Drainage rearrangements and in situ diversification of an endemic freshwater fish genus from northeastern Brazilian rivers

Running title: Phylogeography of *Nematocharax*

Silvia Britto Barreto[a], L. Lacey Knowles[b], Rilquer Mascarenhas[a], Paulo Roberto Antunes de Mello Affonso[c], and Henrique Batalha-Filho[a]

[a]National Institute of Science and Technology in Interdisciplinary and Transdisciplinary Studies in Ecology and Evolution (INCT IN-TREE), Institute of Biology, Federal University of Bahia, 147 Barão de Jeremoabo St., 40170-115, Salvador – BA, Brazil

[b]Department of Ecology and Evolutionary Biology, Museum of Zoology, University of Michigan, 1105 N University Ave., 48109, Ann Arbor – MI, USA

[c]Department of Biological Sciences, State University of Southwestern Bahia, José Moreira Sobrinho Ave., 45208-091, Jequié – BA, Brazil

Corresponding author: Silvia Britto Barreto (https://orcid.org/0000-0001-8780-1959)
Address: Institute of Biology, Federal University of Bahia, 147 Barão de Jeremoabo St., 40170-115, Salvador – BA, Brazil; Telephone: +55 71 3283 6556; Fax: +55 71 3283 6511; E-mail address: sbrittob@gmail.com

This is the author manuscript accepted for publication and has undergone full peer review but has not been through the copyediting, typesetting, pagination and proofreading process, which may lead to differences between this version and the Version of Record. Please cite this article as doi: 10.1111/fwb.13879

ABSTRACT

1. Drainage rearrangements, either headwater captures or coastal paleodrainages formed when sea level was low, are often invoked to explain connectivity and isolation among fish populations. Unraveling these events is crucial for understanding the evolutionary processes that have shaped the genetic diversity and differentiation in freshwater fishes, which is especially relevant in regions with high endemism and species richness.

2. Here, we analyze mitochondrial (COI) and genomic (RADseq) data to test the putative effects of the current configuration of basins and historical drainage rearrangements on the genetic structuring of a characid fish (*Nematocharax*) endemic to a largely overlooked Neotropical freshwater ecoregion – the Northeastern Mata Atlantica (NMA). Bathymetric and geomorphological data were also used to generate hypotheses for two potential routes of dispersal (headwater captures and coastal paleodrainages).

3. We found that the divergence between lineages from the highlands of the Brazilian shield and the lowlands occurred during the Mio-Pliocene (i.e., divergence between *N. varii* and *N. venustus*), followed by divergence events within *N. venustus* in lowland basins during the Pleistocene. The general distribution of genetic variation in *N. venustus* seems to reflect the current configuration of basins, suggesting long-term isolation, but a subset of the inferred drainage rearrangements have facilitated movement among these watersheds, which is supported by both mtDNA and genomic data.

4. Our results suggest that the NMA river basins have had their own independent histories, except for some past temporary connections that allowed dispersal events and multiple independent colonization of basins, as seen in the Contas and Cachoeira river systems.

5. Estimating when and where connections between river basins may have occurred is fundamental to understand the role of different historical processes structuring divergence in freshwater fish species.

Keywords: coastal basins; genetic structure; headwater captures; Neotropical fish; paleodrainages.

INTRODUCTION

River networks represent complex systems whose architecture may vary greatly over both space and time, affecting the evolution and distribution of freshwater organisms (Hughes et al., 2009; Dias et al., 2014). Drainage rearrangements, in particular headwater captures and coastal paleodrainages, are considered potential drivers of diversification or divergence within species by providing periodic connections between previously isolated rivers such that species can expand their ranges into river segments (e.g. Thomaz et al., 2015; Lima et al., 2017; Shelley et al., 2020). Unlike terrestrial and marine species, obligate freshwater fishes depend on stream connectivity to move among river basins and colonize new drainages. Therefore, isolation among populations and species lineages may be pronounced in scenarios with limited dispersal opportunities, resulting in long-term isolation (Carvajal-Quintero et al., 2019). On the other hand, some fish may show shared haplotypes among drainages in cases where past episodic connections allowed genetic exchange between adjacent basins (Thomaz & Knowles, 2018).

The complex evolutionary history of the Neotropical ichthyofauna is still not fully understood, given its high species richness and the variable topography of the hydrographic systems in this region (Albert & Reis, 2011). Not only does the current configuration of disconnected basins represent barriers to gene flow and contribute to divergence processes by allopatry, but recent studies have also shown the effects of past geomorphological events on population genetic diversity and differentiation in Neotropical fishes (e.g. Thomaz et al., 2017; Camelier et al., 2018; Lima et al., 2021). Among these events, past connections between today's isolated basins are forged by either headwater captures or coastal paleodrainages, which can facilitate dispersal between drainages and range expansion, and possibly promote speciation. Headwater capture involves the splitting and merging of drainages through processes influenced by the steepness, amount of rainfall, and type (hardness) of the underlying rock (Bishop, 1995; Albert et al., 2018). Coastal paleodrainages also provided past connections among river basins due to sea level retreats associated with glacial cycles that largely exposed the continental shelf, which occurred mainly during the Pleistocene (Weitzman et al., 1988). This type of paleo-connection is especially relevant in small to mid-size coastal basins, in which the near-shore marine environment can function as a barrier for strictly freshwater fishes. In large river systems, such as the Amazon and Orinoco rivers along the northern coast of South America, the spreading of freshwater plumes is hypothesized to facilitate dispersal of freshwater fishes among coastal basins regardless of potential Pleistocene river extensions (Winemiller & Willis, 2011).

Valuable information on the occurrence of drainage rearrangements can be provided by genetic data, for instance by revealing the presence of closely related groups in adjacent basins or secondary contact between populations that were previously geographically isolated in different basins (Burridge et al., 2006; Schwarzer et al., 2012). However, bathymetric and geomorphological evidence are crucial complementary data that can be used to generate hypotheses of past connections between river systems (e.g. Thomaz & Knowles, 2020). Furthermore, the predominance of headwater captures and coastal paleodrainages in structuring fish communities is expected to vary regionally. For example, the width of the continental shelf differs along the Brazilian coast; the northeast region is considerably narrower compared to the southeast (Martins & Coutinho, 1981). Consequently, connections via coastal paleodrainages are less likely in the northeast (above the Abrolhos Bank) during periods of marine regression (Thomaz & Knowles, 2018). On the other hand, headwater captures may have provided several connections in northeastern Brazil (Ribeiro, 2006).

With limited information on the biogeographic history of freshwater organisms from northeastern Brazilian coastal basins like those of the Northeastern Mata Atlantica (NMA) ecoregion (Abell et al., 2008), it is not clear the extent to which historical events have shaped the evolution of the ichthyofauna. This is especially the case in fish communities that are poorly understood despite high levels of endemism, such as the NMA ecoregion, which includes coastal lowlands and adjacent highlands from the Espinhaço Range, and where up to 61% of species are endemic (Albert et al., 2011). Parallel and isolated river basins of the NMA, which is bordered by the São Francisco River in the north and west, and the Paraíba do Sul River in the south (Hales & Petry, 2015), have no doubt triggered high levels of endemism. The NMA ecoregion is subdivided into three biogeographic regions (Northern, Central, and Southern), with different species compositions and independent histories (Camelier & Zanata, 2014). However, within each region, and especially in the Central region, faunal distributions suggest past connections between presently isolated river basins (i.e., the Contas, Almada, Cachoeira, Una, Pardo, and Jequitinhonha river basins; Figure 1).

One taxon that supports the hypothesis of some shared history among the Central NMA basins is the endemic fish *Nematocharax venustus* (Camelier & Zanata, 2014), a small member (up to 60 mm in standard length) of the Characidae family. The species was once considered vulnerable to extinction due to impacts on its habitats such as removal of riparian vegetation and introduction of exotic species (Menezes & Lima, 2008). Outstanding features of *N. venustus* include sexual dimorphism and morphological and genetic variation among its populations (Menezes et al., 2015; Barreto, Cioffi, et al., 2016; Barreto, Nunes, et al., 2016).

On the other hand, its only congener, *N. varii*, is known for only two nearby localities in a tributary of the Upper Contas River situated around 530 m elevation on the Chapada Diamantina highlands (Barreto et al., 2018). In this sense, reconstructing the evolutionary history of *Nematocharax* populations can help disentangle the role of drainage rearrangements on the evolution of freshwater fishes.

Here we used cytochrome c oxidase subunit I (COI) sequences and restriction site-associated DNA (RADseq) markers to investigate the spatial and temporal patterns of population divergence across the entire range of *Nematocharax* (including both *N. venustus* and *N. varii*). Specifically, by considering the genetic data jointly with bathymetric, geological, and geomorphological information, we tested the putative effects of the current configuration of basins and historical drainage rearrangements, either headwater captures or coastal paleodrainages, on the phylogeographic structuring of *Nematocharax*. We hope that by combining these different data sources to unravel the population history of this endemic fish genus we can shed light on the biogeographic history of the NMA region's aquatic biota.

