

RESEARCH

Open Access



Genome-wide species delimitation analyses of a silverside fish species complex in central Mexico indicate taxonomic over-splitting

Victor Julio Piñeros^{1†}, Carmen del R. Pedraza-Marrón^{2†}, Isaí Betancourt-Resendes³, Nancy Calderón-Cortés^{1*}, Ricardo Betancur-R² and Omar Domínguez-Domínguez^{4,5*}

Abstract

Background: Delimiting species across a speciation continuum is a complex task, as the process of species origin is not generally instantaneous. The use of genome-wide data provides unprecedented resolution to address convoluted species delimitation cases, often unraveling cryptic diversity. However, because genome-wide approaches based on the multispecies coalescent model are known to confound population structure with species boundaries, often resulting in taxonomic over-splitting, it has become increasingly evident that species delimitation research must consider multiple lines of evidence. In this study, we used phylogenomic, population genomic, and coalescent-based species delimitation approaches, and examined those in light of morphological and ecological information, to investigate species numbers and boundaries comprising the *Chirostoma* “*humboltianum* group” (family Atherinidae). The *humboltianum* group is a taxonomically controversial species complex where previous morphological and mitochondrial studies produced conflicting species delimitation outcomes. We generated ddRADseq data for 77 individuals representing the nine nominal species in the group, spanning their distribution range in the central Mexican plateau.

Results: Our results conflict with the morphospecies and ecological delimitation hypotheses, identifying four independently evolving lineages organized in three geographically cohesive clades: (i) *chapalae* and *sphyaena* groups in Lake Chapala, (ii) *estor* group in Lakes Pátzcuaro and Zirahuén, and (iii) *humboltianum* sensu stricto group in Lake Zacapu and Lerma river system.

Conclusions: Overall, our study provides an atypical example where genome-wide analyses delineate fewer species than previously recognized on the basis of morphology. It also highlights the influence of the geological history of the Chapala-Lerma hydrological system in driving allopatric speciation in the *humboltianum* group.

[†]Victor Julio Piñeros and Carmen del R. Pedraza-Marrón contributed equally to this work

*Correspondence: ncalderon@enesmorelia.unam.mx; goodeido@yahoo.com.mx

¹ Laboratorio de Ecología Molecular, Escuela Nacional de Estudios Superiores Unidad Morelia, Universidad Nacional Autónoma de México, Antigua Carretera a Pátzcuaro 8701, Ex-Hacienda de San José de La Huerta, 58190 Morelia, Michoacán, Mexico

⁴ Laboratorio de Biología Acuática, Facultad de Biología, Universidad Michoacana de San Nicolás de Hidalgo, Edificio “R” Planta Baja, Ciudad Universitaria, 58030 Morelia, Michoacán, Mexico

Full list of author information is available at the end of the article



Keywords: *Chirostoma humboldtianum*, Freshwater fishes, Genome-wide data, Genomic structure, SNP loci, Species delimitation

Background

Species delimitation, which consists of determining the boundaries between species, is a fundamental aim of evolutionary biology, not only to document the extent of organismal diversity, but also to conserve and manage biodiversity. Species delimitation is challenging, as speciation is generally not instantaneous, where there is a diffuse area between populations and species—known as the speciation continuum [1]. The inference of species delimitation thus becomes particularly challenging in recently diverged or closely related species, where it is difficult to differentiate the population-level structure from distinct species as they often undergo incomplete lineage sorting (ILS) or hybridization [2].

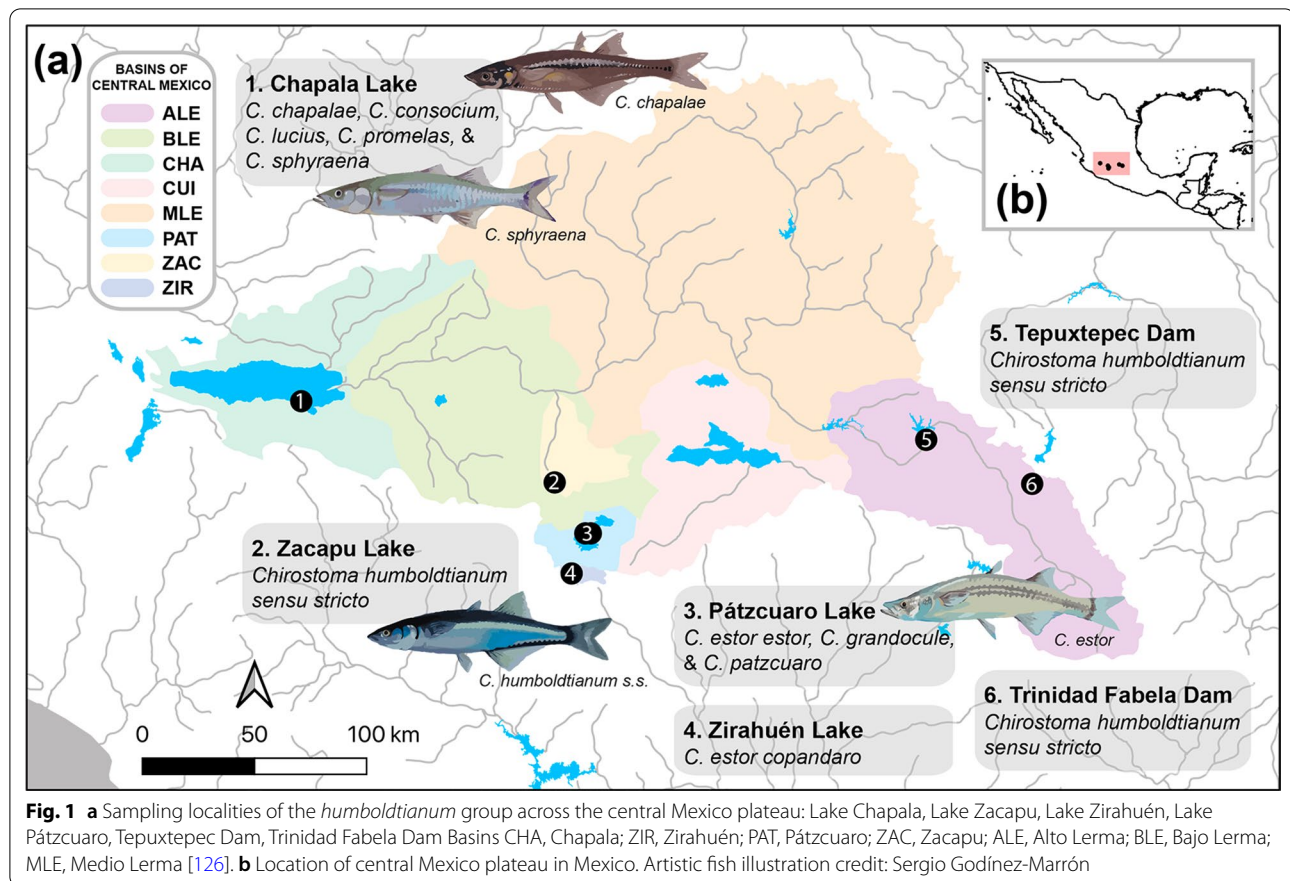
The use of next generation sequencing techniques offers the opportunity to generate robust datasets consisting of many samples at thousands of loci to address species delimitation conflicts with unprecedented resolution, allowing for the detection of genetic structure at a finer scale [3, 4]. Coupled with species delimitation approaches under the multispecies coalescent (MSC) model, genome-wide data provide considerable statistical power to identify recently differentiated species boundaries [4]. However, the MSC model has been recently questioned as it may confound population structure with putative species, often leading to overestimating species diversity [4–6]. Thus, the implementation of MSC analyses combined with other approaches, such as phylogenomic and population genomic methods (e.g., multivariate, assignment, and genetic differentiation analyses), in conjunction with other lines of evidence (e.g., biogeographic, ecological, or life-history), confers a robust framework to assess species delimitations [7].

The New World silverside fishes in the *Chirostoma humboldtianum* species complex (Atherinopsidae; hereafter referred to as the “*humboldtianum* group”), distributed in central Mexico, have been of high economic and cultural importance since pre-Hispanic times [8]. Currently, *Chirostoma* species are considered one of the most important fishery resources in the region, where they have been severely overexploited [8, 9]. The *humboldtianum* group represents an interesting system in terms of species delimitation as their nominal species have been a subject of taxonomic controversy, with varying numbers of species recognized based on morphologic [10] or molecular [11] data. This group is geographically restricted to the lacustrine environments of the central Mexico plateau occurring mainly in Lakes Chapala,

Pátzcuaro, Zirahuén, and Zacapu [10] (Fig. 1)—an area comprising roughly 1716.5 km² [12, 13]—making it feasible to address species delimitation by examining the extent of species diversity based on samples collected across their range.

Nine nominal species in the *humboldtianum* group have been described based on morphological, osteological, meristic, and allozyme data: *C. chapalae*, *C. lucius*, *C. promelas*, *C. sphyraena*, and *C. consocium* from Lake Chapala; *C. grandocule*, and *C. patzcuaro* from Lake Pátzcuaro; *C. humboldtianum* sensu stricto from Lake Zacapu and the Lerma River System; and two subspecies for *C. estor*, *C. e. estor*, and *C. e. copandaro* from Lakes Pátzcuaro and Zirahuén, respectively [10, 14–17] (Fig. 1). While a molecular study using a single mitochondrial marker (NADH dehydrogenase subunit 2; *ND2*) failed to identify support for the monophyly for each of the nine nominal species, and also to resolve the phylogenetic relationships within the species complex [18], a recent phylogeographic study based on two mitochondrial (cytochrome b, *Cytb*; and a fragment of the hypervariable control region, *D-loop*) and one nuclear marker (first intron of the S7 ribosomal protein) identified five genetic groups, suggesting that the diversity in this species complex may be overestimated [11]. The genetic structure in these groupings largely clusters by lakes, indicating that diversification in this group is the result of historical geological and hydrographic processes.

Ecological divergence has also been suggested as a driver of speciation in this group given the high morphological disparity across traits related to different habitats [18, 19]. For instance, ecotypes (‘peces blancos’ and ‘charales’; 117–300 mm and 70–142 mm standard length SL, respectively) based on the size of the species could coexist by feeding on prey of different sizes [10]. Betancourt-Resendes et al. [11] observed intra-lacustrine patterns of evolution in Lakes Chapala and Pátzcuaro (two well-differentiated genetic groups within the Lake Chapala representing five morphospecies, and two genetic groups in Lake Pátzcuaro representing three morphospecies), suggesting that these lineages probably evolved through ecological speciation. Although no strong evidence of segregation of the nominal species by trophic specialization has been reported, these patterns remain to be evaluated [14, 20, 21]. The discrepancy in the recognition of nine morphological *versus* five mitonuclear species (sensu [11]), together with the economic importance and critical conservation status



of these species, highlight the necessity of an accurate estimation of species boundaries in this group.

Here, we used double-digest restriction site-associated DNA sequencing (ddRADseq) to generate genome-wide single nucleotide polymorphism (SNP) data from all nine nominal species in the *humboldtianum* group, sampled throughout their distribution range. We also considered previous studies to interpret the observed genetic structure in the light of multiple lines of evidence. By generating this molecular dataset our study aims to: (i) test whether the morphospecies or ecotypes are concordant with the genomic groupings, (ii) investigate the number and boundaries of species in the *humboldtianum* group, (iii) examine the evolutionary processes driving the divergence in this species complex, and (iv) discuss conservation implications in light of the proposed species boundaries and their observed genetic structure. Ultimately, our study adds to the growing body of work addressing complex species delimitation scenarios with genomic data, while providing critical information to guide conservation and management efforts in the *humboldtianum* group.

Results

ddRADseq assembly and datasets

We obtained $\sim 3.6 \times 10^8$ sequence raw reads, of which $\sim 3.5 \times 10^8$ (97.6%) passed quality filters. On average, there were *ca.* 2 million reads per sample. The de novo assembly using a combination of *m5M4n5* parameters generated 275,533 putative loci with an average coverage of 16.1 per sample. After filtering for missing data and different *maf* thresholds, we generated five final matrices (A–E), which varied from 1887 to 33,716 SNPs and 7.7–15.77% of missing data. These final matrices were also filtered to contain neutral-only and outlier loci, ranging from 1795 to 33,346 and 82–370 SNP loci, respectively. For more information, refer to the “Methods” and Additional file 1, 2 and 3.

Multivariate analyses

The PCAs performed to evaluate the robustness of the genomic results consistently identified four genomic groups regardless of the number of SNPs (1887–33,716), individuals (37–72), morphospecies (4–9), or missing data (0.3–52.1%) included in each of the matrices

(Additional file 2: Fig. S1). The PCAs including either all or neutral-only SNP loci (matrices A–E) also recovered four genomic groups, which align with the geography of the central Mexico plateau, but not with a delineation according to the previously recognized morphospecies. The first group (*humboldtianum* sensu stricto group hereafter) clustered samples from Lake Zacapu, Tepuxtepec, and Trinidad Fabela dams. The second group (*estor* group hereafter) included individuals of *C. e. estor*, *C. e. copandaro*, *C. grandocule*, and *C. patzcuaro* from Lakes Pátzcuaro and Zirahuén. The third group (*chapalae* group hereafter) is composed of samples of *C. chapalae*, *C. consocium*, *C. lucius*, and *C. promelas*, from Lake Chapala. Finally, the fourth group (*sphyraena* group hereafter) is formed by *C. sphyraena*, also from Lake Chapala. The first principal component accounted for 24.76–33.51% of the genetic variation and is congruent with the separation of *estor* and *humboldtianum* sensu stricto from *chapalae* and *sphyraena* groups. The second

principal component represented 6.37–7.47% of the genetic variation, resulting in the clear segregation of the *humboldtianum* sensu stricto group (Fig. 2a; Additional file 2: Fig. S2).

