



Scale(s) matter: Deconstructing an area of endemism for Middle American freshwater fishes

Diego J. Elías¹ | Caleb D. McMahan² | Wilfredo A. Matamoros³ |
Adan E. Gómez-González^{3*} | Kyle R. Piller⁴ | Prosanta Chakrabarty¹

¹LSU Museum of Natural Science (Ichthyology section), Department of Biological Sciences, Louisiana State University, Baton Rouge, LA, USA

²Field Museum of Natural History, Chicago, IL, USA

³Museo de Zoología, Instituto de Ciencias Biológicas, Universidad de Ciencias y Artes de Chiapas, Tuxtla Gutiérrez, Chiapas, México

⁴Department of Biological Sciences, Southeastern Louisiana University, Hammond, LA, USA

Correspondence

Diego J. Elías, LSU Museum of Natural Science (Ichthyology section), Department of Biological Sciences, Louisiana State University, Baton Rouge, LA, USA.
Email: delias@lsu.edu

Funding information

Negaunee Foundation; Society of Systematic Biologists; National Science Foundation, Grant/Award Number: DEB 1354149 and DEB 1354930

Handling Editor: Jani Heino

Abstract

Aim: The processes that generate biodiversity occur at finer scales than are often studied, particularly in freshwater systems of the northern Neotropics. We investigate whether fine-scale biogeographic patterns are present within the larger Grijalva–Usumacinta Area of Endemism (AoE) —a region in Middle America where nearly 60% of freshwater fishes are endemic—and if present-day river basins are single historical units.

Location: Northern Central America and southern Mexico.

Taxon: Freshwater fishes (e.g., Cichlidae, Poeciliidae, Ostariophysi)

Methods: We used fine-scale distributional data for freshwater fishes in northern Middle America (based on museum records and fieldwork) and performed cluster analyses on a presence/absence data matrix of 117 species. We assessed statistical support of geographic clusters using global one-way analysis of similarity and recovered endemic areas that were further supported by species indicator analyses. Endemic areas were diagnosed by the distribution of endemic species within these areas. For a broader scale we also performed phylogeographic analyses for widespread species representing the dominant families of the region.

Results: We found unrecognized biogeographic structure within the Grijalva–Usumacinta AoE. Different lineages possess varying degrees of geographic structuring and endemic species were not homogeneously distributed across the riverscape.

Main conclusions: We redefine the geographic boundaries of two northern Middle American areas of endemism. We identify five endemic areas nested within the larger Grijalva–Usumacinta AoE. We demonstrate that the upper reaches of the Usumacinta and Grijalva river basins are independent historical biogeographic units and can be differentiated from their lower reaches based on their assemblage and molecular diversity. The recognition of these endemic areas can help us define the geographic landscape and develop hypotheses about the processes that generated the ichthyofauna of northern Middle America.

KEYWORDS

biogeography, ecology, endemism, fishes, Northern Neotropics, rivers

* Published posthumously.

Piller and Chakrabarty are joint senior authors.



1 | INTRODUCTION

Biogeographical studies rely on the delimitation of functional geographical units that are hypotheses of natural areas (Morrone, 2009, 2017; Parenti & Ebach, 2009). One way of doing this delimitation is by finding and describing areas of endemism (AoE), which can be key to uncovering contemporary and historical distributions (Parenti & Ebach, 2009). One feature of AoEs is that they are nested entities (see Crother & Murray, 2011) in which larger geographical units have the potential to untangle the evolutionary history at a deeper geological time and higher taxonomic scales (e.g. family/genus level). In contrast, fine-scale geographical units can help to shed light on the evolutionary history of groups that have diversified more recently, potentially at lower taxonomic levels (e.g. species). Therefore, the identification of fine-scale geographical units for biogeographical studies can help to uncover ecological, geological and evolutionary patterns that are overlooked at more coarse scales. Here we test our hypothesis that fine-scale biogeographical patterns exist in one of the most geologically complex areas on Earth, the freshwaters of northern Middle America.

Biogeographical regions for freshwater fishes have been described at continental and regional scales globally (Abell et al. 2008; Dagosta & Pinna, 2017; Leroy et al., 2019; Matamoros, Hoagstrom, Schaefer, & Kreiser, 2016; Matamoros, McMahan, Chakrabarty, Albert, & Schaefer, 2015; Reyjol et al., 2007; Unmack, 2001). Whereas fine-scale patterns (i.e. identification of river networks with shared endemics) within these larger regions are less often investigated (e.g. Lemopoulos & Covain, 2019; Smith & Bermingham, 2005), this is particularly true for freshwater fishes in the northern Neotropics.

Northern Middle America ranges from the Rio Grande in North America to the Polochic–Motagua–Jocotán fault system in Guatemala and Honduras in Central America. The ichthyofauna of the region is primarily hypothesized to have diversified from South American origins between the late Cretaceous to the Miocene (Hrbek, Seckinger, & Meyer, 2007; McMahan, Chakrabarty, Sparks, Smith, & Davis, 2013; Tagliacollo, Duke-Sylvester, Matamoros, Chakrabarty, & Albert, 2017), pre-dating estimated ages (~20 to ~6 Ma) for the closure of the Panamanian Isthmus (Bacon et al., 2015). This region has been subjected to everything from shifting tectonic plates, volcanic activity and sea-level changes to river reversals that have made for a complex geological history, particularly for its freshwaters (Brocard et al., 2012; Iturralde-Vinent & MacPhee, 1999; Rosen, 1975, 1979, 1985).

Ichthyofaunal diversity in the region is dominated (~70%) by the families Cichlidae and Poeciliidae (Bussing, 1985; Matamoros et al., 2015; Miller, 1966), in contrast with other regions of the Neotropics, where Ostariophysan fishes (e.g. Characiformes and Siluriformes) dominate the diversity. This pattern has been explained by the *Ostariophysan vacuum* hypothesis (Myers, 1966) in which cichlids and poeciliids colonized and diversified in Northern Middle America (Hrbek et al., 2007; Ríčan, Piálek, Zardoya, Doadrio, & Zrzavy, 2013) in the absence of Ostariophysan fishes (which are hypothesized to have arrived after the initial closure of the Panamanian isthmus; Reeves & Bermingham, 2006). But the discovery of the monotypic family Lacantunidae in the upper reaches of

the Usumacinta River (Rodiles-Hernández, Hendrickson, Lundberg, & Humpries, 2005) and the hypothesis of its phylogenetic relationships based on molecular data (Lundberg, Sullivan, Rodiles-Hernández, & Hendrickson, 2007) do not support the Ostariophysan vacuum hypothesis and instead suggest that a more diverse group of Ostariophysans fishes (i.e. family Lacantunidae) went extinct in Northern Middle America prior the arrival of the currently distributed Ostariophysan fishes in the region.

Middle America is split into several AoEs, with the Grijalva–Usumacinta (G-U AoE; sensu Matamoros et al., 2015) encompassing 25 river basins across southern Mexico, Guatemala and Belize (Figure 1), and possessing the highest proportion (59.2%) of endemic fishes in Middle America (Matamoros et al., 2015). Climatic events during the Plio-Pleistocene impacted the availability of surface water in the region, resulting in drying and isolation of water bodies followed by reconnection and refilling (Anselmetti et al., 2006; Brenner, Rosenmeier, Hodell, & Curtis, 2002; Hodell et al., 2008). The complex and dynamic history of ancient and recent (Brocard et al., 2012; Iturralde-Vinent & MacPhee, 1999; Lopez-Ramos, 1975) geological and climatic processes have likely played an important role in shaping the diversity and distributional patterns of freshwater fishes (e.g. McMahan, Ginger, et al., 2017). Despite the complex history of the region and its unique species assemblages, only a few comparative studies have been conducted in Northern Middle America (e.g. for fishes: Rosen, 1979; Perdices, Bermingham, Montilla, & Doadrio, 2002; for reptiles: Castoe et al., 2009; Daza, Castoe, & Parkinson, 2010).

The central goal of this study was to investigate whether fine-scale biogeographical structure is present within the G-U AoE. We investigate and describe fine-scale distributional patterns for freshwater fishes in Northern Middle America. We evaluate two general hypotheses: (1) *Unrecognized biogeographical structure exists within the G-U AoE*; we hypothesize that unique assemblages of freshwater fishes will be identifiable into geographical clusters of river sub-basins within the larger G-U AoE and (2) *River basins in Northern Middle America are not single independent biogeographical units*; empirical work at wide scales have identified contrasting composition and distributional patterns of freshwater fishes in the upper and lower position of river networks (e.g. Carvajal-Quintero et al., 2019; Dias et al., 2014) and the utility of river basins as single historical units has been challenged (Dagosta & Pinna, 2017). Therefore, we hypothesize that river sub-basins in the upper portions of the watershed will possess a more similar assemblage of freshwater fishes within each other than with sub-basins in the lower part of the drainage. We expect non-random patterns of biogeographical structure across taxonomic groups at finer-scales (between and across river basins and other watersheds) in the G-U AoE.

2 | MATERIALS AND METHODS

2.1 | Study area and distributional data

Intraspecific dispersal and the maintenance of gene flow within populations of freshwater fishes are closely tied to hydrobasin

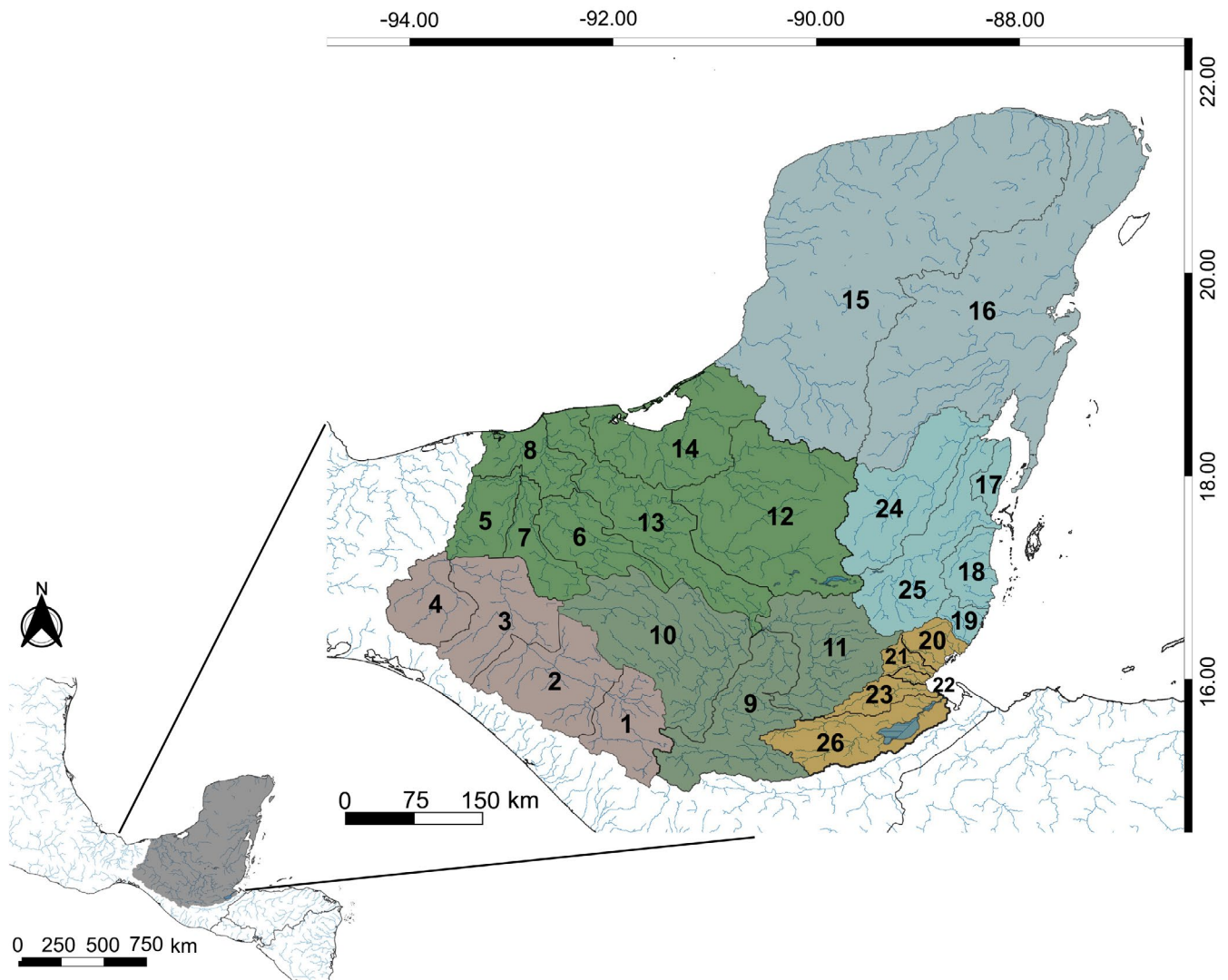


FIGURE 1 Map of the river sub-basins that form the Grijalva–Usumacinta (1–25) and the Polochic–Cahabón (26; light brown) areas of endemism. Upper Grijalva (dark brown), Upper Usumacinta (dark green), Lower Grijalva–Usumacinta (light green), Northern Maya Block (Grey), Eastern Maya Block (teal). Colour scheme maintained through the paper. River sub-basin boundaries follow: INEGI (2010); Suarez (2011); Lehner and Grill (2013); and the river network follow: Lehner and Grill (2013) [Colour figure can be viewed at wileyonlinelibrary.com]

network connectivity, as are changes in connections across the geologic history of these systems (Carvajal-Quintero et al., 2019; Dias et al., 2014). Therefore, we divided the G-U AoE into 25 distinct hydrological units based on river sub-basins and previously published studies (Table 1). The Polochic–Cahabón AoE (sensu Matamoros et al., 2015) was additionally included to test if these adjacent river systems share affinities with the Río Usumacinta proper or other parts of the G-U AoE (Figure 1).

