

# Population Genomics of Premature Termination Codons in Cavefish With Substantial Trait Loss

Emma Y. Roback <sup>1</sup>, Estephany Ferrufino <sup>2</sup>, Rachel L. Moran <sup>1,3</sup>, Devin Shennard <sup>2</sup>,  
Charlotte Mulliniks <sup>2</sup>, Josh Gallop <sup>1,4</sup>, James Weagley <sup>1,5</sup>, Jeffrey Miller <sup>1,6</sup>,  
Yaouen Fily <sup>7</sup>, Claudia Patricia Ornelas-García <sup>8</sup>, Nicolas Rohner <sup>9,10</sup>,  
Johanna E. Kowalko <sup>2</sup>, Suzanne E. McGaugh <sup>\*</sup><sup>1</sup>

<sup>1</sup>Ecology, Evolution, and Behavior, University of Minnesota, Saint Paul, MN 55108, USA

<sup>2</sup>Department of Biological Sciences, Lehigh University, Bethlehem, PA 18015, USA

<sup>3</sup>Department of Biology, Texas A&M University, College Station, TX 77843, USA

<sup>4</sup>Department of Dermatology, Cleveland Clinic, Cleveland, OH 44195, USA

<sup>5</sup>Edison Family Center for Genome Sciences and Systems Biology, Washington University School of Medicine, St. Louis, MO 63110, USA

<sup>6</sup>Department of Molecular, Cellular, and Biomedical Sciences, University of New Hampshire, Durham, NH 03820, USA

<sup>7</sup>Wilkes Honors College, Florida Atlantic University, Jupiter, FL 33458, USA

<sup>8</sup>Colección Nacional de Peces, Departamento de Zoología, Instituto de Biología, Universidad Nacional Autónoma de México, México City CP 04510, Mexico

<sup>9</sup>Stowers Institute for Medical Research, Kansas City, MO 64110, USA

<sup>10</sup>Department of Cell Biology and Physiology, University of Kansas Medical Center, Kansas City, KS 66160, USA

\*Corresponding author: E-mail: [smcgaugh@umn.edu](mailto:smcgaugh@umn.edu).

Associate editor: Kelley Harris

## Abstract

Loss-of-function alleles are a pertinent source of genetic variation with the potential to contribute to adaptation. Cave-adapted organisms exhibit striking loss of ancestral traits such as eyes and pigment, suggesting that loss-of-function alleles may play an outsized role in these systems. Here, we leverage 141 whole genome sequences to evaluate the evolutionary history and adaptive potential of single nucleotide premature termination codons (PTCs) in Mexican tetra. We find that cave populations contain significantly more PTCs at high frequency than surface populations. We also find that PTCs occur more frequently in genes with inherent relaxed evolutionary constraint relative to the rest of the genome. Using SLiM to simulate PTC evolution in a cavefish population, we show that the smaller population size and increased genetic drift is sufficient to account for the observed increase in PTC frequency in cave populations without positive selection. Using CRISPR-Cas9, we show that mutation of one of these genes, *pde6c*, produces phenotypes in surface Mexican tetra that mimic cave-derived traits. Finally, we identify a small subset of candidate genes that contain high-frequency PTCs in cave populations, occur within selective sweeps, and may contribute to beneficial traits such as reduced energy expenditure, suggesting that a handful of PTCs may be adaptive. Overall, our work provides a rare characterization of PTCs across wild populations and finds that they may have an important role in loss-of-function phenotypes, contributing to a growing body of literature showing genome evolution through relaxed constraint in subterranean organisms.

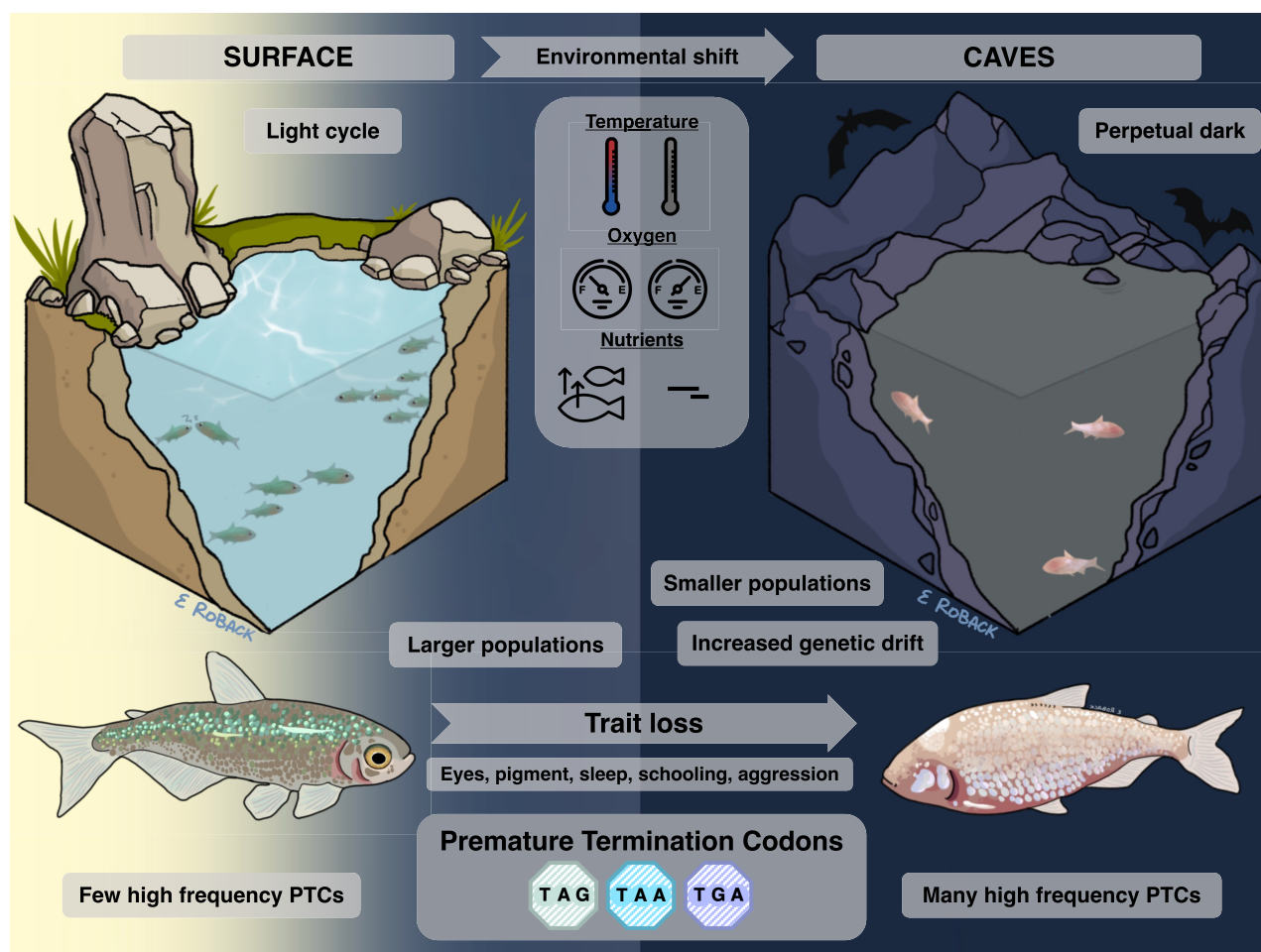
**Keywords:** population genomics, adaptation, stop codons, cavefish, relaxed constraint

Received: August 17, 2024. Revised: January 6, 2025. Accepted: January 7, 2025

© The Author(s) 2025. Published by Oxford University Press on behalf of Society for Molecular Biology and Evolution.

This is an Open Access article distributed under the terms of the Creative Commons Attribution-NonCommercial License (<https://creativecommons.org/licenses/by-nc/4.0/>), which permits non-commercial re-use, distribution, and reproduction in any medium, provided the original work is properly cited. For commercial re-use, please contact [reprints@oup.com](mailto:reprints@oup.com) for reprints and translation rights for reprints. All other permissions can be obtained through our RightsLink service via the Permissions link on the article page on our site—for further information please contact [journals.permissions@oup.com](mailto:journals.permissions@oup.com).

## Graphical abstract



## Introduction

Modern examinations of population genomic diversity have revealed loss-of-function alleles are prevalent across taxa (Saleheen et al. 2017; Xu et al. 2019), but the adaptive significance of these variants in natural populations remains largely unknown (Xu and Guo 2020; Monroe et al. 2021). Loss-of-function alleles represent an underexplored source of potential adaptive variation that may be pertinent during colonization of a new environment where ancestral states are no longer beneficial and may contribute to adaptation more commonly than traditionally appreciated (Hottes et al. 2013; Albalat and Cañestro 2016; Fernández and Gabaldón 2020; Guijarro-Clarke et al. 2020; Monroe et al. 2021).

Colonization of caves by surface-dwelling organisms is an extreme example of species entering a novel environment with distinct selective pressures. Cave-adapted organisms occur across Metazoa and consistently lack eyes, pigmentation, and circadian cycling (Poulson and White 1969; Schilthuizen et al. 2005; Niemiller et al. 2008; Villacorta et al. 2008; Jeffery 2009; Klaus et al. 2013; Zhang and Li 2013; Behrmann-Godel et al. 2017). Thus, cave organisms offer a unique opportunity to test for the evolutionary advantages and phenotypic impact of loss-of-function mutations in a system characterized by trait loss.

The Mexican tetra (*Astyanax mexicanus*) is an emerging model species that consists of at least two separate lineages of surface fish that have colonized cave systems, resulting in

at least two independent evolutionary origins of a cave ecomorph (Bradic et al. 2012; Herman et al. 2018). Independently derived cavefish populations have repeatedly evolved loss of eyes (Yamamoto et al. 2004; Krishnan and Rohner 2017; Sifuentes-Romero et al. 2020), pigment (Protas et al. 2006; Gross et al. 2009), schooling behavior (Kowalko et al. 2013), aggression (Elipot et al. 2013), stress (Chin et al. 2018), and sleep and circadian processes (Duboué et al. 2011; Beale et al. 2013; Jaggard et al. 2017, 2018; Mack et al. 2021; O’Gorman et al. 2021). Given that cave animals are characterized by the reduction of ancestral traits, we hypothesized that loss-of-function alleles may be an important facilitator of cave adaptation.

Premature termination codons (PTCs) can truncate a gene through substitution of a single nucleotide to generate a stop codon prior to the ancestral end of the reading frame. PTCs can result in varying degrees of functional deficit depending on PTC context including the identity of the stop codon and surrounding sequence (Miller and Pearce 2014; Wangen and Green 2020). Interestingly, PTCs have been linked to trait losses relevant to Mexican tetra phenotypes in domestic animals and humans, including pigmentation (Giebel et al. 1991; Everts et al. 2000; Fontanesi et al. 2009; Utzeri et al. 2014) and visual system functioning (Pittler and Baehr 1991; Rosenfeld et al. 1992; Suber et al. 1993; Rio Frio et al. 2008). Despite the potential for PTCs to contribute substantially to phenotypic variation, few studies have characterized PTCs in

genome-wide population genomic datasets or natural systems (Lee and Reinhardt 2011; Hoehn et al. 2012; Yang et al. 2014). Consequently, little is known about the adaptive potential of PTCs.

Here we define the population genetics of PTCs across *A. mexicanus* cave and surface populations by generating a genome-wide catalog of PTCs, and find that the landscape of PTCs was shaped largely through relaxed constraint and genetic drift, in line with other studies showing relaxed constraint is prevalent in subterranean organisms (Niemiller et al. 2008; Calderoni et al. 2016; Tavares and Seuánez 2018; Langille et al. 2022; Zhao et al. 2022; Garduño-Sánchez et al. 2023).

Results

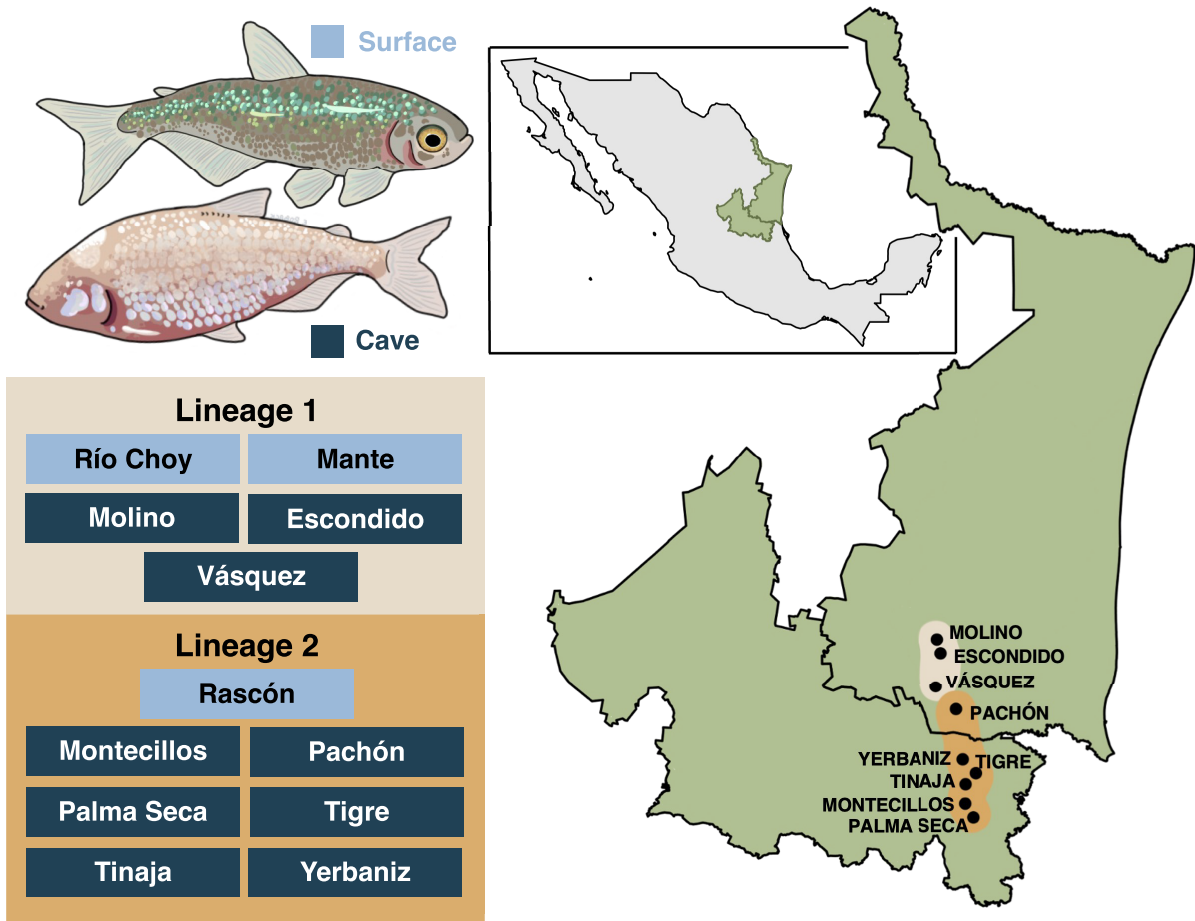
Thousands of PTCs are Present Within Mexican Tetra Populations

To identify PTCs present in Mexican tetra, we analyzed whole-genome resequencing data from 138 individuals across 12 focal populations: three surface populations and nine cave populations from two independent lineages (Fig. 1; supplementary table S1a, Supplementary Material online). Additionally, we used three individuals from two outgroup taxa. In this study, we focus on PTCs generated by single nucleotide changes rather than indels or larger structural variants, because single-nucleotide variants are more reliably

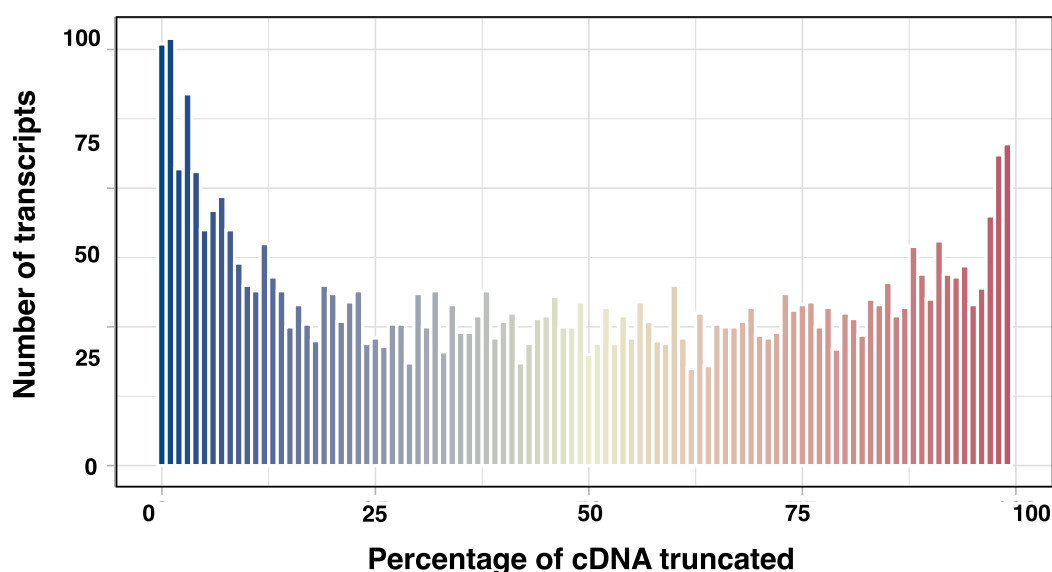
called and confirmed within our sequencing data. We note that our work should be interpreted with the caveat that PTCs based on indels or larger structural variants were omitted, and therefore our dataset of PTCs is likely an underestimate of the true number of PTCs in the genome.

