Article title: Speciation on a complex landscape; phylogeny and demography of Mexican Goodeid fish.

Leeban H. Yusuf[1], Yolitzi Saldívar Lemus[1], Peter Thorpe[2], Constantino Macías Garcia[3], Michael G. Ritchie[1]

[1]Centre for Biological Diversity, School of Biology, University of St Andrews, St Andrews, UK.

[2] School of Medicine, Mackenzie Institute for early diagnosis, University of St Andrews, North Haugh, St Andrews, UK.

[3]Instituto de Ecologia, Universidad Nacional Autónoma de México, Ciudad Universitaria, Circuito exterior s/n anexo al Jardín Botánico C. P. 04510, Mexico City D. F, Mexico.

Keywords: phylogenomics, vicariance, demography, gene flow

Abstract

Understanding the phylogeographic history of a group and identifying the factors contributing to speciation is an important challenge in evolutionary biology. Vicariance has clearly played a major role in diversification across diverse groups of organisms, but how this interacts with intrinsic biological features such as changes in reproductive system or sexual selection is less well understood. The *Goodeinae* are a Mexican endemic group of live-bearing fish. Here, we develop genomic resources for species within the *Goodeinae* and use phylogenomic approaches to characterise their evolutionary history. We sequenced, assembled and annotated the genomes of four new *Goodeinae* genomes, including *A. toweri,* the only matrotrophic live-bearing fish without a trophotaenia in the group. We produced a phylogeny of the *Goodeinae* and estimate timings of species divergence. We determined the extent and timing of introgression between the species to assess if this may have occurred during an early radiation, or in more recent episodes of secondary contact. We also analyse patterns in the changes of effective population size for the species and examine the time course of expansion and decline. We used branch-site models to detect genome-wide positive selection across *Goodeinae*, and we specifically ask if this differs in *A. toweri*, where reversal of placental viviparity has recently occurred. We find that Goodeinae diverged rapidly, with reductions in effective population size after each split. Contrary to expectations, we find evidence of gene flow between geographically isolated species, suggesting vicariant speciation was supplemented by limited post-speciation gene flow, potentially as a result of river piracy. Genes under positive selection in the

group are likely to be associated with the switch to live-bearing. Overall, our studies suggest that both volcanism driven vicariance and changes in reproductive mode influenced radiation in the Goodeinae.

Introduction

Understanding the historical biogeography of species groups has traditionally been dominated by vicariance models of speciation (Coyne & Orr 2004; Mayr 1963). Allopatric divergence facilitates the evolution of unique adaptations and other differences, including incompatibilities that can stimulate speciation. In temperate species, this model is thought to involve climatic fluctuations which have patterned both interspecific and more recent intra-specific differences during repeated changes in distribution leading to cycles of allopatry and secondary contact. Many species groups show biogeographic patterns consistent with recent divergence and speciation being patterned by such vicariance, especially during the Pleistocene (Schmitt, 2007; Kadereit and Abbott, 2021), though divergence may have been initiated earlier (Klicka and Zink, 1997; Ebdon *et al.*, 2021). Although climatic cycles are perhaps less pronounced outside of the temperate regions, they also occurred in tropical systems that show related patterns of relatively recent allopatric divergence (Thom *et al.*, 2020).

Genomic analyses are revealing that many species show recent or historical gene flow, which is compatible with the "repeated allopatry and secondary contact" models of speciation (Harrison and Larson, 2014; Suvorov *et al.*, 2022). Some studies suggest that gene flow can facilitate speciation in multiple ways (Marques, Meier and Seehausen, 2019). One of these is when ancient hybridisation can provide genetic material contributing to adaptive radiation in species groups, either by increasing genetic variation or promoting exchange of adaptive genotypes. Some of the most detailed studies of this concern freshwater lacustrine radiations, especially of the African cichlids (Meier *et al.*, 2017; Malinsky *et al.*, 2018; McGee *et al.*, 2020; Svardal

et al., 2020). Determining the timing and biological significance of gene flow during speciation in species groups with complex evolutionary histories is an important challenge.

Central Mexico is a biodiversity hotspot, but relatively few detailed phylogeographic studies of endemic groups have been completed. Pleistocene climatic fluctuations are often invoked to explain biogeographic patterns, but these can be 'complex' (Colin and Eguiarte, 2016). Central Mexico contains much high altitude upland and has been subject to extensive dynamic volcanism during the Miocene and Pleistocene (Ferrari *et al.,* 2012). The trans-Mexican volcanic belt is implicated in a Mexican Transition Zone where Nearctic and tropical biotas meet (Morrone, 2010). The highlands may be characterised by sky islands with frequent fragmentation. Mastretta-Yanes *et al.* (2015) argued that such changes may have led to opportunities for hybridisation during speciation of Mexican highland species. Freshwater fish, as well as other organisms (Musher *et al.,* 2022) associated with river networks, may have an added complexity to their distributional changes as volcanism has influenced river patterns (Dias *et al.,* 2013; Craw *et al.,* 2016), including river piracy. In *Poeciliopsis* and *Poecilia genera,* phylogeographic studies implicate a Plio-Pleistocene vicariant event stimulated by the Trans-Mexican Volcanic Belt in driving speciation (Mateos, 2005). Other biogeographical studies show concordant patterns of river vicariance driven by volcanic activity in the Trans-Mexican Volcanic Belt in less well-studied freshwater fish taxa including Goodeins (Domínguez-Domínguez, Doadrio and León, 2006; Domínguez-Domínguez *et al.,* 2008; Pérez-Rodríguez *et al.,* 2015).

Here we report the first genomic scale studies of the evolutionary history of an endemic group of Mexican fish, the *Goodeinae*. These are a group of viviparous, freshwater fish comprising 36 species across 16 genera. Previous studies have suggested that they radiated in the Miocene ~16 MYA predominantly due to allopatric speciation (Webb *et al.*, 2004; Pérez-Rodríguez *et al.*, 2015). Their helminth parasites show similar biogeographic patterns (Quiroz-Martínez and Salgado-Maldonado, 2013). It has been proposed that the Goodeins may represent an adaptive radiation following the evolution of live-bearing, and that speciation rate may be influenced by sexual selection. All species within *Goodeinae* are viviparous and matrotrophic, such that offspring depend on a continuous maternal provisioning (Wourms, Grove and Lombardi, 1988; Vega-López *et al.*, 2007). Embryos develop extremely rapidly via maternal provisioning through a placental analogue called the trophotaenia (Wourms, Grove and Lombardi, 1988; Hollenberg and Wourms, 1995). Within the *Cyprinodontiformes* (including the *Goodeidae* and *Poecilidae*) speciation rate is higher in live-bearing fish (Helmstetter *et al.*, 2016). Interestingly, one species of Goodeid, *Ataenobious toweri,* is thought to have lost the trophotaenia, though little is known about any physiological or genetic changes involved.

Sexual selection has influenced behavioral and morphological sexual dimorphism in the *Goodeinae*. Males cannot inseminate females without female co-operation (Bisazza, 1997), so sexual selection acts strongly on female choice and male secondary sexual traits. Females show preference for male colour patterns across the group (Moyaho, Garcia and Ávila-Luna, 2004; Garcia and Ramirez, 2005; Garcia and Lemus, 2012), particularly in fin structures, and there are pronounced differences in

fin size and body shape in males (Garcia, Jimenez and Contreras, 1994; Garcia, Saborío and Berea, 1998; Foster and Piller, 2018). Live-bearing may also give rise to considerable sexual conflict in the group, which can influence speciation rate (Gavrilets, 2000, 2014; Gavrilets and Waxman, 2002). How geography and these intrinsic biological features may interact during speciation is unknown. Previous studies based on microsatellite markers showed that sexually dimorphic species show greater genetic differences between populations (Ritchie *et al.*, 2007), but gene flow between populations of a dimorphic species was unrelated to male morphology, instead showing isolation by distance and vicariance (Macías Garcia *et al.*, 2012). Mitochondrial DNA sequencing across the group supported evidence for an early radiation, but provided equivocal evidence for a faster speciation rate in more sexually dimorphic species (Ritchie *et al.,* 2005).

Here, genomic resources were developed for species within *Goodeinae,* and we used phylogenomic approaches to characterise their evolutionary history. The genomes of four new *Goodeinae* genomes, including *A. toweri* were sequenced, assembled and annotated. We have produced a phylogeny of the species and estimated timings of the divergence. We also analysed the extent and timing of introgression between the species to assess if this may have occurred during an early radiation, or in more recent episodes of secondary contact. We also analysed patterns in the changes of effective population size for the species (many of which are critically endangered) and examine the time course of expansion and decline (Bailey, Garcia and Ritchie, 2007). We used branch-site models to detect genome-wide positive selection across Goodeinae, and

we specifically asked if this differs in *A. toweri*, where reversal of placental viviparity

has recently occurred.

Methods

Read mapping and filtering

Raw genomic reads for each species were trimmed using fastp v.0.20.1 (Chen *et al.*, 2018) with default parameters. Cleaned reads were then mapped to a *G. multiradiatus* reference genome (Du *et al.*, 2022) using bwa (v. 0.7.17). We used samtools (v. 1.11) (Li *et al.*, 2009) to index, sort, mark and remove duplicate reads from each sample and Picard (Institute, 2019) to add read groups to each sample. Subsequently, variant calling was performed on mapped reads using Freebayes (v. 1.3.2) with the following parameters: --report-genotype-likelihood-max --no-population-priors --use-best-n-alleles 4 --hwe-priors-off --use-mapping-quality --ploidy 2 --theta 0.02 --haplotype-length -1 --genotype-qualities. Finally, variants were filtered using GATK hard-filtering best practices guidelines (Van der Auwera *et al.*, 2013) ($QD < 2.0$, $QUAL < 30.0$, $SOR > 3.0$, $FS > 60.0$, $MQ < 40.0$, $MQRankSum < -12.5$, $ReadPosRankSum < -8.0$).

Phylogenetic history and dating of Goodeinae

To infer the phylogenetic history of the *Goodeinae*, we further filtered variants used in our phylogenetic analysis. We removed variants with missing genotypes across any of our samples, non-biallelic positions, and positions with either no alternative allele or only alternative alleles using bcftools. Additionally, we sample SNPs equally across the genome and pruned these by sampling 1 SNP for every 100 SNPs, to try to select independent histories. We used the following parameters in bcftools: bcftools +prune -w 100bp –n 1. The pruned and filtered VCF was then converted into a phylip alignment by concatenating all variants using vcf2phylip (Ortiz, 2019) for all nine

species. We first inferred phylogenetic relationships using a IQTREE2 (v. 2.1.4) using ModelFinder and 1000 bootstraps. Since maximum-likelihood concatenation approaches may fail when incomplete lineage sorting is prominent, we sought to infer a species tree using SVDquartets (Chifman and Kubatko, 2014, 2015) in paup (v.4.0a)(Wilgenbusch and Swofford, 2003) with 100 bootstraps and using filtered, pruned variants.

