

# DNA barcoding highlights taxonomic uncertainties and cryptic lineages of São Francisco River basin fishes



<sup>1</sup>Denis Bruno Santos Marques Nunes<sup>1</sup>, <sup>1</sup>Alany Itala Pontes<sup>1</sup>,  
<sup>1</sup>Leandro Ferreira dos Santos<sup>1</sup>, <sup>2</sup>Leonardo Luís Calado<sup>2</sup>,  
<sup>1</sup>Luciano Freitas Barros-Neto<sup>3</sup>, <sup>1</sup>Silvia Yasmin Lustosa-Costa<sup>3</sup>,  
<sup>1</sup>Sergio Maia Queiroz Lima<sup>3</sup> and <sup>1</sup>Uedson Pereira Jacobina<sup>1</sup>

Correspondence:  
Uedson Pereira Jacobina  
uedson.jacobina@penedo.ufal.br

Submitted June 28, 2024

Accepted October 11, 2024

by Claudio Oliveira

Epub January 10, 2025

The São Francisco River (SFR) is a watershed of extreme socioeconomic importance and rich biodiversity, notable for its unique ichthyofauna and high levels of endemism. However, human interference threatens several fish species, putting them at risk of extinction. Genetic knowledge about the ichthyofauna of the SFR is still lacking. To address this gap, we used DNA barcoding of 94 taxa from six orders, 32 families, and 73 genera, sourced from the BOLD Systems platform and samples in the middle and lower SFR regions. Analyses revealed distinct patterns of genetic divergence. The BOLD platform identified 111 bins, 75 as MATCH, 32 as SPLIT, two as MERGE and two MIXTURE. We detected taxonomic confusions in the genera *Astyanax* and *Psalidodon*. Fourteen taxa exhibited high cryptic diversity, corroborated by lineage delimitation methods (BINs, ASAP, and GMYC). A significant portion of this cryptic diversity (eight lineages) is concentrated in the upper SFR, a region with tectonic faults, geomorphological reactivations, and the transposition of another basin, which likely contributed to genetically divergent lineages. Our data fill shortfalls in Linnaean and Darwinian knowledge about the SFR ichthyofauna, considering its hydrogeological formation. We hope this information will serve as a foundation for future management and conservation plans for this important ichthyofauna.

**Keywords:** Genetic divergence, Molecular delimitation of lineages, Darwinian shortfall, Linnaean shortfall, Species complex.

Online version ISSN 1982-0224

Print version ISSN 1679-6225

Neotrop. Ichthyol.  
vol. 22, no. 4, Maringá 2024

1 Programa de Pós-Graduação em Diversidade Biológica e Conservação nos Trópicos, Laboratório de Sistemática Molecular Integrativa e Análise de Microssistemas, Campus Arapiraca, Universidade Federal de Alagoas (UFAL), Av. Manoel Severino Barbosa, Bom Sucesso, 57309-005 Arapiraca, AL, Brazil. (DBSMN) denis.nunes@icbs.ufal.br, (AIP) alanypontes@gmail.com, (LFS) leandrynho.16@gmail.com, (UPJ) uedson.jacobina@penedo.ufal.br (corresponding author).

2 Instituto Nacional do Semiárido (Governo Federal do Brasil), Av. Francisco Lopes de Almeida, Serrotão, 58434-700 Campina Grande, PB, Brazil. (LLC) leonardocalado@hotmail.com.

3 Laboratório de Sistemática e Ictiologia Evolutiva, Departamento de Botânica e Zoologia, Campus Lagoa Nova, Universidade Federal do Rio Grande do Norte (UFRN), Lagoa Nova, 59078-970 Natal, RN, Brazil. (LFBN) cdp\_luciano@hotmail.com, (SYLC) silviayasmincosta@gmail.com, (SMQL) sergio.lima@ufrn.br.

O rio São Francisco (SFR) é uma bacia hidrográfica de extrema importância socioeconômica e ampla biodiversidade, destacando-se por sua ictiofauna singular e altos níveis de endemismo. Contudo, a interferência humana ameaça várias espécies de peixes, colocando-as em risco de extinção. O conhecimento genético sobre a ictiofauna do SFR ainda é deficitário. Para preencher essa lacuna, utilizamos códigos de barras de DNA em 94 táxons de seis ordens, 32 famílias e 73 gêneros, provenientes da plataforma BOLD Systems e de coletas nas regiões do submédio e baixo SFR. As análises revelaram padrões distintos de divergência genética. A plataforma BOLD identificou 111 bins, 75 como MATCH, 32 SPLITS, dois MERGE e dois MIXTURE. Detectamos confusões taxonômicas nos gêneros *Astyanax* e *Psalidodon*. Quatorze táxons apresentaram alta diversidade críptica, corroborada por métodos de delimitação de linhagens (BINs, ASAP e GMYC). Uma parte significativa dessa diversidade críptica (oito linhagens) está concentrada na parte alta do SFR, uma região com falhas tectônicas, reativações geomorfológicas e a transposição de outra bacia, que provavelmente pode ter contribuído para linhagens geneticamente divergentes. Nossos dados preenchem lacunas de conhecimentos Lineano e Darwiniano sobre a ictiofauna do SFR, considerando sua formação hidrogeológica. Esperamos que essas informações sirvam de base para futuros planos de manejo e conservação dessa importante ictiofauna.

**Palavras-chave:** Complexo de espécies, Delimitação molecular de linhagens, Divergência genética, Lacuna Darwiniana, Lacuna Lineana.

## INTRODUCTION

The Neotropical ichthyofauna comprises approximately 20 to 25% of all freshwater fish on Earth (Dagosta, de Pinna, 2019), with an estimated richness of about 5,160 described species distributed across 739 genera, 69 families, and 20 orders (Reis *et al.*, 2016). However, a significant portion of this diversity remains unknown, with an estimated 3,000 to 4,000 species yet to be described (Reis *et al.*, 2016).

The lack of knowledge about species is influenced by the taxonomic impediment, which includes a shortage of scientific-human resources, such as taxonomists, inadequate infrastructure, and the need for taxonomic revision, especially in diverse and complex groups (Jörger, Schrödl, 2013). This is manifested in several deficits, including the Linnean shortfall, which points to the absence of knowledge about species identification, leading to many remaining undescribed, and the Darwinian shortfall, indicating a lack of knowledge about species phylogeny and evolution (Bini *et al.*, 2006; Diniz-Filho *et al.*, 2013; Hortal *et al.*, 2015).

Additionally, the lack of knowledge about the ichthyofauna in the Neotropical region has contributed to the absence of effective conservation management. This knowledge gap is particularly concerning in the context of the rapid advancement of deforestation and the pollution of rivers, streams, and other bodies of water, which have significantly impacted natural fish populations (Ilha *et al.*, 2019; Freitas *et al.*, 2021). Globally, human activity is rapidly increasing, and it is estimated that more than one-third of all freshwater fish species are at risk of extinction (Harrison *et al.*, 2021).

In Brazil, the São Francisco River basin (SFR) holds significant socioeconomic importance. It is recognized as the national integration basin, spanning approximately 2,700 km in length and covering an area of ~640,000 km<sup>2</sup> (Knoppers *et al.*, 2005). This basin traverses the Cerrado, Caatinga, and the Atlantic Forest, divided into upper, middle, sub-middle, and lower sections based on geomorphological, hydrological, and ecological variations along the river's course (Knoppers *et al.*, 2005; Santos *et al.*, 2014; O'Hanley *et al.*, 2020). The water resources of the SFR are utilized in various human activities, such as energy generation, irrigation, urban and industrial water supply, navigation, agriculture, and aquaculture (Santos *et al.*, 2014; O'Hanley *et al.*, 2020). The floodplain areas of the SFR represent one of the most crucial fishing environments in Brazil's inland and are considered nurseries for several fish species (Sato, Godinho, 2004).

Despite its importance, the SFR has been subjected to various anthropogenic actions, including damming, deforestation, pollution, introduction of exotic species, and, in recent years, water diversion (Knoppers *et al.*, 2005; Jardine *et al.*, 2015; Sabo *et al.*, 2017; Bezerra *et al.*, 2019; Silva *et al.*, 2020). These detrimental actions have caused the disappearance of certain species in specific stretches of the SFR, such as the surubim (*Pseudoplatystoma corruscans* Spix & Agassiz, 1829), pirá (*Conorhynchos conirostris* Valenciennes, 1840), and dourado (*Salminus franciscanus* Lima & Britski, 2007). Meanwhile, non-native species, such as the Nile tilapia (*Oreochromis niloticus* Linnaeus, 1758) and tucunaré (*Cichla kelberi* Kullander & Ferreira, 2006), have seen a substantial increase in their populations in recent years (Assis *et al.*, 2017). This rise in non-native species has resulted in six endemic species listed as threatened of extinction. In response, a National Action Plan was developed for the basin (Instituto Chico Mendes de Conservação da Biodiversidade – ICMBio, 2015).

The knowledge of the ichthyofauna in the SFR basin has expanded through recent studies (*i.e.*, Carvalho *et al.*, 2011; Lima *et al.*, 2017; Silva *et al.*, 2020). An inventory of the SFR resulted in records of 250 native species (ICMBio, 2015). Another study developed a checklist for the sub-middle SFR, identifying a total of 86 species (Silva *et al.*, 2020). Regarding genetic aspects, Carvalho *et al.* (2011), using DNA barcoding techniques, identified some cryptic lineages between the upper and middle sections of the SFR. However, information on the genetic diversity of species in the sub-middle and lower stretches, which suffer from high levels of damming and silting, is still limited, representing a knowledge gap in the region.

