Genetic structure and historical and contemporary gene flow of *Astyanax mexicanus* in the Gulf of Mexico slope: a microsatellite-based analysis

Rodolfo Pérez-Rodríguez [1] , **Saraí Esquivel-Bobadilla** [2] , **Adonaji Madeleine Orozco-Ruíz** [2] , **José Luis Olivas-Hernández** [2] , **Francisco Javier García-De León** Corresp. [2]

[1] 1Laboratorio de Biología Acuática, Facultad de Biología, Universidad Michoacana de San Nicolás de Hidalgo, Morelia, Michoacán, México

[2] Laboratorio de Genética para la Conservación, Centro de Investigaciones Biológicas del Noroeste, La Paz, Baja California Sur, México

Corresponding Author: Francisco Javier García-De León
Email address: fgarciadl@cibnor.mx

Background. *Astyanax mexicanus* from the river basins of the Gulf of Mexico slope are small freshwater fish that usually live in large groups in different freshwater environments. The group is considered successful due to its high capacity for dispersal and adaptation to different habitats, and the species present high morphological variability throughout their distribution in Mexico. This has produced the most extreme morphotype of the group; the hypogeous or troglobite, which has no eyes or coloration, and is probably the cause of taxonomic uncertainty in the recognition of species across the entire range. Most studies of *A. mexicanus* have mainly focused on cave individuals, as well as their adjacent surface locations, providing an incomplete evolutionary history, particularly in terms of factors related to dispersal and the potential corridors used, barriers to gene flow, and distribution of genetic variability. The aim of the present study is to determine the population structure and the degree and direction of genetic flow in this complex taxonomic group, incorporating geographic locations not previously included in analyses using microsatellite loci. Our aim is to contribute to the knowledge of the intricate evolutionary history of *A. mexicanus* throughout most of its range. **Methods.** The present study included a set of several cave and surface locations of *A. mexicanus*, which have been widely sampled along the Gulf of Mexico slope, in a genetic population analysis using 10 microsatellite loci. **Results.** Ten genetic populations or lineages were identified. In these populations, gene flow was recorded at two time periods. Historical gene flow, both inter and intra-basin, was observed among surface populations, from surface to cave populations, and among cave populations, whereas recording of contemporary gene flow was limited to intra-basin exchanges and observed among surface populations, surface to cave populations, and cave populations.

1 **Genetic structure and historical and contemporary gene flow of *Astyanax mexicanus* in the**
2 **Gulf of Mexico slope: a microsatellite-based analysis**
3

5 Rodolfo Pérez-Rodríguez[1], Sarai Esquivel-Bobadilla[2], Adonaji Madeleine Orozco-Ruíz[2], José
6 Luis Olivas-Hernández[2], Francisco J. García-De León[2]
7
8 [1]Laboratorio de Biología Acuática, Facultad de Biología, Universidad Michoacana de San Nicolás
9 de Hidalgo, Francisco J. Mújica s/n, Ciudad Universitaria, Morelia, Michoacán, 58030, México.
10 [2]Laboratorio de Genética para la Conservación, Centro de Investigaciones Biológicas del
11 Noroeste, Instituto Politécnico Nacional 195, Col. Playa Palo de Santa Rita, La Paz, B.C.S. 23096,
12 México.
13
14 Corresponding author:
15 Francisco J. García-De León[2]
16 Laboratorio de Genética para la Conservación, Centro de Investigaciones Biológicas del Noroeste,
17 Instituto Politécnico Nacional 195, Col. Playa Palo de Santa Rita, La Paz, B.C.S. 23096, México.
18 Email address: fgarciadl@cibnor.mx

Abstract

21 **Background.** *Astyanax mexicanus* from the river basins of the Gulf of Mexico slope are small
22 freshwater fish that usually live in large groups in different freshwater environments. The group is
23 considered successful due to its high capacity for dispersal and adaptation to different habitats, and
24 the species present high morphological variability throughout their distribution in Mexico. This
25 has produced the most extreme morphotype of the group; the hypogeous or troglobite, which has
26 no eyes or coloration, and is probably the cause of taxonomic uncertainty in the recognition of
27 species across the entire range. Most studies of *A. mexicanus* have mainly focused on cave
28 individuals, as well as their adjacent surface locations, providing an incomplete evolutionary
29 history, particularly in terms of factors related to dispersal and the potential corridors used, barriers
30 to gene flow, and distribution of genetic variability. The aim of the present study is to determine
31 the population structure and the degree and direction of genetic flow in this complex taxonomic
32 group, incorporating geographic locations not previously included in analyses using microsatellite
33 loci. Our aim is to contribute to the knowledge of the intricate evolutionary history of *A. mexicanus*
34 throughout most of its range.
35 **Methods.** The present study included a set of several cave and surface locations of *A. mexicanus*,
36 which have been widely sampled along the Gulf of Mexico slope, in a genetic population analysis
37 using 10 microsatellite loci.
38 **Results.** Ten genetic populations or lineages were identified. In these populations, gene flow was
39 recorded at two time periods. Historical gene flow, both inter and intra-basin, was observed among
40 surface populations, from surface to cave populations, and among cave populations, whereas

41 recording of contemporary gene flow was limited to intra-basin exchanges and observed among
42 surface populations, surface to cave populations, and cave populations.
43
Introduction
45 The Mexican characid *Astyanax* is a Neotropical genus found in North American basins, reaching
46 as far north as the Bravo River hydrographic system (Strecker, Faundez & Wilkens, 2004). This
47 group includes several species for which the taxonomy is uncertain and currently in dispute, even
48 though various morphological and molecular genetic approaches have been applied to the taxon
49 (Schmitter-Soto, 2017; Wilkens & Strecker, 2017). *Astyanax* from the river basins of the Gulf of
50 Mexico slope refer to a group of three lineages that are mitochondrially well-differentiated and
51 correspond to *Astyanax mexicanus*, *Astyanax aeneus*, and *Astyanax hubbsi* (Ornelas-García,
52 Domínguez-Domínguez & Doadrio, 2008; Ornelas-García & Pedraza-Lara, 2016). The first two
53 taxa coincide with a geographic congruence along the Gulf of Mexico slope, reflecting a vicariant
54 pattern (Miller, 1986), *i.e.*, the taxon located north of the Trans Mexican Volcanic Belt (TMVB)
55 corresponds to *A. mexicanus*, while the taxon located south of the TMVB corresponds to *A. aeneus*,
56 although both species co-occur in a contact zone along the boundaries of their distribution ranges
57 (Ornelas-García, Domínguez-Domínguez & Doadrio, 2008; Ornelas-García & Pedraza-Lara,
58 2016). In the case of *A. hubbsi*, Ornelas-García & Pedraza-Lara (2016) consider it as a relict of the
59 older expansion of *Astyanax* in North America. However, hybridization detected with
60 microsatellite markers has been reported between *A. mexicanus* and *A. aeneus*, and between *A.*
61 *mexicanus* and *A. hubbsi*, calling into question the validity of the species as independent
62 evolutionary units (Hausdorf, Wilkens & Strecker, 2011). The recurrent introgression recorded
63 among these three species or lineages of *Astyanax* of the Gulf of Mexico slope is regarded as the
64 main factor that renders species delimitation based on a single marker (Ornelas-García,
65 Domínguez-Domínguez & Doadrio, 2008; Ornelas-García & Pedraza-Lara, 2016) a difficult task
66 (Hausdorf, Wilkens & Strecker, 2011; Herman *et al.*, 2018). With a more cautious and operational
67 approach, Torres-Paz *et al.* (2018) assign all surface and cave lineages of *Astyanax* spp. of the Gulf
68 of Mexico slope to *A. mexicanus*; therefore, in this study this approach was applied.
69   In addition to the reticulate genetic pattern, *A. mexicanus* shows a high morphological
70 plasticity strongly associated with recurrent morphological convergence (parallel evolution) to
71 similar environments (Ornelas-García, Domínguez-Domínguez & Doadrio, 2008; Garita-
72 Alvarado, Barluenga & Ornelas-García, 2018). The most studied and conspicuous morphotype in
73 *A. mexicanus* is the cave morph (blind and depigmented, Gross, 2012; Garita-Alvarado, Barluenga
74 & Ornelas-García, 2018). Studies on population genetics in *A. mexicanus* have produced several
75 important findings, most of which focus on cavefish populations (Gross, 2012). These studies
76 show that the cavefish populations are derived mainly from several distinct temporary cave
77 invasions (Dowling, Martasian & Jeffery, 2002; Strecker, Bernatchez & Wilkens, 2003; Ornelas-
78 García & Pedraza-Lara, 2016). These invasions have left signs of different degrees of
79 troglomorphy in the cave populations, from individuals with pigmentation and some functional,
80 but slightly reduced, visual systems, to those with pigmentation or visual systems that are

81 significantly reduced or absent (Strecker, Faundez & Wilkens, 2004; Bradic *et al.,* 2012). As a
82 result of these events, some cave populations remain completely isolated, while others indicate
83 gene flow with the surface locations (Bradic *et al.,* 2012). Nevertheless, it should be noted that
84 most of these studies feature partial sampling, limited focus and/or different markers. Other sets
85 of morphological traits that reveal parallel evolution are the body shape and the trophic traits of
86 the head that have been associated with the trophic specialization (tooth shape, dental formula, eye
87 size, snout length, body depth, head profile and mouth orientation) of two species from two
88 independent lakes (Garita-Alvarado, Barluenga & Ornelas-García, 2018; Ornelas-García *et al*.,
89 2020). But unlike cave morphs, there are few population genetic studies that limit themselves to
90 assessing the genetic differentiation between two lacustrine divergent morphs isolated in sympatry
91 within Catemaco Lake, which were originally considered to be different genera (i.e. *Bramocharax*
92 and *Astyanax*), and currently, are even considered to be different species (*A. aeneus* and *Astyanax*
93 *caballeroi*) (Ornelas-García, Bastir & Doadrio, 2014). Therefore, it remains unknown whether the
94 population has historical gene flow with specimens from other nearby basins.
95   *Astyanax mexicanus* constitute an excellent study model group, not only because of their
96 widely studied traits associated with parallel evolution, but also due to their high capacity for
97 dispersal across a wide geographic area that has allowed colonization of different habitats and the
98 occurrence of distinct recurrent introgression events over time (Hausdorf, Wilkens & Strecker,
99 2011; Strecker, Hausdorf & Wilkens, 2012; Bradic *et al*., 2012; Coghill *et al*., 2014).
100 Determination of the magnitude and direction of both the historical and contemporary gene flow
101 of one of the freshwater fish species distributed along the Gulf of Mexico slope therefore represents
102 an important contribution to our knowledge of one of the transition zones of aquatic fauna between
103 Neotropical and Nearctic regions (Miller, 1986; Obregon-Barboza, Contreras-Balderas & Lozano-
104 Vilano, 1994). In addition to the land barriers situated among the different river basins that lie
105 along the Gulf of Mexico slope, there is a marked geographic barrier consisting of a volcanic
106 mountain range. This mountain range forms the eastern limit of the TMVB known as Punta del
107 Morro (PDM), which since its formation has served as a major geographic barrier for several
108 freshwater fish groups (Obregón-Barboza, Contreras-Balderas & Lozano-Vilano, 1994; Contreras-
109 Balderas, Obregón & Lozano-Vilano, 1996; Mateos, Sanjur & Vrijenhoek, 2002; Perdices *et al*.,
110 2002; Husley *et al*., 2004; Ornelas-García, Domínguez-Domínguez & Doadrio, 2008; Agorreta *et*
111 *al*., 2013).
112   The aim of this study is to integrate genetic information from new combination set of
113 microsatellite markers using surface and cave populations that have never been analyzed together,
114 in order to contribute to the evolutionary history of this characid fish. We wish to know how many
115 genetic populations or lineages exist in the different hydrographic basins of the Mexican Atlantic
116 slope. Furthermore, we wish to know whether the hydrological basins of the Mexican Atlantic
117 slope and the TMVB, represent barriers against gene flow for *Astyanax mexicanus*, a highly
118 dispersive species. The PDM is expected to function as a barrier to genetic flow despite the high
119 dispersive capacity of the species and the direction of gene flow is greater in surface locations
120 close to the caves than those that are further away. By including samples taken north of the location