We applied a set of filters limiting consideration to PTCs that were (i) derived in *A. mexicanus* rather than long-segregating variants present in outgroup species, (ii) had a genotype call in at least half of the sampled individuals per population to increase the probability that estimates of a PTC's frequency were representative of the population, (iii) transcribed, and (iv) in an annotated exon as defined in orthologs across other species. These filters reduced the original Variant Effect Predictor (VEP) output to 4,366 computationally filtered PTCs within 3,398 different genes. Each PTC was detected in at least one individual from the 138 *A. mexicanus* samples. Throughout, we compared the 3,398 genes segregating for PTCs to a control set of 13,030 genes that had no PTCs detected by VEP. Hereafter, we refer to these gene sets as “PTC genes” and “control genes”, respectively.

For each transcript that contains a PTC, we determined the amount of the transcript that was truncated. In line with studies of PTCs in *Drosophila* and humans, we find a concentration of PTCs at the beginning and ends of transcripts (Fig. 2) (Ng et al. 2008; Lee and Reinhardt 2011;



**Fig. 1.** Overview of Mexican tetra populations. Focal populations of surface and cave Mexican tetra in which PTCs were identified. The 12 populations represent two genetically distinct lineages, Lineage 1 and Lineage 2, with cave populations repeatedly derived in independent events. Map indicates location of Mexican tetra cave populations, San Luis Potosí and Tamaulipas, Mexico.



**Fig. 2.** Transcript truncation by PTCs. Distribution of the percent of transcript cDNA truncated by PTCs for all transcripts affected by computationally confirmed PTCs. Bars to the left represent less of the cDNA truncated (starting with the bin of 0% to 1% of the gene truncated) and bars to the right represent more of the cDNA truncated (ending with 99% to 100% truncated). Subsequently, bars to the left represent PTCs that occur towards the end of a gene, whereas bars to the right represent PTCs that occur towards the beginning of a gene. PTCs private to cave and surface populations did not have a significantly different distribution from one another.

MacArthur et al. 2012), putatively because PTCs at the beginning and ends of genes may not trigger nonsense mediated decay, potentially allowing truncated proteins to retain some function, or because there is a greater tolerance to truncation close to the end of the coding sequence (Nagy and Maquat 1998; Asselta et al. 2001; Harries et al. 2005). Additionally, we compared the truncation of transcripts by PTCs found only in cave populations to those found only in surface. We do not find significantly different distributions of truncation by cave and surface specific PTCs (Mann–Whitney  $U$ ,  $W = 3806300$ ,  $P$ -value  $> 0.05$ ).

### Cave Populations Contain More High-Frequency PTCs Than Surface Populations

To explore the evolutionary history of PTCs, we compared the distribution and frequency of PTCs between cave and surface populations. Given that the number of individuals sampled was different for each population, we divided the total number of found PTCs in each population by the number of sampled individuals to account for finding more PTCs simply due to sampling more individuals. We find that while the surface populations contain many PTCs, very few PTCs occur at a high frequency (allele frequency  $\geq 0.8$ ) within the populations (Fig. 3a). High-frequency PTCs represented 0.55%, 0.43%, and 1.94% of the per-individual total PTCs present in the Río Choy, Mante, and Rascón surface populations, respectively. In contrast, the nine cave populations contained a much greater proportion of high-frequency PTCs, from 6.98% in Tinaja to 30.25% in Escondido (Fig. 3a, supplementary table S1b, Supplementary Material online). Notably, Tinaja exhibits more gene flow with surface fish populations than other cave populations (Herman et al. 2018) which may explain why we find fewer high-frequency PTCs in this population than in the other cave populations. Of the 236 total PTCs that are present in caves at a high frequency, 46% are putative de novo mutations (not present in any

sampled surface populations), and 54% are present in one or more surface populations.

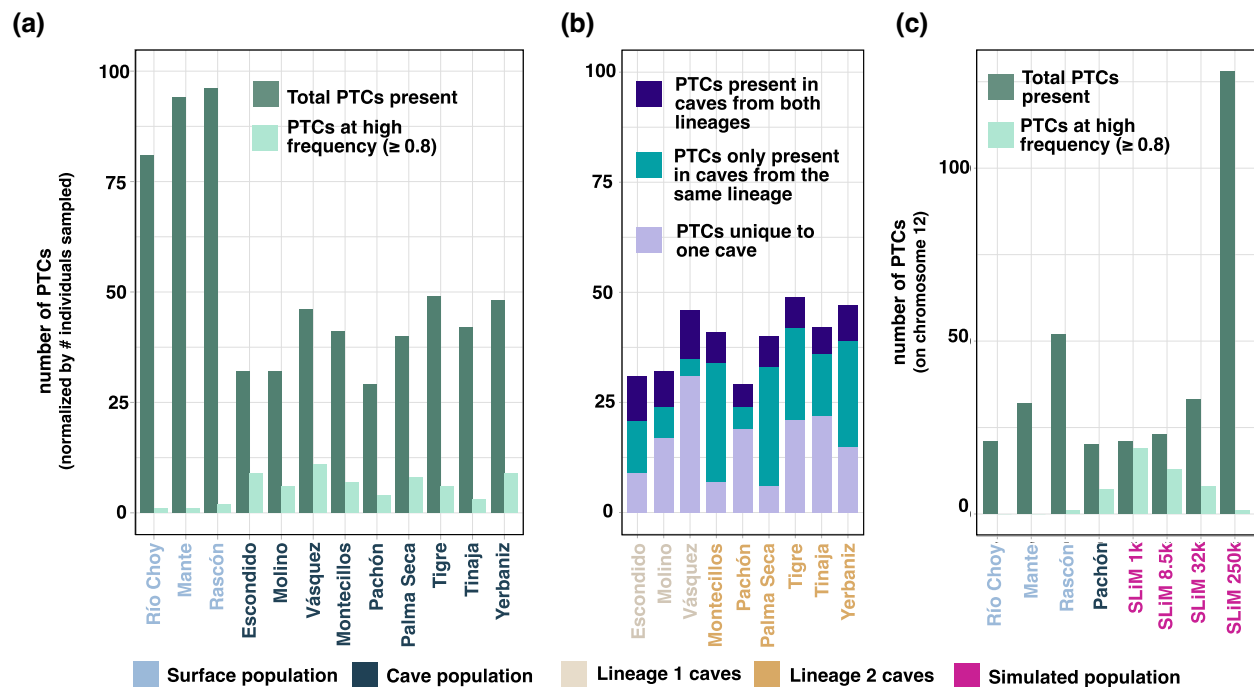
### PTCs are Shared More Among Caves From the Same Evolutionary Lineage Than Between Lineages

Cave populations of Mexican tetra were derived in two independent evolutionary events but undergo some gene flow with surface and other cave populations (Brdic et al. 2012; Herman et al. 2018; Garduño-Sánchez et al. 2023). Thus, we explored the extent to which PTCs present in caves were unique to a cave or shared between caves (Fig. 3b). We found that many cave populations contain numerous unique PTCs that we did not find in any other cave populations. In line with the independent evolutionary origins of the two cave lineages, when PTCs are shared between caves, they were most often found within caves of the same lineage. The caves with the fewest PTCs shared between lineages (e.g. Montecillos, Palma Seca, Tigre, and Yerbaniz) lie further to the south (Fig. 1) and were more geographically separated from caves in the other lineage, suggesting that gene flow and not segregating ancestral polymorphism likely accounts for the observed patterns. Caves from different lineages that occur geographically closer together (e.g. Vásquez from lineage 1, and Pachón from lineage 2) had noticeably more PTCs shared between lineages (supplementary Table S1c, Supplementary Material online).

### Genes With PTCs Show Less Evolutionary Constraint Than Genes Without PTCs

We find several lines of evidence suggesting that genes containing PTCs in our dataset are under less evolutionary constraint than genes that do not harbor segregating PTCs. First, genes with tissue specific expression are less visible to selection (Van Dyken and Wade 2010), while genes that are expressed more broadly likely undergo stronger negative selection (Subramanian and Kumar 2004). The relationship between





**Fig. 3.** PTC distribution by population. a) Distribution of PTCs in cave and surface Mexican tetra populations. The number of PTCs shown for each population is the total PTCs found in the population divided by the number of individuals sampled. While both cave and surface populations have many PTCs within the population and fewer PTCs which occur at high frequency, cave populations harbor significantly more PTCs at high frequency than are present in surface populations. b) Total PTCs found in each cave broken down by presence in other cave populations (number of PTCs normalized by population sample size). PTCs unique to one cave are found only in the focal cave population. PTCs also present in caves from the same lineage are shared in one or more caves from the same evolutionary lineage as the focal cave. Finally, PTCs present in caves from both lineages are found in at least one cave from each lineage. As expected, PTCs are shared more frequently among caves from the same evolutionary lineage than between lineages. c) Comparison of total and high-frequency PTCs found on chromosome 12 in surface, cave, and simulated populations. Values for Río Choy ( $n=9$ ), Mante ( $n=10$ ), Rascón ( $n=14$ ), and Pachón ( $n=19$ ) are the observed PTC counts found in the population genomic dataset. Values for SLiM simulated cave populations (population size 1 thousand, 32 thousand, and 250 thousand individuals, respectively) are from 19 random individuals (same sample size as Pachón) output at the end of the simulation (see [Materials and Methods](#)).

reduced expression breadth and relaxed selection has been shown repeatedly (Snell-Rood et al. 2010; Hoehn et al. 2012; Roberts and Josephs 2023). Using previously collected expression data across 13 tissues from surface fish (supplementary table S2, Supplementary Material online), we found that genes with PTCs are expressed in fewer tissues than non-PTC genes (Kruskal–Wallis  $\chi^2=106.16$ ,  $df=1$ ,  $P$ -value  $<0.00001$ ; supplementary table S3, Supplementary Material online). Thus, genes with existing PTCs may retain potentially deleterious variants because they are expressed less ubiquitously and, therefore, their disruption may have limited fitness consequences. Because the RNA data were sourced from surface fish, for a further analysis, we removed any genes that have a PTC in surface fish from this analysis (thereby only examining the expression of only genes that have PTCs in other populations but do not actively contain a PTC affecting the RNA). Here too, we found expression in significantly fewer tissues in PTC genes compared with control genes (Kruskal–Wallis  $\chi^2=43.76$ ,  $df=1$ ,  $P$ -value  $<0.00001$ ; supplementary table S2, Supplementary Material online), indicating that the reduced expression of PTC genes is not solely due to the PTC itself. Thus, through this test, we confirmed that the expression of these genes is inherently less broad rather than being restricted through the presence of a PTC.

Second, in other systems, genes that occur within a larger gene family have greater robustness to mutation (Conant and Wagner 2004) and genes containing PTCs are more likely

to be part of a gene family with multiple paralogs (Ng et al. 2008; MacArthur and Tyler-Smith 2010). Likewise, we find that genes with PTCs have significantly more paralogs (median gene family size = 10) than genes without PTCs (median gene family size = 5; Kruskal–Wallis  $\chi^2=247.49$ ,  $df=1$ ,  $P$ -value  $<0.0001$ ), implying that the effect of PTCs may be more often buffered through functional redundancy, and potentially reducing the need for negative selection to remove PTCs.

Finally, we find that PTC genes have significantly greater sequence diversity than control genes. We calculated the within-population nucleotide diversity ( $\pi$ ) for each gene within each of our 12 populations and found that genes that contained PTCs in the focal population (the population in which  $\pi$  was calculated) exhibited significantly larger  $\pi$  values relative to control genes (all tests  $P$ -value  $<0.0001$  for all cave populations;  $P$ -value  $<0.05$  for all surface populations; supplementary table S3, Supplementary Material online). This trend was intensified in cave populations. In surface populations, genes that contain PTCs in the focal population were 1.5 to 2.7x more diverse than control genes, while in cave populations, genes where the PTC is present in the focal population were 3.9 to 9.6x more diverse than control genes. Genes containing PTCs may have more sequence diversity because PTCs create pseudogenized genes, which then accumulate mutations. Alternatively, genes may be more diverse because they are inherently under less selective constraint

and, thus, predisposed to accumulate PTCs and other mutations. To assess support for the latter case, we examined the  $\pi$  of genes that segregated for PTCs in other populations, but not in the focal population used to calculate  $\pi$ . For these genes,  $\pi$  was still significantly greater than control genes (all tests  $P < 0.01$ ; [supplementary table S3, Supplementary Material online](#)). Notably,  $\pi$  was also significantly  $< \pi$  for genes where the PTC is present in the focal population (all tests  $P < 0.0001$ , except for Vásquez and Rascón where there was not a significant difference between genes based on the presence or the absence of the PTC in the focal population; [supplementary table S3, Supplementary Material online](#)). Thus, genes with segregating PTCs are a nonrandom sample of genes inherently under relaxed evolutionary constraint relative to the rest of the genome, and pseudogenization following a PTC mutation likely also contributes to further elevated  $\pi$ .

Likewise, we measured between-population divergence ( $D_{XY}$  and  $F_{ST}$ ) for each gene in each cave population against the surface population from the same lineage (Lineage 1: Río Choy and Mante, Lineage 2: Rascón). We observe the same trend as with  $\pi$ :  $D_{XY}$  and  $F_{ST}$  were consistently significantly greatest in PTC genes where the PTC present was in the focal cave population, intermediate in PTC genes where the PTC was present in another population but absent in the focal population used to calculate population genetic statistics, and the lowest in control genes ([supplementary tables S4 and S5, Supplementary Material online](#)).

