Quantitative trait loci concentrate in specific regions of the Mexican cavefish genome and reveal key candidate genes for cave-associated evolution

Jonathan Wiese[1], Emilie Richards[1], Johanna E. Kowalko[2], Suzanne E. McGaugh[1]

1. Department of Ecology, Evolution, and Behavior, University of Minnesota, St. Paul, MN
2. Department of Biological Sciences, Lehigh University, Bethlehem, PA

Corresponding author: Suzanne McGaugh, smcgaugh@umn.edu

Abstract

A major goal of modern biology is connecting phenotype with its underlying genetic basis. The Mexican cavefish (*Astyanax mexicanus*), a characin fish species comprised of a surface ecotype and a cave-derived ecotype, is well suited as a model to study the genetic mechanisms underlying adaptation to extreme environments. Here we map 206 previously published quantitative trait loci (QTL) for cave-derived traits in *A. mexicanus* to the newest version of the surface fish genome assembly, AstMex3. These analyses revealed that QTL cluster in the genome more than expected by chance, and this clustering is not explained by the distribution of genes in the genome. To investigate whether certain characteristics of the genome facilitate phenotypic evolution, we tested whether genomic characteristics associated with increased opportunities for mutation, such as highly mutagenic CpG sites, are reliable predictors of the sites of trait evolution but did not find any significant trends. Finally, we combined the QTL map with previously collected expression and selection data to identify 36 candidate genes that may underlie the repeated evolution of cave phenotypes, including *rgrb,* which is predicted to be involved in phototransduction. We found this gene has disrupted exons in all non-hybrid cave populations but intact reading frames in surface fish. Overall, our results suggest specific regions of the genome may play significant roles in driving adaptation to the cave environment in *Astyanax mexicanus* and demonstrate how this compiled dataset can facilitate our understanding of the genetic basis of repeated evolution in the Mexican cavefish.

Introduction

A major goal of modern evolutionary biology is connecting phenotypic evolution with its underlying genetic basis. Understanding the genetic basis of phenotypes can help reveal whether shared or different genetic mechanisms contribute to repeated trait evolution (Arendt and Reznick, 2008; Manceau, *et al.*, 2010; Martin and Orgogozo, 2013; Waters and McCulloch, 2021) and how often repeated evolution occurs through the same or divergent genetic variants (Courtier-Orgogozo, *et al.*, 2020; Lee and Coop, 2017; Lee and Coop, 2019). Further, while studies have revealed that loci contributing to trait evolution can cluster in specific genomic regions, the extent to which specific regions in the genome are commonly used in trait evolution and what features of these regions make them more likely to harbor genetic variants contributing to phenotypic evolution remains understudied (Storz, 2016; Woodhouse and Hufford, 2019; Wortel, *et al.*, 2023; Xie, *et al.*, 2019).

Linking genotypes to specific phenotypes remains the first step in any query of the repeatability of molecular evolution, and quantitative trait locus (QTL) mapping is a powerful technique for discovering statistical associations between genotypes and phenotypes that has been utilized since the 1980's (Paterson, *et al.*, 1988) for a wide variety of organisms and traits (Nuzhdin, Khazaeli and Curtsinger, 2005; Shaw, Parsons and Lesnick, 2007; Yano, *et al.*, 1997). QTL mapping is often the first step to understanding proximate mechanisms for how traits evolve, however, any single QTL study will underestimate the number of loci that influence a given trait and often overestimate their effect sizes (Beavis, 1994; Kearsey and Farquhar, 1998; Xu, 2003). Combining the results of several QTL studies for the same trait can identify loci that were missed by a single study and give a more comprehensive estimate of the number of loci underlying the trait (Peichel and Marques, 2017).

The Mexican cavefish, *Astyanax mexicanus*, is an ideal model system to investigate questions regarding the genetic basis of repeated adaptation to novel environments. *Astyanax mexicanus* is a species of freshwater characin fish that exists in two distinct ecotypes: a surface form found in streams across southern Texas and northeastern Mexico, and a cave-dwelling form comprised of populations in at least 35 different caves (Elliott, 2018; Espinasa, *et al.*, 2020; Miranda-Gamboa, *et al.*, 2023). Phylogeographic analyses suggest that cave populations arose from at least two distinct colonization events from two distinct surface ancestral lineages within the last 200,000 years, and since the initial colonization most cave populations have evolved relatively independently from one another, although there is evidence of gene flow between some cave populations (Bradic, *et al.*, 2012; Coghill, *et al.*, 2014; Herman, *et al.*, 2018; Moran, *et al.*, 2023). To varying degrees, cave populations have convergently evolved several morphological traits, most notably degenerated eyes and reduced pigmentation (Jeffery, 2001; Jeffery, 2020), as well as behavioral characteristics such as decreased sleep, schooling, and stress behaviors (Chin, *et al.*, 2019; Duboue, Keene and Borowsky, 2011; Kowalko, *et al.*, 2013). The most frequently studied cave populations include the Pachón and Tinaja cave populations, originated from caves located in the Sierra de El Abra region of Mexico, and the Molino cave population, from the nearby Sierra de la Guatemala (Jeffery, 2020). Importantly, fish from cave-dwelling *A. mexicanus* populations can interbreed with fish from surface populations (Şadoğlu, 1957; Şadoğlu, 1957). The ability to cross cave and surface-dwelling fish to generate viable hybrid offspring allows researchers to utilize QTL mapping to identify statistical associations between genotypes and phenotypes, facilitating the study of the genetic basis of phenotypic evolution in *A. mexicanus* (Borowsky and Wilkens, 2002; O'Quin and McGaugh, 2015; Wilkens, 1988).

In the last two decades, 20 QTL studies have been conducted to elucidate the genetic bases of traits in *A. mexicanus* (Table 1). The first QTL study in *A. mexicanus* was published by Borowsky and Wilkens (2002) who identified QTL for three quantitative traits: eye size, pigmentation, and condition factor (a measure of weight scaled by length). Intriguingly, this study identified two clusters of QTL, one consisting of overlapping QTL for condition factor and pigmentation and another consisting of overlapping QTL for condition factor and eye size. In their discussion, Borowsky and Wilkens (2002) highlighted the overlapping QTL and raise two possible explanations: (1) that the same genes underlie both traits due to pleiotropic effects, or (2) that different closely linked genes underlie the traits. A series of QTL studies by Protas and colleagues (Protas*, et al.*, 2007; Protas*, et al.*, 2008; Protas*, et al.*, 2006) corroborated the findings of Borowsky and Wilkens and showed that many QTL for cave-derived traits form clusters in the *A. mexicanus* genome. Since these initial studies, (O'Quin and McGaugh, 2015) mapped all published QTL through 2014 to the first draft of the *A. mexicanus* Pachón cavefish genome (McGaugh*, et al.*, 2014), and found 13 significant QTL clusters across eight linkage groups.

Since this 2015 analysis (O'Quin and McGaugh, 2015), the number of published QTL studies for *A. mexicanus* has nearly doubled, with dozens of new QTL for traits ranging from scleral ossification to heart regeneration (Table 1). Further, a new surface fish chromosome level genome assembly, AstMex3, was made available in July 2022. Mapping QTL to a more complete surface fish genome, instead of the cavefish genome, will yield more comprehensive results if important loci were missing from the fragmented cavefish assembly. The first draft cavefish genome was a scaffold level assembly, comprised of 10,735 scaffolds with a contig N50 of 14.7 kb and a scaffold N50 of 1.775 Mb (McGaugh*, et al.*, 2014). In contrast, the AstMex3 surface fish assembly is chromosome level, containing 25

chromosomes, 109 unplaced scaffolds, and greatly improved N50 values of 47.1 Mb for contigs and 51.9 Mb for scaffolds (Warren*, et al.*, 2024). Thus, linkage between traits as well as additional candidate genes may be uncovered since the new genome is much more contiguous.

Here we present an updated analysis of the distribution of QTL across the *A. mexicanus* genome, incorporating most published QTL studies in this species to date. Genetic markers used in prior studies were re-mapped to the AstMex3 genome to create an integrated genomic map, which was utilized to anchor QTL intervals to the new genome. Putative QTL clusters were then identified. To understand if certain mutational biases predispose regions to contribute to phenotypic evolution, we tested several genomic variables for correlations with the observed QTL. Next, while acknowledging the bias of QTL to identify regions of moderate to large effect (Rockman, 2012), we analyzed the percent variance explained (PVE) values of cave-derived QTL for different phenotypic classes. Lastly, we demonstrate the utility of our compiled QTL analysis for the evolutionary genetics community by cross-referencing QTL for cave-derived phenotypes with other sources of genomic data to identify candidate genes for repeated trait evolution in cavefish. In sum, this study contributes to our understanding of adaptation to an extreme environment and provides a resource for expedited candidate gene discovery and understanding potential links between multiple phenotypes.

