Novel candidate genes underlying extreme trophic specialization in Caribbean pupfishes

Joseph A. McGirr¹, Christopher H. Martin¹.

¹Department of Biology, University of North Carolina at Chapel Hill, 120 South Rd., Chapel Hill, NC 27514-3280

Running Title: Genetic basis of trophic specialization in Caribbean pupfishes

Keywords: candidate gene, trophic morphology, fitness landscape, adaptive radiation, de novo mutation, standing genetic variation, ecological novelty, selective sweep

Correspondence: Christopher H Martin. Department of Biology, University of North Carolina at Chapel Hill, Campus Box 3280, 120 South Rd., NC, 27599, USA

Email: chmartin@unc.edu

Word count abstract: 210 words; Word count main text: 7,586, Table 1, Figures 1-6, Table S1-S2, Figures S1-S5

© The Author 2016. Published by Oxford University Press on behalf of the Society for Molecular Biology and Evolution. All rights reserved. For permissions, please e-mail: journals.permissions@oup.com
Abstract

The genetic changes responsible for evolutionary transitions from generalist to specialist phenotypes are poorly understood. Here we examine the genetic basis of craniofacial traits enabling novel trophic specialization in a sympatric radiation of *Cyprinodon* pupfishes endemic to San Salvador Island, Bahamas. This recent radiation consists of a generalist species and two novel specialists: a small-jawed ‘snail-eater’ and a large-jawed ‘scale-eater.’ We genotyped 12 million SNPs by whole-genome resequencing of 37 individuals of all three species from nine populations and integrated genome-wide divergence scans with association mapping to identify divergent regions containing putatively causal SNPs affecting jaw size – the most rapidly diversifying trait in this radiation. A mere 22 fixed variants accompanied extreme ecological divergence between generalist and scale-eater species. We identified 31 regions (20kb) containing variants fixed between specialists that were significantly associated with variation in jaw size and contained 11 genes annotated for skeletal system effects and 18 novel candidate genes never previously associated with craniofacial phenotypes. Six of these 31 regions showed robust signs of hard selective sweeps after accounting for demographic history. Our data are consistent with predictions based on quantitative genetic models of adaptation, suggesting the effect sizes of regions influencing jaw phenotypes are positively correlated with distance between fitness peaks on a complex adaptive landscape.
Introduction

Identifying genetic changes underlying phenotypic diversity is necessary to understand how these changes drive adaptation and speciation (Coyne and Orr 2004; Moczek 2008; Byers et al. 2016; but see Rauscher and Delph 2015). Adaptive radiations showcase the world’s most dramatic instances of rapid ecological divergence (Turner 1976; Schluter 2000; Seehausen 2006; Losos and Ricklefs 2009; Lamichhaney et al. 2016) making them ideal for investigating the genetic basis of traits influencing novel niche use. Characterizing divergent regions underlying adaptation will address several longstanding questions in evolutionary genomics, such as: how many differentiated regions do we find between closely related species? Is novel trophic specialization driven by selective sweeps? Does the effect size of loci contributing to phenotypic divergence depend on the distance between fitness peaks across an adaptive landscape? (Hermisson and Pennings 2005; Orr 2005; Noor and Feder 2006, Barrett and Schluter 2008; Jensen 2014; Dittmar et al. 2016; Hoban et al. 2016). Genomic divergence scans measuring relative genetic differentiation and genome-wide association mapping are two strategies used to detect candidate gene regions responsible for species differences (Visscher 2012; Gompert et al. 2012; Pallares et al. 2014; Comeault et al. 2014; Puzey 2015; Irwin et al. 2016; Chaves et al. 2016). Together these powerful tools can be used to discover genomic regions that are both highly diverged between species and associated with ecologically important traits (Li et al. 2011; Xia et al. 2013; Byers et al. 2016).

A number of recent genome-wide $F_{st}$ scans comparing closely related species pairs have located small regions (typically < 200kb) that are highly differentiated relative to the rest of the genome (Carneiro et al. 2014; Soria-Carrasco 2014; Poelstra et al. 2014; Malinsky et al. 2015; Lamichhaney et al. 2015), suggesting these regions are responsible for species-specific phenotypes. Recent literature has emphasized the importance of estimating $F_{st}$ alongside within-population nucleotide diversity ($\pi$) and between-population divergence ($D_{xy}$) in order to more accurately interpret the evolutionary significance of genetically differentiated regions (Nachman and Payseur 2012; Cruickshank and Hahn 2014; Irwin 2016). Importantly, any reduction of within-population diversity will necessarily inflate estimates of $F_{st}$ because it is a relative measure of differentiation (reviewed in Noor and Bennett 2009; Nachman and Payseur 2012; Cruickshank and Hahn 2014). Therefore, $F_{st}$ interpretations are heavily dependent on the
interplay of forces acting to reduce within-population diversity, including selective sweeps, purifying selection, background selection, and low recombination rates (Noor and Bennett 2009; Cruickshank and Hahn 2014). Estimating between-population divergence at loci with high $F_{st}$ and low within-population diversity can help distinguish between these possibilities because nucleotide divergence between species increases at loci under different selective regimes (Nachman and Payseur 2012; Cruickshank and Hahn 2014; Irwin 2016). However, between-population divergence can also be influenced by patterns of hitchhiking and background selection (Cruickshank and Hahn 2014). Selection statistics comparing the distribution of allele frequencies across segregating sites can also help determine if reduced diversity at a locus is due to selective sweeps, in which selection has increased the frequency of a single (hard sweep) or multiple haplotypes (soft sweep) (Maynard Smith and Haigh 1974, Tajima 1989; Hermisson and Pennings, 2005; Pavlidis et al. 2013; Jensen 2014). Statistics that rely on the distribution of allele frequencies within and between populations should be interpreted in the context of their demographic history (Galtier et al. 2000; Andolfatto 2001; Nielsen 2005; Hoban et al. 2016). This can be achieved by inferring changes in ancestral population sizes and using these estimates to model a demography-corrected neutral distribution of allele frequencies (Pavlidis et al. 2013; Schiffels & Durbin 2014). Combining $F_{st}$, $\pi$, $D_{xy}$, and selective sweep statistics can reveal functionally diverged regions of the genome; however, these statistics alone are insufficient to determine how such regions might affect phenotypic differences between species.

Genome-wide association studies expand on divergence scans by identifying regions that are directly associated with phenotypic differences between species. The simplest approach involves estimating associations between SNPs and quantitative traits by fitting a linear regression of phenotype on allele frequency (Purcell et al. 2007; Visscher 2012), while more advanced methods account for population structure and estimate the effect size of SNPs associated with traits (Price et al. 2006; Kang et al. 2010; Zhou and Stevens 2012; Zhou et al. 2013). Accounting for population structure can help filter out false positive associations, but may also filter out true associations (Marchini 2004; Zhao et al. 2011). Thus, we implemented both types of association models alongside genome divergence scans. We used this mixed strategy to identify candidate SNPs affecting novel ecological traits in an excellent system for examining rapid adaptive diversification.
Three sympatric *Cyprinodon* species inhabit the hypersaline lakes of San Salvador Island, Bahamas, and radiated within the past 10,000 years based on the most recent drying of these lakes (Mylroie and Hagey 1995; Turner et al. 2008). A generalist species, *C. variegatus*, feeds primarily on algae and detritus, a diet representative of all allopatric Cyprinodontidae (Martin and Wainwright 2011). The first of two specialist species, the ‘snail-eater’ *C. brontotheroides*, expanded its diet to include more gastropods and ostracods (Martin and Wainwright 2013a). Snail-eater oral jaws are smaller with a larger in-lever to out-lever ratio compared to the generalist, increasing mechanical advantage for biting (Martin and Wainwright 2013a). The snail-eater is also defined by a prominent protruding nasal region that may be used for leverage while crushing hard-shelled prey (Martin and Wainwright 2013a, 2013b). The second sympatric specialist, the ‘scale-eater’ *C. desquamator*, expanded its diet to include scales removed from other species during quick strikes. Scale-eaters have greatly enlarged jaws with a smaller in-lever to out-lever ratio, large adductor muscles, and an elongated body compared to the generalist and snail-eater species (Martin and Wainwright 2013a). Phylogenetic analyses of outgroup *Cyprinodon* species and surveys of pupfish populations on neighboring Bahamian islands confirm that scale-eating and snail-eating niches are entirely unique to *C. desquamator* and *C. brontotheroides*, respectively, and that each species is endemic to hypersaline lakes on San Salvador Island, providing strong support that these specialists diverged from a generalist common ancestor during recent adaptive radiation (Martin and Wainwright 2011; Martin 2016a).