121 of the cave populations, we seek to determine whether another, previously undetected, colonization
122 event has taken place in the caves.
123
Materials & Methods
125 **Sampling**
126 The present study included samples of *Astyanax* spp. from at least seven of the main hydrographic
127 basins of the Gulf of Mexico slope in Mexico: The Bravo, San Fernando, Soto la Marina, Pánuco,
128 Tuxpan, Papaloapan, and Grijalva-Usumacinta River basins (Table 1; Fig. 1). Fish specimens were
129 collected using electrofishing equipment and trawl nets. Tissue samples (fin clips) were preserved
130 in 96% ethanol for subsequent DNA extraction. A total of 469 individuals from 16 localities,
131 corresponding to 4 cave and 12 surface locations, were included. Collection of the samples was
132 carried out with the approval of the Mexican government (DGOPA.05003.181010-5003;
133 DGOPA.00570.288108-0291; DAPA/2/130409/0961, 230401-613-03).
134
135 **DNA extraction, PCR and genotyping**
136 Total DNA was extracted from fin tissue using a DNeasy blood and tissue kit from Qiagen. Ten
137 polymorphic microsatellite loci developed for the genus *Astyanax* (Strecker, 2003; Protas *et al.,*
138 2006) were studied (Table S1) through amplification in 15 μl reactions, following the cycling
139 conditions shown in Table S1. Individual genotyping was obtained using 6% denaturing
140 polyacrylamide electrophoresis gels, run in a vertical electrophoresis unit at 1700 V, 40A and 45
141 W for 2.5 h. DNA fragments were visualized by silver staining (Bassam & Caetano, 1993). Allelic
142 size was obtained with gels using the Sequentix Gel Analyzer (http://www.sequentix.de/).
143
144 **Statistical analyses**
145 In order to detect the presence of null alleles at every locus and in each location, the Brookfield 1
146 Index (Brookfield, 1996) and the Expectation-Maximization algorithm (Dempster, Laird & Rubin,
147 1977) were applied using MICRO-CHECKER version 2.2.3 (Oosterhout, Wills & Hutchinson,
148 2004) and FreeNA (Chapuis & Estoup, 2007), respectively. Linkage disequilibrium and significant
149 deviation from Hardy-Weinberg (HW) expectations using an exact test (10,000 dememorization
150 steps, 5000 batches of 10,000 iterations per batch) were determined with GENPOP (Raymond &
151 Rousset, 2001). Statistical significance values were corrected using the sequential Bonferroni
152 method (Rice, 1989), with a significance value of 5%.
153   Genetic variability within populations (as estimated by STRUCTURE) was evaluated by
154 the number of alleles (N), number of alleles per locus (Na), effective number of alleles (Nae),
155 observed heterozygosity (H_O), and the fixation index (F_{IS}), using GENETIX 4.05 (Belkhir *et al.*,
156 2004). Unbiased expected (He) heterozygosity was calculated in GenAlex 6.5 (Peakall & Smouse,
157 2012). Allelic richness (A_R) was calculated in the function allele.rich in PopGeReport. v3.0.4
158 (Adamack & Gruber, 2014), in R (R Core Team, 2017), which uses the rarefaction approach to
159 deal with differences in sample sizes.
160

161 **Clustering and genetic differentiation**

162 The fixation index F_{ST}, based on Weir & Cockerham (1984), was estimated with GENETIX 4.05
163 (Belkhir *et al.,* 2004) to describe genetic differentiation based on the variance in allele frequencies
164 among locations. Bonferroni corrections (adjusted $P = 0.05$) were performed by multiple
165 assessment (Rice, 1989). In addition, F_{ST} were estimated using and not using ENA algorithm as
166 implemented in FreeNA.

167 A Discriminant Analysis of Principal Components (DAPC) was performed using the
168 ADEGENET package (Jombart, 2008) in R Core Team (2012), through cross-validation of the
169 POPPR package (Jombart, Devillard & Balloux, 2010) with the xvalDapc function. This function
170 chooses the number of axes to retain by testing different numbers of PCs, and the DAPC quality
171 is subsequently evaluated by cross-validation. The DAPC is carried out on the set comprised of
172 90% of the observations, and then used to predict the groups of the remaining 10% of observations.
173 The number of PCs associated with the lowest Mean Squared Error is then retained for the DAPC.
174 This method combines the advantages of principal component analysis, by assuming neither
175 Hardy-Weinberg equilibrium nor linkage disequilibrium (Jombart, 2008), and from discriminant
176 analysis, by which it attempts to summarize the genetic differentiation between groups, while
177 omitting variation within groups (Jombart, Devillard & Balloux, 2010). Two independent analyses
178 were carried out, one using all 16 locations and the other using only the 12 surface locations, while
179 excluding the four cave locations.

180 Analysis of molecular variance (AMOVA) was conducted for four different partitions of
181 the 16 locations using ARLEQUIN 3.5 (Excoffier & Lischer, 2010). The following four grouping
182 models were employed under different criteria to maximize variance between the groups and thus
183 reveal the genetic structure contained in the data: A) All 16 locations; B) phylogenetic criteria
184 considering two subclades of *A. mexicanus* [(CC)+(SF, GV, TR, AL, ML, PCH, LCA, SLP, V)],
185 subclades of *A. aeneus* [(CAT), (TE, TA, RTZ)], and the clade of *A. hubbsi* (SAB, TIN); C)
186 hydrographic basin criteria considering the seven independent river basins included in the present
187 study [Bravo (CC), San Fernando (SF), Soto la Marina (GV, TR), Pánuco (AL, ML, PCH, SAB,
188 TIN, LCA, SLP), Túxpan (V), Papaloapan (CAT), Grijalva-Usumacinta (TE, TA, RTZ)]; D)
189 Localities were grouped according to their degree of genetic differentiation as measured by the
190 F_{ST}, which resulted in the grouping of those locations that presented non-significant differentiation
191 levels [(CC), (SF), (GV, TR), (AL, LCA, SLP), (V), (PCH), (ML), (SAB, TIN), (CAT), (TE, TA,
192 RTZ)]. Ten thousand permutations were used in each of these analyses.

193 A Bayesian model-based clustering method was also performed using STRUCTURE 2.3.3
194 software (Pritchard, Stephens & Donnelly, 2000; Pitchard, 2007) in order to evaluate population
195 structure and assign individuals to genetic clusters. This approach was conducted using the
196 admixture model, which excluded information on the location of origin and assumed independence
197 among loci and non-informative priors. Values were tested for $K = 1$ to $K = 16$, including all
198 locations, and for $K = 1$ to $K = 12$, without cave locations, that is, including only surface locations.
199 The mean and variance of the log likelihoods for each K were calculated from 10 independent runs
200 of 100,000 iterations in order to determine the highest posterior probability K. To estimate the true

201   number of clusters represented along the sampled range of *Astyanax* spp. (Evanno, Regnaut &
202   Goudet, 2005), ΔK statistics were calculated by http://taylor0.biology.ucla.edu/struct_harvest/
203   (Earl, 2011). Individual and mean population membership coefficients of ancestry in our inferred
204   demes were presented graphically in DISTRUCT, version 1.1 (Rosenberg, 2004).

205

206   **Isolation by distance and historical contemporary gene flow**
207   The pattern of isolation-by-distance (IBD) among locations was assessed by performing a Mantel
208   test, based on pairwise F_{ST} and geographic distances between distinct locations, using the Isolde
209   algorithm with the on-line version of GENEPOP (http://genepop.curtin.edu.au/). Four distinct
210   Mantel tests were performed as follows: 1) using all 16 locations; 2) using only the surface
211   populations; 3) using only the locations located north of the TMVB and 4) using only the locations
212   south of the TMVB. A total of 10,000 permutations were required to estimate the 95% upper tail
213   probability of the matrix correlation coefficients.
214   Historical gene flow (for approximately the four previous N_e generations; Beerli &
215   Felsenstein, 2001) was estimated using the program MIGRATE-N 3.6.11 (Beerli & Felsenstein,
216   2001). Mutation-scaled migration rates (M = m/μ, where m = migration rate and μ = mutation rate)
217   and mutation-scaled size theta (Θ = $4N_e$μ, where N_e = effective population size, and μ = mutation
218   rate) were estimated between locations. Different migration model parameter distributions were
219   used between every location detected by MIGRATE. Microsatellite mutation was modeled as a
220   continuous Brownian process and the mutation rate was set to constant for all loci. A static heating
221   scheme was used with four chains and temperatures (1.0, 1.5, 3.0 and 1000000), for which the
222   initial Θ and m values were generated from the F_{ST} option. Exponential priors were placed for both
223   Θ (bounded between 0 and 50.0) and M (bounded between 0 and 100.0). The analyses were
224   conducted using 10 replicates of a single long Markov chain, with 100,000 steps recorded for every
225   100 generations and the first 500,000 trees per run discarded as burn-in. Three patterns of
226   migration were estimated in *A. mexicanus* from the Gulf of Mexico slope: among surface
227   populations, among cave populations, and from surface to cave. Gene flow from cave to surface
228   has been regarded as a nonexistent event, as reflected in its negligible estimate (1.5×10^{-6} to 10.2
229    $\times 10^{-4}$, Fumey *et al.*, 2018).
230   Recent gene flow (in the past few generations) was estimated using BAYESASS 3.04
231   (Wilson & Rannala, 2003). This program estimates the posterior probability of an individual's
232   migratory history and thus allows estimation of the rate and direction of recent dispersal. The
233   Markov Chain Monte Carlo method was run for 10,000,000 iterations. Delta values (i.e. maximum
234   parameter change per iteration) were left as default (Beerli & Felsenstein, 2001).

