Major stages of vertebrate adaptive radiation are assembled from a disparate spatiotemporal landscape

Authors: Emilie J. Richards[1,2], Joseph A. McGirr[3], Jeremy R. Wang[4], Michelle E. St. John[1,2], Jelmer W. Poelstra[5], Maria J. Solano[3], Delaney C. O'Connell[3], Bruce J. Turner[6], Christopher H. Martin[1,2*]

Affiliations:
[1]Department of Integrative Biology, University of California, Berkeley, CA
[2]Museum of Vertebrate Zoology, University of California, Berkeley, CA
[3]Department of Biology, University of North Carolina, Chapel Hill, NC
[4]Department of Genetics, University of North Carolina, Chapel Hill, NC
[5]Department of Biology, Duke University, Durham, NC
[6]Department of Biological Sciences, Virginia Tech, VA
*Correspondence to: chmartin@berkeley.edu

Abstract: To investigate the origins and stages of vertebrate adaptive radiation, we reconstructed the spatial and temporal histories of genetic variants underlying major phenotypic axes of diversification from the genomes of 202 Caribbean pupfishes. Ancient standing variation from disparate spatial sources was reassembled into new combinations which are under strong selection for adaptation to novel trophic niches on only a single island throughout the Caribbean. This occurred in three stages: first, standing variation associated with feeding behavior swept, then standing variation regulating craniofacial development and pigmentation, and finally de novo variation for craniofacial development. Our results provide clear support for two longstanding hypotheses about adaptive radiation and demonstrate how ancient alleles maintained for millennia in distinct environmental refugia can be assembled into new adaptive combinations.

One Sentence Summary: Ancient origins of adaptive radiation

Main Text: Adaptive radiations are fundamental to understanding the biodiversity of life. These bursts of phenotypic and ecological diversification may occur in response to ecological opportunity provided by unoccupied niche space (*1*, *2*). However, the origins and major features of this process are still controversial. For example, ecological opportunity does not explain why only some lineages radiate (*3–6*). One hypothesis is that introgression from disparate source populations might be necessary to trigger diversification (*7*, *8*). Despite substantial evidence of adaptive introgression during radiation (*9–12*), no previous studies have compared adaptive introgression between closely related radiating and non-radiating lineages to distinguish introgression as a trigger. A parallel debate centers on whether adaptive diversification proceeds in three distinct temporal stages: shifts in habitat preference first, followed by trophic morphology, and finally sexual communication (*13*). This 'behavior-first' view is also commonly proposed for initiating adaptive trait evolution (*14–16*). However, existing evidence for the temporal stages hypothesis comes from ancestral state reconstructions of rapidly diversifying traits which are unreliable without fossil data (*17–19*).

Here we provide strong support for these two major hypotheses of adaptive radiation using multiple lines of genomic, transcriptomic, and phenotypic evidence in a nascent adaptive radiation of Caribbean pupfishes. This sympatric radiation contains a widespread generalist algae and detritus-eating species (*Cyprinodon variegatus*) and two trophic specialists endemic to 10 ky old hypersaline lakes on San Salvador Island (SSI), Bahamas: a molluscivore *C. brontotheroides* with a unique nasal protrusion and a scale-eater *C. desquamator* with striking two-fold longer oral jaws. This clade exhibits hallmarks of adaptive radiation. First trait diversification rates are up to 1,400 times faster than non-radiating generalist populations on neighboring Bahamian

islands in nearly identical hypersaline lake environments. Second, craniofacial diversity within the radiation is comparable to all other Cyprinodontidae species combined (*5*, *20*).

To investigate the spatiotemporal history of adaptive variants unique to trophic specialists on SSI we first constructed a high quality *de novo* hybrid assembly for *C. brontotheroides* (1.16 Gb size; scaffold N50 = 32 Mb; L50 = 15; 86.4% complete Actinopterygii BUSCOs) and resequenced 202 genomes (7.9x median coverage) from across the range of *Cyprinodon* and the two closest outgroups *Megupsilon* and *Cualac* (Fig. 1A;Table S1; Data S1). Population structure across the Caribbean was largely explained by geographic distance (Fig. 1) and the SSI radiation did not contain higher overall genetic diversity than the rest of the Caribbean (Fig. S1). All Caribbean populations experienced similar declines in effective population size following the last glacial maximum 15 kya when an order of magnitude more Caribbean coastal habitat was above sea level (Fig. 1D).

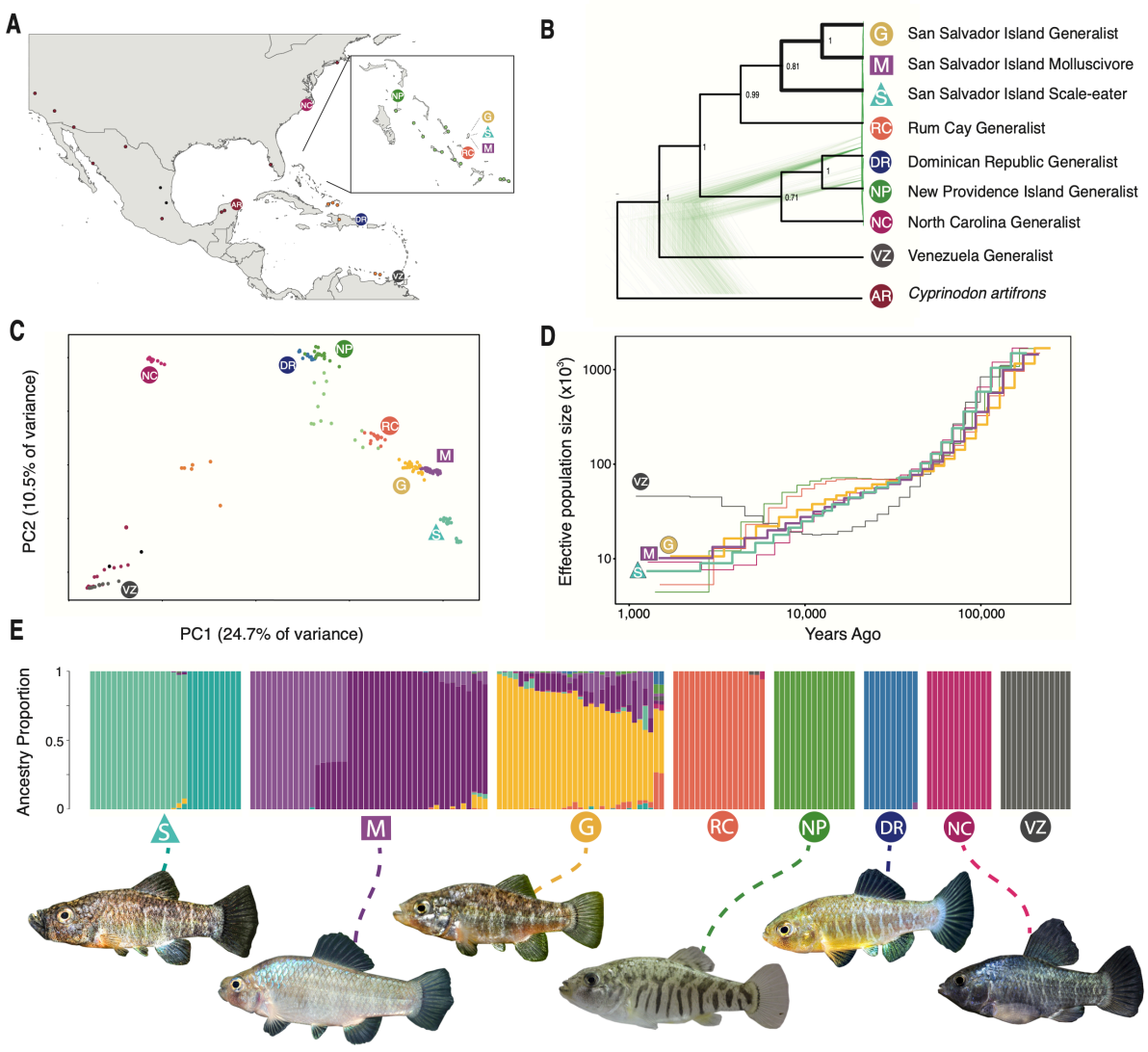


Fig. 1. Genetic diversity of pupfishes across the Caribbean. A) Sample locations of *Cyprinodon* pupfishes (n = 202) with eight focal populations ($n \geq 10$ per population) marked with symbols and single individuals from other locations in the Bahamas (light green), Caribbean (orange), continental North and South America (maroon), and *Megupsilon* and *Cualac* outgroups to *Cyprinodon* (black). B) Maximum clade credibility phylogeny estimated with SNAPP from 10K SNPs for focal populations and the outgroup *Cyprinodon artifrons overlaying 500 gene trees randomly sampled from the posterior distribution and visualized with Densitree*. C) Principal component analysis of *Cyprinodon* pupfishes. D) Changes in effective population size

over time for focal populations in the Caribbean inferred using MSMC. E) Ancestry proportions across individuals in San Salvador and 5 other Caribbean populations. Proportions estimated from a LD-pruned dataset in ADMIXTURE with k=11.

We scanned 5.5 million variants across 202 Caribbean-wide pupfish genomes to identify 3,464 scale-eater and 1,491 molluscivore candidate adaptive variants, respectively, showing evidence of both strong genetic differentiation between trophic specialists ($F_{st} \geq 0.95$) and a hard selective sweep (Fig. 2A; Fig S2; Data S2 and S3). 28% of these candidate adaptive variants were found in cis-regulatory regions, 12% in intronic regions, and 2% in coding regions, resulting in a total of 176 candidate genes for adaptive radiation (Table S2-S3). These genes were enriched (FDR < 0.05) for developmental processes associated with the major axes of phenotypic diversification in this radiation, including feeding behavior, muscle tissue, and craniofacial development (Fig. 2C; Table S4). 45% of these genes were also differentially expressed between trophic specialists (FDR < 0.05; Data S4-S5) in whole embryos at 2 and/or 8 days post fertilization (dpf) (*21*).

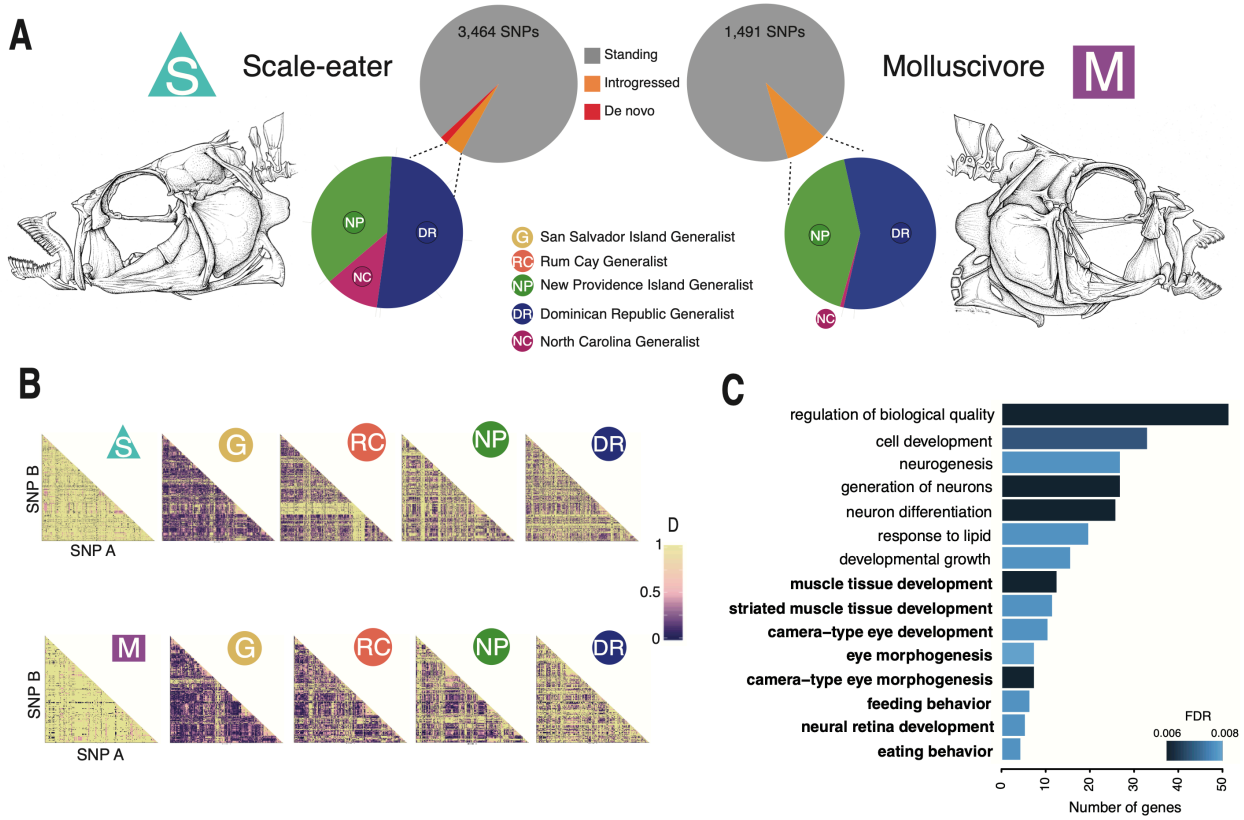


Fig. 2. All candidate adaptive genetic variants underlying trophic specialist pupfishes. A) Origins of all candidate adaptive variants for trophic specialists (F_{st} > 0.95; hard selective sweep CLR≥ 4.47.; see supplementary methods for more details) divided into three categories: SSI only (de novo: red), introgression from a specific source population (introgressed: orange), or observed in multiple Caribbean populations as standing genetic variation (standing: grey). Introgressed variation is further broken down by focal source population (Tables S9-S12). B) Heatmaps of linkage disequilibrium among all pairwise combinations of candidate adaptive variants for scale-eaters (S; top row) and molluscivores (M; bottom row) on SSI in comparison to linkage patterns among these SNPSs in generalists on SSI (G) and three other focal generalist populations across the Caribbean (Venezuela removed from comparison due to its unique bottleneck history among the populations). Note the breakdown in linkage disequilibrium among candidate adaptive variants outside of trophic specialist populations. C) Top 15 GO categories in

which scale-eater candidate adaptive variants were significantly enriched, with relevant terms corresponding to major axes of divergence in this radiation highlighted in bold (FDR < 0.01; full list of terms with FDR < 0.05 in Table S4).

Even though both trophic specialists are endemic to SSI, we also found nearly all their candidate adaptive variants to occur as standing genetic variation across the Caribbean (molluscivore: 100%; scale-eater: 98%; Fig. 2A). Furthermore, nearly half these variants were ancient and also found in *Cualac* or *Megupsilon* outgroups to *Cyprinodon* (41% and 55% of scale-eater and molluscivore adaptive variants, respectively), which diverged over 5 Mya (*22*). However, most adaptive variants did not show any evidence of selection in five other focal high-coverage Caribbean generalist populations (only 2% and 6% of scale-eater and molluscivore variants, respectively; Fig. S3) and strong linkage disequilibrium among adaptive variants in SSI trophic specialists was not observed in these focal populations (Fig. 2B). Thus, novel trophic specialists within a microendemic adaptive radiation were almost entirely assembled from ancient standing genetic variation through strong selection for new adaptive combinations of alleles.

Multiple lines of evidence suggest that more hybridization and adaptive introgression took place in SSI populations than other Caribbean island populations, consistent with the hypothesis that hybridization triggered adaptive radiation. First, the strongest signal of introgression across the Caribbean was into the root node of the SSI radiation (Fig. 3B). Second, trophic specialists on San Salvador experienced at least twice as much adaptive introgression as other generalist populations across the Caribbean ($P < 0.006$; Fig. 3C). Third, based on

126 differences between the distribution of tract lengths for neutral and adaptively introgressed
127 regions, we infer that adaptive introgression only resulted from older hybridization events with
128 generalist source populations, despite evidence of recent and continuous introgression to the
129 present (Fig 3D-E).

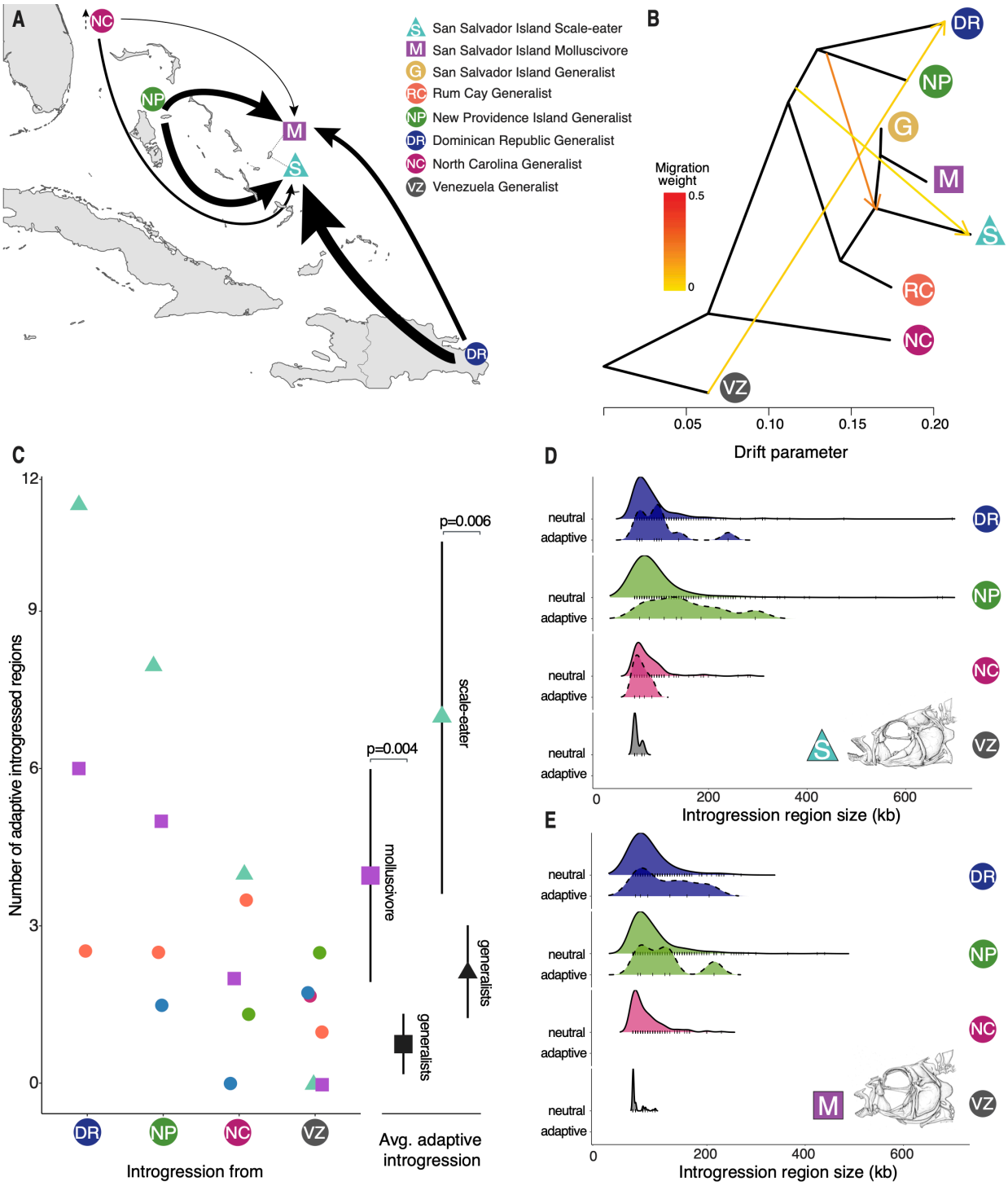


Fig. 3. A history of hybridization across the Caribbean. A) A summary map of adaptive introgression into SSI trophic specialists from focal generalist populations across the Caribbean, with thickness of arrows proportional to the number of outlier regions of the f_d statistic. B)

Genome-wide population graph inferred from *Treemix* with the 3 strongest signals of admixture. Note that the strongest signal is into the root node of the SSI radiation. C) The total number of candidate adaptive introgression regions from outgroup generalist populations is often larger in specialists (triangle: scale-eater and square: molluscivore) than in other outgroup generalist populations (circle, color matches population legend in panel A). The bootstrapped mean and confidence interval for the number of adaptive introgression events into the molluscivore and scale-eater populations compared to other Caribbean generalist populations are shown to the right of the panel (triangle: scale-eater introgression regions and square: molluscivore introgression regions). D-E) Density plots of the size distribution of adaptive introgression regions (dashed line) and neutral introgression regions (solid line) across the genomes of D) scale-eaters and E) molluscivores, divided into the four focal generalist source populations. Tick marks on the bottom of density curves indicate the number of introgression regions observed.

