Genetic differentiation in the southern population of the Fathead Minnow *Pimephales promelas* Rafinesque (Actinopterygii: Cyprinidae)

Nayarit E. Ballesteros-Nova[1], Rodolfo Pérez-Rodríguez[2,3], Rosa G. Beltrán-López[1,4] and Omar Domínguez-Domínguez[2,3]

[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
[3] Laboratorio Nacional de Análisis y Síntesis Ecológica para la Conservación de Recursos Genéticos de México, Escuela Nacional de Estudios Superiores, Unidad Morelia, Universidad Nacional Autónoma de México, Morelia, Michoacán, México
[4] Laboratorio de Ictiología, Centro de Investigaciones Biológicas, Universidad Autónoma del Estado de Morelos, Cuernavaca, Morelos, México

Corresponding author
Omar Domínguez-Domínguez, goodeido@yahoo.com.mx

ABSTRACT

The North American cyprinid *Pimephales promelas* is a species with a wide distribution range, occurring in distinct hydrographic basins in Mexico, Canada, and the United States. Previous morphological and meristic analyses of *P. promelas* concluded that at least three subspecies exist in the midwestern and northeast region of the United States. No studies have been carried out on the Mexican population of *P. promelas*, but the findings of cryptic diversity in United States populations of this species, as well as in other codistributed fish species in Mexico could be an indication that Mexican populations of *P. promelas* consist of cryptic species. Using the mitochondrial gene cyt *b* and the first intron of the *S7* ribosomal protein-coding nuclear gene we carried out phylogenetic and phylogeographic analyses of populations of *P. promelas* across its distribution range in northwestern Mexico. Using this information were analyzed the structure and differentiation level between populations of *P. promelas* from distinct river basins in the region in identifying cryptic diversity. Twenty-four sequences were obtained for cyt *b*, and 30 for *S7*, which included the two heterozygous alleles. The results revealed the existence of four well-differentiated lineages: (1) Yaqui in the Pacific slope; (2) Santa Maria, and (3) Casas Grandes in the Guzman Basin; and (4) Nazas+Conchos in Chihuahua state. This challenges the current taxonomy of *P. promelas*. Differences in the relationships between markers and the small sample size for the Santa Maria population ($n = 1$), indicate that our results must be corroborated with more data and morphological analyses. Biogeographic analysis of these findings suggest that the evolutionary history of *P. promelas* is associated with the fragmentation of the ancestral Rio Grande river system since Miocene in northwestern Mexico consistent with findings for codistributed fish species.

INTRODUCTION

Phylogeographic studies provide useful information that complements biogeographic hypotheses relating to the current distributions of fish species and their link to fragmentation and/or to expansion events (*Devitt, 2006*; *Riddle & Hafner, 2006*; *Vallinoto et al., 2010*; *Schönhuth et al., 2015*).

In North America, tectonic activity and climate changes are the main factors that have contributed to the formation of drainage basins and the evolution of associated species (*Galloway, Whiteaker & Ganey-Curry, 2011*; *Aranda-Gómez et al., 2018*). In northwestern Mexico, the main tecto-volcanic 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 level of tectonic activity associated with the SMO has been proposed as responsible for the high level of endemism of flora and fauna in this region of Mexico, and the main force that has shaped the speciation process (vicariance and/or dispersion) in freshwater fishes of the region (*Echelle et al., 2005*; *Hughes, Rinne & Calamusso, 2005*; *Ceballos, Arroyo-Cabrales & Ponce, 2010*; *Morrone, 2010*; *Schönhuth & Mayden, 2010*; *Domínguez-Domínguez et al., 2011*; *Schönhuth et al., 2011*; *Schönhuth et al., 2012*; *Schönhuth et al., 2014*; *Schönhuth et al., 2015*; *Clements, Bart & Hurley, 2012*). The SMO is also considered an important biogeographic corridor and a Pleistocene refuge (*Hughes, Rinne & Calamusso, 2005*), associated with the expansion and contraction of numerous species ranges in response to climate change during the Pleistocene (*Domínguez-Domínguez et al., 2011*).

Northwestern Mexico is an area including two major physiographic provinces: the highland Chihuahuan desert region in the Mesa del Norte, and the SMO (*Schönhuth et al., 2015*). The Mesa del Norte is located from the United States (US)-Mexico border to Zacatecas and San Luis Potosi states (*Miller, Minkley & Norris, 2005*). These areas include major drainages as Yaqui, Mayo, Fuerte, and Mezquital rivers on the Pacific Slope of the SMO; endorheic drainages as the Nazas and Aguanaval rivers; and the Atlantic drainage Conchos River (*Schönhuth et al., 2015*) (Fig. 1).

Previous studies of fishes in the Mesa del Norte based on molecular data (*Smith & Miller, 1986*; *Mayden, Matson & Hillis, 1992*; *Mayden et al., 1992*; *Echelle et al., 2005*; *Domínguez-Domínguez et al., 2011*; *Schönhuth et al., 2011*; *Schönhuth et al., 2012*; *Schönhuth et al., 2015*; *Corona-Santiago et al., 2018*) and geological information (*Galloway, Whiteaker & Ganey-Curry, 2011*), proposed the existence of an extended Rio Grande system that originated in the Oligocene and persisted for more than 10 million years. Extensive zones of the southwestern highlands of the US (Colorado Plateau, in current Utah, Colorado and Arizona), Chihuahuan Desert (New Mexico, Chihuahua, Coahuila and Durango) and the current Lower Grande River Basin were part of the ancient Rio Grande system (*Galloway, Whiteaker & Ganey-Curry, 2011*; *Schönhuth et al., 2015*).This paleo-river system was proposed to have connected drainages in Mexico that are currently independent (*Mayden, Matson & Hillis, 1992*; *Echelle et al., 2005*; *Schönhuth, Doadrio & Mayden, 2007*; *Schönhuth et al., 2011*; *Schönhuth et al., 2015*; *Domínguez-Domínguez et al., 2011*; *Corona-Santiago et al., 2018*), acting as a hydrological corridor for the ancestors of current freshwater fish
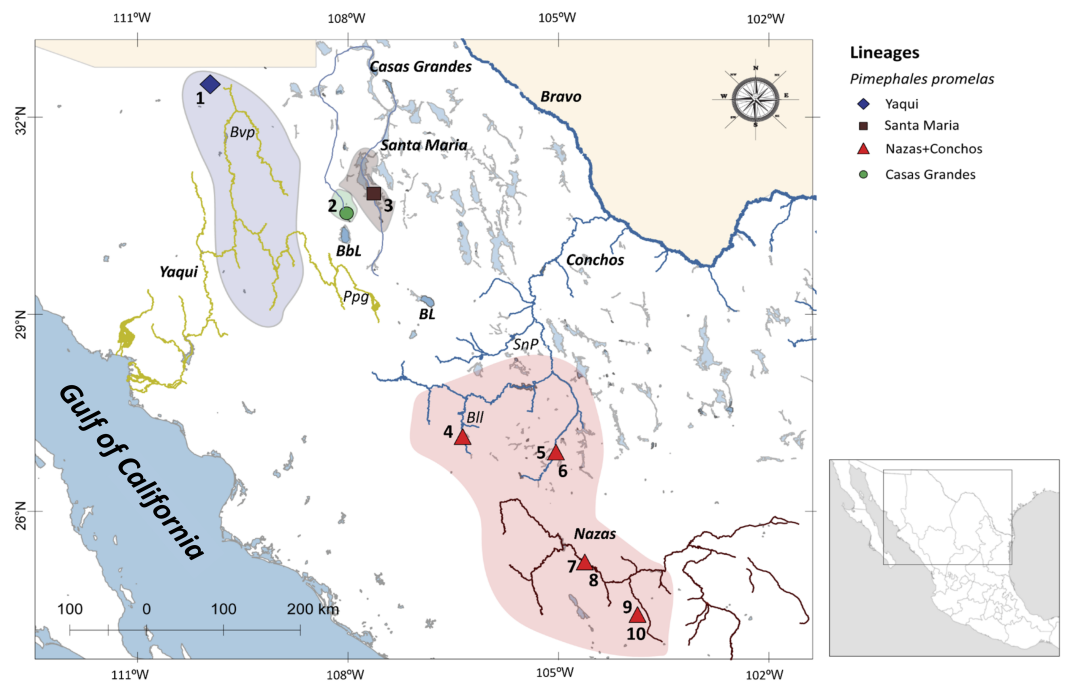
Figure 1 Drainage basins sampled for *P. promelas* and genetic lineages found. Colors and shapes correspond to the four lineages identified in phylogenetic analyses. Numbers and forms indicate different localities where *P. promelas* was collected: 1. Cabullona, (six samples); 2. Casas Grandes, (five samples); 3. Buenaventura, (one sample); 4. Nonoava, (one sample); 5. Villa Coronado, (two samples); 6. Porvenir, (two samples); 7. Jicorica, (two samples); 8. Abasolo, (two samples); 9. Paso Nacional, (two samples); and 10. Covadonga, (two samples), according to the localities in Table S1. Abbreviations: *Bvp*, Bavispe River; *BbL*, Babicora Lagoon; *Ppg*, Papigochic River; *BL*, Bustillos Lagoon; *SnP*, San Pedro River; *Bll*, Balleza River.

species and permitting their expansion across the drainages in the Chihuahuan Desert (*Smith & Miller, 1986*; *Mayden, Matson & Hillis, 1992*; *Domínguez-Domínguez et al., 2011*; *Schönhuth et al., 2011*; *Schönhuth et al., 2015*). This scenario includes the possible transfers and/or integration of headwaters of the Mesa del Norte with drainages from the Pacific slope (Yaqui, Mayo, Fuerte, Piaxtla, Culiacan and Mezquital rivers) (*Minckley, Hendrickson & Bond, 1986*; *Miller, Minkley & Norris, 2005*; *Domínguez-Domínguez et al., 2011*; *Schönhuth et al., 2011*; *Schönhuth et al., 2014*; *Schönhuth et al., 2015*; *Corona-Santiago et al., 2018*) as far south as the Nazas and Aguanaval rivers (*Smith & Miller, 1986*; *Mayden, Matson & Hillis, 1992*; *Domínguez-Domínguez et al., 2011*; *Schönhuth et al., 2011*; *Schönhuth et al., 2014*; *Schönhuth et al., 2015*; *Corona-Santiago et al., 2018*).