235

236   # Results
237   Hardy-Weinberg equilibrium tests following Bonferroni correction revealed heterozygote excess
238   at the following loci and locations: locus *Am145a* for GV/TR, locus *Ast02* for AL/LCA/SLP and
239   SAB/TIN, and loci *Ast10* and *Am106b* for CAT. Nevertheless, the HW multilocus did not show
240   deviation from equilibrium (Table S2). MICRO-CHECKER detected null alleles in only two tests

241 (locus *Ast10* for CAT, and locus *Ast02* for SAB-TIN), and FreeNA detected the presence of null
242 alleles at five loci in different genetic populations: locus *Ast09* for CC, locus *Am241b* for SF,
243 locus *Am145a* for GV/TR, loci *Am122b* and *Am106b* for AL/LCA/SLP, and locus *Am106b* for
244 CAT, but, none of them at frequencies greater than 0.15 (Table S3). There was no evidence of
245 linkage disequilibrium among pairs of loci across all populations (Table S4).
246

247 **Population structure**
248 Pairwise F_{ST} values revealed high, moderate, and low divergences (Table 2). High and significant
249 values were observed among the cave locations (0.77-0.85), except in the comparison between the
250 SAB and TIN caves (0.27). Moderate and significant values were recorded for most of the pairwise
251 comparisons of surface locations, regardless of geographic distance between them (0.11-0.30),
252 while low but significant values were found for the comparisons of surface locations of the Pánuco
253 basin (AL, LCA, SLP) with those of the Papaloapan (CAT) and Grijalva-Usumacinta (TE, TA,
254 RTZ) basins (0.05-0.10). The comparisons made within the Soto la Marina (GV vs TR) and
255 Grijalva-Usumacinta (TE vs TA vs RTZ) basins presented the lowest and non-significant values
256 (0.01-0.02). Pairwise F_{ST} values using and not using ENA did not present significant differences
257 (Table S5).
258   An initial DAPC, with all 16 locations, showed the formation of four main clusters (Fig.
259 2A and A′) in which 15 discriminant functions were retained. The first two functions accounted
260 for 93.3% of the total variance. Cluster *a* corresponded to the PCH cave; cluster *b* included the
261 SAB and TIN caves; cluster *c* was formed by the ML cave, and cluster *d* corresponded to all the
262 surface locations. A second DAPC, with only the 12 surface locations, revealed the formation of
263 seven clusters (Fig. 2B and B′) in which 11 discriminant functions were retained. The first two
264 functions accounted for 94% of the total variance. Cluster *e* was formed by the CC location; cluster
265 *f* by the SF location; cluster *g* included the GV and TR locations; cluster *h* was formed by the AL,
266 LCA, and SLP locations; cluster *i* included the V location; cluster *j* included the CAT location;
267 and finally, cluster *k* included the remaining surface locations, TE, TA, and RTZ.
268   The AMOVA analysis indicated that model D of 10 populations showed the highest inter-
269 group variance (F_{CT} = 0.2449, P = 0.00) and minimal intra-group variance (F_{SC} = 0.0455, P = 0.00)
270 (Table 3). STRUCTURE also revealed 10 genetic groups. Three cave populations were assigned
271 the greatest probability of ancestry (ML: 99%, PCH: 99%, SAB-TIN: 97%), as were the following
272 epigean populations corresponding to distinct river basins: Bravo (CC: 95%), Soto la Marina
273 (GV+TR: 92%), and Tuxpan (V: 91%). Other populations that also were homogeneous, but with
274 lower assignment probability, were the Papaloapan (CAT: 85%), Grijalva-Usumacinta
275 (TE+TA+RTZ: 84%), and Pánuco (AL+LCA+SLP: 70%) basins (Fig. 1). The Pánuco and
276 Grijalva-Usumacinta basins presented a genetic admixture between them (Fig. 1), which was
277 reflected in both analyses (DAPC and F_{ST} values) (Fig. 2; Table 2). Another genetic admixture
278 was recorded between the populations of the Soto la Marina and Papaloapan basins (Fig. 1).

279   In summary, the results of F_{ST}, DAPC, AMOVA, and STRUCTURE agree that the number
280   of homogeneous genetic groups along the Atlantic slope is ten, consisting of three cave
281   populations, and seven surface populations.
282
283   **Isolation by distance, effective sample size, and gene flow**
284   In the Mantel tests for both the 16 (all) and 12 (surface only) locations, the results showed a non-
285   significant correlation ($R^2 = 0.07$, $P = 0.908$; $R^2 = 0.067$, $P = 0.125$, respectively). However, when
286   only locations to the north of the TMVB (CC, SF, GV, TR, AL, LCA, SLP, and V) were included
287   in a Mantel test, a significant correlation was found between genetic and geographic distances (R^2
288    $= 0.696$, $P = 0.0$; Fig. S6A). A fourth Mantel test, using all surface locations except for CAT (11
289   locations), also recorded a significant correlation between genetic and geographic distances ($R^2 =$
290   0.633, $P = 0.0$; Fig. S6B).
291   The smallest values of the effective population sizes (Ne) were presented by caves locations
292   (average values from 1153.334 ML to 3260 for TIN; Fig. 3). For surface locations, the lowest Ne
293   values were recorded in northern localities (average values of 5316.667 for CC and 10816.667 for
294   SF; Fig. 3). Remaining surface locations presented higher Ne values (average values from
295   16677.78 for LCA to 22450.00 for TE; Fig. 3).
296   In general, the historical migration rates (m) among locations and populations show low
297   values (below m= 0.005), except for the highest genetic flow values recorded among distinct
298   basins. Of particular note is the TR location from the Soto la Marina basin that presented a
299   symmetric flow with the CC population, and the same TR that presented an asymmetric flow
300   toward the cave population PCH. All three of these populations are assigned to *A. mexicanus* (Fig.
301   4). Intermediate gene flow values also were recorded among distinct basins (Fig. 4). It should be
302   noted that in the lowest gene flow there were no differences within basins, between basins, or
303   between caves (Fig. 4).
304   Considerable recent gene flow using BAYESASS 3.04 was detected only unidirectionally
305   among the following locations within the same cave or basin system that formed a single
306   population: for cave locations, from TIN to SAB, and for surface locations, from GV to TR in the
307   Soto la Marina River basin, from SLP and LCA to AL in the Pánuco River basin, and from TA to
308   TE in the Grijalva-Usumacinta River basin (Table 4).
309
310   **Characterizing population genetic variability**
311   The ten populations varied in terms of the polymorphism of the microsatellite loci. The number of
312   alleles per locus (N_a) ranged from 1 to 23 (average = 8.30), number of effective alleles (N_{ae}) ranged
313   from 1 to 11.8 (average = 4.0), allelic richness (A_R) ranged from 1 to 13.9 (average = 5.8), observed
314   heterozygosity (Ho) presented values from 0 to 0.923 (average = 0.485), expected heterozygosity
315   (uHe) ranged from 0 to 0.921 (average = 0.556), and the fixation index (F_{IS}) ranged from -0.292
316   to 0.665 (average = 0.145). As expected, the cave populations (ML, PCH, SAB/TIN) presented
317   the lowest values of genetic variability and the highest values of the fixation index, while the

318   surface populations showed higher variability and lower values of the fixation index (Fig. S7 and
319   Table S2).
320
321   **Discussion**
322   The *A. mexicanus* from the Gulf of Mexico slope comprises what is probably one of the most
323   studied groups of cavefish (Gross, 2012; Ornelas-García & Pedraza-Lara, 2016). However, the
324   relationship between these cavefish and their surface counterparts is still poorly understood,
325   especially among populations located in the northernmost ranges. These populations have been
326   sub-sampled only for use in phylogeographic studies using genetic sequences of nuclear (Ornelas-
327   García, Domínguez-Domínguez & Doadrio, 2008) and mitochondrial (Strecker, Faundez &
328   Wilkens, 2004; Ornelas-García, Domínguez-Domínguez & Doadrio, 2008) loci, as well as SNPs
329   (Coghill *et al.*, 2014). Though recent studies based on powerful genomic approaches have provided
330   important insights regarding hybridization between fish lineages (between cave and surface, and
331   among caves), the geographic extension and reticulate evolution of this study model continues to
332   be a challenge in terms of explaining the evolution of the complex *Astyanax* cave-surface
333   populations or lineages (Herman *et al.*, 2018). This confirms the importance of the different efforts
334   and approaches utilized in order to fully understand this model. The present study's extensive
335   sampling throughout the geographic distribution range of the *A. mexicanus* of the Gulf of Mexico
336   slope proved an invaluable advantage. In addition to defining the number and genetic variation of
337   populations within this region, it has revealed the contemporary and historical dynamic of the
338   genetic exchange that determined the populations.
339   This study revealed 10 surface and cave genetic populations or lineages (Fig. 1). According
340   to the F_{ST} values, five levels of genetic differentiation (Table 2) were distinguished along the Gulf
341   of Mexico slope, from the Grijalva-Usumacinta system in southern Mexico, to the Bravo River
342   basin in northern Mexico (Table 2; Table 3; Fig. 1; Fig. 2). These results clearly show that the
343   pattern of population structure obtained is not consistent with the assignment to the three taxa
344   mentioned above (Table 1), and is consistent with the previously stated taxonomic uncertainties
345   (Hausdorf, Wilkens & Strecker, 2011; Herman *et al.,* 2018).
346   At the highest differentiation level (F_{ST}= 0.77 to 0.85), the three cave populations
347   differentiated considerably (Fig. 2A and A´), with the highest probability of assignment based on
348   the STRUCTURE analysis, ML: 99%, PCH: 99% and SAB+TIN: 97%, and no evidence of
349   admixture between them (Fig. 1). This suggests almost complete isolation among cave populations
350   from different geographic areas, such as the case of ML (from Sierra de Guatemala), and PCH,
351   and SAB+TIN (from Sierra de El Abra), as well as among cave populations within the same
352   geographic area (Sierra de El Abra). The pattern of similarity between the SAB and TIN caves
353   could be explained by the physical connectivity between these two, since the SAB+TIN caves
354   form the "*Sistema de los Sabinos*", a cluster comprising the Sabinos, Arroyo, and Tinajas caves,
355   which are connected by sumps (Elliott, 2016). Other studies found a similar differentiation pattern
356   between cave populations; mainly between the regions of El Abra and Sierra de Guatemala
357   (Strecker, Hausdorf & Wilkens, 2012; Bradic *et al.*, 2012), and between caves that are