To infer divergence times and effective population sizes for ancestral nodes, we used the A00 model in BPP (v4.4.0)(Flouri *et al.*, 2018) and putatively selectively-neutral non-coding loci. To generate our selectively-neutral non-coding dataset, we first estimated substitution rates of four-fold degenerate sites (non-conserved regions) in 21 species across *Cyprinodontiformes* to identify regions with conserved substitution rates using PhastCons (Siepel et al. 2005; Pollard et al. 2010; Yusuf et al, in prep). Variants falling within these conserved regions were subsequently removed from our dataset, alongside coding regions, using bedtools (v. 1.12) (Quinlan and Hall, 2010). Additionally, to ensure sampled loci represented different evolutionary histories as required by BPP, we made 10kb windows that were at least 100kb away from each other. This yielded 2,740 loci of variable size that were converted into alignments with no missing sites for any species using vcf2phylip (Ortiz, 2019). To estimate divergence times and effective population sizes, we used A00 analysis with default settings except for amendments to theta (0.003) and tau (0.03) priors. We used a burn-in period of 8,000 and 100,000 samples. Divergence time estimates and effective population size estimates were scaled to geological time using the *Xiphophorus*

maculatus mutation rate (3.5×10^{-9}) from Schumer et al. (2018) and a generation time of 1 via R package bppr (v.0.6.1) (Angelis and Dos Reis, 2015).

Principal component analysis

To supplement the phylogenetic analysis, we sought to understand sample relationships using principal components analysis. We utilized hard-filtered biallelic SNPs that contained no missing data across our samples and were pruned based on distance to nearest SNP. We used plink2 to compute principle components analysis from 8,161,340 SNPs using the following parameters: "--double-id --allow-extra-chr --set-missing-var-ids @:# --vcf-half-call m –pca" (Frazer *et al.*, 2007; Chang *et al.*, 2015).

Demographic inference and population size changes

In order to detect recent changes in effective population size for each species from a single representative sample, we used the pairwise sequentially Markovian coalescent (PSMC) model (Li and Durbin, 2011). We converted mapped reads for each species individually into mpileup format using samtools and bcftools with the -c option (Li *et al.*, 2009). We then filtered out variants with depth lower than 10, a depth higher than 100 and mapping quality lower than 30, and converted variants into *fastq* format using the *vcf2fq* function of the vcfutils.pl program available via samtools (Li *et al.*, 2009). To determine choice of time interval parameters, we tested a range of interval parameters and found no difference in our results. As a result, we used default interval parameters of 4+25*2+4+6 for our final analysis. To scale estimates of effective population size,

we used the *Xiphophorus* maculatus mutation rate (3.5 x 10^{-9}) from Schumer et al. (2018) and a generation time of 1.

To understand how changes in population size affected genetic divergence genome-wide, we computed absolute nucleotide divergence (D_{XY}) in 50 kb non-overlapping windows between three relatively closely-related pairs: *A. splendens* and *X. captivus*; *I. furcidens* and *X. resolanae;* and *G. multiradiatus* and *A. toweri*, using the program pixy (v. 1.2.3.beta1) (Korunes and Samuk, 2021).

Testing global and local patterns of introgression

To test for evidence of introgression between species, we calculated Patterson's D-statistic using the *Dtrios* function in Dsuite v.4 (Green *et al.*, 2010; Malinsky, Matschiner and Svardal, 2020) with hard filtered variants (described above). Specifically, we inferred introgression across 56 trios with trios arranged according to phylogenetic relationships in our inferred species tree. As well as calculating D-statistics for all 56 trios, we also calculated the admixture fraction (*f4-ratio*). To control for false discovery, we applied a Benjamini-Hochberg procedure to p-values for all trios and kept only trios with corrected p-values lower than 0.05 (Benjamini and Hochberg, 1995). To filter out trios with unreliable signals, we removed trios where *D* and the *f4-ratio* was lower than 1%.

Because introgression signals can represent both (a) very recent introgression between extant lineages and (b) ancient introgression between ancestral lineages, we inferred the *f-branch* statistic via Dsuite. The *f-branch* statistic computes allele sharing

using *f4-ratios* between P3, where P3 is an outgroup species to P1 species and P2 species, and the descendants of a branch labelled *b* (i.e., the descendants of P2, for example) relative to a sister branch labelled *a*. This statistic allows for correlated *f4-ratios* to be disentangled and specific branches with strong signals of recent introgression to be identified.

To supplement tests for introgression, we also sought to understand local patterns of introgression and identify genes that may have been introgressed. We estimated f_d genome-wide in 150bp windows using the *Dinvestigate* program in *Dsuite* (Martin, Davey and Jiggins, 2015; Malinsky, Matschiner and Svardal, 2020). Specifically, we surveyed introgression genome-wide only for trios that showed the most consistent signals of introgression across all genome-wide estimates of introgression. To determine genes which may have introgressed across trios, we subsetted our dataset to only include windows within the top 10% of windows based on f_d genome-wide. We then used the *GenomicRanges* package (Lawrence *et al*., 2013) in R to determine whether any windows overlapped with annotated genes in our gff file. This subsetting approach allowed for the conservative identification of high-confidence introgressed genes maintained across multiple independent introgression events or an ancient event(s).

Detecting positive selection in rapid diversification of *Goodeinae*

To test for positive selection in short internal branches leading to the rapid diversification of Goodeids, we used the reference genome from *G.multiradiatus* and our hard-filtered variant calls to produce a consensus sequence alignment for each

gene using the vcf2fasta software (https://github.com/santiagosnchez/vcf2fasta). To ensure gene alignments did not contain de-novo genes not found in our outgroup species *C.baileyi*, alongside the reference genome and hard-filtered VCF file, we provided a filtered gene annotation file containing only annotated genes found in both *G. multiradiatus* and *C. baileyi*. To produce this consensus gene annotation, we lifted annotations from *C. baileyi* to the *G. multiradiatus* annotation; out of 29,739 genes we considered only 20,138. Nucleotide gene alignments were then aligned using MACSE v2.05 and the following parameters: "-prog refineAlignment -optim 2 -local_realign_init 0.001 -local_realign_dec 0.001". Nucleotide alignments were then masked for any remaining internal stop codons and frameshift mutations using the parameters: "-codonForInternalStop NNN -codonForInternalFS --- -charForRemainingFS – ". Finally, nucleotide alignments were translated into amino acid alignments and were subsequently used to produce a codon alignment using the program PAL2NAL and the parameters: "-nogap -nomismatch".

We used two approaches to detect positive selection. First, for each internal branch and codon alignment, we asked whether a proportion of sites were evolving under positive selection relative to background branches using aBSREL (Smith *et al.*, 2015). For each codon alignment, we used the inferred species tree phylogeny. Since all internal branches were tested separately, p-values were corrected for multiple testing and filtered ($p < 0.05$) using the Benjamini-Holm procedure. Secondly, for all internal branches considered together, we tested whether specific codons in each codon alignment (gene) were under positive selection in internal branches, using the codon model MEME (Murrell *et al.*, 2012) and the species tree phylogeny. Again, p-values

were corrected for multiple testing using the Benjamini-Hochberg procedure (Benjamini and Hochberg, 1995). Extraction of results for aBSREL and MEME from JSON files was performed using Python package *phyphy* (J. Spielman 2018). Additionally, we performed gene ontology via R package 'clusterProfiler' on genes found to have evidence of positive selection on at least one internal branch (Yu *et al.*, 2012).

To facilitate understanding of how the trophotaenium was lost in *A. toweri,* we performed a branch-site test via aBSREL on the terminal branch leading to *A. toweri*. We supplemented this test for positive selection in *A. toweri* in two ways. First, by overlapping positively selected genes with genes upregulated in the trophotaenia of *G.multiradiatus*, taken from a recent genomic and transcriptomic analysis of *G.multiradiatus* (Du *et al.,* 2022). Second, we used the Ensembl Variant Effect Predictor (VEP) tool to infer, using genome annotation information, high-impact mutations in protein-coding genes that may drastically alter protein function (McLaren *et al.*, 2016). Altogether, we searched for genes upregulated specifically in the trophotaenium of *G. multiradiatus* that may also have high-impact mutations in *A. toweri* indicating potential pseudogenization.

Results

Rapid diversification of Goodeinae occurred in the middle Miocene

Both approaches to inferring the phylogeny gave a consistent species tree topology with complete bootstrap support and complete posterior probabilities, and recapitulate relationships inferred in previous molecular phylogenies produced using mitochondrial sequences (Webb *et al.*, 2004). We find that following the divergence of *Empetrichthyinae* (represented by *C. baileyi*), a group containing *A. toweri* and *G. multiradiatus* diverged next, supporting previous suggestions that the evolutionary reversal of trophotaenia development occurred in *A. toweri*, as opposed to an ancestral placement of *A . toweri* at the crown of the *Goodeinae.* We resolve the placement of *C. lateralis*, which diverged following the split of *A. toweri* and *G. multiradiatus*. The remaining species fall into two groups; a group containing *I. furcidens* and *X. resolanae*, and another containing *G. atripinnis*, *A. splendens* and *X. captivus,* both of which are concordant with previous phylogenetic reconstructions of *Goodeinae* (Webb *et al.*, 2004) (Figure 1). These patterns are also recovered in principal components analysis of filtered SNPs (Supplementary figure 1). Whilst PC1 likely reflects technical variation separating *G. multiradiatus* from all other species, PC2 and PC3 cluster species reflect relationships described in the phylogenetic inference.

Since classic measures of phylogenetic support tend to overestimate confidence in phylogenomic datasets (Kumar *et al.*, 2012; Salichos and Rokas, 2013; Yang and Zhu, 2018), we inferred gene concordance factors using the species tree inferred with ASTRAL and 19,908 gene trees via IQTREE2. Gene concordance factors quantify the

percentage of gene trees that contain a given branch in the species tree. We find that almost all branches show extremely low gene concordance, except for branches leading to the split of *G. atripinnis*, *A. splendens* and *X. captivus,* and of *I. furcidens* and *X. resolanae* (Figure 1). This gene tree discordance is confirmed by inference of normalized quartet scores in ASTRAL where only 60% of quartets across all gene trees are also found in the species tree. Additionally, around the crown radiation, we observe short branch lengths and broad variation in support for alternate topologies, indicating frequent incomplete lineage sorting, presumably as a result of rapid speciation of *Goodeinae* (Figure 1).

