic and morphological analysis of *Astyanax scabripinnis paranae* (Pisces, Characidae, Tetragonopterinae). Genet. Mo. Biol. 21.2 (1998) 201-206.

<https://doi.org/10.1590/S1415-47571998000200005>.

[29] D.R. Smith, "The past, present and future of mitochondrial genomics: have we sequenced enough mtDNAs?". Brief. Funct Genomics 15.1 (2015) 47-54.

<https://doi.org/10.1093/bfgp/elv027>.

[30] M. Nakatani, M. Miya, K. Mabuchi, K. Saitoh, M. Nishida, Evolutionary history of Otophysi (Teleostei), a major clade of the modern freshwater fishes: Pangaeon

origin and Mesozoic radiation. *BMC Evol. Biol.* 11 (2011) 177.
<https://doi.org/10.1186/1471-2148-11-177>.

[31] D.M.D.A. Silva, R. Utsunomia, F.J. Ruiz-Ruano, C. Oliveira, F. Foresti, The complete mitochondrial genome of *Astyanax paranae* (Teleostei: characiformes). *Mitochondrial DNA B.* 1.1 (2016) 586-587.
<https://doi.org/10.1080/23802359.2016.1222251>.

[32] C.A.V. Barreto, M.M.C. Granja, P.M.P. Vidigal, A.O. Carmo, J.A. Dergam, Complete mitochondrial genome sequence of neotropical fish *Astyanax giton* Eigenmann 1908 (Ostariophysi; Characidae). *Mitochondrial DNA B.* 2.2 (2017) 839-840. <https://doi.org/10.1080/23802359.2017.1403869>.

[33] R.M. Calegari, P.H.M. Rodrigues, R.Z.D. Santos, F. Foresti, R. Utsunomia, F. Porto-Foresti, The complete mitochondrial genome sequences of five Otophysi species (Vertebrata, Teleostei). *Mitochondrial DNA B.* 4.2 (2019) 4198-4199.
<https://doi.org/10.1080/23802359.2019.1693294>.

[34] N. Dierckxsens, P. Mardulyn, G. Smits, NOVOPlasty: de novo assembly of organelle genomes from whole genome data. *Nucleic Acids Res.* 45.4 (2017) e18.
<https://doi.org/10.1093/nar/gkw955>.

[35] W. Iwasaki, T. Fukunaga, R. Isagozawa, K. Yamada, Y. Maeda, T.P. Satoh, T. Sado, K. Mabuchi, H. Takeshima, M. Miya, M. Nishida, MitoFish and MitoAnnotator: a mitochondrial genome database of fish with an accurate and automatic annotation pipeline. *Mol. Biol Evol.* 30.11 (2013) 2531-2540.
<https://doi.org/10.1093/molbev/mst141>.

[36] E. Afgan, D. Baker, B. Batut, M. Van Den Beek, D. Bouvier, M. Čech, J. Chilton, D. Clements, N. Coraor, B.A. Grüning, A. Guerler, J. Hilman-Jackson, S. Hiltemann, V. Jalid, H. Rasche, N. Soranzo, J. Goecks, J. Taylor, A. Nekrutenko, D. Blankenberg, The Galaxy platform for accessible, reproducible and collaborative biomedical analyses: 2018 update. *Nucleic Acids Res.* 46.1 (2018) 537-544.
<https://doi.org/10.1093/nar/gky379>.

[37] S. Chen, Y. Zhou, Y. Chen, J. Gu, fastp: an ultra-fast all-in-one FASTQ preprocessor. *Bioinformatics*. 34.17 (2018) i884–i890.

<https://doi.org/10.1093/bioinformatics/bty560>.

[38] N.F. Alikhan, N.K. Petty, N.L.B. Zakour, S.A. Beatson, BLAST Ring Image Generator (BRIG): simple prokaryote genome comparisons. *BMC Genomics*. 12 (2011) 402. <https://doi.org/10.1186/1471-2164-12-402>.

[39] G. Benson, Tandem repeats finder: a program to analyze DNA sequences. *Nucleic Acid Res.* 27.2 (1999) 573-580. <https://doi.org/10.1093/nar/27.2.573>.