358   geographically distant from each other within the El Abra region (Bradic *et al.*, 2012). In particular,
359   the structure and differentiation between the PCH and SAB+TIN cave populations are consistent
360   with mitochondrial lineages (*sensu* Ornelas-García, Domínguez-Domínguez & Doadrio, 2008)
361   and SNPs markers (Coghill *et al.*, 2014) that suggest two independent invasions within the El Abra
362   region (see multiple invasions section).
363   The second level of genetic differentiation (F_{ST}=0.30-0.67) also revealed a marked
364   separation between cave and surface populations (Table 2), indicating a strong isolation among
365   the surface and cave populations, although there were two cases of historical gene flow from
366   surface to caves (see below). It is important to point out that the present study did not include any
367   caves with an admixture of both surface and cave morphs, as is the case of the Caballo Moro,
368   Chica, Micos, and Yerbaniz caves where a high frequency of surface fish is observed (Strecker,
369   Hausdorf & Wilkens, 2012; Bradic *et al.,* 2012).
370   The third level of genetic differentiation (F_{ST}= 0.11-0.27) was observed among the
371   following seven genetic clusters of surface populations, revealing a hydrographic divergence
372   pattern: CC (Bravo basin), SF (San Fernando basin), GV+TR (Soto la Marina basin), V (Tuxpan
373   basin), AL+SLP+LCA (Pánuco basin), CAT (Papaloapan basin), and TE+TA+RTZ (Grijalva-
374   Usumacinta basin) (Table 2; Fig. 1). Likewise, this called attention to the fact that the genetic
375   divergence pattern lacks geographic congruence, which was supported by the Mantel test
376   (R^2=0.067, *p*=0.125) in which all 12 surface locations were analyzed. Thus, the pattern of
377   differentiation among surface populations is structured by the hydrographic basins along the
378   Atlantic slope of Mexico. Based on SNPs nuclear loci and practically the same location sampling
379   as used in the present study, Coghill *et al.*, (2014) recovered most of these genetic surface
380   populations as independent lineages of well-structured populations. However, we want to highlight
381   the case of the CAT population that was previously considered to be a different genus
382   (*Bramocharax*). Recent findings confirmed that this population belongs to the genus *Astyanax*
383   (Ornelas-García, Bastir & Doadrio, 2014; Garita-Alvarado, Barluenga & Ornelas-García, 2018;
384   Ornelas-García *et al.*, 2020). Although CAT recorded low gene flow in relation to other locations,
385   the high diversity and the heterogeneous structure pattern (Fig. 1) suggest an ancient gene
386   exchange that preceded the isolation of Catemaco Lake. This could explain the conspicuous
387   morphological differences related to trophic specialization (body shape and the trophic traits of the
388   head, Garita-Alvarado, Barluenga & Ornelas-García, 2018; Ornelas-García *et al.*, 2020) within
389   this population. This is in accordance with the lack of genetic structure between the two
390   differentiated morphs within Catemaco Lake, whose origin had been associated with a process of
391   ecological speciation (Ornelas-García *et al.*, 2020). The heterogeneous structure pattern in addition
392   to low gene flow, both found herein, are consistent with the relevant role of ancient hybridization
393   in the repeated evolution of traits (*sensu* Herman *et al.*, 2018).
394   The fourth level of genetic differentiation (F_{ST}=0.05-0.10) involved the AL, SLP, and LCA
395   locations within the Pánuco basin, and locations from the Grijalva-Usumacinta population, in
396   addition to the CAT population in relation to the Grijalva-Usumacinta population (Table 2; Fig. 1;
397   Fig. 2B and B´). These populations, recognized as two mitochondrially well differentiated lineages

398   that correspond to the *A. mexicanus* and *A. aeneus* lineages (Ornelas-García, Domínguez-
399   Domínguez & Doadrio, 2008; Hausdorf, Wilkens & Strecker, 2011), show high levels of genetic
400   mixing and reciprocal introgression (Fig. 1), highlighting the taxonomic uncertainties between
401   these species. Using different microsatellite loci and locations from the Pánuco and Papaloapan
402   River basins, Hausdorf, Wilkens & Strecker (2011) found a similar introgression pattern,
403   confirming genetic mixing between populations of these two independent basins, despite the
404   physical obstacle presented by the TMVB. The fifth level of genetic differentiation, which
405   presented the lowest and non-significant values (F_{ST}= 0.01-0.02), occurred only among surface
406   populations within the same river basin, particularly in the northern Soto la Marina and southern
407   Grijalva-Usumacinta basins (Table 2; Fig. 1; Fig. 2B and B´).
408

Confirming multiple independent cave invasions

410   Until a few years ago, the widely accepted evolutionary scenario regarding the origin of *A.*
411   *mexicanus* cavefish was that some of these populations were ancient relict populations (i.e.
412   hundreds of thousands of years to millions of years) whose surface relatives were extinct, while
413   other cave populations were considered more recent, and therefore closer to current surface fish
414   populations. This scenario involved two types of cave populations derived from an ancient
415   invasion and a recent invasion (revision in Gross, 2012). However, recently Fumey *et al.* (2018)
416   found that the origin of the blind cave fish is much more recent (<20,000 years), thus refuting the
417   hypothesis of the two ancient invasions. In another study based on another independent approach,
418   the findings also refuted the origin theory of two ancient invasions, although they supported at
419   least four independent origins for the cavefish populations (Coghill *et al.*, 2014). Bradic *et al.*
420   (2012), while recognizing the origin of the cave populations through the two ancient invasions,
421   also found four independent invasion events within the Sierra del Abra (Pachón, Chica cave,
422   Micos, and the six adjacent central caves [Yerbaniz, Japones, Arroyo, Tinajas, Curva, and Toro])
423   which correspond to the oldest invasion wave, plus a fifth invasion within the Sierra de Guatemala.
424   The isolation and differentiation pattern of the cave populations included in the present study, ML,
425   PCH, and SAB+TIN caves (Fig. 1; Fig. 2A and B; Table 2), indicate three independent origins,
426   thus, corroborating those findings pointing to multiple independent subterranean invasions.
427   Greater efforts in field research and modern genomic techniques will surely be required to
428   accurately determine the invasion times of this highly dispersive species into caves.
429   The case of the SAB+TIN cave fish, while they were not separated by the STRUCTURE
430   and DAPC analyses (Fig. 1; Fig. 2A and A´), both presented considerable and significant
431   differentiation coefficients (F_{ST} = 0.27, Table 2). Despite the physical connectivity between these
432   caves, they are not adjacent and are in fact located at different elevations, with the Sabinos cave
433   located at a higher elevation (239 masl) than the Tinajas cave (166 masl) (Elliott, 2016). Such
434   inconsistency in the level of structure (Fig. 1; Fig. 2A and A´) could be the result of two possible
435   scenarios regarding how the SAB+TIN complex was formed. The first (*e1*) considers that a cave
436   population within "*Sistema de los Sabinos*" experienced a partial fragmentation, implying the
437   occurrence of an isolation stage with low connectivity as indicated by the historical gene flow (Fig.

438   4). The second (*e2*) scenario consists of independent isolation events in each of these caves derived
439   from a common surface ancestor, which would imply two independent cave invasions in "*Sistema*
440   *de los Sabinos*", with subsequent secondary contact. These scenarios could both reflect the very
441   dynamic karst system in northeastern Mexico, associated with the formation of new caves, and/or
442   incidents of separation or fusion of these systems (Strecker, Hausdorf & Wilkens, 2012) (see below
443   for more discussion).
444
445   **Dispersal pattern of *Astyanax mexicanus* throughout the Gulf of Mexico slope**
446   In concordance with previous studies (Bradic *et al*., 2012; Fumey *et al*., 2018), findings obtained
447   herein showed very low gene flow in both intra-basin locations and populations, and inter-basin
448   populations. However, there were six particular cases that recorded the highest migration rates,
449   which will be discussed next (Fig 4). Recent findings based on analysis of the whole genome,
450   support the occurrence of considerable historical and contemporary gene flow between cave and
451   surface populations (Herman *et al*., 2018).
452
453   *Migration among surface populations*
454   Although most of the historical migratory events between surface populations, including intra and
455   inter-basin gene flow, recorded lower gene flow (below m=0.005), there were four cases of
456   moderate to high gene flow (Fig. 4). The highest and symmetrical rates were found in northern
457   locations between populations in disjunct basins, the CC population on the Bravo basin and the
458   TR river, a tributary of the Soto la Marina River basin (Fig. 4). Considerable differences in
459   migration rates between these two populations in relation to those observed among other remaining
460   surface populations, suggest the existence of an ancestral connection between the Soto la Marina
461   River and the Cuatro Ciénegas valley. Other species of primary freshwater fish distributed among
462   the Soto la Marina, Cuatro Ciénegas valley, and other independent basins (Minckley, 1984;
463   García-De León *et al.*, 2005) support such an ancestral connection, e.g., *Ictalurus lupus* (Girard
464   1858), *Herichthys cyanoguttatus* (Baird & Girard 1854), and *Micropterus salmoides* (Lacepède
465   1802). There are two likely connections: 1) a hypothetical ancestral corridor through the
466   intermediate headwaters of the San Fernando River Basin, and 2) an ancestral coastal system like
467   the current Laguna Madre, an extensive coastal lagoon into which the Bravo, San Fernando, and
468   Soto la Marina rivers flow (De la Lanza-Espino, Ortiz & Carbajal, 2013). The tolerance to salinity
469   observed in some species of *Astyanax* has allowed them to inhabit coastal lagoons (Wilkens, 1982;
470   Avilés-Torres, Schmitter-Soto & Barrientos-Medina, 2001; Zubiria-Rengifo *et al.*, 2009). This
471   indicates that an ancestral coastal system may have functioned as a corridor that connected only
472   the Bravo and Soto la Marina rivers, thus allowing considerable historical gene flow.
473   The other three gene flow estimates were moderate and asymmetric, and all three included
474   the Pánuco River basin. In two cases the direction of flow was from northern populations of the
475   San Fernando (SF population) and Soto la Marina (TR population) basins towards southern
476   populations (AL, LC and SLP) in the Pánuco basin (Fig. 4). The other unidirectional migration
477   event was in the opposite direction, from the population of the Pánuco basin (AL) to the population