To understand when diversification of *Goodeinae* began, we estimated divergence times and effective population sizes using filtered dataset of 2,740 putatively neutral non-coding loci, the species tree estimated above using ASTRAL and a multi-species coalescent approach. We find that crown age of the group and the divergence of *Empetrichthyinae* occurred between 24.7 – 24.4 MYA. We estimate that the diversification of *Goodeinae* began 18 MYA with the diversification of a clade containing *G. multiradiatus* and *A. toweri*. Following the split of *G. multiradiatus* and *A. toweri*, divergence of two clades containing (a) *G. atripinnis*, *A. splendens* and *X. captivus,* and (b) of *I. furcidens* and *X. resolanae* occurred alongside divergence of *C.lateralis* around 17.4 – 17.2 MYA. We estimate that the split of *C.lateralis* from the others was accompanied by considerable reduction (~100 fold) in effective population size, likely owing to vicariance as a result of hypothesized increase in volcanism (Webb *et al.*, 2004). This is supported by similar reduction (~5 fold) in effective population size following the split of *C. baileyi*. Divergence of *A. toweri,* where

evolutionary reversal of trophotaenia is hypothesized, occurred around 13.9 – 13.1 MYA. Altogether, our divergence time and effective population size estimates suggest incomplete lineage sorting is likely explained by rapid vicariant speciation events that occurred early in the diversification of Goodeids.
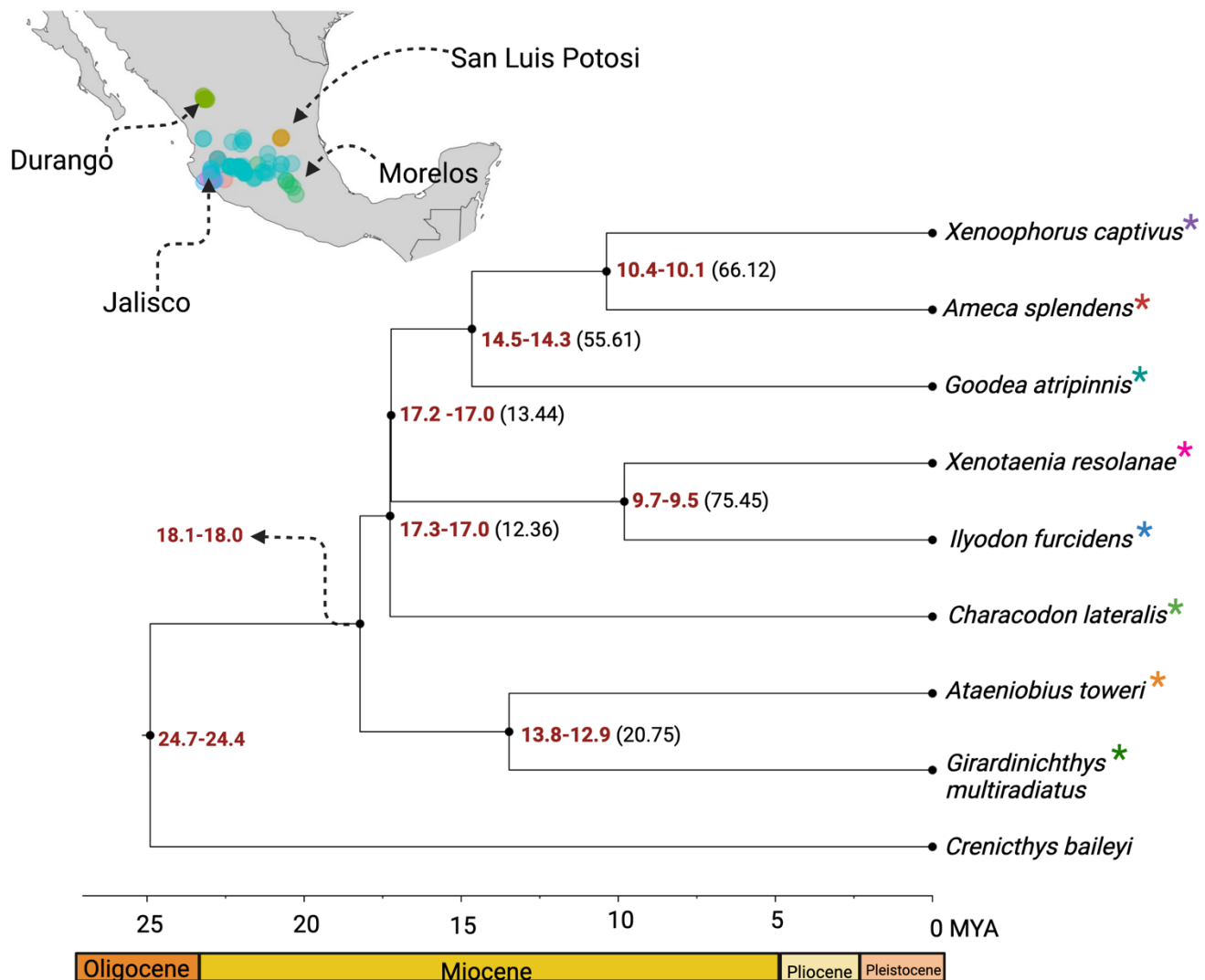
Figure 1: Species tree of the family *Goodeidae.* Scaled divergence times are shown for each node (in red) and gene concordance factors are shown for relevant nodes (in brackets). Colours in asterisks on the phylogeny denote species in map showing distribution of species. Distribution information taken from FishNet2.net.
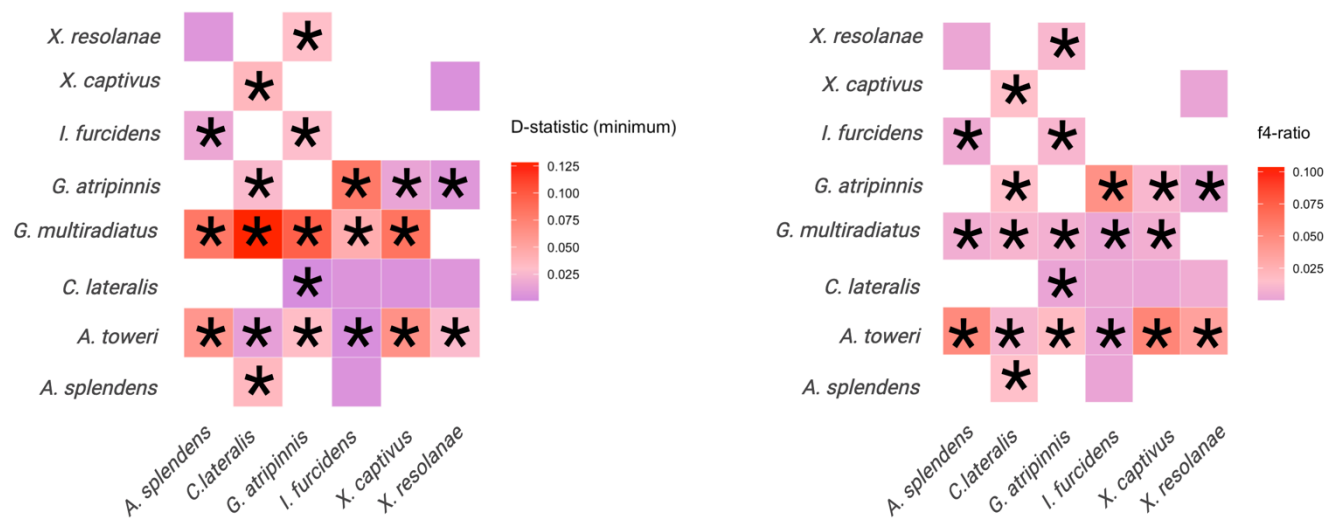
Figure 2: Evidence of gene flow between species in *Goodeinae*. Plots showing D-statistics and f4-ratios in a pairwise manner. Significant allele sharing after multiple-testing correction is indicated by panels with an asterisk.

Limited evidence of introgression supports vicariant speciation of *Goodeids.*

Alongside incomplete lineage sorting, gene flow may also produce patterns of underlying gene tree conflict with the species tree. To assess the potential importance of gene flow in explaining patterns of genetic variation in Goodeids, we computed Patterson's D statistic and the f4-ratios for all possible species combinations, without assuming or arranging species according to their species tree relationships. This approach produces conservative estimates of allele sharing (D_{min}) for all trios across all possible topologies (Malinsky *et al.*, 2018; Malinsky, Matschiner and Svardal, 2020). We find that 41 out of 56 tested trios (73%) have a significantly positive D-statistic value (D_{min}) after correcting for multiple testing, with average genome-wide allele sharing of 4.7% (Figure 2).

However, estimates of introgression are not independent since a single, ancestral gene flow event may result in erroneous interpretations of recent, widespread gene flow. We therefore re-calculated D_{min} by reorganising trios according to the relationships observed in the species tree inferred using ASTRAL. Specifically, we calculated the f-branch statistic fb(C) which utilises correlated f4-ratios to provide branch-specific estimates of introgression. We find support for gene flow events (fb(C) > 10%) between *A. toweri*, which is endemic to San Luis Potosí, and all other species apart from *G. multiradiatus.* Additionally, we find support for weaker introgression between *G. multiradiatus* and an ancestral branch leading to the divergence of *G. atripinnis, X. captivus* and *A. splendens.* Altogether, we suggest there is evidence for

a potential ancient gene flow event between the ancestor of *G. multiradiatus* and *A. toweri* and an ancestor of *G. atripinnis, X. captivus* and *A. splendens.*

To examine patterns of local introgression and identify introgressed genes, we computed f_d for the introgression event(s) with most support. We examined introgression between trios containing *G. multiradiatus* (P1), *A. toweri* (P2) and all other Goodeines (P3), with *C. baileyi* (P4) used as the outgroup (Supplementary Figure 2). We find similar levels of mean admixture proportions across surveyed across comparisons, with highest levels of mean admixture (~7.4%) observed in *G. atripinnis*, which shares its range with *A. toweri,* and in *X. resolanae*, which is currently allopatric from *A. toweri.* In *C. lateralis*, *A. splendens* and *X. captivus*, lower levels of genome-wide admixture with *A. toweri* are observed (5.7 – 5.9%), perhaps owing to differential loss of introgressed variation.

Across all comparisons, we find 64 genes that show evidence of high f_d (top 10%) across all trios. Among those, we find evidence of introgression in class II, major histocompatibility complex, transactivator (CIITA), which mediates expression of other MHC II genes and is under divergent sequence evolution in other viviparous lineages (Roth *et al.*, 2020). We found a significant enrichment of genes related to sugar and fat metabolic processes, as well as activation and regulation of transcription factors (Supplementary table 1). However, gene ontology analysis of introgressed genes maintained across all comparisons showed no significant enrichment of any biological processes, after correcting for multiple testing. Since introgressed genetic variation may be differentially lost by drift or removed by purifying selection, we also conducted

gene ontology analysis on the comparison (introgression between *A. toweri* and *X. resolanae*) with most introgressed genetic variation to recover as much genetic variation that was initially exchanged as possible. For this comparison, we found significant enrichment of genes involved in cilium projection and organization, cell projection and gastrulation (Supplementary table 2).

Evidence for large changes in population size across Goodeid diversification and on contemporary timescales.

To identify contemporary changes in contemporary effective population size for each species following their divergence, we reconstructed the demographic trajectory through time for 9 species in *Goodeinae* including *C. baileyi* using the Pairwise Sequentially Markovian Coalescent (PSMC) approach (Li and Durbin, 2011). We find that most species showed a coincident large increase in population size starting around 5 MYA (well after their initial radiation), except for *G. multiradiatus* and *X. captivus*, which began diverging around 1 MYA. Populations then declined progressively since the middle Pleistocene, though there is substantial variation in the initiation and rate of decline (Figure 3). Decline was persistent until recently, with only *G. atripinnis* and to a lesser extent *G. multiradiatus* showing subsequent growth then decline in more recent times, prior to the Holocene. Altogether, our analyses indicate considerable, recent fluctuation in effective population size in *Goodeinae*, paired with similar fluctuations in effective population size during early diversification of the family.

