Title:

Evolution of the gastrointestinal tract morphology and plasticity in cave-adapted Mexican tetra, *Astyanax mexicanus*

Authors:

Misty R. Riddle[1], Fleur Damen[1], Ariel Aspiras[1,*], Julius A. Tabin[1], Suzanne McGaugh[2], Clifford J. Tabin[1,#]

Affiliations:

1. Department of Genetics, Blavatnik Institute, Harvard Medical School, Boston, MA 02115
2. College of Biological Sciences, University of Minnesota, Saint Paul, MN 55108

* Current affiliation: Department of Stem Cell and Regenerative Biology, Harvard University, Cambridge, MA 02138

\# For correspondence: tabin@genetics.med.harvard.edu

Abstract

The gastrointestinal tract has evolved in numerous ways to allow animals to optimally assimilate energy from different food sources. The morphology and physiology of the gut is plastic and can be greatly altered by diet in some animals. In this study, we investigate the evolution and plasticity of gastrointestinal tract morphology by comparing laboratory raised cave-adapted and river-adapted forms of the Mexican tetra, *Astyanax mexicanus* reared under different dietary conditions. In the wild, river-dwelling populations (surface fish) consume plants and insects throughout the year, while cave-dwelling populations (cavefish) live in a perpetually-dark environment dependent on nutrient-poor food brought in by bats or seasonal floods. We find that multiple cave populations converged on a reduced number of digestive appendages called pyloric caeca and that some cave populations have a lengthened gut while others have a shortened gut. Moreover, we identified differences in how gut morphology and proliferation of the epithelium respond to diet between surface fish and cavefish. Using a combination of quantitative genetic mapping, population genetics, and RNA sequencing we implicate molecular and genetic changes influencing cell proliferation, cell signaling, and immune system function in the evolution of gut morphology.

Introduction

The gastrointestinal tract consists of functional regions distinguished by the organization of muscle and mucosal layers and cellular composition of epithelium lining the lumen. The secretory and absorptive cell types of the epithelium are constantly being replaced by self-renewing stem cells.

Gut homeostasis requires integration of intrinsic signaling pathways set up during development, with external cues. For example, in the mammalian gut, Notch and WNT signaling is essential for establishing the stem cell niche and balancing stem cell renewal and differentiation of cell types (reviewed in Gehart & Clevers, 2019). High-fat diet alters these pathways to favor self-renewal and secretory cell formation (2, 3). Cytokine signals from tissue-resident immune cells can also promote stem cell renewal in response to infection (4, 5). While studies utilizing model organisms and organoid culture have begun to reveal how homeostasis of the epithelium is maintained, there is a limited understanding of how these pathways have evolved in response to different diets.

Evolutionary changes to gut morphology such as expansion of functional domains has been correlated to altered patterns of gene expression during development (6). The organization of the gut epithelium into crypts that support villi is important for stem cell maintenance in mammals, but in other species the epithelium is characterized by irregular folds, zig-zags, honeycomb patterns, or a spiral fold (7, 8). The presence of spatially restricted gut stem cells is common across distant phyla (i.e. arthropods and chordates) (9–11) but although it has been demonstrated that diet influences gut morphology in a number of species (reviewed in 12), how the signaling pathways that integrate internal and external cues evolve is not well understood. For animals that survive months without food like snakes and migrating birds (13, 14), evolution of these pathways was likely critical to maximize nutrient assimilation during times of food abundance, and conserve energy when food is scarce.

A. mexicanus is a species of small fish that exists as river- and cave-adapted populations that evolved on very different diets (15, 16). The river (surface) fish have access to insects and plants in abundance, while in the cave, the absence of light makes cavefish dependent on bat droppings or material brought in by seasonal floods (15, 16). There are a number of cave populations, named for the caves they inhabit (i.e. Tinaja,

Pachón, Molino). Based on whole genome sequencing analysis, these populations reflect two independent derivations from surface fish less than 200,000 years ago; Tinaja and Pachón cavefish populations are a part of a clade separate from the clade that includes Molino cavefish (17). Cavefish have converged on similar morphological changes such as eye loss and increased sensory structures (18–20), behavioral changes like reduced sleep (21–24), and metabolic changes such as increased fat accumulation and starvation resistance (25–27). We previously showed that Pachón cavefish have altered gastrointestinal motility during post-larval growth to slow food transit possibly to achieve increased nutrient absorption.

In this study, we explored whether cave-adaptation has led to changes in the adult gut. First we describe the anatomy and histology of the *A. mexicanus* gastrointestinal tract noting distinct functional regions. Next, we compare cavefish populations fed the same diet and discover differences in pyloric caeca number and gut length. We go on to show that cavefish respond differently to changes in diet composition at the level of gut homeostasis and morphology. We use quantitative trait loci (QTL) analysis to investigate the genetic architecture controlling gut length and identify a significant QTL associated with hindgut length. We investigate genes within the QTL using population genetics and RNA sequencing and reveal genetic pathways that have been altered during the evolution of gut morphology in *A. mexicanus*.

Results

Anatomy and histology of the adult Astyanax mexicanus gastrointestinal (GI) tract

The GI tract of teleost fish is commonly divided into four sections from anterior to posterior: head gut, foregut, midgut, and hindgut. We found that these sections are easily defined in *A. mexicanus* and can be further subdivided based on distinct morphology and histology. The foregut in *A. mexicanus* consists of the esophagus, stomach, and pylorus. The esophagus has two perpendicular layers of striated muscle and a stratified epithelium of mucus secreting cells, identifiable by a lack of staining with hematoxylin and eosin (Figure 1A). It connects to the J-shaped stomach that has three muscle layers: outermost

longitudinal, the middle and thickest circumferential, and innermost oblique (Figure 1B). Gastric glands are evident in the stomach mucosa and at the base of the glands pepsin-secreting chief cells are distinguishable by eosinophilic granules. The lumen-facing epithelium of the stomach consists of columnar mucus-secreting cells. The stomach ends at the pylorus that has thick muscle layers and connects the foregut to the midgut.

The midgut is the longest portion of the GI tract. The proximal end of the midgut contains pyloric caeca: pouch structures that have a thin layer of outer connective tissue and lamina propria, and an inner epithelium organized into ridges that run from the base to the tip of the pouch (Figure 2A). The epithelium is made up of mostly columnar enterocytes and a few mucus secreting goblet cells (Figure 1C). There are four pyloric caeca on the right-hand side of the midgut (Figure 2, "1-4"), two on the left-hand side ("5-6"), and as many as three that are shorter and more proximally restricted ("7-9"). Caeca one through five are present in fish of all four populations (surface, Tinaja, Pachón, and Molino) while six through nine are variable. Surface fish are more likely to have 1-9, Tinaja and Molino 1-6, and Pachón 1-7 (Figure 2C).

The remaining tubular portion of the midgut has inner circumferential and outer longitudinal muscle layers and the epithelium is folded circumferentially at irregular angles. The anterior epithelium (Figure 1D) is mostly mucus secreting goblet cells that completely lack H&E staining (Figure 1D', black arrow) compared to the epithelium of the distal midgut (Figure 1E) that has goblet cells with basophilic granules (Figure 1E'). The lumen of the distal midgut also contains basophilic enteroendocrine cells that are distinguishable by a smaller size and more apical location (Figure 1E' yellow arrow). A greater number of esonophilic immune cells are also evident in the lamina propria of the distal region of the midgut (Figure 1E' black arrow head).

The proximal hindgut is wider than the midgut but has thinner muscle and lamina propria layers (Figure 1F-F'). The enterocytes in the epithelium are highly vacuolated likely representing absorbed hydrophobic molecules (Figure 1F', grey arrow). Apically located enteroendocrine cells are also evident in the epithelium (Figure 1F', yellow arrow). The distal hindgut (rectum) has a thinker muscle wall and connects the GI tract to the anus. Overall, the organization of the *A. mexicanus* gut is similar to other Characiformes. Having

established the general anatomy and histology of the GI tract, we were next able to explore differences between the surface and cave populations.

Phylogenetic differences in gut length

The length of each region of the GI tract is an important variable determining digestive efficiency. This variable is also influenced by diet. To determine the phylogenetic differences in gut length between populations we compared relative gut length (gut length divided by fish length) between fish fed the same diet: 38% protein, 7% fat, 5% fiber (Figure 3). To control for differences in appetite, we individually housed the fish to ensure they consumed the same amount per day (6mg). We found that Tinaja (n=6) and Molino (n=5) cavefish tend to have a longer midgut, and Pachón (n=6) cavefish tend to have a shorter midgut compared to surface fish (n=6) (Figure 3A, B). Tinaja and Molino also tend to have a longer hindgut, and Pachón have a significantly shorter hindgut compared to surface fish (Figure 2C, p=0.05, one-way ANOVA with Tukey's post hoc test). Interestingly, while all fish were dissected 24 hours post-feeding, we noted a difference in the amount of fecal matter in the gut (Figure 2A). All of the Tinaja cavefish guts were entirely full (n=6), compared to one out of six Pachón, and four out of five Molino. Surface fish all had a mostly empty gut (n=6). In summary, there are moderate differences in adult gut length between populations that are independent of diet, and some adult cavefish populations may have slower gastrointestinal transit.

Diet influences gut morphology differently in surface fish and Tinaja cavefish

Surface-adapted *A. mexicanus* consume plants and insects throughout the year while cavefish populations are subject to seasonal fluctuations, including regular periods of starvation, and even in times of abundance consume relatively nutrient-poor food. We hypothesized that cavefish may have evolved increased plasticity in gut morphology to maximize nutrient absorption while conserving energy. To test this hypothesis, we switched the diet of adult surface fish and Tinaja cavefish from moderate nutrient content to either low-nutrient (4% fat, 32% protein, 3% fiber) or high-nutrient (18% fat, 55%

protein, 2% fiber). After four months, we measured the length of the gut segments as well as the circumference and average number of folds in cross section, and compared the values to fish that did not change diets (Figure 4, n=3 fish per population and diet, and 3 histological sections per segment). On the moderate nutrient diet fed *ad libitum*, Tinaja tend to have a shorter midgut that is wider and has more folds compared to surface fish (relative midgut length: 0.45 surface, 0.36 Tinaja, circumference: 0.16cm surface, 0.20 Tinaja, fold number: average 6 in surface, 9 in Tinaja, Figure 4 A-C). Tinaja also tend to have a longer hindgut on this diet as was observed on the controlled diet (0.17 surface, 0.20 Tinaja).

In fish that were switched to a low-fat diet, we observed a 24% increase in the length of the midgut in both populations (0.45 to 0.56 surface, 0.36 to 0.44 Tinaja, Figure 4A, G). The circumference of the midgut also increased significantly; by 89% in surface fish (0.16 to 0.30 cm, $p<.005$) and 78% in Tinaja cavefish (0.20 to 0.36 cm, $p<.005$, Figure 4B, G). The number of folds increased 2-fold in Tinaja (9 to 18) and nearly 2-fold in surface fish (6 to 11). The hindgut length increased by 70% in surface fish (0.17 to 0.28) and only 7% in Tinaja cavefish (0.20 to 0.21, Figure 4D, H). The circumference of the hindgut decreased by 3% in surface fish (0.34 to 0.33cm) and increased by 45% in Tinaja (0.24 to 0.35cm, Figure 4D, H). The number of folds in the hindgut changed only slightly in both populations (13 to 12 Tinaja, 11 to 13 Surface). In summary, midgut length, width and fold number increase similarly in both populations in response to a low-fat diet (Figure 4G). In contrast, the hindgut responds by growing in length in surface fish, and growing in width in cavefish.

