Riverscape properties contribute to the origin and structure of a hybrid zone in a Neotropical freshwater fish

Silvia Britto Barreto[1], L. Lacey Knowles[2], Paulo Roberto Antunes de Mello Affonso[3], Henrique Batalha-Filho[1]

[1]Institute of Biology, Federal University of Bahia, 147 Barão de Jeremoabo St., 40170-115, Salvador – BA, Brazil

[2]Department of Ecology and Evolutionary Biology, Museum of Zoology, University of Michigan, 1109 Geddes Ave., 48109, Ann Arbor – MI, USA

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

Running title: Neotropical freshwater fish hybrid zone

Corresponding author: Silvia Britto Barreto

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; http://orcid.org/0000-0001-8780-1959

Acknowledgments

The authors 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 financed in part by CAPES (Finance Code 001, PDSE proc. n. 88881.186858/2018-01, and proc. n. 23038.000776/2017-54), FAPESB (RED0045/2014 and JCB0026/2016), and CNPq (Research Productivity Fellowship n. 307037/2018-5 and proc. n. 465767/2014-1). The collection license (number 51856-2) was provided by the Instituto Chico Mendes de Conservação da Biodiversidade (ICMBio/SISBIO). This study was approved by the Ethics Committee of Utilization of Animals from the Universidade Estadual do Sudoeste da Bahia (CEUA/UESB, number 71/2014). All authors would like to thank Stuart J. E. Baird and two anonymous reviewers for their helpful comments that improved the manuscript. The authors have no conflicts of interest to declare.

DR SILVIA BRITTO BARRETO (Orcid ID : 0000-0001-8780-1959)

Article type : Research Papers

Abstract

Understanding the structure of hybrid zones provides valuable insights about species boundaries and speciation, such as the evolution of barriers to gene flow and the strength of selection. In river networks, studying evolutionary processes in hybrid zones can be especially challenging, given the influence of past and current river properties along with biological species-specific traits. Here, we suggest that a natural hybrid zone between two divergent lineages of the sexually dimorphic Neotropical fish Nematocharax venustus was probably established by secondary contact as a result of a river capture event between the Contas and Pardo river basins. This putative river capture is supported by hydrogeological evidence of elbows of capture, wind gaps, and geological faults. The morphological (color pattern) and genetic (mtDNA and RADseq) variation reveal a clinal transition between parental lineages along the main river, with predominance of F2 hybrids at the centre of the hybrid zone, absence of early generation backcrosses, and different levels of hybridization in the tributaries. We highlight that different sources of information are crucial for understanding how the riverscape spatial history influences the connectivity between and within rivers systems and, consequently, the dynamics of gene flow between freshwater lineages/species.

Keywords: Characidae, cline analysis, hybridization, RADseq, river capture, secondary contact

1. Introduction

As windows to the evolutionary consequences of genetic exchange between divergent lineages, hybrid zones can provide useful insights about the forces involved in the speciation process, such as the evolution of barriers to gene flow and the strength of selection (Hewitt, 1988; Gompert, Mandeville, & Buerkle, 2017). These zones can arise under a variety of scenarios, ranging from those originating within a continuous population in response to environmental differences or, more commonly, after secondary contact between formerly allopatric populations (Hewitt, 1988). A focus on the origin and structure of a hybrid zone forms the central framework for understanding the consequences of hybridization, including processes related to: (1) reinforcement of reproductive barriers; (2) introgression and fusion between parental groups; or (3) stability of the hybrid zone (Albert, Jónsson, & Bernatchez, 2006; Stewart, Hudson, & Lougheed, 2017).

Most hybrid zones consist of clines (i.e., gradients of variation) maintained by a balance between selection against hybrids within the zone and dispersal of parental individuals into the zone, hence the term "tension zone" (Barton & Hewitt, 1985). However, there are other theoretical models (e.g. bounded hybrid superiority and mosaic models) that seek to explain how these zones are maintained (see Arnold, 1997). The bounded hybrid superiority model is characterized by a smooth transition between parental groups throughout the hybrid zone, in which hybrids can exhibit equivalent or higher fitness compared to parental individuals in intermediate habitats (Moore, 1977). Mosaic hybrid zones, in turn, correspond to parental individuals adapted to different environments that are patchily distributed, and hybrid individuals occurring on the boundaries or in intermediate habitats (Harrison & Rand, 1989). Under these circumstances, endogenous selection (e.g. genetic incompatibilities) may act against hybrids and exogenous selection (e.g. environmental heterogeneity) may favor hybrids in the hybrid zone (De La Torre, Wang, Jaquish, & Aitken, 2014). Thus, each model encompasses a particular condition depending on how selection acts on parental and hybrid individuals, even though some hybrid zones can be represented by a mixture of models (Curry, 2015).

In river systems, applying such models to understand the structure of hybrid zones is particularly challenging because river networks are organized into a spatial hierarchy with downstream flow that makes equal (isotropic) dispersal in all directions unlikely (Hughes, 2007; Fullerton et al., 2010). In this sense, architectural and functional particularities of river

systems should be considered in evolutionary analyses of obligate freshwater organisms because these physical characteristics interact with the way organisms disperse through space thereby influencing both diversity within populations and differentiation among populations (Chaput-Bardy, Fleurant, Lemaire, & Secondi, 2009; Thomaz, Christie, & Knowles, 2016). For instance, although hybridization occurs widely in freshwater species, especially fishes (Scribner, Page, & Bartron, 2000), the position of populations within a river network and the geomorphological features within and between basins can affect the connectivity among populations and, consequently, the spatial distribution of genetic and/or phenotypic variation (Hughes, Schmidt, & Finn, 2009; Duvernell & Schaefer, 2014; Mandeville et al., 2017).

Here we analyze a combination of phenotypic and genomic data, providing an ideal opportunity for investigating the complexity of natural hybrid zones in riverscapes. Specifically, we focus on the sexually dimorphic characid fish Nematocharax venustus Weitzman, Menezes, & Britski, whose males have elongated rays in their dorsal, pelvic, and anal fins. The phenotypic and genetic variation found in this group shows a contact zone between lineages in drainages of the Gongogi River sub-basin, in the Contas River basin (Fig. 1), Northeastern Brazil. The sympatry of the two divergent mitochondrial DNA (mtDNA) lineages was first detected in the Cambiriba Stream, a tributary of the Gongogi River (Barreto et al., 2016). Interestingly, a new putative species (N. costai) was also described for the same locality (Bragança, Barbosa, & Mattos, 2013), although it was later synonymized with its only congener at the time (N. venustus) in a subsequent taxonomic treatment (Menezes, Zanata, & Camelier, 2015). The synonymization was justified by the authors because the species shared overlapped phenotypic features, including secondary sexual traits (Menezes et al., 2015). In sympatry, differences in secondary sexual traits can determine the evolution of reproductive barriers, given the direct link between these characters and processes such as species recognition and mate choice (Questiau, 1999). Therefore, without considering the dynamics of speciation in this putative hybrid zone, the taxonomic issue surrounding N. venustus and N. costai may reflect an incomplete assessment of the phenotypic geographic variation in Nematocharax.

With extended sampling efforts along the Gongogi River sub-basin, we quantify the spatial structure of this hybrid zone. Specifically, we (i) evaluate the extent of hybridization between the divergent populations, (ii) characterize how morphological (color pattern) and genetic variation is structured across the hybrid zone, including genomic data based on RADseq, (iii) verify whether there is coincidence and concordance among clines for the different datasets, and (iv) indirectly infer the strength of selection in the maintenance of this

hybrid zone based on cline analyses. We combined these results with geological and geomorphological data to infer the potential factors responsible for the origin and structure of this freshwater fish hybrid zone.

2. Materials and Methods

2.1. Sampling

We collected 193 individuals of Nematocharax venustus comprising the two previously identified divergent mtDNA lineages (Barreto et al., 2016; Barreto, 2019), hereafter called Northern and Southern lineages, in 14 locations along the Gongogi River sub-basin (Contas River basin); our sampling included the Cambiriba Stream (where the contact zone was first identified), the main river (Gongogi), and four other nearby tributaries (Fig. 1). All individuals were photographed alive in the field under standard conditions (see Fig. 2) to register their color pattern. A small fragment of muscle tissue was removed and preserved in absolute ethanol at -20ºC in the lab and the specimens were deposited in the ichthyological collection of the Universidade Federal da Bahia (UFBA), Brazil (Table S1 – Supporting Information). Collections were authorized by the Instituto Chico Mendes de Conservação da Biodiversidade (ICMBio/SISBIO; license number 51856–2), and the euthanasia and experimental procedures were approved by the Ethics Committee of Utilization of Animals from the Universidade Estadual do Sudoeste da Bahia (CEUA/UESB, number 71/2014).

Twenty individuals of Nematocharax venustus previously collected by Barreto et al. (2016) from two locations (sites 1 and 10 on the map in Fig. 1) were included in our study, consisting of cytochrome c oxidase subunit I (COI) sequences and photographs taken in the field. COI sequences are available in BOLD (Barcode of Life Data Systems; http://www.boldsystems.org/) under accession numbers PIABA028-14 to PIABA035-14 and PIABA050-14 to PIABA061-14.

2.2. DNA extraction and sequencing

Total DNA was extracted from all collected individuals using the Wizard Genomic DNA Purification kit (Promega, Madison, WI, USA). To identify the mtDNA lineage of each individual, a 650 base-pair (bp) fragment of the COI gene was amplified and sequenced with the primers FishF2_t1 and FishR2_t1 (Ward et al., 2005) following Barreto et al. (2016), with

approximately 15 samples per location. Sequencing was performed at the Gonçalo Moniz Research Center (FIOCRUZ-Bahia) using the BigDyeTerminator v3.1 Cycle Sequencing Ready Reaction kit (Applied Biosystems, Foster City, CA, USA). The COI gene was selected as an appropriate molecular marker due to its general utility for studying population-level phenomena in fish (e.g. Thomaz, Malabarba, Bonatto, & Knowles, 2015; Lima et al., 2017; Cunha et al., 2019), its previously proven ability to reveal sympatric divergence in Nematocharax (Barreto et al., 2016), and the availability of published sequences from populations analyzed here.

