scientific reports

Check for updates

OPEN

DNA barcoding and phylogeography of the *Hoplias malabaricus* species complex

Karen Larissa Auzier Guimarães[1,2,3], Marcos Prado Lima[4], Diego José Santana[5], Mendelsohn Fujiie Belém de Souza[3], Rômulo Sarmento Barbosa[1] & Luís Reginaldo Ribeiro Rodrigues[1,2,3]✉

Hoplias malabaricus (Bloch, 1794) is a carnivorous fish species widely distributed from northern to southern South America. This taxon is believed to be a good model for the investigation of biogeographic events that shape the ichthyofauna evolution in the Neotropical freshwater systems. However, many studies have revealed that *H. malabaricus* hides a species complex that hampers its taxonomic identity and limit its practical value for evolutionary and biogeographic studies. In this paper, we used the mitochondrial gene *cytochrome c oxidase subunit I* (COI) to delimit cryptic species and explore the phylogeography of *H. malabaricus* sensu stricto. We found genetic evidence for putative new species in the genus *Hoplias* and showed that *H. malabaricus* (Bloch, 1794) is a major clade assigned to barcode index number (BIN) BOLD:ABZ3047. This species is structured in six subpopulations differentiated by high Fst values and restricts gene flow. The subpopulations of the São Francisco/East Atlantic/Eastern Northeast Atlantic/Parnaíba/Itapecuru River basins and Tapajós River Basin were the most differentiated and showed demographic fluctuations. The present distributional pattern is most likely explained through a scenario from the Pleistocene.

The trahiras are carnivorous fish of the family Erythrinidae (Characiformes) classified into three genera: *Erythrinus* Scopoli 1777, *Hoplerythrinus* Gill, 1896 and *Hoplias* Gill, 1903. The latter encompasses 15 species tentatively arranged in three species groups: the *Hoplias aimara* group, *H. lacerdae* group and *H. malabaricus* group. In this study we adopted the taxonomy of *H. malabaricus* group as Cardoso and colleagues[1]: *H. malabaricus* (Bloch, 1794), *H. microlepis* (Günther, 1864), *H. teres* (Valenciennes, 1847), *H. mbigua* Azpelicueta, Benítez, Aichino & Mendes, 2015, *H. misionera* Rosso, Mabragaña, González-Castro, Delpiani, Avigliano, Schenone & Díaz de Astarloa, 2016, and *H. argentinensis* Rosso, González-Castro, Bogan, Cardoso, Mabragaña, Delpiani & Díaz de Astarloa, 2018.

Hoplias malabaricus is widespread in the South American hydrographic basins and is frequently recorded along the Orinoco, Amazonas, São Francisco, Paraná, Paraguay and Prata river basins, where it inhabits lentic waters from lagoons, rivers and reservoirs[2–4].

Cytogenetic data revealed karyotypic polymorphisms within *H. malabaricus* populations with chromosome morphology and diploid number variation (2n = 39–42), differentiated sex chromosome systems and other cytogenetic markers such as C-bands and Ag-NORs[2,5,6]. In the Amazon basin, five distinct karyomorphs (types A, C, E, F and G) were detected, including some highly dispersed through neighboring basins (types A, C and F) and others restricted (types E and G)[2]. Chromosomes were important to demonstrate intraspecific divergence of *H. malabaricus* populations and raise the hypothesis of a species complex hidden in this taxon[2,5–7] that was corroborated with molecular data[8,9].

Progress in molecular systematics has directly contributed to improving the species discovery and taxonomy resolution[10–12]. DNA barcoding is a high throughput method that uses a standardized sequence of the mitochondrial gene *cytochrome c oxidase subunit I* (COI) to delimit biological species and also to reveal if there is a phylogenetic structure within a species[13], which has been successfully applied to address taxonomic questions

[1]Programa de Pós-Graduação em Recursos Naturais da Amazônia (PPGRNA), Universidade Federal do Oeste do Pará (UFOPA), Santarém, Pará, Brazil. [2]Programa de Pós-Graduação em Biodiversidade e Biotecnologia (REDE BIONORTE), Universidade Federal do Oeste do Pará (UFOPA), Santarém, Pará, Brazil. [3]Laboratório de Genética e Biodiversidade (LGBio), Universidade Federal do Oeste do Pará (UFOPA), Santarém, Pará, Brazil. [4]Laboratório de Biologia Molecular Animal, Instituto de Ciências e Tecnologia das Águas (ICTA), Universidade Federal do Oeste do Pará (UFOPA), Santarém, Pará, Brazil. [5]Instituto de Biociências, Universidade Federal de Mato Grosso do Sul (UFMS), Campo Grande, Mato Grosso do Sul, Brazil. ✉email: luisreginaldo.ufpa@hotmail.com

of the Neotropical ichthyofauna[10,11,14]. DNA barcoding revealed markedly deep divergence of COI sequences between *H. malabaricus* populations, which supports the hypothesis that some mitochondrial lineages may constitute putative independent species[1,9].

Cardoso and colleagues[1] delimited 16 mitochondrial lineages of the *H. malabaricus* species complex to further investigate their taxonomic status. Recently, three of these lineages (BINs: ACO5223, AAZ3734, AAB1732) were recognized as valid species: *H. mbigua*, *H. misionera* and *H. argentinensis*[15–17]. On the other hand, the remaining cryptic diversity enclosed in the *H. malabaricus* complex wait for investigation. At least seven of the *H. malabaricus* mitochondrial lineages were recorded in the Amazon basin (BINs: AAB1732, ABZ3046, ABZ3047, AAB1731, ACF3787, ACK2158 and ADG3393).

Understanding population divergences on a regional scale is invaluable to disentangle intricate questions about the Neotropical ichthyofauna evolution. Widely distributed fish species, such as *H. malabaricus*, can be considered suitable models for biogeographic studies because they are exposed to extensive evolutionary and ecological drives that promote dispersal events[18–20]. On the other hand, the presence of cryptic species in the *H. malabaricus* complex is a confounding factor that limits its value for phylogeographic and population genetic studies. Additionally, a precise delimitation of cryptic species may strengthen the inferences in both basic and applied research fields, for example in conservation measures, ecological risk assessment and climate change effects on biodiversity[21].

In this study, we assembled an extensive database of DNA barcodes including new sequences from the Amazon basin and adjacent drainages, aiming to delimit mitochondrial lineages representative of formally described species and discriminate putative cryptic species of the *H. malabaricus* complex. We delimited *H. malabaricus* (Bloch, 1794) "sensu stricto" following criteria in Cardoso and colleagues[1] and evaluated it for genetic structure, phylogeography and demographic history.

Methods

Ethics statements.
Fish specimens were collected under a Brazilian Government of the Sistema de Autorização e Informação em Biodiversidade (SISBIO), permission N. 24384-1. For tissue and voucher preservation, the specimens were anesthetized and euthanized by exposure in eugenol solution for a few minutes until the complete stop of the opercula beats. All procedures were approved by Comissão de Ética no Uso de Animais (CEUA) of Universidade Federal do Oeste do Pará (N 09003/2016) and followed all relevant guidelines. Additionally, the study was carried out in compliance with ARRIVE guidelines.

Sampling and study area.
We sampled 153 *H. malabaricus* specimens from 38 localities in the Amazon Basin (Brazil, Peru), the Araguaia-Tocantins basin (Brazil), the Western Northeastern Atlantic Basin (Brazil), the Guiana Shield drainages (Guiana) and the Orinoco Basin (Venezuela, Colombia). The fish were collected using seine nets, casting nets and fish hooks. Samples of epaxial muscle were preserved in absolute ethanol and stored at -20 °C. The specimens were fixed with 10% formalin for 48 h, washed and preserved in 70% ethanol. Specimens were later deposited in the Fish Collection of the Instituto de Ciências e Tecnologia das Águas and the Laboratório de Genética e Biodiversidade, Universidade Federal do Oeste do Pará (Brazil).

DNA extraction, PCR and sequencing.
DNA extraction followed a salting-out protocol and the amounts were evaluated in a 1% agarose gel stained with Gelred (Biotium)[22,23]. DNA fragments of *cytochrome c oxidase subunit I* (COI) mitochondrial gene were amplified using the standard DNA barcoding primers Fish F1 and Fish R1[24]. The reactions were assembled in 25 μL, containing 15 μL sterile H_2O, 2.8 μL dNTP mix (1.25 mM), 2.5 μL buffer 10 × (200 mM Tris–HCl (pH 8.4) + 500 mM KCl), 2.5 μL $MgCl_2$ (50 mM), 0.5 μL of each primer (5 μM), 0.2 μL Taq DNA polymerase (5 U/μL) and 1 μL of genomic DNA (50–100 ng). The cycling profile was as follows: 95 °C/2 min, 35 cycles of 94 °C/30 s, 54 °C/30 s and 72 °C/1 min, and a final step of 72 °C/10 min. The amplifications were performed with a Pxe 0.2 thermocycler (Thermo Scientific) and the amplified products were evaluated in a 1% agarose gel stained with Gelred. PCR products were purified with PEG8000 protocol[25]. COI sequences were obtained by the Sanger method using the ABI PRISM Big Dye Terminator V.3 Cycle Sequencing kit (Applied Biosystems, Waltham, Massachusetts, USA). Sequencing reactions were perfomed in 96-well plates with a final volume of 10 μL, containing 5 μL of sterile H_2O, 1.5 μL of sequencing buffer 5 ×, 0.5 μL of primer (10 μM), 1 μL of Big Dye mixture and 2 μL of cleaned PCR. The dye incorporation reactions followed 96 °C/1 min; 35 cycles of 96 °C/15 s, 50 °C/15 s, and 60 °C/4 min. The plates were precipitated in ethanol/EDTA, eluted with 10 μL Formamida Hi-Di and detected with an ABI 3500 genetic analyser (Applied Biosystems), following the manufacturer's instructions.

Molecular data analysis and species delimitation.
The DNA barcode dataset was enriched with 286 COI sequences downloaded from the public repositories: Barcode of Life Database (https://www.boldsystems.org) and GenBank (https://www.ncbi.nlm.nih.gov). To obtain higher accuracy for cryptic species delimitation with barcodes we included sequences representative of *H. malabaricus*, *H. misionera*, *H. mbigua*, *H. microlepis*, *H. argentinensis*, *H. intermedius*, *H. aimara*, *H. australis*, *H. lacerdae* and *H. curupira*. The geographic coverage of the *H. malabaricus* complex was enlarged in the following regions: Lower Amazonas Rivers[9,11], the East Atlantic, the Eastern Northeast Atlantic, Parnaíba River, São Francisco and the Itapecuru River basins[8] the Madeira River in Brazil and Bolivia, and drainages from the Guiana shields in Surinam and French Guiana[1]. Detailed information on the DNA barcoding dataset and specimen metadata is listed in Supplementary Data S1.

