Mitochondrial evidence for a new evolutionary significant unit within the *Gila eremica* lineage (Teleostei, Cyprinidae) in Sonora, Northwest Mexico

C. A. Ballesteros–Córdova, A. Varela–Romero, G. Ruiz–Campos, L. T. Findley, J. M. Grijalva–Chon, L. E. Gutiérrez–Millán

Ballesteros–Córdova, C. A., Varela–Romero, A., Ruiz–Campos, G., Findley, L. T., Grijalva–Chon, J. M., Gutiérrez–Millán, L. E., 2019. Mitochondrial evidence for a new evolutionary significant unit within the *Gila eremica* lineage (Teleostei, Cyprinidae) in Sonora, Northwest Mexico. *Animal Biodiversity and Conservation*, 42.1: 171–186, Doi: https://doi.org/10.32800/abc.2019.42.0171

Abstract

Mitochondrial evidence for a new evolutionary significant unit within the Gila eremica lineage (Teleostei, Cyprinidae) in Sonora, Northwest Mexico. We present the phylogenetic affinities and DNA barcode of *Gila* cf. *eremica,* a geographically isolated and morphologically divergent population from *G. eremica* DeMarais, 1991. Mitochondrial phylogenetic analyses of *cyt–b*, *cox1* and *nd2* show a clades pattern within the *G. eremica* lineage, placing *G.* cf. *eremica* in a clade of specific identity to and sharing a putative common ancestor with *G. eremica* from the Mátape River basin. The barcoding analysis using a character–based approach of CAOS showed seven single pure characters discriminating *G. eremica* from its regional congener *G. purpurea*, and one fixed character in *G.* cf. *eremica* discriminating it from *G. eremica.* These results and the recent detection of diagnostic morphological differences between *G.* cf. *eremica* and *G. eremica* support the hypothesis of *Gila* cf. *eremica* as an significant evolutionary unit within the *G. eremica* lineage.

Key words: *Gila,* Phylogenetic analyses, DNA barcode, Evolutionary significant unit, Northwest México

Resumen

Evidencia mitocondrial de una nueva unidad evolutivamente significativa en el linaje de Gila eremica *(Teleostei, Cyprinidae) en Sonora, Noroeste de México.* Presentamos las afinidades filogenéticas y el código de barras del ADN de *Gila* cf. *eremica*, una población morfológicamente divergente y geográficamente aislada de *G. eremica* DeMarais, 1991. Los análisis filogenéticos mitocondriales de *cyt–b*, *cox1* y *nd2* muestran la existencia de un patrón de clados dentro del linaje de *G. eremica*, que sitúa a *G.* cf. *eremica* en un clado de identidad específica que comparte un supuesto ancestro común con *G. eremica*, originario de la cuenca del río Mátape. El análisis de código de barras, en que se utilizó un método basado en caracteres del programa informático CAOS, mostró siete caracteres puros que diferencian a *G. eremica* de su congénere regional *G. purpurea* y un carácter fijo en *G.* cf. *eremica* que lo diferencia de *G. eremica*. Estos resultados y la reciente detección de diferencias morfológicas diagnósticas entre *G.* cf. *eremica* y *G. eremica* sostienen la hipótesis de que *Gila* cf. *eremica* es una unidad evolutivamente significativa dentro del linaje de *G. eremica*.

Palabras clave: *Gila*, Análisis filogenéticos, Código de barras de ADN, Unidad evolutivamente significativa , Noroeste de México

Received: 28 V 18; Conditional acceptance: 12 IX 18; Final acceptance: 09 X 18

C. A. Ballesteros–Córdova, A. Varela–Romero, J. M. Grijalva–Chon, L. E. Gutiérrez–Millán, Departamento de Investigaciones Científicas y Tecnológicas de la Universidad de Sonora, Blvd. Luis Encinas y Rosales s/n., 83000 Hermosillo, Sonora, México.– G. Ruiz–Campos, Facultad de Ciencias, Universidad Autónoma de Baja California, 22800 Ensenada, Baja California, México.– L. T. Findley, Centro de Investigación en Alimentación y Desarrollo, A. C. Unidad Guaymas, 85480 Guaymas, Sonora, México.

Corresponding author: A. Varela–Romero. E–mail: alejandro.varela@unison.mx

ISSN: 1578–665 X
eISSN: 2014–928 X

Introduction

The genus *Gila*, one of the most widespread groups of the family Cyprinidae in North America, includes morphologically heterogeneous fishes inhabiting waters of arid and semiarid regions of the western United States (USA) and northwestern Mexico (Miller et al., 2005). Using both mitochondrial and nuclear markers it has been observed that recent phylogenetic analyses of cyprinids in North Americaalign *Gila* with ten other nominal genera within a so–called Revised Western Clade (RWC) (Schönhuth et al., 2012), and suggest that *Gila* comprises an evolutionary lineage involving at least 18 species, including species currently recognized taxonomically in the monotypic genera *Acrocheilus* and *Moapa* (Schönhuth et al., 2014). Nevertheless, Schönhuth and colleagues termed he composition of the *Gila* lineage incomplete because of phylogenetic affinities of *G. coerulea* and *Ptychocheilus lucius* that were only resolved within *Gila* by using mitochondrial marker *cyt–b* rather than the concatenated nuclear genes *rag1*, *rhod*, and *s7* or concatenated mitochondrial and nuclear markers (Schönhuth et al., 2012, 2014).

Molecular and morphological analyses of *Gila* species occurring in northern México and western USA suggest that nominal species of the *G. robusta* complex in the Colorado River basin show both allopatric and sympatric distributions, with probable hybrid origins at least in part (DeMarais et al., 1992; Dowling and DeMarais, 1993; Gerber et al., 2001; Schönhuth et al., 2014; Dowling et al., 2015; Page et al., 2017). *Gila* species in Mexican waters of the Atlantic slope Chihuahuan Desert region, and those in the Pacific slope, however, apparently show mainly allopatric distributions associated with major river drainages, suggesting peripatric speciation events (Wiley, 1981; Schönhuth et al., 2014).

Phylogenetic affinities and current distributions of all known species of *Gila* show that those in Mexico include an Atlantic–slope lineage referred to as the Chihuahuan Desert Group, which includes *G. pulchra, G. conspersa, G. nigrescens, G. brevicauda,* an undescribed species, and a lineage composed of *G. modesta* nested within *G. pandora* (Schönhuth et al., 2014). Their phylogenetic affinities, morphological characteristics, and geographical distributions suggest that species in this group share a single common ancestor, unrelated to any species belonging to the *G. robusta* complex of the Colorado River system (Uyeno, 1960; Schönhuth et al., 2014). The remaining nominal species of *Gila* in México (*G. ditaenia, G. minacae, G. purpurea*, *G. eremica*) occur in Pacific–slope drainages. These four species were not resolved in analyses by Schönhuth et al. (2014) as part of their Chihuahuan Desert Group. However, they were resolved as monophyletic within the greater *Gila* lineage, and *G. eremica* and *G. purpurea* were corroborated as sister species (Schönhuth et al., 2012, 2014), as proposed by previous morphological analysis (DeMarais, 1991).

The study of the evolution of *Gila* is important because it is an opportunity to understand the evolution of freshwater fish because it relates to the geological history of western North America. Evaluations of *Gila* species in USA and México suggest a current–day lack of understanding regarding the diversity of the genus (Schönhuth et al., 2014). This is substantiated by paraphyletic groupings obtained by Schönhuth et al. (2012, 2014) and recent records of undescribed populations of the genus in several drainages of central–north and northwest Mexico (DeMarais, 1991; Varela–Romero, 2001; Norris et al., 2003; Minckley and Marsh, 2009; Bogan et al., 2014; Schönhuth et al., 2014).

The rapid development of molecular taxonomic and systematic methods in recent years has provided several tools to study biodiversity. Nowadays, in addition to molecular phylogenetic methods, the DNA barcoding technique (Hebert et al., 2003a, 2003b) has been applied as a molecular taxonomic tool to support species identifications and species discoveries (Hebert et al., 2003a, 2003b; Witt et al., 2006; Hubert et al., 2008; Rach et al., 2008; Lara et al., 2010; Li et al., 2011; Zou and Li, 2016; Yi et al., 2017). Current analytical methods to assign DNA barcodes to taxa can be divided into distance–based, phylogeny–based, and character–based approaches (Hebert et al., 2003a; Pons et al., 2006; Sarkar et al., 2008; Puillandre et al., 2012; Taylor and Harris, 2012). Although genetic distance–based methods for DNA barcoding have been considered useful tools in species discrimination and cryptic species discovery (Ward et al., 2005; Hubert et al., 2008; Lara et al., 2010; April et al., 2011; Lakra et al., 2015, Zou and Li, 2016), this approach has been questioned because of the relatively high rates of evolution of mitochondrial DNA between and within species (and between different groups of species) that can result in overlaps of intra– and interspecific distances, thus suggesting an uncertain existence of a barcoding gap for all species (Kipling and Rubinoff, 2004; Rubinoff, 2006; Rubinoff et al., 2006).

The character–based analytical method known as CAOS (characteristic attribute organization system) has been used to define barcodes of taxa (Rach et al., 2008; Sarkar et al., 2008; Damm et al., 2010; Yassin et al., 2010; Li et al., 2011; Reid et al., 2011; Jörger and Schrödl, 2013; Yu et al., 2014; Zou and Li, 2016; Yi et al., 2017), and has been considered to better approximate a 'real' barcode, as well as providing better resolution to distinguish species than other approaches (Reid et al., 2011; Zou et al., 2011; Yu et al., 2014). Like traditional taxonomy, species identifications using CAOS DNA barcoding operate under the premise that members of a given taxonomic group share character attributes (i. e., putative diagnostic nucleotides) that are absent from other related groups (Rach et al., 2008; Sarkar et al., 2008; Bergmann et al., 2009; Jörger and Schrödl, 2013).