METHODS

Sampling

Tissue samples from 182 specimens of *Nematocharax*, including *N. venustus* and *N. varii*, were obtained from 37 collection sites across the entire distribution of the genus (i.e., the Contas, Almada, Cachoeira (including the Santana River), Una, Pardo, and Jequitinhonha river basins; Figure 1). The collection license (number 51856-2) was provided by the Instituto Chico Mendes de Conservação da Biodiversidade (ICMBio/SISBIO), and the ethical approval for this study was obtained from the Ethics Committee of Utilization of Animals from the Universidade Estadual do Sudoeste da Bahia (CEUA/UESB, number 71/2014). All collected fish are deposited in the ichthyological collection of the Universidade Federal da Bahia (UFBA) (see Table S1 for details).

Sanger sequencing

We investigated the genetic structure of *Nematocharax* at a broad spatial scale using COI sequences for all samples. Total DNA was extracted from ethanol preserved tissues (muscle or fin) using the Wizard Genomic DNA Purification kit (Promega, Madison, USA).

Six hundred and fifty base pairs (bp) of COI were amplified with the primers FishF2_t1 and FishR2_t1 (Ward et al., 2005) following the same conditions and steps described in Barreto, Nunes, et al. (2016). Sequencing in both directions was performed at the Gonçalo Moniz Research Center (FIOCRUZ-Bahia) using the BigDyeTerminator v3.1 Cycle Sequencing Ready Reaction kit (Applied Biosystems, Foster City, USA).

Forward and reverse electropherograms were assembled into contigs using the program CodonCode Aligner version 6.0 (CodonCode Corporation; http://www.codoncode.com/aligner/). Sequences were then aligned with the ClustalW Multiple Alignment tool (Thompson et al., 1994) in BioEdit 7.1.9 and deposited in GenBank (accession numbers MN011168-MN011188 and MN011203-MN011363). In addition to the 182 sequences generated specifically for this study, other 113 COI sequences of *Nematocharax* (including five additional localities) were downloaded from GenBank and BOLD (Barcode of Life Data Systems; http://www.boldsystems.org/) (acc. num. PIABA001-14 to PIABA091-14, MG025937 to MG025944, and MN011189-MN011202 from Barreto, Nunes, et al., 2016, Barreto et al., 2018, and Barreto et al., 2020, respectively; see Table S1 for further details), totaling 295 samples for the COI dataset (274 for *N. venustus* and 21 for *N. varii*).

Analyses of mitochondrial DNA

A time-calibrated phylogeny was estimated in BEAST 1.8.4 (Drummond et al., 2012) to assess the timing of lineage diversification and test whether it is consistent with the Pleistocene, during which glacial cycles reduced the sea level by up to 130 m below the current level in the Last Glacial Maximum (LGM) (Clark et al., 2009). In addition to the sequences of *Nematocharax*, we included as outgroups sequences representing three genera closely related to it (Oliveira et al., 2011), which are: *Hasemania nana* (acc. num. NC_022724), *Hemigrammus marginatus* (acc. num. HM906014), and *Moenkhausia costae* (acc. num. HM405163). The HKY+I+G substitution model was selected in jModelTest 2 (Darriba et al., 2012) according to the Akaike Information Criterion (AIC). We used a substitution rate for fish mtDNA (1% per Myr; e.g. Thomaz et al., 2015) under a strict clock model and speciation Yule process as a tree prior to perform the analysis, considering the presence of different species in the dataset and highly differentiated mtDNA lineages within *Nematocharax*, some of them already known from Barreto, Nunes et al. (2016) and Barreto et al. (2018). Two independent Markov Chain Monte Carlo (MCMC) runs were carried out,

each consisting of 100 million generations, sampling every 1,000 generations, with the first 10% of the runs excluded as burn-in. Convergence was checked using Tracer 1.7.1 (Rambaut et al., 2018), which we also used to ensure that all Effective Sample Size (ESS) values were >200. Independent runs were combined in LogCombiner 1.8.4, and the Maximum Clade Credibility (MCC) tree was generated in TreeAnnotator 1.8.4 and then visualized in FigTree 1.4.3 (http://tree.bio.ed.ac.uk/software/figtree/).

We performed a Bayesian analysis of genetic differentiation among samples of *N. venustus* using the software BAPS 6.0 (Bayesian Analysis of Population Structure; Corander & Marttinen, 2006); *N. varii* was not included in the analysis due to insufficient geographic sampling. We conducted a population mixture analysis based on a "clustering with linked loci" setting the maximum number of clusters (*K*) to 15. In addition, for each mitochondrial lineage, we calculated the nucleotide diversity (π) and number of haplotypes in DnaSP 6 (Rozas et al., 2017) and evaluated putative signs of demographic expansion using Tajima's *D* (Tajima, 1989) and Fu's F_s (Fu, 1997), with default settings. The significance of the tests was obtained from 1,000 coalescent simulations.

Genomic data generation and assembly

Genomic data for a subset of 23 representative individuals of the 10 mtDNA lineages (see Results) were generated using RADseq. Specifically, the ezRAD libraries (one per sample) were prepared according to Toonen et al. (2013) and Knapp et al. (2016) from total DNA extracted from muscle tissues using the DNeasy Blood & Tissue kit (Qiagen, Hilden, Germany). Briefly, the DNA was digested with the *Dpn*II restriction endonuclease (New England Biolabs, Ipswich, USA) and purified by adding a 1.8X volume of AMPure XP magnetic beads (Beckman Coulter, Brea, USA), which were also used to size select 150-350 bp DNA fragments. A-tailing, adapter ligation, enrichment by PCR, and additional clean-up steps were performed using the TruSeq Nano DNA HT Library Prep kit (Illumina, San Diego, USA). All libraries were quantified on a Qubit 2.0 Fluorometer using the Qubit dsDNA BR Assay kit (Thermo Fisher Scientific, Waltham, USA) and an Agilent Bioanalyzer 2100 using DNA 1000 chips (Agilent Technologies, Santa Clara, USA). Paired-end sequencing (2x75 bp) was performed on an Illumina NextSeq 500 sequencer at the Genome Investigation and Analysis Laboratory (GENIAL) core facility (CEFAP-USP, São Paulo, Brazil), using a Mid-Output v2 kit with 150 cycles.

After sequencing, the raw reads of each individual were demultiplexed using blc2fastq 1.8.4 (Illumina; http://support.illumina.com/downloads.html) and processed with ipyrad version 0.7.28 (Eaton, 2014; http://ipyrad.readthedocs.io/). Filtered reads were assembled *de novo* with a clustering threshold of 0.85 and a minimum read depth of 6X, with the most stringent filtering to remove Illumina adapters, and 15 bp trimmed from the 3' ends to reduce error rates due to low quality of reads (Del Fabbro et al., 2013) (see Table S2 for more details on the parameters used in ipyrad). The RADseq data resulted in a total of 19,099 filtered loci and 61,696 SNPs, excluding loci with >50% missing data. The number of filtered reads per sample varied from 1,609,668 to 6,021,549 (Table S3). Individuals from the Jequitinhonha River basin were poorly sequenced and therefore were not included in the genomic analyses.

Analyses of RADSeq data

For the genomic data, we estimated phylogenetic relationships using the concatenated dataset, as well as using the multispecies coalescent (MSC) model to accommodates different genealogical histories among loci due to phenomena such as incomplete lineage sorting (ILS; Rannala et al., 2020). The concatenated data were used to conduct a Maximum Likelihood (ML) analysis under a GTRAC model in RAxML 8.2.10 (Stamatakis, 2014) and a Bayesian analysis in MrBayes 3.2.6 (Ronquist et al., 2012), both through the CIPRES Science Gateway Portal (Miller et al., 2010). For the Bayesian inference, the best model of nucleotide evolution was inferred as GTR+I+G with jModelTest 2. We performed four million generations using two runs of four chains each sampled every 500 generations, with a 10% burn-in. On the other hand, the MSC analysis was conducted in SVDquartets (Chifman & Kubatko, 2014) implemented in PAUP* 4.0 (Swofford, 2003) using a single SNP per locus.

The analysis of population structure within *N. venustus* was conducted using the sparse non-negative matrix factorization (sNMF) algorithm (Frichot et al., 2014). We tested *K*-values (number of ancestral populations) ranging from 1 to 15, running 200 replicates for each *K* under four different alpha regularization parameter values (i.e., 1, 10, 100, and 1000) to test the robustness of the results. The cross-entropy criterion was used to select the best *K*-value.