We compiled presence/absence data for 117 species of freshwater fishes (diadromous fishes were not included in our analysis) from 14 families across 25 river sub-basins (Table S1) occurring in the study region. The distributional dataset of Matamoros et al. (2015) served as a starting point; this dataset was updated and revised to agree with the taxonomy of Eschmeyer's Catalog of Fishes (Fricke, Eschmeyer, & van der Laan, 2019). Additional data were compiled through literature searches and recent distributional publications (Barrientos, Elías, & Quintana, 2015;

Barrientos, Quintana, Elías, & Rodiles-Hernández, 2018; Gómez-González, Velásquez-Velásquez, Anzueto-Calvo, & Maza-Cruz, 2015; Greenfield & Thomerson, 1997; Quintana, Barrientos, & Rodiles-Hernández, 2019). Museum records were compiled and examined based on searches through online aggregators (e.g. FishNet2, GBIF) and material recently collected by the authors. Georeferenced coordinates were all validated and substantiated for questionable museum records and we checked the identifications of questionable records.

2.2 | Identification of fine-scale biogeographical structure

The delineation and diagnosis of natural areas (e.g. areas of endemism, endemic areas) as units of analysis is a key first step in systematic, evolutionary and model-based biogeographical

TABLE 1 Hydrological subdivisions within the Grijalva–Usumacinta AoE used for distributional analyses. Numbers in brackets correspond to sub-basins identified in Figure 1

Río Grijalva sensu Gómez-González et al. (2015)	
Upper:	Lower:
Headwaters [1]	Peñitas [5]
La Angostura [2]	Tulija [6]
Chicoasén [3]	Teapa [7]
Malpaso [4]	Tabascal Plains [8]
Río Usumacinta sensu Miller et al. (2005)	
Upper:	Lower:
Río Chixoy [9]	Río San Pedro-Candelaria [12]
Río Lacantún [10]	Main channel [13]
Río La Pasión [11]	Laguna Términos [14]
Yucatán Peninsula	
Western [15]	Eastern [16]
Belize sensu Greenfield and Thomerson (1997)	
North [17]	Monkey River [19]
Central [18]	Grande River [20]
	Moho River [21]
	Temash River [22]
	Sarstun River [23]
Additional rivers and AoE	
Río Hondo [24]	
Río Mopán [25]	
Polochic–Cahabón AoE [26] sensu Matamoros et al. (2015)	

studies (Ebach & Michaux, 2017; Morrone, 2009, 2017; Murray & Crother, 2016; Nelson & Platnick, 1981; Parenti & Ebach, 2009; Ree, Moore, Webb, & Donoghue, 2005). These natural areas possess fuzzy geographical boundaries and are diagnosable by co-distributed species found nowhere else (Crother & Murray, 2011; Parenti & Ebach, 2009). Herein we use the term *endemic areas* to refer to those regions (i.e. clusters of river sub-basins) containing species endemic to the G-U AoE but that are uniquely distributed within individual portions of the AoE (Domínguez, Roig-Juñent, Tassin, Ocampo, & Flores, 2006) as our operational definition of endemic areas. This is also in line with the definition of the term proposed by Parenti and Ebach (2009): ‘any disjunct or continuous geographical space, through time, that delimits the current and past distribution of one or more taxa’.

To delineate endemic areas within the study region, we used multivariate clustering methods (Kreft & Jetz, 2010) to identify geographical clusters of river sub-basins based on assemblages (sensu Fauth et al., 1996) of freshwater fishes. These methods provide a quantitative and objective first step in identification and delineation of biogeographical regions and have been widely used for aquatic ecosystems (Cousseau et al., 2019; Ennen, Agha, Matamoros, Hazzard, & Lovich, 2016; Ennen et al., 2020; Matamoros et al., 2016; Matamoros, Kreiser, & Schaefer, 2012). First, we performed a hierarchical cluster analysis of the river

sub-basins based on an unweighted pair group method analysis with arithmetic means (UPGMA) on a dissimilarity matrix based on Jaccard index from our presence/absence dataset using the *hclust* function implemented in the R package ‘vegan’. We calculated cophenetic indices (CI) to evaluate the degree of correlation between dissimilarity and raw data matrices to assess if the recovered hierarchical clusters reflect the information contained in the raw data. Cophenetic indices near one indicated a high degree of correlation between matrices (Saraçlı, Dogan, & Dogan, 2013; Sokal & Rohlf, 1962).

We additionally performed a non-hierarchical clustering analysis, partitioning by *K*-means (Legendre & Legendre, 2012; MacQueen, 1967), to objectively assess congruence with the hierarchical approach. This approach a priori requires the number of potential clusters (*K*) in our dataset with a minimum of $K = 2$ (the two AoEs under analysis) and maximum of $K = 12$ (highest number of clusters composed of two sub-basins). To quantitatively identify the optimal number of clusters within the G-U AoE, we used the Calinski–Harabasz (C-H) index (Caliński & Harabasz, 1974; Legendre & Legendre, 2012; Milligan & Cooper, 1985) implemented in the R package ‘vegan’ (version 3.1.0, Oksanen et al., 2017; R Core Team, 2015). This stopping rule method calculates a multivariate *F*-statistic that increases as the numbers of evaluated clusters are more differentiated (Legendre & Legendre, 2012; Oksanen et al., 2017). We evaluated the 11 different potential numbers of *K* in our dataset with the highest C-H index indicating the optimal *K* (Oksanen et al., 2017). We additionally tested whether other potential *K* increased the C-H index as the number of clusters increased using the *cascadeKM* function in ‘vegan’. This evaluation allowed us to identify other optimal *K* values in our dataset (Oksanen et al., 2017); if more than one optimal *K* was recovered, we chose the value that maximized number of clusters while preserving geographical coherence (combination of adjacent river sub-basins) as suggested by Matamoros et al. (2016). The *cascadeKM* function returns a summary table from assignment analysis of river sub-basins to clusters for each potential *K* evaluated (Oksanen et al., 2017). We compared the non-hierarchical sub-basin assignment of the *K*-means analysis with the dendrogram topology from the hierarchical cluster analysis to evaluate congruence among methods and delineate endemic areas.

We used the optimal number of *K* to conduct one-way analyses of similarity (ANOSIM; Clarke & Warwick, 1994) to test the statistical significance of the recovered river sub-basin clusters. ANOSIM calculates an *R* statistic that ranges from 0 to 1. If $R = 0$, no differences exist in species composition within or between clusters. If $R = 1$, all pairs of samples within groups are more similar than comparisons with any samples from different groups. We performed a one-way ANOSIM post-hoc pairwise comparisons among groups in the software PAleontological STatistics (PAST Ver. 3.25; Hammer, 2019). Statistical significance among geographical clusters was assessed using sequential Bonferroni correction (Holm, 1979; Rice, 1989).



Finally, we evaluated distributional patterns specifically for four major fish groups in the region: Families Cichlidae and Poeciliidae, which dominate diversity in the region; superorder Ostariophys (represented by characin tetras and catfishes), which dominate biomass in the region (Myers, 1966); and *non-dominant taxa* (represented by Cyprinodontidae, Fundulidae, Rivulidae, Profundulidae, Synbranchidae, Lepisosteidae and Catostomidae) to test the hypothesis that our recovered general patterns were congruent among these defined groups in the species pool.

To diagnose endemic areas within the larger G-U AoE, we evaluated which endemic species occur in individual geographical clusters. Species exclusively distributed in a single recovered cluster of river sub-basins are diagnostic of that endemic area within the larger G-U AoE. Additionally, we used species indicator analyses as an independent assessment of fidelity of endemism. Values ranged from 0 to 1, with 1 indicating high fidelity (found only in one cluster or endemic area) and high frequency (present in all river sub-basins forming the endemic area; Dufrêne & Legendre, 1997).

2.3 | Phylogeographical data and analyses

We used genetic data as an independent test of biogeographical structure within the G-U AoE. We evaluated phylogeographical patterns for six widespread co-distributed freshwater fishes in Northern Middle America. Three species of the family Cichlidae (*Petenia splendida*, *Trichromis salvini* and *Vieja melanura*) and three species of the family Poeciliidae (*Belonesox belizanus*, *Gambusia sexradiata* and *Xiphophorus hellerii*). Tissue samples across distributions (Figure S1) were provided by the Southeastern Louisiana University Vertebrate Museum (SLU), the LSU Museum of Natural Science (LSUMZ), Museo de Zoología of Universidad de Ciencias y Artes de Chiapas, México (MZ-UNICACH), Museo de Historia Natural de Guatemala (USAC) and the Field Museum of Natural History (FMNH). Museum codes follow Sabaj (2016). Additionally, published sequence data with georeferenced coordinates, all of which were confirmed for accuracy, were included in our datasets (Table S2).

Whole genomic DNA was extracted from tissue samples using a DNeasy kit (Qiagen, Inc.). The mitochondrial cytochrome *b* (cyt *b*) gene was amplified and sequenced for 38 samples of *P. splendida*, 35 of *V. melanura*, 40 of *T. salvini*, 24 of *B. belizanus*, 55 of *X. hellerii* and 56 samples of *G. sexradiata*. Mitochondrial DNA markers (e.g. cyt *b*) provide sufficient data (Bowen et al., 2014) for the objectives of this study and allowed for the inclusion of existing available genetic data from previous studies increasing our geographical coverage. Molecular protocols and primer set for amplification and sequencing of cyt *b* are described in Supplementary Information 1.

Sequence data were aligned and visually inspected using the software GENEIOUS 9.1 (Kearse et al., 2012), and multiple sequence alignments were constructed for each species with the MAFFT algorithm (Katoh & Standley, 2013) using default parameters

in GENEIOUS. Sequences were deposited in GenBank (Accession numbers listed in Table S2). Outgroups and additional sequences were downloaded from GenBank for each species under study. Nucleotide (π ; Nei & Li, 1979), genetic (θ_w ; Watterson, 1975) and haplotype (Hd) diversity indices were calculated in the software DnaSP v.5.10 (Librado & Rozas, 2009). Matrilineal haplotypes were reconstructed using statistical parsimony analysis (SPA; Templeton, Crandall, & Sing, 1992) implemented in the software TCS (Clement, Posada, & Crandall, 2000) with a 95% confidence limit for connected haplotypes. The SPA calculates the probability ($p = >0.95$) that two haplotypes are connected by a single mutational step, the most parsimonious solution (Templeton et al., 1992). The implementation of this algorithm in the software TCS calculates the maximum number of mutations allowed to connect a pair of haplotypes based on the parsimony criterion (Clement et al., 2000). If the number of mutations connecting two haplotypes or two inferred networks is higher than the threshold calculated by the parsimony criterion, then these networks are reconstructed as unconnected. Recovered networks with the TCS software were used to infer gene genealogies using a median-joining network (Bandelt, Foster, & Röhl, 1999) implemented in the software POPART (Leigh & Bryant, 2015). If the SPA implemented in TCS recovered unconnected networks, we inferred the haplotype networks in POPART independently.

The best partition scheme and model selection were selected (Table S3) using a greedy search and the corrected Akaike information criterion-AICc (Hurvich & Tsai, 1989) in the software PARTITIONFINDER (Lanfear, Calcott, Ho, & Guindon, 2012). Phylogenetic hypotheses were inferred in MrBayes 3.2.6 (Huelsenbeck, Ronquist, Nielsen, & Bollback, 2001) implemented in the CIPRES portal (Miller, Pfeiffer, & Schwartz, 2010). Three independent analyses of 12,000,000 generations for *P. splendida*, *V. melanura* and *G. sexradiata*, 14,000,000 generations for *B. belizanus*, and 20,000,000 generations for *T. salvini* and *X. hellerii* were performed with two replicate runs each. Sampling stationarity was assessed via visual inspection in TRACER 1.5 (Rambaut, Drummond, Xie, Baele, & Suchard, 2018) and observed average standard deviations of split frequencies (<0.01) values. Two thousand trees were sampled and the initial 25% were discarded as burn-in. Bayesian posterior probabilities (BPP) were calculated using the post burn-in trees, and the three independent runs were compared to evaluate topological congruence among runs. Subsequently, phylogenetic hypotheses were inferred as above using a reduced dataset that only included one tip per haplotype for each species. Phylogenetic trees and frequencies of haplotypes in geographical areas were visualized using the *dotTree* function in the R package 'phytools' (Revell, 2012).

We used two complementary approaches to objectively determine the number of independent mitochondrial lineages for each species within the G-U AoE. We inferred the number of geographically delimited haplogroups present in our molecular dataset using results of the SPA and conducted a statistical test (Rosenberg's Test; Brown et al., 2012; Rosenberg, 2007) on resolved nodes of the inferred phylogenies. Matrilineal haplotype



networks were reconstructed using the SPA implemented in the software TCS (Templeton et al., 1992) with a 95% confidence limit for connecting haplotypes. This analysis has shown to be robust in identifying mitochondrial lineages with Sanger sequence data (Hart & Sunday, 2007). Therefore, we used this method to identify if there were geographically defined haplogroups (i.e. phylogeographical structure) for each species under study that supported the identified endemic areas. Recovered networks with TCS were used to infer gene genealogies using a median-joining network (Bandelt et al., 1999) implemented in the software POPART (Leigh & Bryant, 2015).

We tested the null hypothesis that all samples belonged to a single panmictic population to determine if phylogeographical structure was stochastic or due to inadequate sample sizes. We performed Rosenberg's test for reciprocal monophyly (Brown et al., 2012) for each bifurcating node, testing the null hypothesis that a single taxonomic entity followed a random branching model (Rosenberg, 2007). If the null hypothesis was supported by the observed bifurcating pattern in the tree, it was interpreted as random branching model expected in populations evolving under the Yule model (Rosenberg, 2007) and we considered the samples to belong to a single mitochondrial lineage. If the null hypothesis was rejected, the recovered clades could be interpreted as geographically delimited haplogroups of the same species (Rosenberg, 2007). Tests were executed in the R package 'spider' (Brown et al., 2012). We took a conservative approach where geographical haplogroups were delimited if they were recovered as an unconnected network in the SPA and supported by Rosenberg's test.

3 | RESULTS

3.1 | Distributional patterns between the G-U and Polochic-Cahabón areas of endemism

Cluster analyses of river sub-basins that comprise the G-U AoE and the Polochic-Cahabón AoE (sensu Matamoros et al., 2015) recovered geographical structuring of freshwater fish assemblages within the region based on the hierarchical analysis (UPGMA) based on dissimilarity (Jaccard Index D_j) and the non-hierarchical analysis (K -means). The cophenetic correlation index was high ($CC = 0.88$), and the Calinski-Harabasz index (C-H) identified two optimal numbers of clusters within the geographical study area including both the G-U AoE and Polochic-Cahabón AoE: $K = 4$ clusters (Table S4; ANOSIM output: $R = 0.9109$, $p = 0.001$) and $K = 6$ clusters (Table S4; ANOSIM output: $R = 0.9139$, $p = 0.001$). We followed the latter scheme because it maximized number of clusters (i.e. $K = 5$ clusters in G-U AoE and $K = 1$ cluster for Polochic-Cahabón AoE) while preserving geographical coherence (see Materials and Methods). In both geographical schemes ($K = 4$ and $K = 6$), the Grande, Moho, Temash and Sarstún river basins located in the southeastern portion of the G-U AoE (Eastern

Guatemala and Southern Belize) shared more affinities with the Polochic-Cahabón AoE than any other river basin(s) within the G-U AoE based on assemblages of freshwater fishes (Figure 2, Table S4). Therefore, our results are presented with the inclusion of these rivers with the Polochic-Cahabón AoE (sensu Matamoros et al., 2015) so that the G-U AoE was treated as an analytically supported entity.