Recent records of unstudied populations of *Gila* in northwestern México include two populations inhabiting a series of large spring–fed pools (*pozas*) of the Arroyo El Tigre sub–basin (Varela–Romero, 2001; Bogan et al., 2014). This sub–basin pertains to the Mátape River basin to the east and includes the adjacent low–elevation subtropical canyons La Balandrona and La Pirinola, both located in the southeastern sector of the Sierra

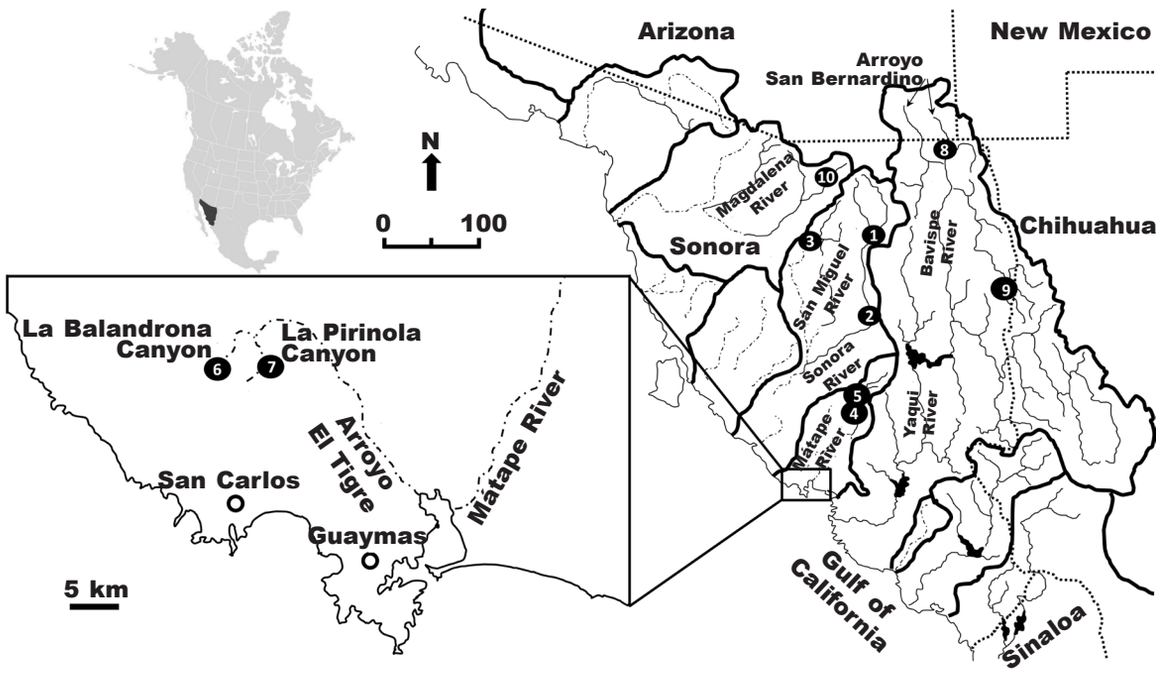


Fig. 1. Collection sites for *Gila* specimens analyzed. Open circles represent towns, and numbers in solid black circles are locations detailed in table 1. Hydrographic drainage divides are indicated by thick lines. Dashed lines indicate contemporary intermittent drainage courses. Dotted lines are state boundaries.

Fig. 1. Sitios de captura de los especímenes de Gila *analizados. Los círculos representan las ciudades. Los números en el interior de los círculos negros son las localidades que figuran en la tabla 1. Las divisiones de las cuencas hidrográficas se indican con líneas gruesas. Las líneas discontinuas indican los cursos de agua intermitentes contemporáneos. Las líneas punteadas son los límites estatales.*

El Aguaje coastal mountain range, near the towns of San Carlos and Guaymas (fig. 1, inset). The proximate geographical location of these newly discovered populations suggest they are part of the *Gila eremica* lineage (Varela–Romero, 2001; Bogan et al., 2014) and thus are referred to herein as *Gila* cf. *eremica*. However, both populations are geographically isolated from populations of the lineage of *G. eremica* inhabiting the Mátape and Sonora River basins to the east and northeast. This isolation was probably promoted by volcanic events occurring in the area during the Miocene (Mora–Álvarez and McDowell, 2000), causing a geographical disconnection of the Arroyo El Tigre sub–basin from the Mátape River sub–basin. Recent morphological evaluations of the *Gila eremica* lineage revealed the *G*. cf. *eremica* populations as distinct compared to all *G. eremica* and other selected congeners analyzed, and showed at least 16 morpho–linear and two meristic characters that distinguish *G.* cf. *eremica* from other *G. eremica* populations (Ballesteros–Córdova et al., 2016). These mensural and meristic differences detected in the *G.* cf. *eremica* populations, as well as their isolated geographic occurrence, suggest a potential evolutionary isolation event within the *G. eremica* lineage (Ballesteros–Córdova et al., 2016), similar to that proposed for *G. eremica* and *G. purpurea* in Sonora (DeMarais, 1991; Schönhuth et al., 2014). The proposal of a *Gila eremica* lineage comprised of populations from the Sonora and Mátape River basins, plus the *G.* cf. *eremica* from the isolated Arroyo El Tigre sub–basin in La Balandrona and La Pirinola canyons, calls for development of molecular analyses to further investigate the evolutionary affinities of *G*. cf. *eremica* within the entire *Gila* lineage, and to potentially detect character attributes that may discriminate it from related groups. Knowledge of the evolutionary history of *Gila* can contribute to elucidating speciation mechanisms involved in this taxonomically problematic genus and other related fishes from arid and semiarid regions in North America. Also, the potential recognition of an evolutionary significant unit within *Gila* in México would enable the development of management strategies for its conservation.

Material and methods

Sample collection and DNA extraction

A total of 178 specimens of five species of the genus *Gila* occurring in four river basins (seven sub–basins)

Table 1. Taxa, sampling localities (drainage depicted in figure 1), molecular markers, accession numbers (No), voucher specimen catalog numbers (USON, Universidad de Sonora, Hermosillo, México; MNCN, Museo Nacional de Ciencias Naturales, Madrid, Spain; UAIC, University of Alabama, Tuscaloosa, Alabama, USA; BYU, Brigham Young University, Provo, Utah, USA), and source references for specimens of *Gila* spp. used for molecular analyses (S: 1, this study; 2, Schönhuth et al., 2014).

Tabla 1. Taxones, localidades de muestreo (drenajes representados en la figura 1), marcadores moleculares, números de accesión (No), números de catálogo de los especímenes de referencia (USON, Universidad de Sonora, Hermosillo, México; MNCN, Museo Nacional de Ciencias Naturales, Madrid, España; UAIC, Universidad de Alabama, Tuscaloosa, Alabama, EE.UU.; BYU, Universidad Brigham Young, Provo, Utah, EE.UU.) y referencia de los especímenes de Gila *spp. utilizados para realizar los análisis moleculares (S: 1, este estudio; 2, Schönhuth et al., 2014).*

Taxon	Locality	Gene	No	Voucher	S
Sonora River basin					
G. eremica	1 - Bacanuchi River sub–basin,	*cyt–b*	KX855966	USON–1301–1	1
	Bacanuchi River at Tahuichopa ford,	*cyt–b*	KX855967	USON–1301–3	1
	Arizpe–Cananea road, Sonora	*cyt–b*	KX855968	USON–1301–15	1
	30º 21' 59.66" N, 110º 09' 24.54" W	*cyt–b*	KX855969	USON–1301–17	1
		cyt–b	KX855970	USON–1301–25	1
		nd2	KX858655	USON–1301–1	1
		nd2	KX858656	USON–1301–15	1
		cox1	KX858664	USON–1301–1	1
		cox1	MH091955	USON–1301–2	1
		cox1	MH091956	USON–1301–3	1
		cox1	MH091957	USON–1301–4	1
		cox1	MH091958	USON–1301–5	1
		cox1	MH091959	USON–1301–6	1
		cox1	MH091960	USON–1301–9	1
		cox1	MH091961	USON–1301–10	1
		cox1	MH091962	USON–1301–11	1
		cox1	MH091963	USON–1301–12	1
		cox1	MH091964	USON–1301–13	1
G. eremica	2 - Sonora River sub–basin,	*cyt–b*	KF514191	MNCN–279687	2
	Sonora River at El Cahui	*cyt–b*	KF514192	MNCN–279687	2
	(1 km from La Labor), Sonora	*cyt–b*	KF514189	MNCN–279686	2
	29º 36' 23.65" N, 110º 7' 29.57" W				
G. eremica	3 - San Miguel River sub–basin,	*cyt–b*	KX855971	USON–1304–1	1
	San Miguel River at Cucurpe, Sonora	*cyt–b*	KX855972	USON–1304–6	1
	30º 20' 26.59" N, 118º 41' 37.09" W	*nd2*	KX858657	USON–1304–6	1
		cox1	MH091965	USON–1304–1	1
		cox1	MH091966	USON–1304–2	1
		cox1	MH091967	USON–1304–4	1
		cox1	MH091968	USON–1304–5	1
		cox1	KX858665	USON–1304–6	1
		cox1	MH091969	USON–1304–7	1
		cox1	MH091970	USON–1304–8	1

Table 1. (Cont.)