### PTCs are More Frequent in Longer Genes

We previously found that genes that were subject to hard or soft selective sweeps tend to be longer than other genes in the genome, consistent with longer genes representing a larger mutation target size ([Moran et al. 2023](#)). Likewise, here we found that genes with PTCs were generally longer than control genes in both total base pairs (Kruskal–Wallis  $\chi^2 = 72.019$ ,  $df = 1$ ,  $P\text{-value} < 0.0001$ ), and coding sequence length of the longest transcript (Kruskal–Wallis  $\chi^2 = 158$ ,  $df = 1$ ,  $P\text{-value} < 0.0001$ ). Further, genes that contain putative de novo cave-derived PTCs were significantly longer than genes that contain PTCs also found in surface populations (Kruskal–Wallis  $\chi^2 = 7.805$ ,  $df = 1$ ,  $P\text{-value} < 0.01$ ). Taken together, these data support that observed PTCs may be more common in longer genes due to their increased mutational opportunity.

### Evidence for Selection on PTC Genes

Despite our overall finding of a strong signature of relaxed selective constraint affecting genes that contain PTCs, PTCs that eliminate expensive, unused traits like eyes may be favored by natural selection in a cave environment ([Moran et al. 2014, 2015](#)). To identify PTCs that may have been influenced by positive selection, we used diploS/HIC, a powerful supervised machine learning approach. DiploS/HIC combines multiple population genetic statistics to characterize windows of the genome which have likely undergone selection and neutral windows which have no evidence of selective sweeps ([Kern and Schrider 2018](#)). Similar to the population genetic statistics described in the previous section, selective sweeps were determined by diploS/HIC for each population independently. For each PTC and control gene, we indicated whether the gene occurred in a region of the genome with or without evidence of a selective sweep in each cave and surface population. Genes that could not be definitively placed in either a selective or

neutral window by diploS/HIC were discarded for this analysis. For PTC genes, we also indicated whether a PTC was present in the focal population or if the gene had a PTC, but in another population. We performed a  $\chi^2$  test between PTC genes (PTC present in focal population) and control genes to determine whether genes with a PTC were more often associated with selective sweeps than control genes. In accordance with our evidence that most PTCs have likely accumulated through relaxed selective constraint, we find that PTC genes (where the PTC is present in the population used to detect sweeps) are not associated with selective sweeps compared with control genes (all  $P\text{-values} \geq 0.05$ , [supplementary table S6, Supplementary Material online](#)). Interestingly, we find that in Escondido, Mante, Pachón, Tinaja, and Yerbaniz, genes which contained PTCs in other populations were more often associated with selective sweeps (all  $P\text{-values} < 0.05$ , [supplementary table S6, Supplementary Material online](#)) than genes that contain PTCs in the focal population, suggesting that genes that contain PTCs may have been under selection, but once a PTC occurs in the gene the pseudogenization process may erase signatures of selection.

### Population Genetic Modeling in SLiM

We used the forward genetic simulation framework SLiM ([Haller and Messer 2019](#)) to explore the evolutionary processes shaping our observed landscape of PTCs. We simulated the evolution of PTCs on a chromosome in the Pachón cave-fish population at the estimated effective population size of 32,000 fish ([Herman et al. 2018](#)). We first explored the fate and frequency of standing and de novo PTCs with selection coefficients sampled from three different distributions of fitness effects (DFE) and compared these patterns to the 20 total and 7 high-frequency (3 standing and 4 de novo) PTCs we observed in Pachón on chromosome 12 ([supplementary table S7a, Supplementary Material online](#)).

In a neutral model where all PTC mutations have the same effect as a neutral mutation (represented by a fixed DFE of 0), we found considerably more mutations present (155) compared with the observed PTCs on chromosome 12 (20 PTCs), suggesting that there is likely negative selection reducing the prevalence of PTCs in nature ([supplementary table S7b, Supplementary Material online](#)). This is in line with the fact that PTCs are a potential loss-of-function variant and could have deleterious fitness effects if they occur in constitutive genes. Additionally, to address the possibility that the finding of 155 PTCs in our neutral simulation is inflated by overestimating the PTC mutation rate, we repeated our neutral model with the mutation rate reduced by a third. With this lower mutation rate, the neutral simulations resulted in 104 PTCs ([supplementary table S7b, Supplementary Material online](#)). Thus, even with a reduced PTC mutation rate, we still find approximately 4 $\times$  more PTCs in the neutral model than observed.

Alternatively, we found that models with a gamma DFE where PTCs ranged from nearly neutral to strongly deleterious resulted in a frequency spectrum of PTCs similar to the observed PTCs on chromosome 12. Our best gamma DFE model with a mean selection coefficient of  $-0.3$  resulted in a group of PTCs at low and intermediate frequencies, with an average of 3.5 high-frequency standing PTCs and 4.2 high-frequency de novo PTCs ([supplementary table S7c, Supplementary Material online](#)). Notably, this result shows that the observed pattern of PTCs in Pachón can be emulated without positive

selection. Finally, we tested models with a normal DFE where PTCs ranged from weakly beneficial to deleterious. Regardless of the mean and standard deviation supplied to the model, the resulting frequency spectrum of both de novo and standing PTCs consisted of only high-frequency PTCs (supplementary table S7d, Supplementary Material online), indicating that even the small amount of positive selection in this model is likely stronger than the selection on most PTCs in Pachón. Given that we are using a quite simple statistic (the number of PTCs present) to evaluate the outcome of these models we would like to highlight that many factors could influence the apparent ability of any model to explain our observed findings. While our models surely do not replicate the full complexity of how PTCs impact Mexican tetra fitness in the wild, the results across our three SLiM models support the overall finding from our population genomic analyses that likely few PTCs are under positive selection.

Our second objective using SLiM was to investigate whether our finding that cave populations of Mexican tetra contain a significantly greater proportion of high-frequency PTCs could be explained by comparatively smaller population size and consequently stronger effects of genetic drift in caves. We reran our gamma distributed model that best approximated the observed PTCs in Pachón with three additional population sizes: (i) a large population of 250,000 individuals, in line with the estimated effective population size of surface population Río Choy (Herman et al. 2018), (ii) a smaller population of 85,000 individuals as estimated in Pachón by (Mitchell et al. 1977) and within the confidence interval of the estimate for Pachón in (Herman et al. 2018), and (iii) an extremely small population of 1,000 individuals as estimated in (Avise and Selander 1972; Bradic et al. 2012; Fumey et al. 2018). In the modeled population of 32,000 individuals, 11% of total de novo PTCs reached high frequency. Comparatively, in a population of 1 k, 8.5 k, and 250 k individuals, 88%, 45%, and approximately 0% of total de novo PTCs reached high frequency, respectively (Fig. 3c, supplementary table S7e, Supplementary Material online). While our SLiM model is likely not an exact replica of the true evolution of PTCs in nature, we find that when the same model is run in a very large population nearly zero PTCs rise to high frequency and in a very small population nearly all PTCs reach high frequency, despite the same selection on PTCs in both populations. Thus, it is likely that the significantly greater proportion of high-frequency PTCs we observe in caves is results from the increased strength of genetic drift from smaller population size. However, if the true effective population size of Pachón is as small as some estimates predict (Avise and Selander 1972; Bradic et al. 2012; Fumey et al. 2018), PTCs would need to be under even greater negative selection than that invoked by the strongly deleterious mean selection coefficient of  $-0.3$  used in the model. Thus, considering our observed PTC allele frequency data, we do not view extremely small effective population sizes of around 1,000 individuals as realistic.

### Enrichment for Functional Groups or PTCs Under Previously Documented QTL

We find no significantly enriched GO terms when comparing our entire set of computationally confirmed PTCs, genes containing PTCs only in cavefish, or genes containing PTCs at high frequency in cavefish to the background of the genome, indicating that PTCs do not consistently occur in a suite of genes with a specific function. Additionally, we used a

dataset of previously established cavefish Quantitative Trait Loci (QTL) for a wide range of traits (Dryad repository, <https://doi.org/10.5061/dryad.3xsj3txmf>) and identified whether each PTC and control gene was found within an established QTL region (supplementary table S8, Supplementary Material online). We find that control genes are more often found within QTL regions than PTC genes ( $\chi^2 = 23.241$ ,  $df = 1$ ,  $P$ -value  $< 0.0001$ , supplementary table S8, Supplementary Material online). Additionally, we performed the same comparison for subsets of QTL that are associated with some of the most pertinent cavefish traits of eyes, pigment, and sensory systems and find that control genes are also more often found within QTL for each of these specific traits than PTC genes (eyes:  $\chi^2 = 8.256$ ,  $df = 1$ ,  $P$ -value  $< 0.01$ , pigment:  $\chi^2 = 6.518$ ,  $df = 1$ ,  $P$ -value  $< 0.01$ , sensory systems:  $\chi^2 = 12.499$ ,  $df = 1$ ,  $P$ -value  $< 0.001$ , supplementary table S8, Supplementary Material online). These results suggest that PTCs are depleted in regions of the genome associated with cavefish traits.

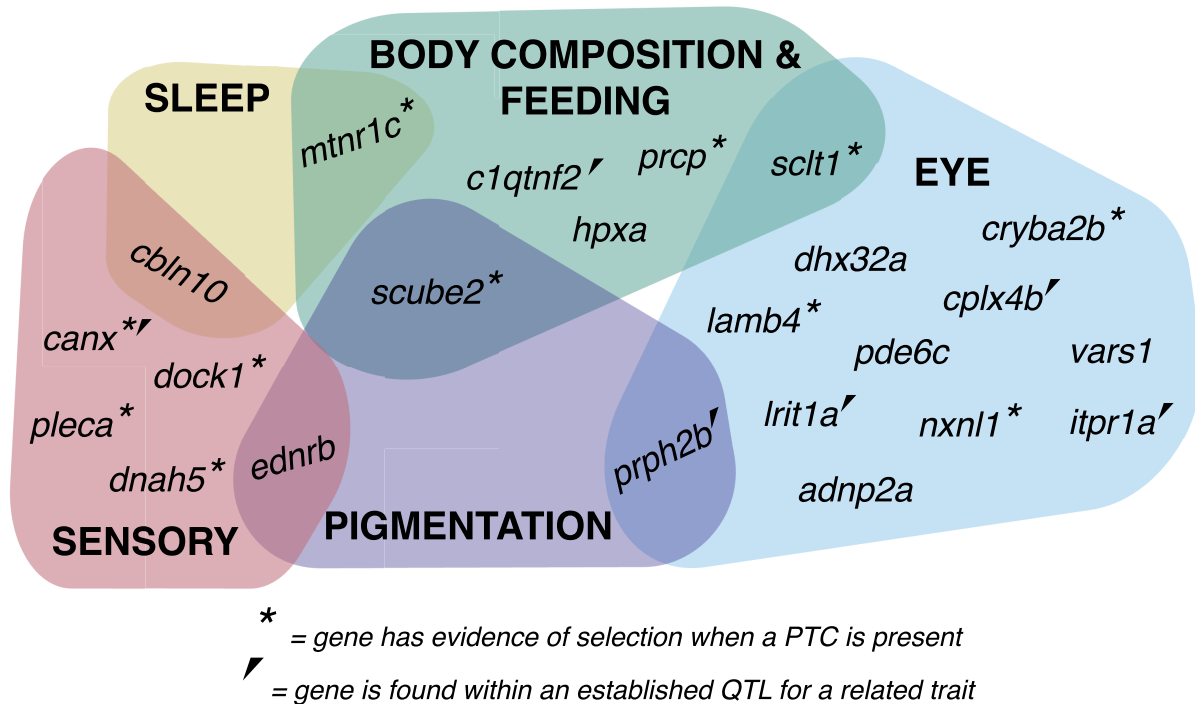
### Candidate PTC Genes Underlying Cave Adaptation

While our data suggest that many PTCs are likely phenotypically buffered and rise in frequency in caves through drift, we identified a subset of PTCs that are candidates for contributing to the evolution of cave traits (Fig. 4). We curated a collection of 105 genes which occurred at high frequency ( $\geq 0.8$ ) in one or more cave populations, never occurred in a homozygous state in surface fish, and for which no transcripts were predicted to splice out the PTC (supplementary table S9a, Supplementary Material online). From this set, 50 genes had a known gene annotation from Ensembl (while the remaining 55 genes were considered “novel genes”). Of these 50 candidate genes, 23 have phenotypic annotations in humans, mice, and zebrafish that correspond to one or more cavefish traits (supplementary table S9b, Supplementary Material online), 23 genes are found within previously established QTL regions (supplementary table S9c, Supplementary Material online) and 21 genes occur within a selective sweep in the population where the PTC is present (supplementary table S9d, Supplementary Material online).

### Functional Impact of the PTC in Candidate Gene *pde6c*

Degradation of eyes is one of the most prevalent trait losses in cavefish, thus the disruption of eye related genes by PTCs may contribute to this loss. We find that candidate gene *pde6c* is repeatedly disrupted by PTCs in other dark-adapted species, including fossorial mammals and nocturnal bats (Fig. 5a) (Emerling and Springer 2014, 2015; Blumer et al. 2022). Given the repeated disruption of *pde6c* across taxa and its documented involvement in photoreception and vision in zebrafish (Saade et al. 2013; Viringipurampeer et al. 2014; Zhang et al. 2016, 2017; Iribarne et al. 2017), we explored the functional impact of the PTC in *pde6c* in Mexican tetra using CRISPR-Cas9 gene editing.

In Mexican tetra, *pde6c* has 1 transcript which has a coding sequence (CDS) length of 2,556 bp. We find that the average CDS length for all transcripts affected by PTCs is 2,541 bp, so *pde6c* is approximately average in length compared with all PTC genes. Comparatively, the average CDS length across all Mexican tetra genes in the genome is 1,560 bp, so *pde6c* is a longer gene than average when considering all genes (consistent with our finding that genes affected by PTCs tend to be



**Fig. 4.** Putative relationships between PTC genes and cave-derived phenotypes. We explored documented functional consequences of 50 annotated candidate PTC genes (those that occurred at high frequency [ $\geq 0.8$ ] in one or more cave populations, never occurred in a homozygous state in surface fish, and for which no transcripts were predicted to splice out the PTC) in zebrafish, mice, and humans by leveraging the databases Zfin, MGI, IMPC, Orphanet, OMIM, and DECIPHER, and review of current literature (supplementary table S9, Supplementary Material online). Each colored bubble represents a suite of phenotypes related to cave-derived traits. The genes listed in each bubble had documented phenotypes related to the phenotype indicated in the bubble. The asterisk denotes candidate genes with evidence of selection in population(s) where the PTC is present, and the triangle denotes candidate genes that overlap a QTL (established by previous studies of Mexican tetra) for a trait related to the phenotype indicated in the bubble. The genes listed here represent the most compelling candidates for PTCs with the potential to contribute to cave adaptation.

longer genes). We used CRISPR-Cas9 gene editing to mutate candidate gene *pde6c* in Mexican tetra surface fish. We injected surface fish with a gRNA targeting *pde6c* exon 7, upstream of the PTC, and a Cas9 protein, then assessed phenotypes in 9 days post fertilization (dpf) injected individuals (CRISPs) compared with uninjected, wild-type siblings. The relative eye size was significantly reduced in *pde6c* CRISPant fish compared with wild-type siblings (Left eye: Kruskal–Wallis  $\chi^2 = 6.7201$ ,  $df = 1$ ,  $P$ -value = 0.009533, Right eye: Kruskal–Wallis  $\chi^2 = 6.5939$ ,  $df = 1$ ,  $P$ -value = 0.01023; Fig. 5b and c). Additionally, we assessed optomotor response (OMR) in *pde6c* CRISPant fish and their wild-type siblings. OMR assays are used in zebrafish and Mexican tetra as visual defect detection assay (LeFauve et al. 2021; Choy et al. 2024). OMR is an innate, orienting behavior evoked by whole-field visual motion, in our case, contrasting black and white lines that move across a screen. Fish that can see the visual cue inherently follow the direction of the lines, while fish with vision deficits do not. *Pde6c* CRISPant fish have significantly reduced OMR compared with wild-type siblings, who consistently orient to the visual cue and swim in the direction of the lines (Kruskal–Wallis  $\chi^2 = 23.606$ ,  $df = 1$ ,  $P$ -value < 0.0001; Fig. 5d), indicating that *pde6c* CRISPants have visual defects. We did not formally assay other behavior abilities in the *pde6c* CRISPant fish beyond OMR; however, we did not see any qualitative differences in locomotor behavior in our CRISPant fish.

