Garduño-Sánchez et al. *Zool. Res.* 2023, 44(4): 761−775
https://doi.org/10.24272/j.issn.2095-8137.2022.437

ZR | Zoological Research

Article Open Access

# Mitochondrial phylogeography and molecular evolution of the rhodopsin visual pigment in troglobitic populations of *Astyanax mexicanus* (De Filippi, 1853)

Marco A.A. Garduño-Sánchez[1,2], Vladimir de Jesus-Bonilla[1,3], Silvia Perea[1], Ramses Miranda-Gamboa[4], Andrea Herrera-García[1], Mauricio de la Maza Benignos[5], Claudia Patricia Ornelas-García[1,*]

[1] *Colección Nacional de Peces*, *Departamento de Zoología*, *Instituto de Biología*, *Universidad Nacional Autónoma de México*, *Ciudad de México*, *C.P.* 04510, *México*

[2] *Posgrado en Ciencias Biológicas*, *Colección Nacional de Peces*, *Departamento de Zoología*, *Instituto de Biología*, *Universidad Nacional Autónoma de México*, *Ciudad de México*, *C.P.* 04510, *México*

[3] *Licenciatura en Ciencias Forenses*, *Facultad de Medicina*, *Universidad Nacional Autónoma de México*, *Ciudad Universitaria*, *Coyoacán*, *México City*, *C.P.* 04510, *México*

[4] *Instituto de Energías Renovables*, *Universidad Nacional Autónoma de México*, *Temixco*, *Morelos C.P.* 62580, *Mexico*

[5] *Facultad de Ciencias Biológicas*, *Universidad Autónoma de Nuevo León*, *Ciudad Universitaria*, *San Nicolás de los Garza*, *Nuevo León*, *C.P.* 66450, *México*

## ABSTRACT

Cave-adapted animals provide a unique opportunity to study the evolutionary mechanisms underlying phenotypic, metabolic, behavioral, and genetic evolution in response to cave environments. The Mexican tetra (*Astyanax mexicanus*) is considered a unique model system as it shows both surface and cave-dwelling morphs. To date, at least 33 different cave populations have been identified, with phylogenetic studies suggesting an origin from at least two independent surface lineages, thereby providing a unique opportunity to study parallel evolution. In the present study, we carried out the most exhaustive phylogeographic study of *A. mexicanus* to date, including cave and surface localities, using two mitochondrial markers (cytochrome *b* (cyt *b*) and cytochrome c oxidase subunit I (*COI*)) and nuclear rhodopsin visual pigment (*rho*). Additionally, we inferred the molecular evolution of *rho* within the two contrasting environments (cave and surface) and across three geographic regions (Sierra de El Abra, Sierra de Guatemala, and Micos). In total, 267 individuals were sequenced for the two mitochondrial fragments and 268 individuals were sequenced for the *rho* visual pigment from 22 cave and 46 surface populations. Phylogeographic results based on the mitochondrial data supported the two-lineage hypothesis, except for the Pachón and Chica caves, whose introgression has been largely documented. The Sierra de El Abra region depicted the largest genetic diversity, followed by the Sierra de Guatemala region. Regarding the phylogeographic patterns of *rho*, we recovered exclusive haplogroups for the Sierra de El Abra (Haplogroup I) and Sierra de Guatemala regions (Haplogroup IV). Moreover, a 544 bp deletion in the *rho* gene was observed in the Escondido cave population from Sierra de Guatemala, reducing the protein from seven to three intramembrane domains. This change may produce a loss-of-function (LOF) but requires further investigation. Regarding nonsynonymous (*d*N) and synonymous (*d*S) substitution rates (omega values ω), our results revealed the prevailing influence of purifying selection upon the *rho* pigment for both cave and surface populations (ω<1), but relaxation at the El Abra region. Notably, in contrast to the other two regions, we observed an increase in the number of *d*N mutations for Sierra de El Abra. However, given that a LOF was exclusively identified in the Sierra de Guatemala region, we cannot dismiss the possibility of a pleiotropic effect on the Rho protein.

**Keywords:** Visual pigments; Selection; Cavefish; Vision loss; Loss of function

## INTRODUCTION

Populations adapted to cave habitats provide a unique opportunity to study the evolutionary mechanisms that underlie the changes in phenotype, metabolism, behavior, and



Received: 10 March 2023; Accepted: 06 July 2023; Online: 06 July 2023
Foundation items: This research was supported by the Project No. 191986, Fronteras de la Ciencia - CONACyT and Programa de Apoyo a Proyectos de Investigación e Innovación Tecnológica (PAPIIT), UNAM No. IN212419
*Corresponding author, E-mail: patricia.ornelas.g@ib.unam.mx

genetics that arise in response to these environments (McGaugh et al., 2020). Mexico is considered a hotspot for troglobitic freshwater fish fauna, with representatives from seven distinct families (Miller et al., 2005&). The blind Mexican tetra (*Astyanax mexicanus jordani*) is one of the most studied cavefish groups worldwide. Its remarkable ability to interbreed with its closest surface-dwelling relative has allowed the study of local adaptation and regressive evolution (Gross, 2012; Gross et al., 2013; Hinaux et al., 2013; Jeffery, 2009; Keene et al., 2016; Ornelas-García & Pedraza-Lara, 2016; Yoshizawa et al., 2012). Blind Mexican tetras inhabit the eastern margin of the Sierra Madre Oriental (Elliott, 2016), while the surface morphs occur in a wide distribution range along the Atlantic slope, encompassing the Tecolutla to the Bravo-Conchos basins (Miller et al., 2005).

To date, 33 cave populations of *A. mexicanus* have been recognized (Elliott, 2016; Espinasa et al., 2020; Miranda-Gamboa et al., 2023), with geographical distribution across three mountain ranges: i.e., the Sierra de El Abra, Sierra de Guatemala, and Sierra de la Colmena (Micos). Previous phylogenetic analyses have suggested that these cave populations originated from independent surface lineages during at least two different colonization events (Herman et al., 2018). Moreover, a geographical correspondence between the different mountain ranges and their phylogenetic lineages has also been established, with the Sierra de El Abra mountain range corresponding to a different lineage (previously known as “old lineage”) compared to the Sierra de Guatemala and Micos mountain ranges (known as “new lineage”) (see Bradic et al., 2012; Coghill et al., 2014; Fumey et al., 2018; Hausdorf et al., 2011; Herman et al., 2018; Strecker et al., 2004, 2012).

In contrast to the closely related surface populations, *Astyanax* cavefish have evolved a series of adaptations that confer selective advantage for life in the dark (Tabin, 2016). These traits can be considered as either regressive, such as vision and pigment loss (Bilandžija et al., 2013; Jeffery, 2009; Moran et al., 2015; Protas et al., 2006), or constructive, such as increased number and size of neuromasts (Yoshizawa et al., 2012) and enhanced olfactory capabilities (Bibliowicz et al., 2013; Blin et al., 2018; Hinaux et al., 2016; Protas et al., 2008).

Among troglobitic adaptations, vision loss in cave populations is one of the most striking evolutionary processes, which has convergently evolved in different taxonomic groups in both invertebrates (Stern & Crandall, 2018) and vertebrates (Culver & Wilkens, 2000; Culver & Pipan, 2009; Protas & Jeffery, 2012). In *Astyanax* cavefish, small eye primordia start to develop during the initial steps of embryogenesis, then undergo a short period of growth before arrest in development, with subsequent degeneration and descent into the orbits (Jeffery, 2009). However, our knowledge regarding the genes and/or regulatory sequences involved in vision loss is somewhat limited (McGaugh et al., 2014; Gross, 2016; McGaugh et al., 2014; Tabin, 2016), particularly given the potential influence of molecular convergent evolution on vision loss across *Astyanax* cavefish lineages (see Protas et al., 2006). Moreover, questions regarding the molecular mechanisms (i.e., neutral vs. selective mechanisms) that have driven vision loss in other cave-adapted animals remain to be answered (see Moran et al., 2023; Niemiller et al., 2013). Genes freed from selective constraints can evolve like pseudogenes, subject to mutation and drift, which may drive loss-of-function (LOF) mutations (Calderoni et al., 2016; Niemiller et al., 2013). However, several studies have found no support for LOF of vision-related genes in distinct troglobitic groups (Crandall & Hillis, 1997; Jeffery, 2009), although such outcomes could be attributed to the recent divergence between surface and cave lineages (Herman et al., 2018), as well as possible pleiotropic effects (Crandall & Hillis, 1997; Niemiller et al., 2013). An example of this is the dim-light-specialized visual pigment photoreceptor gene *rho*, a candidate gene for studying regressive evolution in cave-adapted organisms (Calderoni et al., 2016; Niemiller et al., 2013; Palczewski, 2006), but whose LOF frequency in such animals is very low, suggesting pleiotropic effects regarding circadian rhythms (see Stern & Crandall, 2018).

In *Astyanax* cavefish lineages, the *rho* gene has been implicated in the adaptation to perpetually dark environments, displaying marked differences in expression between surface- and cave-dwelling individuals (Gross et al., 2013). Nevertheless, the underlying evolutionary mechanisms responsible for the differences in expression remain unknown, including whether they are the result of changes in the coding region of the *rho* gene or from differential selective pressures acting upon this gene between cave and surface populations.

To gain insights into the molecular evolution of genes associated with regressive traits in cave populations, different evolutionary processes need to be considered, such as gene flow between surface and cave-dwelling populations (Bradic et al., 2012), time since cave colonization (Calderoni et al., 2016; Niemiller et al., 2013), genetic drift (Stern & Crandall, 2018), and direct or indirect selection (Calderoni et al., 2016; Niemiller et al., 2013). Within the context of *Astyanax*, a substantial body of evidence supports the occurrence of gene flow between the surface and cave populations, suggesting this may be a relevant process for certain cave populations (Herman et al., 2018) , particularly at the extremes of the Sierra de El Abra mountain range, including the northern Pachón cave and southern Chica-Chiquitita and Toro System, which exhibit constant gene flow with surface populations (Bradic et al., 2012; Herman et al., 2018; Strecker et al., 2004, 2012). Additional aspects to consider are the time since cave colonization and the potential demographic patterns affecting genetic drift in cave populations. Previous studies have suggested nearly contemporaneous cave colonization events, with the old and new lineages dating to approximately 162–191k and 114k generations, respectively (Herman et al., 2018). Thus, recognizing both biogeographic and demographic patterns across various cave systems can provide additional evidence regarding the historical contingencies impacting the molecular evolution of cave forms.

In the present study, we evaluated the phylogeographic patterns and molecular evolution of the *rho* visual pigment using the most exhaustive sampling to date, including 22 of the 33 known *Astyanax* cavefish populations and 46 surface populations. We aimed to: (1) Reconstruct the phylogeographic relationships of the *A. mexicanus* cavefish populations based on two mitochondrial markers (cytochrome *b* (cyt *b*) and cytochrome c oxidase subunit I (*COI*)) and the phylogeographic patterns of *rho*; and (2) Determine the molecular evolution of *rho* in two contrasting environments (i.e., cave and surface) and three geographic regions (i.e., Sierra de El Abra, Sierra de Guatemala, and Micos).