We also observed distinct temporal stages of adaptive radiation based on divergence time estimates of candidate loci and timing of selective sweeps. SSI pupfish diversified along the three major axes of vertebrate adaptive radiation (*13*), including their trophic environment (mediated through foraging behavior: scale-eating or snail-eating), trophic morphology, and sexual communication signals. Based on both GO annotations of genes near candidate adaptive variants and genome-wide association mapping (GWAS) in Caribbean pupfishes for these trait axes, we found that the divergence times of loci associated with these three major stages of adaptive radiation exhibited a significant temporal signal overall (ANOVA, $P = 0.03$; Fig. 4A), driven by a 'behavior-first' pattern (permutation test; P=0.0041). Independent adaptive variants

in the *cis*-regulatory regions of two genes associated with feeding behavior (*prlh*, *cfap20;* (*23*, *24*)) were the oldest estimated sweeps out of all adaptive loci associated with these axes and swept from standing genetic variation ten times older than the radiation itself (Fig. 4A,C; 95% HPD sweep ages: 6,747-8,490 and 6,594-9,210 years, respectively). *Cfap20* and *prlh* were also both differentially expressed between trophic specialists at early developmental stages, consistent with a *cis*-regulatory function of these variants (Data S4-S5;(*21*)).

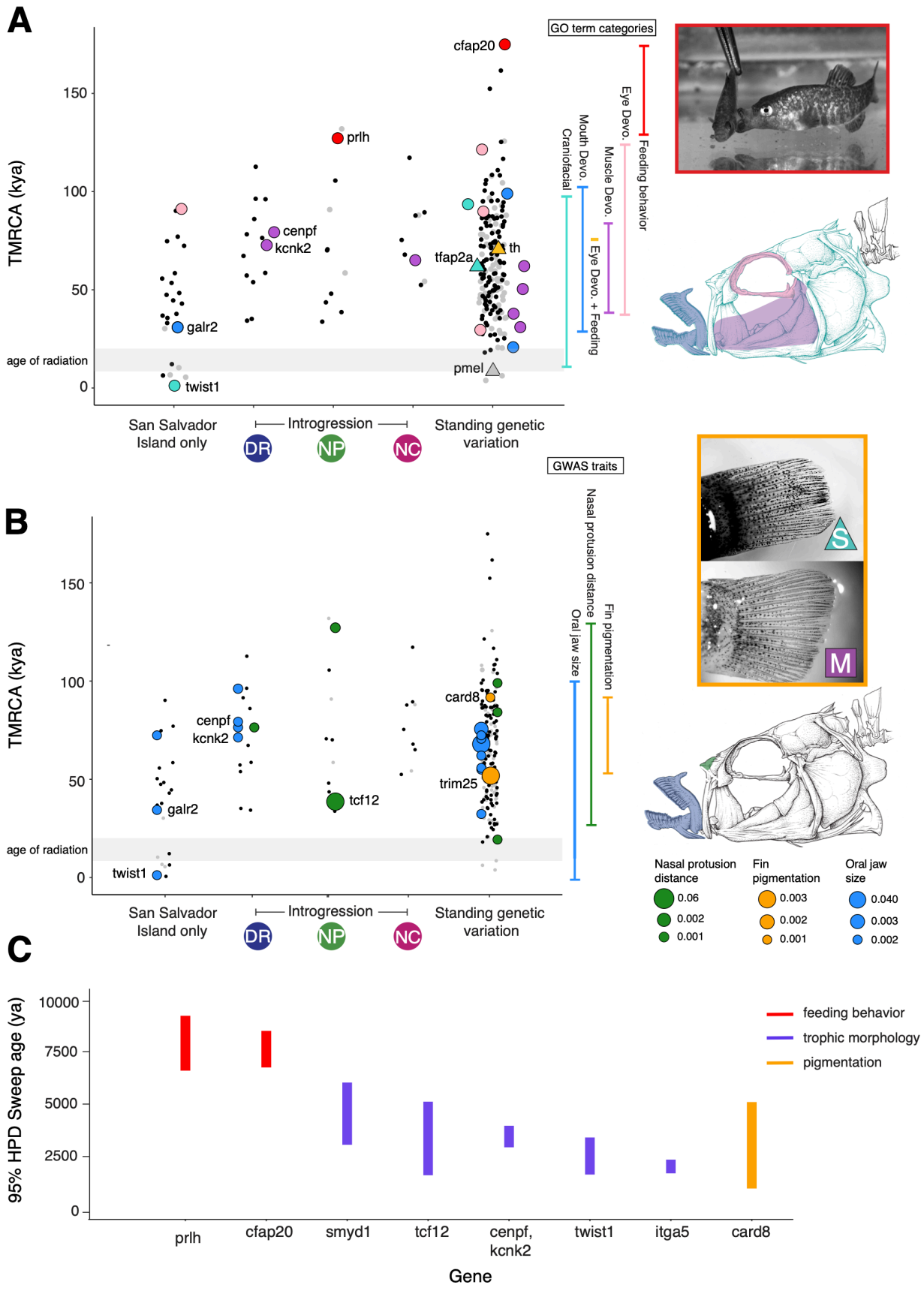
Consistent with the second stage of trophic morphological divergence, adaptive variants in the regulatory regions of genes associated with eye development (*gnat2, tbc1d20, zhx2*), muscle development (*fhod3, smyd1, kcnk2, fhl2, pdim5*), and craniofacial morphology (*bcor, itga5, tfap2a*, *med1*) swept from both standing genetic variation and adaptive introgression from three different focal populations across the Caribbean (Fig. 4). We also observed a final stage of novel refinement in which two *de novo* variants associated with craniofacial morphology arose on SSI and swept in scale-eating populations (Fig 4). One of these variants is a non-synonymous substitution in the second exon of *twist1*, a transcription factor known to influence cleft palate and oral jaw size (*25*), which was significantly associated with SSI pupfish oral jaw size in a genome-wide association scan (99th PIP percentile GWAS; Fig 4B), is highly conserved across ray-finned fishes (GERP score: 2.19; Fig S4), and is one of the most recent sweeps of any adaptive variant detected in our analyses (95% HPD: 1,636-3,413 ya). The second variant is in the regulatory region upstream of the gene *galr2*, which produces a transmembrane receptor for galanin, a peptide known to facilitate bone formation (*26*). This gene was significantly associated with SSI pupfish oral jaw size (99th PIP percentile GWAS; Fig 4B; Data S6) and lies within a significant QTL that accounts for 15% of the variation in oral jaw size in an F2 intercross between SSI specialist species (Table S5; (*27*)). Molluscivores displayed similar temporal stages

with craniofacial variation sweeping recently (Fig. S5), but contained no *de novo* variants. 85% of candidate adaptive regions under selection in molluscivores were also under selection in scale-eaters but contained different fixed variants (Fig. S3), consistent with previously observed patterns of parallel metabolic pathway gene expression in trophic specialists but divergent genotypes (*21*).

There was no evidence for a temporally distinct third stage of diversification in sexual signals, despite the striking divergence in male reproductive pigmentation between SSI trophic specialists. Instead, pigmentation diverged throughout the process of adaptive radiation. Adaptive variants in two genes known to affect pigmentation (*tfap2a*, *th*; Fig 4A) and two additional candidate variants associated with pupfish caudal fin pigmentation (99th PIP percentile GWAS; Fig. 4B; Data S7) indicate that diversification in male reproductive coloration spans a range of divergence times, with some variants more ancient than the radiation itself (Fig 4B). For example, two nearly fixed adaptive variants in the regulatory region of *card8*, a homolog of *nlrp1* which is associated with vitiligo and pigmentation loss in humans [(*28*); 99th PIP percentile GWAS; differentially expressed at 2 and 8 dpf; Fig 4B; Data S4-S5,S7], swept at the same time as many variants associated with craniofacial morphology during the second stage of adaptive radiation. Divergence estimates of loci associated with pigmentation also span a large range: from variants in the regulatory region of *card8*, which is ten times older than the radiation, to a single variant fixed in scale-eaters and in the neighboring Rum Cay generalist population found in the regulatory region of *pmel* (known to affect melanosome development (*29*); differentially expressed between specialists at 8 dpf ;Data S5), which is as young as the radiation itself (Fig. 4A). This broad range of divergence times and sweep ages may indicate that the distinctive light/dark reproductive coloration associated with trophic specialists diverged

throughout the process of adaptive radiation using existing standing genetic variation, rather than only during a final stage.

Intriguingly, along with distinct temporal patterns of adaptive divergence, we also find distinct spatial patterns to the sources of adaptive introgression. Introgression from different regions of the Caribbean brought in adaptive variation for different major axes of phenotypic diversification within the radiation. Adaptive variants associated with feeding behavior (*prlh;* Fig 4A) and nasal protrusion (*tcf12*; 99th PIP percentile GWAS;Fig 4B;Data S8) originated in the northwestern Bahamas (New Providence Island, Exumas, and Cat Island) whereas adaptive introgression of variants associated with muscle development arrived from the southeastern Caribbean in the Dominican Republic (*cenpf, eya2;* Fig 4A). This suggests that the extant SSI radiation of trophic specialists was reassembled from pools of ancient genetic variation contained in at least two distinct environmental refugia in other regions of the Caribbean, perhaps due to previous ephemeral adaptive radiations within these regions, thus connecting micro- and macroevolutionary-scale processes (*30*, *31*).



216

Fig. 4. The spatiotemporal landscape of adaptive radiation in pupfish. Time to most recent common ancestor (TMRCA) of candidate adaptive variants based on *Dxy* for 50-kb region containing candidate adaptive variants. Variants are separated by spatial distribution: SSI only (de novo), introgression, and standing genetic variation. Gray lines highlight the approximate age of the radiation: from the peak of the last glacial maximum (approximately 20 kya; (*32*)) to the youngest age estimate for filling of hypersaline lakes on SSI (~3 kya: (*33*)). A) Variants are colored by their adaptive relevance to this system. Black: adaptive variants annotated for non-focal GO terms or unannotated; Gray: nearly fixed neutral variants between specialists with no signal of hard selective sweep; and triangles represent variants associated with pigmentation. All variants annotated for the GO categories of feeding behavior (red), muscle (purple), eye, mouth development (blue), and craniofacial development (mouth and eye and/or muscle(teal)) are highlighted. Genes highlighted in the text are labeled by their associated variant. B) Variants are colored by significant association (99^{th} percentile of PIP) in GWAS with a larger lower oral jaw size, darker caudal fin pigmentation, and smaller nasal protrusion distance. Dot sizes scale with PIP score. C) 95% HPD interval of the posterior distribution for selective sweep ages for focal gene regions. Regions are colored by GO and/or GWAS annotations relevant to the three major stages of adaptive radiation: habitat preference (feeding behavior), trophic morphology (craniofacial and muscle), and sexual communication (pigmentation).

In conclusion, the SSI adaptive radiation was triggered by the formation of a hybrid swarm of largely ancient alleles maintained within different pools of standing variation in Caribbean and mainland outgroups. Distinct temporal stages of adaptation observed in this nascent radiation are consistent with a behavior-first stage of vertebrate radiation. Much of the

adaptive variation contributing to major phenotypic axes of diversification in this radiation originated over longer timescales preceding the divergence of trophic specialist species on SSI, and across large spatial scales. Our results show that adaptive radiations can occupy expansive evolutionary spaces: spanning the existing radiation itself and the multitude of both past and present ephemeral pools of genetic variation that contributed to rapid diversification. Research into the broader spatiotemporal landscape of vertebrate radiations, including the hominin radiation (*31*), can provide clear answers regarding longstanding hypotheses about their origins and contributions to global patterns of biodiversity.

References and Notes:

1. G. G. Simpson, *Tempo and mode of evolution* (Columbia University Press, ed. 15, 1944).
2. J. T. Stroud, J. B. Losos, Ecological Opportunity and Adaptive Radiation. *Annu. Rev. Ecol. Evol. Syst.* **47**, 507–532 (2016).
3. D. H. Erwin, Novelty and innovation in the history of life. *Curr. Biol.* **25**, R930–R940 (2015).
4. C. E. Wagner, L. J. Harmon, O. Seehausen, Ecological opportunity and sexual selection together predict adaptive radiation. *Nature*. **487**, 366–369 (2012).
5. C. H. Martin, The cryptic origins of evolutionary novelty: 1,000-fold-faster trophic diversification rates without increased ecological opportunity or hybrid swarm. *Evolution (N. Y).*, 1–16 (2016).
6. D. L. Rabosky, Diversity-Dependence, Ecological Speciation, and the Role of Competition in Macroevolution. *Annu. Rev. Ecol. Evol. Syst.* **44**, 481–502 (2013).
7. O. Seehausen, Hybridization and adaptive radiation. *Trends Ecol. Evol.* **19**, 198–207 (2004).
8. D. A. Marques, J. I. Meier, O. Seehausen, A Combinatorial View on Speciation and Adaptive Radiation. *Trends Ecol. Evol.* (2019), doi:https://doi.org/10.1016/j.tree.2019.02.008.
9. E. J. Richards, C. H. Martin, Adaptive introgression from distant Caribbean islands contributed to the diversification of a microendemic adaptive radiation of trophic specialist pupfishes. *PLoS Genet.* **13**, 1–35 (2017).
10. J. I. Meier, D. A. Marques, S. Mwaiko, C. E. Wagner, L. Excoffier, O. Seehausen, Ancient hybridization fuels rapid cichlid fish adaptive radiations. *Nat. Commun.* **8**, 14363 (2017).

11. The Heliconius Genome Consortium, K. K. Dasmahapatra, J. R. Walters, A. D. Briscoe, J. W. Davey, A. Whibley, N. J. Nadeau, A. V. Zimin, D. S. T. Hughes, L. C. Ferguson, S. H. Martin, C. Salazar, J. J. Lewis, S. Adler, S.-J. Ahn, D. a. Baker, S. W. Baxter, N. L. Chamberlain, R. Chauhan, B. a. Counterman, T. Dalmay, L. E. Gilbert, K. Gordon, D. G. Heckel, H. M. Hines, K. J. Hoff, P. W. H. Holland, E. Jacquin-Joly, F. M. Jiggins, R. T. Jones, D. D. Kapan, P. Kersey, G. Lamas, D. Lawson, D. Mapleson, L. S. Maroja, A. Martin, S. Moxon, W. J. Palmer, R. Papa, A. Papanicolaou, Y. Pauchet, D. A. Ray, N. Rosser, S. L. Salzberg, M. a. Supple, A. Surridge, A. Tenger-Trolander, H. Vogel, P. a. Wilkinson, D. Wilson, J. a. Yorke, F. Yuan, A. L. Balmuth, C. Eland, K. Gharbi, M. Thomson, R. A. Gibbs, Y. Han, J. C. Jayaseelan, C. Kovar, T. Mathew, D. M. Muzny, F. Ongeri, L.-L. Pu, J. Qu, R. L. Thornton, K. C. Worley, Y.-Q. Wu, M. Linares, M. L. Blaxter, R. H. ffrench-Constant, M. Joron, M. R. Kronforst, S. P. Mullen, R. D. Reed, S. E. Scherer, S. Richards, J. Mallet, W. Owen McMillan, C. D. Jiggins, Butterfly genome reveals promiscuous exchange of mimicry adaptations among species. *Nature*. **487**, 94–98 (2012).

12. S. Lamichhaney, J. Berglund, M. S. Almén, K. Maqbool, M. Grabherr, A. Martinez-Barrio, M. Promerová, C.-J. Rubin, C. Wang, N. Zamani, B. R. Grant, P. R. Grant, M. T. Webster, L. Andersson, Evolution of Darwin's finches and their beaks revealed by genome sequencing. *Nature*. **518**, 371–375 (2015).

13. T. J. Streelman, P. D. Danley, The stages of vertebrate evolutionary radiation. *Trends Ecol. Evol.* **18**, 126–131 (2003).

14. J. B. Losos, T. W. Schoener, D. A. Spiller, Endurance trials Dietary effects on ontogenic consistency and second analysis of natural selection. *Lett. to Nat.* **432**, 505–508 (2004).

15. J. Diamond, Evolution of ecological segregation in the New Guinea montane avifauna. *Community Ecol.*, 98–125 (1986).

16. R. B. Huey, P. E. Hertz, B. Sinervo, Behavioral drive versus behavioral inertia in evolution: A null model approach. *Am. Nat.* **161**, 357–366 (2003).

17. L. C. Sallan, M. Friedman, Heads or tails: Staged diversification in vertebrate evolutionary radiations. *Proc. R. Soc. B Biol. Sci.* **279**, 2025–2032 (2012).

18. R. E. Glor, Phylogenetic Insights on Adaptive Radiation. *Annu. Rev. Ecol. Evol. Syst.* **41**, 251–270 (2010).

19. S. Duchêne, R. Lanfear, Phylogenetic uncertainty can bias the number of evolutionary transitions estimated from ancestral state reconstruction methods. *J. Exp. Zool. Part B Mol. Dev. Evol.* **324**, 517–524 (2015).

20. C. H. Martin, P. C. Wainwright, Trophic novelty is linked to exceptional rates of morphological diversification in two adaptive radiations of cyprinodon pupfish. *Evolution (N. Y).* **65**, 2197–2212 (2011).

21. J. A. Mcgirr, C. H. Martin, Parallel evolution of gene expression between trophic specialists despite divergent genotypes and morphologies. *Evol. Lett.* **2**, 62–75 (2018).

22. A. A. Echelle, E. W. Carson, A. F. Echelle, D. Bussche, T. E. Dowling, Historical

Biogeography of the New-World Pupfish Genus Cyprinodon (Teleostei: Cyprinodontidae). *Copeia*. **2005**, 320–339 (2005).

23. T. M. Maia, D. Gogendeau, C. Pennetier, C. Janke, R. Basto, Bug22 influences cilium morphology and the posttranslational modification of ciliary microtubules. *Biol. Open*. **3**, 138–151 (2014).

24. Y. Takayanagi, H. Matsumoto, M. Nakata, T. Mera, S. Fukusumi, S. Hinuma, Y. Ueta, T. Yada, G. Leng, T. Onaka, Endogenous prolactin-releasing peptide regulates food intake in rodents. *J. Clin. Invest.* **118**, 4014–4024 (2008).

25. C. S. Teng, M. C. Ting, D. T. Farmer, M. Brockop, R. E. Maxson, J. G. Crump, Altered bone growth dynamics prefigure craniosynostosis in a zebrafish model of Saethre-Chotzen syndrome. *Elife*. **7**, 1–23 (2018).

26. H. W. McGowan, J. A. Schuijers, B. L. Grills, S. J. McDonald, A. C. McDonald, Galnon, a galanin receptor agonist, improves intrinsic cortical bone tissue properties but exacerbates bone loss in an ovariectomised rat model. *J. Musculoskelet. Neuronal Interact.* **14**, 162–172 (2014).