Taxon	Locality	Gene	No	Voucher	S
		cox1	MH091971	USON–1304–11	1
		cox1	MH091972	USON–1304–13	1
		cox1	MH091973	USON–1304–14	1
		cox1	MH091974	USON–1304–15	1
		cox1	MH091975	USON–1304–16	1
		cyt–b	JX443052	UAIC–15297	2
Mátape River basin					
G. eremica	4 - Mátape River sub–basin,	*cyt–b*	KX855973	USON–1376–4	1
	Mátape River at Mazatán, Sonora	*cyt–b*	KX855974	USON–1376–9	1
	29º 59' 56.04" N, 100º 8' 51.25" W	*cyt–b*	KX855975	USON–1376–25	1
		nd2	KX858658	USON–1376–4	1
		nd2	KX858659	USON–1376–9	1
		nd2	KX858660	USON–1376–25	1
		cox1	MH091976	USON–1376–1	1
		cox1	MH091977	USON–1376–2	1
		cox1	MH091978	USON–1376–3	1
		cox1	MH091979	USON–1376–4	1
		cox1	MH091980	USON–1376–5	1
		cox1	MH091981	USON–1376–6	1
		cox1	KX858666	USON–1376–9	1
		cox1	MH091982	USON–1376–11	1
		cox1	MH091983	USON–1376–12	1
		cox1	MH091984	USON–1376–13	1
		cox1	MH091985	USON–1376–14	1
		cox1	MH091986	USON–1376–15	1
G. eremica	5 - Mátape River sub–basin,	*cyt–b*	KF514193	UAIC–15296	2
	Mátape River just W San José de Pimas				
	on Hwy 16, Sonora				
	28º 43' 7.82" N, 110º 20' 53.43" W				
G. cf. *eremica*	6 - Arroyo El Tigre sub–basin,	*cyt–b*	KX855976	USON–1300–17	1
	La Balandrona Canyon,	*cyt–b*	KX855977	USON–1300–18	1
	Sierra El Aguaje mountains, Sonora	*cyt–b*	KX855978	USON–1300–29	1
	28º 2' 38.04" N, 111º 4' 21.98" W	*nd2*	KX858649	USON–1300–17	1
		nd2	KX858650	USON–1300–18	1
		nd2	KX858651	USON–1300–29	1
		cox1	KX858667	USON–1300–17	1
		cox1	MH091928	USON–1300–18	1
		cox1	MH091929	USON–1300–19	1
		cox1	MH091930	USON–1300–20	1
		cox1	MH091931	USON–1300–21	1

Table 1. (Cont.)

Taxon	Locality	Gene	No	Voucher	S
		cox1	MH091932	USON–1300–22	1
		cox1	MH091933	USON–1300–23	1
		cox1	MH091934	USON–1300–24	1
		cox1	MH091935	USON–1300–25	1
		cox1	MH091936	USON–1300–26	1
		cox1	MH091937	USON–1300–27	1
		cox1	MH091938	USON–1300–28	1
		cox1	MH091939	USON–1300–29	1
		cox1	MH091940	USON–1300–30	1
		cox1	MH091941	USON–1300–31	1
G. cf. *eremica*	7 Arroyo El Tigre sub–basin,	*cyt–b*	KX855981	USON–1302–7	1
	La Pirinola Canyon,	*cyt–b*	KX855979	USON–1302–12	1
	Sierra El Aguaje mountains, Sonora	*cyt–b*	KX855980	USON–1302–13	1
	28º 5' 32" N, 111º 2' 15" W	*nd2*	KX858654	USON–1302–7	1
		nd2	KX858652	USON–1302–12	1
		nd2	KX858653	USON–1302–13	1
		cox1	MH091942	USON–1302–2	1
		cox1	MH091943	USON–1302–3	1
		cox1	MH091944	USON–1302–4	1
		cox1	MH091945	USON–1302–5	1
		cox1	MH091946	USON–1302–6	1
		cox1	KX858668	USON–1302–7	1
		cox1	MH091947	USON–1302–8	1
		cox1	MH091948	USON–1302–9	1
		cox1	MH091949	USON–1302–10	1
		cox1	MH091950	USON–1302–11	1
		cox1	MH091951	USON–1302–13	1
		cox1	MH091952	USON–1302–14	1
		cox1	MH091953	USON–1302–15	1
		cox1	MH091954	USON–1302–16	1
Yaqui River basin					
G. purpurea	8 - Bavispe River sub–basin,	*cyt–b*	JX443020	BYU–14072	2
	Arroyo San Bernardino at US/MX border	*nd2*	KX858661	USON–1378–1	1
	31º 19' 57.37" N, 109º 15' 35.17" W	*cox1*	KX858669	USON–1378–1	1
		cox1	MH091987	USON–1378–2	1
		cox1	MH091988	USON–1378–3	1
		cox1	MH091989	USON–1378–4	1
		cox1	MH091990	USON–1378–5	1
		cox1	MH091991	USON–1378–6	1
		cox1	MH091992	USON–1378–7	1
		cox1	MH091993	USON–1378–8	1

Table 1. (Cont.)

Taxon	Locality	Gene	No	Voucher	S
G. minacae	9 - Bavispe River sub–basin,	*cyt–b*	KF514195	UAIC–14983	2
	Arroyo El Largo, 2.5 km E Ejido	*nd2*	KX858663	USON–1224–1	1
	Arroyo El Largo, Sonora	*cox1*	KX858671	USON–1224–1	1
	29º 44' 3.9" N,108º 36' 48.6" W				
De la Concepción River basin					
G. ditaenia	10 - Magdalena River sub–basin,	*cyt–b*	JX443022	UAIC–15299	2
	Magdalena River at road crossing	*nd2*	KX858662	USON–1377–1	1
	to San Ignacio–Terrenate, Sonora	*cox1*	KX858670	USON–1377–1	1
	30º 41' 50.7" N,110º 55' 39.5" W	*cox1*	MH091994	USON–1377–2	1
		cox1	MH091995	USON–1377–3	1
		cox1	MH091996	USON–1377–4	1
		cox1	MH091997	USON–1377–5	1
		cox1	MH091998	USON–1377–6	1
		cox1	MH091999	USON–1377–7	1
		cox1	MH092000	USON–1377–8	1
		cox1	MH092001	USON–1377–9	1
		cox1	MH092002	USON–1377–10	1

in Sonora, Mexico, were collected between April 2000 and November 2015 (fig. 1, table 1). Individuals of the *G. eremica* lineage included samples of *G.* cf. *eremica* collected from large spring–fed pools in the intermittent–flowing arroyos of the two subtropical canyons, La Balandrona (n = 30) and La Pirinola (n = 30), located in the southeastern sector of the Sierra El Aguaje mountains (fig. 1, inset). Specimens of *G. eremica* were collected from its known distribution in the Sonora (Sonora River sub–basin, n = 30; San Miguel River sub–basin, n = 30) and Mátape (Mátape River sub–basin, n = 30) river drainages. Samples of *G. purpurea* were collected from the Arroyo San Bernardino (n = 8) in the Bavispe River sub–basin of the extensive Yaqui River system. Specimens of *G. ditaenia* (n = 10) were collected from the Magdalena River sub–basin of the De la Concepción River basin, and specimens of *G. minacae* (n = 10) were obtained from Arroyo El Largo in the Bavispe River sub–basin, Yaqui River system (fig. 1, table 1). In the field, specimens for genetic analyses were labeled and tissue from a pelvic fin was removed and preserved in absolute ethanol and stored at 4 ºC until DNA extraction. After fin–clipping, the specimens were preserved in 10 % buffered formalin and later transferred to 50 % ethanol for deposition as vouchers in the Native Fish Collection of the Departamento de Investigaciones Científicas y Tecnológicas (DICTUS) of the Universidad de Sonora (USON) in Hermosillo.

DNA extraction and PCR amplification

Total DNA was obtained from fin tissue of all collected specimens following protocols of an extraction kit, the QIAamp DNA Mini Kit (QIAGEN). Amplification reactions were performed in a total volume of 50 µl using GoTaq Colorless Master Mix (Promega). The mitochondrial gene *cyt–b* was totally amplified (1140 bp) using the primers FW–L15058 5'–TGA CTT GAA AAM CCA CCG TTG–3' and RV: H16249 5'–TCA GTC TCC GGT TTA CAA GAC–3' as reported by Kocher et al. (1989). The total sequence (1047 bp) of mitochondrial gene *nd2* was amplified with primers ND2F: 5'–AAC CCA TRC YCA AGA GAT CA–3' and ND2R: 5'–ACT TCT RCT TAR AGC TTT GAA GG–3', designed for other sequences of the genus as reported in GenBank. Conditions for the amplification of *cyt–b* and *nd2* comprised an initial denaturation for 5 min at 94 ºC followed by 35 cycles of 50 s at 94 ºC, 50 s annealing temperature at 50 ºC for *cyt–b* and 60 ºC for *nd2,* and then 2 min at 72 ºC. The final extension was performed at 72 ºC for 7 min. A region of 651 bp of the mitochondrial gene *cox1* was amplified with the primers FishF2_t1: 5'–TCT ACA AAY CAC AAA GAC ATT GGT AC–3' and FishR2_t1: 5'–ACC TCT GGG TGR CCA AAG AAT CAG AA–3', modified of Ivanova et al. (2007) to make them more specific to *Gila*. Conditions for amplification of *cox1* comprised an initial denaturation for 5 min at 94 ºC followed by 34 cycles of 50 s at 94 ºC, 50 s at 50 ºC, and 1 min at

72 ºC. The final extension was performed at 72 ºC for 7 min. The PCR products were sent to Macrogen, Inc., Seoul, South Korea for purification and bidirectional sequencing according to the company's specifications.