The consensus sequences were assembled using Geneious R7 (Biomatters, New Zealand) and aligned with Clustal W v1.4[26]. We used GBLOCKS v0.91b[27] to inspect the alignment and trim the sequence tips and poorly

aligned regions. The sequences generated in this work were deposited in a DNA barcoding repository (http://www.boldsystems.org) linked to the Project "Amazonian Trahiras (AMTRA)" (see Supplementary Data S2).

We delimited species by: (1) barcode index number (BIN)[28], (2) generalized mixed Yule coalescent (GMYC)[29,30] and (3) automatic barcode gap discovery (ABGD)[31]. The BIN analysis is an automated process implemented in the platform www.boldsystems.org, which is based on genetic distances to identify clusters of query sequences against a DNA barcode library; such clusters are coded as BIN numbers and interpreted as operational taxonomic unit (OTUs) that represent a species. To perform the GMYC analysis, we removed haplotype duplicates with ElimDupes (https://www.hiv.lanl.gov/content/sequence/elimdupesv2/elimdupes.html) and made an ultrametric tree using BEAST v1.8.0[32], following these parameters: evolutionary model HKY + I + G chosen with jModelTest[33], molecular clock lognormal relaxed, Yule speciation process. *Leporinus amblyrhynchus* Garavello & Britski 1987 was adopted as the outgroup. We ran a Bayesian reconstruction with 80 million MCMC iterations, sampled every 1000 iterations with a burn-in of 10%. The tree convergence and stability were checked with the software Tracer v.1.7.1[32], retaining an effective sample size (ESS) > 200. The trees were combined with TreeAnotator v1.8.0[32], and the output file was saved in Newick tree format to be used for GMYC delimitation. The analysis of coalescence/speciation (GMYC) was processed following the model single threshold, in the environment R 3.4.3[34] supplemented with libraries *Splits* (Species Limits by Threshold Statistics)[35] and *Ape* (Analyses of Phylogenetics and Evolution in R language)[36]. ABGD was processed on the platform www.bioinfo.mnhn.fr/abi/public/abgd/abgdweb.html using the alignment data set (fasta file) as the input file. We set the parameters: model K80, Pmin. 0.001, Pmax 0.01 and barcoding gap width X = 0.2.

To integrate the phylogenetic information, species delimitation and divergence time we processed a second Bayesian reconstruction following the procedures mentioned above with minor modifications: 200 million MCMC iterations sorted at each 1000 and 10% of burn-in and a strict clock model. Divergence times were calibrated with a mutation rate of 1% per million years (Myr), which is conservative for fish mtDNA[37,38]. The resulting trees were assembled with TreeAnotator and the topology was visualized/edited with FigTree v1.2.2 (http://tree.bio.ed.ac.uk/software/figtree/). Pairwise genetic distances between delimited species were measured following the K80 model[39] using the software MEGA X software[40].

Population genetics and phylogeography. The individuals assumed to be *Hoplias malabaricus* sensu stricto (BIN ABZ3047) following designation proposed in Cardoso and colleagues[1] were investigated for intraspecific genetic diversity and population structure. We used the software GENELAND R package v. ≥ 4.0.0 (http://www2.imm.dtu.dk/gigu/Geneland/)[41] implemented with *R* v3.4.0[34] to investigate population subdivisions and to find the geographic population units, based on Bayesian statistics. The population genetic structure was evaluated through F_{ST} statistics and molecular variance analysis (AMOVA) implemented with Arlequin v.3.1[42]. We assumed populations as the clusters of individuals such as revealed by Geneland analysis. For F_{ST} divergence, we follow Wright and colleagues[43] categories: low (0.00–0.05), moderate (0.05–0.15), high (0.15–0.25) and elevated (> 0.25). Parameters of the population genetics (e.g. haplotypes, nucleotide diversity, polymorphic sites) were analysed with DNAsp v.6[44]. A haplotype network was constructed based on median joining algorithm[45] with the assistance of PopART software[46].

To explore the demographic history, we applied neutrality tests Tajima's D[47] and Fu's Fs[48], implemented with Arlequin v.3.1[42]. Additionally, to detect population size variations we investigated the mismatch distributions and Bayesian skyline plot (BSP). These analyses were implemented with DNAsp v.6[44] and BEAST v.1.8.0[32]. BSP analysis adopted the HKY + I + G model and 100 million MCMC sorted each 1000 iterations.

We constructed an ecological niche model with the maximum entropy algorithm MAXENT version 3.3.3 k[49–51] based on 82 georeferenced occurrence points (Fig. 6a) and 19 bioclimatic variables from WorldClim (https://www.worldclim.org/data/bioclim.html). Such variables were correlated with a 2.5 arc-minute spatial scale[51], and the distributional limits were assessed from the median occurrence with 50 bootstrap pseudoreplicates. The theoretical distributional patterns were visualized with QGIS v.3.16.8-Hannover (Quantum GIS Development Team, www.qgis.org). We used jackknife permutations to evaluate the model performance gain and to identify and retain the most relevant explanatory variables.

Results

DNA barcoding and species delimitation. We analysed 439 COI sequences from 10 species of the genus *Hoplias*. The sequences were 621 bp long without stop codons or indels. The dataset showed a base composition of 29.4% (T), 29.5% (C), 23.3% (A) and 17.7% (G).

Phylogenetic Bayesian inference showed three major clades in *Hoplias*. The largest clade encompasses the species assembled to the *H. malabaricus* group that emerged more recently; the second clade is configured with the species from the *H. lacerdae* group, which surprisingly nested five individuals deposited as *H. malabaricus* (Panamá clade); the most basal clade is represented by *Hoplias curupira* lineages (Fig. 1). All the lineages received strong support from the Bayesian genealogic inference (100% posterior probability—PP), with the exception of the *H. malabaricus* lineages that were assembled to BIN ABZ3047 supported by 65% PP (Fig. 1).

Based on the present DNA barcode library of *Hoplias* we delimited a varied number of species with different methods: ABGD (28sp), BIN (33sp) and GMYC (42 sp). The pairwise genetic distances between groups (BINs) ranged from 1.1 to 21.5%, while the intra-BIN distances ranged from 0 to 1.6% (Table 1). The largest distances were detected for *Hoplias curupira* lineages. We found total congruence for delimiting six groups that match unequivocally to species: *H. lacerdae*, *H. australis*, *H. intermedius*, *H. argentinensis*, *H. misionera* and *H. microlepis*. Additionally, 17 putative new species were delimited with all the methods applied (BIN, ABGD and GMYC). On the other hand, discrepant species delimitation was observed and showed taxonomic uncertainty on the following species or clades: *H. malabaricus* (Panama Clade) was delimited with three species (BIN ADR2428,

Figure 1. COI mitochondrial gene tree recovered from a Bayesian inference analysis in the genus *Hoplias* (Erythrinidae). All clades highlighted are the evolutionary lineages recovered in BIN species delimitation. Black bars on the right show the partitions inferred as putative species delimited through, BIN, ABGD and GMYC analysis. Values in the nodes indicate statistical support of posterior probability. The tree was drawn with FigTree v1.2.2 (http://tree.bio.ed.ac.uk/software/figtree/).

