A peer-reviewed version of this preprint was published in PeerJ on 29 April 2019.

Ballesteros-Nova NE, Pérez-Rodríguez R, Beltrán-López RG, Domínguez-Domínguez O. 2019. Genetic differentiation in the southern population of the Fathead Minnow *Pimephales promelas* Rafinesque (Actinopterygii: Cyprinidae) PeerJ 7:e6224 https://doi.org/10.7717/peerj.6224

Genetic differentiation of the southern population of the fathead minnow *Pimephales promelas* Rafinesque (Actinopterygii: Cyprinidae)

Nayarit E Ballesteros-Nova [1] , **Rodolfo Pérez-Rodríguez** [2] , **Rosa G Beltrán-López** [1] , **Omar Domínguez-Domínguez** [2]
Corresp.

[1] Programa Institucional de Doctorado en Ciencias Biológicas, Facultad de Biología, Universidad Michoacana de San Nicolás de Hidalgo, Morelia, Michoacán, México
[2] Laboratorio de Biología Acuática, Facultad de Biología, Universidad Michoacana de San Nicolás de Hidalgo, Morelia, Michoacán, México

Corresponding Author: Omar Domínguez-Domínguez
Email address: goodeido@yahoo.com.mx

Background. Mexico is one of the most megadiverse countries in the world, with considerable reaches and endemicity in the diversity of fishes compared to other countries and regions. Recent phylogenetic studies in co-distributed species of widespread fishes, besides revealing a subestimation of species richness in the Mesa del Norte, in Mexico, suggest phylogenetic congruence in some species complexes. Previous morphological and meristic analyses concluded that at least three subspecies of *Pimephales promelas* exist in United States populations, suggesting that the richness in *Pimephales* could be underestimated. But no studies have examined the morphologic and genetic diversity in Southern populations of *Pimephales promelas*. We presented analyses of the genetic variation among *P. promelas* populations across its Southern distributional range.

Methods. Phylogenetic reconstruction and genetic distances using cytochrome *b* and *S7* sequences were done.

Results. The results based on phylogenetic trees, species tree, genetic distances and haplotype networks revealed the existence of at least four well-differentiated lineages (Yaqui Lineage, Nazas+Conchos Lineage, Santa Maria Lineage and Casas Grandes Lineage).

Discussion. The four well supported lineages found confirm *Pimephales promelas* as a species complex. Composition and distribution of these major lineages is also consistent with previous biogeographic hypothesis for other fishes in the region, supporting the fragmentation of the ancestral Lake Cabeza de Vaca, possibly due to the combined influence of tectonic events and increasing regional aridity, as well as events of interchange between basins via stream capture.

Genetic Differentiation of the Southern Population of the Fathead Minnow *Pimephales promelas* Rafinesque (Actinopterygii: Cyprinidae)

Nayarit Emérita Ballesteros-Nova[1], Rodolfo Pérez-Rodríguez[2], Rosa Gabriela Beltrán-López[1], Omar Domínguez-Domínguez[2]

[1]Programa Institucional de Doctorado en Ciencias Biológicas, Facultad de Biología, Universidad Michoacana de San Nicolás de Hidalgo, Morelia, Michoacán, México

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

Corresponding Author:

Omar Domínguez-Domínguez[2]

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

Email address: goodeido@yahoo.com.mx

Abstract

Background. Mexico is one of the most megadiverse countries in the world, with considerable reaches and endemicity in the diversity of fishes compared to other countries and regions. Recent phylogenetic studies in co-distributed species of widespread fishes, besides revealing a subestimation of species richness in the Mesa del Norte, in Mexico, suggest phylogenetic congruence in some species complexes. Previous morphological and meristic analyses concluded that at least three subspecies of *Pimephales promelas* exist in United States populations, suggesting that the richness in *Pimephales* could be underestimated. But no studies have examined the morphologic and genetic diversity in Southern populations of *Pimephales promelas*. We presented analyses of the genetic variation among *P. promelas* populations across its Southern distributional range.

Methods. Phylogenetic reconstruction and genetic distances using cytochrome *b* and *S7* sequences were done.

Results. The results based on phylogenetic trees, species tree, genetic distances and haplotype networks revealed the existence of at least four well-differentiated lineages (Yaqui Lineage, Nazas+Conchos Lineage, Santa Maria Lineage and Casas Grandes Lineage).

Discussion. The four well supported lineages found confirm *Pimephales promelas* as a species complex. Composition and distribution of these major lineages is also consistent with previous biogeographic hypothesis for other fishes in the region, supporting the fragmentation of the ancestral Lake Cabeza de Vaca, possibly due to the combined influence of tectonic events and increasing regional aridity, as well as events of interchange between basins via stream capture.

Keywords. Mesa del Norte, Mexico, Genetic lineages.

Introduction

The biogeographic explanation of a species with a widespread distribution pattern is mainly associated with a suitable biological and ecological trait, that allows high dispersal ability and better possibilities to colonize a new habitat (Nathan, 2001). Particularly, for the case of freshwater fishes, besides of the organism´s ability for moving away from their parent source, the continuity of the aquatic habitats works as the main factor that determines their distributional range (Pérez-Rodríguez et al., 2015). Thus, the widespread distributional pattern in freshwater fishes implies at least two elements: 1) the organism´s possession of suitable biological traits to solve the environmental conditions involved in the dispersion process, and 2) an opening route for dispersal between water bodies, which serve as connections between hydrographic systems.

The Fathead Minnow, *Pimephales promelas*, is a widespread cyprinid in Eastern North America, ranging from the Lake Slave and the Hudson Bay, with East Canada as its Northern limit, southward through the Mississippi Valley, the Great Plains and the Gulf slope streams of Alabama and the Grande River Basin, including the Conchos River; it also includes the endorheic basins of Casas Grandes, Nazas, Del Carmen, Santa Maria and Bustillos, and the Pacific slope drainage of the Yaqui River, in Mexico (Vandermeer, 1966; Miller, Minkley & Norris, 2005). The wide distribution pattern of *P. promelas* is associated with the ability to survive and reproduce in different environmental conditions across different regions (Scott & Crossman, 1973). Nevertheless, based on the idea that larger areas present a higher probability to be dissected by geographical barriers, and consequently widely distributed species are more likely to speciate (Letsch, Gottsberger & Ware, 2016), the occurrence of isolation events through the larger range of *P. promelas* is very likely. Even a previous hypothesis of morphological

variation recognizes several subspecies between Northern, Southern, and Eastern populations in the United States (Vandermeer, 1966).

Isolation events due to geographical barriers have been found in other fish species that inhabit the dry region of Northern Mexico, as the case of *Cyprinella formosa* (Wood & Mayden, 2002), *Cyprinella lutrensis* (Schönhuth & Mayden, 2010), *Campostoma anomallum* (Blum et al., 2008), *Campostoma ornatum* (Domínguez-Domínguez et al., 2011; Schönhuth et al., 2011), *Dionda episcopa* (Schönhuth et al., 2012), *Gila* spp. (Schönhuth et al., 2014), *Codoma ornata* (Schönhuth et al., 2015) and *Pantosteus plebeius* (Corona-Santiago et al., 2018, in press). In this case, phylogeographic studies have provided useful information that complements knowledge of biogeographic hypothesis of areas and species, and current distributions of fish species and their link to fragmentation and or expansion events (Devitt, 2006; Riddle & Hafner 2006; Vallinoto et al., 2009; Schönhuth et al., 2015). In North America, tectonic and climatic factors are the main factors that contribute to the evolution of drainages and species (Galloway, Whiteaker & Ganey-Curry, 2011; Aranda-Gómez et al., 2018). In Northern Mexico, the main tectovolcanic activity since the Oligocene is related to the formation of the Sierra Madre Occidental (SMO) mountain range (Ferrari, Valencia-Moreno & Bryan, 2007; Aguirre-Díaz et al., 2008). This high tectonic activity has been proposed as responsible of the high endemism of flora and fauna in this region of the country, and includes the main forces that shape the speciation process (vicariance and or expansion) in freshwater fishes (Echelle et al., 2005; Hughes, Rinne & Calamusso, 2005; Ceballos, Arroyo-Cabrales & Ponce, 2010; Morrone, 2010; Schönhuth & Mayden, 2010; Domínguez-Domínguez et al., 2011; Clements, Bart & Hurley, 2012; Schönhuth et al., 2012, 2014, 2015). The SMO is also considered an important biogeographic corridor and a Pleistocenic refuge (Hughes, Rinne & Calamusso, 2005), in which the expansion and contraction in the

91 distribution of numerous species have been associated with climate change during the
92 Pleistocene. The effects of climate changes on epicontinental waters should have significantly
93 affected the distribution of fish fauna of the SMO (Domínguez-Domínguez et al., 2011).

94 Previous studies based on molecular data of fishes (Smith & Miller, 1986; Mayden,
95 Matson & Hillis, 1992; Mayden et al., 1992; Echelle et al., 2005; Domínguez-Domínguez et al.,
96 2011; Schönhuth et al., 2011, 2012, 2015) and geological information (Galloway, Whiteaker &
97 Ganey-Curry, 2011), proposed the existence of an extended ancestral drainage system in the
98 Mesa del Norte in Mexico, suggesting the existence of an extended Rio Grande system, that
99 persisted for more than 10 Ma and extended from the South-Western highlands of the USA
100 (Colorado Plateau, in current Utah, Colorado and Arizona) across the current Chihuahuan Desert
101 (New Mexico, Chihuahua, Coahuila and Durango) to a confluence with the Lower Rio Grande
102 Basin (Galloway, Whiteaker & Ganey-Curry, 2011; Schönhuth et al., 2015). This paleo-river
103 system was proposed to have originated in the Oligocene at the same time that the SMO grew in
104 volume and area (Schönhuth et al., 2015), connecting current independent drainages in Mexico
105 (Mayden, Matson & Hillis, 1992; Echelle et al., 2005; Schönhuth et al., 2006, 2011, 2015;
106 Domínguez-Domínguez et al., 2011), and acting as a hydrological corridor for the ancestral
107 forms of current freshwater fish species, allowing their radiation/expansion across the drainages
108 in the Chihuahuan Desert (Smith & Miller, 1986; Mayden, Matson & Hillis, 1992; Schönhuth et
109 al., 2011, 2015; Domínguez-Domínguez et al., 2011). It includes possible transfers and/or
110 integration of headwaters of the Mesa del Norte with drainages from the Pacific slope (Minckley,
111 Hendrickson & Bond, 1986; Miller, Minkley & Norris, 2005; Schönhuth et al., 2015), and as far
112 South as the Nazas and Aguanaval rivers (Smith & Miller, 1986; Mayden, Matson & Hillis,
113 1992; Schönhuth et al., 2015).

Tecto-volcanic events since the Miocene, Pleistocenic glacial/interglacial cycles and the increasing regional aridity since Holocene time, gave rise to the fragmentation of the ancestral system (Metcalfe, 2006; Schönhuth et al., 2015), effectively isolating different populations. Smith & Miller (1986) further suggested that rivers in western Mexico draining to the Pacific (Upper Yaqui and Upper Mezquital) across desert regions originated through headwater capture from the ancestral and the extant Grande River system, and shared distributions by stream capture events (Schönhuth et al., 2011). Adittional to this, a more recent paleohydrological hypothesis was proposed by Smith & Miller (1986), related to the fragmentation of the Lake Cabeza de Vaca in the Pleistocene time and the formation of the pluvial Lake Palomas (Rosenthal & Forstner, 2004; Wood & Mayden, 2002; Corona-Santiago et al., 2018, in press), a relict of the Lake Cabeza de Vaca that was fragmented into several endorheic rivers, lagoons, and springs (Corona-Santiago et al., 2018, in press).