27. C. H. Martin, P. A. Erickson, C. T. Miller, The genetic architecture of novel trophic specialists: higher effect sizes are associated with exceptional oral jaw diversification in a pupfish adaptive radiation. *Mol. Ecol.* **26**, 624–638 (2017).

28. A. D. Osualdo, J. C. Reed, NLRP1, a regulator of innate immunity associated with vitiligo. *Pigment Cell Melanoma Res.*, 5–8 (2011).

29. H. B. Schonthaler, J. M. Lampert, J. Von Lintig, H. Schwarz, R. Geisler, S. C. F. Neuhauss, A mutation in the silver gene leads to defects in melanosome biogenesis and alterations in the visual system in the zebrafish mutant fading vision. *Dev. Biol.* **284**, 421–436 (2005).

30. E. B. Rosenblum, B. A. J. Sarver, J. W. Brown, S. Des Roches, K. M. Hardwick, T. D. Hether, J. M. Eastman, M. W. Pennell, L. J. Harmon, Goldilocks Meets Santa Rosalia: An Ephemeral Speciation Model Explains Patterns of Diversification Across Time Scales. *Evol. Biol.* **39**, 255–261 (2012).

31. C. H. Martin, E. J. Richards, The Paradox behind the Pattern of Rapid Adaptive Radiation: How Can the Speciation Process Sustain Itself through an Early Burst? *Annu. Rev. Ecol. Evol. Syst.* (2019).

32. P. U. Clark, A. S. Dyke, J. D. Shakun, A. E. Carlson, J. Clark, B. Wohlfarth, J. X. Mitrovica, S. W. Hostetler, A. M. McCabe, The Last Glacial Maximum. *Science (80-.).* **325**, 710–714 (2009).

33. B. J. Turner, D. D. Duvernell, T. M. Bunt, M. G. Barton, Reproductive isolation among endemic pupfishes (Cyprinodon) on San Salvador Island, Bahamas: Microsatellite evidence. *Biol. J. Linn. Soc.* **95**, 566–582 (2008).

348 **Acknowledgments:** We thank Rebecca Tarvin for helpful comments on the manuscript; the
349 Gerace Research Centre and Troy Day for logistical support; the governments of the Bahamas,
350 Dominican Republic, the National Park Service and U.S. Fish and Wildlife Service for
351 permission to collect and export samples; the Vincent J. Coates Genomics Sequencing Center
352 and Functional Genomics Laboratory at UC Berkeley for performing whole-genome library
353 preparation and sequencing (supported by NIH S10 OD018174 Instrumentation Grant), and the
354 University of North Carolina at Chapel Hill for computational resources. **Funding:** This work
355 was funded by the National Science Foundation DEB CAREER grant #1749764, National
356 Institutes of Health grant 5R01DE027052-02, the University of North Carolina at Chapel Hill,
357 and the University of California, Berkeley to CHM. **Author contributions:** Conceptualization –
358 EJR,CHM; Data Curation: EJR,JRW,MJS, DCOC; Formal Analyses: EJR,JAM,MSJ; Resources:
359 CHM, BJT; Visualization: EJR,CHM; Writing – original draft: EJR; Writing – review & editing:
360 EJR, CHM, JAM, MSJ, JRW, JWP, MJS. **Competing interests:** The authors declare no
361 competing interests. **Data and materials availability:** Genomic and RNA sequence data are
362 archived at the National Center for Biotechnology Information BioProject Database (Accession:
363 XXX; PRJNA394148, PRJNA391309; and PRJNA305422); and additional data and scripts are
364 on the Dryad Digital Repository (XXX) and Github (https://github.com/joemcgirr).

365 **Supplementary Materials:**
366 Materials and Methods
367 Figures S1-S13
368 Tables S1-S14
369 External Databases S1-S9
370 References (*34-86*)
371
372
373

Materials and Methods

Sampling

Pupfishes were collected from across the complete Atlantic and Caribbean range of *Cyprinodon* from Massachusetts to Venezuela. For the three species of the SSI radiation, individual pupfish were collected from 15 isolated hypersaline lakes on San Salvador Island (Table S1;Data S1) and one estuary (Pigeon Creek) using hand and seine nets between 2011 and 2018. 36 *Cyprinodon variegatus*, 47 *C. brontotheroides*, and 39 *C. desquamator* were sampled across these lakes, including six lakes in which one or two specialist species occur in sympatry with the generalists (Crescent Pond, Storr's Lake, Little Lake, Oyster Pond, Osprey Lake, Moon Rock Pond). The sampling of outgroup high-coverage focal populations of generalist pupfish included 17 individuals from *C. laciniatus* from Lake Cunningham, New Providence Island, Bahamas; 18 *C. variegatus* from Lake George, Rum Cay, Bahamas; 12 *C. higuey* from Laguna Bavaro, Dominican Republic; 14 *C. variegatus* from Fort Fisher estuary, North Carolina, United States; and 14 *C. dearborni* from Isla Margarita, Venezuela. 37 individuals were also collected from other islands and localities spanning the range of *Cyprinodon* across the Caribbean and Atlantic coasts, including captive-bred individuals from the extinct species *Megupsilon aporus* and threatened species *Cualac tessellatus,* the most closely related outgroup genera to *Cyprinodon* (Fig. 1A; Table S1;Data S1;(*1*)).

Fishes 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), the University of North Carolina at Chapel Hill Animal Care and Use Committee (#18-061.0), and the University of California, Berkeley Animal Care and Use Committee (AUP-2015-01-7053) and preserved in 95-100% ethanol.

398 Genomic library preparation

399 DNA was extracted from muscle tissue using DNeasy Blood and Tissue kits (Qiagen, Inc.) and
400 quantified on a Qubit 3.0 fluorometer (Thermofisher Scientific, Inc.). Genomic libraries were
401 prepared using the automated Apollo 324 system (WaterGen Biosystems, Inc.) at the Vincent J.
402 Coates Genomic Sequencing Center (QB3). Samples were fragmented using Covaris sonication,
403 barcoded with Illumina indices, and quality checked using a Fragment Analyzer (Advanced
404 Analytical Technologies, Inc.). Nine to ten samples were pooled per lane for 150PE sequencing
405 on four lanes of an Illumina Hiseq4000 and an additional 96 individuals were sequenced on one
406 150PE lane of Illumina Novaseq with S4 chemistry. This included 42 individuals from a
407 previous genomic study (*2*).

408

409 De novo genome assembly and annotation

410 We constructed a hybrid de novo assembly for an inbred lab-raised individual of *C.*
411 *brontotheroides* using three different sequencing technologies. Oxford Nanopore sequencing was
412 performed at UNC's High Throughput Sequencing Facility, a 10X Genomics synthetic long-read
413 library was prepared and sequenced by Hudson Alpha, and Chicago and HiC libraries were
414 prepared and sequenced by Dovetail Genomics. Genomic DNA was extracted from an inbred F4
415 male *C. brontotheroides* individual, an offspring from three generations of full-sib mating in the
416 lab, starting with an F0 generation collected from Crescent Pond, SSI. 10X sequencing was
417 performed on this individual according to 10X Genomics' recommended protocol, sequenced on
418 a HiSeq4000, resulting in 460 million 2x150 bp reads. DNA was extracted from this same
419 molluscivore individual for Nanopore sequencing using modified a phenol:chloroform extraction
420 protocol (*3*). Two libraries were sequenced on R9.4 flow cells on GridION – one using the Rapid

Sequencing Kit (RAD004) and one Ligation Kit (LSK109), producing 4.9 Gbp of sequences with a read length N50 of 4.7 Kbp.

10X Genomics sequences were first assembled using Supernova (v2.0.0, 10X Genomics) to produce a preliminary "pseudohap" assembly. Nanopore reads were corrected using FMLRC (*4*). The Supernova assembly was scaffolded with corrected nanopore reads using LINKS (*5*) using the recommended iterative approach (34 rounds). The nanopore-scaffolded assembly was further scaffolded using HiC and Chicago sequences; we predicted Hi-C contacts using Juicer (v1.6.2; (*6*)), followed by scaffolding with 3D-DNA (v180922) (*7*). We performed a final polishing with four rounds of Racon (v1.3.1; (*8*)) using corrected nanopore reads. The final assembly consists of 1.16 Gbp in 15,698 scaffolds with an N50 of 32,013,756 bp.

To further validate our assembly, we ran BUSCO (v3.0.1) (*9*) to identify known single-copy conserved genes. We found 86.4% of BUSCOs in the Actinopterygii class assembled completely, and 83.4% in a single copy. We annotated this assembly using the Maker pipeline (v3.01.02)(*10*), providing alternate ESTs and protein evidence for ab-initio gene prediction from *C. variegatus* (*11*), which is expected to have very similar genic structure and codon usage. Predicted genes were assigned putative function by aligning (BLASTp) to the UniProt database.

Population genotyping

Raw reads were mapped from 222 individuals to a de-novo assembly of *Cyprinodon brontotheroides* reference genome (v 1.0; total sequence length = 1,162,855,435 bp; number of scaffold = 15,698, scaffold N50 = 32 Mbp) with bwa-mem (v 0.7.12;(*12*)). Duplicate reads were identified using MarkDuplicates and BAM indices were created using BuildBamIndex in the Picard software package (http://picard.sourceforge.net(v.2.0.1)). We followed the best practices

guide recommended in the Genome Analysis Toolkit (v 3.5;(*13*)) to call and refine our single nucleotide polymorphism (SNP) variant dataset using the program HaplotypeCaller. We filtered SNPs based on the recommended hard filter criteria (i.e. QD < 2.0; FS < 60; MQRankSum < -12.5; ReadPosRankSum < -8;(*13*, *14*)) because we lacked high-quality known variants for these non-model species. Variants for San Salvador Island individuals were additionally filtered to remove SNPs with a minor allele frequency below 0.05, genotype quality below 20, or containing more than 20% missing data across all individuals at the site using vcftools (v.0.1.15;(*15*)). This set of 9.3 million variants was then further filtered for variants that had minor allele frequencies above 0.05 and less than 50% missing data across all Caribbean outgroup individuals with population level sampling. Variants in poorly mapped regions were then removed using a mask file generated from the program SNPable (http://bit.ly/snpable; k-mer length =50, and 'stringency'=0.5). Our final dataset after filtering contained 5.5 million variants.

Population genetic analyses

This dataset was first pruned for SNPs in strong linkage disequilibrium using the LD pruning function (--indep-pairwise 50 5 0.5) in plink (v1.9)(*16*), leaving 2.6 million variants. To visualize population structure in our dataset, we ran a principle component analysis using the eigenvectors outputted by plink's pca function (--pca). The first two principal components were plotted in R (R Core Team 2018 v3.5.0). To visualize admixture among the species we estimated the proportion of shared ancestry among individuals in our dataset using ADMIXTURE (v.1.3.0)(*17*). The number of populations (K) was decided upon using ADMIXTURE's cross-validation method (--cv) across 1-20 population values of K. A K of 11 populations was then

chosen using the broken-stick method. Ancestry proportions estimated by ADMIXTURE were plotted in R. Four individuals that appeared to exhibit recent hybrid ancestry between *C. variegatus* and *C. brontotheroides* and two individuals that appeared to exhibit recent hybrid ancestry between *C. variegatus* and *C. desquamator* were removed from downstream analyses. We also excluded 15 individuals that appeared as strong outliers in the PCA and ADMIXTURE analyses (3 *C. variegatus* from San Salvador Island, 1 *C. brontotheroides*, 3 *C. laciniatus*, 2 *C. higuey*, 3 *C. variegatus* from North Carolina, and 3 *C. dearborni* from Venezuela), resulting in 32 *Cyprinodon variegatus*, 44 *C. brontotheroides*, and 26 *C. desquamator* individuals from San Salvador Island, 16 individuals from *C. laciniatus* from Lake Cunningham, New Providence Island in the Bahamas, 17 *C. variegatus* from Lake George, Rum Cay, 10 *C. higuey* from Lake Bavaro, Dominican Republic, 12 *C. variegatus* from Fort Fisher estuary North Carolina, and 11 *C. dearborni* from Isla Margarita, Venezuela (Fig 1E). None of the 37 single individuals from other locations were removed. The final dataset used in downstream analyses included 202 individuals.

Within-population nucleotide diversity (π) was calculated in 50-kb windows across the genome for each of eight focal populations with more than 10 individuals sampled. To ensure an equal comparison among populations, we downsampled individuals from each population to the number of individuals in the focal population with the lowest sampling (n=10). We did this by randomly selecting 10 individuals for each population before calculating π in sliding windows. We repeated this 100 times and averaged π across the replicates (Fig. S1). Due to the large sample size of windows for each population (n=30,762), slight differences in mean genome-wide within-population genetic diversity resulted in statistically significant differences in genome-wide diversity among populations (ANOVA, $P > 2.2 \times 10^{-16}$). However, the effect sizes of the

difference in these means was small in all comparisons except San Salvador Island generalists and North Carolina and Venezuela generalist populations (Cohen's d=0.87 and 1.38 respectively). The much lower within-population genetic diversity in Venezuela than other generalist populations could be explained by the fact that this population seems to have undergone a recent bottleneck in the past that was not observed in the other populations sampled (Fig. 1C and S1).

Genome-wide F_{st} for pairwise San Salvador Island species comparisons was calculated for each variant site and in 50-kb windows with at least 100 variant sites across the genome using the python script popGenWindows.py available from https://github.com/simonhmartin/genomics_general (*18*). Allowing for some admixture among these recent species, we considered divergent SNPs to be those that were nearly fixed between specialist species ($F_{st} \geq 0.95$; Fig. S2; Table S2-S3; Data S2-S3). All divergent variants that were also in a region of a hard selective sweep (see section below) were also in the 90th percentile of D_{xy} across all genomic windows, as calculated for the same 50-kb windows using the python script popGenWindows.py.

A window size of 50-kb for sliding window tests was based on the extent of linkage disequilibrium (LD) along a scaffold. We calculated LD decay from pairwise calculations of LD between all SNPs within 100-kb of each other along the largest scaffold in the genome using PLINK's LD function (--r^2). Linkage disequilibrium decayed to background rates after 50-kb at a threshold of $r^2 \geq 0.1$ (Fig. S6).

Mutation rate estimation

The spontaneous mutation rate for Caribbean pupfishes was estimated from high coverage sequencing (15-69x) of parents and offspring from two independent pedigreed crosses of San Salvador Island species: one cross between a second generation inbred lab-reared generalist and third-generation inbred lab-reared molluscivore individual from Little Lake and another between a second-generation lab-reared generalist and second-generation lab-reared scale-eater from Little Lake. Using the same pipeline for alignment to the reference genome and variant calling with GATK as above, we obtained 9 million variants across 7 individuals from these two crosses after using GATK's recommend hard filter criteria (i.e. QD < 2.0; FS < 60; MQRankSum < -12.5; ReadPosRankSum < -8). We independently called variants for these same individuals again using samtools mpileup (v1.9) with the command line arguments *bcftools mpileup -Ou | bcftools call -m -Ob -f GQ,GP*. For both sets of variants (GATK and samtools), poorly mapped regions were then removed using a mask file generated from the program SNPable (http://bit.ly/snpable; k-mer length =50, and 'stringency'=0.5). We further excluded sequences in which indels were called in any sample, as well as 3 bp of sequence around the indel.

After variant calling, we searched for new mutations in the offspring: sites where an offspring is heterozygous for an allele not found in either of the parents. To determine a set of filters for new mutation sites in the offspring, we first looked for variants which were heterozygous in the offspring and alternatively homozygous in the parents (i.e. known heterozygous sites). Eleven measures of variant quality scores for these known heterozygous sites in the offspring were then used to filter sites for new mutations in the offspring (e.g. depth, genotype quality, mapping quality; see Table S6 for full details) following similar pipelines and filters from several previous studies ((*19–21*)). For example, only new mutation sites that had a depth within 2 standard deviations of the mean depth of the known heterozygous sites in the

offspring were kept (all thresholds reported in Table S6). Additionally, new mutations in the offspring were determined from sites in which parents were homozygous for the reference allele and the offspring were heterozygous with quality scores within those of the known heterozygote sites (Table S6) and had an allele balance score between 0.3 and 0.7. This set of variants was then filtered for those independently called in both GATK and samtools runs (Table S6) with no known segregating sites in the larger Caribbean population dataset (i.e. not sites where the alternative new allele in the offspring was found in four or more copies in the larger dataset).

Using the GATK function *callable loci*, we then determined the 'accessible genome': the total number of base pairs from the genome in which mutations could be confidently called for each cross. This number was estimated using similar variant quality filters as for the new mutation estimate, excluding those filters that were only applicable to the new mutations and heterozygous sites (i.e. filters assessing quality of alternative allele calls). This meant excluding genomic regions where 1) read map depth in any member was not within two standard deviations of the average read map depth (varies by sample; Table S6), 2) regions with mapping quality scores less than 50, and 3) regions with base quality scores less than 30.

Since the new mutations observed could have originated on either chromosome, the point estimate of per site mutation rate is the number of new mutations observed divided by two times the size of the accessible genome. The mutation rates were then averaged across individual offspring for each cross (Table S6) to obtain a mean mutation rate estimate of 1.56×10^{-8} mutations per site per generation. This is faster than mutation rate estimates for other teleosts (*20*, *22*, *23*); however, short-lived smaller species with higher metabolism rates like pupfishes are expected to exhibit faster mutation rates (*24*). We estimate generation times in the field to be approximately one year based on laboratory and field (*25*) longevity studies.

Demographic Inferences

Various demographic histories can shift the distribution of low- and high-frequency derived variants to falsely resemble signatures of hard selective sweeps. In order to account for demography in downstream analyses, we used the MSMC (v. 1.0.1; *24*) to infer historical effective population size (N_e) changes in seven focal populations. We ran MSMC on unphased GATK-called genotypes separately for an individual in each of seven focal populations (excluding generalist *C. higuey* due to poor sequencing quality of our single high-coverage individual) with higher sequencing coverage (17-28x mean coverage across individuals; Fig 1D;Table S7). As recommended in the MSMC documentation, we masked out sites with less than half or more than double the mean coverage for that individual or with a genotype quality below 20. We also excluded sites with <10 reads as recommended by Nadachowska-Brzyska et al. (*27*). To scale the output of MSMC to real time and effective population sizes, we used a one-year generation time (*24*) and the estimated spontaneous mutation rate of 1.56×10^{-8} for Caribbean pupfishes (see previous section).

Introgression

We characterized differential introgression between specialists in the SSI radiation on both a genome-wide and local level. We visualized the directionality of hybridization and introgression on a genome-wide level using *TreeMix* (v 1.13; (*28*)). *TreeMix* estimates a maximum likelihood phylogeny of the focal populations and then fits a user specified number of migration edges to the tree by comparing genetic covariances of allele frequencies among populations. We ran

TreeMix with *C. dearborni* as the root node, and with 0 through 20 migration edges. The most likely number of migration events was chosen using the broken-stick approach (Fig. S7).

We investigated how signatures of hybridization at the genome-wide level contributed variation potentially important to the divergence between species using the f_d statistic, which is designed to look for signatures of introgression across sliding genomic windows (*18*). The f_d statistic, a modified version of the *D*-statistic, looks at allele frequencies fitting two allelic patterns referred to as ABBA and BABA based on the tree (((P1,P2),P3),O), where O is an outgroup species in which no gene flow is thought to occur with the other populations (*18*). We used 2 individuals of *C. artifrons* from Cancun, Mexico as our distantly related outgroup population for this test, which forms the deepest divergence event with *C. variegatus* within the *Cyprinodon* clade (*1*), and focused on introgression between San Salvador Island specialists and Caribbean populations. Based on the tree (((P1,P2),P3), *C. artifrons*), the f_d statistic was calculated for the combinations of populations in which P2 was either the scale-eater or molluscivore and P3 was one of the Caribbean outgroup populations (Table S8 and S9) in 50-kb sliding windows with a minimum of 100 variant sites with no missing data within a population using the ABBABABA.py script (available on https://github.com/simonhmartin/genomics_general;(*18*)). To compare these patterns of introgression into the specialist to patterns of introgression into focal generalist populations on other islands, we also calculated f_d statistics across the genomes of generalist populations on islands where we had population-level sampling and sister groups to fit the relationship expected for the test (Table S9B and S9D).