Adaptive landscapes describe the relative fitness of various trait (or allelic) combinations given a particular environment – where adaptive peaks represent optimal combinations and adaptive valleys represent unfit combinations (Wright 1932; Wright 1988; Schluter 2000). If the scale-eater and snail-eater specialists rapidly ascended to novel adaptive peaks within the past 10,000 years, then we should expect to see high rates of morphological diversification in traits associated with trophic specialization. Indeed, San Salvador *Cyprinodon* pupfishes exhibit morphological diversification rates up to 51 times faster than other Cyprinodontidae clades, with jaw size undergoing the most rapid diversification (Martin and Wainwright 2011; Martin 2016a). The San Salvador pupfish system is one of the few examples of a multi-peak adaptive landscape measured for three species (Martin and Wainwright 2013c; Martin 2016b), presenting an excellent opportunity to test mathematical models of adaptation. This landscape was estimated using F2 hybrids generated from F1 hybrid intercrosses and backcrosses to all three species. This
produced a continuum of phenotypes that were used to estimate relationships between fitness and phenotypic resemblance to parental types. The fitness optima for generalist and snail-eater phenotypes were separated by a small fitness valley, while the phenotypic optimum of the scale-eater presumably exists outside of the range of phenotypic variation tested in the F$_2$ population (Fig. 1) (Martin and Wainwright 2013c). Although this landscape did not measure a scale-eater fitness optimum, it does show that the phenotypic distance is greater between the generalist fitness peak and the fitness valley surrounding hybrid phenotypes most resembling the scale-eaters than between the generalist and snail-eater fitness peaks (Fig. S1A). This greater phenotypic distance is primarily due to the large jaws of scale-eaters (Fig. S1B). Orr’s extension of Fisher’s geometric model predicts that \textit{de novo} mutations with a large effect on phenotypic variation are more likely to be fixed during adaptation toward distant phenotypic optima than nearby optima (Orr 1998; Orr 2005). Based on this model, we predict more large-effect variants mediated the transition from generalist to scale-eater due to the greater phenotypic distance across the fitness valley separating these species.

Here we focus on identifying loci associated with variation in jaw morphology within this radiation due to the strikingly rapid divergence of this trait that has clear ecological fitness consequences. We identified 12 million SNPs from 37 genomes sequenced to 7x coverage across nine populations of all three species on San Salvador Island. We discovered novel candidate genes associated with jaw size along with evidence supporting the role of large-effect alleles in crossing between distant phenotypic optima.

Results

\textit{Estimating phenotypic distances}

Orr’s extension of Fisher’s geometric model predicts that \textit{de novo} mutations with a large effect on phenotypic variation are more likely to be fixed during adaptation toward distant phenotypic optima than nearby optima (Orr 1998; Orr 2005). To test this prediction, we measured the phenotypic distance between hybrids used to estimate the multi-peaked adaptive landscape for San Salvador pupfishes (dataset published in Dryad repository for Martin 2016b and originally used for Martin and Wainwright 2013b). These hybrids were measured for 16 morphological
traits. We visualized the distance between fitness peaks along the first two principal component axes of phenotypic variation and for the key trait of upper jaw length and found that the distance between phenotypic optima is greater between the generalist fitness peak and the fitness valley surrounding hybrid phenotypes most resembling the scale-eaters than between the generalist fitness peak and the neighboring higher fitness peak corresponding to hybrids resembling the snail-eater (Fig S1).

**Population structure and genome scans**

Principal component analysis revealed population structure at the level of species and individual lake population, with the top two principal components together explaining 9.44% of the genetic variation (Fig. 2A). The axes show two distinct clusters of scale-eaters: smaller-jawed individuals from Osprey Lake, Great Lake, and Oyster Pond and the largest-jawed individuals from Crescent Pond and Little Lake. Genome-wide mean estimates of within-species diversity ($\pi$: generalist = 0.00402, snail-eater = 0.00321, scale-eater = 0.00324) and mean between-population divergence ($D_{xy}$: generalist × snail-eater = 0.000166, generalist × scale-eater = 0.000169, scale-eater × snail-eater = 0.000167) were similar for all comparisons, revealing that most variants were shared among species. The similarity between $D_{xy}$ among species suggests that divergence from a generalist ancestor likely occurred near the same time for both specialists.

We used genome-wide $F_{st}$ scans to identify fixed regions associated with each species across nine lake populations on San Salvador and one neighboring island. Very few fixed sites corresponded to the discrete species-specific phenotypes across populations. We found 6,673 sites fixed between specialists, 123 sites fixed between generalist and snail-eater species, and a mere 22 sites fixed between generalist and scale-eater species (Fig. 3, Table S1). Eight of these 22 fixed SNPs were also fixed between specialists. Genome-wide mean $F_{st}$ estimates for each comparison (scale-eater/snail-eater = 0.143, generalist/snail-eater = 0.080, generalist/scale-eater = 0.089) were comparable to previous estimates based on microsatellites (Turner et al. 2008) and RADseq derived SNPs (Martin and Feinstein 2014).

**Association Mapping**
We initially used quantitative trait association mapping in PLINK to identify SNPs associated with jaw length variation among individuals without correcting for population structure, which would remove true positives in addition to false positives. This uncorrected PLINK analysis identified 9,214 variants associated with jaw size variation between the generalist, scale-eater, and snail-eater species ($P < 4.0 \times 10^{-9}$ (Fig. 4)). Of these variants, 556 were fixed in at least one pairwise species comparison. 555 of these SNPs were fixed between the two specialists; nine were fixed between the generalist and scale-eater; zero were fixed between the generalist and snail-eater.

Out of the nine PLINK outlier SNPs significantly associated with jaw size and fixed between the generalist and scale-eater, six were located across four different gene regions (magi3, cabp2, lingo1, and pigr) and three unannotated regions (Table S1). Out of the top 20 outliers fixed between the snail-eater and scale-eater, 13 were located across five different gene regions (galr2, gmds, soga3, tmem30a, plxna2) and seven were located across three unannotated regions (Table 1). Combined, PLINK identified 14 divergent regions (nine genic and five unannotated) significantly associated with jaw size and fixed in scale-eaters.

We further assessed the significance of jaw size associations for these top candidate regions containing fixed SNPs by correcting for population structure using two methods. First, we used PLINK to include the top two principal components as covariates in the model (Price et al. 2006; Hunter et al. 2007). This stringent analysis did not identify any SNPs associated with jaw size at our highly conservative Bonferroni-corrected significance threshold (Table S2). However, this likely reflects the fact that the first principal component is significantly correlated with jaw size ($P = 0.0013$, Fig S2). Next, we performed independent association mapping with GEMMA, which corrects for population structure by incorporating a genetic relatedness matrix as a covariate in a Bayesian sparse linear mixed model (Zhou et al. 2013). This is a more reliable correction for population structure because the relatedness matrix accounts for pairwise relatedness between individuals; whereas principal components only capture broad linear axes of population structure (Novembre and Stephens 2008; Kang et al. 2010). Because the uncorrected PLINK analysis likely identified a subset of true associations in addition to false positives, we chose to combine uncorrected PLINK results with our corrected GEMMA results in order to evaluate the significance of regions associated with jaw size (following Zhao et al. 2011). We
identified 31 regions (20kb each) implicated by uncorrected PLINK analyses that also showed association with jaw size after correcting for population structure in GEMMA (Fig. 4). We assessed the significance of associations based on PIP (posterior inclusion probability) parameters which report the proportion of iterations in which a SNP is estimated to have a non-zero effect on phenotypic variation (effect size $\beta \neq 0$). These 31 regions showed robust association across 10 independent Markov Chain Monte Carlo (MCMC) runs. We used $\beta$ effect size parameters to assess whether regions contributed to larger jaw size ($+\beta$) or decreasing jaw size ($-\beta$) and found slightly more candidate regions increased (16) than decreased jaw size (13).