While the results of the PCAs estimated using outlier SNP loci did not produce a clear demarcation relative to matrices considering all and neutral-only SNPs, these analyses mostly resulted in the same clustering patterns except in two instances (matrices D and E) where the *chapalae* and *sphyraena* groups clustered together. The first two principal components of the PCAs calculated with outlier SNPs accounted for 20.32–86.58% of the genomic variation explained. Overall, PCAs generated with just outlier (82–370) SNP loci delineated between three and four genomic clusters, superimposing populations from *chapalae* with *sphyraena* groups (Fig. 2a; Additional file 2: Fig. S2).

Scatter plots of the DAPC analyses produced results similar to those based on PCA. The first two discriminant

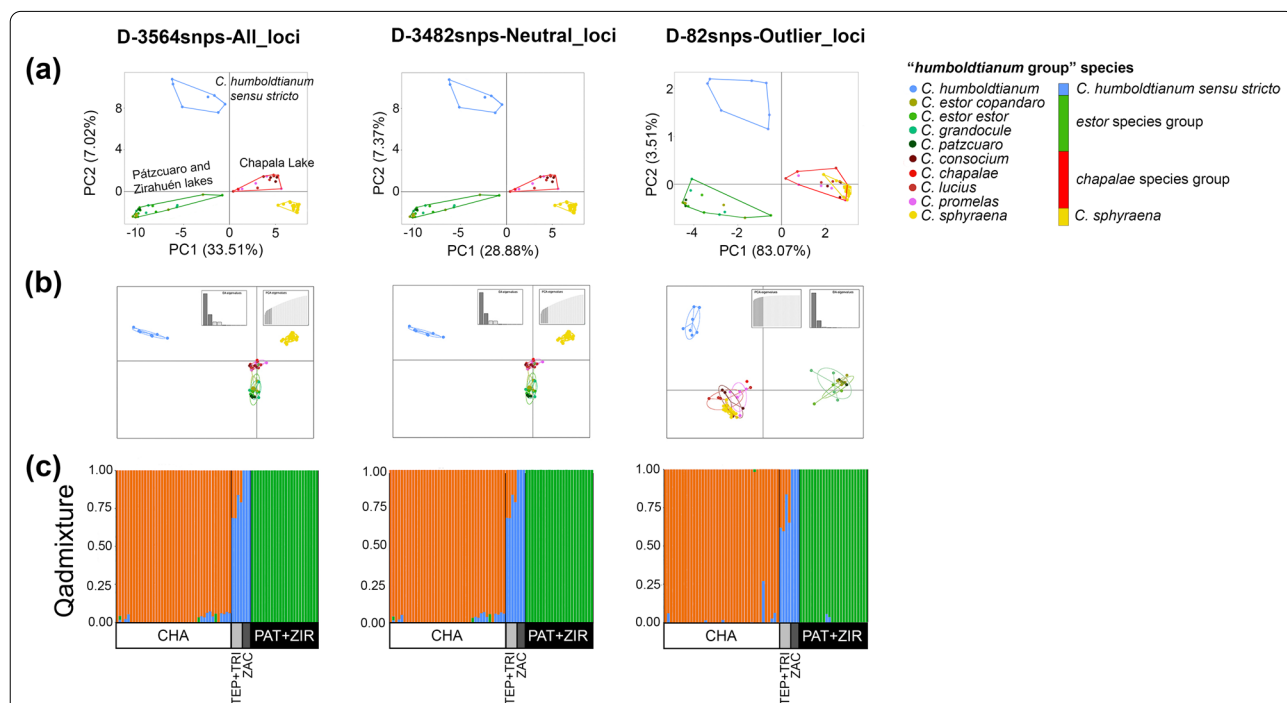


Fig. 2 Multivariate and admixture results based on all, neutral, and outlier SNP loci from the matrix D (3564-snp loci), which explain the highest percentage of explained variation in the analyses. **a** Principal component analyses (PCAs), and **(b)** Discriminant analyses of principal components (DAPCs) consistently recovered four genomic groups ($k=4$) with all and neutral loci that are in agreement with geographic patterns but not with the previously recognized morphospecies: *humboldtianum* sensu stricto group (blue), from Lake Zacapu; *estor* group (green) from Lakes Pátzcuaro and Zirahuén; *chapalae* group (red), from Lake Chapala; and *C. sphyraena* group (yellow), also from Lake Chapala. Outlier loci resolved three groups ($k=3$) where *chapalae* and *sphyraena* groups clustered together. Morphospecies are color-coded according to the genomic groups observed. In PCAs scatterplots the point clustering groups are delimited by convex hulls. In DAPC scatterplots, the point clustering groups are inside their 95% inertia ellipses, and the lines connect points to the mean value for each group. The eigenvalue bar plots are showing in the upper right of each figure. **c** Admixture assignment analyses estimated using all, neutral, and outlier SNPs consistently identified three well-differentiated clusters ($k=3$). Each bar represents the probability of assignment to each cluster. Genomic clusters are color-coded as blue, *humboldtianum* sensu stricto, green, *estor* group; orange, *chapalae-sphyraena* group. CHA, Lake Chapala; TEP, Tepuxtepec Dam; TRI, Trinidad Fabela Dam; PAT, Lake Pátzcuaro; ZIR, Lake Zirahuén

axes identified the same four genomic clusters (BIC: $k=4$) (Fig. 2b; Additional file 2: Figs. S3, S4), retaining 9–12 PCs that represent 48.5–57.7% of the cumulative variance. The DAPC analyses estimated using the outlier loci (matrices C, D and E) also identified three genomic clusters (BIC: $k=3$) where *chapalae* and *sphyraena* form a single group (Fig. 2b; Additional file 2: Figs. S3, S4).

Genomic structure and genomic differentiation analyses

Admixture assignment results of the SNP loci (matrices A–E) considering all, neutral-only, and outlier loci identified three genomic clusters ($k=3$) as the best-fit model (Additional file 2: Fig. S5), each corresponding to the major groups (*humboldtianum* sensu stricto, *estor*, and *chapalae-sphyraena*) obtained with PCA and DAPC analyses using the outlier SNP loci (Fig. 2c; Additional file 2: Fig. S6). As in previous analyses, the admixture results did not delineate morphospecies boundaries.

All individuals of the *humboldtianum* sensu stricto group collected from artificial dams (three from Tepuxtepec and one from Trinidad Fabela) had strong admixture with the *chapalae* group (0.11–0.45), whereas *chapalae* and *estor* groups produced rather low values of shared genetic information (<0.10). While there were higher levels of individual admixture between *humboldtianum* sensu stricto and *chapalae* groups when fewer SNP loci were analyzed, a higher number of outlier SNP loci (99–370) identified more admixture between the three groups than those considering fewer outlier SNPs (82–92) (Fig. 2c; Additional file 2: Fig. S6).

Intra-lake admixture analyses for Lake Chapala identified no evidence of finer genetic structure, as the best-fit model supports a scenario represented by a single metapopulation ($k=1$, Additional file 2: Fig. S7), clustering all individuals of the *sphyraena* and *chapalae* groups. Genetic structure was not observed within the Pátzcuaro-Zirahuén lakes either ($k=1$). By and large, the observed genetic structure based on intra-lake admixture analyses conflict with both morphospecies and ecotype groupings ('peces blancos' vs. 'charales') (Additional file 2: Fig. S7).

Pairwise F_{ST} comparisons were much higher and statistically more significant at inter-lacustrine than at intra-lacustrine levels, reflecting a strong association between genetic structure and geography (Additional file 3: Tables S4–S7). Pairwise F_{ST} values between the nine morphospecies varied from 0.023 to 0.256 and were mostly significant (p values < 0.0012, $\alpha=0.0014$) at inter-lacustrine comparisons, with the exception of *C. chapalae* which showed no significant differentiation with the rest of the morphospecies. By contrast, comparisons among sympatric morphospecies from Lakes Chapala and Pátzcuaro resulted in negative, non-significant values

($F_{ST} = -0.317$ to -0.012 , p values > 0.029), except for *C. sphyraena* which was significantly different than the rest (Additional file 3: Table S4). A greater genetic differentiation was found between the five mitonuclear groups (0.050–0.246), where the majority of the pairwise comparisons were statistically significant (p values < 0.0001, $\alpha=0.005$), except for comparisons between subspecies *C. e. estor* and *C. e. copandaro* (p value = 0.05; Additional file 3: Table S5). F_{ST} values between DAPC clusters were significant and varied from 0.04–0.248 (p values < 0.0001, $\alpha=0.0083$; Additional file 3: Table S6); the three genomic groups detected by the admixture analyses also resulted in significant F_{ST} values (0.120–0.217, p values < 0.0001, $\alpha=0.017$; Additional file 3: Table S7). Finally, the genetic structure between 'peces blancos' and 'charales' ecotypes (-0.005 to 0.001) was not significant in any of the intra-lacustrine comparisons (p values = 0.01–0.027, $\alpha=0.005$; Additional file 3: Table S8).

Phylogenetic analyses

Phylogenetic inferences in a Maximum Likelihood (ML) framework using IQ-TREE based on matrices A, C, D and E (all and neutral) SNP loci resolved three highly supported clades (ultrafast bootstrap or UFBoot = 90–99%; Figs. 3a and Additional file 3: Fig. S8), which are largely in agreement with the results from admixture analyses. Clade I clustered species within *chapalae* and *sphyraena* groups from Lake Chapala (UFBoot = 83–95%); clade II is formed by the *humboldtianum* sensu stricto group from Lake Zacapu and dams (UFBoot = 73–100%); and clade III is represented by the *estor* group from Lakes Pátzcuaro and Zirahuén (UFBoot = 100%). Samples of *C. sphyraena* and the subspecies *C. e. copandaro* were resolved as monophyletic within clade I (UFBoot = 80–100%) and clade III (UFBoot = 94–97%), respectively. While the phylogenetic trees obtained with outlier loci datasets had less resolution than those using the complete matrix, these different analyses resolved the reciprocal monophyly of Clade I and Clades II–III, except for matrix A where Clade I was paraphyletic. Also, the relationship between Clade II and III was more problematic: in some cases these clades were reciprocally monophyletic (matrices A and C), but in other cases they were paraphyletic (databases B, D, and E) (Additional file 2: Fig. S8).

Tree inference based on mitochondrial data failed to resolve the groups identified with our genome-wide RADseq analyses, suggesting mitonuclear discordance (Fig. 3c; Additional file 1: Fig. S9). However, in agreement with the ddRADseq phylogenies, analyses of the mitochondrial genealogies and a phylogeny estimated using the concatenated matrix did not delineate any clear monophyletic groups according to morphospecies boundaries. Although some groupings comprised

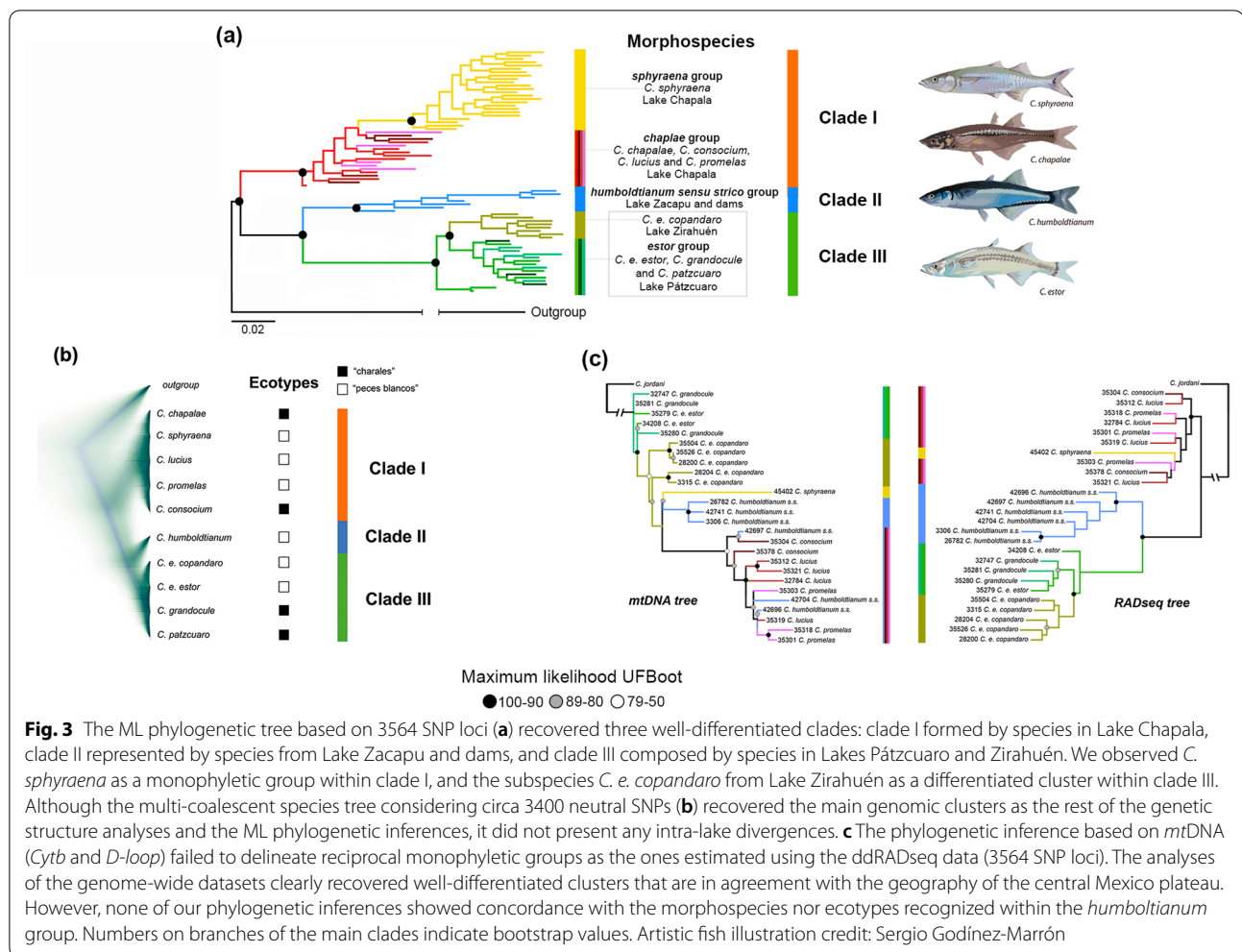


Fig. 3 The ML phylogenetic tree based on 3564 SNP loci (a) recovered three well-differentiated clades: clade I formed by species in Lake Chapala, clade II represented by species from Lake Zacapu and dams, and clade III composed by species in Lakes Pátzcuaro and Zirahuén. We observed *C. sphyraena* as a monophyletic group within clade I, and the subspecies *C. e. copandaro* from Lake Zirahuén as a differentiated cluster within clade III. Although the multi-coalescent species tree considering circa 3400 neutral SNPs (b) recovered the main genomic clusters as the rest of the genetic structure analyses and the ML phylogenetic inferences, it did not present any intra-lake divergences. c The phylogenetic inference based on *mtDNA* (*Cytb* and *D-loop*) failed to delineate reciprocal monophyletic groups as the ones estimated using the ddRADseq data (3564 SNP loci). The analyses of the genome-wide datasets clearly recovered well-differentiated clusters that are in agreement with the geography of the central Mexico plateau. However, none of our phylogenetic inferences showed concordance with the morphospecies nor ecotypes recognized within the *humboldtianum* group. Numbers on branches of the main clades indicate bootstrap values. Artistic fish illustration credit: Sergio Godínez-Marrón

of neighboring localities were resolved with the *mtDNA* trees, resulting clades from these trees are not cohesively clustered by geography. The *Cytb* genealogy resolved the monophyly of *C. sphyraena*, a result that is concordant with the RADseq phylogeny and previous studies [11]. However, these relationships were not resolved with confidence (UFBoot < 95%). Overall, none of the phylogenetic hypotheses were aligned with the morphospecies or ecotypes proposed within the *humboldtianum* group (Fig. 3; Additional file 2: Fig. S9).

