

RESEARCH ARTICLE

Open Access



# Genetic mapping of metabolic traits in the blind Mexican cavefish reveals sex-dependent quantitative trait loci associated with cave adaptation

Misty R. Riddle<sup>1\*</sup> , Ariel Aspiras<sup>3</sup>, Fleur Damen<sup>2</sup>, Suzanne McGaugh<sup>4</sup>, Julius A. Tabin<sup>2</sup> and Clifford J. Tabin<sup>2</sup>

## Abstract

**Background:** Despite a longstanding interest in understanding how animals adapt to environments with limited nutrients, we have incomplete knowledge of the genetic basis of metabolic evolution. The Mexican tetra, *Astyanax mexicanus*, is a species of fish that consists of two morphotypes; eyeless cavefish that have adapted to a low-nutrient cave environment, and ancestral river-dwelling surface fish with abundant access to nutrients. Cavefish have evolved altered blood sugar regulation, starvation tolerance, increased fat accumulation, and superior body condition. To investigate the genetic basis of cavefish metabolic evolution we carried out a quantitative trait loci (QTL) analysis in surface/cave F2 hybrids. We genetically mapped seven metabolism-associated traits in hybrids that were challenged with a nutrient restricted diet.

**Results:** We found that female F2 hybrids are bigger than males and have a longer hindgut, bigger liver, and heavier gonad, even after correcting for fish size. Although there is no difference between male and female blood sugar level, we found that high blood sugar is associated with weight gain in females and lower body weight and fat level in males. We identified a significant QTL associated with 24-h-fasting blood glucose level with the same effect in males and females. Differently, we identified sex-independent and sex-dependent QTL associated with fish length, body condition, liver size, hindgut length, and gonad weight. We found that some of the genes within the metabolism QTL display evidence of non-neutral evolution and are likely to be under selection. Furthermore, we report predicted nonsynonymous changes to the cavefish coding sequence of these genes.

**Conclusions:** Our study reveals previously unappreciated genomic regions associated with blood glucose regulation, body condition, gonad size, and internal organ morphology. In addition, we find an interaction between sex and metabolism-related traits in *A. mexicanus*. We reveal coding changes in genes that are likely under selection in the low-nutrient cave environment, leading to a better understanding of the genetic basis of metabolic evolution.

**Keywords:** *Astyanax mexicanus*, Cavefish, Quantitative trait loci, Blood glucose, Sex, Metabolism, Liver, Gut, Body condition, Teleost, Evolution

## Background

The Mexican tetra (*Astyanax mexicanus*) is a species that consists of river-dwelling surface fish and eyeless cavefish that evolved in starkly different environments (Fig. 1A). Surface fish in rivers from Southern Texas to Central Mexico consume a diet of insects and plants,

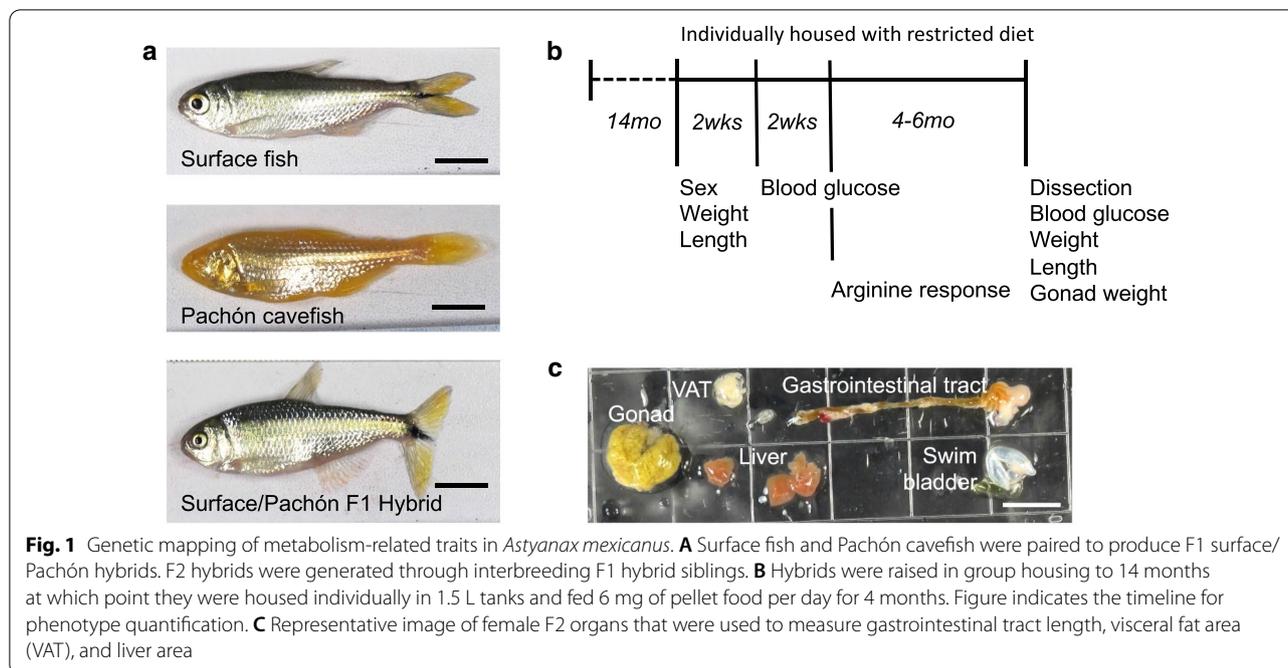
\*Correspondence: mistyriddle@unr.edu

<sup>1</sup> Department of Biology, University of Nevada, Reno, Reno, NV 89557, USA

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



© The Author(s) 2021. **Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>. The Creative Commons Public Domain Dedication waiver (<http://creativecommons.org/publicdomain/zero/1.0/>) applies to the data made available in this article, unless otherwise stated in a credit line to the data.



while cavefish inhabiting perpetually-dark limestone caves in the Sierras of Northeastern Mexico depend predominantly on bat droppings and material brought in by seasonal floods [1]. Some independently evolved cavefish populations (e.g. Tinaja, Molino, Pachón) have converged on similar metabolic adaptations that may be advantageous in caves such as starvation resistance [2, 3], binge-eating (hyperphagia) [3], increased fat accumulation [4], and insulin resistance [5].

The genetic basis of cavefish metabolism has been examined using candidate gene approaches and genomic data to find evidence of cave-specific mutations. For example, hyperphagia and insulin resistance have been associated with genes that are known to regulate appetite and blood glucose in other animals: *melanocortin receptor 4* (*mc4r*) and *insulin receptor alpha* (*insra*) [3, 5]. Comparing gene expression in metabolic tissues such as the liver across wild and lab-raised surface fish and cavefish revealed that environment and phylogeny impact gene activity [6, 7]. The surface fish and cavefish liver transcriptome differs significantly in expression of genes related to carbohydrate metabolism, gluconeogenesis, and glycogen synthesis [6]. Identifying the genetic changes that underlie these shifts in metabolism-related genes is challenging with transcriptome data alone as gene expression differences could be the result of *cis*- or *trans*-regulatory evolution.

A major strength of *A. mexicanus* as a model system is the ability to produce surface/cave F1 hybrids (Fig. 1A) that can be crossed to perform quantitative trait loci

(QTL) studies to associate genomic regions with trait evolution. Such studies in *A. mexicanus* have focused on morphological and behavioral traits and laid the foundation for functional analyses. For example, a QTL associated with albinism revealed that separate deletions in *oculus albinism 2* (*oca2*) underlie pigment loss in independent cavefish populations [8]. A QTL associated with eye size led to the discovery that *cis*-regulatory changes that impact expression of *cystathionine beta-synthase* (*cbsa*) contribute to cavefish eye degeneration [9].

Here we present the results of a genetic mapping study focused on metabolic evolution in *A. mexicanus*. We analyzed blood glucose regulation, changes in weight, and internal organ morphology in F2 surface/Pachón hybrids shifted from eating ad libitum to a fixed amount. We found that female F2s are larger than males and have a bigger liver, longer hindgut, and heavier gonad even after normalizing the values to fish size. Regardless of sex, we observed that small F2 fish tend to gain weight on a nutrient restricted diet, and that weight gain is associated with a bigger gonad and longer hindgut. We used a linkage map with high marker density to identify a significant QTL associated with blood glucose level. We also identified sex-independent and sex-dependent QTL associated with body condition, liver size, relative hindgut length, and relative gonad weight. We found genes within the QTL that show exceptional divergence in Pachón cavefish, suggesting they may be important for cave adaptation. In addition, we report predicted nonsynonymous changes in these genes using sequencing data from wild

caught surface fish and Pachón cavefish. Overall, our study reveals genomic regions associated with evolution of metabolism-related traits in cavefish. Combined with functional studies, these findings will further improve our understanding of how cavefish thrive in the unique cave environment.

## Results

### Sex bias in weight and organ size but not blood glucose regulation

We measured metabolism-related traits in surface/Pachón F2 hybrids that were individually housed and consumed 6 mg of pellet food per day for a period of 4 months (Fig. 1B). We found that most traits are significantly different between sexes (Table 1). At the start of the feeding trial, on average, females were significantly bigger than males (average length = 5.18 cm versus 4.38 cm,  $p = 2.10 \times 10^{-13}$  False discovery rate (FDR)  $p$ -value  $t$ -test, average start weight = 3.53 g versus 1.58 g,  $p = 1.08 \times 10^{-19}$  FDR adjusted Wilcox). Females also had higher body condition factor (average 2.35 versus 1.83,  $p = 4.03 \times 10^{-18}$  FDR Wilcox), which is a standard

measurement of fish nourishment level [10]. During the feeding trial, males gained weight and females lost weight (average weight change = 0.14 g male,  $-0.23$  g female,  $p = 5.60 \times 10^{-5}$  FDR Wilcox). This equated to an average 14.70% weight increase in males and 0.13% increase in females ( $p = 4.99 \times 10^{-6}$  FDR Wilcox). However, on average, females remained significantly larger than males and had better body condition at the end of the trial (average weight 3.33 g versus 1.73 g,  $p = 1.07 \times 10^{-18}$  FDR Wilcox, average length 5.10 cm vs 4.43 cm,  $p = 4.09 \times 10^{-12}$  FDR  $t$ -test, body condition 2.38 versus 1.92,  $p = 2.21 \times 10^{-15}$  FDR Wilcox). Females have a significantly larger gonad compared to males (average gonad weight = 0.42 g versus 0.06 g,  $p = 1.29 \times 10^{-24}$  FDR Wilcox). This alone does not account for the weight difference between sexes, since the body weight after subtracting the gonad is also significantly greater in females (average 2.91 g versus 1.65 g,  $p = 4.72 \times 10^{-13}$  FDR Wilcox). In addition, gonad weight normalized to body weight is significantly greater in females (average 0.13 versus 0.03,  $p = 1.14 \times 10^{-20}$  FDR Wilcox). This is also true for non-hybrid *A. mexicanus*. For example, in 1.5-year-old female fish average

**Table 1** Average values for surface/Pachón F2 traits

Trait	Male	N	female	N	t-test_pval	wilcox_pval
Weight start (g)	<b>1.58</b>	<b>100</b>	<b>3.53</b>	<b>119</b>		<b>1.08E-19</b>
Weight 4 month (g)	<b>1.73</b>	<b>95</b>	<b>3.33</b>	<b>115</b>		<b>1.07E-18</b>
Length start (cm)	<b>4.38</b>	<b>99</b>	<b>5.18</b>	<b>117</b>	<b>2.10E-13</b>	
Length end (cm)	<b>4.43</b>	<b>91</b>	<b>5.10</b>	<b>108</b>	<b>4.09E-12</b>	
Body condition start	<b>1.83</b>	<b>97</b>	<b>2.35</b>	<b>116</b>		<b>4.03E-18</b>
Body condition end	<b>1.92</b>	<b>89</b>	<b>2.35</b>	<b>106</b>		<b>2.21E-15</b>
Weight change (g)	<b>0.14</b>	<b>96</b>	<b>-0.23</b>	<b>117</b>		<b>5.60E-05</b>
% weight change	<b>14.70</b>	<b>95</b>	<b>0.13</b>	<b>115</b>		<b>4.99E-06</b>
Gonad weight (g)	<b>0.06</b>	<b>80</b>	<b>0.42</b>	<b>100</b>		<b>1.29E-24</b>
Weight minus gonad (g)	<b>1.65</b>	<b>81</b>	<b>2.91</b>	<b>100</b>		<b>4.72E-13</b>
Gonad weight/weight	<b>0.03</b>	<b>80</b>	<b>0.13</b>	<b>60</b>		<b>1.14E-20</b>
Blood glucose start (mg/dL)	41.10	48	41.23	111		7.68E-01
Blood glucose end (mg/dL)	60.76	84	59.28	84		4.13E-01
Blood glucose change (mg/dL)	13.41	70	7.55	105		2.68E-01
Arginine blood glucose (mg/dL)	58.46	88	61.26	104		7.08E-01
Arginine response	19.97	69	23.93	50	6.25E-01	
Muscle triglyceride	0.19	83	0.20	102		7.67E-01
Visceral fat area (cm <sup>3</sup> )	0.21	88	0.27	108		7.67E-01
Liver area (cm <sup>3</sup> )	<b>0.31</b>	<b>88</b>	<b>0.50</b>	<b>108</b>		<b>6.67E-12</b>
Midgut length (cm)	<b>1.43</b>	<b>83</b>	<b>1.66</b>	<b>102</b>	<b>2.73E-05</b>	
Hindgut length (cm)	<b>0.64</b>	<b>87</b>	<b>0.78</b>	<b>107</b>		<b>6.77E-08</b>
Liver area/fish length	<b>0.07</b>	<b>87</b>	<b>0.10</b>	<b>107</b>	<b>7.36E-10</b>	
Midgut length/fish length	0.32	70	0.33	84	5.00E-01	
Hindgut length/fish length	<b>0.14</b>	<b>44</b>	<b>0.15</b>	<b>44</b>	<b>3.20E-02</b>	

Traits that are significantly different between sexes in bold. Students  $t$ -test used for normally distributed traits and Wilcox test used for non-normally distributed traits based on Shapiro-Wilks test. Adjusted false discovery rate  $p$ -value is shown [65]

gonad weight relative to body weight (surface = 0.13, Pachón = 0.14) is significantly greater than in males of the same age (surface = 0.04, Pachón = 0.02,  $n = 5$  fish per sex and population,  $t$ -test comparing sex: surface  $p = 0.003$ , Pachón  $p = 0.001$ ). Although F2 hybrid females are larger than males, we did not find evidence of elevated fat content; muscle triglyceride level and the amount of visceral fat surrounding the organs are not different between sexes (Table 1).