## MATERIALS AND METHODS

### Sample collection and DNA extraction

Sampling was conducted from 2015 to 2019 during the dry season. A total of 399 individuals were sampled from 46 surface populations of *A. mexicanus* (Bravo, Conchos, San Juan, San Fernando, Salado, Nazas, and Pánuco) and *A. aeneus* (Balsas, Coahuayana, and Grijalva), and 21 cave populations of *A. mexicanus* (five from Sierra de Guatemala, 15 from Sierra de El Abra, and one from Sierra de La Colmena) and one cave population of *A. aeneus* (Grutas de Granadas) (Figure 1; Supplementary Table S1). All cavefish samples were collected under the auspices of SEMARNAT (permit numbers: SGPA/DGVS/2438/15-16, SGPA/DGVS/05389/17, and SGPA/DGVS/1893/19). Fish specimens were captured using hand nets. Some fish were kept alive and processed under semi-sterile conditions. Fish were euthanized following conditions stated in the collection permits. All samples were preserved as voucher specimens in 95% ethanol and deposited in the Colección Nacional de Peces (CNPE), at Instituto de Biología, la Universidad Nacional Autónoma de México (IBUNAM).

A small fin clip sample was taken from all individuals, and complete individuals were preserved in ethanol as voucher specimens for future morphological analyses. The fin clips were preserved in 90% ethanol and subsequently stored at −20 °C. DNA was extracted using standard Proteinase-K in SDS/EDTA digestion buffer, NaCl (4.5 mol/L), and chloroform, as described previously (Sonnenberg et al., 2007).

### Molecular methods and alignment

A total of 275 samples were amplified for the two mitochondrial fragments, 1 064 bp for cytochrome *b* (cyt *b*), using primers GLUDG (5′-TGACCTGAAR-AACCAYCGTTG-3′) and H1690 (5′-CGAYCTTCGGATTACAAGACCG-3′) (Zardoya & Doadrio, 1998), and 659 bp for mitochondrial cytochrome oxidase subunit I (*COI*), using primers Fish F (5′-TTC TCA ACT AAC CAY AAA GAY ATY GG-3′) and Fish R (5′-TAG ACT TCT GGG TGG CCR AAR AAY CA-3′) (Ward et al., 2005). We also amplified a 796 bp section (starting at the 40$^{th}$ amino acid position of the complete gene) of the *rho* gene for 268 individuals, using primers Rh193 (5′-CNTATGAATAYCCTCAGTACTACC-3′) and Rh1073r (5′-CCRCAGCACA RCGTGGTGATCATG -3′) (Chen et al., 2003). Polymerase chain reaction (PCR) amplifications were performed in 10 μL reaction volume, containing 2 μL of template DNA, 2 μL of buffer, 0.002 mol/L $MgCl_2$, 0.0005 mol/L dNTP, 0.1 μL of *Taq* DNA polymerase, and 0.2 μL of each primer. Cycling conditions included initial denaturation at 94 °C for 4 min, followed by 30 cycles at 94 °C for 45 s, annealing temperature (AT)=48 °C for cyt *b* and AT=55 ºC for *rho* for 30 s, and 72 °C for 30 s, with a final extension at 72 °C for 10 min. For verification of PCR products, electrophoresis with a 2% agarose gel was used. The PCR products were purified with ExoSAP-IT™ (ThermoFisher, USA) and sent to Macrogen Korea, and the Genomics Unit at the IBUNAM, for sequencing. Independent alignments using the ClustalW algorithm for each mitochondria DNA (mtDNA) and *rho* fragment were constructed using Geneious R11 (https://www.geneious.com/). The sequences were deposited in GenBank (Supplementary Table S2).

### Phylogeographic analysis using mitochondrial and nuclear data

We constructed two haplotype networks, one for a concatenated matrix with the *COI* and cyt *b* sequence alignments and one for the *rho* gene. Haplotype networks are useful for analyzing intraspecific relationships when trees are

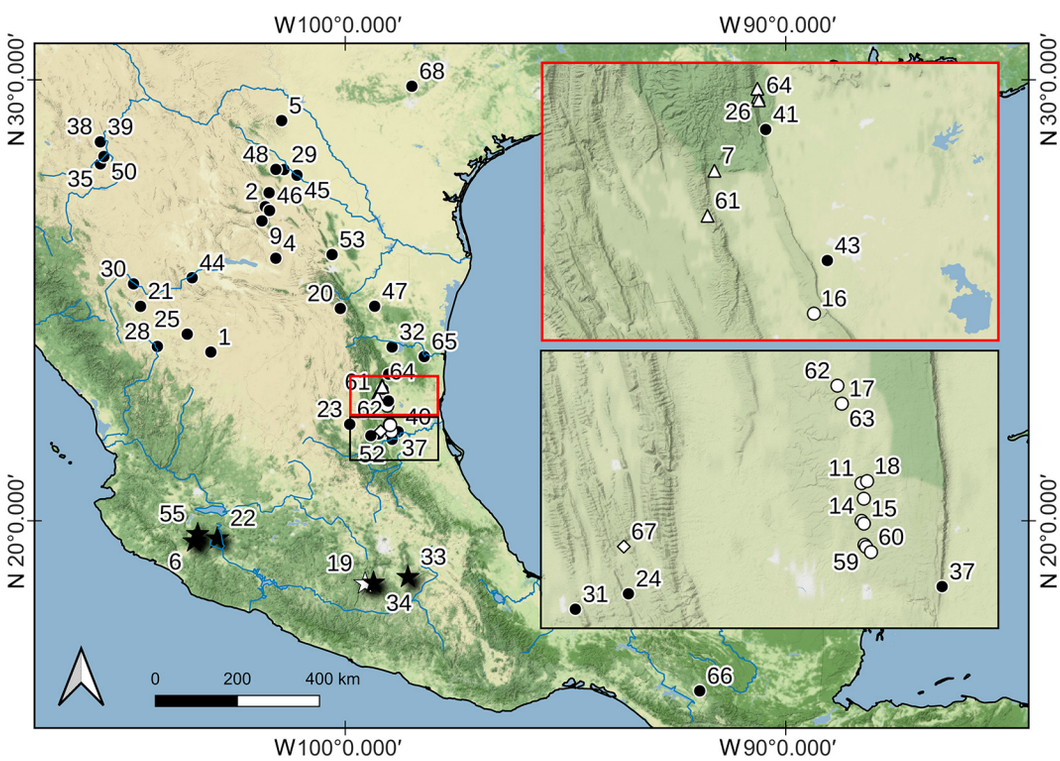


**Figure 1 Sampling points of cave and surface populations of *Astyanax mexicanus***

Boxes correspond to zoomed in cave regions: Sierra de Guatemala (white triangles), Sierra de El Abra (white circles), and Micos (white diamond). Black circles correspond to surface populations.

not sufficiently informative (Templeton, 1998). To explore haplotype relationships among the *A. mexicanus* populations, a genetic distance matrix was calculated using the *diss.dist* function in the *poppr* package v2.9.1 (Kamvar et al., 2014) implemented in R v4.0.3. A minimum spanning network (MSN) was constructed using the *poppr.msn* function from the same package, with *A. aeneus* used as an outgroup.

**Genetic diversity and demographic analyses**

Each cyt *b*, *COI*, and *rho* gene fragment was analyzed independently. We calculated the number of haplotypes (*h*), haplotype diversity (*Hd*), nucleotide diversity ($\pi$), and number of polymorphic sites using DnaSP v6.12.03 (Rozas et al., 2017).

To investigate the demographic history of the *A. mexicanus* populations, we carried out mismatch distribution analysis, as well as Bayesian skyline plot analysis (BSP) (Drummond et al., 2005). BSP analysis infers changes in the effective population size to infer demographic history (Ho & Shapiro, 2011). We ran BSP analyses separately in six groups: i.e., Sierra El Abra, Bravo-Conchos-San Fernando, Sierra de Guatemala, Pánuco, Chica-Chiquitita, and old lineage surface populations (Supplementary Table S2). Micos and Santa Anita were not analyzed due to the low number of haplotypes. A collapsed haplotype matrix was used to run Markov Chain Monte Carlo simulations with BEAST v1.7.2 (Suchard et al., 2018) for 15 million generations, with sampling every 5 000 generations and the first 10% of runs discarded as burn-in. We verified convergence of runs with Tracer v1.5 (ESS>200). We used mitochondrial freshwater fish mean substitution rates, ranging from 0.005 to 0.017 substitutions/site/million year, as priors (Porter et al., 2007). For each group, we ran two analyses, *demMin* and *demMax*, with minimum and maximum substitution rates, respectively.

We used two tests to assess changes in demographic history, i.e., Fu and Li's *D* test of neutrality (Tajima, 1989) and $R_2$ (Ramos-Onsins & Rozas, 2002). $R_2$ uses information from segregating sites and is more powerful when a small number of segregating sites is analyzed (Ramos-Onsins & Rozas, 2002). The mismatch analysis, allows us to evaluate demographic processes (i.e., population decrease or increase), since these changes in the population size leave characteristic signatures in the paired distribution of nucleotide differences by site (Rogers & Harpending, 1992). When plotting mismatch values, the populations that have experienced a recent demographic expansion will present a unimodal distribution, while those that are in demographic equilibrium will present a multimodal distribution. Similar to BSP, the mismatch distribution values were estimated for six groups, i.e., Sierra El Abra, Bravo-Conchos-San Fernando, Sierra de Guatemala, Pánuco, Chica-Chiquitita, and old lineage surface populations, calculated using DnaSP v6.12.03 considering constant and population expansion scenarios.

**Protein three-dimensional (3D) reconstruction**

To compare the effect of selective pressures between cave and surface environments, we traced nonsynonymous (*d*N) mutations from the *A. mexicanus* cavefish populations over a 3D structural model of the complete *rho* gene sequence (~345 amino acid residues) from *A. aeneus* using the Phyre2 web portal (Protein Homology/analogY Recognition Engine v2.0) for protein modeling, prediction, and analysis (http://www.sbg.bio.ic.ac.uk/phyre2/html/page.cgi?id=index; Kelley et al., 2015). Phyre2 is a suite of tools used to predict and analyze protein structure, function, and mutations using advanced remote homology detection methods to build 3D models, predict ligand binding sites, and analyze the effect of amino acid variants on protein sequences (Kelley et al., 2015). The core process employed by Phyre2 to generate a 3D model from a protein sequence consists of several technical stages, starting with the search for homologous sequences against protein sequence databases. Protein sequences that exhibit a high degree of similarity with the query sequence are used as a template to construct crude backbone-only models, with loop modeling and amino acid side chain placement performed to generate a list of 3D models (Kelley et al., 2015). The models generated by Phyre2 are ranked by a raw alignment score based on the number of aligned residues and quality of alignment. This model, in turn, is based on the similarity of residue probability distributions for each position, secondary structure similarity, and presence or absence of insertions and deletions (Kelley et al., 2015). The best rhodopsin model estimated from *A. aeneus* with Phyre2 was visualized with FirstGlance in Jmol (Herráez, 2006), where *dN* mutations from each region were plotted.

Furthermore, we detected a substantial deletion in the Escondido samples, resulting in a sequence length of approximately half the number of amino acids compared to that of the reference sequence. Specifically, the Escondido samples contained only 171 amino acids, whereas the *A. aeneus* rho sequence obtained from the complete genome in GenBank contained 345 amino acids (Supplementary Table S2). Thus, we used Phyre2 to construct a 3D model of the Escondido *rho* amino acid sequence to compare the structural changes between the complete *rho* reconstruction and *rho* with the deletion.