To test which historical demographic scenario is more likely to have occurred with the lineages of *Nematocharax*, we used the coalescent-based program fastsimcoal2 (Excoffier et al., 2013) in order to compare divergence models with and without gene flow. Despite the difficulties in identifying genetic signatures of drainage rearrangement events (Souza et al.,

2020), these models can provide insights into the effects of gene flow on populations, allowing to test whether the lineages have been isolated for long periods of time or historical gene flow has occurred among currently isolated river basins, which may indicate that drainage rearrangements provided effective connections between watersheds.

Because fastsimcoal2 requires no missing data, the tests were conducted in separate analyses of four datasets containing different subsets of individuals (Table S4). Specifically, pairwise divergences were estimated for each of the four divergences based on the relationships from the MSC tree estimated with SVDquartets. Distinct assemblies were made in ipyrad for each of the four datasets (i.e., for individuals from the: i-ii, iii-iv, v-vi, and vii-viii lineages; see Results) retaining only loci without missing data (details on the ipyrad parameters are reported in Table S2). The joint Site Frequency Spectrum (SFS) was obtained using the Python script 'easySFS.py' (http://github.com/isaacovercast/easySFS) for each VCF file for the four datasets. To improve model performance, the effective population size (N_e) of the first lineage of each pair was fixed based on π (π=4Neμ) (calculated in DnaSP 6, based on all variant and invariant sites) (Excoffier & Foll, 2011). The other estimated parameters were the N_e of the second lineage of the pair, the ancestral population size (N_{ANC}), and the divergence time (T_{DIV}) for the scenario of strict divergence. In turn, for the scenario of divergence with gene flow, two additional parameters were estimated: the migration rates backward in time from population 1 to population 2 (MIG_1) and from population 2 to population 1 (MIG_2).

For both coalescent models (divergence with and without gene flow), we assumed a generation time of one year and a mutation rate (μ) of 2.24×10^{-8}, as estimated for a closely related species (Thomaz et al., 2017). This mutation rate is appropriate for *Nematocharax* because it was calculated from the regression formula for cellular organisms (Lynch, 2010) based on the average genome size of Characidae 'clade C', where *Nematocharax* is positioned (Thomaz et al., 2010). A total of 40 replicates were run for each dataset and model, with 250,000 simulations per likelihood estimate, a stopping criterion of 0.001, and 10-40 expectation-conditional maximization (ECM). We obtained 95% confidence intervals of parameters from 100 parametric bootstraps by simulating 100 SFS from the maximum likelihood estimates and re-estimating parameters with 40 runs for each SFS. The best model for each lineage pair was selected based on the AIC values calculated using the R script 'calculateAIC.sh' (http://github.com/speciationgenomics/scripts/blob/master/calculateAIC.sh).

Inferred drainage rearrangements from bathymetric and geomorphological data

To test for a correspondence between genetic structure and the current configuration of basins or historical drainage rearrangements due to headwater captures and coastal paleodrainages, we applied GIS techniques to infer putative drainage rearrangements between the NMA river basins. We reconstructed the putative paleodrainages that would have formed during periods of low sea level using bathymetric and topographic data at 30 arc seconds resolution obtained from the General Bathymetric Chart of the Oceans (GEBCO, http://www.gebco.net/). These data were processed using the Hydrological tools in ArcGis10 following the methods described in Thomaz et al. (2015) (see also the reconstruction in Thomaz & Knowles, 2018 for the entire Brazilian coast). Using the inferred connections among currently isolated rivers implied by the coastal paleodrainage reconstruction, we verified whether populations sampled in contemporary rivers within a paleodrainage are genetically more similar to each other than populations from rivers located in different paleodrainages.

We also inferred putative headwater captures in the Central region of the NMA following the methodology described in Barreto et al. (2020), which uses the QGIS software to analyze geological and geomorphological information that, according to de Oliveira (2010), suggest the occurrence of this type of drainage rearrangement, particularly: (1) elbows of capture (abrupt changes in drainage direction at approximately 90°; Bishop, 1995) detected from the Continuous Cartographic Base of the Brazilian hydrography at 1:250,000 scale (DGC, 2017); (2) wind gaps (dry valleys that cross watershed divides and correspond to ancient river beds; Ollier & Pain, 2000) identified from the SRTM 90 m Digital Elevation Data (Jarvis et al., 2008); and (3) geological faults (areas of possible tectonic reactivation; de Oliveira, 2010) obtained from the SD.24 Salvador sheet (DGC, 2016). These three data sources were analyzed in QGIS 3.4.1 (QGIS Development Team, 2019) and, using the inferred headwater captures, we evaluated whether there is a spatial correspondence between the drainage rearrangements and the distribution of the *Nematocharax* clusters recovered by the mtDNA and genomic data.

RESULTS

Spatial and temporal inferences based on mtDNA

The alignment of 295 COI sequences resulted in 46 haplotypes (93 variable sites) distributed in nine mitochondrial lineages of *N. venustus* and one lineage of *N. varii*, according to the BAPS analysis (Figures 2 and 3a). Based on the time-calibrated mtDNA phylogeny (Figure 2), the earliest divergence between *N. varii* (lineage 10) and *N. venustus* (lineages 1-9) date to the late Pliocene (2.88 Mya; 95% HPD = 3.67-2.16 Mya), while divergences within *N. venustus* date to the Pleistocene. Using the BAPS assignment at the tips of the mitochondrial time-calibrated tree, we found reciprocal monophyly for most lineages, except mtDNA lineage 4 (see Figure 2) that was recovered as paraphyletic but with low support values (<0.7). Genetic diversity varied across mitochondrial lineages (Table S5), with the highest haplotype diversity being found in lineages 1 and 8 (13 and 12 haplotypes, respectively), whereas some lineages were fixed for the same haplotype (i.e., lineages 3 and 10). In addition, significant negative Tajima's *D*-values and Fu's F_s were detected in five lineages (Table S5).

Geographically, we found mtDNA lineages restricted to particular basins (lineage 1 from Contas; lineages 2 and 3 from Almada; lineages 4 and 9 from Cachoeira/Santana; and lineages 6 and 7 from Pardo) and others shared among basins (lineage 5 from Cachoeira and Una, and lineage 8 from Contas, Pardo, and Jequitinhonha; see Figure 1). Regarding the watersheds with multiple mtDNA lineages, our results also show geographically structured populations within basins (in the upstream versus downstream regions of the Contas, Cachoeira, and Pardo river basins). Lineage 10 isolated in the Upper Contas River basin is the only one that corresponds to the species *N. varii* (Figure 1).

Genomic structure and diversification history

The sNMF analysis of genomic data (Figure 3b) indicated three groups ($K = 3$) as the best clustering for *N. venustus*. These groups encompass different mitochondrial lineages with evidence of admixture among them. The first group (light gray) comprises mtDNA lineages 2, 3, 4, and 5, corresponding to the small coastal basins of Almada, Cachoeira/Santana, and Una. Lineage 5 is the only one in this group that occurs in more than one river basin (i.e., Cachoeira and Una). The second group (dark gray) includes mtDNA lineages 7 and 8, from the large basins of Contas, Pardo, and Jequitinhonha. In this case, lineage 8 shows a wider distribution, occurring in these three basins. Lastly, the third group (black) contains the mtDNA lineage 9, which is restricted to the Santana River, a small tributary of the Cachoeira River basin (Figure 3b; also see map in Figure 1). Interestingly,

one sample collected in the Santana River (sample code 0684; Figures 3 and S1), although it presented a higher ancestry coefficient for this group (black), differed from the other individuals from that locality for presenting the mtDNA lineage 4 (which was detected in another portion of the Cachoeira River basin). The remaining individuals for which we obtained RADseq data correspond to mtDNA lineages 1 and 6, which exhibited the highest levels of admixture among the three recovered genomic groups.

Regarding the evolutionary relationships from the MSC tree (Figure 3c), we observed river basins characterized by reciprocally monophyletic groups (e.g. Almada) and river basins where lineages have more than one origin (e.g. Contas in green and Cachoeira in orange), some of them occurring in sympatry (see Figure 1). Both ML and Bayesian phylogenetic trees using concatenated genomic data recovered the mtDNA lineages with high support values, except for lineages 2/3 and 4/9 from Almada and Santana rivers, respectively, which were grouped within mixed clusters (Figure S1).