We examined the morphology of internal organs that play essential roles in metabolism: the liver, midgut, and hindgut. Since larger fish have bigger organs, we compared values relative to standard fish length. We found no difference in relative midgut length (average 0.32 male versus 0.33 female,  $p = 0.50$  FDR Wilcox), but the relative size of the liver and length of the hindgut are significantly greater in female F2 hybrids (Table 1, normalized liver = 0.07 male versus 0.10 female,  $p = 7.36 \times 10^{-10}$  FDR Wilcox, normalized hindgut = 0.14 male versus 0.15 female,  $p = 0.03$  FDR Wilcox). These results reveal a previously unappreciated sex bias in organ size in *A. mexicanus*.

We did not observe a significant difference between male and female 24-h fasting blood glucose levels in surface/Pachón F2 hybrids (Table 1). Overall, blood glucose levels increased after four months on the nutrient restricted diet (starting blood glucose: min = 23 mg/dL, max = 82 mg/dL, mean = 41.17 mg/dL, ending blood glucose: min = 23 mg/dL, max = 120 mg/dL, mean = 59.92 mg/dL). On average, male blood glucose levels increased more than females but not significantly (average 13.41 mg/dL male, 7.55 mg/dL female,  $p = 0.27$  FDR Wilcox). To probe the dynamics of blood glucose regulation, we measured blood glucose levels 5 h after F2 hybrids were injected with arginine. Arginine injection stimulates release of insulin and results in a drop in blood sugar in surface fish, but does not change blood sugar significantly in cavefish [5]. We did not observe a difference between male and female blood glucose levels in response to arginine injection (Table 1).

### Correlations between metabolism-related traits

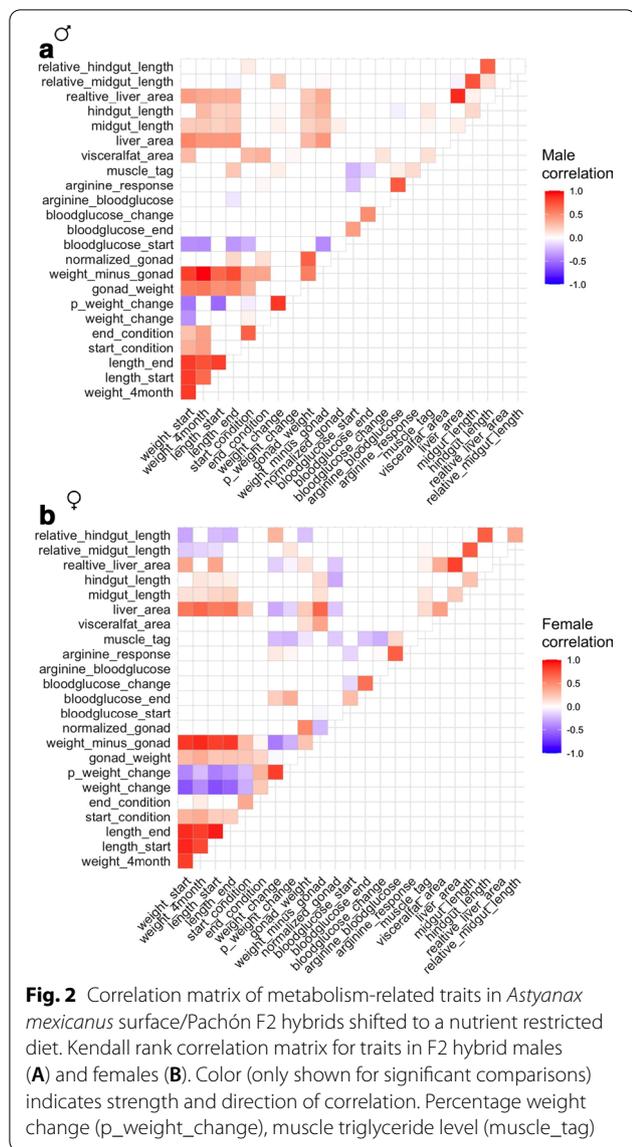
We examined pairwise correlations between the traits measured in our study to explore which factors are associated with tolerance to a nutrient restricted diet (Fig. 2A, B). We found that weight change negatively correlates with starting weight in both sexes (Figs. 2, 3A, B). Males and females that were greater than 2 g at the start of the feeding trial tended to lose weight and smaller fish tended to gain weight. To explore other variables that contribute to weight maintenance we used Akaike's Information Criterion (AIC) to compare linear models for weight change that included the other traits measured in our study. The

model with the lowest AIC value ( $-11.50$ ) suggests that starting weight, gonad weight, and hindgut length are significantly associated with weight change (Table 2). Unlike starting weight, gonad weight and hindgut length had a positive effect on weight change (Estimate = 0.83 and 0.52 respectively). In other words, smaller fish tend to gain weight, regardless of sex, and weight gain is associated with a bigger gonad and longer hindgut.

We found that starting and ending blood glucose level are positively correlated (Fig. 2, combined male and female  $p = 6.385e-07$ , Kendall). In males, starting blood glucose level negatively correlates with starting and ending weight, length, starting body condition, fish weight minus the gonad, and muscle triglyceride level (Fig. 2A). In other words, larger, fatter males have lower fasting blood sugar levels. Fasting blood glucose level at the end of the 4-month restricted diet is also negatively correlated with muscle triglyceride level, but not any other trait in males (Fig. 2A). In females, we observed a weak negative correlation between starting blood glucose level and fish weight subtracting the gonad (Fig. 2B). We found a more dramatic positive correlation between ending blood glucose, weight change, and percent weight change (Fig. 2B). These results suggest that higher blood sugar is associated with weight gain in females.

We analyzed linear models for ending blood sugar level that included sex and all of the traits measured in our study as variables. Examining the model with the lowest AIC value (429), we found that starting blood glucose level, muscle triglyceride level, and end weight and length are significantly associated with fasting blood sugar level on the nutrient restricted diet (Table 3). Higher body weight and blood sugar level are associated with higher ending blood sugar level (Estimate = 11.31, 0.35), while greater muscle triglyceride level and longer length is associated with lower ending blood sugar (Estimate =  $-26.57$ ,  $-23.99$ ). These results are surprising considering that cavefish have higher blood sugar and higher fat content [3, 5]. Our results suggest that the interaction of cave and surface fish genotypes in F2 hybrids may impact the association of fat content with fasting blood glucose level.

Arginine injection stimulates release of insulin which reduces blood sugar in surface fish [5]. Since cavefish are insulin resistant, arginine injection does not change blood sugar significantly [5]. In F2 hybrids, blood glucose level after arginine injection is positively correlated with muscle triglyceride level in males (Fig. 2A, B). The difference between fasting blood glucose and blood glucose after arginine injection (arginine response) is also positively correlated with muscle triglyceride level in males (Fig. 2A). In other words, fatter fish have higher blood sugar levels after arginine injection, suggesting a



possible association with impaired insulin response, similar to what is observed in cavefish. In females, arginine response is also positively correlated with weight change and percent weight change (Fig. 2B). We analyzed a linear model for arginine blood glucose that includes other traits measured in our study as variables and did not observe any significant associations in the model with the lowest AIC value (493, Table 4).

Liver size and gut length positively correlate with fish size in both sexes (Fig. 2A, B). We found that the relative size of the liver and relative length of the hindgut are significantly greater in females (Table 1). Relative hindgut length positively correlates with body condition factor at

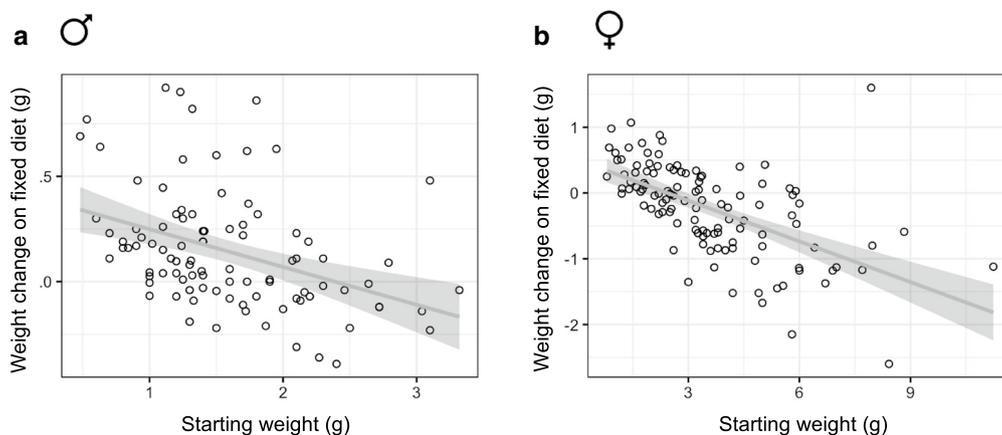
the start of the food-reduction trial in males and weight change in females, further suggesting the length of the hindgut may contribute to weight gain (Fig. 2A, B). We analyzed a linear model for hindgut length including sex, fish length, and other traits measured in our study as variables. Analyzing the model with the lowest AIC value (-184), we found that sex, fish length, weight change, and midgut length contribute significantly to hindgut length (Table 5). These quantitative traits have a positive association with hindgut length. In other words, a longer hindgut is linked with weight gain on a nutrient restricted diet. Combined, these results reveal that organ size is sexually dimorphic in *A. mexicanus* and that visceral organ morphology is associated with maintaining weight on limited nutrients.

### Genetic mapping of metabolic traits

We calculated genome wide logarithm of the odds (LOD) scores for each trait using a linkage map and genotype-by-sequencing markers generated as part of a previous study [11] (Additional files 1, 2). For traits that were significantly different between sexes, we compared genome wide scans for QTL, including sex as an additive covariate (accounting for sex differences, but assuming QTL have the same effect in both sexes) and interactive co-variate (assuming that QTL have a different effect depending on sex). We used a permutation test with 1000 iterations to establish LOD significance thresholds for each trait. Significant sex-interacting QTL were identified by comparing the genome wide LOD thresholds when sex is included as an additive versus interactive covariation (see “Methods” [12]). Table 6 shows the peak LOD score for each trait and model. We observed significant QTL associated with blood glucose level, standard fish length, body condition, weight change, percent weight change, gonad weight, liver area, and hindgut length. Some of these QTL display significant sex-bias as discussed in more detail below.

### QTL associated with blood glucose level

Tinaja and Pachón cavefish carry a coding mutation in *insulin receptor alpha* (*insra* ENSAMX00000001680) that is associated with insulin resistance and weight gain [5]. The mutation is necessary, but not sufficient, for elevated fasting blood sugar levels suggesting that elevated blood sugar in cavefish is likely the result of multiple genetic changes. Consistent with this hypothesis, we found that fasting blood glucose level is associated with significant QTL on linkage group nine of the genetic map used in this study and the *insra* mutation is on linkage group eight (Fig. 4A). The peak marker for fasting blood glucose at the start of the feeding trial at 151 cM on linkage group



**Fig. 3** Starting weight correlates with weight change on nutrient restricted diet in *Astyanax mexicanus* F2 hybrids. Scatterplot showing starting weight and weight change after male (A) and female (B) adult fish were individually housed and consumed a fixed diet for 4 months. Trendline with 95% confidence interval shown in grey

**Table 2** Linear model fitting for weight change

Term	Estimate	Std.error	Statistic	p.value
(Intercept)	0.17	0.17	0.98	3.30E-01
Weight start	<b>- 0.38</b>	<b>0.03</b>	<b>- 12.25</b>	<b>2.18E-16</b>
Gonad weight	<b>0.73</b>	<b>0.15</b>	<b>4.91</b>	<b>1.10E-05</b>
Blood glucose start	- 0.0017	0.0021	- 0.81	4.20E-01
Arginine blood glucose	0.0011	0.0012	0.95	3.46E-01
Visceral fat area	0.33	0.22	1.48	1.45E-01
Hindgut length	<b>0.52</b>	<b>0.20</b>	<b>2.63</b>	<b>1.14E-02</b>

Significant values in bold

**Table 3** Linear model for end blood glucose level

Term	Estimate	Std.error	Statistic	p.value
(Intercept)	<b>125.87</b>	<b>32.55</b>	<b>3.87</b>	<b>3.60E-04</b>
Weight 4 month	<b>11.31</b>	<b>5.47</b>	<b>2.07</b>	<b>4.47E-02</b>
Length end	<b>- 23.99</b>	<b>9.23</b>	<b>- 2.60</b>	<b>1.27E-02</b>
Gonad weight	1.38	10.55	0.13	8.97E-01
Arginine blood glucose	0.09	0.08	1.15	2.58E-01
Blood glucose start	<b>0.35</b>	<b>0.14</b>	<b>2.46</b>	<b>1.81E-02</b>
Muscle triglyceride	<b>- 26.57</b>	<b>11.32</b>	<b>- 2.35</b>	<b>2.35E-02</b>
Visceral fat area	1.72	14.93	0.12	9.09E-01
Liver area	- 4.30	16.76	- 0.26	7.99E-01

Significant values in bold

**Table 4** Linear model for arginine blood glucose

Term	Estimate	Std.error	Statistic	p.value
(Intercept)	28.07	16.50	1.70	9.56E-02
% weight change	0.40	0.24	1.66	1.04E-01
Gonad weight	- 23.30	14.34	- 1.62	1.11E-01
Blood glucose start	0.27	0.25	1.06	2.94E-01
Muscle triglyceride	30.77	19.68	1.56	1.25E-01
Visceral fat area	18.46	26.53	0.70	4.90E-01
Liver area	39.58	23.33	1.70	9.65E-02

**Table 5** Linear model for hindgut length

Term	Estimate	Std.error	Statistic	p.value
(Intercept)	<b>0.17</b>	<b>0.08</b>	<b>2.04</b>	<b>4.32E-02</b>
Sex	<b>0.08</b>	<b>0.03</b>	<b>3.18</b>	<b>1.75E-03</b>
Length start	<b>0.07</b>	<b>0.02</b>	<b>3.55</b>	<b>4.92E-04</b>
Weight change	<b>0.05</b>	<b>0.02</b>	<b>2.26</b>	<b>2.52E-02</b>
Midgut length	<b>0.10</b>	<b>0.03</b>	<b>3.17</b>	<b>1.79E-03</b>

Significant values in bold

9 has a LOD score of 7.15 accounting for 26% of the phenotype variance. The peak marker for fasting blood glucose at the end of the feeding trial is nearby at 152 cM on linkage group 9 with LOD score of 5.85 accounting for 13% of the phenotype variance. As expected, the cave genotypes at the peak markers are associated with the

highest blood glucose levels (Fig. 4B, C). These findings reveal a genomic location with a strong effect on fasting blood glucose level in *A. mexicanus*.