The G-U AoE contained 108 of the overall 117 species of fishes, 59 (~55%) of which were endemic. The Polochic-Cahabón AoE contained 34 species, seven (~21%) of which were endemic, and showed high levels of dissimilarity from the G-U AoE (UPGMA, $D_j = -0.70$; Figure 2). This was highly supported by the distributional patterns of cichlids and poeciliids (Figure 3) with some degree of variation across individual taxa. Seven species were identified as endemic to the Polochic-Cahabón AoE, with three of those identified by species indicator analysis supporting uniqueness of this cluster (Table 2). Furthermore, all recovered geographical clusters ($K = 6$) were statistically supported (Table 3; ANOSIM: $p < 0.05$) with the exception of the Northern Maya Block when compared with non-adjacent geographical clusters (Upper Grijalva $p = 0.0639$, Upper Usumacinta $p = 0.1015$).

3.2 | Distributional patterns within the G-U AoE

3.2.1 | Upper Grijalva

The four river basins from the upper reaches of the Río Grijalva were recovered in a single cluster by the UPGMA and K -means cluster analyses with a diversity of 33 species (Table 2, Figure 2, Table S4). This area showed the highest level of dissimilarity based on the UPGMA analysis ($D_j \sim 0.81$; Figure 2) across the G-U AoE and was driven by the assemblage of the families Poeciliidae and Cichlidae (Figure 3). In all, 11 fish species endemic to the G-U AoE were found in the Upper Grijalva, and five of these were unique to this cluster (Table 2). Species indicator analysis identified four endemic species that supported the Upper Grijalva endemic area (Table S5).

3.2.2 | Upper Usumacinta

Clustering analyses (UPGMA and K -means) recovered the three river basins from the upper reaches of the Río Usumacinta clustered together (Figure 2; Table S4) and showed intermediate levels of dissimilarity when compared with the cluster of river sub-basins of the lower reaches of the Grijalva and Usumacinta rivers (UPGMA, $D_j \sim 0.44$; Figure 2). This cluster was highly supported by poeciliids, moderately by cichlids distributions (Figure 3) and three endemic ostariophysan fishes (Table S5). In all, 58 species of fishes were found to occur in the Upper Usumacinta, with 29 species endemic to the G-U AoE occurring in the Upper Usumacinta and 15 of those were found only in the Upper Usumacinta (Table 2). Species indicator analysis

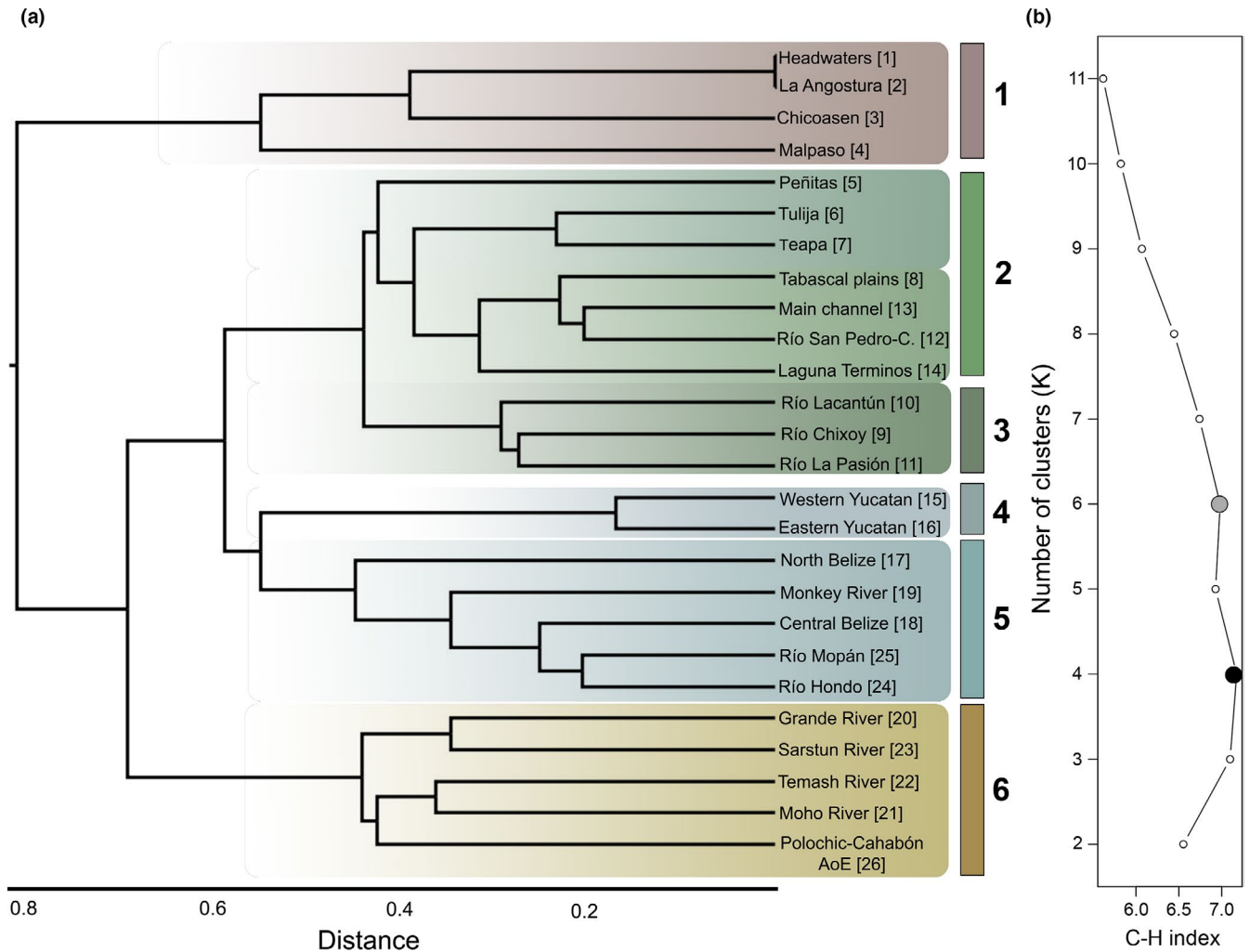


FIGURE 2 Geographical structure recovered based on fine-scale distributional data for assemblages of freshwater fishes in the Grijalva-Usumacinta area of endemism; (a) Dendrogram based on the Jaccard dissimilarity index. 1: Upper Grijalva, 2: Lowlands Grijalva-Usumacinta, 3: Upper Usumacinta, 4: Northern Maya Block, 5: Eastern Maya Block and 6: Polochic-Cahabón AoE. (b) Calinski-Harabasz index to determine the number of clusters (K) present in the data. Black filled circle is optimal number of K and grey filled circle is another informative number of K identified in the dataset (see Materials and Methods) [Colour figure can be viewed at wileyonlinelibrary.com]

identified eight species that supported the upper Usumacinta as a unique cluster (Table S5).

3.2.3 | Lower Grijalva and Usumacinta

Clustering analyses (UPGMA and K -means) recovered a cluster of seven river sub-basins that collectively form the lower portions of the Río Grijalva and Río Usumacinta (Figure 2; Table S4). This cluster was moderately supported by assemblages of the families Cichlidae and Poeciliidae (Figure 3). This area contained the highest diversity (66 spp.) within the Grijalva-Usumacinta AoE. In total, 26 species endemic to the G-U AoE were found in this region, and 8 of those were restricted to river basins in the lower Grijalva and Usumacinta (Table 2; Table S5). Species indicator analysis identified only *Heterophallus echeagarayi* (Poeciliidae) supporting the Lower Grijalva and Usumacinta as an endemic area but this species is not diagnostic of the larger G-U AoE.

3.2.4 | Northern Maya Block

Clustering analyses (UPGMA and K -means) recovered water bodies in the northwestern and northeastern Yucatán Peninsula (Figure 2; Table S4) in a cluster based mainly on the assemblage of cyprinodontid fishes (non-dominant taxa, Figure 3). This cluster showed intermediate levels of dissimilarity (UPGMA, $D_j = 0.55$; Figure 2) when compared with the cluster of river sub-basins from the Eastern Maya Block. In all, 48 species were found to occur in the area; difference in the species composition of this cluster was supported by the ANOSIM (R statistics; Table 3) when compared with the other recovered clusters, however, this geographical cluster was not statistically different from two non-adjacent clusters the Upper Grijalva and Upper Usumacinta (Table 3). In all, 11 of 23 G-U diagnostic endemic species present in the Northern Maya Block are uniquely distributed in this region (Table 2). Species indicator analysis identified nine species supporting the Northern Maya Block as an endemic area (Table S5).

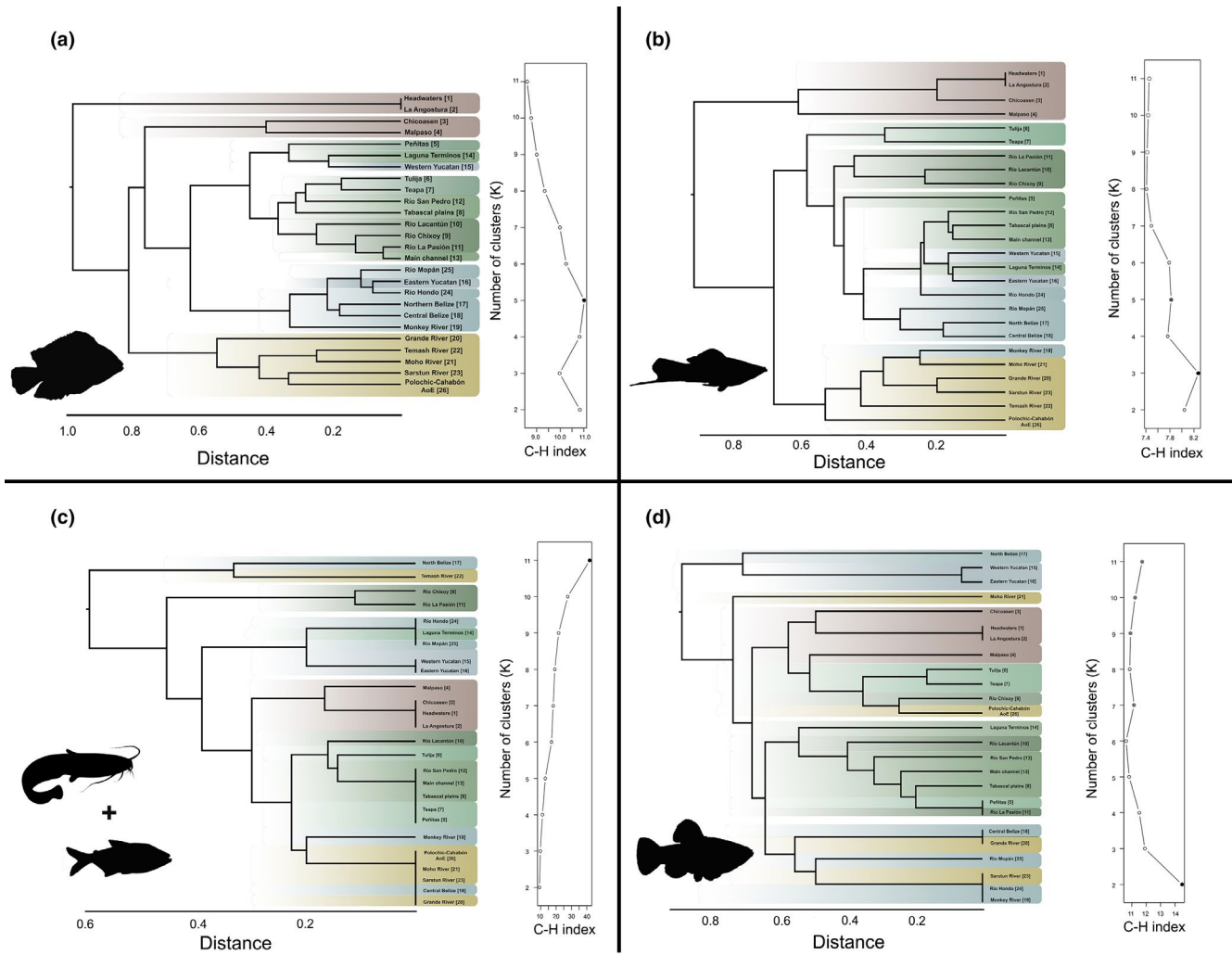


FIGURE 3 Geographical structure recovered for each investigated taxonomic group based on the Jaccard dissimilarity index. CI = cophenetic index. (a) Family Cichlidae, CI = 0.92; (b) Family Poeciliidae, CI = 0.91; (c) Superorder Ostariophysi, CI = 0.85; (d) non-dominant taxa, CI = 0.82. Calinski-Harabasz indices on right side of each panel [Colour figure can be viewed at wileyonlinelibrary.com]

Endemic area	No. species	No. endemics	% endemics	No. diagnostic endemics	% diagnostic endemics
G-U AoE					
Upper Grijalva	33	11	33.3%	5	15.15%
Upper Usumacinta	58	29	50%	15	25.86%
Lowlands Grijalva-Usumacinta	66	26	39.39%	8	12.12%
Northern Maya Block	48	23	47.92%	11	22.92%
Eastern Maya Block	37	11	29.73%	1	2.7%
Polochic-Cahabón AoE	34	7	20.58%	7	20.58%

TABLE 2 Summary of diversity and endemism. *No. endemics* is number of species diagnosing larger G-U AoE that occur in that area. *No. diagnostic endemics* is number of endemics in G-U AoE that uniquely occur in that area

3.2.5 | Eastern Maya Block

Clustering analyses (UPGMA and K-means) recovered five river sub-basins from northeastern Guatemala, southern México and Belize clustered together (Figure 2; Table S4) and was highly supported by

distributional patterns of cichlids and intermediately by poeciliids (Figure 3). In all, 37 species are distributed in this endemic area, 11 of which were G-U AoE endemics (Table 2). *Poecilia teresa* (Poeciliidae) is the only endemic species that is uniquely distributed in the area (Mopan River basin; Figure 1). Species indicator analysis did not



TABLE 3 Analysis of similarity pairwise comparison. R statistic values under the diagonal (Range 0–1 if R = 0, no differences exist in species composition within or between clusters. If R = 1, all pairs of samples within groups are more similar than comparisons with any samples from different groups). *p* values (sequential Bonferroni correction) above the diagonal. Bold values indicate comparisons that are statistical significant ($p < 0.05$)

	Upper Grijalva	Upper Usumacinta	Lowlands Grijalva–Usumacinta	Northern Maya Block	Eastern Maya Block	Polochic–Cahabón AoE
Upper Grijalva	-	0.030	0.003	0.064	0.006	0.008
Upper Usumacinta	0.963	-	0.008	0.102	0.019	0.017
Lowlands Grijalva–Usumacinta	0.995	0.591	-	0.024	0.001	0.001
Northern Maya Block	1.000	1.000	0.955	-	0.049	0.049
Eastern Maya Block	1.000	1.000	0.900	0.891	-	0.006
Polochic–Cahabón AoE	1.000	1.000	1.000	1.000	0.878	-

identify any species supporting the Eastern Maya Block as an endemic area (Table S5).