A previous morphological analysis of *Pimephales promelas* populations in the United States, concluded that there exist at least three subespecies (Vandermeer, 1966), suggesting that the species richness in *P. promelas* could be underestimated. No studies have examined the genetic diversity in *Pimephales promelas* to test this hypothesis and identify independent lineages, in the United States and Southern areas of distribution. The aim of the present study is to assess genetic divergences using mitochondrial and nuclear markers in populations of *P. promelas* through its distributional range in Northwestern Mexico, in order to infer the phylogenetic relationships among Southern populations of *P. promelas* and suggest its biogeographic history.

137 Methods

138 **Taxon sampling**

139 Specimens from five independent basins and ten localities across most of the Southern
140 distributional range of *P. promelas* were collected with electrofishing and seine nets (Fig. 1;
141 Table S1). A piece of fin was fixed in 95% ethanol for DNA extraction and stored at -70°C. A
142 maximum of five specimens per site were preserved in 5% formalin following the protocols
143 approved by the Ministry of Environment and Natural Resources of the Mexican Government.
144 Fish and tissue samples were deposited in the fish collection of the Universidad Michoacana de
145 San Nicolas de Hidalgo, Mexico (SEMARNAT registration number MICH-PEC-227-07-09). All
146 procedures performed were reviewed and approved by a committee of Mexican Ministry of
147 Environmental and Natural Resources (SEMARNAT), collection permit number PPF/DGOPA-
148 262/17. In addition, sequences of *P. promelas* were obtained from the GenBank dataset (Table
149 S1). Sequences of *Pimephales notatus*, *Pimephales tenellus* and *Codoma ornata* were retrieved
150 from the GenBank dataset and used as outgroups (Table S1).

151 **DNA extraction, PCR amplification and sequencing**

152 Total DNA was extracted with the standard Proteinase K/phenol/chloroform protocol (Sambrook
153 et al., 1989). Sequences for a fragment of the mitochondrial cytochrome *b* gene (cyt *b*: 1049 bp)
154 for twenty-four specimens were obtained using the primers LA (5´-
155 GTGACTTGAAAAACCACCGTTG) and HA (3´-CAACGATCTCCGGTTTACAAGAC)
156 (Dowling et al., 2002).

157 A subset of eighteen specimens accounting for all the variation found for the cyt *b* gene
158 were selected for the amplification of the first intron of the gene coding for the *S7* ribosomal

protein (*S7*: 704 bp) with the primers S71F (5′-TGGCCTCTTCCTTGGCCGTC) and S72R (3′-AACTCGTCTGGCTTTTCGCC) (Chow & Hazama, 1998). The final concentrations for the Polymerase Chain Reaction (PCR) per 25 µL reaction were as follows: 50 ng template DNA, 10 µM of each primer, 0.7 units of Taq DNA polymerase, 0.25 mM of each dNTP, 2.5 µL of 10X reaction buffer and 2.5 mM $MgCl_2$. Thermocycling conditions for the amplification of the mitochondrial cyt *b* gene consisted of an initial denaturalization step of 3 min at 94 °C, followed by 35 cycles of 30 s at 94°C, 1 min at 48°C, 1:30 min at 72°C, and a final 5 min extension step at 72°C. The *S7* gene was amplified under the following conditions: initial denaturalization step of 3 min at 94 °C, followed by 35 cycles of 45 s at 94°C, 50 s at 57°C, 1.40 min at 72°C, and a final step of 10 min extension at 72°C. All PCR products were purified with ExoSAP-IT™. Purified PCR products were sent to Macrogen Korea for sequencing.

Sequences were edited and aligned using the default parameters of Clustal X (Thompson et al., 1997) implemented in Mega v6.06 (Tamura et al., 2013), and examined with chromatograms. For *S7*, sequences with point mutation were phased using DNAsp v5.10 (Librado & Rozas, 2009) and a test of recombination was applied using a coalescent algorithm (10000 replicates). The sequences of *S7* showed heterozygous indels, in this case a manual reconstruction of the two allele phases was performed following the procedure described by Sousa-Santos et al. (2005).

Phylogenetic analyses

The evolutionary substitution model based on the Akaike Information Criterion (AIC) and an optimal partition-setting was obtained using PartitionFinder v1.1.0 (Lanfear et al., 2012); the optimal partition setting was obtained by assigning a substitution model to each gene. The models obtained were the Transitional Model (TIM3) (Posada, 2003) + gamma (TIM3+G) for

the cyt *b* gene and the Tamura-Nei (Tamura & Nei, 1993) + gamma (TrN+G) for the individual *S7* gene.

Bayesian Inference analyses were implemented in MrBayes v3.2.1 (Ronquist et al., 2012). Sequences were analysed in two different data sets, one for each gene independently and one for the two genes concatenated. We used the two heterozygous alleles. The substitution models were set according to the selected model for each gene by PartitionFinder v1.1.0 (Lanfear et al., 2012). The analyses were run for 10 million generations, with two independent runs implementing four Markov Chain Monte Carlo (MCMC) processes and sampling every 500 generations. We evaluated the chains convergence with the log-likelihood (-InL) values of the two independent runs on Tracer v1.5 (Rambaut & Drummond, 2007), discarding 10% of generations as burn-in to construct the consensus tree. The trees were visualized in FigTree v1.4.2 (Rambaut, 2014).

Maximum Likelihood trees were constructed in RAxMLGUI v1.3.1 (Silvestro & Michalak, 2012; Stamatakis, 2014), as implemented in CIPRES (Miller, Pfeiffer & Schwartz, 2010), with the default GTR+G model and the fast bootstrap option (CIPRES portal v3.3) at the San Diego Supercomputer Center at http://www.phylo.org/ sub_sections/portal/.

Species Tree Analysis and Divergence Times

We estimated a species tree and divergence times for major nodes in *P. promelas* for the two genes (cyt *b* + *S7*) using the Bayesian Method in *BEAST v1.8.1 (Drummond et al., 2012), and the substitution models were set according to the selected model for each gene by PartitionFinder v1.1.0 (Lanfear et al., 2012). This analysis was carried out with a subset of thirty-six *P. promelas* sequences and two outgroup sequences. Assumptions included a calibrated molecular clock using the mutation rate of cyt *b* in teleosts of 0.76-2.2%/Ma (Zardoya & Doadrio, 1999;

Berendzen, Gamble & Simons, 2008; Nagle & Simons, 2012). We estimated the evolutionary rate of the *S7* gene relative to the cyt *b* gene. A lognormal relaxed clock (Uncorrelated) model on branch length (Drummond et al., 2006) was used. We performed a MCMC analysis with 200 million of generations and sampled every 1000 generations, with a Yule process speciation prior. Analyses were run in the CIPRES Science Gateway v3.3 (http://www.phylo.org/sub_sections/portal/). Convergence was assessed with ESSs in TRACER v1.5 (Rambaut & Drummond, 2007). Ten percent of the generations were discarded (burn-in) using the Tree Anotator v1.8.1 (Drummond et al., 2012). The tree was visualized in FigTree v1.4.2 (Rambaut, 2014).

Genetic distances and haplotype networks

The uncorrected genetic distances were calculated among the recovered monophyletic groups in the phylogenetic trees for both genes independently in Mega v6.06 (Tamura et al., 2013).

Unrooted networks were constructed under the null hypotheses of no genetic differentiation among populations, for each gene, using the median-joining method (Bandelt, Forster & Röhl, 1999) as implemented in PopART v1.7 (available at http://popart.otago.ac.nz).

Results

Sixty sequences were obtained for both genes: 24 for cyt *b*, and 36 for *S7,* which included the two heterozygous alleles. No significant recombination was detected in the nuclear *S7* sequences ($p > 1$). In addition, four mitochondrial sequences for cyt *b* and three nuclear gene sequences for *S7* retrieved from the GenBank were included in the analyses (Table S1).

Sixteen haplotypes were obtained from the 24 sequences of the cyt *b* gene. These haplotypes were defined by 213 polymorphic sites within a 1049 bp sequence fragment (total

number of mutations = 240). Sixty-nine of those sites were singletons, and 164 substitutions were parsimony informative.

Eighteen haplotypes were obtained for the *S7* gene; these haplotypes were defined by 43 polymorphic sites (total number of mutations = 45). Twenty-five of those sites were singletons, and 18 substitutions were parsimony informative.

Phylogenetic relationships

Maximum Likelihood and Bayesian Inference showed the same topology for cyt *b* and concatenated analyses, unlike the mitochondrial and concatenated datasets, the nuclear marker (Fig. S2) recovered the four differentiated groups but with different phylogenetic relationships.

The phylogenetic results for cyt *b* and for the two genes (cyt *b* + *S7*) show the formation of two well-differentiated and supported clades (Clade A and Clade B) based on Maximum Likelihood (Bootstrap (BS): cyt *b*, BS=86%; cyt *b*+*S7*, BS=88%) and Bayesian Inference (Posterior Probabilities (PP): cyt *b*, PP=90%; cyt *b*+*S7*, PP=100%) (Fig. 2; Fig. S1). For Clade A, two sub-clades were recovered in cyt *b* and concatenated data set (BS=100% and PP=100%): A1 clustered specimens from the Casas Grandes River, whereas Sub-clade A2 clustered specimens from the Santa Maria River. For Clade B, two well supported sub-clades were also recovered, for cyt *b* and concatenated data set (BS and PP = 100%), Sub-clade B1 included specimens from localities of the Lower Conchos Drainage (Villa Coronado and El Porvenir), as well as specimens from localities of the Nazas River drainage (Peñon de Covadonga, Abasolo, Paso Nacional and Gicorica), whereas Sub-clade B2 included the samples collected in the Yaqui drainage (Cabullona locality). The phylogeny recovered with *S7* identified the same four well differentiated sub-clades: Casas Grandes (A1), BS=99% and PP=100%; Santa Maria (A2), BS and PP=100%; Nazas+Conchos (B1), BS=98% and PP=100%; and Yaqui (B2), BS=59% and

PP=60%, but with different relationships. The Santa Maria Lineage (Sub-clade A2) was a sister group to a well supported clade including all three Sub-clades, Casas Grandes (A1), Nazas+Conchos (B1) and Yaqui (B2) (Fig. S2).

Species tree analysis and divergence times

The species tree analysis supports the assumption of two well-differentiated clades and four sub-clades with high internal branch-length Bayesian Posterior Probability (>99%), with the same relationships as the cyt *b* and the concatenated phylogenetic tree. This is also consistent with the *S7* gene phylogenetic analyses, in both cases corresponding to the four recovered sub-clades: Casas Grandes (A1), Santa Maria (A2), Nazas+Conchos (B1) and Yaqui (B2) (Fig. 3).

Divergence between the main clades and sub-clades of *P. promelas* was dated at the Pleistocene. The separation of main clades was dated *ca*. 1.39 Ma (95% HPD: 0.53-2.25) (Fig. 3). The split between the two sub-clades within Clade A (Santa Maria *vs* Casas Grandes) was estimated at *ca*. 1.06 Ma (95% HPD: 0.21-1.92), whereas within Clade B the separation event of the two Sub-clades (Yaqui *vs* Nazas+Conchos) was estimated at *ca*. 0.94 Ma (95% HPD: 0.25-1.64) (Fig. 4).