Phylogenetic inferences

Obtained sequences were edited and assembled by overlapping using Chromas Pro 1.6 (Technelysium Pty Ltd, South Brisbane, Queensland, Australia). Each gene was identified using BLAST searches (Altschul et al., 1990) against GenBank data. Sequence divergence for the members of the *G. eremica* lineage, including *G*. cf. *eremica* and *G. purpurea*, was analysed using MEGA v5 (Tamura et al., 2011). Phylogenetic relationships of *Gila* cf. *eremica* from both La Balandrona and La Pirinola canyons were first evaluated via mitochondrial gene *cyt–b*, with sequences used by Schönhuth et al. (2014) for their phylogenetic inferences of the genus, including other members of their Revised Western Clade, plus our sequences of the *G. eremica* lineage. The evolutionary relationships of the two *G.* cf. *eremica* populations were corroborated using concatenated sequences of the mitochondrial genes *cyt–b, nd2* and *cox1* of all specimens obtained in this study plus sequences of congeners available from GenBank.

Phylogenetic trees using Maximum Likelihood (ML) for *cyt–b* and the concatenated mitochondrial gene dataset (*cyt–b*, *nd2* and *cox1*) were estimated using RAxML–HPC2 on XSEDE 8.0.24 (Stamatakis, 2014). The JModeltest2 software (Darriba et al., 2012) was used to find the best nucleotide substitution models for each dataset separately. We defined data blocks based on genes, and used the Akaike information Criterion (Posada and Buckley, 2004), and the best–fit model was used for the subsequent analyses. ML trees were performed on the CIPRES Science Gateway 3.3 (Miller et al., 2010), using GTRGAMMA model, and 1,000 bootstraps pseudoreplications (Felsenstein, 1985) to estimate the node reliability. Bayesian inference (BI) analyses were conducted for each gene data set using MrBayes v3.2.6 (Ronquist and Huelsenbeck, 2003). The Akaike information criterion (AIC) implemented in JModelTest2 (Posada and Buckley, 2004; Darriba et al., 2012) was used to identify the optimal molecular evolutionary model for each partition block on each sequence data set of the analysis. For BI, ten million cycles were implemented in four simultaneous Monte Carlo Markov chains; sampling the Markov chain at intervals of 1,000 generations. Log–likelihood stability was attained after 100,000 generations; the first 1,000 trees were discarded as burn–in in each analysis. The remaining trees were used to compute a 25 % majority rule consensus tree in PAUP* (Swoford, 2002). Support for BI tree nodes was determined based on values of Bayesian posterior probabilities. Final trees of ML and BI were edited using FigTree 1.4.2 (Rambaut, 2014).

DNA barcode and *cox1* sequence analysis

DNA barcode analyses included samples of nominal *Gila eremica* populations from the Sonora and Mátape River basins, samples of *G*. cf. *eremica* from the Arroyo El Tigre sub–basin's La Balandrona and La Pirinola canyons, *G. purpurea* from Arroyo San Bernardino, and *G. ditaenia* as outgroup. *Cox1* sequences were edited and assembled by overlapping using Chromas Pro v1.6 (Technelysium Pty Ltd, South Brisbane, Queensland, Australia). The values of haplotype (H), nucleotide diversity (π) (Nei, 1987), and the polymorphic/variable sites were estimated with DnaSP v5.0 (Librado and Rozas, 2009). Genetic uncorrected *p* distance analysis between groups was performed using MEGA v5 (Tamura et al., 2011). A neighbor–joining (NJ) 50 % majority–rule consensus tree was constructed in PAUP* (Swoford, 2002) over 1,000,000 bootstrap replicates, using K2P (Kimura, 1980). The obtained tree was incorporated into the NEXUS file of the DNA data matrix of *Gila* species using Mesquite v3.10 (Maddison and Maddison, 2016), according to the specifications of Jörger and Schrödl (2014). The NEXUS/TREE file was carried out in CAOS software package (Sarkar et al., 2008) to identify diagnostic single pure characters (sPu) which are present in all members of a clade but absent from all members of another clade. These sPu characters were used for taxa discrimination of our groups of interest.

Results

Phylogenetic relationships within the *Gila eremica* lineage

Amplified sequences of the mitochondrial gene *cyt–b* for all specimens of the *G. eremica* lineage analyzed had a length of 1,140 bp with no insertions or deletions. Thirty–four positions were variable sites, 15 of which were parsimony informative. Individuals of the *G. eremica* lineage using mitochondrial gene *cyt–b* were classified into 19 haplotypes. Haplotypes 8 and 10 were shared between individuals from the Sonora and San Miguel Rivers sub–basins (Sonora River drainage), and none of the remaining haplotypes were shared with any other population. Sequence divergence (uncorrected *p* distance) within the *G. eremica* lineage ranged between 0.26 % and 1.11 % with an average of 0.7 %. Genetic divergence of *G. eremica* populations from the Sonora, San Miguel and Mátape Rivers sub–basins compared to *G.* cf. *eremica* from La Balandrona and La Pirinola canyons (Arroyo El Tigre sub–basin) was 0.44–1.14 %. The mean distance between all individuals of *G. eremica* and *G*. cf. *eremica* was 0.83 %. The genetic divergence for individuals of *G. eremica* (but not including *G.* cf. *eremica*) against *G. purpurea* was 1.93–2.72 %. Genetic divergence between *G.* cf. *eremica* and *G. purpurea* ranged from 2.11–2.46 %. Mean nucleotide frequencies within nominal *G. eremica* populations were 26.30 % A, 29.57 % T, 27.48 % C, and 16.65 % G. Mean nucleotide frequencies for *G*. cf. *eremica* populations were 26.58 % A, 29.59 % T, 27.56 % C, 16.37 % G. The estimated Transition/Transversion bias (*R*) for both groups was 2.27. Substitution pattern and rates were estimated under the General Time Reversible model (GTR) (Nei and Kumar, 2000).

Phylogenetic analyses of *cyt–b* using ML and BI with all species analyzed of the Revised Western Clade of Schönhuth et al. (2014), and including our samples of *G. eremica* and *G*. cf. *eremica*, showed the same topology with variations in nodal support of bootstrap probabilities (BP) for ML and posterior probabilities (PP) for BI (fig. 2). The tree topology was consistent with that obtained by Schönhuth et al. (2014) for members of the *Gila* lineage (excluding *G.* cf. *eremica*). Members of the *G. eremica* lineage, including *G.* cf. *eremica* of the present study, were always resolved as monophyletic, with *G. purpurea* as the sister species (BP = 98 %, PP = 100 %) (fig. 2). The *G. eremica* lineage was resolved with a clade for the Sonora River sub–basins (Sonora and San Miguel Rivers), and a sister clade for the Mátape River sub–basins (Mátape River and Arroyo El Tigre) (fig. 2). Individuals of *G.* cf. *eremica* from La Balandrona and La Pirinola canyons of the Arroyo El Tigre sub–basin were nested together sharing a putative common ancestor, corroborating these two isolated populations as unequivocal members of the *G. eremica* lineage (fig. 2).

Phylogenetic analyses by ML and BI, using the concatenated results of mitochondrial genes *cyt–b*, *nd2*, and *cox1*, included *Gila robusta*, *G. ditaenia*, *G. purpurea*, members of the *G. eremica* lineage (including *G.* cf. *eremica*), with *G. minacae* as outgroup. The tree topology resulting from the analyses was the same for both criteria, with variations in the nodal support values of BP for ML and PP for BI (fig. 3). The analyses showed that members of the *G. eremica* lineage are monophyletic (BP = 100 %, PP = 100 %) with *G. purpurea* as sister species (BP = 100 %, PP = 100 %), and corresponded with our results from *cyt–b* in regards to monophyly within the *G. eremica* lineage (figs. 2, 3). The geographical clade for all members of the Mátape River basin, including those from the two isolated Arroyo El Tigre canyons, was better supported in the concatenated genes' analyses (fig. 3, BP = 76 %, PP = 95 %) compared to using *cyt–b* alone (fig. 2). Individuals of *G.* cf. *eremica* from both canyons in the Arroyo El Tigre sub–basin were supported in a clade of specific identity (fig. 3, BP = 94 %, PP = 100 %) and indicating relationship with *G. eremica* from the Mátape River sub–basin as putative closest relative (fig. 3).

DNA barcoding analysis of the Gila eremica lineage

The analyses of 82 *cox1* sequences of several *Gila* species showed a total of nine haplotypes: five for *G. eremica*, one for *G.* cf. *eremica*, one for *G. purpurea*, and two for *G. ditaenia.* The *G. eremica* lineage, including *G.* cf. *eremica*, showed a haplotype and nucleotide diversity of 0.6093 (SD = 0.034) and 0.00125 (SD = 0.00112), respectively. Individuals of *G. eremica* from the Sonora, San Miguel and Mátape Rivers sub–basins shared at least one haplotype, and also showed unique haplotypes for the Sonora and San Miguel rivers sub–basins. The analysis did not detect shared haplotypes among *G.* cf. *eremica* and all the *G. eremica* populations. The sequences of taxa included in the DNA barcoding analyses showed 46 polymorphic sites (table 2). The genetic uncorrected *p* distances analysis between groups produced a value of 1.27 % for populations of *G. eremica* and *G. purpurea*. The distance value between *G.* cf. *eremica* and *G. purpurea* was 1.38 %, and the value between nominal *G. eremica* and *G.* cf. *eremica* was 0.20 %.