3.3 | Taxa-specific distributional patterns

3.3.1 | Cichlids (38 species)

Fine-scale biogeographical structure was recovered in the G-U AoE (Figure 3), with the hydrological systems within the upper Grijalva recovered as the most distinct geographical clusters across the AoE. A grouping consisting of all aquatic systems of the Río Usumacinta, the lower Grijalva and freshwater systems in the western Yucatán Peninsula. This cluster was most similar to the assemblage of cichlids occurring in the eastern Yucatán Peninsula and Río Hondo south to the Monkey River in Belize. Species composition of cichlids in rivers of southern Belize (Grande, Moho, Temash and Sarstún) is more similar to the assemblage of cichlids distributed in the Polochic–Cahabón AoE (Figure 3) than with any other river system within the G-U AoE.

3.3.2 | Poeciliids (43 species)

Similar distributional patterns were recovered for poeciliids as for cichlids, with the upper Grijalva as the most distinct geographical cluster across the AoE (Figure 3). Hydrological units in the upper Usumacinta (Chixoy, Lacantún and La Pasión rivers) were recovered as a single cluster. The lower Usumacinta shared most species of poeciliids with the Yucatán Peninsula and Río Hondo in contrast with cichlids and analysis of all taxa (Figures 2 and 3). However, the assemblage of poeciliids in the Río Mopán and north and central Belize was clustered together. The Monkey River was more similar to aquatic systems in southern Belize than to central and northern Belize. For poeciliids, rivers in southern Belize also clustered as most similar to the Polochic–Cahabón AoE (Figure 3).

Cluster analyses based on the distribution of ostariophysan fishes ($n = 12$ species) did not recover geographical structuring (Figure 3),

and analysis based on the distribution of the non-dominant taxa group recovered limited geographical structure within the region. The non-dominant taxa ($n = 24$ species) was the only group that supported the Northern Maya Block as a distinct cluster (Figure 3).

3.4 | Phylogeographical structure in cichlids and poeciliids

Due to uncertainty at the beginning and end of sequences, and to avoid missing information in any samples, the final alignment for each species was truncated. The final alignment length consisted of 1,050 bp for *Petenia splendida* ($n = 48$); 1,023 bp for *Vieja melanura* ($n = 41$); 975 bp for *Trichomis salvini* ($n = 53$); 1,101 bp for *Belonesox belizanus* ($n = 59$); 1,053 bp for *Xiphophorus hellerii* ($n = 55$) and 1,092 bp for *Gambusia sexradiata* ($n = 56$).

Our analysis recovered strong phylogeographical structure across the distribution of *B. belizanus* (Figure 4), and we recovered five divergent clades supported by Rosenberg's Test of reciprocal monophyly. Furthermore, SPA recovered five unconnected genealogical networks that were largely congruent with the phylogenetic approach. The difference between phylogenetic and haplotype inferences was not related to the geographical area of focus. Three of the clades fell within the G-U AoE (BPP > 0.95, uncorrected *p*-distance ranges from 0.038 to 0.041, *SE* ranges from 0.05 to 0.06). The first clade (A) contained individuals from the Upper Usumacinta, Lower Grijalva and Usumacinta, and Eastern Maya Block with 10 haplotypes and possessed low levels of genetic (π , θ) but high haplotype (Hd) diversity (Table S6). The second clade (B) contained individuals from West of the G-U AoE, Upper Usumacinta, Northern Maya Block and Eastern Maya Block. Intermediate levels of genetic (π , θ) and high haplotype (Hd) diversity were observed (Table S6) with 20 haplotypes and three unsampled haplotypes estimated. The third clade (C) contained individuals from Eastern Maya Block and the Polochic–Cahabón AoE, with only three haplotypes and one unsampled haplotype estimated with intermediate levels of genetic (π , θ) and high haplotype (Hd) diversity (Table S6).

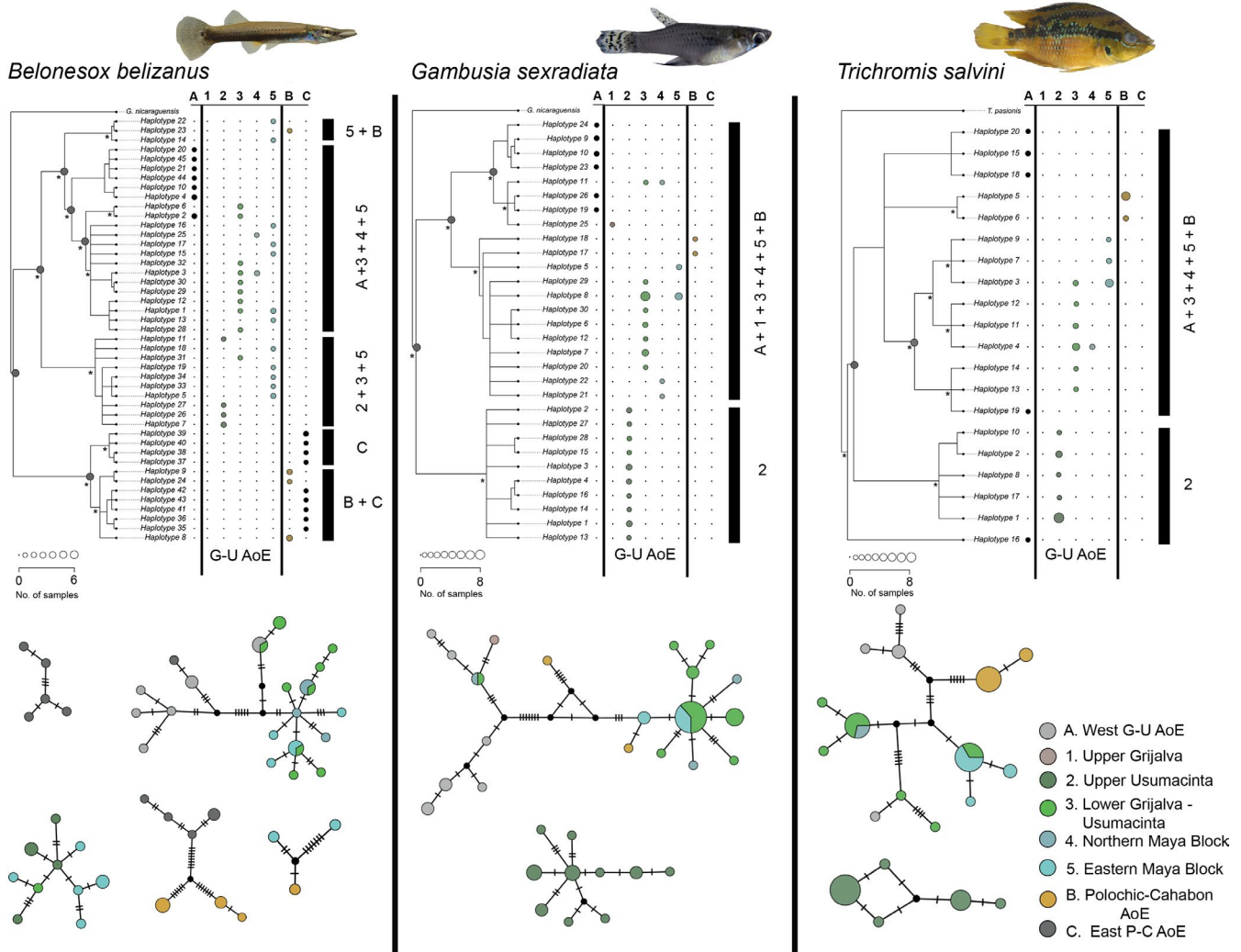


FIGURE 4 Phylogeographic-level analysis of widespread species with phylogeographical structure in the Grijalva–Usumacinta area of endemism. Asterisks on nodes indicate Bayesian posterior probabilities >0.95 . Grey filled circles on nodes indicate support of reciprocal monophyly of sister clades based on Rosenberg's test. Haplotype network based on median joined network, dashes represent mutational steps, black filled circles represent unsampled haplotypes and size of circles corresponds to number of individuals sharing that haplotype [Colour figure can be viewed at wileyonlinelibrary.com]

We recovered phylogeographical structure across the distribution of *G. sexradiata* (Figure 4). Two clades were recovered (BPP > 0.95 ; uncorrected p-distance = 0.56, SE = 0.007) and Rosenberg's Test supported reciprocal monophyly of the two main clades, as well as two additional internal clades. However, SPA only recovered two unconnected genealogical networks that were congruent with the deepest clades recovered with the phylogenetic approach. One geographically delimited haplogroup was restricted to samples from the Upper Usumacinta and possessed low levels of genetic (π , θ) but high haplotype (Hd) diversity (Figure 4; Table S6). In all, 10 haplotypes were recovered with one unsampled haplotype estimated. The second geographically delimited haplogroup included individuals from the remaining river basins within and west of the G-U AoE (Figure 4). Intermediate levels of genetic (π , θ) and haplotype (Hd) diversity were observed (Table S6), and this haplogroup included 20 haplotypes with five unsampled haplotypes estimated (Figure 4).

Only one species of cichlid (*T. salvini*) demonstrated phylogeographical structure with two recovered clades (Figure 4; BPP = 0.84; uncorrected p-distance = 0.030, SE = 0.005) and Rosenberg's Test supported reciprocal monophyly. One haplotype (haplotype 16) from the western portion of the distribution of *T. salvini* was recovered as the sister group to all other haplotypes. Rosenberg's Test did not support reciprocal monophyly at this node, and with only one individual from this locality the sample was excluded from haplotype reconstruction. Two unconnected genealogical networks (Figure 4) were recovered by the SPA and congruent with our phylogenetic approach. One geographically delimited haplogroup was restricted to the Upper Usumacinta, with low genetic (π , θ) and intermediate haplotype (Hd) diversity (Table S6). This geographically delimited haplogroup included five haplotypes and one unsampled haplotype estimated. The second haplogroup included individuals from lowlands in the G-U AoE and localities west of the G-U AoE, with intermediate genetic (π , θ) and haplotype (Hd) diversity (Table S6). In



all, 14 haplotypes were recovered with three unsampled haplotypes estimated (Figure 4).

3.5 | Absence of phylogeographical structure in cichlids and poeciliids

Xiphophorus hellerii possesses high haplotype (Hd) diversity (Table S6) with 27 haplotypes and shallow phylogeographical structure between two clades (Figure 5). Rosenberg's Test supported reciprocal monophyly of the two clades; however, *X. hellerii* does not show patterns of phylogeographical structure within the G-U AoE and the SPA recovered a single genealogical network (Figure 5).

Intermediate levels of genetic (π , θ) and high levels of haplotype (Hd) diversity were observed for the cichlid *V. melanura* (Table S6). In all, 15 haplotypes were recovered with no phylogeographical structure across the distribution (Figure 5) and a single genealogical network was recovered. A second cichlid, *P. splendida*, demonstrated low levels of genetic (π , θ) and haplotype (Hd) diversity (Table S6), with 14 haplotypes and no phylogeographical structure across its distribution. One haplotype was common (present in 63% of samples) across all river basins under study in the G-U AoE (Figure 5).

4 | DISCUSSION

Our analyses of distributional and molecular data uncover previously unrecognized geographical structure for freshwater fishes within Northern Middle America. We recovered unique distributional patterns across river sub-basins at the assemblage (Figure 2), taxonomic (Figure 3) and phylogeographical (Figure 4) levels. Our quantitative assessment (C-H index) of the number of geographical clusters of river sub-basins suggest two optimal number of clusters $K = 4$ and $K = 6$. Under the geographical scheme of $K = 4$, the endemic areas Upper Usumacinta and Lower Grijalva-Usumacinta belong to the same geographical cluster (Table S4), and the Northern and Eastern Maya blocks belong to the same geographical cluster (Table S4). We recognized $K = 6$ as the number of geographical clusters identified within the study area, following our criteria (see Materials and Methods), the redefined Polochic-Cahabon AoE (see below) and five endemic areas nested within the larger Grijalva-Usumacinta AoE (see below). The biogeographical structure we uncovered ($K = 6$) based on clustering analyses of freshwater fishes distributional data is biological meaningful as we identified endemic species diagnosing each geographical region (Table 2, Figure 6).

Our fine-scale distributional and phylogeographical patterns support our hypothesis that there is unrecognized biogeographical structure within the G-U AoE. Furthermore, the recovered distributional and phylogeographical patterns supports the second hypothesis that river basins in the region are not single independent biogeographical units. The lower portions of the two largest river basins (i.e. Grijalva and Usumacinta rivers) within the G-U AoE share

more similarities within each other at the assemblage (Figure 2) and molecular (Figure 4) levels than among any other areas. In contrast, upper portions of both river basins do possess unique assemblages and molecular diversity that are not shared with the lower portion of the same river network. Finally, congruent patterns were recovered for different taxonomic groups (i.e. cichlids and poeciliids) at the assemblage and phylogeographical level within the G-U AoE.

