Comparative phylogeography of two codistributed species of the genus *Herichthys* (Actinopterygii: Cichliformes: Cichlidae) in northeastern Mexico

Fabian PÉREZ-MIRANDA1,2, Omar MEJÍA1, Benjamín LÓPEZ3, Eduardo SOTO-GALERA1, Amairany BERNAL-PORTILLO1, Wilfredo A. MATAMOROS2,4

1 Departmento de Zoología, Escuela Nacional de Ciencias Biológicas, Instituto Politécnico Nacional, Ciudad de México, Mexico
2 Instituto de Ciencias Biológicas, Universidad de Ciencias y Artes de Chiapas, Tuxtla Gutiérrez, Chiapas, Mexico
3 Universidad Michoacana de San Nicolás Hidalgo, Michoacán de Ocampo, Mexico
4 Field Museum of Natural History, Chicago, IL, USA

https://zoobank.org/1302CFAD-D245-496B-AFB1-B6E48E90A786

Corresponding author: Fabian Pérez-Miranda (monodactilo@hotmail.com)

Academic editor: Predrag Simonović ● Received 4 September 2023 ● Accepted 3 November 2023 ● Published 7 December 2023


Abstract

Phylogeographic patterns of freshwater fishes in coastal regions are highly susceptible to eustatic sea level changes associated with Pleistocene glaciations. In this context, the Plain Coastal Gulf in northeastern Mexico represents an ideal study area due to its low elevation. Herein, we compare the phylogeographic structures of two cichlid species of the genus *Herichthys* Baird and Girard, 1854 widely distributed in the Pánuco–Tamesí system in northeastern Mexico using two mitochondrial markers. The species studied were: *Herichthys carpintis* (Jordan et Snyder, 1899) and *Herichthys pantostictus* (Taylor et Miller, 1983). We estimate their genetic diversity, gene flow, and demographic history and perform biogeographic reconstructions using a Bayesian computation approach and environmental niche modeling. The biogeographic reconstruction suggests a different history for each species. Environmental niche modeling indicates that both species experienced a demographic expansion during the Pleistocene but responded differently to Pleistocene climatic changes. In summary, their current sympatric distribution could be the outcome of contemporary and not historical processes reflecting a pseudo-incongruent pattern.

Keywords

Bayesian computation, gene flow, niche modeling, paleohydrography

Introduction

Evolutionary biogeography seeks to reconcile dispersal and vicariance paradigms through five stages that reconstruct a geobiotic scenario that attempts to explain biotic component evolution based on geological data (Morrone 2007). One stage involves dating the vicariant events using molecular clocks and phylogeographic studies. Phylogeography deals with how genetic lineages are arranged through geographic space (Avise 2009). Therefore, comparative phylogeography studies are relevant since they enable us to interpret how different cenocrorns (biotic elements) have been integrated into a horobiota (a snapshot of a biota in a particular time) (Morrone 2020). In a comparative phylogeographic study, three different scenarios could arise. First, in a concerted response, the codistributed species respond similarly to geological and climatic events leading to a congruent pattern. Second, in an
independent response, the codistributed species show independent responses to simultaneous regional processes. Third, in a multiple response, the species could show similar spatial congruence but different temporal frames leading to a pseudo-congruent pattern or different responses both in space and time leading to a pseudo-incongruent pattern (Bagley and Johnson 2014). The cooling and heating periods experienced by the planet during the Pleistocene led to speciation and extinction events and changes in the distributional patterns of species worldwide (Hewitt 1996, 2000, 2004). However, Pleistocene glaciations also affect populations, leaving genetic signatures that are traceable through phylogeographic studies (Hewitt 2003; Lister et al. 2005; Comte and Grenouillet 2015).

These phenomena are more evident in freshwater ecosystems due to climate changes and geological factors, such as volcanism and orogenesis, playing a fundamental role in drainages rearrangement, leading to isolation, reconnection, and formation of new rivers and lakes (Bermingham and Martin 1998; Waters and Wallis 2000; Rincon-Sandoval et al. 2019). Their effects are particularly apparent in species distributed toward the coastal shoreline (Abreu et al. 2020; Pio and Carvalho 2021) due to the Pleistocene’s eustatic sea level changes leading to the exposition and covering of the continental shelf, promoting both population connection and isolation (Hewitt 2004; Lambeck and Chappell 2001).

Due to Mexico’s complex geologic and paleohydrographic history, its freshwater fishes are ideal for phylogeographic studies (Marshall and Liebher 2000; Mronne 2004; Domínguez-Domínguez and Pérez-Ponce de León 2009), their study has already led to several published papers in selected taxonomical groups, including atherinopsids (Bloom et al. 2009; García-Martínez et al. 2020), characids (Strecker et al. 2004; Ornelas-García et al. 2008; Hausdorff et al. 2011; Coghill et al. 2014), cyprinids (García-Andrade et al. 2021), and poecilids (Mateos et al. 2002; Chen and Borowsky 2004; Mexos 2005; Gutiérrez-Rodríguez et al. 2007; Bagley et al. 2013).

Cichlids are one of the most diverse clades of freshwater fishes in Mexico. However, despite their species abundance, phylogeographic studies with them are scarce (e.g., Barluenga and Meyer 2010; Bagley et al. 2017; McMahan et al. 2017). The genus *Herichthys* Baird and Girard, 1854 represents an excellent model for evolutionary and biogeographic studies since it includes species of wide and restricted geographic distributions that have been extensively studied in recent years, from systematics to molecular clocks and biogeography among others (Pérez-Miranda et al. 2018, Pérez-Miranda et al. 2020).