[40] J.D. Thompson, D.G. Higgins, T.J. Gibson, CLUSTALW: Improving the sensitivity of progressive multiple sequence through weighing, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res.* 22 (1994) 4673-4680.

[41] S. Kumar, G. Stecher, M. Li, C. Knyaz, K. Tamura, MEGA X: molecular evolutionary genetics analysis across computing platforms. *Mol. Biol. Evol.* 35.6 (2018) 1547-1549. <https://doi.org/10.1093/molbev/msy096>.

[42] F. Roquist, M. Teslenko, P. van der Mark, D.L. Ayres, A. Darling, S. Höhna, B. Larget, L. Liu, M.A. Suchard, J.P. Huelsenbeck, MrBayes 3.2: efficient Bayesian phylogenetic inference and model choice across a large model space. *Syst. Biol.* 61 (2012) 539-542. <https://doi.org/10.1093/sysbio/sys029>.

[43] R. Lanfear, P.B. Frandsen, A.P. Wright, T. Senfeld, B. Calcott B, PartitionFinder 2: new methods for selecting partitioned models of evolution for molecular and morphological phylogenetic analyses. *Mol. Biol. Evol.* 34 (2016) 772-773. <https://doi.org/10.1093/molbev/msw260>.

[44] A. Rambaut, A.J. Drummond, D. Xie, G. Baele, M.A. Suchard, Posterior summarization in Bayesian Phylogenetics using Tracer 1.7. *Syst. Biol.* 67 (2018) 901-904. <https://doi.org/10.1093/sysbio/syy032>.

[45] B.C. Rossini, C.A. Oliveira, F.A. Melo, V.A. Bertaco, J.M. Astarloa, J.J. Rosso, F. Foresti, C. Oliveira, Highlighting *Astyanax* Species Diversity through DNA

Barcoding. PloS one. 11.12 (2016) e0167203.

<https://doi.org/10.1371/journal.pone.0167203>.

[46] O.A. Ryder, Species conservation and systematics: the dilemma of subspecies. Trends Ecol. Evol. 1 (1986) 1-9. [https://doi.org/10.1016/0169-5347\(86\)90059-5](https://doi.org/10.1016/0169-5347(86)90059-5).

[47] C.A.S. de Lucena, H.G. Soares, Review of species of the *Astyanax bimaculatus* "caudal peduncle spot" subgroup sensu Garutti & Langeani (Characiformes, Characidae) from the rio La Plata and rio São Francisco drainages and coastal systems of southern Brazil and Uruguay. Zootaxa. 4072.1 (2016) 101-125. <http://dx.doi.org/10.11646/zootaxa.4072.1.5>.

[48] M.A. da Silva, G.X. Perazzo, K.F. Kavalco, R. Pasa, Shape diversity of the fish genus *Astyanax* Baird & Girard, 1854 (Teleostei, Characidae) in adjacent basins. Biologia. (2020) 1-9. <https://doi.org/10.2478/s11756-020-00544-5>.

Table 1 Ecological data from the collected specimens for this study.

Species	Sequence	Sex	Country/basin	Lat/Lon	Altitude	Specimen voucher	Tissue type
<i>Astyanax altiparanae</i>	Mitochondrial	Male	Brazil: Paranaíba river basin	-19.3517611111111 -46.7930027777778	843 m	LAGEEVO - 4121	liver and heart
<i>Astyanax lacustris</i>	Mitochondrial	Male	Brazil: Abaeté river basin	-20.1231722222222 -45.9708916666667	708 m	LAGEEVO - 3732	liver and heart
<i>Psalidodon fasciatus</i>	Mitochondrial	Female	Brazil: Bambuí river basin	-19.2099138888889 -46.1093861111111	1013 m	LAGEEVO - 2096	liver and heart
<i>Psalidodon rioparanaibano</i>	Mitochondrial	Male	Brazil: Paranaíba river basin	-19.1877138888889 -46.2361777777778	1074 m	LAGEEVO - 4285	liver and heart
<i>Psalidodon rivularis</i> 2n = 46	Mitochondrial	Male	Brazil: Abaeté river basin	-18.9428 -45.9383888888889	889 m	LAGEEVO - 2030	liver and heart
<i>Psalidodon rivularis</i> 2n = 50	Mitochondrial	Male	Brazil: Abaeté river basin	-19.2099138888888 -46.1102194444444	1017 m	LAGEEVO - 2614	liver and heart

Table 2 General statistics of the sequences analyzed. The abbreviations PR and SF after *P. fasciatus*, indicate Paranaíba river basin and São Francisco river basin respectively. The mitogenome that was not reassembled following our methodology is indicated by * .