Multispecies coalescent analyses

The species tree estimated with SNAPP also evidenced three geographically cohesive clades (I, II, and III; Fig. 3b). Coalescent-based species delimitation analyses using the BFD* method selected species delimitation scenarios of four and five species as the top-ranked models with the highest MLE values (MLE = -30,568.17 and -21,025.86 to -11,910.25, respectively; Table 1). The main difference of the models of four and five species

relies on the subspecies *C. e. copandaro* from Lake Zirahuén forming a different group from *C. e. estor* and the sympatric species in Lake Pátzcuaro. Bayes factors analyses supported models of four (subset 1 and 3) and five species (subset 2), rejecting the nine species that reflects the current taxonomy (BFs = 295.56–850.28; Table 1) and also alternative models of three species (BFs = 158.5–660.4; Table 1).

Genetic diversity and effective population size (N_e) analyses

The genomic diversity of the *humboldtianum* group is summarized in Table 2. The observed and expected heterozygosity (H_o and H_e) across different lineages was 0.097–0.136 and 0.084–0.139, respectively. All lineages distributed within Lake Chapala revealed higher genetic diversity (H_o = 0.123–0.138) than those in the Lakes Zacapu and Pátzcuaro-Zirahuén (H_o = 0.100–0.126) (Table 2). The values of N_e estimator (*HetExcessNe* and *CoancestryNe*) across different lineages were 7.2–317

Table 1 Results of bayes factor delimitation (BFD*) analyses for the *humboldtianum* group using three SNPs subsets (ranging from 39 to 59 individuals, and 411–1102 SNP loci)

Model	Species number	Subset 1: 39ind_411 SNP loci		Subset 2: 39ind_1102 SNP loci		Subset 3: 59ind_548 SNP loci	
		MLE	BF	MLE	BF	MLE	BF
Current morpho-species	9	– 12,058.03	295.56	– 30,993.31	850.28	– 21,279.57	507.42
Mitochondrial groups	5	– 11,910.25	NA	– 30,600.16	63.98	– 21,025.86	NA
DAPC groups	4	– 11,927.68	34.85	– 30,568.17	NA	– 21,161.43	271.14
Admixture groups	3	– 11,989.52	158.53	– 30,898.37	660.41	– 21,278.3	504.88

Bayes factor (BF) calculations are estimated against the model with the best marginal-likelihood estimate (Model 1). Positive BF values indicate support for model 1 (the best model). Overall, these analyses support a species delimitation scenario between four and five species, rejecting the nine morphospecies scheme

MLE: marginal likelihood estimates. Bold values correspond to the best model selected by the analysis. NA correspond to not applied, and its associated marginal likelihood is in bold. $0 < BF < 2$, non-significant; $2 < BF < 6$, positive evidence; $6 < BF < 10$, strong support; $BF > 10$, decisive support

and 2.2–31, respectively (Table 2). Values of *LDNe* estimator were infinite with the exception of *chapalae-sphyraena* group (86.9), while *C. humboldtianum* sensu stricto produced infinite values with both *LDNe* and *HetExcessNe* estimators (Table 2). These infinite values indicate the effect of sampling error.

FST loci outlier analysis

A total of 410 outlier loci (1.1% of the total analyzed, Fig. S10) were compared to GenBank entries using BLAST-n. In summary, 116 sequences did not have a match, while the remaining 294 (71.7%) loci matched fish sequences (E-value = $4E-42$ to $2E-02$), including protein-coding genes (26%) (Additional file 3: Table S9). Out of these coding regions, we identified genes related to a wide array of biological functions such as genes implicated in immune responses (*kpna3*: *Nothobranchius furzeri*, GenBank accession number XM_015966110.1; *bcl11b*: *Cheilinus undulatus*, XM_041804421.1; *trim59*: *Megalops cyprinoides*, XM_036537431.1; *NLR family CARD domain-containing protein 3-like*: *Acanthopogon latus*, XM_037091148.1), sensory systems (*or132-1*: *Danio rerio*, DQ306116.1; *sws2a* and *rh2-1*: *Lucania goodei*, MT850055.1; and *lws2*: *Monopterus albus*, XM_020592984.1), growth (*cand1*: *Poecilia formosa*, XM_007569621.2; and *sgta*: *Archocentrus centrarchus*, XM_030727116.1), and skeletal-muscle system (*neb*: *Gymnodraco acuticeps*, XM_034221710.1; and *ttn*: *Fundulus heteroclitus*, XM_036135404.1) (Additional file 3: Table S9).

Discussion

In this study, we used ~1800 to ~33,000 SNPs to test the morphological- (nine nominal species) and *mtDNA*/*nDNA*-based (five mitochondrial groups) hypotheses to elucidate the number and boundaries of species in the *humboldtianum* group. Our results consistently

identified four independently evolving lineages organized in three well-differentiated clades: *chapalae-sphyraena* in Lake Chapala, *humboldtianum* sensu stricto in Lake Zacapu and dams, and *estor* in Lakes Pátzcuaro and Zirahuén. The genomic clusters obtained were not in agreement with the morphospecies delimitation but rather with geography reflecting ancient isolation events in central Mexico. This scenario suggests that the geologic history of the Lerma-Chapala hydrologic system has played a major role in driving divergence in this species complex. Our analyses also revealed an intra-lake cladogenetic event where *C. sphyraena* ('pez blanco') is distinguished from its sympatric counterparts ('peces blancos' and 'charales') in Lake Chapala. Although body size in *Chirostoma* species in Lake Chapala has been suggested as a promoter of ecological niche partitioning for this species complex [10, 21], we did not find evidence of genetic structure related to the 'peces blancos' and 'charales' ecotypes.

Herein, the use of genome-wide data provided an unprecedented resolution that had not been achieved using a scant number of genetic markers to test species delimitation scenarios within the *humboldtianum* group [11, 18]. While species delimitation studies examining thousands of genetic loci often unveil cryptic diversity [22–25], our study represents one of the few cases where the use of genome-wide SNP data and MSC approaches provide evidence of taxonomic over-splitting [2, 5, 26]. Recently, it has been recognized that the MSC model can potentially confound population structure with species boundaries—particularly when major sampling gaps near the distribution range exist—leading to an over-estimation of the number of species, a bias compounded by the high statistical power that often results from genome-wide analyses [4–6]. However, we account for these caveats by assessing our results in a framework that examines all nominal species in the *humboldtianum*

Table 2 Genomic diversity of 3482 neutral SNPs loci estimates for the *humboldtianum* group, under each species delimitation hypothesis examined in this study (i–iv)

	N	HO (SD)	HE (SD)	F (SD)	LDNe (95% CIs)	HetExcessNe (95% CIs)	CoancestryNe (95% CIs)
9 morphospecies (i; Barbour, 1973)							
<i>C. chapalae</i>	1	–	–	–	–	–	–
<i>C. consocium</i>	4	0.138 (0.004)	0.110 (0.003)	– 0.225 (0.008)	Infinite (infinite–infinite)	8.9 (6.4–14.9)	10.9 (8.1)
<i>C. lucius</i>	7	0.134 (0.004)	0.115 (0.003)	– 0.144 (0.007)	Infinite (infinite–infinite)	12.6 (8.4–25.6)	6 (4.9)
<i>C. promelas</i>	5	0.123 (0.004)	0.099 (0.003)	– 0.222 (0.009)	Infinite (infinite–infinite)	Infinite (20.1–infinite)	31 (19.3)
<i>C. sphyraena</i>	24	0.136 (0.004)	0.121 (0.003)	– 0.083 (0.005)	Infinite (infinite–infinite)	8.9 (7.2–11.6)	2.2 (1.9)
<i>C. humboldtianum</i>	7	0.126 (0.004)	0.117 (0.003)	– 0.062 (0.007)	Infinite (infinite–infinite)	Infinite (108.1–infinite)	3 (2.7)
<i>C. estor</i>	15	0.103 (0.004)	0.096 (0.003)	– 0.041 (0.007)	Infinite (infinite–infinite)	244 (19.9–infinite)	3.6 (3.2)
<i>C. grandocule</i>	6	0.100 (0.004)	0.084 (0.003)	– 0.170 (0.008)	Infinite (infinite–infinite)	10.6 (6.9–24.6)	7.2 (5.4)
<i>C. patzcuaro</i>	3	0.112 (0.004)	0.085 (0.003)	– 0.301 (0.009)	Infinite (infinite–infinite)	7.2 (5.3–11.8)	Infinite (infinite)
5 mitonuclear groups (ii, Betancourt-Resendes et al. 2020)							
<i>C. chapalae</i>	17	0.130 (0.003)	0.139 (0.003)	0.079 (0.007)	Infinite (infinite–infinite)	24.7 (13.5–179.5)	4.5 (3.9)
<i>C. sphyraena</i>	24	0.136 (0.004)	0.121 (0.003)	– 0.083 (0.005)	Infinite (infinite–infinite)	8.9 (7.2–11.6)	2.2 (1.9)
<i>C. humboldtianum</i>	7	0.126 (0.004)	0.117 (0.003)	– 0.062 (0.007)	Infinite (infinite–infinite)	Infinite (108.1–infinite)	3 (2.7)
<i>C. e. estor</i>	16	0.103 (0.004)	0.096 (0.003)	– 0.037 (0.007)	Infinite (infinite–infinite)	10.8 (7.6–18.9)	4.4 (3.6)
<i>C. e. copandaro</i>	8	0.097 (0.004)	0.105 (0.003)	0.082 (0.010)	Infinite (infinite–infinite)	8.9 (6.2–16.4)	5.4 (4.2)
4 multivariate DAPC clusters (iii, This study)							
<i>C. chapalae</i> group ^a	17	0.130 (0.003)	0.139 (0.002)	0.079 (0.007)	Infinite (infinite–infinite)	24.7 (13.5–179.6)	4.5 (3.9)
<i>C. sphyraena</i>	24	0.136 (0.004)	0.121 (0.003)	– 0.083 (0.005)	Infinite (infinite–infinite)	8.9 (7.2–11.6)	2.2 (1.9)
<i>C. humboldtianum</i> sensu stricto	7	0.126 (0.003)	0.117 (0.003)	– 0.062 (0.007)	Infinite (infinite–infinite)	Infinite (108.1–infinite)	3 (2.7)
<i>C. estor</i> group ^b	24	0.101 (0.003)	0.104 (0.003)	0.103 (0.008)	Infinite (infinite–infinite)	106.4 (20.2–infinite)	4.4 (3.8)
Admixture clusters K=3 (iv, This study)							
<i>C. chapalae-sphyraena</i> groups	41	0.134 (0.003)	0.138 (0.003)	0.085 (0.006)	86.9 (84.8–89.1)	317 (40–infinite)	4.2 (3.8)
<i>C. humboldtianum</i> sensu stricto group	7	0.126 (0.003)	0.117 (0.003)	– 0.062 (0.007)	Infinite (infinite–infinite)	Infinite (108.1–infinite)	3 (2.7)
<i>C. estor</i> group	24	0.101 (0.003)	0.104 (0.003)	0.103 (0.008)	Infinite (infinite–infinite)	106.4 (20.2–infinite)	4.4 (3.8)

N number of individuals analyzed per group, observed (H_O) and expected (H_E) heterozygosity, fixation index coefficient (F), linkage disequilibrium effective (LDNe), heterozygote-excess (HetExcessNe), and molecular coancestry (CoancestryNe) population size estimators. SD = standard deviation; CI = confidence intervals. The grey shading in the cells indicates the highest values of genetic diversity

^a *C. chapalae* group conformed by *C. chapalae*, *C. consocium*, *C. lucius*, and *C. promelas*

^b *C. estor* group represented by *C. e. estor*, *C. e. copandaro*, *C. grandocule*, and *C. patzcuaro*

group from across their distribution ranges in the Lerma-Chapala hydrologic system, and by taking into account morphological and ecological lines of evidence. All in all,

we provide a robust analysis to assess the number of species within the *humboldtianum* group and also to examine the evolutionary history of the group.