In recent decades, molecular taxonomy techniques using the mitochondrial gene cytochrome oxidase subunit I (*cox1*) have been particularly useful due to their speed and low cost (Hebert *et al.*, 2003). This technique has several applications, such as aiding in species identification, especially when there are subtle variations in morphological characteristics (Vicente *et al.*, 2020), and in detecting species complexes (Pereira *et al.*, 2021). In the Neotropical region, this strategy has proven to be quite effective in freshwater fish (Pereira *et al.*, 2011; Gomes *et al.*, 2015; Ramirez *et al.*, 2017; Jacobina *et al.*, 2018). However, many studies aiming to evaluate the ichthyofauna of river basins have been limited, focusing on evaluating only one or a few stretches of the rivers, which can result in underestimating the genetic diversity along the basin (Carvalho *et al.*, 2011; Pereira *et al.*, 2011, 2013; Gomes *et al.*, 2015; Pugedo *et al.*, 2016).

Molecular systematics techniques utilizing the *cox1* gene offer valuable insights of understanding taxonomic identity (Linnaean shortfall) and the potential vicariant

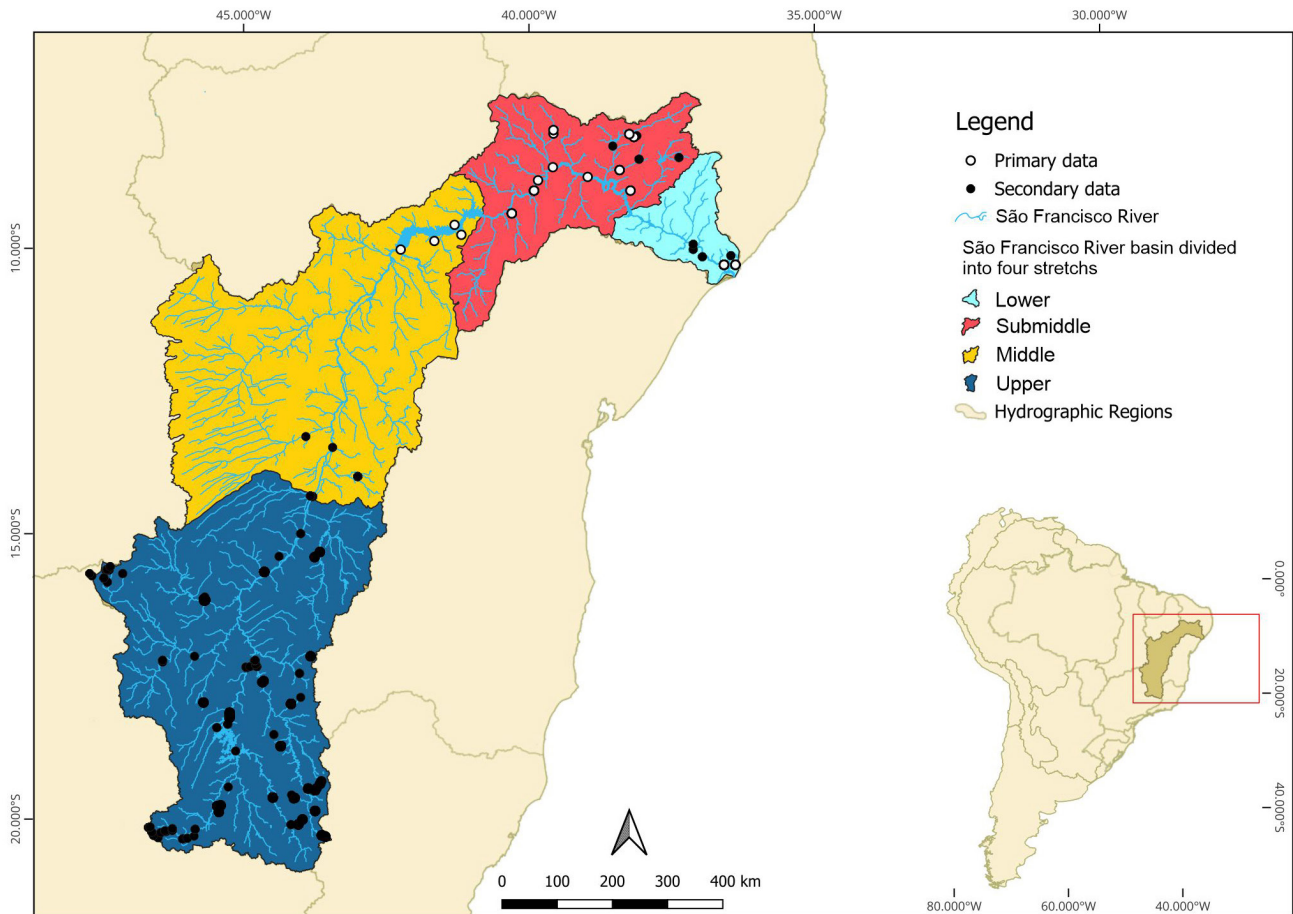
processes (Darwinian shortfall) influencing the genetic diversity of ichthyofauna in the SFR basin. In this context, the objective of this study was to address these knowledge gaps by evaluating the taxonomic identity and evolutionary cohesion of fish across all four stretches. We utilized existing sequence data from the upper and middle sections of the SFR, sourced from online databases, and augmented our analysis by incorporating newly collected sequences from the sub-middle and lower segments of the SFR, employing sequence clusters (BINs) recognized by the Barcode Index Number System and Species Delimitation Methods. Our research sought to clarify the genetic identity of fish species in the SFR basin in different reaches, contributing to a deeper understanding of biodiversity and aiding conservation efforts in the region.

## MATERIAL AND METHODS

**Sampling.** Sampling in SFR took place between 2018 and 2020 in the lower and sub-middle stretches, along the main river channel and its tributaries (Fig. 1). This was done to encompass the broadest diversity of habitats, utilizing two types of fishing gear: active (hand nets and trawling nets) and passive (gill nets and trap nets) to mitigate the selectivity of fishing gears. Gill nets were left overnight whenever feasible. A total of 120 specimens, representing 31 species, were collected, identified, and deposited in the ichthyological collection of the Universidade Federal do Rio Grande do Norte (UFRN) (Tab. S1). All specimens were georeferenced and identified to the lowest taxonomic level using identification keys or taxonomic studies (Reis *et al.*, 2003; Rosa *et al.*, 2003; Lima *et al.*, 2017).

The animals were collected and anesthetized by immersion in a 1% benzocaine solution in water. Fin samples from each individual were collected and stored in absolute ethanol for molecular analysis and deposited in the tissue bank of the Laboratório de Sistemática e Análise Integrativa Molecular de Microssistemas da Universidade Federal de Alagoas - Campus Arapiraca. Specimens were subsequently fixed in a 4% formaldehyde solution and stored in 70% ethanol in the UFRN collection.

**DNA sampling, PCR, and sequencing.** Genomic DNA was extracted from all samples using the Wizard Genomic DNA Purification kit (Promega), following the manufacturer's instructions. The *cox1* gene fragment was amplified via PCR using universal primers BarcFish1 and BarcFish2 (Ward *et al.*, 2005). All PCR reactions were conducted in a final volume of 25  $\mu$ l, containing 12  $\mu$ l of 2X Taq master mix (Vivantis), 2  $\mu$ l of genomic DNA solution with a concentration ranging from 10 to 20 ng/ $\mu$ l, 0.5  $\mu$ l of each primer (10 mM), and 15  $\mu$ l of ultrapure water. The reactions comprised an initial step of 2 minutes at 95°C, followed by 35 cycles of 30 sec at 94°C, 30 sec at 57°C, and 2 min at 72°C, with a final extension of 7 min at the same temperature, adapted from Ward *et al.* (2005). PCR products were purified with 20% polyethylene glycol (PEG). Sequencing was performed using the BigDye™ Terminator v. 3.1 Cycle Sequencing Ready Reaction kit (Applied Biosystems) with M-13 primer and read on an ABI 3130 Model sequencer (Applied Biosystems). Sequencing was conducted at the Fundação Oswaldo Cruz (FIOCRUZ), Bahia. After the evaluation of their electropherograms, all sequences were deposited in the BOLD Systems Platform.



**FIGURE 1** | Sampling map of the ichthyofauna in the São Francisco River basin. The data include specimens from primary collections and secondary data from the BOLD Systems platform. The basin is divided into four sections: upper, middle, lower middle, and lower São Francisco.

**Obtaining SFR sequences from public domain databases.** Initially, searches in public domain databases were conducted to identify species present in the São Francisco River (SFR), based on information from studies such as Carvalho *et al.* (2011), ICMBio (2015), Lima *et al.* (2017), and Silva *et al.* (2020). The species were classified based on endemism and native criteria, considering those restricted to the SFR basin region, according to the distribution patterns found in FishBase (<https://fishbase.se/search.php>) and ICMBio (2015). A total of 606 georeferenced *cox1* gene sequences (representing 94 species) from basin species were obtained from the BOLD Systems Platform. These sequences were integrated into our dataset (primary dataset) for comparative analysis (Tab. S1).

For a more detailed examination of species distribution, all geographic coordinates were plotted on a specific shapefile of the SFR basin, utilizing the freely accessible Quantum Geographic Information System (QGIS), v. 3.28.3 (QGIS Development Team, 2022). In this file, the basin is subdivided into its upper, middle, sub-middle, and lower sections, in accordance with the Agência Nacional de Águas metadata catalog, Hydrographic Base of the SFR basin – 2022.