The peak marker for starting blood glucose maps to scaffold KB871607.1 on the Pachón cavefish genome. To investigate which genes within the QTL may be associated with cave adaptation, we determined several population genetic metrics using the coding regions of all of the genes on this scaffold utilizing whole genome sequencing

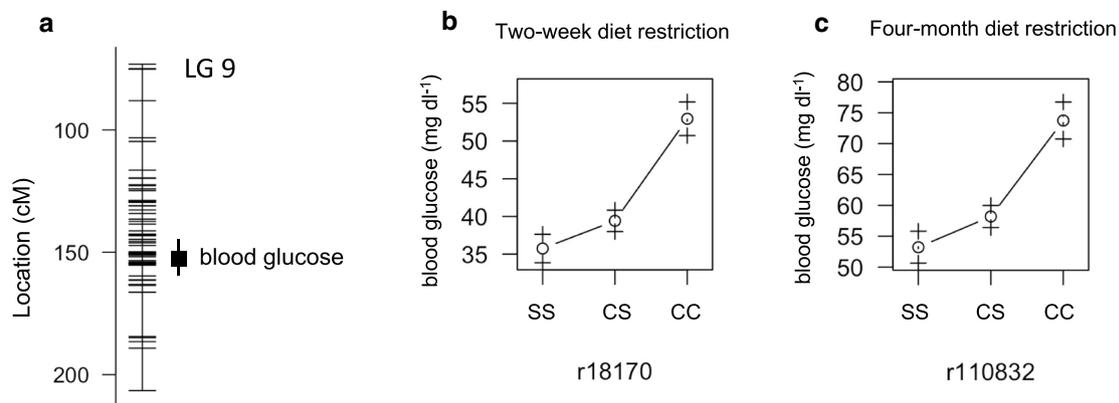
**Table 6** QTL mapping of metabolic traits

Trait	N	Covariate	LOD 5%	LOD 10%	LOD peak	PVE	Peak marker LG position	Scaffold	bp position
Blood glucose start	<b>108</b>	<b>None</b>	<b>4.22</b>	<b>3.81</b>	<b>7.16</b>	<b>26.31</b>	<b>r18170 9 150.9</b>	<b>KB871607</b>	<b>839,363</b>
Blood glucose end	<b>196</b>	<b>None</b>	<b>3.92</b>	<b>3.61</b>	<b>5.85</b>	<b>12.84</b>	<b>r110832 9 151.8</b>	<b>KB872022</b>	<b>216,703</b>
Weight start	219	add_int	7.35	6.59	4.03	8.13	r197874 19 36	KB882103	5,600,729
Weight start	219	Int	7.35	6.59	5.86	11.59	r230947 13 151	KB882132	1,682,036
Weight 4 month	210	add_int	7.00	6.55	4.01	8.42	r91812 19 64.3	KB871887	492,183
Weight 4 month	210	Int	7.00	6.55	5.46	11.28	r230947 13 151.0	KB882132	1,682,036
Length start	216	add_int	5.93	5.41	5.03	10.17	r231495 13 144	KB882132	577,777
Length start	216	Int	5.93	5.41	5.22	10.53	r91812 19 64.3	KB871887	492,183
Length start	<b>216</b>	<b>sex_int</b>	<b>2.77</b>	<b>2.37</b>	<b>3.63</b>	<b>7.45</b>	<b>r113255 25 90.7</b>	<b>KB872047</b>	<b>277,233</b>
Length end	199	add_int	5.75	5.27	4.35	9.58	r197874 19 36	KB882103	5,600,729
Length end	199	Int	5.75	5.27	4.83	10.58	r230947 13 151.0	KB882132	1,682,036
K start	<b>213</b>	<b>add_int</b>	<b>4.00</b>	<b>3.56</b>	<b>4.21</b>	<b>8.70</b>	<b>r132099 2 35.1</b>	<b>KB872277</b>	<b>193,967</b>
K end	<b>195</b>	<b>Add</b>	<b>3.98</b>	<b>3.68</b>	<b>5.65</b>	<b>12.49</b>	<b>r134839 2 97.6</b>	<b>KB872296</b>	<b>1,995,763</b>
K end	<b>195</b>	<b>sex_int</b>	<b>2.17</b>	<b>1.84</b>	<b>3.89</b>	<b>8.78</b>	<b>r113674 6 107</b>	<b>KB872052</b>	<b>163,270</b>
Weight change	213	Int	6.62	6.12	4.8	9.86	r107738 25 87.5	KB871991	413,427
Weight change	<b>213</b>	<b>sex_int</b>	<b>3.31</b>	<b>2.88</b>	<b>3.4</b>	<b>7.09</b>	<b>r107738 25 87.5</b>	<b>KB871991</b>	<b>413,427</b>
% weight change	<b>210</b>	<b>Add</b>	<b>4.03</b>	<b>3.73</b>	<b>4.04</b>	<b>8.48</b>	<b>r231495 13 144</b>	<b>KB882132</b>	<b>577,777</b>
Gonad weight	180	Int	7.96	6.90	5.84	13.88	r5323 1 41.8	KB871579	6,548,071
Gonad weight	180	Int	7.96	6.90	5.65	13.46	r28416 19 75.4	KB871633	1,033,602
Gonad weight	180	Int	7.96	6.90	4.99	11.99	r47899 22 143.9	KB871694	475,774
Gonad weight	<b>180</b>	<b>sex_int</b>	<b>3.79</b>	<b>3.23</b>	<b>3.03</b>	<b>7.46</b>	<b>r5323 1 41.8</b>	<b>KB871579</b>	<b>6,548,071</b>
Weight—gonad	181	Int	6.76	6.14	5.3	12.62	r230947 13 151	KB882132	1,682,036
Normalized gonad	140	Int	6.76	6.29	6.25	18.58	r177364 1 76.5	KB882087	2,900,519
Normalized gonad	140	Int	6.76	6.29	5.37	16.19	r104862 22 178.1	KB871966	141,642
Normalized gonad	140	Int	6.76	6.29	5.4	16.27	r1393 24 29.5	KB871578	4,785,777
Normalized gonad	<b>140</b>	<b>sex_int</b>	<b>3.26</b>	<b>2.97</b>	<b>3.63</b>	<b>11.26</b>	<b>r177364 1 76.5</b>	<b>KB882087</b>	<b>2,900,519</b>
Muscle TAG	192	None	4.10	3.70	na				
Visceral fat area	119	None	4.28	3.89	na				
Liver area	185	add_int	6.07	5.64	4.37	10.31	r154446 7 153	KB873184	12,227
Liver area	185	Int	6.07	5.64	5.38	12.53	r230914 13 150.8	KB882132	1,586,871
Liver area	<b>185</b>	<b>Int</b>	<b>6.07</b>	<b>5.64</b>	<b>6.39</b>	<b>14.71</b>	<b>r28572 19 74.2</b>	<b>KB871633</b>	<b>467,313</b>
Liver area	<b>185</b>	<b>sex_int</b>	<b>2.93</b>	<b>2.50</b>	<b>3.16</b>	<b>7.56</b>	<b>r248014 7 153</b>	<b>KB882149</b>	<b>2,507,088</b>
Midgut length	196	sex_int	5.09	4.75	3.43	7.74	r195661 4 145	KB882102	1,812,243
Hindgut length	<b>196</b>	<b>add_int</b>	<b>4.05</b>	<b>3.65</b>	<b>4.64</b>	<b>10.33</b>	<b>r172341 10 68.3</b>	<b>KB882083</b>	<b>490,666</b>
Hindgut length	<b>196</b>	<b>sex_int</b>	<b>2.30</b>	<b>1.95</b>	<b>3.14</b>	<b>7.11</b>	<b>r33595 6 62.4</b>	<b>KB871649</b>	<b>380,272</b>
R liver area	185	add_int	5.81	5.40	5.62	13.06	r154446 7 153	KB873184	12,227
R liver area	<b>185</b>	<b>sex_int</b>	<b>2.79</b>	<b>2.43</b>	<b>2.95</b>	<b>7.08</b>	<b>r170604 21 12.8</b>	KB882082	730,436
R midgut length	194	Add	5.23	4.78	na				
R hindgut length	<b>194</b>	<b>add_int</b>	<b>5.11</b>	<b>4.79</b>	<b>4.68</b>	<b>10.51</b>	<b>r172341 10 68.3</b>	<b>KB882083</b>	<b>490,666</b>
R hindgut length	<b>194</b>	<b>sex_int</b>	<b>2.29</b>	<b>1.88</b>	<b>3.83</b>	<b>8.69</b>	<b>r59663 4 220</b>	<b>KB871744</b>	<b>301,165</b>

Significant traits in bold. Number of individuals (N), covariate model (add = additive, int = interactive, sex\_int = sex interacting QTL), logarithm of the odds (LOD) significance thresholds, peak marker LOD score, percent variance explained (PVE), peak marker name and position on indicated linkage group (LG) in centimorgans, and position on Pachón cavefish genome scaffold in base pairs (bp). Abbreviations for traits: body condition factor (K), triglyceride (TAG), length relative to standard fish length (R)

data from wild caught cavefish and surface fish populations [13]. This dataset includes Tinaja (n = 10), Pachón (n = 10), and Molino (n = 9) cave populations and two surface fish populations: Río Choy (n = 9), which is

representative of our lab strains and is more closely related to the stock of fish that invaded the Molino cave, and Rascón (n = 8), which is more representative of a separate stock of surface fish that invaded the Tinaja and



**Fig. 4** Quantitative trait loci associated with blood glucose level in *Astyanax mexicanus*. **A** Location of markers (horizontal lines) on linkage group 9 indicating the peak position and bayes credible interval (box and whiskers) for QTL associated with 24 h fasting blood glucose level. **B** Average fasting blood glucose level ( $\pm$  1 SE) of F2 surface/Pachón hybrids of the indicated peak linkage group 9 QTL marker genotype after 2-weeks on the nutrient restricted diet (S = surface allele, C = cave allele). **C** Average fasting blood glucose level ( $\pm$  1 SE) of F2 surface/Pachón hybrids of the indicated peak linkage group 9 QTL after 4 months on the nutrient restricted diet for each genotype (S = surface allele, C = cave allele)

Pachón caves [13]. We found that one gene on scaffold KB871607.1, *ADAM metallopeptidase domain 9* (*adam9*) exhibits evidence of non-neutral evolution and is therefore likely to be under selection (estimation of hapFLK statistic resulted in 20 p-values less than 0.05 for this gene, Additional file 3). We aligned the predicted cDNA sequence of *adam9* using sequences from the surface fish and Pachón cavefish reference genomes and whole genome sequencing data from wild caught fish [13] and discovered four predicted single amino acid substitutions (F87L, Q384R, D553E, K681N) and one predicted frame shift (T255Ffs, Table 7, Additional file 4). The amino acid substitutions are homozygous and different in all samples while the frame shift is present in 5/10 Pachón cavefish and 2/10 Rio Choy surface fish. *Adam9* has not been associated with blood sugar regulation to our knowledge but investigating the impact of these coding differences represents an avenue for future discoveries. The nearest gene to the peak marker, *nuclear repressor corepressor 2* (*ncor2*), is however a part of a gene network that regulates metabolism [14]. While *ncor2* does not display

evidence of non-neutral evolution comparing the coding sequence, there could be regulatory changes that impact *ncor2* expression. Exploring *ncor2* gene regulation and *adam9* gene function represent promising directions for future work on the evolution of cavefish blood glucose regulation.