Inter-lacustrine divergences

Our results suggest that vicariance events during the Pleistocene influenced the early divergence within the *humboldtianum* group. Speciation within this group appear to be strongly related to the complex geologic history of volcanism, tectonism, and climatic events that promoted the connection and disconnection of the Lerma-Chapala hydrological system, and surrounding tributaries—including several lakes and paleolakes such as Chapala, Cuitzeo, Zacapu, Pátzcuaro, and Zirahuén [10].

Although our phylogenetic results are somewhat incongruent with those previously estimated by Betancourt-Resendes et al. [11] (*chapalae*, *sphyraena*, *estor*, and *humboldtianum* sensu stricto groups), their divergence time analyses provide a rough estimate of the timing of cladogenesis in the group, placing the origin of independent evolutionary lineages within the *humboldtianum* group during the Pleistocene < 1 Ma (0.58–0.13 Ma). There are two main biogeographic processes that are synchronous and congruent with the observed genetic patterns recovered in this study, suggesting that these events played an important role as the main drivers of diversification in the *humboldtianum* group. First, the allopatric fragmentation of clade I (Lake Chapala) from II–III (Lake Zacapu–Lakes Pátzcuaro and Zirahuén) may correspond to a biogeographical barrier promoted by the geologic activity of the Penjamillo Graben and the formation of the Tarascan corridor, which started much earlier during the Late Miocene–Early Pliocene [27, 28]. These geological events separated the ancient corridors between the paleo Lerma-Chapala system and the Cuitzeo paleolake plus adjacent tributaries that connected Lakes Zacapu, Pátzcuaro, and Zirahuén [10, 29]. The separation of these hydrologic regions experienced its highest peak during the Early Pleistocene [29], where severe climatic fluctuations promoted a series of connection and disconnection events of small water bodies, with the last dry episode starting *ca.* 0.12 Ma [10, 29]. The second biogeographic event is the fragmentation of the ancestral Villa Morelos and Chucandiro-Huaniqueo corridors, also during the Pleistocene. This episode, promoted by the geologic activity of the Northeast-Southwest fault system *ca.* 0.7–0.5 Ma, separating Lake Zacapu from Lakes Cuitzeo–Pátzcuaro–Zirahuén [30, 31], and correlating with the genetic patterns of divergence between clade II (Lake Zacapu) and III (Lakes Pátzcuaro and Zirahuén).

The same cladogenetic patterns, in agreement with the aforementioned biogeographic events, have been observed in goodeid freshwater species endemic to central Mexico, including the divergence between the sister species *Skiffia multipunctata* and *S. lermiae* and the split of the *Allotoca diazi* complex from *A. zacapuensis*,

although these diversification events do not appear to have occurred in synchrony [32–35].

Intra-lacustrine divergences

Our genome-wide data revealed the presence of two genetic clusters in Lake Chapala (*sphyraena* and *chapalae* groups), as suggested by Betancourt-Resendes et al. [11] using a few *mtDNA* and *nDNA* sequences. The evolutionary processes segregating these populations seem to be related to ecological speciation—divergent specializations promoted by ecological opportunity following reproductive isolation [36, 37]—a major driver of sympatric evolution in lakes [36, 38, 39]. The most iconic examples of ecological speciation are depicted by South American [40–42] and African cichlids [43, 44], and sticklebacks [45, 46], where patterns of morphological divergence are associated with trophic partitioning. For example, preference for soft mobile (e.g., copepods) compared to hard sessile preys (e.g., gastropods) can lead to disruptive selection on skull morphology and body shape [47], where body size can be crucial to succeed in the related foraging mode (e.g., benthic *vs.* limnetic in *Gasterosteus* spp; [48, 49]).

In the case of species of the *humboldtianum* group occurring within the Lake Chapala (*C. sphyraena*, *C. promelas*, *C. consocium*, and *C. lucius*), several hypotheses of ecological speciation suggest the coexistence of ecotypes in agreement with a differentiation of morphological traits (e.g., jaw shape, head length, oral gape, and gill raker structure) related to feeding habits [10, 50], or an ecological partition that correlates with species body sizes (larger ‘peces blancos’ *vs.* smaller ‘charales’) [10, 16, 21]. However, no strong evidence showing clear patterns of differentiation between morphospecies or ecotypes has been documented to date [21].

Herein, we evaluated the hypothesis by Barbour [10] that differentiation in body length would promote niche partitioning, allowing the co-occurrence of different species that feed on distinct prey sizes. Although signals of trophic specialization separating ‘peces blancos’ and ‘charales’ have been documented by Mercado-Silva et al. [21] for three species in Lake Chapala, no analysis conducted here (PCAs, DAPCs, inter- and intra-lake admixture, and phylogenetic trees; Figs. 2 and 3; Additional file 2: Figs. S1–S9) provide support for a scenario of diversification related to body size. Nonetheless, our multivariate and F_{ST} analyses (Fig. 2a, b, Additional file 2: Figs. S1, S2, S4; Additional file 3: Tables S4–S6) clearly demonstrate that the *sphyraena* group—the only entity previously resolved as monophyletic based on a scant number of mitochondrial and nuclear markers [11]—represents a separate genetic cluster with almost no gene flow shared with other members in the *chapalae* group, although in

our intra and inter-lake admixture result did not find evidence of *sphyraena* and *chapalae* groups genetic separation (see “Discussion” below). Based on this evidence, and considering the fact that *C. sphyraena* is the most taxonomically differentiated nominal species (particularly on traits related to trophic specialization; [10]), we hypothesize that *C. sphyraena* is in its early stages of ecological speciation. We note, however, that species-specific habitats are unknown for most of the *Chirostoma* species [21], and thus further ecological studies are necessary to better understand the evolutionary history of the *chapalae* and *sphyraena* groups.

Species delimitation and taxonomic implications

The BFD* species delimitation analyses provided strong support for four- and five-species delimitation scenarios (Table 1), while the coalescent-based species tree identified three major monophyletic groups (Fig. 3b). Recent simulation studies that evaluated the efficiency of MSC methods suggest that this model tends to confound population structure with species boundaries [4–6]. This becomes particularly problematic in recently-diverged and closely-related species, such as the *humboldtianum* group, where processes promoting differentiation lie at the intersection between population structure and species divergence [4], generating gene trees with short branches and with multi coalescent histories that make species tree inference challenging [51, 52]. This is not the case in the *humboldtianum* group, a recent species complex that diverged less than 1 Ma [11], where the genetic structure detected fewer species than those recognized by taxonomic studies.

The finer genetic structure recovered for the *sphyraena* group and the sub-species *C. e. copandaro* favor models of four or five species over a three-species scheme. In this scenario, under a strict reciprocal monophyly criterion, none of the *C. sphyraena* or the sub-species *C. e. copandaro* lineages would represent a species (under the phylogenetic species concept, PSC; [1]). However, our intra-lake admixture analyses suggest that there is no genetic structure among *sphyraena* and *chapalae* groups. We conducted intra-lake admixture analyses to evaluate whether including the rest of the species in the *humboldtianum* group may be potentially masking finer genetic structure within each lake. Our results demonstrate that the best-fit model for each lake is represented by a single population (Additional file 2: Fig. S7). These results, combined with the admixture analyses based on all individuals (matrices A–E; Figs. 2c; Additional file 2: Fig. S6) and phylogenetic trees (Fig. 3; Additional file 2: Fig. S8) suggest that *sphyraena* may be an incipient species given its nested phylogenetic position that otherwise renders

chapalae sensu stricto as paraphyletic (see also multivariate analyses in Fig. 2a, b, Additional file 2: Figs. S2, S4; and pairwise F_{ST} analyses in Additional file 3: Tables S4–S7). Until further evidence becomes available (e.g., ecological, ethological), the *sphyraena* group and the sub-species of *C. e. copandaro* should not be considered species per se as their ancestral polymorphisms have not been fully sorted by genetic drift. Thus, by taking into account multiple lines of evidence (population genomics, phylogenomics, morphology, biogeography, and ecology), we propose that *estor*, *humboldtianum*, and *chapalae* groups constitute three well-differentiated species (*C. estor*, *C. humboldtianum* sensu stricto, and *C. chapalae*, respectively).

The observed discordance between our genomic analyses, which resolved individuals from each lake as monophyletic, and previous studies [11, 18] where clades are not so cohesively clustered by geography, suggests that few mitochondrial and nuclear markers do not have sufficient statistical power to resolve the phylogenetic relationships of this recently diverged group. Here, we show that the use of thousands of SNP loci collectively provide strong power to detect phylogenetic signal while reducing the probability of stochastic error [53], as has also been demonstrated in several species of freshwater [46, 54–57] and marine [58–61] fishes. We observe a discordance between the phylogenetic position of the three main clades, where the *mtDNA* tree places individuals from the *estor* group as early branches while the ddRADseq analyses place them in a more nested position. Such disagreement could be related to the characteristics of the type of markers (e.g., matrilineal inheritance). Mitonuclear discordance can lead to an inaccurate estimation of the evolutionary history of species, ultimately misleading species delimitations [62, 63, 64]. Unlike the *mtDNA* trees, the ddRADseq analyses align with the geography and geology of central Mexico, identifying geographically cohesive clades.

Finally, in all cases, the nine-species model was strongly rejected thus refuting the morphological-based hypothesis sensu Barbour [10]. Our study represents a rare case where genome-wide data evidences an over estimation of species diversity based on morphological characters. The delineation of such morphospecies was based on several characters, particularly head and body traits, that have been considered as the basis for the current taxonomy in the genus. However, trait measurements such as jaw length, jaw shape, teeth size, and body shape are subjected to great environmental plasticity related to the species’ trophic ecology and habitat characteristics, making it difficult to find diagnostic characters among *Chirostoma* species [10, 19, 65].

Selection across lakes

The study of alleles involved in local adaptation can unveil loci responsible for adaptive differences among populations. Our analyses of outlier loci detected 294 putative loci under selection, of which at least 106 are related to important biological processes (Additional file 3: Table S4). For example, we detected two SNPs associated with immune response loci (*kpna3* and *bcl11b* genes). Previous studies in trout [67] and sticklebacks [68–70] have highlighted the importance of these loci in the response to different pathogens. Thus, it is possible that the selection of loci associated with the immune response could be related to exposure to lake-specific pathogens during the colonization of new habitats, promoting the local adaptation and divergence of the *humboldtianum* group once geographic isolation started.

Other genes detected in this analysis were the *or132-1*, which is an odor receptor associated with the detection of food in the aquatic system [71], and *sws2a*, *lws2a*, and *rh2-1* associated with vision. These findings could be related to the photic environment of each lake: whereas Lakes Chapala and Pátzcuaro are shallow eutrophic water bodies with a high sediment charge of turbid water [72, 73], Lake Zacapu is a clear-water lake where the light reaches an average depth of 4.3 m (up to 11.5 m in some places; [13]). Differences in photic conditions can affect the planktonic community within each lake, promoting differential selection as related to olfactory and visual receptors [74]. Finally, loci associated with the skeleton-muscle apparatus, such as nebulin (*neb*) and titin (*ttn*), and genes involved in growth hormone regulation, such as cullin-associated NEDD8-dissociated protein 1 (*cand1*), and a small glutamine-rich tetratricopeptide repeated containing alpha (*sgta*), could also be involved in the phenotypic changes associated with body size.

We note that these inferences need to be taken with caution as there is a limited performance of F_{ST} outlier approaches in non-model organisms to identify candidate genes without a reference genome [75], particularly if the demographic history of the species is not modeled accurately [76]. Reduced representation genomic libraries that use restriction enzymes to cut the DNA may fail to identify key loci as these techniques only capture a small portion of the genome [77, 78], while many loci are lost due to low coverage and filtering [66, 79]. Without a reference genome, these results are highly sensitive to false positives, such as loci linked to sites experiencing purifying selection that can be confounded with processes of local adaptation [66]. Thus, the set of candidate genes identified in this study provides a starting point for further targeted research into underlying genetic bases of selection in the *humboldtianum* group.

Genomic diversity, fishery management and conservation

Our genome-wide SNP analyses revealed that genomic diversity within the *C. humboldtianum* species complex is low (Table 2), a pattern that has been also reported in other lacustrine fish species such as the Nile tilapia [80] and sticklebacks [69]. The low heterozygosity in freshwater fish species may be related to smaller effective population sizes and varying demographic histories involving bottlenecks during the recent colonization of new freshwater habitats [69]. The *humboldtianum* group diverged and colonized the lakes of the central Mexico plateau during recent evolutionary times, between 76 and 580 thousands of years [11]. Such sudden demographic expansion from a smaller number of founders [81] may explain their low genetic diversity. In addition, these fishes have recently decreased demographically due to over-exploitation by commercial fisheries, which could have also influenced their genomic diversity [82]. Although the majority of the results of *LDNe* and *HetExcessNe* estimators produced infinite values indicating the effect of sampling error caused by the small sample size (1–24) [83, 84], the *Ne* values observed for the *humboldtianum* group were similar to or even smaller than those observed in other lacustrine fish species (e.g., *Amphilophus labiatus* and *A. citrinellus* [85]), or species that have experienced a population collapse (e.g., *Oscorhynchus nerka* [86]).

Although we have no evidence of the presence of hybrid individuals in natural lakes, as has been previously reported [10, 87], given the frequent translocation events promoted by aquaculture policies since 1970 [88, 89], we detected some hybrid individuals of *C. humboldtianum* sensu stricto in artificial dams that are genomically closer to the clade I. These hybrids could be the result of deliberate introductions of different species of the “*humboldtianum* group” into several artificial dams of the region [8, 81, 88], which were done to promote artisanal fisheries of this important resource and to improve the local economy. Future studies should evaluate the effectiveness of the introductions in the local fisheries and their impact on natural populations within the *humboldtianum* group. Some hybrid individuals may show hybrid vigor and could become better competitors than local species, making this practice another factor affecting natural populations [8].