**Data analyses.** All sequences collected from the sub-middle and lower reaches of the SFR, encompassing 31 identified species, were submitted to the Canadian DNA Barcoding Center (CCDB, Ontario Biodiversity Institute, University of Guelph) for sequence analysis and Barcode Index Numbers (BINs) generation. The project name “Genetic Diversity of Fishes from the São Francisco River” and the ID number in the BOLD System with the code RSFUJ. All sequences were deposited in April 2023. Subsequently, we downloaded and integrated these georeferenced sequences into the dataset already available in BOLD Systems for the São Francisco River basin (SFR), aiming to conduct comparative analyses among different stretches of the SFR.

We classified each species from the SFR basin according to the MATCH, SPLIT, MERGE, and MIXTURE categories, as defined by Ratnasingham, Hebert (2013). For this study, we applied these categories exclusively to species collected in the São Francisco basin. A species with only a single BIN within the São Francisco basin was classified as MATCH. When a species had two distinct BINs occurring exclusively within the São Francisco basin, it was classified as SPLIT, indicating potential intra-basin genetic divergence. Species within the São Francisco basin with BINs that are associated with different species were categorized as MERGE, suggesting potential taxonomic ambiguity or misidentification. Finally, the MIXTURE category was used for species that presented a combination of SPLIT and MERGE. This means that a species classified as MIXTURE has BINs that reflect both internal subdivisions within the basin (indicating SPLIT) and associations with BINs of other species or from other river basins (indicating MERGE). This category represents a more complex scenario, where there are both taxonomic ambiguities and genetic divergences among different populations or geographical regions.

Subsequently, all sequences of each species were aligned and edited using Geneious software (Kearse *et al.*, 2012). The alignment process utilized the Clustal  $\Omega$  package (Sievers, Higgins, 2014). Intraspecific genetic distances were calculated using the Nearest Neighbor Distance (NND) with the K2P model in MEGA 11 software (Tamura *et al.*, 2021), following the analytical routine of the BOLD Systems Platform (Ratnasingham, Hebert, 2007).

**Lineage delimitation methods.** For cases of SPLITs obtained on the BOLD Systems Platform, we employed two lineage delimitation methods: GMYC (Generalized Mixed Yule Coalescent) and ASAP (Species by Automatic Partitioning). The GMYC analysis was conducted using the GMYC web server (<http://species.l-its.org/gmyc/>), assuming a single threshold (Pons *et al.*, 2006). In GMYC, an uncalibrated Maximum Clade Credibility (MCC) tree was generated in BEAST 1.10.4 (Suchard *et al.*, 2018), containing only distinct haplotypes. The Yule model with a relaxed lognormal molecular clock was applied. The analysis consisted of two independent runs with 10 million generations each. Parameters were sampled every 1,000 generations, and a burn-in of 10% was utilized. Convergence and performance were assessed using Tracer 1.5 (Suchard *et al.*, 2018), ensuring that all Effective Sample Size (ESS) values exceeded 200.

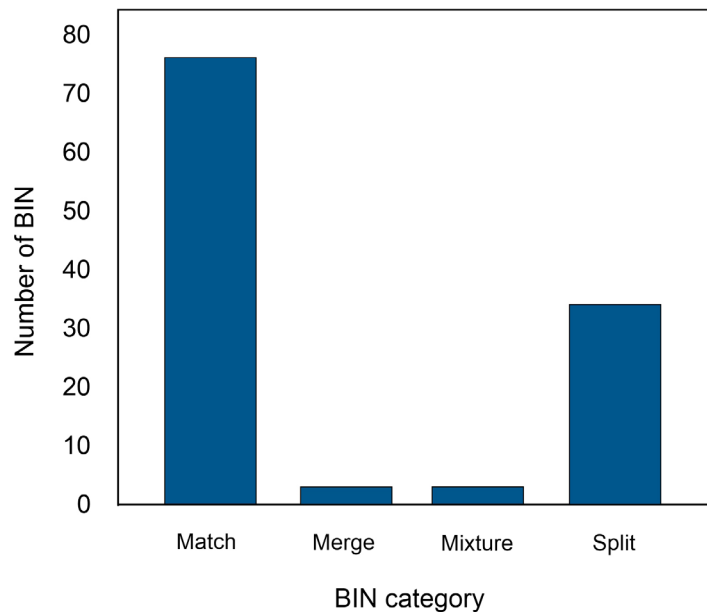
The ASAP method is based on sequence alignments (Puillandre *et al.*, 2021), accessible through the web interface (<https://bioinfo.mnhn.fr/abi/public/asap/>). The K2P nucleotide substitution model was used, and other parameters were set to default. ASAP divides species partitions based on genetic pairwise distances and calculates

panmixia probability (p-val), a relative gap width metric (W), and classifies the results by ASAP score: the lower the score, the better the partition (Puillandre *et al.*, 2021). The numbers of Molecular Operational Taxonomic Units (MOTUs) predicted by the 1<sup>st</sup> and 2<sup>nd</sup> ASAP scores were selected and compared with other methods.

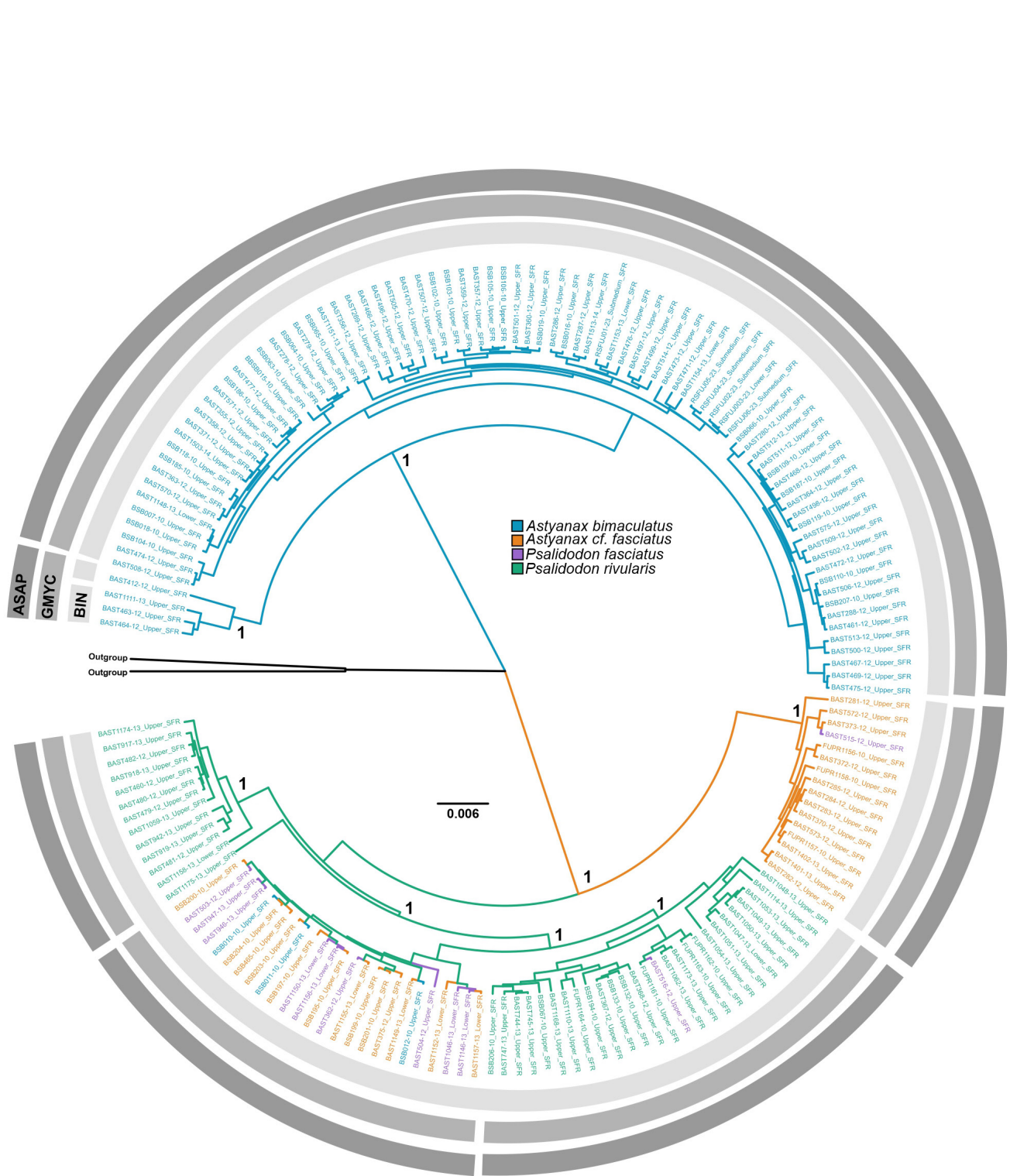
## RESULTS

We obtained *cox1* gene sequences from 726 specimens of the SFR fish fauna, including 120 original (primary data) and 606 from the BOLD Systems Platform (secondary data), representing 54% of the species cataloged for the SFR basin (Fig. 1). The total number of individuals belongs to 94 species, hierarchized within 73 genera, 32 families, and six orders: Characiformes (43 species), Siluriformes (36), Cyprinodontiformes (7), Cichliformes (4), Gymnotiformes (3) and Synbranchiformes (1) (Tab. S2). These 94 species are grouped into 111 Barcode Index Numbers (BINs), where 68 of these are exclusive to the SFR basin, meaning they are not shared with other river basins.