In fish switched from a moderate to a high-fat diet, the length of the midgut did not change in surface fish (0.45 to 0.46, Figure 4A) and decreased in Tinaja cavefish by 20% (0.36 to 0.29). The circumference of the midgut increased in surface fish (0.16 to 0.23cm) and did not change in cavefish (0.20 to 0.20cm, Figure 4C). The average number of folds increased slightly in surface fish (6 to 7) and decreased slightly in cavefish (9 to 8). Hindgut length decreased in both populations in response to a high-fat diet, although the decrease was greater in Tinaja: 39% (0.20 to 0.12) versus 31% in Surface fish (0.17 to 0.12, Figure 4B). The circumference of the hindgut also decreased; by 26% in surface fish (.34 to .25cm) and by 21% in cavefish (0.24 to 0.19cm, Figure 4D). The number of folds in the hindgut decreased by one is surface fish (11 to 10) but more strikingly in Tinaja cavefish reduced

almost by half (13 to 7). In summary, the Tinaja gut is more responsive to a high-fat diet as it exhibits a greater reduction in length and fold number. Overall, the results suggest that plasticity of the gut is dependent on the type of diet and region of the gut, and reveal fundamental differences between the populations.

Differences in the regulation of gut proliferation between surface fish and Tinaja cavefish

The gastrointestinal epithelium undergoes constant turnover. It is replenished by populations of dividing cells that are restricted to the base of villi or epithelial folds. We next sought to understand the how differences in morphology are achieved by investigating homeostasis of the gut epithelium. We injected fish on the low-nutrient diet with a thymadine analog (5-ethynyl-2'-deoxyuridine, EdU) and after 24 hours determined the number of EdU positive cells as a measure of the amount of cell division (Figure 5A,B,F,J). We found that Tinaja cavefish had on average over four times as many EdU positive cells in the midgut (Figure 5F, average 103 surface, 449 Tinaja, p=0.063 two-tailed t-test). Higher proliferation corresponds to a significantly greater number of folds in the midgut epithelium in Tinaja cavefish (average of 11 in surface, 18 in Tinaja p=0.04, two-tailed t-test). However, the number of EdU positive cells per fold is also greater in Tinaja (average of 11 per fold in surface fish, and 26 per fold in cavefish, p=0.09); EdU positive cells appear more restricted to the base of the folds in surface fish, whereas they extended further up the fold in cavefish (Figure 5A, B). In the hindgut, the number of EdU positive cells is also greater in Tinaja (Figure 5J, average 154 in surface, 354 in Tinaja, p=0.17), although there is only a slight difference in the number of folds (average 13 versus 12). These results suggest that altered homeostasis of the gut epithelium could drive morphological differences between the populations.

Cell-turnover is energetically expensive. Cavefish evolved in a nutrient-limited environment and unlike surface fish are subject to periods of starvation. We hypothesized that cavefish may have evolved mechanisms to limit the proliferation in the gut when food is not available. To test this, we fasted the fish that were fed the low-nutrient diet for a period of two-weeks and compared EdU incorporation over a 24 hour period (Figure 5F,J,K,L). The number of EdU positive cells in the Tinaja midgut decreased significantly by

81% (average 449 fed versus 85 fasted p= 0.005, one-way ANOVA comparing diet and population) while the number increased in surface fish (average 103 fed versus 193 fasted, p =0.63). In both populations, midgut length, circumference, and fold number decreased in response to fasting (Figure 5C-D, G-I). In surface fish, the midgut length shortened significantly by 40% (0.56 to 0.34, p=0.01, one-way ANOVA) compared to only 19% in Tinaja cavefish (0.44 to 0.36, p=0.42). The midgut circumference decreased by 25% in surface fish (0.30 to 0.23cm) and significantly in cavefish by 34% (0.36 to 0.24cm, p=0.004). The number of folds decreased from 11 to 7 in surface fish (p=0.53) and more drastically, from 18 to 6 in Tinaja cavefish (p=0.01). In summary, Tinaja cavefish reduce proliferation in response to fasting and this is associated with a substantial decrease in midgut circumference and fold number, but a more moderate decrease in length compared to surface fish.

In the hindgut, the average number of EdU positive cells also reduced significantly in Tinaja cavefish in response to fasting (Figure 5J, average 354 to 130, p=0.005, one-way ANOVA). The number also reduced in surface fish but to a lesser extent (38%, average 154 to 95, p= 0.63 one-way ANOVA). Relative hindgut length reduced by 50% in surface fish (0.28 to 0.14) and only 20% in Tinaja (0.21 to 0.15). We observed a similar reduction in hindgut circumference, by 35% in surface fish (0.33 to 0.21cm) and 34% in Tinaja cavefish (0.35 to 0.23cm). The number of folds reduced from 13 to 8 in surface fish and 12 to 6 in Tinaja cavefish. These results show that similar to what occurs in the midgut, Tinaja decrease proliferation more substantially in the hindgut in response to starvation but have a less drastic decrease in length compared to surface fish. Combined, our findings support the hypothesis that cavefish have mechanisms for tuning gut homeostasis in response to food intake that differ from surface fish.

Genetic mapping reveals quantitative trait loci associated with hindgut length

We next sought to understand the genetic basis of differences in gut morphology between surface fish and cavefish. We carried out a quantitative trait loci (QTL) analysis using F2 surface/Tinaja hybrids (see methods). To eliminate the effect of diet or appetite, we individually housed the hybrids as adults and ensured they consumed the same amount

of nutrient-moderate food per day (6mg, 38% protein, 7% fat, 5% fiber). After four months, we imaged and weighed the fish, dissected out the gut, counted the number of pyloric caeca and measured the length of the midgut and hindgut. We identified a QTL for fish weight on linkage group 13 with a LOD score of 4.37 accounting for 62% of the variance in this trait (Figure 6A). The same region is associated with fish length, although the LOD scores did not rise above a significance threshold (Figure 6B). We did not identify a significant QTL for pyloric caeca number, but did observe a peak on linkage group 24 (Figure 6C).

There is a strong positive correlation between fish length and gut length as expected (midgut p-value=2.2x10^-16, hindgut p-value=5.2x10^-14, Pearson's product-moment correlation). We therefore performed the QTL analysis using relative gut length. We did not identify a significant QTL for relative midgut length, but observed a peak on linkage group 14 (Figure 6D). Differently, we found a significant QTL for relative hindgut length on linkage group 10 with a LOD score of 4.82 at position r172341 accounting for 11% of the variance in this trait. Interestingly, it is the heterozygous genotype at this position that is associated with longest relative hindgut length (Figure 7C). Similarly, F1 hybrids have a longer hindgut than surface fish or Tinaja cavefish (Figure 7B). The midgut of F1 hybrids is also longer, however when we examined the effect plot for the marker with the highest LOD score for relative midgut length, we found that individuals with cave genotype have the longest gut (Figure 7D,E). Combined, the results suggest a complex genetic architecture controlling gut length and presence of antagonistic alleles that when combined result in positive epistasis.

Candidate genes controlling hindgut length

To search for genetic changes that contribute to variation in hindgut length, we used a combination of population genetics and differential gene expression analysis. We first used the Pachón cavefish genome assembly to identify the position of markers that define the hindgut QTL confidence interval and/or have a significant LOD score. We found ten markers that were spread among four scaffolds (Table 1). We determined several population genetic metrics for all of the genes on each scaffold using whole genome sequencing data (17)(Supplemental Table 1). This data set includes Tinaja, Pachón and

Molino cave populations and two surface populations; Rio choy that is representative of our lab strains and is more closely related to the stock of fish that invaded the Molino cave, and Rascón that is more representative of a separate stock of surface fish that invaded the Tinaja and Pachón caves. We found that 45 genes in the QTL have fixed differences in the coding regions between Rio Choy surface fish (n=9) and Tinaja cavefish (n=10) populations (maximum d_{XY} = 1). Ten of these genes are also strongly favored to be under selection in Tinaja since the haplotype surrounding the genes exhibit evidence of non-neutral evolution (HapFLK p-value < 0.05, Table 2, Figure 8a). Two of these ten genes are also favored to be under selection comparing Tinaja to Rascón (n=8). These two genes are predicted to encode complement factor B-like proteins that are a part of the innate immune system (herein referred to as cfb1 and cfb2). We found that cfb1/2 also show evidence of selection in Pachón, but not Molino (Table 2). The genes that are favored to be under selection in Tinaja have varying levels of divergence in the cave populations (Figure 8ab). Genes that show high levels of divergence in all populations are more likely to be generally important for cave adaptation (i.e. complement factor B-like, Figure 8A, B). Genes that show divergence in only the Tinaja population (i.e. klf5l) may be under selection in the Tinaja cave.

Next, we explored which genes in the QTL may have regulatory mutations by comparing levels of expression. For this analysis, we utilized the surface fish genome assembly that is organized into chromosomes. We found that all of the markers we identified in the QTL are on chromosome 14 and span a region of approximately 15KB (Table 1). This region has 342 genes. We determined if any of these genes are differentially expressed in the hindgut using RNA sequencing data from adult surface, Tinaja, and surface/Tinaja F1 hindguts (n=5 hindguts per population). We found that 31 genes are differentially expressed using a likelihood ratio test to compare all three sample types (Table 3). By pairwise comparison, we found 11 additional genes differentially expressed between Tinaja cavefish and surface fish (Table 3). Included in this list is cfb1; Tinaja cavefish have the highest expression of cfb1 and expression in the F1 hybrid is intermediate (Figure 8c). We reasoned that genes showing either the greatest or lowest expression in F1 hybrids may be strong candidates for controlling differences in length since the heterozygous genotype at the QTL is associated with the longest hindgut. Of the

genes that are in the QTL and differentially expressed, five show the highest expression in the hybrid and eight show the lowest expression in the hybrid. Among the genes that are lowest in the hybrid is Notch1A (Figure 8C, Table 3). Notch signaling controls stem cell renewal and formation of secretory versus absorptive cell fate in the intestine (28).

Expression patterns associated with differences in gut length

We next broadened our analysis of the RNA sequencing data to investigate other pathways that may have been altered during the evolution of the cavefish gut. We found that 3550 genes are differentially expressed between surface, F1 hybrids, and Tinaja using a likelihood ratio test (adjusted p-value <0.05). We searched for patterns across the differentially expressed genes between sample groups (degPatterns function in DEseq 2) and identified clusters of genes that show either greatest (group 3, n=55) or lowest (group4, n=53) expression in the F1 hybrid reasoning that these may be more likely to be associated with differences in length (Supplemental Figure 1, Supplemental Table 2). The genes that have the greatest expression in the hybrid include an additional component of the complement system (C8G), a negative regulator of Notch signaling (nrarpa), retinoic acid signaling components (Cyp26A1, rdh1), and a circadian rhythm gene (timeless) (Figure 8C). The genes that show a pattern of lowest expression in the hybrid include tumor suppressor and apoptosis related genes (casp6, brca2, pak1), a suppressor of cytokine activity (not annotated), regulator of retinoic acid production (bco1), and a circadian rhythm gene that typically correlates with timeless expression (per1b) (Figure 8C). In summary, we found surface fish and cavefish have differences in the genetic architecture controlling hindgut length and that cavefish have altered expression of genes controlling cell proliferation, cell signaling, and immune system function.