In this study, we compared the phylogeographic structures of two cichlid species of the genus *Herichthys* that have a wide geographic distribution in the Pánuco–Tamesi system: *Herichthys carpintis* (Jordan et Snyder, 1899) and *Herichthys pantosticus* (Taylor et Miller, 1983). The latter species was included in the past in the genus *Nosferatu* (see De la Maza-Benignos et al. 2015), however, as pointed out by Ričan et al. (2016), the shape of the teeth that are the diagnostic character of the genus is a plesiomorphic state for the *Theraps–Paranneipterus* clade and it is also present in some species of the genus *Herichthys* turning the genus *Nosferatu* in a paraphyletic and not natural group. Therefore, all species previously placed in the genus *Nosferatu* by De la Maza-Benignos et al. (2015) should now be included in the genus *Herichthys* (see Fricke et al. 2023; Froese and Pauly 2023; WoRMS Editorial Board 2023). The Pánuco–Tamesi system drains part of the Gulf Coastal Plain in northeastern Mexico covering a surface of near 157 752 km² and comprises 11 sub-basins (FAO 2022). This vast region is characterized by altitudes ranging from 0 to 600 m above sea level and was prone to the effects of marine transgressions and regressions during the Pleistocene glaciation periods (Álvarez 1961; Bagley et al. 2013). While both species have a sympatric distribution throughout most of their geographic range, their current geographic distribution patterns are likely the result of different evolutionary and biogeographic histories (Pérez-Miranda et al. 2020). While *H. carpintis*’ sister species, *Herichthys tepehua* De la Maza–Benignos, Ornelas-García, Lozano-Vilano, García-Ramírez et Doadrio, 2014 is distributed towards the coastline, *H. pantosticus*’ sister group, comprising *Herichthys bartoni* (Bean, 1982) and *Herichthys labrindens* (Pellegrin, 1903), is distributed inland (Pérez-Miranda et al. 2018).

Therefore, we expected these species to show a pseudo-incongruent pattern due to different past distributions and recent community assemblies with a scarce or null shared history between them. We test our prediction by evaluating the effect of Pleistocene glaciations on the colonization and connectivity of *H. carpintis* and *H. pantosticus* populations by determining the numbers and ages of their genetic populations, the gene flow among them, and their demographic history and colonization processes using two mitochondrial markers, COI, and D-loop.

**Material and methods**

**Sampling and genetic analysis.** The specimens of *Herichthys carpintis* and *Herichthys pantosticus* used were collected between 2000 and 2016 and covered the known geographic distribution of both species (Suppl. material 1). Tissue samples were obtained from 96 *H. carpintis* individuals and 60 *H. pantosticus* individuals. DNA extraction was performed according to the protocol of Aljanabi and Martínez (1997) for amplifying mitochondrial markers cytochrome oxidase subunit 1 (COI) and D-loop. For the mitochondrial COI marker, we complemented our *H. carpintis* data set with an additional 104 previously generated sequences available in two Barcode of Life Data (BOLD) projects (FFPTR and HBGM).

Polymerase chain reactions (PCRs) had a final volume of 25 μL. For the COI marker, we used the primers reported by Ward et al. (2005) with the conditions reported by León-Romero et al. (2012) to amplify a fragment of 589 bp for the D-loop marker, we designed FPM-F
(5'-CTTTGGGAGTGGGTTGA-3') and FPM-R (5'-CACCAGAAGTGTAGCAAGCGG-3') primers to amplified a 687 bp fragment in a reaction mix comprising 1× PCR buffer, 3 mM MgCl₂, 200 μM dNTPs, 0.15 μM of each primer, 40 ng of DNA template, and 1 U of Go-Taq (Invitrogen) using the following conditions: initial preheating at 95°C for 5 min, followed by 35 cycles of denaturing at 96°C for 1 min, annealing at 60°C for 1.5 min, and extension at 72°C for 2 min, with a final 5 min extension at 72°C. PCR products were purified, and both strands were sequenced at the Laboratorio Nacional de Genómica para la Biodiversidad campus (Irapuato, México).

Sequences were aligned with Clustal X versión 2.0 (Larkin et al. 2007) and edited in Seaview (Gouy et al. 2010). The final numbers of available sequences were 195 for COI and 73 for D-loop in H. carpintis and 41 for COI and 54 for D-loop in H. pantostictus.

Population genetics. We are aware of the potential caveats of dealing with sample size and their effect on genetic diversity and other population genetics estimators. For the aforementioned, we evaluate if the number of analyzed individuals was enough using the function HACSim s implemented in the R library HACSim to build an accumulation curve and to estimate the R-value that represents the estimated fraction of species haplotype diversity captured from sampling (Phillips et al. 2020).

After that, we assessed the number of genetic clusters (k) (hereinafter referred to as populations) using the “optimise.baps” option in the fastbaps library (Tonkin-Hill et al. 2019) of the R statistical software version 4.0.4 (RStudio Team 2020). Each population’s diversity parameters, such as haplotypic (h) and nucleotidic (π) diversity, were estimated using Arlequin v.3.5.2 (Excoffier and Lischer 2010).