The maximum distance for both genes occurred between Nazas+Conchos (Sub-clade B1) *vs* Santa Maria (Sub-clade A2) (10.6% for cyt *b* and 4.7% for *S7*) (Table 1). The lower genetic distance for the cyt *b* gene was found between the two sub-clades whithin Clade B, the Nazas+Conchos Lineage *vs* the Yaqui Lineage (3.7%). For *S7* the lower genetic distance occurred between Casas Grandes (Sub-clade A1) *vs* Yaqui (Sub-clade B2) (1.1%). Genetic distances within each sub-clade ranged from 0% to 0.5% with cyt *b*, and 0.1 to 0.4 with *S7* (Table 1).

Haplotype networks

For both genes, the haplotype networks showed the existence of four haplogroups, corresponding to the four sub-clades found in phylogenetic analyses: Casas Grandes (Sub-clade A1), Santa Maria (Sub-clade A2), Nazas+Conchos (Sub-clade B1) and Yaqui (Sub-clade B2), but as in phylogenetic analyses, different relationships were found between the cyt *b* and *S7* groups. In the cyt *b* haplotype network no mixed haplotypes between drainages were found, with the same distribution of samples as in phylogenetic analyses. The mutation steps between sub-clade B2 and sub-clade B1 were 53; 19 between sub-clade A1 and A2, whereas de mutation steps between sub-clades A1 and B1 were 125. For *S7* a mixture of haplotypes between the Nazas and Conchos samples was found. Sub-clade A2 was separated by 22 mutation steps from B2; A1 was separated by three mutation steps from B2, whereas B1 and B2 were separated by four mutation steps with respect to B1.

Discussion

The results of the implemented analyses presented herein showed the existence of four highly divergent and supported clades within Southern populations of *P. promelas* distributed in Mexico. Previous zoogeographic studies support a complex evolutionary history of the freshwater fish species distributed along the Mesa del Norte in Mexico (Wood & Mayden, 2002; Blum et al., 2008; Schönhuth & Mayden, 2010; Domínguez-Domínguez et al., 2011; Schönhuth et al., 2011, 2012, 2014, 2015; Corona-Santiago et al., 2018, in press). Moreover, freshwater fish species previously considered as a single widespread taxon, have been determined as a species complex based on phylogenetic analyses, using mitochondrial and nuclear genes, as is the case of *Campostoma ornatum* (Domínguez-Domínguez et al., 2011; Schönhuth et al., 2011), and

294 *Codoma ornata* (Schönhuth et al., 2015). Even in United States populations of *P. promelas,*
295 previous studies show high geographic variation on morphological and meristic traits, suggesting
296 a species richness subestimation (Vandermeer, 1966).

297 In the current work, the findings of highly divergent genetic groups based on
298 phylogenetic trees, species tree analysis, genetic distances and haplotype networks, highlighted
299 the possibility of the existence of a species complex within Mexican *P. promelas* populations.

300 **Phylogenetic relationships and taxonomic implications**

301 *Pimephales promelas* is considered a widespread species (Miller, Minkley & Norris, 2005),
302 highly susceptible to population fragmentation and speciation processes (Letsch, Gottsberger &
303 Ware, 2016). Althought the current study focused on a part of the species´ distribution, the
304 Chihuahuan Desert and SMO basin drainages, genetic distances (Table 1), phylogenetic
305 relationships (Fig. 2; Figs. S1 and S2), the species tree (Fig. 3) and the results of haplotype
306 networks (Fig. 4), suggest the existence of four highly differentiated and supported groups within
307 *P. promelas* populations. One grouping the specimens from the Casas Grandes River; another
308 grouping the specimens collected in the Santa Maria River; a third group included the specimens
309 collected in the Nazas+Conchos rivers; and the last one contained all samples pertaining to the
310 Yaqui River.

311 The genetic divergences in the cyt *b* among these four lineages ranged between 3.7%
312 among Yaqui and Nazas+Conchos populations, to 10.6% among populations of the
313 Nazas+Conchos rivers and the Santa Maria River drainage. These genetic distances are the same
314 or above the genetic distances found between the seven species of the Chihuahuan Desert Group
315 of the genus *Gila,* that ranges between 3.68 to 5.56 (Schönhuth et al., 2014), the lowest distance
316 found between well recognized Southwest species in the *Dionda* genus of 1.3% (Schönhuth et

al., 2008), the lowest distance of 1.9% found between well recognized species in *Cyprinella* and 4.7% found in species of the *Tampichthys* genus (Schönhuth & Mayden, 2010) and the 2.1-4.0% found between species of *Algansea* (Pérez-Rodríguez et al., 2009), or even above the mean cut-off value of 2% for the recognition of species with the cyt *b* gene (Avise, 1996; Bradley & Baker, 2001). Also the *S7* showed relatively high genetic differences for a nuclear gene, being the low genetic distances (*p*-distance 1.1%) in the comparision between the geographically proximate Casas Grandes Lineage *vs* the Yaqui Lineage drainages, and the higher (*p*-distance 4.7%) Santa Maria *vs* Nazas+Conchos populations. Previous studies have found similar genetic distances in *S7* nuclear gene for a well recognized species in the cyprinid family; as in the *Algansea* genus in which interspecific genetic distances between 0.5 to 3.6% were found (Pérez-Rodríguez et al., 2009).

Pimephales promelas was described by Rafinesque (1820) from a Pond near Lexington, Kentucky, USA. Considering the results of the present work, we suggested that *P. promelas* must be considered as a species complex. Previous phylogenetic and phylogeographic studies also proposed that fish diversity in Northen Mexico has been largely underestimated (Domínguez-Domínguez et al., 2011; Schönhuth et al., 2011, 2012, 2014, 2015). According to the analyses presented herein, the Southern populations of *P. promelas* show at least four evolutionary independent groups that could be recognized as different species. A more integrative taxonomic analysis is pending to complement these results; it should include morphological, morphometric, meristic and pigmentation characters, as well as a more extensive sample size.

Evolutionary history of *P. promelas*

Although with the information and analyses provided herein it is possible to elucidate a general biogeographic and evolutionary process in Southern populations of *P. promelas*, some

incongruences in the phylogenetic signal between markers was found. However, the incongruence found between cyt *b* and *S7* datasets can not be attributed to a retention of ancestral polymorphisms because of incomplete lineage sorting and/or introgression following secondary contact, because we did not recover shared haplotypes between the four well differentiated sub-clades in neither of the two genes (Fig. 4; Figs. S1 and S2). Genetic distances recovered between clades were high (Table 1), and we found the same topology in the cyt *b* phylogeny, concatenated gene tree and species tree (Fig S1; Figs. 2 and 3). The more plausible explanation for gene tree incongruence comes from the use of a small number of sequence data in the Santa Maria Basin, which has been shown to generate phylogenetic hypotheses that are incongruent or lacking support (Satta, Klein & Takahata, 2000; Kopp & True 2002; Rokas et al., 2003; Rokas & Carroll, 2005). Also, the mtDNA mutation rate is typically one order of magnitude higher than the nuclear one (Ballard & Whitlock, 2004; Brown, George & Wilson, 1979; Lynch, 2006; Nabholz, Glémin & Galtier, 2009), as was the case for the studied Southern populations of *P. promelas*. In this way, the incongruence between gene trees in the current study could be due to the presence of only one specimen in the Santa Maria Lineage and different genetic mutation rates between nuclear and mitochondrial DNA.

Cladogenesis of the main clades

Previous biogeographic studies of Chihuahuan Desert and SMO fish fauna have mentioned ancient and recent tecto-volcanic events and climatic factors that explain the geographic distribution and phylogenetic relationships of the freshwater fishes of the region (Smith & Miller, 1986; Wood & Mayden, 2002; Blum et al., 2008; Schönhuth & Mayden, 2010; Domínguez-Domínguez et al., 2011; Schönhuth et al., 2011, 2012, 2014, 2015; Aranda-Gómez et al., 2018). In the current study, the divergence time estimation separating Clade B, distributed

363 in the Yaqui and Nazas+Conchos rivers, to those of the Clade A, distributed in the Guzman
364 Basin, was calculated to have ocurred during the Pleistocene (*ca.* 1.39 Ma, 95% HPD: 0.53-
365 2.25). The most plausible biogeographic scenario for the separation of Clade A *vs* Clade B is the
366 desiccation of an ancient pluvial paleolake of the Pleistocene that connected present Yaqui River
367 and drainages from the current Guzman Basin, with the subsequent disruption of the connection,
368 as has been proposed by Smith & Miller (1986) for co-distributed freshwater fishes (Fig. 5A).
369 This pluvial paleolake could have been formed by the Lake Cabeza de Vaca and the actual
370 Babicora Lagoon Basin. Studies based on allozyme analysis (Wood & Mayden, 2002) and
371 molecular data in codistributed freshwater fishes of the North of Mexico (Kim & Conway, 2014;
372 Corona-Santiago et al., 2018, in press), as well as Chihuahuan Desert herpetofauna (Axtell,
373 1977; Rosenthal & Forstner, 2004), have hypothesized the existence of the Late Pliocene Lake
374 Cabeza de Vaca (Strain, 1966), that formed a basin corridor during cycles of heavy precipitation
375 since Late Pliocene to Middle Pleistocene (Reeves, 1965; 1969; Strain, 1966; Rosenthal &
376 Forstner, 2004). Previous studies indicate that the endorreic basin in the Grande River depression
377 started its formation during the Late Terciary and culminated in the mid-Pleistocene (Ruhe,
378 1962; Metcalf, 1967; Gile, Hawley & Grossman, 1981; Hawley, 1969; Smith & Miller, 1986),
379 whereas the drain of the ancient paleolake Cabeza de Vaca began during the early-middle
380 Pleistocene (Reeves, 1965; 1969), caused by tectonic activity in the Grande River Rift and the
381 glacial era (Kansan, Middle Pleistocene) (Strain, 1966; Corona-Santiago et al., 2018, in press).
382 For ichtyofauna, pluvial paleodrainages may have facilitated a radiation/expansion of ancestral
383 forms across the Chihuahuan Desert basins; including possible transfers and/or integration of
384 headwaters of the Mesa del Norte with Pacific drainages, probably caused by pluvial paleolakes,
385 and later fragmentation of them by climatic ant tecto-volvanic events, as have been previously

386 stated for the connection and subsequent isolation of the Yaqui and Guzman basins supported by
387 other co-distributed fish species as *Cyprinella formosa* (Wood & Mayden, 2002), *Campostoma*
388 *ornatum* (Domínguez-Domínguez et al., 2011; Schönhuth et al., 2011) and *Gila nigrescens*
389 (Schönhuth et al., 2014). Similarly, previous studies in *Cyprinodon*, *Salmo* sp., and *Ictalurus*
390 *pricei* (Smith & Miller, 1986; Miller, 1959) have found a relationship between the Guzman
391 Basin and Yaqui River basins, suggesting a former hydrographic exchange between the Yaqui
392 Drainage and the Guzman Basin.