Groups barcode index number (species) [n]	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34
1 ABZ3047 [247]	**0.016**																																	
2 AAB1733 [12]	0.080	**0.002**																																
3 AAD3630 [2]	0.044	0.074	**0.013**																															
4 AAY4779 [6]	0.087	0.066	0.078	**0.004**																														
5 ACI3811 [8]	0.094	0.084	0.086	0.065	**0.003**																													
6 ACR9466 [46]	0.028	0.082	0.051	0.088	0.091	**0.001**																												
7 AEF4663 [2]	0.043	0.077	0.054	0.087	0.099	0.053	**0**																											
8 AAB1731 [4]	0.025	0.071	0.048	0.076	0.092	0.032	0.043	**0.009**																										
9 ABZ3046 [2]	0.034	0.080	0.042	0.079	0.100	0.039	0.041	0.030	**0.002**																									
10 ACF3787 [22]	0.026	0.075	0.042	0.075	0.085	0.035	0.041	0.019	0.024	**0.004**																								
11 ADL3159 [15]	0.030	0.075	0.043	0.082	0.088	0.031	0.032	0.030	0.029	0.028	**0.001**																							
12 AEA9444 [1]	0.039	0.076	0.051	0.085	0.094	0.047	0.015	0.038	0.035	0.036	0.030	–																						
13 ADG3393 [4]	0.063	0.088	0.067	0.087	0.096	0.066	0.057	0.062	0.064	0.059	0.057	0.056	**0.004**																					
14 ACK2158 [2]	0.061	0.089	0.061	0.092	0.100	0.061	0.052	0.058	0.045	0.053	0.046	0.058	0.074	**0.005**																				
15 ACK8876 [1]	0.058	0.098	0.064	0.096	0.097	0.070	0.055	0.055	0.051	0.050	0.049	0.050	0.079	0.075	–																			
16 AAI8240 [6]	0.040	0.075	0.054	0.079	0.108	0.047	0.051	0.031	0.026	0.036	0.040	0.047	0.072	0.055	0.063	**0.001**																		
17 AAI8239 [2]	0.044	0.071	0.044	0.074	0.072	0.054	0.041	0.041	0.037	0.036	0.037	0.036	0.062	0.059	0.030	0.045	**0**																	
18 AEA5279 [3]	0.032	0.083	0.052	0.087	0.103	0.039	0.046	0.022	0.036	0.030	0.036	0.047	0.064	0.062	0.057	0.043	0.045	**0.003**																
19 AAZ3734 (H. argentinensis) [5]	0.072	0.075	0.070	0.080	0.081	0.074	0.070	0.073	0.075	0.069	0.074	0.072	0.091	0.073	0.094	0.076	0.075	0.077	**0.001**															
20 ACO5223 (H. mbigua) [5]	0.044	0.078	0.046	0.079	0.076	0.052	0.038	0.043	0.035	0.038	0.032	0.034	0.063	0.057	0.031	0.050	0.011	0.046	0.075	**0.003**														
21 AAD3629 (H. microlepis) [5]	0.055	0.081	0.038	0.082	0.080	0.060	0.060	0.055	0.049	0.050	0.057	0.059	0.077	0.057	0.071	0.061	0.050	0.060	0.066	0.052	**0.003**													
22 AAB1732 (H. misionera) [5]	0.071	0.081	0.064	0.078	0.089	0.076	0.060	0.076	0.071	0.070	0.068	0.063	0.077	0.082	0.086	0.081	0.065	0.079	0.074	0.064	0.082	**0.004**												
23 ADR3966 [2]	0.171	0.157	0.158	0.152	0.160	0.169	0.159	0.165	0.167	0.160	0.170	0.171	0.168	0.152	0.193	0.175	0.167	0.157	0.147	0.168	0.155	0.155	**0.002**											
24 AAB1730 [1]	0.169	0.154	0.155	0.151	0.157	0.165	0.156	0.161	0.168	0.158	0.166	0.167	0.173	0.148	0.189	0.172	0.164	0.153	0.148	0.166	0.156	0.154	0.024	–										
25 ADR2428 [2]	0.168	0.153	0.155	0.147	0.150	0.167	0.157	0.160	0.162	0.155	0.168	0.168	0.172	0.147	0.184	0.171	0.159	0.157	0.147	0.159	0.155	0.148	0.015	0.021	**0.002**									
26 AAB1734 (H. intermedius) [10]	0.143	0.162	0.149	0.147	0.138	0.146	0.156	0.144	0.154	0.140	0.150	0.160	0.148	0.162	0.176	0.162	0.154	0.146	0.137	0.152	0.162	0.131	0.158	0.166	0.158	**0.009**								
27 ADG3392 (H. curupira) [2]	0.180	0.200	0.188	0.198	0.199	0.173	0.189	0.183	0.186	0.176	0.171	0.193	0.181	0.190	0.204	0.181	0.186	0.185	0.185	0.183	0.184	0.197	0.209	0.215	0.216	0.184	**0**							
28 ADG3181 (H. curupira) [2]	0.182	0.198	0.186	0.199	0.195	0.173	0.175	0.187	0.187	0.178	0.175	0.184	0.192	0.195	0.199	0.184	0.182	0.190	0.181	0.177	0.180	0.195	0.214	0.215	0.212	0.195	0.035	**0**						
29 ADG3391 (H. curupira) [1]	0.153	0.181	0.163	0.175	0.186	0.150	0.165	0.156	0.158	0.152	0.143	0.164	0.161	0.162	0.182	0.155	0.161	0.156	0.173	0.159	0.161	0.185	0.205	0.203	0.209	0.183	0.021	0.040	–					
30 ACD9164 (H. australis) [5]	0.153	0.162	0.161	0.149	0.152	0.165	0.151	0.159	0.163	0.156	0.166	0.159	0.148	0.161	0.182	0.166	0.158	0.162	0.140	0.156	0.153	0.151	0.147	0.143	0.143	0.112	0.176	0.184	0.179	**0.003**				
31 ABW2258 (H. lacerdae) [5]	0.147	0.147	0.153	0.136	0.134	0.156	0.142	0.150	0.161	0.148	0.155	0.153	0.129	0.165	0.175	0.160	0.154	0.157	0.140	0.153	0.158	0.135	0.146	0.145	0.146	0.113	0.170	0.179	0.170	0.059	**0.003**			
32 ADE1357 (H. aimara) [2]	0.168	0.167	0.167	0.171	0.169	0.179	0.166	0.170	0.160	0.161	0.173	0.172	0.176	0.177	0.192	0.164	0.162	0.172	0.163	0.165	0.171	0.154	0.164	0.171	0.176	0.119	0.204	0.221	0.201	0.149	0.138	**0**		
33 ADG3375 (H. aimara) [2]	0.159	0.168	0.159	0.170	0.159	0.173	0.156	0.160	0.159	0.151	0.164	0.157	0.175	0.177	0.178	0.159	0.149	0.162	0.162	0.151	0.165	0.145	0.172	0.170	0.175	0.126	0.199	0.206	0.190	0.150	0.136	0.022	**0.002**	
34 Outgroup [3]	0.214	0.240	0.196	0.209	0.218	0.211	0.198	0.216	0.215	0.214	0.211	0.214	0.223	0.193	0.227	0.216	0.205	0.214	0.203	0.209	0.210	0.201	0.228	0.226	0.229	0.210	0.224	0.231	0.217	0.213	0.212	0.234	0.223	**0.004**

Table 1. Pairwise genetic distances (K2P) between and within species of *Hoplias* based on COI sequences. Bold values indicate the intra-BIN distance. Number of individuals are shown in brackets.

ADR3966 and AAB1730) but with two species (ABGD, GMYC). *H. mbigua* was delimited with three species (BIN, GMYC) (ACO5223, AAI8239 and ACK8866) but with two species (ABGD). The specimens assigned to *H. malabaricus* were assembled in 11 putative species (BINs: ADG3393, AEA4944, AEF4663, ADL3159, ABZ3046, AAI8240, AAB1731, AEA5279, ACF3787, ACR9466 and ABZ3047); however, ABGD supported eight species while GMYC delimited 21 species hidden in this taxon. Therefore, the *H. malabaricus* species complex was clearly evidenced from distinct DNA barcode analysis.

We recognized *H. malabaricus* (Bloch, 1794) as the clade that nested the individuals from Suriname, since it is the type locality of this species. Such clade was delimited as BIN ABZ3047. To explore the intraspecific genetic relationships, we assumed BIN ABZ3047 as a single species, which is hereinafter referred to as *H. malabaricus* sensu stricto following Cardoso et al. (2018). This species comprises eleven entities widely distributed throughout the Amazon Basin and adjacent drainages, including a subclade that is restricted to the Guyana shield drainages. An updated distribution map of the *Hoplias malabaricus* species group based on records of barcoded individuals is provided and the geographic occurrence of news records of putative species is highlighted (Fig. 2).

Population genetics and phylogeography of *H. malabaricus* sensu stricto.

Based on the DNA barcodes of 247 individuals we observed 79 polymorphic sites and 66 haplotypes (see Supplementary Data S3). The spatial variation of genetic data indicated a distribution of six subpopulations (SPOP): (1) Lower Amazonas River confluences/Xingu River/Rupununi River (SPOP1), (2) Madeira/Purus Rivers (SPOP2), (3) São Francisco/East Atlantic/Eastern Northeast Atlantic/Parnaíba/Itapecuru River basins (SPOP3), (4) Guiana Shields drainages (SPOP4), 5) Western-Northeast Atlantic/Tocantins basin (SPOP5), (6) Tapajós River Basin (SPOP6) (Fig. 3).

The *H. malabaricus* sensu stricto subpopulations presented high genetic diversity counting from 6 to 19 haplotypes and haplotypic diversity (h) from 0.230 (SPOP6) to 0.956 (SPOP2). The nucleotide diversity (π) was higher in SPOP1 (0.0187) and lowest in SPOP6 (0.0005). Two subpopulations (SPOP3 and SPOP6) resulted in negative and statistically significant Fu's Fs and Tajima's D values that were interpreted as evidence of neutrality deviation by purifying selection or population expansion. For the other subpopulations, both the neutrality tests suggested the long-term population stability (Table 2).

The haplotype network showed that only two haplotypes were shared between *H. malabaricus* sensu stricto subpopulations (SPOP1-SPOP2 and SPOP1-SPOP5). Three haplogroups were depicted and coincide with SPOP3, SPOP4 and SPOP6, which showed only private haplotypes demonstrating higher genetic differentiation (Fig. 4).

A pairwise Fst comparison demonstrates that subpopulation SPOP6 is the most differentiated (Table 3). The lowest Fst value (0.407) was recorded for the SPOP6xSPOP1 comparison and the highest differentiation was detected to SPOP6xSPOP2 (0.545). AMOVA results demonstrated 27.20% of variation was detected among populations and 72.80% within populations (Table 4).

In the mismatch distribution analysis among the population groups, the SPOP3 and SPOP6 presented a unimodal curve and were suggestive of demographic fluctuations, whereas the SPOP1, SPOP2, SPOP4 and SPOP5 populations displayed multimodal curves (Fig. 5a–f). Bayesian skyline plots (BSP) provided a signal of long-term demographic stability (Fig. 5a–f). For the SPOP1, the BSP showed continuous and discrete historical population size growth approximately 950 thousand years before present (kybp), with a decrease with the onset of the 150 kybp, after which it remained stable for a short time before increasing to the current population size (Fig. 5a). For SPOP2, the BSP plot suggests expansion time estimated onset 600 kybp, followed by a period of stability starting at 200 kybp and shortly after very slight tendency of decrease approximately 50 kybp (Fig. 5b). The SPOP3 and SPOP4 started a period of expansion at 350 kybp, and evidence of a substantial decline started at 50 kybp (Fig. 5c,d). SPOP5 demonstrated a fluctuation, with rapid population expansion starting at 150 kybp (Fig. 5e). For SPOP6, a remarkably progressive increase in population size started at 35 kybp, and the lightest tendency of increase occurred at the start of the Holocene (Fig. 5f).

For the niche model, the mean diurnal range [mean of monthly (max temp – min temp)] was the most important variable driving the *Hoplias malabaricus* sensu stricto distribution. The other most important variables were the annual temperature range, mean temperature of the wettest quarter, mean temperature of the driest quarter and precipitation seasonality (coefficient of variation). The prediction is *H. malabaricus* was potentially distributed congruent with the present-day geographical distribution. The areas with the highest levels of suitability were eastern Amazon Basin, Guiana Shield and northeast Brazil. Spatial displacements did not seem to occur, but the range size population increased over time (see Fig. 6).

Discussion

The species of *Hoplias* tend to be morphologically conserved and its taxonomy is still under debate[15,17]. Despite its conspicuous karyotypic variation, cytogenetic markers have shown poor resolution for taxonomy and phylogenetic reconstruction[2,8,9]. Recent progress has been achieved with DNA barcoding coupled to morphological traits (integrative taxonomy), which starts to delineate a clearer picture of the evolutionary history of this group[15,17,52].