To assess the potential impact of changes in effective population size on genome divergence between closely related species, we calculated absolute nucleotide

divergence (D_{XY}) in autosomes and the X chromosome between closely related species in Goodeids. An important consequence of population contraction is that genetic variation is differentially affected on the X chromosome compared to the autosomes. Specifically, ancestral population contractions may lead to reduction of genetic diversity on the X chromosome compared to autosomes, resulting in lower levels of D_{XY} on the X chromosome (Van Belleghem *et al.*, 2018). However, since D_{XY} measures nucleotide differences accrued since the divergence time, it is not affected by contemporary changes in population size after species divergence. We find significantly lower D_{XY} on the X chromosome compared to the autosomes between *I. furcidens* and *X. resolanae* (*Welch's t* = 3.1389, df = 92.575, p-value < 0.001), and between *G. multiradiatus* and *A. toweri* (*Welch's t* = 4.7094, df = 87.64, p-value < 0.001), but not between *A. splendens* and *X. captivus* (*Welch's t* = 1.7081, df = 92.675, p-value = 0.09), though we did also observe lower levels of divergence on the X chromosome in this comparison also (Figure 4).
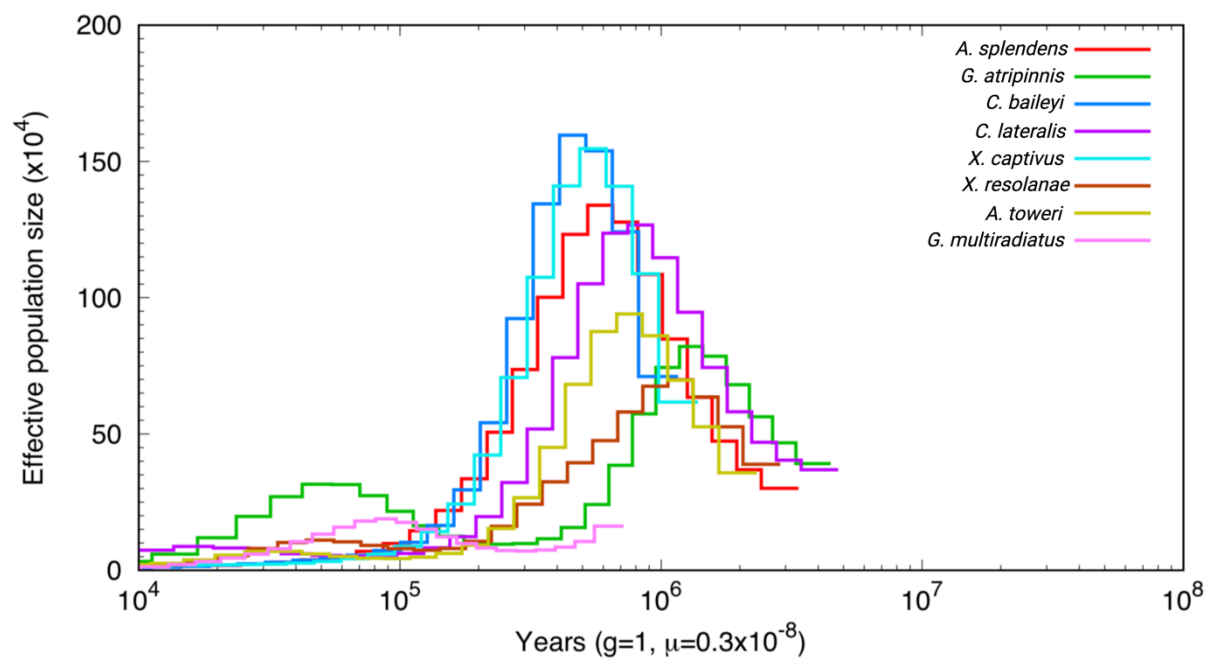
Figure 3: Concordant changes in contemporary effective population size across _Goodeidae_. Inferred effective population sizes through time using pairwise sequentially Markovian coalescent (PSMC) analysis. The X axis indicates scaled time estimates, assuming a mutation rate of 3.5 x 10^{-9} and a generation time of 1.
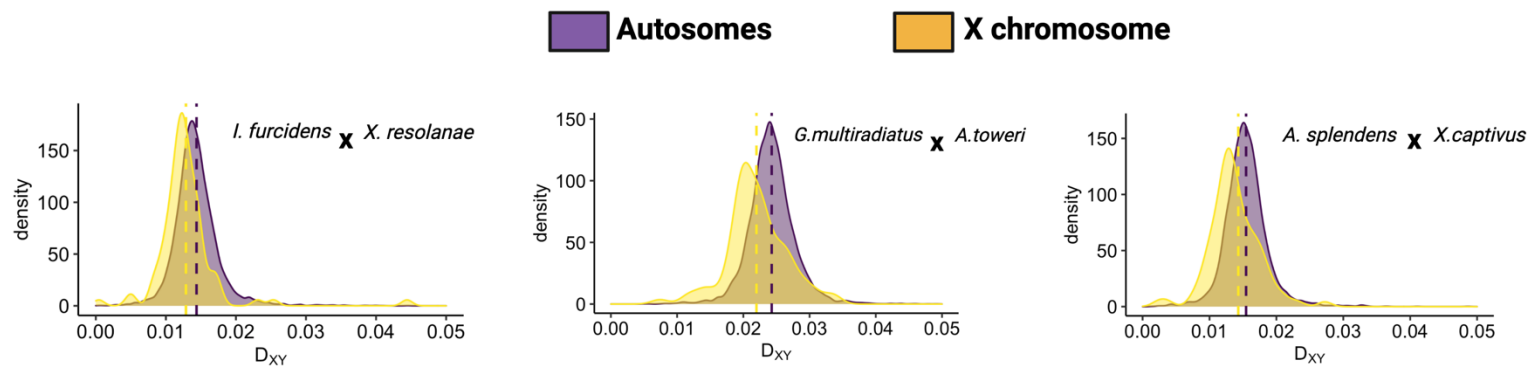
Figure 4: Differences in absolute divergence between autosomes and the X chromosome in three comparisons of closely related species in *Goodeinae*. Distributions of absolute nucleotide divergence (D_{XY}) calculated in non-overlapping 50kb windows. Coloured dotted lines show means for both autosomes and X chromosome for each pair.

Positive selection across *Goodeinae*

To understand what genes may have been associated with diversification across *Goodeinae*, we performed branch-site tests in all internal branches in the species tree. After multiple testing correction, we found 1482 genes (7.3% of all tested) that were under positive selection in at least one internal branch. Of those genes, 1293 (6.4%) of these are only under positive selection in a single branch, with 168 (0.8%) showing evidence of positive selection in two branches and 21 (0.1%) genes evolving under positive selection in three branches. Genes under selection across three internal branches include canonical immune response genes like interleukin-2 receptor subunit gamma (*IL-2RG*), cluster of differentiation 4 (*CD4*) and Lymphocyte activation gene 3 (*LAG3*). Gene ontology across all positively selected genes (1482) shows a significant enrichment (FDR < 0.05) of genes involved in DNA damage and repair, reproduction and cilium movement, among others (Figure 5; Supplementary table 3).

Additionally, we examined the number and identity of genes under positive selection in each internal branch. In the branch leading to the split of *G. multiradiatus* and *A. toweri,* we find 108 genes under positive selection. In subsequent internal branches leading to the split of *C. lateralis* and the ancestor of *I. furcidens*, *X. resolanae*, *G. atripinnis*, *A. splendens* and *X. captivus*, we find even fewer genes under positive selection (73 and 97, respectively). We detect many more genes under positive selection in longer internal branches leading to the split of *G. atripinnis* from *A. splendens* and *X. captivus* (431 genes), *A. splendens* and *X. captivus* (404 genes)*, and *I. furcidens* and *X. resolanae* (579 genes). These include genes presumed to be

important to the evolution of viviparity such as follicle stimulating hormone receptor (*FSHR*), androgen receptor (*AR*), vascular endothelial growth factor a (*VEGFA*), insulin growth factor 1 receptor (*IGF1R*) and insulin-degrading enzyme (*IDE*).

Identifying genes associated with evolutionary reversal of trophotaenia development in *Ataeniobius toweri.*

Based on the phylogenetic inference above and morphological evidence, the trophotaenia was likely lost in *A. toweri* following its split from *G. multiradiatus.* We therefore performed additional branch-site tests for positive selection on the terminal branch leading to *A. toweri*. Altogether, 1177 genes were found to be under positive selection here, after correcting for multiple testing. We found no enrichment of any gene ontology terms for genes under positive selection in *A. toweri.* We overlapped these with genes upregulated ($\log_2$ fold change > 2) in the trophotaenia compared to all other tissues of *G. multiradiatus.* We found 56 genes which showed significant upregulation in trophotaenia and under positive selection in *A. toweri*. Whilst gene ontology analysis showed no significant enrichment of any terms after multiple testing correction, we note that non-significant terms relate to fin and appendage development (adjusted p-value = 0.13), perhaps consistent with a potential role for these genes in trophotaenia developmental reversal (Supplementary table 4).

We also used Variant Effect Predictor (VEP) analysis to detect high-impact mutations that may have disruptive effects on protein structure or function in *A. toweri*. Specifically, we searched for homozygous high-impact mutations specific to *A. toweri* and overlapped these with the 56 trophotaenia-specific genes under positive selection