One of the most widespread fish species in the Mesa del Norte hydrological basins is the Fathead Minnow *Pimephales promelas*. This species is distributed from eastern North America into Northern Mexico. *P. promelas* ranges from Lake Slave to the Hudson Bay at its northern limit, southward through the Mississippi Valley, the Great Plains and the Gulf slope streams of Alabama and the Grande River basin into the Conchos River. The range of *P. promelas* also includes the endorheic basins of Casas Grandes, Nazas, Del Carmen, Santa Maria, and Bustillos in the Mesa del Norte, as well as the Pacific slope drainage of

the Yaqui River in Mexico (*Vandermeer, 1966*; *Miller, Minkley & Norris, 2005*). Previous morphological analyses of *P. promelas* populations in the United States concluded that at least three subspecies of this widespread species exist in the midwestern and northeast region of the US (*Vandermeer, 1966*). No studies have been carried out in Mexican population, but based on the findings of cryptic diversity in United States populations of this species (*Vandermeer, 1966*), as well as in other codistributed fish species in Mesa del Norte (*Campostoma anomallum* (*Blum et al., 2008*), *Cyprinella lutrensis* (*Schönhuth & Mayden, 2010*), *Campostoma ornatum* (*Domínguez-Domínguez et al., 2011*; *Schönhuth et al., 2011*), *Dionda episcopa* (*Schönhuth et al., 2012*), *Codoma ornata* (*Schönhuth et al., 2015*) and *Pantosteus plebeius* (*Corona-Santiago et al., 2018*) we hypothesize that Mexican populations of *P. promelas* have been taxonomically underestimated.

In accordance with the above, the aim of the present study is to assess the genetic divergences in *P. promelas* populations, to tests the current taxonomic status with molecular data, and determine whether populations from distinct basins correspond to different evolutionary lineages. Based on mitochondrial and nuclear markers, we use both, phylogenetic and phylogeographic approaches to examine the populations of *P. promelas* across its distribution range in northwestern Mexico.

METHODS

Taxon sampling

The specimens of *P. promelas* were collected from five independent drainage basins and ten localities across the Mexican distributional range of the species using electrofishing and seine netting techniques: (1) Cabullona, (6 specimens); (2) Casas Grandes, (5 specimens); (3) Buenaventura, (1 specimen); (4) Nonoava, (1 specimen); (5) Villa Coronado, (2 specimens); (6) Porvenir, (2 specimens); (7) Jicorica, (2 specimens); (8) Abasolo, (2 specimens); (9) Paso Nacional, (2 specimens); and (10) Covadonga, (2 specimens) (Fig. 1; Table S1). For each specimen we fixed a piece of the fin in 95% ethanol for extraction of the DNA, which was then stored at −70 °C. We preserved a maximum of five specimens per site in 5% formalin following the protocols approved by the Mexican Ministry of Environment and Natural Resources (SEMARNAT). We deposited fish and tissue samples in the fish collection of the Universidad Michoacana de San Nicolas de Hidalgo, Mexico (SEMARNAT registration number MICH-PEC-227-07-09). All procedures were reviewed and approved by a committee of Mexican Ministry of Environment and Natural Resources, under collection permit number PPF/DGOPA-362/1. *C. ornata*, *Pimephales notatus*, and *Pimephales tenellus* were used as outgroups (based on prior phylogenetic studies (*Schönhuth et al., 2008*; *Schönhuth et al., 2016*; *Schönhuth et al., 2018*) (Table S1).

DNA extraction, PCR amplification, and sequencing

DNA was extracted using the standard proteinase K/phenol/chloroform protocol (*Sambrook, Fritsch & Maniatis, 1989*). We obtained sequences for a fragment of the mitochondrial cytochrome *b* gene (cyt *b*: 1,049 bp) in twenty-four specimens using the primers LA (5′-GTGACTTGAAAAACCACCGTTG) and HA (3′-CAACGATCTCCGGTTTACAAGAC) (*Dowling et al., 2002*). A subset of eighteen

specimens was selected, accounting for all the variation found for the cyt *b* gene, for amplification of the first intron of the nuclear *S7* ribosomal protein-coding gene (*S7*: 704 bp), with the primers S71F (5′-TGGCCTCTTCCTTGGCCGTC) and S72R (3′-AACTCGTCTGGCTTTTCGCC) (*Chow & Hazama, 1998*).

The final concentrations in each 25 μL polymerase chain reaction (PCR) were: 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 Reaction Buffer and 2.5 mM MgCl$_2$. Thermocycling conditions for 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, 90 s 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, 100 s at 72 °C, and a final step of 10 min extension at 72 °C. All PCR products were purified with ExoSAP-IT^{TM}. The purified PCR products were sent to Macrogen Korea for sequencing.

The 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 using chromatograms. The *S7* sequences were phased with point mutation using DNAsp v5.10 (*Librado & Rozas, 2009*). We evaluated nuclear recombination using the pairwise homoplasy index (PHI) test (*Bruen, Philippe & Bryant, 2006*) as implemented in Splitstree4 (*Huson & Bryant, 2006*). No significant recombination was detected in the nuclear *S7* sequences ($p = 0.3741$). The sequences of *S7* showed heterozygous indels; in this case we performed a manual reconstruction of the two allele phases following the procedure described by *Sousa-Santos et al. (2005)*.

Phylogenetic analyses and haplotype networks

We obtained the evolutionary substitution models, based on the corrected Akaike Information Criterion (AICc), and an optimal partition setting using PartitionFinder v1.1.0 (*Lanfear et al., 2012*). We obtained the optimal partition setting 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 model (*Tamura & Nei, 1993*) + gamma (TrN+G) for the *S7* gene.

Bayesian Inference (BI) analyses were applied in MrBayes v3.2.1 (*Ronquist et al., 2012*). The sequences were analyzed in two different data sets, one for each gene independently and one for both genes concatenated. We used the two heterozygous alleles for the *S7* gene. The analyses for 10 million generations were run with two independent runs, implementing four Markov Chain Monte Carlo (MCMC) processes and sampling every 500 generations. We evaluated the convergence of the chains with the log-likelihood (-InL) values of the two independent runs on Tracer v1.5 (*Rambaut & Drummond, 2007*), discarding 10% of the generations as burn-in to construct the consensus tree. We visualized the trees in FigTree v1.4.2 (*Rambaut, 2014*).

Maximum Likelihood (ML) 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+I model using the rapid bootstrap algorithm

with 1000 replicates (CIPRES portal v3.3) at the San Diego Supercomputer Center at
http://www.phylo.org/sub_sections/portal/.

For each gene, we constructed unrooted networks under a null hypothesis of no genetic differentiation among populations, using the median-joining method (*Leigh & Bryant, 2015*) as implemented in PopART v1.7 (available at http://popart.otago.ac.nz).

Species tree analysis, genetic distances and divergence times

We estimated a species tree and divergence times for the major nodes in *Pimephales promelas* for both genes (cyt *b* + *S7*) using the Bayesian Method in *BEAST v1.8.1 (*Drummond et al., 2012*). The substitution models were set according to the model selected for each gene by PartitionFinder v1.1.0 (*Lanfear et al., 2012*), the Generalised Time Reversible model (*Tavare, 1986*) + gamma (GTR+G) for the cyt *b* gene and the TN93 model (*Tamura & Nei, 1993*) + gamma (TN93+G) for the individual *S7* gene. We carried out this analysis with a subset of 33 sequences that include all different haplotypes for all genes, and two outgroup sequences (*P. notatus* and *P. tenellus* based on a previous phylogenetic study (*Schönhuth et al., 2018*)) (Table S1). The lineages in the *BEAST analysis were selected according to the lineages recovered on the phylogenetic trees. The model parameters were unlinked across cyt *b* and *S7* genes. Considering that the performance of the strict clock is virtually identical to the lognormal distribution, of the uncorrelated relaxed clock (*Firth et al., 2010*), and the use of uncorrelated relaxed clocks takes the rate variation among lineages into account (*Baele et al., 2012*; *Hasegawa, Kishino & Yano, 1985*), we selected a lognormal relaxed clock (Uncorrelated) model for branch length (*Drummond et al., 2006*). Because of the lack of reliable fossil data, we calibrated the molecular clock using the mutation rate of cyt *b* in teleosts of 0.76–2.2%/million years (*Zardoya & Doadrio, 1999*; *Berendzen, Gamble & Simons, 2008*; *Nagle & Simons, 2012*), applied in a prior with a uniform distribution. Since the mutation rate is not available for the nuclear gene, we included this gene in the analysis without calibration information. We selected the tree prior-species Tree: Yule process model. We ran Markov chain Monte Carlo analysis for 120 million generations, sampled every 1000 generations. We ran analyses in CIPRES Science Gateway v3.3 (http://www.phylo.org/sub_sections/portal/). We assessed whether parameter values had reached effective sample size and convergence in Tracer v1.5 (*Rambaut & Drummond, 2007*), and built the maximum clade credibility tree using Tree Annotator v1.8.1 (*Drummond et al., 2012*), discarding the first 10% of the trees as burn-in. We visualized the tree in FigTree v1.4.2 (*Rambaut, 2014*).

