

The Population Genomics of Repeated Evolution in the Blind Cavefish *Astyanax mexicanus*

Martina Bradic,^{1,2} Henrique Teotónio,² and Richard L. Borowsky^{*1}

¹Biology Department, New York University

²Instituto Gulbenkian de Ciência, Apartado, Oeiras, Portugal

*Corresponding author: E-mail: rborowsky@nyu.edu; rb4@nyu.edu.

Associate editor: Michael Nachman

Abstract

Distinct populations of *Astyanax mexicanus* cavefish offer striking examples of repeatable convergence or parallelism in their independent evolutions from surface to cave phenotypes. However, the extent to which the repeatability of evolution occurred at the genetic level remains poorly understood. To address this, we first characterized the genetic diversity of 518 single-nucleotide polymorphisms (SNPs), obtained through RAD tag sequencing and distributed throughout the genome, in seven cave and three groups of surface populations. The cave populations represented two distinct lineages (old and new). Thirty-one SNPs were significantly differentiated between surface and old cave populations, two SNPs were differentiated between surface and new cave populations, and 44 SNPs were significantly differentiated in both old and new cave populations. In addition, we determined whether these SNPs map to the same locations of previously described quantitative trait loci (QTL) between surface and cave populations. A total of 25 differentiated SNPs co-map with several QTL, such as one containing a fibroblast growth factor gene (*Fgf8*) involved in eye development and lens size. Further, the identity of many SNPs that co-mapped with QTL was the same in independently derived cave populations. These conclusions were further confirmed by haplotype analyses of SNPs within QTL regions. Our findings indicate that the repeatability of evolution at the genetic level is substantial, suggesting that ancestral standing genetic variation significantly contributed to the population genetic variability used in adaptation to the cave environment.

Key words: SNP, QTL, RAD tag sequencing, convergence, parallelism, standing variation.

Introduction

Astyanax mexicanus cavefish are widely distributed in Northeast Mexico (Mitchell et al. 1977), with many of the cave populations representing independent examples of adaptation to an environment defined by perpetual darkness and limited food. At the morphological, physiological, and behavioral levels evolution from surface ancestors in the cave environment has led to parallel or convergent responses in cave populations, such as loss of vision and pigmentation (Wilkens 1971; Protas et al. 2007, 2008; Gross et al. 2008; Yoshizawa et al. 2010; Duboué et al. 2011). Other striking examples of evolutionary replication are found in freshwater populations of sticklebacks and in dwarf whitefishes (Bell and Foster 1994; Bell and Orti 1994; Pigeon et al. 1997; Peichel et al. 2001; Colosimo et al. 2004, 2005; Cresko et al. 2004; Rogers and Bernatchez 2005, 2007; St-Cyr et al. 2008; Barrett et al. 2009; Schluter and Conte 2009; Hohenlohe, Bassham, et al. 2010; Renaut et al. 2011). Some of these studies demonstrated parallel or convergent evolution at the genetic as well as the phenotypic level (St-Cyr et al. 2008; Hohenlohe, Bassham, et al. 2010; Renaut et al. 2011; Jones et al. 2012; Rogers et al. 2012). However, the studied populations of sticklebacks and whitefish are phylogenetically young, dating to the end of the last period of glaciation (10–12 kya), whereas the cave *Astyanax* populations are believed to be much older and believed to have evolved over the course of hundreds of

thousands of years, or more (Gross 2012). Thus, *A. mexicanus* cavefish are uniquely poised to address how parallel and convergent events at both the phenotypic and genetic levels depend upon time (Stern and Orgogozo 2009; Nadeau and Jiggins 2010).

The past decade has seen great progress in clarifying the phylogeographic relationships among *Astyanax* cave and surface populations of Northeast Mexico, knowledge key to an understanding of the forces driving cave adaptation in this group. The current view, based on mitochondrial DNA and microsatellite (mSat) variation, is that the cave populations were drawn from at least two distinct lineages (Dowling et al. 2002; Strecker et al. 2003, 2004, 2012; Hausdorf et al. 2011) and that individual cave populations have evolved independently at least five times (Bradic et al. 2012; Gross 2012). These studies confirmed earlier conclusions of multiple origins for the cave populations based on hybridization and complementation studies (Wilkens 1971; Strecker et al. 2003).

Phylogeographic studies in *Astyanax* using a variety of molecular markers have shown that extant *A. mexicanus* cavefish were derived from two distinct lineages that diverged approximately 6.7 Ma (Ornelas-Garcia et al. 2008). Surface fish of an older stock appear to have given rise to most of the present day cave populations, but then went locally extinct and were subsequently replaced by the current surface stock (fig. 1A and B; SN1 and SN2). We term the cave populations derived

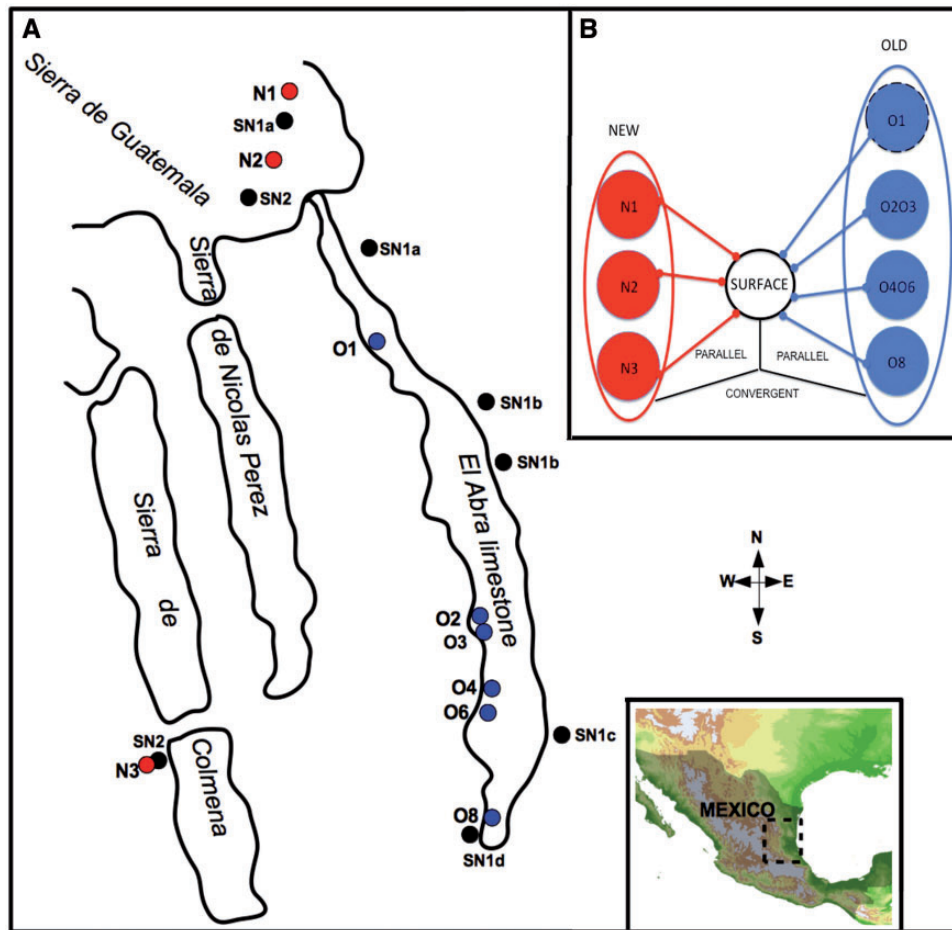


Fig. 1. Geographical distribution of the studied populations from three distinct geographical regions in NE Mexico and a schematic representation of the comparisons between the populations. (A) Collecting sites from which we obtained cave and surface specimens of *Astyanax mexicanus*. Circles represent sampled populations and are labeled as follows: El Abra region: Cueva Pachón (O1), Cuevas Yerbaniz and Japonés (O2O3), Cuevas Arroyo and Curva (O4O6), Cueva Chica (O8*) (blue circles); Guatemala region: Cueva Molino (N1), Cueva Caballo Moro (N2) (red circles); Micos region: Cueva Río Subterráneo (N3*) (red circle); Surface localities: El Abra and Guatemala surface localities (SN1) Rio Subterráneo Valley surface locality (SN2). The populations groups, as defined here, represent independent instances of cave adaptation and the population designations follow prior usage (Bradic et al. 2012). (B) Population comparisons made in this study. We compared individual populations from either the new (red line) or old (blue line) cave lineages with surface populations (SN1 and SN2). Higher level comparisons were performed based on commonalities determined in each of two cave lineages in comparison to surface, representing instances of parallel evolution. We also compared loci identified in new cave and old cave groups to determine instances of convergent evolution (solid black line). The dotted black line designates the old cave population (O1) that was used in the F_2 cross from which SNP markers were derived.

from the old lineage as old cave populations (fig. 1A and B; O1, O2O3, O4O6, and O8) and the ones derived from the current surface lineage as new cave populations (fig. 1A and B; N1, N2, and N3). Despite habitat differences, physical barriers and taxonomic separation, significant levels of migration still occur between cave and surface populations. Nevertheless, cave populations maintain a cave-specific phenotypic syndrome, which implies selection for its maintenance in the face of gene flow (Bradic et al. 2012).

The impact of natural selection on cave evolved regressive traits is still under debate. One hypothesis is that given enough time and a sufficiently high mutation rate, random mutations in eye-forming genes accumulate in cave animals under relaxed selective pressure (Culver and Fong 1986; Wilkens 1988, 2010a, 2010b). Thus, eyes would eventually disappear because they are not necessary for survival in the

dark environment. An alternative hypothesis suggests that eyes are lost in cave animals because they are detrimental, perhaps because of the metabolic cost of maintaining them (Poulson 1964). The role of natural selection on eye regression is supported by earlier quantitative trait loci (QTL) studies (Protas et al. 2007; Borowsky 2010), but the power of those conclusions is obviously constrained by the power of QTL analysis. One promising approach to test the degree to which natural selection is driving the evolution of cave phenotypes is the genome-wide survey of natural variation, especially when multiple independently evolved cave populations as well as related surface populations are available for study.

To address the impact of natural selection on the cavefish genome, we developed new genomic markers for *A. mexicanus* using restriction site associated DNA (RAD)

tag sequencing technology (Baird et al. 2008; Hohenlohe, Bassham, et al. 2010). We detected multiple single-nucleotide polymorphisms (SNPs) and developed a genome-wide panel of 518 SNPs to estimate differentiation between the following: 1) surface and new cave populations; 2) surface and old cave populations; and 3) new and old cave populations (fig. 1B). SNPs significantly differentiated in frequencies in these comparisons are candidate markers for regions having been subject to natural selection during the evolution of the cave populations. Significantly differentiated SNPs that are shared among cave populations of different lineages (new and old) are candidates for molecular convergence while those shared among cave populations of the same lineage are candidates for parallel evolution. This approach has been proven to be a powerful strategy for detecting hallmarks of selection in several organisms, including the aforementioned sticklebacks and whitefish (Peichel et al. 2001; Cresko et al. 2004; Colosimo et al. 2005; Rogers and Bernatchez 2005; Protas et al. 2007; Baird et al. 2008; Hohenlohe, Bassham, et al. 2010; Duboué et al. 2011).

We expanded a previous genetic map from a cross between a new surface and an old cave population. For some phenotypes, this significantly improved our estimates of QTL parameters, thus facilitating our ability to combine the approaches of population genetics and QTL analysis.

In this study, we identify numerous SNPs bearing hallmarks suggestive of directional selection in multiple cave populations. Some of these SNPs exhibit parallel changes within lineages; others exhibit convergent changes between lineages. Importantly, this shows that many alleles in the cave populations were derived from standing variation in the ancestral surface populations. In some, but not all, cases, differentiated SNPs can be aligned with QTL for cave-related traits.

Results

SNP Diversity in Natural Populations

After quality control of the samples from natural populations, we retained 272 out of 281 individuals and 518 out of 745 SNPs markers (see Materials and Methods). Table 1 details

sample locations, sizes, and quality control performed. There were no systematic differences in genotyping success between markers derived from Sanger and RAD tag sequencing (data not shown). We observed the highest numbers of polymorphic SNPs in surface (SN1 and SN2) and in the cave population samples (N3* and O8*) (table 1; supplementary fig. S1, Supplementary Material online). The old cave population (O1) used for RAD Tag SNP discovery had similar levels of polymorphism as populations O2O3 and O8, suggesting that O1 is representative of old cave populations. The great majority of SNPs were in Hardy–Weinberg Equilibrium (HWE) proportions, when samples from all populations are considered together (table 1). Those that were not came primarily from two populations previously identified as admixtures of cave and surface populations (N3* = 27 and O8* = 32) (Bradic et al. 2012).