Discussion

A major challenge for *Astyanax mexicanus* when they invaded caves was adjusting to the drastically limited availability of nutrients compared to what they experienced in the ancestral river environment. In response, they evolved hyperphagia, starvation resistance,

and increased fat accumulation (26, 27, 29). Here we investigated how the morphology and homeostasis of their gastrointestinal tract adapted as a consequence of nutrient restriction. By comparing fish on the same diet in the laboratory, we found that Tinaja and Molino cave populations converged on a lengthened gut while Pachón have a shortened gut. A plausible explanation for this is differences in cave ecology. The Pachón cave is less impacted by seasonal flooding and has cave-adapted micro-crustaceans that could serve as a food source (15). Thus, our data suggest that specializations for individual caves may have occurred. In all cave populations, we observed a reduction in the number of pyloric caeca, specifically absence of caeca 7-9. The pyloric caeca appear to have roles in digestion, absorption, and immune system function (30)(31). It is possible that caeca 7-9 are important for digestion of a food that is only present in the surface fish diet and these structures became vestigial and eventually disappeared in the cave. Other groups of fish like salmonids can have hundreds of pyloric caeca (32). Our study establishes a genetically accessible model to understand pyloric caeca function, development, and evolution.

We hypothesized that cavefish evolved increased plasticity of the gut as an adaptation to save energy during seasonal fluctuation of food. The most striking difference we observed between populations is that cavefish have much greater proliferation in the gut epithelium compared to surface fish under fed conditions and reduce proliferation in response to food deprivation. Cavefish may have increased sensitivity in the pathways that sense or respond to nutrient shortage and it is plausible that this provides an advantage in the cave. On a high-nutrient diet, cavefish exhibited a more dramatic reduction in gut length compared to surface fish. Reducing length when nutrients are abundant could achieve a more optimal balance between energy extraction and storage; the gut is packaged within the body cavity and its size could limit the amount of visceral fat that can accumulate. How the cellular dynamics of the epithelium translate to changes in gut morphology are not entirely clear. For example, proliferation does not change in surface fish during starvation yet gut length decreases. The balance between cell renewal and death is likely different between the populations and in line with this, we observed that Tinaja have lower expression of tumor-suppressor and pro-apoptosis genes (ie brca2, p21 activated kinase, caspase 6). Overall, while we observe plasticity in gut morphology in both

surface fish and cavefish, our data support the hypothesis that cavefish have a more dramatic response to nutrient fluctuations.

We additionally observed differences in the amount of food material present in guts assessed 24 hours after feeding, suggesting transit time for ingested material differs between the populations. In particular, the cave populations appear to have a longer transit time. This could be the result of selection to optimize absorption of nutrients when available and/or reflect greater time needed to digest decaying material mixed with mud, a major way food is obtained in the caves. At post-larval stages, Pachón cavefish have altered gastrointestinal motility resulting in delayed transit of ingested food (33). Optimal assimilation of food can be achieved through different mechanisms and our findings suggest that both gut activity and morphology have been altered in cavefish.

We found evidence that the genetic architecture controlling gut length diverged between surface fish and Tinaja cavefish. We identified a QTL associated with hindgut length and found that F1 and heterozygous F2 hybrids have a longer hindgut compared to surface fish and cavefish. This phenomenon (heterosis) suggests that there are multiple alleles that have antagonistic effects on hindgut length and exhibit positive epistasis. We determined which genes in the QTL showed evidence of selection by analyzing genetic variation in wild river and cave populations and gained insight into the pathways they influence through gene expression analysis. Our results revealed genes that are known to regulate intestinal homeostasis, and those that have not been explored previously.

Two of the genes in the hindgut QTL that are known to regulate homeostasis of the intestinal epithelium are klf5 (krüppel like factor 5) and Notch1A. KLF5 is a transcription factor expressed in intestinal crypts that can promote or inhibit proliferation and is commonly dysregulated in colon cancer (34–36). Notch signaling balances stem cell renewal and differentiation; activation of Notch signaling promotes stem cell proliferation and suppresses secretory cell formation (28). Klf5 is favored to be under selection in Tinaja, but not Pachón or Molino. It is possible that cave populations have altered gut morphology through different genetic pathways. While the Notch coding region is not favored to be under selection, we found that expression of Notch is significantly different comparing surface fish, Tinaja cavefish, and F1 hybrids suggesting there may be regulatory mutations within the QTL that influence Notch expression. Tinaja cavefish tend to have

higher Notch expression in line with greater proliferation in the epithelium. However, we found that notch-regulated ankyrin repeat protein (nrarpa) that suppresses Notch signaling in other cell contexts (37, 38) is more highly expressed in both Tinaja cavefish and F1 hybrids. Since Notch signaling tends to be context dependent, comparing the components of the Notch signaling pathway with cellular resolution may give insight into how differences in proliferation and morphology arise.

Cave-adapted *A. mexicanus* not only experience a different diet in the cave, but also a different microbial landscape (39). We found that complement factor B (cfb1) is within the hindgut QTL, shows significant divergence in all populations, and is differentially expressed in the hindgut. Complement factor B is part of the innate immune system in humans. It is expressed in multiple tissues including the colonic mucosa, and functions in the pathway that recognizes and eliminates bacterial pathogens by controlling immune cell differentiation and (40, 41). Mounting evidences suggests that immune cells alter self-renewal of stem cells in the epithelium through cytokine signaling (5). Interestingly, we found that cavefish and F1 hybrids have lower expression of a suppressor of cytokine signaling in the hindgut. Recent research suggests that Pachón cavefish have fewer pro-inflammatory immune cells in the visceral adipose tissue (42), but the amount and differences in the immune cells in the gut have not be explored. Our results suggest that evolution of the immune system in *A. mexicanus* likely plays a role in how the gut responds to external cues. In the lab, the external microbial landscape is the same for the populations yet it is possible that genetic differences select for specific microbes and this in turn influences intestinal homeostasis.

Cavefish accumulate more carotenoids (retinoic acid (RA) precursors) in the visceral adipose tissue compared to surface fish and this correlates with decreased expression of genes that convert carotenoids into retinoids in the gut epithelium (43). We found that cavefish and F1 hybrids have lower expression of beta-carotene oxygenase (bco1) and higher expression of retinol dehydrogenase (rhd1) in the hindgut. The outcome may be changes in RA production and signaling. RA signaling is important for stem cell maintenance in a number of contexts (44), but how it regulates intestinal stem cell homeostasis *in vivo* is not well understood. It is clear however that RA signaling regulates innate immune cells in the gut (45). It is possible that altered RA signaling in the cavefish

gut influences the cross-talk between immune cells and stem cells resulting in differences in morphology.

We found that expression of the circadian clock genes per1 and timeless that are typically synchronized, show the opposite expression pattern in the hindguts of cavefish and F1 hybrids. Proliferation and gene expression in the intestine is under circadian control and disruption of clock genes alters renewal of the epithelium (46–48). Having evolved in complete darkness, cavefish have altered circadian rhythm as evidenced by developmentally delayed and reduced amplitude of clock gene expression, lack of circadian cycles of metabolism, and reduced sleep (25, 49–51). It is plausible that alterations to the gene regulatory network controlling circadian rhythm influenced homeostasis of the intestinal epithelium during cavefish evolution.

Overall, our study reveals multiple levels at which the *Astyanax mexicanus* gut has evolved in response to ecological differences in food availability including: morphology, homeostasis, and plasticity in response to dietary fluctuations. We found evidence of divergence in the genetic architecture controlling gut length between surface fish and cavefish through genetic mapping and analysis of genetic variation in wild populations. Furthermore, we gained insight into the range of pathways that have been altered at the level of gene expression. This work provides a foundation for identification and functional analysis of mutations that drive evolution of the gastrointestinal tract.

Methods

Fish husbandry and diet

Fish husbandry was performed according to (52). For fixed diet experiments fish were housed individually in 1.5L tanks and fed three pellets (approximately 6mg) of New Life Spectrum TheraA+ small fish formula once per day for greater than 4 months.

Phenotype quantification

Fish were weighed and dissected 24-hours post feeding. Images were taken using a Cannon Powershot D12 digital camera. ImageJ was used to measure fish length and gut length. Data visualization was performed using R (53)

Quantitative trait loci analysis

F2 hybrid population. We bred a surface fish female with a Tinaja cavefish male to produce a clutch of F1 hybrids. We generated a population of surface/Tinaja F2 hybrids by interbreeding individuals from this clutch. The F2 mapping population (n=221) consisted of three clutches produced from breeding paired F1 surface/Tinaja hybrid siblings.

Genotype by sequencing. We extracted DNA from caudal tail fins using DNeasy Blood and Tissue DNA extraction kit (Qiagen). DNA was shipped to Novogene (Chula Vista, CA) for quality control analysis and sequencing. Samples that contained greater than 1.5 ug DNA, minimal degradation (as determined by gel electrophoresis), and OD_{260}/OD_{280} ratio of 1.8 to 2.0 were used for library construction. Each genomic DNA sample (0.3~0.6 μg) was digested with Mse1, HaeIII, and EcoR1. Digested fragments were ligated with two barcoded adapters: a compatible sticky end with the primary digestion enzyme and the Illumina P5 or P7 universal sequence. All samples were pooled and size-selected after several rounds of PCR amplification to obtain the required fragments needed to generate DNA libraries. Concentration and insert size of each library was determined using Qubit® 2.0 fluorometer and Agilent® 2100 bioanalyzer respectively. Finally, quantitative real-time PCR (qPCR) was used to detect the effective concentration of each library. Qualified DNA libraries had an effective concentration of greater than 2 nM and were pooled by effective concentration and expected data production. Pair-end sequencing was then performed on Illumina®

HiSeq platform, with the read length of 144 bp at each end. Raw Illumina genotype-by-sequencing reads were cleaned and processed through the *process_shortreads* command in the Stacks software package (54). The cleaned reads were aligned to the *Astyanax mexicanus* reference genome (AstMex102, INSDC Assembly GCA_000372685.1, Apr 2013) using the Bowtie2 software (55). The aligned reads of 4 surface fish, 4 Tinaja cavefish and 4 F_1 surface/Tinaja hybrids were manually stacked using the *pstacks* command. We then assigned the morphotypic origin of each allele and confirmed heterozygosity in the F_1 samples using the *cstacks* command. Finally, we used this catalog to determine the genotypes at each locus in the F2 samples with the *sstacks* and *genotypes* command. This genotype database was formatted for use in R/qtl (56)

Linkage map. Using R/qtl, we selected for markers that were homozygous and had opposite genotypes in cavefish versus surface fish (based on three surface and three cave individuals). A linkage map was constructed from these loci using only the F2 population. All markers that were genotyped in less than 180 individuals were omitted, as well as all individuals that had poor marker genotyping (<1500 markers). Markers that did not conform to the expected allele segregation ratio of 1:2:1 were also omitted ($p < 1e^{-10}$). These methods produced a map with 1839 markers, 219 individuals, and 29 linkage groups. Unlinked markers and small linkage groups (<10 markers) were omitted until our map consisted of the optimal 25 linkage groups (*Astyanax mexicanus* has 2n = 50 chromosomes (57)). Each linkage group had 20-135 markers and markers were reordered by default within the linkage groups, producing a total map length of > 5000 cM. Large gaps in the map were eliminated by manually switching marker order to the best possible order within each linkage group (error.prob = 0.005). The final linkage map consisted of 1800 markers, 219 individuals and 25 linkage groups, spanning 2871 cM. Maximum spacing was 32.1 cM and average spacing was 1.6 cM.

QTL scan. Genome wide logarithm of the odds (LOD) scores were calculated using a single-QTL model (scanone) and Haley-Knott regression. Fish length and gut length were analyzed using a normal model and fish weight was analyzed using a non-parametric model. Significance threshold was calculated using a thousand permutations.