This study revealed that *Hoplias* encompasses at least 23 species and hides new cryptic species in both the *H. lacerdae* and *H. malabaricus* groups. Additionally, taxonomic confusion possibly caused by misidentifications could be detected. For instance, *H. intermedius* (BIN AAB1734) from the Jequitinhonha River (Minas Gerais, Brazil) was deposited as *H. brasiliensis*. Two independent evolutionary units were recovered for *H. aimara* and three for *H. curupira* as previously demonstrated[1]; herein we show that *H. aimara* belongs to the *H. lacerdae* group. Instead of *H. curupira*, which is the most basal clade and could receive a species-group status for the current *Hoplias* classification.

The conservative interspecific genetic distance of 2% for COI sequences was considered imprecise to delimit species from the megadiverse Neotropical ichthyofauna (see Queiroz and colleagues[53]). Using a single marker
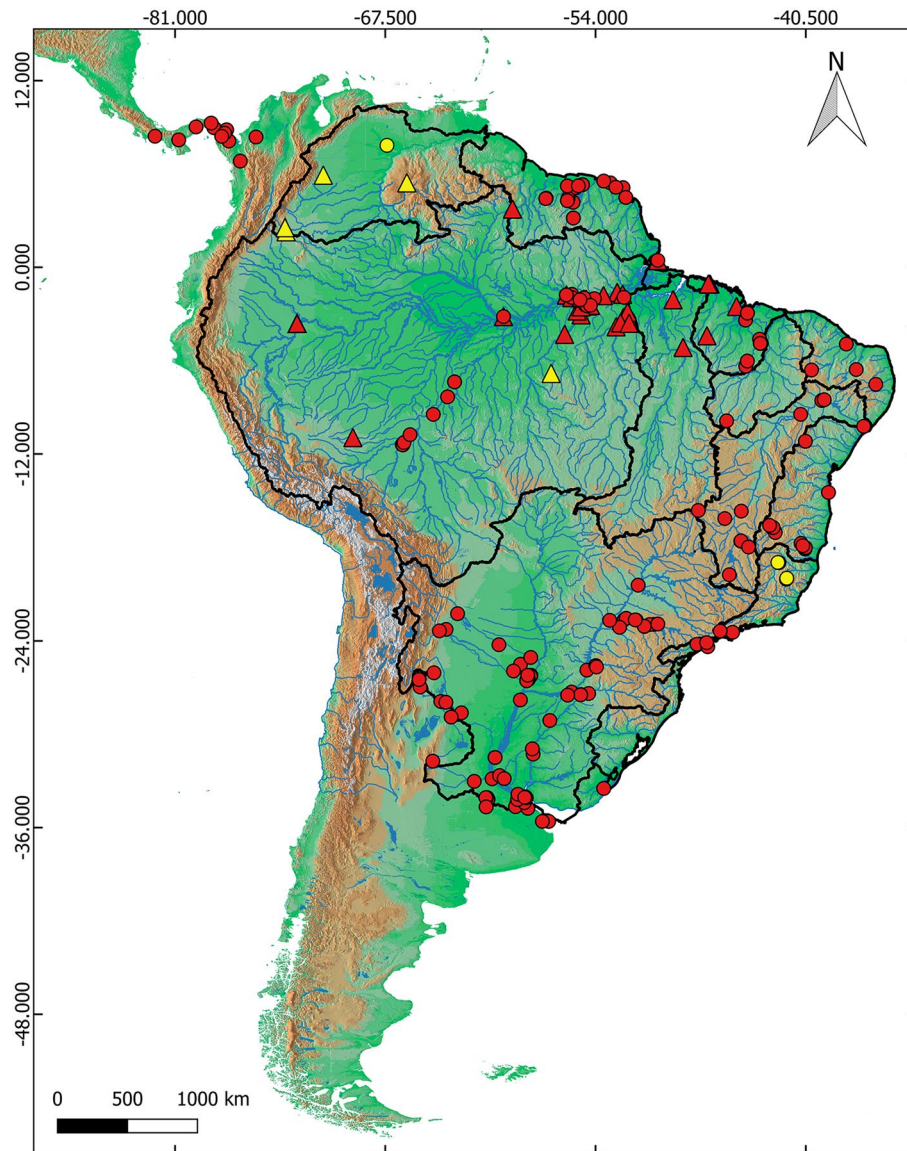
Figure 2. Geographic distribution of all genetic samples from species belonging to the *Hoplias malabaricus* species-group in South America. Localities sampled in this study (triangles) and localities of sequences mined from the internet (circles). The distribution of new putative species found in this study is in yellow. The map was done with QGIS v.3.16.8-Hannover (Quantum GIS Development Team, http://www.qgis.org).

species delimiting approach we observed deep genetic divergence in *Hoplias*. The species in the *H. malabaricus* group diverged by a maximum of 10.3%, which was enough to discriminate between sister species formally described and lineages from the *H. malabaricus* complex, which was assumed to be independent evolutionary units (ESUs). This magnitude of genetic divergence supports cryptic speciation in the *H. malabaricus* complex, as revealed by multiple species delimitation analyses (BIN, ABGD and GMYC). Therefore, our results corroborate the existence of the *H. malabaricus* complex of cryptic species inhabiting the Amazon Basin and adjacent drainages[2,5,7,8]. Some of these lineages were presented in previous studies[1,9,52]; however, we could recognize six new putative species from the *H. malabaricus* complex (BINs: AAY4779, AAD3630, AEA5279, ADL3159, AEF4663 and AEA4944).

COI deep divergences, higher than 2%, are a marked signature of lineage/species differentiation in *Hoplias*[1,8,9]. Such a phenomenon has been observed between populations from the same and distinct hydrographical basins[8] and between populations that share identical karyomorphs[9]. *H. misionera* and *H. argentinensis* diverged from their nearest neighbors by 5.6 and 9.0%[15,17]. Jacobina and colleagues[8] similarly recorded deep divergence (7 to 7.3%) delimiting putative species from *H. malabaricus* lineages in distinct Brazilian hydrographic basins. On the other hand, Cardoso and colleagues[1] demonstrated that speciation in *Hoplias* may be accompanied by a large range of COI genetic distances, between approximately 1% and 20%. Indeed, *H. mbigua* diverged from its nearest neighbor by only 1.13%[1].
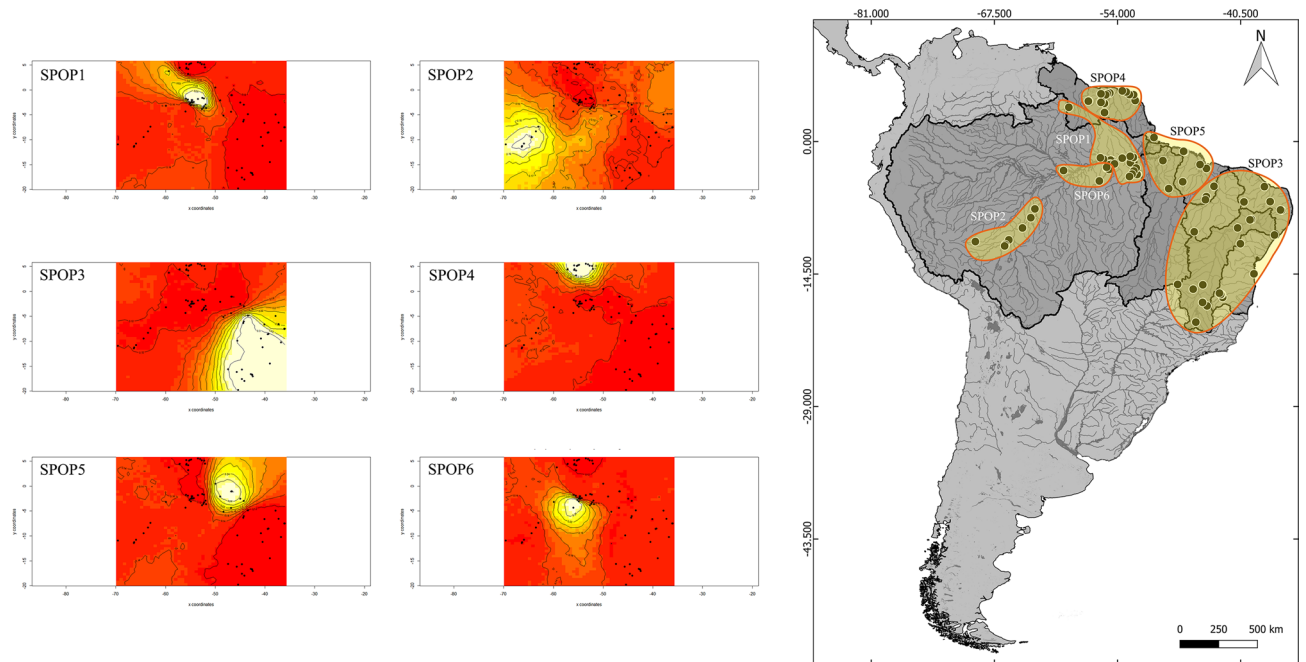
Figure 3. Subpopulations of *Hoplias malabaricus* (BIN ABZ3047) from the Amazon Basin and adjacent drainages inferred with COI haplotypes and geographic data using GENELAND R package v. ≥ 4.0.0 (http://www2.imm.dtu.dk/singrainageseam). The polygons shaded white indicate the largest probabilities of associating haplotypes (black dots) with populations. The map was done with QGIS v.3.16.8-Hannover (Quantum GIS Development Team, www.qgis.org).

Subpopulation	N	Ha	S	*H*	π	Fu's FS	p value	Tajima's D	p value
SPOP1	75	19	39	0.874	0.0187	− 0.09562	0.52300	− 0.21355	0.50100
SPOP2	10	8	16	0.956	0.0114	− 1.69595	0.15100	− 0.06384	0.48100
SPOP3	75	15	20	0.877	0.0047	**− 6.44668**	**0.00800**	**− 1.69915**	**0.01600**
SPOP4	25	9	15	0.760	0.0051	− 2.14428	0.10400	**− 1.50352**	**0.04300**
SPOP5	20	11	21	0.947	0.0156	− 0.99411	0.35500	0.13010	0.59400
SPOP6	42	6	4	0.230	0.0005	**− 4.38750**	**0.00000**	**− 1.76047**	**0.01000**

Table 2. Genetic diversity and values of neutrality tests of *H. malabaricus* subpopulations from the Amazon basin and adjacent drainages, based on mtDNA (COI gene). **N** = individuals, **Ha** = haplotypes, **S** = polymorphic sites, **h** = haplotypic diversity, **π** = nucleotide diversity. Bold values are statistically significant (p < 0.05). Subpopulations: (1) Lower Amazonas River confluences/Xingu River/Rupuruni River (SPOP1), (2) Madeira/Purus Rivers (SPOP2), (3) São Francisco/East Atlantic/Eastern Northeast Atlantic/Parnaíba/Itapecuru River basins (SPOP3), (4) Guiana Shields drainages (SPOP4), (5) Western-Northeast Atlantic/Tocantins basin (SPOP5), (6) Tapajós River Basin (SPOP6).