393 ***Major Clade A: Casas Grandes and Santa Maria lineages split***

394 Within Clade A, the divergence time for the separation between the Casas Grandes and
395 Santa Maria sub-clades was estimated to have occurred *ca.* 1.06 Ma (95% HPD: 0.21-1.92) in
396 Pleistocenic time, associated to the Lake Palomas fragmentation via desertification in the Los
397 Muertos Bolson (Fig. 5B). It is hypothethised that in the mid-Pleistocene, the fragmentation of
398 the Lake Cabeza de Vaca created the current configuration of the Bravo River in Mexico and
399 formed the pluvial Lake Palomas, in the current Guzman Basin (Rosenthal & Forstner, 2004;
400 Smith & Miller, 1986). During the Middle Pleistocene, Lake Palomas covered much of the
401 Guzman Basin (Reeves, 1969), the lake complex was fed by the four existing rivers: Casas
402 Grandes, Santa Maria, Carmen and Mimbres (Reeves, 1969; Hawley, 1969); it was in turn
403 fragmented by climatic changes since the Pleistocene, forming the current configuration of the
404 basins in the area (Strain, 1970; Wood & Mayden, 2002; Corona-Santiago et al., 2018, in press;
405 Kim & Conway, 2014). Previous studies propose that the isolation between the Casas Grandes,
406 the Upper Grande River basins, the Del Carmen and the Santa Maria rivers ocurred *ca.* 1.8 Ma,
407 by dessication of the Los Muertos Bolson (Reeves, 1969; Wood & Mayden, 2002; Domínguez-
408 Acosta & Gill, 2007; Echelle, 2008; Corona-Santiago et al., 2018, in press). The event of

409 isolation within the Guzman Basin, which separated the Casas Grandes and Santa Maria
410 populations, was similar to those found by Domínguez-Domínguez et al. (2011) in *Campostoma*
411 *ornatum*; these showed Pleistocenic event that isolated the Casas Grandes River populations
412 around 1.5 Ma, and associated these events mainly to the aridity increase in the zone (Galloway,
413 Whiteaker & Ganey-Curry, 2011). In *Pantosteus plebeius* the isolation of the Palomas Lagoon
414 from the Casas Grandes River was *ca.* 1.1 Ma, whereas the isolation of the Del Carmen from the
415 Santa Maria drainages was calculated to ocurr at 1.2 Ma (Corona-Santiago et al., 2018, in press),
416 mainly proposed by the desertification process in the Los Muertos Bolson (Reeves, 1969; Wood
417 & Mayden, 2002; Domínguez-Acosta & Gill, 2007; Echelle, 2008), allowing the fragmentation
418 of the current Guzman complex.

419 ***Major Clade B: Yaqui and Nazas+Conchos lineages split***

420 Within major Clade B, the separation between the Yaqui and the Nazas+Conchos sub-
421 clades was dated at the Pleistocene, approximately 0.94 Ma (95% HPD: 0.25-1.64). This event
422 seems to be related to the fragmentation of a pluvial paleolake that was present in the Pleistocene
423 time or a river capture and peripheral isolation, induced by tectonic activity in the SMO (Ferrari,
424 Valencia-Moreno & Bryan, 2007; Aguirre-Díaz et al., 2008) (Fig. 5C), as has been hypothesised
425 for those Pacific slope rivers that have headwaters extending eastward in the SMO to areas of the
426 Mesa del Norte. In which previous studies identified possible transfers of fish species between
427 drainages by integration or fragmentation of drainages, as is the case of the Yaqui River.
428 Previous studies have documented closely related lineages in both sides of the SMO, in the
429 Yaqui and Conchos rivers. Also the climatic changes caused by glacial/interglacial cycles, gave
430 rise to connections/disconnections between interior basins in the area (Galloway, Whiteaker &
431 Ganey-Curry, 2011), even during periods of adequate rainfall, the interior basins were flooded by

pluvial lakes (Smith & Miller, 1986), shaping freshwater species distributions (Minckley et al., 1986; Miller & Smith, 1986). The fish interchange between basins has involved fish fauna widely distributed as *Codoma ornata*, *Campostoma ornatum*, *Gila* spp., *Rhinichthys cataractae* and *Catostomus.* This connection could be related to stream capture events since the Plio-Pleistocene and a later disconnection during Late Pleistocene to Holocene times (Smith & Miller, 1986; Miller, Minkley & Norris, 2005; Domínguez-Domínguez et al., 2011; Schönhuth et al., 2011, 2014, 2015; Kim & Conway, 2014), through the combined influence of tectonic events and increasing regional aridity (Galloway, Whiteaker & Ganey-Curry, 2011).

A mixture of samples from the current Nazas and Conchos isolating basins was found within the Sub-clade B1 of *P. promelas*. This could be related to a recent connection of both basins via river capture or through pluvial lakes, as shown by the cyt *b* gene, whereas the mixture of haplotypes between both populations in *S7* could be related to incomplete lineage sorting due the low mutation rate of the nDNA. The close relationship between samples of both drainages was previously found in *Campostoma ornatum* populations, a relationship explained by a fish interchange as a result of river capture (Schönhuth et al., 2011). In the same way, Burr (1976), hypothesized in a previous review of *Campostoma ornatum*, the formation of a connection between the Nazas River and the Grande River via the Mayran and Viesca lagoons (both currently dry) during the Late Pleistocene (Meek, 1904; Burr, 1976). Similarly, a previous biogeographic study in *Codoma ornata* found a close relationship between the Conchos and Nazas drainage populations, attributed to an incomplete lineage sorting, but no biogeographic scenario was presented (Schönhuth et al., 2015).

Conservation considerations

The conservation concept at the species level using molecular data gained importance when the concept of Evolutionary Significant Units (ESU's) emerged (Ryder, 1986; Domínguez-Domínguez & Vázquez-Domínguez, 2009), and since, genetic diversity was recognized as the baseline level of biodiversity (Domínguez-Domínguez & Vázquez-Domínguez, 2009). An Evolutionary Significant Unit (ESU) has been defined as a group of individuals or populations that present reciprocal monophyly for mitochondrial markers and significant divergences in allelic frequencies at the nuclear loci, considering in turn the divergence time between populations (Moritz, 1994, 2002; Domínguez-Domínguez & Vázquez-Domínguez, 2009). This idea of conservation seeks to identify the management units that are part of the evolutionary history of the lineages within the species, to create effective programs for their conservation, in order to preserve the evolutionary processes, indispensable for the long-term permanence of the species, and the associated evolutionary factors (Avise & Hamrick, 1996; Crandall et al., 2000; Moritz, 2002; Pertoldi, Bijlsma & Loeschcke, 2007; Domínguez-Domínguez & Vázquez-Domínguez, 2009). The results presented herein support the existence of four well differentiated genetic groups within the study populations of *Pimephales promelas,* that may correspond to independent and evolutionary significant units, that must be preserved in order to warrant the long-term conservation of the diferent evolutionary trajectories.

Conclusions

The geographic distribution of the four genetic groups recovered in *P. promelas* (Casas Grandes, Santa Maria, Yaqui and Nazas+Conchos lineages) seems to be similar to the lineage distribution found in other freshwater fishes in the North of Mexico, as *Pantosteus plebeius* (Coronado-Santiago et al., 2018, in press) *Rhinichthys cataractae* (Kim & Conway, 2014), *Codoma ornata* (Schönhuth et al., 2015) and *Campostoma ornatum* (Schönhuth et al., 2011; Domínguez-

Domínguez et al., 2011). Cladogenic events in *P. promelas* appear to have been caused by the combined influence of tectonic events and increasing regional aridity. Particularly, the fragmentation of the ancestral Lake Cabeza de Vaca and interchange events between basins via stream capture. The four genetic groups must be considered as Evolutionary Significant Units that must be conserved.

Acknowledgements

This paper is part of results of N.E.B.-N's PhD research in the Programa Institucional de Doctorado en Ciencias Biológicas, UMSNH. N.E.B.-N. and R.G.B. thank Consejo Nacional de Ciencia y Tecnología for the scholarship received to complete their PhD research in the Posgrado Institucional de Doctorado en Ciencias Biológicas, UMSNH (CONACYT-No. 270686 and CONACYT-No. 270686, respectively). We thank Dr. Isaí Betancourt Resendes, Dr. Victor Julio Piñeros and Dr. Diushi Keri Corona Santiago for their valuable help in the genetic analyses. This investigation was financially supported by grant from the Program Coordinadora de Investigaciones Científicas (CIC-UMSNH) to R.P.R. R.P.R. thanks to PRODEP-UMSNH-PTC-384 for partial financial support.

497 References

498 1. **Aguirre-Díaz GJ, Labarthe-Hernández G, Tristán-González M, Nieto-Obregón J,**
499 **Gutiérrez-Palomares I. 2008.** The ignimbrite flare-up and graben calderas of the Sierra
500 Madre Occidental, Mexico. In: Gottsman J, Marti J, eds. *Caldera Volcanism: analysis,*
501 *modelling and response*. New York: Elsevier, Developments in volcanology. **10**:143-180
502 DOI: 10.1016/S1871-644X(07)00004.
503 *2.* **Aranda-Gómez JJ, Carranza-Castañeda O, Wang X, Tseng ZJ, Pacheco-Castro A.**
504 **2018.** Notes on the origin of extensive endorheic regions in central and northern Mexico,
505 and some implications for paleozoogeography. *Journal of South American Earth*
506 *Sciences* **83**:55-67 DOI: 10.1016/j.jsames.2018.02.003.
507 3. **Avise JC, Hamrick JL. 1996.** *Conservation genetics: case histories from nature.* New
508 York: Chapman and Hall.
509 4. **Axtell RW. 1977.** Ancient playas and their infuence on the recent herpetofauna of the
510 northern Chihuahuan Desert. In: Wauer RH, Riskind DH, eds. *Transactions of the*
511 *Symposium on the Biological Resources of the Chihuahuan Desert Region, United States*
512 *and Mexico.* US Department of the Interior: National Park Service Transactions and
513 Proceedings. **3**:493-511.
514 5. **Ballard J, Whitlock M. 2004.** The incomplete natural history of mitochondria.
515 *Molecular Ecology* **13**(**4**):729-744 DOI: 10.1093/oxfordjournals.molbev.a026036.
516 6. **Bandelt H, Forster P, Röhl A. 1999.** Median-joining networks for inferring intraspecific
517 phylogenies. *Molecular Biology and Evolution* **16**(**1**):37-48 DOI:
518 10.1093/oxfordjournals.molbev.a026036.

7. **Berendzen PB, Gamble T, Simons AM. 2008.** Phylogeography of the bigeye chub *Hybopsis amblops* (Teleostei: Cypriniformes): early Pleistocene diversification and post-glacial range expansion. *Journal of Fish Biology* **73**(**8**):2021-2039 DOI: 10.1111/j.1095-8649.2008.02046.x.
8. **Blum MJ, Neely DA, Harris PM, Mayden RL. 2008.** Molecular systematics of the Cyprinid genus *Campostoma* (Actinopterygii: Cypriniformes): disassociation between morphological and mitochondrial differentiation. *Copeia* **2008**(**2**):360-369 DOI: 10.1643/CI-06-093.
9. **Bradley RD, Baker RJ. 2001.** A test of the genetic species concept: cytochrome *b* sequences and mammals. *Journal of Mammalogy* **82**(**4**):960-973 DOI: 10.1644/1545-1542(2001)082<0960:ATOTGS>2.0.CO;2.
10. **Brown WM, George M, Wilson AC. 1979.** Rapid evolution of animal mitochondrial DNA. *Proceedings of the National Academy of Sciences of the United States of America* **76**(**4**):1967-1971.
11. **Burr BM. 1976.** A review of the Mexican Stoneroller, *Campostoma ornatum* Girard (Pisces: Cyprinidae*). Transactions of the San Diego Society of Natural History* **18**:127-143 DOI: 10.5962/bhl.part.15545.
12. **Ceballos G, Arroyo-Cabrales J, Ponce E. 2010.** Effects of Pleistocene environmental changes on the distribution and community structure of the mammalian fauna of Mexico. *Quaternary Research* **73**(**3**):464-473 DOI: 10.1016/j.yqres.2010.02.006.
13. **Chow S, Hazama K. 1998.** Universal PCR primers for *S7* ribosomal protein gene introns in fish. *Molecular Ecology* **7**(**9**):1255-1256.