All 31 regions contained variants fixed between specialists and showed outlier median parameter values in the 99th percentile for PIP estimated across all SNPs included in the analysis (following Gompert et al. 2012), indicating an association with jaw size after accounting for population structure (Table 1). These regions span 25 scaffolds and contain 29 genes, 11 of which are annotated for skeletal system functions (NCBI Cyprinodon release 100). The top ten regions with the highest PIP implicated three of the same genes identified by PLINK (galr2, gmds, soga3) as well as three additional genes (fam49b, znf664, and pard3) and one large (60kb) unannotated region. The unannotated region and galr2 showed the highest $\beta$ values in the direction of large jaws, while the region containing gmds showed the highest $\beta$ values in the direction of smaller jaws (Figs. 5-6). Encouragingly, galr2 is within a QTL explaining 15% of the variation in jaw size in an F2 intercross between specialist species (Martin et al. accepted).

**History of Selection and Demography**

To determine whether candidate regions were potentially subject to hard selective sweeps, we interrogated the site frequency spectrum using SweeD (Pavlidis et al. 2013) and Tajima’s D (Tajima 1989). Tajima’s D compares observed nucleotide diversity to diversity under a null model assuming genetic drift, where negative values indicate a reduction in diversity across segregating sites (Tajima 1989). SweeD scans across non-overlapping windows to calculate a composite likelihood ratio (CLR), comparing a model assuming selection to a null model calibrated by the observed site frequency spectrum across the entire scaffold. Both of these statistics infer selection based on the shape of the site frequency spectrum, which can also be influenced by changes in effective population size over time (Galtier et al. 2000; Nielsen 2005).
We therefore used the Multiple Sequentially Markovian Coalescent (MSMC) (Schiffels and Durbin 2014) to infer historical population size in all three species, and applied these estimates to analytically calculate the expected neutral site frequency spectrum in SweeD. MSMC results suggest that that the population size of all three species has been decreasing across at least the last 10,000 years (~20,000 generations) (Fig. S3). This model suggests a population decrease that is consistent with changes in sea level during the last glacial maximum when saline lakes on San Salvador Island first appeared (Mylroie and Hagey 1995; Turner et al. 2008). We first looked for signatures of hard sweeps in both specialist populations by analyzing the site frequency spectrum without demographic assumptions. Next, we calculated the expected neutral site frequency spectrum assuming a population decline as suggested by our demographic model. Windows that showed CLRs above the 95th percentile across their respective scaffolds in this second analysis were interpreted as regions that recently experienced a hard sweep.

Out of our 31 candidate regions affecting jaw size, six were consistent with hard selective sweeps. One candidate region was excluded from these analyses because it fell within a small scaffold that could not be used to sample an adequate background distribution of heterogeneity. All six regions also showed negative estimates of Tajima’s D (Figs. 5-6). The 60kb unannotated region associated with large jaws showed the strongest signatures of selection, followed by a 40kb region associated with small jaws. This smaller region contains four genes all annotated for skeletal system effects (hint1, lyrm7, dyn2li1, abcg5) (Fig. 6). Five of the six regions that experienced strong selection also show reduced within-population diversity (π) in the specialist species and increased between-population divergence (D_{xy}) when compared to generalists (Figs. 5-6). This pattern may suggest that strong selection on a beneficial allele reduced diversity within specialists across candidate regions. Importantly, low diversity in these regions is not shared between specialists and generalists, possibly suggesting that selection unique to each specialist was responsible for reduced diversity. This combined evidence implicates divergent regions influencing jaw morphology that experienced strong selection within the specialist linages. Finally, we did not find evidence for hard sweeps in 25 of our 31 candidate regions, possibly suggesting that multiple haplotypes were swept to fixation (Herisson and Pennings 2005; Jensen 2014).
**More large-effect alleles were required to evolve large jaws than small jaws**

Based on differences in the phenotypic distance across fitness valleys separating each specialist species from its putative generalist ancestor (Fig. 1), we predicted to find more large-effect SNPs associated with large jaws than small jaws. There are two lines of evidence supporting this prediction. First, we directly compared positive and negative effect sizes for regions associated with small jaws (-β) and large jaws (+β). Our β outlier threshold included 83 of the regions most strongly associated with jaw size that had the largest effects on jaw size (β > 99.9th percentile). We found more than twice as many outlier SNPs with large effects on increasing jaw size (n = 56) compared to large-effects on decreasing jaw size (n = 27) (Fig. 7). Second, we identified five times fewer SNPs fixed between the generalist and scale-eater (n = 22) than SNPs fixed between the generalist and snail-eater species (n = 123) (Fig. 3), supporting the prediction that SNPs with larger effect sizes should fix faster than SNPs with smaller effects, especially given short divergence times (Griswold 2006; Yeaman and Whitlock 2011).

**Discussion**

Genome-wide divergence scans revealed that the evolution of trophic novelty in two ecological specialists involved surprisingly few genetic variants fixed between species. We determined which of these fixed variants influenced the most rapidly diversifying trait in this radiation – jaw size – using quantitative trait association mapping. We uncovered 31 candidate regions fixed between species and associated with jaw size after correcting for population structure, with six of these regions showing signs of hard selective sweeps. We used these data to test the prediction that more large-effect variants should affect large jawed scale-eaters than small jawed snail-eaters.

**Genetic Basis of Jaw Size Divergence**

We report 31 divergent candidate regions associated with jaw size among San Salvador *Cyprinodon* pupfish. We identified these regions using 37 genomes sequenced to 7x coverage across nine populations. This is significant because much work on the genetic basis of adaptation
has relied on reduced representation strategies (*i.e.* RADseq, RNAseq) that likely overlook loci contributing to adaptation (Hoban et al. 2016). All 31 regions contained SNPs fixed between specialists that were significant in both association mapping approaches. We searched genes listed under the ‘skeletal system’ ontology in the phenotype database Phenoscape (Mabee et al. 2012; Midford et al. 2013; Manda et al. 2015, Edmunds et al. 2016) finding matches for 11 genes within candidate regions (Table 1). The most strongly associated gene annotated for skeletal effects, *galr2*, is interesting for several reasons. The protein product of *galr2* is a transmembrane galanin receptor with a role in numerous physiological functions (Webling et al. 2012). Galanin, the binding substrate of GALR2, has been shown to facilitate bone formation by increasing the size and proliferation of osteoblasts (McDonald et al. 2007; McGowen et al. 2014). Additionally, the scaffold containing *galr2* overlaps with a moderate effect QTL explaining 15% of the variation in jaw size in an independent F<sub>2</sub> mapping cross between the two specialist pupfishes (Martin et al. accepted), increasing confidence in our association mapping strategy. The gene region most associated with smaller jaws was *gmds*, which is important for tagging cell surface proteins involved in many cellular processes such as cell growth, migration, and apoptosis (Moriwaki et al. 2009). This gene represents a novel candidate for craniofacial effects. We identified four genes annotated for skeletal effects spanning a 40kb region that showed significant association with smaller jaws (*hint1, lyrm7, dyn2li1, abcg5*). Mutations in *lyrm7* have been associated with mitochondrial complex III deficiency, a disorder characterized by skeletal muscle weakness and weak muscle tone (hypotonia) (Invernizzi et al. 2013). Mutations in *dyn2li1*, a gene involved in skeletogenesis and expressed in the cartilage of growth plates, have been shown to cause short rib polydactyly skeletal disorders (Taylor et al. 2015). Thus, our candidate regions are associated with genes involved in bone and skeletal muscle development – the two tissues most differentiated in the external anatomy of San Salvador pupfishes. Finally, we identified eight SNPs fixed between the generalist and scale-eater that were also fixed between specialists, possibly indicating that these regions affect traits in both specialists. However, none of these overlapping SNPs showed significant association with jaw size after correcting for population structure.