478   of the Bravo basin (CC). The existence of approximately 15 coastal lagoons and the more than 40
479   estuaries that currently lie between the Bravo and the Pánuco river basins suggests that the
480   historical dynamic of these coastal systems, which involves large changes in their geomorphology,
481   caused extreme fluctuations in environmental conditions (De la Lanza-Espino, Ortiz & Carbajal,
482   2013), and likely had an influence on the gene flow between the populations of the northern and
483   southern basins. In an apparently remote scenario of connection, Herman *et al*. (2018) found
484   historical gene flow between two geographically distant populations, the Rascón surface
485   population (from a Pánuco basin) and one population of *A. aeneus s. stricto* from Guerrero, Mexico
486   in the Pacific slope. Further, in the same study genetic exchange was found among the *A. aeneus*
487   population and the Choy surface population (also from the Pánuco basin) and ML cave population.
488   Rather than regarding these two gene flow events as evidence of direct connection between
489   populations, the authors interpreted them as signals of ancient hybridization in the genome. Based
490   on the above, an alternative explanation to the scenarios of the connections discussed above, is
491   that the historical gene flow signal observed in the present study, indicates a past genetic exchange
492   in a common ancestral area where the lineages that preceded the present populations met. Finally,
493   the low estimates of historical gene flow recorded between populations located north and south of
494   "Punta del Morro" confirm that this volcanic mountain is a biogeographic barrier to freshwater
495   fish (Contreras-Balderas, Obregón & Lozano-Vilano, 1996; Husley *et al*., 2004).
496   The significant and unidirectional contemporary gene flow between surface populations
497   observed only between locations in the same basin (within the Soto la Marina, Pánuco, and
498   Grijalva-Usumacinta basins), suggests a continuous corridor between source and target locations.
499

500   ***Migration among cave populations***
501   The historical gene flow found between the adjacent SAB and TIN caves was symmetrical and
502   low, but slightly higher than gene flow among disjunct caves (Fig. 4). The low historical gene flow
503   is related to the moderate and significant F_{ST} value (0.27). Estimates of both the differentiation
504   index and gene flow gave results similar to those recorded among surface populations from
505   different river basins (Table 2; Fig. 4). Thus, although the genetic structure results (Fig. 1; Fig. 2A
506   and 2A´) indicate that the SAB and TIN caves contain a single population, the level of
507   differentiation between these two caves is striking. The Tinajas and Sabinos caves have temporally
508   independent origins; the former is the oldest and most distant cave of the "*Sistema de los Sabinos*"
509   (Elliot, 2016). Based on the above, scenario *e2* is the most plausible, implying another independent
510   cave invasion within the El Abra region, with subsequent secondary contact through the fusion of
511   caves. With respect to contemporary gene flow, rather than being associated with the rise of a
512   barrier between Sabinos and Tinajas, the unidirectionality from TIN to SAB (Table 4) could be
513   associated with a demographic factor, i.e., the Tinajas cave has two large lake passages (Elliott,
514   2016), which potentially house a larger population that directly affects the proportion of migration
515   rate. Regardless of flow direction, it should be noted that since these two cave populations have
516   become connected, they have experienced continuous gene flow over time, increasing towards the
517   present day.

518
519 ***Migration from surface to cave populations***
520 In this case, significant estimates were only obtained for historical gene flow. There were two
521 cases of dispersal from surface to cave populations (Fig. 4). The first was a moderate intra-basin
522 gene flow from the LC and AL corresponding to a surface population to the ML cave population,
523 both located within the Pánuco River (Fig. 4). The second was a higher inter-basin gene flow in
524 the TR location, from the Soto la Marina basin to the PCH cave population in the Pánuco basin
525 (Fig. 4). The first case, which implicates intra-basin gene flow, is highly consistent with findings
526 obtained in Bradic *et al*. (2012), which recorded higher gene flow from surface populations found
527 very close to the cave populations. Furthermore, contemporary and historical gene flow between
528 surface and cave populations recorded recently is associated with the intermediate phenotypes
529 observed in several caves that resulted from the surface introgression into cave populations. This
530 surface introgression is promoted by flooding in caves during the rainy season and is considered
531 to play an important role in a repeated evolutionary adaption in cavefish (Herman *et al*., 2018). In
532 the second case of gene flow between adjacent basins (Soto la Marina and Pánuco), the high
533 estimate (Fig. 4) could indicate an ancient admixture within a proximate area. An alternative
534 explanation could be the presence of an ancient TR surface migrant within the Bravo and Pánuco
535 basins, suggesting that the Soto la Marina basin functioned as a source population that expanded
536 to the northwest and south, as shown by our estimates of gene flow (Fig. 4).
537

Conclusions

539 The results of this study, which used a particular set of 10 microsatellite loci for the genus
540 *Astyanax*, confirmed previous findings in *A. mexicanus* from the Gulf of Mexico slope, obtained
541 with a different set of microsatellite loci as well as other molecular markers and samples, and
542 provided new findings.
543   Ten discrete populations or lineages consisting of seven surface and three cave populations
544 with different levels of differentiation and structure were detected, with the highest levels
545 presented when cave populations were compared to each other, or with surface populations. This
546 pattern confirms that the cave populations, particularly those included in this study such as the
547 Molino cave from Sierra de Guatemala and the Pachón, Sabinos, and Tinajas caves from the El
548 Abra region, correspond to relics of independent invasions. The findings obtained in this study
549 revealed two independent invasions in the "*Sistema de los Sabinos*" cluster of caves.
550   In addition, the present study confirmed a substantial genetic exchange between distinct
551 populations, including surface and cave populations, mainly between hydrographic systems
552 located north of the volcanic mountain range "Punta del Morro". The inter-basin gene flow, at
553 least the flow to the north of the "Punta del Morro", indicates the occurrence of bidirectional
554 invasions, confirming the strength of this physiographic component as a biogeographic barrier.
555 Consistently, the contemporary gene flow only occurred between populations within the same
556 basin.

557 Finally, the heterogeneous structure pattern, in addition to the low gene flow recorded in
558 the Catemaco Lake population, suggested an ancient hybridization that could have given rise to
559 the two divergent lacustrine morphs. Likewise, the history of reticulate evolution within *A.*
560 *mexicanus* by hybridization is confirmed.
561

Acknowledgements

563 RPR would like to thank the Posgrado de la Facultad de Biología, UMSNH for supporting a
564 research stay that was invaluable for conducting this study and for funding the English edition of
565 the manuscript. This study was developed with the support of our own projects (FJGdL and the
566 Conservation Genetics Laboratory) and with the collaboration of several national and international
567 colleagues. We thank the Mexican government for providing the collection permits
568 (DGOPA.05003.181010-5003; DGOPA.00570.288108-0291; DAPA/2/130409/0961, 230401-
569 613-03). We also thank Dean Hendrickson and Gil Rosenthal for providing samples.
570

References

572 Adamack AT, Gruber B. 2014. PopGenReport: simplifying basic population genetic analyses in
573 R. *Methods in Ecology and Evolution* 5:384-387. DOI: https://doi.org/10.1111/2041-
574 210X.12158.
575

576 Agorreta A, Domínguez-Domínguez D, Reina RG, Miranda R, Bermingham E, Doadrio I. 2013.
577 Phylogenetic relationships and biogeography of *Pseudoxiphophorus* (Teleostei:
578 Poeciliidae) based on mitocondrial and nuclear genes. *Molecular Phylogenetics and*
579 *Evolution* 66: 80-90. DOI: http://dx.doi.org/10.1016/j.ympev.2012.09.010.
580

581 Bassam BJ, & Caetano-Anollés G. 1993. Silver staining of DNA in Polyacrylamide gels. *Applied*
582 *Biochemistry and Biotechnology* 42:181-188. DOI: 10.1038/nprot.2007.330.
583

584 Beerli P, Felsenstein J. 2001. Maximum likelihood estimation of a migration matrix and effective
585 population sizes in n subpopulations by using a coalescent approach. *Proceedings of the*
586 *National Academy of Sciences* 98:4563-4568. DOI: 10.1073/pnas.081068098.
587

588 Belkhir K, Borsa P, Chiki L, Raufaste N, Bonhomme F. 1999-2004. GENETIX 4.05, logiciel sous
589 Windows TM pour la génétique des populations. Laboratoire Genome, Populations,
590 Interactions; CNRS UMR 5000, Universitié de Montpellier, Francia.
591 http://www.genetix.univ-montp2.fr/genetix/intro.htm.
592

593 Brookfield JF. 1996. A simple method for estimating null allele frequency from heterozygote
594 deficiency. *Molecular Ecology* 5:453-455. DOI: 10.1046/j.1365-294X.1996.00098.x
595

596 Bradic M, Beerli P, García-de León FJ, Esquivel-Bobadilla S, Borowsky RL. 2012. Gene flow and
597   population structure in the Mexican blind cavefish complex (*Astyanax mexicanus*). *BMC*
598   *Evolutionary Biology* 12:9. DOI: 10.1186/1471-2148-12-9
599

600 Chapuis MP, Estoup A. 2007. Microsatellite null alleles and estimation of population
601   differentiation. *Molecular Biology and Evolution* 24:621-631. DOI:
602   10.1093/molbev/msl191
603

604

605 Coghill LM, Hulsey CD, Chaves-Campos J, García-de León FJ, Johnson SG. 2014. Next
606   generation phylogeography of cave and surface *Astyanax mexicanus*. *Molecular Biology*
607   *and Evolution* 79:368–374. DOI: 10.1016/j.ympev.2014.06.029.
608

609 Contreras S, Obregón H, Lozano ML. 1996. Punta del morro, an interesting barrier for
610   distributional pattern of continental fishes in north and central Veracruz, Mexico. *Acta*
611   *Biológica Venezuelica* 16:37-42.
612

613 Dempster AP, Laird NM, Rubin DB. 1977. Maximum likelihood from incomplete data via the EM
614   algorithm. *Journal of the Royal Statistical Society B* 39:1-38.
615

616 De la Lanza-Espino G, Ortiz-Pérez MA, Carbajal-Pérez JL. 2013. Diferenciación
617   hidrogeomorfológica de los ambientes costeros del Pacífico, del Golfo de México y del
618   Mar Caribe. *Boletín del Instituto de Geografía UNAM* 2013:33-50. DOI:
619   https://doi.org/10.14350/rig.33375.
620

621 Dowling TE, Martasian DP, Jeffery WR. 2002. Evidence for multiple genetic forms with similar
622   eyeless phenotypes in the blind cavefish, *Astyanax mexicanus*. *Molecular Biology and*
623   *Evolution* 19:446-455. DOI: 10.1093/oxfordjournals.molbev.a004100
624

625 Earl D A. 2011. Structure Harvester v0.6.5 Available at
626   http://alumni.soe.ucsc.edu/~dearl/software/struct_harvest/.
627

628 Elliott WR. 2016. Cave exploration and mapping in the Sierra de El Abra region. In: Keene AC,
629   Yoshizawa M, Mc Gaugh SE, eds. *Biology and evolution of the mexican cavefish*.
630   Academic press, 9-40.
631

632 Evanno G, Regnaut S, Goudet J. 2005. Detecting the number of clusters of individuals using
633   software STRUCTURE: a simulation study. *Molecular Ecology* 14, 2611-2620.
634   https://doi.org/10.1111/j.1365-294X.2005.02553.x.
635

636 Excoffier L, Lischer HEL. 2010. Arlequin suite ver 3.5: A new series of programs to perform
637   population genetics analyses under Linux and Windows. *Molecular Ecology Resources* 10:
638   564-567. DOI: 10.1111/j.1755-0998.2010.02847.x.