Previous phylogenetic studies in freshwater fishes in northwestern Mexico use uncorrected pairwise (*p*)-distances for generating distance matrices (*Domínguez-Domínguez et al., 2011*; *Schönhuth et al., 2011*; *Schönhuth et al., 2014*; *Schönhuth et al., 2015*; *Corona-Santiago et al., 2018*). In order to be able to compare our results we used *p*-distances, calculated among the recovered monophyletic groups in the phylogenetic trees for both genes independently in Mega v6.06 (*Tamura et al., 2013*).

Bayesian species delimitation test

We conducted Bayesian multilocus species delimitation tests using the concatenated dataset in Bayesian Phylogenetics and Phylogeography (BPP v3.1; *Yang & Rannala, 2010*; *Yang &*

Rannala, 2014; *Yang, 2015*). This method uses a species phylogeny represented by a user-specified guide tree and accommodates lineage sorting due to ancestral polymorphism (*Yang & Rannala, 2010*). To generate the guide tree, each lineage recovered in the concatenated phylogenetic analyses (BI, ML, and species tree) was treated as a terminal taxon in *BEAST (*Drummond et al., 2012*), with the resulting species tree used as a guide for the BPP analyses.

For the BPP analyses of the two concatenated genes, we used the reversible-jump Markov Chain Monte Carlo (rjMCMC) (*Yang & Rannala, 2010*) algorithm to determine whether to collapse or retain nodes throughout the phylogeny. Using the entire dataset coded by each gene, we tested with the Analysis A10 algorithm, in which we used the rjMCMC algorithm to move between species delimitation models that were compatible with a fixed guide tree (*Rannala & Yang, 2013*; *Yang & Rannala, 2010*).

To determine whether lineages could be considered as distinct species under a general lineage species concept, the program assessed the probability of the node separating the species (*De Queiroz, 2007*). We used algorithm 0 with values of 5, 10 and 15 for the fine-tuning parameter in order to ensure that the rjMCMC mixed effectively in the species delimitation models. We conducted analyses with prior distributions on the ancestral population size (θ) and root age (τ_0) (*Leaché & Fujita, 2010*) to discern how these parameters influenced the results. We initially set the gamma prior at θ and τ_0 to the values $\alpha = 1$ and 2, and $\beta = 10$, 100, and 2000; and ran five analyses of each with different starting seeds for two independent chains of 500,000 generations with a burn-in of 50,000 and thinning conducted every five generations.

RESULTS

We obtained 54 sequences for both genes: 24 for cyt *b*, and 30 for *S7*, which included the two heterozygous alleles.

We obtained 16 haplotypes 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.

We obtained 18 haplotypes 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 and haplotype networks

Phylogenetic analyses based on the two concatenated genes (cyt *b* +*S7*) (1,753 bp) recovered four well-supported lineages: Santa Maria lineage, clustered one specimen from the Santa Maria River in the Guzman Basin (BS = 100% and PP = 1); Casas Grandes lineage, clustered two specimens from the Casas Grandes River in the Guzman Basin (BS = 100% and PP = 1); Yaqui lineage clustered six specimens from the Yaqui River in the Cabullona locality (BS = 100% and PP = 1); and, Nazas+Conchos lineage, clustered eight specimens from the Nazas and Conchos rivers (BS = 100% and PP = 1) (Fig. 2; Table S1).
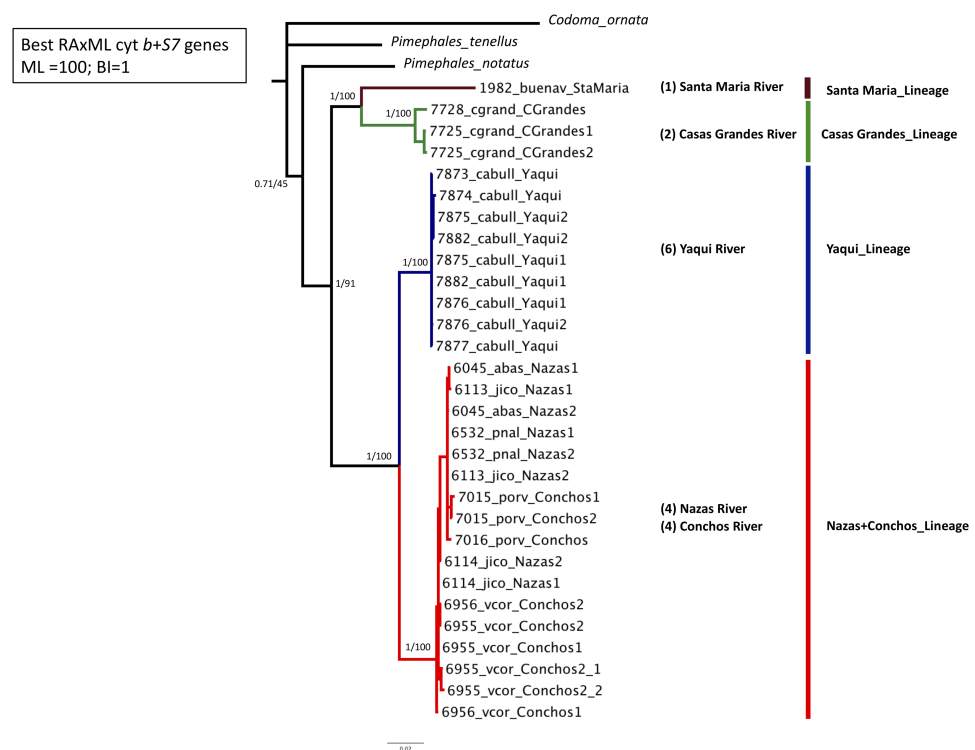
Figure 2 **Maximum Likelihood (ML) phylogeny of *P. promelas* based on the concatenated genes (cyt *b* + *S7*), using a GTR+G+I model with ML bootstrap values (based on 1,000 replicates).** Numbers on the branches separated by a diagonal correspond to Bayesian posterior probabilities and Maximum Likelihood bootstrap values. Numbers in parentheses correspond to the sample size in each drainage basin. Lineages are color-coded according to the distribution areas in the map (Fig. 1).

Phylogenies based on separate analyses for each gene revealed generally consistent results, although they recovered a differing number of lineages. Phylogenies based on cyt *b* and concatenated analyses recovered four major lineages and the same phylogenetic relationships: Santa Maria lineage (BS = 100% and PP = 1); Casas Grandes lineage (BS = 100% and PP = 1); Yaqui lineage (BS = 100% and PP = 1); and Nazas+Conchos lineage, included specimens from localities of the Lower Conchos Drainage (Villa Coronado, El Porvenir and Nonoava), as well as specimens from localities of the Nazas River drainage (Covadonga, Abasolo, Paso Nacional and Jicorica) (BS = 97% and PP = 1) (Fig. 1 and Fig. S1; Table S1). The phylogeny recovered with *S7* identified similar lineages: Casas Grandes, BS = 98% and PP = 1; Santa Maria, BS = 96% and PP = 1; Nazas+Conchos, BS = 97% and PP = 1; and Yaqui, BS = 48% and PP = 0.66. In the *S7* analysis the Santa Maria lineage was recovered as a sister group to a well-supported clade including all three lineages, Casas Grandes, Nazas+Conchos and Yaqui (Fig. S2).