To test whether polymorphisms are largely the result of common or rare SNPs, we eliminated all SNPs with minor allele frequency (MAF < 5%). Most polymorphisms were due to rare alleles (supplementary fig. S2, Supplementary Material online) and the number of polymorphisms was reduced in all of the populations (table 1). The largest reductions were observed in both new and old cave populations (N1, N2, and O4O6). Based on the MAF distributions in surface populations, two classes of SNPs were recognized. Those with MAF more than 5% in the surface populations were termed HMAF SNPs, whereas those with MAF less than 5% were termed LOMAF SNPs. Of the 518 SNPs, 210 were LOMAF SNPs, and 308 were HMAF SNPs. These differences in polymorphic SNP content are primarily the consequence of ascertainment bias caused because SNP discovery was based on an F_1 cross in which a single old cave individual (O1) was mated with a surface individual (supplementary fig. S2, Supplementary Material online).

The mean observed heterozygosity (H_o) averaged over LOMAF and HMAF markers ranged from 0.05 ± 0.11 (N1) to 0.19 ± 0.11 (N3*) and the mean unbiased expected heterozygosity (Nei's H_e) varied from 0.05 ± 0.09 (N1) to 0.19 ± 0.12 (N3*) (supplementary table S1, Supplementary Material online). Mean allelic richness (AR) varied from 1.1 ± 0.08

Table 1. Details on Sampled Populations and Markers.

Population	Code	Origin	<i>N</i>	<i>a</i>	<i>b</i>	Not in HW
El Abra and Guatemala (surface, SN1a, SN1b, SN1c, and SN1d)	SN1	Surface	25	326	251	26
Rio Subterráneo valley (surface)	SN2	Surface	45	381	289	18
Cueva Molino	N1	New cave	22	131	36	4
Cueva Caballo Moro	N2	New cave	26	270	148	12
Cueva Río Subterráneo	N3*	New cave	25	341	236	27
Cueva Pachón	O1	Old cave	31	345	179	9
Cuevas Yerbaniz and Japonés	O2O3	Old cave	22	375	286	8
Cuevas Arroyo and Curva	O4O6	Old cave	29	277	139	21
Cueva Chica	O8*	Old cave	32	396	319	32

NOTE.—Summary of the population locations and abbreviations, their origins, and sample sizes (*N*). Column *a* lists the number of loci, out of the original 518, that remained after excluding monomorphic loci in the population. Column *b* lists the numbers of loci remaining after further eliminating those with MAF < 5% (see Materials and Methods for details). The numbers of SNPs per population that do not follow HW equilibrium (not in HW) are shown in last column. *a* is the number of polymorphic marker in each population, *b* is the number of polymorphic markers in each population after removing MAF < 5%.

*Admixed populations previously described.

(N1) to 1.47 ± 0.41 (O8*). As expected, the percentage of polymorphic loci (%P) was positively correlated with the observed heterozygosity (Spearman's $\rho = 0.737$, $P < 0.0001$). The average percentage of polymorphic loci varied greatly among populations, ranging from $13.45 \pm 0.015\%$ (O4O6) to $49.66 \pm 0.47\%$ (N3*). Overall, we detected the greatest diversities in the surface and admixed cave populations (N3* and O8*). Inbreeding coefficients across populations (F_{IS}) ranged from -0.02 ± 0.04 in the O1 population to 0.31 ± 0.33 in the new cave populations (N2 and N3*), suggesting higher inbreeding in admixed cave populations.

When Nei's H_e was averaged over all SNPs, there was no significant difference between new and old cave populations. Nei's H_e for new caves was 0.127 ± 0.071 and for old caves it was 0.113 ± 0.038 ($t_5 = 0.35$, $P = ns$; data in [supplementary table S1, Supplementary Material online](#)). When HIMAF and LOMAF SNPs were distinguished; however, it was clear that trends in heterozygosities differed between the two cave groups. For LOMAF, SNPs average H_e was 0.037 ± 0.015 for new caves and 0.138 ± 0.046 for old caves ($t_5 = 3.54$, $P = 0.017$). For HIMAF SNPs, the trend was in the opposite direction with average H_e being 0.217 ± 0.127 for new caves and 0.085 ± 0.047 for old caves ($t_5 = 1.96$, $P = 0.11$). We tested the significance of the difference in trends with a two-way ANOVA with dummy variables representing the lineage and SNP set differences. The interaction term was highly significant ($F_{1,10} = 10.18$, $P = 0.010$). This suggests that the HIMAF SNPs (from SN1 and SN2 populations) represents the ancestral polymorphisms in new cave lineages and are shared to some extent with the old cave lineage. In contrast, LOMAF SNPs are specific for the old cave populations and to a much lesser extent in the new cave populations. Trends in heterozygosity were similar in SNPs discovered by Sanger sequencing as well as those revealed by RAD sequencing ([fig. 2](#)).

QTL Mapping in a Surface-Cave Cross

We mapped several phenotypes in an F_2 cross between a new lineage surface fish and an old lineage cavefish (O1; Pachón). After quality control (see methods for details), the data set contained 474 SNPs for 273 F_2 individuals. The integrated genetic map using both SNP and microsatellite markers had 40 linkage groups defined by 450 markers and a total genetic length of 2,646 cM. The map's average resolution was 6.5 cM. Forty-nine QTL were identified on the new map ([supplementary fig. S4, Supplementary Material online](#)). This cross had previously been mapped with microsatellite markers (Protas et al. 2006, 2008), but the availability of the new SNP markers allowed us to generate a more densely populated map and to estimate QTL parameters more precisely. More importantly, it allowed us to associate patterns of SNP variation among populations with the phenotypes influenced by QTL variation throughout the genome.

The number of QTL per trait ranged from one for estimated daily growth rate (G) to eight for relative eye size (E) ([table 2](#)). The positions of the QTL and their one logarithm of odds (LOD) score support intervals are shown in [supplementary figure S3 \(Supplementary Material online\)](#), and the best estimates of QTL parameters are listed in [table 3](#). The LOD profiles were particularly narrow in LG3 and LG14 for lens and eye size QTL. These linkage groups correspond to the original LG14 and LG20, respectively (Protas et al. 2008). The QTL for eye size in our LG14 was the strongest we detected (LOD = 60.2, PEV = $49.2 \pm 2.5\%$). Previously, this QTL had been detected with an LOD score of 30 and explained only 18% of the trait variance (12.5 standard errors of the mean [SEM] below the new value of 49.2% ; $P < 10^{-5}$). The steep increase is centered on SNP marker $m315$ and the effect strongly decreases toward the NYU14 marker, which was previously the center of the QTL (Protas et al. 2008). Another example of the improvement of QTL strength is a

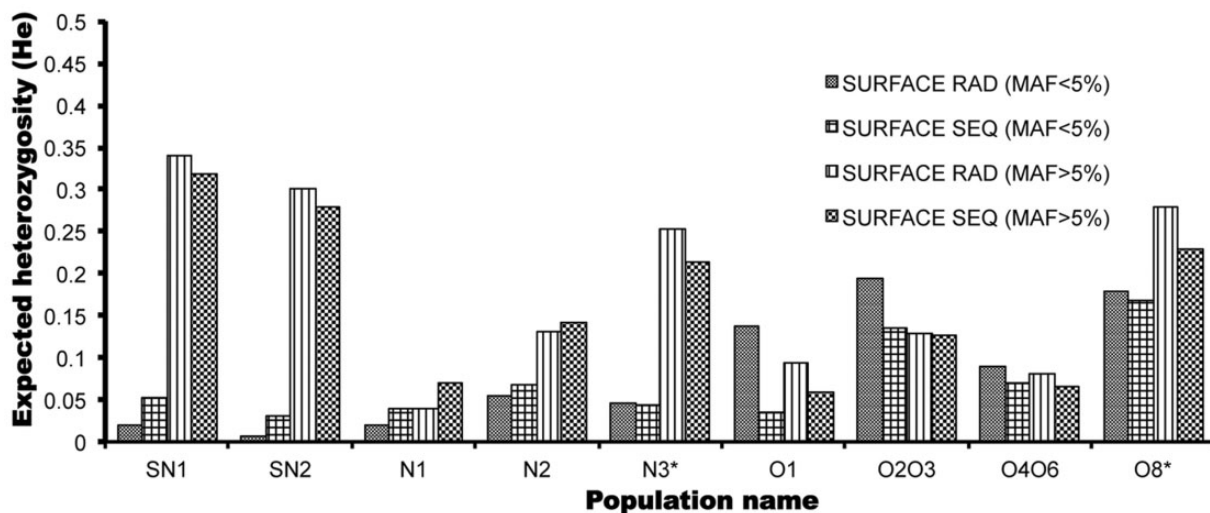


Fig. 2. Trends in heterozygosity for different marker panels. Different histogram patterns represent different marker groups. Markers are divided in four groups; two groups based on the sequencing methodology used (RAD and Sanger sequencing), which are further divided into two more groups based on MAF in surface populations (MAF > 5% or MAF < 5%), as described in the legend. Surface populations: SN1 and SN2; New caves: N1, N2, and N3*; Old caves: O1, O2O3, O4O6, and O8*.

Table 2. Summary of the Measured Phenotypes and Identified QTLs in Surface × Old Cave (O1) Cross.

Trait	Symbol	Description of the Trait
Eye size	E	Observed eye size divided by eye size predicted from the regression of eye size on standard length (SL). Skew corrected in MultiQTL. Cave < surface.
Pigmentation	Me	Count of melanophores in an area ($2.0 \times 0.4 \text{ mm}^2$) located 3.0 mm above the left eye. See supplementary methods in Protas et al. (2007) for illustration. Cave < surface.
Pigmentation	Md	Count of melanophores in an area ($2.0 \times 0.4 \text{ mm}^2$) located 2.0 mm below the insertion of the dorsal fin. Cave < surface.
Pigmentation	Ma	Count of melanophores in an area ($2.0 \times 0.4 \text{ mm}^2$) located on the flanks 4.0 mm dorsal to the insertion of the anal fin. Cave < surface.
Lens size	Le or Ll	Observed lens size standardized by eye size (Le) or by standard length (Ll). Cave < surface
Relative condition	C	Observed weight divided by predicted weight calculated from regression of log weight on log length. Surface < cave
Weight loss	W	Rate of weight loss on fast expressed as percent decrease per day. Weight loss is slower in cave than in surface individuals.
Growth	G	Estimated daily growth rate base on a 4-mm length at hatch and standard length at adulthood. Because four different broods were raised, the metric used was the residual after correction for group differences.
Length	L	Residual after multiple regression of standard length on age and dummy variables to correct for differences among holding tanks. Cave < surface.
Dentary	D	Length of the dentary standardized by standard length.
Chemical sense	A	Threshold sensitivity to dissolved amino acids in the water assessed by searching response triggered by the addition of amino acid solution to the test aquarium. Responses were scored over the full range of concentrations and used to estimate the concentration at which response probability was 50%. This was scored as an integer (6–11) representing the concentration (10^{-6} to 10^{-11} M). Pachón cave > sensitivity than surface.

NOTE.—Description of the measured phenotypes, abbreviations (symbol), and description of their calculation in the F_2 generation as previously reported (Protas et al. 2008).

lens size (Ll) QTL in LG3. This QTL now accounts for $13.4 \pm 3.0\%$ of the variance, exceeding the earlier estimate of 4.2% by 3.07 SEM ($P = 0.0021$). This improvement is clearly the result of multiple SNP markers that were placed between the 16C and 133B microsatellite markers on our LG3 and narrowed down the QTL to the strong peak of LOD = 8, centered at 115.0 cM (95% confidence interval = 108.5–120.5). This new map allowed us to investigate whether outlier SNPs detected in the population genetic survey correspond in location to QTL.

As explained later, we identified outlier SNPs as candidates for having been subjected to directional selection through our population analyses and defined a subset of these that were shared among multiple cave populations. To test whether shared outlier SNPs tend to be associated with QTL, we considered only SNPs that were used to build the linkage map. Furthermore, we minimized the redundancies of multiple SNPs within bacterial artificial chromosomes (BACs) or sequenced genes by counting those features as single nonsignificant SNPs if they contained no outlier SNPs, or as one nonsignificant plus one outlier SNP, if they did. Of the 66 outlier SNPs so defined, four were located exactly on a QTL peak, a total of nine mapped within 1.0 cM, and a total of 12 mapped within 2.0 cM. The total map space spanned by ± 2 cM flanking all of the QTL is 187.7 cM, which is 7.1% of the total map length of 2,646 cM. The expected number of outlier SNPs within 2 cM of a QTL is 4.65, which is significantly lower than the 12 observed (Yates corrected $\chi^2 = 10.8$, df = 1, $P \ll 0.05$). In a second test of association, mapped outlier SNPs were found to coincide in position with QTL (i.e., within 1 cM) significantly more often than mapped nonoutlier SNPs

(10 of 68 vs. 14 of 334; $P = 0.0027$, Fisher's exact test), thus further supporting the hypothesis that significant SNPs are markers for sites in the genome relevant to troglomorphic evolution. Thus, the positions of outlier SNPs and QTL are significantly coincident. Testing for a broader association, however, did not reveal significance (29% of 66 mapped outlier SNPs were within QTL boundaries compared with 23% of the 257 nonsignificant SNPs, $P = \text{ns}$). We conclude that those cases in which the QTL/SNPs are closely paired arise because the two methods are signaling the same evolutionary event; but that the lack of broader scale significant association suggests that the QTL and SNP data sets give mostly complementary information about genomic regions where evolutionarily significant events occurred.