Population genomic metrics and analysis

We performed the following measures with GATK-processed data, including a core set of samples analyzed in previous studies (17, 27) which contained: Pachón, N = 10 (9 newly re-sequenced plus the reference reads mapped back to the reference genome); Tinaja N =10; Molino N = 9; Rascón N = 8; and Río Choy N = 9 and required six or more individuals have data for a particular site. Details of sequencing and sample processing are in (17, 27).

We used VCFtools v0.1.13 (58) to calculate π, F_{ST} and d_{XY} and custom python scripts to calculate these metrics on a per gene basis. We identified the allele counts per population with VCFtools and used these for subsequent d_{XY} and fixed differences (D_F) calculations. We used hapFLK v1.3 https://forge-dga.jouy.inra.fr/projects/hapflk (59) for genome-wide estimation of the hapFLK statistic of across all 44 *Astyanax mexicanus* samples and two *Astyanax aeneus* samples. For all of these metrics, we only used sites that contained six or more individuals per population and calculated the metric or included the p-values (in the case of hapFLK) for the entire gene, including introns and UTRs. Candidate genes were included if (a) the comparison between Río Choy and Tinaja resulted in a maximum d_{XY} = 1, suggesting that there was at least one fixed difference between Río Choy and Tinaja and (b) HapFLK resulted in at least one p value less than 0.05, suggesting that the haplotype surrounding the gene exhibited evidence for non-neutral evolution.

RNA sequencing

RNA extraction and cDNA synthesis. Adult *A. mexicanus* were euthanized in 400ppm Tricane and the hindgut was immediately removed and homogenized in 0.3mL Trizol using a motorized pellet pestle and stored at -80°C. Total RNA was extracted using Zymo Research Direct-zol RNA MicroPrep with DNAse treatment according to the manufacturers protocol. LunaScript RT supermix kit with 1μg of RNA was used to synthesize cDNA. All samples were processed on the same day. Diluted cDNA samples (50ng/μL) were used for sequencing.

RNA sequencing and differential gene expression analysis. HiSeq Illumina sequencing was performed by Novogene (Chula Vista, CA). All samples were indexed and run as pools, providing an estimated 20-30 million single-end reads per sample.

Samples were processed using an RNA-seq pipeline implemented in the bcbio-nextgen project (https://bcbio-nextgen.readthedocs.org/en/latest/). Raw reads were examined for quality issues using FastQC (http://www.bioinformatics.babraham.ac.uk/projects/fastqc/) to ensure library generation and sequencing are suitable for further analysis.
Reads were aligned to Ensembl build 2 of the *Astyanax mexicanus* genome, augmented with transcript information from Ensembl release 2.0.97 using STAR (60) with soft trimming enabled to remove adapter sequences, other contaminant sequences such as polyA tails and low quality sequences. Alignments were checked for evenness of coverage, rRNA content, genomic context of alignments and complexity using a combination of FastQC, Qualimap (61), MultiQC (https://github.com/ewels/MultiQC), and custom tools. Counts of reads aligning to known genes were generated by featureCounts (62) and used for further QC with the R package bcbioRNASeq [Steinbaugh MJ, Pantano L, Kirchner RD et al. bcbioRNASeq: R package for bcbio RNA-seq analysis [version 2; peer review: 1 approved, 1 approved with reservations]. F1000Research 2018, 6:1976 (https://doi.org/10.12688/f1000research.12093.2)]. In parallel, Transcripts Per Million (TPM) measurements per isoform were generated by quasialignment using Salmon (63) for downstream differential expression analysis as quantitating at the isoform level has been shown to produce more accurate results at the gene level (64). Salmon output was imported into R using tximport (65) and differential gene expression analysis and data visualization was performed using R with the DEseq 2 package (53, 66).

Acknowledgments

Brian Martineu and Megan Peavey for fish husbandry. This work was supported by grants from the National Institutes of Health [HD089934, DK108495].

References

1. H. Gehart, H. Clevers, Tales from the crypt: new insights into intestinal stem cells.

Nat. Rev. Gastroenterol. Hepatol. (2019) https:/doi.org/10.1038/s41575-018-0081-y.

2. S. Beyaz, *et al.*, High-fat diet enhances stemness and tumorigenicity of intestinal progenitors. *Nature* (2016) https:/doi.org/10.1038/nature17173.
3. C.-W. Cheng, *et al.*, Ketone Body Signaling Mediates Intestinal Stem Cell Homeostasis and Adaptation to Diet. *Cell* (2019) https:/doi.org/10.1016/j.cell.2019.07.048.
4. C. A. Lindemans, *et al.*, Interleukin-22 promotes intestinal-stem-cell-mediated epithelial regeneration. *Nature* (2015) https:/doi.org/10.1038/nature16460.
5. M. Biton, *et al.*, T Helper Cell Cytokines Modulate Intestinal Stem Cell Renewal and Differentiation. *Cell* (2018) https:/doi.org/10.1016/j.cell.2018.10.008.
6. D. M. Smith, R. C. Grasty, N. A. Theodosiou, C. J. Tabin, N. M. Nascone-Yoder, Evolutionary relationships between the amphibian, avian, and mammalian stomachs. *Evol. Dev.* **2**, 348–359 (2000).
7. K. D. Walton, D. Mishkind, M. R. Riddle, C. J. Tabin, D. L. Gumucio, Blueprint for an intestinal villus: Species-specific assembly required. *Wiley Interdiscip. Rev. Dev. Biol.*, e317 (2018).
8. N. A. Theodosiou, E. Oppong, 3D morphological analysis of spiral intestine morphogenesis in the little skate, Leucoraja erinacea. *Dev. Dyn.* (2019) https:/doi.org/10.1002/dvdy.34.
9. S. Takashima, D. Gold, V. Hartenstein, Stem cells and lineages of the intestine: A developmental and evolutionary perspective. *Dev. Genes Evol.* (2013) https:/doi.org/10.1007/s00427-012-0422-8.
10. I. Miguel-Aliaga, H. Jasper, B. Lemaitre, Anatomy and physiology of the digestive tract of drosophila melanogaster. *Genetics* (2018) https:/doi.org/10.1534/genetics.118.300224.
11. N. Aghaallaei, *et al.*, Identification, visualization and clonal analysis of intestinal stem cells in fish. *Dev.* (2016) https:/doi.org/10.1242/dev.134098.
12. S. C. Leigh, B. Q. Nguyen-Phuc, D. P. German, The effects of protein and fiber content on gut structure and function in zebrafish (Danio rerio). *J. Comp. Physiol. B Biochem. Syst. Environ. Physiol.* (2018) https:/doi.org/10.1007/s00360-017-1122-5.
13. S. M. Secor, Digestive physiology of the Burmese python: Broad regulation of integrated performance. *J. Exp. Biol.* (2008) https:/doi.org/10.1242/jeb.023754.

14. S. R. McWilliams, W. H. Karasov, Phenotypic flexibility in digestive system structure and function in migratory birds and its ecological significance in *Comparative Biochemistry and Physiology - A Molecular and Integrative Physiology*, (2001) https:/doi.org/10.1016/s1095-6433(00)00336-6.
15. L. Espinasa, *et al.*, Contrasting feeding habits of post-larval and adult Astyanax cavefish. *Subterr. Biol.* **21**, 1–17 (2017).
16. G. Spicher, [Cytochrome absorption spectra of bacteria as aid for solving taxonomic problems. 1. Elaboration and description of a method for measuring the redox and carbon monoxide difference spectra of suspensions of intact live bacterial cells (author's transl)]. *Zentralbl. Bakteriol. Orig. A.* **226**, 524–40 (1974).
17. A. Herman, *et al.*, The role of gene flow in rapid and repeated evolution of cave-related traits in Mexican tetra, Astyanax mexicanus. *Mol. Ecol.* **27**, 4397–4416 (2018).
18. M. E. Protas, *et al.*, Genetic analysis of cavefish reveals molecular convergence in the evolution of albinism. *Nat. Genet.* **38**, 107–111 (2006).
19. S. E. McGaugh, *et al.*, The cavefish genome reveals candidate genes for eye loss. *Nat. Commun.* **5** (2014).
20. A. K. Powers, E. M. Davis, S. A. Kaplan, J. B. Gross, Cranial asymmetry arises later in the life history of the blind Mexican cavefish, Astyanax mexicanus. *PLoS One* **12** (2017).
21. H. Hinaux, *et al.*, Sensory evolution in blind cavefish is driven by early embryonic events during gastrulation and neurulation. *Development* **143**, 4521–4532 (2016).
22. J. E. Kowalko, *et al.*, Loss of schooling behavior in cavefish through sight-dependent and sight-independent mechanisms. *Curr. Biol.* **23**, 1874–1883 (2013).
23. M. Yoshizawa, *et al.*, Distinct genetic architecture underlies the emergence of sleep loss and prey-seeking behavior in the Mexican cavefish. *BMC Biol.* **13**, 15 (2015).
24. J. Jaggard, *et al.*, The lateral line confers evolutionarily derived sleep loss in the Mexican cavefish. *J. Exp. Biol.* **220**, 284–293 (2017).
25. D. Moran, R. Softley, E. J. Warrant, Eyeless Mexican cavefish save energy by eliminating the circadian rhythm in metabolism. *PLoS One* **9**, e107877 (2014).
26. A. C. Aspiras, N. Rohner, B. Martineau, R. L. Borowsky, C. J. Tabin, Melanocortin 4 receptor mutations contribute to the adaptation of cavefish to nutrient-poor

conditions. *Proc. Natl. Acad. Sci.* **112**, 9668–9673 (2015).

27. M. R. Riddle, *et al.*, Insulin resistance in cavefish as an adaptation to a nutrient-limited environment. *Nature* **555**, 647–651 (2018).
28. E. S. Demitrack, L. C. Samuelson, Notch regulation of gastrointestinal stem cells. *J. Physiol.* (2016) https:/doi.org/10.1113/JP271667.
29. S. Xiong, J. Krishnan, R. Peuß, N. Rohner, Early adipogenesis contributes to excess fat accumulation in cave populations of Astyanax mexicanus. *Dev. Biol.* **441**, 297–304 (2018).
30. R. K. Buddington, J. M. Diamond, Aristotle revisited: The function of pyloric caeca in fish. *Proc. Natl. Acad. Sci. U. S. A.* (1986) https:/doi.org/10.1073/pnas.83.20.8012.
31. N. A. Ballesteros, *et al.*, The Pyloric Caeca Area Is a Major Site for IgM+ and IgT+ B Cell Recruitment in Response to Oral Vaccination in Rainbow Trout. *PLoS One* (2013) https:/doi.org/10.1371/journal.pone.0066118.
32. L. D. Townsend, Variation in the Number of Pyloric Caeca and Other Numerical Characters in Chinook Salmon and in Trout. *Copeia* (1944) https:/doi.org/10.2307/1438249.
33. M. R. Riddle, W. Boesmans, O. Caballero, Y. Kazwiny, C. J. Tabin, Morphogenesis and motility of the Astyanax mexicanus gastrointestinal tract. *Dev. Biol.* **441**, 285–296 (2018).
34. B. B. McConnell, *et al.*, Krppel-like factor 5 is important for maintenance of crypt architecture and barrier function in mouse intestine. *Gastroenterology* (2011) https:/doi.org/10.1053/j.gastro.2011.06.086.
35. T. Nakaya, *et al.*, KLF5 regulates the integrity and oncogenicity of intestinal stem cells. *Cancer Res.* (2014) https:/doi.org/10.1158/0008-5472.CAN-13-2574.
36. X. Zhang, *et al.*, Somatic superenhancer duplications and hotspot mutations lead to oncogenic activation of the KLF5 transcription factor. *Cancer Discov.* (2018) https:/doi.org/10.1158/2159-8290.CD-17-0532.
37. T. Ishitani, K. Matsumoto, A. B. Chitnis, M. Itoh, Nrarp functions to modulate neural-crest-cell differentiation by regulating LEF1 protein stability. *Nat. Cell Biol.* (2005) https:/doi.org/10.1038/ncb1311.
38. T. J. Yun, M. J. Bevan, Notch-Regulated Ankyrin-Repeat Protein Inhibits Notch1

Signaling: Multiple Notch1 Signaling Pathways Involved In T Cell Development. *J. Immunol.* (2003) https:/doi.org/10.4049/jimmunol.170.12.5834.