Following the criteria of the BOLD Systems platform, 75 BINs were detected presenting cases of MATCH, another 32 classified in the SPLIT category, in addition to two cases of MERGE, and two cases of MIXTURE (Fig. 2). Emblematic cases involving the genera *Astyanax* Baird & Girard, 1854 and *Psalidodon* Eigenmann, 1911 were observed in the MATCH, SPLIT and MIXTURE categories. Additionally, three clusters were found for *Astyanax lacustris* (Lütken, 1875) in the SFR basin (SPLIT). Conversely, a MIXTURE cluster involved representatives from all four taxa, including *Astyanax bimaculatus* (Linnaeus, 1758), *A. cf. fasciatus* (Cuvier, 1819), *Psalidodon rivularis* (Lütken, 1875), and *P. fasciatus* (Fig. 3).



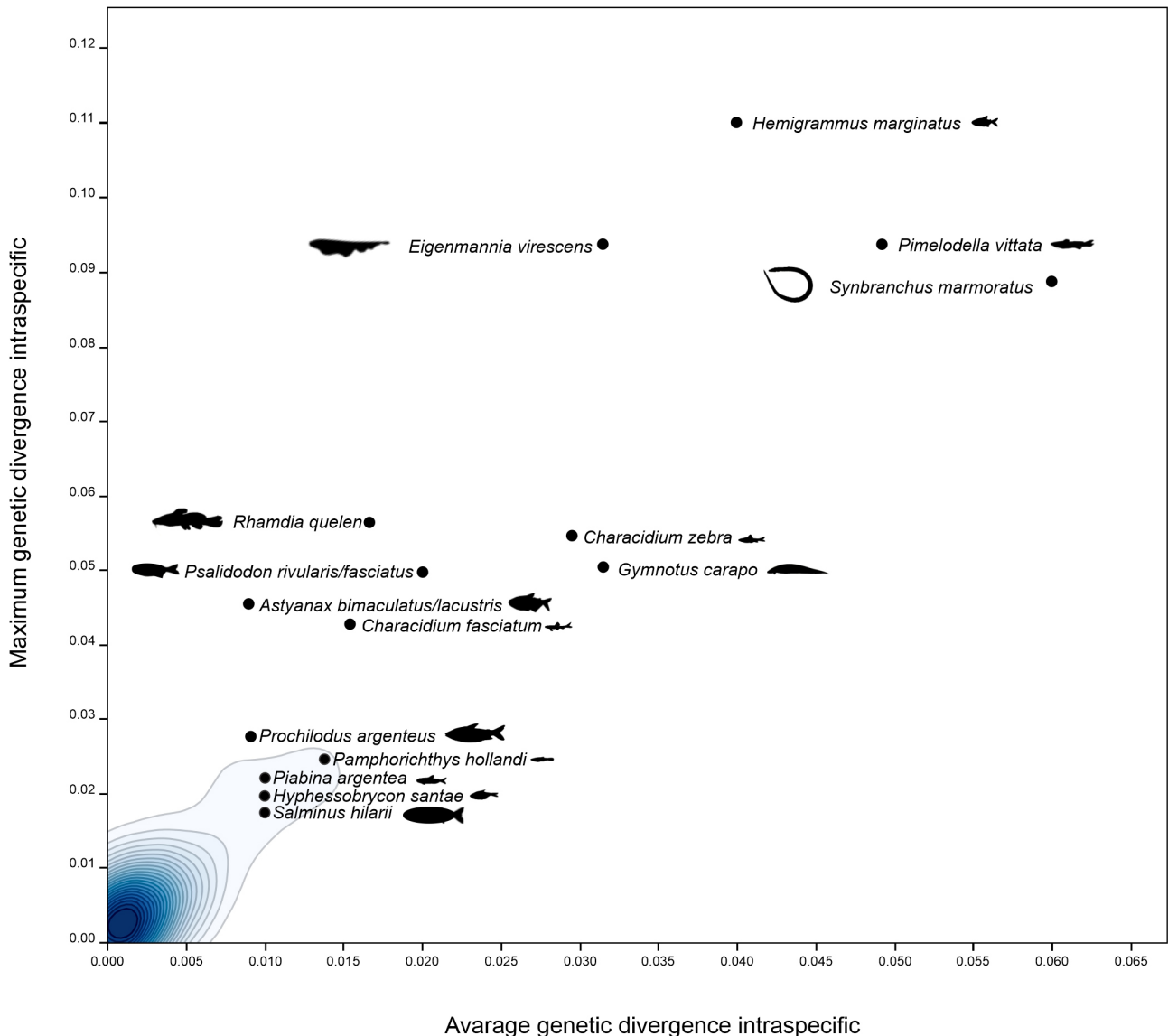
**FIGURE 2 |** Number of BINs classified into MATCH, SPLIT, MERGE, and MIXTURE categories.



**FIGURE 3** | Lineage delimitation using the BINs, ASAP, and GMYC methods in representatives of the genera *Astyanax* and *Psalidodon*. The bars show the different methods. The phylogenetic tree presents branches colored for each possible taxon, indicating cases of merge, mixture, and split. Nodes with high posterior probabilities are shown on the tree.

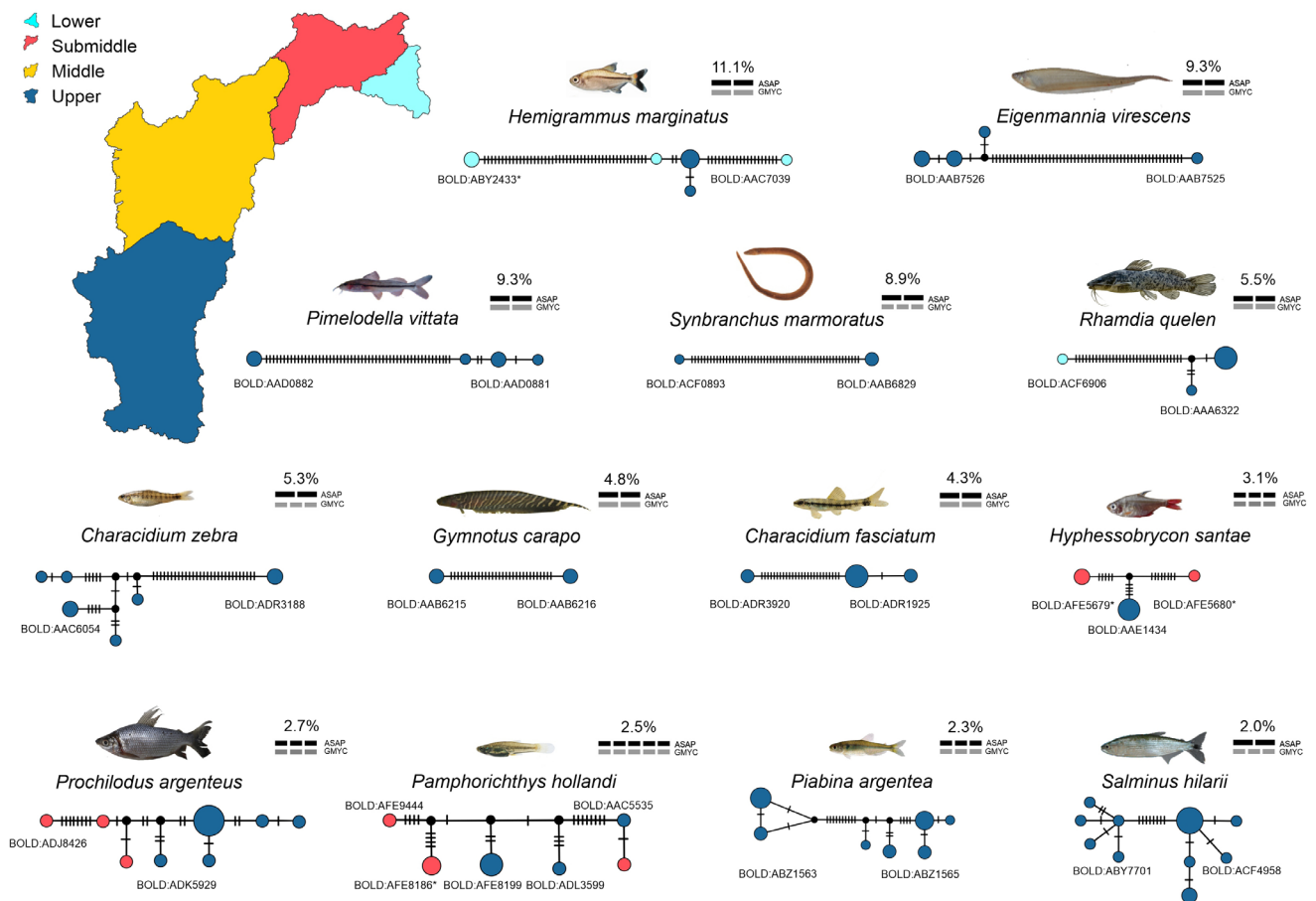


In addition to these categories, genetic divergences, evaluated intra-specifically between the regions (upper, medium, sub-medium and low) of the SFR basin, revealed distinct patterns, from low genetic divergences (less than 1%) for 74 species, moderate (between 1 and 2%) for six species, and high divergences (above 2%) for 14 species (Fig. 4; Tab. S2). Of the 94 species, 32 were evaluated within more than one section of the SFR, seven with high divergences, three with moderate divergences and 22 with genetic distances considered low. Eight species with high divergences stand out, found only in a single section, in the high SFR (Tab. S2).