Genetic Differentiation

We used an analysis based on hierarchical island models of neutral divergence to detect loci significantly differentiated among population samples (Excoffier et al. 2009). The hierarchical portion of our analysis accounted for structure only in the surface populations; comparisons with cave populations were pairwise. Analyses were performed separately for each cave, because each cave represents an independent evolutionary origin. Differentiated SNP markers were termed outliers. The following comparisons were made: 1) old cave populations versus surface populations, followed by comparison of common outliers within the old cave lineage; 2) new cave populations versus surface populations, followed by comparison of common outliers within the new cave lineage; and 3) new cave versus old cave lineages (fig. 1B). F_{ST} values plotted as functions of expected

Table 3. Summary of Identified QTL with Their Trait and Linkage Groups.

Linkage	Trait	Location	Range	LOD	P	PEV	PEVad	Substitution Effect	Heterozygote Effect
LG1	E	88.8	83.0–92.1	5.35	0.001	0.032 ± 0.010	0.028 ± 0.0097	−0.098 ± 0.019	−0.020 ± 0.018
LG2	Ma	63	52.9–80.3	3.51	0.007	0.175 ± 0.093	0.149 ± 0.074	−22.0 ± 6.9	−3.779 ± 5.968
LG3	LI	115	108.5–120.5	8.18	0.001	0.134 ± 0.030	0.058 ± 0.036	−0.261 ± 0.087	0.212 ± 0.064
LG3	W	199	196.0–205.0	4.38	0.003	0.089 ± 0.030	0.023 ± 0.023	0.043 ± 0.030	0.059 ± 0.024
LG4	A	85.3	72.7–92.2	4.83	0.001	0.101 ± 0.037	0.0099 ± 0.012	0.223 ± 0.409	−0.972 ± 0.241
LG4	L	0.5	0.0–10.6	3.41	0.02	0.043 ± 0.019	0.0049 ± 0.0092	−0.226 ± 0.933	−1.746 ± 0.75
LG6	D	47	37.0–57.0	4.17	0.001	0.069 ± 0.026	0.0087 ± 0.0096	−0.009 ± 0.007	0.021 ± 0.006
LG6	E	57.9	46.9–59	4.32	0.001	0.024 ± 0.0092	0.020 ± 0.0085	−0.081 ± 0.018	−0.018 ± 0.020
LG8	Ma	0.5	0.0–2.4	3.74	0.002	0.174 ± 0.071	0.041 ± 0.044	8.933 ± 8.142	−14.245 ± 4.773
LG10	L	13.3	9.1–17.8	4.24	0.001	0.085 ± 0.036	0.053 ± 0.029	−3.016 ± 1.065	−1.411 ± 1.083
LG11	A	0.5	0.0–7.4	5.08	0.001	0.110 ± 0.034	0.054 ± 0.033	1.022 ± 0.351	0.710 ± 0.341
LG12	Md	38.9	36.0–43.3	3.94	0.002	0.240 ± 0.108	0.161 ± 0.079	−12.78 ± 6.18	6.205 ± 3.231
LG12	E	21.5	17.0–25.7	5.42	0.001	.025 ± .0092	0.015 ± .0085	−0.063 ± 0.033	0.038 ± .021
LG13	Md	27.7	23.2–30.7	15.5	0.001	0.292 ± 0.047	0.288 ± 0.046	−18.22 ± 2.01	−0.701 ± 1.310
LG13	Me	30.8	26.9–35.2	15.1	0.001	0.348 ± 0.044	0.344 ± 0.045	−28.42 ± 2.75	0.220 ± 2.060
LG14	LI	0.5	0.0–4.8	8.64	0.001	0.172 ± 0.040	0.161 ± 0.043	−0.449 ± 0.072	−0.067 ± 0.047
LG14	E	34.8	33.2–36.7	60.19	0.001	0.492 ± 0.025	0.481 ± 0.027	−0.411 ± 0.020	−0.041 ± 0.013
LG14	L	18	0.0–38.8	2.99	0.005	0.053 ± 0.022	0.043 ± 0.023	−2.742 ± 0.826	0.431 ± 0.835
LG15	A	0.5	0–2.56	5.52	0.001	0.096 ± 0.030	0.0033 ± 0.0047	0.018 ± 0.266	0.985 ± 0.191
LG16	A	12.8	7.1–22.8	3.59	0.002	0.081 ± 0.026	0.0089 ± 0.010	−0.337 ± 0.283	0.868 ± 0.175
LG17	A	59.4	36.9–63.7	4.54	0.002	0.090 ± 0.031	0.080 ± 0.030	−1.289 ± 0.279	−0.242 ± 0.211
LG20	Le	5.3	0.0–9.0	3.54	0.002	0.071 ± 0.026	0.014 ± 0.014	−0.187 ± 0.115	−0.296 ± 0.068
LG20	C	63.3	51.1–75.0	5.14	0.005	0.068 ± 0.024	0.052 ± 0.022	−0.069 ± 0.016	−0.023 ± 0.015
LG20	L	14.8	4.6–27.5	4.2	0.009	0.084 ± 0.038	0.032 ± 0.023	−2.213 ± 1.109	2.138 ± 0.651
LG22	W	3.9	0.0–20.1	3.56	0.001	0.098 ± 0.026	0.051 ± 0.032	−0.075 ± 0.027	−0.048 ± 0.022
LG25	D	28.5	18.1–42.9	4.05	0.001	0.111 ± 0.043	0.0084 ± 0.012	0.006 ± 0.010	−0.028 ± 0.007
LG25	Me	54.2	49.4–59.0	4.14	0.001	0.099 ± 0.026	0.092 ± 0.028	−14.50 ± 2.97	0.542 ± 2.920
LG26	A	20.2	18.3–22.2	8.36	0.001	0.264 ± 0.061	0.119 ± 0.043	1.558 ± 0.355	−1.230 ± 0.203
LG26	Me	24.6	20.7–30.0	4.1	0.002	0.074 ± 0.031	0.044 ± 0.027	−9.637 ± 3.413	−5.553 ± 2.202
LG26	C	29.5	19.1–30.0	3.83	0.005	0.046 ± 0.017	0.040 ± 0.018	−0.060 ± 0.015	−0.001 ± 0.017
LG27	E	0.5	0.0–3.5	2.94	0.001	0.014 ± 0.0068	0.0016 ± 0.0023	0.009 ± 0.021	0.046 ± 0.013
LG28	C	2.8	0.0–9.8	5	0.005	0.069 ± 0.026	0.062 ± 0.025	0.076 ± 0.017	−0.012 ± 0.014
LG28	W	21.8	8.0–24.6	3.54	0.001	0.071 ± 0.034	0.052 ± 0.037	−0.075 ± 0.029	−0.025 ± 0.024
LG29	C	26	19.2–29.7	5.24	0.005	0.097 ± 0.032	0.024 ± 0.014	0.045 ± 0.015	−0.057 ± 0.013
LG29	E	2.8	0.0–8.3	10.29	0.001	0.066 ± 0.019	0.063 ± 0.018	−0.147 ± 0.023	0.013 ± 0.017
LG30	Md	74.8	67.9–82.5	4.89	0.001	0.105 ± 0.046	0.029 ± 0.024	5.219 ± 2.568	−6.242 ± 2.583
LG30	Me	83.4	69.6–89.9	3.97	0.001	0.084 ± 0.031	0.036 ± 0.027	8.542 ± 3.822	−6.797 ± 3.455
LG30	E	132.5	128.0–138.0	13.2	0.001	0.072 ± 0.017	0.047 ± 0.017	−0.125 ± 0.025	0.065 ± 0.016
LG31	Md	35.3	22.9–44.3	6.88	0.001	0.073 ± 0.019	0.052 ± 0.019	−7.651 ± 1.516	−3.130 ± 1.437
LG31	LI	36.9	30.1–44.2	3.85	0.002	0.079 ± 0.029	0.012 ± 0.013	−0.098 ± 0.079	0.203 ± 0.048
LG31	C	15.5	13.3–22.9	5.78	0.005	0.057 ± 0.017	0.047 ± 0.017	−0.066 ± 0.013	0.016 ± 0.014
LG33	C	9.6	4.0–16.8	4.64	0.005	0.056 ± 0.019	0.050 ± 0.019	0.068 ± 0.014	0.011 ± 0.013
LG33	W	2.8	0.0–3.9	2.91	0.003	0.066 ± 0.036	0.037 ± 0.029	−0.061 ± 0.030	0.039 ± 0.016
LG34	Md	15.3	7.4–25.1	2.95	0.006	0.040 ± 0.018	0.014 ± 0.013	−3.402 ± 2.218	3.684 ± 1.190
LG34	G	7.1	2.6–12.5	3.27	0.002	0.084 ± 0.032	0.051 ± 0.031	0.043 ± 0.016	0.019 ± 0.018
LG34	Ma	19.7	5.1–24.5	2.99	0.005	0.103 ± 0.041	0.050 ± 0.031	−12.02 ± 4.29	−8.545 ± 3.939
LG35	L	5.9	3.5–6.4	5.29	0.001	0.076 ± 0.027	0.064 ± 0.024	3.421 ± 0.717	0.841 ± 0.588
LG37	E	0.5	0.0–8.6	5.39	0.001	0.032 ± 0.011	0.030 ± 0.011	0.101 ± 0.020	0.002 ± 0.015
LG38	Me	22.5	17.9–24.8	3.88	0.001	0.058 ± 0.024	0.032 ± 0.019	−8.320 ± 2.727	−5.231 ± 2.077

NOTE.—Trait: refer table 2; location on LG (cM), Range (= one LOD score support interval), maximum LOD score (LOD) and significance (P) after MIM. PEV refers to total trait variance; PEVad refers to the proportion of additive variance explained by the QTL. Allelic substitution and heterozygote effects and their standard errors are listed in the last two columns.

heterozygosities in cave-surface comparisons are shown in [figure 3](#). Details for each SNP and population are given in the [supplementary table S2](#) ([Supplementary Material](#) online). In subsequent discussion, where appropriate, we distinguish between outlier SNPs that were identified in only one population in either or both lineages, and shared outlier SNPs that were observed in two or more cave populations from the same lineage. Although our criterion for shared outlier SNPs is two or more, the data in [supplementary table S2](#) ([Supplementary Material](#) online) show that on average they were shared among 3.66 populations, more than half of the seven studied.

Parallel Evolution within the Old Lineage

In comparisons between surface and old cave populations, the highest number of outlier SNPs was detected in the O1 population (135 total outliers), an expected consequence of ascertainment bias (see [supplementary material](#), [Supplementary Material](#) online). We also identified numerous outliers in two other old cave populations (108 and 104 in O8 and O2O3, respectively), but a much lower number in O4O6 (78) ([table 4](#)). For subsequent analysis, we only considered shared outlier SNPs ([fig. 4](#)). These were far more common among LOMAF SNPs, 70 out of 216, than HIMAF SNPs, 7 out of 303 ($\chi^2 = 90.4$, $df = 1$, $P < 0.0001$). This suggests any of three possibilities: 1) that these alleles were present in the old surface progenitors, 2) that they arose by new mutation in the old lineage, or 3) that they are present in extant surface populations but at such low frequency as to elude detection.

We also detected shared outlier SNPs in two QTL that contained candidate genes. The first is the *m682* SNP in the *Oca2* gene in populations O1, O2O3, O4O6, and O8*. The second example is a pair of SNP markers, *m595* and *m619*, from a BAC_gh1 containing the growth hormone candidate gene in populations O1, O2O3, and O8* ([table 5](#); [supplementary fig. S2](#), [Supplementary Material](#) online). We found no significantly differentiated SNPs in the QTL containing the other candidates, *Cryaa*, *Mc1r*, or *Prox1*. Note, however, that within surface-single cave comparison some of these SNPs were highly differentiated, even if not shared among multiple caves. Two such examples are *m689* in *Cryaa* and *m685* in *Oca2* in the O2O3 population ([table 5](#)). Another SNP, near the *Fgf8* gene, was exceptional in being a significant outlier in all the sampled cave populations (N1, N2, N3*, O1, O2O3, O4O6, and O8*; [table 5](#)).

To gain insights into the effects of linkage disequilibrium (LD), we reconstructed multi-SNP haplotypes within the QTL containing outlier SNPs and a comparison region with outlier SNPs, but no QTL. To further test the candidate regions for selection identified using F_{ST} statistics, we calculated the differences in the extended haplotype homozygosity decay (EHHS) between sets of populations (ρ_{Rsb} , see [Materials and Methods](#)) ([Tang et al. 2007](#); [Gautier and Naves 2011](#)).

We first counted the numbers of ρ_{Rsb} more than 1.3 ($P < 0.05$) within QTL on LG10, LG16, LG25, and LG33, and in the region on LG7 with outlier SNPs but no QTL. We then

considered a region to be significant if it contained at least one SNP exceeding this threshold. Using this approach, we found five significant QTL supported by eight significant values in the seven linkage groups tested ([fig. 5Ai](#)).