Methods

Literature review

Quantitative trait locus (QTL) mapping studies in *Astyanax mexicanus* were identified by querying the online databases PubMed and Web of Science with the following search terms: "astyanax" AND "qtl"; "cavefish" AND "qtl"; "astyanax" AND "quantitative trait loc*"; "cavefish" AND "quantitative trait loc*". The resulting publications were manually filtered to exclude studies that did not map de novo QTL, as well as studies for which genetic marker sequences were not available. Only studies published before 6/6/22 were included in the analysis.

Mapping QTL to the AstMex3 genome

To map previously published QTL to the most recent version of the *A. mexicanus* genome, AstMex3, all available marker sequences used in prior QTL studies were collected and aligned to the AstMex3 surface genome with BLAST+ version 2.8.1 (Camacho*, et al.*, 2009) (Table S1, S2, S3). For each marker sequence, the best quality BLAST alignment was identified by filtering according to the following criteria: 1. Highest bitscore, 2. Smallest e-value, 3. Highest percent identity. We refer to this match as the "best hit." Markers with multiple BLAST hits that could not be ranked using these criteria were discarded from the analysis. Additionally, markers for which the best BLAST hit mapped to a chromosome discordant with other markers on its original linkage group were discarded. The best BLAST hits for each of the remaining markers were collated into an integrated genomic map, anchored to the AstMex3 surface genome (Table S4). Three genes (*igf1*, *shhb* also known as *twhh,* and *pax6*) that were used as markers in previous QTL studies but did not have

sequences included in our marker database were added manually to the integrated map using the known genomic coordinates given in the AstMex3 genome assembly.

The integrated genomic map was used to identify the location of each QTL in the AstMex3 surface genome assembly. We caution that this is a rough estimate of QTL location, as we were comparing across over 20 years of studies, many of which report different levels of information. Nevertheless, compiling QTL peaks, even if rough approximation across studies, may be useful for future candidate gene identification and to understand traits with potentially co-localizing genomic regions. When available, the genetic markers designating the 95% confidence interval for each QTL (as defined by the study in which the QTL was originally mapped) were used to designate the section of the genome in which loci underlying the trait of interest may be found (see Table S5 for available metadata for each QTL). If a confidence interval was not available, only the marker representing the QTL peak was used to estimate the location of the QTL in the genome. If the markers designating both the 95% confidence interval and the peak of a given QTL were absent from the integrated genomic map, then that QTL was discarded from the rest of the analysis. We manually assigned the phenotypic category of each QTL that was mapped to the AstMex3 genome into one of eight categories (Eye, Pigment, Skeletal, Non-Eye Sensory, Metabolic, Behavior, Cardiac, Other morphological (this category consisted of traits that did not fit clearly into one of the other seven categories, e.g. length); Tables S5, S6), and the QTL intervals and peaks mapped to each chromosome were visualized using the karyoploteR package (v. 1.24.0; Gel and Serra, 2017).

Identifying QTL clusters

To analyze QTL distribution across the genome, QTL were filtered to include only the QTL that were mapped to intervals (not just peaks) in the integrated genomic map. Further, since QTL that were coarsely mapped to vast stretches of the genome may mask any potential trends in genomic QTL distribution, QTL with total interval lengths that were statistical outliers were excluded according to the outlier formula: Upper limit = Upper quartile + (1.5*IQR), where the upper quartile is the 75th percentile of QTL lengths and the IQR is the interquartile range of QTL lengths (75th percentile – 25th percentile). Next, the 25 chromosomes of the AstMex3 surface genome were divided into 118 non-overlapping 10 Mb windows using the tileGenome function from the GenomicRanges R package (v. 1.49.0; Lawrence*, et al.*, 2013). A window size of 10 Mb was chosen because it approximates the typical length of a QTL interval in the filtered dataset (mean length = 11.4 Mb, median length = 9.64 Mb). Windows were "centered" by finding the largest whole number of 10 Mb windows that could fit in each chromosome, subtracting the total length of these windows from the total length of the chromosome to find the "remaining" chromosome length, and shifting the starting position of the first window to a value equal to one half of the remaining length.

The number of QTL overlapping each window was then identified using the bed_intersect tool in the R package valr (v. 0.6.6; Riemondy*, et al.*, 2017). For the purposes of counting QTL within each window, all QTL for phenotypes relating to eye size (e.g., eye size, lens size, pupil area, etc.) were counted as one "distinct" QTL, as were all QTL for a melanophore phenotype (e.g., eye melanophore, dorsal melanophore, anal melanophore, etc.). The number

of QTL overlapping each window after adjusting for these similar phenotypic traits is hereafter referred to as the number of "distinct QTL".

To test whether the number of observed distinct QTL in each window matches a distribution expected by chance, a Monte-Carlo Chi-square test was conducted with the xmonte function from the XNomial R package (v 1.0.4; Engels, 2015). Two expected distributions were tested: (1) a null expectation that QTL are distributed uniformly across the genome, and (2) the expectation that QTL are distributed across the genome proportional to protein-coding gene density. Following the tests, 10 Mb windows with high standardized residual values (>3) against either the uniform distribution or the gene density distribution were grouped together and designated as putative QTL clusters (Table S7) (as in Peichel and Marques, 2017). The observed distribution of distinct QTL across the *A. mexicanus* genome relative to the expected number of QTL under each of the two null expectations was visualized using the karyoploteR package (v. 1.24.0; Gel and Serra, 2017).

Testing potential molecular factors underlying QTL distribution

Mutational biases or opportunities may predispose regions to contribute to phenotypic evolution (Storz*, et al.*, 2019; Wortel*, et al.*, 2023; Xie*, et al.*, 2019). To investigate whether any molecular factors may explain the distribution of cave-associated QTL across the *A. mexicanus* genome, data was generated for each of the 118 10Mb windows of the AstMex3 surface genome assembly (Table S8). For each factor listed below, we conducted a Wilcoxon rank-sum test in R to test for differences between the set of 10 Mb windows that were designated as putative QTL clusters and the set of 10 Mb windows that contain no mapped

QTL. Six 10 Mb windows across four different chromosomes were identified as putative QTL clusters due to having significantly more QTL than expected under either the uniform null distribution or the gene density null distribution.

The following variables were investigated:

- *Repetitive DNA*: Repetitive DNA was quantified for each window as the proportion of all bases in the window that were reported as an "N" base in the hard-masked version of the AstMex3 surface fish genome, which was downloaded from the Ensembl Rapid Release FTP site (Cunningham*, et al.*, 2022). This method quantifies repetitive DNA in a broad sense, as it captures any sequence considered "repetitive" in the original masking procedures of the AstMex3 genome assembly.
- *GC content*: GC content was calculated for each window as the proportion of all bases in the window that are a "C" or "G" in the unmasked version of the AstMex3 surface fish genome.
- *CpG sites*: CpG sites were quantified for each window as the proportion of all dinucleotides in the window that are "CG" in the unmasked genome. Dinucleotide distributions were calculated using the oligonucleotideFrequency function from the Biostrings R package (v 2.66.0; Pages*, et al.*, 2016).
- *TGTGTG repeats*: Similarly, TGTGTG repeats are linked to replication fragile sites (Xie*, et al.*, 2019) and were quantified for each window as the proportion of all oligonucleotides of length 6 that are "TGTGTG" or "CACACA" in the unmasked genome with the oligonucleotideFrequency function.
- *Gene number*: The total number of genes, both protein-coding and non-protein-coding, overlapping each window was obtained with the bed_intersect function in the

valr package (Riemondy, *et al.*, 2017), with the genome annotation file from AstMex3 (Ensembl rapid release, GTF format; (Cunningham, *et al.*, 2022)) providing the genomic coordinates of each gene. Gene counts were also obtained separately for genes of the following biotypes as reported in the GTF file: protein-coding, miRNA, snRNA, lncRNA.

- *Gene length*: The average and median protein-coding gene length in bp was calculated for each window using the start and end points given in the GTF file (i.e., the lengths included the entire length of the gene, not just the coding regions).
- *Transcription factor gene number*: Changes in the coding sequences of transcription factors (TF) may be involved in the evolution of gene regulatory networks, potentially facilitating phenotypic change (Cheatle Jarvela and Hinman, 2015). The list of all genes annotated as transcription factors in *A. mexicanus* was downloaded from the Animal TFDB 4.0 (Shen, *et al.*, 2023). The bed_intersect function was then used to identify the number of TF genes that overlap with each window.
- *Recombination rate*: Linkage maps were generated for Pachón x Surface, Tinaja x Surface, and Molino x Surface F2 crosses as part of a separate ongoing study. For each linkage map, a recombination rate between every consecutive pair of markers was estimated by dividing the distance between the markers in centimorgans (cM) by the distance in Mb, giving a recombination rate in units of cM/Mb. These values were then used to find the average recombination rate across each 10 Mb window using a weighted average, where the weight applied to each individual rate was equal to the number of bp by which the interval overlapped the 10 Mb window.