#### QTL associated with weight, length, and body condition

We found that weight and length at the start and end of the feeding trial are associated with QTL on linkage group 13 and 19 with peak markers in close proximity or at the same position (Fig. 5A, E, Table 6). On linkage group 13, start/end weight and end length produce a peak at 151 cM accounting for approximately 12% of the variance in each trait. The cave genotype is associated with lower weight and shorter length in both males and females (Fig. 5B, C). We found that percent weight change is also associated with a QTL on linkage group 13 at 144 cM that accounts for 8% phenotype variance. In this case, the cave genotype is associated with weight gain (Fig. 5D, E). The peak markers for the linkage group

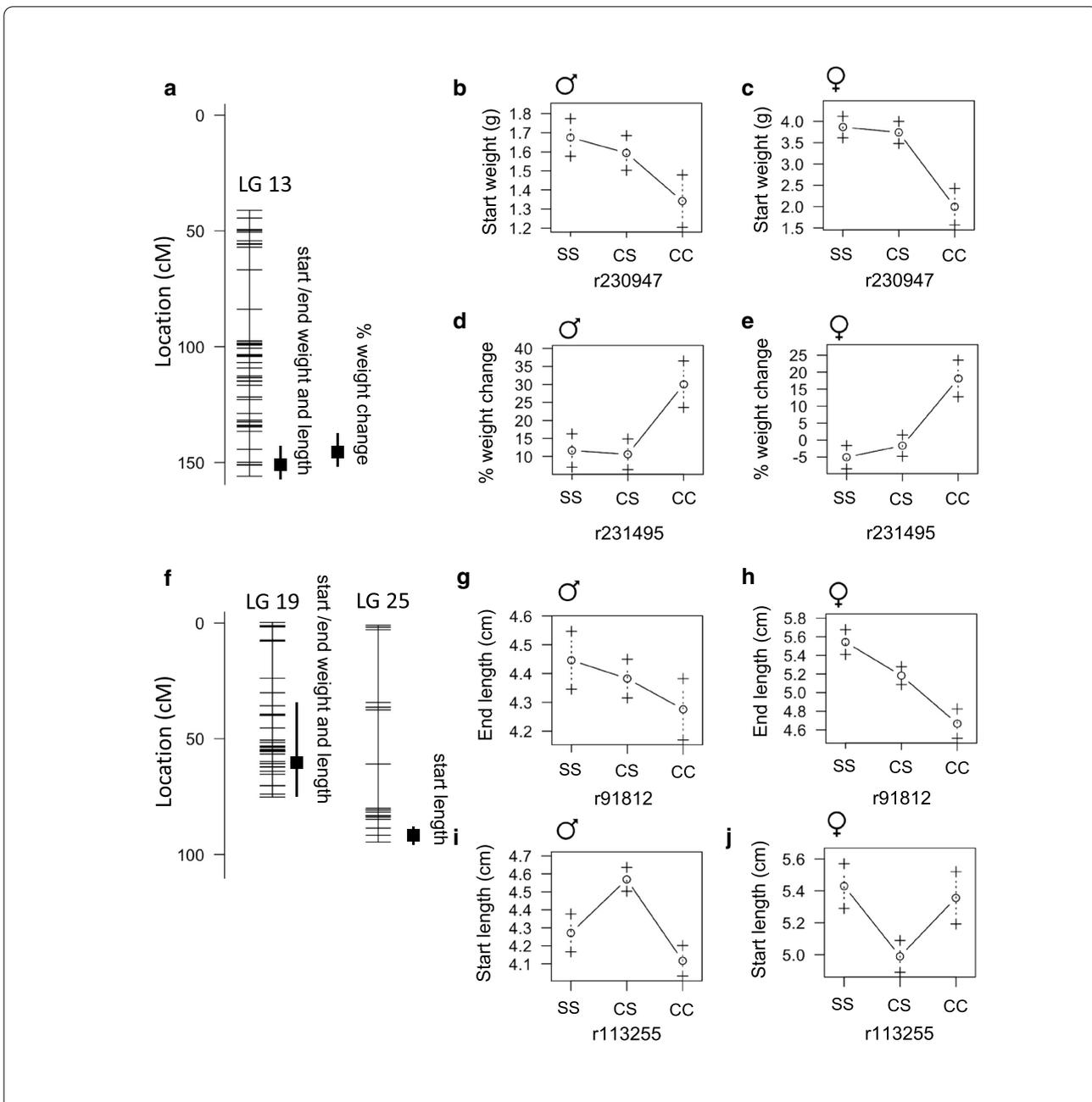
**Table 7** Predicted changes to the amino acid sequence based on cDNA alignments of genes within metabolism-related QTL that display evidence of non-neutral evolution by HapFLK analysis and have fixed differences between Río Choy surface fish and Pachón cavefish (n = 10 fish per population)

Trait	Gene name	Predicted surface to Pachón amino acid changes
Blood glucose	<i>adam9</i>	F87L, T255Ffs*, Q384R, D553E, K681N
Weight and length	<i>ano5b</i>	K67E*, S119L, K639R
Weight and length	<i>dpep2</i>	No gene annotated in surface fish genome
Body condition	<i>ahnak2</i>	E104K*
Body condition	<i>exoc3l4</i>	K431R*, L435M, L483S*, H524Q*, L527E
Gonad weight	<i>amfr</i>	Two synonymous differences
Gonad weight	<i>ano9a</i>	L42W*, E178D*, E362D, T735I*, A811G*, L744Q*, P749H, V752F*, G757S*, E835K
Gonad weight	<i>cd44a</i>	I9M*, Y23_F34del*, Q53_L54insRSCARFWVQHWPV*
Gonad weight	<i>slc1a2b</i>	Q215E
Gonad weight	ENSAMXG00000016170	No gene annotated in surface fish genome
Gonad weight	<i>gzm3</i>	V28A*, R33L*, E131V, D175S, S196A, T201S, K204T, K216Q*, L226G, T227F
Gonad weight	<i>si:dkey-100n23.5</i>	R184del*
Gonad weight	<i>tnsf13b</i>	One synonymous difference
Liver size	<i>si:ch1073-174d20.1</i>	A9V, S101N
Liver size	<i>nfat5a</i>	D43A, T550N, N611K, Q918K, N1079_Q1083del*, Q1084E*, T1093S, S1293A, I1334M
Hindgut length	<i>cbf2</i>	Y389H*, M542T*
Hindgut length	<i>cbf1</i>	R27T*, R46Q*, S48P*, V75A*, F101Y*, H142D, E264Q, M266I, K299N, T387I*, E561D, I621N*, K671T*
Hindgut length	<i>zgc:110782</i>	E92K
Hindgut length	<i>gpr119</i>	No gene annotated for surface fish genome
Hindgut length	<i>rbm14a</i>	Two synonymous differences
Hindgut length	<i>rin1a</i>	S203N, M464Q*, G768S
Hindgut length	<i>mark2a</i>	S392N*, N722K
Hindgut length	<i>adssl</i>	F108Lfs*, N224Y, K338R, N367Tfs*, A404Lfs*, C444R
Hindgut length	<i>spry3</i>	P3L
Hindgut length	<i>rce1b</i>	P116S, I147Lfs*#, F193del*#
Hindgut length	<i>tenm2</i>	W411Ffs*
Hindgut length	<i>zgc:77262</i>	G288_A291del*, P293_P294del*, G327S
Hindgut length	<i>lsm14b</i>	G78C*, A173T*, L266S*
Hindgut length	<i>fam65b/ripor2</i>	D468Efs*#, R664H
Hindgut length	<i>rfx5</i>	G342R*
Hindgut length	<i>txnla</i>	A496T, E504K*, F529C*
Hindgut length	<i>cpsf1</i>	L53P*, E355D*, A702T*
Hindgut length	<i>adck5</i>	M552I, A569V*

See Additional file 4 for gene and transcript IDs. Asterisk (\*) indicates that the change is not homozygous and different between all samples and number sign (#) indicates the change is annotated relative to surface fish reference genome although that is not the most abundant variant in the surface fish samples (see Additional file 4 for details)

13 QTL map to Pachón scaffold KB882132.1. None of the genes on this scaffold exhibit evidence of non-neutral evolution (Additional file 3). On linkage group 19, the peak markers for starting weight, starting length, ending weight, and ending length are at 36 cM, 64.3 cM, 64.3 cM and 62.4 cM respectively likely representing the same overlapping QTL. The cave genotypes are associated with lower weight and shorter length in both males and females (Fig. 5G, H). The peak markers map to Pachón scaffold KB882103.1 and we found that two genes on this

scaffold, *ano5b* (ENSAMXG00000005967) and *dpep2* (ENSAMXG00000005810) exhibit evidence of non-neutral evolution (Additional file 3). The *ano5b* gene encodes anoctamin 5b, a calcium activated chloride channel (zfin.org). Mutations in the human *ANO5* gene are associated with reduced bone mineral density [15]. The *dpep2* gene encodes dipeptidase 2, a protein predicted to have metallopeptidase activity (zfin.org). Morpholino knockdown of *dpep2* reduces body size in zebrafish [16]. We aligned the predicted cDNA sequences of *ano5b* from 10 surface fish



**Fig. 5** Quantitative trait loci associated with weight, length, and percentage weight change on a nutrient restricted diet in *Astyanax mexicanus*. **A** Location of markers (horizontal lines) on linkage group 13 indicating the peak position and bayes credible interval (box and whiskers) for QTL associated with the indicated trait. **B, C** Average weight ( $\pm$  1 SE) of adult F2 surface/Pachón hybrid females and males of the indicated peak marker genotype (S = surface allele, C = cave allele). **D, E** Average weight change as a percentage of total weight ( $\pm$  1 SE) of F2 surface/Pachón hybrids of the indicated linkage group 13 peak marker genotype (S = surface allele, C = cave allele, after 4 months on a nutrient restricted diet). **F** Location of markers (horizontal lines) on linkage group 19 and 25 indicating the peak position and bayes credible interval (box and whiskers) for QTL associated with the indicated trait. **G, H** Average standard length ( $\pm$  1 SE) of adult F2 surface/Pachón hybrid females and males of linkage group 19 peak marker genotype (S = surface allele, C = cave allele). **I, J** Average standard length ( $\pm$  1 SE) of adult F2 surface/Pachón hybrid females and males of linkage group 25 peak marker genotype (S = surface allele, C = cave allele)

and 10 Pachón cavefish and discovered three predicted amino acid substitutions (K67E, S119L, K639R, Table 7, Additional file 4). There is no *dpep2* gene annotated on the surface fish genome, so we were not able to investigate coding changes using available sequencing data.

In addition to the QTL on linkage groups 13 and 19, we found a sex-specific QTL associated with starting length on linkage group 25 with the peak marker accounting for 7.45% of phenotype variance; in males, the heterozygous genotype is associated with the greatest length and, in females, the heterozygous genotype is associated with the shortest length (Fig. 5F, I, J). The QTL maps to Pachón scaffold KB872047.1 and none of the genes on this scaffold exhibit evidence of non-neutral evolution. Exploring the impact of the coding changes we discovered in *ano5b* represents a promising future direction for understanding the genetic basis cavefish length and weight evolution.

Body condition is a measure of fish nourishment level and is a function of length and weight. In the wild, cavefish have better body condition compared to surface fish [17]. We identified QTL associated with body condition that do not overlap the QTL associated with weight and length. Body condition at the start of the feeding trial is associated with a QTL on linkage group 2 with the peak marker at 35.1 cM accounting for 8.7% of phenotype variance. End condition is associated with a QTL on linkage group 2 with the peak marker at 97.6 cM accounting for 12% of phenotype variance. These are likely the same overlapping QTL and the cave genotypes are associated with better body condition in both males and females (Fig. 6A–E). The peak marker for end body condition maps to Pachón scaffold KB872296.1. Based on analysis of coding sequences, we found that four genes on this scaffold exhibit evidence of non-neutral evolution and two of the four genes also have fixed differences between Pachón cavefish and both Río Choy and Rascón surface fish: *ahnak2* (ENSAMXG00000001960) that is predicted to encode a nucleoprotein involved in calcium signaling, and *exoc3l4* (ENSAMXG00000001944) that encodes an exocyst complex component associated with global cortical glucose metabolism in humans [18] (Additional file 3). The single *ahnak2* gene in the Pachón genome is annotated as three separate genes in the surface fish genome. We compared the predicted cDNA sequence from Río Choy surface fish and Pachón cavefish samples using the surface fish gene that corresponds to the first part of the Pachón gene and discovered a single amino acid substitution (E104K) that is homozygous in all of the Pachón cavefish samples ( $n = 10/10$ ) and half of the Río Choy surface fish samples ( $n = 5/10$ , Table 7, Additional file 4). Based on alignments using the same sequencing date we discovered that the Pachón cavefish EXOC3L4 contains five predicted amino acid substitutions (K431R,

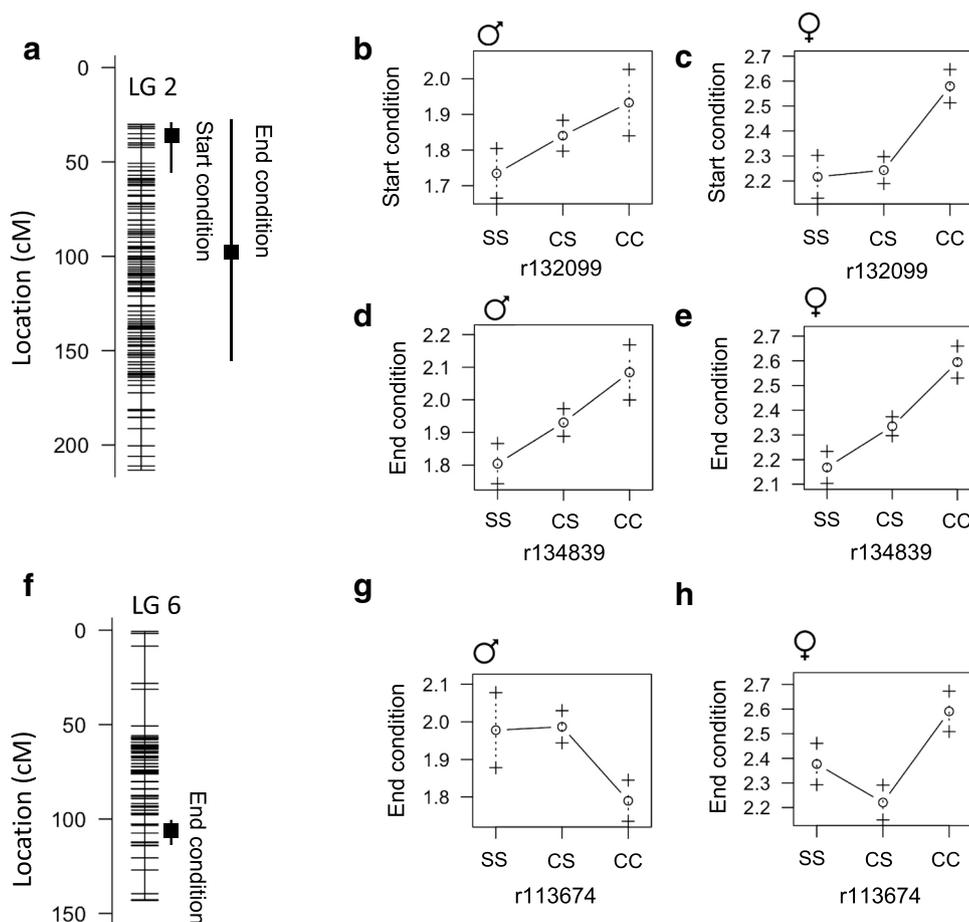
L435M, L483S, H524Q, L527E, Table 7, Additional file 4). Investigating the impact of the predicted coding changes we observed in *ahnak2* and *exoc3l4* may lead to a better understanding of enhanced body condition in cavefish.