39. P. Ornelas-García, S. Pajares, V. M. Sosa-Jiménez, S. Rétaux, R. A. Miranda-Gamboa, Microbiome differences between river-dwelling and cave-adapted populations of the fish astyanax mexicanus (De Filippi, 1853). *PeerJ* (2018) https:/doi.org/10.7717/peerj.5906.
40. J. Laskowski, J. M. Thurman, "Chapter 14 - Factor B" in *The Complement FactsBook (Second Edition)*, Factsbook., Second Edition, S. Barnum, T. Schein, Eds. (Academic Press, 2018), pp. 135–146.
41. A. Andoh, *et al.*, Detection of complement C3 and factor B gene expression in normal colorectal mucosa, adenomas and carcinomas. *Clin. Exp. Immunol.* (1998) https:/doi.org/10.1046/j.1365-2249.1998.00496.x.
42. R. Peuß, *et al.*, Single cell analysis reveals modified hematopoietic cell composition affecting inflammatory and immunopathological responses in Astyanax mexicanus *bioRxiv*, 647255 (2019).
43. M. R. Riddle, *et al.*, Genetic architecture underlying changes in carotenoid accumulation during the evolution of the Blind Mexican cavefish, Astyanax mexicanus *bioRxiv*, 788844 (2019).
44. Mezquita, Mezquita, Two Opposing Faces of Retinoic Acid: Induction of Stemness or Induction of Differentiation Depending on Cell-Type. *Biomolecules* (2019) https:/doi.org/10.3390/biom9100567.
45. P. Czarnewski, S. Das, S. M. Parigi, E. J. Villablanca, Retinoic acid and its role in modulating intestinal innate immunity. *Nutrients* (2017) https:/doi.org/10.3390/nu9010068.
46. P. Karpowicz, Y. Zhang, J. B. Hogenesch, P. Emery, N. Perrimon, The Circadian Clock Gates the Intestinal Stem Cell Regenerative State. *Cell Rep.* (2013) https:/doi.org/10.1016/j.celrep.2013.03.016.
47. K. Parasram, P. Karpowicz, Time after time: circadian clock regulation of intestinal stem cells. *Cell. Mol. Life Sci.* (2019) https:/doi.org/10.1007/s00018-019-03323-x.
48. E. Peyric, H. A. Moore, D. Whitmore, Circadian Clock Regulation of the Cell Cycle in the Zebrafish Intestine. *PLoS One* (2013)

https:/doi.org/10.1371/journal.pone.0073209.

49. A. Beale, *et al.*, Circadian rhythms in Mexican blind cavefish Astyanax mexicanus in the lab and in the field. *Nat. Commun.* (2013) https:/doi.org/10.1038/ncomms3769.

50. I. A. Frøland Steindal, A. D. Beale, Y. Yamamoto, D. Whitmore, Development of the Astyanax mexicanus circadian clock and non-visual light responses. *Dev. Biol.* (2018) https:/doi.org/10.1016/j.ydbio.2018.06.008 (August 8, 2018).

51. E. R. Duboué, A. C. Keene, R. L. Borowsky, Evolutionary convergence on sleep loss in cavefish populations. *Curr. Biol.* **21**, 671–676 (2011).

52. Y. Elipot, L. Legendre, S. Père, F. Sohm, S. Rétaux, *Astyanax* Transgenesis and Husbandry: How Cavefish Enters the Laboratory. *Zebrafish* **11**, 291–299 (2014).

53. R-core-team, R: A Language and Environment for Statistical Computing (2019).

54. J. M. Catchen, A. Amores, P. Hohenlohe, W. Cresko, J. H. Postlethwait, Stacks: building and genotyping Loci de novo from short-read sequences. *G3 (Bethesda).* **1**, 171–82 (2011).

55. B. Langmead, S. L. Salzberg, Fast gapped-read alignment with Bowtie 2. *Nat. Methods* **9**, 357–9 (2012).

56. D. Arends, P. Prins, R. C. Jansen, K. W. Broman, R/qtl: High-throughput multiple QTL mapping. *Bioinformatics* (2010) https:/doi.org/10.1093/bioinformatics/btq565.

57. K. F. Kavalco, L. F. De Almeida-Toledo, Molecular cytogenetics of blind Mexican tetra and comments on the karyotypic characteristics of genus Astyanax (Teleostei, Characidae). *Zebrafish* (2007) https:/doi.org/10.1089/zeb.2007.0504.

58. P. Danecek, *et al.*, The variant call format and VCFtools. *Bioinformatics* (2011) https:/doi.org/10.1093/bioinformatics/btr330.

59. M. I. Fariello, S. Boitard, H. Naya, M. SanCristobal, B. Servin, Detecting signatures of selection through haplotype differentiation among hierarchically structured populations. *Genetics* (2013) https:/doi.org/10.1534/genetics.112.147231.

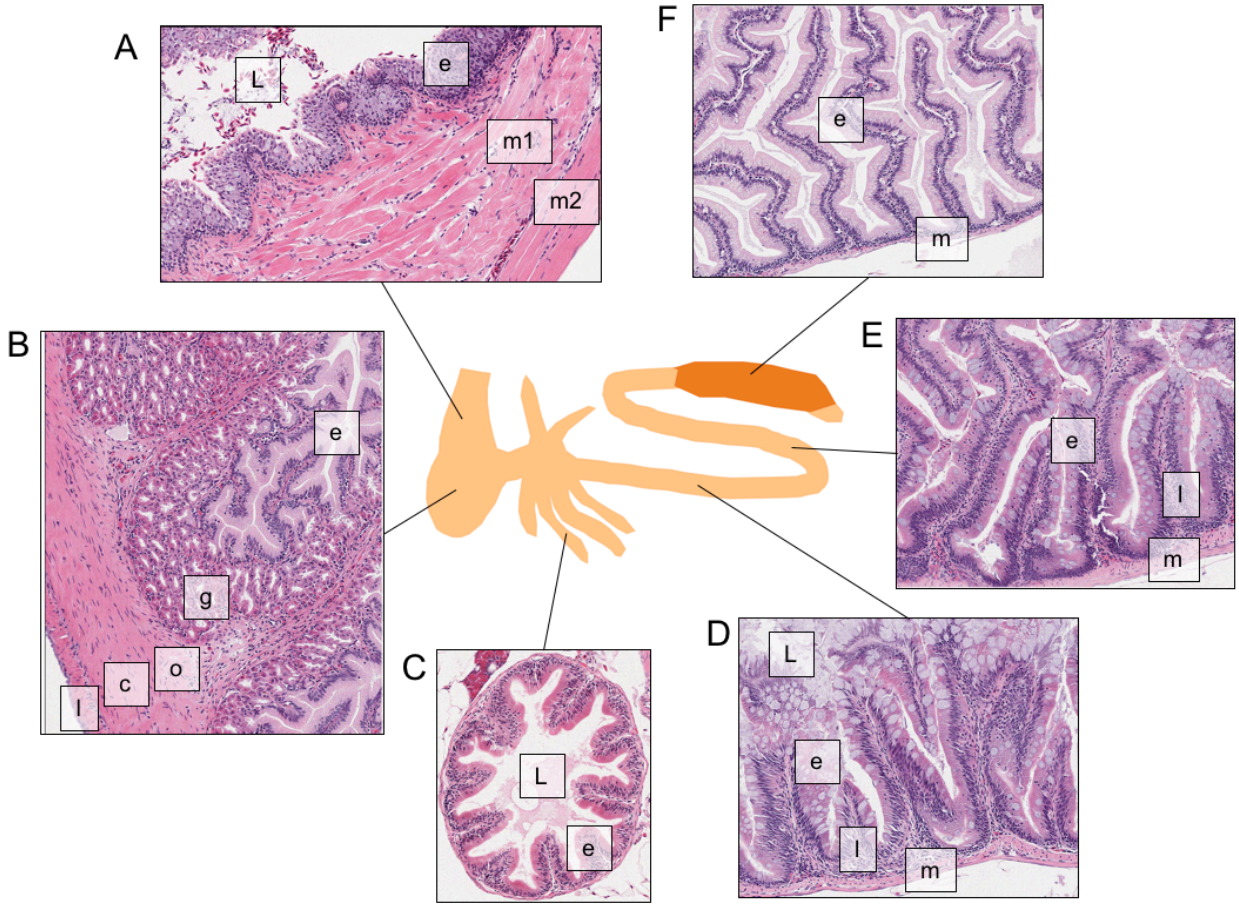
60. A. Dobin, *et al.*, STAR: Ultrafast universal RNA-seq aligner. *Bioinformatics* (2013) https:/doi.org/10.1093/bioinformatics/bts635.

61. F. García-Alcalde, *et al.*, Qualimap: Evaluating next-generation sequencing alignment data. *Bioinformatics* (2012) https:/doi.org/10.1093/bioinformatics/bts503.

62. Y. Liao, G. K. Smyth, W. Shi, FeatureCounts: An efficient general purpose program for

assigning sequence reads to genomic features. *Bioinformatics* (2014) https:/doi.org/10.1093/bioinformatics/btt656.

63. R. Patro, G. Duggal, C. Kingsford, Accurate, fast, and model-aware transcript expression quantification with Salmon. *bioRxiv* (2015) https:/doi.org/10.1101/021592.
64. C. Soneson, M. I. Love, M. D. Robinson, Differential analyses for RNA-seq: Transcript-level estimates improve gene-level inferences [version 2; referees: 2 approved]. *F1000Research* (2016) https:/doi.org/10.12688/F1000RESEARCH.7563.2.
65. C. Soneson, M. I. Love, M. D. Robinson, Differential analyses for RNA-seq: transcript-level estimates improve gene-level inferences. *F1000Research* (2015) https:/doi.org/10.12688/f1000research.7563.1.
66. M. I. Love, W. Huber, S. Anders, Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* (2014) https:/doi.org/10.1186/s13059-014-0550-8.