The BINs AAD3630 and AEA5279 are from Colombia and are composed of individuals from Orinoco basin drainages. The BINs AEA4944 and AEF4663 from the Ventuari-Orinoco were delimited as a single species with ABGD and GMYC. Herein, we consider the taxonomic status of these lineages most imprecise, and their delimitations as putative species must be regarded cautiously. Putative species delimited from a small number of sequences (< 5 individuals) are highly susceptible to bias[31,54]. The ADL3159 lineage was restricted to the Crepori River, a tributary of the middle Tapajós drainage, distantly more than 500 km from the confluence zone between the Tapajós and Amazonas Rivers. This group was supported by distinct methods, and we believe it is a new undescribed species within the *H. malabaricus* complex. Recently, an integrative taxonomic analysis by our team formally describes this taxon (article accepted).

Because of the taxonomic uncertainty of the genus *Hoplias*, particularly in the *H. malabaricus* complex, few studies have focused on population genetics and micro evolutionary processes, which may be caused by imprecise species discrimination under field conditions due to the high genetic diversity (karyomorphs, DNA), morphological similarity and poor taxonomic knowledge. Herein, we successfully identified *H. malabaricus* from cryptic congeners based on DNA barcoding sequences. Although the COI gene has been routinely used for species delimitation[13,55], it can be useful to explore population genetic structuring in freshwater fishes[12,56].
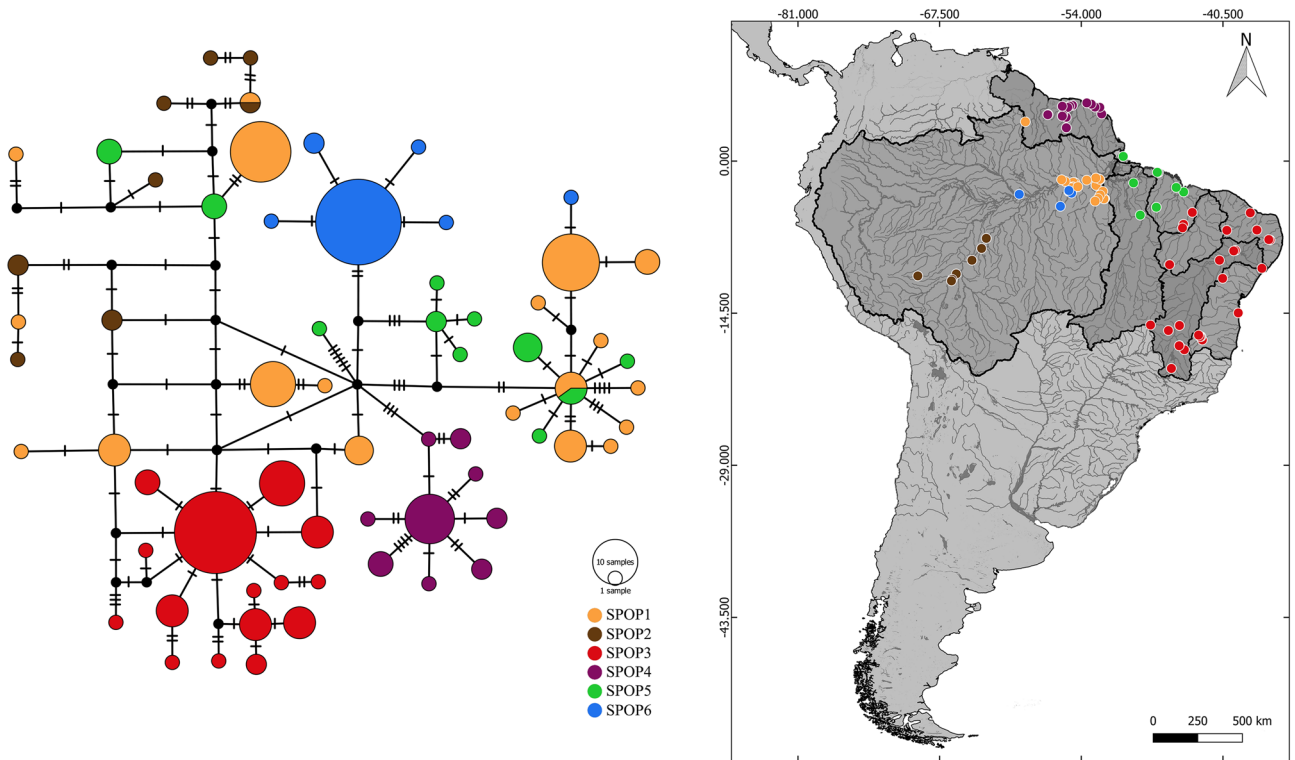
Figure 4. Median joining haplotype network of *Hoplias malabaricus* (BIN ABZ3047) based on the mitochondrial gene COI. The size of circles is proportional to haplotype frequency. Black circles indicate not sampled or possibly extinct haplotypes. Perpendicular bars show the mutational steps. Subpopulations: (1) Lower Amazonas River confluences/Xingu River/Rupuruni River (SPOP1), (2) Madeira/Purus Rivers (SPOP2), (3) São Francisco/East Atlantic/Eastern Northeast Atlantic/Parnaíba/Itapecuru River basins (SPOP3), (4) Guiana Shields drainages (SPOP4), 5) Western-Northeast Atlantic/Tocantins basin (SPOP5), (6) Tapajós River Basin (SPOP6). The network was drawn with PopART (http://popart.otago.ac.nz) and the map was done with QGIS v.3.16.8-Hannover (Quantum GIS Development Team, http://www.qgis.org).

Hierarchical level	F_{ST} matrix				
	SPOP1	SPOP2	SPOP3	SPOP4	SPOP5
SPOP1					
SPOP2	0.09022 (0.00391)				
SPOP3	0.17423 (0.00000)	0.15050 (0.00000)			
SPOP4	0.17667 (0.00000)	0.15419 (0.00000)	0.22995 (0.00000)		
SPOP5	0.09871 (0.00000)	0.06009 (0.01270)	0.15725 (0.00000)	0.15910 (0.00000)	
SPOP6	0.40667 (0.00000)	0.54463 (0.00000)	0.45698 (0.00000)	0.54259 (0.00000)	0.48650 (0.00000)

Table 3. Pairwise F_{ST} values between subpopulations of *Hoplias malabaricus stricto* sensu (BOLD:ABZ3047) based on mtDNA (COI gene) haplotypes. The numbers in brackets are p-values at a significance level of 0.05. Subpopulations: (1) Lower Amazonas River confluences/Xingu River/Rupuruni River (SPOP1), (2) Madeira/Purus Rivers (SPOP2), (3) São Francisco/East Atlantic/Eastern Northeast Atlantic/Parnaíba/Itapecuru River basins (SPOP3), (4) Guiana Shields drainages (SPOP4), (5) Western-Northeast Atlantic/Tocantins basin (SPOP5), (6) Tapajós River Basin (SPOP6).

Source of variation	d.f.	Sum of squares	Variance components	Percentage of variation	p value
Among populations	5	27.681	0.13687 Va	27.20	0.000
Within populations	240	87.936	0.36640 Vb	72.80	0.000
Total	245	115.618	0.50327		