4.1 | Redefining edges of the G-U AoE

The inclusion of the Polochic-Cahabón AoE (sensu Matamoros et al., 2015) in our analysis allows us to re-define the geographical boundaries of the G-U AoE and the Polochic-Cahabón AoE based on fine-scale distributional data. Matamoros et al. (2015) recognized all rivers in Belize as part of the G-U AoE; however, our results strongly support the inclusion of the southern-most rivers south of the Maya Mountains (i.e. Grande, Moho, Temash and Sarstún) as part of the Polochic-Cahabón AoE instead of the G-U AoE (Figure 2, Table 3). In addition to the species listed by Matamoros et al. (2015), two cichlids, *Chuco godmanni* and *Cincelichthys bocourti*, now additionally diagnose this AoE (Table S5).

The G-U AoE is the most species-rich area with the highest level of endemism for freshwater fishes in all of the Northern Neotropics. Our overall results agree with previous estimates (Matamoros et al., 2015), but we removed endemic species that were found to occur outside the geographical boundaries of this AoE in addition to moving four rivers to the Polochic-Cahabón AoE (Table S4). The G-U AoE ranges eastward from the Grijalva river basin in México and Guatemala to the Yucatán Peninsula in the northeast, then south to the Monkey River and small basins in Central Belize. The Southwestern boundaries encompass the Chixoy and La Pasion river basins that form the upper reaches of the Río Usumacinta (Figure 1).

4.2 | Assemblage-level patterns within the G-U AoE

Areas of endemism are our core functional units in biogeography and these areas can be nested entities in both space and time (Crother & Murray, 2011; Parenti & Ebach, 2009). Thus, their discovery, diagnosis and utility for understanding the evolutionary and biogeographical history of a region or taxa depend on scale (time and space) of the question of interest.

The G-U AoE shares a complex history with other diagnosed areas of endemism in the region (e.g. Polochic-Cahabón; Matamoros et al., 2015) and its utility as a single coarse biogeographical unit is likely greater at deeper evolutionary and larger spatial scales and at higher taxonomic levels. Our results highlight that endemic species diagnosing the G-U AoE are not homogeneously distributed across the riverscape (Figure 6, Supp. Table 3). Furthermore, we provided empirical evidence that not all groups of freshwater fishes possess congruent distributional patterns in the region. We identified five endemic areas nested within the larger G-U AoE: Upper Grijalva,

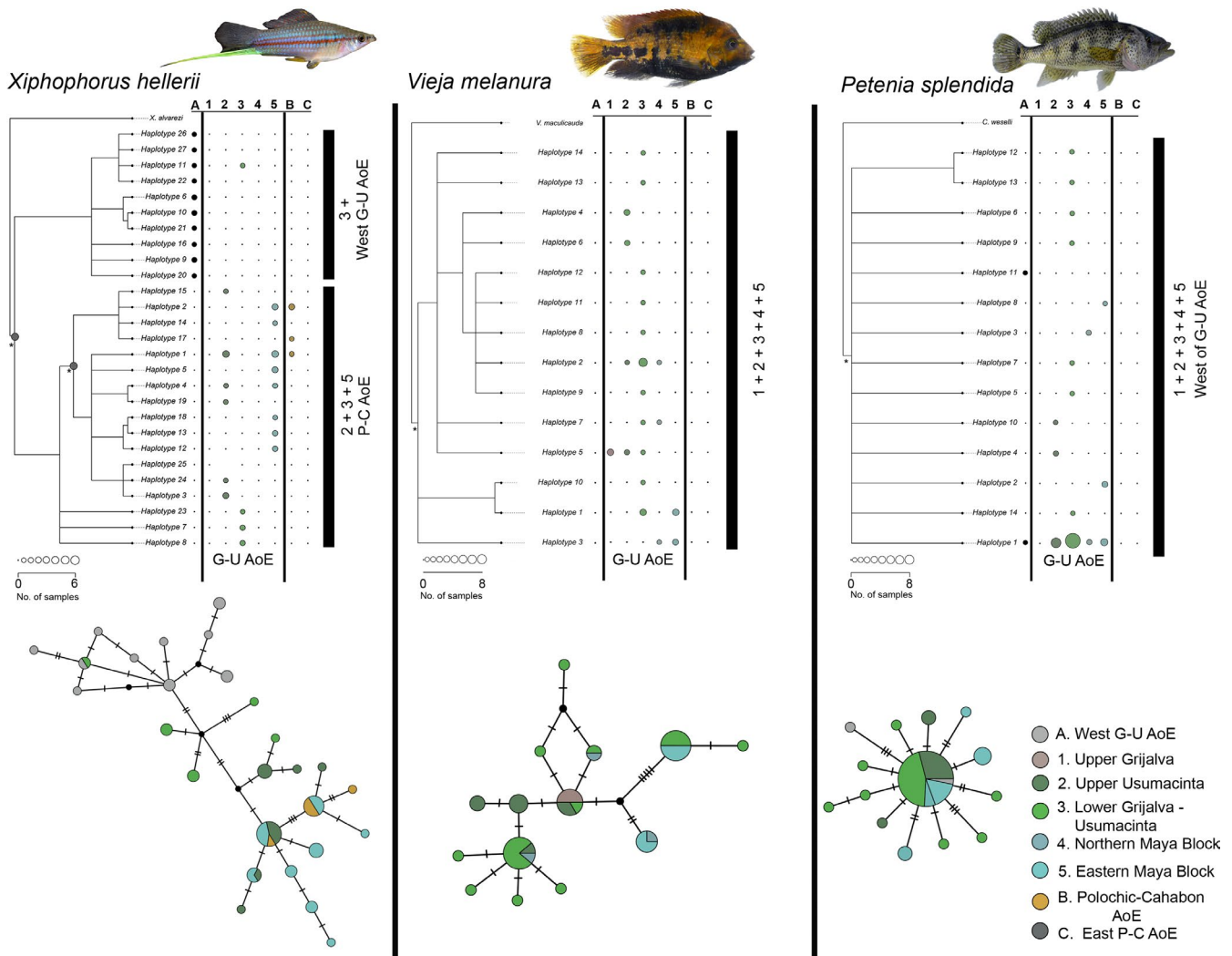


FIGURE 5 Phylogeographic-level analysis of widespread species with no phylogeographical structure in the Grijalva–Usumacinta area of endemism. Asterisks on nodes indicate high Bayesian posterior probabilities >0.95 . Grey filled circles on nodes indicate support of reciprocal monophyly of sister clades based on Rosenberg's test. Haplotype network based on median joined network, dashes represent mutational steps, black filled circles represent unsampled haplotypes and size of circles corresponds to number of individuals sharing that haplotype [Colour figure can be viewed at wileyonlinelibrary.com]

Upper Usumacinta, Lower Grijalva–Usumacinta, Northern Maya Block and Eastern Maya Block (Figure 2). The Upper Grijalva and Upper Usumacinta fall within the Maya highlands physiographical province (sensu Marshall, 2007) while the Lower Grijalva–Usumacinta and Northern Maya Block fall largely within the Yucatán platform physiographical province extending their boundaries into the Maya highlands (sensu Marshall, 2007).

The five endemic areas recovered were primarily supported by the distribution of poeciliids and cichlids but only the non-dominant taxon category provided strong support for the distinctiveness of the Northern Maya Block. Cichlids and poeciliids dominate the species pool of freshwater fishes in the region (Chakrabarty & Albert, 2011; Matamoros et al., 2015; Miller, 1966) and it is hypothesized that these taxa colonized and diversified in the region prior the initial closing of the Panamanian isthmus (Chakrabarty & Albert, 2011; Hrbek et al., 2007; Ričan et al., 2013; Tagliacollo et al., 2017) in contrast with other taxa like modern Ostariophysi, which began dispersing

into the Northern Neotropics only after initial closing of the Isthmus of Panama (Reeves & Bermingham, 2006) up to 20 million years ago (Bacon et al., 2015). Such differences in the timing of colonization of Northern Middle America by the present-day distributed lineages likely influenced contemporary differences in species richness and distributional patterns among taxonomic groups. Notably, ostariophysans showed no pattern of geographical structure within the G-U AoE, while the other groups did (Figure 3). Despite their lack of geographical structure, ostariophysans do exhibit unique distributions among endemic areas (Figure 6) in the Upper Grijalva (*Rhamdia laluchensis*), Upper Usumacinta (*Astyanax baileyi*, *A. dorioni* and *Lacantunia enigmatica*), Lower Grijalva–Usumacinta (*Rhamdia macuspanensis*) and Northern Maya Block (*A. altior*).

Assemblages of cichlids and poeciliids are the most unique in the Upper Grijalva endemic area (Figure 3). Both lineages additionally support distinctiveness of the adjacent Upper Usumacinta, where poeciliids exhibit uniqueness but cichlids share more species with

Proportion of endemic species

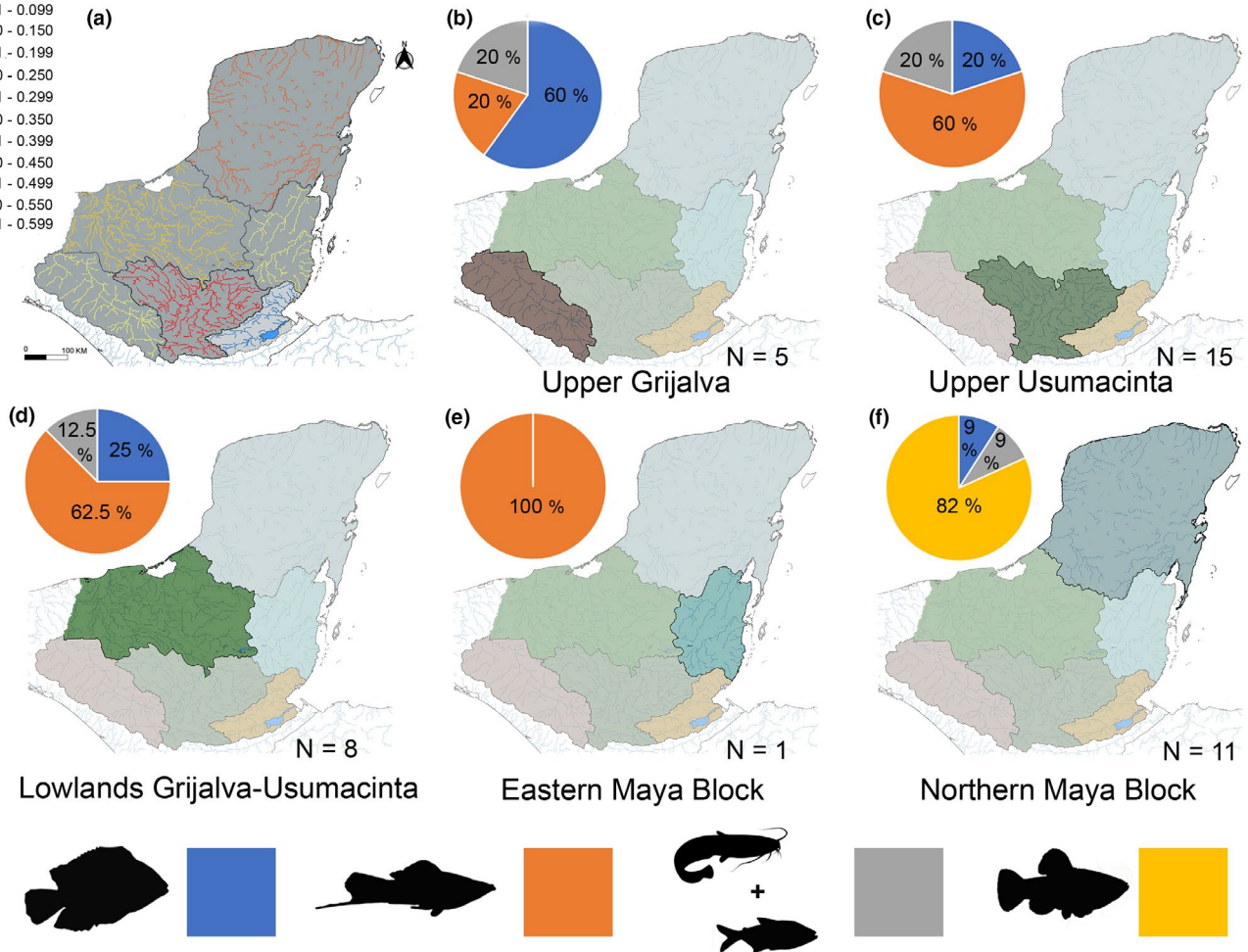


FIGURE 6 Distributional patterns of endemic species that diagnose the Grijalva–Usumacinta area of endemism. (a) Heatmap depicting the proportion of endemics species diagnostic of the Grijalva–Usumacinta area of endemism (darker grey) in each of the five endemic areas identified and the Polochic–Cahabón area of endemism (light grey). Number of diagnostic endemics and its proportion by taxonomic group (see Materials and Methods) for each endemic area: Upper Grijalva (b), Upper Usumacinta (c), Lowlands Grijalva–Usumacinta (d), Eastern Maya Block (e) and Northern Maya Block (f) [Colour figure can be viewed at wileyonlinelibrary.com]

the main channel of the Usumacinta River. While the Upper Grijalva contains the most unique assemblages of freshwater fishes in the G-U AoE, the Upper Usumacinta harbours the highest proportion of endemic species ($n = 29$, Figure 6). Furthermore, the species pool of the Upper Usumacinta contains relict species of freshwater fishes, remnant of old faunal elements in Northern Middle America. The Chiapas Catfish, *Lacantunia enigmatica*, with closest living relatives in Africa (Lundberg et al., 2007) and *Xenodexia ctenolepis*, a species of livebearer (family Poeciliidae) that has been hypothesized to be the sister to all other New World members of the family based on molecular data (Hrbek et al., 2007), are both endemic to the Upper Usumacinta. Several geological events, for example reversal of flow and collapse of stream beds (Rosen, 1979), have led to a complex topography and connectivity (e.g. underground connectivity among river sub-basins; Rosen, 1967; Rosen, 1970; Rosen, 1979) in the aquatic systems of the Upper Usumacinta. These geological events

have likely played a role in isolating the Upper Usumacinta and provide a plausible explanation for the high levels of endemism of freshwater fishes in this endemic area.