**Selection analyses**

To assess potential differences in selective pressures on the visual pigment *rho* between cave and surface environments, we studied the *d*N and synonymous (*d*S) substitution rates at the codon level. The ratio of *d*N/*d*S or ω is a commonly used approach to assess the strength of selection on coding genes, with ω>1 indicating positive selection, ω=1 indicating neutral selection, and ω<1 indicating purifying selection (Yang, 2007). In the present work, we used the contrast fixed effects likelihood (Contrast-FEL) approach implemented in HyPhy (Pond & Frost, 2005), which estimates the *d*N and *dS* rates and assumes that selection pressures are constant per site across phylogeny. This method compares ω rates across sites using sets of branches partitioned into groups (e.g., caves vs. surface). A significant result at a site indicates that ω differs between the two sets of branches, with either an increase or a decrease between the two sets of branches. This can assess whether any particular site within a gene is associated with adaptation between different branch groups in a phylogenetic reconstruction (Pond et al., 2020). For each site and each branch group, nucleotide frequencies, branch lengths, and three different rates (*d*S and *d*N rates between reference branches, and *d*N in test branches) are estimated. In our study, four different models were tested to detect selection pressure on sites, which have previously been used to evaluate selection on visual pigments, such as *rho* (Porter et al., 2007) under a phylogenetic framework. The models included M0, a random sampling of branches with a single ω value, M1, a two-ratio model with ω values for surface and cave populations ($\omega_{Surface} \neq \omega_{Caves}$). M2, a four-ratio model with

one ω value for surface regions and three values for cave regions ( $\omega_{Surface}$≠ωSierra de El Abra≠ωMicos≠ωSierra de Guatemala). And M3, a five-ratio model with different ω values for surface lineages and cave regions (ωSurface New Lineage≠ωSurface Old Lineage≠ωSierra de El Abra≠ ωMicos≠ωSierra de Guatemala). Corrected Akaike information criterion (AICc) was used to compare the different models.

Using the best model recovered based on FEL (i.e., $\omega_{Surface}$≠$\omega_{Caves}$), we estimated the ω rates in PAML v4.9d using two different methods (Yang, 2007): a count-based method (YN00) (Yang & Nielsen, 2000) implemented in "yn00", and a maximum-likelihood (ML) method (Goldman & Yang, 1994; Muse & Gaut, 1994) implemented in "codeml". The YN00 count-based method estimates the number of *d*S and *d*N substitutions per site, considering two major features of DNA sequence evolution: i.e., transition/transversion bias (*k*) and base (codon) frequency bias (Yang & Nielsen, 2000). The ML method for estimating *d*S and *d*N between two sequences is based on an explicit model of codon substitution, offering flexibility to incorporate various aspects of the substitution process, including *k*, codon usage bias, and even chemical differences between amino acids, into the model (Goldman & Yang, 1994; Yang & Nielsen, 2000). As both methods use paired comparisons to evaluate ω, comparisons were carried out using the complete *A. aeneus rho* sequence against the *rho* sequences of the *A. mexicanus* cave and surface populations. For real data, calculations were based on multiple sequence alignments, with sites containing missing data removed. The *d*N/*d*S ratios were then plotted according to *rho* haplotypes using *ggplot2* (Wickham, 2016).

For both HyPhy and PAML analyses, we estimated an ML tree using IQ-TREE (Nguyen et al., 2015). Nucleotide model selection was performed with ModelFinder (Kalyaanamoorthy et al., 2017) implemented in IQ-TREE to determine the best-fit model. ModelFinder supports a wide range of substitution models for DNA, protein, codon, binary and morphological alignments (Minh et al., 2022). For each topology, we used ultrafast bootstrap approximation (UFBoot) (Hoang et al., 2018) and SH-like approximate likelihood ratio test implementation (Guindon et al., 2010) as support values.

## RESULTS

### Phylogeographic analyses using mitochondrial and nuclear data

We sequenced a total of 399 individuals from 46 surface populations of *A. mexicanus* (Bravo, Conchos, San Juan, San Fernando, Salado, Nazas, and Pánuco) and *A. aeneus* (Balsas, Coahuayana, and Grijalva), and 21 cave populations of *A. mexicanus* (five from Sierra de Guatemala, 15 from Sierra de El Abra, and one from Sierra de La Colmena) and one cave population of *A. aeneus* (Grutas de Granadas) (Supplementary Table S1). Regarding the markers, we sequenced 275 and 268 individuals for the mtDNA and *rho* fragments, respectively (Supplementary Table S2). The amplified *rho* fragment spanned from amino acid 40 to 305, based on the complete reference gene. The haplotype network derived from mitochondrial markers (cyt *b* and *COI*) showed a bush-like pattern, with no dominant haplotype regarding frequency (Figure 2). The network showed correspondence between the haplogroups and their geographic distribution. Thus, haplogroup A clustered specimens from the old lineage surface populations, including populations from Tamasopo, Río Gallinas, and Peroles, with two haplotypes from the Bravo and Conchos basins and the Arroyo and Sabinos caves in the Sierra de El Abra region (old lineage cave populations). In haplogroup B, most of the new lineage surface specimens from northern Mexico clustered with haplotypes from the Bravo-Conchos basins, together with two *A. aeneus* haplotypes from the Ahuehuevo and Horcones basins. Haplogroup C included old lineage cave individuals from the Sierra de El Abra region clustered with haplotypes from the Piedras, Arroyo, Sabinos, Tinaja, Jos, Palma Seca, and Montecillos cave populations (Supplementary Table S2). Finally, intercalated haplotypes from the Soto-La Marina, Pánuco, and Santa Anita basins were recovered with new lineage cave populations, with no apparent geographical or environmental-based structure.

For the nuclear marker, among the 22 cave populations analyzed, we detected a 544 bp deletion in the *rho* gene in the Escondido cave population, but a complete *rho* gene in the remaining cave populations from Guatemala, Sierra de El Abra, and Micos. Thus, the Escondido sequences were excluded from the haplotype network, but were included in the protein 3D reconstruction (see below). In the rho haplotype network, which was constructed using 267 individuals (Figure 3), we observed the presence of four distinct haplogroups. These haplogroups had a central haplotype from which other haplotypes were derived. In addition, we identified haplogroups unique to caves, such as haplogroup IV from the Guatemala region. Haplogroup I was mostly comprised of haplotypes from the El Abra region (Cuates, Piedras, Pichihumo, Arroyo, Tinaja, Jos, Montecillos, Palma Seca, Tigre, Yerbaniz, and Sabinos), with the central haplotype shared by surface specimens from northern Mexico (Bravo and Conchos). Haplogroup II comprised a central haplotype primarily consisting of old lineage surface specimens (Rascon, Peroles, and Tamasopo), but also several new lineage surface specimens (Choy) and cave individuals from El Abra region (Pichihumo). Haplogroup III consisted of a central haplotype that included specimens from different regions, mostly new lineage surface specimens (Bravo, Conchos, Soto La Marina, Lajia, and Pánuco) and cave individuals (El Abra region; Chiquitita and Cuates). This central haplotype also included cave specimens from Micos and Sierra de El Abra, and surface individuals from the Bravo, Conchos, and Pánuco river basins. Finally, haplogroup IV consisted exclusively of cave individuals from Sierra de Guatemala.

### Genetic diversity and demographic analyses

In general, haplotype diversity was high in both mitochondrial markers (cyt *b* and *COI*, *h*>0.5) for all geographic regions, except for the Pánuco Basin for cyt *b* (*h*=0.39), with seven different haplotypes from 28 different individuals analyzed. When comparing surface and cave population lineages, nucleotide diversity was larger for the old lineage ($\pi$: 0.0095–0.0225). For the nuclear marker (*rho*), both haplotype diversity (*h*: 0.211–0.619) and nucleotide diversity ($\pi$: 0.00042–0.0018) were lower than that of the mitochondrial markers (Table 1).

Under different environments (cave vs. surface), both haplotype and nucleotide diversities tended to be higher in the surface than in the cave lineages for both mitochondrial markers. In contrast, for the *rho* gene, diversities were similar for both surface and old lineage cave populations (i.e. Sierra

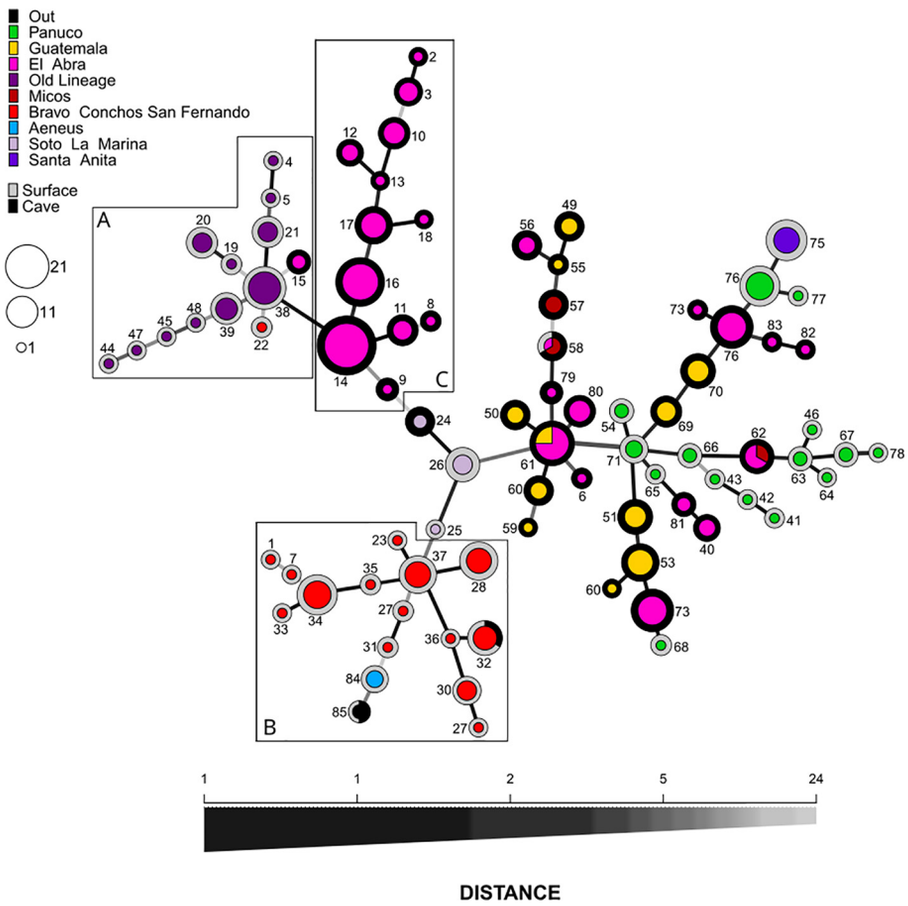


**Figure 2 Haplotype network for mitochondrial markers (cyt *b* and *COI*) in *Astyanax mexicanus***

Circle size indicates frequency of each haplotype. Exterior circle indicates environment (cave and surface) and inner circle indicates region in which haplotypes were collected.

de El Abra), while genetic diversity was two-fold lower in Sierra de Guatemala than in Sierra de El Abra, despite the latter group having double the number of specimens. The unique haplotype in the Micos region prevented estimation of genetic diversity in this cave system for the mitochondrial and *rho* markers.

The mismatch curves were multimodal for the Sierra de El Abra, Bravo-Conchos-San Fernando, Pánuco, and old lineage surface population (i.e., Río Gallinas and Peroles), but unimodal for Sierra de Guatemala (Figure 4A). Fu and Li's *D* test was negative and significant for Bravo-Conchos-San Fernando, indicating an excess of singletons, suggesting a population under expansion. However, we did not recover any significant $R_2$ test values (Table 1).