*Caveats to our association mapping approach*
The significance of our association mapping results should be interpreted with caution. Our principal component analysis revealed significant population structure associated with four different clusters of jaw sizes across species and between two different clusters of large and short-jawed scale-eaters among lake populations (Fig. 2A), which likely created a bias toward false positive associations implicated by PLINK. Furthermore, when we accounted for this structure by incorporating the first two principal components as covariates in the model, we did not find any SNPs reaching significance at our conservative Bonferroni-corrected level of significance. However, this analysis almost certainly filtered out true associations because the first PC is highly correlated with jaw size. We reassessed the significance of these associations by using GEMMA – a complementary mapping approach that corrects for population structure by incorporating a genetic relatedness matrix into a Bayesian sparse linear mixed model (BSLMM) (Zhao et al. 2013). We used the BSLMM to investigate the genetic architecture of jaw size – a complex polygenic trait (Helms and Schneider 2003; Albertson et al. 2003; Pallares et al. 2014; Porto et al. 2016; Martin et al. accepted). Our PIP estimates for regions associated with jaws size variation suggest that jaw shape is controlled by many loci of relatively small effect (see Comeault et al. 2016 for an example of BSLMMs used for a simple Mendelian color locus; see Gompert et al. 2012, Chaves et al. 2013 for complex traits). Indeed, a linkage mapping analysis of phenotypic diversity in an F\textsuperscript{2} intercross between specialists identified QTL with only moderate effects explaining up to 15% of the variation in jaw size (Martin et al. accepted).

While uncommonly implemented across species, association mapping techniques have proven successful at identifying associations across ‘varieties,’ ‘subspecies,’ and ‘ecotypes’ with greater genetic differentiation (Fournier-Level et al. 2011; Zhao et al. 2011; Pallares et al. 2014) or minimal divergence similar to that of San Salvador pupfishes (Comeault et al. 2014). Association mapping within populations may result in spurious associations due to background population structure (Kang et al. 2010; Marchini et al. 2011), but our sampling of multiple, relatively isolated populations may have provided greater resolution of candidate regions due to sampling a diversity of genetic backgrounds. We do not expect false associations due to sequencing error biases because mean coverage across candidate SNPs mirrored coverage across individuals (range: 4.9x – 6.6x). It is possible that our methods excluded significant SNPs as false negatives. We examined the position of all 22 SNPs fixed between the generalist and scale-eater for gene annotations (Table S1), finding four within the gene \textit{coll11a1}. None of these four
SNPs showed a significant association with jaw size in either mapping approach; however, *col11a1* has been associated with jaw skeleton phenotypes in humans (Hufnagel et al. 2014). It is unclear whether *col11a1* variants influence jaw divergence in pupfish but escaped detection in both mapping analyses.

**Variants with relatively large effects drive divergence across a large fitness valley**

Orr’s extension of Fisher’s geometric model of adaptation predicts that *de novo* mutations with a large effect on phenotypic variation are more likely to be fixed during adaptation toward distant phenotypic optima than nearby optima (Orr 1998; Orr 2005). This distribution of effect sizes for mutations fixed during adaptation has been supported by QTL mapping analyses in multiple systems (Baxter et al. 2009; Rogers et al. 2012; Conte et al. 2015; Martin et al. accepted). We show that the phenotypic distance across the fitness valley is larger between the generalist and large-jawed scale-eater species than between the generalist and small-jawed snail-eater species (Figs. 1, S1) (Martin and Wainwright 2013c; Martin 2016b). Based on this adaptive landscape, we predicted more large-effect variants associated with large jaws than small jaws. Adaptive landscapes are not static, and the distance between fitness optima may have fluctuated over the past 10,000 years of divergence in this system (Merrell 1994; Hansen et al. 2008). However, scale-eater prey has been available since the initial colonization of generalists on San Salvador. Furthermore, the availability of hard-shelled prey (ostracods, gastropods), is likely not substantially depleted in these lakes due to the rarity of snail-eater specialists (<5% of the total pupfish population) and high productivity of eutrophic saline lakes (Martin and Wainwright 2013a).

Although Orr’s model assumes a single population and ignores standing genetic variation (Orr, 1998; Dittmar et al. 2016) and thus may not apply here, we present two lines of evidence supporting the model in this system. First, we found twice as many outlier regions with the largest effect sizes associated with larger jaws than smaller jaws (Fig. 7). Second, there are more than five times as many fixed SNPs between the generalist and snail-eater than between the generalist and scale-eater (Fig. 3). Divergent demographic histories could account for this pattern; however, similar changes in population size over 20,000 generations for each species (Fig. S3), combined with evidence for gene flow between species in sympatry (Martin and
Feinstein 2014), suggest that this is not the case. Large-effect variants are predicted to become fixed between species more quickly than variants with smaller effects in the presence of gene flow, especially when divergence time is short (Griswold 2006; Yeaman and Whitlock 2011). This difference suggests that more large-effect alleles influencing jaw size were necessary to evolve the specialized scale-eating phenotype, while smaller jaw phenotypes may result from more alleles with small to moderate effect sizes. Further support for this prediction within the San Salvador pupfish system comes from a complementary linkage mapping study which found moderate effect QTL explaining up to 15% of variance in jaw size within an F2 intercross between both specialists but no significant QTL with effects on nasal protrusion — a trait unique to the snail-eater species (Martin et al. accepted). Overall these data agree with Orr’s model, suggesting that large effect loci are used to cross larger distances between fitness optima (Orr 1998; Orr 2005).

**Strong selection on candidate regions**

We reasoned that strong selection on variants within candidate genes would be necessary for extreme shifts in ecological specialization. This can result in a pattern of hard selective sweeps resulting from a single haplotype rising quickly to fixation in a population derived from *de novo* mutation or standing variation (Orr and Betancourt 2001; Jensen 2014). Alternatively, a soft sweep occurs when selection drives multiple adaptive haplotypes to fixation — a pattern that can only result from selection on standing variation (Hermisson and Pennings 2005; Jensen 2014). Currently there are no theoretical predictions about the likelihood of adaptation from standing genetic variants versus *de novo* mutation for populations with small values of within-population divergence such as ours (Dittmar et al. 2016), and the relative importance of hard sweeps versus soft sweeps during adaptation is a subject of much debate (Hermisson and Pennings 2005; Pritchard et al. 2010; Jensen 2014; Garud et al. 2015; Shrider et al. 2015). In order to investigate whether regions associated with large jaws experienced hard sweeps, we examined the site frequency spectrum across candidate regions looking for signature shifts in variant frequencies across scaffolds.

Changes in ancestral population size can produce similar signals to hard selective sweeps. To account for this we first estimated the effective population size of all three species over the
past 20,000 generations and observed a 100-fold population decrease occurring within the same time as we predict ancestral populations colonized lakes on San Salvador Island (Mylroie and Hagey 1995; Turner et al. 2008; Martin and Wainwright 2013a). We next calculated a neutral site frequency spectrum under this bottleneck scenario and still detected hard sweeps in six of our candidate regions (three contributing to smaller jaws and three to larger jaws) (Fig. 5-6). Regions containing hint1, lyrm7, dyn2li1, and abcg5 along with a large unannotated region showed the strongest signs of hard sweeps after accounting for demographic history (Fig. 6). Low estimates of Tajima’s D, low nucleotide diversity in specialists, and high divergence between specialists and generalists lend further support for past selection at these loci (Tajima 1989; Nielsen 2005; Cruickshank and Hahn 2014). Alternatively, low recombination rates could account for low nucleotide diversity and high divergence at these loci (Nachman and Payseur 2012). A decrease in population size can also reduce genome-wide nucleotide diversity (Tajima 1989; Galtier et al. 2000). However, our demographic analysis indicates comparable decreases in population size for the generalist and specialist populations. Interestingly, 25 of our 31 strongest candidate regions do not show signs of hard selective sweeps. This may support a history of soft selective sweeps, where beneficial standing genetic variants were swept to fixation resulting in multiple haplotypes at candidate loci (Herisson and Pennings 2005; Jensen 2014).