Table 4. AMOVA results of *Hoplias malabaricus stricto* sensu from the Amazon basin and adjacent drainages.
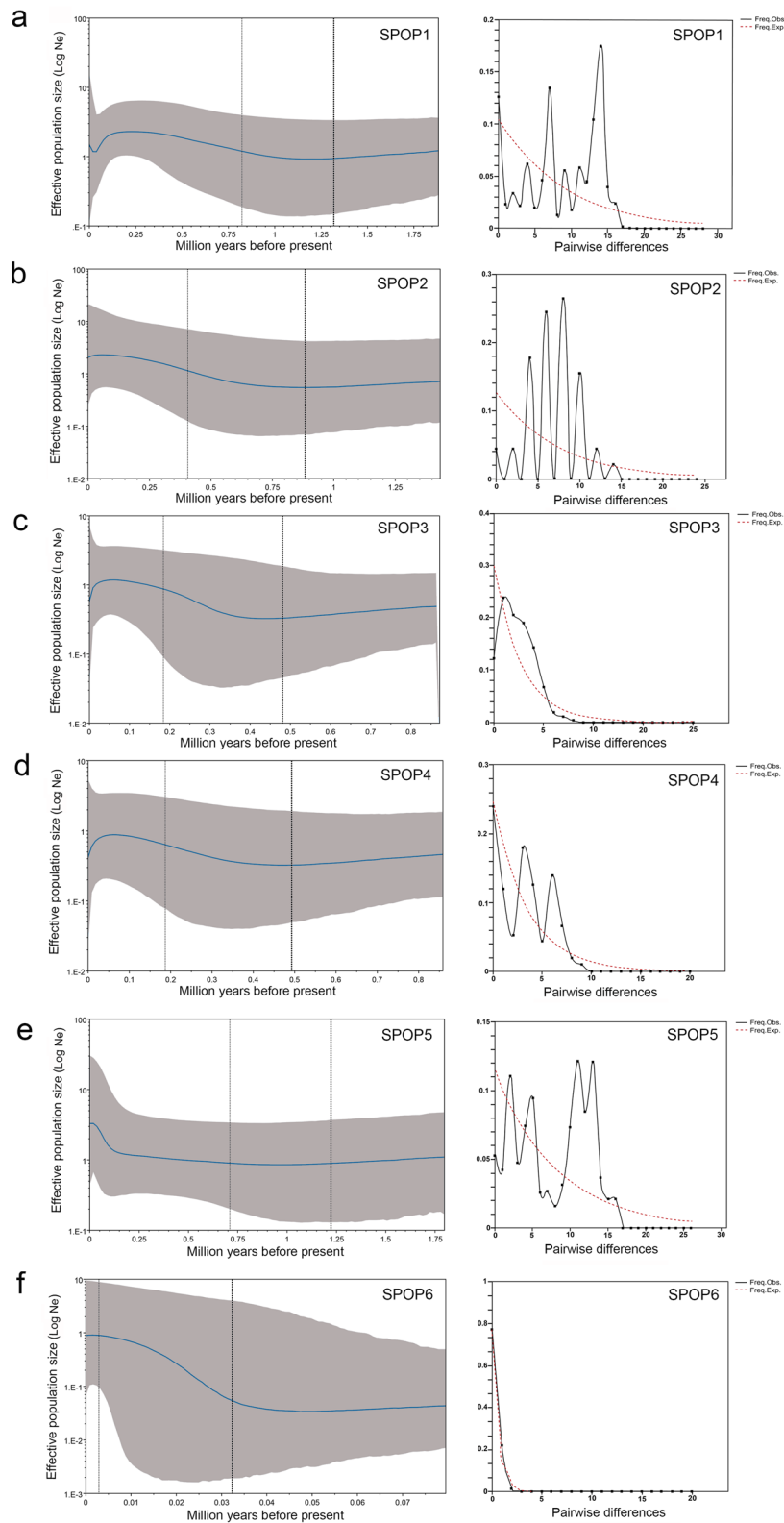
Figure 5. Mismatch distribution (right) and Bayesian skyline plot (BSP) (left) for the six population subdivisions of *H. malabaricus* (BIN BOLD:ABZ3047): (**a**) Lower Amazonas River confluences/Xingu River/Rupuruni River (SPOP1), (**b**) Madeira/Purus Rivers (SPOP2), (**c**) São Francisco/East Atlantic/Eastern Northeast Atlantic/Parnaíba/Itapecuru River basins (SPOP3), (**d**) Guiana Shields drainages (SPOP4), (**e**) Western-Northeast Atlantic/Tocantins basin (SPOP5), (**f**) Tapajós River Basin (SPOP6). BSPs show changes in the effective population size. The thick solid line represents the median estimate and the margins of the surrounding area represent the largest posterior density ranges of 95%.
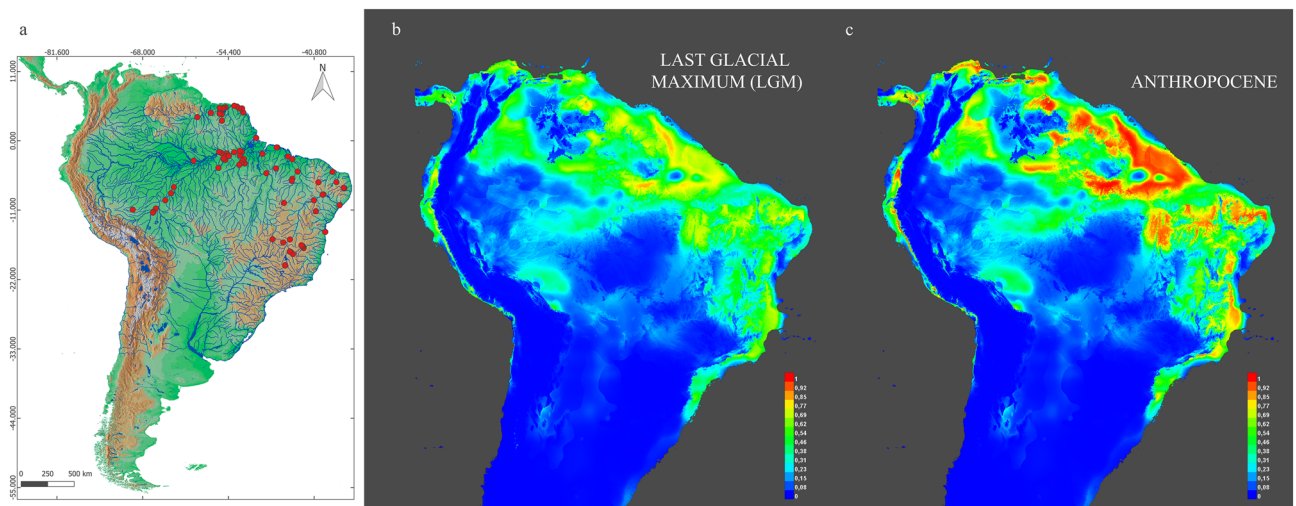
Figure 6. Present distribution of *H. malabaricus* (BIN ABZ3047) (**a**). Bioclimatic model of the most suitable geographic areas for the occurrence of this species in the Last Glacial Maximum (**b**) and Anthropocene (**c**). The warm colors (red, yellow) indicate high probability. The map was done with QGIS v.3.16.8-Hannover (Quantum GIS Development Team, http://www.qgis.org).

From the *H. malabaricus* complex, we detached a clade supported with 65% PP and delimited to BIN ABZ3047, which was interpreted as a single evolutionary unit and tested for genetic structure and taxonomic status validation. This putative species was assumed to be *H. malabaricus* stricto sensu and believed to be representative of *H. malabaricus* (Bloch, 1794). We found a clear genetic structure pattern, initially evidenced by the Bayesian topology that showed multiple subclades, and spatially divergent subpopulations shown with Geneland analyses. High Fst differentiation values and almost 50% of the genetic variation among populations support the genetic structure hypothesis. For comparison, the Fst values of the Tapajos River Basin population (SPOP6), the most differentiated, ranged from 40 to 54%. Aguirre and colleagues[57] measured a maximum F_{ST} divergence of 20% between *H. microlepis* populations from rivers and artificial impoundments in Ecuador. Due to its wide distribution range and sedentary habits, *Hoplias* species tend to exhibit high population genetic differentiation[9,58].

Genetic structuring in nonmigratory fish from the Amazon basin dispersed to adjacent drainages has been previously observed in Arapaimidae species, *Osteoglossum bicirrhosum*[59] and *Arapaima gigas*[60]. We detected a genetic link between populations from Orinoco drainage and the Amazon basin, in the Trombetas River (northern Amazonas drainage). Although, Amazonas and Orinoco systems remain currently isolated, past contacts between them are well documented and such gene flow exchange has been observed in other fishes (*e.g.,* needle fish genus *Potamorrhaphis*)[61]. Therefore, such connections between *H. malabaricus* populations may be explained by lineage sorting taking to the maintenance of ancestral haplotypes.

The SPOP 3 and 6 presented demographic instability and neutrality deviation that was interpreted as a signature of recent expansion. The biological drives in such populations may be adjusted with Pleistocenic phenomena, which span the most recent period of repeated glaciations. The SPOP 3 flows through portions of morphoclimatic domains of the Cerrado, Caatinga and part of the Atlantic Forest, where there is an extensive field of sand dunes in the São Francisco River basin which represents a testimony drier climate in this region in the past. The São Francisco River changed to its present course during the Mindel glaciation (approximately 450 kybp). Presently, it flows towards the north, curving towards the southeast and to the Atlantic Ocean, but this river previously flowed in a different direction, connecting with the Parnaíba River to the Atlantic Ocean[62], which may explain why this subpopulation was also be formed by samples from the Parnaíba River and the Western Northeast Atlantic basin. Similarly, the highlands in Guiana shields and Andes Mountains accumulated thick ice caps during the Pleistocene glaciations and the Tapajós region experienced long periods of erosion[63,64]. The SPOP6, formed by samples from the lower Tapajós River basin, is consistent within the limits of the Pleistocene (2.5 mybp–11.7 kybp), as shown by with mismatch distribution and population expansion analyses.

Modern Amazonian biodiversity dimensions were achieved during the Neogene (23–2.6 mybp), but the most ancient lineages are probably present since the Paleogene (66–23 mybp) and Late Cretaceous (146–66 mybp)[65–67]. During the middle to late Cenozoic, the Western Amazon basin was a lacustrine habitat, while the eastern and central portions were repeatedly invaded by marine incursions, resulting in the isolation of Guiana and Brazilian Shield tributaries[68]. The river capture dynamics during the Neogene have been proposed as the main force driving the last diversification of aquatic and terrestrial Amazonian taxa that are ecologically restricted to water bodies and riparian forests[65,69].

The proportion in which animals and plants retain ancestral ecological traits during time scales, that is, niche conservation, is still a controversial theme[70]. The finding of niche conservation is based on many variable aspects of the fundamental niche, taking into account that the rate of adaptation to conditions outside the niche is slower than the rate of extinction. The observation of conservation or niche change is an area of great relevance in the study of niche biogeography and in historical biogeography and in the exploration of patterns and mechanisms

of species richness in wide spatial areas[71]. Similar to the study by McNyset[72], which shows ecological niche conservation for six freshwater fish from North America under a moderate time scale, the present study points out that from the last glacial maximum (LGM) to the Anthropocene the niche remains very close in this taxonomic group. The *H. malabaricus* populations showed demographic evidence of expansion-decline cycles during the Pleistocene, immediately before the LGM. We hypothesize that such demographic fluctuations in *H. malabaricus* may be synchronized with the recent glaciation and the geologic events linked to the formation of the modern Amazonas River system[65,73].

Currently, *H. malabaricus* sensu stricto is dispersed through the Guiana drainages, Amazon basin, Western Atlantic Northeast basin and São Francisco River basin[1]. These systems experienced connection-isolation cycles during the Pleistocene period driven by climatic fluctuations and geomorphologic forces[1,18,74]. Our paleogeographic model revealed that the hypothetical distribution in the Pleistocene was mostly congruent with the present-day distribution and was influenced by climate change. The geographical distribution of the complex was restricted in the North and Northeast at South America during the LGM, and in the Anthropocene there was a reduction in potential areas. The drainages from the eastern portion of the Amazon basin and coastal drainages in the western Northeast Atlantic basin seem to be important routes for population dispersal.

Geomorphologic forces, mainly plate tectonics, marine incursions/regressions and climate fluctuations, recurrently shaped the Amazon landscapes, driving important processes in aquatic systems and leading to intensive changes in the river courses and river captures. These combined phenomena may be involved in the taxonomic radiation and geographic dispersion of the fish populations throughout the Amazon basin[65–67,75–78].

To conclude, the data herein indicate that *Hoplias malabaricus* sensu stricto presents a remarkable population structure, which seems to have been caused by past events. The cryptic diversity observed among the *Hoplias malabaricus* species group has conservation and taxonomic implications. Although morphological data have not been evaluated here, our molecular data are necessary to determine whether the mitochondrial lineages can represent distinct species. If to ABZ3047 subpopulations are not shown in the future to be different species, at least the SPOP 3, SPOP 4 and SPOP 6 can be considered as different evolutionarily significant units for conservation purposes. Additionally, even whether *Hoplias malabaricus* sensu stricto constitutes a well-distributed metapopulation in northern South America, its distribution would be much smaller than what is currently attributed to the species; thus, the taxonomy and geographic distribution of *H. malabaricus* should be reevaluated.