Although the Bayesian Skyline Plot (BSP) analysis took into account the maximum and minimum substitution rates (0.017 to 0.005 substitutions/site/million year, respectively), the results showed inconsistent time scales. The estimated time frames ranged from hundreds of thousands of years to thousands of years (Figure 4B, C). The Bravo-Conchos, Pánuco, and Sierra del Abra lineages exhibited a decline in effective population size, followed by expansion. Based on the slower substitution rate, this demographic decline occurred between 300–150 thousand years ago (kya), with an expansion between 250–100 kya; based on the fastest substitution rate, the contraction occurred between 100–25 kya, with an expansion between 50–10 kya. Under both substitution rates, the contraction and subsequent demographic expansion were older in the surface populations (i.e., Bravo-Conchos and Pánuco) than in the cave populations (i.e., Sierra de El Abra). Interestingly, the old lineage surface population did not show these demographic changes. Lastly, the BSP analysis conducted for the Guatemala region indicated a recent expansion, which aligns with the findings from the mismatch analysis, as evidenced by the unimodal distribution curve (Figure 4A). The contraction and expansion dates agreed with timings observed for the surface new lineage area and the Sierra de El Abra region using both substitution rates. Nevertheless, none of the tests (Fu and Li's *D* and R2) yielded significant results.

### Protein 3D reconstruction

The rho model of *A. aeneus* was constructed using Phyre2, utilizing the crystal structure of rho bound to arrestin from *Mus musculus* as a template (see methods) which served as the basis for constructing the final model (PBD code: 4ZWJ, Kang et al., 2015). Phyre2 modeled 335 residues of the *rho* sequence (95% of the coding region) with 100.0% confidence using the single highest scoring template (Figure 5A). The identity percentage between the *rho* sequence of *A. aeneus* and the template was 76%. According to JSmol, the model contained 67.5% helix structures (62.7% alpha helices and 4.8% pi helices), 3.6% beta strands, and 8.1% turns. Following *rho* reconstruction, *d*N substitutions of the Rho protein from the cave and surface populations were mapped on the 3D protein model reconstruction. Of note, more nonsynonymous mutations were present in the cave populations along the entire protein and were more numerous

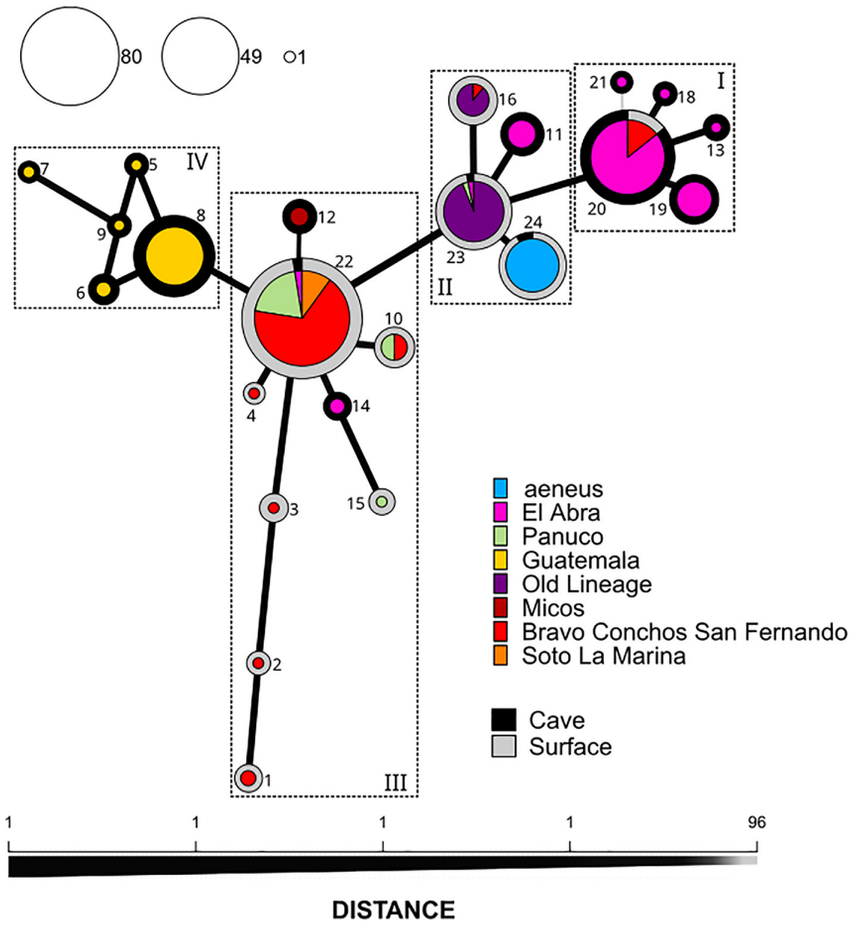


**Figure 3 Haplotype network for nuclear marker (*rho*) in *Astyanax mexicanus***

Circle size indicates frequency of each haplotype. Exterior circle indicates environment (cave and surface), and inner circle indicates region in which haplotypes were collected.

**Table 1 Genetic diversity of six geographic regions, considering both cave and surface populations from old and new lineages**

| Lineage | Geographic region | *N* | *h* | *Hd* | *S* | $\pi$ | *k* | Theta /site | *D* | $R_2$ |
|---|---|---|---|---|---|---|---|---|---|---|
| *COI* | | | | | | | | | | |
| Old lineage cave | Sierra del Abra | 80 | 11 | 0.61 | 21 | 0.0095 | 5.95 | 0.007 | 1.74895 | 0.141742 |
| New lineage cave | Sierra de Guatemala | 36 | 5 | 0.60 | 4 | 0.0014 | 0.79 | 0.002 | 1.05441 | 0.10226 |
| | Micos | 7 | 1 | – | – | – | – | – | – | – |
| New lineage surface | BCSF | 41 | 12 | 0.86 | 29 | 0.0053 | 3.43 | 6.778 | –3.5354* | 0.109696 |
| | Pánuco Basin | 28 | 12 | 0.87 | 18 | 0.00653 | 4.24 | 4.626 | –1.7265 | 0.104854 |
| Old lineage surface | Rascón+Peroles | 35 | 8 | 0.79 | 43 | 0.0178 | 11.48 | 10.441 | 1.3577 | 0.131801 |
| cyt *b* | | | | | | | | | | |
| Old lineage cave | Sierra del Abra | 80 | 8 | 0.66 | 18 | 0.0142 | 5.49 | 3.634 | 1.21952 | 0.151945 |
| New lineage cave | Sierra de Guatemala | 36 | 2 | 0.51 | 1 | 0.0012 | 0.51 | 0.001 | 0.57434 | 0.256349 |
| | Micos | 7 | 1 | – | – | – | – | – | – | – |
| New lineage surface | BCSF | 41 | 5 | 0.51 | 20 | 0.0060 | 2.29 | 0.012 | –2.505* | 0.101897 |
| | Pánuco Basin | 28 | 7 | 0.39 | 16 | 0.0065 | 2.61 | 4.112 | 0.0964 | 0.07573 |
| Old lineage surface | Rascón+Peroles | 35 | 10 | 0.76 | 35 | 0.0225 | 9.21 | 8.499 | 1.2448 | 0.129959 |
| *rho* | | | | | | | | | | |
| Old lineage cave | Sierra del Abra | 68 | 7 | 0.406 | 5 | 0.00097 | – | – | – | – |
| New lineage cave | Sierra de Guatemala | 36 | 4 | 0.211 | 5 | 0.00042 | – | – | – | - |
| | Micos | 3 | 1 | – | – | – | – | – | – | – |
| New lineage surface | BCSF | 107 | 7 | 0.461 | 8 | 0.00118 | – | – | – | – |
| | Pánuco Basin | – | – | – | – | – | – | – | – | – |
| Old lineage surface | Rascón+Peroles | 52 | 4 | 0.619 | 3 | 0.00093 | – | – | – | – |

*N*: Sample size; *h*: Number of haplotypes; *Hd*: Gene diversity; *S*: Number of polymorphic sites; $\pi$: Nucleotide diversity; *k*: Average pairwise nucleotide differences; *D*: Fu and Li's *D*; $R_2$: Ramos-Onsins and Rozas. –: Not estimated due to the lack of variation (i.e., Micos), or because *D* and $R_2$ were not evaluated for the *rho* fragment.

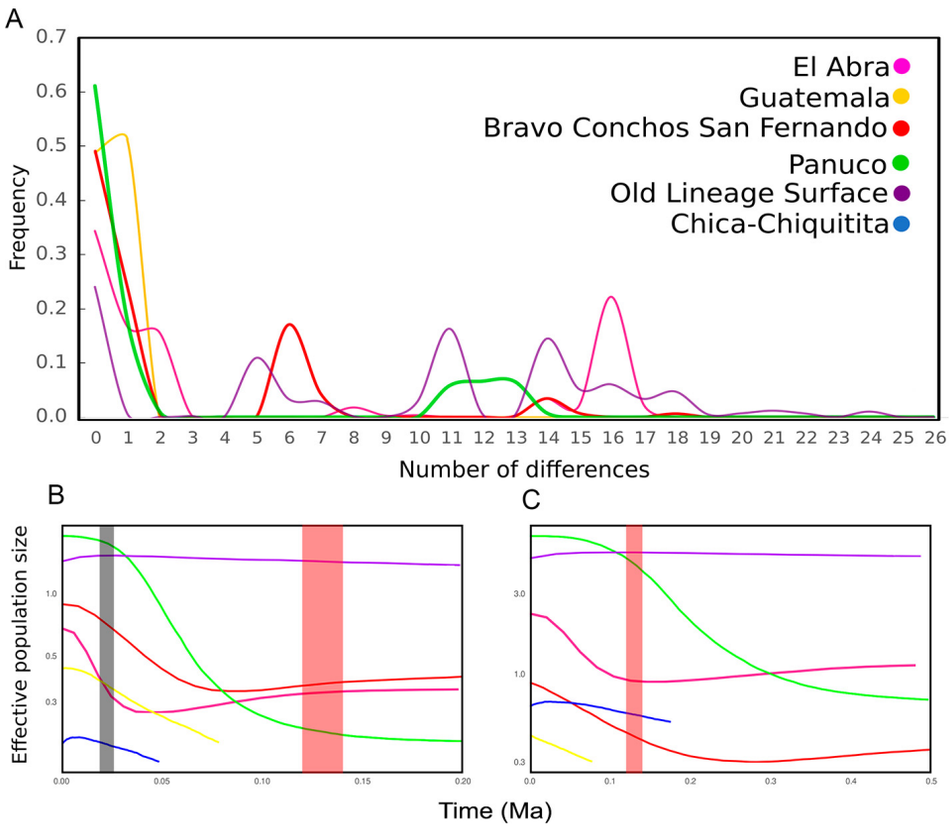


**Figure 4 Analysis of mismatch distribution and Bayesian skyline plot for demographic inferences**

A: Mismatch distribution analysis of *A*styanax *mexicanus* populations. B, C: Bayesian skyline plots of *A. mexicanus* groups considering maximum (B) and minimum (C) mutation rates reported in the literature. Colored lines represent median of effective population size. Gray bar represents last glacial maximum (26–19 kya), and red bars represent last interglacial (140–120 kya).