[Figure 5Bi](#) illustrates an example of EHHS within a candidate region in LG25 that is associated with a QTL for size of the dentary bone. Its core haplotype had EHHS values of approximately 0.6 over an interval region of 25 cM ([fig. 5Bi](#)); this suggests high homozygosity of this particular haplotype and that it was selected in old cave populations. This is likely to be an example of parallel molecular evolution in the old cave lineage.

Parallel Evolution within the New Cavefish Lineage

The percentages of total outlier SNPs in the new cave lineage ranged from 9.6% in N1 to 23.4% in N2. We identified two outliers unique for new cave populations, one of which was present within a QTL for eye size and length on LG14 and another in LG12 that is not associated with a QTL. The numbers of outlier loci detected comparing old cave with surface populations were not significantly greater than for new cave to surface populations ([table 4](#)). In the new cave populations, we only detected one outlier SNP near (1.76 cM) a candidate QTL for lens size ([table 5](#), gene *Fgf8*).

Within the new cave lineage, we detected six significant regions in four linkage groups (LG3, LG7, LG10, and LG16) ([fig. 5Aii](#)). Analysis of EHHS in LG3 showed fixation of the new cave haplotype (EHHS = 1) that was extended over 10cM from the core haplotype ([fig. 5Bii](#)). In contrast, surface and old cave populations had significantly lower EHHS values that quickly decayed. This suggests positive selection of this haplotype in the new cave populations and is an example of parallel molecular evolution in the new cave lineage.

Convergent Evolution of *Astyanax* Cavefish

Forty-four SNPs that were outliers common to both old and new cave populations are candidates for convergent evolution, although their ancestral state is unknown. Twelve of these markers were present in seven QTL regions. Thirty-two other markers that were in common were not present within the QTL ([supplementary table S2](#), [Supplementary Material](#) online).

The sharing of outlier SNPs between lineages was significantly greater than predicted by chance. Counting all SNPs in [supplementary table S2](#) ([Supplementary Material](#) online), we noted how often there was at least one P value of 0.05 or less in both the old and new cave population set. The observation was 119 times out of 518, compared with an expectation of 92.8, based on the proportions of hits among old and new caves, separately (Yates corrected $\chi^2 = 8.67$, $df = 1$, $P = 0.005$).

Significant ρ_{Rsb} values were common to old and new caves in two instances. In the first case, we detected three significant regions in QTL for sensitivity to dissolved odorants in LG16 ([fig. 5Aiii](#)). Furthermore, two of the three regions were not significantly different between the two cave groups, suggesting that the same allele may be present in both new and old lineages. In the second case, significance was detected in LG7, which has no QTL ([fig. 5Aiii](#)).

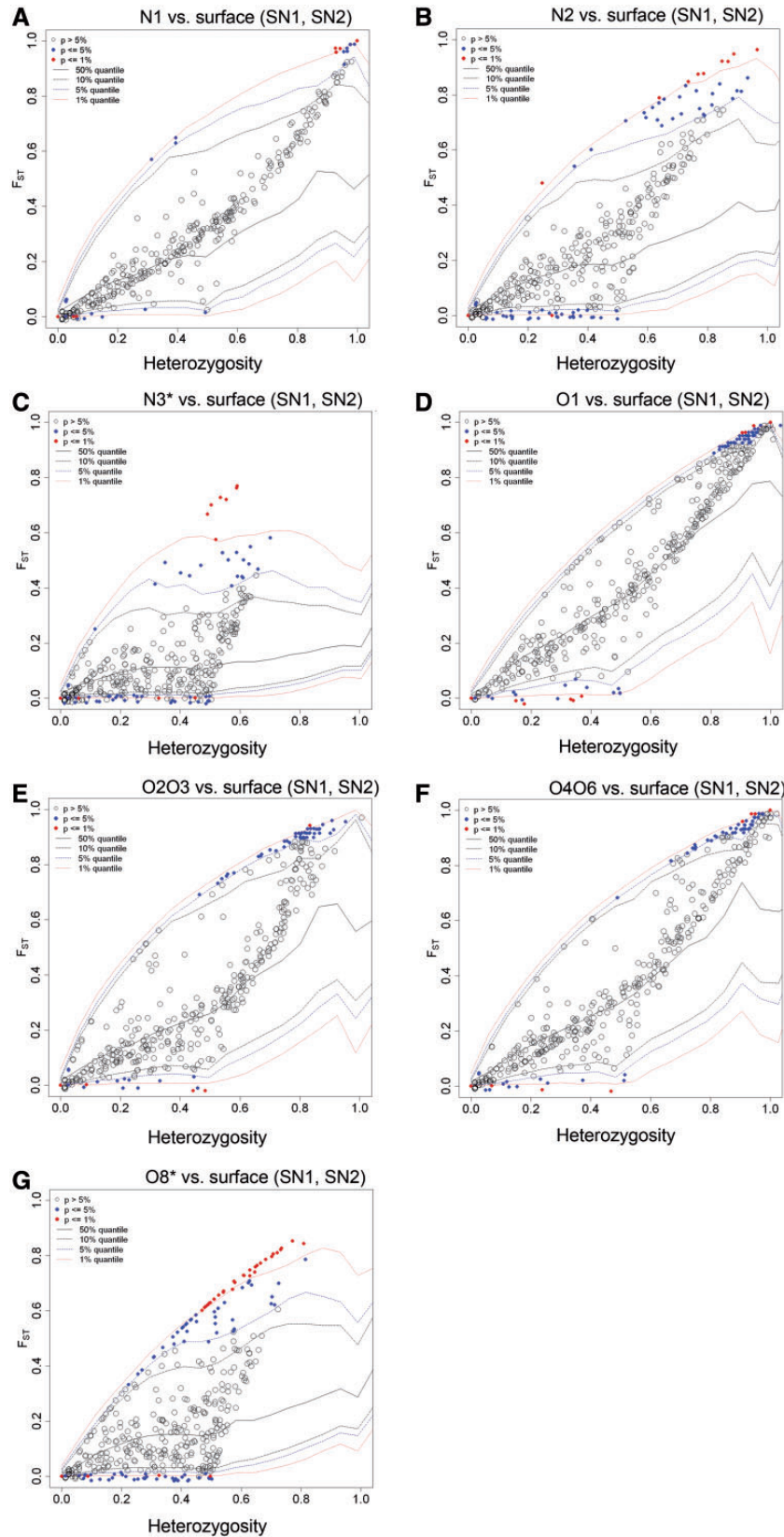


FIG. 3. SNP locus-specific F_{ST} as a function of expected heterozygosity in cave vs. surface comparisons for 518 loci. Confidence envelopes corresponding to the 50%, 10%, 5%, and 1% quantiles are drawn as solid and dashed lines. Individual SNPs deviating from null expectations (derived from the hierarchical island model) are shown in blue (5% significance) or in red (1% significance). Open circles represent SNPs that were not significant. Per locus F_{ST} values were calculated at individual markers, with the following population structure: (A) N1 vs. surface (SN1 and SN2); (B) N2 vs. surface (SN1 and SN2); (C) N3* vs. surface (SN1 and SN2); (D) O1 vs. surface (SN1 and SN2); (E) O2O3 vs. surface (SN1 and SN2); (F) O4O6 vs. surface (SN1 and SN2); (G) O8* vs. surface (SN1 and SN2).

Table 4. Numbers of F_{ST} Outliers Identified in Each Individual Population Based on the Hierarchical Model.

Type of Outlier	New Caves			Old Caves			
	N1	N2	N3*	O1	O2O3	O4O6	O8*
Total	30	77	73	135	104	78	108
Inside QTL	10	28	19	37	37	19	39
Outside QTL	20	49	54	98	67	59	69
HIMAF	6	43	43	17	11	15	43
LOMAF	24	34	30	118	93	63	65
All SNPs analyzed**	313	329	331	403	450	369	396

NOTE.—Outliers are further characterized by their location within or outside of QTL and by SNP category (LOMAF SNPs vs. HIMAF SNPs).

*Admixed populations previously described.

**Number of markers that are polymorphic in surface population and/or individual cavefish population from which F_{ST} values were calculated.

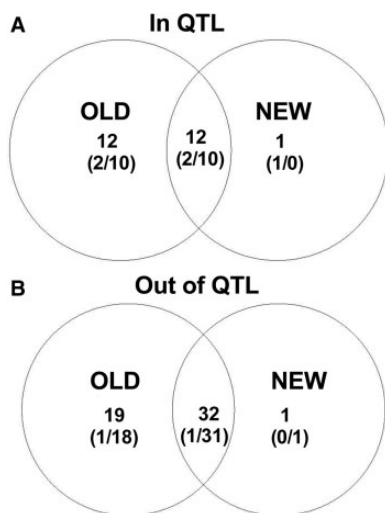


Fig. 4. Venn diagram of the outlier loci identified in each cave lineage relative to surface populations and the number of shared outlier loci between the two lineages. (A) The number of outliers within QTL. (B) The number of outliers out of QTL. Outliers are also separated in both figures based on HIMAF SNPs and LOMAF SNPs (HIMAF SNPs/LOMAF SNPs, first and the second number in the bracket, respectively).

Core haplotypes within QTL in LG16 exhibited the highest homozygosity in both the old and new cave groups (fig. 5Biii). Values of EHHS in both new and old cave populations were as high as 0.6 (new cave) and 1 (old cave) and decayed slowly across the region. In contrast, EHHS in surface populations decayed quickly to 0.6 and to 0. This suggests convergence of those haplotypes in both old and new caves.

Discussion

Numerous SNPs Are Significantly Diverged in Multiple Cave Populations

Advances in population genomics made genome-wide screens for adaptive loci feasible even in species lacking a sequenced genome (Barchi et al. 2011; Bradbury et al. 2011; Helyar et al. 2011; Seeb, Carvalho, et al. 2011; Seeb, Pascal, et al. 2011; Seeb, Templin, et al. 2011; Peterson et al. 2012; Scaglione et al. 2012).

Our study investigated the genetic framework of adaptation to a cave environment by means of a genome scan for

the loci with outlier behavior based on 518 SNP markers. We looked for loci diverging from neutral expectations when comparing populations between the old and new lineages. Altogether, 77 outlier loci were identified as potentially involved in adaptation to the cave environment because they were detected in two or more independent populations from the same lineage. This result confirmed the power of the outlier method to reveal potentially selected loci when multiple comparisons are available and when the association between demography and selection may be complex and/or cryptic as in this system.

The forces driving the evolution of cave phenotypes are still debated, with strong advocacy for 1) neutral mutation and drift, 2) direct selection, and 3) indirect selection through pleiotropy (Protas et al. 2007; Jeffery 2010; Wilkens 2010a, 2010b; Yoshizawa et al. 2012). Although we believe that all three mechanisms are involved in cave evolution, the present results strongly support a pervasive role for selection, direct or indirect, in the process.

Other population genetic causes besides natural selection, such as admixture and the Wahlund's effect, can also cause outlier behavior (Lewontin and Krakauer 1973; Beaumont and Nichols 1996; Luikart et al. 2003; Beaumont and Balding 2004; Hess et al. 2011). For example, microsatellite surveys of the N3* and O8* populations showed admixture between cave and surface individuals (Brdic et al. 2012); consistent with this we identified many loci out of HW in both of these populations as well as many loci with F_{IS} more than 0 (supplementary table S3, Supplementary Material online). Thus, there might be some influence of admixture on the estimation of the outliers, due to Wahlund effects. To test whether SNP differentiation could be caused by demography rather than natural selection, we examined the correlation between F_{IS} and F_{ST} over all populations at each polymorphic locus. Cave populations in which F_{IS} could be calculated exhibited rather extreme values, either $F_{IS} = 1$ or $F_{IS} = 0$ (supplementary table S3, Supplementary Material online). Intermediate values were mostly present in admixed populations (N3* and O8*) and in the old population (O4O6). Again, we did not see significant correlations between F_{ST} and F_{IS} (supplementary fig. S6, Supplementary Material online). Thus, inbreeding is not a major contributing factor in SNP differentiation.