Significance of f_d values in sliding windows across the genome was evaluated using simulations with no migration using ms-move (*29*) using estimates of effective population size

changes from our MSMC analyses and a divergence time between the two specialists set to 10,000 years, the age of the hypersaline lakes on SSI (Table S8). Significant regions were determined by calculating the f_d statistic across a genome simulated under a coalescent model with no gene flow and consisting of 150,000 windows, each containing the mean number of variants found across 50-kb windows in the empirical genomes. Empirical windows were considered candidate introgressed regions if the f_d statistic was above the maximum simulated f_d value (Table S8). Consecutive 50-kb f_d outlier windows were merged to estimate the range in the sizes of introgressed regions to approximate the age of introgression events (Fig. 2D).

Candidate adaptive variants underlying San Salvador Island specialist phenotypes

In the specialists, we also looked for regions that appeared to be under strong divergent selection in the form of a hard selective sweep from the site frequency spectrum calculated with SweeD (v.3.3.4;(*30*)). In this calculation of the composite likelihood ratio (CLR) of a sweep, we incorporated our empirical estimate of the decrease in population size for each focal population estimated from MSMC analyses in 50-kb windows across scaffolds that were at least 100-kb in length (99 scaffolds; 85.6% of the genome). We also calculated CLRs across 100,000 scaffolds consisting of neutrally evolving sequences simulated with ms-move (*29*), controlling for the impact of the inferred population size decreases over time for each population from MSMC runs mentioned above (Fig. 1D; Table S7). The CLR ratios for the simulated datasets were then used to assess outlier CLR ratios from the empirical dataset. Regions with CLR ratios above the 95th percentile value of CLR from the neutral simulated dataset were considered candidate hard selective sweep regions (scale-eater: CLR > 5.28; molluscivore: CLR > 4.47; Table S7). Candidate hard selective sweep regions were also independently inferred for the five focal

Caribbean generalist populations (sample size ≥ 10) following the same method outlined above for the specialists (Table S7).

Spatial distribution of candidate adaptive variants

To determine candidate adaptive variants underlying the divergence observed on San Salvador Island, we looked for regions of the genome that contained signatures of strong genetic divergence and hard selective sweeps. Variants that were nearly fixed between the species on San Salvador (F_{st} ≥0.95) and located in a candidate selective sweep region (empirical CLR > demographic simulations CLR; Table S7) were considered candidate adaptive variants (Table S2-S3; Data S2-S3). We then surveyed all pupfish individuals sampled from outside these populations for the specialist allele at their respective candidate variant sites (e.g. whether or not the scale-eater allele was found in other populations across the Caribbean). Variants were then separated into two categories of genetic variation: *de novo* (the specialist allele was found only on San Salvador Island) or standing (the specialist allele was also found in at least one generalist population sampled outside of San Salvador Island). Standing genetic variation that was found in candidate introgression regions from the f_d tests for introgression (see Introgression section) was further parsed into the category of introgressed variation and separated by geographic region of introgression (North Carolina (NC), New Providence Island (NP), or Dominican Republic (DR)). Given the amount of candidate adaptive variation that exists as standing genetic variation across the Caribbean (Fig. 2A), we looked for how many of these candidate regions also showed evidence of hard selective sweeps in focal generalist populations outside of San Salvador Island and found that very few of these regions exhibited signatures of a hard selective sweeps outside of San Salvador Island populations (Fig. S3).

653 Adaptive introgression
654
655 We examined all introgressed regions for evidence of hard selective sweeps on SSI and genetic
656 divergence between trophic specialists, which would provide evidence that secondary gene flow
657 brought in variation potentially important for speciation. We looked for overlap between the f_d
658 introgression outliers, SweeD selective sweep outliers, and whether these overlapping regions
659 contained at least one of the nearly fixed variants ($F_{st} \geq 0.95$) between specialists (*31*) (Table
660 S10-S14).
661 We were interested in whether San Salvador Island specialist genomes exhibited more
662 introgression in regions undergoing hard selective sweeps on SSI than other generalist
663 populations. In the absence of a clear null expectation for the number of introgressed regions, we
664 calculated the number of these candidate adaptive regions for the specialists that were also
665 outlier *fd* regions in other combinations of populations across the Caribbean (Table S9), to
666 compare to the number of adaptive introgression regions observed in the specialists. Since
667 several outgroup generalist populations had multiple values for the number of adaptive
668 introgression regions (due to different combinations of sister lineages (P1) available for testing
669 against: Table S9), only the average number of adaptive introgression regions per generalist
670 population was shown for ease of visualization (Table S9; Fig. 3C). We then performed a Mann-
671 Whitney U test to determine if the mean number of adaptive introgression regions in each
672 specialist was greater than the mean from the rest of the Caribbean (Table S9A v. S9B and Table
673 S9C v. S9D) and calculated 95% confidence intervals around these means using the boot.ci
674 function in the R package boot (v1.3; Fig. 3C). Since neither of the SSI specialists appear to have
675 experienced adaptive introgression from the Venezuela *C. dearborni* population, it was excluded

as a potential donor population for the focal generalist populations on other islands as well in these comparative analyses.

Functional analysis of candidate adaptive variants

GO analysis of candidate variant regions

We performed gene ontology (GO) enrichment analyses for genes near candidate adaptive variants using ShinyGo (v.0.51;(*32*)). In the *C. brontotheroides* reference genome annotations (described in *de novo* genome assembly and annotation section), gene symbols largely match human gene symbols. Thus, we searched for enrichment across biological process ontologies curated for human gene functions.

For genes with focal GO terms (e.g. feeding behavior, muscle, mouth, eye and craniofacial development) relevant to stages of diversification in this system (i.e. habitat preference, trophic morphology, and pigmentation; Fig. 2C; Fig. 4; Table S4), we checked other annotation databases and studies for verification of putative function, including Phenoscape Knowledgebase (https://kb.phenoscape.org/#/home), NCBI's PubMed (https://www.ncbi.nlm.nih.gov/pubmed), and the Gene Ontology database using AMIGO2 (*33*) . All genes had consistent annotations across databases, except *galr2*. *Galr2* was annotated for feeding behavior in the Biological Processes database (Ensemble 92), but to our knowledge, the most recent studies indicate that it does not play a role in feeding behavior (*34*, *35*). Thus, we removed its annotation as a candidate gene for feeding behavior, but kept it as a candidate for trophic morphology.

Differential gene expression

Additionally, we looked for overlap between genes associated with candidate variants and genes differentially expressed between the two specialists in whole embryos at two early developmental stages (2 and 8 days post-fertilization (dpf)) reported in a previous study (*36*). Tables with differentially expressed genes at 2 and 8 dpf from this study are provided in Data S4 and S5.

Morphometrics and caudal fin pigmentation

We measured two key morphological traits associated with the major axes of phenotypic diversification in the SSI radiation, lower jaw length and nasal protrusion distance. Ethanol-preserved specimens from SSI were measured from external landmarks on the skull using digital calipers. Measurements were repeated on both lateral sides and averaged for each specimen. Lower jaw length was measured from the quadrate-articular jaw joint to the tip of the most anterior tooth on the dentary (Data S6). Nasal protrusion distance was measured by placing a tangent line from the dorsal surface of the neurocranium to the tip of the premaxilla and measuring the perpendicular distance that the nasal region protrudes from this tangent (Fig. S8A; Data S6). Each specimen was also measured for standard length using digital calipers in order to remove the effects of variation in body size on the craniofacial trait measurements among individuals and species. We log-transformed morphological measurements and regressed them against log-transformed standard length (Fig. S9; Data S6) and used the residuals for association mapping analyses.

The major axis of divergence in reproductive coloration and patterning between trophic specialists on SSI is the overall lightness or darkness of breeding males. Scale-eaters reach a nearly jet black coloration in the wild while guarding a breeding territory whereas molluscivore

726 males remain paler throughout their body and fins. This pair of sympatric specialists exceeds the
727 lightness contrast in male reproductive breeding coloration observed across all other *Cyprinodon*
728 pupfishes. Females of each species show the same general pattern of lightness/darkness. We
729 detected no difference in the total number of melanocytes on the caudal, anal, or pectoral fins
730 among the SSI species (data not shown). Instead, we found that scale-eater individuals were
731 significantly darker on their caudal fins (two-tailed *t*-test, t=5.25, df=45.5, *P* -value= 3.8×10^{-6};
732 Fig. 4B; Data S6), perhaps due to the larger melanocyte areas relative to molluscivores. We
733 found similar patterns for anal and pectoral fins and used only caudal fin lightness values for
734 GWAS (data not shown). A Meiji EMZ-8TR stereomicroscope with standardized external
735 illumination and an OMAX 18 Mp digital microscope camera was used to take lateral
736 photographs of the caudal fin of each individual against the same white reference background in
737 each image (Fig. 4B;Data S6). Adobe Photoshop (Creative Cloud) was used to select a
738 rectangular area from inside the caudal fin, not including the caudal peduncle region or terminal
739 marginal band and measure the mean overall lightness of this region relative to a control region
740 selected from the illuminated white background. Standardized caudal fin pigmentation was then
741 calculated as the proportion of the caudal fin lightness value relative to the control background
742 lightness value for downstream GWAS analyses.
743
744 *Genome-wide association mapping analyses*
745
746 We employed a Bayesian Sparse Linear Mixed Model (BSLMM) implemented in the GEMMA
747 software package (v. 0.94.1; (*37*)) to identify genomic regions associated with variation in lower
748 oral jaw length, caudal fin pigmentation, and nasal protrusion distance. The BSLMM uses
749 Markov Chain Monte Carlo sampling to estimate the proportion of phenotypic variation

explained by all SNPs included in the analysis (proportion of phenotypic variance explained (PVE);Fig. S10A-C), explained by SNPs of large effect (proportion of genetic variance explained by sparse effects (PGE); Fig. S10D-F), and the number of large-effect SNPs needed to explain PGE (nSNPs;Fig. S10G-I). GEMMA also estimates a posterior inclusion probability (PIP) for each SNP. We used PIP (the proportion of iterations in which a SNP is estimated to have a non-zero effect on phenotypic variation) to assess the significance of regions associated with jaw size variation. We performed 10 independent runs of the BSLMM using a step size of 100 million with a burn-in of 50 million steps for three traits (lower oral jaw size (n = 78), caudal fin pigmentation (n = 61), and nasal protrusion distance (n = 65)). We chose to only include SSI individuals in these analyses given extensive Caribbean-wide population structure that could confound significant associations (Fig. 1C). We summed PIP parameter estimates across 20-kb windows to avoid dispersion of the posterior probability density across SNPs in linkage disequilibrium due to physical linkage. Independent runs were consistent in reporting the strongest associations for the same 20-kb windows. We identified regions with strong association with our traits of interests as ones that had a PIP score in the 99th percentile of PIP scores across all regions (Data S7-9). Our PIP estimates for strongly associated windows suggest that jaw length may be controlled predominantly by a few loci of moderate effect (see bimodal PGE distribution, Fig. S10H). This is consistent with a previous QTL mapping study in an F2 intercross between SSI trophic specialists which detected one significant QTL with moderate effects on oral jaw size explaining up to 15% of the variation and three to four additional potential quantitative trait loci (QTL) with similar moderate effects (*38*).

QTL analysis for jaw size

We also investigated candidate variants for effects on craniofacial morphology by overlapping scaffolds with a previously published linkage map and QTL analysis of an F2 intercross between specialist species (*38*). We overlapped markers from this study that spanned the 95% Bayesian credible interval for a significant QTL for lower jaw length (LG15; taken from Fig S2 in (*38*)). The fasta sequences for these two markers bookending the QTL region on a single scaffold were then blasted against the *Cyprinodon brontotheriodes* genome using the blastn function in BLAST+ (*39*) and we selected the result with the highest percent identity and lowest e-value (Table S5). We then looked at all the genic regions within the interval between these two markers to investigate overlap between the QTL region and the candidate variants in this current study. The top hits for overlap between the sequences of two markers that spanned the LG15 QTL region and the *Cyprinodon brontotheroides* reference genome showed that this QTL corresponds to a 18 Mb region on scaffold c_bro_v1_0_ scaf8 (Table S5). However, this large region contained only a few candidate adaptive variants associated with the genes *map2k6* (3 variants), *galr2* (2 variants), and *grid2ip* (4 variants).

Timing of divergence for candidate adaptive variants

If adaptive diversification in this radiation of pupfishes followed the temporal stages hypothesis (habitat preference first, then trophic morphology, and finally sexual communication; (*40*)), we predicted an ordering of divergence times among candidate regions containing genes annotated for traits related to these axes of diversification, with regions containing genes related to habitat preference having older divergence times, followed by regions with genes related to trophic morphology, and finally regions with pigmentation genes having the youngest divergence times. In order to determine if there have been stages of adaptive diversification in this adaptive

radiation of pupfishes, we first estimated divergence times between molluscivores and scale-eaters for each candidate adaptive region (i.e. each 50 kb window containing signatures of a hard selective sweep in either specialist and a nearly fixed variant between specialists).

Many methods for estimating divergence times and allele ages rely on the pattern of variation in the haplotype background surrounding the allele of interest. Heuristic approaches, particularly those that use point estimates of number of derived mutations within a chosen distance of the site are accessible, quick ways to approximate divergence times among regions and allele ages without extensive haplotype data (*41*, *42*). We estimated sequence divergence in regions surrounding candidate variants using D_{xy}, an absolute measure of genetic divergence. To get a heuristic estimate of divergence time between specialists at these regions, we used this D_{xy} count of the number of variants that have accumulated between specialists and the approximation that the observed genetic differences between two lineages should be equal to $2\mu t$: t, the time since their divergence and μ, the mutation rate (*43*). Using the per generation mutation rate estimated above (1.56×10^{-8}), we calculated the time since divergence for candidate regions and compared that time to the estimated 10-15 kya age of the radiation (based on estimates of the last period of drying of hypersaline lake basins on San Salvador Island during the last glacial maximum (*44*)).

We plotted the divergence time estimates for candidate adaptive regions by the spatial categories they were assigned to (de novo, introgression, standing genetic variation), and against a background of neutral regions that contained a nearly fixed variant, but no signature of a hard selective sweep (Fig 4, S5 and S11). So that each point in the plot was independent, we pruned variants by randomly selecting one from the group of variants that fell within the same 50-kb window. Some windows had multiple variants with different spatial distributions (e.g. de novo

vs. standing genetic variation), so to make sure all points on the figure were independent, we made alternative plots for alternative spatial distributions of variants that occurred within a single 50 kb window (the smaller vs larger spatial distribution; Fig. 4 and Fig. S11). This applied to several variants that were characterized as either introgressed or standing genetic variation in four regions containing genes with relevant adaptive annotations (*galr2, gmp6a, and kcnk2*) from the plot with the smaller spatial distribution (Fig. 4). The alternative, larger spatial distribution (i.e. the regions containing these genes were assigned to the standing genetic variation column instead of the int rogression column) for each of these four annotated 50 kb windows is shown in Figure S11.

To look for stages of diversification along different trait axes, we matched regions to potential phenotypes in two ways: 1) from our GO enrichment analyses that implicated genes in the region as relevant to the major axes of adaptive radiation in this system (e.g. craniofacial development, pigmentation, feeding behavior), and 2) with regions strongly associated with either lower jaw size, nasal protrusion distance, or caudal fin pigmentation in the GWAS for San Salvador Island pupfish species. Regions associated with traits in the GWAS were then polarized for effects on trait values that were in directions relevant to each specialist. For scale-eaters, this meant filtering for regions with significant association with darker caudal fin pigmentation, larger lower oral jaws, and smaller nasal protrusion distances. For molluscivores, this meant filtering for regions with significant associations for lighter caudal fin pigmentation, larger lower oral jaws, and larger nasal protrusion distances. We found 23 regions containing genes with relevant GO terms and 24 regions containing variants significantly associated with traits in the GWAS.

Based on relevant GO terms, we found that regions with genes annotated for feeding behavior have some of the oldest divergence times (Fig 4A and S11) while regions with genes related to craniofacial morphology and pigmentation had younger divergence times (Fig 4A and S11). Similarly, we found younger divergence times among regions with genes annotated for traits related to trophic morphology based on GWAS annotations (Fig 4B).

Considering there are only 23 and 24 observations used to test for stages of adaptive diversification, we also checked that this pattern was statistically robust and not due to chance. First, we used linear models to determine if divergence times among adaptive variant regions could be predicted by the associated adaptive traits. In our first linear model used to test for stages, major GO categories relevant to divergent traits in this system (feeding behavior, eye, mouth, muscle, and craniofacial development) were used as the predictor variable and the rank ordering of divergence times was the response variable. We found that relevant GO categories significantly predicted the rank ordering of divergence times among regions (df=5, F= 3.03, P-value= 0.039). From this, we then tested all pairwise comparisons of different categories of relevant traits to determine which were driving the significant difference in ordering using Tukey's HSD test. For the stages of trait diversification based on relevant GO terms (feeding behavior, eye, mouth, muscle, and craniofacial development): diversification in feeding behavior was significantly older than mouth development (P = 0.022), and moderately older than muscle and craniofacial development (P = 0.05 for both). This conservative test supports a 'behavior-first' ordering to trait diversification in this system (*45*). Divergence in behavioral traits as a major stage of radiation in pupfish makes sense because both specialists are more aggressive than generalist species (*46*), and adaptive kinematic traits important for scale-eating performance are behaviorally mediated (*47*).

867 Evidence for the behavior first stage of adaptive diversification is largely driven by the
868 fact that the two oldest divergences times occur in regions with genes annotated for feeding
869 behavior. Therefore, we further investigated the probability that this 'behavior first' pattern could
870 have occur by chance given our small sample size. We used a permutation test to calculate the
871 probability of the two regions with the oldest divergence times both being related to feeding
872 behavior can occur randomly. We first ranked the 23 candidate regions based on their
873 corresponding age and randomly reassigned these rankings for each region 10,000 times. We
874 then assessed how many of each of these iterations had the two regions with the oldest
875 divergence times assigned to feeding behavior (as observed in the empirical data). We found that
876 only 41 of the iterations matched the observed value, suggesting that the probability of observing
877 the two regions with the oldest divergence times related to feeding behavior by chance alone is
878 extremely small (*P-value* = 0.0041).
879 In our second linear model to test for stages of adaptive radiation, significant association
880 with three key adaptive traits measured for our GWAS (lower oral jaw size, caudal fin
881 pigmentation, and nasal protrusion distance) was used as the predictor variable and ordering of
882 divergence times was the response variable. We found that GWAS trait association did not
883 significantly predict the ordering of divergence times among regions (*df*=3, *F*= 1.67, *P*-value =
884 0.197). However, this is qualitatively similar to our results from the GO analyses considering
885 that we lack a behavioral trait related to feeding behavior in our GWAS analyses.
886
887 Timing of selection for candidate adaptive variants
888
889 We also estimated the age of hard selective sweeps using the R package *McSwan* (v1.1.1;;
890 https://github.com/sunyatin/McSwan; (*48*)). *McSwan* detects hard selective sweeps by comparing

local site frequency spectra (SFS) simulated under neutral and selective demographic models, which it uses to assign selective sweeps to regions of the genome and predict the age of selection events (*48*). By using information from the SFS, *McSwan* is advantageous for estimating selective sweep ages in non-model organisms because it does not require high quality haplotype data to detect sweeps and predict their ages. However, this flexibility comes at the cost of not jointly estimating the selection coefficient of a particular sweep, so it assumes the strength of selection is equal across all sweeps. With the input of a mutation rate estimate, neutral demographic model (effective population size changes, divergence events), and variant file, *McSwan* generates simulated and observed SFSs and a prior of sweep ages, whose upper bound is determine by the divergence time estimate specified in the demographic model (in our case: 10,000 years). *McSwan* uses these simulated selective and neutral SFSs to scan the input variant file for selective sweep regions and produce a posterior distribution of sweep ages for each sweep region it detects.