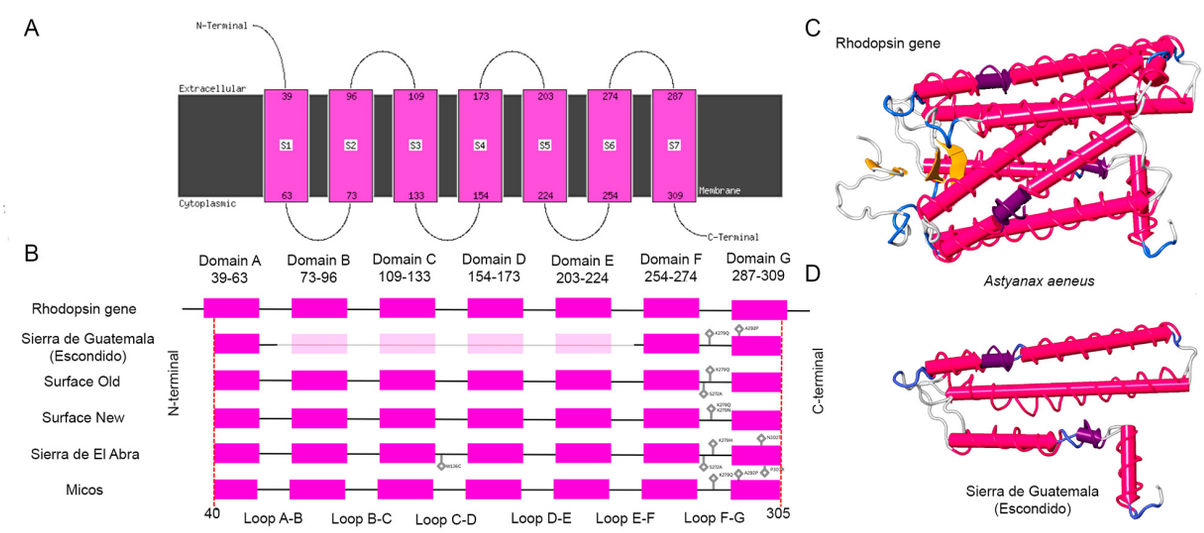


**Figure 5 Two-dimensional representation of *rho* gene in *Astyanax mexicanus*, with seven transmembrane domains (S1–S7)**

A: Extracellular domain contains N-terminus and three extracellular loops. Cytoplasmic domain contains C-terminus and three cytoplasmic loops. B: Comparison of *rho* gene using *A. aeneus rho* sequence as a reference in contrast to *rho* sequences of *A. mexicanus* from surface and cave populations. Gray diamonds correspond to nonsynonymous amino acid substitutions. First representation corresponds to *A. aeneus* reference, followed by Escondido cave population from Sierra de Guatemala with four transmembrane deletions and two amino acid changes (K297Q and A292P). Old lineage surface population showed two amino acid changes (S272A and K297Q). New lineage surface population showed two amino acid changes (K297N and K297Q). Sierra de El Abra nonsynonymous mutations showed five amino acid changes (W136C, S272A (shared with surface populations from old lineage), K279H, N302T, and P303R). Sierra de Guatemala nonsynonymous mutations showed two amino acid changes (K279Q and A292P). Micos showed one amino acid change (A292P; shared with Sierra de Guatemala). C: 3D *rho* gene model of *A. aeneus*. Alpha helices are shown as “rockets” and beta strands are shown as “ribbons”. Arrowheads indicate carboxy termini. Colors represent secondary structures over 3D model (alpha helices pink, *π* helices purple, 310 helices magenta, beta strands orange, turns blue, and coils white). D: 3D *rho* gene model for Escondido cave population considering four-domain deletion.

in Sierra de El Abra compared to the Sierra de Guatemala and Micos regions (Figure 5). The substitutions in the Sierra de El Abra region included W136C, S272A (shared with old lineage surface populations), K279H, N302T, and P303R. The cave populations from Sierra de Guatemala showed two nonsynonymous substitutions, i.e., K279Q and A292P (shared with cave populations from Micos). The Micos populations exhibited one nonsynonymous substitution (A292P). The new lineage surface populations contained two nonsynonymous mutations (K279Q and K279N), as did the old lineage surface populations (S272A and K279Q). As seen in the model, most changes occurred in the F-G loop (extracellular loop) and G transmembrane domain (Figure 5).

The Rho 3D model obtained for the Escondido cave population by Phyre2 showed a contrasting structure relative to the other cave populations and the reference *A. aeneus rho* model (Figure 5). In this case, only 171 residues from the complete sequence (49.5% of the coding region) were recovered due to the presence of a deletion in this region. The model showed 100% confidence and percentage identity with the template of 45%. According to JSmol, the model contained 76.4% helices of three different types (65.4% alpha helices, 7.1% 310 helices, and 3.9% pi helices) and 7.1% turns. In the Escondido population, nearly 50% of the coding region of Rho protein was found to be missing. This resulted in the loss of four out of the seven intramembrane domains, specifically from the transmembrane domain B to E. The missing regions also included parts of the cytoplasmic loop A-B and the extracellular loop E-F (Figure 5).

### Selection analyses

To evaluate the presence of sites under selection, we used the Contrast-FEL test (Pond et al., 2021) within the HyPhy suite (Pond & Frost, 2005). Four different hypotheses were constructed for comparison (Table 2). Based on AICc, M1 showed the lowest value and was selected as the best model, suggesting a two-ω ratio ($\omega_{Surface} \neq \omega_{Caves}$), regardless of the geographic cave regions. The overall ω values between cave and surface populations showed that branches containing cave populations had a higher ω value (ω=0.3481) than branches containing surface populations (ω=0.0869). These results indicated that both environments were under purifying selection pressure (ω<1), with a relaxation in the cave populations. Among all comparisons, no sites were detected as under selection in the Contrast-FEL test.

Analyses were carried out considering surface and cave populations independently, as well as the three cave regions (El Abra, Guatemala, and Micos) and two surface lineages (old and new).

We recovered consistent ω values using the CodeML and YN00 methods implemented in PAML (Table 3), indicating that *rho* evolved in a non-neutral fashion as the codons were more likely under purifying selection (ω<1). The lowest ω values were obtained for the old lineage surface population, with no observed substitutions across the individuals analyzed, followed by cave populations from the Micos (CodeML ω=0.001 and YN00=0.0) and Guatemala regions (CodeML $\omega_{median}$=0.001 and YN00 $\omega_{median}$=0.0). The new lineage surface population showed intermediate values (CodeML ω=0.1033 and YN00=0.1275), while the El Abra region showed the highest values (CodeML $\omega_{median}$=0.1329 and YN00 $\omega_{median}$=0.1593). Comparing the surface and cave populations, ω showed a bimodal distribution using the CodeML and YN00 methods (Figure 6), with a more heterogeneous mode in the surface populations compared to the cave populations. However, in general terms, the surface populations showed the lowest ω value, with a mainly unimodal distribution close to zero, while the cave populations showed a bimodal distribution. For the distribution of ω values among the hybrid cave populations (Figure 6), the Sierra de El Abra and Sierra de Guatemala populations both showed bimodal distribution.

## DISCUSSION

The Mexican tetra is recognized as a model organism for studying regressive evolution (Jeffery, 2009), where different surface ancestors have evolved troglobitic forms (Herman et al., 2018). This has made it possible to evaluate the parallel evolution of characteristics such as vision loss (Tabin, 2016).

**Table 2 AICc scores and ω estimates for various branch-based models assessing heterogeneous selection pressures for *rho* using Contrast-FEL test**

| Model | Descripción | AIC-c | ω estimates |
|---|---|---|---|
| M0 | Random sampling of the branches (null hypothesis) | 2 894.19 | Background=0.3316, Test=0.2314 |
| **M1** | **Two-ratio model with surface ω and a single ωfor cave branches** | **2 893.46** | **Surface=0.0869, Caves=0.3481** |
| M2 | Four-ratio model considering a single ω for the surfaceand a different ω for each cave region | 2 894.81 | Sierra de El Abra=0.4161, Sierra de Guatemala=0.3466, Micos=0.0, Surface=0.1733 |
| M3 | Five-ratio model, with a different ω value for eachsurface lineage and three ω values for each cave region | 2 896.02 | Sierra de El Abra=0.4173, Sierra de Guatemala=0.3478, Micos=0.0,Surface new lineage=0.291, Surface old lineage=0.0 |

Best-fitting model is in bold.

**Table 3 Global ω rate of *rho* gene in *Astyanax mexicanus* sequence using ML (Goldman & Yang, 1994; Muse & Gaut, 1994) and YN00 methods (Yang & Nielsen, 2000)**

| Region | CodeML | YN00 |
|---|---|---|
| | Mean ω paired value | Mean ω paired value |
| El Abra | 0.1227 (±0.046) | 0.1488 (±0.055) |
| Old lineage surface | 0 | 0 |
| Guatemala | 0.06965 (±0.02) | 0.08077(± 0.02) |
| Micos | 0.062 (±0.001) | 0.0724 (±0.001) |
| New lineage surface | 0.1033 (±0.045) | 0.1275 (±0.056) |

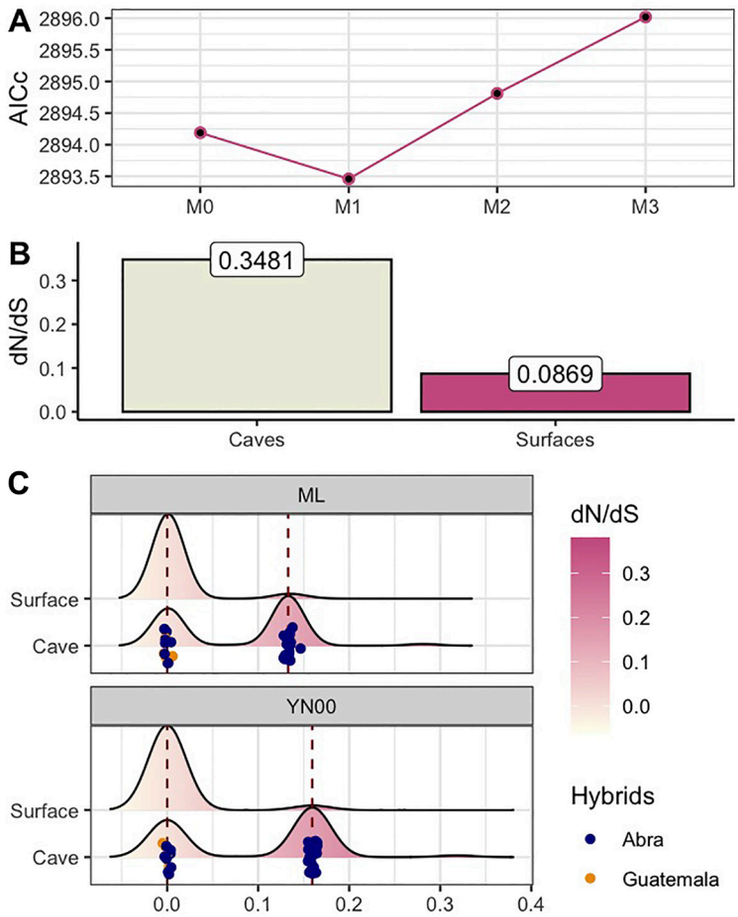


**Figure 6 Selection results from CodeML and YN00 methods**

A: Different selection models tested for *rho* gene (see Table 2), with respective AICc values considering Contrast-FEL test. B: ω values for cave (ω=0.3481) and surface branches (ω=0.0869) considered in M1. C: Ridgeplot of distribution of ω ratio (*d*N/*d*S), derived from pairwise comparison using the *A. aeneus rho* reference sequence and M1 ($\omega_{caves}$ *vs.* $\omega_{surface}$) with the ML (Goldman & Yang, 1994; Muse & Gaut, 1994) and YN00 methods (Yang & Nielsen, 2000). Dots correspond to cave populations with previously reported surface introgression (Bradic et al., 2012; Herman et al., 2018), dark-blue dots are from Sierra de El Abra (Arroyo, Chiquitita, Cuates, Pachon, Tigre, and Yerbaniz), and orange dots are from Sierra de Guatemala (Caballo Moro).