Conclusions

The San Salvador Cyprinodon pupfish radiation has proven itself as an excellent system for investigating the genetic basis of novel trophic specialization. The extensive phenotypic diversity among these species results from low levels of genetic divergence and very few fixed variants. 31 regions with fixed variants showed significant associations with jaw size – the most rapidly diversifying trait in this system. Selection scans across regions associated with jaw size revealed a history of novel adaptation driven in part by hard selective sweeps. Additionally, we identified more variants with larger effects used to adapt to a more distant phenotypic optimum – consistent with Orr’s model of adaptation. Our evidence for the evolution of larger jaw size raises an alluring question with broad implications for research on adaptation: why has trophic novelty evolved exclusively on San Salvador Island? It is surrounded by islands with comparable physiochemistry, lake areas, macroalgae communities, and generalist Cyprinodon pupfish.
populations that exhibit similar genetic, phenotypic, and dietary diversity to generalist populations on San Salvador Island. This is consistent with similar levels of ecological opportunity on neighboring islands without specialists (Martin 2016a). Nonetheless, scale-eating and snail-eating species appear to be endemic to a single island. Answering this question will require continued exploration of the ecological and genetic factors shaping this exceptional case of rapid ecological specialization.

**Materials and Methods**

*Study system and sample collection*

Individuals were caught from hypersaline lakes on San Salvador Island, Bahamas using a hand net or seine net. 14 scale-eaters were sampled from six populations; ten snail-eaters were sampled from four populations; and 11 generalists were sampled from nine populations on San Salvador and a neighboring island. Samples were collected from nine isolated lakes on San Salvador (Great Lake, Stout’s Lake, Oyster Lake, Little Lake, Crescent Pond, Moon Rock, Mermaid’s Pond, Osprey Lake, Pigeon Creek, and one closely related outgroup *C. variegatus* population from Lake Cunningham, New Providence Island, Bahamas). Fish were euthanized in an overdose of buffered MS-222 (Finquel, Inc.) following approved protocols from the University of California, Davis Institutional Animal Care and Use Committee (#17455) and University of California, Berkeley Animal Care and Use Committee (AUP-2015-01-7053) and stored in 95-100% ethanol.

*Morphometrics*

Upper jaw lengths were measured using digital calipers from external landmarks on ethanol-preserved tissue specimens from the point of rotation on the quadroarticular joint (lower jaw joint), to the tip of the most anterior tooth on the dentigerous arm of the premaxilla. Body length was measured from the midline of the posterior margin of the caudal peduncle to the tip of the lower jaw (the nasal protrusion on some preserved *C. brontotheroides* samples obscured the upper jaw). In order to remove the effects of size variation, all measurements were log
transformed and regressed against log-transformed body length. We fit a log-transformed trait by log-transformed body length linear regression and used the residuals for association mapping.

**Genomic sequencing and bioinformatics**

DNA was extracted from muscle tissue using DNeasy Blood and Tissue kits (Qiagen, Inc.) and quantified on a Qubit 3.0 fluorometer (Thermofisher Scientific, Inc.). PCR-free Truseq-type genomic libraries were prepared using the automated Apollo 324 system (WaferGen BioSystems, Inc.) at the Vincent J. Coates Genomic Sequencing Center (QB3). Samples were fragmented using Covaris sonication, barcoded with Illumina indices, and quality checked using a Fragment Analyzer (Advanced Analytical Technologies, Inc.). 9-10 samples were pooled in four different libraries for sequencing on four lanes of Illumina 150PE Hiseq4000.

We mapped raw reads from 37 individuals to the *Cyprinodon* reference genome (NCBI, *Cyprinodon variegatus* annotation aeclease 100; total sequence length = 1,035,184,475; number of scaffolds = 9,259; scaffold N50 = 835,301; contig N50 = 20,803) with the Burrows-Wheeler Alignment Tool (Li and Durbin 2009 (v. 0.7.12)). The Picard software package (http://picard.sourceforge.net (v. 2.0.1)) was used to identify duplicate reads (MarkDuplicates) and create BAM indexes (BuildBamIndex). We followed the best practices guide recommended by the Genome Analysis Toolkit (DePristo et al. 2011; Van der Auwera et al. 2013 (v. 3.5)) in order to call and refine our SNP variant dataset using Haplotype Caller. Filtering SNP variants in GATK for model organisms conventionally requires high-quality known variants to act as a reference. Instead we called SNPs in our dataset using conservative hard-filtering parameters following GATK guidelines (DePristo et al. 2011; Marsden 2014): Phred-scaled variant confidence divided by the depth of non-reference samples >2.0, Phred-scaled P-value using Fisher's exact test to detect strand bias > 60, Mann-Whitney rank sum test for mapping qualities (z > 12.5), Mann-Whitney rank sum test for distance from the end of a read for those with the alternate allele (z > 8.0). Further filtering was performed using VCFtools (Danecek et al. 2011 (v. 0.1.14)) to only include individuals with a genotyping rate above 90% (no individuals were excluded by this filter) and SNPs with minor allele frequencies higher than 5%. Our final filtered dataset included 12,586,315 variant sites across 37 individuals with a mean aligned read sequencing depth of 7.19 per individual (range: 5.15 – 9.28).
Population genetic analyses

Our filtered dataset was converted from Variant Call Format to PED and MAP files using VCFtools. In order to visualize population structure in our samples (McVean 2009), we performed principal component analyses using eigenvectors output by PLINK’s ‘pca’ function (Purcell et al. 2007 (v. 1.9)). We plotted the first two principal components in R (R Core Team 2016 (v. 3.2.4)).

Genome wide $F_{st}$ for pairwise species comparisons was calculated for each variant site using VCFtools’ ‘weir-fst-pop’ function. Within-population nucleotide diversity ($\pi$) was estimated across 10kb windows using VCFtools’ ‘window-pi’ function. We used a custom python script to extract allele frequencies from the VCF files which were then used to estimate between population divergence ($D_{xy}$) with a separate R script (provided by A. Comeault). We calculated $D_{xy}$ across 10kb windows for ten scaffolds (totaling 9.7Mb) containing candidate SNPs for jaw size variation.

Association Mapping

We first estimated SNP × trait associations for jaw size variation using the PLINK ‘assoc’ function which fits a standard linear regression of phenotype on allele frequency and subsequently estimates $P$-values for each SNP with an asymptotic Wald test. We set a genome-wide level of significance using Bonferroni correction ($0.05 / 12,586,315 = 4.0 \times 10^{-9}$). Although this correction is highly conservative (Johnson et al. 2010), we are concerned here with only the most significant outliers. We then used the first two principal components explaining 9.44% of the variance in our dataset to correct for population structure by incorporating them into the model as covariates. We also performed an alternative method of mapping using a Bayesian sparse linear mixed model (BSLMM) implemented in the GEMMA software package (Zhou et al. 2013 (version 0.94.1)). GEMMA’s BSLMM combines linear mixed models, which assume every genetic variant has an effect on phenotype, and sparse regression models, which assume few variants will affect the phenotype. Importantly, GEMMA controls for background population structure by estimating and incorporating a kinship relatedness matrix as a covariate.
in the regression model. The BSLMM uses Markov Chain Monte Carlo (MCMC) to estimate the proportion of phenotypic variation explained by every SNP included in the analysis (PVE), the proportion of phenotypic variation explained by SNPs of large effect (PGE), which are defined as SNPs with a non-zero effect on the phenotype, and the number of large-effect SNPs needed to explain PGE (nSNPs). GEMMA calculates an effect size coefficient ($\beta$) and a posterior inclusion probability (PIP) for each SNP. Markers with non-zero values of $\beta$ are inferred to affect phenotypic variation in one iteration of the MCMC sampler. $\beta$ can be a positive or negative integer based on the direction of association, so we present estimates of this parameter in terms of its absolute value. PIP reports the proportion of iterations in which a SNP is estimated to have a non-zero effect on phenotypic variation ($\beta \neq 0$). This estimate might be difficult to interpret for SNPs in high linkage disequilibrium (LD) because tightly linked neutral and causal SNPs could each have a high probability of inclusion in separate iterations. We estimated pairwise LD ($r^2$) between SNPs on the largest scaffold (4.5 Mb) and found that linkage dropped to background levels between SNPs separated by more than 20kb ($r^2 < 0.1$) (Fig. S4). Thus, we summed $\beta$ and PIP parameters across 20kb windows to account for any unwanted dispersion of these values across SNPs in LD.