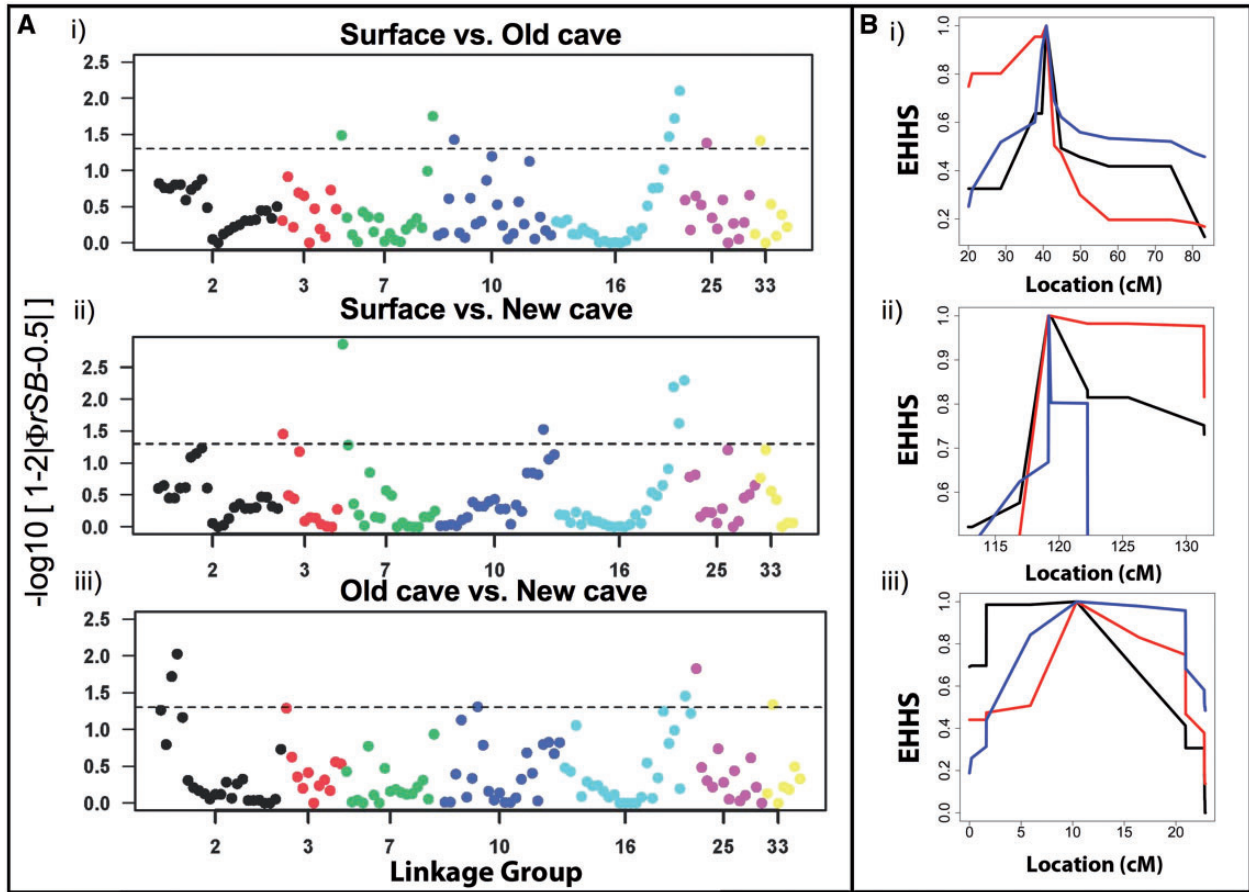


FIG. 5. Haplotype structure within linkage groups of interest. These include all linkage groups having QTL associated with outlier SNPs, plus one linkage group without QTL (LG7) for contrast. LG7 was chosen because it had the most outlier SNPs of any linkage group. For each region, haplotypes estimated in each population group were pooled. (A) Three sets of comparisons and their P values are given for (i) surface vs. new cave, (ii) surface vs. old cave, and (iii) old vs. new cave over the *Astyanax* linkage groups. Significance of Rsb statistics between groups is shown as $-\log_{10} P$ value. The dotted lines on the graphs represent the 5% significance limit; observations exceeding that limit represent significant differentiation of the compared groups. Significant P values indicate much slower EHHS decay in one population than the other, and therefore represent possible evidence of selection. Different colors represent markers in different linkage groups. (B) Extended haplotype homozygosity of an individual SNP site (EHHS) across three linkage groups. (i) EHHS across LG25, (ii) EHHS across LG3, (iii) EHHS across LG16. The red line represents EHHS in new cave populations, the blue line represents EHHS in old cave populations and the black line represents EHHS in surface populations. New cave populations are N1, N2, and N3*, old cave populations are O1, O2O3, O4O6, and O8* and surface populations are SN1 and SN2.

Table 5. Summary of F_{ST} Outliers Identified from Candidate Genes.

Candidate Gene	QTL Status	No. of Markers	No. of Markers after QC	Populations with Outliers
<i>GH1</i>	+	5	5	—
<i>Oca2</i>	+	13	3	O2O3
<i>Cryaa</i>	+	9	8	O1
<i>Mc1r</i>	+	5	4	O2O3
<i>Fgf8</i>	+	5	2	N1, N2, N3*, O1, O2O3, O4O6, O8*
<i>Prox1</i>	?	3	2	—
<i>Mch1R</i>	+	1	0	—
<i>Hsp90</i>	—	1	0	—

NOTE.—Abbreviations represent the following genes: *GH*, growth hormone; *Oca2*, oculocutaneous albinism type 2; *Cryaa*, α A-crystallin; *Mc1r*, melanocortin 1 receptor; *Fgf8*, fibroblast growth factor 8; *Prox1*, prospero homeobox 1; *Mch1R*, melanin concentrating hormone 1; *Hsp90*, heat shock protein 90; QTL status, presence in a QTL; Markers typed, number of markers genotyped in the study; markers after QC, number of markers after quality control of the markers; Outliers, significant F_{ST} value (0.05) and population in which the marker was observed. QC, quality control (MAF >0.05); ?= unknown QTL status.

*Admixed populations previously described in Bradic et al. (2012).

We detected 77 outlier SNPs shared across independently derived cave populations (31 exclusive to old and two exclusive to new cave populations). Although some of these cases might arise through chance events, the most parsimonious explanation is that in most cases the same genomic regions were involved in parallel adaptations within each lineage. This suggests that parallel genetic changes are correlated with the level of the relationship between the populations, and that evolutionary history is a very important factor in the process of adaptation (Cohan 1984; Travisano et al. 1995; Bull et al. 1997; Crill et al. 2000; Teotonio and Rose 2000; Teotonio et al. 2009).

Forty-four convergent loci across different lineages were also present and they suggest that despite the different evolutionary history of the two lineages the same alleles were used in the evolution of independent populations. The repeated usage of the same genes across distantly related populations and taxonomic groups is well established for some genes in vertebrates (i.e., *Mc1r*) (Nachman et al. 2003; Hoekstra 2006; Arendt and Reznick 2008) as well as usage of same loci in distantly related mimetic butterflies (Baxter et al. 2010; Counterman et al. 2010). This suggests that there are constraints on which genomic regions can evolve to generate particular adaptive phenotypic variants and thus that the evolutionary response may be predictable to some extent (Orgogozo and Stern 2009).

Earlier studies suggested that old cave populations may be specifically distinct from surface fish of the El Abra, but, as of yet, there have been no systematic studies of this question and it remains unresolved. However, given the long history of divergence between surface and old cave populations, limited gene-flow (allopatric populations) and periodical changes in population size, some genome-wide genetic barriers such as hybrid sterility between those populations could be expected. In this case, evolutionary divergence between those populations detected as outlier loci might be related to genomic regions that are impermeable to introgression between taxa (Dobzhansky–Muller interactions) due to the association with a barrier (Noor and Feder 2006; Wolf et al. 2010). As a result, some of the F_{ST} outliers could be false positives and not necessarily indicative of natural selection (Bierne et al. 2013). Because of this possibility, as emphasized earlier, we only considered statistically significant outliers as potentially biologically significant when they occurred in two or more independently derived populations.

We observed instances of long distance LD in both old and new cave populations while surface populations showed low LD (supplementary fig. S7, Supplementary material online). It is possible that these LD structures reflect unidirectional gene flow from surface to cave populations or epistatic interactions among loci. Our haplotype analysis within QTL regions shows different levels of EHH in different loci (fig. 5), which further suggests the existence of extended LD islands, perhaps reflecting significant population structure or epistasis, as has been observed in other natural systems (white fish and sticklebacks) (Hohenlohe et al. 2012; Renaut et al. 2012). Epistasis and gene flow could create false positive outliers.

Previous work suggests that there is little gene flow from surface to the old cave populations, but perhaps there is a greater influx of surface genes into the new cave populations (Bradic et al. 2012; Strecker et al. 2012) so this is, at worst, a minor distorting factor.

Repeated Evolution from Standing Genetic Variation

The data suggest that standing variation was an important source of adaptive variants facilitating cave adaptation. We base this conclusion on the recurrent patterns of heterozygosity reduction at the same outlier loci, and divergence of the same haplotypes in independent cavefish populations. Independent repeated mutations and their subsequent fixation in moderately sized or small populations, such as we have in *Astyanax*, are unlikely to occur in independently derived populations (Cohan and Hoffmann 1989; Travisano et al. 1995; Bull et al. 1997; Crill et al. 2000; Teotonio and Rose 2000; Teotonio et al. 2009).

Other studies have come to similar conclusions. In sticklebacks, alleles at the Ectodysplasin (*Ectd*) locus responsible for putatively adaptive reductions in bony lateral plates are present in low frequencies in the ancestral populations, and rapidly increased in frequency subsequent to the populations' invasion of freshwater habitats (Colosimo et al. 2005; Barrett et al. 2009; Schluter and Conte 2009). Recent genome wide studies of sticklebacks and whitefish also suggest that most of the adaptive variation is present in a very low frequency in the ancestral population (Hohenlohe, Phillips, et al. 2010; Renaut et al. 2011).

One limitation in our ability to reconstruct the evolutionary events leading to present day population structure is that the genetic states in the ancestral surface fish are unknown. This limitation is common to all such studies. In our case, we use extant surface fish from the El Abra region as surrogates for the ancestors, because these are the most closely related surface stocks that have yet been identified. For many of the cave related traits, the counterpart surface phenotypes, such as functional eyes and pigmentation, are under strong stabilizing selection, and it may be that their current genetics closely reflects the ancestral state.

We also observed a large reduction of variation in cave populations that can be explained by demographic events (i.e., bottleneck) during cave colonization or by ascertainment bias introduced by our SNP discovering approach; these largely decreased our power to detect outlier loci. This resulted in small numbers of variants that were present in either new or old lineages (LOMAF SNPs) but at low frequencies or undetected in the surface populations (HIMAF and LOMAF SNPs).

Although it is possible that some of the increases in frequencies of the same haplotypes that occurred in independent populations could have been driven by selection for independent mutations that arose close in time to when the cave colonization events occurred, we consider this unlikely. Furthermore, some of the adaptive haplotypes (i.e., LG3) were also identified in surface populations. Taken together, the sum of the data suggest that at least some of the commonly selected loci in independent cave populations

represent cases of standing genetic variation, sorted out by natural selection during parallel cave adaptation.

The multiple independent instances of cave adaptation within two different lineages allowed us to explore the possibility of reuse of the same alleles in the adaptation of those populations. We observed more outlier loci to be shared within each lineage than between the lineages. This suggests that the parallel reuse of the same variation is very likely. This allows for possible understanding of evolutionary paths, because the structure of the variation, epistasis, and linkage in the genome that is required for the adaptive shift to happen is still shared in high degree, especially among recently diverged lines (Conte et al. 2012). Thus, we suggest that sharing of population history between phylogenetically close populations of new and old lineages plays a significant role in the response to a new environment. However, we do not know if the same mutations in the parallel cases are involved in the adaptation because our study could not distinguish among individual mutations in the same or closely linked loci.

Integrating Population Genetics with QTL Mapping to Detect Repeated Evolution

We expected that outlier loci would tend to map within QTL and candidate genes if their divergences were correlated with the evolution of cave phenotypes. This expectation has met mixed success in previous studies. For example, recent studies in sticklebacks showed that multiple peaks of differentiation fell within previously found QTL (Peichel et al. 2001; Cresko et al. 2004; Baird et al. 2008; Hohenlohe, Bassham, et al. 2010; Rogers et al. 2012), and similar results have been found in *Heliconius* butterflies (Baxter et al. 2010; Counterman et al. 2010; Joron et al. 2011; Wark et al. 2012). In other studies, however, little correlation between differentiation and QTL was detected in pea aphids, whitefish, and sunflowers (Rogers and Bernatchez 2007; Yatabe et al. 2007; Via and West 2008; Renaut et al. 2011). In our case, a subset of outlier SNPs were significantly closely linked to QTL, but there were many outlier SNPs not near QTL and many QTL not near outlier SNPs. Thus, in our comparison of cavefish and surface stock, QTL and outlier analyses give complementary, rather than coincident information. There are several reasons why this might be so.

QTL analyses typically begin with small numbers of line founders. As such, they usually capture only a subset of relevant loci. Further, the power of QTL analysis to detect and localize evolutionarily relevant loci is limited by the relatively small number of progeny that can be raised and by the low number of recombinants generated between close markers (Stinchcombe and Hoekstra 2008). Thus, the subset of QTL detected in our previous studies may have missed a proportion of the loci under selection pressure in natural populations. Also, it may be that F_{ST} -based signatures of selection we detected are correlated with other troglomorphic traits not included in our QTL studies. It may also be that ancient correlations between QTL and F_{ST} outliers have decayed through time; adaptive QTL have longer persistence times than population genetic signatures of selection, which eventually

narrow through recombination. In this light, we must note, however, that we detected more loci as outliers in old cave populations. This unanticipated result might largely reflect the influence of ascertainment bias from sampling SNPs from hybrids between one old cave population and one surface population. Finally, it is possible that our analysis missed outlier loci through too stringent a criterion for their detection (i.e., we defined an outlier by its presence in two or more populations from the same lineage) and that some of the candidate loci do not have any adaptive significance (Protas et al. 2008; Wilkens 2010b).