in *A. toweri*. Only 6 genes showed evidence high impact mutations, including a *cadherin related family member 5* (CDHR5), which modulates assembly of the intestinal brush border (Crawley *et al.*, 2014), and *Fraser extracellular matrix complex subunit 1* (FRAS1), which is involved in with epithelial-mesenchymal formation during embryonic development (Pavlakis, Chiotaki and Chalepakis, 2011).
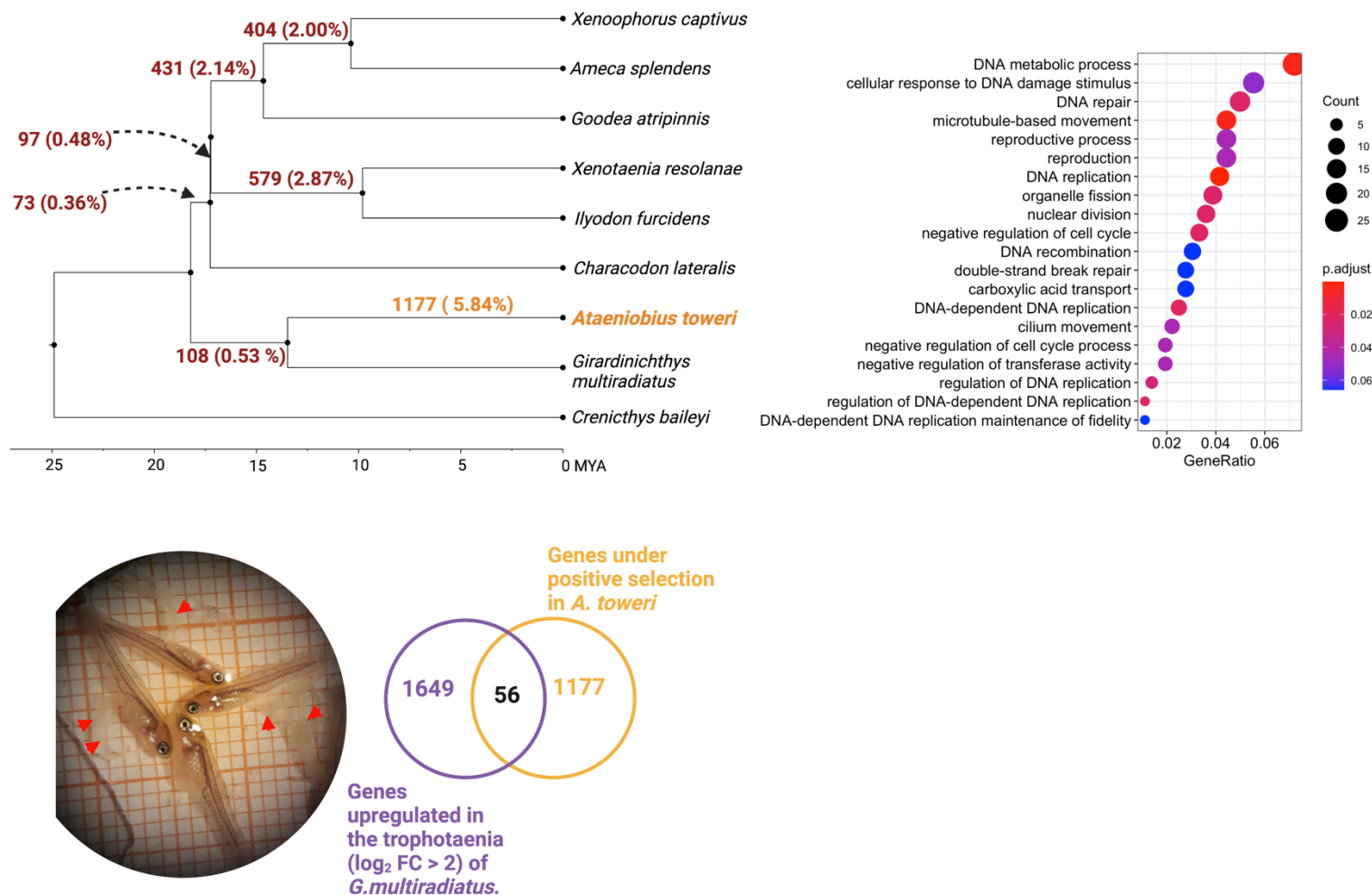
Figure 5: Testing for evidence of positive selection in branches across *Goodeinae*. A) Phylogeny of *Goodeinae* showing the number of genes and the percentage of genes under positive selection at tested branches. B) Gene ontology terms associated with genes under positive selection across *Goodeinae*. Colours denote adjusted p-value (for multiple testing) and size of data point denotes number of under positive selection. GeneRatio indicates the ratio of the number of genes under positive selection and the background gene set for each enrichment term. C) Pregnant *G. multiradiatus*. Red lines point towards exposed trophotaenia. Venn diagram shows the number of genes upregulated in the trophotaenia of *G. multiradiatus*, the number of genes under positive selection in *A. toweri*, and their overlap.

Discussion

Understanding the spatial context of speciation and identifying the factors contributing to this is an important challenge in evolutionary biology. Since demography and evolutionary history interact with evolutionary forces, inferring species relationships alongside estimates of demography is helping to resolve the relative importance of different evolutionary processes. In *Goodeinae,* the development of live-bearing leads to females provision offspring throughout development and males contribute fewer resources, setting the stage for sexual conflict, and there are varying levels of sexual dimorphism observed throughout the clade. Live-bearing has influenced speciation rates in the Cyprinodontiformes (Helmstetter *et al.*, 2016). Analyses of sexual dimorphism and speciation rate in the Goodeinae have provided ambiguous results (Ritchie *et al.*, 2005, 2007). Extrinsic factors, primarily recurrent volcanic activity throughout Central Mexico, have also been proposed as an important driver of speciation rates in *Goodeinae*. Our results provide detailed insight into the evolutionary history of *Goodeinae* and suggest a joint role for vicariant speciation with gene flow, and the importance of matrotrophic viviparity in the diversification of *Goodeinae*.

Overall, we found support for rapid divergence early on in the evolutionary history of *Goodeinae*, largely confirming previous phylogenetic analyses (Webb *et al.*, 2004; Ritchie *et al.*, 2005; Dominguez-Dominguez *et al.*, 2010). We find that diversification of *Goodeinae* began 18 MYA with the split of *Ataeniobius* and *Girardinichthys* and was followed by rapid divergence of *Characodon* and lineages including *Goodea*, *Xenotaenia*, *Xenoophorus*, *Ilyodon* and *Ameca* around 17 MYA. One noticeable

change in our tree is that *C. lateralis* is no longer basal, as seen in trees based on mtDNA (Webb *et al.,* 2004; Dominguez-Dominguez *et al.,* 2010) but within the main body of the *Goodeinae*. The position of *C. lateralis* is supported in both of our independent coalescent analyses (SVDquartets and ASTRAL) of whole-genome sequence data, despite underlying gene conflict. However, biogeographic analyses have supported an earlier branching off of the *Characodontini* (Dominguez-Dominguez *et al.,* 2010)*,* so this may need to be reconsidered in future.

These radiations coincide with the beginning of activity in the Trans-Mexican Volcanic Belt, and are consistent with previous phylogenetic analysis and fossil evidence (Webb *et al.,* 2004; Ferrari *et al.,* 2012). Our results disagree with previous analysis supporting basal divergence of *Characodon* in Durango, and instead support diversification of *Goodeinae* that likely occurred east to west with divergence of *Ataeniobius* and *Girardinichthys* (Webb *et al.,* 2004). Whilst we find complete statistical support for these early divergence events, we also find extensive underlying gene tree conflict, with only 12-13% of protein-coding gene trees supporting these branches and most gene trees supporting polyphyletic relationships. Here, gene tree conflict is expected to be generated by incomplete lineage sorting as a result of the rapid early diversification. Divergence of *Goodea*, *Xenotaenia*, *Xenoophorus*, *Ilyodon* and *Ameca* occurred 10-9 MYA, likely caused by another bout of volcanic activity beginning in the late Miocene (11 MYA)(Ferrari *et al.,* 2012). Altogether, the phylogeographic history of *Goodeinae* inferred here suggests early divergence of the group was characterised by repeated volcanic events occurring throughout the Miocene facilitating bursts of vicariant speciation, as opposed to potential

idiosyncratic, intrinsic factors specific to each divergence event (Webb *et al.*, 2004; Ritchie *et al.*, 2005; Crisp and Cook, 2007).

Whilst geological changes likely contributed to early diversification of *Goodeinae,* recurrent volcanic activity may have also facilitated secondary contact and gene flow. Recent work has shown that classic models of vicariant speciation may be relatively simplistic and obscure more complex speciation histories (Kopuchian *et al.*, 2020; Naka and Pil, 2020). Here, we examined evidence of introgression between species in *Goodeinae*, characterizing the degree of allele sharing genome-wide and at particular genomic regions. We found limited evidence of introgression between *A. toweri* and species within *Goodea*, *Xenotaenia*, *Xenoophorus*, *Ilyodon* and *Ameca,* which likely represent a single gene flow event between an ancestral lineage leading to the diversification of the aforementioned species and the ancestor of *A. toweri* and *G. multiradiatus*. This is partially supported by biogeographical evidence from multiple species of *Goodeinae* likely co-occurring with *A. toweri* in the tributaries of Rio Panuco at some point in their evolutionary history (Webb *et al.*, 2004). Some previous analyses of genetic differentiation between populations of dimorphic and monomorphic *Goodeinae* species found gene flow to be occurring, but less readily between populations of monomorphic species (Ritchie *et al.*, 2007). Recent genetic analysis of populations within *Goodea* and *Ilyodon* suggest connectivity and gene flow between species and populations is highly variable and mediated by complex topographical changes (Beltrán-López *et al.*, 2017, 2021). Whilst we detect moderate levels of ancient gene flow, it is not possible here to tell whether gene flow is adaptive and has contributed to phenotypic diversity and diversification, as observed in other systems

(Malinsky *et al.*, 2015; Meier *et al.*, 2017; Svardal *et al.*, 2020), but some of these introgressed genes were maintained.

Changes in contemporary and ancient population size throughout the evolutionary history of *Goodeinae* may be common, and several species of the group are endangered. We found dramatic concordant changes in effective population size (N_e) for almost all species in *Goodeinae* around 500 KYA, followed by concordant and steady reduction in effective population size in all species. This is largely consistent with distributions and the current conservation status of Goodeinae found here (Bailey, Garcia and Ritchie, 2007; Lyons *et al.*, 2019). However, it is important to note that absolute time estimates should be interpreted with caution given mutation rates from *Xiphophorus* that were used to scale parameter estimates in this study may be different, and mutation variation within Goodeids may vary. Alongside changes in contemporary changes in effective population sizes, we also find considerable reduction in population size following the split of *C. baileyi* and *C. lateralis,* suggesting possible vicariance as a result of volcanic activity resulted in uneven population sizes in ancestral lineages. These results are further supported by greater genetic divergence observed in autosomes compared to X chromosomes in three comparisons of closely related Goodeinae species. Ordinarily, genetic divergence is expected to be greater on the X chromosome than on the autosomes due to a disproportionate role of the X chromosome in driving reproductive isolation (Meisel and Connallon, 2013; Charlesworth, Campos and Jackson, 2018). However, marked differences in levels of genetic divergence between the X chromosome and autosomes may also be explained by differences in effective population size (Meisel

and Connallon, 2013; Wolf and Ellegren, 2017). For example, when absolute genetic divergence between closely-related species is higher on autosomes relative to the X chromosome, this may be best explained by ancestral population declines (Van Belleghem *et al.*, 2018). Given these predictions, our results are consistent with ancestral population contractions generating confounding signals of lower levels of genetic divergence on the X chromosome compared to the autosomes in closely related species in *Goodeinae*.

The evolution of viviparity could drive diversification rates in Goodeinae (Helmstetter *et al.,* 2016), and genes involved in placental viviparity could be expected to show rapid evolution (Crespi and Semeniuk, 2004). At a broad level, we find limited support for this idea. Gene ontology for all genes under positive selection in at least one internal branch in *Goodeinae* show enrichment for reproduction and cilium movement. The lack of a clear link indicated by gene ontology may be due to difficulty in detecting positive selection at short branches where incomplete lineage sorting is pervasive. We also surveyed genes under positive selection across multiple branches and found a number of canonical immune response genes including interleukin-2 receptor subunit gamma (*IL-2RG*), an essential constituent signalling component of many interleukin receptors (Russell *et al.*, 1993); cluster of differentiation 4 (*CD4*), a central component of adaptive immune response (Luckheeram *et al.*, 2012); and Lymphocyte activation gene 3 (*LAG3*), a gene important in modulating adaptive immune responses (Grosso *et al.*, 2007). Positive selection on immune response genes has been shown to be a relatively common feature of surveys of positive selection across mammals, birds, and invertebrates (Jiggins and Kim, 2007; Kosiol *et al.*, 2008; Shultz and Sackton, 2019).