A
L
e
m1
m2
B
e
g
o
c
l
C
L
e
D
L
e
l
m
E
e
l
m
F
e
m

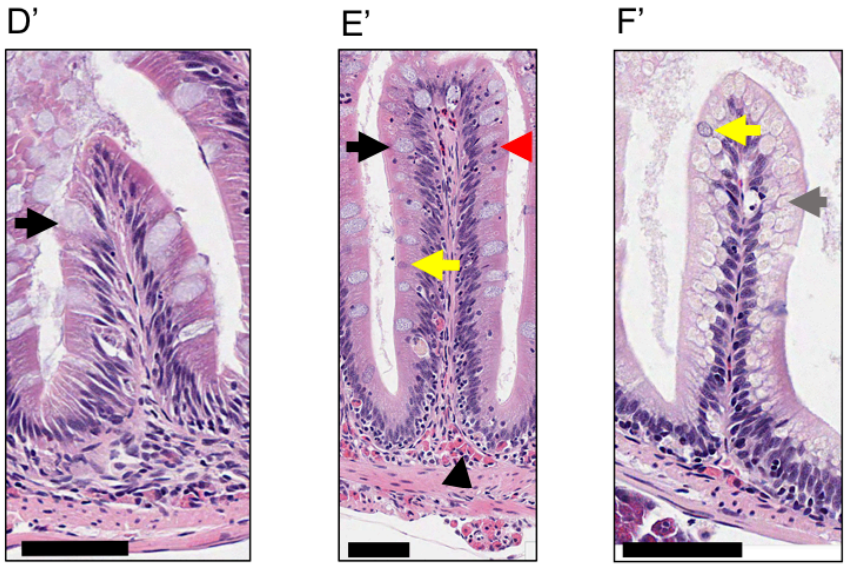

D'
E'
F'

Figure 1. Histology of the *Astyanax mexicanus* gastrointestinal tract as shown by hematoxylin and eosin staining. A, Transverse section of esophageal region (L: lumen, e: epithelium, m1-2: muscle layers). B, Stomach section showing epithelium (e), eosenophilic gastric glands (g), and layers of muscle (o: orthogonal, c: circumferential, l: longitudinal). C, Cross section of a pyloric caecum showing the lumen (L) and ridged epithelium (e). D, Transverse section of the proximal midgut showing lumen filled with mucus (L), epithelium (e), lamina propria (l), and circumferential inner and longitudinal outer muscle layers (m). E, Transverse section of a more posterior section of the midgut showing epithelium (e), lamina propria (l), and circumferential inner and longitudinal outer muscle layers (m). F, Transverse section of the hindgut showing epithelium (e), and muscle layers (m). A-F image magnification is 10X. D-F' 20X images (scale bar is 50µM) showing the cell types of the gut epithelium in the proximal midgut (D'), distal midgut (E'), and hindgut (F'). Black arrows highlight the differences in secretory goblet cell types. Black arrow head in E' indicates group of immune cells (pink). Red arrow head shows pyknotic nuclei. Grey arrow in F' highlights cells with hydrophobic (mostly clear) structures within the cytoplasm. D-E are images from the same slide.

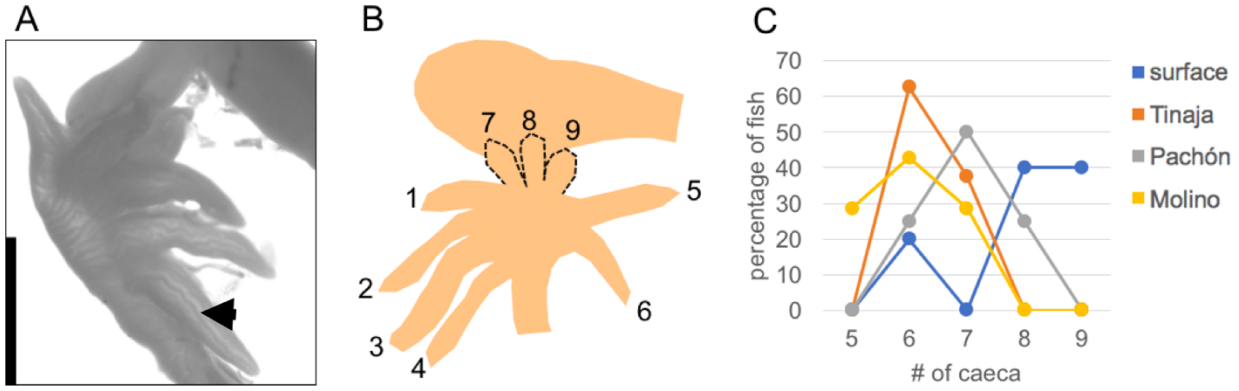


Figure 2. Pyloric caeca structure and number in of *Astyanax mexicanus* surface and cave morphs. A, Image of the region of gut containing pyloric caeca. The darker lines show the ridged morphology of the epithelium (black arrow, scale bar is 2 mm). B, Drawing of the pyloric caeca indicating the position and number of caeca. The caeca at position 7, 8, and 9 are variable. C, Graph comparing percentage of fish with the indicated number of total caeca for each population (n=5, 6, 8, 6).

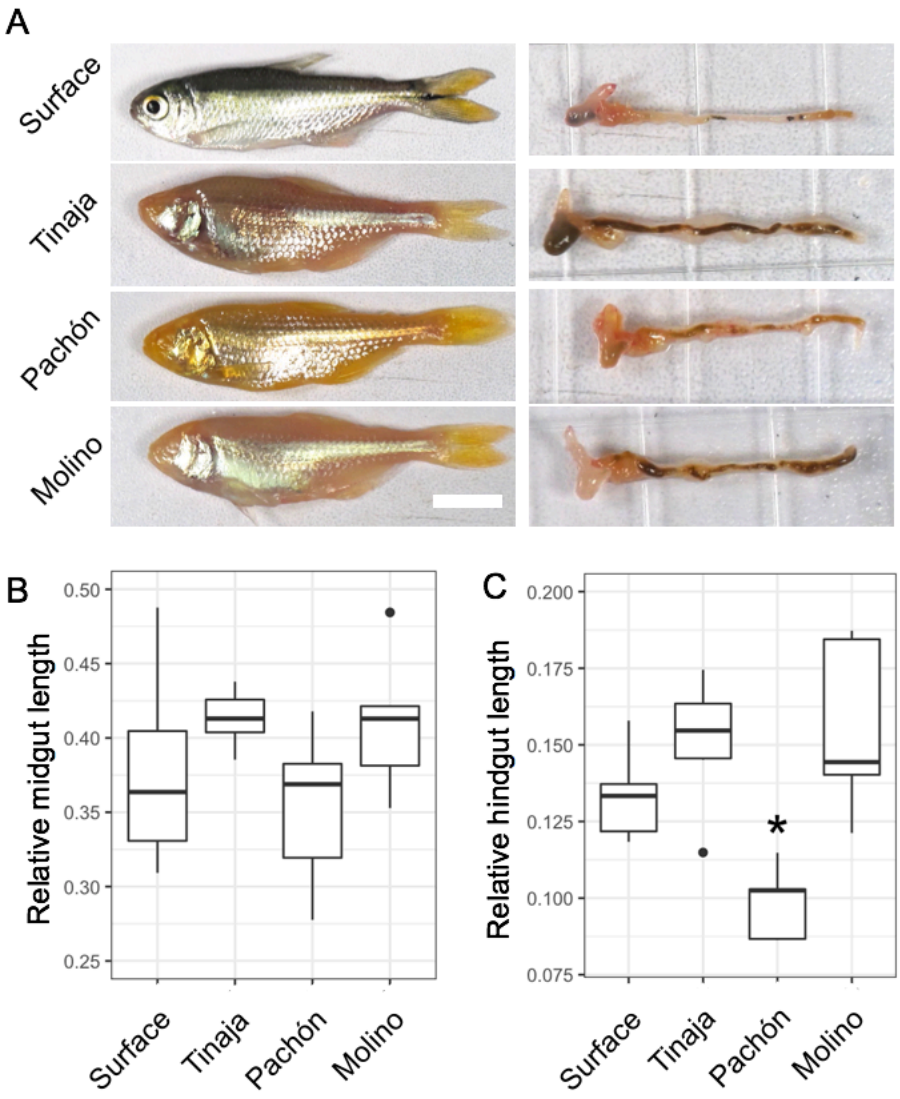


Figure 3. Relative length of the gastrointestinal (GI) tract of *Astyanax mexicanus* surface and cave morphs. A, Images of fish from the indicated populations that were fed 6 mg of pellet food per day for greater than 8 months and their GI tracts (scale bars is 1cm). B, Boxplots showing relative midgut and hindgut length of fish from the indicated populations (n=6 fish per population). For box plots, median, 25th, 50th, and 75th percentiles are represented by horizontal bars and vertical bars represent 1.5 interquartile ranges. Significance codes from one-way ANOVA with HSD post hoc test, *$p<0.05$.

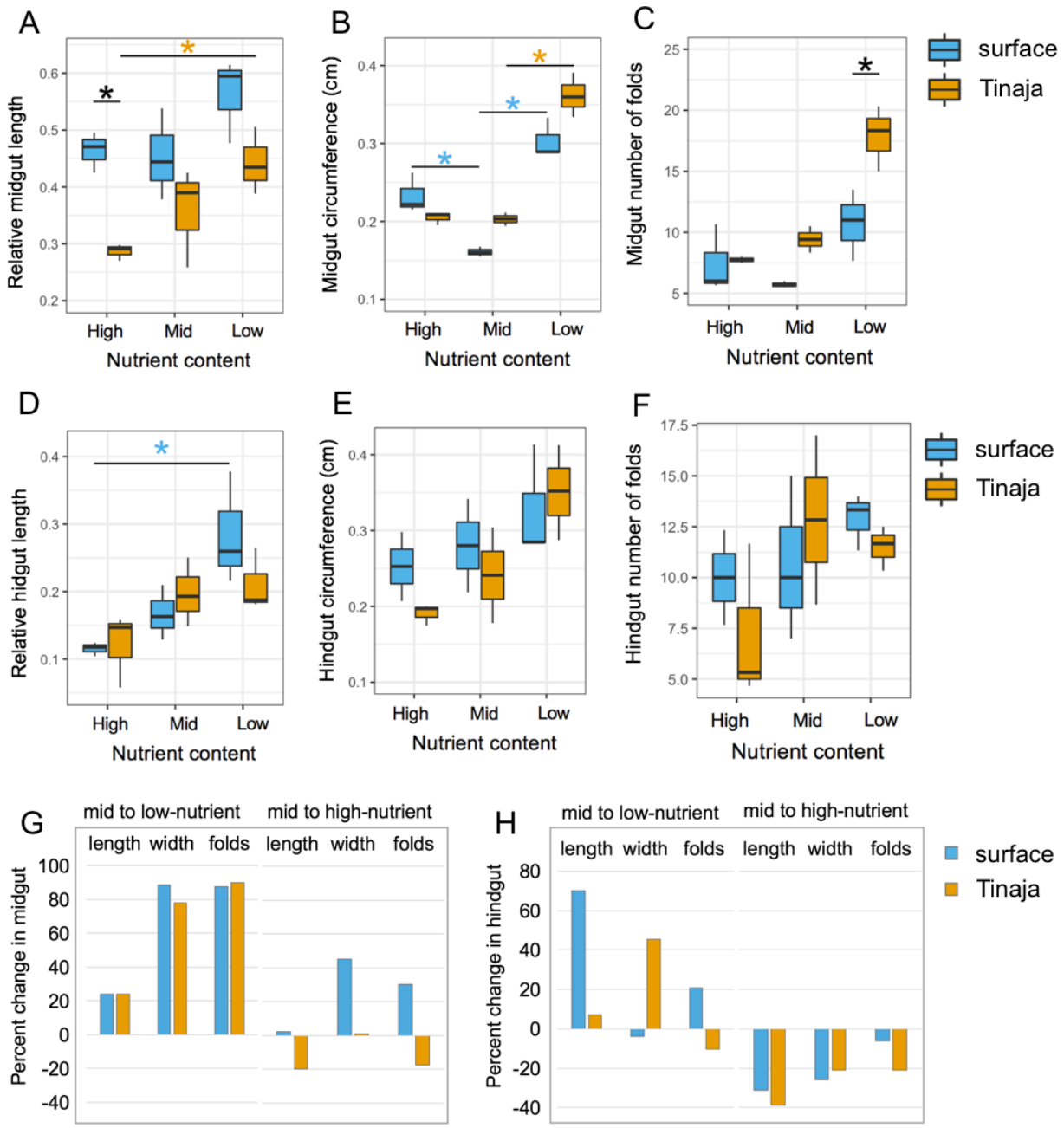
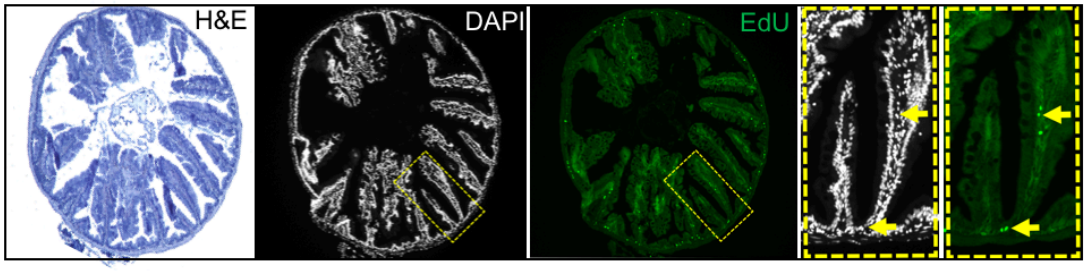
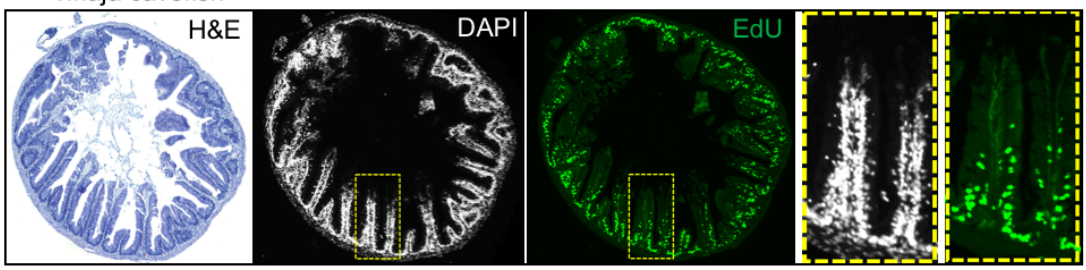