Analyzing effect sizes of cave-derived QTL

The distributions of reported PVE values for all QTL, as well as QTL in the Eye, Pigment, Skeletal, Non-Eye Sensory, Metabolic, and Behavior categories separately, were visualized as a histogram using ggplot2. These distributions were tested for skewness (a measure of the symmetry of a distribution) and kurtosis (a measure of whether the distribution is light-tailed or heavy-tailed) using the respective functions in the R package moments (v. 0.14.1; Komsta and Novomestky, 2015).

Identifying candidate genes underlying repeated cave-derived trait evolution

Lastly, we demonstrate the utility of our compiled QTL analysis for identifying candidate genes for future functional analysis. By cross-referencing QTL with other sources of genomic data, we identified candidate genes for cave-derived phenotypes in cavefish. First, the set of all genes overlapping with a QTL was obtained using the valr package in R (v. 0.6.6; Riemondy, *et al.*, 2017). Second, the set of all genes exhibiting evidence of selective sweeps in three independent cave populations (Pachón, Tinaja, and Molino) yet no evidence of selection in two surface populations (i.e., Río Choy, Rascón) was defined as in Moran, *et al.* (2023). The Ensembl gene IDs published by Moran, *et al.* (2023) were given for version 2 of the *A. mexicanus* genome. These IDs were converted to the version 3 equivalents using the AstMex3 homology table available on the Ensembl Rapid Release FTP site. Third, RNAseq data, originally collected by (Mack, *et al.*, 2021) from whole body samples of larval Pachón, Tinaja, and Molino cavefish as well as Río Choy surface fish, were reanalyzed. Raw reads were downloaded from the NCBI Sequence Read Archive (accession numbers given in Table S9 (Leinonen, *et al.*, 2010)), aligned to the AstMex3 surface fish genome using STAR (v. 2.7.1; Dobin, *et al.*, 2013), and aligned reads were counted using HTSeq (v. 0.11.0; Anders,

Pyl and Huber, 2015). Read counts were analyzed for differential expression between Pachón and Río Choy, Tinaja and Río Choy, and Molino and Río Choy, all at time 0 in the original study, using the DESeq2 package in R (v. 1.38.1; Love, Huber and Anders, 2014) (Tables S10, S11, S12), and differentially expressed genes were identified using a cutoff of a Benjamini-Hochberg adjusted p-value < 0.1. The set of genes that were differentially expressed in all three cave populations relative to the surface population, and for which the change in expression followed the same pattern in each population (i.e., upregulated in all three caves or downregulated in all three caves), were designated as genes with shared differential expression. Candidate genes for cave-derived phenotypes (Table S13) were identified by finding the subset of genes that 1) Have undergone shared selective sweeps in three separate cave populations and not in surface populations as defined by Moran*, et al.* (2023), 2) Have shared differential expression between three cave populations relative to a surface population, and 3) Are found within a QTL for cave-derived traits.

While the whole-body RNAseq reads from (Mack*, et al.*, 2021) provided a starting point to identify candidate genes, we were also interested in investigating gene expression specifically in eye tissue, as eye loss is a major evolutionary change that has occurred in cavefish. To this end, we accessed RNAseq reads that were originally collected from eye tissue of developing embryos at 54 hpf (Table S9) (Gore*, et al.*, 2018), and re-mapped these reads to the AstMex3 genome assembly as described above for the whole-body RNAseq experiment. Aligned reads were counted using HTSeq and analyzed for differential expression using DESeq2 as described above for the whole-body RNAseq experiment. From these data, genes were identified that were significantly differentially expressed in the developing eye tissue of Pachón cavefish relative to surface fish (Table S14).

Results

A map of cave-derived QTL across the Astyanax mexicanus genome

The first goal of this study was to provide a comprehensive map of previously published QTL for cave-derived traits in *A. mexicanus* to a new high-quality genome assembly, AstMex3, to facilitate analyses of trait evolution in this species. To this end, a total of 9023 genetic marker sequences from prior QTL studies (Tables 1, S1, S2) were collected and aligned to the AstMex3 surface fish genome using BLASTn (Table S3). Two-hundred markers did not have any BLAST hits, 126 markers had a BLAST hit that mapped to a chromosome that was discordant with other markers on the original linkage group, and 96 markers did not have a best BLAST hit that could be unambiguously resolved. The final integrated genomic map consisted of the best BLAST hit for the remaining 8486 genetic markers (representing 94% of the collected markers) (Table S4).

This genomic map was used to associate 206 previously published QTL for cave-derived traits to the AstMex3 surface genome (Tables S4, S5). For 168 of the QTL, genetic markers marking the ends of the LOD confidence interval were part of the integrated genomic map, and the genomic range spanning the beginning of the first marker to the end of the second marker was denoted as a QTL interval. For the remaining 38 QTL, only a genetic marker associated with the peak of the LOD curve was present in the integrated genomic map, and the genomic range of this individual marker was denoted as a QTL peak. Table S6 summarizes the phenotypic category of each QTL that was mapped to the AstMex3 genome in this analysis. The resulting QTL map shows a general tendency for QTL for cave-derived traits to overlap with each other across the *A. mexicanus* genome (Figure S1, Tables S4, S7).

QTLs cluster in the Astyanax mexicanus genome

To test whether QTLs cluster in the *A. mexicanus* genome more than expected by chance, the observed QTL distribution across 10 Mb genomic windows was tested for fit against two expected distributions. First, we found that the observed QTL distribution does not fit the expected distribution under a null model which assumes a uniform distribution of QTL across the genome (chi-square = 192.9, p = 0.00002; Figure 1). Second, the observed QTL distribution does not fit the expected distribution under a model in which QTL density is proportional to protein-coding gene density (chi-square = 216.1, $p < 0.00001$; Figure 1). Together, these results support the hypothesis that QTL for cave-derived traits cluster in the genome. Six 10 Mb windows across four different chromosomes were identified as putative QTL clusters due to having significantly more QTL than expected under either distribution (Table S7).

Distribution of putative QTL clusters is not explained by molecular factors

Increased mutation rates lead to increased opportunity for evolution (Bromham, 2009; Kimura and Ohta, 1971). To investigate whether the distribution of putative QTL clusters across the *A. mexicanus* genome are associated with known characteristics of the genome that are potentially associated with increased opportunity for mutation, 13 variables including recombination rate, gene length, gene number, transcription factor number, CpG sites, TGTGTG repeats, and repetitive DNA were tested for significant differences between putative QTL clusters (n=6) compared to 10 Mb windows that contain no known QTL (n=22; Table S8). None of the variables tested were significantly different between the two groups

(Wilcoxon rank-sum test, $p>0.05$ for all tests), suggesting that none of these factors are associated with cave-derived evolution at this coarse scale (Figure S2).

Distribution of effect sizes depends on type of trait

While acknowledging that we will miss many small and medium effect loci, the distribution of PVE values reported for all previously mapped cave-derived QTL in *A. mexicanus* was compiled to understand more about the distributions of effect size across different trait classifications (Figure 2). The distribution is extremely positively skewed (Skewness coefficient = 2.04) and light-tailed (kurtosis coefficient = 8.02). This indicates that most alleles detected to be associated with cave-derived phenotypes explain a relatively small proportion of phenotypic variation, with just a few outlying mutations that contribute more substantially to cave-derived phenotypes.

This trend generally holds true when cave-derived QTL are divided into trait categories – all distributions are positively skewed and light-tailed, but there is variation in the degrees of skewness and kurtosis. The most highly positive skewed trait categories are Eye (skewness coefficient = 2.08) and Skeletal (skewness coefficient = 2.55), suggesting that a strong majority of mutations affecting these traits are of small effect with a few mutations having larger effects. The least strongly positively skewed trait categories are Non-Eye Sensory (skewness coefficient = 0.77) and Behavior (skewness coefficient = 0.73), suggesting that the distribution of effect sizes for these trait categories are somewhat larger than other traits. Notably, traits in the Pigment category have some of the largest effect sizes across all traits

(Figure 2), and this may be because multiple studies mapped albinism which is a Mendelian trait.

A robust list of candidate genes underlying repeated cave-derived trait evolution

By utilizing multiple avenues of genomic data, the pool of candidate genes underlying traits of interest can be narrowed. We identified 36 candidate genes (Table S13) underlying the repeated evolution of cave-derived traits in *A. mexicanus* by finding the subset of protein-coding genes that are present in at least one cave-derived QTL, have evidence of repeated selective sweeps across three populations of cavefish (Moran*, et al.*, 2023), and are differentially expressed between cavefish and surface fish in all of the same three cave populations (Mack*, et al.*, 2021) (Tables S10-S12). In addition, 14 of the 36 candidate genes (Table S13) also exhibit down regulation in Pachón cavefish eyes relative to surface fish at 54hpf (Tables S13, S14). Notably, none of the 36 candidate genes were significantly upregulated in Pachón cavefish eyes at 54hpf relative to surface fish.