Although at least 33 troglobitic populations of *A. mexicanus* are known (Elliott, 2016; Espinasa et al., 2020; Miranda-Gamboa et al., 2023), most studies have focused on a very restricted number, thereby limiting accurate interpretations of historical and recent events. In the present study, we evaluated the phylogeographic patterns of *A. mexicanus* based on two mitochondrial markers (cyt *b* and *COI*) and comprehensive sampling, consisting of 46 surface populations (i.e., Bravo, Conchos, San Juan, San Fernando, Salado, Nazas, and Pánuco) and 21 cave populations (five populations from Sierra de Guatemala, 15 from Sierra de El Abra, and one from Sierra de La Colmena). This study provides a phylogeographic framework for certain cave populations, such as Jineo, Vasquez, and Escondido, resulting in the most complete sampling of cave and surface populations of *A. mexicanus* to date. In addition, we characterized the molecular variation of the visual pigment rhodopsin across surface and cave populations, evaluating different selection scenarios and their biogeographic regions.

**Phylogeographic patterns based on mitochondrial and nuclear data**

The phylogeographic patterns recovered in the mtDNA data showed geographic congruence, consistent with previous studies (e.g., Dowling et al., 2002; Ornelas-García et al., 2008; Strecker et al., 2004), with the cave and surface populations recovered in two independent mitochondrial lineages, and genetic segregation associated with geographic regions in each lineage. Thus, two different haplogroups were found to correspond to the previously named old lineage. Notably, haplogroup A corresponded to the surface populations in the Tamasopo River, Río Gallinas, and Peroles spring, and haplogroup C corresponded with cave-dwelling populations in Sierra de El Abra. A new population was found within the old lineage from the Rioverde Basin, corresponding with Peroles spring.

The new lineage included both surface and cave populations with geographical congruence. The surface populations of haplogroup B (Figure 2) included populations from northern Mexico and southern USA, clustering surface populations from Río Bravo, Conchos, and San Fernando, similar to our previous study (Ornelas-García et al., 2008). Within haplogroup B, the Granadas, Horcones, and Pitayos populations from the Balsas basin corresponded to sister species *A. aeneus*, forming a paraphyletic group, as suggested in previous study (e.g., Ornelas-García et al., 2008). However, recent phylogenomic studies recovered both

species as reciprocally monophyletic, and this mitochondrial pattern may correspond to an ancient introgression event between *A. mexicanus* and *A. aeneus* at their contact zone at the Gulf of Mexico, as evidenced by migration analysis (Herman et al., 2018), or to incomplete lineage sorting (ILS) within the mitogenome.

Our network also showed a mixture of both surface and cave populations, encompassing distinct geographical regions. These groups predominantly consisted of populations from the Pánuco River basin, in association with certain populations from Sierra de Guatemala, Sierra de la Colmena, and Sierra de El Abra (i.e., Chica, Chiquitita, and Toro). Similar results have been recovered for Sierra de Guatemala and Sierra de la Colmena in previous studies (Garduño-Sánchez et al., 2022) . For Sierra de Guatemala and Sierra de la Colmena, we suggest the occurrence of ILS between surface and cave populations, as evidenced in recent genomic studies differentiating between Molino cave and Choy River populations (Herman et al., 2018). In contrast, recent research has revealed the hybrid nature of several populations located at Sierra de El Abra, such as that found in Chica cave (Moran et al., 2022). Rather than aligning with the old lineage, like most other populations in the El Abra region, the population in the southern part of Sierra de El Abra is subject to a constant influx of surface-dwelling individuals, as evidenced by the strong admixture between surface and cave lineages (old lineage), resulting in inconsistencies between the nuclear and mitochondrial data. Finally, the Soto la Marina population was associated with both haplogroups from northern Mexico and the Pánuco area, with considerable geographical congruence, consistent with our previous study (Ornelas-García et al., 2008).

The rho network analysis revealed distinct patterns of haplotype segregation among the cave and surface populations, as well as different geographic regions. Specifically, haplogroup I was found to be more prevalent in the Sierra del Abra region, while haplogroup IV was exclusive to the Sierra de Guatemala region. Haplogroups II and III showed a mixture of haplotypes from various hydrological basins (Bravo-Conchos-San Fernando, Pánuco, Sierra de Guatemala) and environments (cave vs. surface). Some of this segregation could be attributed to their evolutionary history, where certain haplotypes were shared due to ancestry, while other haplotypes may be specific to cave environments.

Based on the above findings, we posit that the haplotypes exclusive to the cave environments may be associated with the selective pressures inherent to these environments, such as the variants in Sierra de Guatemala (i.e., haplogroup IV) and higher haplotypic diversity in Sierra de El Abra. Previous studies have demonstrated that the *rho* gene can exhibit adaptation under contrasting light conditions (Kelley et al., 2016; Niemiller et al., 2013; Yokoyama et al., 1995). Niemiller et al. (2013) described the loss of *rho* functionality across different Amblyopsidae cave lineages. Furthermore, independent amino acid replacements in the *rho* gene have been identified in diverse rockfish species of the *Sebastes* genus, which inhabit varying ocean depths (Kelley et al., 2016). Therefore, our model system provides the opportunity to discriminate between polymorphisms shared by ancestry and those variants exclusive to cave environments.

**Demographic analyses**

Although no significant expansion values were detected in both the Fu and Li's *D* test and R2 test for any region, differences in demographic patterns between lineages and regions were evident in both the mismatch analysis and the Bayesian Skyline Plot (BSP). These differences suggest varying demographic dynamics among the lineages and regions under investigation. First, mismatch analysis indicated that the old lineage surface population was in equilibrium (Rogers & Harpending, 1992), displaying a multimodal distribution, while BSP analysis did not provide evidence of changes in population size. While Sierra de El Abra showed a multimodal distribution, an expansion in the effective population size between 100 kya and 25 kya was possible to observe based on BSP analysis, depending on the substitution rate. Similarly, the new lineage surface population showed an expansion in population size, with relatively older expansion times estimated between 250 kya and 100 kya. The cave-dwelling populations in Sierra de Guatemala showed an expansion pattern in both mismatch and BSP analyses, although the expansion was not always significant in the former (Table 1). For Sierra de Guatemala, the estimated expansion dates were between 100 kya and 75 kya based on BSP analysis. Although these estimates were obtained from mitochondrial data, and thus warrant cautious interpretation due to potential bias (see Crandall et al., 2012), they provide important biogeographical insights.

Our study region is situated within the Chihuahuan Desert, the largest desert in North America and the second most biodiverse desert on the planet. It extends from southern USA, encompassing the Bravo and Conchos basins, to the Pánuco River basin in central Mexico. Geological and climatic events within this expansive region have played a pivotal role in shaping its current diversity, resulting in a high degree of endemism (Scheinvar et al., 2017). Notably, climatic fluctuations during the Pleistocene, such as the last glacial maximum and interglacial periods, have strongly impacted present-day diversity, leaving as evidence regions identified as fauna and flora refuges during the Quaternary (Gavin et al., 2014; Scheinvar et al., 2017). Our results support changes in population demography in the Bravo-Conchos and Pánuco basins, both previously identified as Quaternary refuges (Scheinvar et al., 2017; Schönhuth et al., 2011, 2015). Based on the slowest substitution rate, expansion would have overlapped with an interglacial period between 120–140 kya (Scheinvar et al., 2017). Alternatively, based on the fastest substitution rate, expansion would have occurred at 30 kya ago, primarily coincident with the last interglacial period (approximately 20 kya, Shakun & Carlson, 2010). Future work addressing these phylogeographic patterns with nuclear data could help to better understand the history of these populations and the geological and climatic factors that have shaped them.

Previous research has suggested that the effective population size of the Guatemala cave population is very low, i.e., less than 1 000 (Bradic et al., 2012). Consistently, our study found the lowest diversity values in Sierra de Guatemala, with a *Hd* of 0.51 for cyt *b*, 0.60 for *COI*, and 0.21 for *rho*. Field observations indicated that Jineo cave harbored the smallest population, as suggested by Elliot (2016). However, based on our results, we cannot discard the possibility that the Escondido cave population may also be under strong drift, as evidenced by the presence of a large deletion in the *rho* gene that was not shared with other populations from Sierra de Guatemala.

Thus, the cave population sizes in the three regions may be the result of both historical and recent factors, such as population drift (i.e., Escondido cave) and migration events (e.g., Chica cave see Herman et al., 2018; Moran et al., 2022), contributing to the complex demographic history of the cavefish populations.

**Repeated evolution of *rho* variation patterns in cave-adapted morphs of *A. mexicanus***

The contrasting light environments experienced by the cave and surface populations of Astyanax have given rise to distinct selective pressures. Our study provides support for the repeated occurrence of molecular variations in the rho gene within cave-adapted morphs in two separate lineages of A. mexicanus. This pattern consistently differs from the genetic variation observed in surface populations, which aligns with similar findings in other cavefish species as well as cave-adapted invertebrates (see Hart et al., 2020; Stern & Crandall, 2018). Our results demonstrated a predominance of purifying selection (ω<1) on the *rho* gene in surface-dwelling morphs, a common pattern found in nature to avoid deleterious mutations (e.g., visual pigments Crandall & Cronin, 1997, immune system genes, Mukherjee et al., 2009; mtDNA, Palozzi et al., 2018). However, we observed differences in the distribution of ω values between the environments ($\omega_{surface}$=0.086 and $\omega_{cave}$=0.3481), suggesting a relaxation of purifying selection in caves. Notably, the Sierra de El Abra region exhibited the highest number of nonsynonymous (β) substitutions compared to the other cave regions and surface lineages (i.e., β Sierra de Abra caves>β Sierra de Guatemala and Micos>β surface lineages). The best branch model favored the two-ratio cave model with different ω values for the surface branches (i.e., M1, Table 2; Figure 6A, B), as opposed to other models with the same ω value for all branches or different ω values for each cave geographic system.

Notably, the ω values obtained using PAML for specific cave populations from Sierra de Guatemala and Micos were lower compared to those of surface populations (Table 3). This observation is consistent with the suggestion that other cave organisms have not experienced a loss of purifying selection on the rho gene. It is hypothesized that the rho gene may have pleiotropic effects on circadian rhythms and other functions, making it essential for maintaining fitness in cave environments. Further investigations across different cave lineages are needed to deepen our understanding of this phenomenon (Crandall & Hillis, 1997; Stern & Crandall, 2018; Shin et al., 2012).

Three competing, non-exclusive hypotheses for regressive evolution of troglomorphic characteristics have been suggested, including natural selection, recurrent mutation with genetic drift, and pleiotropy (Borowsky, 2018; Culver & Wilkens, 2000; Protas et al., 2007). Several theories state that the loss of a trait may directly or indirectly promote an increase in fitness in the new colonized environments, as maintaining "non-necessary" structures in the new environment may be energetically costly (Culver & Wilkens, 2000; Espinasa & Espinasa, 2008). The accumulation of new mutations in the *rho* gene due to relaxed purifying selection, as evidenced in the cave-dwelling organisms from Sierra de El Abra, may potentially lead to non-functional proteins (Calderoni et al., 2016; Li & He, 2009; Niemiller et al., 2013; Policarpo et al., 2021), as well as local adaptation, suggesting that variants caused by this relaxation may provide the raw material for evolution (Baduel et al., 2019).