Our divergence model tests showed which scenario best fits the data of each lineage pair within *Nematocharax* (see Figure 3c and Table 1). Specifically, divergence without gene flow (i.e., strict divergence) showed a better fit than divergence with gene flow for pair 1 (i-ii; divergence between *N. varii* and *N. venustus*) and pair 4 (vii-viii; divergence between small coastal basins – Almada, Cachoeira, and Una – and the Contas). In the first case (pair 1), we recovered much older divergence times between the two species of *Nematocharax* than those estimated by mtDNA. On the other hand, pair 4 represents the most recent divergence within *N. venustus*, during the Late Pleistocene (~17 kya). With respect to pair 2 (iii-iv; divergence between the lineage from the Cachoeira/Santana River basin and the others) and pair 3 (v-vi; divergence between the lineage from the large river basins – Contas, Pardo, and Jequitinhonha – and the remaining groups), divergence with gene flow was inferred as the best fit model (Table 1). When analyzing the migration rates and direction of movement across drainages (Table 1; Figure 3c), we observe that they were probably asymmetric over time, occurring mainly from small coastal basins to the Santana River and large river basins.

Comparison between genetic structure and drainage rearrangements

The reconstruction of coastal paleodrainages for the Central NMA shows that the past and current configuration of rivers are largely congruent (see Figure 1; also see Thomaz & Knowles, 2018). One exception is the Una and Pardo river basins, which were connected on

the exposed continental shelf during sea level retreats. Our paleodrainage reconstruction also revealed an additional paleo-connection within the Cachoeira River basin, between the main river and the Santana River, which are currently connected only by the estuary (Figure 1).

Our geomorphological analysis identified nine putative headwater captures between drainages across the distribution of *Nematocharax* (Figures 1 and 4). Specifically, these points represent past connections between: Contas and Almada (captures a and b), Almada and Cachoeira (capture c), Contas and Cachoeira (captures d and e), Contas and Pardo (capture f), Cachoeira and Pardo (capture g), Cachoeira and Una (capture h), and Pardo and Jequitinhonha (capture i). This type of event may have occurred multiple times during the geological history of these basins, allowing dispersal/isolation of fish lineages throughout the area.

The comparison between the genetic structure found in *N. venustus* and the location of putative drainage rearrangements shows that both mitochondrial and genomic data (see Figure 3) do not reflect the reconstructed coastal paleo-connection between Una and Pardo river basins (Figure 1) because the populations sampled within this paleodrainage do not share a common ancestry. On the other hand, the presence of one sample in the Santana River basin (sample code 0684) that has evidence of genomic admixture and belongs to the mtDNA lineage 4 corroborates the more extensive coastal paleo-connection between the Santana and Cachoeira rivers that allowed dispersal of freshwater fish across the continental shelf (Figures 1 and 3). Likewise, we observed the best fit of the 'divergence with gene flow' model for this pair (pair 2; Table 1). Note that because of the match between current basin boundaries and coastal paleodrainage boundaries (Figure 1) the effect of the remaining coastal paleodrainages cannot be tested.

Among the nine inferred headwater captures (Figures 1 and 4), three of them are supported by the geographic distribution of mtDNA lineages (specifically, headwater captures labeled f, h, and i), whereas four putative past connections are suggested by the genomic clusters; headwater captures c and h (Figure 1) would explain the composition of the light gray cluster (Figure 3b), which includes the small coastal basins of Almada, Cachoeira, and Una, and headwater captures f and i (Figure 1) help understand the geographic distribution of the dark gray cluster (Figure 3b), including the secondary contact of lineages (1-8) in the Contas River basin. Our RADseq data indicate that both the Contas and Cachoeira river basins (green and orange, respectively) were colonized twice (see Figure 3c).

DISCUSSION

The geographic isolation of river basins within the NMA has no doubt been a major factor in promoting differentiation, as evidenced by the clear spatial structuring of genetic diversity in *Nematocharax*. However, our analyses suggest a dynamic history in which past connections among these currently isolated basins have also shaped the geographic structuring of genetic variation within *Nematocharax*. Our work then shows that paleo-landscapes, as with terrestrial organisms, are an important factor in understanding the current distribution of genetic structure in riverine fishes, especially considering the obvious constraints imposed by a freshwater lifestyle. Additionally, as predicted by some models that consider the river architecture (see Thomaz et al., 2016), we find evidence of genetic differentiation within river basins, which illustrates the effect of the complex architecture of rivers on the evolutionary dynamics of local populations.

The role of drainage rearrangements in *Nematocharax* diversification

Our results indicate a mixed history for the central basins of the NMA ecoregion, in which long-term isolation among watersheds largely dictated the genetic structuring in *Nematocharax*, but drainage rearrangements that allowed historical movement across drainage divides probably also left a noticeable signature on patterns of genetic divergence. It is known that other geomorphological processes can facilitate dispersal of organisms between basins, such as divide inundation during flooding and swamps on low drainage divides that form intermittent wet connections (Burridge et al., 2008). However, this may not be the case in the NMA, because tectonic reactivation of ancient faults and erosive processes are known to have caused several events of headwater capture along the eastern margin of the Brazilian crystalline shield, where the watershed divides are generally formed by mountainous landscapes (Ribeiro, 2006; Buckup, 2011). It is estimated that tectonic reactivations in this region are as recent as <1.6 Mya (Saadi et al., 2002).

In this context, the first divergence within the genus (between *N. varii* and *N. venustus*) took place in the Mio-Pliocene, a period during which tectonic events reactivated old faults in the Espinhaço Range, where the Upper Contas River basin is located (Saadi, 1995). Given the geological and geomorphological processes related to the mountainous relief in the Upper Contas River, the evolutionary history of *Nematocharax* is perhaps marked by the complete isolation of *N. varii* in the highlands of the Chapada Diamantina

(Barreto, Nunes et al., 2016), which explain the best fit of the 'divergence without gene flow' model for pair 1.

Within *N. venustus*, we identified mtDNA lineages distributed across currently isolated basins and divergent genomic lineages within a single basin (e.g. those associated with the river basins in green and orange). These divergence histories probably took place in the Pleistocene, although estimates based on COI sequences differed significantly from those based on RADseq data. This is not unexpected given the methodological differences between divergence time estimates of calibrated phylogenetic trees and coalescent-based models (Edwards & Beerli, 2000; Carsten & Knowles, 2007). Regarding the Santana River, in the Cachoeira River basin (in orange), we found two divergent lineages in both mitochondrial (lineages 4 and 9) and genomic dataset, thus likely reflecting the effect of glacial cycles that resulted in repeated sea level shifts and, consequently, repeated events of colonization through coastal paleodrainages. Currently, the connection between the Cachoeira and Santana rivers is possible only by the estuary, which represents an effective barrier to dispersal of obligate freshwater fish from one river to another. This seems to be the case of *Nematocharax*, given its little or no tolerance for saltwater (Wilzbach & Cummins, 2008). Thus, sea level fluctuations during the Pleistocene may have allowed dispersal from the Cachoeira to Santana River due to the expansion of this freshwater connection in periods of sea regression, with subsequent isolation in periods of transgression. This hypothesis is reinforced by the best fit of the 'divergence with gene flow' model (pair 2), in which the migration rate was higher to the Santana River, in addition to the geomorphological evidence of paleo-connection between the Cachoeira and Santana rivers on the continental shelf. We also found one individual in the Santana River whose mtDNA is more related to individuals from other portions of the Cachoeira River basin.

For the remaining river basins, the paleodrainage reconstruction reveals limited connectivity between currently isolated rivers systems during sea level retreats because of the relatively narrow continental shelf of the Central NMA, whereas the evidence of headwater captures is widely distributed in the studied area. These results agree with the assumption that the narrower and shallower northeastern Brazilian continental shelf would have imposed conspicuous geographic isolation and eventually high levels of endemism in freshwater and estuarine fish species (e.g. Baggio et al., 2017; Thomaz & Knowles, 2018; Argolo et al., 2020). Therefore, putative dispersal routes via headwater captures are presented herein based on evidence of elbows of capture (sudden shifts in the course of a river), wind gaps (dry valleys once occupied by a river), and geological faults (areas possibly associated with

tectonic reactivations; see de Oliveira, 2010; Barreto et al., 2020). The signs of demographic expansion for five mtDNA lineages also support that headwater capture events provided opportunities for population expansion (Burridge et al., 2006; Waters et al., 2020).

By comparing the genetic structure recovered by mtDNA and the genomic loci, we can assume that past connections allowed mitochondrial movement, and the paths of genomic divergence were also redirected. For instance, individuals of *N. venustus* with morphological and genetic differences were found at the point of sympatry between mtDNA lineages 1 and 8 (see sampling point '1*8' in the Contas River basin; Barreto et al., 2020). This locality is close to the division between the Contas and Pardo river basins, where a putative headwater capture was suggested (see capture f), which may have allowed the secondary contact and hybridization between these lineages (Barreto et al., 2020), with strong support for a population expansion in lineage 8. In fact, our divergence model test recovered 'divergence with gene flow' for pair 3, which is in accordance with the hypothesis of connection between these groups.