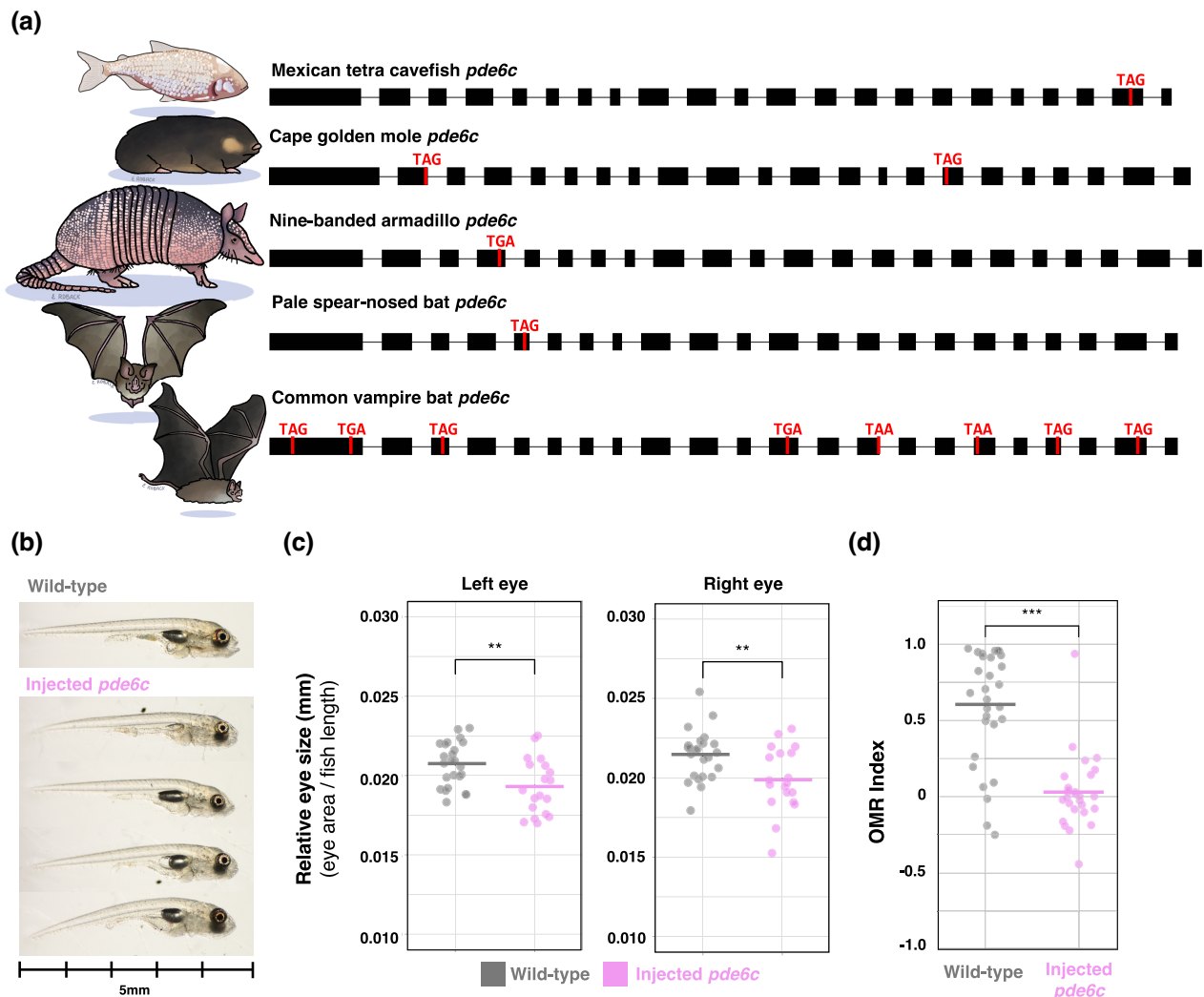
While we note that our CRISPants are likely not an exact phenocopy of the cavefish *pde6c* because the CRISPR target

was upstream of the PTC, we show that interruption of *pde6c* may regress the eye size and therefore, disruption of *pde6c* by a PTC may be a portion of the genetic architecture underlying eye loss in cavefish. Thus, while our work found that Mexican tetra PTCs are not enriched in genes with evidence of selection or in established QTL, we demonstrate the power of this dataset to identify functionally relevant candidate genes which are impacted by PTCs.

## Discussion

Mounting evidence over the past two decades indicates that loss-of-function alleles are an evolutionarily important and pervasive source of genetic variation across taxa and can underlie evolutionary novelty (Albalat and Cañestro 2016; Murray 2020; Monroe et al. 2021). In experimental evolution, >200 instances of loss-of-function variants have been associated with adaptations to altered environmental conditions, showing that substantial adaptation can proceed through loss (Hottes et al. 2013). Similarly, loss-of-function variants including frameshifts, splice site mutations, and PTCs have contributed to adaptation and phenotypic diversification in populations of *Arabidopsis* (Xu et al. 2019). In cave animals, loss or reduction of pigment, baseline sleep duration, and eyes characterize some of the most iconic traits, thus, cave animal are excellent systems for examining the potential role of loss-of-function alleles in phenotypic evolution. Here, we explore the population genetics of one specific loss-of-function variant, PTCs.





**Fig. 5.** Functional Impacts of PTCs in *pde6c*. a) Location of PTCs in *pde6c* across taxa. PTCs are attributed to the regression of visual protein networks in species that burrow underground or are active strictly at night, including the cape golden mole (*C. asiatica*) (Emerling and Springer 2014), nine-banded armadillo (*Dasypus novemcinctus*) (Emerling and Springer 2015), and multiple bat species (Blumer et al. 2022). Black bars represent each exon, and all introns have been scaled to 100 bp. b) Visual comparison of 9 dpf injected *pde6c* CRISPRants with reduced eyes compared with uninjected, wild-type siblings. c) Comparison of relative eye size (eye area/fish length) in 9 dpf injected *pde6c* CRISPRants compared with uninjected, wild-type siblings. *Pde6c* CRISPRant surface fish show significantly reduced eye size;  $P = 0.01$  for both the left and right eyes. d) Comparison of OMR between injected *pde6c* CRISPRants and uninjected, wild-type siblings. OMR index represents the proportion of the total possible distance traveled in the direction of the visual cue. Positive values indicate positive directional response to visual cue, while negative values indicate an inability to discern and orient towards visual cues. *Pde6c* CRISPRants have significantly lower OMR response and often do not orient to the visual cue in the OMR assay and therefore likely have visual defects compared with wild-type siblings ( $P < 0.0001$ ).

One of the predominant trends in our data is that genes containing segregating PTCs appear to be inherently less constrained than the remainder of the genome. We find that PTC genes have greater sequence diversity both within and between populations of Mexican tetra, which cannot be attributed solely to mutation accumulation after pseudogenization. Compared with control genes, PTC genes have significantly more paralogs, indicating that their functions may be redundant or overlap those other genes in the genome. This redundancy may allow PTC genes to accumulate mutations without significantly impairing the overall fitness of the organism. Additionally, and in line with characterization of PTCs in *Drosophila* (Lee and Reinhardt 2011; Hoehn et al. 2012; Yang et al. 2014) and *Arabidopsis* (Roberts and Josephs 2023), PTC genes are expressed in fewer tissues than control genes and may, therefore, be less visible to selection than broadly expressed genes (Subramanian and Kumar 2004). These findings are

consistent with investigation of other subterranean organisms including other species of cavefish (Niemiller et al. 2008; Calderoni et al. 2016; Zhao et al. 2022; Garduño-Sánchez et al. 2023), fossorial rodents (Tavares and Seuánez 2018), and aquifer dwelling beetles (Langille et al. 2022) which also have strong evidence of relaxed evolutionary constraint.

While PTC genes are inherently under relaxed constraint, we find evidence that constraint is even further relaxed in cave populations. In our measures of nucleotide diversity ( $\pi$ ), genes with PTCs in surface were 1.5 to 2.7x more diverse than control genes, while genes with PTCs in caves were 3.9 to 9.6x more diverse, indicating that in caves, PTC genes likely undergo even further drift and mutation accumulation. Genes that formerly served important functions in the surface environment may become obsolete in the cave environment and undergo a release from negative selection that maintained their function. This is documented in fruit bats, which lose

function of fat metabolism genes after adaptation to a low-fat, frugivorous diet (Subramanian and Kumar 2004) and in both giant and red pandas, which have convergent pseudogenization of umami receptors following the shift to obligate bamboo feeding from meat-eating ancestors (Hu et al. 2017). In our dataset, we find a high-frequency PTC in the Pachón cavefish population in *lrit1a*, a retinal transmembrane protein. Notably, we find no selective sweeps surrounding this PTC. *lrit1* is essential for the development of accurate visual function in mice, and loss of *lrit1* impairs discrimination of visual signals in daylight (Sarria et al. 2018; Ueno et al. 2018). In the cave environment, cavefish are in perpetual darkness, and thus this may be an example of a previously useful gene incurring a loss-of-function allele that no longer carries negative fitness consequences.

While we find that many PTCs are likely phenotypically buffered, relaxed constraint can also contribute to trait loss through the degradation of genes underlying the trait (Kimura 1979; Wilkens 2020). For example, subterranean beetles exhibit neutral molecular evolutionary processes leading to parallel pseudogenization of genes involved in phototransduction, resulting in loss of light detection (Langille et al. 2022). We detected a high-frequency PTC in *pde6c* in cavefish and show that *pde6c* CRISPRs had a reduced eye structure and vision defects, indicating that this PTC may contribute to the genetic architecture of eye loss in Mexican tetra. Notably, PTCs in *pde6c* are an element of the regression of visual protein networks in other species that burrow underground or are active strictly at night, including the cape golden mole (*Chrysochloris asiatica*) (Emerling and Springer 2014), nine-banded armadillo (*Dasypus novemcinctus*) (Emerling and Springer 2015), and five unique bat species (Blumer et al. 2022). Other forms of inactivating mutations are found in *pde6c* in the fossorial naked mole rat (*Heterocephalus glaber*) (Emerling and Springer 2014), and three species of whales which occupy low light conditions underwater (Springer et al. 2016). Therefore, it seems that degradation of visual proteins has occurred in a repeatably manner across organisms adapted to darkness.

We also observe significantly more high-frequency PTCs in cave populations compared with surface populations. While this could be due, in part, to positive selection on a subset of PTCs, results from evolutionary modeling of PTCs in SLiM suggest that the smaller population size of cavefish and consequently increased effects of genetic drift are sufficient to account for the observed increase in PTC frequency without positive selection. Notably, our simulations show that an especially small effective population size of 1,000 individuals or fewer in Pachón cave suggested by several studies (Avisé and Selander 1972; Bradic et al. 2012; Fumey et al. 2018) would require unrealistically high negative selection coefficients to replicate the moderate level of fixation of PTCs in observed data, supporting the effective population sizes estimated in (Herman et al. 2018).

While drift likely shapes the molecular evolution of many of the genes in our study, we identify a small subset of PTCs under selection that may be adaptive and are excellent candidates for further research. We identify one unique PTC gene that may have been targeted repeatedly in independent bouts of cavefish evolution. We find that *dock1* contains independent, high-frequency PTCs at different CDS positions between the two cavefish lineages and has evidence of selection in every population where the PTC is present. This finding is in line

with previous work that selection has targeted the same genes repeatedly across cavefish lineages (Moran et al. 2023). *Dock1* plays a key role in the functioning of the lateral line system through myelination of Schwann cells in zebrafish (Cunningham et al. 2018), and further research may illuminate why *dock1* is consistently disrupted in cavefish.

Mexican tetra cavefish have evolved attributes that promote survival in the unpredictable conditions of caves with scarce nutrients, including increased appetite, fat accumulation, insulin resistance, and other metabolic changes (Aspiras et al. 2015; Riddle et al. 2018; Xiong et al. 2018, 2022; Medley et al. 2022). Nutrient limitation in caves may also favor the loss of tissues and biological processes that are energetically expensive. Specifically, eyes and their downstream processes levy high metabolic costs on cavefish (Moran et al. 2014, 2015).

In cave populations, several genes, including *mtnr1c*, *cryba2b*, *lamb4*, and *nxnl1*, contain high-frequency PTCs in caves within selective sweeps. Moreover, phenotypes documented in other species indicate that they may play a role in adaptation to nutrient limitation through metabolic changes or reduction in energy expenditure of eyes. *Mtnr1c* (also known as *mel1c*) is a melatonin receptor orthologous to mammalian *gpr50* (Dufourny et al. 2008) which is photoperiod sensitive, (Barrett et al. 2006), results in alterations to metabolism, including hyperphagia and hyperactivity (Ivanova et al. 2008), and impacts transcriptional responses to leptin signaling (Bechtold et al. 2012). One of the key targets of leptin is melanocortin 4 receptor (*mc4r*), which is thought to contribute to the metabolic adaptation of Mexican tetra cavefish to nutrient poor conditions in caves (Aspiras et al. 2015). *Cryba2b* is important for development of a transparent and light-reflective lens and when disrupted, results in reduced lens size and cataracts in both mice and humans (Puk et al. 2011; Reis et al. 2013). *Lamb4* is involved in the morphogenesis and distribution of Müller glia which act as optical fibers for the passage of light from the retinal surface to the photoreceptor cells (Franze et al. 2007; Charlton-Perkins et al. 2019). Finally, *nxnl1* is involved in the maintenance of the retina and results in progressive loss of function in cone and rod photoreceptors when disrupted in mice (Cronin et al. 2010). Notably, selection on the degeneration of a gene may be difficult to detect given that once the gene function is destroyed it is free to accumulate mutations which may obscure the signature of a past selective sweep, especially if PTCs rose in frequency during the earliest stages of cavefish phenotypic evolution. Additionally, ancestral traits may have a complex genetic architecture and can be destroyed by other means than the fixation of a single variant, making the selective signatures of loss-of-function alleles especially difficult to detect (Monroe et al. 2021). Thus, we suspect that our data underestimate the role of positive selection on loss-of-function variants in the Mexican tetra evolutionary history.

In sum, this analysis represents one of the first genome-wide examinations of PTCs across naturally evolving populations. Our work demonstrates that relaxed constraint and drift are predominant forces in shaping the landscape of an important class of loss-of-function mutations, even in a system that is characterized by pervasive disruption of key ancestral phenotypes.

## Materials and Methods

### Sampling, Sequencing, and Variant Calling

Details of DNA extraction, Illumina sequencing, and variant calling are given in (Moran et al. 2022, 2023). In brief,

we conducted whole genome resequencing of *A. mexicanus* cavefish and surface fish collected from the Sierra de El Abra and Sierra de Guatemala regions in San Luis Potosí and Tamaulipas, Mexico. Reads were aligned to the surface *A. mexicanus* genome (Astyanax\_mexicanus-2.0) and genotype calling was performed following the GATK Best Practices. For this study, we focused on 138 individuals from populations with the most extensive sampling from the [Moran et al. \(2023\)](#) study. These samples included three surface populations (Río Choy  $n=9$ , Río Mante  $n=10$ , and Rascón  $n=14$ ) and nine cave populations (Escondido  $n=9$ , Molino  $n=15$ , Montecillos  $n=9$ , Pachón  $n=19$ , Palma Seca  $n=8$ , Tigre  $n=10$ , Tinaja  $n=20$ , Vásquez  $n=8$ , and Yerbaniz  $n=7$ ) ([Fig. 1](#), [supplementary table S1](#), [Supplementary Material](#) online). The caves represent at least two independent origins of cave morph: Lineage 1 (Surface: Río Choy, Mante; Cave: Escondido, Molino, Vásquez) and Lineage 2 (Surface: Rascón, Cave: Montecillos, Pachón, Palma Seca, Tigre, Tinaja, and Yerbaniz). Caves experience gene flow between caves and between cave and surface populations ([Bradic et al. 2012](#); [Herman et al. 2018](#)). We used sequences of three individuals from two outgroup species (*Astyanax aeneus*  $n=1$ , *Astyanax nicaraguensis*  $n=2$ ) to filter out ancestral segregating variants (described below). Thus, this study leverages 141 resequenced genomes in total.