A subset of 55 samples of the sequenced individuals was selected for the restriction site-associated DNA sequencing (RADseq), with 2-5 individuals per location. Whenever possible, this selection included males and females for each location and individuals from both mtDNA lineages in sites of sympatry. Locations from the main river course (sites 1, 4, 6, 7, 11, and 14; see Fig. 1), which correspond to the transect of the cline analyses, represented the largest sample sizes (i.e., five each). We generated the ezRAD libraries according to Toonen et al. (2013) and Knapp et al. (2016), extracting DNA using the DNeasy Blood and Tissue kit (Qiagen, Hilden, Germany), digesting the DNA with the restriction endonuclease DpnII (New England Biolabs, Ipswich, MA, USA), and preparing libraries using the Illumina TruSeq Nano kit (Illumina, San Diego, CA, USA), selecting fragments between 150 and 350 bp. Quantitative and qualitative validation of libraries was carried out using Qubit dsDNA BR (Broad Range) Assay kit (Thermo Fisher Scientific, Waltham, MA, USA) and Agilent 2100 Bioanalyzer (Agilent Technologies, Santa Clara, CA, USA), respectively. Paired-end sequencing of fragments (2x75 bp) was performed using the Illumina NextSeq 550 System with two Mid Output v2 kits (150 cycles) (Illumina, San Diego, CA, USA) at the Genome Investigation and Analysis Laboratory (GENIAL) core facility (CEFAP-USP, São Paulo, Brazil).

2.3. Sequence processing

Consensus COI sequences were obtained by comparing forward and reverse electropherograms in the software CodonCode Aligner 7.1.1 (CodonCode Corporation) and aligned using the ClustalW Multiple Alignment tool (Thompson, Higgins, & Gibson, 1994) in BioEdit 7.2.6.1 (Hall, 1999). All new COI sequences generated as part of this study are deposited in GenBank (accession numbers MN011189-MN011202 and MN011364-MN011542).

For the RADseq data, reads of each individual were demultiplexed using bcl2fastq 1.8.4 (Illumina; http://support.illumina.com/downloads.html). The toolbox ipyrad 0.7.28 (Eaton, 2014; http://ipyrad.readthedocs.io) was used to process the genomic sequences; details regarding read filtering, clustering within samples, joint estimation of heterozygosity and error rate, consensus base calls, and clustering across samples are given in the Supporting Information (Table S2). We used a de novo assembly method to filter data with reads trimmed 15 bp from each 3' edge to reduce low quality bases, retaining all loci with less than 30% missing data.

2.4. Coloration data acquisition

The coloration of both the pelvic-fin filament and the horizontal mark on the caudal peduncle represents the main differences found in individuals of Nematocharax venustus from the Gongogi River sub-basin (Bragança et al., 2013; Menezes et al., 2015). We quantified the variation in these two characters, as well as in two other color-based characters that exhibited noticeable variation in field-caught individuals (both males and females), specifically the color of pectoral fins and a mark in the eye (see Fig. 2). However, only the marks on the caudal peduncle and in the eye showed distinctive color patterns with high frequencies in parental populations (≥ 0.8 in one parental population and ≤ 0.2 in the other), thus being useful as diagnostic traits for the hybrid zone analysis.

Variation in the red mark in the eye was described as five categories based on its intensity, ranging from intense to absent; a similar approach was applied to the pink mark on the caudal peduncle, which ranged from dark pink to lack of pink (see Fig. 2). The color assignment into categories was visually performed by the same observer from photographs taken in the field for all specimens. These data were used to estimate geographic clines across the hybrid zone (described below).

2.5. Hybrid zone analyses

Relationships among mtDNA haplotypes were evaluated using a median-joining network built with the COI sequences in PopART (Leigh & Bryant, 2015). We identified the mtDNA lineage (Northern or Southern) of each individual and, based on this information, we mapped the locations of sympatry and allopatry of mitochondrial lineages.

Structure in the RADseq data was estimated using the variational Bayesian framework implemented in fastSTRUCTURE (Raj, Stephens, & Pritchard, 2014). We tested the number of clusters (K) from 1 to 5 with 10 independent runs each and using the default prior. The algorithm 'chooseK.py' was run to choose the appropriate number of model components explaining the structure in the dataset.

To estimate the posterior probability (pp) that each individual belongs to a distinct hybrid class (i.e., pure1, pure2, F1, F2, BC1, and BC2; Anderson & Thompson, 2002), we used the R package parallelnewhybrid (Wringe et al., 2017a) to run NewHybrids in parallel with a burn-in of 50,000 and 100,000 sweeps. This analysis was performed with a panel of the 200 most informative SNPs (based on global Weir and Cockerham (1984)'s FST) generated by the R package hybriddetective (Wringe et al., 2017b). The z option in NewHybrids was set as prior information for individuals from parental populations (i.e., locations 1 and 14 for the Southern and Northern lineages, respectively). We confirmed the convergence of the Markov Chain Monte Carlo (MCMC) chains using the hybriddetective.

Measures of genomic admixture within individuals were also obtained from the R package gghybrid (Bailey, 2018) by calculating a hybrid index that estimates the proportion of alleles that were inherited from one of the two parental groups (Anderson, 1949; Buerkle, 2005). To ensure that we only capture loci that show significant variation along the hybrid zone, we set the gghybrid parameters 'max.S.MAF' and 'min.diff' to 0.2 and 0.6, respectively. This implies that we only kept loci for which allele frequencies were no greater than 0.2 in one parental set and not lower than 0.8 in the other, and for which the difference in allele frequency between parental sets was greater than 0.6; parental populations correspond to sites 14 and 1 on the map in Fig. 1 for the Northern and Southern lineages, respectively.

To test how genetic and morphological characters vary along the Gongogi River, we used HZAR (hybrid zone analysis for R), an R package that provides functions for fitting the traits to equilibrium geographic cline models (Szymura & Barton, 1986; Gay, Crochet, Bell, & Lenormand, 2008) and allows cline parameters to be estimated using the Metropolis-Hastings algorithm (Derryberry, Derryberry, Maley, & Brumfield, 2014). We fit a combination of equations (15 models) that describe the shape of each cline as defined by Derryberry et al. (2014) following Szymura & Barton (1986, 1991), with a sigmoidal curve at the centre and two exponential decay curves (i.e., tails) on either side. The fit of the 15 models were compared to a null model with no clinal transition using Akaike Information Criterion (AIC) corrected for small sample size (AICc). The maximum likelihood parameters were extracted from the best-fitting model (i.e., the model with the lowest AICc score). Clinal

coincidence (same centre, c, measured in km from the sampling location 1) and concordance (same width, w, measured as 1/maximum slope) were evaluated based on the confidence intervals for each trait and SNP. The mtDNA cline was also modeled along the Gongogi River by using a site of the COI segregating between parental lineages.

We used the ancestry coefficients (q) estimated in fastSTRUCTURE considering the optimal value of K (i.e., K = 2, see Results) to estimate the cline that represents all genomic loci combined. The genomic loci were also analyzed separately based on two filters: (1) exclusion of mtDNA loci after sequence similarity analysis with the BLAST tool in the NCBI database (http://www.ncbi.nlm.nih.gov/BLAST/) and (2) taking into account allele frequencies present in at least two individuals per location and partially diagnostic loci (i.e., those with frequency differences ≥0.6 between parental populations). This procedure resulted in a final dataset of 99 putatively unlinked SNPs (i.e., one SNP per locus) from which we estimated cline parameter values independently. This subsampling was made to strictly analyze diagnostic SNPs (i.e., those that are distinct between parental populations) and to avoid estimates of allele frequency using less than four alleles per locus per location, thus reducing potential biases related to high amounts of missing data.

The hybrid zone transect was established by considering only the locations from the main river course (sites 1, 4, 6, 7, 11, and 14; see Fig. 1) because HZAR requires data collected along one-dimensional transects, with minimal variation perpendicular to the cline (Derryberry et al., 2014). MCMC sampling was run considering the length of this transect along the river course (63 km, including its bends) and using three independent chains with 1.0×10^6 generations for each model.

2.6. Geomorphological inference

Considering that the Southern lineage is also known to occur in a watershed adjacent to the Contas River basin (Barreto, 2019), we used GIS techniques to infer putative topographic changes that may have affected the surveyed area over time. We focused on investigating the occurrence of drainage rearrangements (i.e., river captures) between the Gongogi River sub-basin and neighboring basins. River captures correspond to the natural diversion of waters from one river to another adjacent one, which connects and isolates drainages and, consequently, aquatic populations (Bishop, 1995; Albert, Craig, Tagliacollo, & Petry, 2018). Thus, to infer such captures we used the following data sources: (1) the Continuous Cartographic Base of the Brazilian hydrography at 1:250,000-scale (DGC, 2017)

to visualize the current configuration of rivers and detect elbows of capture (abrupt changes in the river course at the point of capture; Bishop, 1995); (2) SRTM 90m Digital Elevation Data (Jarvis et al., 2008) to detect putative wind gaps (dry areas that correspond to ancient river beds; Ollier & Pain, 2000); and (3) the SD.24 Salvador sheet (DGC, 2016) to identify the presence and location of geological faults (areas subjected to tectonic reactivations; de Oliveira, 2010). All these data were analyzed using the QGIS 3.4.1 software (QGIS Development Team, 2019).