Discussion

The genetic architecture underlying adaptive evolution is a central topic in evolutionary biology. QTL studies can identify regions of the genome important for trait evolution and, thus, are a powerful method for investigating genetic mechanisms of adaptation. In this study, 206 quantitative trait loci (QTL) from 16 previously published studies (Carlson*, et al.*, 2018; Gross, Borowsky and Tabin, 2009; Gross, Krutzler and Carlson, 2014; Kowalko*, et al.*, 2013; Kowalko*, et al.*, 2013; Lyon*, et al.*, 2017; O'Quin*, et al.*, 2015; O'Quin*, et al.*, 2013; Powers, Boggs and Gross, 2020; Protas*, et al.*, 2007; Protas*, et al.*, 2008; Protas*, et al.*, 2006;

Stockdale, *et al.*, 2018; Warren, *et al.*, 2021; Yoshizawa, *et al.*, 2015; Yoshizawa, *et al.*, 2012) were mapped to the newly published *A. mexicanus* surface genome assembly, AstMex3. Collectively, these QTL represent a variety of morphological, physiological, and behavioral traits of interest that differentiate Mexican cavefish from their surface-dwelling counterparts. Our study organizes the disparate maps onto one high quality genome assembly, allowing for easy comparison and analysis of QTL features across studies. Further, this work is an improvement on a previous QTL review in *A. mexicanus* because it maps QTL to a more contiguous surface fish genome, capturing loci in QTL intervals that may have been fragmented or may not have been present in the cavefish genome (O'Quin and McGaugh, 2015).

This integrated QTL map showed that the observed distribution of QTL for cave-derived traits across the *A. mexicanus* genome is non-uniform and is not proportional to protein-coding gene density. As a result of this analysis, six putative QTL clusters that contain more unique QTL than expected by chance were identified on four of the 25 chromosomes in the *A. mexicanus* genome. The most striking of these clusters occurs on Chromosome 1 from base pairs ~2Mb-42Mb. This 40 Mb region contains QTL for 10 distinct phenotypic traits including eye size, vibration attraction behavior, relative condition, tastebuds, melanophore, and several others, suggesting that Chromosome 1 is particularly enriched for cave-derived trait evolution in *A. mexicanus*. This result is very similar to the conclusion of a similar analysis in three-spined stickleback, which found that five out of 21 chromosomes contained more QTL than expected by chance (Peichel and Marques, 2017), suggesting that specific genomic regions drive adaptation in both systems.

Clustering of QTL for seemingly distinct traits, such as what is observed on Chromosome 1, could be due to pleiotropic loci that influence multiple phenotypic traits of interest (O'Quin and McGaugh, 2015; Peichel and Marques, 2017). Pleiotropy could help explain adaptation to the cave environment that has independently occurred in multiple populations of this species, as selective pressure on one trait could drive coordinated change in other traits and bring about extensive phenotypic change (Jeffery, 2005; Yamamoto*, et al.*, 2009). The putative QTL clusters identified in this study provide an ideal starting point to search for genes that may have pleiotropic effects (Jeffery, 2010; Protas*, et al.*, 2007; Yoshizawa, Kelly and Jeffery, 2013). One such pleiotropic gene, *oca2*, has already been identified in the literature and is known to affect both albinism and sleep duration (Klaassen*, et al.*, 2018; O'Gorman*, et al.*, 2021; Protas*, et al.*, 2006). Additionally, tightly linked genes could underlie the traits implicated in QTL clusters (Borowsky, 2013). Tight linkage of locally adapted alleles is predicted under theoretical scenarios of adaptation in the face of gene flow (Yeaman, Aeschbacher and Bürger, 2016; Yeaman and Whitlock, 2011) and appears to be a common theme as more evolutionary models are empirically interrogated (Wilder*, et al.*, 2020; Yeaman, 2022). Indeed, loci associated with suites of traits in other systems have been found to be in tight physical linkage or genomic inversions (Hess*, et al.*, 2020; Noor*, et al.*, 2001; Wellenreuther and Bernatchez, 2018; Wilder*, et al.*, 2020) and future work will explore the potential for inversions to contribute to ecotype differences in this system. Notably, these two mechanisms for the clustering of phenotypes in the genome, pleiotropy and linkage of locally adapted alleles, are not mutually exclusive and can work synergistically (Archambeault*, et al.*, 2020; Ferris*, et al.*, 2017) and could also work independently, but it is beyond the scope of this manuscript to distinguish if one or both of these mechanisms are impacting the statistical clustering of QTL.

Functional interrogation, for example with CRISPR-Cas9 gene editing (Klaassen, *et al.*, 2018), of genes contained in QTL clusters also represents a promising next step toward understanding the extent pleiotropy and linkage may be involved in shaping traits in this system. For instance, if a CRISPR knockout of a single gene is found to affect multiple traits of interest, as was the case for *oca2*, then this experimental result would further support a role for pleiotropy (O'Gorman, *et al.*, 2021). Mutating tightly linked genes within a QTL cluster and finding multiple genes that impact the traits of interest would support roles for linkage.

More broadly, the integrated QTL map is useful in analyzing how the distribution of cave-derived QTL correlates with other factors on a genomic scale. Certain molecular attributes may influence which regions of the genome are more likely to mutate and contribute to phenotypic evolution, including recombination rate (Halldorsson, *et al.*, 2019; Lercher and Hurst, 2002) abundance of CpG sites (Storz, *et al.*, 2019), TGTGTG repeats (Xie, *et al.*, 2019), and gene length (Mei, *et al.*, 2018; Moran, *et al.*, 2023; Woodhouse and Hufford, 2019). If supported, these hypotheses suggest mutational biases predispose certain regions of the genome to phenotypic evolution. In our study, however, none of the 13 variables that were tested were significantly correlated with the distribution of QTL on a genomic scale. While these variables exhibit no significant correlations to QTL clusters, the corresponding QTL intervals in the integrated map presented here are quite large and overlap several hundred genes. Thus, only a subsection of the QTL region may contribute in a meaningful way to the phenotype and the scale at which we examined the correlations may be too broad to accurately address correlations between genomic features and cave-derived evolution.

The collection of all previously published QTL into one dataset facilitates the analysis of effect sizes of cave-derived QTL across a range of traits. While most adaptative walks occur with predominantly small effect mutations, large-effect mutations may be utilized when responding to large environmental shifts (Orr, 2006), especially when, as in the cavefish system, migration with non-adapted populations occurs (Yeaman and Whitlock, 2011). To assess whether large-effect mutations may play an outsized role in the evolution of certain cave-derived phenotypes, the PVE reported for each QTL was recorded for all studies identified in the literature review. The QTL with the highest PVE value is a QTL for albinism, which is a Mendelian trait attributed to *oca2* (Protas*, et al.*, 2006), and pigmentation related traits seem to have several large-effect outliers compared to other trait classes. The phenomenon of a single gene explaining a large amount of variation in a trait, like *oca2* does for albinism, is the exception, similar to the QTL review of stickleback fish (Peichel and Marques, 2017). Additionally, QTL mapping experiments tend to overestimate PVE (Beavis, 1994) and be biased in detecting mostly larger-effect alleles (Rockman, 2012), therefore, some of the outlying QTL with high PVE values may not have a true effect size as great as their position in the distribution suggests. A final caveat is that the bias in overestimating PVE values increases with decreasing sample size, so comparing PVE effects of studies with different sample sizes is inexact (Beavis, 1994; Xu, 2003). Despite the bias in detecting large effect alleles, the shape of the distribution of PVE values across several traits and studies suggests that adaptation is generally attributable to many mutations of small effect rather than few mutations of large effect in *A. mexicanus*.

Likely most important for evolutionary geneticists, the integrated QTL map presented in this study provides an ideal starting point for identifying candidate genes underlying cave-derived traits. As proof of principle, here we utilized three sources of data – QTL, differential

expression, and population genomics – to curate a list of just 36 genes that are likely to be involved in the repeated evolution of cave-derived traits in *A. mexicanus* (Table S13). One exciting candidate gene is *rgrb*, which encodes retinal G protein-coupled receptor b. The protein encoded by this gene is predicted to be an opsin and the gene is widely expressed across zebrafish tissues (Davies*, et al.*, 2015). We find that *rgrb* maps to a QTL for retinal thickness (specifically the outer plexiform layer (OPL)), and that *rgrb* is downregulated in cavefish populations relative to surface populations in whole body samples from 30dpf fry. Furthermore, *rgrb* is downregulated in developing eye tissue of 54 hpf Pachón cavefish relative to surface fish (Gore*, et al.*, 2018). As with all candidates on our list, *rgrb* exhibits evidence of selective sweeps across three cave populations, suggesting that *rgrb* is a target of adaptation to the cave environment. Interestingly, exon 4 and exon 5 are nearly completely deleted across multiple cave populations (Lineage 1: Escondido, Molino, Jineo; Lineage 2: Jos, Monticellos, Pachón, Palma Seca, Tinaja, Yerbaniz), while reading frames are intact for all surface populations (see supplementary alignment). Together, these lines of evidence strongly suggest that *rgrb* has been a target of repeated selection in *A. mexicanus*, and that exon deletions and downregulation of *rgrb* may be involved in convergent abnormal retinal development in *A. mexicanus* cavefish (Emam*, et al.*, 2020).