For both genes, the haplotype networks showed the existence of four haplogroups, corresponding to the four lineages found in the phylogenetic analyses: Casas Grandes, Santa Maria, Nazas+Conchos and Yaqui lineages; however, as in the phylogenetic analyses, different relationships were found between the cyt *b* and *S7* groups (Fig. 3).
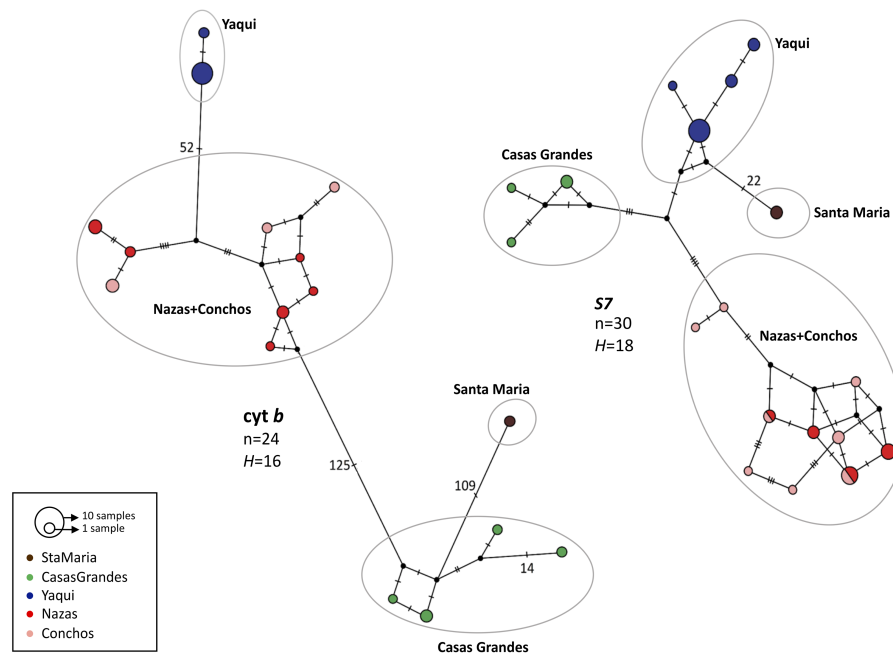
Figure 3 Median-joining haplotype network for mitochondrial (cyt *b*) and nuclear (intron *S7*) genes for *P. promelas*. Each circle represents a different haplotype; circle sizes are proportional to the number of individuals with a particular haplotype. Small black dots represent missing (unsampled or extinct) haplotypes. Lines between circles represent one mutational step, and numbers presented are the number of mutations between haplotypes. Haplogroups are color-coded according to the distribution areas in the map (Fig. 1).

For cyt *b* we found 52 mutation steps (MS) between Yaqui and Nazas+Conchos lineages, 109 MS between Casas Grandes and Santa Maria lineages, and 125 MS between Nazas+Conchos and Casas Grandes lineages. For the Nazas+Conchos population we found nine haplotypes: three in Conchos Basin (Villa Coronado, Nonoava and El Porvenir localities), six in Nazas Basin, one in Covadonga and Paso Nacional localities and two in each of the Jicorica and Abasolo localities. For the Yaqui population we found two haplotypes in the Cabullona locality. In the Casas Grandes population we found four haplotypes. In the Santa Maria population we found one haplotype (Fig. 3).

For *S7* we found a mixture of haplotypes among the Nazas and Conchos samples. We found the Santa Maria lineage to be separated by 22 MS from the Yaqui lineage. For the Nazas+Conchos population we found ten haplotypes: six in Conchos Basin (two in the El Porvenir locality and four in the Villa Coronado locality), two in Nazas Basin (Abasolo and Jicorica localities). For the Yaqui population we found four haplotypes in the Cabullona locality. For the Casas Grandes population we found three haplotypes (Fig. 3).

Species tree analysis, divergence times and genetic distances

The species tree analysis with the concatenated dataset (cyt *b* +*S7)* supports the assumption of four lineages of high internal branch-length Bayesian Posterior Probability (>99%), as seen in the cyt *b* and the concatenated (cyt *b* +*S7*) phylogenetic trees. This result is also
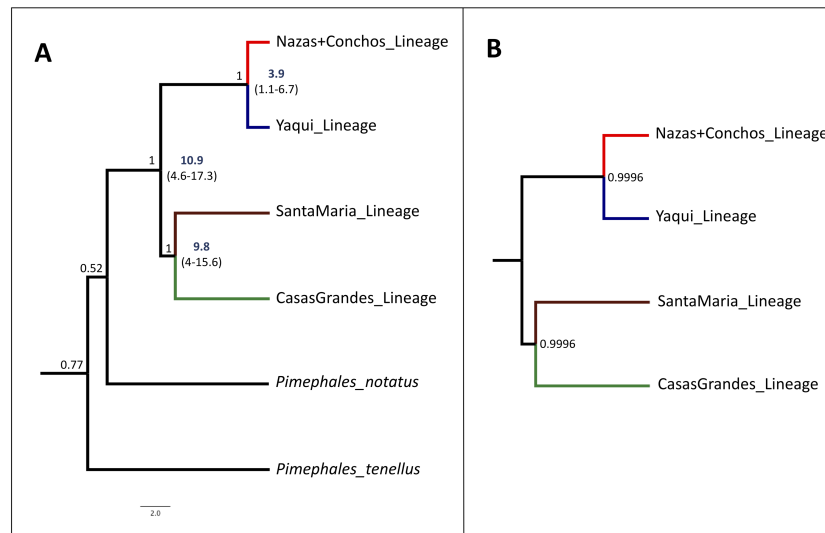
Figure 4 Time-Calibrated Species-Tree and Bayesian species delimitation test results for *P. promelas*. Both nuclear and mitochondrial gene sequences were used. Lineages are color-coded according to the distribution areas in the map (Fig. 1). *(A)* Time-calibrated Species-Tree phylogeny and Divergence Time estimates for nodes based on substitution rates for the cyt *b* gene of 0.76–2.2%/ million of years (*Zardoya & Doadrio, 1999*; *Berendzen, Gamble & Simons, 2008*; *Nagle & Simons, 2012*). Since the mutation rate is not available for the nuclear gene we included in the analysis without calibration information. Numbers in bold and values in parentheses represent the 95% highest posterior density of divergence time estimates. Values on the branches represent the posterior probability. *(B)* Bayesian species delimitation test, assuming four species according to the guide tree obtained from *BEAST. We considered speciation probability values > 0.95 as strong support for a speciation event. For the five analyses, we applied different combinations of θ and t_0 priors: 1. $\theta = (\alpha : 2, \beta : 2000)$, $t_0 = (\alpha : 2, \beta : 20000)$; 2. $\theta = (\alpha : 2, \beta : 20)$, $t_0 = (\alpha : 2, \beta : 200)$; 3. $\theta = (\alpha : 2, \beta : 20)$, $t_0 = (\alpha : 2, \beta : 20000)$; 4. $\theta = (\alpha : 2, \beta : 200)$, $t_0 = (\alpha : 2, \beta : 20000)$ and, 5. $\theta = (\alpha : 2, \beta : 20)$, $t_0 = (\alpha : 2, \beta : 200)$. We obtained strong support (posterior probability ≥ 0.99) in all the five analyses. The number in the nodes correspond to the major posterior probability value obtained in all the five analyses.

consistent with the *S7* gene phylogenetic analyses, corresponding to the four recovered lineages (Fig. 4A).

The divergence between the main clades of *P. promelas* was dated to the Mid-Miocene and Mid-Pliocene. The separation of the main clades (Nazas+Conchos and Yaqui *vs.* Santa Maria and Casas Grandes) was dated at *ca.* 10.9 Million years ago (Mya) (95% Highest Posterior Density (HPD): 4.6–17.3) (Fig. 4A). The split between the Santa Maria and Casas Grandes lineages was estimated at *ca.* 9.8 Mya (95% HPD: 4–15.6), whereas the separation event of the Yaqui and Nazas+Conchos lineages was estimated at *ca.* 3.9 Mya (95% HPD: 1.1–6.7) (Fig. 4A).

The maximum genetic distance for both genes occurred between the Nazas+Conchos and Santa Maria lineages (10.6% for cyt *b* and 4.7% for *S7*) (Tables 1A and 1B). The minimum genetic distance for the cyt *b* gene was found between the Nazas+Conchos and Yaqui lineages (3.7%). For *S7* the minimum genetic distance occurred between the Casas Grandes and Yaqui lineages (1.1%). Genetic distances within each lineage ranged from 0% to 0.5% with cyt *b* and 0.1 to 0.4 with *S7* (Tables 1A and 1B).

Table 1 Genetic divergences within *P. promelas*. Ranges of uncorrected pairwise sequence divergence (%) between major lineages within Pimephales promelas. Values in parentheses correspond to the sample size in each lineage. Values in bold correspond to the genetic distance within each lineage. (A) Genetic divergences in mitochondrial cyt *b* gene. (B) Genetic divergences in the first intron of the S7 ribosomal protein-coding gene.

(A)

Lineages	Nazas+Conchos (13)	Yaqui (6)	Casas Grandes (5)	Santa Maria (1)
Nazas+Conchos	**0.4%**			
Yaqui	3.7%	**0%**		
Casas Grandes	9.4%	9.6%	**0.5%**	
Santa Maria	10.6%	10.5%	7.8%	–

(B)

Lineages	Nazas+Conchos (8)	Yaqui (6)	Casas Grandes (2)	Santa Maria (1)
Nazas+Conchos	**0.4%**			
Yaqui	1.4%	**0.1%**		
Casas Grandes	1.7%	1.1%	**0.3%**	
Santa Maria	4.7%	3.6%	4.4%	–

Bayesian species delimitation test

The speciation model based on the species tree estimate strongly supported the assumption of four putative species. We obtained strong support (posterior probability ≥ 0.99) for the tested speciation model of four *a priori* defined species within Mexican populations of *P. promelas* (Casas Grandes lineage, Santa Maria lineage, Yaqui lineage, and Nazas+Conchos lineage (Fig. 4B)). The BPP was not sensitive to species delimitation, and no alteration of posterior probabilities of the speciation model was observed when we applied different values of root age (τ_0) and population size (θ), demonstrating high posterior probabilities for model tested with the A10 algorithm.