3. Results

3.1. Population structure

Two divergent mtDNA lineages were identified from the haplotype network for the COI gene in Nematocharax venustus, separated by 24 mutational steps (4.1% of divergence) (Fig. S1 – Supporting Information). The Northern lineage includes a larger number of individuals (N = 164) and a broader distribution in the sampled area. Population assignment indicated sympatry of mtDNA lineages in six locations including the main river (Gongogi River) and a single tributary (Cambiriba Stream). Moreover, we found a frequency gradient for the two lineages along the main river, with the Northern and Southern lineages predominating at opposite stretches of the Gongogi River and co-occurring near the confluence with the Cambiriba Stream (Figs. 1-2).

Regarding the RADseq data, we obtained a total of 133.8 M raw reads (1.3-4.3 M per sample) which were filtered into 129.5 M reads encompassing 1,141 loci and 4,226 SNPs (see Table S3 for a final summary of the statistics provided by ipyrad). FastSTRUCTURE results indicated that K = 2 best fit the data, with clusters (taxa) clearly corresponding to the two mtDNA lineages. The analysis showed individuals of mixed ancestry (Fig. 3A), particularly in locations 6-13, whereas significant admixture was absent in the locations assumed as parental (1 and 14 for the Southern and Northern lineages, respectively).

The NewHybrids analysis assigned nine, 12, and 33 individuals to the pure 1, pure 2, and F2 categories, respectively (Fig. 3B), with high posterior probabilities (>0.99 pp) for all samples except one (sample code 0022, location 4), which had pp values equal to 0.4 and 0.6 for the pure 1 and F2 categories, respectively. Geographically, samples classified as pure 1 were mostly detected in locations 1 and 4, whereas pure 2 individuals were found in locations 14, 5, and 3; F2 offspring were shown to occur in locations 2 and 6-13, particularly near the

confluence with the Cambiriba Stream. No evidence of F1 hybrids or early generation backcrosses was detected.

The hybrid index estimated with gghybrid varied from zero to one, ranging from pure individuals of the Southern lineage to pure individuals of the Northern lineage and following an almost linear transect of hybridization along the main river (Fig. 4). Note that the tributaries were not considered in the clinal analyses, especially those associated with locations 2, 3, and 5, also because of their geographic disjunction (i.e., they are geographically closer to location 1, where the Southern lineage predominates, but are genetically similar to location 14, where the Northern lineage predominates).

3.2. Hybrid zone dynamics

According to the AICc calculated by HZAR, the best-fit model for the color of the mark on the caudal peduncle was model IV (which sets trait interval [pMin and pMax] to observed values with two exponential tails mirrored on the cline centre; Fig. 5A), and model I (pMin and pMax set to observed values with no exponential tails fitted) had the lowest AICc score for both the mark in the eye (Fig. 5B) and the mtDNA (Fig. 5C). Cline centre and width for these traits were: caudal peduncle c = 33.72 (27.46–43.40) km and w = 33.18 (18.44–50.58) km; eye c = 40.23 (35.88–45.27) km and w = 20.36 (11.70–32.51) km; and mtDNA c = 28.60 (24.21–33.22) km and w = 24.65 (16.34–39.17) km (Fig. 5A-C). These values place the cline centre for the three traits around location 7, in the confluence area with the Cambiriba Stream (Fig. 1), with the narrowest cline observed for the eye color trait, followed by the mtDNA. The other two morphological traits (color of the pectoral fin and color of the tip of the pelvic fin) were not informative because they showed no clinal variation across the hybrid zone (data not shown).

Considering the cline estimated from the ancestry coefficients (q) inferred by fastSTRUCTURE, AICc score in HZAR indicated model I as the best-fit model, which recovered a smoother and wider cline compared to the mtDNA and color traits, with centre estimated at 35.55 (28.43–43.98) km (near location 7) and width of 40.13 (25.30–70.28) km (Fig. 5D). The geographic cline analysis for the 99 RADseq loci revealed smooth clines for most of them, but 10 loci showed non-concordant abrupt changes in frequency consistent with stepped clines along the transect (Fig. S2 – Supporting Information). Overall, we found broad confidence intervals for cline widths and, consequently, extensive overlap among them

(details on cline centre and width values for each SNP are shown in Table S4 – Supporting Information).

3.3. Riverscape evolution

Based on the investigation of geological and geomorphological data around the hybrid zone, we found evidence of topographical changes in the studied area, specifically in the headwaters of the Gongogi River sub-basin (Fig. 6). These changes suggest a past connection between the Contas and Pardo river basins through a river capture event. It is important to notice that the Pardo River basin is a separate but adjacent basin where populations belonging to the Southern lineage of Nematocharax also occur (Barreto, 2019). Thus, the drainage rearrangement may have allowed the dispersal of the Southern lineage from the Pardo to the Contas river basin. The putative river capture between these adjacent watersheds was inferred from three sources of evidence: (1) presence of drainages with abrupt changes in the flow direction (elbows of capture); (2) wind gaps at the boundary between river basins; and (3) two geological faults near the capture area (Fig. 6).

4. Discussion

Our analyses based on genomic data and color traits support the occurrence of hybridization between lineages of Nematocharax that clearly correspond to two divergent gene pools, with clinal transition between parental forms across the Gongogi River and different levels of hybridization in the tributaries (Figs. 1 and 3). Our findings also help elucidate the origin of this hybrid zone, given the geological and geomorphological evidence of a river capture event that probably caused secondary contact in that stretch of the Contas River basin (Fig. 6).

Although population genomics studies with freshwater fishes have provided valuable insights into the ecological and evolutionary contexts of hybridization (e.g. McKelvey et al., 2016; Mandeville et al., 2017; Sotola et al., 2019), investigations on the influence of past and current river properties, especially using geomorphological data, are still scarce. In addition to inferences about the origin and structure of the hybrid zone in Nematocharax, we discuss the taxonomic implications of these results, including what our genomic data suggest about species boundaries in this fish genus.

4.1. Origin of the hybrid zone

Distinguishing between differentiation in a continuous population as a direct response to the environment vs. divergence in allopatry followed by secondary contact is not an easy task when determining the origin of a hybrid zone (Harrison & Larson, 2016). In this study, the evidence points to a scenario of secondary contact and interbreeding between lineages of N. venustus that diverged in allopatry, forming a hybrid zone around the Cambiriba Stream, in the Gongogi River sub-basin. The most likely explanation is that the current configuration of the rivers was not stable over time, with the Southern lineage reaching the Contas River basin (where the Northern lineage was already present) via dispersal allowed by a river capture (Fig. 6). Our geomorphological data show typical evidence of this type of event, such as elbows of capture, wind gaps, and geological faults (de Oliveira, 2010), thus providing a potential route for dispersal of the Southern lineage from the Pardo to the Contas river basin. This hypothesis is reinforced by the sharing of the same mtDNA haplotype of the Southern lineage between both river basins and by evidence of demographic expansion for the Southern lineage when including populations outside the Contas River basin (Fig. 6; Barreto et al., in prep.).

Phylogeographic analyses encompassing lineages of Nematocharax throughout the entire distribution of the genus and using both COI sequences and RADseq data show that the Northern and Southern lineages are not sister groups (Barreto et al., in prep.). According to the COI, the time to the most recent common ancestor (TMRCA) is 0.48 (0.27-0.62) Mya for the Northern lineage and 0.43 (0.31-0.67) Mya for the Southern lineage. This period corresponds to the Pleistocene, during which tectonic reactivations of geological faults caused several topographic changes in coastal drainages of eastern Brazil, including drainage rearrangements (Saadi et al., 2002; Ribeiro, 2006), which may have promoted secondary contact between aquatic lineages.

Intriguingly, the hybrid zone analyzed here includes the type locality (site 10 on the map; Fig. 1) of a putative new species of Nematocharax (N. costai) that was synonymized with N. venustus (Bragança et al., 2013; Menezes et al., 2015). Despite the clear differentiation between the Northern and Southern lineages in mtDNA, RADseq, and color pattern, our data do not support N. costai in the way it was described by Bragança et al. (2013) based on morphological and morphometric data from five specimens that our analyses show to occur in an area with a high frequency of hybrids (site 10; Fig. 1). As such, our work corroborates the synonymization of N. costai and N. venustus (see Menezes et al., 2015), but

highlights the need for a taxonomically focused study on what is currently recognized as a single taxon, N. venustus.

4.2. Dynamics of the hybrid zone

Geographic clines allow verification of how the frequency of alleles or phenotypes vary along a transect, where smooth transitions suggest weak selection and less effective reproductive barriers, whereas stepped clines indicate strong selective pressures and substantial barriers to gene flow between lineages (Barton & Gale, 1993). Indeed, the clines estimated from the ancestry coefficients (Fig. 5D) and for most individual RADseq loci (Fig. S2 – Supporting Information) are more gradual and largely concordant, indicating that the strength of selection is probably weak and uniform across these loci. The presence of a few abruptly stepped clines (Fig. S2 – Supporting Information) along the hybrid zone, in turn, could indicate that some loci are in regions of strong selection, which may be preventing genetic homogenization between parental groups (Feder et al., 2013). These results exemplify the "semi-permeable" nature of hybrid zones, that is, the idea that gene flow and reproductive isolation can be thought of in terms of genomic regions rather than entire genomes (Kawakami & Butlin, 2012; Harrison & Larson, 2014).

Regarding the color traits (Figs. 2 and 5A, B), it is possible that the differences between parental lineages have diverged before the secondary contact. Thus, the presence of individuals with mixed color patterns (Fig. 2ii) at the centre of the hybrid zone would be a clear evidence of hybridization. However, because the abundance of F2 hybrids and absence of F1 offspring and early generation backcrosses, we may have detected a hybrid lineage that is becoming reproductively independent from the parental lineages, which means that the rates of hybridization and introgression have decreased over time. Color traits would be of fundamental importance in this case, given their potential role on species recognition and mating success (e.g. Houde, 1987; Couldridge & Alexander, 2002).