14. **Clements MD, Bart HL, Hurley DL. 2012.** A different perspective on the phylogenetic relationships of the Moxostomatini (Cypriniformes: Catostomidae) based on cytochrome-*b* and Growth Hormone intron sequences. *Molecular Phylogenetics and Evolution* **63**(**1**):159-167 DOI: 10.1016/j.ympev.2012.01.001.
15. **Corona-Santiago DK, Domínguez-Domínguez O, Tovar-Mora Ll, Pardos-Blas JR, Herrerías-Diego Y, Pérez-Rodríguez R, Doadrio I. 2018.** Historical biogeography of the *Pantosteus plebeius-nebuliferus* complex (Actinopterygii: Catostomidae). *BMC Evolutionary Biology* (In press).
16. **Crandall KA, Bininda-Emonds ORP, Mace GM, Wayne RK. 2000.** Considering evolutionary processes in conservation biology. *Trends in Ecology and Evolution* **15**(**7**):290-295 DOI: 10.1016/S0169-5347(00)01876-0.
17. **Devitt TJ. 2006.** Phylogeography of the Western Lyresnake (*Trimorphodon biscutatus*): testing aridland biogeographical hypotheses across the Neartic-Neotropical transition. *Molecular Ecology* **15**(**14**):4387-4407 DOI: 10.1111/j.1365-294X.2006.03015.x.
18. **Doadrio I, Perdices A, Machordom A. 1996.** Allozymic variation of the endangered killifish *Aphanius iberus* and its application to conservation. *Environmental Biology of Fishes* **45**:259-271.
19. **Domínguez-Acosta M, Gill TE. 2007.** PIXE Based Geochemical Characterization of the Pluvial Lake Palomas System, Chihuahua, Mexico. In: Miranda J, Ruvalcaba-Sil JL, De Lucio OG, eds. *Procceedings of the 11th International Conference on PIXE and its Analytical Applications*. Puebla, Mexico. PII-33.

562 **20. Domínguez-Domínguez O, Vázquez-Domínguez E. 2009**. Phylogeography:
563 Applications in taxonomy and conservation. *Animal Biodiversity and Conservation*
564 **32**:59-70.
565 **21. Domínguez-Domínguez O, Vila M, Pérez-Rodríguez R, Remón N, Doadrio I. 2011.**
566 Complex evolutionary history of the Mexican Stoneroller *Campostoma ornatum* Girard,
567 1856 (Actinopterygii:Cyprinidae). *BMC Evolutionary Biology* **11**:153 DOI:
568 10.1186/1471-2148-11-153.
569 **22. Dowling TE, Tibbets CA, Minckley WL, Smith GR. 2002.** Evolutionary relationships
570 of the plagopterins (Teleostei: Cyprinidae) from cytochrome *b* sequences. *Copeia*
571 **2002**(**3**):665-678 DOI: 10.1643/0045-8511(2002)002[0665:EROTPT]2.0.CO;2.
572 **23. Drummond AJ, Ho SYW, Phillips MJ, Rambaut A. 2006.** Relaxed Phylogenetics and
573 Dating with Confidence. *PLoS Biology* **4**(**5**):e88 DOI: 10.1371/journal.pbio.0040088.
574 **24. Drummond AJ, Suchard MA, Xie D, Rambaut A. 2012.** Bayesian phylogenetics with
575 BEAUti and the BEAST 1.7. *Molecular Biology and Evolution* **29**(**8**):1969-1973 DOI:
576 10.1093/molbev/mss075.
577 **25. Echelle AA, Carson EW, Echelle AF, Van Der Bussche RA, Dowling TE, Meyer A.**
578 **2005.** Historical biogeography of the New-World pupfish genus *Cyprinodon* (Teleostei:
579 Cyprinodontidae). *Copeia* **2005**(**2**):320-339 DOI: 10.1643/CG-03-093R3.
580 **26. Echelle AA. 2008.** The western North American pupfish clade (Cyprinodontidae:
581 Cyprinodon): mitochondrial DNA divergence times and drainage history. *Geological*
582 *Society of America Special Papers* **439**:27-38 DOI: 10.1130/2008.2439(02).

583 **27. Ferrari L, Valencia-Moreno M, Bryan S. 2006**. Magmatismo y tectónica en la Sierra
584 Madre Occidental y su relación con la evolución de la margen occidental de
585 Norteamérica. *Boletín De La Sociedad Geológica Mexicana* **57**(**3**):343-378.
586 **28. Galloway WE, Whiteaker TL, Ganey-Curry P. 2011.** History of Cenozoic North
587 American drainage basin evolution, sediment yield, and accumulation in the Gulf of
588 Mexico basin. *Geosphere* **7**(**4**):938-973 DOI: 10.1130/GES00647.1.
589 **29. Gile LH, Hawley JW, Grossman RB. 1981.** Soils and geomorphology in the Basin and
590 Range area of southern New Mexico-guidebook to the desert project. *New Mexico*
591 *Bureau of Mines and Mineral Resources* **1981**(**39**):222.
592 **30. Hawley JW. 1969.** Notes on the Geomorphology and late Cenozoic geology of
593 northwestern Chihuahua. In: Cordoba DA, Wengerd SA, Shomaker JW, eds. *The border*
594 *region (Chihuahua, Mexico and USA)*. New Mexico Geological Society 20th Annual Fall
595 Field Conference Guidebook. 131-142.
596 **31. Hughes RM, Rinne JA, Calamusso B. 2005.** Historical changes in large river fish
597 assemblages of the Americas: a synthesis. In: Rinne JN, Hughes RM, Calamusso B, eds.
598 *Historical changes in large river fish assemblages of the Americas.* Bethesda, Maryland:
599 American Fisheries Society, Symposium 45. 603-612.
600 **32. Kim D, Conway KW. 2014.** Phylogeography of *Rhinichthys cataractae* (Teleostei:
601 Cyprinidae): pre-glacial colonization across the Continental Divide and Pleistocene
602 diversification within the Rio Grande drainage. *Biological Journal of the Linnean Society*
603 **111**(**2**):317-333 DOI: 10.1111/bij.12209.
604 **33. King PB. 1935.** Outline of structural development of trans-Pecos Texas. *AAPG Bulletin*
605 **19**(**2**):221-261.

606 **34. Kopp A, True JR. 2002.** Phylogeny of the oriental *Drosophila melanogaster* species
607 group: a multilocus reconstruction. *Systematic Biology* **51**(**5**):786-805 DOI:
608 10.1080/10635150290102410.
609 **35. Lanfear R, Calcott B, Ho SY, Guindon S. 2012.** PartitionFinder: Combined selection of
610 partitioning schemes and substitution models for phylogenetic analyses. *Molecular*
611 *Biology and Evolution* **29**(**6**):1695-1701 DOI: 10.1093/molbev/mss020.
612 **36. Letsch H, Gottsberger B, Ware JL. 2016.** Not going with the flow: a comprehensive
613 time-calibrated phylogeny of dragonflies (Anisoptera: Odonata: Insecta) provides
614 evidence for the role of lentic habitats on diversification. *Molecular Ecology* **25**(**6**):1340-
615 1353 DOI: 10.1111/mec.13562.
616 **37. Librado P, Rozas J. 2009.** DnaSP v5: a software for comprehensive analysis of DNA
617 polymorphism data. *Bioinformatics* **25**(**11**):1451-1452 DOI:
618 10.1093/bioinformatics/btp187.
619 **38. Lynch M. 2006.** The origins of eukaryotic gene structure. *Molecular Biology and*
620 *Evolution* **23**(**2**):450–468 DOI: 10.1093/molbev/msj050.
621 **39. Mayden RL, Matson RH, Hillis DM. 1992.** Speciation in the North American genus
622 *Dionda* (Teleostei: Cypriniformes). In: Mayden RL, ed. *Systematics, Historical Ecology,*
623 *and North American Freshwater Fishes*. Stanford, California: Stanford University Press.
624 710-746.
625 **40. Mayden RL, Burr BM, Page LM, Miller RR. 1992.** The native freshwater fishes of
626 North America. In: Mayden RL, ed. *Systematics, Hystorical Ecology, and North*
627 *American Freshwater Fishes*. Stanford, California: Stanford University Press. 827-863.

628 **41. Meek SE. 1904.** The fresh-water fishes of Mexico north of the Isthmus of Tehuantepec.
629 *Field Columbian Museum Publication*, *Zoological Series* **5**:1-252.

630 **42. Metcalf AL. 1967.** Late Quaternary mollusks of the Rio Grande Valley, Caballo Dam,
631 New Mexico to El Paso, Texas. *Science Series, University of Texas at El Paso* **1**:1-62.

632 **43. Metcalfe SE. 2006.** Late quaternary environments of the Northern deserts and central
633 transvolcanic belt of Mexico. *Annals of the Missouri Botanical Garden* **93**(**2**): 258-273.

634 **44. Miller RR, Minkley WL, Norris SM. 2005.** *Freshwater Fishes of Mexico*. Chicago:
635 University of Chicago Press.

636 **45. Miller MA, Pfeiffer W, Schwartz T. 2010.** Creating the CIPRES Science Gateway for
637 inference of large phylogenetic trees. *Gateway Computing Environments Workshop*
638 *(GCE)* **2010**:1-8 DOI: 10.1109/GCE.2010.5676129.

639 **46. Minckley WL, Hendrickson DA, Bond CE. 1986.** Geography of Western North
640 American freshwater fishes: description and relationships to intracontinental tectonism.
641 In: Hocutt CH, Wiley EO, eds. *The Zoogeography of North American Freshwaters*
642 *Fishes*. New York: John Wiley and Sons. 519-614.

643 **47. Moritz C. 1994.** Defining 'evolutionarily significant units' for conservation. *Trends in*
644 *Ecology and Evolution* **9**(**10**): 373-375 DOI: 10.1016/0169-5347(94)90057-4.

645 **48. Moritz C. 2002**. Strategies to protect biological diversity and the evolutionary processes
646 that sustain it. *Systematic Biology* **51**(**2**):238-254 DOI: 10.1080/10635150252899752.

647 **49. Morrone JJ. 2010.** Fundamental biogeographic patterns across the Mexican Transition
648 Zone: an evolutionary approach. *Ecography* **33**(**2**):355-361 DOI: 10.1111/j.1600-
649 0587.2010.06266.x.

650 **50. Nabholz B, Glémin S, Galtier N. 2009.** The erratic mitochondrial clock: variations of
651 mutation rate, not population size, affect mtDNA diversity across birds and mammals.
652 *BMC Evolutionary Biology* **9**:54 DOI: 10.1186/1471-2148-9-54.
653 **51. Nagle BC, Simons AM. 2012.** Rapid diversification in the North American minnow
654 genus *Nocomis*. *Molecular Phylogenetics and Evolution* **63(3)**:639-649 DOI:
655 10.1016/j.ympev.2012.02.013.
656 **52. Nathan R. 2001.** The challenges of studying dispersal. *Trends in Ecology and Evolution*
657 **16(9)**:481-483 DOI: 10.1016/S0169-5347(01)02272-8.
658 **53. Pérez-Rodríguez R, Domínguez-Domínguez O, Doadrio I, Cuevas-García E, Pérez-**
659 **Ponce de León G. 2015.** Comparative historical biogeography of three groups of Neartic
660 freshwater fishes across central Mexico. *Journal of Fish Biology* **86(3)**:993-1015 DOI:
661 10.1111/jfb.12611.