Notably, the ichthyofauna present in the endorheic Río Comitan basin and Laguna Montebellos, in the edge of the Upper Usumacinta, is composed of species from both the Upper Usumacinta and Upper Grijalva endemic areas. Three species unique to the Upper Grijalva (*Poeciliopsis hnlickai*, *Chiapaheros grammodes* and *Vieja hartwegi*) and one from the Upper Usumacinta (*Xiphophorus alvarezi*) co-occur within this isolated system located between the two endemic areas. A probable explanation is that historical river capture events allowed the colonization of this system with both Upper Grijalva and Upper Usumacinta species. Additional detailed work in the region is needed to test this hypothesis. Past work has recognized this endorheic region as derived from the Río Lacantún, Upper Usumacinta (Alcocer et al., 2016; Lozano-Vilano & Contreras-Balderas, 1987). However,



our results clearly show that the region shares a history with both systems.

The Lower Grijalva–Usumacinta is hydrologically the largest and most diverse endemic area within the G-U AoE. Much of the diversity in this area is widespread throughout the lowlands of the larger AoE, with some species occurring outside the boundaries of the G-U AoE. Our results are consistent with recent work demonstrating that river sub-basins located in the lower reaches of river networks show connectivity due to historical (e.g. Last Glacial Maximum) and present-day (e.g. seasonal flooding) climatic events across the globe (Carvajal-Quintero et al., 2019; Dias et al., 2014). This connectivity has played a role in shaping present-day patterns of distribution observed for freshwater fishes at different scales (e.g. species, assemblages) in aquatic environments of the G-U AoE and neighbouring regions.

The Eastern Maya Block endemic area is interesting because no unique assemblage defines it. While this endemic area can be recognized based on the unique occurrence of *Poecilia teresae*, the area is largely recovered as a cluster given its remarkably similar species composition to the Lower Grijalva and Usumacinta Rivers (i.e. dominance of widespread species) and dissimilarity to the Northern Maya Block and rivers of southern Belize now part of the Polochic–Cahabón AoE. Previous hypotheses have proposed that the Maya Mountains in Belize acted as refugia for freshwater fishes (Strecker, Faúndez, & Wilkens, 2004) and other taxa (see Gutiérrez-García & Vázquez-Domínguez, 2013) in Northern Middle America during the Plio-Pleistocene Last Glacial Maximum. Our distributional and molecular phylogeographical evidence (see below) rejects this hypothesis (at least for freshwater fishes) given the dominance of widespread species in the area and prevalence of widespread, common haplotypes shared among endemic areas.

The Northern Maya Block is the only endemic area that was not statistically supported by the ANOSIM analysis when contrasted with the upper Usumacinta and upper Grijalva (Table 3) which are not geographically adjacent. Despite this we recognize the Northern Maya Block as part of the geographical structure present in the G-U AoE based on geological and biological (distribution of endemics) evidence. The karst physiography, climatic conditions and geological history of the Yucatán Peninsula are quite different from the rest of the G-U AoE, with few large rivers but numerous cenotes (sinkholes; McMahan, Davis, et al., 2013; Vázquez-Domínguez & Arita, 2010). The Yucatán platform experienced marine introgression and portions of the peninsula were submerged until the Pliocene (Lopez-Ramos, 1975; Vázquez-Domínguez & Arita, 2010) and colonization by Neotropical fishes (e.g. cichlids, poeciliids and characids) was recent (Miller, Minckley, & Norris, 2005; Strecker et al., 2004). While depauperate in diversity of Neotropical fishes, this endemic area is diagnosed by the distribution of 11 endemic species (Table S5) of which ~72% are cyprinodontid fishes (non-dominant taxa category) with Nearctic affinities and relatives in the Greater Antilles. Furthermore, species indicator analysis identified nine species (Table S3) supporting the recognition of the Northern Maya Block. Therefore, recognition of this endemic area is useful

to understand and describe the evolutionary history of some taxonomic groups (e.g. cyprinodontids: see Miller et al., 2005) that are part of the assemblage of freshwater fishes in this geologically complex region.

4.3 | Phylogeographic-level patterns within the G-U AoE

Our comparative phylogeographical assessment provides independent evidence of fine-scale geographical structure within the G-U AoE. Using six widespread species, we recovered two general phylogeographical patterns, with three widespread (panmictic; Figure 5) and three phylogeographically structured species across Northern Middle America (Figure 4). Two species of cichlids (*V. melanura* and *P. splendida*) and one species of poeciliid (*X. hellerii*) lack phylogeographical structure within the G-U AoE. These results are congruent with other findings for cichlids and poeciliids in the region (Alda, Reina, Doadrio, & Bermingham, 2013; Harrison et al., 2014; McMahan, Geheber, & Piller, 2010; McMahan, Kutz, et al., 2017). In contrast, two species of poeciliids (*B. belizanus* and *G. sexradiata*) and one cichlid (*T. salvini*) exhibit congruent and complex phylogeographical structure within the G-U AoE. These three species possess distinct mitochondrial lineages in the Upper Usumacinta endemic area (Figure 4). A second widespread mitochondrial lineage is recovered for all three species across the lowlands of Northern Middle America from west of the G-U AoE across the Lower G-U, Northern Maya Block and the Eastern Maya Block endemic areas. Our comparative phylogeographical analysis of all six species supports the recognition of widespread distributions across the lowlands of Northern Middle America (Figures 4 and 5) that extend beyond the boundaries of the G-U AoE, as additionally recovered based on distributional data. Furthermore, despite widespread distributions, the species studied are rare in the Upper Grijalva and seem to occur mostly in transitional zones between the Lower-Grijalva–Usumacinta and Upper Grijalva endemic areas or near manmade dams suggesting that these populations might be established recently in the lower edges of the Upper Grijalva endemic area.

The congruent phylogeographical structure recovered supports distinctiveness of the Upper Usumacinta endemic area (as identified by the assemblage level; Figure 2) and highlight unrecognized diversity of freshwater fishes within the region. A plausible hypothesis to explain the phylogeographical structure recovered are the climatic events that took place during the Pleistocene within the region. Approximately 18,000 to 11,000 years before present (bp), the environmental conditions in northern Central America resulted in aridity in the region and a reduced amount of surface water availability (Anselmetti et al., 2006; Hodell et al., 2008; Hulsey & López-Fernández, 2011). This resulted in desiccation or near desiccation of several bodies of water in the region (Anselmetti et al., 2006; Brenner et al., 2002; Hodell et al., 2008). These environmental changes likely created geographical complexity in the region



that potentially lead to the isolation of populations in the upper reaches of the Usumacinta River. During the Holocene, 10,500 and 8,000 bp, after the driest period, a wet phase occurred (Hodell et al., 2008), which increased water availability in the region. This increase in water availability likely resulted in the creation of flood plains, and reconnection of formerly and present-day disconnected water bodies and river basins in the lowlands of Northern Middle America.

One plausible explanation for the contrasting phylogeographical patterns recovered in our study (phylogeographical structure versus no structure) is that not all co-distributed species of freshwater fishes in Northern Middle America responded to the same geological or climatic events (Platnick & Nelson, 1978) that lead to unique lineages in the Upper reaches of the Usumacinta River in some species. Incongruent phylogeographical patterns have been recovered with other taxa in different geographical regions (e.g. Arbogast & Kenagy, 2001; Bowen et al., 2014; Soltis, Morris, McLachlan, Manos, & Soltis, 2006). Our comparative phylogeographical analysis let us to infer a general pattern for the freshwater fish fauna of Northern Middle America based on molecular data in which we identified unique lineages in the Upper Usumacinta endemic area and a widespread lineage in the lowlands of Northern Middle America. This general pattern can be tested using molecular data of other widespread species across the larger G-U AoE.

4.4 | Influence of scale on biogeographical inference

One of the most fundamental aspects to the study of biogeography is the consideration of scale: temporal, geographic and taxonomic (Crother & Murray, 2011; Murray & Crother, 2016). At higher scales, such as illustrated for the region of Middle America, the G-U AoE provides useful resolution for differentiating among biogeographical hypotheses (e.g. timing of colonization of Neotropical elements) for freshwater fishes (Matamoros et al., 2015). But to understand and to propose hypotheses regarding the processes that have generated and maintain these unique and diverse assemblages of freshwater fishes in the northern Neotropics, fine-scale biogeographical units can provide better resolution. Our results support complexity, with up to five endemic areas within the G-U AoE depending on the taxa under study, and underscore the nested nature of biogeographical units.

The combination of similarity analyses based on fine-scale species-level distributions and phylogeographic-level genetic analyses support our hypotheses that the evolutionary and biogeographical history of freshwater fishes in the G-U AoE is more complex than a single geographical unit. Furthermore, our results agree with recent work that highlight the reticulated history of river networks in the Neotropics and we demonstrate that present-day river basins (i.e. Usumacinta and Grijalva rivers) are not single historical entities (see Dagosta & Pinna, 2017). Thus, depending on scale of inquiry and taxa under study the use of river basins as single geographical units might not be adequate in all cases when evaluating biogeographical patterns in this region.

Historical isolation has led to unique biodiversity, assemblages and molecular diversity in the upper reaches of the Grijalva and the Usumacinta rivers. Additionally, historical and present-day connections in lower reaches of the Grijalva and Usumacinta river basins have likely resulted in high similarity of assemblages, widespread species and gene flow shaping and maintaining the Lower Grijalva-Usumacinta as a single historical unit in Northern Middle America.

Finally, our proposed classification of endemic areas allows for elimination of the use of subjective names for the same geographical units (e.g. Mendoza-Carranza et al., 2018; Soria-Barreto, González-Díaz, Castillo-Domínguez, Álvarez-Pliego, & Rodiles-Hernández, 2018) that only cause confusion and hinder communication of scientific findings among scientists, conservation managers, policymakers and the general public. We encourage the use of these endemic areas in future studies that range from evolutionary, biogeographic, systematic, ecological and conservation biology of freshwater biota in the region (México, Guatemala and Belize). We reiterate that the utility of these endemic areas are scale- and taxa-dependent. Ultimately, our proposed area classification provides a testable hypothesis that will help to shed light on the geography of speciation and complex evolutionary history of freshwater fishes in the Northern Neotropics.

ACKNOWLEDGEMENTS

This work was supported, in part, by a graduate student research award from the Society of Systematic Biologists to DJE, funds from the National Science Foundation (DEB 1354149) to PC and (DEB 1354930) and a Schlieder Professorship to KRP. CDM thanks the Negaunee Foundation and Bruce and Mary Feay for their support of his work. We thank Susan Mochel and Kevin Swagel (FMNH) for assistance with specimen access at FMNH and Bill Ludt (NHMLA) and Hernán López-Fernández and Randy Singer (UMMZ) for assisting with confirming specimen identifications. We would like to thank Consejo Nacional de Areas Protegidas (CONAP) for granting research (DRM/002/2018), collecting (No. I-001-13) and export (No. 212/2018 and 213/2018) permits in Guatemala, especially Airam López and Héisel Arriola for their instrumental assistance through this process. Mercedes Barrios (CECON), Sergio Pérez and Rosalito Barrios (Escuela de Biología USAC), Josue García (CEMA), César Fuentes, Elisa Blanda, Lucia López, Elyse Parker, Francis López, and Marvin Xajil contributed invaluable collaboration to this project. DJE thanks Fernando Alda (University of Tennessee at Chattanooga), Jeremy Brown (LSU) and Pavel García (University of Montana) for their feedback on the analytical framework on early versions of this work. We also thank Dr. Rosemary Gillespie, Dr. Jani Heino and the anonymous reviewers for their insightful comments and suggestions that improved our work.

DATA AVAILABILITY STATEMENT

The dataset used in this publication is available in Table S1. Sequence data have been submitted to the NIH genetic sequence database, GenBank® - www.ncbi.nlm.nih.gov/genbank, under accession numbers MT675127-188, MT680044-134, and MT702635-731 (see Table S2).