We observed a sex-specific QTL on linkage group 6 associated with end body condition; in males, the cave genotype is associated with lower body condition and, in females, the cave genotype is associated with higher body condition (Fig. 6F–H). The peak marker maps to Pachón scaffold KB872052.1 and none of the genes on this scaffold exhibit evidence of non-neutral evolution. To our knowledge, *ahnak2* or *exoc3l4* have not been previously implicated in regulating body size, but these genes represent promising candidates for future analysis.

In summary, we found that surface fish genotypes at QTL for weight and length are associated with bigger fish. However, cave genotypes are associated with better body condition and greater weight gain on a limited diet. These genetic differences reflect the growth strategies that evolve in the river versus cave environment.

#### QTL associated with gonad weight

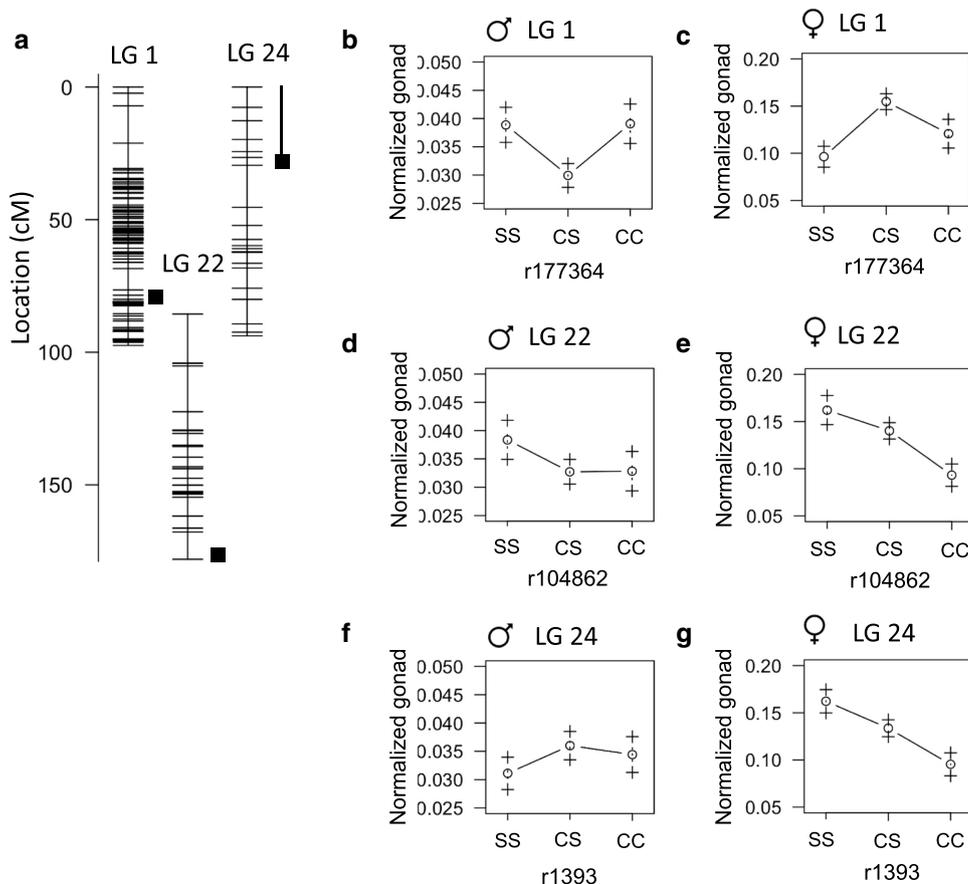
Gonad development in *A. mexicanus* has not been explored, but it is conceivable that the cave environment may have selected for altered reproductive strategies. Cavefish eggs are larger than surface fish eggs and cavefish tend to have smaller clutch sizes [19, 20]. The markers with the highest LOD scores for gonad weight appear on linkage group 1 (41.8 cM), 19 (75.4 cM), and 22 (143.9 cM) and account for 14%, 13%, and 12% of phenotype variance respectively (Table 6). The peak on linkage group 19 is near the QTL for fish length and weight indicating this likely represents the effect of fish size. Normalized gonad weight is also associated with peak markers on linkage group 1 and 22 albeit at different positions (76.5 cM and 178.1 cM) explaining 19% and 16% of phenotype variance, respectively (Fig. 7A). The peak marker on linkage group 1 rises above the LOD threshold for a sex-associated trait and the heterozygous genotype is associated with smaller gonads in males and larger gonads in females (Fig. 7B, C). The peak marker maps to Pachón scaffold KB871579.1. We found nine genes on this scaffold exhibit evidence of non-neutral evolution and five of the nine have fixed differences between Pachón cavefish and both river populations: *ano9a* (ENSAMXG000000017855) predicted to encode a protein involved in calcium activated phospholipid scrambling, *amfr* (ENSAMXG000000017688) that encodes an E3 Ubiquitin Ligase, *cd44a* (ENSAMXG000000017603) that encodes a cell surface receptor, ENSAMXG000000016170 that encodes a NACHT, LRR and PYD domain-containing protein, and *slc1a2b* (ENSAMXG000000017604) which is expressed in the nervous system and is predicted



**Fig. 6** Quantitative trait loci associated with body condition factor in *Astyanax mexicanus*. **A** Location of markers (horizontal lines) on linkage group 2 indicating the peak position and bayes credible interval (box and whiskers) for QTL associated with body condition factor (fish weight (g) × 100/ fish length (cm)<sup>3</sup>). **B, C** Average body condition factor (± 1SE) of adult F2 surface/Pachón hybrid females and males of the indicated linkage group 2 peak marker genotype (S = surface allele, C = cave allele, fish fed ad libitum). **D, E** Average body condition factor (± 1SE) of adult F2 surface/Pachón hybrid females and males of the indicated linkage group 2 peak marker genotype (S = surface allele, C = cave allele, fish fed restricted diet for 4 months). **F** Location of markers (horizontal lines) on linkage group 6 indicating the peak position and bayes credible interval (box and whiskers) for sex-dependent QTL associated with body condition factor after 4 months on a nutrient restricted diet. **G, H** Average body condition factor (± 1SE) of adult F2 surface/Pachón hybrid females and males of the indicated linkage group 6 peak marker genotype (S = surface allele, C = cave allele, fish fed restricted diet for 4 months)

to encode a solute transporter (zfn.org, Additional file 3). CD44 is involved in oocyte maturation in mammals and expressed in ovarian cancer stem cells [21, 22], but a role for CD44 in fish gonads has not been explored

to our knowledge. We discovered one predicted amino acid substitution (I9M), an eleven amino acid deletion (Y23\_F34del), and a thirteen amino acid insertion (Q53\_L54insRSCARFWVQHPV) comparing the predicted



**Fig. 7** Quantitative trait loci associated with gonad weight in *Astyanax mexicanus*. **A** Location of markers (horizontal lines) on linkage group 1, 22, and 24 indicating the peak position and bayes credible interval (box and whiskers) for QTL associated with the weight of the gonad normalized to fish weight. **B–G** Average normalized gonad weight ( $\pm$  1SE) of adult F2 surface/Pachón hybrid females and males of the indicated linkage group peak marker genotype (S = surface allele, C = cave allele)

cDNA sequence of *cd44a* between Río Choy surface fish and Pachón cavefish (Table 7,  $n=10$  fish per population). Using the same sequencing data, we found 10 predicted amino acid substitutions in Pachón Ano9a (L42W, E178D, E362D, T735I, A811G, L744Q, P749H, V752F, G757S, E835K, Table 7, Additional file 4), one amino acid substitution in Pachón Slc1a2b (Q215E), and two cDNA sequence differences in *amfr* that are predicted to cause synonymous changes. There is no gene on the surface fish genome that corresponds to Pachón cavefish ENSAMXG00000016170. Our results reveal possible genetic changes that contribute to size differences in the gonads of surface and cave-adapted *A. mexicanus*.

We found an additional peak marker associated with normalized gonad weight on linkage group 24 that explains 16% of phenotype variance (Fig. 7A). The homozygous surface genotype is associated with the

lowest normalized gonad weight in males, and greatest gonad weight in females (Fig. 7E, G). The peak marker maps to Pachón scaffold KB871578.1 and three genes on this scaffold exhibit evidence of non-neutral evolution: *gzm3* (ENSAMXG00000002176) that encodes a protein with endopeptidase activity, *si:dkey-100n23.5* (ENSAMXG00000001983) that encodes a G-protein coupled receptor, and *tnfsf13b* (ENSAMXG00000003788) that encodes a cytokine which functions in B cell activation (zfin.org, Additional file 3). We discovered ten predicted amino acid substitutions in Pachón Gzm3 (V28A, R33L, E131V, D175S, S196A, T201S, K204T, K216Q, L226G, T227F) and one predicted single amino acid deletion in Si:dkey-100n23.5 (R184del) by comparing cDNA sequences from Río Choy surface fish and Pachón cavefish ( $n=10$  fish per population, Table 7, Additional file 4). The Pachón cavefish *tnfsf13b* cDNA sequence contained

one base pair change that is predicted to result in a synonymous change. Exploring a link between the genes that contain coding changes and gonad development may reveal the genetic basis of increased egg size and reduced clutch size in Pachón cavefish.

#### QTL associated with liver size

Cavefish livers are enlarged and accumulate more fat compared to surface fish livers [3]. We found that liver area is associated with a QTL that overlaps with weight and length QTL on linkage groups 13 and 19 (Fig. 8A, Table 6, 151 cM and 74.2 cM accounting for 13% and 15% of phenotype variance, respectively). The heterozygous genotype on linkage group 13 is associated with the largest liver in both males and females (Fig. 8B, C). This is contrary to the fish weight/length QTL, for which the homozygous surface genotype is associated with the largest fish (Fig. 8B, C). Also, different than the QTL for length on linkage group 19, the cave genotype is associated with greater liver size in males (Fig. 8D). We observed an additional QTL for liver size on linkage group 7 at 153 cM accounting for 10% of phenotype variance (Fig. 8F). The liver size QTL at linkage group 7 is significant and sex specific. In males, the heterozygous genotype is associated with intermediate liver size and, in females, the heterozygous genotype is associated with the smallest liver size (Fig. 8G, H). Liver size normalized to fish length is also associated with the same peak marker on linkage group 7 that approaches significance (Table 6). The peak marker maps to Pachón scaffold KB873184.1 and none of the genes on the scaffold exhibit evidence of non-neutral evolution.