We performed 10 independent runs of the BSLMM for all 37 individuals (following Comeault et al. 2016) using a step size of 100 million with a burn-in of 50 million steps. We used GEMMA to assess the significance of regions associated with jaw size variation and report the median $\beta$ and PIP summed across windows for the 10 independent MCMC runs. Independent runs were consistent in reporting the strongest associations for the same 20kb windows. In order to compare the abundance and effect size of candidate loci between specialist species, we plotted the frequency of $\beta$ estimates for regions with effects on smaller jaws (negative $\beta$) and larger jaws (positive $\beta$).

**Identification of candidate genes**

We restricted our search to those regions both fixed between species and associated with jaw size. Accordingly, candidate regions met two rigorous criteria: 1) they must contain one or more SNPs that are fixed in at least one pairwise species comparison and 2) show significant association with jaw size in both association mapping analyses ($P < 4.0 \times 10^{-9}$ and outlier PIP
estimates above the 99\textsuperscript{th} percentile). We also took advantage of a recent linkage mapping analysis of phenotypic diversity in San Salvador \emph{Cyprinodon} pupfish by comparing our candidate regions for overlap with the four scaffolds containing QTL with moderate effects on jaw size in an F\textsubscript{2} intercross between specialists (Martín et al. accepted).

In addition to our candidate regions, we also report association mapping statistics and gene annotations for all 22 SNPs fixed between the generalist and scale-eater species. We used the Phenoscape Knowledgebase (phenoscape.org; Mabee et al. 2012; Midford et al. 2013; Manda et al. 2015, Edmunds et al. 2016) to determine if any of the annotated genes within fixed SNP regions were associated with skeletal system phenotypes across model taxa.

\textbf{Detecting Selection and Demographic History}

We first calculated Tajima’s D for each species in 10kb genomic windows using VCFtools’ ‘TajimaD’ function. This statistic compares observed nucleotide diversity to diversity under a null model assuming genetic drift, where negative values indicate a reduction in diversity across segregating sites that may be due to positive selection (Tajima 1989). Second, we used the SweepFinder method first developed by Nielsen et al. (2005b) and implemented in the software package SweeD (Pavlidis 2013). SweeD scans across non-overlapping windows to calculate a composite likelihood ratio (CLR) using a comparison between two contrasting models. The first assumes a window has undergone a recent selective sweep, while the second assumes a null model where the site frequency spectrum of the window does not differ from that of the entire scaffold. Windows with high CLR suggest a history of selective sweeps because the site frequency spectrum is shifted toward low and high frequency derived variants (Nielsen 2005b; Pavlidis 2013).

Various demographic histories can shift the distribution of low and high frequency derived variants to falsely resemble signatures of hard selection (Galtier et al. 2000; Nielsen, 2005). In order to account for demography, we used the Multiple Sequentially Markovian Coalescent (MSMC) (Schiffels and Durbin 2014) to infer historical effective population sizes (\(N_e\)) in all three species. MSMC is an extension of the Pairwise Sequentially Markovian Coalescent (PSMC) (Li and Durbin 2011), which uses a hidden Markov model to scan genomes.
analyzing patterns of heterozygosity where long DNA segments with low heterozygosity reflect recent coalescent events. The rate of coalescent events is then used to estimate \( N_e \) at a given time. We ran MSMC on unphased GATK-called genotypes from the 100 largest scaffolds for each individual separately, thus using only two haplotypes as in PSMC (the analysis of multiple individuals simultaneously would inform on more recent timescales, but requires phasing). As recommended in the MSMC documentation, we masked out sites with less than half or more than double the mean coverage for that individual, with a genotype quality below 20. We also excluded sites with less than 10 reads as recommended by Nadachowska-Brzyska et al. (2016). Nadachowska-Brzyska et al. (2016) also recommend to only use individuals with a mean coverage of at least 18x. However, all our individuals were sequenced at a lower coverage and we included only the seven individuals with a coverage of at least 7.5x. This means that our MSMC results should be interpreted with caution; however, the consistency among individuals of the same species (see Fig. S3) suggest that the general patterns of the analysis are likely to be robust.

To scale the output of MSMC to real time and population sizes, we assumed a six-month generation time (Martin et al. 2016c) and a mutation rate measured for cichlids \((6.6 \times 10^{-8} \text{ mutations per site per year})\) (Recknagel et al. 2013), one of the most closely related fish groups with an available estimate of spontaneous mutation rates.

We used ancestral population sizes determined by MSMC to analytically calculate the expected neutral site frequency spectrum with SweeD. We used the ‘-eN’ flag to model a 100-fold population decrease around 10,000 years ago (20,000 generations). We used a grid size of 1kb across our folded SNP dataset which defined sites as ancestral or derived variants based on the major and minor allele frequencies. We also ran SweeD without demographic assumptions for comparison. Because the significance of the CLR depends on the background site frequency spectrum of each scaffold, we compared the percentile of each likelihood estimate across unique scaffolds for candidate regions. Windows that showed CLRs above the 95\(^{th}\) percentile across their respective scaffolds under the assumptions of a population decrease determined by MSMC were interpreted as regions that recently experienced a hard sweep.

The size of the scaffolds containing jaw size candidate loci should be large enough to discover regions under strong selection. Out of our 31 candidate regions, we excluded one
because it fell within a small scaffold that could not be used to sample an adequate background distribution of heterogeneity. Of the 25 scaffolds containing the 31 regions we analyzed with SweeD, the mean scaffold length was 863,416bp. Furthermore, we set a conservative threshold (>95th percentile) to define regions that have experienced hard sweeps. We plot π, D_{xy}, and Tajima’s D across 10kb windows using a cubic smoothing spline in R.

Acknowledgements

This study was funded by the UNC Department of Biology, the Center for Population Biology, and ARCS Foundation awards to CHM. We thank Daniel Matute for helpful comments during the preparation of this manuscript; Jelmer Poelstra for help with PSMC analyses; Aaron Comeault, David Turissini, and Sara Suzuki for valuable discussions and computational assistance; Katelyn Gould for performing the DNA extractions; the Vincent J. Coates Genomic Sequencing Center and Functional Genomics Laboratory at UC Berkeley, supported by NIH S10 OD018174 Instrumentation Grant, for whole-genome resequencing; the Gerace Research Centre for accommodation; and the Bahamian government for permission to conduct this research. All datasets used for this study will be deposited in Dryad. All Illumina reads sequenced will be deposited in the NCBI Short Read Archive.

References


Comeault AA, Carvalho CF, Dennis S, Nosil P. 2016. Color phenotypes are under similar genetic control in two distantly related species of Timema stick insect. Evolution 70-6: 1283–1296


Cruickshank TE, Hahn MW. 2014. Reanalysis suggests that genomic islands of speciation are due to reduced diversity, not reduced gene flow. Mol. Ecol. 23:3133–3157.


Martin CH. 2016b. The cryptic origins of evolutionary novelty: 1,000-fold-faster trophic diversification rates without increased ecological opportunity or hybrid swarm. Evolution. 70-11: 2504–2519.


Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MAR, Bender D, Maller J, Sklar P, de Bakker PIW, Daly MJ, et al. 2007. PLINK: a tool set for whole-genome association and population-based linkage analyses. Am. J. Hum. Genet. 81:559–575.


Author Contributions

JAM wrote the manuscript, measured trait data, and conducted all bioinformatic and population genetic analyses. Both authors contributed to the conception and development of the ideas and revision of the manuscript.
Table 1. Jaw size association statistics and gene annotations for SNPs fixed between *C. desquamator* (scale-eater) and *C. brontotheroides* (snail-eater). Fixed SNPs fall within 20kb windows showing significant association with jaw size after controlling for population structure (Median PIP > 99th percentile). Asterisks (*) show SNPs in gene regions (bold) annotated for skeletal system effects. A cross (Ɨ) indicates overlap with a scaffold within a QTL affecting jaw size (Martin et al. accepted).