Chirostoma species represent a highly important economic and cultural resource since pre-Hispanic times [8]. However, species in this genus have been heavily overfished, leading to the collapse of several populations and severe conservation problems where some species are now considered extinct or in danger of extinction across several locations [8, 9]. Currently, the *humboldtianum* group is threatened by several factors, particularly

habitat loss, pollution, the introduction of exotic species, and overfishing [50]. The decrease of silverside fish populations has also caused the collapse of their fisheries, negatively impacting the local fishermen's communities [8]. Another consequence of the over-exploitation of these fishes is the decrease in the capture size and the age of maturity size [8, 9]. Thus, the delimitation of operational genomic units is critical for fisheries management and conservation plans [8]. Additionally, information on the genomic structure and genetic diversity within and between natural populations of the *humboldtianum* group are crucial to understanding their ability to cope with environmental changes over evolutionary time [90]. Our results support the presence of four genomic groups within the *humboldtianum* group, distributed in the Lakes Chapala, Pátzcuaro-Zirahuén, and Zacapu. We strongly recommend revising management and conservation plans taking into consideration our proposed species boundaries.

Conclusions

Our genome-wide analyses based on ~2 K to ~33 K SNP loci, examined in light of morphological and ecological lines of evidence, provides remarkable resolution to address a convoluted case of species delimitation within the *humboldtianum* group, a result not previously achieved with the use of a small number of *mtDNA* and *nDNA* markers. We resolved four genomic clusters arranged into three geographically cohesive clades (clade I, *chapalae*, and *sphyraena* groups from Lake Chapala; clade II, *humboldtianum* sensu stricto group from Lake Zacapu and dams; and clade III, *estor* group from Lakes Pátzcuaro and Zirahuén). These groups conflict with the previously described morphospecies, and also with the 'peces blancos' and 'charales' ecotypes. Notably, our genomic analyses reject both morphology- and ecotype-based delineations.

Our results suggest that the main cladogenetic events that gave rise to the three clades within the *humboldtianum* group resulted from allopatric processes generated by the complex geologic history of the Lerma-Chapala paleo system, while intra-lake divergence of the *sphyraena* group could be the product of ecological and incipient speciation. It is important to highlight that lumping the nine morphospecies into three does not imply reducing conservation efforts but enforcing the inclusion of molecular information to create management strategies and conservation plans. Critically, the low levels of genetic diversity and *Ne* values observed inside each genomic cluster should be considered in any further conservation efforts. All in all, our study represents a rare case where the use of genome-wide data evidence taxonomic over-splitting based on morphological

information, while it emphasizes the use of approaches that take into account multiple lines of evidence to address complex species delimitation scenarios.

Methods

Sample collections

Our research is a follow-up study of a recently published phylogeographic analysis of the *humboldtianum* group based on two mitochondrial genes (cytochrome b [*Cytb*], and a fragment of the hypervariable control region [*D-loop*]), and one nuclear locus (first intron of the S7 ribosomal protein gene [*S7*]) [11]. We carefully selected 77 individuals representing the nine nominal species of the *humboldtianum* group and the genetic diversity observed by Betancourt-Resendes et al. [11] to build a genomic dataset. We added two individuals of *Chirostoma jordani*, and one of *Chirostoma attenuatum* as outgroups. We collected the samples from local fishermen bycatch in 2014–2018, following the ethical capture methods and regulations approved by the Official Mexican Norm NOM-032-SAG/PESC-2015 and NOM-036-SAG/PESC-2015 for fishing in the lakes of central Mexico. Fishes were euthanized by applying rapid chilling protocols that consisted in placing them in ice water-cold baths (2–4 °C), as recommended by the American Veterinary Medical Association (AVMA; <https://www.avma.org/resources-tools/avma-policies/avma-guide-lines-euthanasia-animals>). Voucher specimens were fixed in 10% formaldehyde and preserved in 70% ethanol. Our sampling spans six localities, which together cover the distribution range of the *humboldtianum* group in the central Mexico plateau (Additional file 3: Table S1; Fig. 1). We sampled fin clips, preserved them in 95% ethanol, and stored them at –76 °C. We deposited tissue and voucher at the fish collection of the Universidad Michoacana de San Nicolás de Hidalgo (UMSNH), Mexico (Additional file 1: Appendix 1). None of our experiments included living organisms.

Molecular protocols, de novo assembly, and SNP genotyping

We extracted high-molecular-weight DNA using the Qia-gen DNeasy Blood and Tissue kit (Qiagen, Inc.) following the manufacturer's protocol. We prepared the ddRAD sequencing libraries at the Sequencing and Genotyping Facility (SGF) at the University of Puerto Rico-Río Piedras (UPR-RP), following the protocol of Peterson et al. [91]. We used the *PstI* and *MseI* restriction enzymes with a size selection window of 300–600 bp. We sequenced ddRADseq libraries in two lanes of Illumina HiSeq 4000 PE 100 bp at the Knapp Center of Biomedical Discovery (KCBD) Genomics Facility, University of Chicago.

We verified the quality of the raw reads using the software FastQC v0.11.5 (<http://www.bioinformatics.babraham.ac.uk/projects/fastqc/>). We demultiplexed the sequenced libraries and removed the restriction sites using the `process_radtags.pl` script implemented in Stacks v2.4 [92, 93]. We trimmed all demultiplexed reads to 86 bp after removal of the restriction site overhangs. We then applied a second quality filter using a Phred score of 33, producing a total of 3.5×10^8 reads that we retained. We deposited all demultiplexed raw reads in GenBank (NCBI) (accession no. SAMN26725252–SAMN26725331, SRA BioProject PRJNA816865).

As there is no *Chirostoma* genome that we could use as a reference, we conducted a de novo assembly of putative loci using Stacks. To select the assembly parameters that best-fit the data, we first performed an exploratory analysis using default parameters. We then used a subset of 15 samples, formed by the individuals of each nominal species with the highest coverage to conduct a second assembly by applying different combinations of parameters as reported by Mastretta-Yanes et al. [94], Paris et al. [95], and Pedraza-Marrón et al. [96]. Details of the protocol for the de novo assembly are given in the Supplementary Material (SM, Additional file 2: Figs. S11–S13). We conducted the final locus assembly using a minimum of five raw reads required to form a stack ($m=5$), with a maximum of four mismatches between stacks ($M=4$), and five mismatches between loci of different individuals ($n=5$).

SNP filtering and database selection

SNP filtering parameters play an important role in the number of recovered loci and the inferred degree of genetic differentiation [97, 98]. We thus implemented a series of exhaustive steps to select the final SNP loci that were included in further analyses (Additional file 2: Fig. S14). (*Step 1*) We filtered biallelic loci according to the number of individuals, populations, and nominal species (Additional file 3: Table S2). We selected four databases that ranged between ~ 1000 and $\sim 105,000$ SNPs to apply further filters. (*Step 2*) We removed low-frequency alleles and potential paralogous loci using a minor allele frequency (*maf*) of 0.01 and 0.05. (*Step 3*) We selected sites with different tolerance thresholds for missing data (0.05, 0.25, 0.50, and 0.75). (*Step 4*) We removed individuals with various cutoffs of missing data (0.05–0.99). Collectively, these filters produced 24 additional datasets, with 150–33,800 SNP loci, 0.3–49% of missing data, and 37–80 individuals representing 4–9 morphospecies. (*Step 5*, Additional file 3: Table S3). We then selected 19 (out of the 24) matrices to assess the robustness of our analyses to differences in numbers of individuals (37–72), nominal species (4–9), SNPs (~ 2 k– ~ 33 k), and

proportions of missing data (0.3–49%). The matrices generated during this step were further used to evaluate the sensitivity of the analyses to the exclusion of nominal species with high levels of missing data (see SM). For more exhaustive analyses, we also selected five (out of the 24) datasets with a final set of 72 individuals representing the nine morphospecies: A-33716snps (33,716 SNPs, 15.7% missing data), B-10517snps (10,517 SNPs, 7.7% missing data), C-4821snps (4821 SNPs, 12.3% missing data), D-3564snps (3564 SNPs, 10.4% missing data), and E-1887snps (1887 SNPs, 12.4% missing data). (*Step 6*) Finally, to conduct F_{ST} outlier analyses (see below), the five matrices (A–E) from previous step were partitioned into all, neutral, and outlier loci, for a total of 15 databases. For details refer to the SM (see also Additional file 2: Fig. S14).

Multivariate analyses

To determine the number of genetic clusters within the *humboldtianum* group, we conducted a principal component analysis (PCA)—designed to identify genetic groups through eigenvector decomposition of allele frequencies [99]—using matrices A–E with all, neutral, and outlier SNP loci. In addition, to identify de novo structure, we conducted an assessment of a priori designations using a discriminant analysis of principal components (DAPC). DAPC combines discriminant (DA) with principal component (PC) analyses to maximize genetic variance among groups while minimizing within-group variance [100, 101]. We selected the a priori groups by configuring the nine nominal *Chirostoma* morphospecies. We assessed the most likely number of clusters ($k=1-9$) using the *find.clusters* function within the *adegenet* v2.1.1. package [102] by selecting the k model with the lowest Bayesian Information Criterion (BIC) score. We calculated the number of PCs to be retained using the *dapcCross* validation function, which is based on the highest successful assignment that presented the lowest mean squared error [100].

Genomic structure and genomic differentiation analyses

To evaluate the number of genetic clusters within the *humboldtianum* group, we analyzed the final matrices (A–E) with all, neutral, and outlier loci under a maximum-likelihood framework using the software ADMIXTURE v1.3.0 [103]. This program estimates the proportion of the ancestral population (Q estimates) for each individual to calculate the number of genetic clusters (k) under a cross-validation procedure, where lower values represent the optimal number of populations [104]. All analyses were run with k values ranging from 1 to 12, which represent the nine nominal morphospecies described for the group, the sub-species

from Zirahuén Lake, *C. e. copandaro*, and two additional genomic groups that could be potentially be detected if a higher number of species exist. To further assess population clustering, we conducted admixture analyses within Chapala and Pátzcuaro-Zirahuén lakes, corresponding to clade I and clade III, respectively (see Results). The analysis for Chapala lake sample ($n=41$) was run with k values ranging from 1 to 5, which represent the five morphospecies of *humboltianum* group described from this lake. The analysis for Pátzcuaro-Zirahuén lakes sample ($n=24$) was run with k values ranging from 1 to 4, which represent the three morphospecies of *humboltianum* group from Pátzcuaro plus the subspecies *C. e. copandaro* from Zirahuén. We plotted all cross-validation and Q estimates using the *ggplot2* [105] and *pop-helper* [106] packages in R studio.

We computed analyses of pairwise F_{ST} comparisons to test the genetic differentiation among groups using the D-3564snps-all_loci matrix. We selected this matrix as it showed the general patterns observed with other matrices, but also because it had the highest percentage of variation explained obtained with multivariate analyses (see “Results”). We examined four alternative species delimitation schemes: (i) nine morphospecies (morphological-based hypothesis; sensu Barbour [10]), (ii) five mitonuclear groups (*mtDNA/nDNA*-based hypothesis; sensu Betancourt-Resendes et al. [11]), (iii) four genomic clusters ($k=4$) observed with the DAPC analyses (see “Results”), and (iv) three clusters detected by the admixture analyses ($k=3$) (see “Results”). To further test whether an agreement between species’ body size and trophic speciation exists [10], we calculated pairwise F_{ST} comparisons among ecotypes by lake (‘peces blancos’ vs. ‘charales’). Details on ecotype discrimination are provided in the SM. We calculated these analyses in ARLEQUIN v3.5.1.2 [107], and significance was determined using 10,000 permutations and a Bonferroni correction [108] to adjust p values for these calculations (significant levels of α are provided in Additional file 3: Tables S4–S8).

Phylogenetic analyses

We estimated phylogenetic trees under a maximum likelihood (ML) framework using the software IQ-TREE v2.0.6 [109]. We conducted phylogenetic inference for the five matrices (A–E) including all, neutral, and outlier SNP loci, and using *C. jordani* as the outgroup. All analyses used the GTR model with gamma distribution and the ascertainment bias correction (ASC) to account for acquisition discrepancies related to SNP datasets. We estimated branch support using IQ-TREE’s ultrafast bootstrap algorithm (UFBoot) with 500 replicates [110].

To evaluate patterns of incongruent evolutionary histories between mitochondrial and nuclear loci—mitonuclear discordance—we used the sequence data of two mitochondrial genes (cytochrome b or *Cytb*, and a fragment of the hypervariable control region, *D-loop*) from Betancourt-Resendes et al. [11]. Using the GTR model in IQ-TREE as explained above, we estimated ML *mtDNA* trees for each gene separately and also for the concatenated *mtDNA* matrix. We extracted tips shared between the *mtDNA* trees and our ddRADseq (D-3482snps-neutral_loci matrix) using the *keep.tip* function implemented in the R package *phytools* [111].

Multispecies coalescent analyses

We inferred a multi-coalescent species tree for the D-3482snps-neutral_loci matrix using the SNAPP v1.3.0 plug-in [112] implemented in BEAST2 v2.6.3 [113]. SNAPP allows the inference of a species tree from unlinked SNP data while bypassing gene tree inference [112]. For this analysis we used two chains of 30 million steps, sampling every 500 trees, with a burn-in of 10%. We set default priors for coalescent and mutation rates, as well as ancestral population size parameters. We visualized all results in the software Tracer v1.7 [114] to confirm that the analyses had converged, reached stationary, and that all effective sample sizes (ESS) were higher than 200. Lastly, we summarized the conflict of posterior distribution trees into the species tree using DENSITREE v2.01 [115].

To test the previously outlined species delimitation scenarios (i–iv) in a coalescent framework, we applied the Bayes Factor Delimitation (BFD*) approach [116] using SNAPP and BEAST2. To overcome computational burden, we generated three subsets encompassing the nine morphospecies with 39–59 individuals and 411–1102 SNP loci that were generated from the D-3482snps-neutral_loci matrix (see Additional files). With BFD* candidate species delimitation scenarios are evaluated according to the marginal likelihood estimates (MLE). The different scenarios are then compared using Bayes Factors (BF) [117], estimated by subtracting MLE values for two models and multiplying the difference by two ($BF = 2 \times (MLE_1 - MLE_2)$). As only two models are compared at the time, we evaluated all possible combinations among the species delimitation scenarios (see Table 1; Additional files). To set up the priors and MCMC runs, we followed the recommendations provided in the BFD* manual [118]. The speciation rate (λ), which represents the birth-rate on the Yule tree prior, was set to follow a gamma distribution with $\alpha = 2$ and $\beta = 200$.