### Variant Filtering and Generation of Control Gene set

We annotated genomic variants in the 138 Mexican tetra individuals from 12 populations and three outgroup individuals using the Ensembl Variant Effect Predictor (VEP) v.100 ([McLaren et al. 2016](#)), with the *A. mexicanus* gtf file Ensembl release v100, and the *A. mexicanus* reference genome v2.0 (GCA\_000372685.2). VEP-classified polymorphisms which result in the addition of a stop codon (consequence = “stop\_gained”) relative to the reference and found 10,556 PTCs in 6,512 genes ([supplementary table S10](#), [Supplementary Material](#) online). Through further examination of these PTCs in the DNA sequence, we identified that VEP mis-annotated PTCs if either (i) a codon experienced two separate single nucleotide changes that individually could result in a PTC but did not produce a PTC when both changes occurred together or (ii) an insertion occurred relative to the reference genome. We corrected these errors through a custom Perl script (available at <https://github.com/robbackem/PTC-perls>) which examined each putative PTC codon and retained only those that produced a true single-nucleotide PTC regardless of phase. This additional curation step resulted in 6,384 PTCs in 4,585 genes across our dataset ([supplementary table S11](#), [Supplementary Material](#) online).

Several other filtering steps ensured that (i) we evaluated only PTCs derived in *A. mexicanus*, (ii) at least half of the sampled individuals per population had sequence coverage over the PTC, (iii) the PTC was transcribed, and (iv) the PTC was in an annotated exon as defined in orthologs across other species.

First, to focus on evolutionary processes within *A. mexicanus* rather than segregating ancestral variants, we used three individuals from two closely related congener species, *A. aeneus* ( $n=1$ ) and *A. nicaraguensis* ( $n=2$ ), as outgroups and retained only PTCs absent in outgroup individuals, suggesting that the PTCs we included in our analyses arose as mutations in *A. mexicanus* lineages. PTCs located at sites for which there was no information in *A. aeneus* and *A. nicaraguensis* were

discarded, resulting in 5,781 PTCs in 4,305 genes remaining from the originally annotated set ([supplementary table S12](#), [Supplementary Material](#) online).

Second, we calculated PTC frequency in each population using a custom Perl script. To increase the probability that estimates of a PTC's frequency were representative of the real population, we required that half the individuals or more in each population have genotype data across the PTC. This step resulted in 5,221 PTCs in 3,961 genes remaining from the originally annotated set ([supplementary table S13](#), [Supplementary Material](#) online).

Third, we confirmed the PTC was transcribed by examining overlaps of the 20 bp region surrounding the PTC (10 bp upstream and 10 bp downstream) to a database of RNAseq reads downloaded from the Ensembl FTP site ([https://ftp.ensembl.org/pub/release-100/data\\_files/astyanax\\_mexicanus/Astyanax\\_mexicanus-2.0/rnaseq/](https://ftp.ensembl.org/pub/release-100/data_files/astyanax_mexicanus/Astyanax_mexicanus-2.0/rnaseq/)). This file contains samples taken from both Pachón cave and surface fish (embryo, ovary, brain, gills, heart, muscle, testis, kidney, bone, intestine, liver, skin, and olfactory epithelium/nasal) and eye tissue from surface fish. We determined whether the PTC was likely transcribed using bedtools overlap ([Quinlan 2014](#)) and retained only PTCs for which the 20 bp region was covered by at least one read from the RNAseq data. This step resulted in 5,106 PTCs in 3,883 genes remaining from the originally annotated dataset ([supplementary table S14](#), [Supplementary Material](#) online).

As a final step in computationally validating PTCs, and to account for potential reference annotation errors, PTC-containing genes were aligned to orthologs from five closely related fish species within Otomorpha: *Pygocentrus nattereri* (red-bellied piranha), *Ictalurus punctatus* (channel catfish), *Cyprinus carpio carpio* (common carp), *Carassius auratus* (goldfish), *Danio rerio* (zebrafish), *Denticops clupeoides* (denticle herring), and *Clupea harengus* (Atlantic herring). We required that the codon was within the annotated coding region in at least one other species. This filter resulted in 4,366 PTCs in 3,398 genes remaining from the originally annotated set ([supplementary table S15](#), [Supplementary Material](#) online).

We considered two additional filtering steps for PTCs based on the amount of transcript truncated by the PTC and whether the PTC occurred in a constituent or alternatively spliced exon. For each transcript containing a PTC, we determined the amount of the transcript truncated by the PTC. For genes with more than one PTC, the amount of each transcript that was truncated was considered independently for each PTC. All analyses and results reported in this manuscript were repeated after applying a filter to exclude PTCs that occur in the first or last 5% of the transcript and may be less likely to result in true loss-of-function. Removing PTCs that occur in the first or last 5% of the transcript from our analyses did not cause any reported results to change directionality or become statistically insignificant. As a result, we chose not to filter PTCs based on their position in the transcript. Additionally, we found that for 81% of PTC genes, the PTC is present in every alternative transcript that can be generated. Although a subset of our identified PTCs may not be full loss-of-function alleles, given that they affect only some of the known transcripts of the gene where they occur, we also chose not to discard these PTCs given that (i) it is difficult to determine the relative functional importance of different transcripts for most genes and (ii) partial loss-of-function alleles may still have some phenotypic impact. This follows



the best practices for loss-of-function alleles as outlined in (MacArthur et al. 2012).

Finally, we identified a set of 13,030 genes which contained no stop codon polymorphisms according to VEP. Genes which contained variants that were discarded at intermediate steps in our filtering pipeline are also excluded from this set. For comparisons of genes with PTCs to a genome-wide representation, we used these genes and hereafter refer to them as “control genes.”

### Expression Breadth of PTCs

In line with the hypothesis that genes with PTCs may experience less selective constraint, less widely expressed genes may be less visible to selection and may be under less selective constraint (Van Dyken and Wade 2010). Previous studies indicate that genes containing PTCs were expressed less broadly than genes without PTCs (Hoehn et al. 2012; Yang et al. 2014; Rivas et al. 2015). Thus, we hypothesized that genes with PTCs will be expressed in fewer tissues than genes with no PTCs were detected. We obtained RNA-seq fasta data from the NCBI SRA for 13 tissues (bone: SRR2045433, brain: SRR2045427, eye: SRR6456921, gills: SRR2045428, heart: SRR2045429, intestine: SRR2045434, kidney: SRR2045432, liver: SRR2045431, muscle: SRR2045430, nasal epithelium: SRR689670, ovary: SRR2045426, skin: SRR690472, testis: SRR2045435). We mapped the tissue-specific fasta files to the surface *A. mexicanus* reference genome v2.0 using STAR (version 2.7.1a, [Dobin and Gingeras 2015]) and counted the number of reads from each tissue that mapped to each PTC and control gene using HTseq (version 0.11.0, [Anders et al. 2014]). Genes with  $\geq 10$  reads mapped from a tissue specific RNAseq data were considered expressed in that tissue. We used nonparametric Kruskal–Wallis tests in R to test for significant differences between control genes and genes containing PTCs (supplementary table S2, Supplementary Material online).

### Gene Family Size

We hypothesized that genes from large gene families may tolerate segregating PTCs because the genes may be somewhat redundant in the genome. Thus, we assayed whether genes containing a PTC were associated with a larger number of paralogs than genes without PTCs. We cataloged the number of paralogs for the 3,634 genes containing PTC and 13,030 genes that did not contain a PTC using Ensembl Biomart Perl API. Kruskal–Wallis tests were performed in R to determine if gene family size was greater for genes with PTCs.

### Population Genomic Summary Statistics

We hypothesized that genes with PTCs would exhibit larger diversity/divergence than control genes because (i) they may be pseudogenized and accumulating mutations or (ii) they may be under less selective constraint than control genes and able to harbor PTCs without large fitness consequences to the organism. To distinguish between these two possibilities, we tested for significant differences between genes without PTCs detected in them (control genes), genes that contained a PTC in the focal pop (PTC-present), and genes that have a PTC in another population but not in the population being assayed (PTC-present in nonfocal pop). By comparing the “PTC-present in nonfocal pop” to the control genes, we can specifically address whether genes which contain segregating PTCs are under less selective constraint than the rest of the

genes in the genome, without the potential confounding effects of PTCs increasing diversity by pseudogenization.

We calculated absolute genetic divergence ( $D_{XY}$ ) and relative genetic divergence ( $F_{ST}$ ) between cave populations and a paired surface population from the same lineage, and nucleotide diversity ( $\pi$ ) within each population as described in (Moran et al. 2022, 2023).  $D_{XY}$ ,  $F_{ST}$ , and  $\pi$  were calculated for each the 3,634 genes with segregating PTCs and 13,030 non-PTC control genes. We used nonparametric Kruskal–Wallis tests in R to test for significant differences between non-PTC, PTC-present in focal pop, and PTC-present in non-focal pop genes for each statistic ( $\pi$ : supplementary table S3, Supplementary Material online,  $D_{XY}$ : supplementary table S4, Supplementary Material online,  $F_{ST}$ : supplementary table S5, Supplementary Material online).

### Overlap With Selective Sweeps and Established QTL

To investigate the extent to which PTC genes were associated with regions of the genome that have evidence of historical selection or are predicted to be involved in the production of cavefish traits, we leveraged a dataset from Moran et al. (2023), available at (Dryad repository, <https://doi.org/10.5061/dryad.3xsj3txmf>). This dataset contains information on the presence of selective sweeps and overlap with QTL from previous studies for each gene in the *A. mexicanus* surface reference genome (V2).

### Population Genetic Modeling in SLiM

We used the forward genetic simulation framework SLiM (Haller and Messer 2019) to explore patterns of PTC fixation and loss across a chromosome during cave evolution under different selection scenarios and compared these patterns to the PTC frequency distribution we observed in our data. We note that our work should be interpreted with the caveat that we only modeled PTCs and not additional types of variation which may impact selection across the genome. We modeled the SLiM chromosome after the real chromosome 12 of the Mexican tetra genome because it is of intermediate length and number of coding genes amongst the 25 total Mexican tetra chromosomes. Additionally, chromosome 12 is representative of an overall trend in our data, which is that of PTCs that are present in caves at a high frequency (allele frequency  $\geq 0.8$ ) approximately half are putative de novo mutations (not present in any surface populations) and half are likely sourced from standing genetic variation (present in at least one surface populations) (supplementary table S7, Supplementary Material online). We chose to model our SLiM population after the Pachón cave population because we had one of the largest sample in this study ( $n = 19$ ) from this cave, increasing the likelihood that our observations of PTCs were representative of the real population, and we have previously estimated demographic parameters for this cave, such as effective population size (Herman et al. 2018).

We generated a chromosome in SLiM equal to the length of chromosome 12 (31,403,898 bp) and added a SLiM genomic element onto our simulated chromosome for each real exon to produce accurate coding regions where PTCs could occur during our simulations. Additionally, we generated an estimated mutation rate specific to PTCs using a custom perl script (available at <https://github.com/robackem/PTC-SLiM>) and the cDNA sequences for each gene on chromosome 12 from Ensembl. For each nucleotide in the cDNA sequence, we calculated the probability that a substitution at that nucleotide



would change the existing codon into a PTC. We used a general per-bp mutation rate (i.e. the probability that a substitution occurs at a given nucleotide) of  $3.5 \times 10^{-9}$ , obtained from data on cichlids (Malinsky et al. 2018) given that a genome-wide mutation rate has not yet been established for Mexican tetra. For each nucleotide, we took the general per-bp mutation rate and multiplied it by the probability that a PTC-producing substitution occurs at the focal nucleotide. This process enabled us to generate an estimated mutation rate representing the probability that (i) a substitution occurs at a given nucleotide and (ii) the substitution produces a PTC. Subsequently, we obtained a chromosome-wide average PTC-specific mutation rate across all genes on chromosome 12 of  $1.51 \times 10^{-10}$  (bp/generation) which was used as the mutation rate in our simulation of PTCs. Generating a PTC-specific mutation rate was essential for our SLiM models given that our modeled exons do not have a nucleotide sequence to give context for where PTCs can occur, and not all SNPs produce PTCs. However, we note that our PTC-specific mutation rate is only an estimation, as it does not account for more complex mutational situations such as more than one nucleotide change occurring simultaneously within a codon. Additionally, our calculation considers each potential substitution equally and does not account for potential biases in mutational spectra such as the rate of transitions versus transversions. In addition to our modeled chromosome and PTC-specific mutation rate, we used a median genome-wide recombination rate of  $1.16 \times 10^{-6}$  cM/bp obtained from a previously published linkage map for Mexican tetra (O'Quin et al. 2013). In all simulations, we instructed SLiM not to discard mutations that reached fixation; therefore, fixed mutations were present in the final output. These core elements of our simulations remain the same for all models described below.

Our first objective using SLiM was to explore the potential strength and direction of selection on PTCs in caves. In the Pachón cave population, we documented 20 total PTCs on chromosome 12, including three high-frequency PTCs likely sourced from standing genetic variation found in surface populations and four high frequency, putatively de novo PTCs. We used this pattern of PTC presence and frequency as our observed outcome that we compared with experimental patterns of PTC frequency from SLiM simulations with different selective pressures. Additionally, we addressed the possibility that the observed finding of 20 PTCs is an underestimate due to variant filtering by identifying the number of PTCs in Pachón prior to filtering for number of individuals genotyped, expression, and checking intron/exon structure across species. We found that there are 29 PTCs in Pachón on chromosome 12 before these filtering steps (PTCs at this level of filtering can be found on [supplementary table S12](#), [Supplementary Material](#) online), indicating that our variant filtering pipeline did not substantially reduce the number of PTCs.

Across our three surface populations in this study (Río Choy, Mante, and Rascón), there are a total of 86 PTCs on chromosome 12 ([supplementary table S7a](#), [Supplementary Material](#) online). We took the average frequency of these PTCs across the surface populations where the PTC was present and seeded our starting SLiM population with these 86 PTCs at the average frequency so that they were present as initial standing genetic variation in our models. This mimics the primary way Mexican tetra are thought to have colonized caves—through stream capture of a surface drainage into a subterranean system (Mitchell et al. 1977). We refer to this group of PTCs as “standing PTCs” going forward. Each of our

models begins with the 86 standing PTCs (which may increase or decrease from their initial frequency throughout time in the model) and gains de novo PTCs within modeled exons at a rate of  $1.51 \times 10^{-10}$  bp/generation over the course of 175,000 generations. This time period was selected as our simulation length because the divergence of cave populations from their phylogenetically closest surface population likely occurred between approximately 161 and 191 thousand generations ago (Herman et al. 2018). We then generated a population of 32,000 individuals in SLiM according to the estimated effective population size of Pachón (Herman et al. 2018) whose genomes consisted of our modeled chromosome. For all models, we instructed SLiM to output 19 random individuals from the population after 175,000 generations whose genomes we then compared with the observed 19 Pachón cavefish samples.