**FIGURE 4 |** Contour plot showing the distribution of cox1 genetic divergences among fish species in the São Francisco River basin. Most species are concentrated near zero, with only a small portion exceeding the co-specific threshold for DNA barcoding, which is set at 0.02 (2%).

Among the fourteen species that exhibited high cryptic diversity, the lineage delimitation methods, ASAP and GMYC, demonstrated congruence with the clusters (BINs) in eight taxa: *Characidium fasciatum* Reinhardt, 1867, *Eigenmannia virescens* (Valenciennes, 1836), *Gymnotus carapo* Linnaeus, 1758, *Hemigrammus marginatus* Ellis, 1911, *Hyphessobrycon santae* (Eigenmann, 1907), *Pamphorichthys hollandi* (Henn, 1916), *Pimelodella vittata* (Lütken, 1874), *Rhamdia quelen* (Quoy & Gaimard, 1824). Notably, five clusters involved *P. hollandi*. Conversely, ASAP and GMYC indicated an additional cluster in *Piabina argentea* Reinhardt, 1867 and *Prochilodus argenteus* Spix & Agassiz, 1829. The GMYC method exhibited the most variability compared to the other methods. For *Characidium zebra* Eigenmann, 1909, *Salminus hilarii* Valenciennes, 1850, and *Synbranchus marmoratus* Bloch, 1795, GMYC detected one additional cluster (Fig. 5).



**FIGURE 5** | Genetic divergence analysis of fish species from the São Francisco River basin using BINs, ASAP, and GMYC methods. The figure displays the percentage of genetic divergence for each species, represented by their BOLD identification codes. The analysis shows significant genetic variation within species, highlighting potential cryptic diversity.

## DISCUSSION

**DNA barcoding effectiveness.** Considering the total number of sequences analyzed (726) in our study, the DNA barcoding methodology proved to be highly effective in accurately discriminating approximately 97% of the 94 fish species analyzed in the SFR. This efficiency underscores the utility of DNA barcoding as an effective tool for species identification, as also observed in other studies of Neotropical fishes (Carvalho *et al.*, 2011; Pereira *et al.*, 2011, 2013; Gomes *et al.*, 2015; Pugedo *et al.*, 2016). Among the 250 species described for the SFR basin, 45% have genetic sequences available in public databases, although the majority is concentrated in the upper and middle stretches of the SFR. In this context, our data expands the genetic knowledge of SFR fish species across different stretches (sub-middle and lower), providing new genetic sequences for 31 different species, 25% of which are endemic to the basin.

Regarding intraspecific genetic divergences, the highest variations were observed in Characiformes (11.1% - *Hemigrammus marginatus*), Gymnotiformes (9.3% - *Eigenmannia virescens*) and Synbranchiformes (5.9% - *Synbranchus marmoratus*). Our data revealed distinct patterns of genetic divergence in fish species inhabiting different stretches of the SFR. Additionally, it was possible to identify potential misidentifications that inflated some genetic distances, suggesting taxonomic uncertainties, possibly related to species complexes.

**The case of SPLIT and MIXTURE in the species complexes of *Psalidodon* and *Astyanax*.** Our results identified emblematic cases involving the genera *Astyanax* and *Psalidodon* within the MATCH, SPLIT, and MIXTURE categories, mitigating the Linnaean deficiencies related to representatives of these genera. Three clusters and three BINs (SPLIT) were identified for *Astyanax lacustris* in the SFR basin, with two exclusives to the upper section and one spread across all four sections, revealing a high cryptic diversity in this nominal taxon. A single cluster and one BIN (MATCH) for *A. cf. fasciatus*, including an individual of *P. fasciatus*, demonstrate the confusion and lack of updates in the taxonomic nomenclature of this group. Recently, *A. fasciatus* was reclassified as *P. fasciatus* (Terán *et al.*, 2020). Morphological studies indicate that this specific epithet should be applied only to specimens from the SFR, in accordance with the original description (Cuvier, 1819). However, Melo, Buckup (2006) reported a species complex, including *A. fasciatus*, occurring outside the São Francisco basin, in the Paraná river basin. Until a complete taxonomic revision is published, the valid epithet for these populations remains *P. fasciatus*.

On the other hand, a single BIN (MIXTURE) involved four species (*A. bimaculatus*, *A. cf. fasciatus*, *P. rivularis*, and *P. fasciatus*). However, Bayesian analyses and the two species delimitation methods, GMYC and ASAP, separated a cluster exclusively for *P. rivularis* from another that is shared with the other mentioned taxa. Recently, the genus *Psalidodon* was reestablished as a valid monophyletic genus by Terán *et al.* (2020). This genus includes many taxa previously classified as *Astyanax*, encompassing most species from the *A. eigenmanniorum*, *A. fasciatus*, *A. scabripinnis*, and *A. dissimilis* groups. In the SFR basin, *P. rivularis* and *P. fasciatus* still present confused identifications across different sections. Therefore, our data highlight significant challenges in taxonomy and the urgent need to revise the classifications related to these species in the São Francisco River basin (Rossini *et al.*, 2016; Terán *et al.*, 2020).

**Genetic divergences within and between stretches of the SFR.** In this study, we evaluated the impact of geographic scale on the genetic divergence among fish populations inhabiting various stretches of the SFR basin. Among the 32 taxa analyzed across multiple sections, three species exhibited moderate genetic divergences, ranging from 1% to 2.0%, while thirteen species demonstrated profound genetic divergences, between 2.1% and 11.1%. By employing a combination of species delimitation methods, including BINs, GMYC, and ASAP, we identified new Molecular Operational Taxonomic Units (MOTUs) among these nominal taxa. This discovery suggests the presence of cryptic lineages within the SFR basin.

Genetic divergences of 1% to 2% may indicate genetic structuring across different stretches of the basin. These divergence processes could have been influenced by historical events, such as alterations in the São Francisco River's course during the Pleistocene (Jacobina *et al.*, 2018; Pontes *et al.*, 2024). Similar genetic variations within this range have also been observed in various species across other Neotropical basins, including the Amazon, Doce, Paraná, and Paranaíba (Abreu *et al.*, 2009; Pazza *et al.*, 2016; Hrbek *et al.*, 2018). These structuring effects predominantly impact rheophilic species, as evidenced by our results of *Leporinus piau* Fowler, 1941 (1.9%), *Salminus hilari* (2%), and *Prochilodus argenteus* (2.5%). Notably, *P. argenteus* has been associated with discrepancies between morphological and molecular identification when compared to its congeners, *Prochilodus costatus* Valenciennes, 1850 and *P. hartii* (Gomes *et al.*, 2015). This suggests the potential occurrence of *Prochilodus* hybrids in basins where these species coexist, challenging the existence of new MOTUs for this taxon (Gomes *et al.*, 2015). Such processes may also be happening within the SFR basin, considering that *P. costatus* and *P. argenteus* coexist, though the use of this mitochondrial marker alone does not allow for definitive confirmation.

Although possible cases of moderate and high genetic divergence, *i.e.*, above 2%, were detected in the present study, a notable finding is that the majority of both moderate and deep genetic divergences occur within the upper reaches of the basin, in contrast to the middle, sub-middle, and lower reaches of the river. These cases of deep genetic divergences in the upper SFR region have already been documented previously in some species, as evidenced by Carvalho *et al.* (2011), and which covers *P. argentea*, *S. marmoratus*, *G. carapo*, *C. zebra* and *H. marginatus*, which are corroborated by our results. This observation suggests that the upper São Francisco region is home of genetically distinct fish lineages, possibly due to specific environmental and/or geomorphological factors in this section of the basin. It is important to highlight that this region has a system of reactivation of geological faults present between the headwaters of the São Francisco River basin, known as the Upper São Francisco Crustal Discontinuity (DCARSF) (Albert, Reis, 2011). Such orogenic movements may have caused headwater capture events between some tributaries that drain the SFR basin, resulting in the isolation of some populations and, consequently, the formation of divergent lineages that accumulated significant genetic mutations. These continuous reactivations of DCARSF have been associated with diversification processes in genera such as *Salminus* Agassiz, 1829 and *Hypostomus* Lacepède, 1803 (Montoya-Burgos, 2003; Machado *et al.*, 2018).

However, it is important to highlight that the divergent lineages found in the upper São Francisco may not be solely derived from historical geological processes, but also from the impact caused by the transposition event of the Piumhi River, which originally

belongs to the upper Paraná basin. This event, as reported in the literature, may have significantly influenced the current composition of the fish fauna in the upper São Francisco River basin (Moreira Filho, Buckup, 2005). Thus, the transposition of the Piumhi River should be considered as a potential factor in the mixture of the genetically distinct lineages observed in the region. Future studies investigating the impact of this transposition on the local fauna could provide additional support for this possibility.