Genetic diversity, effective population size (N_e), and F_{ST} outlier analyses

We estimated the genomic diversity of the *humboldtianum* group considering the four species delimitation scenarios outlined above. We did not test a scheme that included ecotypes as we did not find any support for this hypothesis (see Results). We used GenAEx v6.5 software [119] to calculate the observed (H_o) and expected (H_e) heterozygosity for each cluster using D-3482snps-neutral_loci matrix. Additionally, we evaluated the effective population size (N_e) of each species delimitation scenario by estimating linkage disequilibrium [83], heterozygote excess [120], and molecular co-ancestry values [121] with the software NeEstimator v2 [122].

To detect loci with high levels of genetic differentiation, we examined the A–E matrices formed by all SNP loci under a Bayesian framework using the program BayeScan [123]. In this analysis, F_{ST} coefficients are decomposed into a population-specific component (β) shared by all loci, and a locus-specific component (α) shared by all populations. When a locus-specific component is necessary to explain the data, selection is assumed to play a role at that locus [123]. To reduce the risk of false positives without reducing the power to detect loci evolving under selection, we set the default parameters and priors for a neutral model according to the total number of SNP loci in each matrix (100 for matrices with >1000 loci, and 1000 for matrices with >10,000 loci) [124]. Posterior Odds (PO) values were used to detect loci under selection in the context of multiple testing. PO are simply the ratio of posterior probabilities and indicate how more likely the model with selection is compared to the neutral model (Jeffrey's scale of evidence can also be used with posterior odds). A big advantage of PO values over posterior probabilities is that they directly allow the control of the False Discovery Rate [124]. Statistical significance for outlier loci was assumed if q values ≤ 0.05 . To determine the strength and direction of selection, we estimated the α parameter, where a positive value suggests diversifying selection and a negative value indicates balancing or purifying selection [124]. Lastly, to identify the approximate genomic region within which the outlier loci occur, we searched the consensus stack sequences on the NCBI server (<https://blast.ncbi.nlm.nih.gov/Blast.cgi>) using the Blastn algorithm. To choose the most similar candidate sequence, we used an e-value threshold lower than 0.05, a maximum target sequence of 5000 to retain all possible hits, and a sequence similarity above 75% to filter among the retained sequences. We ran all R analyses using R studio v1.2.1335 and R v4.1.1 [125].

Supplementary Information

The online version contains supplementary material available at <https://doi.org/10.1186/s12862-022-02063-0>.

Additional file 1: Appendix S1. Samples included in this study indicated the groups to which they belong in each of the different species hypotheses, analyses results, ecotypes, locality, and lake.

Additional file 2: Supporting Methods. Figure S1. Principal Components Analyses 19 matrices with different numbers of SNPs loci, missing data, individuals, and species of the *humboldtianum* group. **Figure S2.** Principal component analyses (PCAs) based on all, neutral, and outlier SNP loci from matrices A, B, C, and E. PCAs estimated using all and neutral SNPs (~2 k–33 k) consistently recovered four genomic groups that are in agreement with geographic patterns but not with the previously recognized morphospecies; these are delimited with convex hulls: *humboldtianum* sensu stricto group (blue), from Lake Zacapu; *estor* group (green) from Lakes Patzcuaro and Zirahuén; *chapalae* group (red), from Lake Chapala; and *C. sphyraena* group (yellow), also from Lake Chapala. PCA analyses based on ~100–400 SNPs also resolved four genomic groups in matrices A and B, and in matrices C and E, *chapalae* and *sphyraena* groups clustered together. Morphospecies are color-coded according to the genomic groups observed. **Figure S3.** Plots of Bayesian Information Criterion (BIC) vs. number of clusters (k) of DAPC analyses for all, neutral, and outlier SNP loci from matrices A–E. **Figure S4.** Discriminant analyses of principal components (DAPCs) estimated using ca. 1800–33,700 SNPs resolve four well-differentiated clusters. DAPCs based on outlier SNPs recover three to four groups. These results are largely consistent across analyses based on matrices A, B, C, and E (a–d), and are also concordant with PCA analyses. In neither case, the observed genomic clusters do correspond to the morphology-based species delimitation scenario. Morphospecies are color-coded according to the genomic clusters recovered: blue, *humboldtianum* sensu stricto, green, *estor* group; red, *chapalae* group; yellow, *C. sphyraena*. **Figure S5.** Plots of cross-validation error of each k (number of clusters) analyzed in Admixture analyses for all, neutral, and outlier SNP loci from matrices A–E. The cross-validation procedure was performed with the folds value = 5 (the default), a block relaxation algorithm as point estimation method, and the point estimation terminated with the objective function delta < 0.0001. **Figure S6.** Admixture assignment analyses estimated using ca. 80–33,700 neutral and outlier SNPs consistently identified three well-differentiated clusters ($k=3$) using matrices A, B, C, and E (a–d). Outlier SNPs show an increased intermingling of individuals among clusters compared with analyses based on all or neutral loci. Each bar represents the probability of assignment to each cluster. Genomic clusters are color-coded as blue, *humboldtianum* sensu stricto, green, *estor* group; orange, *chapalae-sphyraena* group. CHA, Lake Chapala; TEP, Tepuxtepec Dam; TRI, Trinidad Fabela Dam; PAT, Lake Patzcuaro; ZIR, Lake Zirahuén. **Figure S7.** Admixture assignment analyses estimated using ca. 33,700 SNP loci a) in Chapala Lake, and b) within Lakes Patzcuaro–Zirahuén. **Figure S8.** Phylogenetic trees of ca. 1800–33,700 SNP loci of the *humboldtianum* group estimated under a maximum likelihood framework in IQ-TREE. Phylogenetic trees were estimated using all (~1900–33,700), neutral-only (~1800–33,300), and outlier (~80–350) SNPs. Individuals are color-coded by genetic clusters according to Fig. 2. **Figure S9.** Mitochondrial trees (left) and ddRADseq phylogeny (right) of the *humboldtianum* group. The tips in the RADseq inferences were pruned to include the same individuals present in the mitochondrial trees. **Figure S10.** F_{ST} versus log₁₀-transformed posterior odds (PO) values for the global outlier detection calculated in BayeScan. The analyses were estimated considering the nine morphospecies and (A) 33,716 SNPs, where the vertical line represents the FDR threshold of $q=0.037$; (B) 10,157 SNPs, $q=0.040$; (C) 4821 SNPs, $q=0.025$; (D) 3564 SNPs, $q=0.034$; and (E) 1887 SNPs, $q=0.015$. **Figure S11.** Mean coverage of the subset with 15 samples using different values of the minimum raw reads required to form a stack (m_1 – m_6). **Figure S12.** Putative loci at different combinations of de novo assembly parameters: m , minimum reads required to form a stack; M , allowed SNPs in a stack required to form a putative locus in an individual; n , allowed SNPs in a stack required to form a locus in the population. Each parameter was changed one at the time ($m=2$ – 6 , $M=0$ – 6 , and $n=0$ – 11) while keeping the others at default values ($m_3M_2n_1$). **Figure S13.** Genotyped loci and

variant sites using a constraint on the number of minimum individuals in a population required to have that locus ($r = 40, 60, \text{ and } 80$) based on different cutoff, as follow: (A) minimum raw reads required to form a stack ($m = 2-6$), (B) maximum mismatches allowed between stacks of the same individual ($M = 0-6$), and (C) mismatches allowed between loci of different individuals ($n = 0-11$). **Figure S14.** Flow chart of the quality control and SNP filtering steps applied to generate the final datasets.

Additional file 3: Table S1. Information of the sampling localities in the central Mexico plateau: localities, lakes, number of samples (n), geographic coordinates (Latitude and Longitude) and Altitude (meters over sea level, m.o.s.l.). **Table S2.** Filters applied to SNP databases. The databases highlighted in bold were selected for further filtering. **Table S3.** Databases generated with Tassel filters used to evaluate the consistency of the genomic patterns to the number of SNPs, individuals, species, and missing data percentage. The matrices used in the genomic analyzes are indicated in bold. **Table S4.** Genomic pairwise F_{ST} comparisons for 3482 SNP loci among the nine morphospecies proposed by Barbour [10]. Values in bold indicate significance at $\alpha = 0.0014$ for pairwise comparisons, following sequential Bonferroni corrections. **Table S5.** Genomic pairwise F_{ST} comparisons for 3482 SNP loci among the five mitonuclear groups proposed by Betancourt-Resendes et al. [11]. Values in bold indicate significance at $\alpha = 0.005$ for pairwise comparisons, following sequential Bonferroni corrections. **Table S6.** Genomic pairwise F_{ST} comparisons for 3842 SNP loci among the four genomic clusters detected in *humboldtianum* group by DAPC analyses. Values in bold indicate significance at $\alpha = 0.0083$ for pairwise comparisons, following sequential Bonferroni corrections. **Table S7.** Genomic pairwise F_{ST} comparisons for 3482 SNP loci among the three genomic clusters detected in the *humboldtianum* group by Admixture and phylogenetic analyses. Values in bold indicate significance at $\alpha = 0.017$ for pairwise comparisons, following sequential Bonferroni corrections. **Table S8.** Genomic pairwise F_{ST} comparisons for 3482 SNP loci among the ecotypes in each lake. Values in bold indicate significance at $\alpha = 0.005$ for pairwise comparisons, following sequential Bonferroni corrections. **Table S9.** Outlier loci detected by the F_{ST} outlier analysis performed in BayeScan related to gene regions.

Acknowledgements

We are thankful to Yadira Ortiz (Universidad de Puerto Rico-Río Piedras) for helping with the DNA extractions and library preparation. Support was provided by Dirección General de Asuntos del Personal Académico—Universidad Nacional Autónoma de México (DGAPA-UNAM), postdoc fellowship Comunicado #161/2019 and Comunicado #80/2021 (V.J.P.), and Consejo Nacional de Ciencia y Tecnología (CONACYT), doctorate scholarship 411629 (C.d.R.P.-M.). Bioinformatic analyses were supported by the computational cluster at the Laboratorio Nacional de Análisis y Síntesis Ecológica (project CONACYT LN315810), ENES unidad Morelia (UNAM), the High Computing Performance Facility at UPR-PP (supported by INBRE P20GM103475 Grant) and the University of Oklahoma Supercomputing Center for Education & Research (OSCER). Financial support was provided by the University of Oklahoma Libraries' Open Access Fund. We thank Matthew Wershebe for his comments to the manuscript. Finally, we thank Sergio Godínez-Marrón for providing artistic illustrations of Figs. 1 and 3.

Author contributions

IB and ODD designed the study; IB and ODD provided samples; VJP, CdRP-M, and RB-R generated data; VJP and CdRP-M conducted ran analyses; VJP, CdRP-M, NCC, RB-R, and ODD drafted the manuscript; all authors contributed to the writing. All authors read and approved the final manuscript.

Funding

This project was funded by Coordinación de la Investigación Científica of the Universidad Michoacana de San Nicolás de Hidalgo (CIC-UMSNH 2018–2020). It was also supported by National Science Foundation (NSF) grants to R.B.R. (DEB-1932759 and DEB-1929248).

Availability of data and materials

Raw sequence data have been deposited in the NCBI SRA BioProject no. PRJNA816865 (accession no. SAMN26725252–SAMN26725331). SNPs genomic matrices are available from the Dryad Digital Repository <https://doi.org/10.5061/dryad.w0vt4b8rz>.

All corresponding sample codes and information are included in Additional file 1: Appendix 1.

Declarations

Ethics approval and consent to participate

All *Chiostoma* samples were collected from local fishermen's bycatch under the collection permit PPF/DGOPA-262/17 authorized by the Mexican Ministry of Environmental and Natural Resources (SEMARNAT). Sampling procedures followed the ethical capture methods and regulations approved by the Official Mexican Norm NOM-032-SAG/PESC-2015 and NOM-036-SAG/PESC-2015 for fishing in the lakes of central Mexico. Euthanasia protocols applied consisted of rapid chilling methods by placing the fish in ice water-cold baths (2–4 °C) as recommended by the American Veterinary Medical Association (AVMA; <https://www.avma.org/resources-tools/avma-policies/avma-guide-lines-euthanasia-animals>). None of our experiments included living organisms.

Consent for publication

Not applicable.

Competing interests

The authors declare that they have no competing interests.

Author details

¹Laboratorio de Ecología Molecular, Escuela Nacional de Estudios Superiores Unidad Morelia, Universidad Nacional Autónoma de México, Antigua Carretera a Pátzcuaro 8701, Ex-Hacienda de San José de La Huerta, 58190 Morelia, Michoacán, Mexico. ²Department of Biology, The University of Oklahoma, 730 Van Vleet Oval, Norman, OK 73019, USA. ³Facultad de Ciencias Naturales, Universidad Autónoma de Querétaro, Avenida de Las Ciencias S/N Juriquilla, Delegación Santa Rosa Jáuregui, 76230 Querétaro, Mexico. ⁴Laboratorio de Biología Acuática, Facultad de Biología, Universidad Michoacana de San Nicolás de Hidalgo, Edificio "R" Planta Baja, Ciudad Universitaria, 58030 Morelia, Michoacán, Mexico. ⁵Laboratorio Nacional de Análisis y Síntesis Ecológica Para la Conservación de Recursos Genéticos de México, Escuela Nacional de Estudios Superiores, Unidad Morelia, Universidad Nacional Autónoma de México, Apartado Postal 27-3 (Xangari), 58089, Michoacán, Morelia, Mexico.