662 **54. Pérez-Rodríguez R, Domínguez-Domínguez O, Pérez-Ponce de León G, Doadrio I.**
663 **2009.** Phylogenetic relationships and biogeography of the genus *Algansea* Girard
664 (Cypriniformes: Cyprinidae) of central Mexico inferred from molecular data. *BMC*
665 *Evolutionary Biology* **9**:223 DOI: 10.1186/1471-2148-9-223.
666 **55. Pertoldi C, Bijlsma R, Loeschcke V. 2007.** Conservation genetics in a globally
667 changing environment: present problems, paradoxes and future challenges. *Biodiversity*
668 *and Conservation* **16(14)**:4147-4163.
669 **56. Posada D. 2003.** Unit 6.5: Using MODELTEST and PAUP* to select a model of
670 nucleotide substitution. In: Baxevanis AD, *et al.*, eds. *Current Protocols in*
671 *Bioinformatics*. John Wiley and Sons **1**:6.5.1-6.5.28
672 DOI: 10.1002/0471250953.bi0605s00.

673 **57. Rambaut A, Drummond A. 2007.** Tracer v.1.5 software. Available at
674 http://tree.bio.ed.ac.uk/software/tracer (accessed 11 November 2016).
675 **58. Rambaut A. 2014.** FigTree v.1.4.2 software. Available at
676 https://tree.bio.ed.ac.uk/software/figtree (accessed 11 November 2016).
677 **59. Reeves C. 1965.** Pluvial Lake Palomas, northwestern Chihuahua, Mexico and Pleistocene
678 geologic history of south-central New Mexico. In: Fitzimmons JP, Balk CL, eds.
679 *Guidebook 16: Southwestern New Mexico II*. Socorro, New Mexico: New Mexico
680 Geological Society. 199-203.
681 **60. Reeves C. 1969.** Pluvial Lake Palomas northwestern Chihuahua, Mexico. In: Cordoba
682 DA, Wengerd SA, Shomaker JW, eds. *Guidebook 20: The Border Region (Chihuahua,*
683 *Mexico, & USA)*. Socorro, New Mexico: New Mexico Geological Society. 143–154.
684 **61. Riddle BR, Hafner DJ. 2006.** A step-wise approach to integrating phylogeographic and
685 phylogenetic biogeographic perspectives on the history of a core North American warm
686 deserts biota. *Journal of Arid Environments* **66(3)**:435-461 DOI:
687 10.1016/j.jaridenv.2006.01.014.
688 **62. Rokas A, Carroll SB. 2005.** More genes or more taxa? The relative contribution of gene
689 number and taxon number to phylogenetic accuracy. *Molecular Biology and Evolution*
690 **22(5)**:1337-1344 DOI: 10.1093/molbev/msi121.
691 **63. Rokas A, Williams BL, King N, Carroll SB. 2003.** Genome-scale approaches to
692 resolving incongruence in molecular phylogenies. *Nature* **425(6960)**:798-804 DOI:
693 10.1038/nature02053.
694 **64. Ronquist F, Teslenko M, Van der Mark P, Ayres DL, Darling A, HÖhna S, Larget**
695 **B, Liu L, Suchard MA, Huelsenbeck JP. 2012.** MrBayes 3.2: Efficient Bayesian

696 Phylogenetic Inference and Model Choice Across a Large Model Space. *Systematic*
697 *Biology* **61**(**3**):539-542 DOI: 10.1093/sysbio/sys029.
698 **65. Rosenthal J, Forstner MRJ. 2014.** Effects of Plio-Pleistocene barrier on Chihuahuan
699 Desert herpetofauna. In: Hoyt CA, Karges J, eds. *Proceedings of the Sixth Symposium on*
700 *the Natural Resources of the Chihuahuan Desert Region*. 269-282.
701 **66. Ruhe RV. 1962**. Age of the Rio Grande Valley in Southern New Mexico. *The Journal of*
702 *Geology* **70**(**2**):151-167 DOI: 10.1086/626805.
703 **67. Ryder OA. 1986.** Species conservation and systematic: the dilemma of subspecies.
704 *Trends in Ecology and Evolution* **1**(**1**):9-10 DOI: 10.1016/0169-5347(86)90059-5.
705 **68. Sambrook J, Fritsch EF, Maniatis T. 1989.** *Molecular cloning: A laboratory manual*.
706 New York: Cold Spring Harbor Laboratory Press.
707 **69. Satta Y, Klein J, Takahata N. 2000.** DNA Archives and Our Nearest Relative: The
708 Trichotomy Problem Revisited. *Molecular Phylogenetics and Evolution* **14**(**2**):259-275
709 DOI: 10.1006/mpev.2000.0704.
710 **70. Schönhuth S, Doadrio I, Mayden RL. 2007.** A biogeographic perspective on the
711 phylogeny of Mexican cyprinids (Actinopterygii: Cyprinidae). In: Lozano-Vilano ML,
712 Contreras-Balderas AJ, eds. *Studies of North American Desert Fishes in Honor of E. P.*
713 *(Phil) Pister, Conservationist.* Monterrey: Universidad Autónoma de Nuevo León. 102-
714 124.
715 **71. Schönhuth S, Doadrio I, Dominguez-Dominguez O, Hillis DM, Mayden RL. 2008.**
716 Molecular evolution of southern North American Cyprinidae (Actinopterygii), with the
717 description of the new genus Tampichthys from central Mexico. *Molecular Phylogenetics*
718 *and Evolution* **47**(**2**):729-756 DOI: 10.1016/j.ympev.2007.11.036.

719 **72. Schönhuth S, Mayden RL. 2010.** Phylogenetic relationships in the genus *Cyprinella*
720 (Actinopterygii: Cyprinidae) based on mitochondrial and nuclear gene sequences.
721 *Molecular Phylogenetics and Evolution* **55**(**1**):77-98 DOI: 10.1016/j.ympev.2009.10.030.
722 **73. Schönhuth S, Blum MJ, Lozano-Vilano L, Neely DA, Varela-Romero A, Espinosa,**
723 **H, Perdices A, Mayden RL. 2011.** Inter-basin exchange and repeated headwater capture
724 across the Sierra Madre Occidental inferred from the phylogeography of Mexican
725 stonerollers. *Journal of Biogeography* **38**(**7**):1406-1421 DOI: 10.1111/j.1365-
726 2699.2011.02481.x.
727 **74. Schönhuth S, Hillis DM, Neely DA, Lozano-Vilano L, Perdices A, Mayden RL. 2012.**
728 Phylogeny, diversity, and species delimitation of the North American Round-Nosed
729 Minnows (Teleostei: *Dionda*), as inferred from mitochondrial and nuclear DNA
730 sequences. *Molecular Phylogenetics and Evolution* **62**(**1**):427-446 DOI:
731 10.1016/j.ympev.2011.10.011.
732 **75. Schönhuth S, Perdices A, Lozano-Vilano L, García-de-León FJ, Espinosa H,**
733 **Mayden RL. 2014.** Phylogenetic relationships of North American western chubs of the
734 genus *Gila* (Cyprinidae, Teleostei), with emphasis on southern species. *Molecular*
735 *Phylogenetics and Evolution* **70**:210-230 DOI: 10.1016/j.ympev.2013.09.021.
736 **76. Schönhuth S, Lozano-Vilano L, Perdices A, Espinosa H, Mayden RL. 2015.**
737 Phylogeny, genetic diversity and phylogeography of the genus *Codoma* (Teleostei,
738 Cyprinidae). *Zoologica Scripta* **44**(**1**):11-28 DOI: 10.1111/zsc.12083.
739 **77. Scott WB, Crossman EJ. 1998.** *Freshwater Fishes of Canada*. Oakville (Canada): Galt
740 House.

741 **78. Silvestro D, Michalak I. 2012.** RaxmlGUI: A graphical front-end for RAxML.
742 *Organisms Diversity and Evolution* **12**(**4**):335-337 DOI: 10.1007/s13127-011-0056-0.
743 **79. Smith ML, Miller RR. 1986.** The evolution of the Rio Grande basin as inferred from its
744 fish fauna. In: Hocutt CH, Wiley EO, eds. *The Zoogeography of North American*
745 *Freshwaters Fishes*. New York: John Wiley and Sons. 457-485 DOI: 10.1086/415476.
746 **80. Sousa-Santos C, Robalo JI, Collares-Pereira MJ, Almada VC. 2005.** Heterozygous
747 indels as useful tools in the reconstruction of DNA sequences and in the assessment of
748 ploidy level and genomic constitution of hybrid organisms. *DNA Sequence* **16**(**6**):462-467
749 DOI: 10.1080/10425170500356065.
750 **81. Stamatakis A. 2014.** RAxML version 8: a tool for phylogenetic analysis and post-
751 analysis of large phylogenies. *Bioinformatics* **30**(**9**):1312-1313 DOI:
752 10.1093/bioinformatics/btu033.
753 **82. Strain WS. 1966.** Blancan mammalian fauna and Pleistocene formations, Hudspeth
754 County, Texas. *Bulletin of the Texas Memorial Museum* **10**:1-55.
755 **83. Strain WS. 1970.** Bolson integration in the Texas-Chihuahuan border region. In:
756 *Geology of the Southern Quitman Mountains Area, Trans-Pecos, Texas; Symposium and*
757 *Guidebook, 1970 Biennial Field Trip.* Midland, Texas: Society of Economic
758 Paleontologists and Minerologists, Permian Basin Section. 88-90.
759 **84. Tamura K, Nei M. 1993.** Estimation of the number of nucleotide substitutions in the
760 control region of mitochondrial DNA in humans and chimpanzees. *Molecular Biology*
761 *and Evolution* **10**(**3**):512-526 DOI: 10.1093/oxfordjournals.molbev.a040023.

762 **85. Tamura K, Stecher G, Peterson D, Filipski A, Kumar S. 2013.** MEGA6: Molecular
763 Evolutionary Genetics Analysis Version 6.0. *Molecular Biology and Evolution*
764 **30(12)**:2725-2729 DOI: 10.1093/molbev/mst197.
765 **86. Thompson JD, Gibson TJ, Plewniak F, Jeanmougin F, Higgins DG. 1997.** The
766 CLUSTAL_X windows interface: flexible strategies for multiple sequence alignment
767 aided by quality analysis tools. *Nucleic Acids Research* **25(24)**:4876-4882 DOI:
768 10.1093/nar/25.24.4876.
769 **87. Vandermeer HJ. 1966.** Statistical Analysis of Geographic Variation of the Fathead
770 Minnow, *Pimephales promelas. Copeia* **1966(3)**:457-466 DOI: 10.2307/1441065.
771 **88. Vallinoto M, Sequeira F, Sodré D, Bernardi JAR, Sampaio I, Schneider H. 2009.**
772 Phylogeny and biogeography of the *Rhinella marina* species complex (Amphibia,
773 Bufonidae) revisited: implications for Neotropical diversification hypotheses. *Zoologica*
774 *Scripta* **39(2)**:128-140 DOI: 10.1111/j.1463-6409.2009.00415.x.
775 **89. Wood RM, Mayden RL. 2002.** Speciation and anagenesis in the genus *Cyprinella* of
776 Mexico (Teleostei: Cyprinidae): a case study of Model III allopatric speciation. *Reviews*
777 *in Fish Biology and Fisheries* **12(2-3)**:253-271 DOI: 10.1023/A:1025018118339.
778 **90. Zardoya R, Doadrio I. 1999.** Molecular Evidence on the Evolutionary and
779 Biogeographical Patterns of European Cyprinids. *Journal of Molecular Evolution.*
780 **49(2)**:227-237 DOI: 10.1007/PL00006545.