On the other hand, in our study, a widely distributed species, *Rhamdia quelen*, exhibited a notable genetic divergence of 5.5% between the upper and lower reaches of the SFR. Significant genetic divergences were also detected in small-sized and low-vagility species, such as *P. holandi* (2.5%) and *Hyphessobrycon santae* (3.1%), between the high and sub-middle stretches of the SFR. In recent years, some species with a wide geographic distribution in the Neotropical region have undergone taxonomic revisions, being recognized as species complexes (Milhomem *et al.*, 2008; Martinez *et al.*, 2011; Utsunomia *et al.*, 2014), as well as those with apparent restricted distributions. Costa *et al.* (2012, 2018), for example, identified a species complex among annual fish of the genus *Hypsolebias* Costa, 2006, endemic to the São Francisco River basin. In this context, different methods of delimiting lineages with DNA barcode, associated with diagnostic morphological characteristics, that is, through integrative approaches, have made it possible to differentiate between morphologically similar species, contributing to the reduction of Linnean shortfalls (Costa *et al.*, 2018; Andrade *et al.*, 2020; Ota *et al.*, 2020; Costa, 2021).

The profound genetic divergences found in the present work, as well as in other studies, suggest that speciation processes may have occurred along stretches of the SFR basin, shaped by historical geomorphological processes. In other Neotropical basins, there is evidence of speciation resulting from multiple invasions from different basins or population segmentation within the same basin (Abreu *et al.*, 2020). Such processes have been observed in the Amazon and in the Paraná River basins (Benzaquem *et al.*, 2015; Guimarães *et al.*, 2018; Rueda *et al.*, 2023). In this context, we suggest that similar processes may have occurred in the SFR basin. In freshwater environments, hydrographic connectivity does not necessarily imply facilitated dispersal for aquatic organisms (Winemiller *et al.*, 2008; Hirschmann *et al.*, 2015; Bessonart *et al.*, 2021). Therefore, specific genetic signals resulting from barriers and geohydrological dynamics between basins or their tributaries can store relicts that are often difficult to detect based on species morphology alone (Machado *et al.*, 2018; Bignotto *et al.*, 2020). In this context, genetic studies can shed light on these divergent processes.

Risk assessment and biodiversity protection in the face of hidden genetic diversity. In summary, species exhibiting moderate to profound genetic divergences are classified as 'Least Concern', 'Data deficient' or of 'Unknown' status according to the IUCN (2022) criteria. This scenario suggests a potential underestimation of the risks faced by such species, due to inadequate assessment. Within this context, overlooking cryptic diversity may lead to ineffective strategies for protecting endemic species that exhibit wide genetic variations and possess specific adaptations to their environments (Bickford *et al.*, 2007). This is particularly concerning in a basin that has experienced various anthropogenic impacts, including pollution, deforestation, introduction of exotic species, and, more recently, the transposition of its waters (Bezerra, 2019). Consequently, this lack of awareness can result in the irreversible loss of unique genetic diversity, with

potential adverse effects on biodiversity, as observed in the upper and middle reaches of the SFR in the species *Conorhynchos conirostris* (Cuvier, 1829) (ICMBio, 2015). Ignoring these genetic variations can therefore create voids in the functional roles these lineages fulfill within ecosystems, affecting the stability and resilience of these complex systems.

Our results bridge shortfalls in Linnean and Darwinian knowledge, offering understandings into the genetic and taxonomic diversity of the ichthyofauna of the São Francisco River. These findings underscore the complexity of genetic diversity among fish species, influenced by geomorphological, and anthropogenic processes. Consequently, this emphasizes the importance of adopting future integrative approaches in the taxonomy and conservation of these species. It also highlights the necessity of management strategies that consider the unique characteristics of each group. Understanding these processes of genetic differentiation is crucial for the conservation of the biodiversity of the SFR basin and for implementing effective protection measures of its resident species.

## ACKNOWLEDGMENTS

We thank the Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq) and Fundação de Amparo do Estado de Alagoas (FAPEAL) for the financial support to UPJ (#425080/2016–1; #APQ2022021000055). SMQL and UPS also thank the Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq) for the productivity grants (312066/2021–0 and 12880/2023–5, respectively). LFBN thanks the CNPq and Fundação de Amparo e Promoção da Ciência Tecnologia e Inovação do Rio Grande do Norte (FAPERN) (Processo no 10959064–720.000038/2022–89). We dedicate this work to the memory of our esteemed colleague, Leandro Ferreira dos Santos, whose contributions were vital to the completion of this study. Unfortunately, Leandro passed away during the preparation of the manuscript for submission. His enthusiasm for research, unwavering dedication, and collaborative spirit were instrumental in the development of this work. We offer our deepest condolences to Leandro's family and friends. His absence will be felt not only as a colleague but also as an inspiring friend who has left an indelible mark on our lives. May his memory continue to serve as a wellspring of inspiration for future generations of researchers.

## REFERENCES

- **Abreu JMS, Waltz BT, Albert JS, Piorski NM.** Genetic differentiation through dispersal and isolation in two freshwater fish species from coastal basins of Northeastern Brazil. *Neotrop Ichthyol.* 2020; 18(3):e190114. <https://doi.org/10.1590/1982-0224-2019-0114>
- **Abreu MM, Pereira LHG, Vila VB, Foresti F, Oliveira C.** Genetic variability of two populations of *Pseudoplatystoma reticulatum* from the Upper Paraguay River Basin. *Genet Mol Biol.* 2009; 32(4):868–73. <https://doi.org/10.1590/S1415-47572009005000075>
- **Albert JS, Reis RE, editors.** Historical biogeography of Neotropical freshwater fishes. Berkeley: University of California Press; 2011.

- **Assis DAS, Dias-Filho VA, Magalhães ALB, Brito MFG.** Establishment of the non-native fish *Metynnis lippincottianus* (Cope 1870) (Characiformes: Serrasalminidae) in lower São Francisco River, northeastern Brazil. *Stud Neotrop Fauna E.* 2017; 52(3):228–38. <https://doi.org/10.1080/01650521.2017.1348057>
- **Benzaquem DC, Oliveira C, Silva BJ, Zuanon J, Porto JIR.** DNA barcoding in pencilfishes (Lebiasinidae: Nannostomus) reveals cryptic diversity across the Brazilian Amazon. *PLoS ONE.* 2015; 10(2):e0112217. <https://doi.org/10.1371/journal.pone.0112217>
- **Bessonart J, Loureiro M, Guerrero JC, Szumik C.** Distribution of freshwater fish from the Southern Neotropics reveals three new areas of endemism and show diffuse limits among freshwater ecoregions. *Neotrop Ichthyol.* 2021; 19(4):e200153. <https://doi.org/10.1590/1982-0224-2020-0153>
- **Bezerra LAV, Freitas MO, Daga VS, Occhi TVT, Faria L, Costa APL *et al.*** A network meta-analysis of threats to South American fish biodiversity. *Fish Fish.* 2019; 20(4):620–39. <https://doi.org/10.1111/FAF.12365>
- **Bickford D, Lohman DJ, Sodhi NS, Ng PK, Meier R, Winker K *et al.*** Cryptic species as a window on diversity and conservation. *Trends Ecol Evol.* 2007; 22(3):148–55. <https://doi.org/10.1016/j.tree.2006.11.004>
- **Bignotto TS, Maniglia TC, Gomes VN, Oliveira IJ, Agostinho CS, Prioli SMAP *et al.*** Genetic evidence for a species complex within the piranha *Serrasalmus maculatus* (Characiformes, Serrasalminidae) from three neotropical river basins based on mitochondrial DNA sequences. *Genet Mol Biol.* 2020; 43(1):e20190131. <https://doi.org/10.1590/1678-4685-GMB-2018-0131>
- **Bini LM, Diniz-Filho JAF, Rangel TFLVB, Bastos RP, Pinto MP.** Challenging Wallacean and Linnean shortfalls: knowledge gradients and conservation planning in a biodiversity hotspot. *Divers Distrib.* 2006; 12(5):475–82. <https://doi.org/10.1111/j.1366-9516.2006.00286.x>
- **Carvalho DC, Oliveira DAA, Pompeu PS, Leal CG, Oliveira C, Hanner R.** Deep barcode divergence in Brazilian freshwater fishes: the case of the São Francisco river basin. *Mitochondr DNA.* 2011; 22:80–86. <https://doi.org/10.3109/19401736.2011.588214>
- **Costa WJEM.** Comparative osteology, phylogeny and classification of the Eastern South American catfish genus *Trichomycterus* (Siluriformes: Trichomycteridae). *Taxonomy.* 2021; 1(2):160–91. <https://doi.org/10.3390/taxonomy1020013>
- **Costa WJEM, Amorim PF, Mattos JLO.** Species delimitation in annual killifishes from the Brazilian Caatinga, the *Hypsolebias flavicaudatus* complex (Cyprinodontiformes: Rivulidae): implications for taxonomy and conservation. *Syst Biodivers.* 2012; 10(1):71–91. <https://doi.org/10.1080/14772000.2012.664177>
- **Costa WJEM, Amorim PF, Mattos JLO.** Cryptic species diversity in the *Hypsolebias magnificus* complex, a clade of endangered seasonal killifishes from the São Francisco river basin, Brazilian Caatinga (Cyprinodontiformes, Aplocheilidae). *Zookeys.* 2018; (777):141–58. <https://doi.org/10.3897/zookeys.777.25058>
- **Dagosta FCP, de Pinna M.** The fishes of the Amazon: distribution and biogeographical patterns, with a comprehensive list of species. *B Am Mus Nat Hist.* 2019; 2019(431):1–163. <https://doi.org/10.1206/0003-0090.431.1.1>
- **Diniz-Filho JAF, Loyola RD, Raia P, Mooers AO, Bini LM.** Darwinian shortfalls in biodiversity conservation. *Trends Ecol Evol.* 2013; 28(12):689–95. <https://doi.org/10.1016/j.tree.2013.09.003>
- **Freitas PV, Montag LFA, Ilha P, Torres NR, Maia C, Deegan L *et al.*** Local effects of deforestation on stream fish assemblages in the amazon-savannah transitional area. *Neotrop Ichthyol.* 2021; 19(3):e210098. <https://doi.org/10.1590/1982-0224-2021-0098>
- **QGIS Development Team.** QGIS geographic information system. Chicago, IL: Open Source Geospatial Foundation Project; 2022. Available from: <https://qgis.org/en/site/>
- **Gomes LC, Pessali TC, Sales NG, Pompeu PS, Carvalho DC.** Integrative taxonomy detects cryptic and overlooked fish species in a neotropical river basin. *Genetica.* 2015; 143(5):581–88. <https://doi.org/10.1007/S10709-015-9856-Z>