Although our study design does not allow for the confirmation of loss-of-function (LOF) mutations in the rho gene, we did identify a deletion in this gene within one of the 22 cave populations examined, specifically the Escondido cave in the Sierra region of Guatemala. This deletion accounted for approximately 57% of the amino acid residues, resulting in the loss of four transmembrane domains, as depicted in the reconstructed protein model (Figure 5B, D). A similar phenomenon of smaller amino acid deletions (31% of the coding region) in the rho gene has also been reported in three amblyopsid cave lineages, demonstrating repeated LOF in the rho gene within cave environments (i.e., *Troglichthys rosae*, *Amblyopsis spelaea*, and *Typhlichthys* cf. *subterraneus*; Niemiller et al., 2013). In bovine rho, a three amino acid deletion (Δ236-239; Franke et al., 1992) resulted in a 57% reduction in transducin activation, while a deletion between amino acids 244-249 completely inhibited transducin activation (Franke et al., 1992). These findings highlight the functional implications of deletions within the rho gene and support the notion of repeated LOF in cave environments across different species. The presence of only one potential LOF mutation in our study may be a consequence of recent cave colonization (Fumey et al., 2018; Herman et al., 2018), as well as possible pleiotropic effects of the *rho* gene. Colonization time to cave environments by S*inocyclocheilus* and *Lucifuga* fish, and the species *Phreatichthys andruzzii*, show a higher number of LOF mutations, exceeding millions of years (Calderoni et al., 2016; Policarpo et al., 2021; Yang et al., 2016), in contrast to hundreds of thousands of generations for *A. mexicanus* (Herman et al., 2018).

Furthermore, our observations revealed that most *rho* gene variations in the cave populations involved 1–5 amino acid substitutions (i.e., Micos and Sierra de El Abra, respectively, Figure 6B), which may affect protein function. For instance, a previous study examining the *rho* gene in *A. mexicanus* cavefish found that an amino acid change from tyrosine 261 to phenylalanine (Y261F) resulted in a shift in maximum absorbance from 496 nm in wild-type pigment to 504 nm (Yokoyama et al., 1995). Similarly, *Sebastes* rockfish exhibit differences in amino acid sequences in *rho* associated with habitat depth changes, with divergence between shallow (<500 m) and deep (>550 m) ecotypes, often involving nonsynonymous substitutions in the transmembrane domains (Kelley et al., 2016). Thus, future studies evaluating the effects of these variations in protein function are necessary to better understand the evolution of visual pigment in cave environments.

In conclusion, our study represents the first analysis of several populations previously known only from their original descriptions, allowing us to estimate their phylogeographic affinity. This study included individuals from the Vazquez, Jineo, and Escondido cave populations, greatly contributing to our understanding of regressive evolution in the *Astyanax* genus. This study also sheds light on the demographic patterns of the two *A. mexicanus* lineages and provides important insight into changes in the population size of cave and surface lineages during the Quaternary between 250 kya and 100 kya. Notably, Sierra de Guatemala exhibited the lowest diversity values compared to other cave regions, supporting previous research regarding their low effective population size (Bradic et al., 2012). Our field observations

also revealed that the Jineo cave population was the smallest. Based on our results, we cannot exclude the possibility that the Escondido cave population may be under genetic drift, as evidenced by the presence of a four-domain transmembrane deletion in the *rho* gene, representing a potential LOF mutation, which was not shared with any other population from Sierra de Guatemala. Finally, our analysis of molecular variation in the rhodopsin visual pigment across the 22 *Astyanax* cave populations provided evidence of relaxation in purifying selection, suggesting that the best-fitting model was the two-ratio model (M1), with one ω value for surface populations and another ω value for cave populations ($\omega_{Surface} \neq \omega_{Caves}$).

## SUPPLEMENTARY DATA

Supplementary data to this article can be found online.

## COMPETING INTERESTS

The authors declare that they have no competing interests.

## AUTHOR'S CONTRIBUTIONS

M.A.A.G.S. carried out laboratory protocols, data analysis, and contributed to the writing of the manuscript. V.J.B. carried out laboratory protocols, data analysis, and contributed to the writing of the manuscript. S.P. carried out data analysis and contributed to the writing of the manuscript. R.M.G. carried out sample collection and contributed to the writing of the manuscript. A.H.G. carried out laboratory protocols and contributed to the writing of the manuscript. M.D.M. carried out sample collection and contributed to the writing of the manuscript. C.P.O.G. designed the study, carried out sample collection and data analysis, and contributed to the writing of the manuscript. All authors read and approved the final version of the manuscript.

## ACKNOWLEDGMENTS

We thank Andrea Jiménez, Laura Márquez, and Nelly López from LaNaBio for their help with molecular work.

## REFERENCES

Baduel P, Quadrana L, Hunter B, et al. 2019. Relaxed purifying selection in autopolyploids drives transposable element over-accumulation which provides variants for local adaptation. *Nature Communications*, **10**(1): 5818.

Bibliowicz J, Alié A, Espinasa L, et al. 2013. Differences in chemosensory response between eyed and eyeless *Astyanax mexicanus* of the Rio Subterráneo cave. *EvoDevo*, **4**(1): 25.

Bilandžija H, Ma L, Parkhurst A, et al. 2013. A potential benefit of albinism in *Astyanax* cavefish: downregulation of the *oca2* gene increases tyrosine and catecholamine levels as an alternative to melanin synthesis. *PLoS One*, **8**(11): e80823.

Blin M, Tine E, Meister L, et al. 2018. Developmental evolution and developmental plasticity of the olfactory epithelium and olfactory skills in Mexican cavefish. *Developmental Biology*, **441**(2): 242−251.

Borowsky R. 2018. Cavefishes. *Current Biology*, **28**(2): R60−R64.

Bradic M, Beerli P, García-de León FJ, et al. 2012. Gene flow and population structure in the Mexican blind cavefish complex (*Astyanax mexicanus*). *BMC Evolutionary Biology*, **12**: 9.

Calderoni L, Rota-Stabelli O, Frigato E, et al. 2016. Relaxed selective constraints drove functional modifications in peripheral photoreception of the cavefish *P. andruzzii* and provide insight into the time of cave colonization. *Heredity*, **117**(5): 383−392.

Chen WJ, Bonillo C, Lecointre G. 2003. Repeatability of clades as a criterion of reliability: a case study for molecular phylogeny of Acanthomorpha (Teleostei) with larger number of taxa. *Molecular Phylogenetics and Evolution*, **26**(2): 262−288.

Coghill LM, Darrin Hulsey C, Chaves-Campos J, et al. 2014. Next generation phylogeography of cave and surface *Astyanax mexicanus*. *Molecular Phylogenetics and Evolution*, **79**: 368–374.

Crandall ED, Sbrocco EJ, DeBoer TS, et al. 2012. Expansion dating: calibrating molecular clocks in marine species from expansions onto the Sunda Shelf following the Last Glacial Maximum. *Molecular Biology and Evolution*, **29**(2): 707−719.

Crandall KA, Cronin TW. 1997. The molecular evolution of visual pigments of freshwater crayfishes (Decapoda: Cambaridae). *Journal of Molecular Evolution*, **45**(5): 524−534.

Crandall KA, Hillis DM. 1997. Rhodopsin evolution in the dark. *Nature*, **387**(6634): 667−668.

Culver DC, Pipan T. 2009. The Biology of Caves and Other Subterranean Habitats. Oxford: Oxford University Press.

Culver DC, Wilkens H. 2000. Critical review of the relevant theories of the evolution of subterranean animals. *In*: Wilkens H, Culver DC, Humphreys WF. Ecosystems of the World: Subterranean Ecosystems. Amsterdam: Elsevier, 381–398.

Dowling TE, Martasian DP, Jeffery WR. 2002. Evidence for multiple genetic forms with similar eyeless phenotypes in the blind cavefish. *Astyanax mexicanus. Molecular Biology and Evolution*, **19**(4): 446−455.

Drummond AJ, Rambaut A, Shapiro B, et al. 2005. Bayesian coalescent inference of past population dynamics from molecular sequences. *Molecular Biology and Evolution*, **22**(5): 1185−1192.

Elliott WR. 2016. Cave exploration and mapping in the Sierra de El Abra Region. *In*: Keene AC, Yoshizawa M, McGaugh SE. Biology and Evolution of the Mexican Cavefish. Boston: Academic Press, 9–40.

Espinasa L, Ornelas-García CP, Legendre L, et al. 2020. Discovery of two new *Astyanax* cavefish localities leads to further understanding of the species biogeography. *Diversity*, **12**(10): 368.

Espinasa M, Espinasa L. 2008. Losing sight of regressive evolution. *Evolution:Education and Outreach*, **1**(4): 509−516.

Franke RR, Sakmar TP, Graham RM, et al. 1992. Structure and function in rhodopsin. Studies of the interaction between the rhodopsin cytoplasmic domain and transducin. *Journal of Biological Chemistry*, **267**(21): 14767−14774.

Fumey J, Hinaux H, Noirot C, et al. 2018. Evidence for late Pleistocene origin of *Astyanax mexicanus* cavefish. *BMC Evolutionary Biology*, **18**(1): 43.

Garduño-Sánchez M, Hernandez-Lozano J, Moran R, et al. 2022. Phylogeographic relationships and morphological evolution between cave and surface *Astyanax mexicanus* populations (De Fillipi 1853) (Actinopterygii, Characidae). *Authorea*,doi: 10.22541/au.166979535.59484815/v1.

Gavin DG, Fitzpatrick MC, Gugger PF, et al. 2014. Climate refugia: joint inference from fossil records, species distribution models and phylogeography. *New Phytologist*, **204**(1): 37−54.

Goldman N, Yang Z. 1994. A codon-based model of nucleotide substitution for protein-coding DNA sequences. *Molecular Biology and Evolution*, **11**(5): 725−736.

Gross JB. 2012. The complex origin of *Astyanax* cavefish. *BMC Evolutionary Biology*, **12**(1): 105.

Gross JB. 2016. Convergence and parallelism in *Astyanax* cave-dwelling fish. *In*: Pontarotti P. Evolutionary Biology: Convergent Evolution, Evolution of Complex Traits, Concepts and Methods. Cham: Springer, 105–119.

Gross JB, Furterer A, Carlson BM, et al. 2013. An integrated transcriptome-wide analysis of cave and surface dwelling *Astyanax mexicanus*. *PLoS One*, **8**(2): e55659.

Guindon S, Dufayard JF, Lefort V, et al. 2010. New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. *Systematic Biology*, **59**(3): 307−321.