We also find evidence of introgression of genes involved in adaptive immune response between *A. toweri* and an ancestral lineage leading to the divergence of *Ameca*, *Xenoophorus*, *Goodea*, *Xenotaenia* and *Ilyodon*, which, taken with evidence of positive selection, may indicate adaptive introgression of genes involved in MHC II pathway. Recurrent positive selection on immune-related genes, particularly genes involved in MHC II pathway, have also been associated with the evolution of viviparity in pipefish and seahorses (Roth *et al.*, 2020), and broadly across Cyprinodontiformes (Yusuf *et al.*, 2022).

In specific branches within *Goodeinae*, we identified genes known to be important in viviparity that were under positive selection. For example, we find positive selection in *ar* and *fshr* well known mediators of hormone synthesis and regulation in mammalian viviparity and reproduction. We also find evidence of positive selection on *igfr1*, but not *igfr2*, where parent-of-origin expression has been shown to mediate drastic changes to offspring size and positive selection has been observed in placental fish (DeChiara, Efstratiadis and Robertsen, 1990; DeChiara, Robertson and Efstratiadis, 1991; O'Neill *et al.*, 2007). Saldivar Lemus *et al.* (2017) show similar parent-of-origin methylation effects on *igf2* are also present in *G. multiradiatus*, suggesting that insulin-like growth factor genes mediating sexual conflict over offspring size may be under rapid evolution within *Goodeinae*.

Additionally, our phylogenetic analysis suggests an evolutionary reversal in the form of trophotaenia reversal occurred in *A. toweri*. In trying to understand the genes that may have been involved in this, we inferred positive selection in the terminal branch

leading to *A. toweri* and also identified genes that were expressed in the trophotaenia of *G. multiradiatus* (Du *et al.*, 2022) and looked for loss of function mutations in these. Whilst we found no significant enrichment of any gene ontology terms, we found some genes involved in fin and appendage development, suggesting a potential link between genes identified using our approach and trophotaenia reversal in *A. toweri*. However, genes implicated in trophotaenia reversal may also be under relaxed selection leading to elevated gene-wide divergence (Hiller *et al.*, 2012; Wertheim *et al.*, 2015).

Conclusion

Geological and climactic changes can have profound effects on demography, often resulting in complex speciation histories. The *Goodeinae* radiated rapidly during the Miocene and experienced constant fluctuations in population size throughout divergence, owing to periodic volcanism. We found evidence that ancient and contemporary expansions and crashes in effective population size have affected genetic variation genome-wide in closely related pairs. Volcanic activity and changes in species distribution likely facilitated ancient gene flow between ancestral lineages. We show that the same introgressed genetic variation is maintained in *A. toweri, and G. atripinnis, X. captivus* and *A. splendens,* suggesting a potential adaptive role for gene flow during the radiation of the group. Genes repeatedly under positive selection likely play a role in the evolution of matrotrophic viviparity, and we highlight specific genes that may be implicated in the reversal of trophotenium in *A. toweri*. This study highlights the importance of environmental factors, alongside internal factors such as viviparity and sexual selection, in driving species divergence, and lays the groundwork for understanding the evolution of viviparity in light of the complex demographic changes that have occurred in *Goodeinae*.

Acknowledgements

Data availability

Scripts used to generate the assemblies and annotation can be found here: https://github.com/peterthorpe5/fish_genome_assembly. Scripts used in the comparative analysis of convergent evolution can be found here: https://github.com/LeebanY/GoodeidsPhylogenomicsComplexDemography.

Genomes and whole-genome sequencing reads are available at the NCBI under BioProject PRXXXXXXXX.

References

Angelis, K. and Dos Reis, M. (2015) 'The impact of ancestral population size and incomplete lineage sorting on Bayesian estimation of species divergence times', *Current Zoology*, 61(5), pp. 874–885. doi: 10.1093/czoolo/61.5.874.

Van der Auwera, G. A. *et al.* (2013) 'From fastQ data to high-confidence variant calls: The genome analysis toolkit best practices pipeline', *Current Protocols in Bioinformatics*. doi: 10.1002/0471250953.bi1110s43.

Bailey, N. W., Garcia, C. M. and Ritchie, M. G. (2007) 'Beyond the point of no return? A comparison of genetic diversity in captive and wild populations of two nearly extinct species of Goodeid fish reveals that one is inbred in the wild', *Heredity*, 98(6). doi: 10.1038/sj.hdy.6800947.

Van Belleghem, S. M. *et al.* (2018) 'Patterns of Z chromosome divergence among Heliconius species highlight the importance of historical demography', *Molecular Ecology*, 27(19), pp. 3852–3872. doi: 10.1111/mec.14560.

Beltrán-López, R. G. *et al.* (2017) 'Phylogeny and taxonomy of the genus Ilyodon Eigenmann, 1907 (Teleostei: Goodeidae), based on mitochondrial and nuclear DNA sequences', *Journal of Zoological Systematics and Evolutionary Research*, 55(4), pp. 340–355. doi: 10.1111/jzs.12175.

Beltrán-López, R. G. *et al.* (2021) 'Genetic differentiation among populations of the blackfin goodea Goodea atripinnis (Cyprinodontiformes: Goodeidae): implications for its evolutionary history', *Journal of Fish Biology*, 98(5), pp. 1253–1266. doi: 10.1111/jfb.14654.

Benjamini, Y. and Hochberg, Y. (1995) 'Controlling the false discovery rate: a practical and powerful approach to multiple testing', *Journal of the Royal Statistical Society.*

Series B (Methodological). doi: 10.2307/2346101.

Bisazza, A. (1997) 'Sexual selection constrained by internal fertilization in the livebearing fish Xenotoca eiseni', *Animal Behaviour*, 54(6). doi: 10.1006/anbe.1997.0539.

Chang, C. C. *et al.* (2015) 'Second-generation PLINK: Rising to the challenge of larger and richer datasets', *GigaScience*, 4(1), pp. 1–16. doi: 10.1186/s13742-015-0047-8.

Charlesworth, B., Campos, J. L. and Jackson, B. C. (2018) 'Faster-X evolution: Theory and evidence from Drosophila', *Molecular Ecology*, 27(19), pp. 3753–3771. doi: 10.1111/mec.14534.

Chen, S. *et al.* (2018) 'Fastp: An ultra-fast all-in-one FASTQ preprocessor', in *Bioinformatics*. doi: 10.1093/bioinformatics/bty560.

Chifman, J. and Kubatko, L. (2014) 'Quartet inference from SNP data under the coalescent model', *Bioinformatics*. doi: 10.1093/bioinformatics/btu530.

Chifman, J. and Kubatko, L. (2015) 'Identifiability of the unrooted species tree topology under the coalescent model with time-reversible substitution processes, site-specific rate variation, and invariable sites', *Journal of Theoretical Biology*. doi: 10.1016/j.jtbi.2015.03.006.

Colin, R. and Eguiarte, L. E. (2016) 'Phylogeographic analyses and genetic structure illustrate the complex evolutionary history of Phragmites australis in Mexico', *American Journal of Botany*, 103(5). doi: 10.3732/ajb.1500399.

Craw, D. *et al.* (2016) 'Rapid biological speciation driven by tectonic evolution in New Zealand', *Nature Geoscience*, 9(2). doi: 10.1038/ngeo2618.

Crawley, S. W. *et al.* (2014) 'Intestinal brush border assembly driven by protocadherin-based intermicrovillar adhesion', *Cell*, 157(2). doi: 10.1016/j.cell.2014.01.067.

Crespi, B. and Semeniuk, C. (2004) 'Parent-offspring conflict in the evolution of vertebrate reproductive mode', *American Naturalist*, 163(5), pp. 635–653. doi: 10.1086/382734.

Crisp, M. D. and Cook, L. G. (2007) 'A congruent molecular signature of vicariance across multiple plant lineages', *Molecular Phylogenetics and Evolution*, 43(3), pp. 1106–1117. doi: 10.1016/j.ympev.2007.02.030.

DeChiara, T. M., Efstratiadis, A. and Robertsen, E. J. (1990) 'A growth-deficiency phenotype in heterozygous mice carrying an insulin-like growth factor II gene disrupted by targeting', *Nature*, 345(6270). doi: 10.1038/345078a0.

DeChiara, T. M., Robertson, E. J. and Efstratiadis, A. (1991) 'Parental imprinting of the mouse insulin-like growth factor II gene', *Cell*, 64(4). doi: 10.1016/0092-8674(91)90513-X.

Dias, M. S. *et al.* (2013) 'Natural fragmentation in river networks as a driver of speciation for freshwater fishes', *Ecography*, 36(6). doi: 10.1111/j.1600-0587.2012.07724.x.

Dominguez-Dominguez, O. *et al.* (2010) 'Historical Biogeography of the Goodeinae (Cyprinodontiforms)', *Viviparous fishes II*, (January), pp. 13–30.

Domínguez-Domínguez, O. *et al.* (2008) '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 Evolutionary Biology*, 8(1). doi: 10.1186/1471-2148-8-161.

Domínguez-Domínguez, O., Doadrio, I. and León, G. P. P. De (2006) 'Historical biogeography of some river basins in central Mexico evidenced by their goodeine freshwater fishes: A preliminary hypothesis using secondary Brooks parsimony

analysis', *Journal of Biogeography*, 33(8). doi: 10.1111/j.1365-2699.2006.01526.x.

Du, K. *et al.* (2022) ' Genome biology of the darkedged splitfin, Girardinichthys multiradiatus , and the evolution of sex chromosomes and placentation ', *Genome Research*, 32(3), pp. 583–594. doi: 10.1101/gr.275826.121.

Ebdon, S. *et al.* (2021) 'The Pleistocene species pump past its prime: Evidence from European butterfly sister species', *Molecular Ecology*, 30(14). doi: 10.1111/mec.15981.

Ferrari, L. *et al.* (2012) 'The dynamic history of the Trans-Mexican Volcanic Belt and the Mexico subduction zone', *Tectonophysics*. Elsevier B.V., 522–523, pp. 122–149. doi: 10.1016/j.tecto.2011.09.018.

Flouri, T. *et al.* (2018) 'Species tree inference with BPP using genomic sequences and the multispecies coalescent', *Molecular Biology and Evolution*, 35(10), pp. 2585–2593. doi: 10.1093/molbev/msy147.

Foster, K. L. and Piller, K. R. (2018) 'Disentangling the drivers of diversification in an imperiled group of freshwater fishes (Cyprinodontiformes: Goodeidae)', *BMC Evolutionary Biology*, 18(1). doi: 10.1186/s12862-018-1220-3.