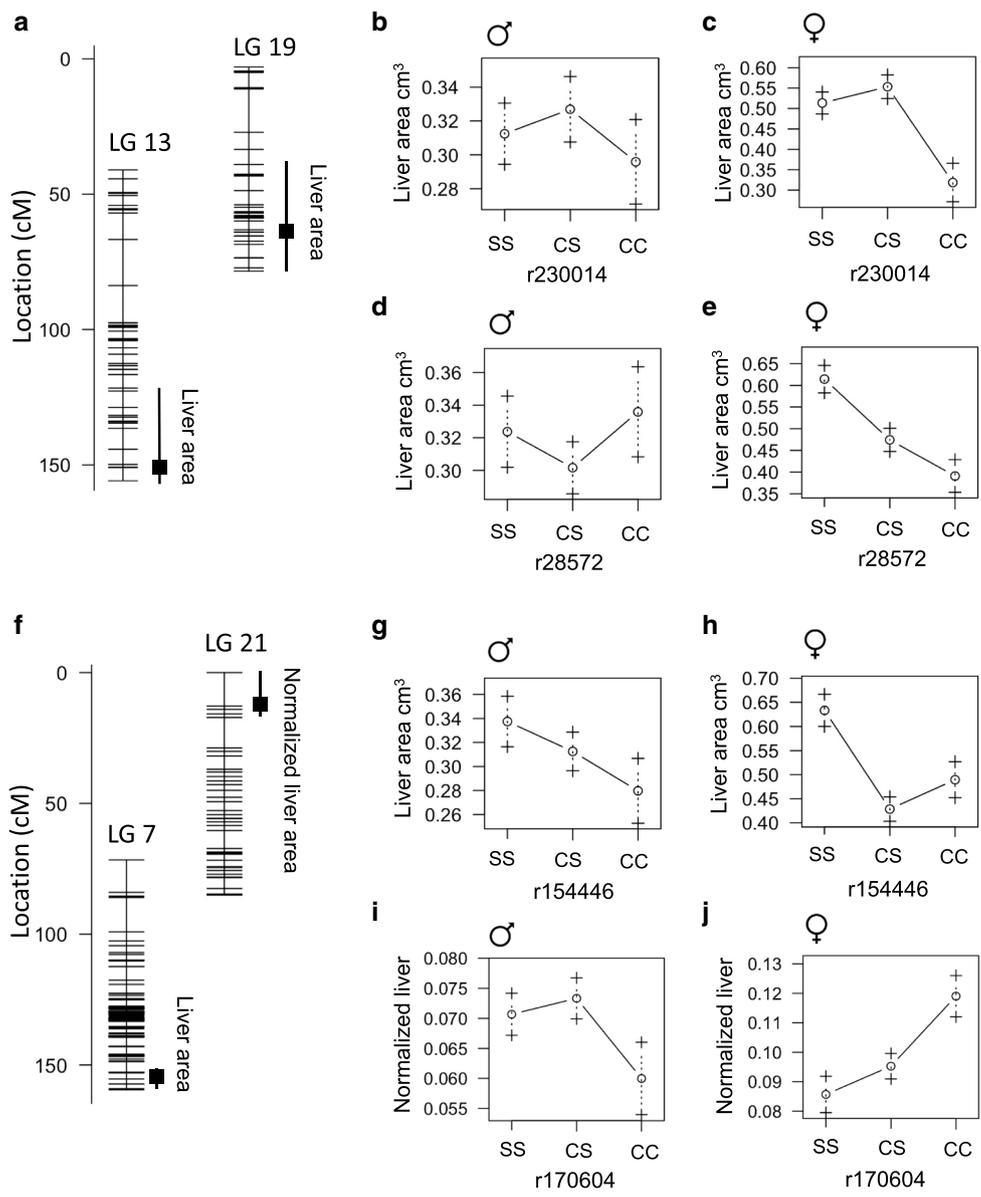
We found an additional sex-interacting QTL on linkage group 21 (12.8 cM, 7% variance) associated with normalized liver size; the homozygous cave genotype is associated with the largest liver in females and the smallest liver in males (Table 6, Fig. 8I, J). The overlap in fish length and liver size QTL on linkage groups 13 and 19 suggest that the association between fish size and liver size may contribute to the QTL. However, the QTL on linkage groups 7 and 21 may be specifically associated with sexual dimorphism of the liver. The peak marker on linkage group 21 maps to Pachón scaffold KB882082.1; five of the genes on this scaffold exhibit evidence of non-neutral evolution and two have fixed differences between Pachón cavefish and both surface fish populations: *nfat5a* (ENSAMXG00000013845) that encodes a transcription factor involved in immune response [23], and *dbn1* (ENSAMXG00000013840) a gene that is mutated in humans with Hermansky–Pudlak Syndrome 7 which is characterized by hypopigmentation and lysosomal storage defects [24] (Additional file 3). We discovered eight predicted amino acid substitutions (D43A, T550N,

N611K, Q918K, Q1084E, T1093S, S1293A, I1334M) and a five amino acid deletion (N1079\_Q1083del) in *Nfat5a*, and two amino acid substitutions in *Dbn1* comparing predicted cDNA sequences between Río Choy surface fish and Pachón cavefish (n = 10 fish per population, Table 7, Additional file 4). Investigating the impact of the coding changes may lead to insights into sex- and morphotype-specific liver development.

#### QTL associated with length of the hindgut

We found that hindgut length and relative hindgut length are associated with a QTL on linkage group 10 at 68.3 cM accounting for 10% of phenotype variance (Fig. 9A, Table 6). The heterozygous genotype is associated with the longest hindgut length in both males and females (Fig. 9E, G). The homozygous cave genotype is associated with a longer hindgut in males and shorter hindgut in females compared to the homozygous surface genotype (Fig. 9E, G). The peak marker maps to scaffold KB882083.1 on the Pachón cavefish genome and we found 16 genes on this scaffold exhibit evidence of non-neutral evolution (Additional file 2). Twelve of these genes have fixed differences in the coding regions comparing Pachón cavefish to both Río Choy and Rascón surface fish populations (Tables 7, 8). Two genes are predicted to encode complement factor B-like proteins that are involved in recognizing and eliminating bacteria (herein referred to as *cfb1* and *cfb2*). We identified thirteen predicted amino acid substitutions in Pachón cavefish *Cfb1* (R27T, R46Q, S48P, V75A, F101Y, H142D, E264Q, M266I, K299N, T387I, E561D, I621N, K671T) and three predicted amino acid substitutions in Pachón *Cfb2* (Y389H, M542T) comparing cDNA sequences from Río Choy surface fish and Pachón cavefish (n = 10 fish per population, Table 7, Additional file 4). Changes to *cfb1/cfb2* may reflect adaptation to the cave microbial landscape [25].

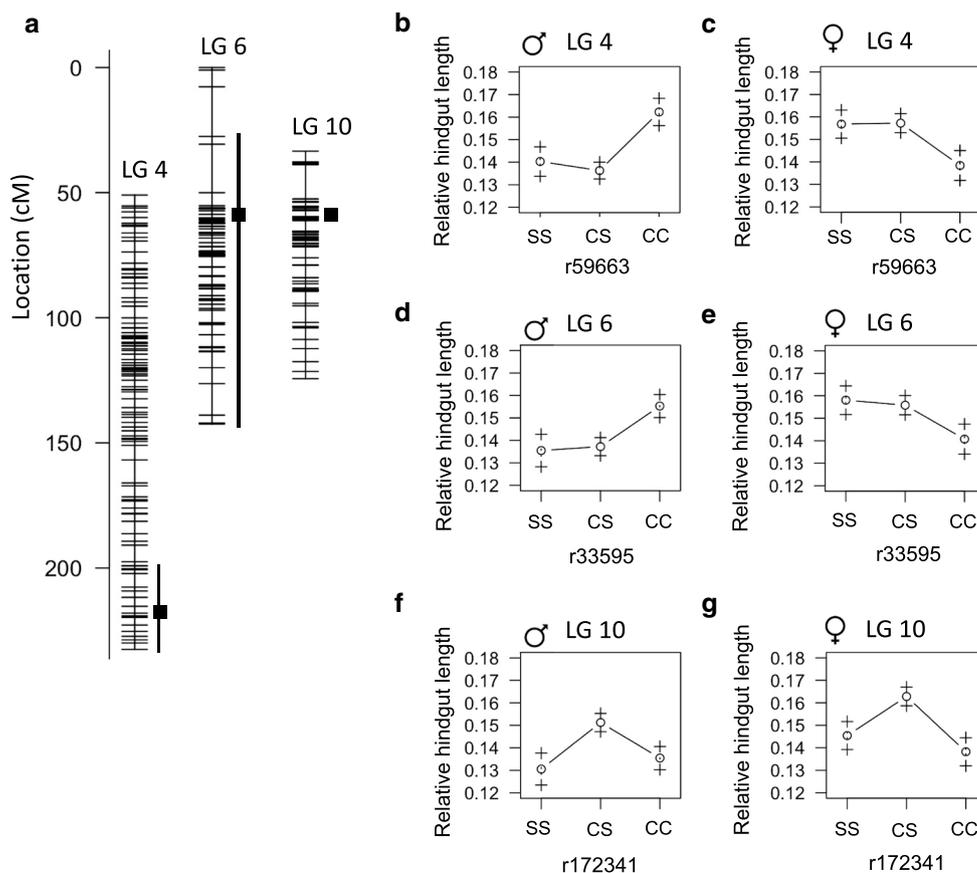
Genes with high divergence rank in only Pachón cavefish compared to Río Choy surface fish include *zgc:77262 Lark-like RNA binding protein* (ENSAMXG00000025173), *Zgc:110782* (ENSAMXG00000007312), which has similarity to *prostaglandin F synthase*, and *G-protein coupled receptor 119* (*gpr119*, ENSAMXG00000007067) (Table 7). Each of these genes has reported roles in the gut: the Lark RNA binding protein controls infection-induced RNA splicing [26], prostaglandins protect the gastrointestinal mucosal barrier from noxious chemicals [27], and GPR119, a glucose-dependent insulinotropic receptor, senses fat and regulates appetite and gut motility via the release of glucagon-like peptide 1 (GLP-1), peptide YY (PYY), and glucose-dependent insulinotropic peptide (GIP) from enteroendocrine cells [28–30]. We discovered one



**Fig. 8** Quantitative trait loci associated with liver area in *Astyanax mexicanus*. **A** Location of markers (horizontal lines) on linkage group 13 and 19 indicating the peak position and bayes credible interval (box and whiskers) for QTL associated with liver area. **B, C** Average liver area ( $\pm$  1SE) of adult F2 surface/Pachón hybrid females and males of the indicated linkage group 13 peak marker genotype (S = surface allele, C = cave allele). **D, E** Average liver area ( $\pm$  1SE) of adult F2 surface/Pachón hybrid females and males of the indicated linkage group 19 peak marker genotype (S = surface allele, C = cave allele, fish fed restricted diet for 4 months). **F** Location of markers (horizontal lines) on linkage group 7 indicating the peak position and bayes credible interval (box and whiskers) for QTL associated with liver area. **G, H** Average liver area ( $\pm$  1SE) of adult F2 surface/Pachón hybrid females and males of the indicated linkage group 7 peak marker genotype (S = surface allele, C = cave allele). **I, J** Average normalized liver area ( $\pm$  1SE) of adult F2 surface/Pachón hybrid females and males of the indicated linkage group 21 peak marker genotype (S = surface allele, C = cave allele)

predicted amino acid substitution in the Pachón Lark-like RNA binding protein (E92K) comparing the cDNA sequences from Río Choy surface fish and Pachón cave-fish (n = 10 fish per population, Table 7). Using the same

sequencing data, we found one predicted amino acid substitution (G327S), a four amino acid deletion (G288\_A291del), and two amino acid deletion (P293\_P294del) in the Pachón Prostaglandin F synthase-like protein



**Fig. 9** Quantitative trait loci associated with hindgut length in *Astyanax mexicanus*. **A** Location of markers (horizontal lines) on linkage group 4, 6, and 10 indicating the peak position and bayes credible interval (box and whiskers) for QTL associated with relative hindgut length. **B–G** Average relative hindgut length ( $\pm$  1SE) of adult F2 surface/Pachón hybrid females and males of the indicated linkage group peak marker genotype (S = surface allele, C = cave allele)

(Table 7, Additional file 4). There is no *gpr119* gene annotated on the surface fish genome. Potential links between gut length and the function of Lark, prostaglandins, or GPR119 have not been previously appreciated. It will be informative to test if these genes impact gut length by controlling proliferation in the intestinal epithelium.

We identified significant sex-specific QTL associated with hindgut length on linkage group four and six (Fig. 9A, Table 6). The genotypes at the peak marker positions of the sex-specific QTL account for 9% and 6% of hindgut length variance, respectively. The homozygous cave genotypes are associated with the shortest hindgut length in females and longest hindgut length in males (Fig. 9C, E). The peak markers for the sex-dependent QTL on linkage group 4 and 6 map to scaffold KB871744.1 and KB871649.1, respectively. We examined all of the genes on each scaffold and found that seven exhibit evidence of

selection via HapFLK; two from linkage group four and five from linkage group six (Table 7). One of the genes on linkage group four, *lsm14b* (ENSAMXG00000002865), has fixed differences in the coding region comparing Pachón cavefish to both Río Choy and Rascón surface fish and has the highest Pachón/Río Choy divergence rank among the sex-specific QTL genes that exhibit non-neutral evolution (Tables 7, 8). LSM14B is an RNA binding protein that likely regulates translation. In humans, *Lsm14b* is most highly expressed in the testis, brain, and ovary [31]. In zebrafish, it is highly expressed in the ovary [32]. A role for LSM14B in modulating gut length has not been investigated. However, a link between sex and gut length has been uncovered in the fruit fly, *Drosophila melanogaster*, as gut length is influenced by the sexual identity of intestinal stem cells [33]. We discovered three predicted amino acid substitutions in Pachón *Lsm14b*

**Table 8** Hindgut QTL candidate genes

Gene ID	Gene name	Divergence from Río Choy			Divergence from Rascón		
		Pachón	Tinaja	Molino	Pachón	Tinaja	Molino
ENSAMXG00000007067	<i>gpr119</i>	0.71	0.44	0.53	0.41	0.1	0.53
ENSAMXG00000007167	<i>cbf2</i>	0.81	0.93	0.65	0.79	0.93	0.65
ENSAMXG00000007196	<i>cbf1</i>	0.95	0.94	0.93	0.94	0.94	0.93
ENSAMXG00000007233	<i>rbm14a</i>	0.29	0.1	0.11	0.18	0.01	0.11
ENSAMXG00000007243	<i>rin1a</i>	0.47	0.3	0.18	0.22	0.12	0.18
ENSAMXG00000007297	<i>mark2a</i>	0.2	0.27	0.23	0.32	0.34	0.23
ENSAMXG00000007312	<i>zgc:110782</i>	0.82	0.38	0.32	0.79	0.29	0.32
ENSAMXG00000007379	<i>adssl</i>	0.86	0.85	0.11	0.47	0.5	0.11
ENSAMXG00000007382	<i>rce1b</i>	0.85	0.76	0.72	0.81	0.76	0.72
ENSAMXG00000007446	<i>tenm2</i>	0.47	0.47	0.56	0.11	0.13	0.56
ENSAMXG000000025170	<i>spry3</i>	0.13	0.3	0.49	0.5	0.39	0.49
ENSAMXG000000025173	<i>zgc:77262</i>	0.73	0.55	0.58	0.25	0.05	0.58
ENSAMXG00000002865	<i>lsm14b</i>	0.82	0.60	0.61	0.82	0.69	0.47
ENSAMXG000000011121	<i>fam65b</i>	0.65	0.73	0.56	0.67	0.72	0.62
ENSAMXG000000011858	<i>Txnla</i>	0.64	0.50	0.29	0.73	0.45	0.60
ENSAMXG000000008444	<i>adck5</i>	0.42	0.26	0.28	0.29	0.14	0.35
ENSAMXG000000008359	<i>rfx5</i>	0.24	0.63	0.29	0.82	0.60	0.73
ENSAMXG000000008485	<i>cpsf1</i>	0.22	0.20	0.16	0.23	0.12	0.21

Shown in the table are the proportion of genes in the genome falling below the specified gene for  $D_{xy}$  when comparing indicated surface and cave populations. Some gene IDs from Pachón cavefish genome Ensembl database version 89 may be deprecated in most current Ensembl database. See Additional file 4 for updated gene IDs

(G78C, A173T, L266S) comparing predicted cDNA sequences from Río Choy surface fish and Pachón cavefish ( $n = 10$  fish per population, Table 7, Additional file 4). LSM14B therefore represents a promising candidate for future studies aimed at understanding a link between sex and gut length in *A. mexicanus*.