<table>
<thead>
<tr>
<th>SNP</th>
<th>Scaffold</th>
<th>Median PIP</th>
<th>PIP percentile</th>
<th>Median β</th>
<th>P</th>
<th>Gene region</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>KL652649.1</td>
<td>0.01795</td>
<td>1.0000</td>
<td>7.764633</td>
<td>1.82E-10</td>
<td>-</td>
</tr>
<tr>
<td>2</td>
<td>KL652649.1</td>
<td>0.0124</td>
<td>0.9999</td>
<td>4.10637</td>
<td>3.29E-10</td>
<td>-</td>
</tr>
<tr>
<td>3</td>
<td>KL653712.1</td>
<td>0.00975</td>
<td>0.9999</td>
<td>1.036102</td>
<td>6.65E-11</td>
<td>FAM49B/ZNF664</td>
</tr>
<tr>
<td>4</td>
<td>KL653062.1</td>
<td>0.0076</td>
<td>0.9999</td>
<td>-2.32365</td>
<td>3.82E-13</td>
<td>GMDS</td>
</tr>
<tr>
<td>5* Ɨ</td>
<td>KL652786.1</td>
<td>0.0069</td>
<td>0.9998</td>
<td>2.207843</td>
<td>6.66E-12</td>
<td>GALR2</td>
</tr>
<tr>
<td>6</td>
<td>KL652758.1</td>
<td>0.0066</td>
<td>0.9998</td>
<td>-1.15222</td>
<td>1.60E-11</td>
<td>SOGA3</td>
</tr>
<tr>
<td>7* Ɨ</td>
<td>KL652786.1</td>
<td>0.00625</td>
<td>0.9998</td>
<td>2.018056</td>
<td>1.41E-10</td>
<td>GALR2</td>
</tr>
<tr>
<td>8</td>
<td>KL652715.1</td>
<td>0.0058</td>
<td>0.9998</td>
<td>-2.21671</td>
<td>1.62E-09</td>
<td>PARD3</td>
</tr>
<tr>
<td>9</td>
<td>KL652649.1</td>
<td>0.0052</td>
<td>0.9998</td>
<td>2.223139</td>
<td>1.05E-10</td>
<td>-</td>
</tr>
<tr>
<td>10*</td>
<td>KL653271.1</td>
<td>0.0052</td>
<td>0.9996</td>
<td>0.291561</td>
<td>2.05E-10</td>
<td>ELN</td>
</tr>
<tr>
<td>11*</td>
<td>KL652666.1</td>
<td>0.0043</td>
<td>0.9995</td>
<td>-0.33468</td>
<td>5.30E-10</td>
<td>DYNC2LI1/ABCG5</td>
</tr>
<tr>
<td>12*</td>
<td>KL654513.1</td>
<td>0.00405</td>
<td>0.9994</td>
<td>-0.99172</td>
<td>1.24E-11</td>
<td>PLAUR</td>
</tr>
<tr>
<td>13*</td>
<td>KL653122.1</td>
<td>0.004</td>
<td>0.9994</td>
<td>1.029314</td>
<td>5.63E-10</td>
<td>ATP8A1</td>
</tr>
<tr>
<td>14</td>
<td>KL653046.1</td>
<td>0.0039</td>
<td>0.9993</td>
<td>1.189392</td>
<td>3.43E-09</td>
<td>LRP1B</td>
</tr>
<tr>
<td>15</td>
<td>KL653805.1</td>
<td>0.0038</td>
<td>0.9991</td>
<td>0.473089</td>
<td>1.23E-09</td>
<td>-</td>
</tr>
<tr>
<td>16*</td>
<td>KL652666.1</td>
<td>0.0037</td>
<td>0.9986</td>
<td>0.368651</td>
<td>1.98E-09</td>
<td>LYRM7/DYNC2LI1/HINT1</td>
</tr>
<tr>
<td>17</td>
<td>KL652617.1</td>
<td>0.0035</td>
<td>0.9983</td>
<td>1.517635</td>
<td>1.48E-12</td>
<td>PLXNA2</td>
</tr>
<tr>
<td>18</td>
<td>KL652527.1</td>
<td>0.0034</td>
<td>0.9983</td>
<td>0.140283</td>
<td>8.12E-13</td>
<td>TMEM30A/FILIP1L</td>
</tr>
<tr>
<td>19*</td>
<td>KL652983.1</td>
<td>0.0034</td>
<td>0.9981</td>
<td>-0.76248</td>
<td>1.06E-09</td>
<td>SKI</td>
</tr>
<tr>
<td>20</td>
<td>KL653291.1</td>
<td>0.00335</td>
<td>0.9977</td>
<td>-0.0425</td>
<td>4.74E-11</td>
<td>-</td>
</tr>
<tr>
<td>21</td>
<td>KL652991.1</td>
<td>0.0032</td>
<td>0.9967</td>
<td>-0.66796</td>
<td>3.12E-09</td>
<td>DLGAP1</td>
</tr>
<tr>
<td>22</td>
<td>KL653356.1</td>
<td>0.003</td>
<td>0.9961</td>
<td>0.979591</td>
<td>3.95E-12</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>---</td>
<td>---</td>
<td>---</td>
<td>---</td>
<td>---</td>
<td>---</td>
<td></td>
</tr>
<tr>
<td>23</td>
<td>KL653356.1</td>
<td>0.00295</td>
<td>0.9952</td>
<td>1.580411</td>
<td>4.39E-10</td>
<td></td>
</tr>
<tr>
<td>24</td>
<td>KL653706.1</td>
<td>0.00285</td>
<td>0.9947</td>
<td>-0.80369</td>
<td>1.60E-09</td>
<td>PLECKHG6</td>
</tr>
<tr>
<td>25</td>
<td>KL653420.1</td>
<td>0.0028</td>
<td>0.9940</td>
<td>-0.93815</td>
<td>7.16E-11</td>
<td></td>
</tr>
<tr>
<td>26</td>
<td>KL652585.1</td>
<td>0.00275</td>
<td>0.9936</td>
<td>1.384922</td>
<td>8.98E-11</td>
<td>FAM172A</td>
</tr>
<tr>
<td>27</td>
<td>KL654513.1</td>
<td>0.0027</td>
<td>0.9927</td>
<td>-0.41968</td>
<td>5.95E-12</td>
<td></td>
</tr>
<tr>
<td>28*</td>
<td>KL653925.1</td>
<td>0.00265</td>
<td>0.9927</td>
<td>-0.50498</td>
<td>1.15E-10</td>
<td>B3BNT3/B3GNT2</td>
</tr>
<tr>
<td>29</td>
<td>KL652727.1</td>
<td>0.00265</td>
<td>0.9927</td>
<td>0.075912</td>
<td>1.25E-09</td>
<td>RABGAP1</td>
</tr>
<tr>
<td>30</td>
<td>KL653654.1</td>
<td>0.00265</td>
<td>0.9919</td>
<td>0.056305</td>
<td>1.96E-09</td>
<td>COL15A1</td>
</tr>
<tr>
<td>31*</td>
<td>KL652717.1</td>
<td>0.0026</td>
<td>0.9919</td>
<td>-0.22127</td>
<td>4.65E-10</td>
<td>ASH1L/DAP3/GBA</td>
</tr>
</tbody>
</table>
Fig. 1. Survival Fitness Landscape for San Salvador *Cyprinodon* Pupfish. A) *C. variegatus* (red), *C. desquamator* (blue), and *C. brontotheroides* (green) from each lake population were intercrossed in every direction to produce F_2_ hybrids which were left for three months in an enclosure on San Salvador. Survival probability is plotted against two axes of the discriminant morphospace, indicating a wide range of jaw phenotypes in the F_2_ hybrids (modified from Martin and Wainwright, 2013). Heat colors correspond to survival probability (with blue being low and red being high). MicroCT scans of the cranial skeleton of each species modified from Hernandez et al. (in revision).

Fig. 2. Standardized jaw size and population structure. A) Principal component analysis showing axes accounting for a combined 9.45% of the total genetic variation between samples from 12 million SNPs genotyped from 37 whole-genome sequences. B) Log-transformed upper jaw length (mm) standardized by log-transformed body length for *C. variegatus* (red), *C. desquamator* (blue), and *C. brontotheroides* (green). Symbols represent individual lake of origin. MicroCT scans of the cranial skeleton of each species, modified from Hernandez et al. (in revision). + = Crescent Pond, × = Lake Cunningham, ▲ = Mermaid’s Pond , ■ = Little Lake, ○ = Osprey Lake , ● = Stout Lake , * = Great Lake, ◆ = Moon Rock, ▽ = Pigeon Creek, △ = Oyster Lake).