Received: 18 March 2022 Accepted: 6 September 2022

Published online: 14 September 2022

References

- De Queiroz K. Species concepts and species delimitation. *Syst Biol.* 2007;56:879–86.
- Parker E, Dornburg A, Struthers CD, Jones CD, Near TJ. Phylogenomic species delimitation dramatically reduces species diversity in an antarctic adaptive radiation. *Syst Biol.* 2021;1–20.
- Schunter C, Garza JC, Macpherson E, Pascual M. SNP development from RNA-seq data in a nonmodel fish: how many individuals are needed for accurate allele frequency prediction? *Mol Ecol Resour.* 2014;14:157–65.
- Sukumaran J, Knowles LL. Multispecies coalescent delimits structure, not species. *Proc Natl Acad Sci USA.* 2017;114:1607–11.
- Chambers EA, Hillis DM. The multispecies coalescent over-splits species in the case of geographically widespread taxa. *Syst Biol.* 2020;69:184–93.
- Hillis DM, Chambers EA, Devitt TJ. Contemporary methods and evidence for species delimitation. *Ichthyol Herpetol.* 2021;109:895–903.
- Wiens JJ. Species delimitation: new approaches for discovering diversity. *Syst Biol.* 2007;56:875–8.
- Rojas-Carrillo P, Sasso-Yada LF. El pescado blanco. *Rev Digit Univ.* 2005;6:2–18.
- Alaye-Rahy N, Meléndez-Galicia C, Romero-Acosta C, Hernández-Montaño D. Estado actual de la pesquería de pescado blanco *Chiostoma estor* del Lago de Pátzcuaro, Michoacán, México. *Cienc Pesq.* 2017;25:5–16.
- Barbour CD. The systematics and evolution of the genus *Chiostoma* Swainson (Pisces, Atherinidae). *Tulane Stud Zool Bot.* 1973;18:97–141.

11. Betancourt-Resendes I, Perez-Rodríguez R, Barriga-Sosa IDLA, Piller KR, Domínguez-Domínguez O. Phylogeographic patterns and species delimitation in the endangered silverside "humboldtianum" clade (Pisces: Atherinopsidae) in central Mexico: understanding their evolutionary history. *Org Divers Evol*. 2020;20:313–30.
12. Barbour CD. A biogeographical history of chirostoma (Pisces: Atherinidae): a species flock from the Mexican Plateau. *Copeia*. 1973;1973:533–56.
13. Ayala-Ramírez GL, Ruiz-Sevilla G, Chacón-Torres A. La laguna de Zacapu, Michoacán. In: De la Lanza-Espino G, Hernández-Pulido S, editors. *Las Aguas Interiores de México: Conceptos y Casos*. México D.F.: AGT EDITOR, S.A.; 2007. p. 268–84.
14. Soria-Barreto M, Paulo-Maya J. Morfometría comparada del aparato mandibular en especies de Chirostoma (Atheriniformes: Atherinopsidae) del Lago de Pátzcuaro, Michoacán, México Morphometric comparison of the mandibular region in species of Chirostoma (Atheriniformes: Atherinopsid). *Hidrobiológica*. 2005;15:161–8.
15. Barbour CD, Chernoff B. Comparative morphology and morphometric of the pescados blancos (genus Chirostoma) from Lake Chapala Mexico. In: Echelle AA, Kornfield I, editors. *Evolution of fish species flock*. Orono, Me.: Univ. Maine Orono; 1984. p. 111–27.
16. De Los Angeles Barriga-Sosa I, Ibáñez-Aguirre AL, Arredondo-Figueroa JL. Morphological and genetic variation in seven species of the endangered Chirostoma "humboldtianum species group" (Atheriniformes: Atherinopsidae). *Rev Biol Trop*. 2002;50:199–216.
17. Echelle AA, Echelle AF. Evolutionary genetics of a "species flock" Atherinid fishes on mesa central of Mexico. In: Echelle AA, Kornfield I, editors. *Evolution of fish species flock*. Orono, Me.: Univ. Maine Orono; 1984. p. 93–110.
18. Bloom DD, Piller KR, Lyons J, Mercado-Silva N, Medina-Nava M. Systematics and biogeography of the silverside tribe menidiini (Teleostomi: Atherinopsidae) based on the mitochondrial ND2 gene. *Copeia*. 2009;408–17.
19. Foster K, Bower L, Piller K. Getting in shape: habitat-based morphological divergence for two sympatric fishes. *Biol J Linn Soc*. 2015;114:152–62.
20. Rodríguez Ruiz A, Granado Lorenzo C. Características del aparato bucal asociadas al régimen alimenticio en cinco especies coexistentes del género Chirostoma (lago de Chapala; México). 1988:35–51.
21. Mercado-Silva N, Lyons J, Moncayo-Estrada R, Gesundheit P, Krabbenhoft TJ, Powell DL, et al. Stable isotope evidence for trophic overlap of sympatric Mexican Lake Chapala silversides (Teleostei: Atherinopsidae: Chirostoma spp.). *Neotrop Ichthyol*. 2015;13:389–400.
22. Chaplin K, Sumner J, Hipsley CA, Melville J. An integrative approach using phylogenomics and high-resolution X-ray computed tomography for species delimitation in cryptic taxa. *Syst Biol*. 2020;69:294–307.
23. Hosegood J, Humble E, Ogden R, de Bruyn M, Creer S, Stevens GMW, et al. Phylogenomics and species delimitation for effective conservation of manta and devil rays. *Mol Ecol*. 2020;29:4783–96.
24. Nieto-Montes de Oca A, Barley AJ, Meza-Lázaro RN, García-Vázquez UO, Zamora-Abrego JG, Thomson RC, et al. Phylogenomics and species delimitation in the knob-scaled lizards of the genus *Xenosaurus* (Squamata: Xenosauridae) using ddRADseq data reveal a substantial underestimation of diversity. *Mol Phylogenet Evol*. 2017;106:241–53.
25. Leaché AD, Fujita MK. Bayesian species delimitation in West African forest geckos (*Hemidactylus fasciatus*). *Proc R Soc B Biol Sci*. 2010;277:3071–7. <https://doi.org/10.1098/rspb.2010.0662>.
26. Poelstra JW, Salmons J, Tiley GP, Schülöder D, Blanco MB, Andriambelosen JB, et al. Cryptic patterns of speciation in cryptic primates: microendemic mouse lemurs and the multispecies coalescent. *Syst Biol*. 2021;70:203–18.
27. Ferrari PL. Avances en el conocimiento de la Faja Volcánica Transmexicana durante la última década. *Boletín la Soc Geol Mex*. 2000;53:84–92.
28. Quintero-Legorreta O. Análisis estructural de fallas potencialmente activas. *Boletín la Soc Geol Mex*. 2002;55:12–29.
29. Israde-Alcántara I, Velázquez-Durán R, Lozano-García MS, Bischoff J, Domínguez-Vázquez G, Victor Hugo GM. Evolución Paleolimnológica del Lago Cuitzeo, Michoacán durante el Pleistoceno-Holoceno. *Boletín la Soc Geol Mex*. 2010;62:345–57.
30. Israde-Alcantara I, Garduño-Monroy VH. Lacustrine record in a volcanic intra-arc setting: the evolution of the Late Neogene Cuitzeo basin system (central-western Mexico, Michoacán). *Palaeogeogr Palaeoclimatol Palaeoecol*. 1999;151:209–27.
31. Israde-Alcántara I. Lagos Volcánicos y Tectónicos de Michoacán. In: Garduño-Monroy VH, Corona-Chávez P, Israde-Alcántara I, Menella L, Arreygüe E, B B, et al., editors. *Carta Geológica de Michoacán Escala 1:250000*. Universidad Michoacana de San Nicolás de Hidalgo; 1999. p. 45–72.
32. Domínguez-Domínguez O, Alda F, De León GPP, García-Garitaigoitia JL, Doadrio I. Evolutionary history of the endangered fish *Zoogoneticus quitzeoensis* (Bean, 1898) (Cyprinodontiformes: Goodeidae) using a sequential approach to phylogeography based on mitochondrial and nuclear DNA data. *BMC Evol Biol*. 2008;8:1–19.
33. Doadrio I, Domínguez O. Phylogenetic relationships within the fish family Goodeidae based on cytochrome *b* sequence data. *Mol Phylogenet Evol*. 2004;31:416–30.
34. Corona-Santiago DK, Doadrio I, Domínguez-Domínguez O. Evolutionary history of the live-bearing endemic *Allotoca diazi* species complex (Actinopterygii, Goodeinae): evidence of founder effect events in the Mexican pre-Hispanic period. *PLoS ONE*. 2015;10:1–21.
35. Domínguez-Vázquez G, Osuna-Vallejo V, Castro-López V, Israde-Alcántara I, Bischoff JA. Changes in vegetation structure during the Pleistocene-Holocene transition in Guanajuato, central Mexico. *Veg Hist Archaeobot*. 2019;28:81–91. <https://doi.org/10.1007/s00334-018-0685-8>.
36. Seehausen O, Wagner CE. Speciation in freshwater fishes. *Annu Rev Ecol Evol Syst*. 2014;45:621–51.
37. Mayr E. *Systematics and the origin of species*. New York: Columbia University Press; 1942.
38. Rundle HD, Nosil P. Ecological speciation. *Ecol Lett*. 2005;8:336–52.
39. Bernardi G. Speciation in fishes. *Mol Ecol*. 2013;22:5487–502.
40. Barluenga M, Stölting KN, Salzburger W, Muschick M, Meyer A. Sympatric speciation in Nicaraguan crater lake cichlid fish. *Nature*. 2006;439:719–23.
41. Elmer KR, Fan S, Kusche H, Luise Spreitzer M, Kautt AF, Franchini P, et al. Parallel evolution of Nicaraguan crater lake cichlid fishes via non-parallel routes. *Nat Commun*. 2014;5.
42. Burruss ED. Ecological diversification associated with the pharyngeal jaw diversity of Neotropical cichlid fishes. *J Anim Ecol*. 2016;85:302–13.
43. Salzburger W. Understanding explosive diversification through cichlid fish genomics. *Nat Rev Genet*. 2018;19:705–17. <https://doi.org/10.1038/s41576-018-0043-9>.
44. Takeyama T. Feeding ecology of lake Tanganyika cichlids. In: Abate ME, Noakes DLG, editors. *The behavior, ecology and evolution of cichlid fishes*. Dordrecht: Springer Netherlands; 2021. p. 715–51. https://doi.org/10.1007/978-94-024-2080-7_19.
45. Härer A, Bolnick DI, Rennison DJ. The genomic signature of ecological divergence along the benthic-limnetic axis in allopatric and sympatric threespine stickleback. *Mol Ecol*. 2021;30:451–63.
46. Jones FC, Grabherr MG, Chan YF, Russell P, Mauceli E, Johnson J, et al. The genomic basis of adaptive evolution in threespine sticklebacks. *Nature*. 2012;484:55–61.
47. Puebla O. Ecological speciation in marine v. freshwater fishes. *J Fish Biol*. 2009;75:960–96.
48. Nagel L, Schluter D. Body size, natural selection, and speciation in sticklebacks. *Evolution (N Y)*. 1998;52:209–18.
49. Bay RA, Arnegard ME, Conte GL, Best J, Bedford NL, McCann SR, et al. Genetic coupling of female mate choice with polygenic ecological divergence facilitates stickleback speciation. *Curr Biol*. 2017;27:3344–3349.e4. <https://doi.org/10.1016/j.cub.2017.09.037>.
50. Miller RR, Minckley WL, Norris SM, of Michigan. *Museum of Zoology U. Freshwater Fishes of México*. University of Chicago Press; 2005. <https://books.google.com.mx/books?id=MZXG-9jKyQC>.
51. Yang Z, Rannala B. Bayesian species delimitation using multilocus sequence data. *Proc Natl Acad Sci*. 2010;107:9264–9. <https://doi.org/10.1073/pnas.0913022107>.
52. Rannala B, Edwards SVS V, Leaché A, Yang Z. The multi-species coalescent model and species tree inference. In: Scornavacca C, Delsuc F, Galtier N, editors. *Phylogenetics in the Genomic Era*. No commercial publisher | Authors open access book; 2020. p. 3.3:1–3.3:21. <https://hal.archives-ouvertes.fr/hal-02535622>.