We mapped the gene *Fgf8* to within the QTL for lens size in LG3. Haplotype analysis of the markers in this region revealed significant divergence of new cave populations, suggesting a history of directional selection (fig. 5B). A recent study in cavefish described signaling modifications of *Fgf8* and suggested that its early embryonic expression pattern is crucial to the normal development of the eye and lens (Pottin et al. 2011).

We also identified parallel evolution in old cave populations as significant haplotypes near QTL for length (LG10), for the size of the dentary (LG25), and for condition factor (LG33) (supplementary fig. S2, Supplementary Material online). One phenotype for which we have identified potential convergence is related to sensitivity to dissolved amino acids, a presumed feeding adaptation, (LG16-A, supplementary fig. S2, Supplementary Material online). However, no potential candidate genes could be identified in this region because no genome sequence is yet available for this species and BLAST searches of other organisms failed to find informative alignments.

Conclusion

The *Astyanax* cavefish system has a unique combination of properties that ideally suit it for the study of the molecular bases of adaptive evolution. It has a complex phylogeny with independent convergences on a cave-adapted phenotype occurring in numerous cave populations in two long diverged lineages (6.7 Ma). This permits investigation of potential parallel and convergent molecular changes. From comparisons of SNP variation among multiple cave and surface fish populations, we identified 77 outlier SNPs that diverged in frequency in two or more independently evolved cave populations out of a total of 518 SNPs surveyed. SNPs were shared among cave populations of the same lineage and sometimes between lineages, providing evidence of both parallel and convergent molecular evolution, and suggesting that standing variation in ancestral populations, rather than new mutation, provided most of the alleles involved in cave adaptation. Reuse of the same alleles was more common in populations the more closely they were related, in agreement with other recent studies (Conte et al. 2012). For the most part, outlier SNPs did not co-map with the QTL we identified for cave adapted phenotypic variation, perhaps because of the antiquity of the evolutionary events. As such, the SNP and QTL data sets are complementary; both provide important information for the ultimate identification of the causative loci and alleles.

Materials and Methods

DNA Sampling

*F*₂ Cross Samples

We used DNA samples from a previously described *F*₂ mapping progeny obtained by crossing two full sib *F*₁ individuals derived from a cross between a wild caught surface fish (Río Valles, San Luis Potosí) and an individual from the O1 (Pachón) cave population (Protas et al. 2006, 2007).

Wild Specimens Derived from Two Diverged Lineages

Fish specimens used in this study were collected in March 2008 and preserved in 70% ethanol in the field. Subsequent DNA extraction was done by standard methods described elsewhere (Protas et al. 2006). We collected samples from multiple geographical locations in Northeast Mexico: El Abra region: Cueva Pachón (O1, *n* = 32), Cueva Yerbaniz (O2, *n* = 10), Japonés (O3, *n* = 12), Cueva Arroyo (O4, *n* = 12), Curva (O6, *n* = 17), Cueva Chica (O8*, *n* = 32, * denotes an admixed population); Guatemala region: Cueva Molino (N1, *n* = 21), Cueva Caballo Moro (N2, *n* = 26); Micos region: Cueva Río Subterráneo (N3*, *n* = 25, * denotes an admixed population); from caves and from surface localities on the eastern face of the Sierra de El Abra: SN1: SN1a (*n* = 8), SN1b (*n* = 10), SN1c (*n* = 7), and SN1d (*n* = 20), to the south and west of the El Abra, Río Subterráneo Valley surface locality (SN2, *n* = 25). The distribution of the populations and their geographical positions are shown in the map (fig. 1A).

SNP Discovery

RAD tag sequencing (Baird et al. 2008; Davey and Blaxter 2010; Stapley et al. 2010; Davey et al. 2011) was done on three *Astyanax* *F*₁ individuals from a cross between a surface fish and a cavefish from population O1 (Pachón) to identify SNPs informative for linkage map construction in the *F*₂ mapping progeny. DNA from each of the three *F*₁ hybrids was digested with high fidelity SbfI (New England Bio labs) and RAD tag libraries were created as in Baird et al. (2008). The multiplexed *Astyanax* library was placed on a single channel of the Illumina GAI system using 2 × 36 bp sequencing chemistry (Paired End Sequencing).

To develop additional polymorphic markers, we resequenced eight candidate genes and five BACs from a BAC genomic library (Di Palma et al. 2007). Four BAC clones were randomly picked from the library and one clone was selected because it contained the potential candidate gene, growth hormone (GH). A standard BAC library screening method was used to extract the GH clone (Di Palma et al. 2007). We obtained flanking sequences around the candidate genes *Mc1r*, *Oca2*, *Fgf8*, and *Cryaa* (Strickler et al. 2001; Protas et al. 2006; Retaux et al. 2008; Gross et al. 2009) using the Genome Walker kit (Clontech) and gene-specific primers designed using the online software program Primer3 (frodo.wi.mit.edu, last accessed August 26, 2013).

SNPs markers were developed based on sequencing of candidate genes and BAC fragments using a panel of 24 individuals from surface and 12 from cave populations (six from old and six from new cave populations). Sequencing of

polymerase chain reaction products was performed on an ABI3730 automated DNA sequencer (Applied Biosystems). Polymorphic loci were detected by alignment of sequences in BioEdit v7.0.9 (Hall 1999) and verified by visual inspection of the chromatograms. All the SNPs for BAC fragments and candidate gene regions were chosen such that the MAF of a given SNP was at least 5% (3 surface individuals out of 24 sequenced). This data set contained a total of 188 SNP markers. All the sequences with the assigned SNPs are present in the NCBI database with the submitted SNP (ss) numbers: 825690719–825691453.

SNP Genotyping

We genotyped SNPs derived using the MassARRAY mass spectrometry platform (Sequenom, San Diego, CA) to genotype 272 samples from 12 natural populations (table 1) for 745 SNPs and 276 *F*₂ individuals for 675 SNPs (not all SNPs within the same gene or BAC clone were typed in the *F*₂ because the low recombination rate made them redundant (supplementary fig. S4B, Supplementary Material online).

Assays were multiplexed up to 40 SNPs per reaction. All SNP genotyping was performed according to the standard Sequenom iPLEX protocol (Bradic et al. 2011). Genotypes were assigned based on the presence of mass peaks by the MassARRAY Typer v4.0 software (Sequenom) (Bradic et al. 2011). Results were manually inspected and verified, using the MassARRAY Typer Analyzer v4.0 software (Sequenom).

Quality Control of Surveyed Markers

Basic quality control was performed on all the markers used either in linkage map construction or in population genomics analysis. Low quality markers, defined as those with 20% missing data or with singleton SNPs, were discarded. Individuals in which more than half the genotypes were missing were discarded from further analysis. Additional classification was performed on the genotypes from natural populations.

SNPs were pooled in two groups based on their polymorphism in the surface population: MAF more than 5% versus MAF less than 5%. Those in the first category are termed HIMAF SNPs while those in the latter are termed LOMAF SNPs. This grouping was based on our expectation that the majority of the variability should be present in the surface population in the form of standing genetic variation at least for the more recent cave lineages (new caves).

Genetic Diversity

Measures of genetic diversity were estimated for each marker and each population or population group (discussed later) by calculating the percentage of polymorphic SNPs (PO), the mean number of alleles per SNP (*A*), and the observed (*H*_o) and unbiased expected heterozygosities defined as $H_e = (n / [n - 1]) \times 2pq$ where *n* represents the sample size and *p* and *q* frequencies of each allele (Nei 1973; Nei and Roychoudhury 1974). AR and private AR, were calculated for all populations and corrected (rarefied) based on the smallest sample size (*n* = 22) using HP-Rare (Kalinowski 2005). The within-population fixation index $F_{IS} = 1 - (H_o/H_e)$ was calculated as previously described (Nei 1977; Weir and Cockerham 1984).

Deviations from HWE were estimated for each polymorphic marker in each population using a Fisher's exact test as described in Wigginton et al. (2005).

We verified that pooling across populations did not cause spurious departures from HWE in our data; populations pooled together were generally very similar in their allelic frequency distributions (data not shown). Thus, we combined groups as follows: for old cave populations, O2 + O3 = O2O3 and O4 + O6 = O4O6. For surface populations SN1a + SN1b + SN1c + SN1d = SN1. All the other populations were treated as independent units. Designations of pooled groups are used consistently in this article. Bonferroni correction was applied in all analyses in which multiple testing was employed (Dunn 1961).

Outlier Loci Detection

To detect potentially adaptive regions in the *Astyanax* genome, we tested for loci with atypical behavior (outlier loci), defined herein as those exhibiting higher F_{ST} values than expected based on drift alone (Lewontin and Krakauer 1973; Weir and Cockerham 1984; Beaumont and Nichols 1996; Luikart et al. 2003; Beaumont and Balding 2004; Beaumont 2005; Bonin 2008; Holsinger and Weir 2009; Narum and Hess 2011).

We calculated the Wright–Fisher estimator of population differentiation (F_{ST}) using the hierarchical island model as implemented in ARELQUIN v3.5 (Beaumont and Balding 2004; Excoffier et al. 2009; Excoffier and Lischer 2010). This model allows for nested analysis of variance and thus accounts for replicated population samples within each hierarchically structured population (Beaumont and Nichols 1996; Beaumont and Balding 2004; Excoffier et al. 2009). We accounted for the structuring in the surface populations (SN1 and SN2) and contrasted their allelic diversity with each of the cave population groups separately (N1, N2, N3*, O1, O2O3, O4O6, and O8*). The analysis was performed assuming the presence of 20 groups (k) of 100 demes (d) with 100,000 simulated loci under the model of hierarchical structure. Prior work has suggested that these parameters give reliable results (Beaumont and Balding 2004; Excoffier et al. 2009; Excoffier and Lischer 2010). Changing these parameters had little effect on the results of the analyses.

Divergences of the individual loci based on F_{ST} measures were calculated for each cave-surface pair. F_{ST} values are shown as functions of expected average heterozygosities (H_e) and F_{ST} . P values were estimated based on the joint distribution between F_{ST} and H_e using a kernel density function.

All the markers (HIMAF SNPs and LOMAF SNPs), a total of 518, were used in ARLEQUIN v3.5 simulations, and the significant outliers were sorted after analysis; only markers that were repeatedly identified in two or more populations from the same lineage were considered for further analysis. This conservative approach minimized false positives reflecting possible errors in outlier detection methodology. For example, there could be discrepancies when the number of immigrants is asymmetrical between populations (Murakami et al. 2007), which is the case of some of our populations

(e.g., migration between the surface population and the O8* cave population). In this analysis, we limited our search to patterns of divergence suggesting directional selection, due to documented limitations of these methods to detect balancing selection (Excoffier et al. 2009; Narum and Hess 2011).

All the detected loci were classified by their origin: as across lineage outliers (present in old and new cave populations) or as lineage-specific outliers (present only in old or new cave populations).

Linkage Map Construction

The previously published linkage map was based on genotyping 539 F_2 individuals using microsatellite markers. The map had 29 linkage groups defined by 259 markers, a total map length of 2,148 cM, and an average resolution of 9.3 cM (Protas et al. 2006, 2008).

We genotyped a subset ($n = 270$ individuals) of the original mapping progeny for 451 SNPs and constructed a new, integrated map by including 259 microsatellite markers. For map construction, we used Multipoint software, as previously described (Korol et al. 1998). Data were checked for segregation patterns and genotypic phase based on parental genotypes. As the progenitors of the mapping progeny were collected from outbred populations, a number of the SNP loci (45) were heterozygous in both P_1 individuals. These were uninformative for mapping and omitted from further analysis. The detailed approach for map generation was previously described (Protas et al. 2008). Briefly, we performed quality control on the loci that showed abnormal segregation as determined using a chi-square goodness of fit test, and removed those where P was less than 0.005. When two or more markers mapped to nearly identical positions, redundant markers were discarded from the map. The final map contained 450 markers and defined 40 linkage groups with a total map length of 2,646 cM (the haploid number in this species is $n = 25$).

QTL mapping with our new map was repeated as done previously (Protas et al. 2008) to establish the regions of interest related to the quantitative trait. We recalculated the QTL for 11 traits previously phenotyped: relative eye size (E), melanophore numbers on the flanks (MI), dorsal melanophore numbers (Md), melanophore numbers in the eye area (Me), condition factor (C), sensitivity to dissolved amino acids (A), rate of weight loss on fast (W), body length (L), rib numbers (B), estimated daily growth rate (G), and relative length of the dentary bone (D) (Protas et al. 2008). Detailed descriptions of each trait are given in table 2. MultiQTL (www.multiqtl.com, last accessed August 26, 2013) was used to search for single QTL for the traits listed earlier to determine the LOD scores and proportions of phenotypic variances in the F_2 mapping progeny explained by the total and additive effects of the QTL (table 3).