640 Fumey J, Hinaux H, Noirot C, Thermes C, Rétaux S, Casane D. 2018. Evidence for late Pleistocene
641   origin of *Astyanax mexicanus* cavefish. *BMC Evolutionary biology* 18:43. DOI:
642   10.1186/s12862-018-1156-7 7

644 García-de León FJ, Gutiérrez D, Hendrickson D, Espinosa H. 2005. Fish of the continental waters
645   of Tamaulipas: Diversity and conservation status. In: Carton JL, Ceballos G, Felger RS,
646   eds. *Biodiversity, ecosystem and conservation in Northern Mexico*. New York: Oxford
647   University Press.138-166.
648 Garita-Alvarado CA, Barluenga M, Ornelas-García, CP. 2018. Parallel evolution of morphs of
649   *Astyanax* species (Teleostei: Characidae) in México and Central America. *Biological
650   Journal of the Linnean Society* 124:706-717. DOI: 10.1093/biolinnean/bly082

652 Gross JB. 2012. The complex origin of *Astyanax* cavefish. *BMC Evolutionary Biology* 12:105.
653   DOI: 10.1186/1471-2148-12-105.

655 Hausdorf B, Wilkens H, Strecker U. 2011. Population genetic patterns revealed by microsatellite
656   data challenge the mitochondrial DNA based taxonomy of *Astyanax* in Mexico
657   (Characidae, Teleostei). *Molecular Phylogenetics and Evolution* 60: 89–97. DOI:
658   10.1016/j.ympev.2011.03.009.

660 Herman A, Brandvain Y, Weagley J, Jeffery WR, Keene AC, Kono TJY, Bilandžija H, Borowsky
661   R, Espinasa L, O'Quin K, Ornelas-García CP, Yoshizawa M, Carlson B, Maldonado E,
662   Gross JB, Cartwright RA, Rohner N, Warren WC, McGaugh SE. 2018. The role of gene
663   flow in rapid and repeated evolution of cave- related traits in Mexican tetra, *Astyanax
664   mexicanus*. *Molecular Ecology* 27:4397–4416. DOI: 10.1111/mec.14877.

666 Hulsey CD, García-De León FJ, Sánchez J, Hendrickson DA, TJ Near. 2004. Temporal
667   diversification of Mesoamerican cichlid fishes across a major biogeographic boundary.
668   *Molecular Phylogenetics and Evolution* 31:754-764

670 Jombart T. 2008. Adagent: A R package for the multivariate analysis of genetic markers.
671   *Bioinformatics* 24:1403-1405. DOI: 10.1093/bioinformatics/btn129.

673 Jombart T, Devillard S, Balloux F. 2010. Discriminant analysis of principal components: a new
674   method for the analysis of genetically structured populations. *BMC Genetics* 11: 94. DOI:
675   10.1186/1471-2156-11-94.

676

677 Lara-Lara JR, Arreola-Lizárraga JA, Calderón-Aguilera LE, Camacho-Ibar VF, de la Lanza-
678   Espino G, Escofet-Giansone A, Espejel-Carbajal MI, Guzmán-Arroyo M, Ladah LB,
679   López-Hernández M, Meling-López EA, Moreno-Casasola Barceló P, Reyes-Bonilla H,
680   Ríos-Jara E, Zertuche-González JA. 2008. Los ecosistemas costeros, insulares y
681   epicontinentales. In Capital natural de México, *vol. I: Conocimiento actual de la*
682   *biodiversidad*. CONABIO, México, 109-134.

683

684 Mateos M, Sanjur OI, Vrijenhoek. 2002. Historical biogeography of the livebearing fish genus
685   *Poeciliopsis* (Poeciliidae:Cyprinodontiformes). *Evolution*, 56: 972-984. DOI:
686   https://doi.org/10.1111/j.0014-3820.2002.tb01409.x.

687

688 Miller RR, 1986. Composition and derivation of the freshwater fish fauna of Mexico. *Anales de la*
689   *Escuela Nacional de Ciencias Biológicas* 301: 121-153.

690

691 Minckley WL. 1984. Cuatro Cienegas Fishes: Research Review and a Local Test of Diversity
692   Versus Habitat Size. *Journal of the Arizona-Nevada Academy of Science* 19:13-21.

693

694 Obregon-Barboza H, Contreras-Balderas S, Lozano-Vilano ML. 1994. The fishes of northern and
695   central Veracruz, Mexico. *Hydrobiologia* 286: 79-95.

696

697 Oosterhout CV, Wills DP, Hutchinson WFD. 2004. MICRO-CHECKER: Software for identifying
698   and correcting genotyping errors in microsatellite data. *Molecular Ecology Notes* 4: 535-
699   538. DOI: 10.1111/j.1471-8286.2004.00684.x.

700

701 Ornelas-García CP, Bastir M, Doadrio I. 2014. Morphometric variation between two morphotypes
702   within the *Astyanax* Baird and Girard, 1854 (Actinopterygii: Characidae) genus, from a
703   Mexican tropical lake. *Journal of Morphology* 275: 721–731.

704

705 Ornelas-García CP, Dominguez-Dominguez O, Doadrio I. 2008. Evolutionary history of the fish
706   genus *Astyanax* Baird & Girard (1854) (Actinopterygii, Characidae) in Mesoamerica
707   reveals multiple morphological homoplasies. *BMC Evolutionary Biology* 2008:340. DOI:
708   10.1186/1471-2148-8-340

709

710 Ornelas-García CP, González EG, Tautz D, Doadrio I. 2020. Lack of genetic differentiation
711   between two sympatric species of *Astyanax* (Characidae:Teleostei) in Lake Catemaco,
712   Mexico. *Research Square* in press. DOI: 10.21203/rs.2.22839/v1

713

714 Ornelas-García CP, Pedraza-Lara C. 2016. Phylogeny and evolutionary history of *Astyanax*
715   *mexicanus*. In: Keene AC, Yoshizawa M, Mc Gaugh SE. Eds. *Biology and evolution of the*
716   *mexican cavefish*. Academic press, 77-90.
717

718 Peakall R, Smouse PE. 2012. GenAlEx 6.5: genetic analysis in Excel. Population genetic software
719   for teaching and research-an update. *Bioinformatics* 28:2537-2539.
720   http://bioinformatics.oxfordjournals.org/content/28/19/2537.
721

722 Perdices A, Bermingham E, Montilla A, Doadrio I. 2002. Evolutionary history of the genus
723   *Rhamdia* (Teleostei: Pimelodidae) in Central America, *Molecular Phylogenetics and*
724   *Evolution* 25: 172-189. DOI: https://doi.org/10.1016/S1055-7903(02)00224-5.
725

726 Pitchard JK. 2007. Documentation for STRUCTURE software: Version 2.3.3
727   https://web.stanford.edu/group/pritchardlab/structure_software/release_versions/v2.3.4/ht
728   ml/WhatsNew.html.
729

730 Pritchard JF, Stephens M, Donnelly P. 2000. Inference of populations structure using multilocus
731   genotype data. *Genetics* 155: 945-959.
732

733 Protas EM, Hersey C, Kochanek D, Zhou Y, Wilkens H, Jeffery WR, Zon LI, Borowsky R, Tabin
734   C. 2006. Genetic analysis of cavefish reveals molecular convergence in the evolution of
735   albinism. *Nature Genetics* 38:107-111.
736

737 Quiroz-Martínez B, Álvarez F, Espinosa H, 2014. Concordant biogeographic patterns among
738   multiple taxonomic groups in the Mexican freshwater biota. *Plos One* 9 e105510. DOI:
739   10.1371/journal.pone.0105510.
740

741 R Core Team. 2017. R: A language and environment for statistical computing.
742

743 Raymond M, Rousset F. 2001. GENEPOP (version 4.0) update version described in: populations
744   genetics software for exact test and ecumenicism. *Journal of Heredity* 86: 248-249. DOI:
745   10.1111/j.1471-8286.2007.01931.x
746

747 Rice WR. 1989. Analyzing tables of statistical test. *Evolution* 43: 223-225. DOI:10.1111/j.1558-
748   5646.1989.tb04220.x
749

750 Rosenberg NA. 2004. Distruct: a program for graphical display of population structure. *Molecular*
751   *Ecology Notes* 4:137-138. DOI: 10.1046/j.1471-8286.2003.00566.x
752

753 Schmitter-Soto JJ. 2017. A revision of *Astyanax* (Characiformes: Characidae) in Central and North
754   America, with the description of nine new species. *Journal of Natural History* 51: 1331-
755   1424. DOI: 10.1080/00222933.2017.1324050
756

757 Strecker U. 2003. Polymorphic microsatellites isolated from the cave fish *Astyanax fasciatus*.
758   *Molecular ecology* 3:150-151. DOI: 10.1046/j.1471-8286.2003.00386.x
759

760 Strecker U, Bernatchez L, Wilkens H. 2003 Genetic divergence between cave and surface
761   populations of *Astyanax* in Mexico (Characidae, Teleostei). *Molecular ecology* 12: 699-
762   710. DOI: 10.1046/j.1365-294X.2003.01753.x
763

764 Strecker U, Faundez VH, Wilkens H. 2004. Phylogeography of surface and cave *Astyanax*
765   (Teleostei) from Central and North America based on cytochrome b sequence data.
766   *Molecular Phylogenetics and Evolution* 33:469-481. DOI: 10.1016/j.ympev.2004.07.001
767

768 Strecker U, Hausdorf B, Wilkens H. 2012. Parallel speciation in *Astyanax* cave fish (Teleostei) in
769   Northern Mexico. *Molecular Phylogenetics and Evolution* 62: 62–70. DOI:
770   10.1016/j.ympev.2011.09.005.
771

772 Torres-Paz J, Hyacinthe C, Pierre C, Rétaux S. 2018. Towards an integrated approach to
773   understand Mexican cavefish evolution. *Biology Letters* 14: 20180101. DOI:
774   https://doi.org/10.1098/rsbl.2018.0101.
775

776 Weir BS, Cockerham CC. 1984. Estimation F-statistics for the analysis of population structure.
777   *Evolution* 38:1358-1370. DOI: 10.1111/j.1558-5646.1984.tb05657.x
778

779 Wilkens H, Strecker U. 2017. Convergent evolution of the cavefish *Astyanax* (Characidae,
780   Teleostei): genetic evidence from reduced eye-size and pigmentation. *Biological Journal*
781   *of the Linnean Society* 80: 545-554. DOI: 10.1111/j.1095-8312.2003.00230.x.
782

783 Wilson GA, Rannala B. 2003. Bayesian inference of recent migration rates using multilocus
784   genotypes. *Genetics* 163:177-1191.
785

786 Zubiria-Rengifo W, Lacayo-Morelli A, Acero A, Narvaez-Barandica JC. 2009. Diversidad y
787   abundancia de la ictiofauna de un complejo de lagunas costeras en una reserve natural del
788   Caribe colombiano. *Memoria de la Fundación La Salle de Ciencias Naturales* 170: 125-
789   139.
790

Sampling localities and their respective basins. n = sample size

n = sample size.