It is worth noting that MCMC chains in the NewHybrids software can sometimes fail to converge, causing nearly all individuals to be recovered as F2 hybrids (Wringe et al., 2017b). However, our MCMC chains converged properly, as shown by hybriddetective, and the predominance of F2 hybrids is in accordance with our other findings, such as the clear spatial structuring of hybrid individuals within the transition zone between areas where parental lineages are found. Although the scale of dispersal for Nematocharax is unknown, it is plausible to assume that the spatial segregation between hybrids and parental populations is

due to the small scale of movement generally described for small characins (Lucas & Baras, 2001). Therefore, the diagnosable genetic and morphological differences between the Southern and Northern lineages could be justified. Additionally, clinal variation across the hybrid zone may reflect isolation by distance in a linear riverine system and gradual admixture of the divergent lineages.

Even though mating behavior has not yet been studied in Nematocharax, sexual selection might also play a role on the maintenance of this hybrid zone. Indeed, empirical and theoretical studies have demonstrated that assortative mating can influence the incidence and rate of hybridization (e.g. MacCallum, Nürnberger, Barton, & Szymura, 1998; Vines, 2002; Culumber, Ochoa, & Rosenthal, 2014). The sexual dimorphism found in Nematocharax, including the presence of hooks and spinules on fins and elongated fins in maturing and mature males, is also commonly reported in small characid species (e.g. Zanata & Camelier, 2009; Dagosta, Marinho, & Camelier, 2014; Marinho, Dagosta, & Birindelli, 2014) being usually related to male display behavior and female mate choice (Bischoff, Gould, & Rubenstein, 1985). Further investigation can directly test whether assortative mating of hybrids contributes to reproductive isolation from the parental lineages.

4.3. Possible role of riverscape properties

Small South American characins have been described as resident or small-scale migrants (Lucas & Baras, 2001). If this is true for Nematocharax, then it could imply restricted movement relative to the width of the hybrid zone, meaning that few pure individuals are likely to be found in the centre. Supporting this idea, field and aquarium observations (pers. obs.) indicate that Nematocharax has a territorial behavior (particularly males), spending most of its lifetime in a very restricted area (Gerking, 1953). For these species, riverscape architecture may have a pronounced effect on the structure of the hybrid zone, influencing the spatial distribution of hybrids and pure individuals. For example, the Southern lineage probably entered the Gongogi River sub-basin via river capture (Fig. 6) and dispersed mainly along the main river channel. However, hybridization only occurs in three of the five tributaries (Figs. 1 and 3), which might be associated with differential dispersal opportunities related to the altitudinal profile and the angle of connection to the main river. To illustrate this hypothesis, Figure 1 shows that, at the confluence with the Cambiriba Stream (site 7), the Gongogi River forms a sharp bend, thus creating a counterflow and favoring the formation of flooded and shallow areas particularly suitable for N. venustus (pers. obs.). This

may help understand why secondary contact was predominant in tributaries such as the Cambiriba Stream, but absent in others (Figs. 1 and 3).

Interestingly, artificial reservoirs have been built in the hybrid zone analyzed here, specifically in location 10 on the map (Fig. 1; Balneário Guaíra, type locality of the synonymized species) and in the connection point between the main river and the tributary where the location 5 is situated (Fig. 1; Balneário Beach Park). Thus, although they are very recent, these places may have been responsible for imposing additional physical barriers or habitat disturbances (e.g. changes in river depth, temperature, and waterflow), with potential influence on fish population structure (e.g. Valenzuela-Aguayo et al., 2020).

Overall, our findings highlight that biological traits and the past and current riverscape architectures can interact to originate and structure hybrid zones in freshwater fish species, even though it is difficult to disentangle the individual effects of a particular factor. Also, our data shed light on the dynamics of hybridization between lineages of N. venustus, with direct relevance to the study of interspecific boundaries in Nematocharax. Our work illustrates that the understanding of a hybrid zone is not dependent on the taxonomic status, but rather on the nature of differences between lineages or groups (Mallet, 1995; Gompert & Buerkle, 2016). Additional studies using, for example, behavioral data and experimental measures of relative fitness of hybrid and parental individuals are required to expand our knowledge on the ecological and evolutionary consequences of hybridization in this system.

5. References

Albert, J. S., Craig, J. M., Tagliacollo, V. A., & Petry, P. (2018). Upland and lowland fishes: a test of the river capture hypothesis. In C. Hoorn, A. Antonelli, & A. Antonelli (Eds.), Mountains, climate and biodiversity (pp. 273-294). New York, NY: Wiley-Blackwell.

Albert, V., Jónsson, B., & Bernatchez, L. (2006). Natural hybrids in Atlantic eels (Anguilla anguilla, A. rostrata): evidence for successful reproduction and fluctuating abundance in space and time. Molecular Ecology, 15(7), 1903-1916. http://doi.org/10.1111/j.1365-294X.2006.02917.x

Anderson, E. (1949). Introgressive hybridization. New York, NY: J. Wiley.

Anderson, E. C., & Thompson, E. A. (2002). A model-based method for identifying species hybrids using multilocus genetic data. Genetics, 160(3), 1217-1229.

Arnold, M. L. (1997). Natural hybridization and evolution. Oxford University Press: New York.

Bailey, R. I. (2018). gghybrid: Evolutionary Analysis of Hybrids and Hybrid Zones. R package version 0.0.0.9000. Retrieved from http://github.com/ribailey/gghybrid

Barreto, S. B. (2019). Filogeografia, zona híbrida e especiação em linhagens de Nematocharax (Characiformes: Characidae). PhD Thesis. Federal University of Bahia, Salvador, Brazil.

Barreto, S. B., Knowles, L. L., Mascarenhas, R., Affonso, P. R. A. M., & Batalha-Filho, H. (n.d.). Drainage rearrangements drove diversification in an endemic fish genus from the Northeastern Mata Atlantica freshwater ecoregion. Manuscript in preparation.

Barreto, S. B., Nunes, L. A., da 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. http://doi.org/10.1139/gen-2015-0166

Barton, N. H., & Gale, K. S. (1993). Genetic analysis of hybrid zones. In R. G. Harrison (Ed.). Hybrid zones and the evolutionary process (pp. 13-45). New York, NY: Oxford University Press.

Barton, N. H., & Hewitt, G. M. (1985). Analysis of hybrid zones. Annual review of Ecology and Systematics, 16(1), 113-148. http://doi.org/10.1146/annurev.es.16.110185.000553

Bischoff, R. J., Gould, J. L., & Rubenstein, D. I. (1985). Tail size and female choice in the guppy (Poecilia reticulata). Behavioral Ecology and Sociobiology, 17(3), 253-255. http://doi.org/10.1007/BF00300143

Bishop, P. (1995). Drainage rearrangement by river capture, beheading and diversion. Progress in Physical Geography, 19(4), 449-473. http://doi.org/10.1177/030913339501900402

Bragança, P. D., Barbosa, M. A., & Mattos, J. L. (2013). A new Nematocharax species from the middle Contas River basin, northeastern Brazil (Characiformes: Characidae). Vertebrate Zoology, 63(1), 3-8.

Buerkle, C. A. (2005). Maximum-likelihood estimation of a hybrid index based on molecular markers. Molecular Ecology Notes, 5(3), 684-687. http://doi.org/10.1111/j.1471-8286.2005.01011.x

Chaput-Bardy, A., Fleurant, C., Lemaire, C., & Secondi, J. (2009). Modelling the effect of in-stream and overland dispersal on gene flow in river networks. Ecological Modelling, 220(24), 3589-3598. http://doi.org/10.1016/j.ecolmodel.2009.06.027

Couldridge, V. C., & Alexander, G. J. (2002). Color patterns and species recognition in four closely related species of Lake Malawi cichlid. Behavioral Ecology, 13(1), 59-64. http://doi.org/10.1093/beheco/13.1.59

Culumber, Z. W., Ochoa, O. M., & Rosenthal, G. G. (2014). Assortative mating and the maintenance of population structure in a natural hybrid zone. The American Naturalist, 184(2), 225-232. http://doi.org/10.1086/677033

Cunha, M. S., Fregonezi, A. R., Fava, L., Hilsdorf, A. W., Campos, L. A., & Dergam, J. A. (2019). Phylogeography and historical biogeography of the Astyanax bimaculatus species complex (Teleostei: Characidae) in coastal southeastern South America. Zebrafish, 16(1), 115-127. http://doi.org/10.1089/zeb.2018.1668

Curry, C. M. (2015). An integrated framework for hybrid zone models. Evolutionary Biology, 42(3), 359-365. http://doi.org/10.1007/s11692-015-9332-9

Dagosta, F. C., Marinho, M. M., & Camelier, P. (2014). A new species of Hyphessobrycon Durbin (Characiformes: Characidae) from the middle rio São Francisco and upper and middle rio Tocantins basins, Brazil, with comments on its biogeographic history. Neotropical Ichthyology, 12(2), 365-375. http://doi.org/10.1590/1982-0224-20130179

De La Torre, A. R., Wang, T., Jaquish, B., & Aitken, S. N. (2014). Adaptation and exogenous selection in a Picea glauca × Picea engelmannii hybrid zone: implications for forest management under climate change. New Phytologist, 201(2), 687-699. http://doi.org/10.1111/nph.12540

de Oliveira, D. (2010). Capturas fluviais como evidências da evolução do relevo: uma revisão bibliográfica. Revista do Departamento de Geografia, 20, 37-50. http://doi.org/10.7154/RDG.2010.0020.0003