Other intriguing candidate genes from our list span a variety of traits. For instance, we find that the retinoic acid receptor *stra6* is downregulated in cavefish and falls under a QTL for maxillary bone length, which tends to be longer in cavefish than surface fish (Atukorala*, et al.*, 2013). Retinoic acid signaling is known to play a role in skull and tooth development in fish (Draut, Liebenstein and Begemann, 2019), so it is possible that *stra6* downregulation may influence maxillary bone morphology in *A. mexicanus* through this avenue. As another example, the rab-GTPase *rab8a* has been implicated in the maintenance of glucose homeostasis through its regulation of glucose transporters (Sun*, et al.*, 2010), which could

connect to our finding that *rab8a* is upregulated in cavefish and is under a QTL for relative condition, a measure of observed vs predicted weight in which cavefish tend to be higher than surface fish (Protas*, et al.*, 2008). *Rab8a* is also associated with a QTL for peduncle depth. Many of the candidate genes on our list are associated with QTL for more than one phenotype, raising the possibility of pleiotropic effects like what has been observed in *oca2* (O'Gorman*, et al.*, 2021). While we have highlighted just a few examples, they demonstrate the utility of the data presented here in guiding future functional characterization.

In summary, the results presented in this study suggest that QTL contributing to adaptation to the cave environment in populations of *A. mexicanus* cluster in a statistically non-random pattern throughout the genome. While there are no molecular factors that reliably predict the regions of cave-derived trait evolution on a genomic scale, candidate genes for repeated trait evolution represent a promising avenue to study repeated targets in adaptation to extreme environments. These results contribute to the goal of understanding adaptation to extreme environments on a genomic scale and will serve as a useful resource to guide future studies on the genetic basis of cave-derived trait evolution in *A. mexicanus*.

Data availability

All data and analyses are provided in the supplemental materials.

Acknowledgements

We thank the Minnesota Supercomputing Institute without which this work would not be possible. We thank Jeff Gralnick, Yaniv Brandvain, and the McGaugh lab for comments on earlier drafts of this manuscript.

Funding

This work was supported by NSF 2316784/2316783 and R35 GM138345. This work was funded by the EDGE award NSF 1923372 to E. Duboue, J. Kowalko, and S. McGaugh which supported J. Wiese during the preparation of this work. J.W. was also supported by the University of Minnesota URS program.

Figure 1. Observed distribution of the number of distinct cave-associated QTL in each 10 Mb window of the *Astyanax mexicanus* genome (cyan histogram) compared to the expected distribution under a null model in which the QTL distribution is uniform across the genome (black line), and a model in which the QTL distribution is proportional to protein-coding gene density (orange line). Asterisks denote specific 10 Mb windows which contain more QTL than expected under either of the models (standardized residual > 3).

Figure 2. Distribution of reported percent variance explained (PVE) values for previously published cave-associated QTL in *Astyanax mexicanus* by trait category. Skewness coefficients indicate the degree of asymmetry in the distribution; in this case, all distributions are positively skewed, indicating that there are more mutations of below average effect size than of above average effect size. Kurtosis coefficients indicate the degree to which the distribution is light-tailed relative to a normal distribution, (i.e., higher values of kurtosis mean that outlying mutations with high PVE are more frequent).

References Cited

ANDERS S, PYL PT, HUBER W. 2015 HTSeq—a Python framework to work with high-throughput sequencing data. Bioinformatics. 31(2):166-169.

ARCHAMBEAULT SL, BÄRTSCHI LR, MERMINOD AD, PEICHEL CL. 2020 Adaptation via pleiotropy and linkage: association mapping reveals a complex genetic architecture within the stickleback Eda locus. Evolution letters. 4(4):282-301.

ARENDT J, REZNICK D. 2008 Convergence and parallelism reconsidered: what have we learned about the genetics of adaptation? Trends in Ecology & Evolution. 23(1):26-32.

ATUKORALA AD, HAMMER C, DUFTON M, FRANZ-ODENDAAL TA. 2013 Adaptive evolution of the lower jaw dentition in Mexican tetra (*Astyanax mexicanus*). EvoDevo. 4(1):28.

. The power and deceit of QTL experiments: lessons from comparative QTL studies; 1994.

BOROWSKY R. 2013 Eye regression in blind *Astyanax* cavefish may facilitate the evolution of an adaptive behavior and its sensory receptors. BMC Biol. 11(1):81.

BOROWSKY R, WILKENS H. 2002 Mapping a cave fish genome: Polygenic systems and regressive evolution. Journal of Heredity. 93:19–21.

BRADIC M, BEERLI P, GARCIA-DE LEON FJ, ESQUIVEL-BOBADILLA S, BOROWSKY RL. 2012 Gene flow and population structure in the Mexican blind cavefish complex (*Astyanax mexicanus*). BMC evolutionary biology. 12(1):9.

BROMHAM L. 2009 Why do species vary in their rate of molecular evolution? Biology letters. 5(3):401-404.

CAMACHO C, COULOURIS G, AVAGYAN V, MA N, PAPADOPOULOS J, BEALER K, MADDEN TL. 2009 BLAST+: architecture and applications. BMC bioinformatics. 10:1-9.

CARLSON BM, KLINGLER IB, MEYER BJ, GROSS JB. 2018 Genetic analysis reveals candidate genes for activity QTL in the blind Mexican tetra. PeerJ. 6:e5189.

CHEATLE JARVELA AM, HINMAN VF. 2015 Evolution of transcription factor function as a mechanism for changing metazoan developmental gene regulatory networks. EvoDevo. 6:1-11.

CHIN JSR, ALBERT LT, LOOMIS CL, KEENE AC, DUBOUÉ ER. 2019 Behavioral approaches to studying innate stress in Zebrafish. Journal of Visualized Experiments. 2019.

COGHILL LM, DARRIN HULSEY C, CHAVES-CAMPOS J, GARCIA DE LEON FJ, JOHNSON SG. 2014 Next generation phylogeography of cave and surface *Astyanax mexicanus*. Mol Phylogenet Evol. 79:368-374.

COURTIER-ORGOGOZO V, ARNOULT L, PRIGENT SR, WILTGEN S, MARTIN A. 2020 Gephebase, a database of genotype–phenotype relationships for natural and domesticated variation in Eukaryotes. Nucleic Acids Research. 48(D1):D696-D703.

CUNNINGHAM F, ALLEN JE, ALLEN J, ALVAREZ-JARRETA J, AMODE MR, ARMEAN IM, AUSTINE-ORIMOLOYE O, AZOV AG, BARNES I, BENNETT R. 2022 Ensembl 2022. Nucleic Acids Research. 50(D1):D988-D995.

DAVIES WI, TAMAI TK, ZHENG L, FU JK, RIHEL J, FOSTER RG, WHITMORE D, HANKINS MW. 2015 An extended family of novel vertebrate photopigments is widely expressed and displays a diversity of function. Genome research. 25(11):1666-1679.

DOBIN A, DAVIS CA, SCHLESINGER F, DRENKOW J, ZALESKI C, JHA S, BATUT P, CHAISSON M, GINGERAS TR. 2013 STAR: ultrafast universal RNA-seq aligner. Bioinformatics. 29(1):15-21.

DRAUT H, LIEBENSTEIN T, BEGEMANN G. 2019 New insights into the control of cell fate choices and differentiation by retinoic acid in cranial, axial and caudal structures. Biomolecules. 9(12):860.

DUBOUE ER, KEENE AC, BOROWSKY RL. 2011 Evolutionary convergence on sleep loss in cavefish populations. Curr Biol. 21(8):671-676.

ELLIOTT WR. 2018 *The Astyanax caves of Mexico: Cavefishes of Tamaulipas, San Luis Potosí, and Guerrero*. Association for Mexican Cave Studies.