Fig. 3. *F*\_sub*st* distribution across 9,259 scaffolds. Upper panels show the distribution of genome-wide per-site *F*\_sub*st* estimates for 12,586,315 SNPs across all *Cyprinodon* scaffolds for A) *C. variegatus* vs. *C. desquamator* (28 individuals from ten lake populations), B) *C. variegatus* vs. *C. brontotheroides* (24 individuals from nine lake populations), and C) *C. brontotheroides* vs. *C. desquamator* (23 individuals from six lake populations). Lower panels show the distribution of SNPs with *F*\_sub*st* estimates greater than 0.80.
Fig. 4. **Quantitative Trait Association Mapping.** Log-transformed $P$ values for 12,586,315 SNP associations with jaw size variation estimated by PLINK ($n = 37$ individuals). Dotted blue line shows Bonferroni-corrected level of significance ($P < 4.0 \times 10^{-9}$). Red squares show the 31 SNPs spread across 25 scaffolds most strongly associated with jaw size that are also fixed between specialists.

Fig. 5. **Candidate regions associated with large jaw size.** Row 1 shows individual SNP $F_{st}$ values between *C. variegatus/C. desquamator*. Row 2 shows composite likelihood ratios estimated by SweeD using an analytical site frequency spectrum assuming a population bottleneck (magenta) and a frequency spectrum calculated without demographic assumptions (cyan) for *C. desquamator*. Row 3 shows Tajima’s D (dark yellow) for *C. desquamator*. Row 4 shows within-population diversity ($\pi$) (red: *C. variegatus*, green: *C. brontotheroides*, blue: *C. desquamator*). Row 5 shows between-population divergence ($D_{xy}$, black) for *C. variegatus/C. desquamator*. Black bars in row 4 show windows containing fixed SNPs that showed significant
association with jaw size in both PLINK and GEMMA association mapping analyses. Red bars in row 1 show exonic and intronic gene regions within windows.

**Fig. 6. Candidate regions associated with large and small jaw size.** Row 1 shows individual SNP $F_{st}$ values between *C. variegatus/C. brontotheroides*. Row 2 shows composite likelihood ratios estimated by SweeD using an analytical site frequency spectrum assuming a population bottleneck (magenta) and a frequency spectrum calculated without demographic assumptions (cyan) for *C. brontotheroides*. Row 3 shows Tajima’s D (dark yellow) for *C. brontotheroides*. Row 4 shows within-population diversity ($\pi$) (red: *C. variegatus*, green: *C. brontotheroides*, blue: *C. desquamator*). Row 5 shows between-population divergence ($D_{xy}$, black) for *C. variegatus/C. brontotheroides*. Black bars in row 4 show windows containing fixed SNPs that showed significant association with jaw size in both PLINK and GEMMA association mapping analyses. Red bars in row 1 show exonic and intronic gene regions within windows.

**Fig. 7. More Large-Effect Regions Control Large Jaw Phenotypes.** Distribution of effect size posterior parameters ($\beta$) estimated using GEMMA for 20kb regions with a posterior inclusion probability (PIP) greater than zero. We report median $\beta$ and PIP taken across 10 independent MCMC runs. Association mapping analysis shows twice as many outlier regions with large effects ($\beta > 99^{th}$ percentile (dotted red line)) on increasing jaw size ($n = 56$) compared to large-effects on decreasing jaw size ($n = 27$).
Fig. 1. Survival Fitness Landscape for San Salvador Cyprinodon Pupfish. A) C. variegatus (red), C. desquamator (blue), and C. brontotheroides (green) from each lake population were intercrossed in every direction to produce F2 hybrids which were left for three months in an enclosure on San Salvador. Survival probability is plotted against two axes of the discriminant morphospace, indicating a wide range of jaw phenotypes in the F2 hybrids (modified from Martin and Wainwright, 2013). Heat colors correspond to survival probability (with blue being low and red being high). MicroCT scans of the cranial skeleton of each species modified from Hernandez et al. (in revision)

89x85mm (300 x 300 DPI)
Fig. 2. Standardized jaw size and population structure. A) Principal component analysis showing axes accounting for a combined 9.45% of the total genetic variation between samples from 12 million SNPs genotyped from 37 whole-genome sequences. B) Log-transformed upper jaw length (mm) standardized by log-transformed body length for C. variegatus (red), C. desquamator (blue), and C. brontotheroides (green). Symbols represent individual lake of origin. MicroCT scans of the cranial skeleton of each species, modified from Hernandez et al. (in revision). + = Crescent Pond, × = Lake Cunningham, ▲ = Mermaid’s Pond, ■ = Little Lake, ○ = Osprey Lake, ● = Stout Lake, * = Great Lake, υ = Moon Rock, σ = Pigeon Creek, ρ = Oyster Lake.

152x85mm (300 x 300 DPI)
Fig. 3. Fst distribution across 9,259 scaffolds. Upper panels show the distribution of genome-wide per-site Fst estimates for 12,586,315 SNPs across all Cyprinodon scaffolds for A) C. variegatus vs. C. desquamator (28 individuals from ten lake populations), B) C. variegatus vs. C. brontotheroides (24 individuals from nine lake populations), and C) C. brontotheroides vs. C. desquamator (23 individuals from six lake populations). Lower panels show the distribution of SNPs with Fst estimates greater than 0.80.
Fig. 4. Quantitative Trait Association Mapping. Log-transformed P values for 12,586,315 SNP associations with jaw size variation estimated by PLINK (n = 37 individuals). Dotted blue line shows Bonferroni-corrected level of significance (P < 4.0 x 10^-9). Red squares show the 31 SNPs spread across 25 scaffolds most strongly associated with jaw size that are also fixed between specialists.
Fig. 5. Candidate regions associated with large jaw size. Row 1 shows individual SNP Fst values between C. variegatus/C. desquamator. Row 2 shows composite likelihood ratios estimated by SweeD using an analytical site frequency spectrum assuming a population bottleneck (magenta) and a frequency spectrum calculated without demographic assumptions (cyan) for C. desquamator. Row 3 shows Tajima’s D (dark yellow) for C. desquamator. Row 4 shows within-population diversity (π) (red: C. variegatus, green: C. brontotheroides, blue: C. desquamator). Row 5 shows between-population divergence (Dxy, black) for C. variegatus/C. desquamator. Black bars in row 4 show windows containing fixed SNPs that showed significant association with jaw size in both PLINK and GEMMA association mapping analyses. Red bars in row 1 show exonic and intronic gene regions within windows.
Fig. 6. Candidate regions associated with large and small jaw size. Row 1 shows individual SNP Fst values between C. variegatus/C. brontotheroides. Row 2 shows composite likelihood ratios estimated by SweeD using an analytical site frequency spectrum assuming a population bottleneck (magenta) and a frequency spectrum calculated without demographic assumptions (cyan) for C. brontotheroides. Row 3 shows Tajima’s D (dark yellow) for C. brontotheroides. Row 4 shows within-population diversity (π) (red: C. variegatus, green: C. brontotheroides, blue: C. desquamator). Row 5 shows between-population divergence (Dxy, black) for C. variegatus/C. brontotheroides. Black bars in row 4 show windows containing fixed SNPs that showed significant association with jaw size in both PLINK and GEMMA association mapping analyses. Red bars in row 1 show exonic and intronic gene regions within windows.
Fig. 7. More Large-Effect Regions Control Large Jaw Phenotypes. Distribution of effect size posterior parameters ($\beta$) estimated using GEMMA for 20kb regions with a posterior inclusion probability (PIP) greater than zero. We report median $\beta$ and PIP taken across 10 independent MCMC runs. Association mapping analysis shows twice as many outlier regions with large effects ($\beta > 99$th percentile (dotted red line)) on increasing jaw size ($n = 56$) compared to large-effects on decreasing jaw size ($n = 27$).