We explored the fate and frequency of standing and de novo PTCs under three different DFE. The specific model parameters and outcomes for all models are given in [supplementary table S7](#), [Supplementary Material](#) online and the code for each SLiM model can be found at [www.github.com/robackem/PTC-SLiM](https://github.com/robackem/PTC-SLiM). First, we modeled a scenario where all PTCs are completely neutral and the selection coefficient for all PTCs was fixed at a value of 0 (no effect on individual fitness) ([supplementary table S7b](#), [Supplementary Material](#) online). Second, we modeled scenarios where each PTC received a selection coefficient sampled from a gamma DFE with a specified mean and shape parameter. Under this distribution, PTCs never receive a selection coefficient  $< 0$ , (i.e. PTCs range from nearly neutral to strongly deleterious as influenced by the distribution mean) ([supplementary table S7c](#), [Supplementary Material](#) online). Finally, we modeled scenarios where each PTC received a selection coefficient sampled from a normal DFE with a specified mean and standard deviation. Under this distribution, PTCs range from beneficial (selection coefficient  $< 0$ ) to deleterious (selection coefficient  $> 0$ ) as influenced by the distribution mean and standard deviation ([supplementary table S7d](#), [Supplementary Material](#) online). For each scenario, we repeated the simulation 100 times, and the presence and frequency of the modeled PTCs was compared with the observed frequency distribution on chromosome 12 in the Pachón population.

Our second objective using SLiM was to explore whether our observation that cave populations contain a significantly higher proportion of high-frequency PTCs than surface populations could be simply due to the smaller population size, and subsequently increased genetic drift, in caves. To investigate this, we re-ran the model that fit our observed PTC frequency distribution on chromosome 12 most closely with different population sizes (compared with our initial models using 32,000 individuals) to see how the proportion of high-frequency PTCs would change ([supplementary table S7e](#), [Supplementary Material](#) online). For our large population scenario, we used the estimated effective population size of surface population Río Choy (250,000 individuals). Additionally, we explored two scenarios with smaller populations; the first, with 8,500 individuals as predicted by historical mark-recapture census estimate (Mitchell et al. 1977), which is also within the 95% confidence interval for Pachón in (Herman et al. 2018) and the second with 1,000 individuals, as some estimates place the effective population size of Pachón cave is 150 to 1,150 individuals (Avisé and Selander 1972; Bradic et al. 2012; Fumey et al. 2018).

To further test the robustness of our SLiM modeling of PTCs we also simulated the accumulation of synonymous

SNPs in Pachón cave from the Río Choy surface reference (supplementary table S7f, Supplementary Material online). First, we identified the number of unique synonymous SNPs present in our sampling of Pachón fish using VEP v.100 (McLaren et al. 2016), with the *A. mexicanus* gtf file Ensembl release v100, and the *A. mexicanus* reference genome v2.0 (GCA\_000372685.2). We then calculated an estimated synonymous mutation rate using the same process as described for the PTC-specific mutation rate, above (code available at <https://github.com/robackem/PTC-SLiM>) and used a runtime of 250,000 generations as per the estimated divergence time between the Pachón cave and Río Choy surface populations (Herman et al. 2018). The full details of the parameters tested, and model outcomes can be found in supplementary table S7f, Supplementary Material online.

### Identification & Functional Validation of Candidate PTCs Underlying Cave-derived Traits

To identify genes containing PTCs that are potentially involved in cave-derived phenotypes, we filtered the 4,366 computationally confirmed PTCs through several additional steps. First, we required that the PTC be at a frequency  $\geq 0.80$  in one or more caves, given that alleles which are close to fixation in cave populations represent particularly strong candidates for cave-derived phenotypes (236 PTCs in 231 genes) (supplementary table S16, Supplementary Material online). Second, to identify PTCs that could be solely responsible for large phenotypic effects in cavefish, we included only PTCs for which surface populations were not homozygous (bringing our candidate gene list to 167 PTCs within 163 genes) (supplementary table S17, Supplementary Material online). Finally, we curated a final set of candidate genes by requiring that the gene cannot create an alternate transcript that splices out the PTC. These steps produced a set of 105 genes that we considered as candidates for functional validation (supplementary table S18, Supplementary Material online). We explored documented functional consequences of 105 candidate PTC genes in zebrafish, mice, and humans by leveraging the databases Zfin (Bradford et al. 2022), MGI (Baldarelli et al. 2024), IMPC (Groza et al. 2022), Orphanet (Pavan et al. 2017), OMIM (Hamosh et al. 2005), and DECIPHER (Firth et al. 2009), and review of current literature (supplementary table S9, Supplementary Material online).

### Sanger Sequencing for Functional Validation

To ensure the PTCs identified computationally were not a result of sequencing error, we designed primers using Primer 3 and obtained generated primers from Integrated DNA Technologies for six genes from our candidate gene list. PCR was conducted using 12.5  $\mu$ L AccuPrime SuperMix II, 1  $\mu$ M each of forward and reverse primer, 8  $\mu$ L of molecular biology grade water, and 2.5  $\mu$ L of DNA (from fish in the population containing the PTC according to our population genetic data). We used a program of 94 °C initial denaturation followed by 32 cycles of: 2 min 94 °C 30 s, 57 °C 30 s, 68 °C 1 min, followed by a 68 °C final extension for 5 min. After sequencing, we confirmed five PTCs and were unable to confirm the PTC in one gene (supplementary table S9e, Supplementary Material online).

### CRISPR-Cas9

Functional experiments were performed in CRISPR-Cas9 injected (CRISPRant) larval fish where the *pde6c* gene was targeted.

gRNAs were designed using ChopChop software (<https://chopchop.cbu.uib.no>) to target the coding region of each gene. For *pde6c*, the gRNA was 5'- GGCCACTAACAAG AGCCCAG -3' (targeting exon 7, upstream of PTC). The gRNA oligos were synthesized (IDT) containing the gRNA target site, T7 promoter, and an overlap sequence. A DNA template was generated by annealing the gene-specific gRNA oligo and a second oligo containing an overlap sequence and the remainder of the gRNA, and the gRNA was in vitro transcribed using the T7 Megascript kit (Ambion) as described previously (Varshney et al. 2015). The gRNAs were purified using a miRNeasy mini kit (Qiagen). Cas9 protein (PNA Bio) was used to make injection mixes. 150 pg Cas9 protein and 25 pg *pde6c* gRNA were injected into single-cell surface fish embryos (Jao et al. 2013; Kowalko et al. 2016; Klaassen et al. 2018; Stahl et al. 2019).

Surface fish were maintained in the lab under standard conditions, and were derived from fish originally collected in Mexico, and bred in the lab for multiple generations. Fish were bred following standard protocols by overfeeding followed by increasing the water temperature 1 to 2 °C (Borowsky 2008). Embryos were collected following spawning and either injected with CRISPR reagents or raised as wild-type control fish. Fish were raised in glass bowls in an incubator at 25 °C until 6 dpf, and then transferred to 2 L tanks in the aquatics facility (23 °C) from 6 to 9 dpf. Fish are maintained under a 14:10 Light:dark cycle and fed brine shrimp beginning at 6 dpf.

At 9 d post-fertilization, larvae were euthanized in MS-222, placed in methyl cellulose, and imaged using a dissecting microscope (Zeiss). After imaging, genomic DNA was extracted from CRISPRants and wild-type (uninjected) controls by placing fish in 50 mM NaOH for 30 min at 95 °C, followed by addition of 1/10th volume of 1 M Tris-HCl pH 8.0. Genotyping was performed by PCR (*Pde6c* forward primer: CTGTTATGTTGAAACTGAGC, *Pde6c* reverse primer: CGTTTTCAGCTACATAAGTC) followed by gel electrophoresis to determine if mutagenesis was achieved at the locus. Fish carrying mutant alleles were identified by the presence of more than one band, appearing like a smeary band, while wild-type fish were identified by the presence of a single band.

### Optomotor Response Assay

Fish were fed and then acclimated to the assay room for a minimum of 1 h prior to starting the assay. Untreated 4-well plates (Thermo Scientific Nunc Rectangular Dishes) were filled 90% with fresh system water and placed on top of a tablet (Samsung Galaxy Tab Active PRO). Fish were placed in individual wells and acclimated to a white background on the tablet for 1 min prior to starting the assay. Following acclimation, fish are subjected to moving black and white lines (2 cm period at a speed of 1 cm/s). Lines move for 30 s in one direction before switching direction, for a total of 5 switches. The program used to create the white background and moving lines is available at <https://github.com/yffily/moving-stripes>. Assays were filmed from above using a 1,920 × 1,200 Mono USB 3.0 Grasshopper Camera and 12 mm HP Series lens, 1.1" at a rate of 30 fps. After recording, CRISPRant and uninjected sibling fish were collected and imaged for eye size the next day. Fish were then euthanized and genotyped to verify mutagenesis in CRISPRants.

OMR was analyzed using Fiji (Schneider et al. 2012). The distance that the fish moved in the direction of the line was

measured across five direction changes. The distances were then averaged, and the proportion of distance traveled in the direction of the lines was calculated by dividing the average distance traveled by the arena length. Fish that had an OMR followed the direction of the lines for greater distance and therefore have a higher proportion of distance traveled following the lines.

## Phenotyping

Measurements of eye area, pupil area, and standard length were taken from images taken at 9 dpf. Measurements were performed in Fiji. Eye area was measured by tracing the outline of the eye. Similarly, the pupil area was measured tracing the pupil. Standard length was measured from snout to caudal peduncle. Eye area was corrected for standard length by dividing eye diameter by fish length. Only fish confirmed to be mutagenic, and for which we had both the right and left side imaged, were used in data analysis comparing to wild type fish. All analyses were performed using R Statistical Software.

## Supplementary Material

[Supplementary material](#) is available at *Molecular Biology and Evolution* online.

## Acknowledgments

The Minnesota Supercomputing Institute (MSI) at the University of Minnesota provided computing resources that contributed to the research results reported within this article.

## Funding

Funding was supported by National Institutes of Health (1R01GM127872-01 to S.E.M. and N.R.; R15HD099022, and R35GM138345 to J.K.) and National Science Foundation (IOS 1933076/IOS 2202359, and IOS EDGE 1923372 to J.K., S.E.M., and N.R.; DEB 2147597, and DEB 2316784 to J.K.; and DEB 00109505 to J.K. and S.E.M.). Fish were collected under CONAPESCA permit SGPA/DGVS/266/15 to C.P.O.-G.

## Data Availability

The data underlying this article, including all custom scripts and modeling, are available in the article and in its online [Supplementary Material](#).

## References

- Albalat R, Cañestro C. Evolution by gene loss. *Nat Rev Genet*. 2016;17(7):379–391. <https://doi.org/10.1038/nrg.2016.39>.
- Anders S, Pyl PT, Huber W. HTSeq—a Python framework to work with high-throughput sequencing data. *Bioinformatics*. 2014;31(2):166–169. <https://doi.org/10.1093/bioinformatics/btu638>.
- Aspiras AC, Rohner N, Martineau B, Borowsky RL, Tabin CJ. Melanocortin 4 receptor mutations contribute to the adaptation of cavefish to nutrient-poor conditions. *Proc Natl Acad Sci U S A*. 2015;112(31):9668–9673. <https://doi.org/10.1073/pnas.1510802112>.
- Asselta R, Duga S, Spena S, Santagostino E, Peyvandi F, Piseddu G, Targhetta R, Malcovati M, Mannucci PM, Tenchini ML. Congenital afibrinogenemia: mutations leading to premature termination codons in fibrinogen A $\alpha$ -chain gene are not associated with the decay of the mutant mRNAs. *Blood*. 2001;98(13):3685–3692. <https://doi.org/10.1182/blood.V98.13.3685>.
- Awise JC, Selander RK. Evolutionary genetics of cave-dwelling fishes of the genus *Astyanax*. *Evolution*. 1972;26(1):1–19. <https://doi.org/10.1111/j.1558-5646.1972.tb00170.x>.
- Baldarelli RM, Smith CL, Ringwald M, Richardson JE, Bult CJ. Mouse genome informatics: an integrated knowledgebase system for the laboratory mouse. *Genetics*. 2024;227(1):iyae031. <https://doi.org/10.1093/genetics/iyae031>.
- Barrett P, Ivanova E, Graham ES, Ross AW, Wilson D, Plé H, Mercer JG, Ebling FJ, Schuhler S, Dupré SM, *et al*. Photoperiodic regulation of cellular retinoic acid-binding protein 1, GPR50 and nestin in tanyocytes of the third ventricle ependymal layer of the Siberian hamster. *J Endocrinol*. 2006;191(3):687–698. <https://doi.org/10.1677/joe.1.06929>.
- Beale A, Guibal C, Tamai TK, Klotz L, Cowen S, Peyric E, Reynoso VH, Yamamoto Y, Whitmore D. Circadian rhythms in Mexican blind cavefish *Astyanax mexicanus* in the lab and in the field. *Nat Comm*. 2013;4(1):2769. <https://doi.org/10.1038/ncomms3769>.
- Bechtold DA, Sidibe A, Saer BR, Li J, Hand LE, Ivanova EA, Darras VM, Dam J, Jockers R, Luckman SM, *et al*. A role for the melatonin-related receptor GPR50 in leptin signaling, adaptive thermogenesis, and torpor. *Curr Biol*. 2012;22(1):70–77. <https://doi.org/10.1016/j.cub.2011.11.043>.
- Behrmann-Godel J, Nolte AW, Kreiselaier J, Berka R, Freyhof J. The first European cave fish. *Curr Biol*. 2017;27(7):R257–R258. <https://doi.org/10.1016/j.cub.2017.02.048>.
- Blumer M, Brown T, Freitas MB, Destro AL, Oliveira JA, Morales AE, Schell T, Greve C, Pippel M, Jebb D, *et al*. Gene losses in the common vampire bat illuminate molecular adaptations to blood feeding. *Sci Adv*. 2022;8(12):eabm6494. <https://doi.org/10.1126/sciadv.abm6494>.
- Borowsky R. Breeding *Astyanax mexicanus* through natural spawning. *CSH Protoc*. 2008;2008:pdb.prot5091. <https://doi.org/10.1101/pdb.prot5091>.
- Bradford YM, Van Slyke CE, Ruzicka L, Singer A, Eagle A, Fashena D, Howe DG, Frazer K, Martin R, Paddock H, *et al*. Zebrafish information network, the knowledgebase for Danio rerio research. *Genetics*. 2022;220(4):iyac016. <https://doi.org/10.1093/genetics/iyac016>.
- Bradic M, Beerli P, Garcia-de Leon FJ, Esquivel-Bobadilla S, Borowsky RL. Gene flow and population structure in the Mexican blind cavefish complex (*Astyanax mexicanus*). *BMC Evol Biol*. 2012;12(1):9. <https://doi.org/10.1186/1471-2148-12-9>.
- Calderoni L, Rota-Stabelli O, Frigato E, Panziera A, Kirchner S, Foulkes NS, Kruckenhauser L, Bertolucci C, Fuselli S. Relaxed selective constraints drove functional modifications in peripheral photoreception of the cavefish *P. andruzzii* and provide insight into the time of cave colonization. *Heredity (Edinb)*. 2016;117(5):383–392. <https://doi.org/10.1038/hdy.2016.59>.
- Charlton-Perkins M, Almeida AD, MacDonald RB, Harris WA. Genetic control of cellular morphogenesis in Müller glia. *Glia*. 2019;67(7):1401–1411. <https://doi.org/10.1002/glia.23615>.
- Chin JSR, Gassant CE, Amaral PM, Lloyd E, Stahl BA, Jaggard JB, Keene AC, Duboue ER. Convergence on reduced stress behavior in the Mexican blind cavefish. *Dev Biol*. 2018;441(2):319–327. <https://doi.org/10.1016/j.ydbio.2018.05.009>.
- Choy S, Thakur S, Polyakov E, Abdelaziz J, Lloyd E, Enriquez M, Jayan N, Fily Y, McGaugh S, Keene AC, *et al*. Mutations in the albinism gene *oca2* alter vision-dependent prey capture behavior in the Mexican tetra. Unpublished data. bioRxiv. 2024. <https://doi.org/10.1101/2024.06.17.599419>, 20 June 2024, preprint: not peer reviewed.
- Conant GC, Wagner A. Duplicate genes and robustness to transient gene knock-downs in *Caenorhabditis elegans*. *Proc R Soc Lond B Biol Sci*. 2004;271(1534):89–96. <https://doi.org/10.1098/rspb.2003.2560>.
- Cronin T, Raffelsberger W, Lee-Rivera I, Jaillard C, Niepon ML, Kinzel B, Clérin E, Petrosian A, Picaud S, Poch O, *et al*. The disruption of the rod-derived cone viability gene leads to photoreceptor