Derryberry, E. P., Derryberry, G. E., Maley, J. M., & Brumfield, R. T. (2014). HZAR: hybrid zone analysis using an R software package. Molecular Ecology Resources, 14(3), 652-663. http://doi.org/10.1111/1755-0998.12209

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 ftp://geoftp.ibge.gov.br/cartas_e_mapas/bases_cartograficas_continuas/bc250/versao2017

Duvernell, D. D., & Schaefer, J. F. (2014). Variation in contact zone dynamics between two species of topminnows, Fundulus notatus and F. olivaceus, across isolated drainage systems. Evolutionary Ecology, 28(1), 37-53. http://doi.org/10.1007/s10682-013-9653-z

Eaton, D. A. (2014). PyRAD: assembly of de novo RADseq loci for phylogenetic analyses. Bioinformatics, 30(13), 1844-1849. http://doi.org/10.1093/bioinformatics/btu121

570 Feder, J. L., Flaxman, S. M., Egan, S. P., Comeault, A. A., & Nosil, P. (2013). Geographic
571 mode of speciation and genomic divergence. Annual Review of Ecology, Evolution, and
572 Systematics, 44, 73-97. http://doi.org/10.1146/annurev-ecolsys-110512-135825
573 Fullerton, A. H., Burnett, K. M., Steel, E. A., Flitcroft, R. L., Pess, G. R., Feist, B. E., ... &
574 Sanderson, B. L. (2010). Hydrological connectivity for riverine fish: measurement
575 challenges and research opportunities. Freshwater Biology, 55(11), 2215-2237.
576 http://doi/org/10.1111/j.1365-2427.2010.02448.x
577 Gay, L., Crochet, P. A., Bell, D. A., & Lenormand, T. (2008). Comparing clines on molecular
578 and phenotypic traits in hybrid zones: a window on tension zone models. Evolution:
579 International Journal of Organic Evolution, 62(11), 2789-2806.
580 http://doi.org/10.1111/j.1558-5646.2008.00491.x
581 Gerking, S. D. (1953). Evidence for the concepts of home range and territory in stream fishes.
582 Ecology, 34(2), 347-365. http://doi.org/10.2307/1930901
583 Gompert, Z., & Buerkle, C. A. (2016). What, if anything, are hybrids: enduring truths and
584 challenges associated with population structure and gene flow. Evolutionary Applications,
585 9(7), 909-923. http://doi.org/10.1111/eva.12380
586 Gompert, Z., Mandeville, E. G., & Buerkle, C. A. (2017). Analysis of population genomic
587 data from hybrid zones. Annual Review of Ecology, Evolution, and Systematics, 48, 207-
588 229. http://doi.org/10.1146/annurev-ecolsys-110316-022652
589 Hall, T. A. (1999). BioEdit: a user-friendly biological sequence alignment editor and analysis
590 program for Windows 95/98/NT. Nucleic Acids Symposium Series, 41(41), 95-98.
591 Harrison, R. G., & Larson, E. L. (2014). Hybridization, introgression, and the nature of
592 species boundaries. Journal of Heredity, 105(S1), 795-809.
593 http://doi.org/10.1093/jhered/esu033
594 Harrison, R. G., & Larson, E. L. (2016). Heterogeneous genome divergence, differential
595 introgression, and the origin and structure of hybrid zones. Molecular Ecology, 25(11),
596 2454-2466. http://doi.org/10.1111/mec.13582
597 Harrison, R. G., & Rand, D. M. (1989). Mosaic hybrid zones and the nature of species
598 boundaries. In J. Endler, & D. Otte (Eds.), Speciation and its consequences (pp. 111-113).
599 Sunderland, MA: Sinauer Associates.
600 Hewitt, G. M. (1988). Hybrid zones-natural laboratories for evolutionary studies. Trends in
601 Ecology & Evolution, 3(7), 158-167. http://doi.org/10.1016/0169-5347(88)90033-X

602 Houde, A. E. (1987). Mate choice based upon naturally occurring color-pattern variation in a
603 guppy population. Evolution, 41(1), 1-10. http://doi.org/10.1111/j.1558-
604 5646.1987.tb05766.x
605 Hughes, J. M. (2007). Constraints on recovery: using molecular methods to study connectivity
606 of aquatic biota in rivers and streams. Freshwater Biology, 52(4), 616-631.
607 http://doi.org/10.1111/j.1365-2427.2006.01722.x
608 Hughes, J. M., Schmidt, D. J., & Finn, D. S. (2009). Genes in streams: using DNA to
609 understand the movement of freshwater fauna and their riverine habitat. BioScience, 59(7),
610 573-583. http://doi.org/10.1525/bio.2009.59.7.8
611 Jarvis, A., Reuter, H.I., Nelson, A., & Guevara, E. (2008). Hole-filled seamless SRTM data
612 V4, International Centre for Tropical Agriculture (CIAT). Retrieved from
613 http://srtm.csi.cgiar.org
614 Kawakami, T., & R. T. Butlin (2012). Hybrid zones. Retrieved from
615 http://doi.org/.10.1002/9780470015902
616 Knapp, I. S. S., Puritz, J., Bird, C., Whitney, J., Sudek, M., Forsman, Z., & Toonen, R.
617 (2016). ezRAD - an accessible next-generation RAD sequencing protocol suitable for non-
618 model organisms_v3.2. Retrieved from http://doi.org/10.17504/protocols.io.e9pbh5n
619 Leigh, J. W., & Bryant, D. (2015). POPART: full-feature software for haplotype network
620 construction. Methods in Ecology and Evolution, 6(9), 1110-1116.
621 http://doi.org/10.1111/2041-210X.12410
622 Lima, S. M., Berbel-Filho, W. M., Araújo, T. F., Lazzarotto, H., Tatarenkov, A., & Avise, J.
623 C. (2017). Headwater capture evidenced by paleo-rivers reconstruction and population
624 genetic structure of the armored catfish (Pareiorhaphis garbei) in the Serra do Mar
625 mountains of southeastern Brazil. Frontiers in Genetics, 8, 199.
626 http://doi.org/10.3389/fgene.2017.00199
627 Lucas, M., & Baras, E. (2001). Migration of freshwater fishes. Oxford, UK: Blackwell
628 Science Ltd.
629 MacCallum, C. J., Nürnberger, B., Barton, N. H., & Szymura, J. M. (1998). Habitat
630 preference in the Bombina hybrid zone in Croatia. Evolution, 52(1), 227-239.
631 http://doi.org/10.1111/j.1558-5646.1998.tb05156.x
632 Mallet, J. (1995). A species definition for the modern synthesis. Trends in Ecology &
633 Evolution, 10(7), 294-299. http://doi.org/10.1016/0169-5347(95)90031-4
634 Mandeville, E. G., Parchman, T. L., Thompson, K. G., Compton, R. I., Gelwicks, K. R., Song,
635 S. J., & Buerkle, C. A. (2017). Inconsistent reproductive isolation revealed by interactions

between Catostomus fish species. Evolution Letters, 1(5), 255-268. http://doi.org/10.1002/evl3.29

Marinho, M. M., Dagosta, F. C., & Birindelli, J. L. (2014). Hemigrammus ataktos: a new species from the rio Tocantins basin, central Brazil (Characiformes: Characidae). Neotropical Ichthyology, 12(2), 257-264. http://doi.org/10.1590/1982-0224-20130091.

McKelvey, K. S., Young, M. K., Wilcox, T. M., Bingham, D. M., Pilgrim, K. L., & Schwartz, M. K. (2016). Patterns of hybridization among cutthroat trout and rainbow trout in northern Rocky Mountain streams. Ecology and Evolution, 6(3), 688-706. http://doi.org/10.1002/ece3.1887

Menezes, N. A., Zanata, A. M., & Camelier, P. (2015). Nematocharax costai Bragança, Barbosa & Mattos a junior synonym of Nematocharax venustus Weitzman, Menezes & Britski (Teleostei: Characiformes: Characidae). Zootaxa, 3920(3), 453-462. http://doi.org/10.11646/zootaxa.3920.3.4

Moore, W. S. (1977). An evaluation of narrow hybrid zones in vertebrates. The Quarterly Review of Biology, 52(3), 263-277. http://doi.org/10.1086/409995

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

Questiau, S. (1999). How can sexual selection promote population divergence?. Ethology Ecology & Evolution, 11(4), 313-324. http://doi.org/10.1080/08927014.1999.9522816

Raj, A., Stephens, M., & Pritchard, J. K. (2014). fastSTRUCTURE: variational inference of population structure in large SNP data sets. Genetics, 197(2), 573-589. http://doi.org/10.1534/genetics.114.164350

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. http://doi.org//10.1590/S1679-62252006000200009

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. Retrieved from http://pubs.usgs.gov/of/2002/ofr-02-230

Scribner, K. T., Page, K. S., & Bartron, M. L. (2000). Hybridization in freshwater fishes: a review of case studies and cytonuclear methods of biological inference. Reviews in Fish Biology and Fisheries, 10(3), 293-323. http://doi.org/10.1023/A:1016642723238