A time-calibrated phylogenetic tree was estimated for each species and molecular marker in BEAST v 1.7.5 (Drummond et al. 2205), to date the trees, we used secondary calibration points previously estimated for H. carpintis and H. pantostictus (see Pérez-Miranda et al. 2020) which used different calibration points including fossil record and vicariant events (but see also Rícan et al. 2013); for H. carpintis, we used the split with her sister species H. tepelhua (outgroup in the phylogeny) occurred 2.5 Ma, meanwhile, for H. pantostictus we used its divergence time against her sister clade (H. bartoni + H. labridens) (outgroup in the phylogeny) occurred 6.5 Ma. Four independent runs of 10 million generations, sampling every 10 000 generations, were performed assuming a GTR substitution model, a strict molecular clock, and a Yule speciation model; convergence among chains was assumed if the ESS values were higher than 200, then, the trees were summarized using LogCombiner v.1.7.4 (Drummond and Rambaut 2007), and a consensus tree was constructed after a 25% burn-in using TreeAnotator v2.6.6 (Drummond and Rambaut 2007).

To evaluate gene flow levels among populations, we used the Bayesian approach implemented in migrate v.4.4.2 (Beerli 1998; Beerli and Felsenstein 2001) using a static heating scheme with four temperature chains (1, 1.5, 3, and 1 000 000), with each analysis comprising 10 000 000 genealogies sampled every 1000 generations after a 10% burn-in. We used the full matrix model as a null hypothesis for each species and molecular marker and postulated several gene flow models following Miller et al. (2005). Since H. pantostictus is mainly distributed inland, its colonization process should be from inland towards the coastline. In contrast, since H. carpintis is mainly distributed along the coastal shoreline, its colonization process should be from the coastline towards the inland.

Ten gene flow models were evaluated for H. pantostictus (five per molecular marker), and 11 models were evaluated for H. carpintis (six for COI and five for D-loop; see Suppl. material 2 for complete details). We compared gene flow models using a Bayes Factor test with the Bezier approach’s marginal likelihood (Beerli et al. 2019) using the BF function in R’s mtraceR library (Pacioni et al. 2015).

Effective population size changes were inferred using Tajima’s D and Fu’s Fₜ tests in Arlequin v.3.5.2 (Excoffier and Lischer 2010). In addition, effective population size changes through time were inferred from Bayesian skyline plots (BSPs) created using BEAST v 1.7.5 (Drummond et al. 2005) with a lumping approach since coalescent-based tests are extremely sensitive to sample size (Heller et al. 2013). The BSPs plots were created using four Markov chains of 10 000 000 generations, sampled every 1000 generations, and a strict molecular clock with the same calibration points mentioned above. The chains’ results were combined using LogCombiner v 1.7.4 after a 25% burn-in (Drummond and Rambaut 2007), and the BSPs were plotted in Tracer v 1.5 (Rambaut et al. 2018).

Biogeographic scenarios and niche modeling. We tested two biogeographic scenarios for the colonization routes of each species using the approximate Bayesian computation (ABC) approach implemented in the DIYABC software (Cornuet et al. 2010). Scenario one assumed a coastal-to-inland colonization process. Scenario two assumed an inland-to-coastal colonization process. The scenarios were compared using reference tables simulating 1 × 10⁶ datasets based on haplotype numbers and the Tajima and Fu test values. First, considering the reference tables’ first 10 000 scenarios, we used a principal component analysis to evaluate whether the generated dataset’s distribution approached that of the observed dataset. Then, a normalized Euclidean distance between the simulated and the observed datasets was calculated to determine the most plausible scenario. Finally, considering the 1% of generated datasets closest to the observed datasets, a direct and logistic regression was used to estimate the posterior probability and the type I and type II errors for each scenario with a 95% highest posterior distribution (HPD). Therefore, the most probable scenario was chosen based on the highest posterior probability and the absence of overlap in the HPD intervals (Cornuet et al. 2010).
Finally, we used an environmental niche model (ENM) approach to evaluate the possible effects of Pleistocene glaciations on both species’ demographic history. The suitability areas for each species were determined using the maximum entropy algorithm implemented in Maxent v.3.2.19 (Phillips et al. 2006) applied to the Mejía et al. (2022) collecting dataset. First, the spThin R library (Aiello-Lammens et al. 2015) was used to reduce spatial autocorrelation by pruning collection records that are <1 kilometer, leaving a total of 205 collection records for *H. carpintis* and a total of 124 collection records for *H. pantostictus*. Then, the geographic space *M* for each species was defined with the Pfafstetter HydroBasin levels 6 and 7 (Lehner and Grill 2013) in QGIS v.3.16.5 (QGIS Development Team 2009). The 19 WorldClim bioclimatic variables (Hijmans et al. 2005) were downloaded for the current period. The paleoclimatic projections used the MPI-ESM-P general circulation model in three temporal frames: the last interglacial period (LIG; 120 ka), the last glacial maximum (LGM; 21 ka), and the mid-Holocene (6 ka). Additionally, three topographic variables (aspect, topographic position index, and slope) were calculated based on a digital elevation model from Hydrosheds (Lehner et al. 2008) using the terrain function of the raster R library (Hijmans and Van Etten 2012). The number of variables in the ENM was reduced using the variance inflation factor (VIF) with the vifcor and vifstep functions in the usdm R library (Naimi et al. 2014).