Frazer, K. A. *et al.* (2007) 'A second generation human haplotype map of over 3.1 million SNPs.', *Nature*, 449(7164), pp. 851–61. doi: 10.1038/nature06258.

Garcia, C. M., Jimenez, G. and Contreras, B. (1994) 'Correlational evidence of a sexually-selected handicap', *Behavioral Ecology and Sociobiology*, 35(4). doi: 10.1007/BF00170706.

Garcia, C. M. and Lemus, Y. S. (2012) 'Foraging costs drive female resistance to a sensory trap', *Proceedings of the Royal Society B: Biological Sciences*, 279(1736). doi: 10.1098/rspb.2011.2611.

Garcia, C. M. and Ramirez, E. (2005) 'Evidence that sensory traps can evolve into honest signals', *Nature*, 434(7032). doi: 10.1038/nature03363.

Garcia, C. M., Saborío, E. and Berea, C. (1998) 'Does male-biased predation lead to male scarcity in viviparous fish?', *Journal of Fish Biology*, 53(SUPPL. A). doi: 10.1111/j.1095-8649.1998.tb01021.x.

Gavrilets, S. (2000) 'Rapid evolution of reproductive barriers driven by sexual conflict', *Nature*, 403(6772), pp. 886–889. doi: 10.1038/35002564.

Gavrilets, S. (2014) 'Is sexual conflict an "Engine of speciation"?', *Cold Spring Harbor Perspectives in Biology*, 6(12), pp. 1–14. doi: 10.1101/cshperspect.a017723.

Gavrilets, S. and Waxman, D. (2002) 'Sympatric speciation by sexual conflict', *Proceedings of the National Academy of Sciences of the United States of America*, 99(16), pp. 10533–10538. doi: 10.1073/pnas.152011499.

Green, R. E. *et al.* (2010) 'A draft sequence of the Neandertal genome.', *Science (New York, N.Y.)*, 328(5979), pp. 710–22. doi: 10.1126/science.1188021.

Grosso, J. F. *et al.* (2007) 'LAG-3 regulates CD8+ T cell accumulation and effector function in murine self- and tumor-tolerance systems', *Journal of Clinical Investigation*, 117(11), pp. 3383–3392. doi: 10.1172/JCI31184.

Harrison, R. G. and Larson, E. L. (2014) 'Hybridization, introgression, and the nature of species boundaries', in *Journal of Heredity*. doi: 10.1093/jhered/esu033.

Helmstetter, A. J. *et al.* (2016) 'Viviparity stimulates diversification in an order of fish', *Nature Communications*. doi: 10.1038/ncomms11271.

Hiller, M. *et al.* (2012) 'A "Forward Genomics" Approach Links Genotype to Phenotype using Independent Phenotypic Losses among Related Species', *Cell Reports*. The Authors, 2(4), pp. 817–823. doi: 10.1016/j.celrep.2012.08.032.

Hollenberg, F. and Wourms, J. P. (1995) 'Embryonic growth and maternal nutrient sources in goodeid fishes (Teleostei: Cyprinodontiformes)', *Journal of Experimental Zoology*, 271(5). doi: 10.1002/jez.1402710508.

Institute, B. (2019) 'Picard toolkit', *Broad Institute, GitHub repository*.

J. Spielman, S. (2018) 'phyphy: Python package for facilitating the execution and parsing of HyPhy standard analyses', *The Journal of Open Source Software*. doi: 10.21105/joss.00514.

Jiggins, F. M. and Kim, K. W. (2007) 'A screen for immunity genes evolving under positive selection in Drosophila', *Journal of Evolutionary Biology*, 20(3), pp. 965–970. doi: 10.1111/j.1420-9101.2007.01305.x.

Kadereit, J. W. and Abbott, R. J. (2021) 'Plant speciation in the Quaternary', *Plant Ecology and Diversity*. doi: 10.1080/17550874.2021.2012849.

Klicka, J. and Zink, R. M. (1997) 'The importance of recent ice ages in speciation: A failed paradigm', *Science*, 277(5332). doi: 10.1126/science.277.5332.1666.

Kopuchian, C. *et al.* (2020) 'A test of the riverine barrier hypothesis in the largest subtropical river basin in the Neotropics', *Molecular Ecology*, 29(12), pp. 2137–2149. doi: 10.1111/mec.15384.

Korunes, K. L. and Samuk, K. (2021) 'pixy: Unbiased estimation of nucleotide diversity and divergence in the presence of missing data', *Molecular Ecology Resources*, 21(4), pp. 1359–1368. doi: 10.1111/1755-0998.13326.

Kosiol, C. *et al.* (2008) 'Patterns of positive selection in six mammalian genomes', *PLoS Genetics*, 4(8). doi: 10.1371/journal.pgen.1000144.

Kumar, S. *et al.* (2012) 'Statistics and truth in phylogenomics', *Molecular Biology and Evolution*. doi: 10.1093/molbev/msr202.

Lawrence, M. *et al.* (2013) 'Software for Computing and Annotating Genomic Ranges', *PLoS Computational Biology*. doi: 10.1371/journal.pcbi.1003118.

Li, H. *et al.* (2009) 'The Sequence Alignment/Map format and SAMtools', *Bioinformatics*, 25(16), pp. 2078–2079. doi: 10.1093/bioinformatics/btp352.

Li, H. and Durbin, R. (2011) 'Inference of human population history from individual whole-genome sequences', *Nature*. doi: 10.1038/nature10231.

Luckheeram, R. V. *et al.* (2012) 'CD4 +T cells: Differentiation and functions', *Clinical and Developmental Immunology*, 2012. doi: 10.1155/2012/925135.

Lyons, J. *et al.* (2019) 'Distribution and current conservation status of the mexican goodeidae (actinopterygii, cyprinodontiformes)', *ZooKeys*, 2019(885), pp. 115–158. doi: 10.3897/zookeys.885.38152.

Macías Garcia, C. *et al.* (2012) 'Variation in sexual dimorphism and assortative mating do not predict genetic divergence in the sexually dimorphic Goodeid fish Girardinichthys multiradiatus', *Current Zoology*, 58(3). doi: 10.1093/czoolo/58.3.440.

Malinsky, M. *et al.* (2015) 'Genomic islands of speciation separate cichlid ecomorphs in an East African crater lake', *Science*, 350(6267), pp. 1493–1498. doi: 10.1126/science.aac9927.

Malinsky, M. *et al.* (2018) 'Whole-genome sequences of Malawi cichlids reveal multiple radiations interconnected by gene flow', *Nature Ecology and Evolution*, 2(12), pp. 1940–1955. doi: 10.1038/s41559-018-0717-x.

Malinsky, M., Matschiner, M. and Svardal, H. (2020) 'Dsuite - Fast D-statistics and related admixture evidence from VCF files', *Molecular Ecology Resources*. doi: 10.1111/1755-0998.13265.

Marques, D. A., Meier, J. I. and Seehausen, O. (2019) 'A Combinatorial View on

Speciation and Adaptive Radiation', *Trends in Ecology and Evolution*. Elsevier Ltd, 34(6), pp. 531–544. doi: 10.1016/j.tree.2019.02.008.

Martin, S. H., Davey, J. W. and Jiggins, C. D. (2015) 'Evaluating the use of ABBA-BABA statistics to locate introgressed loci', *Molecular Biology and Evolution*, 32(1), pp. 244–257. doi: 10.1093/molbev/msu269.

Mastretta-Yanes, A. *et al.* (2015) 'Biodiversity in the Mexican highlands and the interaction of geology, geography and climate within the Trans-Mexican Volcanic Belt', *Journal of Biogeography*, 42(9). doi: 10.1111/jbi.12546.

Mateos, M. (2005) 'Comparative phylogeography of livebearing fishes in the genera Poeciliopsis and Poecilia (Poeciliidae: Cyprinodontiformes) in central Mexico', *Journal of Biogeography*, 32(5). doi: 10.1111/j.1365-2699.2005.01236.x.

McGee, M. D. *et al.* (2020) 'The ecological and genomic basis of explosive adaptive radiation', *Nature*. doi: 10.1038/s41586-020-2652-7.

McLaren, W. *et al.* (2016) 'The Ensembl Variant Effect Predictor', *Genome Biology*. Genome Biology, 17(1), pp. 1–14. doi: 10.1186/s13059-016-0974-4.

Meier, J. I. *et al.* (2017) 'Ancient hybridization fuels rapid cichlid fish adaptive radiations', *Nature Communications*. Nature Publishing Group, 8(May 2016), pp. 1–11. doi: 10.1038/ncomms14363.

Meisel, R. P. and Connallon, T. (2013) 'The faster-X effect: Integrating theory and data', *Trends in Genetics*. Elsevier Ltd, 29(9), pp. 537–544. doi: 10.1016/j.tig.2013.05.009.

Morrone, J. J. (2010) 'Fundamental biogeographic patterns across the Mexican transition zone: An evolutionary approach', *Ecography*, 33(2). doi: 10.1111/j.1600-0587.2010.06266.x.

Moyaho, A., Garcia, C. M. and Ávila-Luna, E. (2004) 'Mate choice and visibility in the expression of a sexually dimorphic trait in a goodeid fish (Xenotoca variatus)', *Canadian Journal of Zoology*, 82(12). doi: 10.1139/z05-004.

Murrell, B. *et al.* (2012) 'Detecting individual sites subject to episodic diversifying selection', *PLoS Genetics*. doi: 10.1371/journal.pgen.1002764.

Musher, L. J. *et al.* (2022) 'River network rearrangements promote speciation in lowland Amazonian birds', *Science Advances*, 8(14). doi: 10.1126/sciadv.abn1099.

Naka, L. N. and Pil, M. W. (2020) 'Moving beyond the riverine barrier vicariant paradigm', *Molecular ecology*, 29(12), pp. 2129–2132. doi: 10.1111/mec.15465.

O'Neill, M. J. *et al.* (2007) 'Ancient and continuing Darwinian selection on insulin-like growth factor II in placental fishes', *Proceedings of the National Academy of Sciences of the United States of America*, 104(30). doi: 10.1073/pnas.0705048104.

Ortiz, E. M. 2019 (2019) *vcf2phylip v2.0: convert a VCF matrix into several matrix formats for phylogenetic analysis. (Zenodo).*, *GitHub*.

Pavlakis, E., Chiotaki, R. and Chalepakis, G. (2011) 'The role of Fras1/Frem proteins in the structure and function of basement membrane', *International Journal of Biochemistry and Cell Biology*. doi: 10.1016/j.biocel.2010.12.016.

Pérez-Rodríguez, R. *et al.* (2015) 'Comparative historical biogeography of three groups of Nearctic freshwater fishes across central Mexico', *Journal of Fish Biology*, 86(3). doi: 10.1111/jfb.12611.

Pollard, K. S. *et al.* (2010) 'Detection of nonneutral substitution rates on mammalian phylogenies', *Genome Research*. doi: 10.1101/gr.097857.109.

Quinlan, A. R. and Hall, I. M. (2010) 'BEDTools: A flexible suite of utilities for comparing genomic features', *Bioinformatics*. doi: 10.1093/bioinformatics/btq033.