Species	Size (bp)	% GC	A	T	C	G	N	D-loop (bp)	Bankit	Genbank
<i>Astyanax aeneus</i>	16769	41.1	5038	4832	4318	2581	0	1090	BankIt2340537	BK013055
<i>Astyanax altiparanae</i>	16730	41.7	4952	4795	4397	2586	0	1066	BankIt2341301	MT428072
<i>Astyanax lacustris</i>	16763	41.8	4959	4797	4415	2592	0	1099	BankIt2340713	MT428067
<i>Astyanax mexicanus</i>	16768	41	5046	4848	4304	2566	0	1091	BankIt2341953	BK013062
<i>Deuterodon giton</i> *	16643	40.7	4950	4911	4182	2600	0	976	-	NC_044970.1
<i>Psalidodon fasciatus PR</i>	16729	42.5	4951	4663	4466	2647	0	1054	BankIt2340825	BK013061
<i>Psalidodon fasciatus SF</i>	16626	42.7	4916	4615	4460	2635	0	951	BankIt2340824	MT428071
<i>Psalidodon paranae</i>	16743	42.8	4946	4626	4526	2643	0	1067	-	SRR5461470
<i>Psalidodon rioparanaibano</i>	16714	42.9	4932	4607	4526	2648	0	1039	BankIt2340806	MT428068
<i>Psalidodon rivularis 2n=46</i>	16779	42.9	4952	4622	4541	2663	0	1102	BankIt2340815	MT428069
<i>Psalidodon rivularis 2n=50</i>	16812	42.8	4972	4642	4536	2661	0	1136	BankIt2340820	MT428070
Average	16734.2	41	4964.9	4723.5	4424.6	2620.2	0	1061		

Table 3 Summary of the properties from the repeated sequence found in the D-loop region. The abbreviations PR and SF after *P. fasciatus*, indicate Paranaiba river basin and São Francisco river basin respectively. The mitogenome that was not reassembled following our methodology is indicated by *.

Species	Position	Period Size	Copy Number	Consensus Size	Percent Matches	Percent Indels	Score	A	C	G	T	Entropy (0-2)	Consensus Repeat Sequence
<i>Astyanax aeneus</i>	15801-15922	35	3.5	35	100	0	244	42	5	9	42	1.60	TATGTATTAGTACATAA TATGTAATATATTACAT A
<i>Astyanax altiparanae</i>	15780-15918	35	4.0	35	89	1	199	43	8	7	41	1.63	TATGTACTAGTACATAA TATGTATAATTATACAT A
<i>Astyanax lacustris</i>	15780-15951	35	5.0	35	79	7	181	40	8	11	40	1.70	TATGTATTAGTACATAA TATGTATAATTATACAT A
<i>Astyanax mexicanus</i>	15793-15925	35	3.8	35	98	1	257	42	6	9	42	1.63	TACATATATGTATTAGT ACATAATATGTAATATA T
<i>Deuterodon giton*</i>	no repeats												
<i>Psalidodon fasciatus PR</i>	15728-15907	35	5.1	35	98	0	342	40	8	8	42	1.66	ACATATATGTATTAGTA CATATTATGCATAATTA T
<i>Psalidodon fasciatus SF</i>	15734-15803	35	2.0	35	100	0	140	40	8	8	42	1.66	TATGTATTAGTACATAT TATGCATAATTATACAT A
<i>Psalidodon paranae</i>	15788-15913	35	3.6	35	98	0	243	39	8	9	42	1.68	ATGCATAATTATACATA TATGTATTAGTACATAT T
<i>Psalidodon rioparanaibano</i>	15797-15876	35	2.3	35	100	0	160	41	8	7	42	1.64	AATTATACATATATGTA TTAGTACATATTATGCA T
<i>Psalidodon rivularis 2n=46</i>	15788-15945	35	4.5	35	98	0	298	39	8	9	42	1.69	TATGCATAATTATACAT ATATGTATTAGTACATA T
<i>Psalidodon rivularis 2n=50</i>	15795-15980	35	5.3	35	97	0	336	40	9	8	41	1.67	TAATTATACATATATGT ATTAGTACATATTATGC A

Table 4 Genetic distances. The abbreviations PR and SF after *P. fasciatus*, indicate Paranaíba river basin and São Francisco river basin respectively.