- dysfunction and susceptibility to oxidative stress. *Cell Death Differ.* 2010;17(7):1199–1210. <https://doi.org/10.1038/cdd.2010.2>.
- Cunningham RL, Herbert AL, Harty BL, Ackerman SD, Monk KR. Mutations in *dock1* disrupt early Schwann cell development. *Neural Dev.* 2018;13(1):17. <https://doi.org/10.1186/s13064-018-0114-9>.
- Dobin A, Gingeras TR. Mapping RNA-seq reads with STAR. *Curr Protoc Bioinform.* 2015;51(1):11.14.11–11.14.19. <https://doi.org/10.1002/0471250953.bi1114s51>.
- Duboué ER, Keene AC, Borowsky RL. Evolutionary convergence on sleep loss in cavefish populations. *Curr Biol.* 2011;21(8):671–676. <https://doi.org/10.1016/j.cub.2011.03.020>.
- Dufourny L, Levesaur A, Migaud M, Callebaut I, Pontarotti P, Malpoux B, Monget P. GPR50 is the mammalian ortholog of *Mel1c*: evidence of rapid evolution in mammals. *BMC Evol Biol.* 2008;8(1):1–14. <https://doi.org/10.1186/1471-2148-8-105>.
- Eliot Y, Hinaux H, Callebaut J, Rétaux S. Evolutionary shift from fighting to foraging in blind cavefish through changes in the serotonin network. *Curr Biol.* 2013;23(1):1–10. <https://doi.org/10.1016/j.cub.2012.10.044>.
- Emerling CA, Springer MS. Eyes underground: regression of visual protein networks in subterranean mammals. *Mol Phylogen Evol.* 2014;78:260–270. <https://doi.org/10.1016/j.ympev.2014.05.016>.
- Emerling CA, Springer MS. Genomic evidence for rod monochromacy in sloths and armadillos suggests early subterranean history for *Xenarthra*. *Proc R Soc Lond B Biol Sci.* 2015;282:20142192. <https://doi.org/10.1098/rspb.2014.2192>.
- Everts RE, Rothuizen J, van Oost BA. Identification of a premature stop codon in the melanocyte-stimulating hormone receptor gene (*MC1R*) in Labrador and Golden retrievers with yellow coat colour. *Anim Genet.* 2000;31(3):194–199. <https://doi.org/10.1046/j.1365-2052.2000.00639.x>.
- Fernández R, Gabaldón T. Gene gain and loss across the metazoan tree of life. *Nat Ecol Evol.* 2020;4(4):524–533. <https://doi.org/10.1038/s41559-019-1069-x>.
- Firth HV, Richards SM, Bevan AP, Clayton S, Corpes M, Rajan D, Van Vooren S, Moreau Y, Pettett RM, Carter NP. DECIPHER: database of chromosomal imbalance and phenotype in humans using ensembl resources. *Am J Hum Genet.* 2009;84(4):524–533. <https://doi.org/10.1016/j.ajhg.2009.03.010>.
- Fontanesi L, Beretti F, Riggio V, Dall'Olio S, González EG, Finocchiaro R, Davoli R, Russo V, Portolano B. Missense and nonsense mutations in melanocortin 1 receptor (*MC1R*) gene of different goat breeds: association with red and black coat colour phenotypes but with unexpected evidences. *BMC Genet.* 2009;10(1):47. <https://doi.org/10.1186/1471-2156-10-47>.
- Franze K, Grosche J, Skatchkov SN, Schinkinger S, Foja C, Schild D, Uckermann O, Travis K, Reichenbach A, Guck J. Müller cells are living optical fibers in the vertebrate retina. *Proc Natl Acad Sci U S A.* 2007;104(20):8287–8292. <https://doi.org/10.1073/pnas.0611180104>.
- Fumey J, Hinaux H, Noirot C, Thermes C, Rétaux S, Casane D. Evidence for late Pleistocene origin of *Astyanax mexicanus* cavefish. *BMC Evol Biol.* 2018;18(1):43. <https://doi.org/10.1186/s12862-018-1156-7>.
- Garduño-Sánchez MAA, De Jesus-Bonilla V, Perea S, Miranda-Gamboa R, Herrera-García A, De la Maza Benignos M, Ornelas-García CP. Mitochondrial phylogeography and molecular evolution of the rhodopsin visual pigment in troglolitic populations of *Astyanax mexicanus* (De Filippi, 1853). *Zool Res.* 2023;44(4):761–775. <https://doi.org/10.24272/z.issn.2095-8137.2022.437>.
- Giebel LB, Musarella MA, Spritz RA. A nonsense mutation in the tyrosinase gene of Afghan patients with tyrosinase negative (type IA) oculocutaneous albinism. *J Med Genet.* 1991;28(7):464–467. <https://doi.org/10.1136/jmg.28.7.464>.
- Gross JB, Borowsky R, Tabin CJ. A novel role for *Mc1r* in the parallel evolution of depigmentation in independent populations of the cavefish *Astyanax mexicanus*. *PLoS Genet.* 2009;5(1):e1000326. <https://doi.org/10.1371/journal.pgen.1000326>.
- Groza T, Gomez FL, Mashhadi HH, Muñoz-Fuentes V, Gunes O, Wilson R, Cacheiro P, Frost A, Keskivali-Bond P, Vardal B, et al. The international mouse phenotyping consortium: comprehensive knockout phenotyping underpinning the study of human disease. *Nucleic Acids Res.* 2022;51(D1):D1038–D1045. <https://doi.org/10.1093/nar/gkac972>.
- Guijarro-Clarke C, Holland PWH, Paps J. Widespread patterns of gene loss in the evolution of the animal kingdom. *Nat Ecol Evol.* 2020;4:519–523. <https://doi.org/10.1038/s41559-020-1129-2>.
- Haller BC, Messer PW. SLiM 3: forward genetic simulations beyond the wright–fisher model. *Mol Biol Evol.* 2019;36(3):632–637. <https://doi.org/10.1093/molbev/msy228>.
- Hamosh A, Scott AF, Amberger JS, Bocchini CA, McKusick VA. Online Mendelian inheritance in man (OMIM), a knowledgebase of human genes and genetic disorders. *Nucleic Acids Res.* 2005;33(Database issue):D514–D517. <https://doi.org/10.1093/nar/gki033>.
- Harries LW, Bingham C, Bellanne-Chantelot C, Hattersley AT, Ellard S. The position of premature termination codons in the hepatocyte nuclear factor –1 beta gene determines susceptibility to nonsense-mediated decay. *Hum Genet.* 2005;118(2):214–224. <https://doi.org/10.1007/s00439-005-0023-y>.
- Herman A, Brandvain Y, Weagley J, Jeffery WR, Keene AC, Kono TJY, Bilandžija H, Borowsky R, Espinosa L, O'Quin K, et al. The role of gene flow in rapid and repeated evolution of cave-related traits in Mexican tetra, *Astyanax mexicanus*. *Mol Ecol.* 2018;27(22):4397–4416. <https://doi.org/10.1111/mec.14877>.
- Hoehn KB, McGaugh SE, Noor MAF. Effects of premature termination codon polymorphisms in the *Drosophila pseudoobscura* subclade. *J Mol Evol.* 2012;75(3-4):141–150. <https://doi.org/10.1007/s00239-012-9528-x>.
- Hottes AK, Freddolino PL, Khare A, Donnell ZN, Liu JC, Tavazoie S. Bacterial adaptation through loss of function. *PLoS Genet.* 2013;9(7):e1003617. <https://doi.org/10.1371/journal.pgen.1003617>.
- Hu Y, Wu Q, Ma S, Ma T, Shan L, Wang X, Nie Y, Ning Z, Yan L, Xiu Y, et al. Comparative genomics reveals convergent evolution between the bamboo-eating giant and red pandas. *Proc Natl Acad Sci U S A.* 2017;114(5):1081–1086. <https://doi.org/10.1073/pnas.1613870114>.
- Iribarne M, Nishiwaki Y, Nakamura S, Araragi M, Oguri E, Masai I. Aipl1 is required for cone photoreceptor function and survival through the stability of Pde6c and Gc3 in zebrafish. *Sci Rep.* 2017;7(1):45962. <https://doi.org/10.1038/srep45962>.
- Ivanova EA, Bechtold DA, Dupré SM, Brennand J, Barrett P, Luckman SM, Loudon ASI. Altered metabolism in the melatonin-related receptor (GPR50) knockout mouse. *Am J Physiol Endocrinol Metabol.* 2008;294(1):E176–E182. <https://doi.org/10.1152/ajpendo.00199.2007>.
- Jaggard J, Robinson BG, Stahl BA, Oh I, Masek P, Yoshizawa M, Keene AC. The lateral line confers evolutionarily derived sleep loss in the Mexican cavefish. *J Exp Biol.* 2017;220(2):284–293. <https://doi.org/10.1242/jeb.145128>.
- Jaggard JB, Stahl BA, Lloyd E, Prober DA, Duboué ER, Keene AC. Hypocretin underlies the evolution of sleep loss in the Mexican cavefish. *Elife.* 2018;7:e32637. <https://doi.org/10.7554/eLife.32637>.
- Jao L-E, Wentz SR, Chen W. Efficient multiplex biallelic zebrafish genome editing using a CRISPR nuclease system. *Proc Natl Acad Sci U S A.* 2013;110(34):13904–13909. <https://doi.org/10.1073/pnas.1308335110>.
- Jeffery WR. Regressive evolution in *Astyanax* cavefish. *Annu Rev Genet.* 2009;43(1):25–47. <https://doi.org/10.1146/annurev-genet-102108-134216>.
- Kern AD, Schrider DR. Diplos/HIC: an updated approach to classifying selective sweeps. *G3.* 2018;8(6):1959–1970. <https://doi.org/10.1534/g3.118.200262>.
- Kimura M. The neutral theory of molecular evolution. *Sci Am.* 1979;241(5):98–129. <https://doi.org/10.1038/scientificamerican1179-98>.
- Klaassen H, Wang Y, Adamski K, Rohner N, Kowalko JE. CRISPR mutagenesis confirms the role of *oca2* in melanin pigmentation in