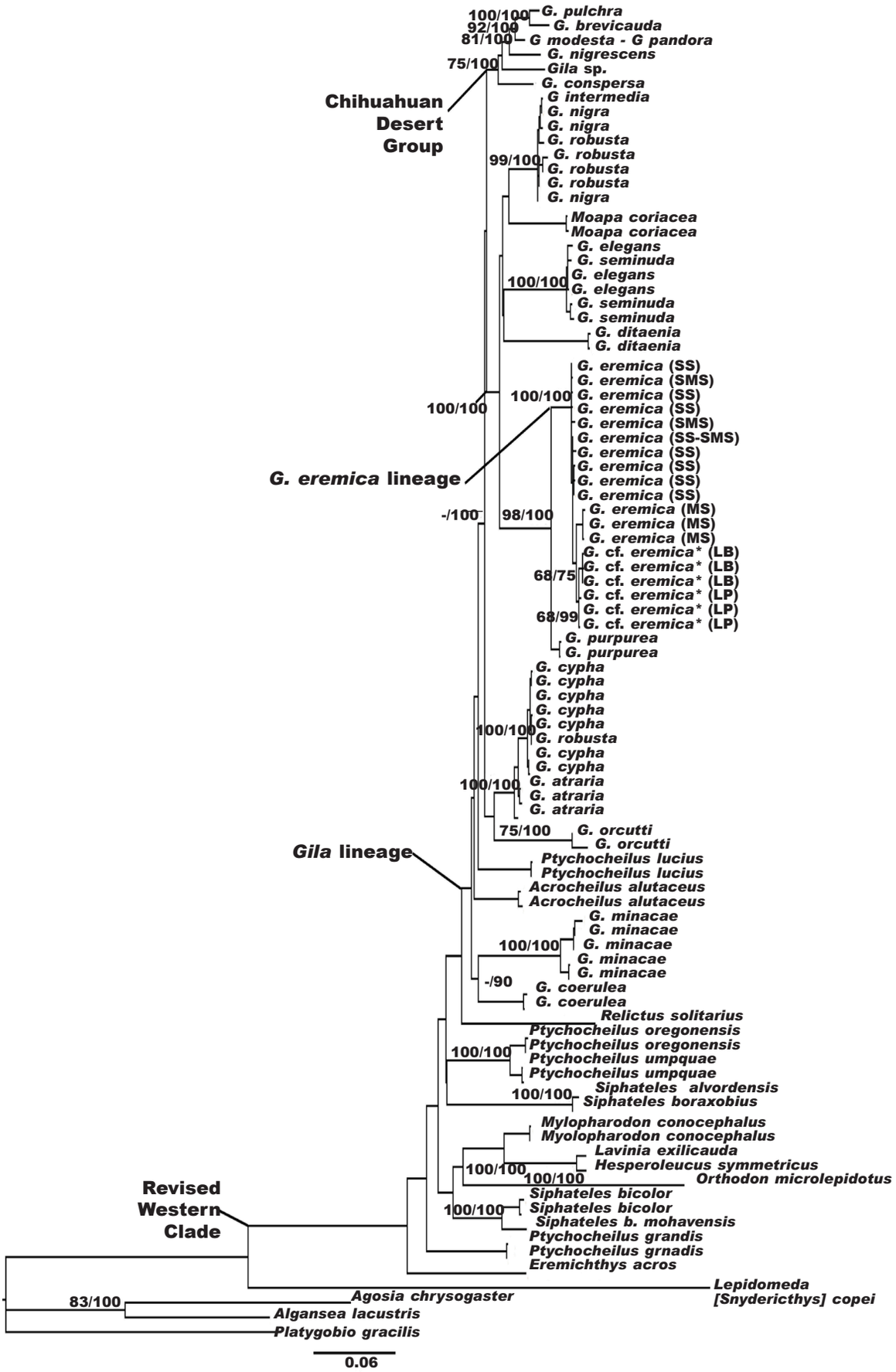
The DNA barcoding analysis using the character–based approach with CAOS showed eight single pure characters (sPu) to discriminate *G. eremica* from *G. purpurea*, nine sPu to discriminate *G.* cf. *eremica* from *G. purpurea,* and one fixed sPu in the 29 analyzed sequences of *G.* cf. *eremica*, discriminating it from *G. eremica* (table 2).

Discussion

The monophyly of the *G. eremica* lineage (Sonora and Mátape River basins) and *G. purpurea* (Bavispe River in northern headwaters of the Yaqui River basin) obtained in the present study was highly supported by both ML and BI criteria, as previously suggested by morphological analyses by DeMarais (1991) and molecular data by Schönhuth et al. (2014). Phylogenetic relationships inferred here for both ML and BI using the *cyt–b* gene alone and the concatenated set of genes *cyt–b*, *nd2* and *cox1,* support with high values of posterior probabilities, the monophyly of the *G. eremica* lineage for all its members, and corroborated *G*. cf. *eremica* as a member of the lineage (figs. 2, 3). The monophyly, geographical clades, and low genetic divergence detected here within the *G. eremica* lineage may be explained by relatively recent isolation of once–connected drainages inhabited by this lineage, as suggested for other nominal species of *Gila* occurring in México (Schönhuth et al., 2014). In addition, we provide evidence for the existence of two geographical clades for the Sonora and Mátape river basins, with high scores for both ML and BI criteria (figs. 2, 3).

The close relationship and low genetic divergence between *G. eremica* from the Mátape River and *G*. cf. *eremica* from Arroyo El Tigre sub–basin (figs. 2, 3) suggest a putative common ancestor for these populations and indicate a relatively recent connection between the two sub–basins. The phylogenetic analysis also supports *G.* cf. *eremica* as a clade of specific identity, apart from other members of the lineage, but closely related to populations of *G. eremica* in the Mátape River sub–basin. The morphological differences detected between *G.* cf. *eremica* and *G. eremica* (Ballesteros–Córdova et al., 2016), along with its phylogenetic position resolved in the present study, suggest that *G.* cf. *eremica* is an evolutionary significant unit within the *G. eremica* lineage that requires additional species delimitation methods such as DNA barcoding for further discrimination.

The effectiveness of character–based methods (e. g., CAOS) for taxa detection relies on the use of diagnostic characters, as those used in traditional taxonomy. The character–based method is based on the premise that members of a given taxon share a distinctive combination or combinations of diagnostic


100/100
92/100
81/100
75/100
G. pulchra
G. brevicauda
G modesta - G pandora
G. nigrescens
Gila sp.
G. conspersa
Chihuahuan Desert Group
G intermedia
G. nigra
G. nigra
G. robusta
99/100
G. robusta
G. robusta
G. robusta
G. nigra
Moapa coriacea
Moapa coriacea
G. elegans
G. seminuda
100/100
G. elegans
G. elegans
G. seminuda
G. seminuda
G. ditaenia
G. ditaenia
G. eremica (SS)
G. eremica (SMS)
100/100
100/100
G. eremica (SS)
G. eremica (SS)
G. eremica (SMS)
G. eremica (SS-SMS)
G. eremica (SS)
G. eremica (SS)
G. eremica (SS)
G. eremica (SS)
G. eremica lineage
-/100
98/100
G. eremica (MS)
G. eremica (MS)
G. eremica (MS)
G. cf. eremica* (LB)
68/75
G. cf. eremica* (LB)
G. cf. eremica* (LB)
G. cf. eremica* (LP)
68/99
G. cf. eremica* (LP)
G. cf. eremica* (LP)
G. purpurea
G. purpurea
G. cypha
G. cypha
G. cypha
G. cypha
G. cypha
100/100
G. robusta
G. cypha
G. cypha
100/100
G. atraria
G. atraria
G. atraria
75/100
G. orcutti
G. orcutti
Gila lineage
Ptychocheilus lucius
Ptychocheilus lucius
Acrocheilus alutaceus
Acrocheilus alutaceus
G. minacae
G. minacae
100/100
G. minacae
G. minacae
G. minacae
-/90
G. coerulea
G. coerulea
Relictus solitarius
Ptychocheilus oregonensis
100/100
Ptychocheilus oregonensis
Ptychocheilus umpquae
Ptychocheilus umpquae
100/100
Siphateles alvordensis
Siphateles boraxobius
Mylopharodon conocephalus
Myolopharodon conocephalus
Lavinia exilicauda
100/100
Hesperoleucus symmetricus
100/100
Orthodon microlepidotus
Revised Western Clade
Siphateles bicolor
100/100
Siphateles bicolor
Siphateles b. mohavensis
Ptychocheilus grandis
Ptychocheilus grnadis
Eremichthys acros
Lepidomeda [Snyderichthys] copei
83/100
Agosia chrysogaster
Algansea lacustris
Platygobio gracilis
0.06

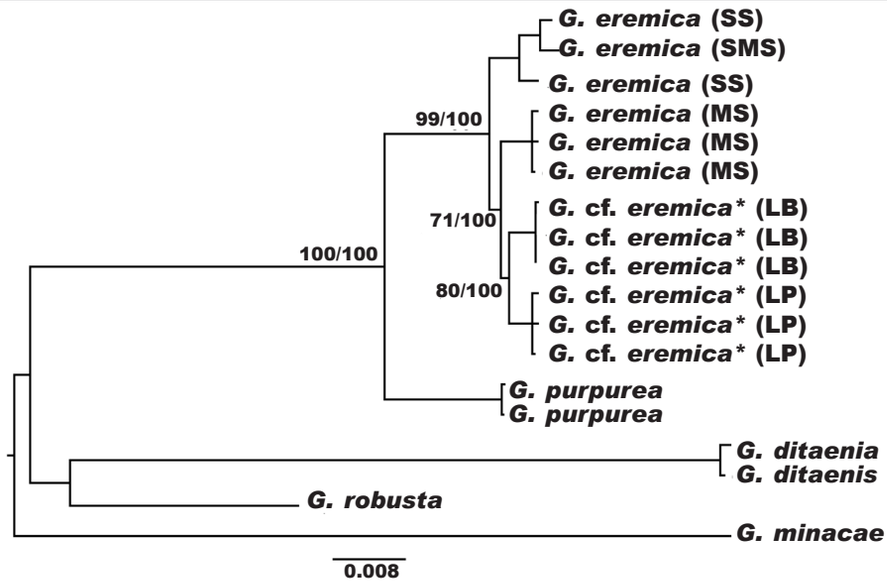


Fig. 3. Recovered phylogenetic tree via maximum likelihood and posterior probabilities of Bayesian inference of the concatenated genes *cyt–b*, *cox1*, and *nd2* for populations of the *Gila eremica* lineage plus other three species examined in this study (SS, Sonora River sub-basin; SMS, San Miguel River sub-basin; MS, Mátape River sub-basin; LB, La Balandrona Canyon; LP, La Pirinola Canyon. Numbers on branches represent values of ML bootstap probabilities (BP > 65 %)/BI posterior probabilities (PP > 70 %). (* specimens from Arroyo El Tigre sub–basin).