	<i>Astyan aeneus</i>	<i>Astyanax altiparanae</i>	<i>Astyanax lacustris</i>	<i>Astyanax mexicanus</i>	<i>Deuterodon giton</i>	<i>Psalidodon fasciatus PR</i>	<i>Psalidodon fasciatus SF</i>	<i>Psalidodon paranae</i>	<i>Psalidodon rioparanaibano</i>	<i>Psalidodon rivularis 2n = 50</i>	<i>Psalidodon rivularis 2n = 46</i>
<i>Astyanax altiparanae</i>	0.15159										
<i>Astyanax lacustris</i>	0.15168	0.01802									
<i>Astyanax mexicanus</i>	0.02153	0.15208	0.15278								
<i>Deuterodon giton</i>	0.20154	0.20706	0.20697	0.20196							
<i>Psalidodon fasciatus PR</i>	0.15402	0.16102	0.16128	0.15425	0.19751						
<i>Psalidodon fasciatus SF</i>	0.15413	0.16078	0.16104	0.1541	0.19716	0.00516					
<i>Psalidodon paranae</i>	0.14967	0.15824	0.15911	0.15016	0.1983	0.03719	0.03761				
<i>Psalidodon rioparanaibano</i>	0.15107	0.15833	0.1585	0.15182	0.19725	0.03938	0.0398	0.02344			
<i>Psalidodon rivularis 2n = 50</i>	0.15089	0.1592	0.15938	0.15252	0.19795	0.03763	0.03796	0.02012	0.02161		
<i>Psalidodon rivularis 2n = 46</i>	0.15072	0.15911	0.15955	0.15217	0.19883	0.03816	0.0384	0.02064	0.02318	0.01758	
<i>Brycon orbignyanus</i>	0.27482	0.27684	0.27605	0.27579	0.27149	0.26684	0.26725	0.26566	0.26654	0.26663	0.26681