Data availability
All data generated or analysed during this study are included in this published article and its Supplementary Information files.

References

1. Cardoso, Y. P. *et al.* A continental-wide molecular approach unraveling mtDNA diversity and geographic distribution of the Neotropical genus Hoplias. *PLoS ONE* **13**(8), e0202024. https://doi.org/10.1371/journal.pone.0202024 (2018).
2. Bertollo, L. A. C., Born, G. G., Dergam, J. A., Fenocchio, A. S. & Moreira-Filho, O. A biodiversity approach in the Neotropical Erythrinidae fish, Hoplias malabaricus: Karyotypic survey, geographic distribution of karyomorphs and cytotaxonomic considerations. *Chrom. Res.* **8**(7), 603–613 (2000).
3. Oyakawa, O. T. Family Erythrinidae (Trahiras). in *Check list of the freshwater fishes of South and Central Americ*a (Reis, R. E., Kullander, S. O. & Ferraris, C.). *Edipucrs* 238–240 (Porto Alegre, 2003).
4. Dagosta, F. C. P. & de Pinna, M. C. C. The fishes of the Amazon: distribution and biogeographical patterns, with a comprehensive list of species. *Bull. Am. Museum Nat. Hist.* **431**, 1–163 (2019).
5. Da Rosa, R., Vicari, M. R., Dias, A. L. & Giuliano-Caetano, L. New insights into the biogeographic and Karyotypic Evolution of *Hoplias Malabaricus*. *Zebrafish* **11**(3), 198–206. https://doi.org/10.1089/zeb.2013.0953 (2014).
6. Santos, U. *et al.* Molecular and karyotypic phylogeography in the neotropical *Hoplias malabaricus* (Erythrinidae) fish in eastern Brazil. *J. Fish Biol.* **75**(9), 2326–2343. https://doi.org/10.1111/j.1095-8649.2009.02489.x (2009).
7. Blanco, D. R., Lui, R. L., Bertollo, L. A. C., Diniz, D. & Filho, O. M. Characterization of invasive fish species in a river transposition region: Evolutionary chromosome studies in the genus *Hoplias* (Characiformes, Erythrinidae). *Rev. Fish Biol. Fish.* **20**(1), 1–8. https://doi.org/10.1007/s11160-009-9116-3 (2010).
8. Jacobina, U. P. *et al.* DNA barcode sheds light on systematics and evolution of neotropical freshwater trahiras. *Genetica* **146**, 505. https://doi.org/10.1007/s10709-018-0043-x (2018).
9. Marques, D. F., Santos, F. A., da Silva, S. S., Sampaio, I. & Rodrigues, L. R. R. Cytogenetic and DNA barcoding reveals high divergence within the trahira, *Hoplias malabaricus* (Characiformes: Erythrinidae) from the lower Amazon River. *Neotrop. Ichthyol.* **11**(2), 459–466. https://doi.org/10.1590/S1679-62252013000200015 (2013).
10. Paz, F. P. C., Batista, J. S. & Porto, J. I. R. DNA barcodes of rosy tetras and allied species (Characiformes: Characidae: Hyphessobrycon) from the Brazilian Amazon Basin. *PLoS ONE* **9**(5), e98603. https://doi.org/10.1371/journal.pone.0098603 (2014).
11. Guimarães, K. L. A., de Sousa, M. P. A., Ribeiro, F. R. V., Porto, J. I. R. & Rodrigues, L. R. R. DNA barcoding of fish fauna from low order streams of Tapajós River basin. *PLoS ONE* **13**(12), e0209430. https://doi.org/10.1371/journal.pone.0209430 (2018).
12. Machado, V. N. *et al.* One thousand DNA barcodes of piranhas and pacus reveal geographic structure and unrecognized diversity in the Amazon. *Sci. Rep.* **8**, 8387. https://doi.org/10.1038/s41598-018-26550-x (2018).
13. Hebert, P. D. N., Cywinska, A., Ball, S. L. & Dewaard, J. R. Biological identifications through DNA barcodes. *Philos. Trans. R. Soc. B* **270**(1512), 313–321. https://doi.org/10.1098/rspb.2002.2218 (2003).
14. Pugedo, M. L., de Andrade Neto, F. R., Pessali, T. C., Birindelli, J. L. O. & Carvalho, D. C. Integrative taxonomy supports new candidate fish species in a poorly studied neotropical region: the Jequitinhonha River Basin. *Genetica* **144**(3), 1–9. https://doi.org/10.1007/s10709-016-9903-4 (2016).
15. Rosso, J. J. *et al.* Integrative taxonomy reveals a new species of the *Hoplias malabaricus* species complex (Teleostei: Erythrinidae). *Ichthyol. Explor. Freshw.* **1**, 1–18. https://doi.org/10.23788/IEF-1076 (2018).
16. Azpelicueta, M. M., Benítez, M., Aichino, D. & Mendez, C. M. D. A new species of the genus Hoplias (Characiformes, Erythrinidae), a tararira from the lower Paraná River, in Missiones, Argentina. *Acta Zool. Lilloana* **59**(1–2), 71–82 (2015).

17. Rosso, J. J. *et al.* A new species of the *Hoplias malabaricus* species complex (Characiformes: Erythrinidae) from the La Plata River basin. *Cybium* **40**(3), 199–208 (2016).

18. Cardoso, Y. P. & Montoya-Burgos, J. I. Unexpected diversity in the catfish *Pseudancistrus brevispinis* reveals dispersal routes in a Neotropical center of endemism: The Guyanas Region. *Mol. Ecol.* **18**(5), 947–964. https://doi.org/10.1111/j.1365-294X.2008.04068.x (2009).

19. Hoorn, C., Wesselingh, F. P., Hovikoski, J. & Guerrero, J. The development of the Amazonian mega-wetland (Miocene; Brazil, Colombia, Peru, Bolivia). *Amazon. Landsc. Species Evol.* https://doi.org/10.1002/9781444306408.ch8 (2010).

20. Albert, J. S. & Reis, R. E. Introduction to neotropical freshwaters. In *Historical Biogeography of Neotropical Freshwater Fishes* (eds Albert, J. S. & Reis, R. E.) 3–19 (University of California Press, 2011).

21. Leys, M., Keller, I., Räsänen, K., Gattolliat, J.-L. & Robinson, C. T. Distribution and population genetic variation of cryptic species of the Alpine mayfly *Baetis alpinus* (Ephemeroptera: Baetidae) in the Central Alps. *BMC Evol. Biol.* https://doi.org/10.1186/s12862-016-0643-y (2016).

22. Aljanabi, S. M. & Martinez, I. Universal and rapid salt-extraction of high quality genomic DNA for PCR-based techniques. *Nucleic Acids Res.* **25**(22), 4692–4693 (1997).

23. Vitorino, C. A., Oliveira, R. C. C., Margarido, V. P. & Venere, P. C. Genetic diversity of Arapaima gigas (Schinz, 1822) (Osteoglossiformes: Arapaimidae) in the Araguaia-Tocantins basin estimated by ISSR marker. *Neotrop. Ichthyol.* **13**, 557–568. https://doi.org/10.1590/1982-0224-20150037 (2015).

24. Ward, R. D., Zemlak, T. S., Innes, B. H., Last, P. R. & Hebert, P. D. N. DNA barcoding Australia's fish species. *Philos. Trans. R. Soc. B* **359**, 1847–1857. https://doi.org/10.1098/srtb.2005.1716 (2005).

25. Dunn, I. S. & Blattner, F. R. Sharons 36 to 40: Multienzyme, high capacity, recombination deficient replacement vectors with polylinkers and polystuffers. *Nucleic Acids Res.* **15**, 2677–2698 (1987).

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

27. Castresana, J. Selection of conserved blocks from multiple alignments for their use in phylogenetic analysis. *Mol. Biol. Evol.* **17**, 540–552. https://doi.org/10.1093/oxfordjournals.molbev.a026334 (2000).

28. Ratnasingham, S. & Hebert, P. D. N. DNA-Based registry for all animal species: The Barcode Index Number (BIN) system. *PLoS ONE* **8**(7), e66213. https://doi.org/10.1371/journal.pone.0066213 (2013).

29. Pons, J. *et al.* Sequence-based species delimitation for the DNA taxonomy of undescribed insects. *Syst. Biol.* **55**(4), 595–609. https://doi.org/10.1080/10635150600852011 (2006).

30. Fujisawa, T. & Barraclough, T. G. Delimiting species using single-locus data and the generalized mixed yule coalescent approach: A revised method and evaluation on simulated data sets. *Syst. Biol.* **62**(5), 707–724. https://doi.org/10.1093/sysbio/syt033 (2013).

31. Puillandre, N., Lambert, A., Brouillet, S. & Achaz, G. ABGD, automatic barcode gap discovery for primary species delimitation. *Mol. Ecol.* **21**(8), 1864–1877. https://doi.org/10.1111/j.1365-294X.2011.05239.x (2012).

32. Drummond, A. & Rambaut, A. BEAST: Bayesian evolutionary analysis by sampling trees. *BMC Evol. Biol.* **7**, 214. https://doi.org/10.1186/1471-2148-7-214 (2007).

33. Posada, D. jModelTest: Phylogenetic model averaging. *Mol. Biol. Evol.* **25**, 1253–1256. https://doi.org/10.1093/molbev/msn083 (2008).

34. R Core Team. *R: A Language and Environment for Statistical Computing*. https://www.R-project.org/ (2017).

35. Ezard, T., Fujisawa, T. & Barraclough, T. *splits: Species Limits by Threshold Statistics. R package version 1.0–19/r52.* https://R-Forge.R-project.org/projects/splits/ (2017).

36. Paradis, E. & Schliep, K. ape 5.0: An environment for modern phylogenetics and evolutionary analyses in R. *Bioinformatics* **35**, 526–528 (2018).

37. Bermingham, E., McCafferty, S. S. & Martin, A. P. Fish biogeography and molecular clocks: Perspectives from the Panamanian Isthmus. In *Molecular Systematics of Fishes* (eds Kocher, T. D. & Stepien, C. A.) 113–128 (Academic Press, 1997).

38. Thomaz, A. T., Malabarba, L. R., Bonatto, S. L. & Knowles, L. L. 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. *J. Biogeogr.* **42**, 2389–2401. https://doi.org/10.1111/jbi.12597 (2015).