Fig. 3. Árbol filogenético recuperado mediante el método de la máxima verosimilitud y probabilidades a posteriori de inferencia bayesiana de los genes concatenados cyt–b, cox1 *y* nd2 *para las poblaciones del linaje de* Gila eremica *y otras tres especies examinadas en este estudio (SS, subcuenca del río Sonora; SMS, subcuenca del río San Miguel; MS, subcuenca del río Mátape; LB, cañón La Balandrona; LP, cañón La Pirinola). Los números sobre las ramas representan las probabilidades de remuestreo de ML (BP > 65%)/ probabilidades a posteriori BI (PP > 70%). (* especímenes de la subcuenca del arroyo El Tigre).*

character attributes (e. g., polymorphisms) that are absent in related groups, and these attributes can be used for species discrimination (Rach et al., 2008; Sarkar et al., 2008; Bergmann et al., 2009; Zou et al., 2011; Jörger and Schrödl, 2013; Zou and Li, 2016). Despite the proposal to use more than three characters as a DNA barcoding gap to separate natural taxonomic groupings (Rach et al., 2008, Yassin et al., 2010; Zou et al., 2011; Yu et al., 2014), Jörger and Schrödl (2013) argue that CAOS does not

Fig. 2. Recovered tree of phylogenetic relationships via Maximum Likelihood (GTR + G + I model) for haplotypes of all members of the *Gila* lineage, including other species of the Revised Western Clade (Schönhuth et al., 2012, 2014), using mitochondrial gene *cyt–b.* Numbers on branches represent ML bootstap probabilities (BP > 65 %)/BI posterior probabilities (PP > 70 %): SS, Sonora River sub-basin; SMS, San Miguel River sub-basin; MS, Mátape River sub-basin; LB, La Balandrona Canyon; LP, La Pirinola Canyon (* specimens from Arroyo El Tigre sub–basin).

Fig. 2. Árbol recuperado de relaciones filogenéticas mediante el método de la máxima verosimilitud (Modelo GTR + G + I) de los haplotipos de todos los miembros del linaje de Gila*, incluidas otras especies del Revised Western Clade (Schönhuth et al., 2012, 2014), utilizando el gen mitocondrial* cyt–b. *Los números sobre las ramas representan la probabilidad de remuestreo de ML (BP > 65%)/probabilidad a posteriori BI (PP > 70%): SS, subcuenca del río Sonora; SMS, subcuenca del río San Miguel; MS, subcuenca del río Mátape; LB, cañón La Balandrona; LP, cañón La Pirinola (* especímenes de la subcuenca del arroyo El Tigre)*

Table 2. Polymorphic sites and composite of attribute characters obtained by CAOS in the first 651 bp of the mitochondrial gene *cox1* for *G*. cf. *eremica* compared to nominal *G*. *eremica* populations, *G*. *purpurea* and *G*. *ditaenia*. Numbers at the top indicate variable sites of the fragment studied in this study. The far right column shows the number of individuals sharing each haplotype. Pure diagnostic characters among *G*. cf. *eremica*, *G*. *eremica* populations and *G*. *purpurea* are in bold. Single pure characters between *G*. cf. *eremica* and *G*. *eremica* are shaded and in bold: LB, La Balandrona Canyon; LP, La Pirinola Canyon; S, Sonora River; SM, San Miguel River; M, Mátape River.

	0	0	0	0	0	0	0	0	0	1	1	1	1	1	1	1	2	2	2	2	3
	0	2	2	3	3	4	4	8	8	2	4	6	7	7	7	9	9	3	6	7	2
	4	0	3	4	7	0	6	2	5	4	5	6	0	5	8	0	3	2	2	4	5
G. cf. *eremica* (LB)	**C**	G	T	G	G	A	**T**	C	A	T	**G**	G	C	**T**	G	T	A	C	C	G	G
G. cf. *eremica* (LP)	**C**	G	T	G	G	A	**T**	C	A	T	**G**	G	C	**T**	G	T	A	C	C	G	G
G. eremica (S, SM, M)	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–
G. eremica (SR, SMR)	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–
G. eremica (SR)	–	C	C	–	–	–	–	–	–	–	–	–	–	–	–	–	G	–	–	–	–
G. eremica (SR)	–	–	–	–	–	–	–	–	–	–	–	–	G	–	–	–	–	–	–	–	A
G. eremica (SMR)	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	C	–	–	–	–	–
G. purpurea	**T**	–	–	–	–	–	**C**	–	–	–	**A**	–	–	**C**	–	–	–	–	–	–	–
G. ditaenia	T	–	–	A	A	G	–	T	G	C	A	A	–	–	C	C	–	T	T	A	A
G. ditaenia	T	–	–	–	A	G	–	T	G	C	A	A	–	–	C	C	–	T	T	A	A

possess an objective criterion with which to delimit a threshold number of distinguishing nucleotides that would indicate a species boundary (i. e., to delimit a probable new species) or an independent population belonging to the same species. On the other hand, the supposition that a characteristic attribute has been fixed within a population gains more confidence if a higher number of samples is analyzed (Rach et al., 2008). According to Zhang et al. (2010), the desired sample size for a DNA barcoding analysis should range from 9.5 to 216.6, which is within the range of samples of *G*. cf. *eremica* and *G. eremica* analyzed here. Our analysis revealed a single fixed polymorphism in all 29 examined sequences of *G.* cf. *eremica* that was absent from all analyzed specimens of the *G. eremica* populations. The unique polymorphism detected here in *G.* cf. *eremica* with respect to *G*. *eremica* represents a sPu character and thus reveals an apomorphy within the *G. eremica* lineage. This indicates a different pattern of genetic variation for *G*. cf. *eremica* compared with *G. eremica* and also with the close congener *G. purpurea*. The study by Rach et al. (2008) identified less than three diagnostic characters in ten very closed related sister taxa of the insect order Odonata using mitochondrial gene *ndh1*, showing similar results to those obtained here between populations of *G*. cf. *eremica* and *G*. *eremica*.

The close relationship and low genetic divergence of the two *G.* cf. *eremica* populations with respect to *G. eremica* obtained in our several molecular analyses bolsters the morphological differences previously detected (Ballesteros–Córdova et al., 2016). Such differences may reflect a general phenotypic plasticity of freshwater fishes to adapt to environmental alterations, or variation in stream size, flow and substrate (Hubbs, 1940), leading in our case to a morphological variant within the *Gila eremica* lineage. However, the morphological differences seen in the *G. eremica* lineage members (Ballesteros–Córdova et al., 2016) are consistent with the establishment of a fixed polymorphism in the geographically isolated *G.* cf. *eremica*. Moreover, none of our phylogenetic analyses nested individuals of *G*. cf. *eremica* with samples of *G*. *eremica,* thus revealing this potentially ESU as a natural group. Similarly, the character–based method using *cox1*, and sequences analyses of *cyt–b* and its concatenation with *nd*2 and *cox1* showed a different pattern of variation in *G*. cf. *eremica* compared with *G. eremica*, and with *G. purpurea.* However, current results will need to be further tested using nuclear data.