Influence of past and current riverscape on genetic structuring

Given that shared lineages are found among small coastal basins (i.e., Almada, Cachoeira, and Una) as well as among large basins (i.e., Contas, Pardo, and Jequitinhonha), it suggests that past connections along the Central NMA ecoregion fostered widespread movement (i.e., they are not restricted geographically). If we consider that the elbow of capture is close to the captor river (Bishop, 1995), our inferred headwater captures show that most captor rivers belong to the small coastal basins (see captures a, b, c, d, and h), which agrees with the direction of movement inferred for the lineage pairs. Moreover, the close relationship recovered among basins is mirrored by faunal distribution patterns in other fish species, including groups with different ecologies (e.g. Camelier & Zanata, 2014; Rodrigues et al., 2016; de Sousa et al., 2021). This congruence in distribution patterns give insights about the past configuration of the NMA rivers, suggesting that historical connections may have structured communities more generally (i.e., the temporary connections are not species-specific).

Despite support for nine putative headwater captures, correspondence between geomorphological and genetic data was not observed for all putative headwater captures, meaning that not all putative headwater captures appear to have facilitated movement. This can be explained by a possible mismatch in timing between the occurrence of the river

capture event and the presence of fish. Dispersal may also simply not be likely depending on the ephemerality of headwater capture events or species dispersal ability. In either case, future investigations with other fish species will be important to test the generality of these potential explanations. This is relevant because, although the drainage rearrangements presented here may have similarly influenced the evolution of other co-distributed fish species, whether temporary connections are available to fish may depend upon species-specific ecologies (Tschá et al., 2017; Thomaz & Knowles, 2020; Waters et al., 2020). Particularly in *Nematocharax*, some biological features such as restricted dispersal capacity typical of small-sized characins (Lucas & Baras, 2001) and the potential influence of sexual selection (Menezes et al., 2015; Barreto, Nunes, et al., 2016) may help explain the occurrence of endemic groups and significant genetic differentiation among lineages in different river basins, despite the evidence of temporary connections.

Although little is known about the *Nematocharax* biology, field and aquarium observations indicate that territorial behavior can be observed especially in males (Barreto et al., 2020), and the species probably occupies microenvironments within streams, completing its life cycle in restricted geographic areas (Cetra et al., 2009). These features help understand the reduced gene flow between populations, with local structuring and co-occurrence of distinct lineages within river basins, as in the Cachoeira and Pardo. This structuring is accentuated by the characteristics of the riverscape, such as the differential degree of branching and basin slope at different portions of the river, with direct impacts on travel distances and dispersal costs for fishes (Campbell Grant et al., 2007; Carvajal-Quintero et al., 2019). Given that long-term population persistence is mainly dependent on connectivity, investigating how riverscape heterogeneity influences spatial genetic structure (see Thomaz et al., 2016) is essential to infer the evolutionary consequences of habitat loss and fragmentation (Davis et al., 2018), factors that have continuously threatened the biodiversity in basins along the Atlantic coast of Brazil (Menezes et al., 2007; Nogueira et al., 2010). Moreover, elucidating the historical interconnectedness of these isolated basins becomes a source of information for guiding conservation efforts, especially because the boundaries of unique evolutionary units may not necessarily correspond to the current configuration of basins.

CONCLUSIONS

Our study shows that a hidden genetic diversity in an endemic fish group has been shaped by the isolated basins within the NMA ecoregion, as well as by some historical temporary connections forged between them. However, in general, these river basins have their own independent evolutionary histories, according to the geographic distribution of *Nematocharax* lineages, with a few notable exceptions in which drainage rearrangements structured divergence histories that are estimated here. These results contrast with genetic studies of Neotropical fish from southern and southeastern Brazilian coastal basins (e.g. Thomaz et al., 2015), reflecting the particular features of the NMA region, especially its narrow continental shelf. As such, our findings demonstrate the importance of integrating different data sources to generate hypotheses about when and where connections formed between isolated river systems, and how they may have structured evolutionary divergence and diversification of aquatic lineages.

ACKNOWLEDGEMENTS

We thank Angela Zanata (UFBA), Priscila Camelier (UFBA), and Daniel Carvalho (PUC Minas) for providing some tissue samples used in this study, and the Knowles Lab for hosting SSB at the University of Michigan, as well as providing computational resources. We also thank the Rede de Plataformas Tecnológicas for the use of its Sequencing Facility in FIOCRUZ-Bahia, the Core Facility for Scientific Research – University of São Paulo (CEFAP-USP/GENIAL) for the use of its Illumina NextSeq sequencer, and the National Laboratory for Scientific Computing (LNCC/MCTI, Brazil) for providing HPC resources of the SDumont supercomputer, which have contributed to the research results reported within this paper. This study was financially supported by CAPES (Finance Code 001, 23038.000776/2017-54, PDSE 88881.186858/2018-01, and 88882.453923/2019-01), FAPESB (RED0045/2014 and JCB0026/2016), and CNPq (465767/2014-1). HB-F acknowledges the CNPq Research Productivity fellowship (307037/2018-5). SBB thanks CAPES for the postdoctoral fellowship (grant 88887.474651/2020-00).

DATA AVAILABILITY STATEMENT

COI sequences: GenBank accession numbers MN011168-MN011188 and MN011203-MN011363.

CONFLICT OF INTEREST STATEMENT

The authors have no conflicts of interest to declare.

REFERENCES

Abell, R., Thieme, M. L., Revenga, C., Bryer, M., Kottelat, M., Bogutskaya, N., … Petry, P. (2008). Freshwater ecoregions of the world: a new map of biogeographic units for freshwater biodiversity conservation. *BioScience*, 58(5), 403–414. DOI: 10.1641/B580507

Albert, J. S., & Reis, R. E. (2011). Historical biogeography of Neotropical freshwater fishes. Berkeley, CA: University of California Press.

Albert, J. S., Craig, J. M., Tagliacollo, V. A., & Petry, P. (2018). Upland and lowland fishes: a test of the river capture hypothesis. Chapter 19. In: Mountains, climate and biodiversity. (Eds: C. Hoorn, A. Perrigo & A. Antonelli). Wiley-Blackwell, New York.

Albert, J. S., Petry, P., & Reis, R. E. (2011). Major biogeographic and phylogenetic patterns. Chapter 2. In: Historical biogeography of Neotropical freshwater fishes. (Eds: J. S. Albert & R. E. Reis). University of California Press, Berkeley.

Argolo, L. A., López-Fernández, H., Batalha-Filho, H., & Affonso, P. R. A. M. (2020). Unraveling the systematics and evolution of the '*Geophagus*' *brasiliensis* (Cichliformes: Cichlidae) species complex. *Molecular Phylogenetics and Evolution*, 150, 106855. DOI: 10.1016/j.ympev.2020.106855

Baggio, R. A., Stoiev, S. B., Spach, H. L., & Boeger, W.A. (2017). Opportunity and taxon pulse: the central influence of coastal geomorphology on genetic diversification and endemism of strict estuarine species. *Journal of Biogeography*, 44(7), 1626–1639. DOI: 10.1111/jbi.12934

Barreto, S. B., Cioffi, M. B., Medrado, A. S., Silva, A. T., Affonso, P. R. A. M., & Diniz, D. (2016). Allopatric chromosomal variation in *Nematocharax venustus* Weitzman, Menezes & Britski, 1986 (Actinopterygii: Characiformes) based on mapping of repetitive sequences. *Neotropical Ichthyology*, 14(2), e150141. DOI: 10.1590/1982-0224-20150141

Barreto, S. B., Knowles, L. L., Affonso, P. R. A. M., & Batalha-Filho, H. (2020). Riverscape properties contribute to the origin and structure of a hybrid zone in a Neotropical freshwater fish. *Journal of Evolutionary Biology*, 33(11), 1530–1542. DOI: 10.1111/jeb.13689

Barreto, S. B., Nunes, L. A., Silva, A. T., Jucá-Chagas, R., Diniz, D., Sampaio, I., ... Affonso, P. R. A. M. (2016). Is *Nematocharax* (Actinopterygii, Characiformes) a monotypic fish genus?. *Genome*, 59(10), 851–865. DOI: 10.1139/gen-2015-0166

Barreto, S. B., Silva, A. T., Batalha-Filho, H., Affonso, P. R. A. M., & Zanata, A. M. (2018). Integrative approach reveals a new species of *Nematocharax* (Teleostei: Characidae). *Journal of Fish Biology*, 93(6), 1151–1162. DOI: 10.1111/jfb.13834

Bishop, P. (1995). Drainage rearrangement by river capture, beheading and diversion. *Progress in Physical Geography*, 19(4), 449–473. DOI: 10.1177/030913339501900402

Buckup, P. A. (2011). The Eastern Brazilian Shield. Chapter 12. In: Historical biogeography of Neotropical freshwater fishes. (Eds: J. S. Albert & R. E. Reis). University of California Press, Berkeley.