Five of the genes on scaffold KB871649.1 exhibit evidence of non-neutral evolution via HapFLK and have fixed differences comparing the coding regions between Pachón cavefish and Río Choy and Rascón surface fish (Tables 7, 8). The gene *fam65b* (also known as *ripor2*) shows the second highest ranked divergence between Pachón cavefish and Río Choy surface fish. It encodes an inhibitor of the small GTPase RhoA. Protein levels of FAM65B are highest in the duodenum, compared to other human tissues, but a role in the gut has not been investigated, to our knowledge [34]. We discovered one predicted amino acid substitution in Pachón Fam65b (R664H) comparing cDNA sequences between Río Choy surface fish and Pachón cavefish ( $n = 10$  fish per population, Table 7, Additional file 4). Using the same sequencing data, we found that half of the surface fish samples (5/10) contained a seven base pair deletion that is predicted to impact the amino acid sequence, resulting in a frame shift (D468Efs).

Another candidate gene on linkage group 6, *rfx5*, shows high divergence rank in Pachón compared to Río Choy (Table 8). RFX5 binds to the X box of MHC-II promoters to regulate translation [35, 36]. It is expressed in the developing duodenum in humans and, in adults, is highly expressed in immunomodulatory tissues. Mutations in RFX5 in humans lead to severe immunodeficiency, but, interestingly, mutations in the similar protein RFX6 cause malformations of the gut [37]. We discovered a predicted single amino acid substitution in Pachón Rfx5 (G342R) comparing cDNA sequences from Río Choy surface fish and Pachón cavefish ( $n = 10$  fish per population, Table 7, Additional file 4). Investigating the functional impact of this coding change in Rfx5 and the coding changes we discovered in the other candidates associated with hindgut length will further our understanding of the genetic basis of gut length evolution in *A. mexicanus*. Since hindgut length is sex biased and linked to weight change on a nutrient limited diet, these studies will lead to insights into the connections between sex differentiation and metabolism.

## Discussion

Sex dramatically impacts phenotypes, yet, sex has not been considered as a covariate in *Astyanax* genetic mapping studies (see [8, 38–46]). Female *A. mexicanus* have

been reported to be larger than males [47], but no other feature has been noted to be impacted by sex (eye size, melanophore number, relative condition, weight loss, tooth number, peduncle depth, fin placement, anal fin rays, SO3 bone width, thoracic rib number, chemical sense, oxygen consumption [2], feeding posture [48], stress behavior [49], and basal sleep [50]). In this study, we observed a sex bias in standard length, weight, body condition, relative liver size, relative gut length, and relative gonad weight in surface/Pachón F2 hybrid *A. mexicanus*. We found that although females are larger than males, males tend to gain weight on a restricted diet and females lose weight. The physiological basis of this phenomenon requires further investigation, but we hypothesize that a combination of behavioral and hormonal differences between the sexes could be at play. In a group setting, smaller fish (which are mostly male) may experience aggression from larger fish (which are mostly female), obtain less food, and remain smaller. When moved to individual housing, smaller fish may obtain more food than they had been able to obtain in a group setting, and larger fish receive less food. Consistent with this hypothesis we found that regardless of sex, smaller fish gained more weight on the nutrient restricted diet compared to larger fish. However, based on a resident-intruder assay, there are no differences in aggression between male and female *A. mexicanus* [51].

Another possibility for the difference in growth between males and females on the restricted diet is that growth was influenced by environmental space; teleost fishes, including *A. mexicanus* [52], exhibit restricted growth when housed in small spaces possibly due to increased stress hormone levels [53]. In our study, hybrid fish were shifted from being housed as a group in a 37 L tank to being individually housed in 1.5 L tanks. It is possible that the larger fish (that were mostly female) experienced more stress in this space compared to smaller fish (that were mostly male) which restricted their growth. How the sex-bias in size arises initially when fish are in the same space is unclear but in the majority of fish species that have been examined, females grow larger than males [54]. Our results warrant a more thorough examination of the impact of sex on phenotypes and how sex influences behavior under different contexts in *A. mexicanus*.

We found that the relative size of the liver and relative length of the hindgut are greater in female *A. mexicanus*. In other species, characteristics of the liver and colon are sexually dimorphic. For example, in humans, the weight of the liver is greater in males and liver metabolism differs between sexes, resulting in sex-specific drug metabolism and disease susceptibility [55]. Human females have a longer colon than males and increased risk of

developing right-sided colon cancer [56–58]. Sex hormones are linked to variation in these traits in humans although the mechanisms of action are not entirely clear. The details of sex determination and differentiation in *A. mexicanus* have yet to be elucidated. Our observations nevertheless provide an opportunity to investigate the evolution of sexually dimorphic traits that relate to growth and metabolism using this species.

Surface fish are longer and weigh more than cavefish, but cavefish have better body condition and lose weight more slowly when fasted [2, 3, 17]. We identified overlapping QTL associated with length, weight, and percent weight change on a nutrient limited diet. We also identified a QTL associated with body condition. Surface alleles are associated with bigger fish and cave alleles are associated with weight increase on a nutrient limited diet and better body condition. These findings highlight the differences in growth strategies between morphotypes. Cavefish have adapted their metabolism to survive starvation [2, 3], while surface fish have adapted to grow larger, perhaps providing an advantage in hunting, competition, and avoiding predators. Previous QTL mapping studies have identified genomic regions associated with body condition and starvation resistance [47]. However, a limited number of genetic markers and unavailable genome annotations made identifying genes in the region difficult. In addition, sex was not considered a covariate in any previous QTL study in *A. mexicanus* to our knowledge.

The dense linkage map we present, combined with availability of the annotated Pachón cavefish and surface fish genomes and genomic data from wild populations allowed us to identify a short list of candidate genes associated with metabolism-related QTL. We mapped traits to the genome and used population genomic analysis to determine if genes within each region display evidence of selection. Although genetic and phenotypic variation must be present for the trait to be mapped, importantly, it does not necessarily mean that the trait is adaptive. It is possible that traits are neutral and therefore genetic changes underlying the trait would not be identified by searching for genes that show evidence of selection. In line with this, genes that display evidence of selection may not control the mapped trait although they happen to be within the QTL. Establishing a connection between the genes within the QTL, the traits, and cavefish evolution will require functional analysis. We identified a number of coding differences between surface fish and Pachón cavefish in genes within metabolic QTL that can be tested for their impact on protein function and role in cavefish trait evolution.

## Conclusions

*Astyanax mexicanus* cavefish have morphological and metabolic adaptations that allow them to thrive in a harsh environment. Utilizing QTL mapping, we identified genomic regions associated with cavefish metabolic traits. We discovered a QTL with strong effect on blood glucose level, in addition to the previously identified *insra* mutation, highlighting that multiple approaches can lead to insights on the genetic basis of cavefish evolution. We found that a number of traits, including fish size, gonad size, and organ size, are sex-biased in *A. mexicanus*. We identified sex-independent and -dependent QTL associated with these traits and genes within the QTL that exhibit evidence of non-neutral evolution suggesting they may be important for cavefish adaptation. Furthermore, we reveal predicted amino acid changes in the proteins the genes encode. Functionally testing the role of the identified changes will provide a better understanding of the unique survival and reproductive strategies that evolve in a nutrient limited environment.

## Methods

### Biological specimens

*Astyanax mexicanus* surface fish and Pachón cavefish used to generate F2 hybrids were derived from fish collected from the Río Choy river and Pachón cave in Mexico that have been bred in the laboratory of Clifford Tabin at Harvard Medical School for multiple generations. For population genomic analysis, surface fish were collected from Río Gallinas (Rascón) and Río Choy, and cavefish were collected from Pachón, Tinaja, and Molino Caves in Mexico. *Astyanax aenus* were collected in Guerrero, Mexico. Additional details are available as part of a previous study [13].

### F2 hybrid population and trait quantification

F1 surface/Pachón hybrids were generated from paired breeding of a male surface fish with a female Pachón cavefish (Fig. 1A). The F2 mapping population ( $n=219$ ) consisted of three clutches produced from breeding paired F1 surface/Pachón hybrid siblings. Fish were housed at densities of less than one adult per 2 L until they were 14 months old and then were moved to individual 1.5 L tanks and fed three pellets (approximately 6 mg) of New Life Spectrum TheraA+ small fish formula once per day for at least 4 months and as long as 6 months (Fig. 1B). Fasting blood glucose was measured using FreeStyle Lite blood glucose meter and test strips 24 h after the fish were fed. To measure arginine response, arginine (6.6  $\mu\text{M}$ /gram fish) was injected into the intraperitoneal cavity using a U-100 insulin needle and blood glucose was measured after 5 h. For non-lethal

blood glucose measurement, blood was collected from the caudal tail vein using a U-100 insulin needle. For final blood glucose measurement blood was collected from the caudal tail vein after the tail was removed. Fish were euthanized in 1400 ppm Tricane. Images were taken using a Cannon Powershot D12 digital camera (Fig. 1A, C). Measurements were made using ImageJ. Body condition factor was calculated as fish weight (g)  $\times$  100/fish length (cm)<sup>3</sup> [10].

### Muscle triglyceride quantification

Fillets of skeletal muscle were flash frozen in liquid nitrogen and stored in  $-80\text{ }^{\circ}\text{C}$ . Frozen fillets were ground to powder in liquid nitrogen with a mortar and pestle. The muscle powder (100 mg/mL) was resuspended with 5% NP-40/ddH<sub>2</sub>O and slowly heated to 100  $^{\circ}\text{C}$  for 5 min and cooled back to room temperature. The heating and cooling were repeated two times after which the samples were centrifuged at max speed using a microcentrifuge for 2 min. The supernatant was then extracted and diluted tenfold with ddH<sub>2</sub>O before calorimetrically measured using a Triglyceride Assay Kit (Abcam, ab65336).

### Statistical analysis

Statistical analysis and data visualization were performed using R version 3.6.1 [59]. To find the appropriate model for comparing male and female F2 hybrids, correlation analysis, and QTL mapping we first determined if each trait is normally distributed using a Shapiro–Wilks test. For phenotypes that were normally distributed in both sexes ( $p>0.05$ ), we tested the impact of sex on each phenotype using a two-sample t-test. For non-normally distributed phenotypes ( $p<0.05$ ), we used a Wilcoxon signed-rank test. We examined correlations between phenotypes with data subset by sex using Kendall rank test. Correlations were visualized with the ggcorrplot R package version 0.1.3.

### Quantitative trait loci analysis

R/qrtl version 1.46–2 [60] was used to calculate genome wide logarithm of the odds (LOD) scores for each trait using a linkage map and genotype-by-sequencing markers generated as part of a previous study [11]. A single-QTL model and Haley–Knott regression was used to assess statistical significance of LOD scores by calculating the 95th percentile of genome-wide maximum penalized LOD score using 1000 random permutations (scanone). For phenotypes that were significantly different between males and females we compared the results of genome wide scans using sex as an additive and interactive co-variate. Significant sex-by-QTL interactions were identified by comparing the genome wide LOD thresholds according to the methods of [12]. Confidence

intervals for identified QTL were estimated using the 95% Bayesian credible interval (bayesint) and 1.5 LOD support interval (lodint) expanded to the nearest genotyped markers. Marker positions on the cavefish genome were determined using the Pachón cavefish assembly (AstMex102, INSDC Assembly GCA\_000372685.1, Apr 2013).

### Population genomic metrics and analysis

We performed the following measures with GATK-processed data, including a core set of samples analyzed in previous studies [5, 13] which contained wild caught: Pachón, N=10 (9 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 [5, 13]. We used VCFtools v0.1.13 [61] to calculate  $\pi$ ,  $F_{ST}$  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 (DF) calculations. We used hapFLK v1.3 <https://forge-dga.jouy.inra.fr/projects/hapflk> [62] 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 only the coding region of the gene. Genes were included in Table 8 if (a) the comparison between Río Choy and Pachón resulted in a maximum  $d_{XY}=1$ , suggesting that there was at least one fixed difference between Río Choy and Pachón 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.

### Identification of coding changes

Predicted cDNA sequence alignments were obtained from a variant call format (VCF) file of wild-caught surface fish (n=9, [13]) and Pachón cavefish genomes (reference genome [63] and 9 wild-caught samples [13]) aligned to the surface fish reference genome [64]. VCF tools was used to produce a tab format which was formatted into FASTA with custom python scripts with two sequences per individual to accommodate diploids. Alignments were viewed using CLC sequence Viewer 7.7.1 to identify predicted nonsynonymous changes that are homozygous and different in at least four out of ten surface fish and Pachón cavefish samples (Table 7, Additional file 4). We annotated changes relative to the most

common sequence in the surface fish samples unless otherwise indicated in Table 7.

### Abbreviations

QTL: Quantitative trait loci; LOD: Logarithm of the odds; AIC: Akaike's Information Criterion.

### Supplementary Information

The online version contains supplementary material available at <https://doi.org/10.1186/s12862-021-01823-8>.