The ENMs were constructed using the dismo (Hijmans et al. 2020), ENMeval (Muscarella et al. 2014), rmaxent (Baumgartner et al. 2017), and kuenm (Cobos et al. 2019) R libraries. We used 30% of the collection records for model construction and the remaining 70% for model training with several combinations of feature classes (linear, quadratic, product, and threshold) and regularization multipliers (1, 2, 5, 10, 15, and 20; Warren and Seifert 2011). Finally, the best ENM models were selected using the Akaike information criteria (Akaike 1974) and partial receiver operating characteristic (ROC) curve values over 1000 bootstraps (Peterson and Nyari 2008).

**Results**

**Population genetics.** The new sequences of *Herichthys carpintis* and *Herichthys pantostictus*, generated for this study, were deposited in GenBank under the accession numbers OP738891–OP738896; OP738385–OP738395, and OP751419–OP751525 (Suppl. material 3). The *R* values recovered from the HACSim curves were higher than 0.9 for both species and molecular markers (Suppl. material 4) and suggest that the number of individuals analyzed was enough to recover the genetic diversity of both species.

The fastbaps analysis recovered three genetic clusters (populations) for each molecular marker in *H. pantostictus*, based on the COI marker, population one (Guayalejo) comprised five individuals and had a geographic centroid near the coastal shoreline at the northern limit of its distribution. Population two (Pánucos) comprised 20 individuals with a geographic centroid located near the center of its geographic distribution. Population three (Tanconchín) comprised 16 individuals with a geographic centroid in the south of its distribution (Fig. 1). Based on the D-loop marker, population four (Tamesí) comprised 23 individuals and had a geographic centroid toward the north of its distribution. Population five (Valles) comprised 19 individuals and had a geographic centroid toward the center of its distribution. Population six (Naranjos) comprised 12 individuals and had a geographic centroid toward the South of its distribution (Fig. 1).

In *H. carpintis*, based on the COI marker, population one (Ozuluama) comprised 90 individuals and had a geographic centroid in the south of its distribution. Population two (Adjuntas) comprised 15 individuals and had a geographic centroid in the center of its distribution. Population three (Mante) comprised 90 individuals and had a geographic centroid in the north of its distribution. For the D-Loop marker, population four (Jaumave) comprised 50 individuals and had a geographic centroid in the north of its distribution. Population five (Tempoal) comprised 12 individuals and had a geographic centroid in the center of its distribution. Population six (Gallinas) comprised 11 individuals and had a geographic centroid in the center of its distribution toward the inland (Fig. 1).

The *h* values in *H. pantostictus* populations ranged from 0.542 in Tanconchín to 0.900 in Guayalejo for the COI marker and from 0.544 in Valles to 0.573 in Tamesí for the D-loop. In *H. carpintis* populations, *h* values were low for the COI marker, ranging from 0.391 in Mante to 0.664 in Ozuluama, but high for the D-loop marker, ranging from 0.818 in Gallinas to 1.000 in Tempoal. The π values were low in the majority of populations, ranging from 0.001 for the COI marker in the *H. carpintis* Adjuntas and Mante populations to 0.026 for the D-loop marker in the *H. carpintis* Jaumave population (Table 1).

The molecular clock analysis inferred similar root ages for both markers in *H. pantostictus* (6.9 Ma, HPD 5.3–8.4 Ma). Population ages inferred based on the COI marker were 0.69 Ma (HPD 0.2–1.1 Ma) for Guayalejo, 1.91 Ma (0.8–3.0 Ma) for Pánucos, and 2.68 Ma (HPD 1.1–4.6 Ma) for Tanconchín. Population ages inferred from the D-loop marker were 2.55 Ma (HPD 1.2–3.8 Ma) for Naranjos, 2.15 Ma (HPD 1.0–3.3 Ma) for Valles, and 2.53 Ma (HPD 1.2–3.8 Ma) for Tamesí (Fig. 2). However, they provided opposing ancient population geographic locations. The youngest population was located toward the coastline for the COI marker (Guayalejo) but inland for the D-loop marker (Valles). In *H. carpintis*, the inferred root ages were 1.9 Ma (HPD 1.5–2.3 Ma) with the COI marker and 2.3 Ma (HPD 1.6–2.9 Ma) with the D-loop marker. Population ages inferred based on the COI were 1.47 Ma (HPD 0.9–1.9 Ma) for Ozuluama, 0.82 Ma (HPD 0.3 1–3 Ma) for Adjuntas, and 1.24 Ma (0.7–1.7 Ma) for Mante. Population ages inferred based on the D-loop marker were 1.9 Ma (HPD 1.5–2.3 Ma) for
Jaumave, 1.2 Ma (HPD 0.8–1.6 Ma) for Tempoal, and 0.7 Ma (0.3–1.0 Ma) for Gallinas (Fig. 2). Therefore, the youngest *H. carpintis* population was toward the inland with both markers: Adjuntas with the COI marker and Gallinas with the D-loop marker (Fig. 2).