Table 1. Sampling localities of *A. mexicanus* and their respective basins. n = sample size. Mitochondrial assignment refers to the presumed species to which the literature assigns those sampling sites, see text for more explanation.

Location	Acronym	Mitochondrial assignment	n	Latitude	Longitude	Basin
Cuatro Ciénegas	CC	*A. mexicanus*	26	26.878	- 102.137	Bravo
San Fernando	SF	*A. mexicanus*	30	24.431	- 98.431	San Fernando
Garza Valdez	GV	*A. mexicanus*	24	24.711	- 99.384	Soto la Marina
Troncones	TR	*A. mexicanus*	40	23.723	- 99.308	Soto la Marina
Arroyo Lagartos	AL	*A. mexicanus*	36	22.804	- 98.941	Pánuco
Molino Cave	ML	*A. mexicanus*	15	23.031	- 99.150	Pánuco
Pachón Cave	PCH	*A. mexicanus*	44	22.607	- 99.044	Pánuco
Sabinos Cave	SAB	*A. hubbsi*	23	22.024	- 98.9	Pánuco
Tinajas Cave	TIN	*A. hubbsi*	12	21.968	- 98.902	Pánuco
La Cañada	LCA	*A. mexicanus*	26	21.867	- 99.151	Pánuco
Huichihuayan	SLP	*A. mexicanus*	19	21.479	- 98.966	Pánuco
El Zapotal	V	*A. aeneus*	30	21.276	- 98.146	Túxpan
						******
Catemaco	CAT	*A. aeneus**	26	18.433	- 95.032	Papaloapan
Teapa	TE	*A. aeneus*	50	17.426	- 92.751	Grijalva-Usumacinta
Tapijulapa	TA	*A. aeneus*	27	17.45	- 92.75	Grijalva-Usumacinta
Rio Tzendales	RTZ	*A. aeneus*	41	16.216	- 90.841	Grijalva-Usumacinta

2   * Previously considered *Bramocharax caballeroi*

3   ****** Punta del Morro biogeography barrier

Pairwise F$_{ST}$ between *Astyanax* spp. locations. Values with asterisk are non-significant

Values in bold and italic font are the highest; values in bold only are the second highest;

black normal font values are moderate and gray normal font values are low but significant.

For definition of the acronyms, see Table 1.

1   Table 2. Pairwise F_{ST} between *Astyanax* locations.

	CC	SF	GV	TR	AL	ML	PCH	SAB	TIN	LCA	SLP	V	CAT	TE	TA	RTZ
CC		0.27	0.25	0.25	0.21	0.59	0.66	0.53	0.49	0.23	0.23	0.29	0.24	0.16	0.18	0.18
SF	0.27		0.19	0.18	0.21	0.56	0.63	0.49	0.42	0.20	0.19	0.24	0.20	0.16	0.17	0.18
GV	0.26	0.20		0.01	0.14	0.45	0.57	0.44	0.38	0.17	0.12	0.18	0.09	0.13	0.12	0.15
TR	0.26	0.18	0.01*		0.14	0.42	0.52	0.40	0.35	0.15	0.12	0.17	0.10	0.13	0.13	0.15
AL	0.21	0.21	0.14	0.13		0.43	0.45	0.33	0.27	0.05	0.05	0.10	0.12	0.07	0.08	0.09
ML	**0.58**	**0.57**	**0.44**	**0.41**	**0.4**		0.85	0.81	0.81	0.47	0.49	0.50	0.43	0.40	0.45	0.43
PCH	**0.67**	**0.64**	**0.57**	**0.52**	**0.45**	*0.85*		0.75	0.76	0.50	0.54	0.52	0.53	0.44	0.52	0.48
SAB	**0.55**	**0.51**	**0.45**	**0.41**	**0.34**	*0.82*	*0.77*		0.27	0.38	0.37	0.38	0.39	0.34	0.37	0.36
TIN	**0.50**	**0.44**	**0.39**	**0.36**	**0.27**	*0.8*	*0.77*	**0.27**		0.30	0.29	0.30	0.33	0.29	0.31	0.30
LCA	0.24	0.20	0.17	0.15	0.05	**0.47**	**0.51**	**0.4**	**0.31**		0.07	0.11	0.11	0.06	0.07	0.06
SLP	0.23	0.19	0.12	0.12	0.05	**0.48**	**0.54**	**0.39**	**0.3**	0.07		0.09	0.10	0.06	0.07	0.07
V	0.30	0.25	0.19	0.18	0.11	**0.49**	**0.53**	**0.4**	**0.31**	0.12	0.1		0.14	0.11	0.11	0.12
CAT	0.24	0.20	0.09	0.1	0.12	**0.42**	**0.53**	**0.4**	**0.33**	0.12	0.1	0.14		0.07	0.07	0.08
TE	0.16	0.16	0.13	0.13	0.07	**0.39**	**0.44**	**0.35**	**0.3**	0.05	0.06	0.11	0.07		0.01	0.01
TA	0.18	0.17	0.13	0.14	0.08	**0.45**	**0.52**	**0.38**	**0.32**	0.08	0.07	0.11	0.07	0.01*		0.02
RTZ	0.19	0.18	0.15	0.15	0.09	**0.43**	**0.49**	**0.37**	**0.31**	0.06	0.08	0.13	0.08	0.01*	0.02*	

2   * non-significant values. Values in italic font are the highest; values in bold only are the second highest; black normal font values are

3   moderate and gray normal font values are low but significant. For definition of the acronyms, see Table 1. Diagonal upper values

4   correspond to F_{ST} values estimated by ENA algorithm as implemented in FreeNA.

Summary of migration analysis (only 95% confidence intervals that did not overlap with 0 are shown)

Median values (with their 0.025 and 0.975 posterior distribution values as 95% confidence interval estimates) of M= migration rate and ɵ= size theta. Values in bold correspond to higher M values. For definition of the acronyms, see Table 1.

1   Table 3. Four distinct AMOVA models assessed.

Tested model	Source of variation	Variance	Percentage of the total (%)	Statistics F	*P* Value
A). 16 groups	Between groups	1.16376	26.55		
				$F_{ST} = 0.2654$	**0.00**
	Within populations	3.21974	73.45		
B). 5 groups	Between groups	0.29026	6.47	$F_{CT} = 0.0647$	0.13
	Within groups	0.97395	21.72	$F_{SC} = 0.2322$	**0.00**
	Within populations	3.21974	71.81	$F_{ST} = 0.2819$	**0.00**
C). 7 groups	Between groups	0.12916	2.93	$F_{CT} = 0.0293$	0.30
	Within groups	1.05758	24.00	$F_{SC} = 0.2472$	**0.00**
	Within populations	3.21974	73.07	$F_{ST} = 0.2693$	**0.00**
D). 10 groups	Between groups	1.09414	24.49	$F_{CT} = 0.2449$	**0.00**
	Within groups	0.15367	3.44	$F_{SC} = 0.0455$	**0.00**
	Within populations	3.21974	72.02	$F_{ST} = 0.2793$	**0.00**

4   A) Each collection location considered as a different population (CC, SF, GV, TR, AL, ML, PCH, SAB, TIN, LCA,

5   SLP, V, CAT, TE, TA and RTZ).

6   B) Localities were grouped by phylogenetic groups based on the results obtained from Ornelas-García-García et al.,

7   (2008) [(CC), (SF, GV, TR, AL, ML, PCH, LCA, SLP, V), (SAB, TIN), (CAT), (TE, TA, RTZ)].

8   C) Localities were grouped by hydrological basins [Bravo (CC), San Fernando (SF), Soto la Marina (GV, TR), Pánuco

9   (AL, ML, PCH, SAB, TIN, LCA, SLP), Tuxpan (V), Papaloapan (CAT), Grijalva-Usumacinta (TE, TA, RTZ)].

10   D) Localities were grouped according to their degree of genetic differentiation as measured by the F_{ST}, grouping

11   those populations that presented non-significant differentiation levels [(CC), (SF), (GV, TR), (AL, LCA, SLP), (V),

12   (PCH), (ML), (SAB, TIN), (CAT), (TE, TA, RTZ)].

Summary of BAYESASS analysis

Considerably higher mean values of migration rates (m) with their 95% confidence intervals are shown. m= migration rate, CI= confidence interval. For definition of the acronyms, see Table 1.

1 Table 4. Summary of BAYESASS analysis. Considerably higher mean values of migration rates

2 (m) with their 95% confidence intervals are shown. m= migration rate, CI= confidence interval.

3 For definition of the acronyms, see Table 1.

Source site	Target site	m	Lower 95% CI	Upper 95% CI
GV	TRO	0.2057	0.1664	0.2450
TIN	SAB	0.1552	0.0977	0.2127
LCA	AL	0.2083	0.1685	0.2481
SLP	AL	0.1807	0.1360	0.2254
RTZ	TE	0.2453	0.2737	0.2169
TA	TE	0.2178	0.2532	0.1824

6
7
8

Figure 1

Geographic distribution of the genetic variation in populations and locations of *A. mexicanus* sampling in this study along the Gulf of Mexico slope.