Burridge, C. P., Craw, D., & Waters, J. M. (2006). River capture, range expansion, and cladogenesis: the genetic signature of freshwater vicariance. *Evolution*, 60(5), 1038–1049. DOI: 10.1111/j.0014-3820.2006.tb01181.x

Burridge, C. P., Craw, D., Jack, D. C., King, T. M., & Waters, J. M. (2008). Does fish ecology predict dispersal across a river drainage divide?. *Evolution*, 62(6), 1484–1499. DOI: 10.1111/j.1558-5646.2008.00377.x

Camelier, P., & Zanata, A. M. (2014). Biogeography of freshwater fishes from the Northeastern Mata Atlântica freshwater ecoregion: distribution, endemism, and area relationships. *Neotropical Ichthyology*, 12(4), 683–698. DOI: 10.1590/1982-0224-20130228

Camelier, P., Menezes, N. A., Costa-Silva, G. J., & Oliveira, C. (2018). Molecular and morphological data of the freshwater fish *Glandulocauda melanopleura* (Characiformes: Characidae) provide evidences of river captures and local differentiation in the Brazilian Atlantic Forest. *PLOS One*, 13(3), e0194247. DOI: 10.1371/journal.pone.0194247

Campbell Grant, E. H., Lowe, W. H., & Fagan, W. F. (2007). Living in the branches: population dynamics and ecological processes in dendritic networks. *Ecology Letters*, 10(2), 165–175. DOI: 10.1111/j.1461-0248.2006.01007.x

Carstens, B. C., & Knowles, L. L. (2007). Shifting distributions and speciation: species divergence during rapid climate change. *Molecular Ecology*, 16(3), 619–627. DOI: 10.1111/j.1365-294X.2006.03167.x

Carvajal-Quintero, J., Villalobos, F., Oberdorff, T., Grenouillet, G., Brosse, S., Hugueny, B., ... Tedesco, P. A. (2019). Drainage network position and historical connectivity explain

673 global patterns in freshwater fishes' range size. *Proceedings of the National Academy of*
674 *Sciences*, 116(27), 13434–13439. DOI: 10.1073/pnas.1902484116
675 Cetra, M., Ferreira, F. C., & Carmassi, A. L. (2009). Caracterização das assembléias de
676 peixes de riachos de cabeceira no período chuvoso na bacia do rio Cachoeira (SE da
677 Bahia, NE do Brasil). *Biota Neotropica*, 9(2), 107–115. DOI: 10.1590/S1676-
678 06032009000200010
679 Chifman, J., & Kubatko, L. (2014). Quartet inference from SNP data under the coalescent
680 model. *Bioinformatics*, 30(23), 3317–3324. DOI: 10.1093/bioinformatics/btu530
681 Clark, P. U., Dyke, A. S., Shakun, J. D., Carlson, A. E., Clark, J., Wohlfarth, B., ... McCabe,
682 A. M. (2009). The Last Glacial Maximum. *Science*, 325(5941), 710–714. DOI:
683 10.1126/science.1172873
684 Corander, J., & Marttinen, P. (2006). Bayesian identification of admixture events using
685 multilocus molecular markers. *Molecular Ecology*, 15(10), 2833–2843. DOI:
686 10.1111/j.1365-294X.2006.02994.x
687 Darriba, D., Taboada, G. L., Doallo, R., & Posada, D. (2012). jModelTest 2: more models,
688 new heuristics and parallel computing. *Nature Methods*, 9(8), 772–772. DOI:
689 10.1038/nmeth.2109
690 Davis, C. D., Epps, C. W., Flitcroft, R. L., & Banks, M. A. (2018). Refining and defining
691 riverscape genetics: how rivers influence population genetic structure. *Wiley*
692 *Interdisciplinary Reviews: Water*, 5(2), e1269. DOI: 10.1002/wat2.1269
693 de Oliveira, D. (2010). Capturas fluviais como evidências da evolução do relevo: uma revisão
694 bibliográfica. *Revista do Departamento de Geografia*, 20, 37–50. DOI:
695 10.7154/RDG.2010.0020.0003
696 de Sousa, J. L. P., Bitencourt, J. A., Sampaio, I., Schneider, H., & Affonso, P. R. A. M.
697 (2021). "More than meets the eye": phylogeographic inferences and remarkable cryptic
698 diversity and in endemic catfish *Parotocinclus* (Loricariidae: Hypoptopomatinae) from
699 neglected and impacted basins in South America. *Conservation Genetics*, 22, 411–425.
700 DOI: 10.1007/s10592-021-01336-3
701 Del Fabbro, C., Scalabrin, S., Morgante, M., & Giorgi, F. M. (2013). An extensive evaluation
702 of read trimming effects on Illumina NGS data analysis. *PLOS One*, 8(12), e85024. DOI:
703 10.1371/journal.pone.0085024
704 Dias, M. S., Oberdorff, T., Hugueny, B., Leprieur, F., Jézéquel, C., Cornu, ... Tedesco, P. A.
705 (2014). Global imprint of historical connectivity on freshwater fish biodiversity. *Ecology*
706 *Letters*, 17(9), 1130–1140. DOI: 10.1111/ele.12319

Diretoria de Geociências (IBGE/DGC) (2016). Falhas Geológicas da Folha SD.24 – Salvador. Retrieved from http://dados.gov.br/dataset/cren_geologiafalhasd24

Diretoria de Geociências (IBGE/DGC) (2017). BC250 - Base cartográfica contínua do Brasil - 1:250 000. Retrieved from http://geoftp.ibge.gov.br/cartas_e_mapas/bases_cartograficas_continuas/bc250/versao2017

Drummond, A. J., Suchard, M. A., Xie, D., & Rambaut, A. (2012). Bayesian phylogenetics with BEAUti and the BEAST 1.7. *Molecular Biology and Evolution*, 29(8), 1969–1973. DOI: 10.1093/molbev/mss075

Eaton, D. A. R. (2014). PyRAD: assembly of de novo RADseq loci for phylogenetic analyses. *Bioinformatics*, 30(13), 1844–1849. DOI: 10.1093/bioinformatics/btu121

Edwards, S. V., & Beerli, P. (2000). Perspective: gene divergence, population divergence, and the variance in coalescence time in phylogeographic studies. *Evolution*, 54(6), 1839–1854. DOI: 10.1111/j.0014-3820.2000.tb01231.x

Excoffier, L., & Foll, M. (2011). Fastsimcoal: A continuous-time coalescent simulator of genomic diversity under arbitrarily complex evolutionary scenarios. *Bioinformatics*, 27(9), 1332–1334. DOI: 10.1093/bioinformatics/btr124

Excoffier, L., Dupanloup, I., Huerta-Sánchez, E., Sousa, V. C., & Foll, M. (2013). Robust demographic inference from genomic and SNP data. *PLOS Genetics*, 9(10), e1003905. DOI: 10.1371/journal.pgen.1003905

Frichot, E., Mathieu, F., Trouillon, T., Bouchard, G., & François, O. (2014). Fast and efficient estimation of individual ancestry coefficients. *Genetics*, 196(4), 973–983. DOI: 10.1534/genetics.113.160572

Fu, Y. X. (1997). Statistical tests of neutrality of mutations against population growth, hitchhiking and background selection. *Genetics*, 147(2), 915–925.

Hales, J., & Petry, P. (2015). Northeastern Mata Atlantica ecoregion. Retrieved from http://www.feow.org/ecoregions/details/328

Hughes, J. M., Schmidt, D. J., & Finn, D. S. (2009). Genes in streams: using DNA to understand the movement of freshwater fauna and their riverine habitat. *BioScience*, 59(7), 573–583. DOI: 10.1525/bio.2009.59.7.8

Jarvis, A., Reuter, H. I., Nelson, A., & Guevara, E. (2008). Hole-filled seamless SRTM data V4. International Centre for Tropical Agriculture (CIAT). Retrieved from http://srtm.csi.cgiar.org