The migrate analysis, which estimated gene flow, suggested that the most probable model for *H. pantostictus* based on the COI marker was colonization from inland towards the coastline (maximum likelihood [ML] = −1101.44, *P* = 0.964; Suppl. material 5), with Tanconchin
Table 1. Summary of the genetic diversity statistics recovered in *Herichthys pantostictus* and *Herichthys carpintis* for the mitochondrial molecular markers COI and D-loop.

| Species     | Marker | Cluster | k  | n   | h     | π  | D   | F_s
|-------------|--------|---------|----|-----|-------|----|-----|-----
| *H. pantostictus* | COI    | Guayalejo | 5  | 0.900 | 0.003 | −1.094 | −3.578^{2}\%
|              |       | Pánuco   | 20 | 0.447 | 0.002 | −1.719^{\ast}\% | −3.408^{2}\%
|              |       | Tanconchín | 16 | 0.542 | 0.003 | −2.003^{\ast}\% | −2.768^{2}\%
|              | D-loop| Tamesí   | 23 | 0.573 | 0.003 | −2.006^{\ast}\% | −2.762^{2}\%
|              |       | Valles   | 19 | 0.544 | 0.003 | −1.894^{\ast}\% | −2.811^{2}\%
|              |       | Naranjos | 12 | 0.848 | 0.007 | 0.483 | −11.383^{2}\%
| *H. carpintis*    | COI    | Ozuluama | 90 | 0.664 | 0.004 | 1.209 | 1.530
|              |       | Adjuntas | 15 | 0.562 | 0.001 | 0.139 | −10.727^{2}\%
|              |       | Mante    | 90 | 0.391 | 0.001 | −0.886 | −14.130^{2}\%
|              | D-loop| Jaumave  | 50 | 0.995 | 0.026 | −1.099 | −24.227^{2}\%
|              |       | Tempoal  | 12 | 1.000 | 0.018 | 0.487 | −4.241^{2}\%
|              |       | Gallinas | 11 | 0.818 | 0.007 | −0.712 | −2.855^{2}\%

<sup>\ast</sup> = P < 0.05, <sup>\ast\ast</sup> = P < 0.001; k = number of genetic clusters recovered by fastbaps, n = number of individuals examined, h = haplotypic diversity, π = nucleotidic diversity, D = Tajima D test values, F_s = Fu test values.

Figure 2. Molecular dated phylogeny of the genetic clusters recovered in *Herichthys carpintis* and *Herichthys pantostictus* for the mitochondrial markers COI and D-loop. The median divergence time (Ma) and the HPD intervals (Ma) are shown below the nodes. The colors represent the genetic clusters depicted in Fig. 1.

giving rise to Pánuco (M = 138; 95% HPD: 0–420), followed by Pánuco giving rise to Guayalejo (M = 138; 95% HPD: 0–937; Fig. 3). However, for the D-loop marker, the colonization process occurred from the distribution’s center to the North and then to the South (ML = −1040.74, P = 0.668; Suppl. material 5), with Valles giving rise to Tamesí (M = 322; 95% HPD: 83–357), followed by Tamesí giving rise to Naranjos (M = 415; 95% HPD: 0–483; Fig. 3).

The gene flow estimation for *H. carpintis* suggests that the colonization process occurred from the coastline toward the inland for the COI marker (ML = −3099.68, P = 1.000), with Ozuluama giving rise to Mante (M = 82;
95% HPD: 0–193), followed by Mante giving rise to Adjuntas ($M = 315; 95\% \text{ HPD}: 0–677; \text{Fig. 3}$). Finally, for the D-loop, the colonization process occurred from inland to the coastline ($\text{ML} = -4002.55, P = 1.000$), with Gallinas giving rise to Tempoal ($M = 448; 95\% \text{ HPD}: 67–286$), followed by Tempoal giving rise to Jaumave ($M = 275; 95\% \text{ HPD}: 3–530; \text{Fig. 3}$).

The Tajima’s D and Fu’s F tests used to evaluate effective population size changes suggest a demographic expansion of the $H. \text{pantostictus}$ populations, except the Tajima tests for both COI and D-loop markers in the Naranjos and Guayalejo populations (Table 1). Both markers’ BSPs provided similar results (Fig. 4). For the COI marker, a slight demographic decrease occurred at ~1 Ma, followed by a sudden expansion at ~250 kya. Similarly, for the D-loop marker, a demographic expansion occurred at ~250 kya (Fig. 4). However, in $H. \text{carpintis}$, the Tajima test was non-significant in all populations with both markers. In contrast, the Fu test suggested an expansion in all populations with both markers, except for Ozuluama with the COI marker (Table 1). Finally, both markers’ BSPs suggested a demographic expansion starting at 100 kya for the COI marker and 1 Ma for the D-loop marker (Fig. 4).

**Figure 3.** Gene flow among populations of *Herichthys carpintis* and *Herichthys pantostictus* according to the most probable model recovered in Bayes Factor analysis (see Suppl. materials 2, 4 for complete details). Theta values and 95% HPD intervals are shown within each genetic cluster. The values on the arrows show the number of migrants from one population to another with the minimum and maximum values maximum in parentheses; solid arrows indicate a single gene flow event, dashed arrows indicated recurrent gene flow (values below), after the initial migrant event (values above).
Biogeographic scenarios and niche modeling. The ABC analysis inferred opposing biogeographic histories for *H. pantostictus* with each molecular marker. For the COI marker, the most probable scenario suggests colonization from the coastline to inland populations (scenario one; logistic posterior probability = 0.596; 95% HPD: 0.587–0.604). In contrast, scenario two was the most plausible for the D-loop marker, suggesting an inland to coastline colonization (*P* = 0.804; 95% HPD = 0.797–0.813; Table 2; Suppl. material 6). However, for *H. carpintis*, scenario one was inferred as the most probable scenario for both molecular markers, suggesting colonization from the coastline to inland (Table 2; Suppl. material 6).