DISCUSSION

The results presented herein revealed that southern populations of *P. promelas* distributed across northwestern Mexico correspond to at least four well-supported independent evolutionary lineages. The four lineages show geographic congruence, with the occurrence of each divergent lineage in independent hydrographic basins, as have been found in other widely distributed fish species (*Blum et al., 2008*; *Schönhuth & Mayden, 2010*; *Domínguez-Domínguez et al., 2011*; *Schönhuth et al., 2011*; *Schönhuth et al., 2012*; *Schönhuth et al., 2014*; *Schönhuth et al., 2015*; *Corona-Santiago et al., 2018*). This study is consistent with other studies of the freshwater fish species distributed along the Mesa del Norte in Mexico that have found that many species previously considered as a single widespread taxon, are in fact a species complex (e.g., *C. ornatum* (*Domínguez-Domínguez et al., 2011*; *Schönhuth et al., 2011*), and *C. ornata* (*Schönhuth et al., 2015*)). This study is also consistent with other studies of *P. promelas*, such as those carried out on the US populations by *Vandermeer (1966)* with morphological characters, that show the species to be made up of complexes of cryptic species.

Phylogenetic relationships and taxonomic implications

In this study of Mexican *P. promelas* populations the results of the concatenated phylogenetic trees (Fig. 2), species tree (Fig. 4A) and BPP analyses (Fig. 4B), supports a strong genetic differentiation pattern, that corresponds to four well-supported lineages, that must be considered independent evolutionary lineages and even undescribed species (Fig. 4). These four evolutionary lineages were recovered in both genes in spite of the incongruence in the phylogenetic relationships between markers (Figs. S1 and S2) and the small sample size in the case of the Santa Maria Basin ($n = 1$) population. The incongruence found between concatenated and independent genes genealogies could not be attributed to retention of ancestral polymorphisms due to incomplete lineage sorting and/or introgression following secondary contact, since we did not recover shared haplotypes between the four well-differentiated lineages in either of the two genes (Fig. 3, Figs. S1 and S2). But the low sample size in the Santa Maria Basin ($n = 1$) could generate gene trees bias results, (*Satta, Klein & Takahata, 2000*; *Kopp & True, 2002*; *Rokas et al., 2003*; *Rokas & Carroll, 2005*), accordingly we recommend a revaluation of the phylogenetic relationships with a larger sample size and more nuclear loci.

In the case of the genetic divergences, we also found congruences in the differentiation pattern of highly differentiated lineages. For cyt *b* among the four lineages the genetic divergences ranged between 3.7% among the Yaqui and Nazas+Conchos lineages, to 10.6% among populations of the Nazas+Conchos and Santa Maria lineages. These genetic distances are the same or greater than those found among the six species of the Chihuahuan Desert Group of the genus *Gila,* that range from 3.68 to 5.56 (*Schönhuth et al., 2014*) and minimum distances of 1.3% have been found between southwestern species in the *Dionda* genus (*Schönhuth et al., 2008*), 1.9% between species in *Cyprinella*, and 4.7% in species of the genus *Tampichthys* (*Schönhuth & Mayden, 2010*) and 2.1–4.0% between species of *Algansea* (*Pérez-Rodríguez et al., 2009*). Moreover, the *S7* gene showed relatively high genetic differences for a nuclear gene, with minimum genetic distances (*p*-distance 1.1%) found in the comparison between the geographically proximate Casas Grandes and Yaqui lineages, and maximum distances (*p*-distance 4.7%) in the Santa Maria and Nazas+Conchos lineages. A previous study has found similar genetic distances in the *S7* nuclear gene for a species in the Cyprinidae family and in species of the genus *Algansea*, in which interspecific genetic distances between 0.5 to 3.6% were found for *S7* (*Pérez-Rodríguez et al., 2009*).

Accordingly, the Mexican populations of *P. promelas* shows four independent evolutionary lineages that could be recognized as different species. A more integrative taxonomic analysis is pending in order to complement the results presented herein.

Biogeographic implications derived from the cladogenetic pattern

Widespread species are associated with suitable biological and ecological traits that permit a high dispersal ability and an ability to colonize new habitats (*Nathan, 2001*). In the case of freshwater fishes, the discontinuity of aquatic habitats represents the main barrier to range expansion (*Pérez-Rodríguez et al., 2015*). In the case of northwestern Mexico, tecto-volcanic events since the Miocene and Pleistocene, glacial/interglacial cycles, and increasing regional aridity since the Holocene, fragmented the ancestral Rio Grande system

(*Metcalfe, 2006*; *Schönhuth et al., 2015*), effectively isolating different populations and shaping 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*; *Schönhuth et al., 2012*; *Schönhuth et al., 2014*; *Schönhuth et al., 2015*; *Aranda-Gómez et al., 2018*; *Corona-Santiago et al., 2018*).

Our divergences times are in agreement with the paleohydrological history of the region in space and time. Accordingly, in the present study, the most recent common ancestor of the four *P. promelas* lineages and the separation of the Santa Maria and Casas Grandes lineages were dated to the Mid-Miocene (*ca.* 10.9 Mya, 95% HPD: 4.6–17.3; *ca.* 9.8 Mya, 95% HPD: 4.0–15.6 respectively) (Fig. 4A). The most plausible biogeographic scenario for the isolation of these lineages is the fragmentation of the ancestral Rio Grande system (*Smith, Song & Miller, 1984*; *Albritton, 1958*; *Miller & Smith, 1986*; *Minckley, Hendrickson & Bond, 1986*; *Smith & Miller, 1986*; *Schönhuth et al., 2015*), caused by the regional patterns of uplift and subsidence during the Miocene and Pliocene, while an arid climate extended across the western interior, causing the reorganization of drainage configurations (*Galloway, Whiteaker & Ganey-Curry, 2011*; *Schönhuth et al., 2015*) (Fig. 5A). This pattern of isolation has also been proposed for codistributed species such as the genus *Tampichthys* (endemic to central-east Mexico) and its sister group *Codoma* (north-western distribution) occurring during the Mid-Miocene, around 9.17 Mya or 12.19 Mya (*Schönhuth et al., 2015*).

The separation between the Yaqui and the Nazas+Conchos lineages was dated to the Pliocene, approximately 3.9 Mya (95% HPD: 1.1–6.7) (Fig. 4A). This event seems to be related to the tecto- volcanic episodes in SMO evolution, including repeat alkaline basalt events (*Henry & Aranda-Gómez, 2000*; *Aranda-Gómez et al., 2005*; *Ferrari, Valencia-Moreno & Bryan, 2007*), river capture and peripheral isolation (*Ferrari, Valencia-Moreno & Bryan, 2007*; *Aguirre-Díaz et al., 2008*) (Fig. 5B), as has been hypothesized for those Pacific slope rivers that have headwaters extending eastward in the SMO to areas of the Mesa del Norte. Previous studies identified the possible transfers of fish species between drainage basins by integration or fragmentation of drainages, as is the case of the Yaqui River, and have documented closely related lineages on both sides of the SMO, in the Yaqui and Conchos rivers. *Smith & Miller (1986)* suggested that rivers in Northwestern Mexico, draining to the Pacific (Upper Yaqui and Upper Mezquital) across desert regions, originated through headwater capture from the ancestral and extant Rio Grande system, and caused fish dispersal by stream capture events (*Schönhuth et al., 2011*).