To simulate SFSs, we used the mutation rate estimated above ($1.56x10^{-8}$) , the same demographic models of changes in effective population sizes used in our SweeD runs for the generalists and scale-eater populations (Table S7), and a divergence time estimate between San Salvador Island generalist and scale-eater of 10,000 years (based on dating of last glacial maximum when the hypersaline lakes were dry on San Salvador Island). We first simulated neutral and selection SFSs that were each comprised of 2,000 simulations (default recommendation) across sequences 50-kb in length. To look for selective sweeps in the specialists, we then generated empirical SFSs from scans across the 500-kb region surrounding each of the 23 sets of candidate variants highlighted in Figure 4. In order to precisely determine the boundaries of hard selective sweep, McSwan iterates its genomic scans over adjacent

914 windows of various lengths and offsets and compares the empirical SFS to the simulated SFS
915 under selection to assign regions as selective sweeps. We set up the iterative scans across these
916 500-kb regions in sliding windows that ranged from 1000 bp to 200-kb in length and a minimum
917 of 50 variants required per window. Each sliding scan of the 500-kb region was done in 100
918 overlapping steps (default setting). We then looked for overlap between the regions detected as
919 hard selective sweeps by McSwan to the candidate regions previously detected with SweeD and
920 F_{st} (Table S2-S3). For the scale-eater population only 9 of the 23 sets of candidate variants
921 detected as hard selective sweeps using SweeD were also detected as hard selective sweeps using
922 *McSwan* and given age estimates. In Tournebize et al. 2019 (*48*), investigations of McSwan's
923 performance noted poor performance to detect selective sweeps when selection was relatively
924 weak ($s \leq 0.05$) and recent (≤ 16 kya; Supplemental information Section 2 of (*48*)). Thus, the
925 thirteen additional candidate adaptive regions undetected by McSwan may be under weaker
926 selection or more recent. The candidate variants surrounding the consecutively spaced genes
927 *cenpf* and *kcnk2* were detected as within the same selective sweep in *McSwan* and thus have the
928 same age estimates (Fig. 4C and Fig. S12).

929 For these 8 regions, we then filtered these distributions of sweep ages for estimates that
930 had a stability value (a parameter that represents the strength of support for a selective sweep
931 model over a neutral model) in the 95th percentile. To get a likely range of selective sweep age
932 estimates for each region, we calculated the 95% high posterior density (HPD) region with the R
933 package HDIntervals (v0.2; https://cran.r-project.org/web/packages/HDInterval/index.html) from
934 their respective posterior distributions. We repeated this process for the 6 sets of candidate
935 regions found in the molluscivore, only three of which were also detected as being under a
936 selective sweep in *McSwan* and given age estimates. The 95% HPD of these age estimates for

the scale-eater and molluscivore populations are presented in Figure 4C, S5 and Table S14 and the full posteriors are shown in Figure S12 and S13.

References

1. A. A. Echelle, E. W. Carson, A. F. Echelle, D. Bussche, T. E. Dowling, Historical Biogeography of the New-World Pupfish Genus Cyprinodon (Teleostei: Cyprinodontidae). *Copeia*. **2005**, 320–339 (2005).
2. E. J. Richards, C. H. Martin, Adaptive introgression from distant Caribbean islands contributed to the diversification of a microendemic adaptive radiation of trophic specialist pupfishes. *PLoS Genet.* **13**, 1–35 (2017).
3. M. R. Green, J. Sambrook, Isolation of high-molecular-weight DNA using organic solvents. *Cold Spring Harb. Protoc.* **2017**, 356–359 (2017).
4. J. R. Wang, J. Holt, L. McMillan, C. D. Jones, FMLRC: Hybrid long read error correction using an FM-index. *BMC Bioinformatics*. **19**, 1–11 (2018).
5. R. L. Warren, C. Yang, B. P. Vandervalk, B. Behsaz, A. Lagman, S. J. M. Jones, I. Birol, LINKS: Scalable, alignment-free scaffolding of draft genomes with long reads. *Gigascience*. **4** (2015), doi:10.1186/s13742-015-0076-3.
6. N. C. Durand, M. S. Shamim, I. Machol, S. S. P. Rao, M. H. Huntley, E. S. Lander, E. L. Aiden, Juicer Provides a One-Click System for Analyzing Loop-Resolution Hi-C Experiments. *Cell Syst.* **3**, 95–98 (2016).
7. O. Dudchenko, S. S. Batra, A. D. Omer, S. K. Nyquist, M. Hoeger, N. C. Durand, M. S. Shamim, I. Machol, E. S. Lander, A. P. Aiden, E. L. Aiden, *Science (80-.).*, in press, doi:10.1126/science.aal3327.
8. R. Vaser, I. Sovic, N. Nagarajan, Š. Mile, Fast and accurate de novo genome assembly from long uncorrected reads. *Genome Res.*, 1–10 (2017).
9. F. A. Simão, R. M. Waterhouse, P. Ioannidis, E. V Kriventseva, E. M. Zdobnov, BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. *Bioinformatics*. **31**, 3210–3212 (2015).
10. B. L. Cantarel, I. Korf, S. M. C. Robb, G. Parra, E. Ross, B. Moore, C. Holt, A. S. Alvarado, M. Yandell, MAKER: An easy-to-use annotation pipeline designed for emerging model organism genomes. *Genome Res.* **18**, 188–196 (2008).
11. E. S. Lencer, W. C. Warren, R. Harrison, A. R. McCune, The Cyprinodon variegatus genome reveals gene expression changes underlying differences in skull morphology among closely related species. *BMC Genomics*. **18**, 424 (2017).
12. H. Li, R. Durbin, Inference of human population history from individual whole-genome sequences. *Nature*. **475**, 493–496 (2011).
13. M. A. DePristo, E. Banks, R. Poplin, K. V Garimella, J. R. Maguire, C. Hartl, A. A. Philippakis, G. del Angel, M. A. Rivas, M. Hanna, A. McKenna, T. J. Fennell, A. M. Kernytsky, A. Y. Sivachenko, K. Cibulskis, S. B. Gabriel, D. Altshuler, M. J. Daly, A framework for variation discovery and genotyping using next-generation DNA sequencing data. *Nat. Genet.* **43**, 491–8 (2011).
14. C. D. Marsden, Y. Lee, K. Kreppel, A. Weakley, A. Cornel, H. M. Ferguson, E. Eskin, G.

C. Lanzaro, Diversity, differentiation, and linkage disequilibrium: prospects for association mapping in the malaria vector Anopheles arabiensis. *G3 (Bethesda).* **4**, 121–31 (2014).

15. P. Danecek, A. Auton, G. Abecasis, C. A. Albers, E. Banks, M. A. DePristo, R. E. Handsaker, G. Lunter, G. T. Marth, S. T. Sherry, G. McVean, R. Durbin, 1000 Genomes Project Analysis Group, The variant call format and VCFtools. *Bioinformatics*. **27**, 2156–2158 (2011).
16. S. Purcell, B. Neale, K. Todd-Brown, L. Thomas, M. A. R. Ferreira, D. Bender, J. Maller, P. Sklar, P. I. W. de Bakker, M. J. Daly, P. C. Sham, PLINK: A Tool Set for Whole-Genome Association and Population-Based Linkage Analyses. *Am. J. Hum. Genet.* **81**, 559–575 (2007).
17. D. H. Alexander, J. Novembre, K. Lange, Fast model-based estimation of ancestry in unrelated individuals. *Genome Res.* **19**, 1655–1664 (2009).
18. S. H. Martin, J. W. Davey, C. D. Jiggins, Evaluating the use of ABBA-BABA statistics to locate introgressed loci. *Mol. Biol. Evol.* **32**, 244–257 (2015).
19. F. L. Wu, A. Strand, C. Ober, J. D. Wall, P. Moorjani, M. Przeworski, A comparison of humans and baboons suggests germline mutation rates do not track cell divisions. *bioRxiv*, 844910 (2019).
20. M. Malinsky, H. Svardal, A. M. Tyers, E. A. Miska, M. J. Genner, G. F. Turner, R. Durbin, Whole-genome sequences of Malawi cichlids reveal multiple radiations interconnected by gene flow. *Nat. Ecol. Evol.* **2**, 1940–1955 (2018).
21. C. Feng, M. Pettersson, S. Lamichhaney, C. J. Rubin, N. Rafati, M. Casini, A. Folkvord, L. Andersson, Moderate nucleotide diversity in the Atlantic herring is associated with a low mutation rate. *Elife*. **6**, 1–14 (2017).
22. A. F. Kautt, G. Machado-Schiaffino, A. Meyer, Multispecies Outcomes of Sympatric Speciation after Admixture with the Source Population in Two Radiations of Nicaraguan Crater Lake Cichlids. *PLoS Genet.* **12**, 1–33 (2016).
23. B. Guo, F. J. J. Chain, E. Bornberg-Bauer, E. H. Leder, J. Merilä, Genomic divergence between nine- and three-spined sticklebacks. *BMC Genomics*. **14**, 756 (2013).
24. C. H. Martin, J. E. Crawford, B. J. Turner, L. H. Simons, Diabolical survival in Death Valley: recent pupfish colonization, gene flow and genetic assimilation in the smallest species range on earth. *Proc. R. Soc. B Biol. Sci.* **283**, 20152334 (2016).
25. C. H. Martin, K. Gould, C. Bocklage, Surprising spatiotemporal stability and frequency-independence across multiple fitness peaks driving adaptive radiation in the wild (2019).
26. S. Schiffels, R. Durbin, Inferring human population size and separation history from multiple genome sequences. *Nat. Genet.* **46**, 919 (2014).
27. K. Nadachowska-Brzyska, R. Burri, L. Smeds, H. Ellegren, PSMC analysis of effective population sizes in molecular ecology and its application to black-and-white Ficedula flycatchers. *Mol. Ecol.* **25**, 1058–1072 (2016).
28. J. K. Pickrell, J. K. Pritchard, Inference of Population Splits and Mixtures from Genome-Wide Allele Frequency Data. *PLoS Genet.* **8**, e1002967 (2012).
29. D. Garrigan, A. Geneva, msmove: A modified version of Hudson's coalescent simulator ms allowing for finer control and tracking of migrant genealogies (2014), , doi:10.6084/m9.figshare.1060474.
30. P. Pavlidis, D. Živković, A. Stamatakis, N. Alachiotis, SweeD: Likelihood-based detection of selective sweeps in thousands of genomes. *Mol. Biol. Evol.* **30**, 2224–2234

(2013).
31. E. J. Richards, M. R. Servedio, C. H. Martin, Searching for Sympatric Speciation in the Genomic Era. *BioEssays*. **41**, 1900047 (2019).
32. S. X. Ge, D. Jung, ShinyGO: a graphical enrichment tool for animals and plants. *bioRxiv*, 2 (2018).
33. E. Balsa-Canto, D. Henriques, A. Gabor, J. R. Banga, AMIGO2, a toolbox for dynamic modeling, optimization and control in systems biology. *Bioinformatics*. **32**, 1–2 (2016).
34. M. E. Anderson, J. Runesson, I. Saar, Ü. Langel, J. K. Robinson, Galanin, through GalR1 but not GalR2 receptors, decreases motivation at times of high appetitive behavior. *Behav. Brain Res.* **239**, 90–93 (2013).
35. S. Wang, L. Ghibaudi, T. Hashemi, C. He, C. Strader, M. Bayne, H. Davis, J. J. Hwa, The GalR2 galanin receptor mediates galanin-induced jejunal contraction, but not feeding behavior, in the rat: Differentiation of central and peripheral effects of receptor subtype activation. *FEBS Lett.* **434**, 277–282 (1998).
36. J. A. Mcgirr, C. H. Martin, Parallel evolution of gene expression between trophic specialists despite divergent genotypes and morphologies, 62–75 (2018).
37. X. Zhou, P. Carbonetto, M. Stephens, Polygenic modeling with bayesian sparse linear mixed models. *PLoS Genet.* **9**, e1003264 (2013).
38. C. H. Martin, P. A. Erickson, C. T. Miller, The genetic architecture of novel trophic specialists: higher effect sizes are associated with exceptional oral jaw diversification in a pupfish adaptive radiation. *Mol. Ecol.* **26**, 624–638 (2017).
39. C. Camacho, G. Coulouris, V. Avagyan, N. Ma, J. Papadopoulos, K. Bealer, T. L. Madden, BLAST+: Architecture and applications. *BMC Bioinformatics*. **10**, 1–9 (2009).
40. T. J. Streelman, P. D. Danley, The stages of vertebrate evolutionary radiation. *Trends Ecol. Evol.* **18**, 126–131 (2003).
41. R. R. Hudson, The Variance of Coalescent Time Estimates from DNA Sequences. *J. Mol. Evol.* **64**, 702–705 (2007).
42. G. Coop, K. Bullaughey, F. Luca, M. Przeworski, The Timing of Selection at the Human FOXP2 Gene. *Mol. Biol. Evol.* **25**, 1257–1259 (2008).
43. N. Masatoshi, Genetic distance between populations. *Am. Nat.* **106**, 283–292 (1972).
44. F. M. Hagey, J. . Mylroie, Pleistocene lake and lagoon deposits, San Salvador Island, Bahamas. *Geol. Soc. Am.* **30**, 77–90 (1995).
45. R. B. Huey, P. E. Hertz, B. Sinervo, Behavioral drive versus behavioral inertia in evolution: A null model approach. *Am. Nat.* **161**, 357–366 (2003).
46. M. E. St. John, J. A. McGirr, C. H. Martin, The behavioral origins of novelty: did increased aggression lead to scale-eating in pupfishes? *Behav. Ecol.* (2018), doi:10.1093/beheco/ary196.
47. M. E. St. John, C. H. Martin, Scale-eating specialists evolved adaptive feeding kinematics within a microendemic radiation of San Salvador Island pupfishes. *bioRxiv*. **648451**, 648451 (2019).
48. R. Tournebize, V. Poncet, M. Jakobsson, Y. Vigouroux, S. Manel, McSwan: A joint site frequency spectrum method to detect and date selective sweeps across multiple population genomes. *Mol. Ecol. Resour.* **19**, 283–295 (2019).
49. M. M. Stevenson, Karyomorphology of Several Species of Cyprinodon. *Copeia*, 494–498 (1981).
50. Y. Ofran, B. Rost, ISIS: Interaction sites identified from sequence. *Bioinformatics*. **23**,

13–16 (2007).

51. E. V. Davydov, D. L. Goode, M. Sirota, G. M. Cooper, A. Sidow, S. Batzoglou, Identifying a high fraction of the human genome to be under selective constraint using GERP++. *PLoS Comput. Biol.* **6** (2010), doi:10.1371/journal.pcbi.1001025.

52. M. Nei, W. H. Li, *Proc. Natl. Acad. Sci.*, in press, doi:10.1073/pnas.76.10.5269.

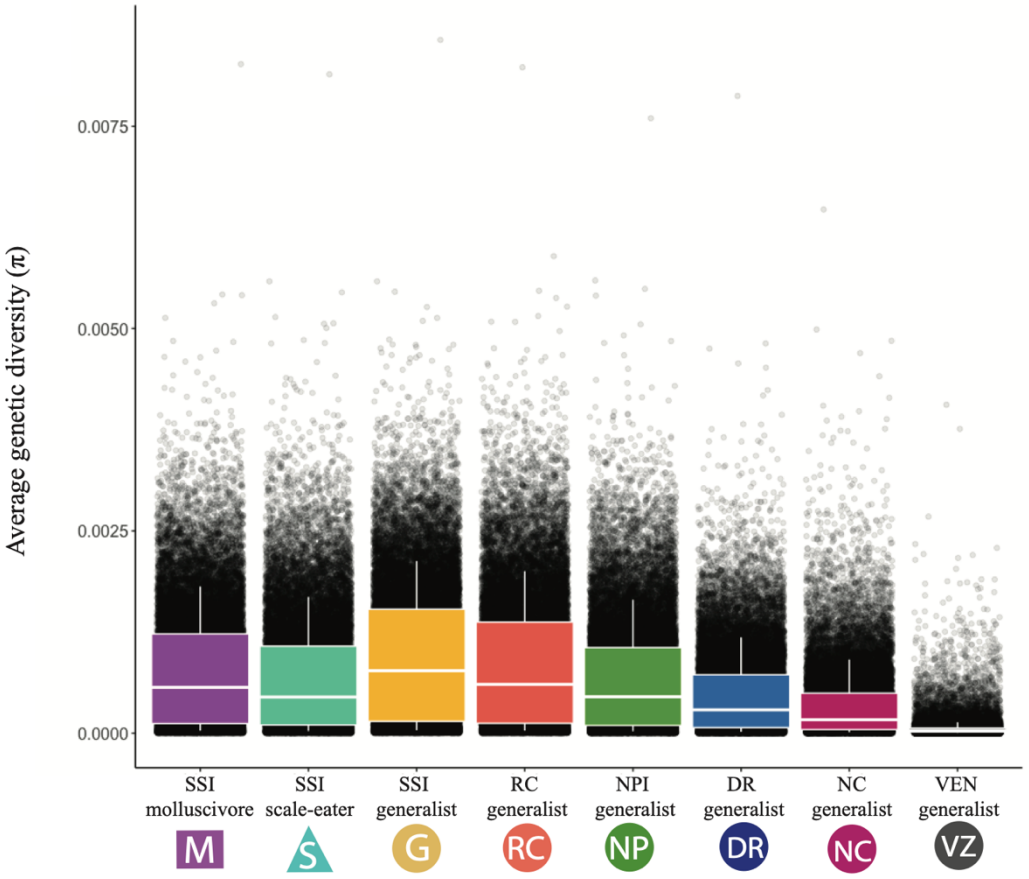


Fig. S1.

Average genetic diversity is similar across Caribbean pupfish populations. Within population (π) nucleotide diversity in 50-kb sliding windows across the genomes of the San Salvador Island (SSI) species and generalist species on Rum Cay (RC), New Providence Island (NPI), Dominican Republic (DR), North Carolina (NC) and Venezuela (VZ). π values are averaged across 100 random samples of 10 individuals from each population in order to down-sample from populations with larger sample sizes and compare π across populations.