Finally, in the ENM, eight bioclimatic (BIO2, BIO3, BIO5, BIO9, BIO13, BIO14, BIO15, and BIO18) and three topographic variables were retained after the VIF test. The potential geographic distribution models for *H. pantostictus* and *H. carpintis* showed good performance (area under the ROC curve [AUC] >0.8; AUC [partial ROC curve] >1.1; Table 3) without over-adjustment (AUC difference = 0.143–0.190 for *H. pantostictus* and 0.159–0.178 for *H. carpintis*; Table 3). For both species, most of the observed variance was explained by the warmest month’s maximum

---

**Figure 4.** Bayesian skyline plots (BSP) of *Herichthys carpintis* and *Herichthys pantostictus* recovered from the analysis of the lumping populations of each of the molecular markers analyzed in this study.

**Table 2.** Scenarios used in the Approximate Bayesian computation (ABC) for the biogeographic history of *Herichthys pantostictus* and *Herichthys carpintis* for the two molecular markers used in this study. The posterior probability for the direct and the logistic regression as well as probability of type I and type II errors are indicated for each one of the postulated scenarios. Selected scenarios are set in bold typeface.
Table 3. Best Maxent model selected for *Herichthys pantostictus* and *Herichthys carpintis* for each one of the four periods considered in this study. LIG (Last interglacial 120 ky), LGM (Last Glacial Maximum (21 ky), mid Holocene (6 ky), Current. Best fit MaxEnt model select for each of species for the four periods considered.

<table>
<thead>
<tr>
<th>Period</th>
<th>Parameter</th>
<th>AUC</th>
<th>Diff. AUC</th>
<th>AIcC</th>
<th>pROC</th>
<th>w</th>
<th>Area [km²]</th>
<th>Area [%]</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td><em>H. pantostictus</em></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>LIG</td>
<td></td>
<td>0.898</td>
<td>0.172</td>
<td>2912.548</td>
<td>1.458</td>
<td>0.750</td>
<td>148262.6</td>
<td>90.0</td>
</tr>
<tr>
<td>LGM</td>
<td></td>
<td>0.873</td>
<td>0.190</td>
<td>2954.457</td>
<td>1.267</td>
<td>0.895</td>
<td>129617</td>
<td>80.0</td>
</tr>
<tr>
<td>Mid Holocene</td>
<td></td>
<td>0.876</td>
<td>0.143</td>
<td>2941.143</td>
<td>1.555</td>
<td>0.889</td>
<td>105585</td>
<td>68.0</td>
</tr>
<tr>
<td>Current</td>
<td></td>
<td>0.885</td>
<td>0.174</td>
<td>2942.630</td>
<td>1.294</td>
<td>0.817</td>
<td>154917</td>
<td>96.0</td>
</tr>
<tr>
<td></td>
<td><em>H. carpintis</em></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>LIG</td>
<td></td>
<td>0.912</td>
<td>0.161</td>
<td>4079.705</td>
<td>1.246</td>
<td>1.000</td>
<td>119371.4</td>
<td>81.4</td>
</tr>
<tr>
<td>LGM</td>
<td></td>
<td>0.871</td>
<td>0.167</td>
<td>4129.669</td>
<td>1.395</td>
<td>0.968</td>
<td>118783</td>
<td>81.0</td>
</tr>
<tr>
<td>Mid Holocene</td>
<td></td>
<td>0.878</td>
<td>0.159</td>
<td>4118.810</td>
<td>1.294</td>
<td>0.998</td>
<td>113358</td>
<td>77.3</td>
</tr>
<tr>
<td>Current</td>
<td></td>
<td>0.859</td>
<td>0.178</td>
<td>4138.734</td>
<td>1.142</td>
<td>0.970</td>
<td>117648</td>
<td>80.3</td>
</tr>
</tbody>
</table>

**Notes:** ROC = receiver operating characteristic, AUC = area under the curve, Diff. AUC = mean difference in the AUC values between the training data and the test data, AIcC = Akaike information criterion, pROC = partial ROC values, w = Akaike weight, Area = total suitability area available for the taxon in km² and expressed in km² as a percentage of the total geographic extent (M).

Temperature (BIO5 = 14.441% for *H. pantostictus* and 15.506% for *H. carpintis*), the warmest quarter’s precipitation (BIO18 = 44.390% for *H. pantostictus* and 20.216% for *H. carpintis*), and altitude (13.361% for *H. pantostictus* and 50.890% for *H. carpintis*).

The geographic extension of suitable climatic areas decreased by 22% from the LIG to the mid-Holocene for *H. pantostictus* (Table 3; Fig. 5 A–C). However, while a similar pattern was inferred for *H. carpintis*, it had a smaller decrease (Table 3; Fig. 5 E–G). Finally, an increase in suitable areas was inferred for both species from the mid-Holocene (6 ky) to the present, which was small (3%) for *H. carpintis* and larger (28%) for *H. pantostictus*.

Discussion

Population structure and diversity. The number of genetic populations recovered in a phylogeographic study depends on several factors, such as the choice of molecular marker; the size of the fragment; sample size; evolutionary processes such as selection, mutation, and genetic drift; and the species’ intrinsic attributes related to habitat and life history traits (Gavrilets 2003; Sofia et al. 2006; Matsumoto and Hilsdorf 2009). In cichlids, we must expect that species distributed in lakes will show a pattern does not always occur. For example, there are riverine species without genetic structure, such as the South American *Gymnogeophagus setequedas* Reis, Malabarba et Pavanelli, 1992 (see Souza-Shibatta et al. 2018) and other Neotropical cichlids found in rivers in South Mexico (Elías et al. 2020). However, like this study’s results, structured populations in riverine environments have been recovered in species distributed in Mexico, such as *Trichromis salvini* ( Günther, 1862) (see Elías et al. 2020) and other South American species (Abreu et al. 2020; Pio and Carvalho 2021).