53. Betancur R, Naylor GJP, Ortí G. Conserved genes, sampling error, and phylogenomic inference. *Syst Biol.* 2014;63:257–62.
54. Deagle BE, Jones FC, Chan YF, Absher DM, Kingsley DM, Reimchen TE. Population genomics of parallel phenotypic evolution in stickleback across stream-lake ecological transitions. *Proc R Soc B Biol Sci.* 2012;279:1277–86.
55. Fagbémi MNA, Pigneur LM, André A, Smitz N, Gennotte V, Michaux JR, et al. Genetic structure of wild and farmed Nile tilapia (*Oreochromis niloticus*) populations in Benin based on genome wide SNP technology. *Aquaculture.* 2021;535.
56. Burress ED, Alda F, Duarte A, Loureiro M, Armbruster JW, Chakrabarty P. Phylogenomics of pike cichlids (Cichlidae: Crenicichla): the rapid ecological speciation of an incipient species flock. 2018;31:14–30.
57. Torati LS, Taggart JB, Varela ES, Araripe J, Wehner S, Migaud H. Genetic diversity and structure in *Arapaima gigas* populations from Amazon and Araguaia-Tocantins river basins. *BMC Genet.* 2019;20:1–13.
58. Carreras C, Ordóñez V, Zane L, Kruschel C, Nasto I, MacPherson E, et al. Population genomics of an endemic Mediterranean fish: differentiation by fine scale dispersal and adaptation. *Sci Rep.* 2016;2017(7):1–12.
59. Jackson AM, Semmens BX, De Mitcheson YS, Nemeth RS, Heppell SA, Bush PG, et al. Population structure and phylogeography in Nassau grouper (*epinephelus striatus*), a mass-aggregating marine fish. *PLoS One.* 2014;9.
60. Laconcha U, Iriando M, Arrizabalaga H, Manzano C, Markaide P, Montes I, et al. New nuclear SNP markers unravel the genetic structure and effective population size of albacore tuna (*Thunnus alalunga*). *PLoS ONE.* 2015;10:1–19.
61. Del Pedraza-Marrón CR, Silva R, Deeds J, Van Belleghem SM, Mastretta-Yanes A, Domínguez-Domínguez O, et al. Genomics overrules mitochondrial DNA, siding with morphology on a controversial case of species delimitation. *Proc R Soc B Biol Sci.* 2019;286.
62. Hughes LC, Cardoso YP, Sommer JA, Cifuentes R, Cuello M, Somoza GM, et al. Biogeography, habitat transitions and hybridization in a radiation of South American silverside fishes revealed by mitochondrial and genomic RAD data. *Mol Ecol.* 2020;29:738–51.
63. Marshall TL, Chambers EA, Matz MV, Hillis DM. How mitonuclear discordance and geographic variation have confounded species boundaries in a widely studied snake. *Mol Phylogenet Evol.* 2021;162:107194. <https://doi.org/10.1016/j.ympev.2021.107194>.
64. Shultz AJ, Baker AJ, Hill GE, Nolan PM, Edwards SV. SNPs across time and space: population genomic signatures of founder events and epizootics in the House Finch (*Haemorrhous mexicanus*). *Ecol Evol.* 2016;6:7475–89.
65. Alarcón-Durán I, Castillo-Rivera MA, Figueroa-Lucero G, Arroyo-Cabrales J, Barriga-Sosa IÁ. Diversidad morfológica en 6 poblaciones del pescado blanco *Chirostoma humboldtianum*. *Rev Mex Biodivers.* 2017;88:207–14. <https://doi.org/10.1016/j.rmb.2017.01.018>.
66. Hoban S, Kelley JL, Lotterhos KE, Antolin MF, Bradburd G, Lowry DB, et al. Finding the genomic basis of local adaptation: Pitfalls, practical solutions, and future directions. 2016.
67. Amish SJ, Ali O, Peacock M, Miller M, Robinson M, Smith S, et al. Assessing thermal adaptation using family-based association and FST outlier tests in a threatened trout species. *Mol Ecol.* 2019;28:2573–93.
68. Robertson S, Bradley JE, MacColl ADC. Eda haplotypes in three-spined stickleback are associated with variation in immune gene expression. *Sci Rep.* 2016;2017(7):1–9.
69. Jones FC, Chan YF, Schmutz J, Grimwood J, Brady SD, Southwick AM, et al. A genome-wide SNP genotyping array reveals patterns of global and repeated species-pair divergence in sticklebacks. *Curr Biol.* 2012;22:83–90. <https://doi.org/10.1016/j.cub.2011.11.045>.
70. Scharfack JP, Kalbe M, Harrod C, Rauch G. Habitat-specific adaptation of immune responses of stickleback (*Gasterosteus aculeatus*) lake and river ecotypes. *Proc R Soc B Biol Sci.* 2007;274:1523–32.
71. Leck KJ, Zhang S, Hauser CAE. Study of bioengineered zebra fish olfactory receptor 131–2. Receptor purification and secondary structure analysis. *PLoS One.* 2010;5.
72. Jimenez Baltazar J, Hernández Morales R, Domínguez DO. Caracterización fisicoquímica y microbiológica del vaso norte de Lago de Pátzcuaro, Michoacán, México. *Rev Latinoam el Ambient y las Ciencias.* 2018;9:400–15.
73. Lind OT, Dávalos-Lind L. An Introduction to the Limnology of Lake Chapala, Jalisco, Mexico. In: Hansen AM, van Afferden M, editors. The Lerma-Chapala watershed: evaluation and management. Boston, MA: Springer US; 2001. p. 139–49. https://doi.org/10.1007/978-1-4615-0545-7_6.
74. Hollenbeck CM, Portnoy DS, Gold JR. Evolution of population structure in an estuarine-dependent marine fish. *Ecol Evol.* 2019;9:3141–52.
75. Theodorou P, Radzevičiūtė R, Kahnt B, Soro A, Grosse I, Paxton RJ. Genome-wide single nucleotide polymorphism scan suggests adaptation to urbanization in an important pollinator, the red-tailed bumblebee (*Bombus lapidarius* L.). *Proc R Soc B Biol Sci.* 2018;285.
76. Whitlock MC, Lotterhos KE. Reliable detection of loci responsible for local adaptation: inference of a null model through trimming the distribution of FST. *Am Nat.* 2015;186:S24–36.
77. Andrews KR, Good JM, Miller MR, Luikart G, Hohenlohe PA. Harnessing the power of RADseq for ecological and evolutionary genomics. *Nat Publ Gr.* 2016. <https://doi.org/10.1038/nrg.2015.28>.
78. Schweyen H, Rozenberg A, Leese F. Detection and removal of PCR duplicates in population genomic ddRAD studies by addition of a degenerate base region (DBR) in sequencing adapters. *Biol Bull.* 2014;227:146–60.
79. Lowry DB, Hoban S, Kelley JL, Lotterhos KE, Reed LK, Antolin MF, et al. Breaking RAD: an evaluation of the utility of restriction site-associated DNA sequencing for genome scans of adaptation. *Mol Ecol Resour.* 2017;17:142–52.
80. Kajungiro RA, Palaikostas C, Pinto FAL, Mmochi AJ, Mtolera M, Houston RD, et al. Population structure and genetic diversity of Nile Tilapia (*Oreochromis niloticus*) strains cultured in Tanzania. *Front Genet.* 2019;10:1–12.
81. García Martínez RM, Mejía O, García-De León FJ, Barriga-Sosa IDLA. Extreme genetic divergence in the endemic fish *Chirostoma humboldtianum*: implications for its conservation. *Hidrobiologica.* 2015;25:95–106.
82. Navarrete Salgado NA. *Chirostoma* (Menidia): Ecología Y Utilización Como Especie De Cultivo En Estanques Rústicos. *BIOCYT Biol Cienc y Tecnol.* 2017;10:736–48.
83. Waples RS, Do C. Linkage disequilibrium estimates of contemporary Ne using highly variable genetic markers: A largely untapped resource for applied conservation and evolution. *Evol Appl.* 2010;3:244–62.
84. England PR, Cornuet JM, Berthier P, Tallon DA, Luikart G. Estimating effective population size from linkage disequilibrium: Severe bias in small samples. *Conserv Genet.* 2006;7:303–8.
85. Sowersby W, Cerca J, Wong BBM, Lehtonen TK, Chapple DG, Leal-Cardín M, et al. Pervasive admixture and the spread of a large-lipped form in a cichlid fish radiation. *Mol Ecol.* 2021;30:5551–71.
86. Setzke C, Wong C, Russello MA. Genotyping-in-Thousands by sequencing of archival fish scales reveals maintenance of genetic variation following a severe demographic contraction in kokanee salmon. *Sci Rep.* 2021;11:1–10. <https://doi.org/10.1038/s41598-021-01958-0>.
87. Alaye-Rahy N. Híbridos entre especies del género *Chirostoma* del Lago de Pátzcuaro, Michoacán. *México Cienc Pesq.* 1996;13:10–7.
88. Medina-Nava M. Ictiofauna de la subcuenca del Río Angulo cuenca del Lerma-Chapala. Michoacán *Zool Inf.* 1997;35:25–52.
89. Rojas-Carrillo PM, Fuentes-Castellanos D. Historia y avances del cultivo del pescado blanco. Instituto Nacional de Pesca, Dirección General de Investigación y Acuicultura; 2003.
90. Reed DH, Frankham R. Correlation between fitness and genetic diversity. *Conserv Biol.* 2003;17:230–7.
91. Peterson BK, Weber JN, Kay EH, Fisher HS, Hoekstra HE. Double digest RADseq: an inexpensive method for de novo SNP discovery and genotyping in model and non-model species. *PLoS One.* 2012;7.
92. Catchen JM, Hohenlohe PA, Bassham S, Amores A, Cresko WA. Stacks: an analysis tool set for population genomics. *Mol Ecol.* 2013;22:3124–40.
93. Catchen JM, Amores A, Hohenlohe P, Cresko W, Postlethwait JH. Stacks: building and genotyping loci de novo from short-read sequences. *Genes Genomes.* 2011;1:171–82.
94. Mastretta-Yanes A, Arrigo N, Alvarez N, Jorgensen TH, Piñeros D, Emerson BC. Restriction site-associated DNA sequencing, genotyping error estimation and de novo assembly optimization for population genetic inference. *Mol Ecol.* 2015:28–41.

95. Paris JR, Stevens JR, Catchen JM. Lost in parameter space: a road map for Stacks. *Methods Ecol Evol.* 2017;8:1360–73.
96. Pedraza-Marrón C del R, Silva R, Deeds J, Van Belleghem SM, Mastretta-Yanes A, Domínguez-Domínguez O, et al. Genomics overrules mitochondrial DNA, siding with morphology on a controversial case of species delimitation. *Proc R Soc B Biol Sci.* 2019;286.
97. Díaz-Arce N, Rodríguez-Ezpeleta N. Selecting RAD-Seq data analysis parameters for population genetics: the more the better? *Front Genet.* 2019;10:1–10.
98. Linck E, Battey CJ. Minor allele frequency thresholds strongly affect population structure inference with genomic data sets. *Mol Ecol Resour.* 2019;19:639–47.
99. Patterson N, Price AL, Reich D. Population structure and eigenanalysis. *PLoS Genet.* 2006;2:2074–93.
100. Jombart T, Devillard S, Balloux F. Discriminant analysis of principal components: a new method for the analysis of genetically structured populations. *BMC Genet.* 2010;11:1–15.
101. Miller JM, Cullingham CI, Peery RM. The influence of a priori grouping on inference of genetic clusters: simulation study and literature review of the DAPC method. *Heredity (Edinb).* 2020;125:269–80. <https://doi.org/10.1038/s41437-020-0348-2>.
102. Jombart T. Adegenet: a R package for the multivariate analysis of genetic markers. *Bioinformatics.* 2008;24:1403–5.
103. Alexander DH, Novembre J, Lange K. Fast model-based estimation of ancestry in unrelated individuals. *Genome Res.* 2009;19:1655–64.
104. Alexander DH, Lange K. Enhancements to the ADMIXTURE algorithm for individual ancestry estimation. *BMC Bioinformatics.* 2011;12.
105. Wickham H. ggplot2: elegant graphics for data analysis. Springer, New York; 2016. <http://ggplot2.org>.
106. Francis RM. pophelper: an R package and web app to analyse and visualize population structure. *Mol Ecol Resour.* 2017;17:27–32.
107. Excoffier L, Laval G, Schneider S. Arlequin (version 3.0): an integrated software package for population genetics data analysis. *Evol Bioinforma.* 2005;1:117693430500100.
108. Rice WR. Analyzing tables of statistical tests. *Evolution (NY).* 1989;43:223–5.
109. Nguyen LT, Schmidt HA, Von Haeseler A, Minh BQ. IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. *Mol Biol Evol.* 2015;32:268–74.
110. Hoang DT, Chernomor O, Von Haeseler A, Minh BQ, Vinh LS. UFBoot2: improving the ultrafast bootstrap approximation. *Mol Biol Evol.* 2018;35:518–22.
111. Revell LJ. phytools: an R package for phylogenetic comparative biology (and other things). *Methods Ecol Evol.* 2012;3:217–23.
112. Bryant D, Bouckaert R, Felsenstein J, Rosenberg NA, RoyChoudhury A. Inferring species trees directly from biallelic genetic markers: bypassing gene trees in a full coalescent analysis. *Mol Biol Evol.* 2012;29:1917–32.
113. Bouckaert R, Heled J, Kühnert D, Vaughan T, Wu CH, Xie D, et al. BEAST 2: a software platform for Bayesian evolutionary analysis. *PLoS Comput Biol.* 2014;10:1–6.
114. Rambaut A, Drummond AJ, Xie D, Baele G, Suchard MA. Posterior summarization in Bayesian phylogenetics using Tracer 1.7. *Syst Biol.* 2018;67:901–4.
115. Bouckaert RR. DensiTree: making sense of sets of phylogenetic trees. *Bioinformatics.* 2010;26:1372–3.
116. Leaché A, Fujita M, Minin V, Bouckaert R. Species delimitation using genome-wide SNP data. *Syst Biol.* 2014. <https://doi.org/10.1101/001172>.
117. Kass RE, Raftery AE. Bayes factors. *J Am Stat Assoc.* 1995;90:773–95.
118. Leaché AD, Bouckaert RR. Species trees and species delimitation with SNAPP: a tutorial and worked example. 2018.
119. Peakall R, Smouse PE. GENALEX 6: genetic analysis in Excel. Population genetic software for teaching and research. *Mol Ecol Notes.* 2006;6:288–95.
120. Zhdanova OL, Pudovkin AI. Nb_HetEx: a program to estimate the effective number of breeders. *J Hered.* 2008;99:694–5.
121. Nomura T. Estimation of effective number of breeders from molecular coancestry of single cohort sample. *Evol Appl.* 2008;1:462–74.
122. Do C, Waples RS, Peel D, Macbeth GM, Tillett BJ, Ovenden JR. NeEstimator v2: Re-implementation of software for the estimation of contemporary effective population size (Ne) from genetic data. *Mol Ecol Resour.* 2014;14:209–14.
123. Foll M, Gaggiotti O. A genome-scan method to identify selected loci appropriate for both dominant and codominant markers: a Bayesian perspective. *Genetics.* 2008;180:977–93.
124. Foll M. BayeScan v2.0 User Manual. 2010;1–9.
125. R Core Team. R: a language and environment for statistical computing. 2020. <https://www.r-project.org/>.
126. Del Pedraza-Marrón CR. Cambios en la distribución de los peces de agua dulce del Centro de México y sus posibles causas antropogénicas. Universidad Michoacana de San Nicolás de Hidalgo; 2012.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Ready to submit your research? Choose BMC and benefit from:

- fast, convenient online submission
- thorough peer review by experienced researchers in your field
- rapid publication on acceptance
- support for research data, including large and complex data types
- gold Open Access which fosters wider collaboration and increased citations
- maximum visibility for your research: over 100M website views per year

At BMC, research is always in progress.

Learn more biomedcentral.com/submissions