The process of species identification through DNA barcoding has often been confused with species discovery (DeSalle, 2006). Species identification has been considered a valid use for the DNA barcode, which does not rely on any particular species concept (Rach et al., 2008). This appears to be because species identification using DNA barcode is consistent with any concept of species a taxonomist may

Tabla 2. Sitios polimórficos y composición de los atributos de carácter obtenidos con CAOS en las primeras 651 pb del gen mitocondrial cox1 *de* G. *cf.* eremica *comparados con las poblaciones de* G. eremica, G. purpurea *y* G. ditaenia*. Los números de la parte superior indican los sitios variables del fragmento estudiado. La columna de la derecha muestra el número de individuos que comparten cada haplotipo. Los caracteres diagnósticos puros entre las poblaciones de* G. *cf.* eremica*,* G. eremica *y* G. purpurea *están en negrita. Los caracteres únicos puros entre* G. *cf.* eremica *y* G. eremica *están sombreados, en negrita: LB, cañón La Balandrona; LP, cañón La Pirinola; S, río Sonora; SM, río San Miguel; M, río Mátape.*

3	3	3	3	3	3	3	3	4	4	4	5	5	5	5	5	5	5	6	6	6	6	6	6	6	
5	5	6	7	7	8	8	9	5	6	6	0	1	1	4	4	6	8	0	1	1	1	2	3	4	
2	5	7	6	9	2	3	7	7	0	3	5	4	7	1	4	2	9	1	3	6	9	8	4	3	
C	C	G	G	**A**	T	C	**G**	T	T	A	**G**	G	C	A	G	T	A	**C**	T	A	A	**A**	C	C	15
C	C	G	G	**A**	T	C	**G**	T	T	A	**G**	G	C	A	G	T	A	**C**	T	A	A	**A**	C	C	14
–	T	–	–	–	–	–	–	–	–	–	**A**	–	–	–	–	–	–	–	–	–	–	–	–	–	27
–	–	–	–	–	–	–	–	–	–	–	**A**	–	–	–	–	–	–	–	–	–	–	–	–	–	4
–	–	–	–	–	–	–	–	–	–	–	**A**	–	–	–	–	–	–	–	–	–	–	–	–	–	1
–	–	–	–	–	–	–	–	–	–	–	**A**	–	–	–	–	–	–	–	–	–	–	–	–	–	1
–	–	–	–	–	–	–	–	–	–	–	**A**	–	–	–	–	–	–	–	–	–	–	–	–	–	1
–	–	–	–	**G**	–	–	**A**	–	–	–	**A**	–	–	–	–	–	–	**T**	–	–	–	**G**	–	–	8
T	–	A	A	–	C	–	A	C	C	G	–	A	T	G	A	C	G	–	C	C	G	–	T	T	1
T	–	A	A	–	C	–	A	C	C	G	–	A	T	G	A	C	G	–	C	C	G	–	T	T	9

use when identifying a named species (Rach et al., 2008). The CAOS approach in DNA barcoding is considered by many as an accurate and available method for testing species boundaries. However, this method requires *a priori* defined groups, making it difficult or unsuitable for species discovery (Jörger and Schrödl, 2013; Kekkonen et al., 2015). Species discovery involves a more complicated task because it requires a recognized species concept along with a traditional taxonomic corroboration system (DeSalle et al., 2005; DeSalle, 2006). Accordingly, a single source of data, be it molecular, morphological, ecological or ethological, is not able by and of itself to be used for species discovery (Rach et al., 2008). However, comparison of DNA sequences can be used to detect potentially new species which then need corroboration by an integrated taxonomic approach using a species concept (Rubinoff, 2006; Rach et al., 2008).

The phylogenetic data and DNA barcoding results obtained here, coupled with those from the morphological analyses for *G.* cf. *eremica* (Ballesteros–Córdova et al., 2016) and its geographic isolation supports it as a natural evolutionary significant unit within *G. eremica*. The low genetic distance detected between *G*. cf. *eremica* and nominal *G*. *eremica*, along with their phylogenetic affinities indicates a relatively recent disruption within this lineage. Our data thus contribute to the knowledge of systematics and evolution of the greater *Gila* lineage. The recognition of an isolated taxon of *Gila* in the Arroyo El Tigre sub–basin of the Sierra El Aguaje reveals a potential microendemism for the genus in subtropical canyons of this region of Northwest Mexico. The evidence presented here calls for further studies aimed at clarifying the biology, origin and history of *G.* cf. *eremica* populations and contributes to increased understanding of the evolution and conservation of fish species inhabiting arid and semiarid regions in Mexico and the USA.

Acknowledgements

Many thanks to Ramón Villafaña and José Ines–Ramírez for facilitating access to sampling sites for *Gila* cf. *eremica*. We also thank Michael T. Bogan, Nohemí Noriega–Félix, Sylvette L. Vidal–Aguilar, Celso Haros–Méndez, Emmanuel M. Bernal–Loaiza, and Dylann Córdova–Martínez for assistance in sampling. Dr. José Said Gutiérrez–Ortega provided advice in the phylogenetical analyses. Field collections were made under Mexican government permits DGOPA.03947.250406.1606 and SGPA/DGVS/00505/10. The first author received a fellowship from the Consejo Nacional de Ciencia y Tecnología (CONACyT) for doctoral studies. This work was funded primarily by CONACyT. Additional thanks to the Desert Fishes Council for partial funding for gene sequencing.

References

Altschul, F., Gish, G., Miller, W., Myers, E. W., Lipman, D. J., 1990. Basic local alignment search tool. *Journal of Molecular Biology,* 215: 403–410.

April, J., Mayden, R. L., Hanner, R. H., Bernatchez, L., 2011. Genetic calibration of species diversity among North America's freshwater fishes. *Proceeding of the National Academy of Science,* 108(26): 10602–10607.

Ballesteros–Córdova, C. A., Ruiz–Campos, G., Findley, L. T., Grijalva–Chon, J. M. Gutiérrez–Millán, L. E., Varela–Romero, A., 2016. Morphometric and meristic characterization of the endemic Desert chub *Gila eremica* (Cyprinidae, Teleostei), and its related congeners in Sonora, Mexico. *Revista Mexicana de Biodiversidad,* 87: 390–398.

Bergmann, T., Hadrys, H., Breves, G., Schierwater, B., 2009. Character–based DNA barcoding: a superior tool for species classification. *Berliner und Munchener Tierarztliche Wochenschrift,* 122: 446–450.

Bogan, M. T., Noriega–Felix, N., Vidal–Aguilar, S. L., Findley, L. T., Lytle, D. A., Gutiérrez–Raucho, O. G., Alvarado–Castro, J. A., Varela–Romero, A., 2014. Biogeography and conservation of aquatic fauna in spring–fed tropical canyons of the southern Sonoran Desert, Mexico. *Biodiversity and Conservation,* 23: 2705–2748.

Darriba, D., Taboada, G. L., Doallo, R., Posada, D., 2012. JModelTest 2: more models, new heuristics and high–performance computing. *Nature Methods,* 9(8): 772, doi: 10.1038/nmeth.2109

Damm, S., Schierwater, B., Hadrys, H., 2010. An integrative approach to species discovery in odonates: from character–based DNA barcoding to ecology. *Molecular Ecology,* 19: 3881–3893.

DeMarais, B. D., 1991. *Gila eremica*, a new cyprinid fish from northwestern Sonora, México. *Copeia,* 1991(1): 179–189.

DeMarais, B. D., Dowling, T. E., Douglas, M. E., Minckley, W.L., Marsh, P. C., 1992. Origin of *Gila seminuda* (Teleostei: Cyprinidae) through introgressive hybridization: implications for evolution and conservation. *Proceedings of the National Academy of Science USA,* 89: 2747–2751.

DeSalle, R., 2006. Species discovery versus species identification in DNA barcoding efforts: response to Rubinoff. *Conservervation Biology,* 20(5): 1545–1547.

DeSalle, R., Egan, M. G., Siddall, M., 2005. The unholy trinity: taxonomy, species delimitation and DNA barcoding. *Philosophical Transactions of the Royal Society B*, 360(1462): 1905–1916.

Dowling, T. E., Anderson, C. D., Marsh, P. C., Rosenberg, M. S., 2015. Population structure in the Roundtail Chub (*Gila robusta* complex) of the Gila River basin as determined by microsatellites: evolutionary and conservation implications. *PLOS One,* 10(10): e0139832, doi:10.1371/journal.pone.0139832.

Dowling, T. E., DeMarais, B. D., 1993. Evolutionary significance of introgressive hybridization in cyprinid fishes. *Nature,* 362: 444–446.

Felsenstein, J., 1985. Confidence limits on phylogenies: an approach using the bootstrap. *Evolution,* 39: 783–791.

Gerber, A. S., Tibbets, C. A., Dowling, T. E., 2001. The role of introgressive hybridization in the evolution of the *Gila robusta* complex (Teleostei: Cyprinidae). *Evolution*, 55: 2028–2039.

Hebert, P. D. N., Cywinska. A., Ball, S. L., Dewaard, J. R., 2003a. Biological identification through DNA barcodes. *Proceedings of the Royal Society of London B,* 270: 313–322.

– 2003b. Barcoding animal life: cytochrome c oxidase subunit 1 divergences among closely related species. *Proceedings of the Royal Society of London B,* 270: S96–S99.

Hubbs, C. L., 1940. Speciation of fishes. *American Naturalist* ,74(752): 198–211.

Hubert, N., Hanner, R., Holm, E., Mandrak, N. E., Taylor, E., Burridge, M., Watkinson, D., Dumont, P., Curry, A., Bentzen, P., Zhang, J., April, J., Bernatchez, L., 2008. Identifying Canadian freshwater fishes through DNA barcodes. *PLOS One,* 3(6): e2490.

Ivanova, N. V., Zemlak, T. S., Hanner, R. H., Hebert, P. D. N., 2007. Universal primer cocktails for fish DNA barcoding. *Molecular Ecology Notes,* 7: 544–548

Jörger, K. M., Schrödl, M., 2013. How to describe a cryptic species? Practical challenges of molecular taxonomy. *Frontiers in Zoology,* 10: 59, doi: 10.1186/1742–9994–10–59.

– 2014. How to use CAOS software for taxonomy? A quick guide to extract diagnostic nucleotides or amino acids for species descriptions. *Spixiana,* 37(1): 21–26

Kekkonen, M., Mutanen, M., Kaila, L., Nieminen, M., Hebert, P. D., 2015. Delineating species with DNA barcodes: a case of taxon dependent method performance in moths. *PLOS One,* 10(4), e0122481.

Kimura, M., 1980. A simple method for estimating evolutionary rates of base substitutions through comparative studies of nucleotide sequences. *Journal of Molecular Evolution,* 16(2): 111–120.