A mixture of samples from the current Nazas and Conchos drainage basins was found within the Nazas+Conchos lineage 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 to the low mutation rate of nDNA. The close relationship between samples of both drainages was previously found in *C. ornatum* populations, a relationship explained by a fish interchange as a result of river capture (*Schönhuth et al., 2011*). Similarly, *Burr (1976),* in a previous review of *C. ornatum*, hypothesized the formation of a connection between the Nazas River and the Grande River via the Mayran
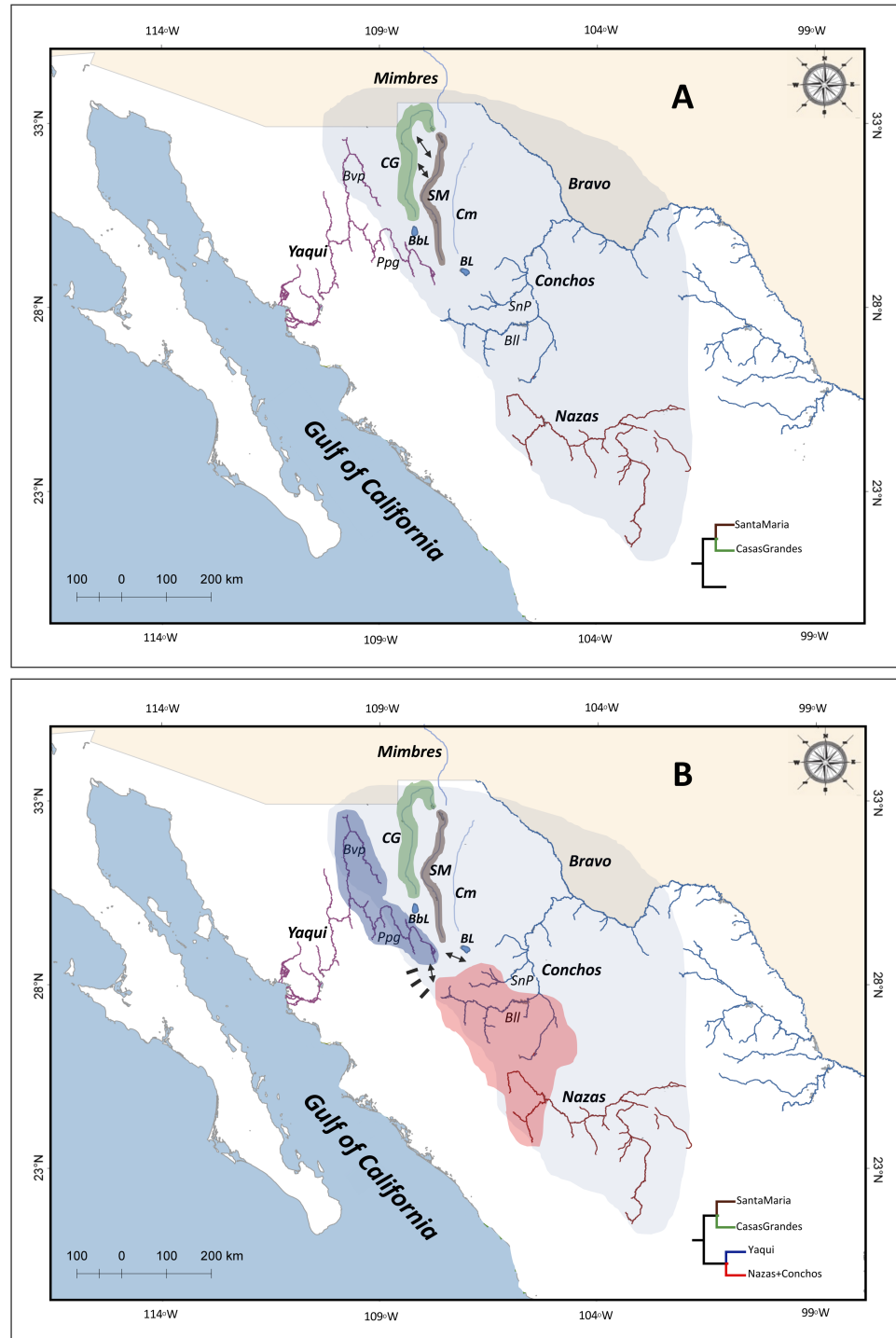
Figure 5 **Chronosequence of cladogenetic events, and suggested drainage basin configurations and distributions of the common ancestor of the four *P. promelas* lineages.** (A) Ancient 'Rio Grande system'; separation of the Santa Maria Lineage, related to the regional patterns of uplift and subsidence during the Miocene and Pliocene. (B) Hydrological connectivity between areas across the SMO (Yaqui-Conchos rivers); separation between Yaqui *vs.*

Figure 5 (…continued)
Nazas+Conchos lineages, related to the tecto-volcanic episodes in SMO evolution. Green, brown, blue and red shaded correspond to the distribution of the lineages as shown in the phylogenetic trees. Black rectangles represent tecto-volcanic events. Solid arrows represent river capture. The colored shade of light blue represents the hypothetical area of the ancient 'Rio Grande System', as described by *Schönhuth et al. (2011)*. Abbreviations: *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.

and Viesca lagoons (both currently dry) during the Late Pleistocene (*Meek, 1904*; *Burr, 1976*). While a previous biogeographic study in *C. ornata* found a close relationship between the Conchos and Nazas drainage populations, which was attributed to incomplete lineage sorting, no biogeographic scenario was presented (*Schönhuth et al., 2015*).

CONCLUSIONS

The geographic distribution of the four genetic lineages recovered in *P. promelas* (Casas Grandes, Santa Maria, Yaqui, and Nazas+Conchos) is similar to the lineage distributions found in other freshwater fishes in the North of Mexico, such as *C. ornatum* (*Domínguez-Domínguez et al., 2011*; *Schönhuth et al., 2011*), *Rhinichthys cataractae* (*Kim & Conway, 2014*), *C. ornata* (*Schönhuth et al., 2015*), and *P. plebeius* (*Corona-Santiago et al., 2018*). Cladogenic events in *P. promelas* are hypothesized to have been caused by the combined influence of tectonic events and increasing regional aridity; in particular, the fragmentation of the ancestral Rio Grande system and interchange events between basins via stream capture.

The phylogenetic analyses, species tree and Bayesian species delimitation tests results validate the presence of four genetic lineages. In the future, it would be interesting to increase both the sample size and the number of genes, and to evaluate the morphological diversity in the other Mexican populations of *P. promelas*. This would allow the taxonomic status of the genetic lineages found in the present study to be established and the determination of their relationship with other populations of *P. promelas* in North America.

ACKNOWLEDGEMENTS

We thank Dr. Isaí Betancourt Resendes, Dr. Victor Julio Piñeros and Dr. Diushi Keri Corona Santiago for their valuable help with the genetic analyses. We sincerely acknowledge Dr. Tom Jameson of Chester Zoo for reviewing the English of the manuscript.

Funding

This research was financially supported by a grant from the Program Coordinadora de Investigaciones Científicas (CIC-UMSNH) awarded to Rodolfo Pérez Rodríguez and Omar Domínguez Domínguez and Ministerio de Economía y Competitividad, Spain (CGL-2013 41375-P). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Grant Disclosures

The following grant information was disclosed by the authors:
Program Coordinadora de Investigaciones Científicas (CIC-UMSNH).
Omar Domínguez Domínguez and Ministerio de Economía y Competitividad, Spain: CGL-2013 41375-P.

Competing Interests

The authors declare there are no competing interests.

Author Contributions

- Nayarit E. Ballesteros-Nova conceived and designed the experiments, performed the experiments, analyzed the data, prepared figures and/or tables, authored or reviewed drafts of the paper, approved the final draft.
- Rodolfo Pérez-Rodríguez conceived and designed the experiments, contributed reagents/materials/analysis tools, authored or reviewed drafts of the paper, approved the final draft, teaching phylogenetic analysis programs.
- Rosa G. Beltrán-López analyzed the data.
- Omar Domínguez-Domínguez conceived and designed the experiments, contributed reagents/materials/analysis tools, authored or reviewed drafts of the paper, approved the final draft.

Field Study Permissions

The following information was supplied relating to field study approvals (i.e., approving body and any reference numbers):

All procedures were reviewed and approved by a committee of Mexican Ministry of Environment and Natural Resources (SEMARNAT), under collection permit number PPF/DGOPA-362/1.

Data Availability

The raw data are provided in the Supplemental Files and in GenBank: MK821253–MK821306.

Supplemental Information

Supplemental information for this article can be found online at http://dx.doi.org/10.7717/peerj.6224#supplemental-information.

REFERENCES

Aguirre-Díaz GJ, Labarthe-Hernández G, Tristán-González M, Nieto-Obregón J, Gutiérrez-Palomares I. 2008. The ignimbrite flare-up and graben calderas of the Sierra Madre Occidental, Mexico. In: Gottsman J, Marti J, eds. *Caldera Volcanism: analysis, modelling and response*. Vol. 10. New York: Elsevier, developments in volcanology, 143–180 DOI 10.1016/S1871-644X(07)00004.

Albritton CC. 1958. Quaternary stratigraphy of the Guadiana Valley, Durango, Mexico. *Bulletin of the Geological Society of America* **69**:1197–1216 DOI 10.1130/0016-7606(1958)69[1197:QSOTGV]2.0.CO;2.

Aranda-Gómez JJ, Carranza-Castañeda O, Wang X, Tseng ZJ, Pacheco-Castro A. 2018. Notes on the origin of extensive endorheic regions in central and northern Mexico, and some implications for paleozoogeography. *Journal of South American Earth Sciences* **83**:55–67 DOI 10.1016/j.jsames.2018.02.003.

Aranda-Gómez JJ, Luhr JF, Housh TB, Valdez-Moreno G, Chávez-Cabello G. 2005. El vulcanismo tipo intraplaca del Cenozoico tardío en el centro y norte de México: una revisión. *Boletin de la Sociedad Geológica Mexicana, Volumen Conmemorativo del Centenario Temas Selectos de la Geología Mexicana* **57(3)**:187–225 DOI 10.18268/bsgm2005v57n3a1.

Baele G, Lemey P, Bedford T, Rambaut A, Suchard MA, Alekseyenko AV. 2012. Improving the accuracy of demographic and molecular clock model comparison while accommodating phylogenetic uncertainty. *Molecular Biology and Evolution* **29(9)**:2157–2167 DOI 10.1093/molbev/mss084.

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.

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.

Bruen TC, Philippe H, Bryant D. 2006. A simple and robust statistical test for detecting the presence of recombination. *Genetics* **172(4)**:2665–2681 DOI 10.1534/genetics.105.048975.

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.

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.

Chow S, Hazama K. 1998. Universal PCR primers for *S7* ribosomal protein gene introns in fish. *Molecular Ecology* **7(9)**:1255–1256.

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.