ORCID

Diego J. Elías  <https://orcid.org/0000-0003-4215-0384>

Caleb D. McMahan  <https://orcid.org/0000-0003-2113-8554>

Wilfredo A. Matamoros  <https://orcid.org/0000-0002-6241-5354>

Kyle R. Piller  <https://orcid.org/0000-0003-1289-9351>

Prosanta Chakrabarty  <https://orcid.org/0000-0003-0565-0312>

REFERENCES

- Alcocer, J., Oseguera, L. A., Sánchez, G., González, C. G., Martínez, J. R., & González, R. (2016). Bathymetric and morphometric surveys of the Montebello Lakes, Chiapas. *Journal of Limnology*, 75, 56–65. <https://doi.org/10.4081/jlimnol.2016.1343>
- Abell, R., Thieme, M. L., Revenga, C., Bryner, M., Kottelat, N., Bogutskaya, N., Petry, P. (2008). Freshwater ecoregions of the world: A new map of biogeographic units for freshwater biodiversity conservation. *BioScience*, 58, 403–414. <https://doi.org/10.1641/B580507>
- Alda, F., Reina, R. G., Doadrio, I., & Bermingham, E. (2013). Phylogeny and biogeography of the *Poecilia sphenops* species complex (Actinopterygii, Poeciliidae) in Central America. *Molecular Phylogenetics and Evolution*, 66, 1011–1026. <https://doi.org/10.1016/j.ympev.2012.12.012>
- Anselmetti, F. S., Ariztegui, D., Hodell, D. A., Hillesheim, M. B., Brenner, M., Gilli, A., ... Mueller, A. D. (2006). Late Quaternary climate-induced lake level variations in Lake Petén Itzá, Guatemala, inferred from seismic stratigraphic analysis. *Palaeogeography, Palaeoclimatology, Palaeoecology*, 230, 52–69. <https://doi.org/10.1016/j.palaeo.2005.06.037>
- Arbogast, B. S., & Kenagy, G. J. (2001). Comparative phylogeography as an integrative approach to historical biogeography. *Journal of Biogeography*, 28, 819–825. <https://doi.org/10.1046/j.1365-2699.2001.00594.x>
- Bacon, C. D., Silvestro, D., Jaramillo, C., Smith, B. T., Chakrabarty, P., & Antonelli, A. (2015). Biological evidence supports an early and complex emergence of the Isthmus of Panama. *Proceedings of the National Academy of Sciences of the United States of America*, 112, 6110–6115. <https://doi.org/10.1073/pnas.1423853112>
- Bandelt, H. J., Foster, P., & Röhl, A. (1999). Median-joining networks for inferring intraspecific phylogenies. *Molecular Phylogenetics and Evolution*, 16, 37–48. <https://doi.org/10.1093/oxfordjournals.molbev.a026036>
- Barrientos, C., Elías, D., & Quintana, Y. (2015). Fishes from lake Yaxhá, Mayan Biosphere reserve, Petén, Guatemala. *Check List*, 11(5), 1751. <https://doi.org/10.15560/11.5.1751>
- Barrientos, C., Quintana, Y., Elías, D. J., & Rodiles-Hernández, R. (2018). Peces nativos y pesca artesanal en la cuenca Usumacinta, Guatemala. *Revista Mexicana De Biodiversidad*, 89(suplem. 2018), S118–S130. <https://doi.org/10.22201/ib.20078706e.2018.4.2180>
- Bowen, B. W., Shanker, K., Yasuda, N., Celia, M., Malay, M. C. D., von der Heyden, S., ... Toonen, R. J. (2014). Phylogeography unplugged: Comparative surveys in the genomic era. *Bulletin of Marine Science*, 90, 13–46. <https://doi.org/10.5343/bms.2013.1007>
- Brenner, M., Rosenmeier, M. F., Hodell, D. A., & Curtis, J. H. (2002). Paleolimnology of the Maya lowlands: Long-term perspectives on interactions among climate, environment, and humans. *Ancient Mesoamerica*, 13, 141–157. <https://doi.org/10.1017/S0956536102131063>
- Brocard, G., Willenbring, J., Suski, B., Audra, F., Authermayou, C., Cosenza-Murales, B., ... Teyssier, C. (2012). Rate and processes of river network rearrangement during incipient faulting: The case of the Cahabón River, Guatemala. *American Journal of Science*, 312, 449–507. <https://doi.org/10.2475/05.2012.01>
- Brown, S. D. J., Collins, R. A., Boyer, S., Lefort, M., Malumbres-Olarte, J., Vink, C. J., & Cruickshank, R. H. (2012). SPIDER: An R package for the analysis of species identity and evolution, with particular reference to DNA barcoding. *Molecular Ecology Resources*, 12, 562–565. <https://doi.org/10.1111/j.1755-0998.2011.03108.x>
- Bussing, W. A. (1985). Patterns of distribution of the Central America ichthyofauna. In F. G. Stehli, & S. D. Webb (Eds.), *The great American biotic interchange* (pp. 453–473). New York: Springer US.
- Calinski, T., & Harabasz, J. (1974). A dendrite method for cluster analysis. *Communications in Statistics*, 3, 1–27. <https://doi.org/10.1080/03610927408827101>
- Carvajal-Quintero, J., Villalobos, F., Oberdorff, T., Grenouillet, G., Brosse, S., Hugueny, B., ... Tedesco, P. A. (2019). Drainage network position and historical connectivity explain global patterns in freshwater fishes' range size. *Proceedings of the National Academy of Sciences of the United States of America*, 116, 13434–13439. <https://doi.org/10.1073/pnas.1902484116>
- Castoe, T. A., Daza, J. M., Smith, E. N., Sasa, M. M., Kuch, U., Campbell, J. A., ... Parkinson, C. L. (2009). Comparative phylogeography of pitvipers suggests a consensus of ancient Middle American highland biogeography. *Journal of Biogeography*, 36, 88–103. <https://doi.org/10.1111/j.1365-2699.2008.01991.x>
- Chakrabarty, P., & Albert, J. S. (2011). Not so fast a new take on the great American biotic interchange. In J. S. Albert, & R. E. Reis (Eds.), *Historical biogeography of Neotropical freshwater fishes* (pp. 293–305). Berkeley, CA: University of California Press.
- Clarke, K. R., & Warwick, R. M. (1994). Similarity-based testing for community pattern: The two-way layout with no replication. *Marine Biology*, 118, 167–176. <https://doi.org/10.1007/BF00699231>
- Clement, M., Posada, D., & Crandall, K. A. (2000). TCS: A computer program to estimate gene genealogies. *Molecular Ecology*, 9, 1657–1659. <https://doi.org/10.1046/j.1365-294x.2000.01020.x>
- Cousseau, M. B., Pequeño, G., Mabrugaña, E., Lucifora, L. O., Martínez, P., & Giussi, A. (2019). The Magellanic province and its fish fauna (South America): Several provinces or one? *Journal of Biogeography*, 47, 220–234. <https://doi.org/10.1111/jbi.13735>
- Crother, B. I., & Murray, C. M. (2011). Ontology of areas of endemism. *Journal of Biogeography*, 38, 1009–1015. <https://doi.org/10.1111/j.1365-2699.2011.02483.x>
- Dagosta, F. C. P., & Pinna, M. (2017). Biogeography of Amazonian fishes: Deconstructing river basins as biogeographic units. *Neotropical Ichthyology*, 15(3), e170034. <https://doi.org/10.1590/1982-0224-20170034>
- Daza, J. M., Castoe, T. A., & Parkinson, C. L. (2010). Using regional comparative phylogeographic data from snake lineages to infer historical processes in Middle America. *Ecography*, 33, 343–354. <https://doi.org/10.1111/j.1600-0587.2010.06281.x>
- Dias, M. S., Oberdorff, T., Hugueny, B., Leprieux, F., Jézéquel, C., Cornu, J.-F., ... Tedesco, P. A. (2014). Global imprint of historical connectivity on freshwater fish biodiversity. *Ecology Letters*, 17, 1130–1140. <https://doi.org/10.1111/ele.12319>
- Domínguez, M. C., Roig-Juñent, S., Tassin, J. J., Ocampo, F. C., & Flores, G. E. (2006). Areas of endemism of the Patagonian steppe: An approach based on insect distributional patterns using endemicity analysis. *Journal of Biogeography*, 33, 1527–1537. <https://doi.org/10.1111/j.1365-2699.2006.01550.x>
- Dufrène, M., & Legendre, P. (1997). Species assemblages and indicator species: The need for a flexible asymmetrical approach. *Ecological Monographs*, 67, 345–366. [https://doi.org/10.1890/0012-9615\(1997\)067\[0345:SAAIST\]2.0.CO;2](https://doi.org/10.1890/0012-9615(1997)067[0345:SAAIST]2.0.CO;2)
- Ebach, M. C., & Michaux, B. (2017). Establishing a framework for a natural area taxonomy. *Acta Biotheoretica*, 65, 167–177. <https://doi.org/10.1007/s10441-017-9310-y>
- Ennen, J. R., Agha, M., Matamoros, W. A., Hazzard, S. C., & Lovich, J. E. (2016). Using climate, energy, and spatial-based hypotheses to interpret macroecological patterns of North America chelonians. *Canadian Journal of Zoology*, 94, 453–461. <https://doi.org/10.1139/cjz-2016-0033>



- Ennen, J. R., Agha, M., Sweat, S. C., Matamoros, W. A., Lovich, J. E., Rhodin, A. G. J., ... Hoagstrom, C. W. (2020). Turtle biogeography: Global regionalization and conservation priorities. *Biological Conservation*, 241, 108323. <https://doi.org/10.1016/j.biocon.2019.108323>
- Fauth, T. E., Bernardo, J., Camara, M., Resetarits, W. J. Jr, Van Buskirk, J., & McCollum, S. A. (1996). Simplifying the jargon of community ecology: A conceptual approach. *The American Naturalist*, 147, 282–286. <https://doi.org/10.1086/285850>
- Fricke, R., Eschmeyer, W. N., & van der Laan, R. (2019). Eschmeyer's catalog of fishes: Genera, species, references. Retrieved from <http://researcharchive.calacademy.org/research/ichthyology/catalog/fishcatmain.asp>
- Gómez-González, A. E., Velásquez-Velásquez, E., Anzueto-Calvo, M. J., & Maza-Cruz, M. F. (2015). Fishes of the Grijalva River basin of Mexico and Guatemala. *Check List*, 11(5), 1726. <https://doi.org/10.15560/11.1726>
- Greenfield, D. W., & Thomerson, J. E. (1997). *Fishes of the continental waters of Belize*. Gainesville, FL: University Press of Florida.
- Gutiérrez-García, T. A., & Vázquez-Domínguez, E. (2013). Consensus between genes and stones in the biogeographic and evolutionary history of Central America. *Quaternary Research*, 79, 311–324. <https://doi.org/10.1016/j.yqres.2012.12.007>
- Hammer, Ø. (2019). *Paleontological statistics version 3.25, reference manual*. Natural History Museum, University of Oslo. 275 pp. Retrieved from <http://folk.uio.no/ohammer/past>
- Harrison, E., Trexler, J. C., Collins, T. M., Vázquez-Domínguez, E., Razo-Mendivil, U., Matamoros, W. A., & Barrientos, C. (2014). Genetic evidence for multiple sources of the non-native fish *Cichlasoma urophthalmus* (Günther; Mayan cichlids) in Southern Florida. *PLoS One*, 9, e104173. <https://doi.org/10.1371/journal.pone.0104173>
- Hart, M. W., & Sunday, J. (2007). Things fall apart: Biological species form unconnected parsimony networks. *Biology Letters*, 3, 509–512. <https://doi.org/10.1098/rsbl.2007.0307>
- Hodell, D. A., Anselmetti, F. S., Ariztegui, D., Brenner, M., Curtis, J. H., Gilli, A., ... Bush, M. B. (2008). An 85-ka record of climate change in lowland Central America. *Quaternary Science Reviews*, 27, 1152–1165. <https://doi.org/10.1016/j.quascirev.2008.02.008>
- Holm, S. (1979). A simple sequential rejective multiple test procedure. *Scandinavian Journal of Statistics*, 6, 65–70.
- Hrbek, T., Seckinger, J., & Meyer, A. (2007). A phylogenetic and biogeographic perspective on the evolution of poeciliid fishes. *Molecular Phylogenetics and Evolution*, 43, 986–998. <https://doi.org/10.1016/j.ympev.2006.06.009>
- Huelsenbeck, J. P., Ronquist, F., Nielsen, R., & Bollback, J. P. (2001). Bayesian inference of phylogeny and its impact on evolutionary biology. *Science*, 294, 2310–2314. <https://doi.org/10.1126/science.1065889>
- Hulse, C. D., & López-Fernández, H. (2011). Nuclear Central America. In J. S. Albert, & R. E. Reis (Eds.), *Historical biogeography of Neotropical freshwater fishes* (pp. 279–291). Berkeley, CA: University of California Press.
- Hurvich, C. M., & Tsai, C. (1989). Regression and time series model selection in small samples. *Biometrika*, 76, 297–307. <https://doi.org/10.1093/biomet/76.2.297>
- INEGI (Instituto Nacional de Estadística y Geografía). (2010). *Documento técnico descriptivo de la red hidrográfica escala 1:50 000*. Aguascalientes.
- Iturralde-Vinent, M. A., & MacPhee, R. D. E. (1999). Paleography of the Caribbean region: Implications for Cenozoic biogeography. *Bulletin of the American Museum of Natural History*, 238, 1–95. <http://hdl.handle.net/2246/1642>
- Katoh, K., & Standley, D. M. (2013). MAFFT Multiple sequence alignment software version 7: Improvements in performance and usability. *Molecular Biology and Evolution*, 30, 772–780. <https://doi.org/10.1093/molbev/mst010>
- Kearse, M., Moir, R., Wilson, A., Stones-Havas, S., Cheung, M., Sturrock, S., ... Drummond, A. (2012). Geneious Basic: An integrated and extendable desktop software platform for the organization and analysis of sequence data. *Bioinformatics*, 28, 1647–1649. <https://doi.org/10.1093/bioinformatics/bts199>
- Kreft, H., & Jetz, W. (2010). A framework for delineating biogeographical regions based on species distributions. *Journal of Biogeography*, 37, 2029–2053. <https://doi.org/10.1111/j.1365-2699.2010.02375.x>
- Lanfear, R., Calcott, B., Ho, S., & Guindon, S. (2012). PartitionFinder: Combined selection of partitioning schemes and substitution models for phylogenetic analyses. *Molecular Biology and Evolution*, 29, 1695–1701. <https://doi.org/10.1093/molbev/mss020>
- Legendre, P., & Legendre, L. (2012). *Numerical ecology* (3rd edn). Amsterdam: Elsevier.
- Lehner, B., & Grill, G. (2013). Global river hydrography and network routing: baseline data and new approaches to study the world's large river systems. *Hydrological Processes*, 27, 2171–2186.
- Leigh, J. W., & Bryant, D. (2015). POPART: Full-feature software for haplotype network construction. *Methods in Ecology and Evolution*, 6, 1110–1116. <https://doi.org/10.1111/2041-210X.12410>
- Lemopoulos, A., & Covain, R. (2019). Biogeography of the freshwater fishes of the Guianas using a partitioned parsimony analysis of endemicity with reappraisal of ecoregional boundaries. *Cladistics*, 35, 106–124. <https://doi.org/10.1111/cla.12341>
- Leroy, B., Dias, M. S., Giraud, E., Huguéy, B., Jézéquel, C., Leprieur, F., ... Tedesco, P. A. (2019). Global biogeographical regions of freshwater fish species. *Journal of Biogeography*, 46(11), 2407–2419. <https://doi.org/10.1111/jbi.13674>
- Librado, P., & Rozas, J. (2009). DnaSP v5: A software for comprehensive analysis of DNA polymorphism data. *Bioinformatics*, 25, 1451–1452. <https://doi.org/10.1093/bioinformatics/btp187>
- Lopez-Ramos, E. (1975). Geological summary of the Yucatan Peninsula. In A. E. M. Nairn, & F. G. Stehli (Eds.), *The Gulf of Mexico and the Caribbean* (pp. 257–282). Boston, MA: Springer. https://doi.org/10.1007/978-1-4684-8535-6_7
- Lozano-Vilano, M. L., & Contreras-Balderas, S. (1987). Lista zoográfica y ecológica de la ictiofauna continental de Chiapas, Mexico. *The Southwestern Naturalist*, 32, 223–236.
- Lundberg, J. G., Sullivan, J. P., Rodiles-Hernández, R., & Hendrickson, D. A. (2007). Discovery of African roots for the Mesoamerican Chiapas catfish, *Lacantunia enigmatica*, require and ancient intercontinental passage. *Proceedings of the Academy of Natural Sciences of Philadelphia*, 156, 39–53. [https://doi.org/10.1635/0097-3157\(2007\)156\[39:DO-ARFT\]2.0.CO;2](https://doi.org/10.1635/0097-3157(2007)156[39:DO-ARFT]2.0.CO;2)
- MacQueen, J. (1967). Some methods for classification and analysis of multivariate observations. In L. Lecam, & J. Neyman (Eds.), *Proceedings of the fifth Berkeley symposium on mathematical statistics and probability* (pp. 281–297). Berkeley, CA: University of California Press.
- Marshall, J. S. (2007). The geomorphology and physiographic provinces of Central America. In J. Bundschuh, & G. E. Alvarado (Eds.), *Central America: Geology, resources and hazards* (Vol. 1, pp. 75–122). London, UK: CRC Press.
- Matamoros, W. A., Hoagstrom, C. W., Schaefer, J. F., & Kreiser, B. R. (2016). Fish faunal provinces of the conterminous United States of America reflect historical geography and familial composition. *Biological Reviews*, 91, 813–832. <https://doi.org/10.1111/brv.12196>
- Matamoros, W. A., Kreiser, B. R., & Schaefer, J. F. (2012). A delineation of Nuclear Middle America biogeographical provinces based on river basin faunistic similarities. *Reviews in Fish Biology and Fisheries*, 22, 351–365. <https://doi.org/10.1007/s11160-011-9232-8>
- Matamoros, W. A., McMahan, C. D., Chakrabarty, P., Albert, J. S., & Schaefer, J. F. (2015). Derivation of the freshwater fish fauna of Central America revisited: Myers's hypothesis in the twenty first century. *Cladistics*, 31, 177–188. <https://doi.org/10.1111/cla.12081>
- McMahan, C. D., Chakrabarty, P., Sparks, J. S., Smith, W. L., & Davis, M. P. (2013). Temporal patterns of diversification across global cichlid biodiversity (Acanthomorpha: Cichlidae). *PLoS One*, 8(8), e71162. <https://doi.org/10.1371/journal.pone.0071162>