One possible explanation for the presence of genetic structure in this study could be associated with the Pleistocene glacial periods with more arid environments, leading to water body desiccation and a lack of connectivity and gene flow (Strecker et al. 2004; Gutiérrez-Rodríguez et al. 2007). Then, during the interglacial period, wetter conditions increase water levels, habitat connectivity, and gene flow (Elías et al. 2020). Another possible explanation could be that both species analyzed in this study had populations distributed toward the coastline. In other South American tropical fish species, it has been postulated that high sea level fluctuation due to Pleistocene climatic changes led to sea level transgression and regression. Therefore, when the sea level rose, the mouth of the rivers was displaced inland, promoting population isolation. In contrast, when the sea level dropped, the influence of the freshwater could extend into the continental shelf, promoting connections between previously isolated populations and gene flow (Abreu et al. 2020, Pio and Carvalho 2021).

The potential effect of sea level fluctuations and Pleistocene climatic changes could be masked by ENM caveats. Most ENM algorithms identify suitability areas through correlations with species occurrence data, ignoring factors such as habitat saturation, response to environmental changes, and biotic interactions (Cordellier and Pfenninger 2009; Wiens et al. 2009). Moreover, they cannot deal with aquatic organisms since climatic variables are optimized for terrestrial organisms, leading to suitability area overestimations (Domisch et al. 2011; Elith et al. 2011). Nevertheless, a potential Pleistocene climatic change effect was effectively detected in *Herichthys pantostictus* through the 12% reduction in its suitability area from LGM to mid-Holocene compared to the 4% reduction for *Herichthys carpintis* in the same period (Table 3). Indeed, a close inspection of Fig. 5 showed that...
Figure 5. Potential geographic distribution of *Herichthys carpintis* and *Herichthys pantostictus* identified using an ecological niche modelling under current bioclimatic conditions (1950–2000), as well as projections to three paleoclimatic periods (LIG, 120 kya; LGM, 21 kya; Mid Holocene, 6 kya). Darker areas represent the pixels of major suitability conditions for the presence of the taxon.
most of *H. pantostictus*’ lost areas were located toward the coastline. Therefore, despite sympatry across most of their geographic distribution, both species showed different responses to environmental variables, as has been previously proposed based on the inequivalence in their environmental niches (see table S3 in Mejía et al. 2022).

Low genetic diversity levels have been associated with the absence (Olivieri et al. 2008; Craul et al. 2009; McMahan et al. 2017) or the presence of genetic structure (Agbebi et al. 2016). In this study, *h* values ranged from medium (0.390) to high (1.000) with low *π* values (0.001–0.026), suggesting a sudden population expansion (Dufnner et al. 2006; Barluenga and Meyer 2010; Ferreira et al. 2015; Azevedo et al. 2017) that could cause the absence of population genetic structure. However, this apparent absence of genetic structure can be counteracted by the cichlids’ life history traits, such as territoriality, parental care, and philopatric behavior, which restrict the displacement of individuals and allow new genetic variants to arise (Budaev et al. 1999; Sofia et al. 2008; Pereira et al. 2009; Seif 2011). Therefore, the genetic structures of *H. carpintis* and *H. pantostictus* we inferred in this study could be the result of populations that experienced a demographic expansion and life-history-related traits.

**Demographic history and ENMs.** The BSPs for both markers and species (Fig. 4) suggest a population expansion, consistent with other neotropical cichlids (Barluenga and Meyer 2010; Bagley et al. 2017; McMahan et al. 2017). The demographic expansion found in this study could be attributable to the Pleistocene’s heating and cooling periods, as has been proposed for other neotropical fishes (Bagley et al. 2013; McMahan et al. 2017, Beltrán-López et al. 2018). Nevertheless, our study results do not correspond with the results found in the ENM, where a reduction in the available suitable areas was inferred for both species since the LIG (Table 3). In *H. carpintis*, the results suggest a slight reduction (4%) in the suitability area from LIG to mid-Holocene (Table 3) but a drastic reduction (22%) for *H. pantostictus* that must have impacted its demographic history. However, this pattern was not present in the BSP (Fig. 4).

These results could be attributable to different factors. For *H. carpintis*, the ENM showed that since the LIG, the most suitable areas for this species were toward the coastline and have existed up to the present (Fig. 5). Therefore, sea level invasions and regressions during glacial periods had a negligible effect on this species’ effective population but did affect its spatial expansion. The gene flow analysis and DIYABC scenarios suggest a coast-to-inland colonization process for *H. carpintis*, consistent with that proposed for other cichlids in Middle America (Bagley et al. 2013; McMahan et al. 2017).