EMAM A, YOFFE M, CARDONA H, SOARES D. 2020 Retinal morphology in Astyanax mexicanus during eye degeneration. Journal of Comparative Neurology. 528(9):1523-1534.
ENGELS B. 2015 XNomial: exact goodness-of-fit test for multinomial data with fixed probabilities. R package version. 1(4).
ESPINASA L, ORNELAS-GARCÍA CP, LEGENDRE L, RÉTAUX S, BEST A, GAMBOA-MIRANDA R, ESPINOSA-PÉREZ H, SPROUSE P. 2020 Discovery of two new Astyanax cavefish localities leads to further understanding of the species biogeography. Diversity. 12(10):368.
FERRIS KG, BARNETT LL, BLACKMAN BK, WILLIS JH. 2017 The genetic architecture of local adaptation and reproductive isolation in sympatry within the Mimulus guttatus species complex. Molecular ecology. 26(1):208-224.
GEL B, SERRA E. 2017 karyoploteR: an R/Bioconductor package to plot customizable genomes displaying arbitrary data. Bioinformatics. 33(19):3088-3090.
GORE AV, TOMINS KA, IBEN J, MA L, CASTRANOVA D, DAVIS AE, PARKHURST A, JEFFERY WR, WEINSTEIN BM. 2018 An epigenetic mechanism for cavefish eye degeneration. Nat Ecol Evol. 2(7):1155-1160.
GROSS JB, BOROWSKY R, TABIN CJ. 2009 A novel role for *Mc1r* in the parallel evolution of depigmentation in independent populations of the cavefish *Astyanax mexicanus*. PLoS genetics. 5(1):e1000326-e1000326.
GROSS JB, KRUTZLER AJ, CARLSON BM. 2014 Complex craniofacial changes in blind cave-dwelling fish are mediated by genetically symmetric and asymmetric loci. Genetics. 196(4):1303-1319.
HALLDORSSON BV, PALSSON G, STEFANSSON OA, JONSSON H, HARDARSON MT, EGGERTSSON HP, GUNNARSSON B, ODDSSON A, HALLDORSSON GH, ZINK F. 2019 Characterizing mutagenic effects of recombination through a sequence-level genetic map. Science. 363(6425):eaau1043.
HERMAN A, BRANDVAIN Y, WEAGLEY J, JEFFERY WR, KEENE AC, KONO TJY, BILANDZIJA H, BOROWSKY R, ESPINASA L, O'QUIN K, ORNELAS-GARCIA CP, YOSHIZAWA M, CARLSON B, MALDONADO E, GROSS JB, CARTWRIGHT RA, ROHNER N, WARREN WC, MCGAUGH SE. 2018 The role of gene flow in rapid and repeated evolution of cave-related traits in Mexican tetra, Astyanax mexicanus. Mol Ecol. 27(22):4397-4416.
HESS JE, SMITH JJ, TIMOSHEVSKAYA N, BAKER C, CAUDILL CC, GRAVES D, KEEFER ML, KINZIGER AP, MOSER ML, PORTER LL. 2020 Genomic islands of divergence infer a phenotypic landscape in Pacific lamprey. Molecular ecology. 29(20):3841-3856.
JEFFERY WR. 2001 Cavefish as a model system in evolutionary developmental biology. Developmental Biology. 231(1):1-12.
JEFFERY WR. 2005 Adaptive evolution of eye degeneration in the Mexican blind cavefish. Journal of Heredity. 96(3):185-196.
JEFFERY WR. 2010 Pleiotropy and eye degeneration in cavefish. Heredity. 105(5):495-496.
JEFFERY WR. 2020 Astyanax surface and cave fish morphs. EvoDevo. 11.
KEARSEY MJ, FARQUHAR A. 1998 QTL analysis in plants; where are we now? Heredity. 80(2):137-142.
KIMURA M, OHTA T. 1971 On the rate of molecular evolution. Journal of Molecular Evolution. 1:1-17.
KLAASSEN H, WANG Y, ADAMSKI K, ROHNER N, KOWALKO JE. 2018 CRISPR mutagenesis confirms the role of oca2 in melanin pigmentation in Astyanax mexicanus. Developmental Biology. 441(2):313-318.
KOMSTA L, NOVOMESTKY F. 2015. moments: CRAN.
KOWALKO JE, ROHNER N, LINDEN TA, ROMPANI SB, WARREN WC, BOROWSKY R, TABIN CJ, JEFFERY WR, YOSHIZAWA M. 2013 Convergence in feeding posture occurs through different genetic loci in independently evolved cave populations of. Proc Natl Acad Sci U S A. 110(42):16933-16938.
KOWALKO JE, ROHNER N, ROMPANI SB, PETERSON BK, LINDEN TA, YOSHIZAWA M, KAY EH, WEBER J, HOEKSTRA HE, JEFFERY WR. 2013 Loss of schooling behavior in

cavefish through sight-dependent and sight-independent mechanisms. Current Biology. 23(19):1874-1883.
LAWRENCE M, HUBER W, PAGÈS H, ABOYOUN P, CARLSON M, GENTLEMAN R, MORGAN MT, CAREY VJ. 2013 Software for computing and annotating genomic ranges. PLoS computational biology. 9(8):e1003118.
LEE KM, COOP G. 2017 Distinguishing Among Modes of Convergent Adaptation Using Population Genomic Data. Genetics. 207(4):1591-1619.
LEE KM, COOP G. 2019 Population genomics perspectives on convergent adaptation. Philos Trans R Soc Lond B Biol Sci. 374(1777):20180236.
LEINONEN R, SUGAWARA H, SHUMWAY M, COLLABORATION OBOTINSD. 2010 The Sequence Read Archive. Nucleic Acids Research. 39(suppl_1):D19-D21.
LERCHER MJ, HURST LD. 2002 Human SNP variability and mutation rate are higher in regions of high recombination. Trends in genetics. 18(7):337-340.
LOVE MI, HUBER W, ANDERS S. 2014 Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome biology. 15(12):550.
LYON A, POWERS AK, GROSS JB, O'QUIN KE. 2017 Two–three loci control scleral ossicle formation via epistasis in the cavefish Astyanax mexicanus. PloS one. 12(2):e0171061.
MACK KL, JAGGARD JB, PERSONS JL, ROBACK EY, PASSOW CN, STANHOPE BA, FERRUFINO E, TSUCHIYA D, SMITH SE, SLAUGHTER BD, KOWALKO J, ROHNER N, KEENE AC, MCGAUGH SE. 2021 Repeated evolution of circadian clock dysregulation in cavefish populations. PLoS Genet. 17(7):e1009642.
MANCEAU M, DOMINGUES VS, LINNEN CR, ROSENBLUM EB, HOEKSTRA HE. 2010 Convergence in pigmentation at multiple levels: mutations, genes and function. Philosophical Transactions of the Royal Society B: Biological Sciences. 365:2439-2450.
MARTIN A, ORGOGOZO V. 2013 The loci of repeated evolution: a catalog of genetic hotspots of phenotypic variation. Evolution. 67(5):1235-1250.
MCGAUGH SE, GROSS JB, AKEN B, BLIN M, BOROWSKY R, CHALOPIN D, HINAUX H, JEFFERY WR, KEENE A, MA L, MINX P, MURPHY D, O'QUIN KE, RETAUX S, ROHNER N, SEARLE SM, STAHL BA, TABIN C, VOLFF JN, YOSHIZAWA M, WARREN WC. 2014 The cavefish genome reveals candidate genes for eye loss. Nat Commun. 5:5307.
MEI W, STETTER MG, GATES DJ, STITZER MC, ROSS-IBARRA J. 2018 Adaptation in plant genomes: Bigger is different. In: Vol. 105), pp. 16-19. Wiley Online Library.
MIRANDA-GAMBOA R, ESPINASA L, VERDE-RAMÍREZ MD, HERNÁNDEZ-LOZANO J, LACAILLE JL, ESPINASA M, ORNELAS-GARCFA CP. 2023 A new cave population of from Northern Sierra de El Abra, Tamaulipas, Mexico. Subterranean Biology. 45:95-117.
MORAN RL, RICHARDS EJ, ORNELAS-GARCIA CP, GROSS JB, DONNY A, WIESE J, KEENE AC, KOWALKO JE, ROHNER N, MCGAUGH SE. 2023 Selection-driven trait loss in independently evolved cavefish populations. Nat Commun. 14(1):2557.
NOOR MA, GRAMS KL, BERTUCCI LA, REILAND J. 2001 Chromosomal inversions and the reproductive isolation of species. Proceedings of the National Academy of Sciences. 98(21):12084-12088.
NUZHDIN SV, KHAZAELI AA, CURTSINGER JW. 2005 Survival analysis of life span quantitative trait loci in Drosophila melanogaster. Genetics. 170(2):719-731.
O'GORMAN M, THAKUR S, IMRIE G, MORAN RL, CHOY S, SIFUENTES-ROMERO I, BILANDZIJA H, RENNER KJ, DUBOUE E, ROHNER N, MCGAUGH SE, KEENE AC, KOWALKO JE. 2021 Pleiotropic function of the *oca2* gene underlies the evolution of sleep loss and albinism in cavefish. Curr Biol. 31(16):3694-3701 e3694.
O'QUIN K, MCGAUGH SE. 2015 The genetic bases of troglomorphy in *Astyanax*: How far we have come and where do we go from here? In: *Biology and Evolution of the Mexican Cavefish*. (A. Keene MY, S.E. McGaugh, ed). Elsevier.
O'QUIN KE, DOSHI P, LYON A, HOENEMEYER E, YOSHIZAWA M, JEFFERY WR. 2015 Complex Evolutionary and Genetic Patterns Characterize the Loss of Scleral Ossification in the Blind Cavefish Astyanax mexicanus. PloS one. 10(12):e0142208.
O'QUIN KE, YOSHIZAWA M, DOSHI P, JEFFERY WR. 2013 Quantitative genetic analysis of retinal degeneration in the blind cavefish *Astyanax mexicanus*. PloS one. 8(2):e57281.