Corona-Santiago DK, Domínguez-Domínguez O, Tovar-Mora Ll, Pardos-Blas JR, Herrerías-Diego Y, Pérez-Rodríguez R, Doadrio I. 2018. Historical biogeography reveals new independent evolutionary lineages in the *Pantosteus plebeius-nebuliferus*

species-group (Actinopterygii: Catostomidae). *BMC Evolutionary Biology* **18**:173 DOI 10.1186/s12862-018-1286-y.

De Queiroz K. 2007. Species concepts and species delimitation. *Systematic Biology* **56(6)**:879–886 DOI 10.1080/10635150701701083.

Devitt TJ. 2006. Phylogeography of the Western Lyresnake (*Trimorphodon biscutatus*): testing aridland biogeographical hypotheses across the Nearctic-Neotropical transition. *Molecular Ecology* **15(14)**:4387–4407 DOI 10.1111/j.1365-294X.2006.03015.x.

Domínguez-Domínguez O, Vila M, Pérez-Rodríguez R, Remón N, Doadrio I. 2011. Complex evolutionary history of the Mexican Stoneroller *Campostoma ornatum* Girard, 1856 (Actinopterygii: Cyprinidae). *BMC Evolutionary Biology* **11**:153 DOI 10.1186/1471-2148-11-153.

Dowling TE, Tibbets CA, Minckley WL, Smith GR. 2002. Evolutionary relationships of the plagopterins (Teleostei: Cyprinidae) from cytochrome *b* sequences. *Copeia* **2002(3)**:665–678 DOI 10.1643/0045-8511(2002)002[0665:EROTPT]2.0.CO;2.

Drummond AJ, Ho SYW, Phillips MJ, Rambaut A. 2006. Relaxed phylogenetics and dating with confidence. *PLOS Biology* **4(5)**:e88 DOI 10.1371/journal.pbio.0040088.

Drummond AJ, Suchard MA, Xie D, Rambaut A. 2012. Bayesian phylogenetics with BEAUti and the BEAST 1.7. *Molecular Biology and Evolution* **29(8)**:1969–1973 DOI 10.1093/molbev/mss075.

Echelle AA, Carson EW, Echelle AF, Van Der Bussche RA, Dowling TE, Meyer A. 2005. Historical biogeography of the New-World pupfish genus *Cyprinodon* (Teleostei: Cyprinodontidae). *Copeia* **2005(2)**:320–339 DOI 10.1643/CG-03-093R3.

Ferrari L, Valencia-Moreno M, Bryan S. 2007. Magmatism and tectonics of the Sierra Madre Occidental and its relation with the evolution of the western margin of North America. *Geological Society of America Special Papers* **422**:1–39 DOI 10.1130/2007.2422(01).

Firth C, Kitchen A, Shapiro B, Suchard MA, Holmes EC, Rambaut A. 2010. Using time-structured data to estimate evolutionary rates of double-stranded DNA viruses. *Molecular Biology and Evolution* **27(9)**:2038–2051 DOI 10.1093/molbev/msq088.

Galloway WE, Whiteaker TL, Ganey-Curry P. 2011. History of Cenozoic North American drainage basin evolution, sediment yield, and accumulation in the Gulf of Mexico basin. *Geosphere* **7(4)**:938–973 DOI 10.1130/GES00647.1.

Hasegawa M, Kishino H, Yano T. 1985. Dating the human-ape split by a molecular clock of mitochondrial DNA. *Journal of Molecular Evolution* **22(2)**:160–174 DOI 10.1007/BF02101694.

Henry CD, Aranda-Gómez JJ. 2000. Plate interactions control middle-late Miocene proto-Gulf and Basin and Tange extension in the southern Basin and Range. *Tectonophysics* **318(1)**:1–16 DOI 10.1130/0016-7606(2001)113<0760:EEAAMA>2.0.CO;2.

Hughes RM, Rinne JA, Calamusso B. 2005. Historical changes in large river fish assemblages of the Americas: a synthesis. In: Rinne JN, Hughes RM, Calamusso B, eds. *Historical changes in large river fish assemblages of the Americas*. 45. Bethesda: American Fisheries Society, Symposium, 603–612.

Huson DH, Bryant D. 2006. Application of phylogenetic networks in evolutionary studies. *Molecular Biology and Evolution* **23(2)**:254–267 DOI 10.1093/molbev/msj030.

Kim D, Conway KW. 2014. Phylogeography of *Rhinichthys cataractae* (Teleostei: Cyprinidae): pre-glacial colonization across the Continental Divide and Pleistocene diversification within the Rio Grande drainage. *Biological Journal of the Linnean Society* **111(2)**:317–333 DOI 10.1111/bij.12209.

Kopp A, True JR. 2002. Phylogeny of the oriental *Drosophila melanogaster* species group: a multilocus reconstruction. *Systematic Biology* **51(5)**:786–805 DOI 10.1080/10635150290102410.

Lanfear R, Calcott B, Ho SY, Guindon S. 2012. PartitionFinder: combined selection of partitioning schemes and substitution models for phylogenetic analyses. *Molecular Biology and Evolution* **29(6)**:1695–1701 DOI 10.1093/molbev/mss020.

Leaché AD, Fujita MK. 2010. Bayesian species delimitation in West African forest geckos (*Hemidactylus fasciatus*). *Proceedings of the Royal Society B: Biological Sciences* **277(1697)**:3071–3077 DOI 10.1098/rspb.2010.0662.

Leigh JW, Bryant D. 2015. Popart: full-feature software for haplotype network construction. *Methods in Ecology and Evolution* **6(9)**:1110–1116 DOI 10.1111/2041-210X.12410.

Librado P, Rozas J. 2009. DnaSP v5: a software for comprehensive analysis of DNA polymorphism data. *Bioinformatics* **25(11)**:1451–1452 DOI 10.1093/bioinformatics/btp187.

Mayden RL, Burr BM, Page LM, Miller RR. 1992. The native freshwater fishes of North America. In: Mayden RL, ed. *Systematics, historical ecology, and North American freshwater fishes.* Stanford: Stanford University Press, 827–863.

Mayden RL, Matson RH, Hillis DM. 1992. Speciation in the North American genus *Dionda* (Teleostei: Cypriniformes). In: Mayden RL, ed. *Systematics, historical ecology, and North American freshwater fishes.* Stanford: Stanford University Press, 710–746.

Meek SE. 1904. The fresh-water fishes of Mexico north of the Isthmus of Tehuantepec. *Field Columbian Museum Publication, Zoological Series* **5**:1–252 DOI 10.5962/bhl.title.2229.

Metcalfe SE. 2006. Late Quaternary environments of the Northern deserts and central transvolcanic belt of Mexico. *Annals of the Missouri Botanical Garden* **93(2)**:258–273 DOI 10.3417/0026-6493(2006)93.

Miller MA, Pfeiffer W, Schwartz T. 2010. Creating the CIPRES science gateway for inference of large phylogenetic trees. *Gateway Computing Environments Workshop* **2010**:1–8 DOI 10.1109/GCE.2010.5676129.

Miller RR, Minkley WL, Norris SM. 2005. *Freshwater fishes of Mexico.* Chicago: University of Chicago Press.

Miller RR, Smith ML. 1986. Origin and geography of the fishes of central Mexico. In: Hocutt CH, Wiley EO, eds. *The Zoogeography of North American Freshwaters Fishes.* New York: John Wiley and Sons, 485–517.

Minckley WL, Hendrickson DA, Bond CE. 1986. Geography of Western North American freshwater fishes: description and relationships to intracontinental tectonism.

In: Hocutt CH, Wiley EO, eds. *Zoogeography of Western North American freshwater fishes.* New York: John Wiley and Sons, 519–613.

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

Nagle BC, Simons AM. 2012. Rapid diversification in the North American minnow genus *Nocomis. Molecular Phylogenetics and Evolution* **63(3)**:639–649 DOI 10.1016/j.ympev.2012.02.013.

Nathan R. 2001. The challenges of studying dispersal. *Trends in Ecology and Evolution* **16(9)**:481–483 DOI 10.1016/S0169-5347(01)02272-8.

Pérez-Rodríguez R, Domínguez-Domínguez O, Doadrio I, Cuevas-García E, Pérez-Ponce de León G. 2015. Comparative historical biogeography of three groups of Nearctic freshwater fishes across central Mexico. *Journal of Fish Biology* **86(3)**:993–1015 DOI 10.1111/jfb.12611.

Pérez-Rodríguez R, Domínguez-Domínguez O, Pérez-Ponce de León G, Doadrio I. 2009. Phylogenetic relationships and biogeography of the genus *Algansea* Girard (Cypriniformes: Cyprinidae) of central Mexico inferred from molecular data. *BMC Evolutionary Biology* **9**:223 DOI 10.1186/1471-2148-9-223.

Posada D. 2003. Unit 6.5: using MODELTEST, and PAUP* to select a model of nucleotide substitution. In: Baxevanis AD, et al., eds. *Current protocols in bioinformatics.* Vol. 1. John Wiley and Sons, New York, 6.5.1–6.5.28 DOI 10.1002/0471250953.bi0605s00.

Rambaut A. 2014. FigTree v.1.4.2 software. *Available at* http://tree.bio.ed.ac.uk/software/figtree/ (accessed on 11 November 2016).