669 Sotola, V. A., Ruppel, D. S., Bonner, T. H., Nice, C. C., & Martin, N. H. (2019). Asymmetric
670 introgression between fishes in the Red River basin of Texas is associated with variation in
671 water quality. Ecology and Evolution, 9(4), 2083-2095. http://doi.org/10.1002/ece3.4901
672 Stewart, K. A., Hudson, C. M., & Lougheed, S. C. (2017). Can alternative mating tactics
673 facilitate introgression across a hybrid zone by circumventing female choice?. Journal of
674 Evolutionary Biology, 30(2), 412-421. http://doi.org/10.1111/jeb.13017
675 Szymura, J. M., & Barton, N. H. (1986). Genetic analysis of a hybrid zone between the
676 fire-bellied toads, Bombina bombina and B. variegata, near Cracow in southern Poland.
677 Evolution: International Journal of Organic Evolution, 40(6), 1141-1159.
678 http://doi.org/10.1111/j.1558-5646.1986.tb05740.x
679 Szymura, J. M., & Barton, N. H. (1991). The genetic structure of the hybrid zone between the
680 fire-bellied toads Bombina bombina and B. variegata: comparisons between transects and
681 between loci. Evolution: International Journal of Organic Evolution, 45(2), 237-261.
682 http://doi.org/10.1111/j.1558-5646.1991.tb04400.x
683 Thomaz, A. T., Christie, M. R., & Knowles, L. L. (2016). The architecture of river networks
684 can drive the evolutionary dynamics of aquatic populations. Evolution: International
685 Journal of Organic Evolution, 70(3), 731-739. http://doi.org/10.1111/evo.12883
686 Thomaz, A. T., Malabarba, L. R., Bonatto, S. L., & Knowles, L. L. (2015). Testing the effect
687 of palaeodrainages versus habitat stability on genetic divergence in riverine systems: study
688 of a Neotropical fish of the Brazilian coastal Atlantic Forest. Journal of Biogeography,
689 42(12), 2389-2401. http://doi.org/10.1111/jbi.12597
690 Thompson, J. D., Higgins, D. G., & Gibson, T. J. (1994). CLUSTAL W: improving the
691 sensitivity of progressive multiple sequence alignment through sequence weighting,
692 position-specific gap penalties and weight matrix choice. Nucleic Acids Research, 22(22),
693 4673-4680. http://doi.org/10.1093/nar/22.22.4673
694 Toonen, R. J., Puritz, J. B., Forsman, Z. H., Whitney, J. L., Fernandez-Silva, I., Andrews, K.
695 R., & Bird, C. E. (2013). ezRAD: a simplified method for genomic genotyping in non-
696 model organisms. PeerJ, 1, e203. http://doi.org/10.7717/peerj.203
697 Valenzuela-Aguayo, F., McCracken, G. R., Manosalva, A., Habit, E., & Ruzzante, D. E.
698 (2020). Human-induced habitat fragmentation effects on connectivity, diversity, and
699 population persistence of an endemic fish, Percilia irwini, in the Biobío River basin
700 (Chile). Evolutionary Applications, 13(4), 794-807. http://doi.org/10.1111/eva.12901
701 Vines, T. H. (2002). Migration, habitat choice and assortative mating in a Bombina hybrid
702 zone. PhD Thesis. University of Edinburgh, Edinburgh, UK.

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. http://doi.org/10.1098/rstb.2005.1716

Weir, B. S., & Cockerham, C. C. (1984). Estimating F-statistics for the analysis of population-structure. Evolution, 38(6), 1358-1370. http://doi.org/10.2307/2408641

Wringe, B. F., Stanley, R. R., Jeffery, N. W., Anderson, E. C., & Bradbury, I. R. (2017a). parallelnewhybrid: an R package for the parallelization of hybrid detection using NEWHYBRIDS. Molecular Ecology Resources, 17(1), 91-95. http://doi.org/10.1111/1755-0998.12597

Wringe, B. F., Stanley, R. R., Jeffery, N. W., Anderson, E. C., & Bradbury, I. R. (2017b). HYBRIDDETECTIVE: A workflow and package to facilitate the detection of hybridization using genomic data in R. Molecular Ecology Resources, 17(6), e275-e284. http://doi.org/10.1111/1755-0998.12704

Zanata, A. M., & Camelier, P. (2009). Astyanax vermilion and Astyanax burgerai: new characid fishes (Ostariophysi: Characiformes) from Northeastern Bahia, Brazil. Neotropical Ichthyology, 7(2), 175-184. http://doi.org/10.1590/S1679-62252009000200007

Figure legends

Fig. 1. Map of the Contas River basin, Bahia, Brazil (at the top right corner, in green) showing the stretch of the Gongogi River sub-basin where the sampling was performed. The 14 sampled locations are situated in the Gongogi River (sites 1, 4, 6, 7, 11, and 14), Cambiriba Stream (sites 8, 9, and 10), and other four tributaries (sites 2, 3, 5, 12, and 13). Sampled locations are color-coded according to which of the divergent mitochondrial lineages are present, with the Southern lineage in yellow and the Northern lineage in green. The circles are divided proportionally according to the number of individuals belonging to each mitochondrial lineage in cases of sympatry (i.e., sites 6, 7, 8, 9, 10, and 11), corresponding to the putative hybrid zone between divergent lineages. Small black arrows indicate the water flow direction of the Gongogi River and Cambiriba Stream.

Fig. 2. Male representatives of Nematocharax venustus highlighting some of the variation in the eye color (a), caudal peduncle (b), pectoral fins (c), and tip of the pelvic fins (d) based on

photographs taken in the field at sites 14, 7, and 4 (see Fig. 1), corresponding to phenotypes representative of (A) the Northern lineage, (B) a putative hybrid, and (C) the Southern lineage, respectively.

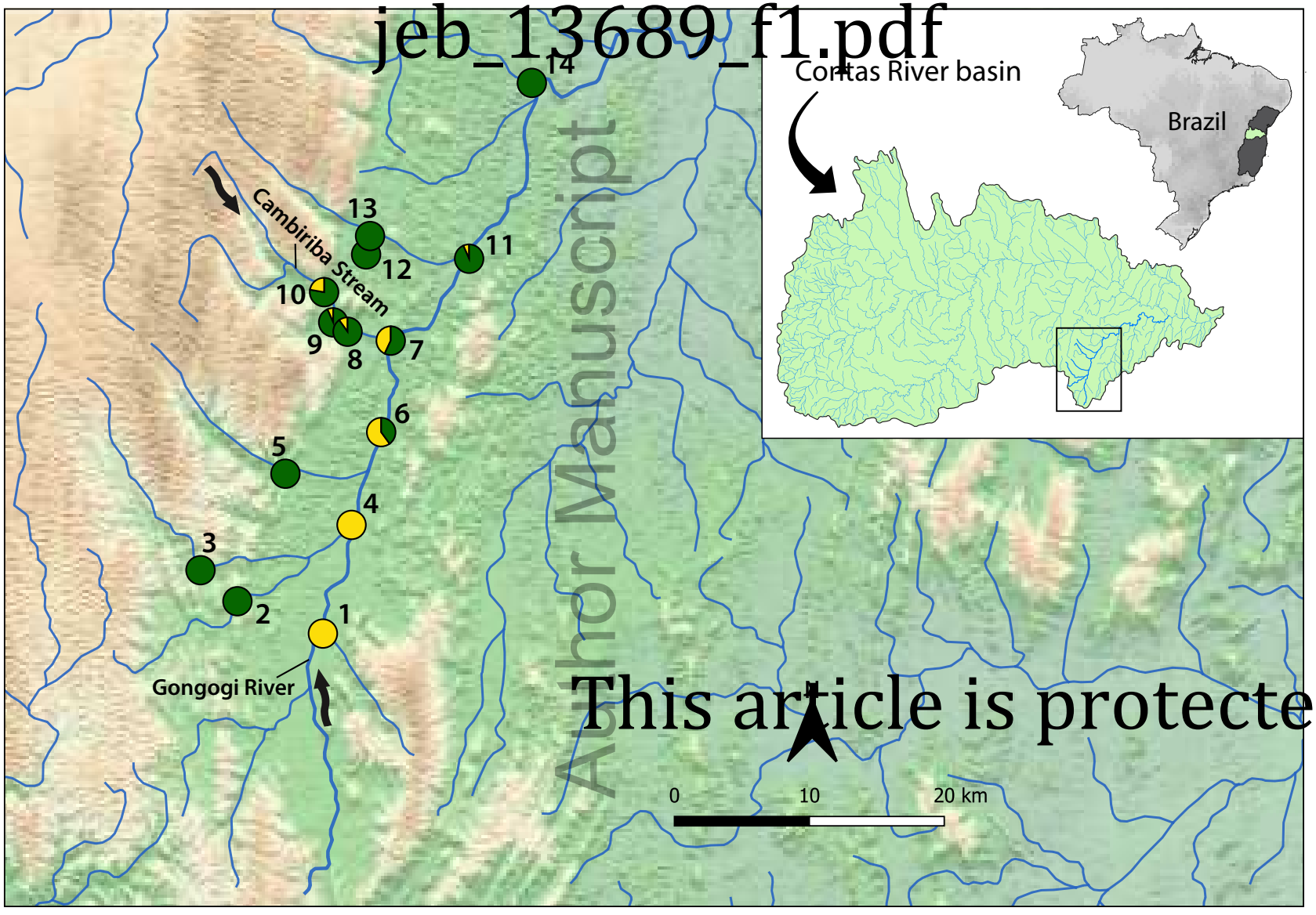
Fig. 3. Results of fastSTRUCTURE (A) and NewHybrids (B) analyses based on RADseq data. In the Structure plot (A) for the best K-value (=2), each bar represents the proportion of ancestry of each individual with respect to each potential ancestor; yellow and green correspond to the divergent lineages. Squares below the bars indicate the mitochondrial group of each individual. In the NewHybrids plot (B), each bar shows the probability of each individual to belong to each hybrid category (i.e., pure 1, pure 2, F1, F2, backcross 1, and backcross 2). Numbers (4492-0005) on the top of the figure and within hexagons (1-14) are the sample codes and localities, respectively (see Fig. 1). Black and gray hexagons refer to sites from the main river and tributaries, respectively.