ORR HA. 2006 The distribution of fitness effects among beneficial mutations in Fisher's geometric model of adaptation. Journal of theoretical biology. 238(2):279-285.
PAGES H, ABOYOUN P, GENTLEMAN R, DEBROY S. 2016 Biostrings: String objects representing biological sequences, and matching algorithms. R package version. 2(0):10.18129.
PATERSON AH, LANDER ES, HEWITT JD, PETERSON S, LINCOLN SE, TANKSLEY SD. 1988 Resolution of quantitative traits into Mendelian factors by using a complete linkage map of restriction fragment length polymorphisms. Nature. 335(6192):721-726.
PEICHEL CL, MARQUES DA. 2017 The genetic and molecular architecture of phenotypic diversity in sticklebacks. Philosophical Transactions of the Royal Society B: Biological Sciences. 372(1713):20150486.
POWERS AK, BOGGS TE, GROSS JB. 2020 An Asymmetric Genetic Signal Associated with Mechanosensory Expansion in Cave-Adapted Fish. Symmetry. 12(12):1951.
PROTAS M, CONRAD M, GROSS JB, TABIN C, BOROWSKY R. 2007 Regressive evolution in the Mexican cave tetra, Astyanax mexicanus. Curr Biol. 17(5):452-454.
PROTAS M, TABANSKY I, CONRAD M, GROSS JB, VIDAL O, TABIN CJ, BOROWSKY R. 2008 Multi-trait evolution in a cave fish, *Astyanax mexicanus*. Evol Dev. 10(2):196-209.
PROTAS ME, HERSEY C, KOCHANEK D, ZHOU Y, WILKENS H, JEFFERY WR, ZON LI, BOROWSKY R, TABIN CJ. 2006 Genetic analysis of cavefish reveals molecular convergence in the evolution of albinism. Nat Genet. 38(1):107-111.
RIEMONDY KA, SHERIDAN RM, GILLEN A, YU Y, BENNETT CG, HESSELBERTH JR. 2017 valr: Reproducible genome interval analysis in R. F1000Research. 6.
ROCKMAN MV. 2012 The QTN program and the alleles that matter for evolution: all that's gold does not glitter. Evolution. 66(1):1-17.
ŞADOĞLU P. 1957 A Mendelian gene for albinism in natural cave fish. Experientia. 13(10):394-394.
ŞADOĞLU P. 1957 Mendelian inheritance in the hybrids between the Mexican blind cave fishes and their overground ancestor. Verh. Dtsch. Zool. Ges. Graz 1957. Zool. Anz. Suppl. 21:432-439.
SHAW KL, PARSONS YM, LESNICK SC. 2007 QTL analysis of a rapidly evolving speciation phenotype in the Hawaiian cricket Laupala. Molecular ecology. 16(14):2879-2892.
SHEN W-K, CHEN S-Y, GAN Z-Q, ZHANG Y-Z, YUE T, CHEN M-M, XUE Y, HU H, GUO A-Y. 2023 AnimalTFDB 4.0: a comprehensive animal transcription factor database updated with variation and expression annotations. Nucleic Acids Research. 51(D1):D39-D45.
STOCKDALE WT, LEMIEUX ME, KILLEN AC, ZHAO J, HU Z, RIEPSAAME J, HAMILTON N, KUDOH T, RILEY PR, VAN AERLE R, YAMAMOTO Y, MOMMERSTEEG MTM. 2018 Heart Regeneration in the Mexican Cavefish. Cell reports. 25:1997-2007.
STORZ JF. 2016 Causes of molecular convergence and parallelism in protein evolution. Nature Reviews Genetics. 17(4):239-250.
STORZ JF, NATARAJAN C, SIGNORE AV, WITT CC, MCCANDLISH DM, STOLTZFUS A. 2019 The role of mutation bias in adaptive molecular evolution: insights from convergent changes in protein function. Philosophical Transactions of the Royal Society B. 374(1777):20180238.
SUN Y, BILAN PJ, LIU Z, KLIP A. 2010 Rab8A and Rab13 are activated by insulin and regulate GLUT4 translocation in muscle cells. Proceedings of the National Academy of Sciences. 107(46):19909-19914.
WARREN WC, BOGGS TE, BOROWSKY R, CARLSON BM, FERRUFINO E, GROSS JB, HILLIER L, HU Z, KEENE AC, KENZIOR A, KOWALKO JE, TOMLINSON C, KREMITZKI M, LEMIEUX ME, GRAVES-LINDSAY T, MCGAUGH SE, MILLER JT, MOMMERSTEEG MTM, MORAN RL, PEUS R, RICE ES, RIDDLE MR, SIFUENTES-ROMERO I, STANHOPE BA, TABIN CJ, THAKUR S, YAMAMOTO Y, ROHNER N. 2021 A chromosome-level genome of *Astyanax mexicanus* surface fish for comparing population-specific genetic differences contributing to trait evolution. Nature communications. 12:1447.
WARREN WC, RICE ES, ROBACK E, KEENE A, MARTIN F, OGEH D, HAGGERTY L, CARROLL RA, MCGAUGH S, ROHNER N. 2024 Astyanax mexicanus surface and cavefish

chromosome-scale assemblies for trait variation discovery. G3: Genes, Genomes, Genetics.jkae103.
WATERS JM, MCCULLOCH GA. 2021 Reinventing the wheel? Reassessing the roles of gene flow, sorting and convergence in repeated evolution. Mol Ecol. 30(17):4162-4172.
WELLENREUTHER M, BERNATCHEZ L. 2018 Eco-evolutionary genomics of chromosomal inversions. Trends in Ecology & Evolution. 33(6):427-440.
WILDER AP, PALUMBI SR, CONOVER DO, THERKILDSEN NO. 2020 Footprints of local adaptation span hundreds of linked genes in the Atlantic silverside genome. Evolution letters. 4(5):430-443.
WILKENS H. 1988 Evolution and genetics of epigean and cave *Astyanax fasciatus* (Characidae, Pisces): Support for the neutral mutation theory. In: *Evolutionary Biology*, Vol. 23: (Hecht MK, Wallace B, eds), pp. 271-367. Plenum Publishing Corporation, Boston, MA.
WOODHOUSE M, HUFFORD M. 2019 Parallelism and convergence in post-domestication adaptation in cereal grasses. Philosophical Transactions of the Royal Society B. 374(1777):20180245.
WORTEL MT, AGASHE D, BAILEY SF, BANK C, BISSCHOP K, BLANKERS T, CAIRNS J, COLIZZI ES, CUSSEDDU D, DESAI MM, VAN DIJK B, EGAS M, ELLERS J, GROOT AT, HECKEL DG, JOHNSON ML, KRAAIJEVELD K, KRUG J, LAAN L, LASSIG M, LIND PA, MEIJER J, NOBLE LM, OKASHA S, RAINEY PB, ROZEN DE, SHITUT S, TANS SJ, TENAILLON O, TEOTONIO H, DE VISSER J, VISSER ME, VROOMANS RMA, WERNER GDA, WERTHEIM B, PENNINGS PS. 2023 Towards evolutionary predictions: Current promises and challenges. Evol Appl. 16(1):3-21.
XIE KT, WANG G, THOMPSON AC, WUCHERPFENNIG JI, REIMCHEN TE, MACCOLL AD, SCHLUTER D, BELL MA, VASQUEZ KM, KINGSLEY DM. 2019 DNA fragility in the parallel evolution of pelvic reduction in stickleback fish. Science. 363(6422):81-84.
XU S. 2003 Theoretical basis of the Beavis effect. Genetics. 165(4):2259-2268.
YAMAMOTO Y, BYERLY MS, JACKMAN WR, JEFFERY WR. 2009 Pleiotropic functions of embryonic *sonic hedgehog* expression link jaw and taste bud amplification with eye loss during cavefish evolution. Developmental Biology. 330(1):200-211.
YANO M, HARUSHIMA Y, NAGAMURA Y, KURATA N, MINOBE Y, SASAKI T. 1997 Identification of quantitative trait loci controlling heading date in rice using a high-density linkage map. Theoretical and Applied Genetics. 95:1025-1032.
YEAMAN S. 2022 Evolution of polygenic traits under global vs local adaptation. Genetics. 220(1):iyab134.
YEAMAN S, AESCHBACHER S, BÜRGER R. 2016 The evolution of genomic islands by increased establishment probability of linked alleles. Molecular ecology. 25(11):2542-2558.
YEAMAN S, WHITLOCK MC. 2011 The genetic architecture of adaptation under migration–selection balance. Evolution. 65(7):1897-1911.
YOSHIZAWA M, KELLY E, JEFFERY WR. 2013 Evolution of an adaptive behavior and its sensory receptors promotes eye regression in blind cavefish: response to Borowsky (2013). BMC Biol. 11(1):82.
YOSHIZAWA M, ROBINSON BG, DUBOUE ER, MASEK P, JAGGARD JB, O'QUIN KE, BOROWSKY RL, JEFFERY WR, KEENE AC. 2015 Distinct genetic architecture underlies the emergence of sleep loss and prey-seeking behavior in the Mexican cavefish. BMC Biol. 13(1):15.
YOSHIZAWA M, YAMAMOTO Y, O'QUIN KE, JEFFERY WR. 2012 Evolution of an adaptive behavior and its sensory receptors promotes eye regression in blind cavefish. BMC Biol. 10:108.