- **Guimarães KLA, Sousa MPA, Ribeiro FRV, Porto JIR, Rodrigues LRR.** DNA barcoding of fish fauna from low order streams of Tapajós River basin. *PLoS ONE*. 2018; 13(12):e0209430. <https://doi.org/10.1371/journal.pone.0209430>
- **Harrison I, Darwall W, Lee RR, Muruven D, Revenga C, Claussen J et al.** The world's forgotten fishes. *World Wide Fund Nature (WWF)*. 2021:48. Available from: [https://wwf.panda.org/discover/our\\_focus/freshwater\\_practice/the\\_world\\_s\\_forgotten\\_fishes/](https://wwf.panda.org/discover/our_focus/freshwater_practice/the_world_s_forgotten_fishes/)
- **Hebert PDN, Cywinska A, Ball SL, Waard JR.** Biological identifications through DNA barcodes. *Proc R Soc B*. 2003; 270(1512):313–21. <https://doi.org/10.1098/RSPB.2002.2218>
- **Hirschmann A, Malabarba LR, Thomaz AT, Fagundes NJR.** Riverine habitat specificity constrains dispersion in a Neotropical fish (Characidae) along Southern Brazilian drainages. *Zool Scr*. 2015; 44(4):374–82. <https://doi.org/10.1111/zsc.12106>
- **Hortal J, Bello F, Diniz-Filho JAF, Lewinsohn TM, Lobo JM, Ladle RJ.** Seven shortfalls that beset large-scale knowledge of biodiversity. *Annu Rev Ecol Evol Syst*. 2015; 46:523–49 <https://doi.org/10.1146/annurev-ecolsys-112414-054400>
- **Hrbek T, Meliciano NV, Zuanon J, Farias IP.** Remarkable geographic structuring of rheophilic fishes of the lower Araguaia River. *Front Genet*. 2018; 9:295. <https://doi.org/10.3389/fgene.2018.00295>
- **Instituto Chico Mendes de Conservação da Biodiversidade (ICMBio).** Portaria nº 34, de 27 de maio de 2015. Plano de Ação Nacional para Conservação das Espécies Ameaçadas de Extinção da Fauna Aquática da Bacia do São Francisco - PAN São Francisco. 2015.
- **International Union for Conservation of Nature (IUCN).** Standards and petitions committee. Guidelines for using the IUCN Red List categories and criteria. Version 15 [Internet]. Gland; 2022. Available from: <http://www.iucnredlist.org/documents/RedListGuidelines.pdf>
- **Ilha P, Rosso S, Schiesari L.** Effects of deforestation on headwater stream fish assemblages in the Upper Xingu River Basin, Southeastern Amazonia. *Neotrop Ichthyol*. 2019; 17(1):e180099. <https://doi.org/10.1590/1982-0224-20180099>
- **Jacobina UP, Lima SMQ, Maia DG, Souza G, Batalha-Filho H, Torres RA.** DNA barcode sheds light on systematics and evolution of neotropical freshwater trahiras. *Genetica*. 2018; 146(6):505–15. <https://doi.org/10.1007/S10709-018-0043-X>
- **Jardine TD, Woods R, Marshall J, Fawcett J, Lobegeiger J, Valdez D et al.** Reconciling the role of organic matter pathways in aquatic food webs by measuring multiple tracers in individuals. *Ecology*. 2015; 96(12):3257–69. <https://doi.org/10.1890/14-2153.1>
- **Jörger KM, Schrödl M.** How to describe a cryptic species? Practical challenges of molecular taxonomy. *Front Zool*. 2013; 10(1):59. <https://doi.org/10.1186/1742-9994-10-59>
- **Kearse M, Moir R, Wilson A, Stones-Havas S, Cheung M, Sturrock S et al.** Geneious basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data. *Bioinformatics*. 2012; 28(12):1647–49. <https://doi.org/10.1093/bioinformatics/bts199>
- **Knoppers B, Medeiros PRP, Souza WFL, Jennerjahn T.** The São Francisco Estuary, Brazil 2005:51–70. [https://doi.org/10.1007/698\\_5\\_026](https://doi.org/10.1007/698_5_026)
- **Lima SMQ, Ramos TPA, Silva MJ, Rosa RS.** Diversity, distribution, and conservation of the Caatinga fishes: advances and challenges. In: Silva JMC, Leal IR, Tabarelli M, editors. *Caatinga*. Springer International Publishing; 2017. p.97–131. [https://doi.org/10.1007/978-3-319-68339-3\\_4](https://doi.org/10.1007/978-3-319-68339-3_4)
- **Machado CB, Galetti Jr. PM, Carnaval AC.** Bayesian analyses detect a history of both vicariance and geodispersal in Neotropical freshwater fishes. *J Biogeogr*. 2018; 45(6):1313–25. <https://doi.org/10.1111/jbi.13207>
- **Martinez JF, Lui RL, Blanco DR, Traldi JB, Silva LF, Venere PC et al.** Comparative cytogenetics of three populations from the *Rhamdia quelen* species complex (Siluriformes, Heptapteridae) in two Brazilian hydrographic basins. *Caryologia*. 2011; 64(1):121–28. <https://doi.org/10.1080/00087114.2011.10589772>
- **Melo FAG, Buckup PA.** *Astyanax henseli*, a new name for *Tetragonopterus aeneus* Hensel, 1870 from southern Brazil (Teleostei: Characiformes). *Neotrop Ichthyol*. 2006; 4(1):42–52. <https://doi.org/10.1590/s1679-62252006000100003>



- **Milhomem SSR, Pieczarka JC, Crampton WGR, Silva DS, Souza ACP, Carvalho JR *et al.*** Chromosomal evidence for a putative cryptic species in the *Gymnotus carapo* species-complex (Gymnotiformes, Gymnotidae). *BMC Genet.* 2008; 9:75. <https://doi.org/10.1186/1471-2156-9-75>
- **Montoya-Burgos JI.** Historical biogeography of the catfish genus *Hypostomus* (Siluriformes: Loricariidae), with implications on the diversification of Neotropical ichthyofauna. *Mol Ecol.* 2003; 12(7):1885–67. <https://doi.org/10.1046/j.1365-294X.2003.01857.x>
- **Moreira-Filho O, Bertollo LAC.** *Astyanax scabripinnis* (Pisces, Characidae): a species complex. *Rev Bras Genet.* 1991; 14(2):331–57.
- **Moreira-Filho O, Buckup PA.** A poorly known case of watershed transposition between the São Francisco and upper Paraná river basins. *Neotrop Ichthyol.* 2005; 3(3):449–52. <https://doi.org/10.1590/S1679-62252005000300015>
- **O’Hanley JR, Pompeu PS, Louzada M, Zambaldi LP, Kemp PS.** Optimizing hydropower dam location and removal in the São Francisco river basin, Brazil to balance hydropower and river biodiversity tradeoffs. *Landsc Urban Plan.* 2020; 195:103725. <https://doi.org/10.1016/j.LANDURBPLAN.2019.103725>
- **Ota RP, Machado VN, Andrade MC, Collins RA, Farias IP, Hrbek T.** Integrative taxonomy reveals a new species of pacu (Characiformes: Serrasalminae: *Myloplus*) from the Brazilian Amazon. *Neotrop Ichthyol.* 2020; 18(1):e190112. <https://doi.org/10.1590/1982-0224-20190112>
- **Pazza R, Argento RFM, Mariano IHS, Kavalco KF.** Phenotypic plasticity and karyotype conservation in allopatric populations of *Astyanax aff. fasciatus* (Teleostei, Characiformes) at the Upper Paraná river basin. *Acta Zool.* 2016; 97(2):241–45. <https://doi.org/10.1111/azo.12120>
- **Pereira LHG, Castro JRC, Vargas PMH, Gomez JAM, Oliveira C.** The use of an integrative approach to improve accuracy of species identification and detection of new species in studies of stream fish diversity. *Genetica.* 2021; 149(2):103–16. <https://doi.org/10.1007/s10709-021-00118-6>
- **Pereira LHG, Hanner R, Foresti F, Oliveira C.** Can DNA barcoding accurately discriminate megadiverse Neotropical freshwater fish fauna? *BMC Genet.* 2013; 14(1):20. <https://doi.org/10.1186/1471-2156-14-20>
- **Pereira LHG, Maia GMG, Hanner R, Foresti F, Oliveira C.** DNA barcodes discriminate freshwater fishes from the Paraíba do Sul River Basin, São Paulo, Brazil. *Mitochondr DNA.* 2011; 22:71–79. <https://doi.org/10.3109/19401736.2010.532213>
- **Pons J, Barraclough TG, Gomez-Zurita J, Cardoso A, Duran DP, Hazell S *et al.*** Sequence-based species delimitation for the DNA taxonomy of undescribed insects. *Syst Biol.* 2006; 55(4):595–609. <https://doi.org/10.1080/10635150600852011>
- **Pontes AI, Loeb MV, Mai ACG, Mott T, Batalha-Filho H, Jacobina UP.** Pleistocene sea level variation and changes in the São Francisco river drove the evolution of Atlantic sabretooth anchovy. *Estuar Coast Shelf Sci.* 2024; 298:108628. <https://doi.org/10.1016/j.ecss.2024.108628>
- **Pugedo ML, Neto FRA, Pessali TC, Birindelli JLO, Carvalho DC.** Integrative taxonomy supports new candidate fish species in a poorly studied neotropical region: the Jequitinhonha River Basin. *Genetica.* 2016; 144(3):341–49. <https://doi.org/10.1007/S10709-016-9903-4>
- **Puillandre N, Brouillet S, Achaz G.** ASAP: assemble species by automatic partitioning. *Mol Ecol Resour.* 2021; 21(2):609–20. <https://doi.org/10.1111/1755-0998.13281>
- **Ramirez JL, Birindelli JL, Carvalho DC, Affonso PRAM, Venere PC, Ortega H *et al.*** Revealing hidden diversity of the underestimated neotropical ichthyofauna: DNA barcoding in the recently described genus *Megaleporinus* (characiformes: Anostomidae). *Front Genet.* 2017; 8:149. <https://doi.org/10.3389/fgene.2017.00149>
- **Ratnasingham S, Hebert PDN.** BOLD: the barcode of life data system. *Mol Ecol Notes.* 2007; 7(3):355–64. <https://doi.org/10.1111/J.1471-8286.2007.01678.X>
- **Ratnasingham S, Hebert PDN.** A DNA-based registry for all animal species: the barcode index number (BIN) system. *PLoS ONE.* 2013; 8(7):e66213. <https://doi.org/10.1371/journal.pone.0066213>