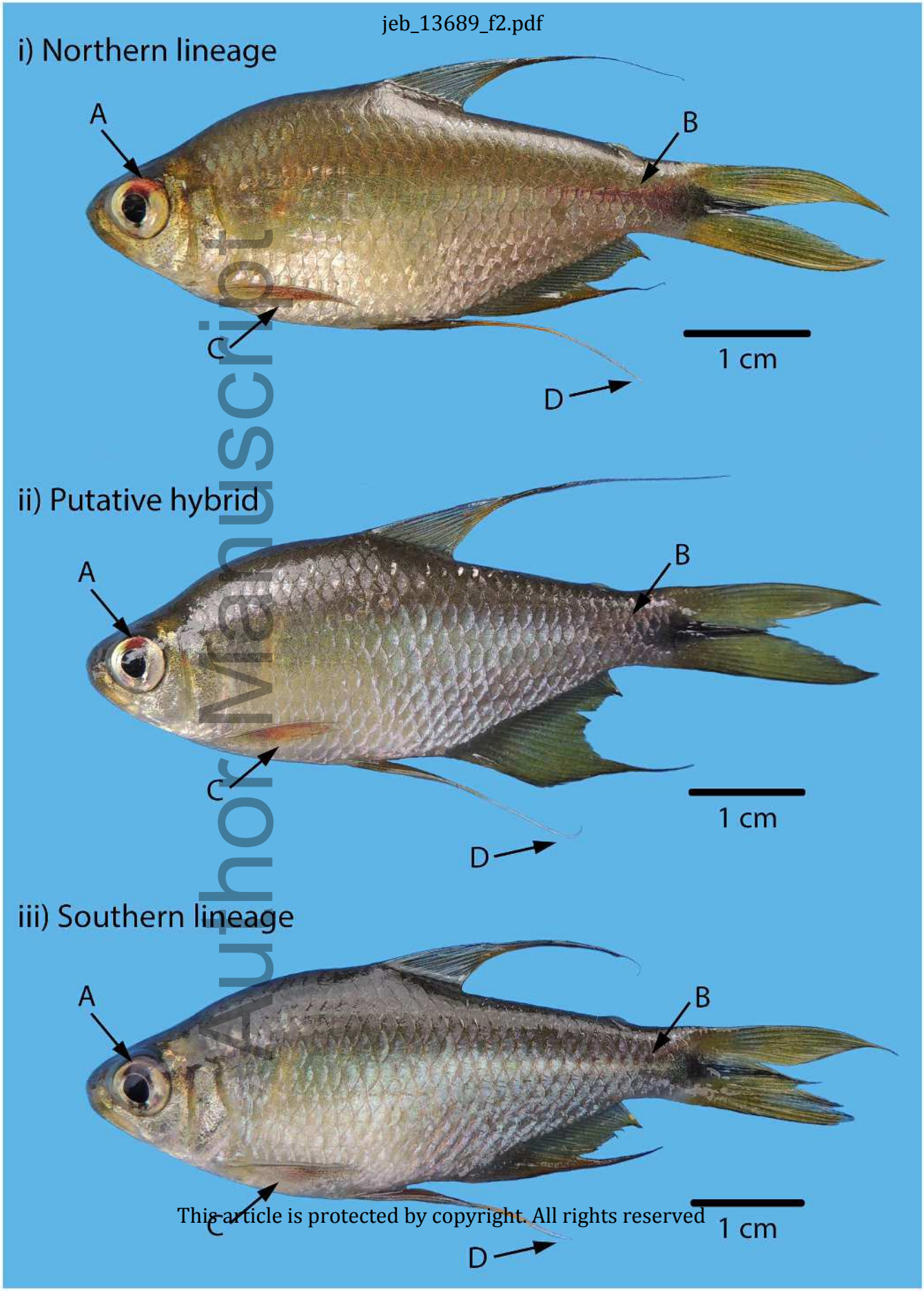
Fig. 4. Distribution of the hybrid index score (and 95% credible intervals) estimated with the gghybrid for the 55 individuals across the hybrid zone. Horizontal dashed lines mark the innermost credible interval for each parental reference set. Numbers within hexagons correspond to the locations according to the map (see Fig. 1). Black hexagons specify locations from the main river (Gongogi) whereas gray hexagons indicate locations from the tributaries.

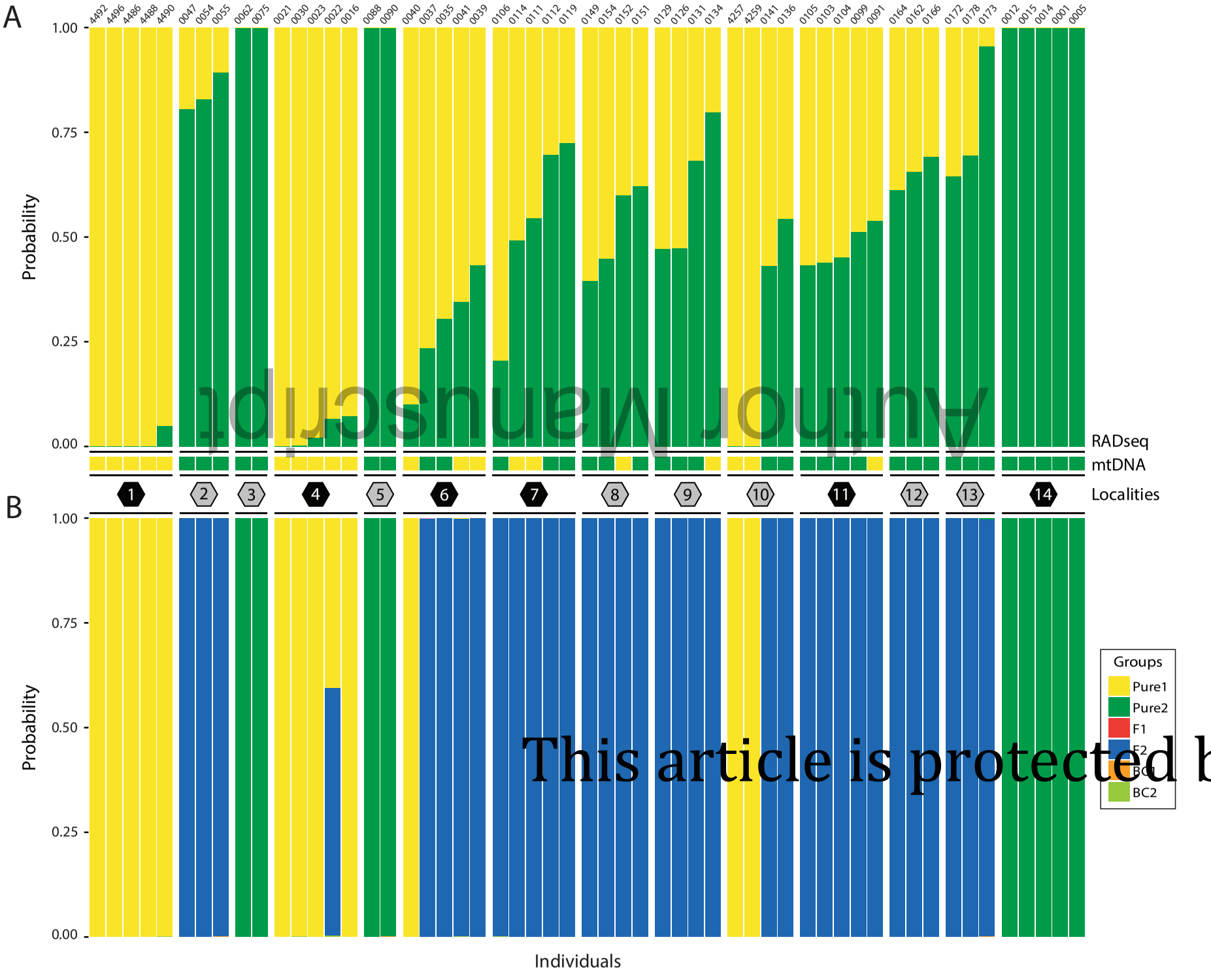
Fig. 5. Plots of the maximum likelihood clines and the 95% credible cline region estimated with the HZAR for the (A) mark on the caudal peduncle and (B) the mark in the eye, as well as (C) the mitochondrial haplotypes, and (D) the ancestry coefficients estimated in fastSTRUCTURE.