Quiroz-Martínez, B. and Salgado-Maldonado, G. (2013) 'Patterns of Distribution of the Helminth Parasites of Freshwater Fishes of Mexico', *PLoS ONE*, 8(1). doi: 10.1371/journal.pone.0054787.

Ritchie, M. G. *et al.* (2005) 'Patterns of speciation in endemic Mexican Goodeid fish: Sexual conflict or early radiation?', *Journal of Evolutionary Biology*. doi: 10.1111/j.1420-9101.2005.00919.x.

Ritchie, M. G. *et al.* (2007) 'Sex and differentiation: Population genetic divergence and sexual dimorphism in Mexican goodeid fish', *Journal of Evolutionary Biology*, 20(5), pp. 2048–2055. doi: 10.1111/j.1420-9101.2007.01357.x.

Roth, O. *et al.* (2020) 'Evolution of male pregnancy associated with remodeling of canonical vertebrate immunity in seahorses and pipefishes', 117(17). doi: 10.1073/pnas.1916251117.

Russell, S. M. *et al.* (1993) 'Interleukin-2 receptor γ chain: A functional component of the interleukin-4 receptor', *Science*, 262(5141), pp. 1880–1883. doi: 10.1126/science.8266078.

Saldivar Lemus, Y. *et al.* (2017) 'Asymmetric paternal effect on offspring size linked to parent-of-origin expression of an insulin-like growth factor', *Ecology and Evolution*, 7(12), pp. 4465–4474. doi: 10.1002/ece3.3025.

Salichos, L. and Rokas, A. (2013) 'Inferring ancient divergences requires genes with strong phylogenetic signals', *Nature*. Nature Publishing Group, 497(7449), pp. 327–331. doi: 10.1038/nature12130.

Schmitt, T. (2007) 'Molecular biogeography of Europe: Pleistocene cycles and postglacial trends', *Frontiers in Zoology*. doi: 10.1186/1742-9994-4-11.

Schumer, M. *et al.* (2018) 'Natural selection interacts with recombination to shape the

evolution of hybrid genomes', *Science*. doi: 10.1126/science.aar3684.

Shultz, A. J. and Sackton, T. B. (2019) 'Immune genes are hotspots of shared positive selection across birds and mammals', *eLife*, 8, pp. 1–33. doi: 10.7554/eLife.41815.

Siepel, A. *et al.* (2005) 'Evolutionarily conserved elements in vertebrate, insect, worm, and yeast genomes', *Genome Research*. doi: 10.1101/gr.3715005.

Smith, M. D. *et al.* (2015) 'Less is more: An adaptive branch-site random effects model for efficient detection of episodic diversifying selection', *Molecular Biology and Evolution*. doi: 10.1093/molbev/msv022.

Suvorov, A. *et al.* (2022) 'Widespread introgression across a phylogeny of 155 Drosophila genomes', *Current Biology*, 32(1). doi: 10.1016/j.cub.2021.10.052.

Svardal, H. *et al.* (2020) 'Ancestral hybridization facilitated species diversification in the lake malawi cichlid fish adaptive radiation', *Molecular Biology and Evolution*, 37(4), pp. 1100–1113. doi: 10.1093/molbev/msz294.

Thom, G. *et al.* (2020) 'Quaternary climate changes as speciation drivers in the Amazon floodplains', *Science Advances*, 6(11). doi: 10.1126/sciadv.aax4718.

Vega-López, A. *et al.* (2007) 'The role of vitellogenin during gestation of Girardinichthys viviparus and Ameca splendens; two goodeid fish with matrotrophic viviparity', *Comparative Biochemistry and Physiology - A Molecular and Integrative Physiology*, 147(3), pp. 731–742. doi: 10.1016/j.cbpa.2006.10.039.

Webb, S. A. *et al.* (2004) 'Molecular phylogeny of the livebearing Goodeidae (Cyprinodontiformes)', *Molecular Phylogenetics and Evolution*. doi: 10.1016/S1055-7903(03)00257-4.

Wertheim, J. O. *et al.* (2015) 'RELAX: Detecting relaxed selection in a phylogenetic framework', *Molecular Biology and Evolution*. doi: 10.1093/molbev/msu400.

Wilgenbusch, J. C. and Swofford, D. (2003) ' Inferring Evolutionary Trees with PAUP
* ', *Current Protocols in Bioinformatics*. doi: 10.1002/0471250953.bi0604s00.

Wolf, J. B. W. and Ellegren, H. (2017) 'Making sense of genomic islands of differentiation in light of speciation', *Nature Reviews Genetics*. Nature Publishing Group, 18(2), pp. 87–100. doi: 10.1038/nrg.2016.133.

Wourms, J. P., Grove, B. D. and Lombardi, J. (1988) 'The maternal-embryonic relationship in viviparous fishes', *Fish Physiology*, 11. doi: 10.1016/S1546-5098(08)60213-7.

Yang, Z. and Zhu, T. (2018) 'Bayesian selection of misspecified models is overconfident and may cause spurious posterior probabilities for phylogenetic trees', *Proceedings of the National Academy of Sciences of the United States of America*. doi: 10.1073/pnas.1712673115.

Yu, G. *et al.* (2012) 'ClusterProfiler: An R package for comparing biological themes among gene clusters', *OMICS A Journal of Integrative Biology*. doi: 10.1089/omi.2011.0118.

Yusuf, L. H. *et al.* (2022) 'Genomic signatures associated with transitions to viviparity in Cyprinodontiformes', *bioRxiv*. Cold Spring Harbor Laboratory. doi: 10.1101/2022.05.25.493474.

Supplementary Figures
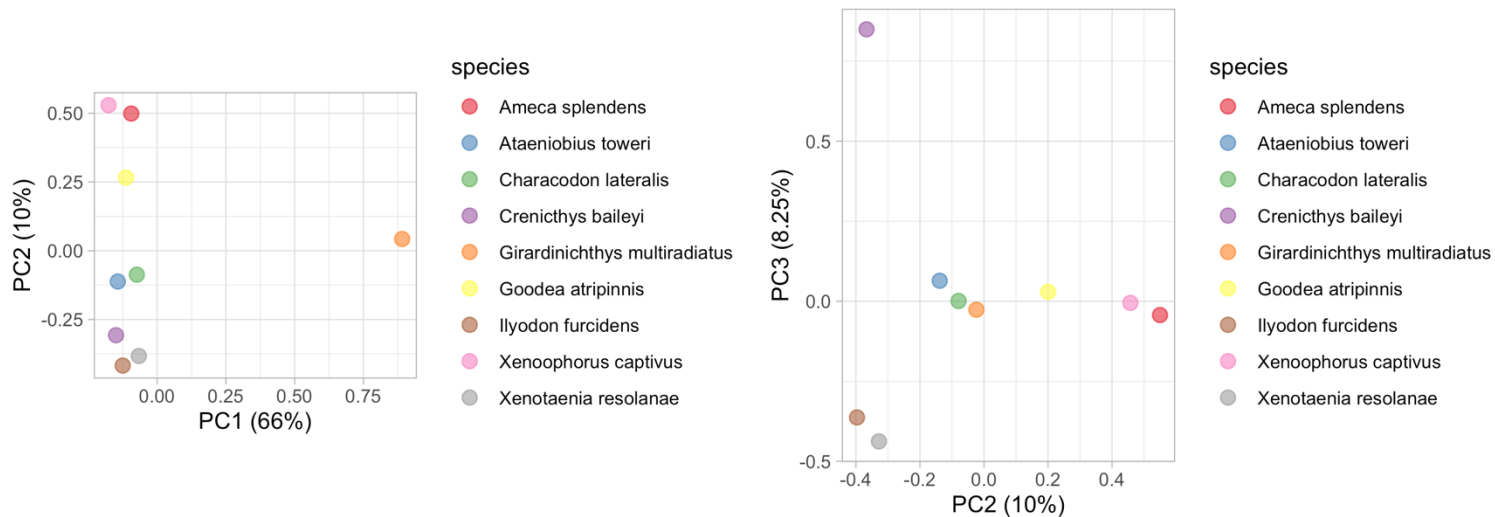
Supplementary Figure 1: Principal component analysis (PCA) of pruned whole-genome data for all sampled species within *Goodeidae*. Axes represent PC axes explaining most variation in the data. Coloured data points in PCA correspond to species shown in legend.
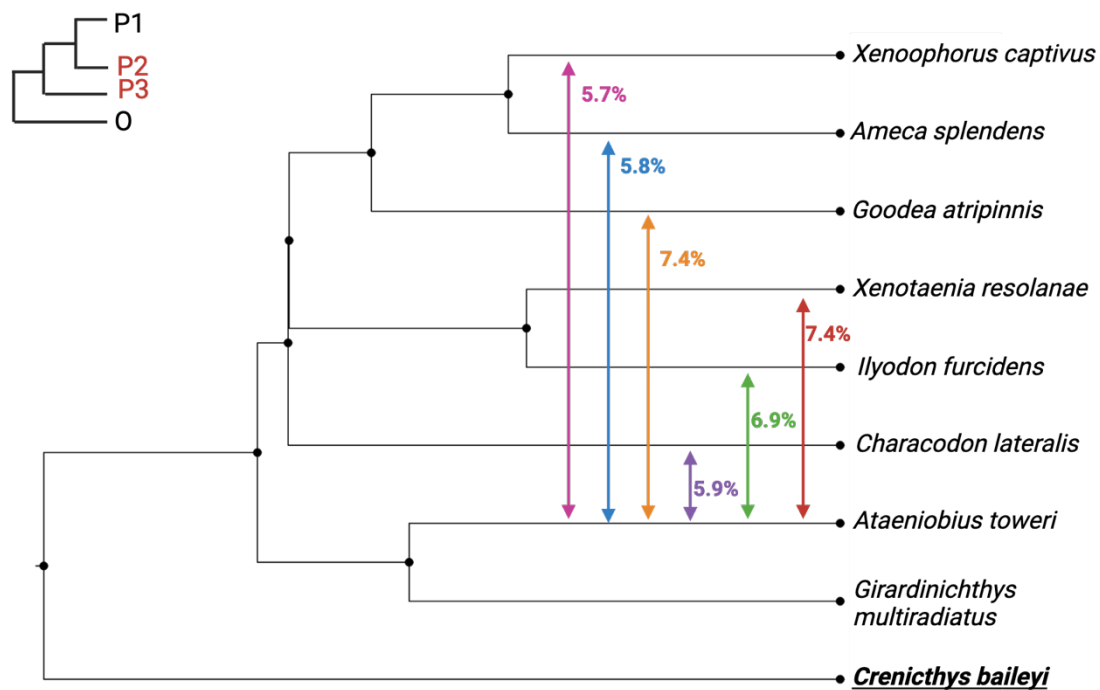
Supplementary Figure 2: Mean genome-wide admixture proportions (f_d) between *A. toweri (P2)* and other Goodeines (P3). *Crenicthys baileyi* used as outgroup species in all contrasts and *G. multiradiatus* was used an in-group species (P1). Phylogeny created using Figtree.