- **Reis RE, Albert JS, Di Dario F, Mincarone MM, Petry P, Rocha LA.** Fish biodiversity and conservation in South America. *J Fish Biol.* 2016; 89(1):12–47. <https://doi.org/10.1111/JFB.13016>
- **Reis RE, Kullander SO, Ferraris CJ, editors.** Check list of the freshwater fishes of south and Central America. Porto Alegre: Edipucrs; 2003.
- **Rosa RS, Menezes NA, Britski HA, Costa WJEM, Groth F.** Diversidade, padrões de distribuição e conservação dos peixes da Caatinga. In: Leal IR, Tabarelli M, Silva JMC, editors. *Ecologia e conservação da Caatinga*. Recife: Editora Universitária da UFPE; 2003. p.135–81.
- **Rossini BC, Oliveira CAM, Melo FAG, Bertaco VA, Astarloa JMD, Rosso JJ et al.** Highlighting *Astyanax* species diversity through DNA barcoding. *PLoS ONE.* 2016; 11(12):e0167203. <https://doi.org/10.1371/journal.pone.0167203>
- **Rueda EC, Machado CB, Castro V, Braga-Silva A, Ojeda G, Vargas F et al.** Genetic population structure of *Pseudoplatystoma corruscans* (Siluriformes: Pimelodidae) and evidence of temporal variation in structure. *J Fish Biol.* 2023; 102(5):1040–48. <https://doi.org/10.1111/jfb.15346>
- **Sabo JL, Ruhi A, Holtgrieve GW, Elliott V, Arias ME, Ngor PB et al.** Designing river flows to improve food security futures in the Lower Mekong Basin. *Science.* 2017; 358(6368). <https://doi.org/10.1126/SCIENCE.AAO1053>
- **Santos LCM, Matos HR, Schaeffer-Novelli Y, Cunha-Lignon M, Bitencourt MD, Koedam N et al.** Anthropogenic activities on mangrove areas (São Francisco River Estuary, Brazil Northeast): a GIS-based analysis of CBERS and SPOT images to aid in local management. *Ocean Coast Manag.* 2014; 89:39–50. <https://doi.org/10.1016/j.ocecoaman.2013.12.010>
- **Sato Y, Godinho HP.** Migratory fishes of the São Francisco River. In: Carolsfeld J, Harvey B, Ross C, Baer A, editors. *Migratory fishes of South America biology, fisheries and conservation status*, 1st edition. Victoria, BC: The World Bank; 2003. p.195–232.
- **Sievers F, Higgins DG.** Clustal Omega. *Curr Protoc Bioinformatics.* 2014. <https://doi.org/10.1002/0471250953.bi0313s48>
- **Silva MJ, Ramos TPA, Carvalho FR, Brito MFG, Ramos RTC, Rosa RS et al.** Freshwater fish richness baseline from the São Francisco Interbasin Water Transfer Project in the Brazilian Semiarid. *Neotrop Ichthyol.* 2020; 18(4):e200063. <https://doi.org/10.1590/1982-0224-2020-0063>
- **Suchard MA, Lemey P, Baele G, Ayres DL, Drummond AJ, Rambaut A.** Bayesian phylogenetic and phylodynamic data integration using BEAST 1.10. *Virus Evol.* 2018; 4(1):vey016. <https://doi.org/10.1093/ve/vey016>
- **Tamura K, Stecher G, Kumar S.** MEGA11: molecular evolutionary genetics analysis version 11. *Mol Biol Evol.* 2021; 38(7):3022–27. <https://doi.org/10.1093/MOLBEV/MSAB120>
- **Terán GE, Benitez MF, Mirande JM.** Opening the Trojan horse: phylogeny of *Astyanax*, two new genera and resurrection of *Psalidodon* (Teleostei: Characidae). *Zool J Linn Soc.* 2020;190(4):1217–34. <https://doi.org/10.1093/zoolinnea/zlaa019>
- **Utsunomia R, Pansonato-Alves JC, Costa-Silva GJ, Mendonça FF, Scacchetti PC, Oliveira C et al.** Molecular and cytogenetic analyses of cryptic species within the *Synbranchus marmoratus* Bloch, 1795 (Synbranchiformes: Synbranchidae) grouping: species delimitations, karyotypic evolution and intraspecific diversification. *Neotrop Ichthyol.* 2014; 12(4):903–11. <https://doi.org/10.1590/1982-0224-20140039>
- **Vicente F, Loeb MV, Paiva ACG, Sampaio CLS, Argolo LA, Jacobina UP.** Integrative systematics unveils the controversial identity of Engraulidae fishing stocks in a Neotropical estuary, northeast Brazil. *Neotrop Ichthyol.* 2020; 18(4):e200037. <https://doi.org/10.1590/1982-0224-2020-0037>
- **Ward RD, Zemplak TS, Innes BH, Last PR, Hebert PD.** DNA barcoding Australia's fish species. *Philos Trans R Soc B Biol Sci.* 2005; 360(1462):1847–57. <https://doi.org/10.1098/RSTB.2005.1716>
- **Winemiller KO, López-Fernández H, Taphorn DC, Nico LG, Duque AB.** Fish assemblages of the Casiquiare River, a corridor and zoogeographical filter for dispersal between the Orinoco and Amazon basins. *J Biogeogr.* 2008; 35(9):1551–63. <https://doi.org/10.1111/j.1365-2699.2008.01917.x>

### AUTHORS' CONTRIBUTION

**Denis Bruno Santos Marques Nunes:** Conceptualization, Formal analysis, Investigation, Methodology, Software, Validation, Visualization, Writing—original draft.

**Alany Itala Pontes:** Conceptualization, Data curation, Formal analysis, Investigation, Methodology, Software, Visualization, Writing—review and editing.

**Leandro Ferreira dos Santos:** Conceptualization, Formal analysis, Investigation, Methodology, Software, Validation, Visualization, Writing—review and editing.

**Leonardo Luís Calado:** Data curation, Formal analysis, Investigation, Methodology, Software, Validation, Writing—review and editing.

**Luciano Freitas Barros-Neto:** Data curation, Formal analysis, Methodology, Validation, Visualization, Writing—review and editing.

**Sílvia Yasmin Lustosa-Costa:** Conceptualization, Data curation, Formal analysis, Methodology, Visualization, Writing—review and editing.

**Sergio Maia Queiroz Lima:** Data curation, Formal analysis, Funding acquisition, Investigation, Methodology, Validation, Visualization, Writing—original draft, Writing—review and editing.

**Uedson Pereira Jacobina:** Conceptualization, Data curation, Formal analysis, Funding acquisition, Investigation, Methodology, Project administration, Resources, Software, Supervision, Validation, Visualization, Writing—original draft, Writing—review and editing.

## Neotropical Ichthyology



This is an open access article under the terms of the Creative Commons Attribution License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

Distributed under Creative Commons CC-BY 4.0

© 2024 The Authors. Diversity and Distributions Published by SBI



Official Journal of the Sociedade Brasileira de Ictiologia

### ETHICAL STATEMENT

All fish were collected under license number 32656–1/2012 issued by ICMBio.

### COMPETING INTERESTS

The authors declare no competing interests.

### HOW TO CITE THIS ARTICLE

- Nunes DBSM, Pontes AI, Santos LF, Calado LL, Barros-Neto LF, Lustosa-Costa SY, Lima SMQ, Jacobina UP. DNA barcoding highlights taxonomic uncertainties and cryptic lineages of São Francisco River basin fishes. *Neotrop Ichthyol.* 2024; 22(4):e240059. <https://doi.org/10.1590/1982-0224-2024-0059>