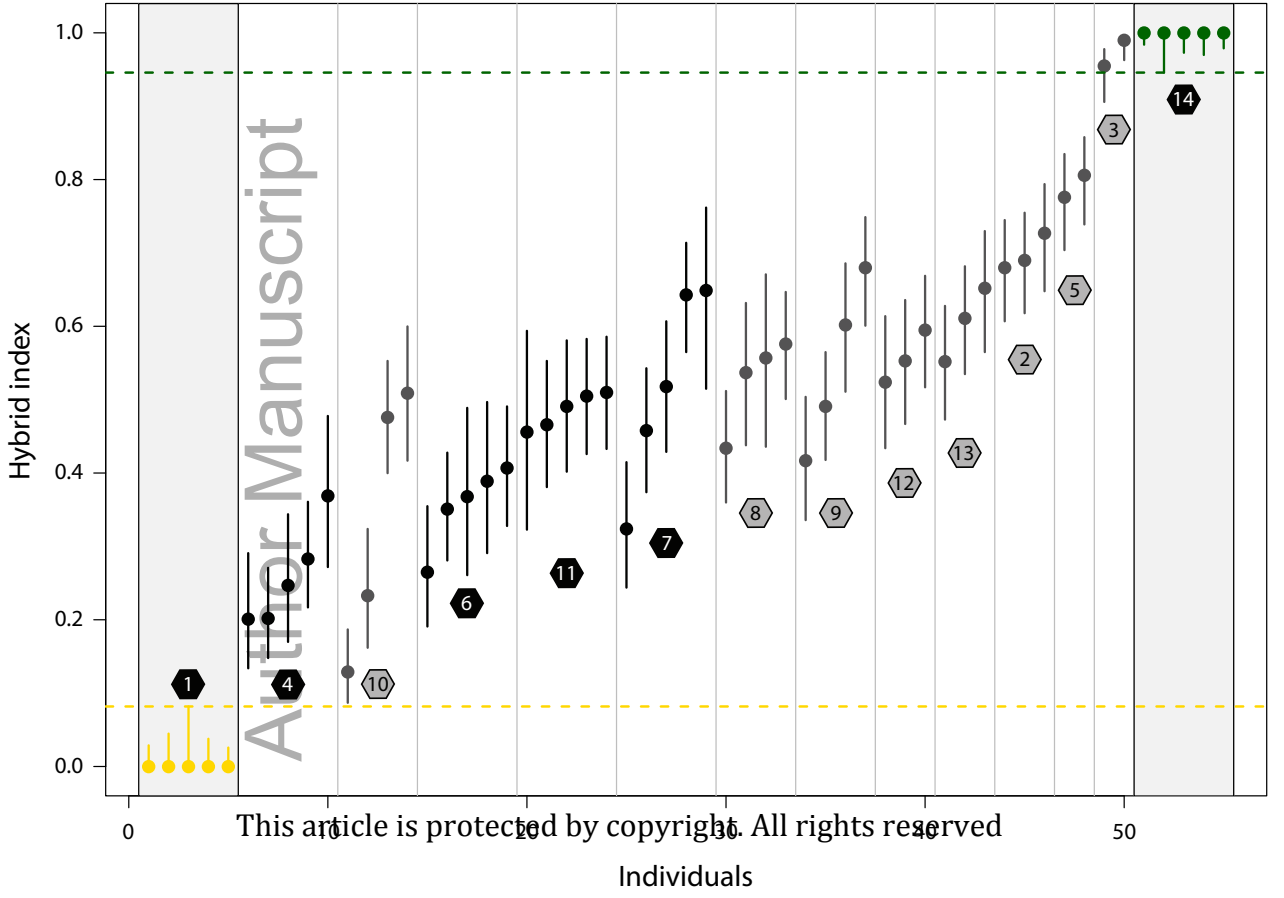
Fig. 6. Putative river capture (highlighted by the red circle) between the Contas and Pardo river basins (with boundaries shown in black; see inset for map of the entire region) in the focal area around the Gongogi River (i.e., sampled locations 1-5 from the hybrid zone; see also Fig. 1). Evidence of drainage rearrangement includes changes in the relief (based on the location of geological faults, represented by the white dashed lines), elbows of capture (based on hydrographic layers, in blue), and wind gaps (based on Digital Elevation Model, DEM, where areas of low and high elevation are shown in dark and light grey, respectively). Note that the Pardo River basin is associated with the yellow lineage (shown here by two dots),

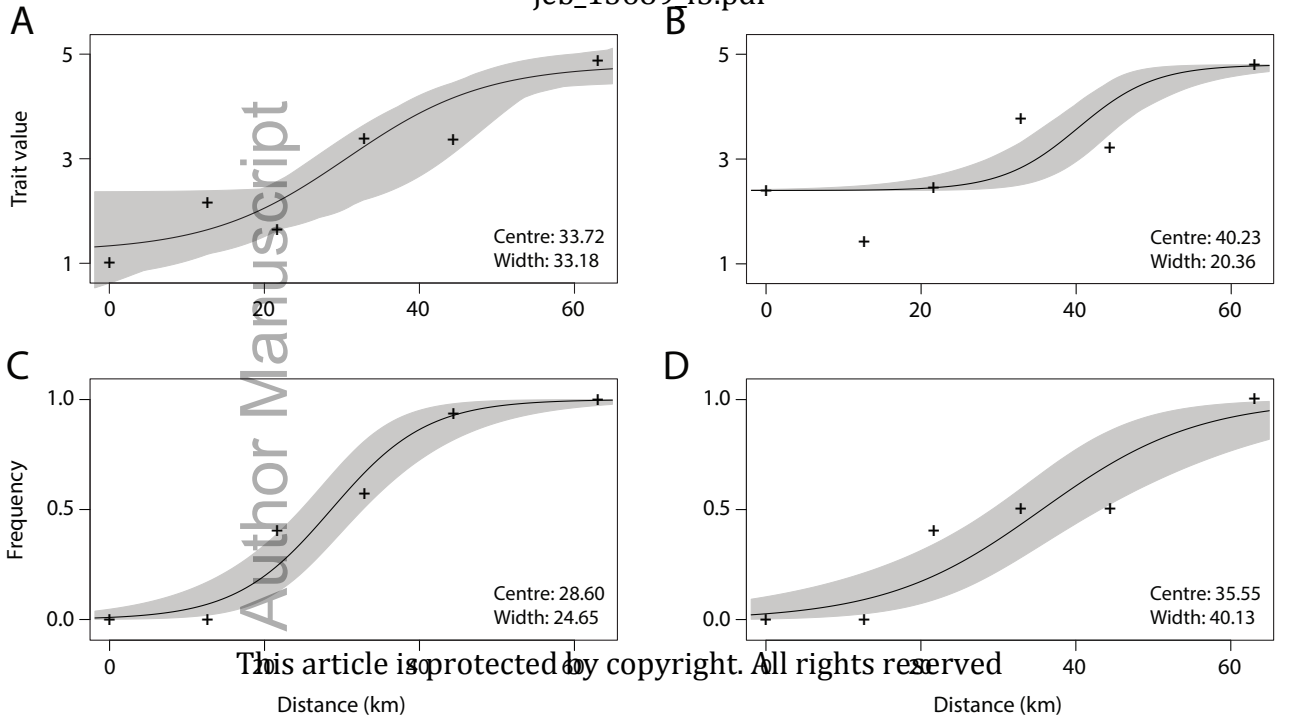
771 which is the presumed source of the Southern lineage in the Contas River basin via river
772 capture.

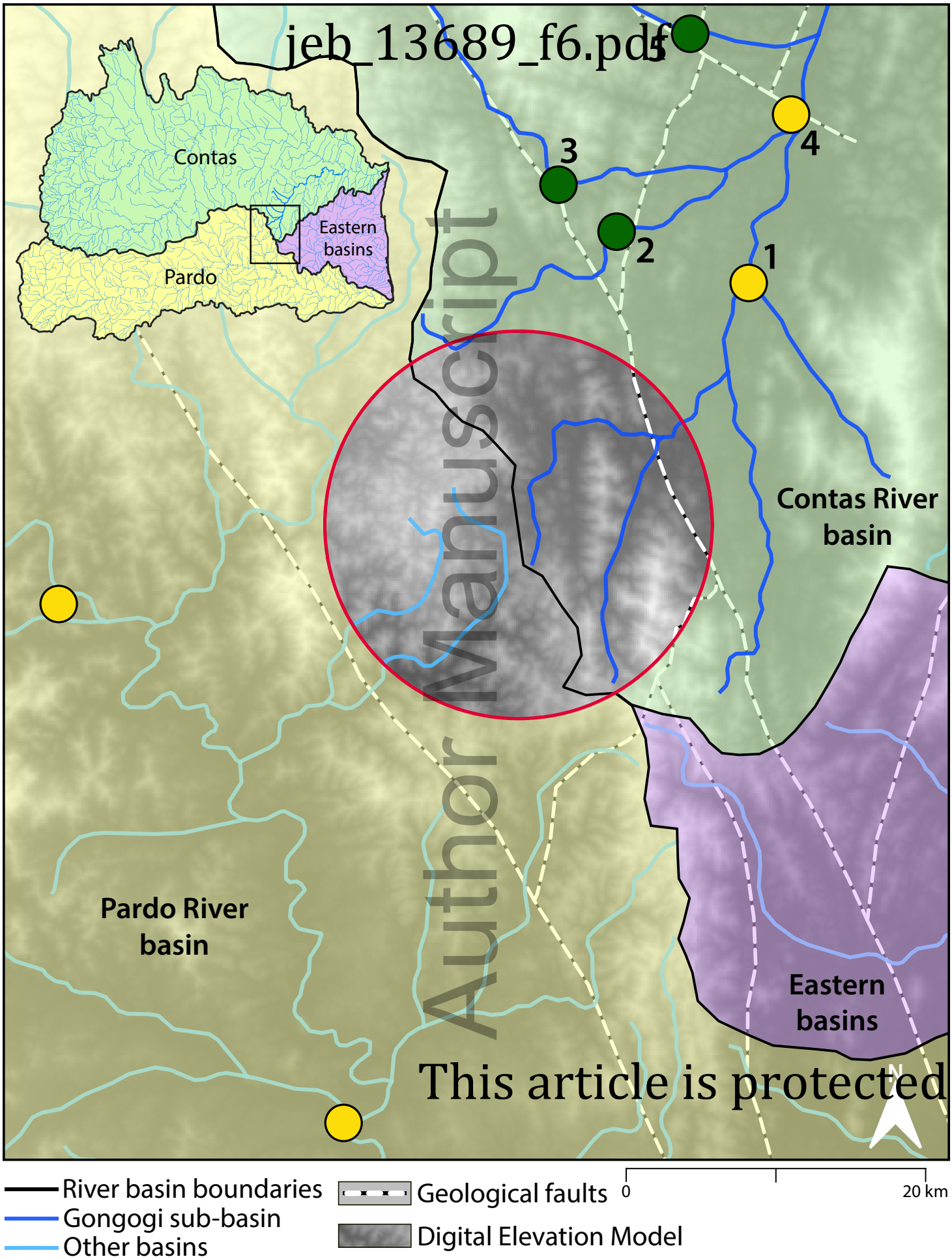











jeb_13689_f6.pdf
Contas
Eastern basins
Pardo
1
2
3
4
5
Contas River basin
Pardo River basin
Eastern basins
Author Manuscript
This article is protected b
River basin boundaries
Gongogi sub-basin
Other basins
Geological faults
Digital Elevation Model
0
20 km