781

782

Figure 1

Basins sampled for *Pimephales promelas* and lineages found using mitochondrial gen cyt *b* and *S7* ribosomal protein gen intron 1.

Bvp, Bavispe River; *BbL*, Babicora Lagoon; *Ppg*, Papigochic River; *BL*, Bustillos Lagoon; *SnP*, San Pedro River; *Bll*, Balleza River. Colors and shapes correspond to the four Sub-clades identified in phylogenetic analyses.

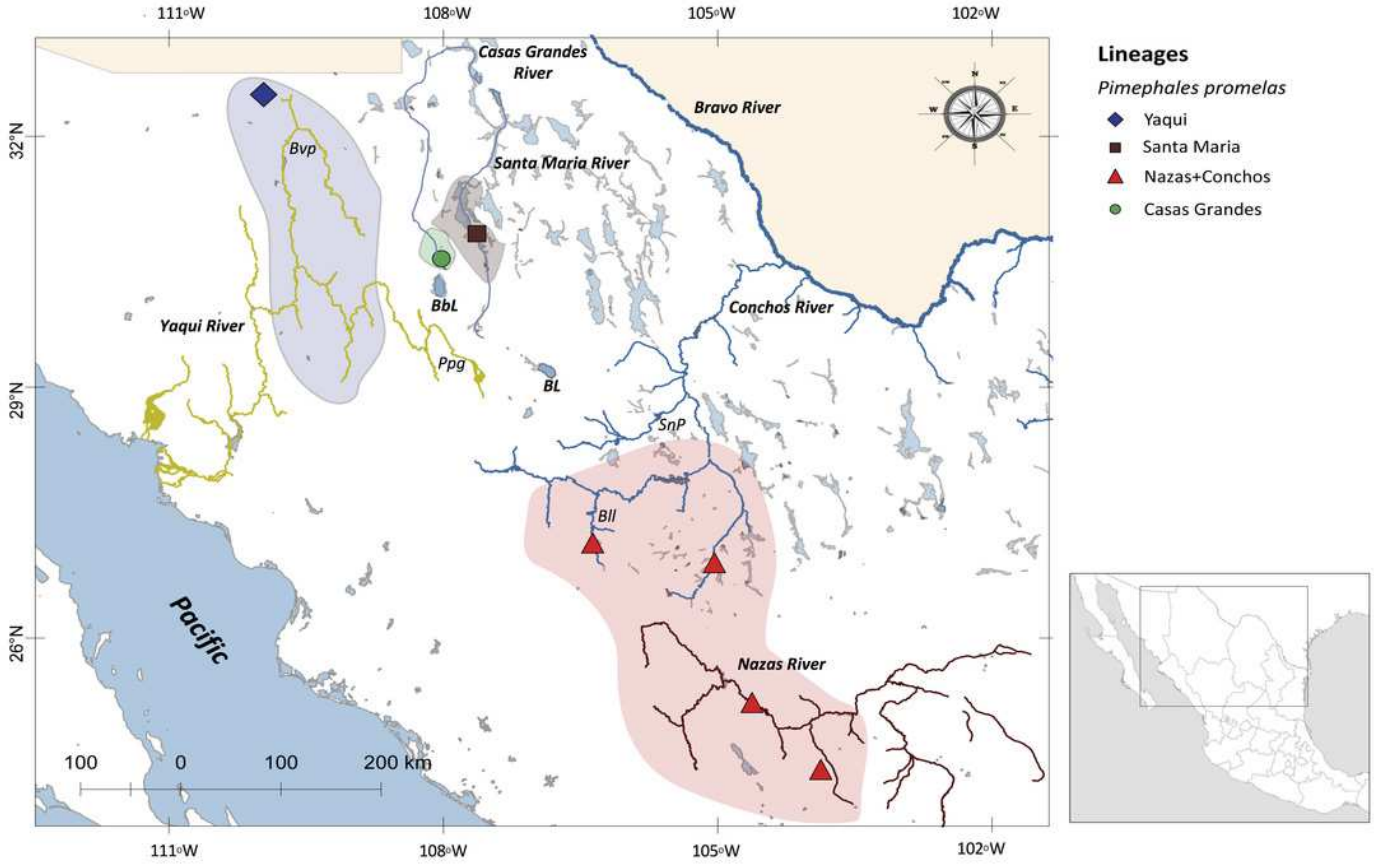

Figure 2

Phylogeny of *Pimephales promelas* based on the concatenated two genes (cyt *b* and *S7*).

Numbers on the branches are Bayesian posterior probabilities (above) and Maximum Likelihood (below). Sub-clades are color-coded to the distribution areas in the map (Fig. 1).

PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.26873v1 | CC BY 4.0 Open Access | rec: 19 Apr 2018, publ: 19 Apr 2018

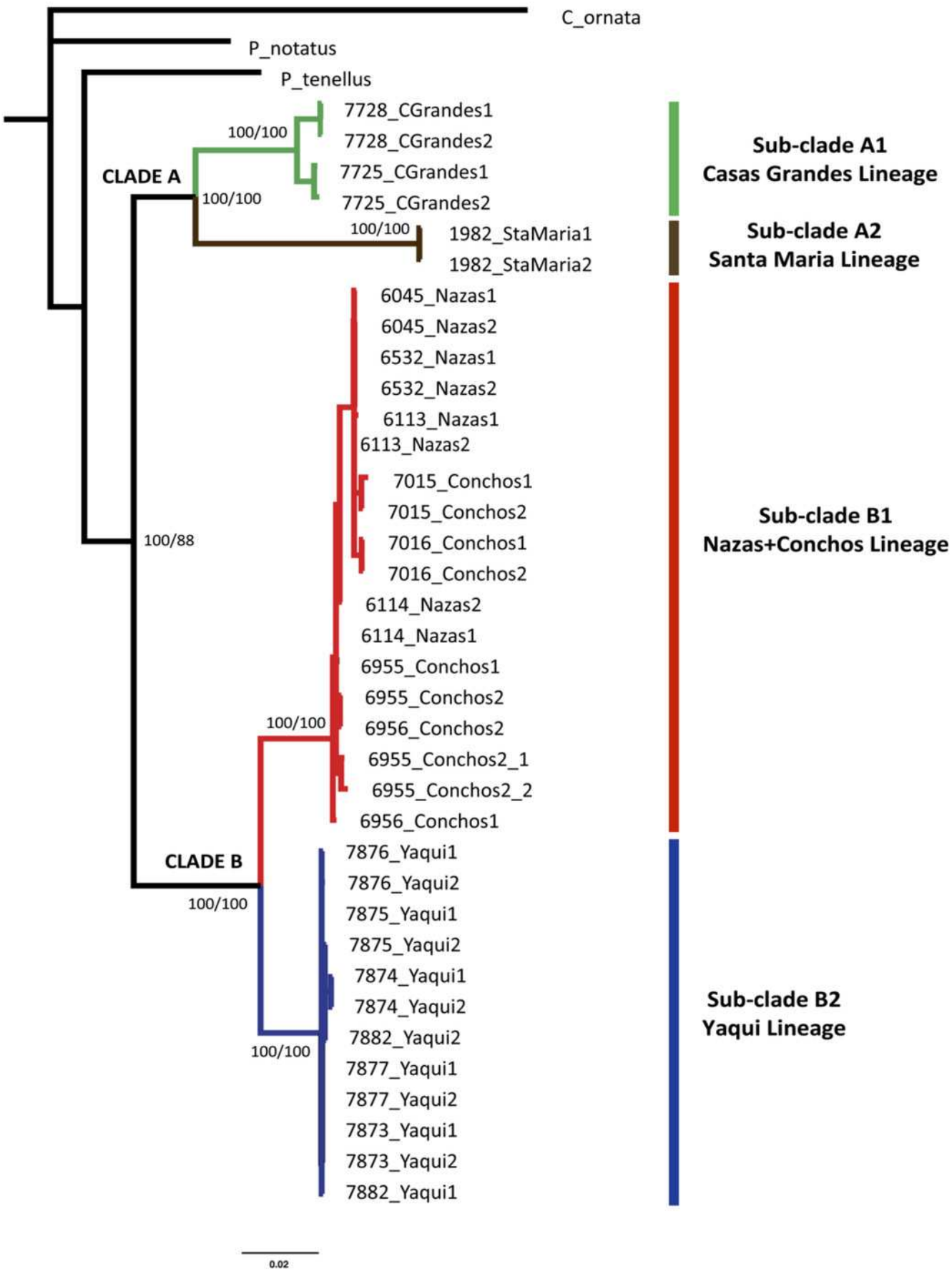

Figure 3

Time-calibrated Species-Tree phylogeny and Divergence Time estimates for nodes based on substitution rates for the cyt *b* gene of 0.76-2.2%/ million of years.

Values in blue color and values in parentheses represent the 95% highest posterior density of divergence time estimates. Values on the branchs represents the highest posterior probability. Lineages are color-coded to the distribution areas in the map (Fig. 1).