Hart PB, Niemiller ML, Burress ED, et al. 2020. Cave‐adapted evolution in the North American amblyopsid fishes inferred using phylogenomics and geometric morphometrics. *Evolution*, **74**(5): 936−949.
Hausdorf B, Wilkens H, Strecker U. 2011. Population genetic patterns revealed by microsatellite data challenge the mitochondrial DNA based taxonomy of *Astyanax* in Mexico (Characidae, Teleostei). *Molecular Phylogenetics and Evolution*, **60**(1): 89−97.
Herman A, Brandvain Y, Weagley J, et al. 2018. The role of gene flow in rapid and repeated evolution of cave-related traits in Mexican tetra. *Astyanax mexicanus. Molecular Ecology*, **27**(22): 4397−4416.
Herráez A. 2006. Biomolecules in the computer: JmoL to the rescue. *Biochemistry and Molecular Biology Education*, **34**(4): 255−261.
Hinaux H, Devos L, Blin M, et al. 2016. Sensory evolution in blind cavefish is driven by early embryonic events during gastrulation and neurulation. *Development*, **143**(23): 4521−4532.
Hinaux H, Poulain J, Da Silva C, et al. 2013. De novo sequencing of *Astyanax mexicanus* surface fish and Pachón cavefish transcriptomes reveals enrichment of mutations in cavefish putative eye genes. *PLoS One*, **8**(1): e53553.
Ho SYW, Shapiro B. 2011. Skyline-plot methods for estimating demographic history from nucleotide sequences. *Molecular Ecology Resources*, **11**(3): 423−434.
Hoang DT, Chernomor O, Von Haeseler A, et al. 2018. UFBoot2: improving the ultrafast bootstrap approximation. *Molecular Biology and Evolution*, **35**(2): 518−522.
Jeffery WR. 2009. Regressive evolution in *Astyanax* cavefish. *Annual Review of Genetics*, **43**: 25−47.
Kalyaanamoorthy S, Minh BQ, Wong TKF, et al. 2017. ModelFinder: fast model selection for accurate phylogenetic estimates. *Nature Methods*, **14**(6): 587−589.
Kamvar ZN, Tabima JF, Grünwald NJ. 2014. *Poppr*: an R package for genetic analysis of populations with clonal, partially clonal, and/or sexual reproduction. *PeerJ*, **2**: e281.
Kang YY, Zhou XE, Gao X, et al. 2015. Crystal structure of rhodopsin bound to arrestin by femtosecond X-ray laser. *Nature*, **523**(7562): 561−567.
Keene AC, Yoshizawa M, McGaugh SE. 2016. Biology and Evolution of the Mexican Cavefish. Boston: Academic Press.
Kelley JL, Brown AP, Therkildsen NO, et al. 2016. The life aquatic: advances in marine vertebrate genomics. *Nature Reviews Genetics*, **17**(9): 523−534.
Kelley LA, Mezulis S, Yates CM, et al. 2015. The Phyre2 web portal for protein modeling, prediction and analysis. *Nature Protocols*, **10**(6): 845−858.
Li ZQ, He SP. 2009. Relaxed purifying selection of rhodopsin gene within a Chinese endemic cavefish genus *Sinocyclocheilus* (Pisces: Cypriniformes). *Hydrobiologia*, **624**(1): 139−149.
McGaugh SE, Gross JB, Aken B, et al. 2014. The cavefish genome reveals candidate genes for eye loss. Nature communications, **5**(1), 5307.
McGaugh SE, Kowalko JE, Duboué E, et al. 2020. Dark world rises: the emergence of cavefish as a model for the study of evolution, development, behavior, and disease. *Journal of Experimental Zoology Part B: Molecular and Developmental Evolution*, **334**(7–8): 397–404.
Minh BQ, Lanfear R, Ly-Trong N, et al. 2022(2022-03-25). IQ-TREE version 2.2. 0: tutorials and Manual Phylogenomic software by maximum likelihood. http://www.iqtree.org/doc/iqtree-doc.pdf.
Miranda-Gamboa R, Espinasa L, de los Angeles Verde-Ramírez M, et al. 2023. A new cave population of *Astyanax mexicanus* from Northern Sierra de El Abra, Tamaulipas, Mexico. *Subterranean Biology*, **45**: 95−117.
Miller RR, Minckley WL, Norris SM. 2005. Freshwater fishes of Mexico (No. QL 629. M54 2005).
Moran D, Softley R, Warrant EJ. 2015. The energetic cost of vision and the evolution of eyeless Mexican cavefish. *Science Advances*, **1**(8): e1500363.
Moran RL, Jaggard JB, Roback EY, et al. 2022. Hybridization underlies localized trait evolution in cavefish. *iScience*, **25**(2): 103778.
Moran RL, Richards EJ, Ornelas-García CP, et al. 2023. Selection-driven trait loss in independently evolved cavefish populations. *Nature Communications*, **14**(1): 2557.
Mukherjee S, Sarkar-Roy N, Wagener DK, et al. 2009. Signatures of natural selection are not uniform across genes of innate immune system, but purifying selection is the dominant signature. *Proceedings of the National Academy of Sciences of the United States of America*, **106**(17): 7073−7078.
Muse SV, Gaut BS. 1994. A likelihood approach for comparing synonymous and nonsynonymous nucleotide substitution rates, with application to the chloroplast genome. *Molecular Biology and Evolution*, **11**(5): 715−724.
Nguyen LT, Schmidt HA, Von Haeseler A, et al. 2015. IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. *Molecular Biology and Evolution*, **32**(1): 268−274.
Niemiller ML, Fitzpatrick BM, Shah P, et al. 2013. Evidence for repeated loss of selective constraint in rhodopsin of amblyopsid cavefishes (Teleostei: Amblyopsidae). *Evolution*, **67**(3): 732−748.
Ornelas-García CP, Domínguez-Domínguez O, Doadrio I. 2008. Evolutionary history of the fish genus *Astyanax* Baird & Girard (1854) (Actinopterygii, Characidae) in Mesoamerica reveals multiple morphological homoplasies. *BMC Evolutionary Biology*, **8**: 340.
Ornelas-García CP, Pedraza-Lara C. 2016. Phylogeny and evolutionary history of *Astyanax mexicanus*. *In*: Keene AC, Yoshizawa M, McGaugh SE. Biology and Evolution of the Mexican Cavefish. Boston: Academic Press, 77–90.
Palczewski K. 2006. G protein-coupled receptor rhodopsin. *Annual Review of Biochemistry*, **75**: 743−767.
Palozzi JM, Jeedigunta SP, Hurd TR. 2018. Mitochondrial DNA purifying selection in mammals and invertebrates. *Journal of Molecular Biology*, **430**(24): 4834−4848.
Policarpo M, Fumey J, Lafargeas P, et al. 2021. Contrasting gene decay in subterranean vertebrates: insights from cavefishes and fossorial mammals. *Molecular Biology and Evolution*, **38**(2): 589−605.
Pond SLK, Frost SDW. 2005. Datamonkey: rapid detection of selective pressure on individual sites of codon alignments. *Bioinformatics*, **21**(10): 2531−2533.
Pond, SLK, Poon AF, Velazquez R, et al. 2020. HyPhy 2.5—a customizable platform for evolutionary hypothesis testing using phylogenies. *Molecular Biology and Evolution*, **37**(1): 295−299.
Pond SLK, Wisotsky SR, Escalante A, et al. 2021. Contrast-FEL—a test for differences in selective pressures at individual sites among clades and sets of branches. *Molecular Biology and Evolution*, **38**(3): 1184−1198.
Porter ML, Dittmar K, Pérez-Losada M. 2007. How long does evolution of the troglomorphic form take? Estimating divergence times in *Astyanax mexicanus*. *Acta Carsologica*, **36**(1): 173–182.
Protas M, Conrad M, Gross JB, et al. 2007. Regressive evolution in the Mexican cave tetra. *Astyanax mexicanus. Current Biology*, **17**(5): 452−454.
Protas M, Jeffery WR. 2012. Evolution and development in cave animals: from fish to crustaceans. *Wiley Interdisciplinary Reviews:Developmental Biology*, **1**(6): 823−845.
Protas M, Tabansky I, Conrad M, et al. 2008. Multi-trait evolution in a cave fish. *Astyanax mexicanus. Evolution & Development*, **10**(2): 196−209.
Protas ME, Hersey C, Kochanek D, et al. 2006. Genetic analysis of cavefish reveals molecular convergence in the evolution of albinism. *Nature Genetics*, **38**(1): 107−111.
Ramos-Onsins SE, Rozas J. 2002. Statistical properties of new neutrality tests against population growth. *Molecular Biology and Evolution*, **19**(12): 2092−2100.
Rogers AR, Harpending H. 1992. Population growth makes waves in the

distribution of pairwise genetic differences. *Molecular Biology and Evolution*, **9**(3): 552−569.

Rozas J, Ferrer-Mata A, Sánchez-DelBarrio JC, et al. 2017. DnaSP 6: DNA sequence polymorphism analysis of large data sets. *Molecular Biology and Evolution*, **34**(12): 3299−3302.

Scheinvar E, Gámez N, Castellanos-Morales G, et al. 2017. Neogene and Pleistocene history of *Agave lechuguilla* in the Chihuahuan desert. *Journal of Biogeography*, **44**(2): 322−334.

Schönhuth S, Blum MJ, Lozano-Vilano L, et al. 2011. Inter-basin exchange and repeated headwater capture across the Sierra Madre Occidental inferred from the phylogeography of Mexican stonerollers. *Journal of Biogeography*, **38**(7): 1406−1421.

Schönhuth S, Lozano-Vilano L, Perdices A, et al. 2015. Phylogeny, genetic diversity and phylogeography of the genus *Codoma* (Teleostei, Cyprinidae). *Zoologica Scripta*, **44**(1): 11−28.

Shakun JD, Carlson AE. 2010. A global perspective on Last Glacial Maximum to Holocene climate change. *Quaternary Science Reviews*, **29**(15–16): 1801–1816.

Shin HS, Kim NN, Choi YJ, et al. 2012. Differential expression of rhodopsin and Exo-rhodopsin genes in the retina and pineal gland of olive flounder (*Paralichthys olivaceus*). *Journal of Applied Animal Research*, **40**(3): 229−246.

Sonnenberg R, Nolte AW, Tautz D. 2007. An evaluation of LSU rDNA D1-D2 sequences for their use in species identification. *Frontiers in Zoology*, **4**: 6.

Stern DB, Crandall KA. 2018. The evolution of gene expression underlying vision loss in cave animals. *Molecular Biology and Evolution*, **35**(8): 2005−2014.

Strecker U, Faúndez VH, Wilkens H. 2004. Phylogeography of surface and cave *Astyanax* (Teleostei) from Central and North America based on cytochrome *b* sequence data. *Molecular Phylogenetics and Evolution*, **33**(2): 469−481.

Strecker U, Hausdorf B, Wilkens H. 2012. Parallel speciation in *Astyanax* cave fish (Teleostei) in northern Mexico. *Molecular Phylogenetics and Evolution*, **62**(1): 62−70.

Suchard MA, Lemey P, Baele G, et al. 2018. Bayesian phylogenetic and phylodynamic data integration using BEAST 1.10. *Virus Evolution*, **4**(1): vey016.

Tabin CJ. 2016. Introduction: the emergence of the Mexican cavefish as an important model system for understanding phenotypic evolution. *In*: Keene AC, Yoshizawa M, McGaugh SE. Biology and Evolution of the Mexican Cavefish. Boston: Academic Press, 1–5.

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

Templeton AR. 1998. Nested clade analyses of phylogeographic data: testing hypotheses about gene flow and population history. *Molecular Ecology*, **7**(4): 381−397.

Ward RD, Zemlak TS, Innes BH, et al. 2005. DNA barcoding Australia's fish species. *Philosophical Transactions of the Royal Society B:Biological Sciences*, **360**(1462): 1847−1857.

Wickham H. 2016. ggplot2: Elegant Graphics for Data Analysis. Cham: Springer.

Yang JX, Chen XL, Bai J, et al. 2016. The *Sinocyclocheilus* cavefish genome provides insights into cave adaptation. *BMC Biology*, **14**: 1.

Yang ZH. 2007. PAML 4: phylogenetic analysis by maximum likelihood. *Molecular Biology and Evolution*, **24**(8): 1586−1591.

Yang ZH, Nielsen R. 2000. Estimating synonymous and nonsynonymous substitution rates under realistic evolutionary models. *Molecular Biology and Evolution*, **17**(1): 32−43.

Yokoyama R, Knox BE, Yokoyama S. 1995. Rhodopsin from the fish, *Astyanax*: role of tyrosine 261 in the red shift. *Investigative Ophthalmology & Visual Science*, **36**(5): 939−945.

Yoshizawa M, Yamamoto Y, O'Quin KE, et al. 2012. Evolution of an adaptive behavior and its sensory receptors promotes eye regression in blind cavefish. *BMC Biology*, **10**: 108.

Zardoya R, Doadrio I. 1998. Phylogenetic relationships of Iberian cyprinids: systematic and biogeographical implications. *Proceedings of the Royal Society of London. Series B:Biological Sciences*, **265**(1403): 1365−1372.