39. Kimura, M. A simple method for estimating evolutionary rate of base substitutions through comparative studies of nucleotide sequences. *J. Mol. Evol.* **16**, 111–120 (1980).

40. Kumar, S., Stecher, G., Li, M., Knyaz, C. & Tamura, K. MEGA X: Molecular evolutionary genetics analysis across computing platforms. *Mol. Biol. Evol.* **35**, 1547–1549. https://doi.org/10.1093/molbev/msy096 (2018).

41. Guillot, G., Renaud, S., Ledevin, R., Michaux, J. & Claude, J. A unifying model for the analysis of phenotypic, genetic and geographic data. *Syst. Biol.* **61**(6), 897–911. https://doi.org/10.1093/sysbio/sys038 (2012).

42. Excoffier, L., Laval, G. & Schneider, S. *Arlequin: A Software for Population Data Analysis.* Version 3.1. http://cmpg.unibe.ch/software/arlequin3 (2007).

43. Wright, S. Evolution and the genetics of populations: Variability within and among natural populations. *Univ. Chicago* **4**, 580 (1978).

44. Rozas, J. *et al.* DnaSP 6: DNA sequence polymorphism analysis of large datasets. *Mol. Biol. Evol.* **34**, 3299–3302. https://doi.org/10.1093/molbev/msx248 (2017).

45. Bandelt, H. J., Forster, P. & Röhl, A. Median-joining networks for inferring intraspecific phylogenies. *Mol. Biol. Evol.* **16**(1), 37–48 (1999).

46. Leigh, J. W. & Bryant, D. POPART: Full-feature software for haplotype network construction. *Methods Ecol. Evol.* **6**, 1110–1116. https://doi.org/10.1111/2041-210X.12410 (2015).

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

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

49. Austin, M. P. Continuum concept, ordination methods, and niche theory. *Annu. Rev. Ecol. Syst.* **16**(1), 39–61. https://doi.org/10.1146/annurev.es.16.110185.000351 (1985).

50. Graham, A., Atkinson, P. & Danson, F. Spatial analysis for epidemiology. *Acta Trop.* **91**(3), 219–225. https://doi.org/10.1016/j.actatropica.2004.05.001 (2004).

51. Phillips, S. J., Anderson, R. P. & Schapire, R. E. Maximum entropy modeling of species geographic distributions. *Ecol. Model.* **190**(3–4), 231–259. https://doi.org/10.1016/j.ecolmodel.2005.03.026 (2006).

52. Guimarães, K. L. A., Rosso, J. J., Souza, M. F. B., de Astarloa, J. M. D. & Rodrigues, L. R. R. Integrative taxonomy reveals disjunct distribution and first record of Hoplias misionera (Characiformes: Erythrinidae) in the Amazon River basin: Morphological, DNA barcoding and cytogenetic considerations. *Neotrop. Ichthyol.* **19**(2), e200110. https://doi.org/10.1590/1982-0224-2020-0110 (2021).

53. Queiroz, L. J. *et al.* Evolutionary units delimitation and continental multilocus phylogeny of the hyperdiverse catfish genus *Hypostomus*. *Mol. Phylogenet. Evol.* **145**, 106711. https://doi.org/10.1016/j.ympev.2019.106711 (2020).

54. Phillips, J. D., Gillis, D. J. & Hanner, R. H. Incomplete estimates of genetic diversity within species: Implications for DNA barcoding. *Ecol. Evol.* https://doi.org/10.1002/ece3.4757 (2019).

55. Blaxter, M. L. The promise of a DNA taxonomy. *Philos. Trans. R. Soc. B.* **359**(1444), 669–679. https://doi.org/10.1098/rstb.2003.1447 (2004).

56. Nwani, C. D. *et al.* DNA barcoding discriminates freshwater fishes from southeastern Nigeria and provides river system-level phylogeographic resolution within some species. *Mitochondrial DNA* **22**(1), 43–51. https://doi.org/10.3109/19401736.2010.536537 (2011).

57. Aguirre, W. E., Shervette, V. R., Navarrete, R., Calle, P. & Agorastos, S. Morphological and genetic divergence of *Hoplias microlepis* (Characiformes: Erythrinidae) in rivers and artificial impoundments of Western Ecuador. *Copeia* **2013**(2), 312–323. https://doi.org/10.1643/ci-12-083 (2013).

58. Pires, W. M. M., Barros, M. C. & Fraga, E. C. DNA Barcoding unveils cryptic lineages of *Hoplias malabaricus* from Northeastern Brazil. *Braz. J. Biol.* **81**(4), 917–927. https://doi.org/10.1590/1519-6984.231598 (2020).

59. Souza, F. H. S. *et al.* interspecific genetic differences and historical demography in South American Arowanas (Osteoglossiformes, Osteoglossidae, Osteoglossum). *Genes* **10**(9), 693. https://doi.org/10.3390/genes10090693 (2019).

60. Torati, L. S. *et al.* Genetic diversity and structure in Arapaima gigas populations from Amazon and Araguaia-Tocantins river basins. *BMC Genet.* https://doi.org/10.1186/s12863-018-0711-y (2019).

61. Lovejoy, N. R. & Araujo, M. L. G. Molecular systematics, biogeography and population structure of Neotropical freshwater needle-fishes of the genus Potamorrhaphis. *Mol. Ecol.* **9**(3), 259–268. https://doi.org/10.1046/j.1365-294x.2000.00845.x (2000).

62. Mabesoone, J. M. *Sedimentary Basins of Northeast Brazil* (Federal University of Pernambuco, 1994).

63. Haffer, J. & Prance, G. T. Impulsos climáticos da evolução na Amazônia durante o Cenozóico: Sobre a teoria dos Refúgios da diferenciação biótica. *Estudos Avançados USP* **46**, 175–208. https://doi.org/10.1590/S0103-40142002000300014 (2002).

64. Riker, S. R. L., Lima, F. J. C., Motta, M. B. Evidências de glaciação Pleistocênica na Amazônia Brasileira. *Anais do 14° Simpósio de Geologia da Amazônia, Sociedade Brasileira de Geologia* 15–18 (2015).

65. Albert, J. S., Val, P. & Hoorn, C. The changing course of the Amazon River in the Neogene: Center stage for Neotropical diversification. *Neotrop. Ichthyol.* **16**(3), e180033. https://doi.org/10.1590/1982-0224-20180033 (2018).

66. Lundberg, J. G. *et al.* The stage for Neotropical fish diversification: a history of tropical South American rivers. (eds. Malabarba, L. R., Reis, R. E., Vari, R. P., Lucena, Z. M., Lucena, C. A. S. Phylogeny and classification of Neotropical fishes). *Edipucrs* 13–48 (1998).

67. Hubert, N. & Renno, J. F. Historical biogeography of South American freshwater fishes. *J. Biogeogr.* **33**(8), 1414–1436. https://doi.org/10.1111/j.1365-2699.2006.01518.x (2006).

68. Farias, I. P. & Hrbek, T. Patterns of diversification in the discus fishes (*Symphysodon* spp. Cichlidae) of the Amazon basin. *Mol. Phylogenet. Evol.* **49**, 32–43. https://doi.org/10.1016/j.ympev.2008.05.033 (2008).

69. Tagliacollo, V. A., Bernt, M. J., Craig, J. M., Oliveira, C. & Albert, J. S. Model-based total evidence phylogeny of Neotropical electric knifefishes (Teleostei, Gymnoti-formes). *Mol. Phylogenet. Evol.* **95**, 20–33. https://doi.org/10.1016/j.ympev.2015.11.007 (2015).

70. Hutchinson, G. E. Concluding remarks. Cold Spring Harbor Symposium. *Quant. Biol.* **22**, 415–427 (1957).

71. Wiens, J. J. & Graham, C. H. Niche conservatism: Inte-grating evolution, ecology, and conservation biology. *Annu. Rev. Ecol. Evol. Syst.* **36**, 519–539 (2005).

72. McNyset, K. M. Ecological niche conservatism in North American freshwater fishes. *Biol. J. Lin. Soc.* **96**, 282–295 (2009).

73. Silva, W. C., Marceniuk, A. P., Sales, J. B. L. & Araripe, J. Early pleistocene lineages of Bagre bagre (Linnaeus, 1766) (Siluriformes: Ariidae), from the Atlantic coast of South America, with insights into the demography and biogeography of the species. *Neotrop. Ichthyol.* https://doi.org/10.1590/1982-0224-20150184 (2016).

74. Lemopoulos, A. & Covain, R. Biogeography of the freshwater fishes of the Guianas using a partitioned parsimony analysis of endemicity with reappraisal of ecoregional boundaries. *Cladistics* **35**(2019), 106–124. https://doi.org/10.1111/cla.12341 (2018).

75. Hoorn, C. Marine incursions and the influence of Andean tectonics on the Miocene depositional history of northwestern Amazonia: Results of a palynostratigraphic study. *Palaeogeogr. Palaeoclimatol. Palaeoecol.* **105**, 267–309. https://doi.org/10.1016/0031-0182(93)90087-Y (1993).

76. Hoorn, C., Guerreiro, J. & Sarmiento, G. Andean tectonics as a cause for changing drainage patterns in Miocene Northern South America. *Geology* **23**(3), 237–240. https://doi.org/10.1130/0091-7613(1995)023%3c0237:ATAACF%3e2.3.CO;2 (1995).

77. Ribeiro, A. C. 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. *Neotrop. Ichthyol.* **4**(2), 225–246. https://doi.org/10.1590/S1679-62252006000200009 (2006).

78. Lovejoy, N. R., Albert, J. S. & Crampton, W. G. R. Miocene marine incursions and marine/freshwater transitions: Evidence from Neotropical fishes. *J. S. Am. Earth Sci.* **21**(1–2), 5–13. https://doi.org/10.1016/j.jsames.2005.07.009 (2006).

Acknowledgements

Author contributions

K.L.A.G. and L.R.R.R. conceived the project idea, wrote the main manuscript text, performed experiments and generated sequence data; R.S.B. performed the ecological niche model analysis and aided in interpreting the results; M.F. devised the figures of maps 1, 3, 4 and 6 with help from K.L.A.G.; D.J.S. contributed to the design and to the analysis of the results. M.P.L. helped with financial support and worked on the manuscript. All authors reviewed the manuscript.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1038/s41598-022-09121-z.

Correspondence and requests for materials should be addressed to L.R.R.R.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.