- Astyanax mexicanus*. *Dev Biol*. 2018;441(2):313–318. <https://doi.org/10.1016/j.ydbio.2018.03.014>.
- Klaus S, Mendoza JC, Liew JH, Plath M, Meier R, Yeo DC. Rapid evolution of troglomorphic characters suggests selection rather than neutral mutation as a driver of eye reduction in cave crabs. *Biol Lett*. 2013;9(2):20121098. <https://doi.org/10.1098/rsbl.2012.1098>.
- Kowalko JE, Ma L, Jeffery WR. Genome editing in *Astyanax mexicanus* using transcription activator-like effector nucleases (TALENs). *J Vis Exp*. 2016;112:e54113. <https://doi.org/10.3791/54113>.
- Kowalko JE, Rohner N, Rompani SB, Peterson BK, Linden TA, Yoshizawa M, Kay EH, Weber J, Hoekstra HE, Jeffery WR, et al. Loss of schooling behavior in cavefish through sight-dependent and sight-independent mechanisms. *Curr Biol*. 2013;23(19):1874–1883. <https://doi.org/10.1016/j.cub.2013.07.056>.
- Krishnan J, Rohner N. Cavefish and the basis for eye loss. *Philos Trans R Soc Lond B Biol Sci*. 2017;372(1713):20150487. <https://doi.org/10.1098/rstb.2015.0487>.
- Langille BL, Tierney SM, Bertozzi T, Beasley-Hall PG, Bradford TM, Fagan-Jeffries EP, Hyde J, Leijes R, Richardson M, Saint KM, et al. Parallel decay of vision genes in subterranean water beetles. *Mol Phylogenet Evol*. 2022;173:107522. <https://doi.org/10.1016/j.ympev.2022.107522>.
- Lee YCG, Reinhardt JA. Widespread polymorphism in the positions of stop codons in *Drosophila melanogaster*. *Genome Biol Evol*. 2011;4(4):533–549. <https://doi.org/10.1093/gbe/evr113>.
- LeFauve MK, Rowe CJ, Crowley-Perry M, Wiegand JL, Shapiro AG, Connaughton VP. Using a variant of the optomotor response as a visual defect detection assay in zebrafish. *J Biol Methods*. 2021;8(1):e144. <https://doi.org/10.14440/jbm.2021.341>.
- MacArthur DG, Balasubramanian S, Frankish A, Huang N, Morris J, Walter K, Jostins L, Habegger L, Pickrell JK, Montgomery SB, et al. A systematic survey of loss-of-function variants in human protein-coding genes. *Science*. 2012;335(6070):823–828. <https://doi.org/10.1126/science.1215040>.
- MacArthur DG, Tyler-Smith C. Loss-of-function variants in the genomes of healthy humans. *Hum Mol Genet*. 2010;19(R2):R125–R130. <https://doi.org/10.1093/hmg/ddq365>.
- Mack KL, Jaggard JB, Persons JL, Roback EY, Passow CN, Stanhope BA, Ferrufino E, Tsuchiya D, Smith SE, Slaughter BD, et al. Repeated evolution of circadian clock dysregulation in cavefish populations. *PLoS Genet*. 2021;17(7):e1009642. <https://doi.org/10.1371/journal.pgen.1009642>.
- Malinsky M, Svardal H, Tyers AM, Miska EA, Genner MJ, Turner GF, Durbin R. Whole-genome sequences of Malawi cichlids reveal multiple radiations interconnected by gene flow. *Nat Ecol Evol*. 2018;2(12):1940–1955. <https://doi.org/10.1038/s41559-018-0717-x>.
- McLaren W, Gil L, Hunt SE, Riat HS, Ritchie GRS, Thormann A, Flicek P, Cunningham F. The Ensembl variant effect predictor. *Genome Biol*. 2016;17(1):122. <https://doi.org/10.1186/s13059-016-0974-4>.
- Medley JK, Persons J, Biswas T, Olsen L, Peuss R, Krishnan J, Xiong S, Rohner N. The metabolome of Mexican cavefish shows a convergent signature highlighting sugar, antioxidant, and ageing-related metabolites. *Elife*. 2022;11:e74539. <https://doi.org/10.7554/eLife.74539>.
- Miller JN, Pearce DA. Nonsense-mediated decay in genetic disease: friend or foe? *Mutat Res Rev Mutat Res*. 2014;762:52–64. <https://doi.org/10.1016/j.mrrrev.2014.05.001>.
- Mitchell RW, Russell WH, Elliott WR. *Mexican eyeless characin fishes, genus Astyanax: environment, distribution, and evolution*. Texas Tech University: The Museum Special Publications No. 12. 1977.
- Monroe JG, McKay JK, Weigel D, Flood PJ. The population genomics of adaptive loss of function. *Heredity (Edinb)*. 2021;126(3):383–395. <https://doi.org/10.1038/s41437-021-00403-2>.
- Moran D, Softley R, Warrant EJ. Eyeless Mexican cavefish save energy by eliminating the circadian rhythm in metabolism. *PLoS One*. 2014;9(9):e107877. <https://doi.org/10.1371/journal.pone.0107877>.
- Moran D, Softley R, Warrant EJ. The energetic cost of vision and the evolution of eyeless Mexican cavefish. *Sci Adv*. 2015;1(8):e1500363. <https://doi.org/10.1126/sciadv.1500363>.
- Moran RL, Jaggard JB, Roback EY, Kenzior A, Rohner N, Kowalko JE, Ornelas-García CP, McGaugh SE, Keene AC. Hybridization underlies localized trait evolution in cavefish. *iScience*. 2022;25(2):103778. <https://doi.org/10.1016/j.isci.2022.103778>.
- Moran RL, Richards EJ, Ornelas-García CP, Gross JB, Donny A, Wiese J, Keene AC, Kowalko JE, Rohner N, McGaugh SE. Selection-driven trait loss in independently evolved cavefish populations. *Nat Comm*. 2023;14(1):2557. <https://doi.org/10.1038/s41467-023-37909-8>.
- Murray AW. Can gene-inactivating mutations lead to evolutionary novelty? *Curr Biol*. 2020;30(10):R465–R471. <https://doi.org/10.1016/j.cub.2020.03.072>.
- Nagy E, Maquat LE. A rule for termination-codon position within intron-containing genes: when nonsense affects RNA abundance. *Trends Biochem Sci*. 1998;23(6):198–199. [https://doi.org/10.1016/S0968-0004\(98\)01208-0](https://doi.org/10.1016/S0968-0004(98)01208-0).
- Ng PC, Levy S, Huang J, Stockwell TB, Walenz BP, Li K, Axelrod N, Busam DA, Strausberg RL, Venter JC. Genetic variation in an individual human exome. *PLoS Genet*. 2008;4(8):e1000160. <https://doi.org/10.1371/journal.pgen.1000160>.
- Niemiller ML, Fitzpatrick BM, Miller BT. Recent divergence with gene flow in Tennessee cave salamanders (Plethodontidae: Gyrinophilus) inferred from gene genealogies. *Mol Ecol*. 2008;17(9):2258–2275. <https://doi.org/10.1111/j.1365-294X.2008.03750.x>.
- O’Gorman M, Thakur S, Imrie G, Moran RL, Choy S, Sifuentes-Romero I, Bilandžija H, Renner KJ, Duboué E, Rohner N, et al. Pleiotropic function of the oca2 gene underlies the evolution of sleep loss and albinism in cavefish. *Curr Biol*. 2021;31(16):3694–3701.e4. <https://doi.org/10.1016/j.cub.2021.06.077>.
- O’Quin KE, Yoshizawa M, Doshi P, Jeffery WR. Quantitative genetic analysis of retinal degeneration in the blind cavefish *Astyanax mexicanus*. *PLoS One*. 2013;8(2):e57281. <https://doi.org/10.1371/journal.pone.0057281>.
- Pavan S, Rommel K, Mateo Marquina ME, Höhn S, Lanneau V, Rath A. Clinical practice guidelines for rare diseases: the orphanet database. *PLoS One*. 2017;12(1):e0170365. <https://doi.org/10.1371/journal.pone.0170365>.
- Pittler SJ, Baehr W. Identification of a nonsense mutation in the rod photoreceptor cGMP phosphodiesterase beta-subunit gene of the rd mouse. *Proc Natl Acad Sci U S A*. 1991;88(19):8322–8326. <https://doi.org/10.1073/pnas.88.19.8322>.
- Poulson TL, White WB. The cave environment: limestone caves provide unique natural laboratories for studying biological and geological processes. *Science*. 1969;165(3897):971–981. <https://doi.org/10.1126/science.165.3897.971>.
- Protas ME, Hersey C, Kochanek D, Zhou Y, Wilkens H, Jeffery WR, Zon LI, Borowsky R, Tabin CJ. Genetic analysis of cavefish reveals molecular convergence in the evolution of albinism. *Nat Genet*. 2006;38(1):107–111. <https://doi.org/10.1038/ng1700>.
- Puk O, Ahmad N, Wagner S, Hrabé de Angelis M, Graw J. First mutation in the βA2-crystallin encoding gene is associated with small lenses and age-related cataracts. *Invest Ophthalmol Vis Sci*. 2011;52(5):2571–2576. <https://doi.org/10.1167/iovs.10-6443>.
- Quinlan AR. BEDTools: the Swiss-army tool for genome feature analysis. *Curr Protoc Bioinform*. 2014;47(1):11.12.11–11.12.34. <https://doi.org/10.1002/0471250953.bi1112s47>.
- Reis LM, Tyler RC, Muheisen S, Raggio V, Salvati L, Han DP, Costakos D, Yonath H, Hall S, Power P, et al. Whole exome sequencing in dominant cataract identifies a new causative factor, CRYBA2, and a variety of novel alleles in known genes. *Hum Genet*. 2013;132(7):761–770. <https://doi.org/10.1007/s00439-013-1289-0>.
- Riddle MR, Aspiras AC, Gaudenz K, Peuß R, Sung JY, Martineau B, Peavey M, Box AC, Tabin JA, McGaugh S, et al. Insulin resistance in cavefish as an adaptation to a nutrient-limited environment. *Nature*. 2018;555(7698):647–651. <https://doi.org/10.1038/nature26136>.
- Rio Frio T, Wade NM, Ransijn A, Berson EL, Beckmann JS, Rivolta C. Premature termination codons in PRPF31 cause retinitis pigmentosa via haploinsufficiency due to nonsense-mediated mRNA decay.

- J Clin Invest.* 2008;118(4):1519–1531. <https://doi.org/10.1172/JCI34211>.
- Rivas MA, Pirinen M, Conrad DF, Lek M, Tsang EK, Karczewski KJ, Maller JB, Kukurba KR, DeLuca DS, Fromer M, *et al.* Effect of predicted protein-truncating genetic variants on the human transcriptome. *Science.* 2015;348(6235):666–669. <https://doi.org/10.1126/science.1261877>.
- Roberts M, Josephs EB. Weaker selection on genes with treatment-specific expression consistent with a limit on plasticity evolution in *Arabidopsis thaliana*. *Genetics.* 2023;224(2):iyad074. <https://doi.org/10.1093/genetics/iyad074>.
- Rosenfeld PJ, Cowley GS, McGee TL, Sandberg MA, Berson EL, Dryja TP. A null mutation in the rhodopsin gene causes rod photoreceptor dysfunction and autosomal recessive retinitis pigmentosa. *Nat Genet.* 1992;1(3):209–213. <https://doi.org/10.1038/ng0692-209>.
- Saade CJ, Alvarez-Delfin K, Fadool JM. Rod photoreceptors protect from cone degeneration-induced retinal remodeling and restore visual responses in zebrafish. *J Neurosci.* 2013;33(5):1804–1814. <https://doi.org/10.1523/JNEUROSCI.2910-12.2013>.
- Saleheen D, Natarajan P, Armean IM, Zhao W, Rasheed A, Khetarpal SA, Won H-H, Karczewski KJ, O'Donnell-Luria AH, Samocha KE, *et al.* Human knockouts and phenotypic analysis in a cohort with a high rate of consanguinity. *Nature.* 2017;544(7649):235–239. <https://doi.org/10.1038/nature22034>.
- Sarria I, Cao Y, Wang Y, Ingram NT, Orlandi C, Kamasawa N, Kolesnikov AV, Pahlberg J, Kefalov VJ, Sampath AP, *et al.* LRIT1 modulates adaptive changes in synaptic communication of cone photoreceptors. *Cell Rep.* 2018;22(13):3562–3573. <https://doi.org/10.1016/j.celrep.2018.03.008>.
- Schilthuizen M, Cabanban A, Haase M. Possible speciation with gene flow in tropical cave snails. *J Zool Syst Evol Res.* 2005;43(2):133–138. <https://doi.org/10.1111/j.1439-0469.2004.00289.x>.
- Schneider CA, Rasband WS, Eliceiri KW. NIH image to ImageJ: 25 years of image analysis. *Nat Methods.* 2012;9(7):671–675. <https://doi.org/10.1038/nmeth.2089>.
- Sifuentes-Romero I, Ferrufino E, Thakur S, Laboissonniere LA, Solomon M, Smith CL, Keene AC, Trimarchi JM, Kowalko JE. Repeated evolution of eye loss in Mexican cavefish: evidence of similar developmental mechanisms in independently evolved populations. *J Exp Zool B Mol Dev Evol.* 2020;334(7-8):423–437. <https://doi.org/10.1002/jez.b.22977>.
- Snell-Rood EC, Van Dyken JD, Cruickshank T, Wade MJ, Moczek AP. Toward a population genetic framework of developmental evolution: the costs, limits, and consequences of phenotypic plasticity. *Bioessays.* 2010;32(1):71–81. <https://doi.org/10.1002/bies.200900132>.
- Springer MS, Emerling CA, Fugate N, Patel R, Starrett J, Morin PA, Hayashi C, Gatesy J. Inactivation of cone-specific phototransduction genes in rod monochromatic cetaceans. *Front Ecol Evol.* 2016;4:61. <https://doi.org/10.3389/fevo.2016.00061>.
- Stahl BA, Jaggard JB, Chin JSR, Kowalko JE, Keene AC, Duboué ER. Manipulation of gene function in Mexican cavefish. *J Vis Exp.* 2019;(146):e59093. <https://doi.org/10.3791/59093>.
- Suber ML, Pittler SJ, Qin N, Wright GC, Holcombe V, Lee RH, Craft CM, Lolley RN, Baehr W, Hurwitz RL. Irish setter dogs affected with rod/cone dysplasia contain a nonsense mutation in the rod cGMP phosphodiesterase beta-subunit gene. *Proc Natl Acad Sci U S A.* 1993;90(9):3968–3972. <https://doi.org/10.1073/pnas.90.9.3968>.
- Subramanian S, Kumar S. Gene expression intensity shapes evolutionary rates of the proteins encoded by the vertebrate genome. *Genetics.* 2004;168(1):373–381. <https://doi.org/10.1534/genetics.104.028944>.
- Tavares WC, Seuánez HN. Changes in selection intensity on the mitogenome of subterranean and fossorial rodents respective to above-ground species. *Mamm Genome.* 2018;29(5-6):353–363. <https://doi.org/10.1007/s00335-018-9748-5>.
- Ueno A, Omori Y, Sugita Y, Watanabe S, Chaya T, Kozuka T, Kon T, Yoshida S, Matsushita K, Kuwahara R, *et al.* Lrit1, a retinal transmembrane protein, regulates selective synapse formation in cone photoreceptor cells and visual acuity. *Cell Rep.* 2018;22(13):3548–3561. <https://doi.org/10.1016/j.celrep.2018.03.007>.
- Utzeri VJ, Ribani A, Fontanesi L. A premature stop codon in the *TYRP1* gene is associated with brown coat colour in the European rabbit (*Oryctolagus cuniculus*). *Anim Genet.* 2014;45(4):600–603. <https://doi.org/10.1111/age.12171>.
- Van Dyken JD, Wade MJ. The genetic signature of conditional expression. *Genetics.* 2010;184(2):557–570. <https://doi.org/10.1534/genetics.109.110163>.
- Varshney GK, Pei W, LaFave MC, Idol J, Xu L, Gallardo V, Carrington B, Bishop K, Jones M, Li M, *et al.* High-throughput gene targeting and phenotyping in zebrafish using CRISPR/Cas9. *Genome Res.* 2015;25(7):1030–1042. <https://doi.org/10.1101/gr.186379.114>.
- Villacorta C, Jaume D, Oromí P, Juan C. Under the volcano: phylogeography and evolution of the cave-dwelling *Palmorchestia hypogaea* (Amphipoda, Crustacea) at La Palma (Canary Islands). *BMC Biol.* 2008;6(1):7. <https://doi.org/10.1186/1741-7007-6-7>.
- Viringipurampeer IA, Shan X, Gregory-Evans K, Zhang JP, Mohammadi Z, Gregory-Evans CY. Rip3 knockdown rescues photoreceptor cell death in blind *pde6c* zebrafish. *Cell Death Differ.* 2014;21(5):665–675. <https://doi.org/10.1038/cdd.2013.191>.
- Wangen JR, Green R. Stop codon context influences genome-wide stimulation of termination codon readthrough by aminoglycosides. *Elife.* 2020;9:e52611. <https://doi.org/10.7554/eLife.52611>.
- Wilkins H. The role of selection in the evolution of blindness in cave fish. *Biol J Linn Soc.* 2020;130(3):421–432. <https://doi.org/10.1093/biolinnean/blaa054>.
- Xiong S, Krishnan J, Peuß R, Rohner N. Early adipogenesis contributes to excess fat accumulation in cave populations of *Astyanax mexicanus*. *Dev Biol.* 2018;441(2):297–304. <https://doi.org/10.1016/j.ydbio.2018.06.003>.
- Xiong S, Wang W, Kenzior A, Olsen L, Krishnan J, Persons J, Medley K, Peuß R, Wang Y, Chen S, *et al.* Enhanced lipogenesis through Pparγ helps cavefish adapt to food scarcity. *Curr Biol.* 2022;32(10):2272–2280.e6. <https://doi.org/10.1016/j.cub.2022.03.038>.
- Xu YC, Guo YL. Less is more, natural loss-of-function mutation is a strategy for adaptation. *Plant Commun.* 2020;1(6):100103. <https://doi.org/10.1016/j.xplc.2020.100103>.
- Xu Y-C, Niu X-M, Li X-X, He W, Chen J-F, Zou Y-P, Wu Q, Zhang YE, Busch W, Guo Y-L. Adaptation and phenotypic diversification in *Arabidopsis* through loss-of-function mutations in protein-coding genes. *Plant Cell.* 2019;31(5):1012–1025. <https://doi.org/10.1105/tpc.18.00791>.
- Yamamoto Y, Stock DW, Jeffery WR. Hedgehog signalling controls eye degeneration in blind cavefish. *Nature.* 2004;431(7010):844–847. <https://doi.org/10.1038/nature02864>.
- Yang H, He BZ, Ma H, Tsaur S-C, Ma C, Wu Y, Ting C-T, Zhang YE. Expression profile and gene age jointly shaped the genome-wide distribution of premature termination codons in a *Drosophila melanogaster* population. *Mol Biol Evol.* 2014;32(1):216–228. <https://doi.org/10.1093/molbev/msu299>.
- Zhang L, Xiang L, Liu Y, Venkatraman P, Chong L, Cho J, Bonilla S, Jin Z-B, Pang CP, Ko KM, *et al.* A naturally-derived compound Schisandrin B enhanced light sensation in the *pde6c* Zebrafish model of retinal degeneration. *PLoS One.* 2016;11(3):e0149663. <https://doi.org/10.1371/journal.pone.0149663>.
- Zhang L, Zhang X, Zhang G, Pang CP, Leung YF, Zhang M, Zhong W. Expression profiling of the retina of *pde6c*, a zebrafish model of retinal degeneration. *Sci Data.* 2017;4(1):170182. <https://doi.org/10.1038/sdata.2017.182>.
- Zhang Y, Li S. Ancient lineage, young troglobites: recent colonization of caves by *Nesticella* spiders. *BMC Evol Biol.* 2013;13(1):1–10. <https://doi.org/10.1186/1471-2148-13-183>.
- Zhao Q, Shao F, Li Y, Yi SV, Peng Z. Novel genome sequence of Chinese cavefish (*Triplophysa rosa*) reveals pervasive relaxation of natural selection in cavefish genomes. *Mol Ecol.* 2022;31(22):5831–5845. <https://doi.org/10.1111/mec.16700>.