739 Knapp, I. S. S., Puritz, J., Bird, C., Whitney, J., Sudek, M., Forsman, Z., & Toonen R. (2016).
740 ezRAD - an accessible next-generation RAD sequencing protocol suitable for non-model
741 organisms_v3.2. Retrieved from http://doi.org/10.17504/protocols.io.e9pbh5n
742 Lima, S. M., Berbel-Filho, W. M., Araújo, T. F., Lazzarotto, H., Tatarenkov, A., & Avise, J.
743 C. (2017). Headwater capture evidenced by paleo-rivers reconstruction and population
744 genetic structure of the armored catfish (*Pareiorhaphis garbei*) in the Serra do Mar
745 mountains of southeastern Brazil. *Frontiers in Genetics*, 8, 199. DOI:
746 10.3389/fgene.2017.00199
747 Lima, S. M., Berbel-Filho, W. M., Vilasboa, A., Lazoski, C., de Assis Volpi, T., Lazzarotto,
748 H., ... Solé-Cava, A. M. (2021). Rio de Janeiro and other palaeodrainages evidenced by the
749 genetic structure of an Atlantic Forest catfish. *Journal of Biogeography*, 48(6), 1475–
750 1488. DOI: 10.1111/jbi.14091
751 Lucas, M., & Baras, E. (2001). Migration of freshwater fishes. Oxford, UK: Blackwell
752 Science Ltd.
753 Lynch, M. (2010). Evolution of the mutation rate. *Trends in Genetics*, 26(8): 345–352. DOI:
754 10.1016/j.tig.2010.05.003
755 Martins, L. R., & Coutinho, P. N. (1981). The Brazilian continental margin. *Earth-Science*
756 *Reviews*, 17(1-2), 87–107. DOI: 10.1016/0012-8252(81)90007-6
757 Menezes, N. A., & Lima, F. C. T. (2008). *Nematocharax venustus* Weitzman, Menezes &
758 Britski, 1986. In: Livro Vermelho da Fauna Brasileira Ameaçada de Extinção (Eds: A. B.
759 M. Machado, G. M. Drummond & A. P. Paglia). MMA, Brasília.
760 Menezes, N. A., Weitzman, S. H., Oyakawa, O. T., Lima, F. C. T. D., Correa e Castro, R. M.,
761 & Weitzman, M. J. (2007). Peixes de água doce da Mata Atlântica: lista preliminar das
762 espécies e comentários sobre conservação de peixes de água doce neotropicais. São Paulo,
763 SP: Museu de Zoologia da Universidade de São Paulo.
764 Menezes, N. A., Zanata, A. M., & Camelier, P. (2015). *Nematocharax costai* Bragança,
765 Barbosa & Mattos a junior synonym of *Nematocharax venustus* Weitzman, Menezes &
766 Britski (Teleostei: Characiformes: Characidae). *Zootaxa*, 3920(3), 453–462. DOI:
767 10.11646/zootaxa.3920.3.4
768 Miller, M. A., Pfeiffer, W., & Schwartz, T. (2010). Creating the CIPRES Science Gateway
769 for inference of large phylogenetic trees. In: Proceedings of the Gateway Computing
770 Environments Workshop. GCE, New Orleans.

Nogueira, C., Buckup, P. A., Menezes, N. A., Oyakawa, O. T., Kasecker, T. P., Neto, M. B. R., & da Silva, J. M. C. (2010). Restricted-range fishes and the conservation of Brazilian freshwaters. *PLOS One*, 5(6), e11390. DOI: 10.1371/journal.pone.0011390

Oliveira, C., Avelino, G. S., Abe, K. T., Mariguela, T. C., Benine, R. C., Ortí, G., ... Castro, R. M. C. (2011). Phylogenetic relationships within the speciose family Characidae (Teleostei: Ostariophysi: Characiformes) based on multilocus analysis and extensive ingroup sampling. *BMC Evolutionary Biology*, 11, 275. DOI: 10.1186/1471-2148-11-275

Ollier, C., & Pain, C. (2000). The origin of mountains. London, UK: Routledge.

QGIS Development Team (2019). QGIS Geographic Information System, Open Source Geospatial Foundation. Retrieved from http://qgis.osgeo.org

Rambaut, A., Drummond, A. J., Xie, D., Baele, G., & Suchard, M. A. (2018). Posterior summarization in Bayesian phylogenetics using Tracer 1.7. *Systematic Biology*, 67(5), 901–904. DOI: 10.1093/sysbio/syy032

Rannala, B., Edwards, S. V., Leaché, A., & Yang, Z. (2020). The Multispecies Coalescent Model and Species Tree Inference. Chapter 3.3. In: Phylogenetics in the Genomic Era. (Eds: C. Scornavacca, F. Delsuc, & N. Galtier), No commercial publisher, Authors open access book. Retrieved from https://hal.inria.fr/PGE

Ribeiro, A. C. (2006). Tectonic history and the biogeography of the freshwater fishes from the coastal drainages of eastern Brazil: an example of faunal evolution associated with a divergent continental margin. *Neotropical Ichthyology*, 4(2), 225–246. DOI: 10.1590/S1679-62252006000200009

Rodrigues, A. D. S., Brandão, J. H. S. G., Bitencourt, J. A., Jucá-Chagas, R., Sampaio, I., Schneider, H., & Affonso, P. R. A. M. (2016). Molecular identification and traceability of illegal trading in *Lignobrycon myersi* (Teleostei: Characiformes), a threatened Brazilian fish species, using DNA barcode. *The Scientific World Journal*, 2016, 9382613. DOI: 10.1155/2016/9382613

Ronquist, F., Teslenko, M., Van Der Mark, P., Ayres, D. L., Darling, A., Höhna, S., … Huelsenbeck, J. P. (2012). MrBayes 3.2: efficient Bayesian phylogenetic inference and model choice across a large model space. *Systematic Biology*, 61(3), 539–542. DOI: 10.1093/sysbio/sys029

Rozas, J., Ferrer-Mata, A., Sánchez-DelBarrio, J. C., Guirao-Rico, S., Librado, P., Ramos-Onsins, S. E., & Sánchez-Gracia, A. (2017). DnaSP 6: DNA sequence polymorphism analysis of large data sets. *Molecular Biology and Evolution*, 34(12), 3299–3302. DOI: 10.1093/molbev/msx248

Saadi, A. (1995). A geomorfologia da Serra do Espinhaço em Minas Gerais e de suas margens. *Geonomos*, 3(1), 41–63. DOI: 10.18285/geonomos.v3i1.215

Saadi, A., Machette, M. N., Haller, K. M., Dart, R. L., Bradley, L., & Souza, A. M. P. D. (2002). Map and database of Quaternary faults and lineaments in Brazil. U.S. Geological Survey Open-File Report 02-230. Version 1.0. Retrieved from http://pubs.usgs.gov/of/2002/ofr-02-230/

Schwarzer, J., Swartz, E. R., Vreven, E., Snoeks, J., Cotterill, F. P. D., Misof, B., & Schliewen, U. K. (2012). Repeated trans-watershed hybridization among haplochromine cichlids (Cichlidae) was triggered by Neogene landscape evolution. *Proceedings of the Royal Society B: Biological Sciences*, 279(1746), 4389–4398. DOI: 10.1098/rspb.2012.1667

Shelley, J. J., Swearer, S. E., Dempster, T., Adams, M., Le Feuvre, M. C., Hammer, M. P., & Unmack, P. J. (2020). Plio-Pleistocene sea-level changes drive speciation of freshwater fishes in north-western Australia. *Journal of Biogeography*, 47(8), 1727–1738. DOI: 10.1111/jbi.13856

Souza, M. S., Thomaz, A. T., & Fagundes, N. J. (2020). River capture or ancestral polymorphism: an empirical genetic test in a freshwater fish using approximate Bayesian computation. *Biological Journal of the Linnean Society*, 131(3), 575–584. DOI: 10.1093/biolinnean/blaa140

Stamatakis, A. (2014). RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics*, 30(9), 1312–1313. DOI: 10.1093/bioinformatics/btu033

Swofford, D. L. (2003). PAUP*: phylogenetic analysis using parsimony and other methods, version 4.0 b10. Sunderland, MA: Sinauer Associates.

Tajima, F. (1989). Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. *Genetics*, 123(3), 585–595.

Thomaz, A. T., & Knowles, L. L. (2018). Flowing into the unknown: inferred paleodrainages for studying the ichthyofauna of Brazilian coastal rivers. *Neotropical Ichthyology*, 16(3), e180019. DOI: 10.1590/1982-0224-20180019

Thomaz, A. T., & Knowles, L. L. (2020). Common barriers, but temporal dissonance: Genomic tests suggest ecological and paleo-landscape sieves structure a coastal riverine fish community. *Molecular Ecology*, 29(4), 783–796. DOI: 10.1111/mec.15357

Thomaz, A. T., Christie, M. R., & Knowles, L. L. (2016). The architecture of river networks can drive the evolutionary dynamics of aquatic populations. *Evolution*, 70(3), 731–739. DOI: 10.1111/evo.12883

Thomaz, A. T., Malabarba, L. R., & Bonatto, S. L. (2010). The phylogenetic placement of *Hollandichthys* Eigenmann 1909 (Teleostei: Characidae) and related genera. *Molecular Phylogenetics and Evolution*, 57(3), 1347–1352. DOI: 10.1016/j.ympev.2010.10.006

Thomaz, A. T., Malabarba, L. R., & Knowles, L. L. (2017). Genomic signatures of paleodrainages in a freshwater fish along the southeastern coast of Brazil: genetic structure reflects past riverine properties. *Heredity*, 119(4), 287–294. DOI: 10.1038/hdy.2017.46