Kipling, W. W., Rubinoff, D., 2004. Myth of the molecule: DNA barcodes for species cannot replace morphology for identification and classification. *Cladistics,* 20: 47–55.

Kocher,T. D., Thomas., W. K., Meyer, A., Edwards, S. V., Pääbo, S., Villablanca, F. X., Wilson, A. C., 1989. Dynamics of mitochondrial DNA evolution in animals: amplification and sequencing with conserved primers. *Proceedings of the National Academy of Science USA*, 86: 6196–6200.

Lakra, W. S., Singh, M., Goswami, M., Gopalakrishnan, A., Lal, K. K., Mohindra, V., Sarkar, U. K., Punia, O. O., Singh, K. V., Bhatt, J. P., Ayyappan, S., 2015. DNA barcoding Indian freshwater fishes. *Mitochondrial DNA,* 27: 4510–4517.

Lara, A., Rodríguez, R., Casane, D., Côté, G., Bernatchez, L., García–Machado, E., 2010. DNA barcoding of Cuban freshwater fishes: evidence for cryptic species and taxonomic conflicts. *Molecular Ecology Resources,* 10(3): 421–430.

Li, Q. Q., Li, D. Y., Ye, H., Liu, X. F., Shi, W., Cao, N., Duan, Y. Q., 2011. Using COI gene sequence to barcode two morphologically alike species: the cotton bollworm and the oriental tobacco budworm

(Lepidoptera: Noctuidae). *Molecular Biology Reports,* 38(8): 5107–5113.

Librado, P., Rozas, J., 2009. DnaSP v5: a software for comprehensive analysis of DNA polymorphism data. *Bioinformatics,* 25: 1451–1452.

Maddison, W. P., Maddison, R., 2016. Mesquite: a modular system for evolutionary analysis. Version 3.10, http://mesquiteproject.org

Miller, R. R., Minckley, W. L., Norris, S., 2005. Freshwater Fishes of M*éxico*. University of Chicago Press, Chicago, IL.

Miller M. A., Pfeiffer W., Schwartz T., 2010. *Creating the CIPRES Science Gateway for inference of large phylogenetic trees.* Gateway Computing Environments Workshop (GCE), 14 November 2010: 1–8. IEEE, New Orleans, LA, USA, Doi: 10.1109/GCE.2010.5676129

Minckley, W. L., Marsh, P. C., 2009. *Inland Fishes of the Greater Southwest: Chronicle of a Vanishing Biota*. University of Arizona Press, Tucson.

Mora–Álvarez, G., McDowell, F. W., 2000. Miocene volcanism during late subduction and early rifting in the Sierra Santa Ursula of western Sonora, Mexico. In: *Cenozoic tectonics and volcanism of Mexico*: 123–141 (J. M. Stock, H. Delgado–Granados, G. Aguirre–Díaz). *Geological Society of America, Special Paper* 334: 123–141.

Nei, M., 1987. *Molecular Evolutionary Genetics*. Columbia University Press, New York.

Nei, M., Kumar, S., 2000. *Molecular Evolution and Phylogenetics*. Oxford University Press, New York.

Norris, S. M., Fischer, J. M., Minckley, W. L., 2003. *Gila brevicauda* (Teleostei: Cyprinidae), a new species of fish from the Sierra Madre Occidental of México. *Ichthyological Exploration of Fresh–waters,* 14: 19–30.

Page, L. M., Baldwin, C. C., Espinoza–Pérez, H., Findley, L. T., Gilbert, C. R., Hartel, K. E., Lea, R. N., Mandrak, N. E., Schmitter–Soto, J. J., Walker, H. J., 2017. Taxonomy of *Gila* in the lower Colorado River basin of Arizona and New Mexico. *Fisheries* 42(9): 456–460.

Pons, J., Barraclough, T., Gomez–Zurita, J., Cardoso, A., Duran, D. P., Hazell, S., Kamoun, S., Sumlin, W. D., Vogler, A. P., 2006. Sequence–based species delimitation for the DNA taxonomy of undescribed insects. *Systematic Biology,* 55: 595–609.

Posada, D., Buckley, T. R., 2004. Model selection and model averaging in phylogenetics: Advantages of Akaike Information Criterion and Bayesian approaches over likelihood ratio tests. *Systematic Biology,* 53: 793–808.

Puillandre, N., Lambert, A., Brouillet, S., Achaz, G., 2012. ABGD, automatic barcode gap discovery for primary species delimitation. *Molecular Ecology,* 21: 1864–1877.

Rach, J., DeSalle. R., Sarkar. I. N., Schierwater, B., Hadrys, H., 2008. Character–based DNA barcoding allows discrimination of genera, species and populations in Odonata. *Proceedings of the Royal Society of London B,* 275: 237–247.

Rambaut, A., 2014. FigTree v1.4.2: tree figure drawing tool, http://tree.bio.ed.ac.uk

Reid, B. N., Le, M., McCord, W. P., Iverson, J. B., Georges, A., Bergmann, T., Amato, G., DeSalle, R., Naro–Maciel, E., 2011. Comparing and combining distance–based and character–based approaches for barcoding turtles. *Molecular Ecology Resources,* 11: 956–967.

Ronquist, F., Huelsenbeck J. P., 2003. MRBAYES 3: Bayesian phylogenetic inference under mixed models. *Bioinformatics,* 19: 1572–1574.

Rubinoff, D., 2006. Utility of mitochondrial DNA barcodes in species conservation. *Conservation Biology,* 20: 1026–1033.

Rubinoff, D., Cameron, S., Will, K., 2006. A genomic perspective on the shortcomings of mitochondrial DNA for 'barcoding' identification. *Journal of Heredity,* 97: 581–594.

Sarkar, I. N., Planet, P. J., Desalle, R. O. B., 2008. CAOS software for use in character–based DNA barcoding. *Molecular Ecology Resources,* 8(6): 1256–1259.

Schönhuth, S., Perdices, A., Lozano–Villano, L., García–de–León, F. J., Espinoza, H., Mayden, R. L., 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.

Schönhuth, S., Shiozawa, D. K., Dowling, T. E., Mayden, R. L., 2012. Molecular systematics of western North American cyprinids (Cypriniformes: Cyprinidae). Zootaxa, 3586: 281–303.

Stamatakis, A., 2014. RAxML Version 8: A tool for Phylogenetic Analysis and Post Analysis of Large Phylogenies. *Bioinformatics:* 10.1093/bioinformatics/btu033, http://bioinformatics.oxfordjournals.org/content/early/2014/01/21/bioinformatics.btu033.abstract

Swoford, D. L., 2002. PAUP* 4.0: Phylogenetic analysis using parsimony and other methods. Sinauer Associates, Sunderland, MA.

Tamura, K., Peterson, D., Peterson, N., Stecher, G., Nei, M., Kumar, S., 2011. MEGA5: Molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. *Molecular Biology and Evolution,* 28: 2731–2739.

Taylor, H. R., Harris, W. E., 2012. An emergent science on the brink of irrelevance: A review of the past 8 years of DNA barcoding. *Molecular Ecology Resources,* 12(3): 377–388.

Uyeno, T., 1960. Osteology and phylogeny of American cyprinid fishes allied to the genus *Gila*. PhD dissertation, Univ. Michigan, Ann Arbor, MI.

Varela–Romero, A., 2001. Newly documented localities for desert chub, *Gila eremica*, in tropical canyons, Río Mátape basin, Sonora, México. *Proceedings of the Desert Fishes Council,* 32: 33.

Ward, R. D., Zemlak, T. S., Innes, B. H., Last, P. R., Hebert, P. D., 2005. DNA barcoding Australia's fish species. *Philosophical Transactions of the Royal Society B,* 360(1462): 1847–1857.

Wiley, E. O., 1981. Phylogenetics. *The Theory and Practice of Phylogenetic Systematics*. Wiley–Interscience, New York.

Witt, J. D., Threloff, D. L., Hebert, P. D., 2006. DNA barcoding reveals extraordinary cryptic diversity in an amphipod genus: implications for desert spring conservation. *Molecular Ecology,* 15(10): 3073–3082.

Yassin, A., Markow, T. A., Narechania, A., O'Grady, P. M., DeSalle, R., 2010. The genus *Drosophila* as a model for testing tree– and character–based methods of species identification using DNA barcoding. *Molecular Phylogenetics and Evolution,* 57: 509–517.

Yi, S., Zhong, J., Huang, S., Wang, S., Wang, W., 2017. Morphological comparison and DNA barcoding of four closely related species in the genera *Misgurnus* and *Paramisgurnus* (Cypriniformes: Cobitidae). *Biochemistry and Systematic Ecology,* 70: 50–59.

Yu, Z., Li, Q., Kong, L., Yu, H., 2014. Utility of DNA barcoding for Tellinoidea: a comparison of distance, coalescent and character–based methods on multiple genes. *Marine Biotechnology,* 17(1): 55–65.

Zhang, A. B., He, L. J., Crozier, R. H., Muster, C., Zhu, C. D., 2010. Estimating sample sizes for DNA barcoding. *Molecular Phylogenetics and Evolution,* 54(3): 1035–1039.

Zou, S., Li, Q., 2016. Pay attention to the overlooked cryptic diversity in existing barcoding data: the case of Mollusca with character–based DNA barcoding. *Marine Biotechnology,* 18: 1–9.

Zou, S., Li, Q., Kong, L., Yu, H., Zheng, X., 2011. Comparing the usefulness of distance, monophyly and character–based DNA barcoding methods in species identification: a case study of Neogastropoda. *PLOS One,* 6: e26619.