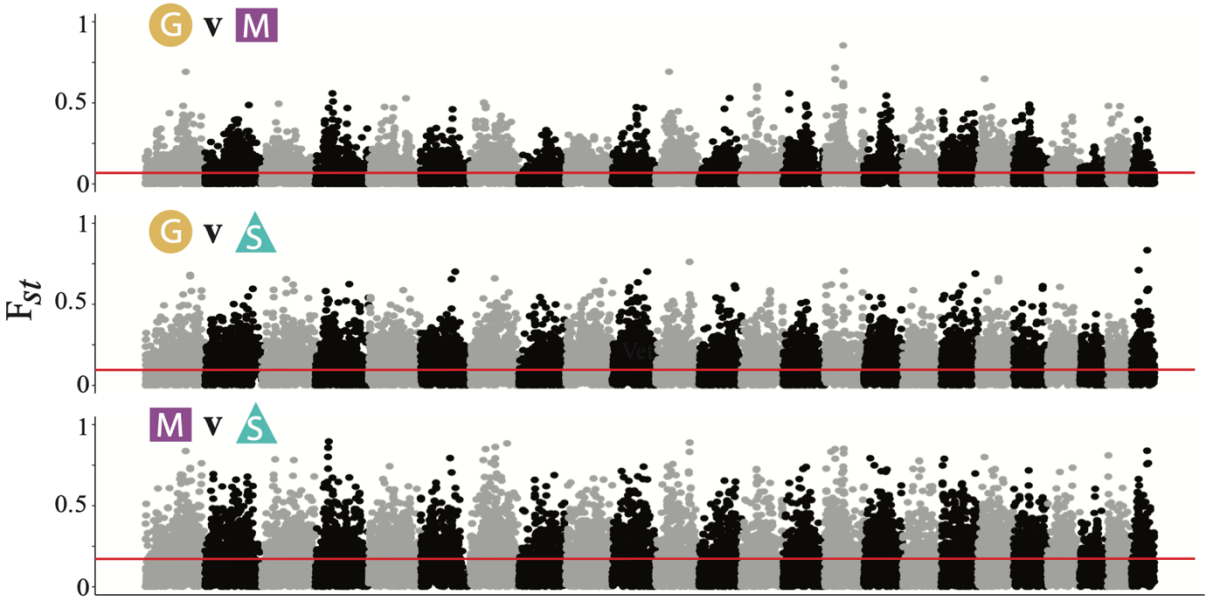


Fig. S2.

Genetic divergence among SSI species. Manhattan plot of F_{st} in 50-kb windows across the genome for the three SSI species on the largest 24 scaffolds in the molluscivore (*C. brontotheroides*) genome corresponding to the 24 chromosomes in *Cyprinodon* (*49*). Solid red line represents the average F_{st} values for each comparison (generalist vs. molluscivore; 0.07; generalist vs. scale-eater: 0.11; molluscivore vs. scale-eater: 0.15).

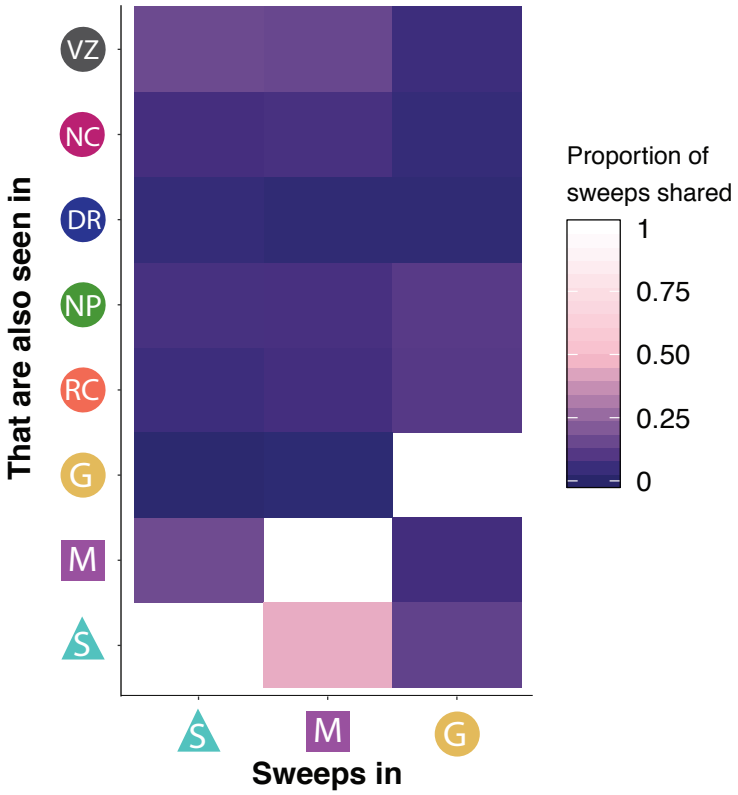


Fig. S3.

Selective sweeps in SSI population shared with other Caribbean populations. The proportion of hard selective sweeps in the SSI species that are also found sweeping in other Caribbean populations. Note that 42% of selective sweeps in the molluscivore population also showed signs of a sweep in the scale-eater population.

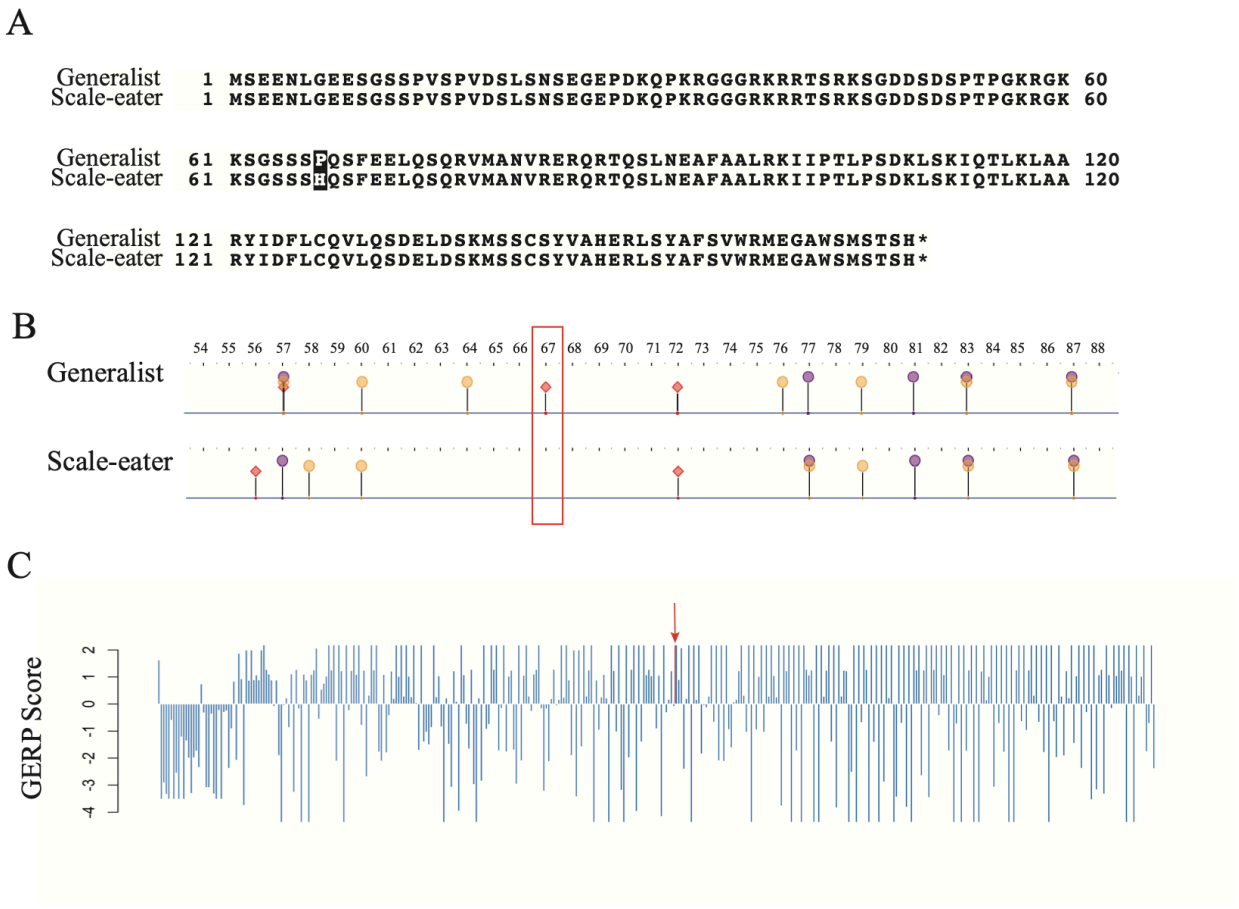
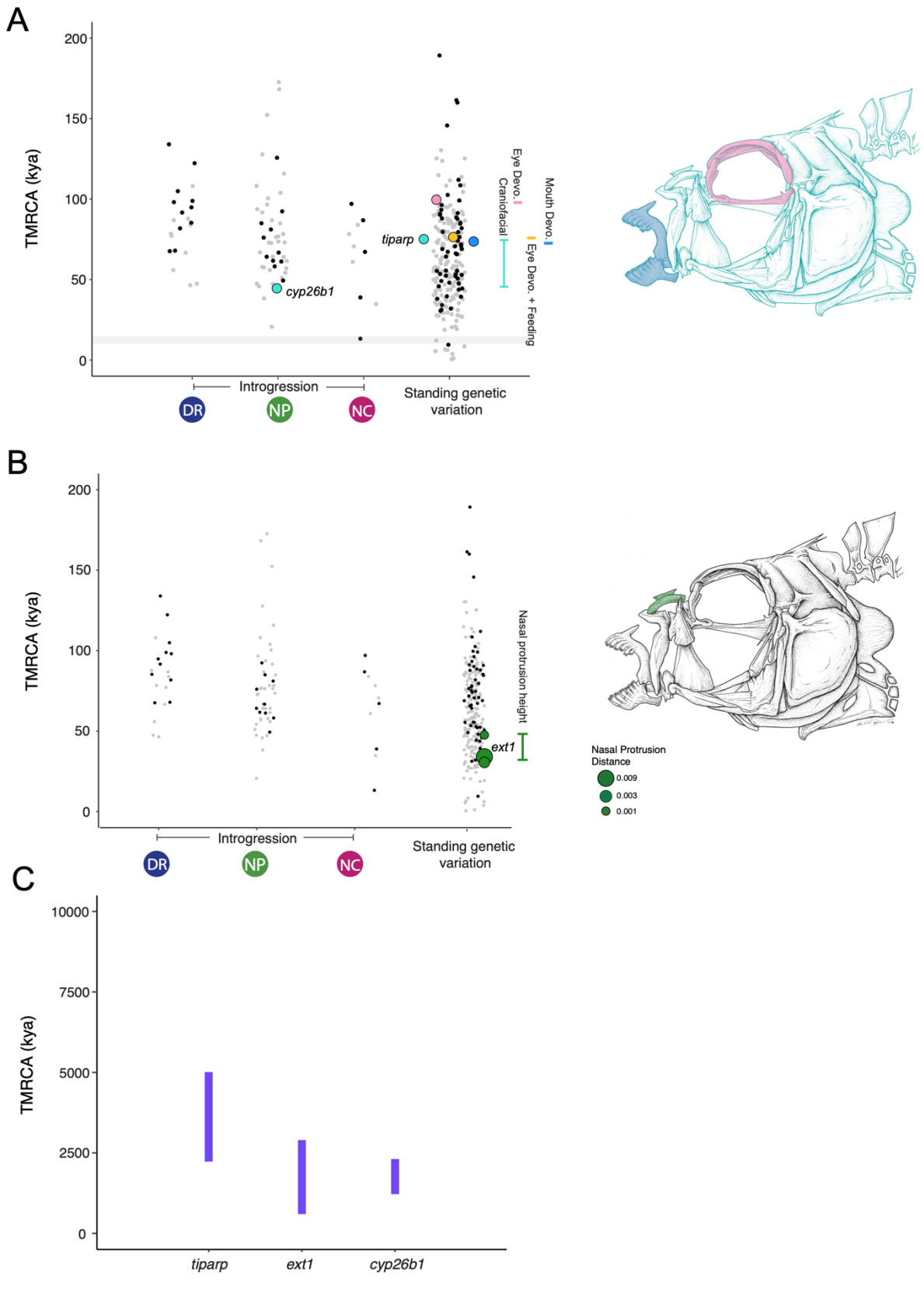


Fig. S4.

Sequence conservation among fishes around candidate gene *twist1*. A) Amino acid sequence of *twist1* protein for SSI generalists and scale-eaters. The non-synonymous substitution that is nearly fixed between the two species changes the amino acid from a proline to histidine (highlighted in black). B) This amino acid substitution alters a protein binding site (highlighted in red box) predicted and visualized with Predict Protein Open (https://open.predictprotein.org) using the machine-learning prediction method PPsites2 (*50*). C) GERP scores for the 500 base pair region surrounding the non-synonymous coding substitution in *twist1* (red arrow) found only on SSI. Conservation scores were obtained from aligning scale-eater genomes to the 60 fish EPO low coverage genome alignment on Ensembl (release 98) and a conservation score above 2 is considered highly conserved (*51*).



1113

Fig. S5.

The spatiotemporal landscape of genetic variation for the molluscivore. Time to most recent common ancestor (TMRCA) of the region surrounding candidate adaptive variants (LD-pruned so that each point is independent) based on relative genetic divergence metric *Da* (*52*) which captures only the amount divergence that has accumulated since the two populations diverged. Variants are separated by spatial distribution into: introgressed from a single population and standing genetic variation observed in more than two populations. No variants existed as de novo variation in the molluscivore populations on SSI. No nearly fixed variants introgressed from the Venezuela population. Points are colored by their adaptive relevance to this system (black points: adaptive variants annotated for non-focal GO terms or unannotated; gray points: nearly fixed variants between specialists that are not under hard selective sweeps). A) All variants annotated for GO categories from the Biological Processes database (Ensemble 92) for feeding behavior, muscle, eye, and mouth development, or craniofacial (mouth and/or eye, muscle) are shown. B) All variants in the 99th percentile of PIP scores for association with either smaller oral jaw size, lighter caudal fin pigmentation, or premaxillary nasal protrusion from GWAS are shown. Dot sizes represent PIP score of each GWAS hit within a trait. C) The 95% high posterior density of the posterior distribution of ages of selective sweeps in focal gene regions. Focal regions are colored based on GO and GWAS annotations relevant to the stages of adaptive radiation hypothesis: habitat preference (feeding behavior: red), trophic morphology (craniofacial and muscle: blue-violet), and sexual communication (pigmentation: orange).

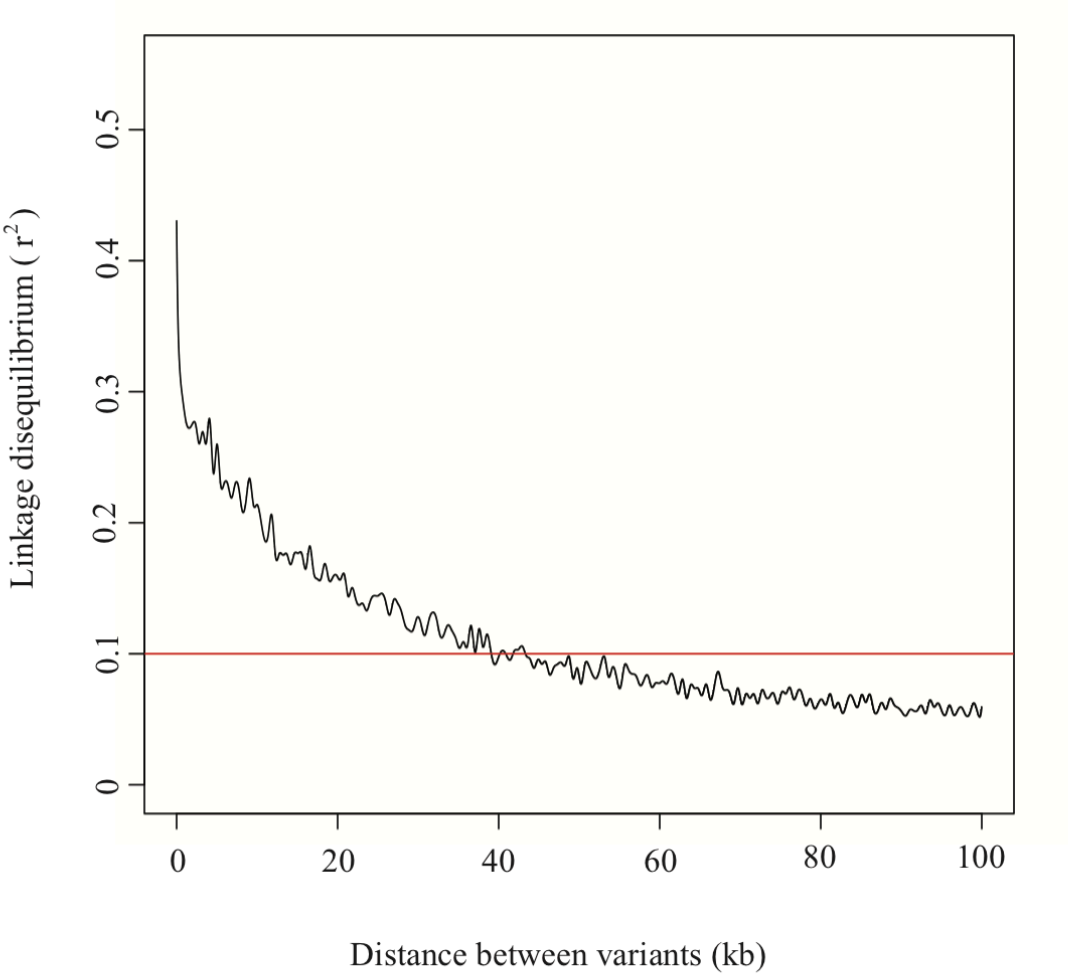


Fig. S6.

Linkage disequilibrium decay along the genome. LD decay over pairwise combinations of variants within 100 kb of each other on the longest scaffold in the genome (49,059,223 bp), with r^2=0.1 marked for reference. From this pattern of decay, we chose a window size of 50-kb for sliding windows analyses used in this study.

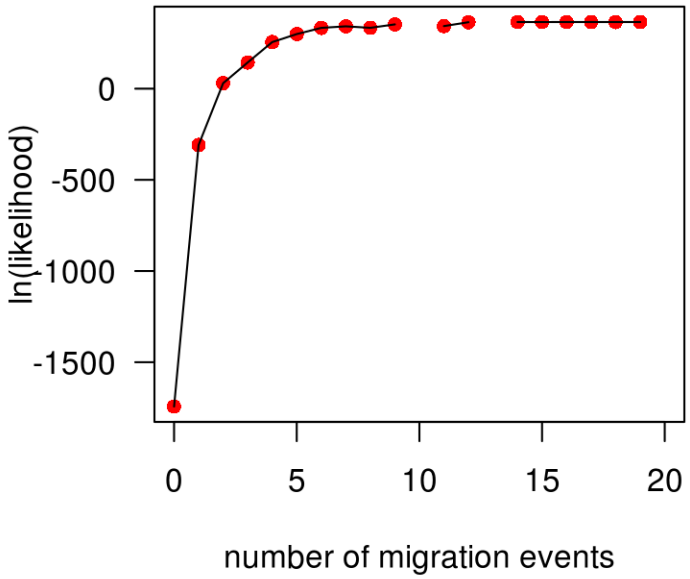


Fig. S7.

The log likelihood of number of migration events on the population graph inferred using TREEMIX (*28*)**.** The rate of change in the likelihood began to decline after three migration events, so three migration arrows were included in the population graph in Fig. 4B.