- McMahan, C. D., Davis, M. P., Domínguez-Domínguez, O., García-de-León, F. J., Doadrio, I., & Piller, K. R. (2013). From the mountains to the sea: Phylogeography and cryptic diversity within the mountain mullet, *Agonostomus monticola*, (Teleostei: Mugilidae). *Journal of Biogeography*, 40, 894–904. <https://doi.org/10.1111/jbi.12036>
- McMahan, C. D., Geheber, A. D., & Piller, K. R. (2010). Molecular systematics of the enigmatic Middle American genus *Vieja* (Teleostei: Cichlidae). *Molecular Phylogenetics and Evolution*, 57, 1293–1300. <https://doi.org/10.1016/j.ympev.2010.09.005>
- McMahan, C. D., Ginger, L., Cage, M., David, K. T., Chakrabarty, P., Johnston, M., & Matamoros, W. A. (2017). Pleistocene to holocene expansion of the Black-belt Cichlid in Central America, *Vieja maculicauda* (Teleostei:Cichlidae). *PLoS One*, 12(5), e0178439. <https://doi.org/10.1371/journal.pone.0178439>
- McMahan, C. D., Kutz, J., Murray, C. M., Chakrabarty, P., Geheber, A. D., & Elias, D. (2017). Objectively measuring subjectively described traits: Geographic variation in body shape and caudal coloration pattern within *Vieja melanura* (Teleostei: Cichlidae). *Revista De Biología Tropical*, 65, 623–631. <https://doi.org/10.15517/rbt.v65i2.25500>
- Mendoza-Carranza, M., Arévalo-Frías, W., Espinoza-Tenorio, A., Hernández-Lazo, C. C., Álvarez-Merino, A. M., & Rodiles-Hernández, R. (2018). La Importancia y diversidad de los recursos pesqueros del río Usumacinta. *México. Revista Mexicana De Biodiversidad*, 89(supl. dic), 131–146. <http://doi.org/10.22201/ib.20078706e.2018.4.2182>
- Miller, M., Pfeiffer, W., & Schwartz, T. (2010). Creating the CIPRES Science Gateway for inference of large phylogenetic trees. In Proceedings of the Gateway Computing Environments Workshop (GCE). 1–8. <https://doi.org/10.1109/GCE.2010.5676129>
- Miller, R. R. (1966). Geographical distribution of Central American freshwater fishes. *Copeia*, 1966, 773–802. <https://doi.org/10.2307/1441406>
- Miller, R. R., Minckley, W. L., & Norris, S. M. (2005). *Freshwater fishes of México*. Chicago, IL: University of Chicago Press.
- Milligan, G. W., & Cooper, M. C. (1985). An examination of procedures for determining the number of clusters in a data set. *Psychometrika*, 50, 159–179.
- Morrone, J. J. (2009). *Evolutionary biogeography an integrative approach with case studies*. New York, NY: Columbia University Press.
- Morrone, J. J. (2017). *Neotropical biogeography regionalization and evolution*. Boca Raton, FL: CRC Press.
- Murray, C. M., & Crother, B. I. (2016). Entities on a temporal scale. *Acta Biotheoretica*, 64, 1–10. <https://doi.org/10.1007/s10441-015-9269-5>
- Myers, G. (1966). Derivation of the freshwater fish fauna of Central America. *Copeia*, 1966, 766–773. <https://doi.org/10.2307/1441405>
- Nei, M., & Li, W. H. (1979). Mathematical model for studying genetic variation in terms of restriction endonucleases. *Proceedings of the National Academy of Sciences of the United States of America*, 76, 5269–5273. <https://doi.org/10.1073/pnas.76.10.5269>
- Nelson, G., & Platnick, N. I. (1981). *Systematics and biogeography: Cladistics and vicariance*. New York, NY: Columbia University Press.
- Oksanen, J., Kindt, R., Legendre, P., O'Hara, B., Simpson, G. L., Solymos, P., & Wagner, H. (2017). *Vegan: Community ecology package*. R package version 2.0-2.
- Parenti, L. R., & Ebach, M. C. (2009). *Comparative biogeography: Discovering and classifying biogeographical patterns of a dynamic earth*. Berkeley, CA: University of California Press.
- Perdices, A., Bermingham, E., Montilla, A., & Doadrio, I. (2002). Evolutionary history of the genus *Rhamdia* (Teleostei: Pimelodidae) in Central America. *Molecular Phylogenetics and Evolution*, 25, 172–189. [https://doi.org/10.1016/S1055-7903\(02\)00224-5](https://doi.org/10.1016/S1055-7903(02)00224-5)
- Platnick, N. I., & Nelson, G. (1978). A method of analysis for historical biogeography. *Systematic Biology*, 27, 1–16. <https://doi.org/10.2307/2412808>
- Quintana, Y., Barrientos, C., & Rodiles-Hernández, R. (2019). Range extension for *Lacantunia enigmatica* Rodiles-Hernández, Hendrickson & Lundberg, 2005 (Siluriformes, Lacantuniidae) in the Usumacinta river basin, Guatemala. *Check List*, 15(1), 161–167. <https://doi.org/10.15560/15.1.161>
- R Core Team. (2015). *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing.
- Rambaut, A., Drummond, A. J., Xie, D., Baele, G., & Suchard, M. A. (2018). Posterior summarization in Bayesian phylogenetics using Tracer 1.7. *Systematic Biology*, 67, 901–904. <https://doi.org/10.1093/sysbio/syy032>
- Ree, R. H., Moore, B. R., Webb, C. O., & Donoghue, M. J. (2005). A likelihood framework for inferring the evolution of geographic range on phylogenetic trees. *Evolution*, 59, 2299–2311. <https://doi.org/10.1111/j.0014-3820.2005.tb00940.x>
- Reeves, R. G., & Bermingham, E. (2006). Colonization, population expansion, and lineage turnover: Phylogeography of Mesoamerican characiform fish. *Biological Journal of the Linnean Society*, 88, 235–255. <https://doi.org/10.1111/j.1095-8312.2006.00619.x>
- Revell, L. (2012). phytools: An R package for phylogenetic comparative biology (and other things). *Methods in Ecology and Evolution*, 3, 217–223. <https://doi.org/10.1111/j.2041-210X.2011.00169.x>
- Reyjol, Y., Hugueny, B., Pont, D., Bianco, P. G., Beier, U., Caiola, N., ... Virbickas, T. (2007). Patterns in species richness and endemism of European freshwater fish. *Global Ecology and Biogeography*, 17, 65–75. <https://doi.org/10.1111/j.1466-8238.2006.00264.x>
- Ričan, O., Piálek, L., Zardoya, R., Doadrio, I., & Zrzavy, J. (2013). Biogeography of the Mesoamerican Cichlidae (Teleostei: Heroini): Colonization through the GAARlandia land bridge and early diversification. *Journal of Biogeography*, 40, 579–593. <https://doi.org/10.1111/jbi.12023>
- Rice, W. R. (1989). Analyzing tables of statistical tests. *Evolution*, 43, 223–225. <https://doi.org/10.1111/j.1558-5646.1989.tb04220.x>
- Rodiles-Hernández, R., Hendrickson, D. A., Lundberg, J. G., & Humpries, J. M. (2005). *Lacantunia enigmatica* (Teleostei: Siluriformes) a new and phylogenetically puzzling freshwater fish from Mesoamerica. *Zootaxa*, 1000, 1–24. <https://doi.org/10.11646/zootaxa.1000.1.1>
- Rosen, D. E. (1967). New Poeciliid fishes from Guatemala, with comments on the origins of some South and Central American forms. *American Museum Novitates*, 2303.
- Rosen, D. E. (1970). A new Tetragonopterine characid fish from Guatemala. *American Museum Novitates*, 2435.
- Rosen, D. E. (1975). A vicariance model of Caribbean biogeography. *Systematic Biology*, 24, 431–464. <https://doi.org/10.1093/sysbio/24.4.431>
- Rosen, D. E. (1979). Fishes from the uplands and intermontane basins of Guatemala: Revisionary studies and comparative geography. *Bulletin of the American Museum of Natural History*, 162, 267–376. <http://hdl.handle.net/2246/1281>
- Rosen, D. E. (1985). Geological hierarchies and biogeographic congruence in the Caribbean. *Annals of the Missouri Botanical Garden Press*, 72, 636–659. <https://doi.org/10.2307/2399218>
- Rosenberg, N. A. (2007). Statistical test for taxonomic distinctiveness from observations of monophyly. *Evolution*, 61, 317–323. <https://doi.org/10.1111/j.1558-5646.2007.00023.x>
- Sabaj, M. H. (2016). Standard symbolic codes for institutional resource collections in Herpetology and Ichthyology: An online reference. Version 6.5. (16 August 2016).
- Saraçlı, S., Dogan, N., & Dogan, I. (2013). Comparison of hierarchical cluster analysis methods by cophenetic correlation. *Journal of Inequalities and Applications*, 2013, 203. <https://doi.org/10.1186/1029-242X-2013-203>
- Smith, S. A., & Bermingham, E. (2005). The biogeography of lower Mesoamerican freshwater fishes. *Journal of Biogeography*, 32, 1835–1854. <https://doi.org/10.1111/j.1365-2699.2005.01317.x>
- Sokal, R. R., & Rohlf, F. J. (1962). The comparison of dendrograms by objective methods. *Taxon*, 11, 33–40. <https://doi.org/10.2307/1217208>



- Soltis, D. E., Morris, A. B., McLachlan, J. S., Manos, P. S., & Soltis, P. S. (2006). Comparative phylogeography of unglaciated eastern North America. *Molecular Ecology*, 15, 4261–4293. <https://doi.org/10.1111/j.1365-294X.2006.03061.x>
- Soria-Barreto, M., Gonzáles-Díaz, A. A., Castillo-Domínguez, A., Álvarez-Pliego, N., & Rodiles-Hernández, R. (2018). Diversidad íctica en la Cuenca del Usumacinta, México. *Revista Mexicana De Biodiversidad*, 89(supl. dic), 100–117. <https://doi.org/10.22201/ib.20078706e.2018.4.2462>
- Strecker, U., Faúndez, V. H., & Wilkens, H. (2004). Phylogeography of surface and cave *Astyanax* (Teleostei) from Central and North America based on cytochrome b sequence data. *Molecular Phylogenetics and Evolution*, 33, 469–481. <https://doi.org/10.1016/j.ympev.2004.07.001>
- Suarez, G. A. (2011). *Cuencas hidrográficas de Guatemala*. Guatemala: Ministerio de Ambiente y Recursos Naturales.
- Tagliacollo, V. A., Duke-Sylvester, S. M., Matamoros, W. A., Chakrabarty, P., & Albert, J. S. (2017). Coordinated dispersal and pre-Isthmian assembly of the Central American ichthyofauna. *Systematic Biology*, 66, 183–1996. <https://doi.org/10.1093/sysbio/syv064>
- Templeton, A. R., Crandall, K. A., & Sing, C. F. (1992). A cladistics analysis of phenotypic association with haplotypes inferred from restriction endonuclease mapping and DNA sequence data. III. Cladogram estimation. *Genetics*, 132, 619–633.
- Unmack, P. J. (2001). Biogeography of Australian freshwater fishes. *Journal of Biogeography*, 28, 1053–1089. <https://doi.org/10.1046/j.1365-2699.2001.00615.x>
- Vázquez-Domínguez, E., & Arita, H. T. (2010). The Yucatan peninsula: Biogeographical history 65 millions years in the making. *Ecography*, 33, 212–219.
- Watterson, G. A. (1975). On the number of segregating sites in genetical models without recombination. *Theoretical Population Biology*, 7, 256–276. [https://doi.org/10.1016/0040-5809\(75\)90020-9](https://doi.org/10.1016/0040-5809(75)90020-9)

BIOSKETCH

Diego J. Elías is a Ph.D. student in the Ichthyology Section of the Museum of Natural Science at Louisiana State University. His research interests centre on systematics and comparative phylogeography and biogeography of Neotropical fishes. All authors are interested in the systematics and biogeography of Middle American freshwater fishes.

Authors' contributions: D.J.E conceived the ideas with all authors. Samples were collected by all authors. D.J.E. compiled and curated distributional data and collected molecular data. D.J.E. analysed the data; D.J.E and C.D.M led the writing with contributions from all authors.

SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section.

How to cite this article: Elías DJ, McMahan CD, Matamoros WA, Gómez-González AE, Piller KR, Chakrabarty P. Scale(s) matter: Deconstructing an area of endemism for Middle American freshwater fishes. *J Biogeogr.* 2020;47:2483–2501. <https://doi.org/10.1111/jbi.13941>