Figures

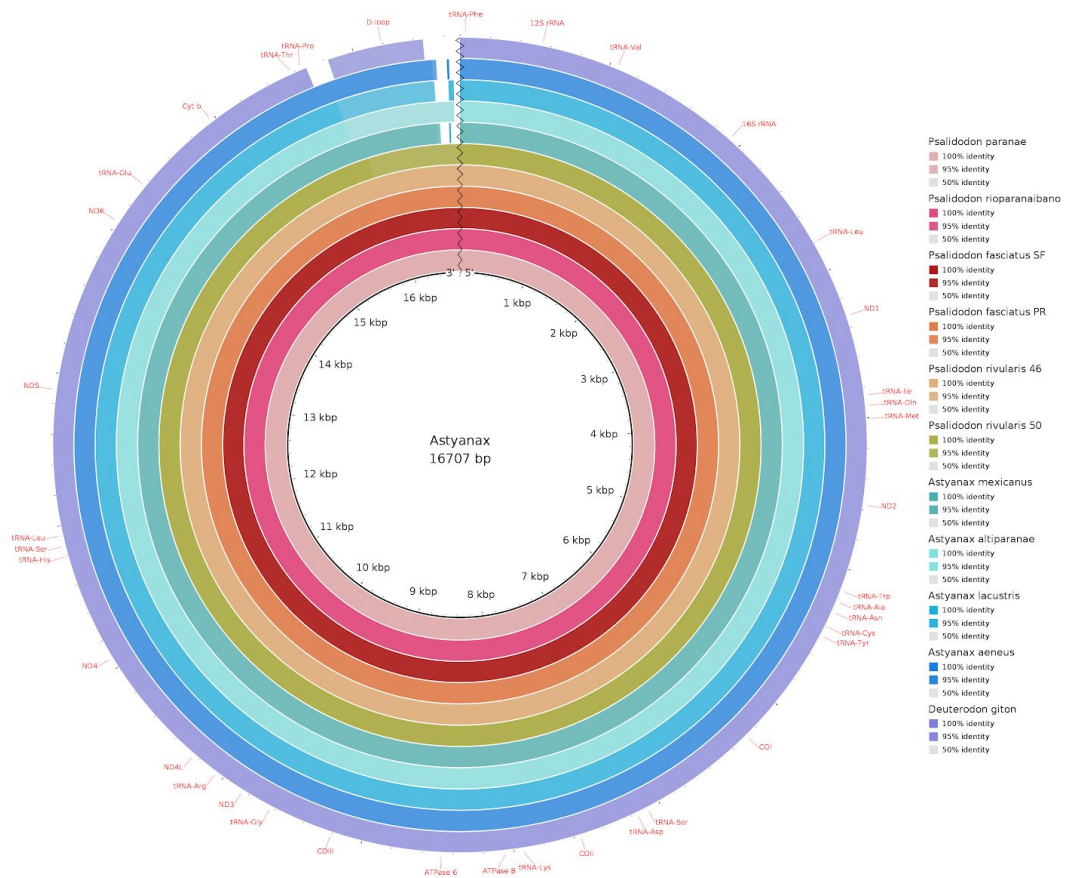


Figure 1: Comparative genomics analysis of all 10 Stethaprioninae fish. BLAST comparison of all mitochondrial genomes against a reference (*Astyanax paranae*) generated by Blast Ring Image Generator (BRIG). Gaps in rings correspond to regions with less than 50% identity to the reference sequence.

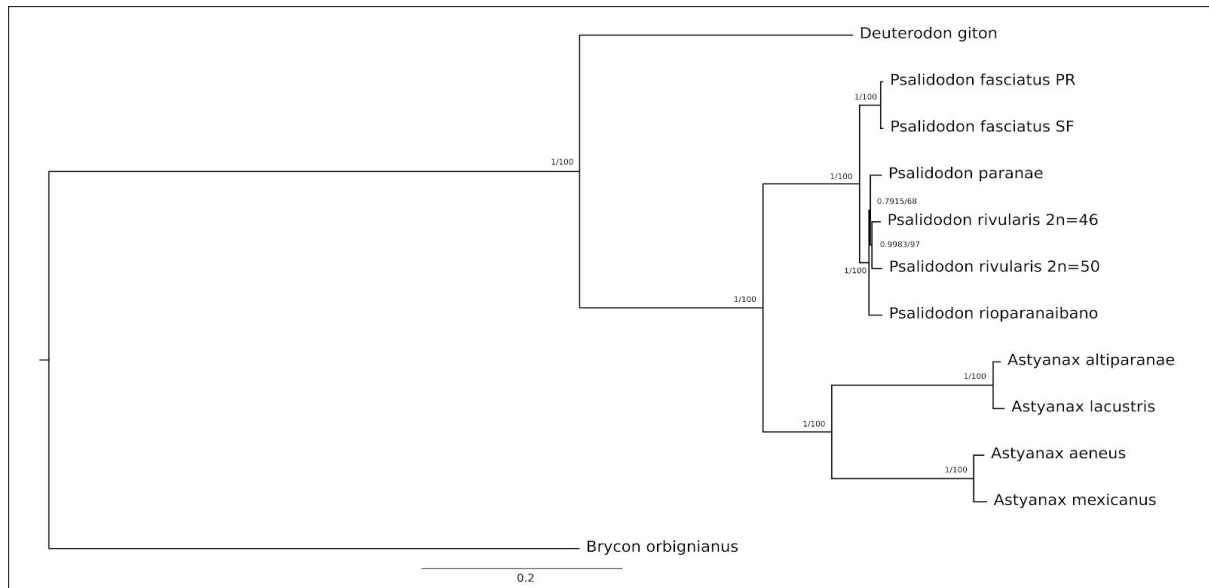


Figure 2: Phylogenetic tree based on 13 protein-coding genes (PCGs) showing the relationships among the Stethaprioninae fish using *Brycon orbignianus* as outgroup. The topology was the same in Maximum Likelihood (100 bootstrap replicates) after the test of the best model (General Time Reversible +G) and Bayesian Inference after calculation of best evolutionary models for each segment, using four independent chains with 10-million-generations (The first 25% of the generations were discarded as burn-in). The posterior probabilities/bootstrap are on the branches.