A) corresponds to the clustering analysis performed in STRUCTURE, using 10 microsatellite loci of *A. mexicanus* sampled; the phylogenetic classification *sensu* Ornelas et al. (2008) and resulting populations were matched. B) Map of distribution range sampled in present study; colored pies charting on the map corresponds to the same genotypic clusters of STRUCTURE. For definition of the acronyms see Table 1.
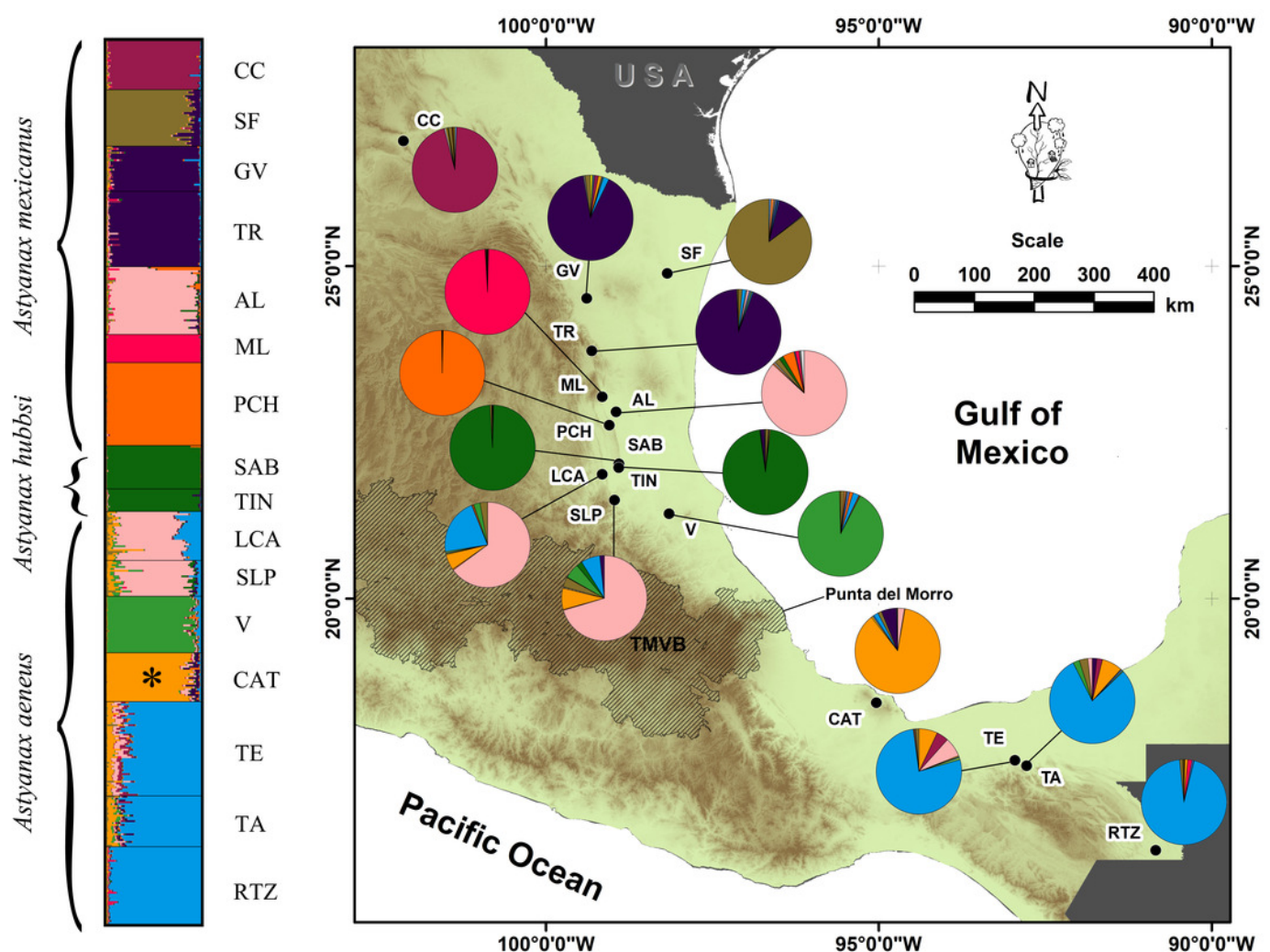
Figure 2

Discriminant Analysis of Principal Components (DAPC).

A= analysis including all 16 locations; B= analysis of 12 populations (excluding the cave populations). the phylogenetic classification *sensu* Ornelas et al. (2008) and resulting populations were matched. For definition of the acronyms see Table 1.
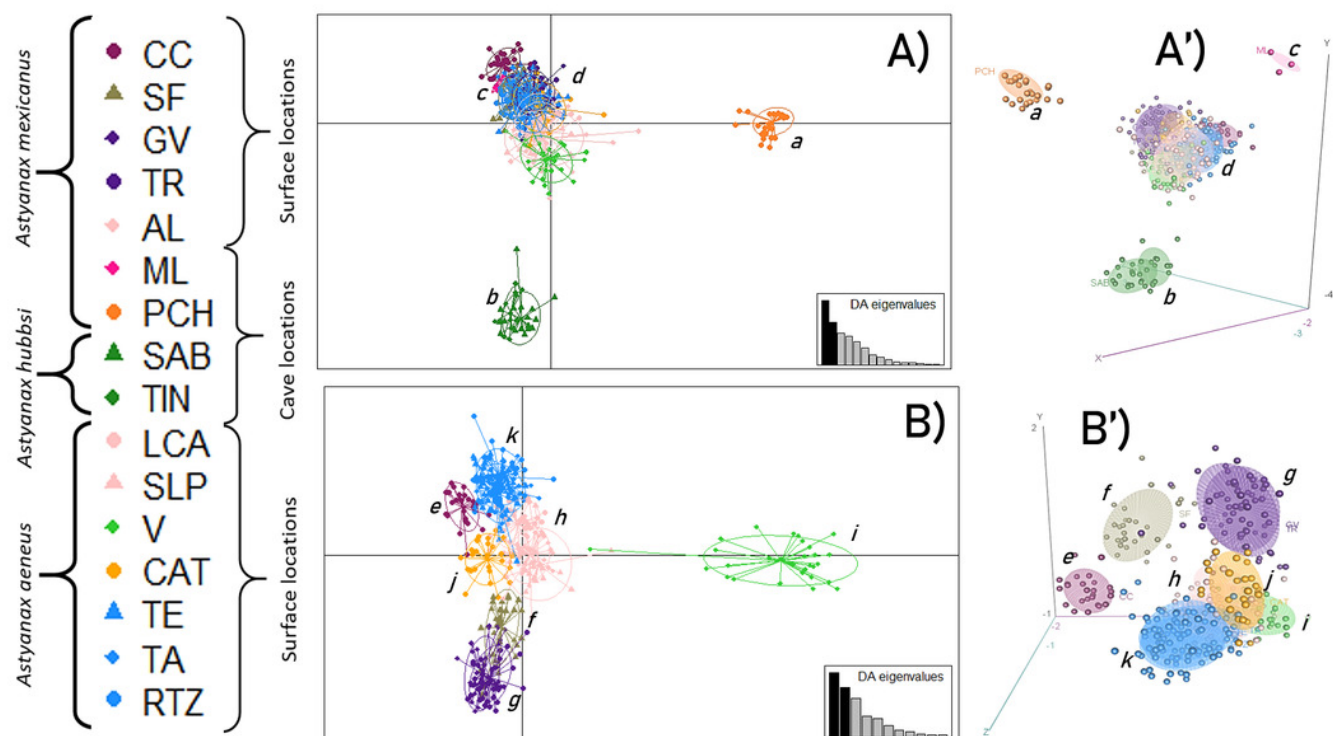
Figure 3

Effective population size of 16 locations sampled based on Bayesian inferences and obtained by the product with the mutation rate = 5 x 10^{-4} (Fumey *et al*. 2018).

For definition of the acronyms see Table 1. The central box of the plots represents the values from the lower to upper quartile (25 to 75 percentile). The middle line represents the median posterior values over all loci. The horizontal line extends from the 2.5% percentile to the 97.5 percentile.
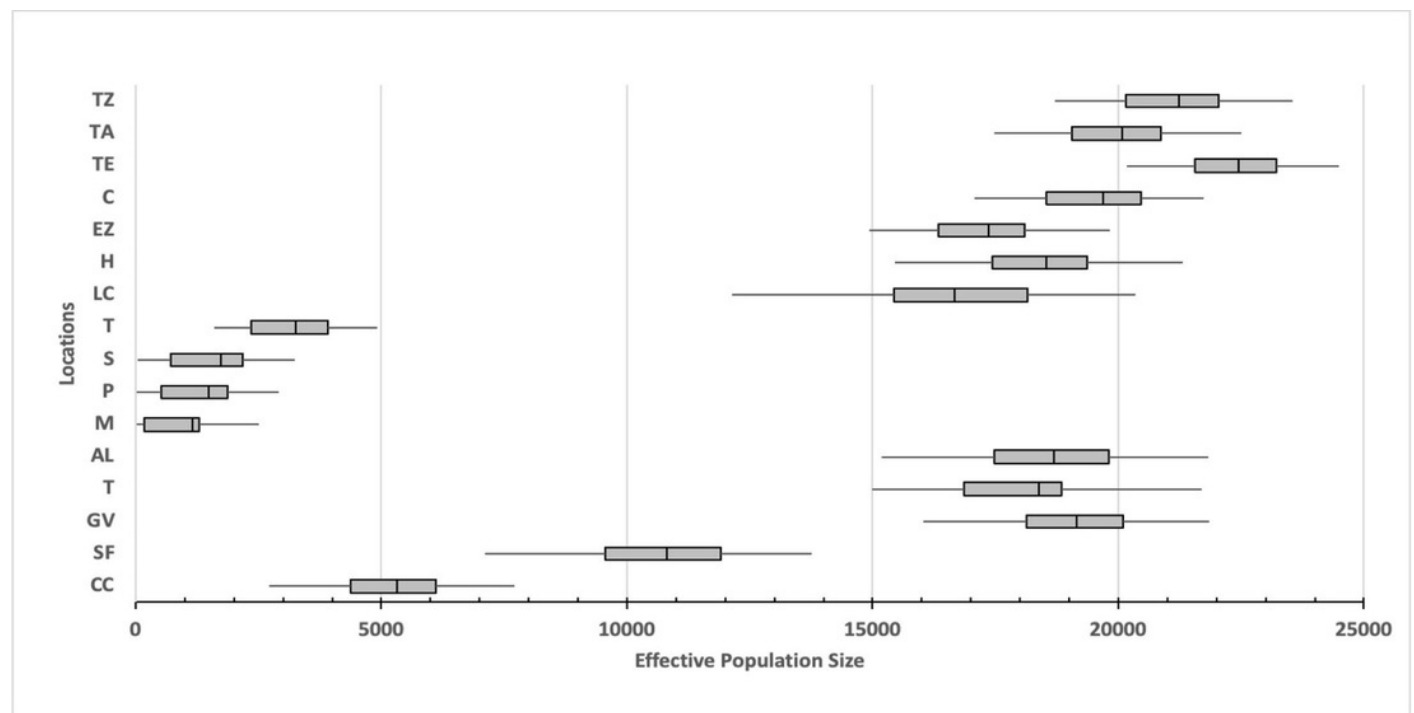
Figure 4

Migration rate (m) based on Bayesian inference among distinct locations of *A. mexicanus* and obtained by the product with the mutation rate = 5 x 10^{-4} (Fumey *et al*. 2018).

For definition of the acronyms see Table 1. The central box of the plots represents the values from the lower to upper quartile (25 to 75 percentile). The middle black line represents the median posterior values over all loci. The horizontal line extends from the 2.5% percentile