In brief, we first identified linkage groups with significant or suggestive trait associations ($P < 0.10$) using simple interval mapping (Lander and Botstein 1989). Next, all linkage groups with putative QTL were used as starting sets for multiple interval mapping (MIM) using the MIM functions of

MultiQTL for each trait (Korol et al. 1998). MIM models each QTL, adjusting for the effects of all other detected QTL for the same trait through an iterative process. This procedure minimizes the effects of background variation and optimizes the estimates of QTL parameters (Kao et al. 1999). Many of the linkage groups in the starting sets had no significant QTL and were eliminated in the first round of MIM. Repeating the MIM analysis generally led to an improvement in significance and precision of the remaining QTL, although some occasionally lost significance. When QTL lost significance, they were eliminated and the procedure was repeated until a stable set was obtained. The final rounds of MIM analysis for each trait with parameters set at the highest stringency confirmed that the estimates of QTL were consistent. Permutation was used to assess significances of all QTL, and confidence intervals on their positions were determined by bootstrap analyses. The significance threshold was set at $P = 0.05$ for individual QTL, with a genome-wide false detection rate of 10% (FDR = 0.10). These methods detect at most one QTL per trait per linkage group. Although it is possible to test hypotheses of two linked QTL using MultiQTL, the data set is small and there is insufficient power to reliably detect linked QTL. Thus, the tests were not performed.

Phasing and Haplotype Analysis

Regions of interest were defined based on the QTL position in the linkage map. Only QTL regions in which SNP outliers were detected were used in the haplotype phasing. The size of the phased region was defined as the one-LOD support interval (Lander and Botstein 1989). All the markers within the support interval were used in phasing individual genotypes into haplotypes. Genotypes of individuals from each natural population were phased into haplotypes using fastPHASE v1.2. (Scheet and Stephens 2006). We used 20 random starts, 200 sampled haplotypes from the posterior distribution, and 10 cross-validation numbers of clusters in the fastPHASE settings algorithm. All the genotypes with posterior probabilities lower than 90% were treated as missing data and were not included in further analysis. We also performed additional quality control by removing haplotypes if the phase could not be inferred for all markers. Haplotypes with frequencies less than 5% were discarded from further analysis.

To contrast the length of haplotype homozygosity regions between surface and new cave, surface and old cave, and new versus old cave populations, we calculated *Rsb* statistics using the *calc_ehhs* and *ies2rsb* functions of the *rehh* package in R (Tang et al. 2007; Gautier and Naves 2011). This statistic permits the contrasting of differences in decay of extended haplotype homozygosity at individual SNP sites (EHHS). EHHS is defined as the decay of identity of haplotypes starting from the core SNP site of a population as a function of distance (Sabeti et al. 2002; Tang et al. 2007; Gautier and Naves 2011). To identify regions in which EHHS was significantly higher in cave populations than in the surface population, we calculated *Rsb* statistics as a ratio of the integrated EHHS relative to map distances (cave population over surface population) standardized as described previously (Tang et al. 2007; Gautier and Naves 2011).

Supplementary Material

Supplementary materials, tables S1–S3, and figures S1–S7 are available at *Molecular Biology and Evolution* online (<http://www.mbe.oxfordjournals.org/>).

Acknowledgments

M.B. designed the study, performed the experiments, analyzed the data, and wrote the manuscript. H.T. designed the study, supervised the project, and co-wrote the manuscript. R.B. designed the study, organized the collection plan and field work, and co-wrote the manuscript. The authors thank the Mexican government for providing the collecting permit (DGOPA.00570.288108-0291). They thank R. Nipper at Floragenex and J. Costa and I. Marques at Instituto Gulbenkian de Ciência, Oeiras, Portugal, for support with SNP discovery and genotyping, and I. Chelo for support with data analysis. They thank F.J. García-de León, S. Esquivel-Bobadilla, G. Hoese, and J. Lacaille Muzquiz for their help during the fieldwork. They thank S. Kryazhimskiy and M.A. Aboim for the useful comments on the manuscript. They also thank two anonymous reviewers for their useful comments and improvement of the manuscript. This project is partial completion of requirements for the Ph.D. degree at Universidade Nova de Lisboa, Oeiras, Portugal, for M.B. This work was supported by Fundação para a Ciência e a Tecnologia PhD grant (SFRH/BD/32982/2006) to M.B., an NSF IOS-0821982 award to R.B., and by the start-up funds from the Instituto Gulbenkian de Ciência, Oeiras, Portugal, to H.T.

References

- Arendt J, Reznick D. 2008. Convergence and parallelism reconsidered: what have we learned about the genetics of adaptation? *Trends Ecol Evol.* 23:26–32.
- Baird NA, Etter PD, Atwood TS, Currey MC, Shiver AL, Lewis ZA, Selker EU, Cresko WA, Johnson EA. 2008. Rapid SNP discovery and genetic mapping using sequenced RAD markers. *PLoS One* 3:e3376.
- Barchi L, Lanteri S, Portis E, Acquadro A, Vale G, Toppino L, Rotino GL. 2011. Identification of SNP and SSR markers in eggplant using RAD tag sequencing. *BMC Genomics* 12:304.
- Barrett RD, Rogers SM, Schluter D. 2009. Environment specific pleiotropy facilitates divergence at the *Ectodysplasin* locus in threespine stickleback. *Evolution* 63:2831–2837.
- Baxter SW, Nadeau NJ, Maroja LS, et al. (20 co-authors). 2010. Genomic hotspots for adaptation: the population genetics of Mullerian mimicry in the *Heliconius melpomene* clade. *PLoS Genet.* 6:e1000794.
- Beaumont MA. 2005. Adaptation and speciation: what can F_{st} tell us? *Trends Ecol Evol.* 20:435–440.
- Beaumont MA, Balding DJ. 2004. Identifying adaptive genetic divergence among populations from genome scans. *Mol Ecol.* 13:969–980.
- Beaumont MA, Nichols RA. 1996. Evaluating loci for use in the genetic analysis of population structure. *Proc R Soc B Biol Sci.* 263:1619–1626.
- Bell MA, Foster SA. 1994. The evolutionary biology of the threespine stickleback. Oxford: Oxford University Press.
- Bell MA, Orti G. 1994. Pelvic reduction in threespine stickleback from Cook Inlet lakes: geographical distribution and intrapopulation variation. *Copeia* 1994:314–325.
- Bierne N, Roze D, Welch JJ. 2013. Pervasive selection or is it . . . ? Why are F_{ST} outliers sometimes so frequent? *Mol Ecol.* 22:2061–2064.
- Bonin A. 2008. Population genomics: a new generation of genome scans to bridge the gap with functional genomics. *Mol Ecol.* 17:3583–3584.

- Borowsky R. 2010. The evolutionary genetics of Cave fishes: convergence, adaptation and pleiotropy. In: Trajano E, Bichuette ME, Kapoor BG, editors. *Biol Subterranean Fishes*. Enfield: Science Publishers. p. 141–168.
- Bradbury IR, Hubert S, Higgins B, et al. (11 co-authors). 2011. Evaluating SNP ascertainment bias and its impact on population assignment in Atlantic cod, *Gadus morhua*. *Mol Ecol Resour*. 11:218–225.
- Bradic M, Beerli P, Garcia-de Leon FJ, Esquivel-Bobadilla S, Borowsky RL. 2012. Gene flow and population structure in the Mexican blind cavefish complex (*Astyanax mexicanus*). *BMC Evol Biol*. 12:9.
- Bradic M, Costa J, Chelo IM. 2011. Genotyping with sequenom. *Methods Mol Biol*. 772:193–210.
- Bull JJ, Badgett MR, Wichman HA, Huelsenbeck JP, Hillis DM, Gulati A, Ho C, Molineux IJ. 1997. Exceptional convergent evolution in a virus. *Genetics* 147:1497–1507.
- Cohan FM. 1984. Can uniform selection retard random genetic divergence between isolated conspecific populations. *Evolution* 38: 495–504.
- Cohan FM, Hoffmann AA. 1989. Uniform selection as a diversifying force in evolution—evidence from *Drosophila*. *Am Nat*. 134:613–637.
- Colosimo PF, Hosemann KE, Balabhadra S, Villarreal G Jr, Dickson M, Grimwood J, Schmutz J, Myers RM, Schluter D, Kingsley DM. 2005. Widespread parallel evolution in sticklebacks by repeated fixation of Ectodysplasin alleles. *Science* 307:1928–1933.
- Colosimo PF, Peichel CL, Nereng K, Blackman BK, Shapiro MD, Schluter D, Kingsley DM. 2004. The genetic architecture of parallel armor plate reduction in threespine sticklebacks. *PLoS Biol*. 2:E109.
- Conte GL, Arnegard ME, Peichel CL, Schluter D. 2012. The probability of genetic parallelism and convergence in natural populations. *Proc Biol Sci*. 279:5039–5047.
- Counterman BA, Araujo-Perez F, Hines HM, et al. (18 co-authors). 2010. Genomic hotspots for adaptation: the population genetics of Mullerian mimicry in *Heliconius erato*. *PLoS Genet*. 6:e1000796.
- Cresko WA, Amores A, Wilson C, Murphy J, Currey M, Phillips P, Bell MA, Kimmel CB, Postlethwait JH. 2004. Parallel genetic basis for repeated evolution of armor loss in Alaskan threespine stickleback populations. *Proc Natl Acad Sci U S A*. 101:6050–6055.
- Crill WD, Wichman HA, Bull JJ. 2000. Evolutionary reversals during viral adaptation to alternating hosts. *Genetics* 154:27–37.
- Culver DC, Fong DW. 1986. Why all cave animals look alike. *Stygologia* 2: 208–216.
- Davey JW, Blaxter ML. 2010. RADSeq: next-generation population genetics. *Brief Funct Genomics* 9:416–423.
- Davey JW, Hohenlohe PA, Etter PD, Boone JQ, Catchen JM, Blaxter ML. 2011. Genome-wide genetic marker discovery and genotyping using next-generation sequencing. *Nat Rev Genet*. 12:499–510.
- Di Palma F, Kidd C, Borowsky R, Kocher TD. 2007. Construction of bacterial artificial chromosome libraries for the Lake Malawi cichlid (*Metriaclichia zebra*), and the blind cavefish (*Astyanax mexicanus*). *Zebrafish* 4:41–47.
- Dowling TE, Martasian DP, Jeffery WR. 2002. Evidence for multiple genetic forms with similar eyeless phenotypes in the blind cavefish, *Astyanax mexicanus*. *Mol Biol Evol*. 19:446–455.
- Duboué E, Keene A, Borowsky R. 2011. Evolutionary convergence on sleep loss in cavefish populations. *Curr Biol*. 21:671–676.
- Dunn OJ. 1961. Multiple comparisons among means. *J Am Stat Assoc*. 56:52–64.
- Excoffier L, Hofer T, Foll M. 2009. Detecting loci under selection in a hierarchically structured population. *Heredity* 103:285–298.
- Excoffier L, Lischer HEL. 2010. Arlequin suite ver 3.5: a new series of programs to perform population genetics analyses under Linux and Windows. *Mol Ecol Resour*. 10:564–567.
- Gautier M, Naves M. 2011. Footprints of selection in the ancestral admixture of a New World Creole cattle breed. *Mol Ecol*. 20:3128–3143.
- Gross JB. 2012. The complex origin of *Astyanax* cavefish. *BMC Evol Biol*. 12:105.
- Gross JB, Borowsky R, Tabin CJ. 2009. A novel role for Mc1r in the parallel evolution of depigmentation in independent populations of the cavefish *Astyanax mexicanus*. *PLoS Genet*. 5:e1000326.
- Gross JB, Protas M, Conrad M, Scheid PE, Vidal O, Jeffery WR, Borowsky R, Tabin CJ. 2008. Synteny and candidate gene prediction using an anchored linkage map of *Astyanax mexicanus*. *Proc Natl Acad Sci U S A*. 105:20106–20111.
- Hall TA. 1999. BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. *Nucleic Acids Symp Ser*. 41:95–98.
- Hausdorf B, Wilkens H, Strecker U. 2011. Population genetic patterns revealed by microsatellite data challenge the mitochondrial DNA based taxonomy of *Astyanax* in Mexico (Characidae, Teleostei). *Mol Phylogenet Evol*. 60:89–97.
- Helyar SJ, Hemmer-Hansen J, Bekkevold D, et al. (11 co-authors). 2011. Application of SNPs for population genetics of nonmodel organisms: new opportunities and challenges. *Mol Ecol Resour*. 11: 123–136.
- Hess JE, Matala AP, Narum SR. 2011. Comparison of SNPs and microsatellites for fine-scale application of genetic stock identification of Chinook salmon in the Columbia River Basin. *Mol Ecol Resour*. 11: 137–149.
- Hoekstra HE. 2006. Genetics, development and evolution of adaptive pigmentation in vertebrates. *Heredity* 97:222–234.
- Hohenlohe PA, Bassham S, Currey M, Cresko WA. 2012. Extensive linkage disequilibrium and parallel adaptive divergence across threespine stickleback genomes. *Philos Trans R Soc B Biol Sci*. 367: 395–408.
- Hohenlohe PA, Bassham S, Etter PD, Stiffler N, Johnson EA, Cresko WA. 2010. Population genomics of parallel adaptation in threespine stickleback using sequenced RAD tags. *PLoS Genet*. 6:e1000862.
- Hohenlohe PA, Phillips PC, Cresko WA. 2010. Using population genomics to detect selection in natural populations: key concepts and methodological considerations. *Int J Plant Sci*. 171:1059–1071.
- Holsinger KE, Weir BS. 2009. Fundamental concepts in genetics genetics in geographically structured populations: defining, estimating and interpreting F_{ST}. *Nat Rev Genet*. 10:639–650.
- Jeffery WR. 2010. Pleiotropy and eye degeneration in cavefish. *Heredity* 105:495–496.
- Jones FC, Grabherr MG, Chan YF, et al. (33 co-authors). 2012. The genomic basis of adaptive evolution in threespine sticklebacks. *Nature* 484:55–61.
- Joron M, Frezal L, Jones RT, et al. (23 co-authors). 2011. Chromosomal rearrangements maintain a polymorphic supergene controlling butterfly mimicry. *Nature* 477:203–206.
- Kalinowski ST. 2005. HP-RARE 1.0: a computer program for performing rarefaction on measures of allelic richness. *Mol Ecol Notes*. 5:187–189.
- Kao CH, Zeng ZB, Teasdale RD. 1999. Multiple interval mapping for quantitative trait loci. *Genetics* 152:1203–1216.
- Korol AB, Ronin YI, Nevo E, Hayes PM. 1998. Multi-interval mapping of correlated trait complexes. *Heredity* 80:273–284.
- Lander ES, Botstein D. 1989. Mapping Mendelian factors underlying quantitative traits using Rflp linkage maps. *Genetics* 121:185–199; erratum in *Genetics* 1994;136:705.
- Lewontin RC, Krakauer J. 1973. Distribution of gene frequency as a test of theory of selective neutrality of polymorphisms. *Genetics* 74: 175–195.
- Luikart G, England PR, Tallmon D, Jordan S, Taberlet P. 2003. The power and promise of population genomics: from genotyping to genome typing. *Nat Rev Genet*. 4:981–994.
- Mitchell RW, Russell WH, Elliott WR. 1977. Mexican eyeless characin fishes, genus *Astyanax*: environment, distribution, and evolution. Lubbock (TX): Texas Tech Press.
- Murakami YI, Yatabe Y, Sakaguchi T, et al. (17 co-authors). 2007. c-Maf expression in angioimmunoblastic T-cell lymphoma. *Am J Surg Pathol*. 31:1695–1702.
- Nachman MW, Hoekstra HE, D'Agostino SL. 2003. The genetic basis of adaptive melanism in pocket mice. *Proc Natl Acad Sci U S A*. 100: 5268–5273.
- Nadeau NJ, Jiggins CD. 2010. A golden age for evolutionary genetics? Genomic studies of adaptation in natural populations. *Trends Genet*. 26:484–492.