Table 1. Summary of published Quantitative Trait Loci mapping studies in *Astyanax mexicanus*. SF = surface fish. BC = backcross (F1 hybrids x Parental generation), F2 = second filial generation (F1 hybrids x F1 hybrids), F3 = third filial generation (F2 hybrids x F2 hybrids). In some cases, the number of pigmentation QTL that are mapped may be inflated because pigmentation at different places on the body were mapped as separate traits and the QTL overlapped, yet, we count them separately for this table.

Study	Populations	Cross	Trait	# of QTL identified
Borowsky and Wilkens (2002)	Pachón x SF	BC	Albinism	1
			Condition Factor	2
			Eye Size	3
			Pigmentation	2
Protas et al. (2006)	Molino x SF[1], Pachón x SF[2]	BC, F2	Albinism	1
Protas et al. (2007)	Pachón x SF[2]	F2	Eye Size	8
			Jaw bone length	7
			Lens size	6
			Maxillary tooth number	6
			Pigmentation	18
			Taste bud number	6
Protas et al. (2008)	Pachón x SF[2]	F2	Eye size	8
			Pigmentation	4
			Condition factor	9
			Weight loss	1
			Tooth number	5
			Peduncle depth	3
			Fin placement	9
			Anal fin rays	2
			SO3 width	6
			Rib number	6
			Length	5
			Chemical sense	4
Gross et al. (2009)	Molino x SF[1], Pachón x SF[2]	BC, F2	Brown melanophore	1
Yoshizawa et al. (2012)	Pachón x SF[3]	F2, F3	Albinism	1
			Eye size	5
			Vibration attraction	2
			Superficial neuromasts	2
O'Quin et al. (2013)	Pachón x SF[3]	F2	Retinal thickness	4
Kowalko et al. (2013a)	Pachón x SF, Tinaja x SF	F2	Feeding angle	2
			Jaw angle	1
			Eye size	5
			Head depth	1
			Feeding angle	1
			Tastebuds	2
			Body depth	1
			Eye orbit diameter	5
Kowalko et al. (2013b)	Tinaja x SF	F2	Schooling behavior	2
			Dark preference	1
			Eye size	1
			Pupil size	1
Gross et al. (2014)	Pachón x SF[2]	F2	Albinism	1
			Eye size	2
			Craniofacial bones	12

			Sex	1
			Standard length	1
Carlson et al. (2015)	Pachón x SF[4]	F2	Albinism	1
O'Quin et al. (2015)	Pachón x SF[3]	F2	Scleral ossification	1
			Scleral width	1
			Eye size	8
Yoshizawa et al. (2015)	Pachón x SF[3]	F2, F3	Vibration attraction	2
			Locomotor activity	2
Lyon et al. (2017)	Pachón x SF[2,3]	F2	Scleral ossification	3
Carlson et al. (2018)	Pachón x SF[4]	F2	Velocity	8
			Top usage	2
			Bottom usage	5
Stockdale et al. (2018)	Pachón x SF[6]	F2	Heart regeneration	3
Powers et al. (2020)	Pachón x SF[4]	F2	Eye orbit neuromasts	1
			Eye orbit area	1
Riddle et al. (2020)	Pachón x SF[5]	F2	Carotenoid accumulation	1
Riddle et al. (2021)	Pachón x SF[5]	F2	Blood glucose	2
			Weight	7
			Length	5
			Condition factor	3
			Gonad weight	9
			Liver area	6
			Gut length	5
Warren et al. (2021)	Pachón x SF[5,6]	F2	Albinism	2
			Eye size	2

1. Protas et al. (2006) and Gross et al. (2009) used the same Molino x SF backcross (n=111)
2. Protas et al. (2006), Protas et al. (2007), Protas et al. (2008), Gross et al. (2009), Gross et al. (2014), and Lyon et al. (2017) used the same Pachón x SF F2 mapping population (n=539)
3. Yoshizawa et al. (2012), O'Quin et al. (2013), O'Quin et al. (2015), Yoshizawa et al. (2015), and Lyon et al. (2017) used the same Pachón x SF F2 mapping population (n=384)
4. Carlson et al. (2015), Carlson et al. (2018), and Powers et al. (2020) used the same Pachón x SF F2 mapping population (n=129).
5. Riddle et al. (2020), Riddle et al. (2021), and Warren et al. (2021) used the same Pachón x SF F2 mapping population (n=219).

Stockdale et al. (2018) and Warren et al. (2021) used the same Pachón x SF F2 mapping population (n=188)

Figure 1

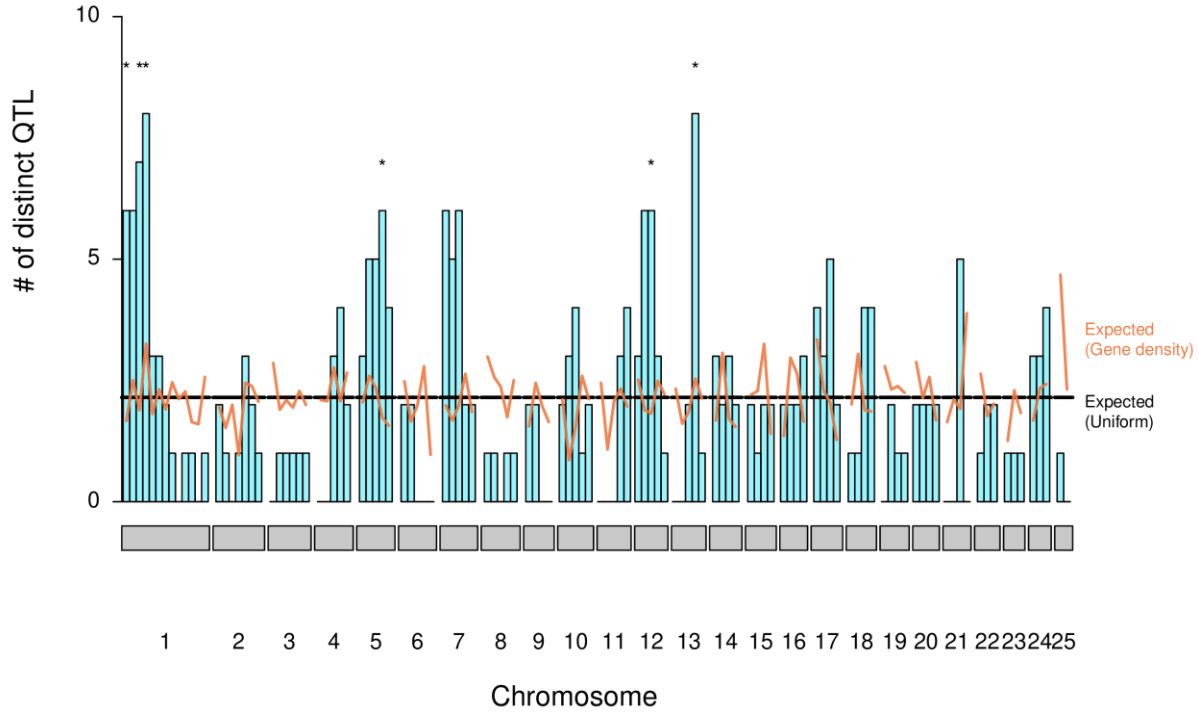

Figure 2