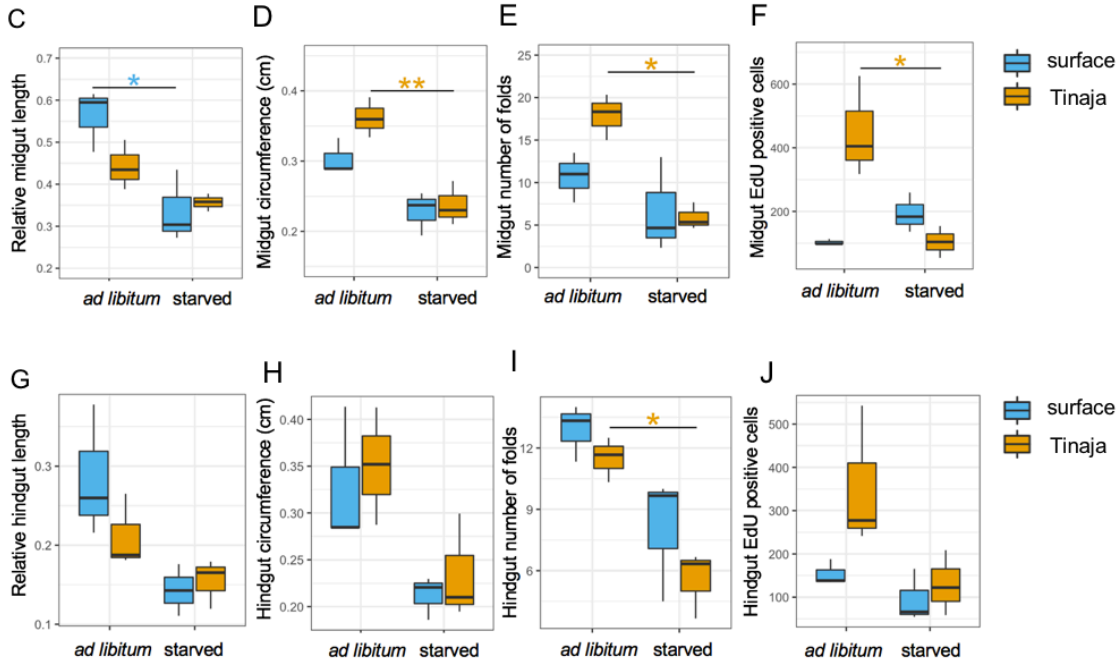
Figure 4. Diet influences *A. mexicanus* gut morphology. Relative length, circumference, and number of folds in the epithelium in the midgut (A-C) and hindgut (D-E) in fish continuously fed a moderate diet or switched to a high-nutrient or low nutrient diet as adults (n=3 fish per population and diet, and 3 tissue sections per sample). In boxplots, median, 25th, 50th, and 75th percentiles are represented by horizontal bars and vertical bars represent 1.5 interquartile ranges. Significance code (*p<0.05) from one-way ANOVA with HSD post hoc test comparing between populations (black asterisks) or within population by diet (orange asterisks Tinaja, blue asterisks surface fish). Bar graph shows percent change in the indicated phenotype in the midgut (G) and hindgut (H).

A surface fish



B Tinaja cavefish





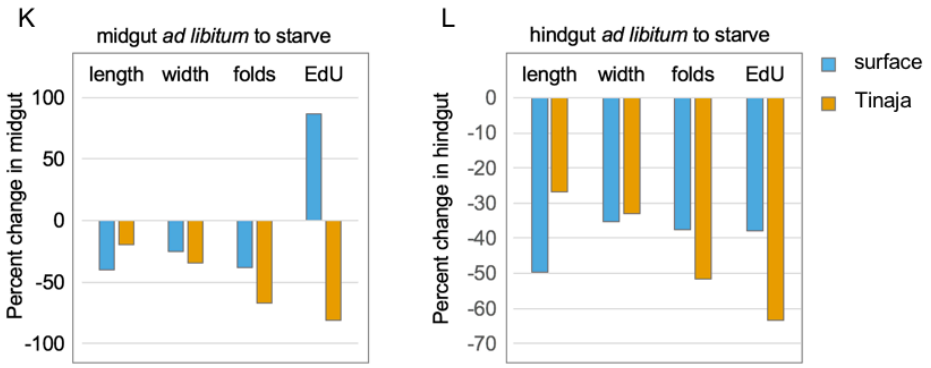


Figure 5. Cavefish have more proliferation in the intestinal epithelium compared to surface fish and reduce proliferation in response to starvation. Cross section of midgut in surface fish (A) and Tinaja cavefish (B) stained with H&E (left) and serial section showing DAPI staining and EdU positive cells; larger view of yellow boxed region shown in right panels. Yellow arrows point to EdU positive cells. C-F, Quantification of midgut length, circumference, number of folds, and EdU positive cells in fish fed *ad libitum* versus fasted for two weeks. G-J, Quantification of hindgut length, circumference, number of folds, and EdU positive cells in fish fed *ad libitum* versus fasted for two weeks (n=3 fish per population and diet and 3 tissue sections averaged per individual fish). In boxplots, median, 25th, 50th, and 75th percentiles are represented by horizontal bars and vertical bars represent 1.5 interquartile ranges. Significance code (*$p<0.05$, **$p<.005$) from one-way ANOVA with HSD post hoc test. K-L, Bar graphs showing percent change in the indicated phenotypes in the midgut (K) and hindgut (L).

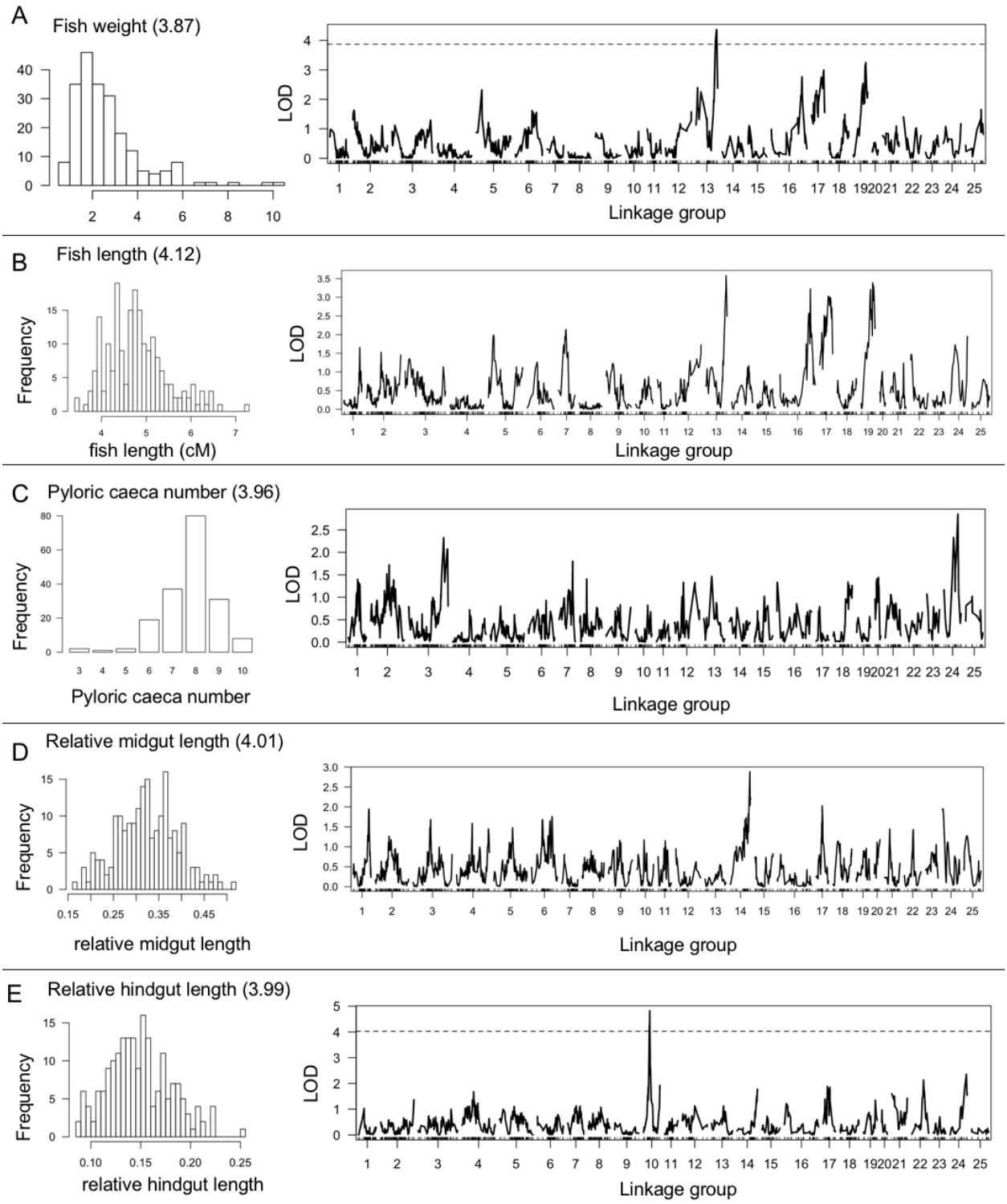


Figure 6. Genetic mapping of gut morphology in *A. mexicanus* surface/Tinaja F2 hybrids. A-E, Histograms showing distribution of the indicated phenotypes and (right) genome wide scan showing logarithm of the odds (LOD) score for the indicated phenotypes using Haley-Knott regression mapping. Significance thresholds at p=.05 are indicated in parenthesis following the phenotype.