Rambaut A, Drummond A. 2007. Tracer v.1.5 software. *Available at* http://tree.bio.ed.ac.uk/software/tracer (accessed on 11 November 2016).

Rannala B, Yang Z. 2013. Improved reversible jump algorithms for Bayesian species delimitation. *Genetics* **194(1)**:245–253 DOI 10.1534/genetics.112.149039.

Riddle BR, Hafner DJ. 2006. A step-wise approach to integrating phylogeographic and phylogenetic biogeographic perspectives on the history of a core North American warm deserts biota. *Journal of Arid Environments* **66(3)**:435–461 DOI 10.1016/j.jaridenv.2006.01.014.

Rokas A, Carroll SB. 2005. More genes or more taxa? The relative contribution of gene number and taxon number to phylogenetic accuracy. *Molecular Biology and Evolution* **22(5)**:1337–1344 DOI 10.1093/molbev/msi121.

Rokas A, Williams BL, King N, Carroll SB. 2003. Genome-scale approaches to resolving incongruence in molecular phylogenies. *Nature* **425(6960)**:798–804 DOI 10.1038/nature02053.

Ronquist F, Teslenko M, Van der Mark P, Ayres DL, Darling A, Höhna S, Larget B, Liu L, Suchard MA, Huelsenbeck JP. 2012. MrBayes 3.2: efficient bayesian phylogenetic inference and model choice across a large model space. *Systematic Biology* **61(3)**:539–542 DOI 10.1093/sysbio/sys029.

Sambrook J, Fritsch EF, Maniatis T. 1989. *Molecular cloning: a laboratory manual.* New York: Cold Spring Harbor Laboratory Press.

Satta Y, Klein J, Takahata N. 2000. DNA archives and our nearest relative: the trichotomy problem revisited. *Molecular Phylogenetics and Evolution* **14(2)**:259–275 DOI 10.1006/mpev.2000.0704.

Schönhuth S, Beachum CE, Knouft JE, Mayden RL. 2016. Phylogeny and genetic variation within the widely distributed Bluntnose Minnow, *Pimephales notatus* (Cyprinidae), in North America. *Zootaxa* **4168(1)**:38–60 DOI 10.11646/zootaxa.4168.1.2.

Schönhuth S, Blum MJ, Lozano-Vilano L, Neely DA, Varela-Romero A, Espinosa H, Perdices A, Mayden RL. 2011. Inter-basin exchange and repeated headwater capture across the Sierra Madre Occidental inferred from the phylogeography of Mexican stonerollers. *Journal of Biogeography* **38(7)**:1406–1421 DOI 10.1111/j.1365-2699.2011.02481.x.

Schönhuth S, Doadrio I, Domínguez-Domínguez O, Hillis DM, Mayden RL. 2008. Molecular evolution of southern North American Cyprinidae (Actinopterygii), with the description of the new genus *Tampichthys* from central Mexico. *Molecular Phylogenetics and Evolution* **47(2)**:729–756 DOI 10.1016/j.ympev.2007.11.036.

Schönhuth S, Doadrio I, Mayden RL. 2007. A biogeographic perspective on the phylogeny of Mexican cyprinids (Actinopterygii: Cyprinidae). In: Lozano-Vilano ML, Contreras-Balderas AJ, eds. *Studies of North American Desert Fishes in Honor of E. P. (Phil) Pister, Conservationist.* Monterrey: Universidad Autónoma de Nuevo León, 102–124.

Schönhuth S, Hillis DM, Neely DA, Lozano-Vilano L, Perdices A, Mayden RL. 2012. Phlogeny, diversity, and species delimitation of the North American Round-Nosed Minnows (Teleostei: *Dionda*), as inferred from mitochondrial and nuclear DNA sequences. *Molecular Phylogenetics and Evolution* **62(1)**:427–446 DOI 10.1016/j.ympev.2011.10.011.

Schönhuth S, Lozano-Vilano L, Perdices A, Espinosa H, Mayden RL. 2015. Phylogeny, genetic diversity and phylogeography of the genus *Codoma* (Teleostei, Cyprinidae). *Zoologica Scripta* **44(1)**:11–28 DOI 10.1111/zsc.12083.

Schönhuth S, Mayden RL. 2010. Phylogenetic relationships in the genus *Cyprinella* (Actinopterygii: Cyprinidae) based on mitochondrial and nuclear gene sequences. *Molecular Phylogenetics and Evolution* **55(1)**:77–98 DOI 10.1016/j.ympev.2009.10.030.

Schönhuth S, Perdices A, Lozano-Vilano L, García-de León FJ, Espinosa H, Mayden RL. 2014. Phylogenetic relationships of North American western chubs of the genus *Gila* (Cyprinidae, Teleostei), with emphasis on southern species. *Molecular Phylogenetics and Evolution* **70**:210–230 DOI 10.1016/j.ympev.2013.09.021.

Schönhuth S, Vukic J, Sanda R, Yang L, Mayden RL. 2018. Phylogenetic relationships and classification of the Holarctic family Leuciscidae (Cypriniformes: Cyprinoidei). *Molecular Phylogenetics and Evolution* **127**:781–799 DOI 10.1016/j.ympev.2018.06.026.

Silvestro D, Michalak I. 2012. RaxmlGUI: a graphical front-end for RAxML. *Organisms Diversity and Evolution* **12(4)**:335–337 DOI 10.1007/s13127-011-0056-0.

Smith ML, Miller RR. 1986. The evolution of the Rio Grande basin as inferred from its fish fauna. In: Hocutt CH, Wiley EO, eds. *The zoogeography of north american freshwaters fishes.* New York: John Wiley and Sons, 457–485 DOI 10.1086/415476.

Smith ML, Song J, Miller RR. 1984. Redescription, variation, and zoogeography of the Mexican darter *Etheostoma pottsi* (Pisces: Percidae). *The Southwestern Naturalist* **29**:395–402 DOI 10.2307/3670991.

Sousa-Santos C, Robalo JI, Collares-Pereira MJ, Almada VC. 2005. Heterozygous indels as useful tools in the reconstruction of DNA sequences and in the assessment of ploidy level and genomic constitution of hybrid organisms. *DNA Sequence* **16(6)**:462–467 DOI 10.1080/10425170500356065.

Stamatakis A. 2014. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* **30(9)**:1312–1313 DOI 10.1093/bioinformatics/btu033.

Tamura K, Nei M. 1993. Estimation of the number of nucleotide substitutions in the control region of mitochondrial DNA in humans and chimpanzees. *Molecular Biology and Evolution* **10(3)**:512–526 DOI 10.1093/oxfordjournals.molbev.a040023.

Tamura K, Stecher G, Peterson D, Filipski A, Kumar S. 2013. MEGA6: molecular evolutionary genetics analysis version 6.0. *Molecular Biology and Evolution* **30(12)**:2725–2729 DOI 10.1093/molbev/mst197.

Tavare S. 1986. Some probabilistic and statistical problems in the analysis of DNA sequences. *Lectures on Mathematics in the Life Sciences. American Mathematical Society* **17**:57–86.

Thompson JD, Gibson TJ, Plewniak F, Jeanmougin F, Higgins DG. 1997. The CLUSTAL_X windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. *Nucleic Acids Research* **25(24)**:4876–4882 DOI 10.1093/nar/25.24.4876.

Vallinoto M, Sequeira F, Sodré D, Bernardi JAR, Sampaio I, Schneider H. 2010. Phylogeny and biogeography of the *Rhinella marina* species complex (Amphibia, Bufonidae) revisited: implications for Neotropical diversification hypotheses. *Zoologica Scripta* **39(2)**:128–140 DOI 10.1111/j.1463-6409.2009.00415.x.

Vandermeer HJ. 1966. Statistical analysis of geographic variation of the fathead minnow, *pimephales promelas. Copeia* **1966(3)**:457–466 DOI 10.2307/1441065.

Wood RM, Mayden RL. 2002. Speciation and anagenesis in the genus *Cyprinella* of Mexico (Teleostei: Cyprinidae): a case study of Model III allopatric speciation. *Reviews in Fish Biology and Fisheries* **12(2–3)**:253–271 DOI 10.1023/A:1025018118339.

Yang ZH. 2015. The BPP program for species tree estimation and species delimitation. *Current Zoology* **61(5)**:854–865 DOI 10.1093/czoolo/61.5.854.

Yang ZH, Rannala B. 2010. Bayesian species delimitation using multilocus sequence data. *Proceedings of the National Academy of Sciences of the United States of America* **107(20)**:9264–9269 DOI 10.1073/pnas.0913022107.

Yang ZH, Rannala B. 2014. Unguided species delimitation using DNA sequence data from multiple loci. *Molecular Biology and Evolution* **31(12)**:3125–3135 DOI 10.1093/molbev/msu279.

Zardoya R, Doadrio I. 1999. Molecular evidence on the evolutionary and biogeographical patterns of european cyprinids. *Journal of Molecular Evolution* **49(2)**:227–237 DOI 10.1007/PL00006545.

FURTHER READING

Ho SY, Phillips MJ, Cooper A, Drummond AJ. 2005. Time dependency of molecular rate estimates and systematic overestimation of recent divergence times. *Molecular Biology and Evolution* **22(7)**:1561–1568 DOI 10.1093/molbev/msi145.