**Additional file 1.** Name and position on linkage group and Pachón reference genome of the QTL markers used in this study.

**Additional file 2.** Phenotype and genotype data for the F2 mapping population.

**Additional file 3:** Population genetic analysis of genes on the scaffold with the peak marker of QTLs identified in this study.

**Additional file 4:** Predicted changes to the amino acid sequence based on cDNA alignments of genes within metabolism-related QTL that display evidence of non-neutral evolution by HapFLK analysis, have differences between Río Choy surface fish and Pachón cavefish, but are not homozygous and different in all samples (n = 10 fish per population). Table indicates number of individuals from each population that are homozygous and heterozygous for the indicated change.

### Acknowledgements

We would like to acknowledge Brian Martineu and Megan Peavey for fish husbandry and Robert Peuß for assisting in fish dissections.

### Authors' contributions

MR, AA, and CT conceived and designed the study. MR, AA, FD, and JT collected and analyzed data. MR wrote the manuscript draft. SM performed population genetic analysis and edited the manuscript. AA, FD, JT and CT edited the manuscript. All authors read and approved the manuscript.

### Funding

This work was supported by Grants from the National Institutes of Health [HD089934, DK108495]. Funding for genomic work was supported by NIH (1R01GM127872-01 to SEM, A. Keene, and N. Rohner). The funding bodies played no role in the design of the study and collection, analysis, and interpretation of data and in writing the manuscript.

### Availability of data and materials

The linked genotype and phenotype data analyzed in this study is included within the article (Additional files 1, 2). Sequencing data analyzed in this study were submitted to the SRA. Project Accession Number: SRP046999, Bioproject: PRJNA260715.

### Declarations

#### Ethics approval and consent to participate

Not applicable.

#### Consent for publication

Not applicable.

#### Competing interests

The authors declare that they have no competing interests.

#### Author details

<sup>1</sup>Department of Biology, University of Nevada, Reno, Reno, NV 89557, USA.

<sup>2</sup>Department of Genetics, Blavatnik Institute, Harvard Medical School, Boston, MA 02115, USA. <sup>3</sup>Department of Stem Cell and Regenerative Biology, Harvard

University, Cambridge, MA 02138, USA. <sup>4</sup>Ecology, Evolution, and Behavior, University of Minnesota, Saint Paul, MN 55108, USA.

Received: 8 December 2020 Accepted: 12 May 2021

Published online: 21 May 2021

## References

- Elliott WR. Cave biodiversity and ecology of the Sierra de El Abra Region. In: Biology and evolution of the Mexican Cavefish. Elsevier; 2016. pp. 59–76.
- Hüppop K. Oxygen consumption of *Astyanax fasciatus* (Characidae, Pisces): a comparison of epigeal and hypogean populations. *Environ Biol Fishes*. 1986;17:299–308.
- 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*. 2015;112:9668–73.
- Xiong S, Krishnan J, Peuß R, Rohner N. Early adipogenesis contributes to excess fat accumulation in cave populations of *Astyanax mexicanus*. *Dev Biol*. 2018;441:297–304.
- Riddle MR, et al. Insulin resistance in cavefish as an adaptation to a nutrient-limited environment. *Nature*. 2018;555:647–51.
- Krishnan J, et al. Comparative transcriptome analysis of wild and lab populations of *Astyanax mexicanus* uncovers differential effects of environment and morphotype on gene expression. *J Exp Zool Part B Mol Dev Evol*. 2020;334:530–9.
- Marandel L, et al. Nutritional regulation of glucose metabolism-related genes in the emerging teleost model Mexican tetra surface fish: a first exploration. *R Soc Open Sci*. 2020;7:191853.
- Protas ME, et al. Genetic analysis of cavefish reveals molecular convergence in the evolution of albinism. *Nat Genet*. 2006;38:107–11.
- Ma L, et al. A hypomorphic cystathionine  $\beta$ -synthase gene contributes to cavefish eye loss by disrupting optic vasculature. *Nat Commun*. 2020;11:2772.
- Froese R. Cube law, condition factor and weight-length relationships: history, meta-analysis and recommendations. *J Appl Ichthyol*. 2006;22:241–53.
- Riddle MR, et al. Genetic architecture underlying changes in carotenoid accumulation during the evolution of the blind Mexican cavefish, *Astyanax mexicanus*. *J Exp Zool Part B Mol Dev Evol*. 2020;334:405–22.
- Broman KW. A brief tour of R/qtl. 2012.
- Herman A, et al. The role of gene flow in rapid and repeated evolution of cave-related traits in Mexican tetra *Astyanax mexicanus*. *Mol Ecol*. 2018;27:4397–416.
- Kang Z, Fan R. PPAR $\alpha$  and NCOR/SMRT corepressor network in liver metabolic regulation. *FASEB J*. 2020;34:8796–809.
- Otaify GA, et al. Gnathodiaphyseal dysplasia: severe atypical presentation with novel heterozygous mutation of the anoctamin gene (ANO5). *Bone*. 2018;107:161–71.
- Fitzgerald TW, et al. Large-scale discovery of novel genetic causes of developmental disorders. *Nature*. 2015;519:223–8.
- Wilkins H, Hüppop K. Sympatric speciation in cave fishes? *J Zool Syst Evol Res*. 2009;24:223–30.
- Miller JE, et al. Rare variants in the splicing regulatory elements of EXOC3L4 are associated with brain glucose metabolism in Alzheimer's disease. *BMC Med Genomics*. 2018;1:1:76.
- Riddle M, Martineau B, Peavey M, Tabin C. Raising the Mexican Tetra *Astyanax mexicanus* for analysis of post-larval phenotypes and whole-mount immunohistochemistry. *J Vis Exp*. 2018. <https://doi.org/10.3791/58972>.
- Hüppop K, Wilkens H. Bigger eggs in subterranean *Astyanax fasciatus* (Characidae, Pisces). *J Zool Syst Evol Res*. 2009;29:280–8.
- Yokoo M, et al. Role of the hyaluronan receptor CD44 during porcine oocyte maturation. *J Reprod Dev*. 2007;53:263–70.
- Zhou J, et al. CD44 expression predicts prognosis of ovarian cancer patients through promoting epithelial-mesenchymal transition (EMT) by regulating Snail, ZEB1, and Caveolin-1. *Front Oncol*. 2019;9:802.
- Lee N, Kim D, Kim W-U. Role of NFAT5 in the immune system and pathogenesis of autoimmune diseases. *Front Immunol*. 2019;10:270.
- Li W, et al. Hermansky-Pudlak syndrome type 7 (HPS-7) results from mutant dysbindin, a member of the biogenesis of lysosome-related organelles complex 1 (BLOC-1). *Nat Genet*. 2003;35:84–9.
- Ornelas-García P, Pajares S, Sosa-Jiménez VM, Rétaux S, Miranda-Gamboa RA. Microbiome differences between river-dwelling and cave-adapted populations of the fish *Astyanax mexicanus* (De Filippi, 1853). *PeerJ*. 2018;6:e5906.
- Bou Sleiman M, et al. Enteric infection induces Lark-mediated intron retention at the 5' end of Drosophila genes. *Genome Biol*. 2020;21:4.
- Wilson DE. Role of prostaglandins in gastroduodenal mucosal protection. *J Clin Gastroenterol*. 1991;13:S65–71.
- Chu Z-L, et al. A role for intestinal endocrine cell-expressed G protein-coupled receptor 119 in glycemic control by enhancing glucagon-like peptide-1 and glucose-dependent insulinotropic peptide release. *Endocrinology*. 2008;149:2038–47.
- Cox HM, et al. Peptide YY Is critical for acylethanolamine receptor Gpr119-induced activation of gastrointestinal mucosal responses. *Cell Metab*. 2010;11:532–42.
- Tough IR, et al. Bidirectional GPR119 agonism requires peptide YY and glucose for activity in mouse and human colon mucosa. *Endocrinology*. 2018;159:1704–17.
- Fagerberg L, et al. Analysis of the human tissue-specific expression by genome-wide integration of transcriptomics and antibody-based proteomics. *Mol Cell Proteomics*. 2014;13:397–406.
- Zhao C-L, et al. Identification of zRAP55, a gene preponderantly expressed in Stages I and II oocytes of zebrafish. *Dong wu xue yan jiu Zool Res*. 2010;31:469–75.
- Hudry B, Khadayate S, Miguel-Aliaga I. The sexual identity of adult intestinal stem cells controls organ size and plasticity. *Nature*. 2016;530:344–8.
- Uhlen M, et al. Tissue-based map of the human proteome. *Science* (80—). 2015;347:1260419–1260419.
- Reith W, et al. Congenital immunodeficiency with a regulatory defect in MHC class II gene expression lacks a specific HLA-DR promoter binding protein, RF-X. *Cell*. 1988;53:897–906.
- Reith W, et al. Cloning of the major histocompatibility complex class II promoter binding protein affected in a hereditary defect in class II gene regulation. *Proc Natl Acad Sci*. 1989;86:4200–4.
- Smith SB, et al. Rfx6 directs islet formation and insulin production in mice and humans. *Nature*. 2010;463:775–80.
- Yoshizawa M, Gorički Š, Soares D, Jeffery WR. Evolution of a behavioral shift mediated by superficial neuromasts helps cavefish find food in darkness. *Curr Biol*. 2010;20:1631–6.
- Yoshizawa M, Yamamoto Y, O'Quin KE, Jeffery WR. Evolution of an adaptive behavior and its sensory receptors promotes eye regression in blind cavefish. *BMC Biol*. 2012;10:108.
- Yoshizawa M, et al. Distinct genetic architecture underlies the emergence of sleep loss and prey-seeking behavior in the Mexican cavefish. *BMC Biol*. 2015;13:15.
- Protas M, Conrad M, Gross JB, Tabin C, Borowsky R. Regressive evolution in the Mexican Cave Tetra *Astyanax mexicanus*. *Curr Biol*. 2007;17:452–4.
- Carlson BM, Klingler IB, Meyer BJ, Gross JB. Genetic analysis reveals candidate genes for activity QTL in the blind Mexican tetra *Astyanax mexicanus*. *PeerJ*. 2018;6:e5189.
- 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:e1000326.
- Moran D, Softley R, Warrant EJ. Eyeless Mexican cavefish save energy by eliminating the circadian rhythm in metabolism. *PLoS ONE*. 2014;9:e107877.
- Frøland Steindal IA, Beale AD, Yamamoto Y, Whitmore D. Development of the *Astyanax mexicanus* circadian clock and non-visual light responses. *Dev Biol*. 2018;441:345–54.
- Moran D, Softley R, Warrant EJ. The energetic cost of vision and the evolution of eyeless Mexican cavefish. *Sci Adv*. 2015;1:e1500363.
- Protas M, et al. Multi-trait evolution in a cave fish, *Astyanax mexicanus*. *Evol Dev*. 2008;10:196–209.
- Kowalko JE, et al. Convergence in feeding posture occurs through different genetic loci in independently evolved cave populations of *Astyanax mexicanus*. *Proc Natl Acad Sci*. 2013;110:16933–8.
- Chin JSR, et al. Convergence on reduced stress behavior in the Mexican blind cavefish. *Dev Biol*. 2018;441:319–27.

50. Jaggard J, et al. The lateral line confers evolutionarily derived sleep loss in the Mexican cavefish. *J Exp Biol*. 2017;220:284–93.
51. Elipot Y, Hinaux H, Callebert J, Rétaux S. Evolutionary shift from fighting to foraging in blind cavefish through changes in the serotonin network. *Curr Biol*. 2013;23:1–10.
52. Gallo ND, Jeffery WR. Evolution of space dependent growth in the teleost *Astyanax mexicanus*. *PLoS ONE*. 2012;7:e41443.
53. Costas B, Aragão C, Mancera JM, Dinis MT, Conceição LEC. High stocking density induces crowding stress and affects amino acid metabolism in Senegalese sole *Solea senegalensis* (Kaup 1858) juveniles. *Aquac Res*. 2007;39:1–9.
54. Pauly D. Female fish grow bigger—let's deal with it. *Trends Ecol Evol*. 2019;34:181–2.
55. Kur P, Kolasa-Wolosiuk A, Misiakiewicz-Has K, Wiszniewska B. Sex hormone-dependent physiology and diseases of liver. *Int J Environ Res Public Health*. 2020;17:2620.
56. Saunders BP, et al. Why is colonoscopy more difficult in women? *Gastrointest Endosc*. 1996;43:124–6.
57. Sadahiro S, Ohmura T, Yamada Y, Saito T, Taki Y. Analysis of length and surface area of each segment of the large intestine according to age, sex and physique. *Surg Radiol Anat*. 1992;14:251–7.
58. Rustgi AK. The genetics of hereditary colon cancer. *Genes Dev*. 2007;21:2525–38.
59. R-Core-Team. R: a language and environment for statistical computing. 2019.
60. Broman KW, Wu H, Sen S, Churchill GA. R/qtl: QTL mapping in experimental crosses. *Bioinformatics*. 2003;19:889–90.
61. Danecek P, et al. The variant call format and VCFtools. *Bioinformatics*. 2011;27:2156–8.
62. Fariello MI, Boitard S, Naya H, SanCristobal M, Servin B. Detecting signatures of selection through haplotype differentiation among hierarchically structured populations. *Genetics*. 2013;193:929–41.
63. McGaugh SE, et al. The cavefish genome reveals candidate genes for eye loss. *Nat Commun*. 2014;5:5307.
64. Warren WC, et al. A chromosome-level genome of *Astyanax mexicanus* surface fish for comparing population-specific genetic differences contributing to trait evolution. *Nat Commun*. 2021;12:1447.
65. Benjamini Y, Hochberg Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J R Stat Soc Ser B*. 1995;57:289–300.

### Publisher's Note

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

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

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

At BMC, research is always in progress.

Learn more [biomedcentral.com/submissions](https://biomedcentral.com/submissions)