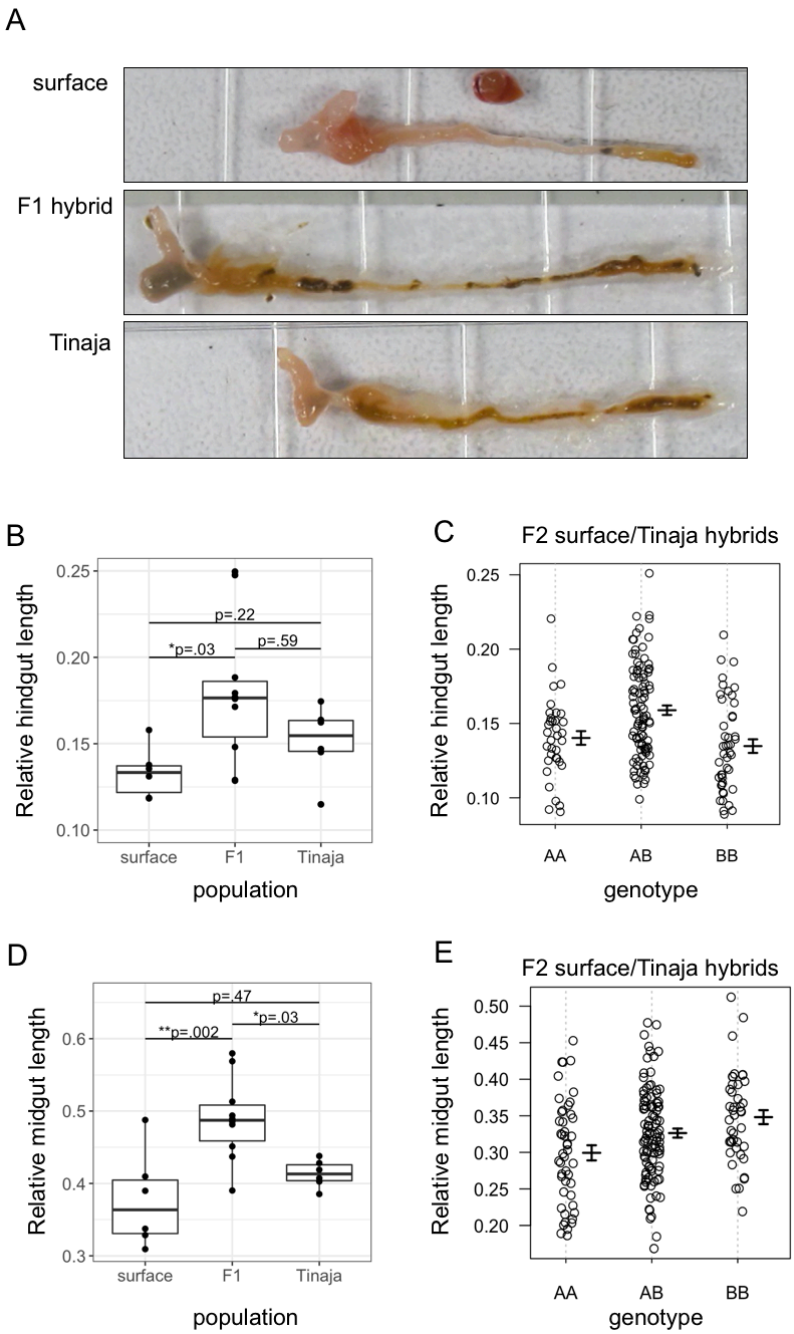


Figure 7. Heterosis of gut length in surface/Tinaja hybrids. A, Image of gut from surface fish, surface/Tinaja F1 hybrid, and Tinaja cavefish of similar lengths. B, Quantification of relative hindgut length in the indicated populations. C, Quantification of relative hindgut length of surface/Tinaja F2 hybrids with the indicated genotype at the marker with the highest LOD score (AA: homozygous surface, AB: heterozygous, BB: homozygous cave). D, Quantification of relative midgut length in the indicated populations. C, Quantification of relative midgut length of surface/Tinaja F2 hybrids with the indicated genotype at the marker with the highest LOD score (AA: homozygous surface, AB: heterozygous BB: homozygous cave).

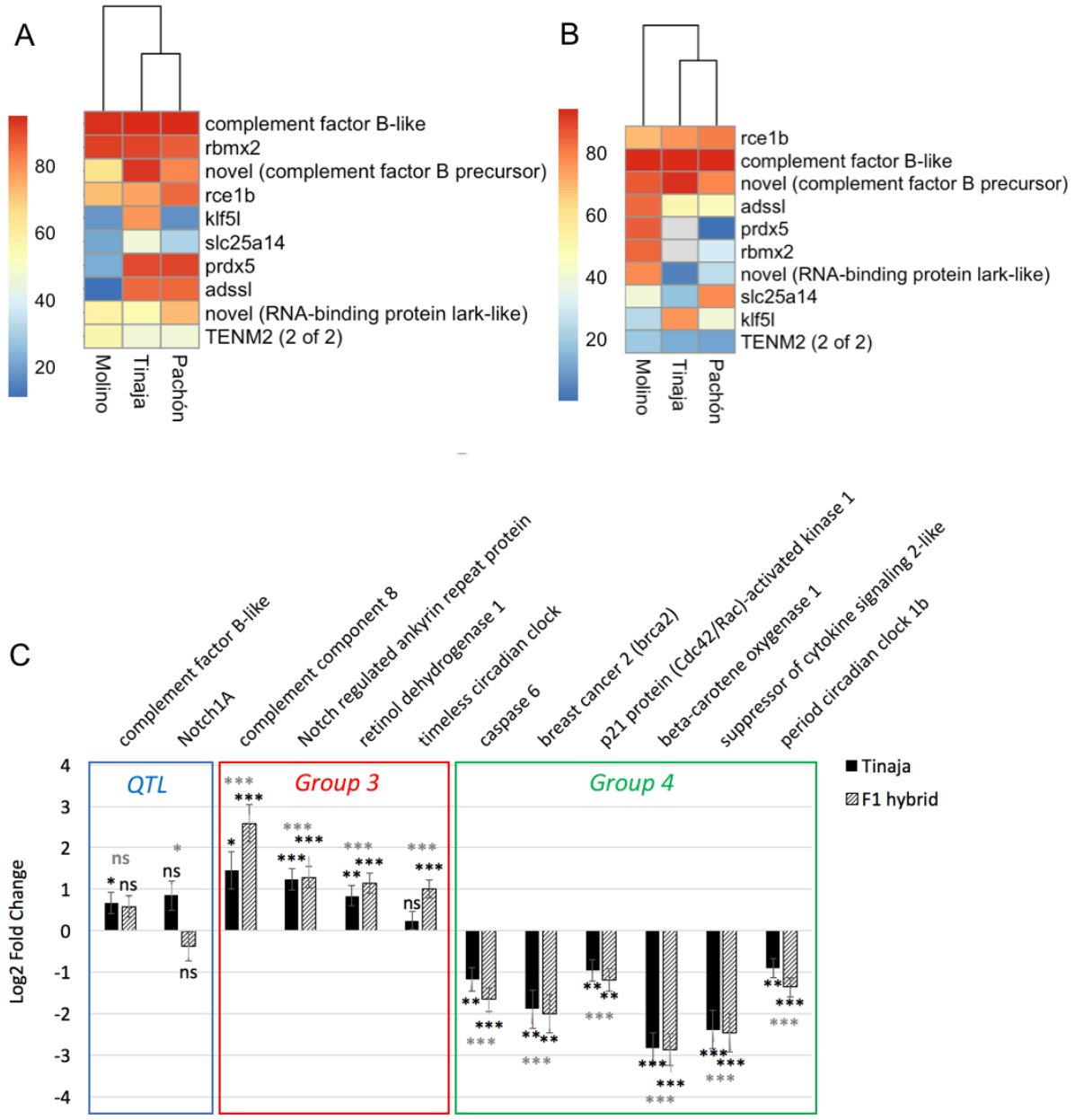
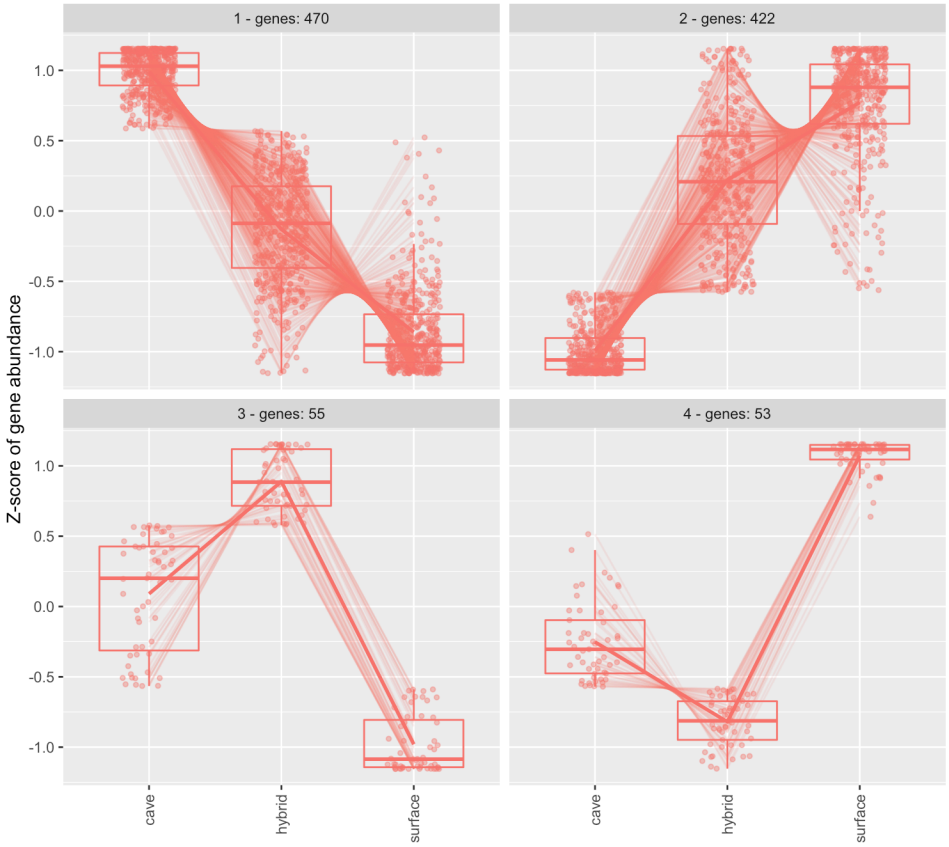


Figure 8. Population genetics and RNA sequencing analysis of candidate genes controlling hindgut length. A, B, Heat maps showing the divergence percentile (Dxy/Max Dxy) of genes within the hindgut QTL that are favored to be under selection in Tinaja (len.p.values.0.05 >0). A, Comparing divergence from Rio Choy surface fish to cavefish, and B, Rascón surface fish to cavefish. C, Log2 Fold expression change compared to surface fish of a subset of genes within the hindgut QTL (blue), or that are a part of the gene cluster that shows lowest (group3, red) and highest (group4, green) expression in F1 hybrids. Grey significance codes from likelihood ratio test comparing all three sample types and black asterisks next to bars indicate significance from Wald test comparing to only surface fish (Significance code for adjusted p-value: *<.05 **<.005 ***<.0005, ns>.05). Error bars indicate standard error estimate for log2 fold change.



Supplemental Figure 1: Gene expression clusters of genes that are differentially expressed between Tinaja, Surface, and F1 hybrid hindguts (n=5 hindguts per population).