Thomaz, A. T., Malabarba, L. R., Bonatto, S. L., & Knowles, L. L. (2015). Testing the effect of palaeodrainages versus habitat stability on genetic divergence in riverine systems: study of a Neotropical fish of the Brazilian coastal Atlantic Forest. *Journal of Biogeography*, 42(12), 2389–2401. DOI: 10.1111/jbi.12597

Thompson, J. D., Higgins, D. G., & Gibson, T. J. (1994). CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Research*, 22(22), 4673–4680. DOI: 10.1093/nar/22.22.4673

Toonen, R. J., Puritz, J. B., Forsman, Z. H., Whitney, J. L., Fernandez-Silva, I., Andrews, K. R., & Bird, C. E. (2013). ezRAD: a simplified method for genomic genotyping in non-model organisms. *PeerJ*, 1, e203. DOI: 10.7717/peerj.203

Tschá, M. K., Baggio, R. A., Marteleto, F. M., Abilhoa, V., Bachmann, L., & Boeger, W. A. (2017). Sea-level variations have influenced the demographic history of estuarine and freshwater fishes of the coastal plain of Paraná, Brazil. *Journal of Fish Biology*, 90(3), 968–979. DOI: 10.1111/jfb.13211

Ward, R. D., Zemlak, T. S., Innes, B. H., Last, P. R., & Hebert, P. D. (2005). DNA barcoding Australia's fish species. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 360(1462), 1847–1857. DOI: 10.1098/rstb.2005.1716

Waters, J. M., Burridge, C. P., & Craw, D. (2020). River capture and freshwater biological evolution: a review of galaxiid fish vicariance. *Diversity*, 12(6), 216. DOI: 10.3390/d12060216.

Weitzman, S. H., Menezes, N. A., & Weitzman, M. J. (1988). Phylogenetic biogeography of the Glandulocaudini (Teleostei: Characiformes, Characidae) with comments on the distributions of other freshwater fishes in eastern and southeastern Brazil. In: Proceedings

of Workshop on Neotropical Distribution Patterns. (Eds: P. E. Vanzolini & W. R. Heyer). Academia Brasileira de Ciências, Rio de Janeiro.

Wilzbach, M. A., & Cummins, K. W. (2008). Rivers and streams: Physical setting and adapted biota. In: Encyclopedia of Ecology. (Eds: S. E. Jørgensen & B. D. Fath). Elsevier, Oxford.

Winemiller, K. O., & Willis, S. C. (2011). The Vaupes Arch and Casiquiare Canal: Barriers and Passages. Chapter 14. In: Historical biogeography of Neotropical freshwater fishes. (Eds: J. S. Albert & R. E. Reis). University of California Press, Berkeley.

878 **FIGURE CAPTIONS**

879

880 **Figure 1.** Map of the Central region of the Northeastern Mata Atlantica (NMA) ecoregion
881 showing the collection sites of *Nematocharax* in the Contas, Almada, Cachoeira, Una, Pardo,
882 and Jequitinhonha river basins, which are color coded to demarcate their boundaries; small
883 coastal basins with no record of *Nematocharax* are shown in gray. The distribution of
884 different mitochondrial lineages (see Figures 2 and 3) is shown as numbers in black boxes;
885 note that an asterisk between numbers indicates collection sites with two mitochondrial
886 lineages (i.e., sympatry). Nine putative headwater captures inferred from geomorphological
887 data (which are detailed in Figure 4) are shown in red letters (a-i), and the reconstructed
888 coastal paleodrainages on the continental shelf during the period of largest drop in sea level
889 (LGM; up to 130 m) are shown by the dashed dark blue lines.

890

891 **Figure 2.** Time-calibrated Bayesian tree reconstructed by BEAST using COI sequences, with
892 lineages of *Nematocharax* color coded according to their basins (see map in Figure 1 for
893 details). Tips were collapsed into the 10 mitochondrial lineages according to the BAPS
894 assignment, with sample sizes (*n*) shown for each lineage. Numbers and bars on the nodes
895 show the estimated age and 95% highest probability density (HPD) intervals of the node age,
896 respectively.

897

898 **Figure 3.** Genetic structure within *N. venustus* based on (a) a Bayesian Analysis of
899 Population Structure (BAPS) using COI sequences ($n = 274$), with dashed lines separating
900 the nine mtDNA clusters (in which individuals are colored according to the river basins
901 where they were collected; Figure 1); b) *K*-genetic clusters ($n = 20$) inferred from the
902 genomic RADseq data with sNMF (light gray, dark gray, and black represent the three groups
903 recovered by the ancestry coefficients and which are separated by dashed lines; dashed lines
904 also show the correspondence between the genomic and mtDNA clusters. Some individuals
905 (those that correspond to mtDNA lineages 1 and 6) were not assigned to any specific
906 genomic cluster due to high levels of mixture (samples not delimitated by dashed lines); c)
907 evolutionary relationships among individuals ($n = 20$) recovered by SVDquartets under the
908 multispecies coalescent (MSC) model using genomic data; the groups used in separate
909 fastsimcoal2 analyses for divergence model tests are marked at the nodes (i.e., the roman

numerals i through viii). In the same way as BAPS (a), the tips of the tree were colored according to the river basins where the specimens were collected (see Figure 1).
*Sample collected in the Santana River with higher ancestry coefficient for the black group but with mtDNA lineage 4
**Individuals from Jequitinhonha were poorly sequenced and therefore were not included in the genomic analyses

Figure 4. Putative headwater captures in the Central portion of the Northeastern Mata Atlantica (NMA) ecoregion inferred from geomorphological data (i.e., a-i) and which are also marked in Figure 1 (red letters) for an overview of the entire NMA. In these maps, current rivers are in blue and red lines define the current watershed divides. The black and white lines mark the location of geological faults (i.e., tectonic reactivation areas), and low to high elevation is shown in dark to light shading, respectively. Wind gaps (i.e., dry valleys that cross watershed divides) are highlighted in yellow, showing how currently isolated river basins were connected in the past.

Table 1. Fastsimcoal2 results for the two tested divergence models (strict divergence and divergence with gene flow) for each lineage pair of Nematocharax (pairs i-ii, iii-iv, v-vi, and vii-viii; see Figure 3c) including the point estimate and 95% confidence interval in parentheses for each parameter. Specifically, the effective size of the first population of the pair (N_1, fixed value calculated in DnaSP 6 based on all variant and invariant sites), effective size of the second population of the pair (N_2), ancestral population size (N_{ANC}), divergence time (T_{DIV}), and migration rates backward in time from population 1 to population 2 (MIG_1) and from population 2 to population 1 (MIG_2) are reported. Also shown are the number of loci used to calculate the site frequency spectrum (SFS) of each pair, for which the most likely divergence model, according to the lowest Akaike Information Criterion (AIC) value, is highlighted in bold.

Lineage pair	Loci	Model	N_1 (fixed)	N_2	N_{ANC}	T_{DIV} (in generations)	MIG_1	MIG_2	AIC
Pair 1 (i-ii)	1,686	**Strict divergence**	3,414	**10,312,417 (5,597,783–6,142,663)**	**9,682,982 (9,551,329–10,417,296)**	**19,860,219 (6,052,532–6,955,249)**	**na**	**na**	**8514.61**
		Divergence with gene flow		5,676,862 (276,563–3,911,478)	4,608,404 (4,421,946–10,322,358)	13,081,732 (236,948–8,783,860)	1.48e-4 (3.40e-5–1.93e-4)	1.37e-4 (7.43e-4–0.10)	8659.22
Pair 2 (iii-iv)	873	Strict divergence	4,103	10,407,940 (3,900,968–4,716,178)	10,144,034 (8,702,434–9,234,337)	5,587,253 (1,601,304–2,061,663)	na	na	5760.33
		Divergence with gene flow		**557,737 (139,608–5,014,358)**	**10,315,367 (7,443,758–8,414,510)**	**9,092 (6,709–1,353,093)**	**0.17 (0.07–0.21)**	**5.58e-3 (2.04e-4–4.79)**	**3871.94**
Pair 3 (v-vi)	3,820	Strict divergence	16,781	7,442,637 (3,233,745–3,608,029)	10,131,818 (8,037,969–8,374,713)	3,445,100 (1,359,940–1,596,286)	na	na	40322.87
		Divergence		**8,839,016 (169,147–**	**10,294,640**	**161,609 (12,643–**	**0.63 (0.16–0.64)**	**0.03 (3.41e-4–**	**21202.70**

		with gene flow		**4,650,537)**	**(7,701,936–8,318,601)**	**1,382,357)**			**3.63)**
Pair 4 (vii-viii)	8,841	**Strict divergence**	15,517	**15,702 (14,791–17,040)**	**36,782 (34,846–38,255)**	**16,931 (15,975–17,956)**	**na**	**na**	**33650.76**