For *H. pantostictus*, the ENM suggested that the most suitable areas were inland since the LGM. Therefore, we expect that eustatic sea level changes must have impacted this species’ population size over time. Dry conditions and low precipitation characterized the LGM in Central Mexico, with the evaporation of some water bodies (Caballero et al. 2010) explaining the drastic reduction in this species’ available suitable areas from the LGM to mid-Holocene and the apparent slight reduction in its effective population size inferred in the BSP with the COI marker (Fig. 4). Another possible scenario is that the ancestral *H. pantostictus* population was on the coastline and the increase in sea level forced the species to move to inland, as suggested by the DIYABC results with the COI marker (Table 2). Therefore, we could suggest that, contrary to expectations, glacial periods did not significantly affect the demographic history of these cichlid species in Northeast Mexico, as has been previously documented in other co-distributed species (Rocamontes-Morales et al. 2021).

**Conclusions**

This study’s results suggest that both species (*Herichthys carpintis* and *Herichthys pantostictus*) arose and developed in different geographic areas: *H. carpintis* in lacustrine environments and low-flow rivers in coastal regions and *H. pantostictus* in clear water inland rivers with medium to high flows. The current sympatry of both species is more compatible with a pseudo-incongruence pattern resulting from dispersal events during the Pleistocene, supported by the DIYABC results for *H. carpintis* with both markers and *H. pantostictus* with the D-loop marker, than an independent response to paleoclimatic events, supported by the ENMs and the DIYABC results for *H. pantostictus* with the COI marker. However, further analysis with other molecular markers is required to disentangle both species’ biogeographic histories and to fully understand the complex history of this region of northeastern Mexico.

**Acknowledgments**

Fabian Pérez-Miranda acknowledges CONACyT for a postdoctoral fellowship, and we also thank Gerardo Zúñiga for his comments.

**References**


FAO (2022) Hydrological basins of the world. https://data.apps.fao.org/catalog/dataset/26215a41-6383-4aa4-a21-74333e0b03ae


Supplementary material 3

GenBank numbers

Authors: Fabian Pérez-Miranda, Omar Mejía, Benjamín López, Eduardo Soto-Galera, Amairany Bernal-Portillo, Wilfredo A. Matamoros
Data type: xlsx
Explanation note: List of accession numbers in GenBank for the unique haplotypes generated in this study.
Copyright notice: This dataset is made available under the Open Database License (http://opendatacommons.org/licenses/odbl/1.0). The Open Database License (ODbL) is a license agreement intended to allow users to freely share, modify, and use this Dataset while maintaining this same freedom for others, provided that the original source and author(s) are credited.
Link: https://doi.org/10.3897/ai.ep.53.112183.suppl3

Supplementary material 4

Accumulation curves

Authors: Fabian Pérez-Miranda, Omar Mejía, Benjamín López, Eduardo Soto-Galera, Amairany Bernal-Portillo, Wilfredo A. Matamoros
Data type: pdf
Explanation note: Haplotype accumulation curves to estimate the fraction of haplotype diversity captured from sample size used in this study. A) Haplotype accumulation curve for the COI marker in H. pantostictus. B) Haplotype accumulation curve for the D loop marker in H. pantostictus. C) Haplotype accumulation curve for the COI marker in H. carpintis. D) Haplotype accumulation curve for the D loop marker in H. carpintis. The R value represent the fraction of haplotype diversity recovered by sample size.
Copyright notice: This dataset is made available under the Open Database License (http://opendatacommons.org/licenses/odbl/1.0). The Open Database License (ODbL) is a license agreement intended to allow users to freely share, modify, and use this Dataset while maintaining this same freedom for others, provided that the original source and author(s) are credited.
Link: https://doi.org/10.3897/ai.ep.53.112183.suppl4

Supplementary material 5

Bayes factor comparison among gene flow models

Authors: Fabian Pérez-Miranda, Omar Mejía, Benjamín López, Eduardo Soto-Galera, Amairany Bernal-Portillo, Wilfredo A. Matamoros
Data type: xlsx
Explanation note: Model comparison of the different gene flow models analysed in the populations of Herichthys pantostictus and Herichthys carpintis for the mitochondrial markers COI and D-loop. The models were compared through a Bayes Factor (BF) using the Bezier marginal likelihood obtained in Migrate. See Suppl. material 2: table S2 for complete details of models. IML (Bezier likelihood), LBF (Likelihood Bayes Factor), mod. rank (model ranking), mod.prob (model probability).
Copyright notice: This dataset is made available under the Open Database License (http://opendatacommons.org/licenses/odbl/1.0). The Open Database License (ODbL) is a license agreement intended to allow users to freely share, modify, and use this Dataset while maintaining this same freedom for others, provided that the original source and author(s) are credited.
Link: https://doi.org/10.3897/ai.ep.53.112183.suppl5

Supplementary material 6

Competing scenarios

Authors: Fabian Pérez-Miranda, Omar Mejía, Benjamín López, Eduardo Soto-Galera, Amairany Bernal-Portillo, Wilfredo A. Matamoros
Data type: pdf
Explanation note: Schematic representation of the different competing scenarios considering to test the colonization routes of H. carpintis and H. pantostictus by Approximate Bayesian Computation (ABC) of the mitochondrial DNA markers COI and D-loop. In the first scenario a costal to inland colonization process was inferred, on the other hand, in the second scenario, an inland to coast colonization process was assumed.
Copyright notice: This dataset is made available under the Open Database License (http://opendatacommons.org/licenses/odbl/1.0). The Open Database License (ODbL) is a license agreement intended to allow users to freely share, modify, and use this Dataset while maintaining this same freedom for others, provided that the original source and author(s) are credited.
Link: https://doi.org/10.3897/ai.ep.53.112183.suppl6