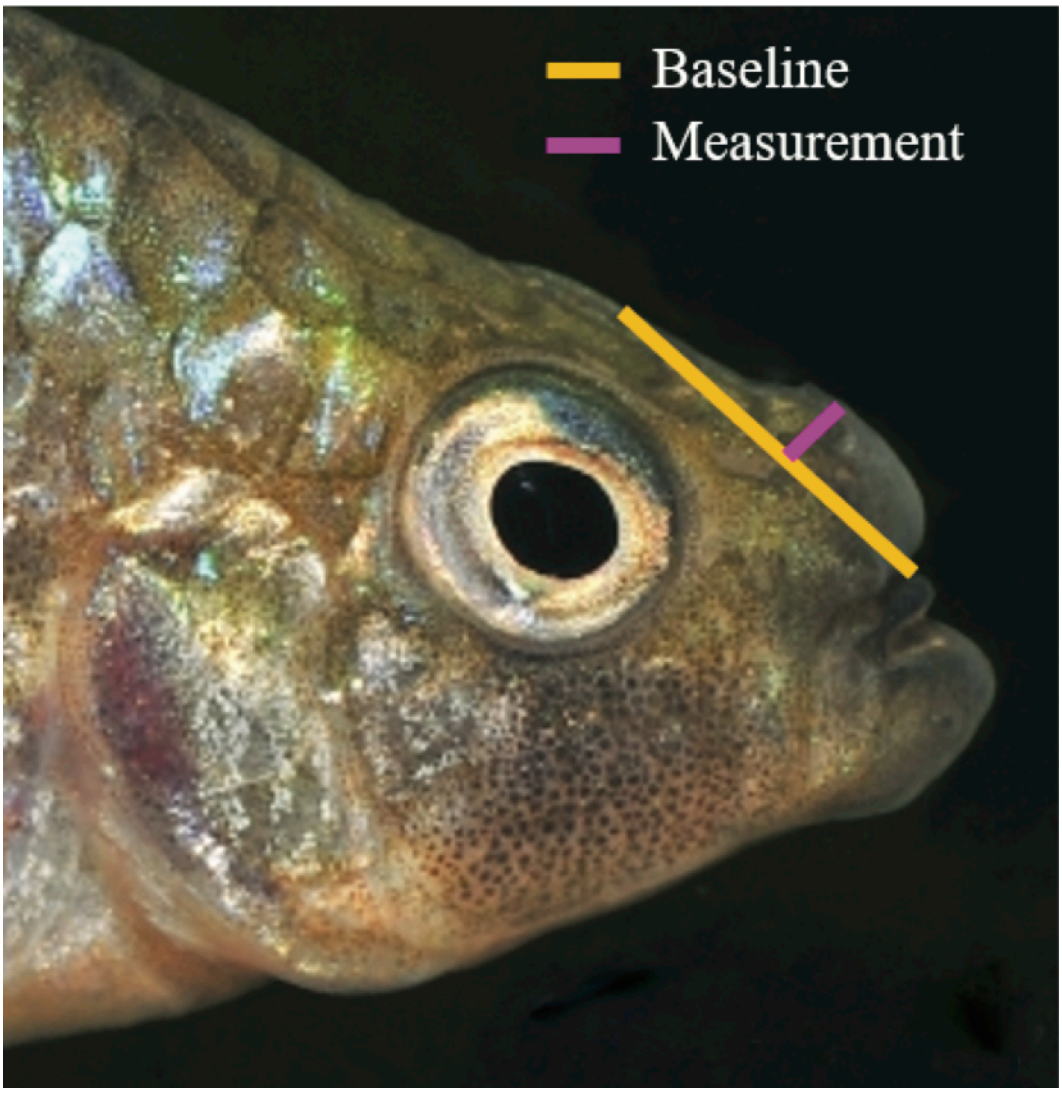


Fig S8. Example image of nasal protrusion distance measurement for GWAS. The Purple line represents the nasal protrusion distance. The yellow line represents a baseline tangent line from the dorsal surface of the neurocranium to the tip of the premaxilla used for reference.

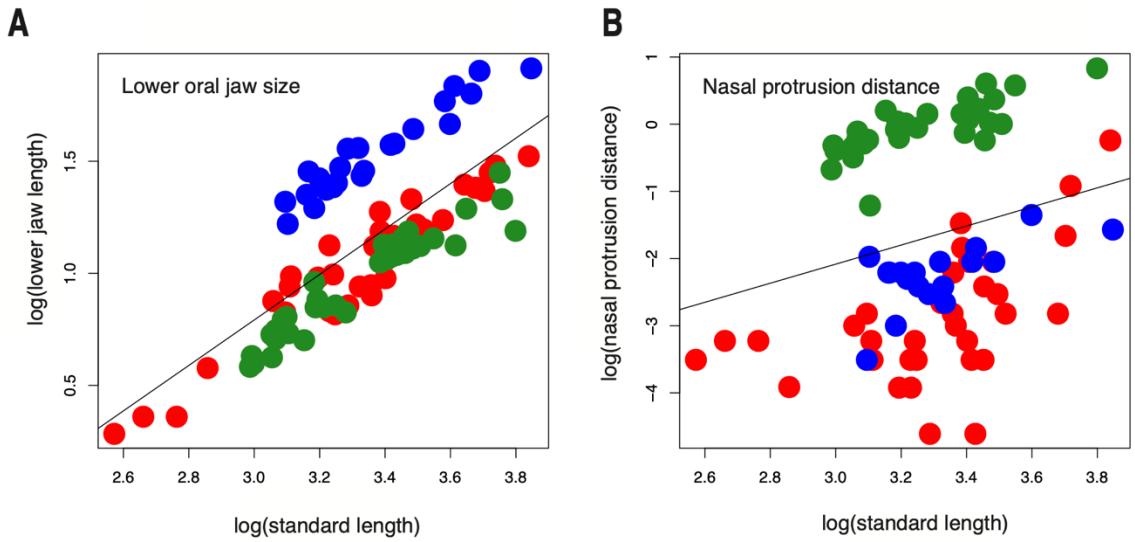


Fig. S9. Standardized craniofacial trait measurements in SSI species. Log-transformed A) lower oral jaw length (mm) and B) nasal protrusion distance (mm) standardized by log-transformed standard length (mm) for SSI generalist (red), molluscivore (green), and scale-eater (blue).

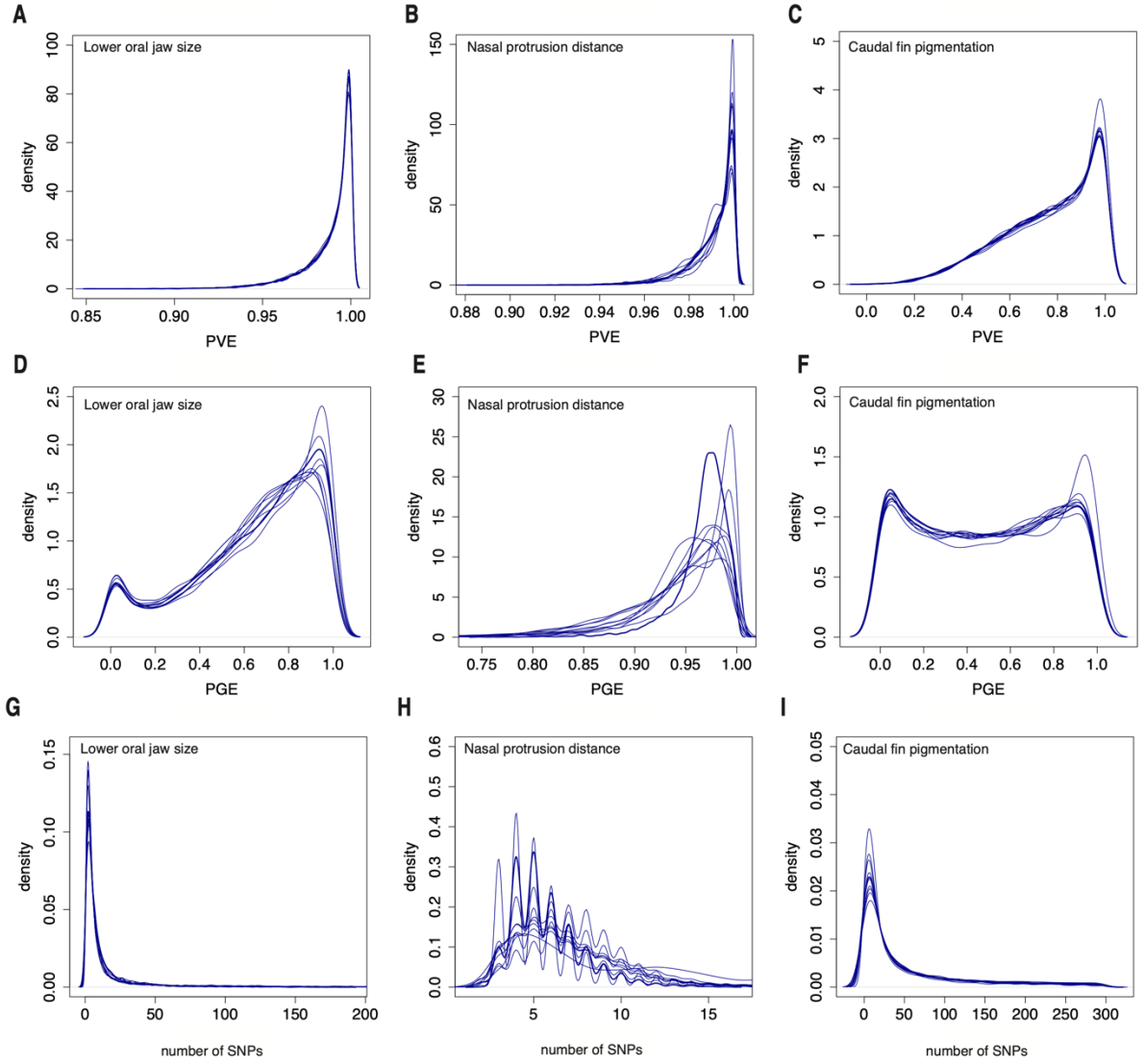


Fig. S10.

Posterior density distributions for hyper-parameters describing the proportion of variance in phenotypes for the three focal traits (lower jaw size, nasal protrusion distance, and caudal fin pigmentation) explained by A-C) every SNP (proportion of phenotype variance explained (PVE)), D-F) SNPs of large effect (proportion of genetic variance explained by sparse effects) PGE)), and G-H) the number of large effect SNPS required to explain PGE. Individual lines represent 10 independent MCMC runs of GEMMA's Bayesian sparse linear mixed model.

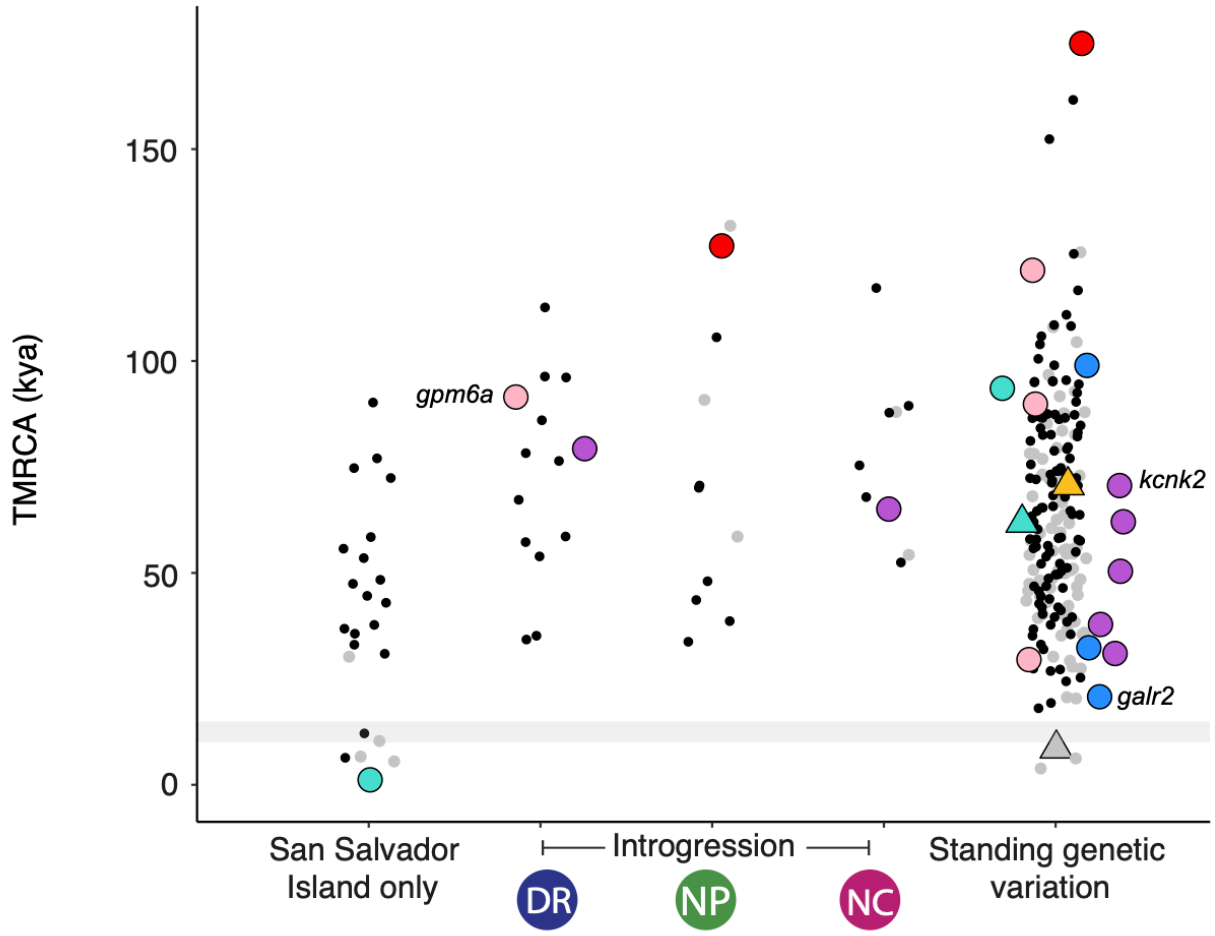


Fig. S11. The spatiotemporal landscape of adaptive radiation from the perspective of the alternative allele (larger spatial distribution). Time to most recent common ancestor (TMRCA) of the region surrounding candidate adaptive variants (LD-pruned so that each point is independent) based on relative genetic divergence metric *Dxy* (*52*) which captures only the amount divergence that has accumulated since the two populations diverged. Variants are separated by spatial distribution into: de novo, introgressed from a single population and standing genetic variation observed in more than two populations. No nearly fixed variants introgressed from the Venezuela population. Points are colored by significantly enriched GO terms chosen for their adaptive relevance to this system (black points: adaptive variants annotated for non-focal GO terms or unannotated; gray points: nearly fixed variants between specialists that are not under hard selective sweeps) and triangle points represent those variants also associated with pigmentation. All variants annotated for GO categories from the Biological Processes database (Ensemble 92) for feeding behavior, muscle, eye, and mouth development, or craniofacial (mouth and/or eye, muscle) are shown. Points labeled with gene names indicate the three regions which have alternative spatial distributions shown in Fig. 4.

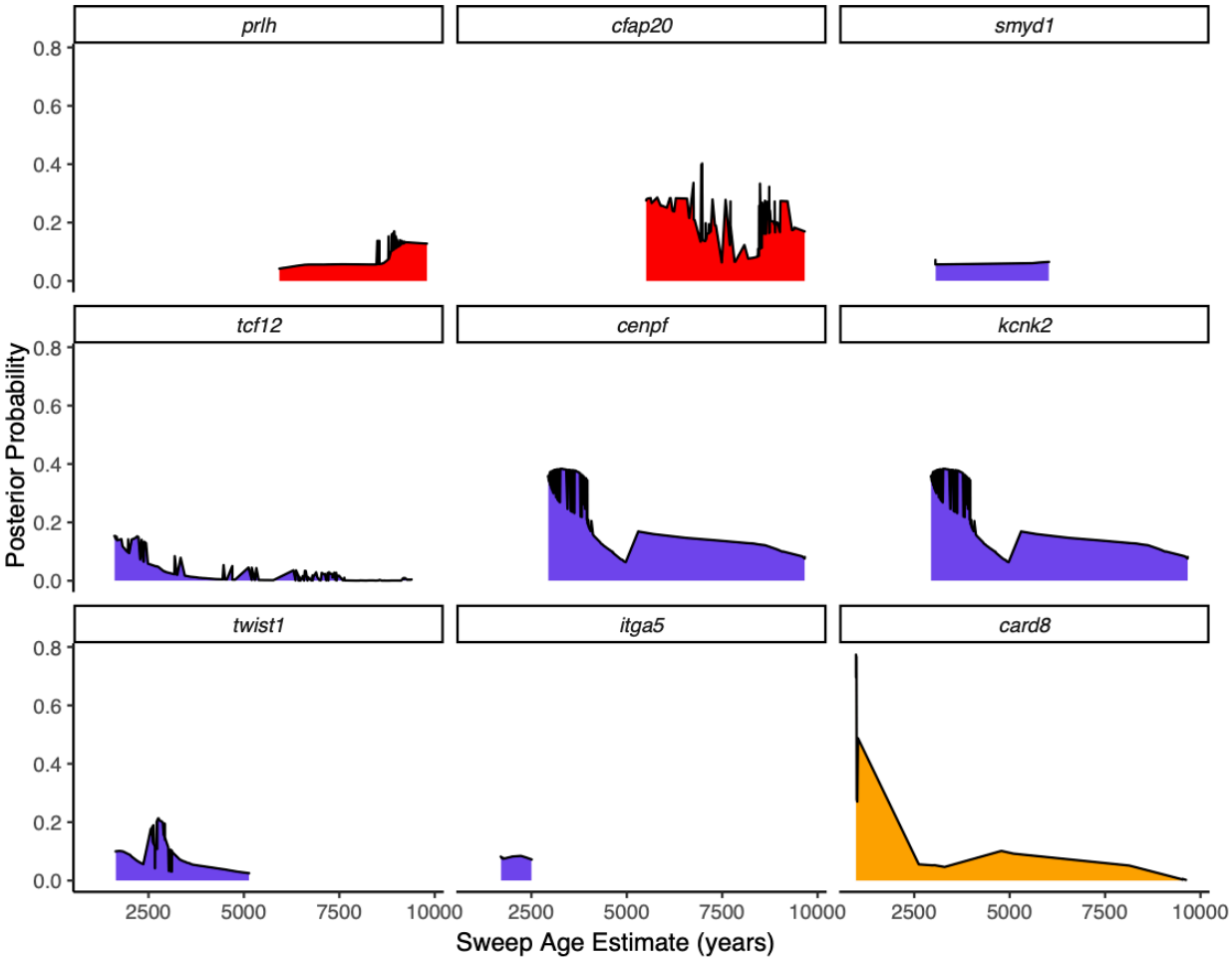


Fig. S12. Posterior distributions for scale-eater sweeps. The posterior distributions of sweep ages estimated from focal regions (Table S13). These nine regions were found under a hard selective sweep using both SweeD and McSwan. Posterior distributions of ages were calculated in McSwan. Focal regions are colored based on GO and GWAS annotations relevant to the stages proposed in the stages of adaptive radiation hypothesis: habitat preference (feeding behavior: red), trophic morphology (craniofacial and muscle: blue-violet), and sexual communication (pigmentation: orange).