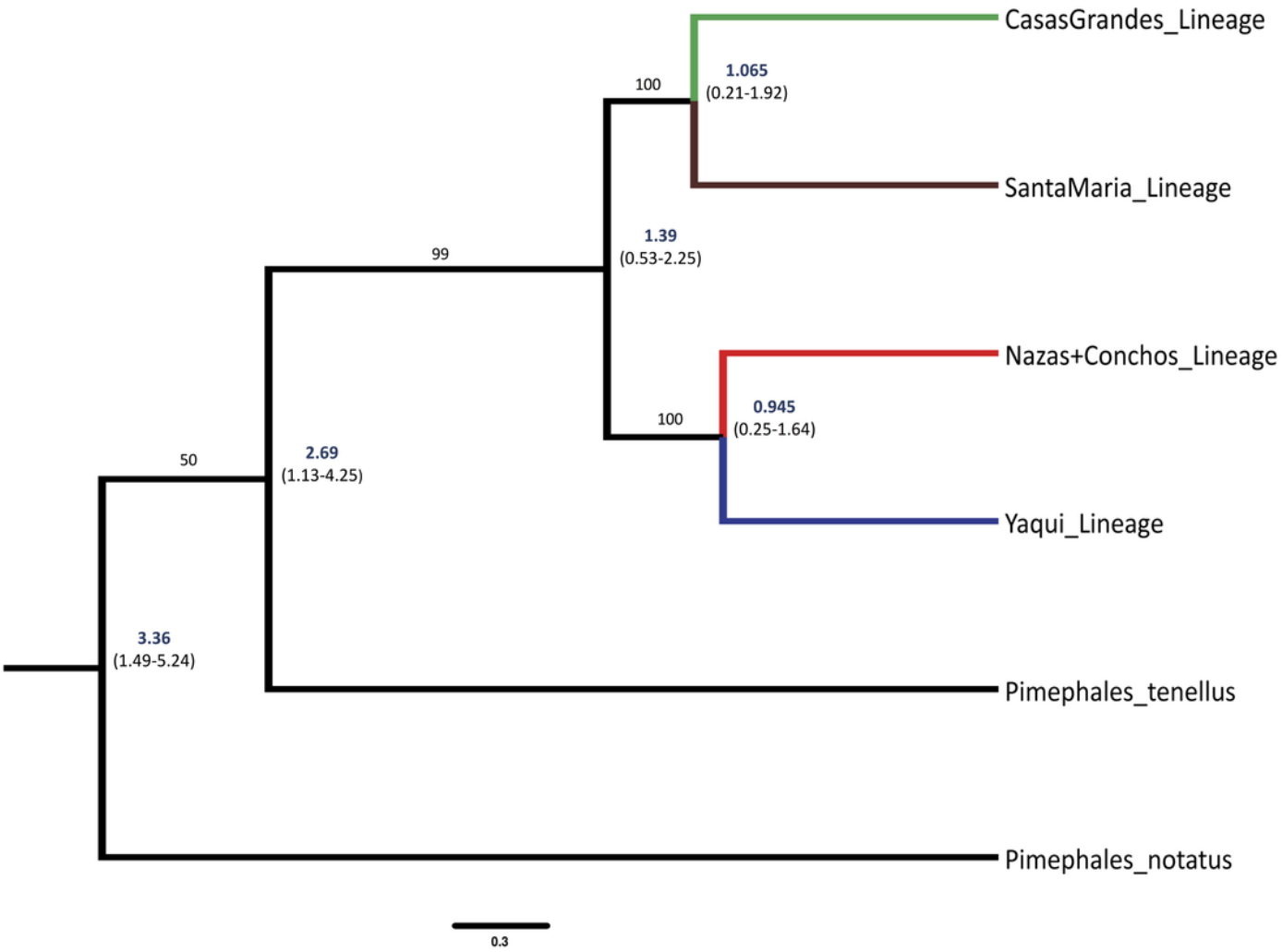

Figure 4

Haplotype network for *Pimephales promelas*.

Median-joining haplotype network for mitochondrial (cyt *b*) and nuclear (intron *S7*) genes for *Pimephales promelas*. Each circle represents a different haplotype; circle sizes are proportional to the number of individuals possessing a particular haplotype. Lineages are represented by different colors and are in accordance with the color of the distribution area in figure 1.

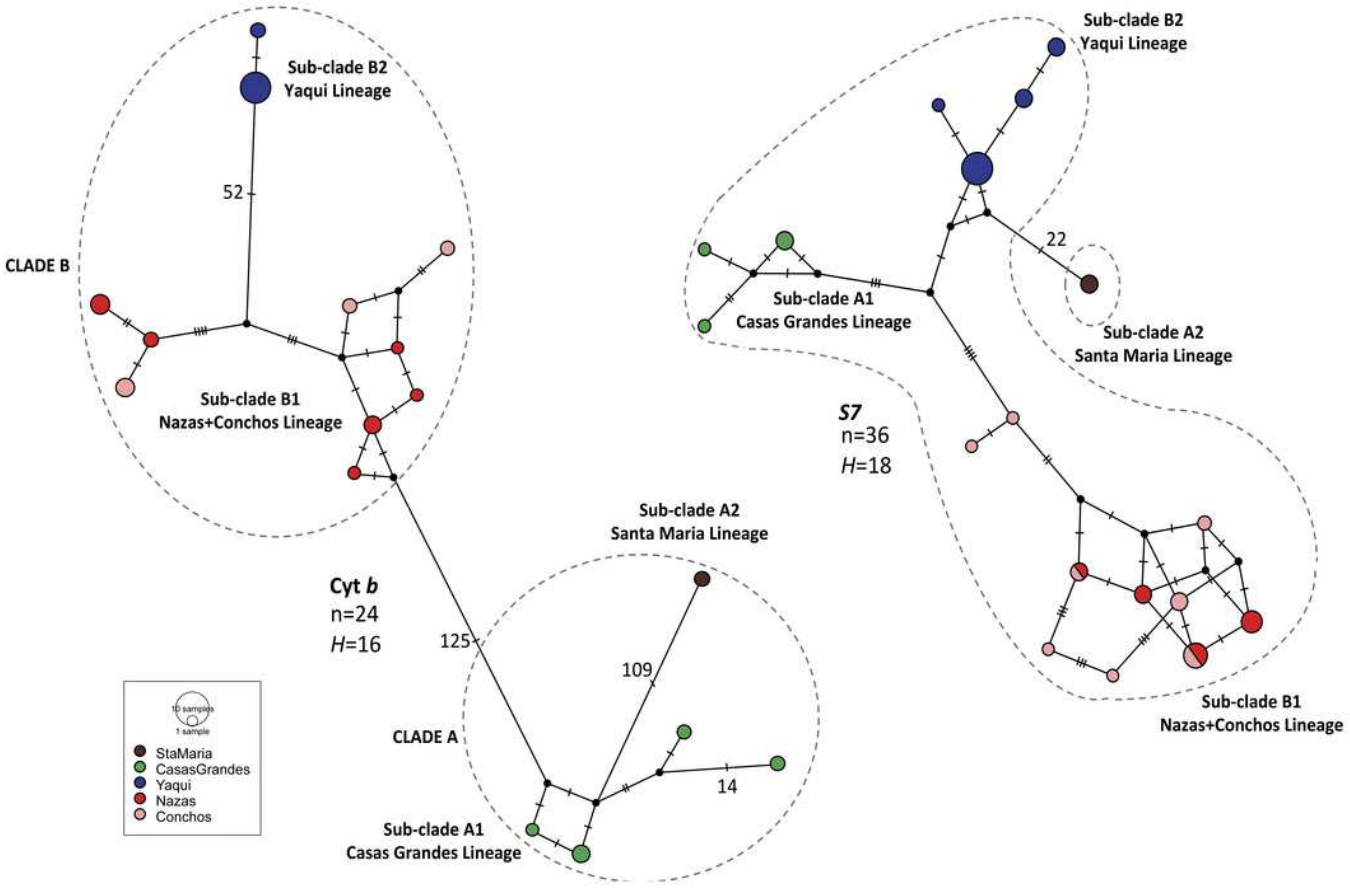

Figure 5

Chronosequence of cladogenetic events, and suggested drainage configurations and distributions of the common ancestor of the four *Pimephales promelas* lineages.

(A) Connection and disconnection between the current Guzman Basin and Yaqui River. *(B)* Separation of rivers within Guzman Basin via Los Muertos Bolson desiccation. *(C)* Recent headwater capture and hydrological connectivity between areas across the SMO (Yaqui-Conchos rivers). *LP*, Lake Palomas; *LCV*, Lake Cabeza de Vaca; *CG*, Casas Grandes River; *SM*, Santa Maria River; *Cm*, Del Carmen River; *Bvp*, Bavispe River; *BbL*, Babicora Lagoon; *Ppg*, Papigochic River; *BL*, Bustillos Lagoon; *Bll*, Balleza River; *SnP*, San Pedro River. Green, purple, red and Brown shaded colors correspond to the distribution of the Sub-clades as shown in the phylogenetic trees. Solid arrows represent expansion range by lake expansion. Doted arrow represent river capture.

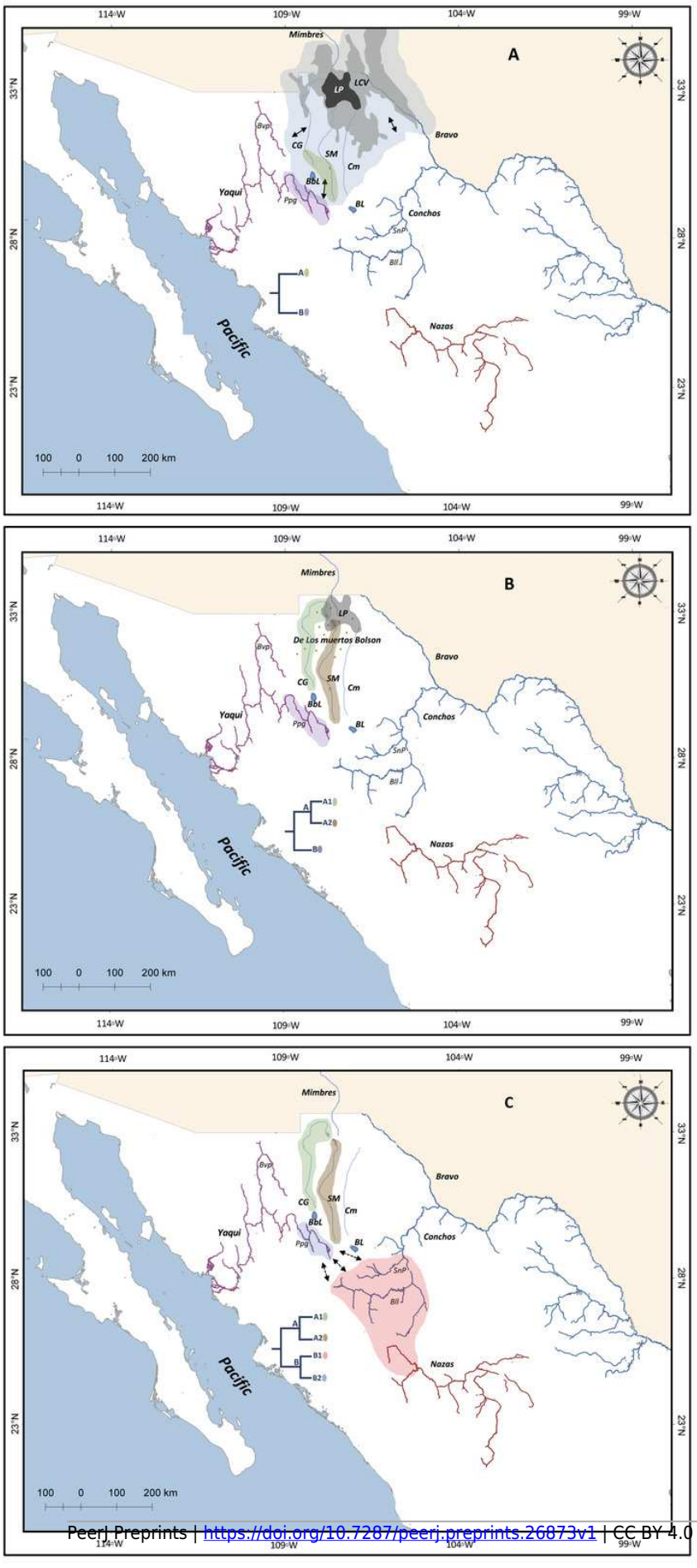

Table 1(on next page)

Genetic divergences within *Pimephales promelas* populations.

Below diagonal cyt *b* and above diagonal *S7* uncorrected pairwise sequence divergence (%). Value inside brackets correspond to the intragroup genetic distance (cyt *b* / *S7*).

1

Lineages	Nazas+Conchos	Yaqui	Casas Grandes	Santa Maria
Nazas+Conchos	(0.4%)/(0.4%)	1.4%	1.7%	4.7%
Yaqui	3.7%	(0%)/(0.1%)	1.1%	3.6%
Casas Grandes	9.4%	9.6%	(0.5%)/(0.3%)	4.4%
Santa Maria	10.6%	10.5%	7.8%	--

2

3

4
5