- Narum SR, Hess JE. 2011. Comparison of F(ST) outlier tests for SNP loci under selection. *Mol Ecol Resour.* 11(1 Suppl), 184–194.
- Nei M. 1973. Analysis of gene diversity in subdivided populations. *Proc Natl Acad Sci U S A.* 70:3321–3323.
- Nei M. 1977. F-statistics and analysis of gene diversity in subdivided populations. *Ann Hum Genet.* 41:225–233.
- Nei M, Roychoudhury AK. 1974. Sampling variances of heterozygosity and genetic distance. *Genetics* 76:379–390.
- Noor MA, Feder JL. 2006. Speciation genetics: evolving approaches. *Nat Rev Genet.* 7:851–861.
- Orgogozo V, Stern DL. 2009. How different are recently diverged species?: more than 150 phenotypic differences have been reported for the *D. melanogaster* species subgroup. *Fly (Austin)* 3:117.
- Ornelas-García CP, Dominguez-Dominguez O, Doadrio I. 2008. Evolutionary history of the fish genus *Astyanax* Baird & Girard (1854) (Actinopterygii, Characidae) in Mesoamerica reveals multiple morphological homoplasies. *BMC Evol Biol.* 8:340.
- Peichel CL, Nereng KS, Ohgi KA, Cole BL, Colosimo PF, Buerkle CA, Schluter D, Kingsley DM. 2001. The genetic architecture of divergence between threespine stickleback species. *Nature* 414:901–905.
- Peterson BK, Weber JN, Kay EH, Fisher HS, Hoekstra HE. 2012. Double digest RADseq: an inexpensive method for de novo SNP discovery and genotyping in model and non-model species. *PLoS One* 7: e37135.
- Pigeon D, Chouinard A, Bernatchez L. 1997. Multiple modes of speciation involved in the parallel evolution of sympatric morphotypes of lake whitefish (*Coregonus clupeaformis*, Salmonidae). *Evolution* 51: 196–205.
- Pottin K, Hinaux H, Retaux S. 2011. Restoring eye size in *Astyanax mexicanus* blind cavefish embryos through modulation of the Shh and Fgf8 forebrain organising centres. *Development* 138:2467–2476.
- Poulson TL. 1964. Animals in aquatic environments, animals in caves. In: Dill DB, editor. *Handbook of physiology*. Washington (DC): American Physiological Society. p. 749–771.
- Protas M, Conrad M, Gross JB, Tabin C, Borowsky R. 2007. Regressive evolution in the Mexican cave tetra, *Astyanax mexicanus*. *Curr Biol.* 17:452–454.
- Protas M, Tabansky I, Conrad M, Gross JB, Vidal O, Tabin CJ, Borowsky R. 2008. Multi-trait evolution in a cave fish, *Astyanax mexicanus*. *Evol Dev.* 10:196–209.
- Protas ME, Hersey C, Kochanek D, Zhou Y, Wilkens H, Jeffery WR, Zon LI, Borowsky R, Tabin CJ. 2006. Genetic analysis of cavefish reveals molecular convergence in the evolution of albinism. *Nat Genet.* 38:107–111.
- Renaut S, Maillat N, Normandeau E, Sauvage C, Derome N, Rogers SM, Bernatchez L. 2012. Genome-wide patterns of divergence during speciation: the lake whitefish case study. *Philos Trans R Soc B Biol Sci.* 367:354–363.
- Renaut S, Nolte AW, Rogers SM, Derome N, Bernatchez L. 2011. SNP signatures of selection on standing genetic variation and their association with adaptive phenotypes along gradients of ecological speciation in lake whitefish species pairs (*Coregonus* spp.). *Mol Ecol.* 20:545–559.
- Retaux S, Pottin K, Alunni A. 2008. Shh and forebrain evolution in the blind cavefish *Astyanax mexicanus*. *Biol Cell.* 100:139–147.
- Rogers SM, Bernatchez L. 2005. Integrating QTL mapping and genome scans towards the characterization of candidate loci under parallel selection in the lake whitefish (*Coregonus clupeaformis*). *Mol Ecol.* 14: 351–361.
- Rogers SM, Bernatchez L. 2007. The genetic architecture of ecological speciation and the association with signatures of selection in natural lake whitefish (*Coregonus* sp. Salmonidae) species pairs. *Mol Biol Evol.* 24: 1423–1438.
- Rogers SM, Tamkee P, Summers B, Balabhadra S, Marks M, Kingsley DM, Schluter D. 2012. Genetic signature of adaptive peak shift in threespine stickleback. *Evolution* 66:2439–2450.
- Sabeti PC, Reich DE, Higgins JM, et al. (17 co-authors). 2002. Detecting recent positive selection in the human genome from haplotype structure. *Nature* 419:832–837.
- Scaglione D, Acquadro A, Portis E, Tirone M, Knapp SJ, Lanteri S. 2012. RAD tag sequencing as a source of SNP markers in *Cynara cardunculus* L. *BMC Genomics* 13:3.
- Scheet P, Stephens M. 2006. A fast and flexible statistical model for large-scale population genotype data: applications to inferring missing genotypes and haplotypic phase. *Am J Hum Genet.* 78:629–644.
- Schluter D, Conte GL. 2009. Genetics and ecological speciation. *Proc Natl Acad Sci U S A.* 106:9955–9962.
- Seeb JE, Carvalho G, Hauser L, Naish K, Roberts S, Seeb LW. 2011. Single-nucleotide polymorphism (SNP) discovery and applications of SNP genotyping in nonmodel organisms. *Mol Ecol Resour.* 11:1–8.
- Seeb JE, Pascal CE, Grau ED, Seeb LW, Templin WD, Harkins T, Roberts SB. 2011. Transcriptome sequencing and high-resolution melt analysis advance single nucleotide polymorphism discovery in duplicated salmonids. *Mol Ecol Resour.* 11:335–348.
- Seeb LW, Templin WD, Sato S, Abe S, Warheit K, Park JY, Seeb JE. 2011. Single nucleotide polymorphisms across a species' range: implications for conservation studies of Pacific salmon. *Mol Ecol Resour.* 11: 195–217.
- St-Cyr J, Derome N, Bernatchez L. 2008. The transcriptomics of life-history trade-offs in whitefish species pairs (*Coregonus* sp.). *Mol Ecol.* 17:1850–1870.
- Stapley J, Reger J, Feulner PG, Smadja C, Galindo J, Ekblom R, Bennison C, Ball AD, Beckerman AP, Slate J. 2010. Adaptation genomics: the next generation. *Trends Ecol Evol.* 25:705–712.
- Stern DL, Orgogozo V. 2009. Is genetic evolution predictable? *Science* 323:746–751.
- Stinchcombe JR, Hoekstra HE. 2008. Combining population genomics and quantitative genetics: finding the genes underlying ecologically important traits. *Heredity* 100:158–170.
- Strecker U, Bernatchez L, Wilkens H. 2003. Genetic divergence between cave and surface populations of *Astyanax* in Mexico (Characidae, Teleostei). *Mol Ecol.* 12:699–710.
- Strecker U, Faundez VH, Wilkens H. 2004. Phylogeography of surface and cave *Astyanax* (Teleostei) from Central and North America based on cytochrome *b* sequence data. *Mol Phylogenet Evol.* 33: 469–481.
- Strecker U, Hausdorf B, Wilkens H. 2012. Parallel speciation in *Astyanax* cave fish (Teleostei) in Northern Mexico. *Mol Phylogenet Evol.* 62: 62–70.
- Strickler AG, Yamamoto Y, Jeffery WR. 2001. Early and late changes in Pax6 expression accompany eye degeneration during cavefish development. *Dev Genes Evol.* 211:138–144.
- Tang K, Thornton KR, Stoneking M. 2007. A new approach for using genome scans to detect recent positive selection in the human genome. *PLoS Biol.* 5:1587–1602.
- Teotonio H, Chelo IM, Bradic M, Rose MR, Long AD. 2009. Experimental evolution reveals natural selection on standing genetic variation. *Nat Genet.* 41:251–257.
- Teotonio H, Rose MR. 2000. Variation in the reversibility of evolution. *Nature* 408:463–466.
- Travisano M, Mongold JA, Bennett AF, Lenski RE. 1995. Experimental tests of the roles of adaptation, chance, and history in evolution. *Science* 267:87–90.
- Via S, West J. 2008. The genetic mosaic suggests a new role for hitchhiking in ecological speciation. *Mol Ecol.* 17:4334–4345.
- Wark AR, Mills MG, Dang LH, et al. (12 co-authors). 2012. Genetic architecture of variation in the lateral line sensory system of three-spine sticklebacks. *G3 (Bethesda)* 2:1047–1056.
- Weir BS, Cockerham CC. 1984. Estimating F-statistics for the analysis of population-structure. *Evolution* 38:1358–1370.
- Wigginton JE, Cutler DJ, Abecasis GR. 2005. A note on exact tests of Hardy-Weinberg equilibrium. *Am J Hum Genet.* 76:887–893.
- Wilkens H. 1971. Genetic interpretation of regressive evolutionary processes—studies on hybrid eyes of 2 *Astyanax* cave populations (Characidae, Pisces). *Evolution* 25:530–544.
- Wilkens H. 1988. Evolution and genetics of Epigeal and Cave *Astyanax-Fasciatus* (Characidae, Pisces)—support for the neutral mutation theory. *Evol Biol.* 23:271–367.

- Wilkens H. 2010a. Genes, modules and the evolution of cave fish. *Heredity* 105:413–422.
- Wilkens H. 2010b. Variability and loss of functionless traits in cave animals. Reply to Jeffery (2010). *Heredity* 106:707–708.
- Wolf JB, Lindell J, Backstrom N. 2010. Speciation genetics: current status and evolving approaches. *Philos Trans R Soc Lond B Biol Sci*. 365:1717–1733.
- Yatabe Y, Kane NC, Scotti-Saintagne C, Rieseberg LH. 2007. Rampant gene exchange across a strong reproductive barrier between the annual sunflowers, *Helianthus annuus* and *H. petiolaris*. *Genetics* 175:1883–1893.
- Yoshizawa M, Goricki S, Soares D, Jeffery WR. 2010. Evolution of a behavioral shift mediated by superficial neuromasts helps cavefish find food in darkness. *Curr Biol*. 20:1631–1636.
- Yoshizawa M, Yamamoto Y, O'Quin KE, Jeffery WR. 2012. Evolution of an adaptive behavior and its sensory receptors promotes eye regression in blind cavefish. *BMC Biol*. 10:108.