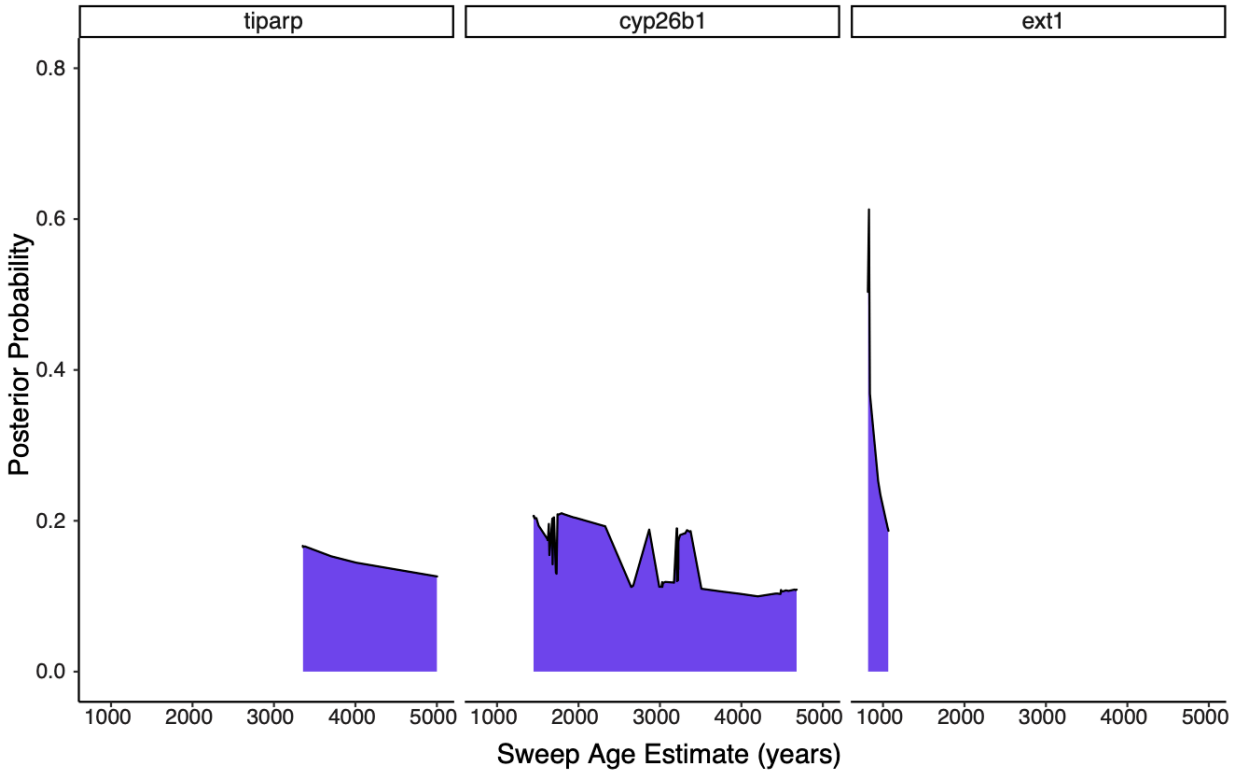


Fig. S13. Posterior distributions for molluscivore sweeps. The posterior distributions of sweep ages estimated from focal regions (Table S13). These three regions were found under a hard selective sweep using both SweeD and McSwan. Posteriors distributions of ages were calculated in McSwan. Focal regions are colored based on GO and GWAS annotations relevant to the stages proposed in the stages of adaptive radiation hypothesis: trophic morphology (craniofacial and muscle: blue-violet).

Table S1.

Summary of Caribbean pupfish sampling. The sampling localities of individuals that were whole genome resequenced from San Salvador Island radiation (SSI), other *Cyprinodon* across the Caribbean, Mexico, and United States, and two outgroups. Full details including sample codes, collector identities, GPS coordinates are included in Data S1 table.

Group	Species	Lake/Site	Island/Nation	Sample size
SSI generalist	*Cyprinodon variegatus*	Clear Pond	San Salvador Island, Bahamas	1
SSI generalist	*Cyprinodon variegatus*	Crescent Pond	San Salvador Island, Bahamas	4
SSI generalist	*Cyprinodon variegatus*	Granny Lake	San Salvador Island, Bahamas	1
SSI generalist	*Cyprinodon variegatus*	Great Lake	San Salvador Island, Bahamas	2
SSI generalist	*Cyprinodon variegatus*	Little Lake	San Salvador Island, Bahamas	1
SSI generalist	*Cyprinodon variegatus*	Stout's Pond	San Salvador Island, Bahamas	2
SSI generalist	*Cyprinodon variegatus*	Mermaid Pond	San Salvador Island, Bahamas	1
SSI generalist	*Cyprinodon variegatus*	Moon Rock Pond	San Salvador Island, Bahamas	1
SSI generalist	*Cyprinodon variegatus*	North Little Lake	San Salvador Island, Bahamas	1
SSI generalist	*Cyprinodon variegatus*	Osprey Lake	San Salvador Island, Bahamas	12
SSI generalist	*Cyprinodon variegatus*	Oyster Lake	San Salvador Island, Bahamas	2
SSI generalist	*Cyprinodon variegatus*	Oyster Lake	San Salvador Island, Bahamas	1
SSI generalist	*Cyprinodon variegatus*	Pain Pond	San Salvador Island, Bahamas	1
SSI generalist	*Cyprinodon variegatus*	Reckley Hill Pond	San Salvador Island, Bahamas	1
SSI generalist	*Cyprinodon variegatus*	Six Pack Pond	San Salvador Island, Bahamas	1
SSI generalist	*Cyprinodon variegatus*	Wild Dilly Pond	San Salvador Island, Bahamas	1
SSI molluscivore	*Cyprinodon brontotheroides*	Crescent Pond	San Salvador Island, Bahamas	12
SSI molluscivore	*Cyprinodon brontotheroides*	Little Lake	San Salvador Island, Bahamas	5
SSI molluscivore	*Cyprinodon brontotheroides*	Moon Rock Pond	San Salvador Island, Bahamas	6
SSI molluscivore	*Cyprinodon brontotheroides*	Osprey Lake	San Salvador Island, Bahamas	12
SSI molluscivore	*Cyprinodon brontotheroides*	Oyster Pond	San Salvador Island, Bahamas	8
SSI scale-eater	*Cyprinodon desquamator*	Crescent Pond	San Salvador Island, Bahamas	10
SSI scale-eater	*Cyprinodon desquamator*	Little Lake	San Salvador Island, Bahamas	5
SSI scale-eater	*Cyprinodon desquamator*	Osprey Lake	San Salvador Island, Bahamas	10

SSI scale-eater	*Cyprinodon desquamator*	Oyster Lake	San Salvador Island, Bahamas	1
Dominican Republic	*Cyprinodon higuey*	Laguna Bavaro	Dominican Republic	10
New Providence Island	*Cyprinodon laciniatus*	Lake Cunningham	New Providence Island, Bahamas	16
Rum Cay	*Cyprinodon variegatus*	Lake George - main lake	Rum Cay, Bahamas	17
North Carolina	*Cyprinodon variegatus*	Fort Fisher estuary	NC, USA	11
Venezuela	*Cyprinodon dearborni*	Isla Margarita	Venezuela	11
Caribbean outgroup generalist	*Cyprinodon artifrons*	Cancun	Mexico	2
Caribbean outgroup generalist	*Cyprinodon variegatus*	North Salt Pond	Acklins Island, Bahamas	1
Caribbean outgroup generalist	*Cyprinodon dearborni*	--	Bonaire	1
Caribbean outgroup generalist	*Cyprinodon variegatus*	--	Caicos Island	1
Caribbean outgroup generalist	*Cyprinodon variegatus*	Great Lake	Cat Island, Bahamas	1
Caribbean outgroup generalist	*Cyprinodon dearborni*	--	Curacao	2
North American outgroup generalist	*Cyprinodon albivelis*	Rio Yaqui basin	Mexico	1
North American outgroup generalist	*Cyprinodon eremus*	Quitobaquito Spring	AZ, USA	1
North American outgroup generalist	*Cyprinodon eximius*	Rio Conchos basin	Mexico	1
North American outgroup generalist	*Cyprinodon fontinalis*	Ojo de Carbonera Spring	Mexico	1
North American outgroup generalist	*Cyprinodon longidorsalis*	Charco Palma	Mexico	1
North American outgroup generalist	*Cyprinodon macularius*	Coachella	CA, USA	1
North American outgroup generalist	*Cyprinodon macrolepis*	Ojo de Hacienda Delores	Mexico	1
North American outgroup generalist	*Cyprinodon radiosus*	Owens Valley	CA, USA	1
Caribbean outgroup generalist	*Cyprinodon veronicae*	Ojo de Agua Charco Azul	Mexico	1
North American outgroup generalist	*Cyprinodon variegatus*	Salt pond near Dean's blue hole	Long Island, Bahamas	1
Caribbean outgroup generalist	*Cyprinodon variegatus*	Unnamed lake ‘near Rokers Point’	Exumas, Bahamas	2
Caribbean outgroup generalist	*Cyprinodon variegatus*	Unnamed lake ‘Ephemeral’	Exumas, Bahamas	1
Caribbean outgroup generalist	*Cyprinodon bondi*	Etang Saumautre	Dominican Republic	1
Caribbean outgroup generalist	*Cyprinodon variegatus*	Unnamed lake	Mayaguana	1
Caribbean outgroup generalist	*Cyprinodon variegatus*	Sarasota estuary	Florida, United States	1

Caribbean outgroup generalist	*Cyprinodon variegatus*	Lake Kilarney	New Providence Island, Bahamas	1
Caribbean outgroup generalist	*Cyprinodon variegatus*	Great Lake in the south	Long Island, Bahamas	1
Caribbean outgroup generalist	*Cyprinodon ovinus*	Falmouth River	Massachusetts, USA	1
Caribbean outgroup generalist	*Cyprinodon variegatus*	New Bight	Cat Island, Bahamas	
Caribbean outgroup generalist	*Cyprinodon variegatus*	Pirate's Well Lake	Mayaguana, Bahamas	1
Caribbean outgroup generalist	*Cyprinodon variegatus*	Salt Pond	Exumas, Bahamas	1
Caribbean outgroup generalist	*Cyprinodon variegatus*	Scully Lake	Mayaguana, Bahamas	1
Lake Chichancab pupfish radiation outgroup	*Cyprinodon maya*	Laguna Chichancanab	Quintana Roo, Mexico	1
Lake Chichancab pupfish radiation outgroup	*Cyprinodon simus*	Laguna Chichancanab	Quintana Roo, Mexico	1
Cualac outgroup	*Cualac tessellatus*	Media Luna	Mexico	1
Megupsilon outgroup	*Megupsilon aporus*	El Potosi	Mexico	1

1224
1225

Table S2.

Candidate adaptive regions for the San Salvador Island scale-eater. Location of the genic regions that contained signatures of a strong selective sweep in the scale-eater (SweeD CLR ≥ 5.28) and at least one divergent variant between the specialists ($F_{st} \geq 0.95$). Full list of variants, including unannotated candidate regions provided in Data S2. Regions highlighted in Figure 4 are listed in bold.

Gene	Scaffold	Gene Start	Gene End	Number of Variants
coq7	c_bro_v1_0_scaf1	28974409	28979038	3
gpr83	c_bro_v1_0_scaf1	38351481	38355816	2
klf1	c_bro_v1_0_scaf1	29239984	29242454	13
notum2	c_bro_v1_0_scaf1	28950946	28957848	1
rbm20	c_bro_v1_0_scaf1	15024176	15044016	1
rps15a	c_bro_v1_0_scaf1	28942599	28947456	2
ube2k	c_bro_v1_0_scaf1	41168936	41171561	2
atp8a2	c_bro_v1_0_scaf11	13000335	13035561	92
cd226	c_bro_v1_0_scaf11	10936603	10941232	7
cdk8	c_bro_v1_0_scaf11	13057400	13067971	1
cmbl	c_bro_v1_0_scaf11	9934853	9938096	11
crispld1	c_bro_v1_0_scaf11	11066268	11081938	7
dok6	c_bro_v1_0_scaf11	10963193	10972277	50
fbxl7	c_bro_v1_0_scaf11	21351783	21356510	6
hnf4g	c_bro_v1_0_scaf11	8350195	8354295	1
med1	c_bro_v1_0_scaf11	21393330	21400087	26
mtrr	c_bro_v1_0_scaf11	9943625	9954042	2
ncoa2	c_bro_v1_0_scaf11	11949666	11977882	4
prlh	**c_bro_v1_0_scaf11**	**9494231**	**9495565**	**18**
rnf6	c_bro_v1_0_scaf11	13047328	13052736	4
shisa2	c_bro_v1_0_scaf11	12945178	12953040	38
slc51a	c_bro_v1_0_scaf11	9862250	9873650	29
spice1	c_bro_v1_0_scaf11	12934206	12942196	2
zfhx4	c_bro_v1_0_scaf11	8078834	8095610	1
zbed1	c_bro_v1_0_scaf14	23383635	23383982	9
abhd8	c_bro_v1_0_scaf16	13452740	13457468	24
b3gnt3	c_bro_v1_0_scaf16	10003286	10004410	15
bmb	c_bro_v1_0_scaf16	10649637	10654441	38
brinp3	c_bro_v1_0_scaf16	11738302	11756508	33
crocc	c_bro_v1_0_scaf16	32985892	33009791	1
dda1	c_bro_v1_0_scaf16	13466708	13470377	2
eef1d	c_bro_v1_0_scaf16	10028318	10042958	30
ptprs	c_bro_v1_0_scaf16	8205473	8246024	56

pycr3	c_bro_v1_0_scaf16	10045452	10047013	8
rfc4	c_bro_v1_0_scaf16	35817866	35832867	38
serpinb1	c_bro_v1_0_scaf16	10634868	10638000	14
tdrd5	c_bro_v1_0_scaf16	12808042	12822317	47
tjp3	c_bro_v1_0_scaf16	35777675	35795399	21
tsta3	c_bro_v1_0_scaf16	10641946	10647463	23
zfp2	c_bro_v1_0_scaf16	35859060	35860865	8
zfp26	c_bro_v1_0_scaf16	35907423	35909825	2
znf271	c_bro_v1_0_scaf16	35840463	35842592	7
znf45	c_bro_v1_0_scaf16	35879283	35880581	7
anks1a	c_bro_v1_0_scaf18	18164811	18167681	1
gnat2	c_bro_v1_0_scaf18	13731762	13735798	2
itga5	c_bro_v1_0_scaf18	28908235	28944244	2
mybph	c_bro_v1_0_scaf18	26461834	26474649	15
nfasc	c_bro_v1_0_scaf18	17031686	17047770	1
sarg	c_bro_v1_0_scaf18	18185730	18187828	2
slc16a1	c_bro_v1_0_scaf18	29586755	29599009	1
nap1l4	c_bro_v1_0_scaf19	7836170	7842620	1
smap	c_bro_v1_0_scaf19	2027249	2028419	2
th	**c_bro_v1_0_scaf19**	**7787018**	**7794685**	**1**
trim44	c_bro_v1_0_scaf19	6431393	6435783	13
aasdhppt	c_bro_v1_0_scaf21	26911467	26919394	1
b3gat1	c_bro_v1_0_scaf21	29988110	29992848	1
cntn5	c_bro_v1_0_scaf21	10012673	10063457	1
col26a1	c_bro_v1_0_scaf21	20284619	20287102	12
emid1	c_bro_v1_0_scaf21	20254093	20266161	2
ifi44	c_bro_v1_0_scaf21	32843968	32848322	9
irf8	c_bro_v1_0_scaf21	41216201	41218789	1
mrm3	c_bro_v1_0_scaf21	15198156	15201994	1
nipsnap2	c_bro_v1_0_scaf21	24482700	24491999	33
nxn	c_bro_v1_0_scaf21	15204991	15221395	8
pde4d	c_bro_v1_0_scaf21	32298408	32320844	1
slc35e1	c_bro_v1_0_scaf21	31978195	31986378	1
tiparp	c_bro_v1_0_scaf21	33709833	33728566	1
trarg1	c_bro_v1_0_scaf21	25190856	25191383	1
atad2	c_bro_v1_0_scaf22	7942666	7961336	3
cyp26b1	c_bro_v1_0_scaf24	20457960	20473004	8
dysf	c_bro_v1_0_scaf24	20196578	20211497	1
ext1	c_bro_v1_0_scaf26	271389	272345	8
ext1b	c_bro_v1_0_scaf26	241224	252635	1

ppp1r3a	c_bro_v1_0_scaf26	8473965	8479904	4
soga3	c_bro_v1_0_scaf26	428526	434421	23
washc5	c_bro_v1_0_scaf26	301047	314009	1
zdhhc14	c_bro_v1_0_scaf2748	17727	21969	1
bri3bp	c_bro_v1_0_scaf33	12638129	12642531	28
gnaq	c_bro_v1_0_scaf33	12884125	12889121	9
pip5k1b	c_bro_v1_0_scaf33	2845282	2870905	6
wdr31	c_bro_v1_0_scaf33	12650071	12652945	20
cadps	c_bro_v1_0_scaf34	25394387	25411387	3
eya2	c_bro_v1_0_scaf34	32387513	32410375	2
srgap3	c_bro_v1_0_scaf34	26044753	26082456	2
st7l	c_bro_v1_0_scaf34	31252675	31262720	1
tfap2a	**c_bro_v1_0_scaf34**	**32260190**	**32264933**	**6**
znf362	c_bro_v1_0_scaf34	27775403	27792854	1
arhgap29	c_bro_v1_0_scaf37	30354970	30373446	1
atp5if1a	c_bro_v1_0_scaf37	3215186	3217688	1
cfap20	**c_bro_v1_0_scaf37**	**5089635**	**5093234**	**24**
chrna7	c_bro_v1_0_scaf37	3585852	3605137	8
dgat1	c_bro_v1_0_scaf37	5067735	5086382	37
dlx6a	c_bro_v1_0_scaf37	12742190	12744024	1
gpr20	c_bro_v1_0_scaf37	5101678	5107779	6
kcnn3	c_bro_v1_0_scaf37	3554189	3565883	4
mylipa	c_bro_v1_0_scaf37	8279827	8292615	1
slc45a4	c_bro_v1_0_scaf37	5115894	5125512	11
tbc1d20	c_bro_v1_0_scaf37	5047715	5065680	25
trim46	c_bro_v1_0_scaf37	3671825	3693120	1
trps1	c_bro_v1_0_scaf37	5649512	5665892	2
rmi1	c_bro_v1_0_scaf39	4258986	4266819	1
smyd1	c_bro_v1_0_scaf39	1675166	1684412	7
ubox5	c_bro_v1_0_scaf39	1621625	1630916	7
c1d	c_bro_v1_0_scaf43	30356623	30357420	6
dst	c_bro_v1_0_scaf43	16259900	16336750	2
ppp3r1	c_bro_v1_0_scaf43	30309740	30313100	4
sertad2	c_bro_v1_0_scaf43	10397273	10398532	4
sptlc3	c_bro_v1_0_scaf43	12316311	12350292	3
tmem26	c_bro_v1_0_scaf43	26556107	26570766	5
znf451	c_bro_v1_0_scaf43	16192481	16198948	15
atp8a1	c_bro_v1_0_scaf44	14934291	14999736	19
cenpf,kcnk2	**c_bro_v1_0_scaf44**	**12548021**	**12569724**	**16**
gpm6a	c_bro_v1_0_scaf44	24566260	24570802	9

kcnk2	**c_bro_v1_0_scaf44**	**12526223**	**12538276**	**23**
tsc22d3	c_bro_v1_0_scaf44	11339700	11340952	3
tstd1	c_bro_v1_0_scaf44	12012710	12013203	1
card8	**c_bro_v1_0_scaf46**	**1328324**	**1329460**	**10**
ccdc178	c_bro_v1_0_scaf46	15536795	15561009	1
xrn1	c_bro_v1_0_scaf46	25988805	26007498	39
dnm1	c_bro_v1_0_scaf47	21986865	22007761	1
map1b	c_bro_v1_0_scaf47	16222149	16245672	9
pdlim5	c_bro_v1_0_scaf47	24141068	24152322	1
ptger4	c_bro_v1_0_scaf47	16158956	16164333	4
aldh1a2	c_bro_v1_0_scaf5	27683247	27700000	1
esrp2	c_bro_v1_0_scaf5	34229725	34252121	1
gse1	c_bro_v1_0_scaf5	28378694	28397287	1
tcf12	**c_bro_v1_0_scaf5**	**27885956**	**27895543**	**15**
bcor	c_bro_v1_0_scaf52	5564938	5578475	2
chpf	c_bro_v1_0_scaf52	21895691	21907353	1
nr4a2	c_bro_v1_0_scaf52	13846770	13849514	4
st6gal2	c_bro_v1_0_scaf52	6730438	6731400	2
vgll3	c_bro_v1_0_scaf52	23953279	23956671	1
cox6b1	c_bro_v1_0_scaf53	24790612	24793003	8
cyp21a2	c_bro_v1_0_scaf53	18529622	18536111	2
eva1b	c_bro_v1_0_scaf53	29794772	29795353	2
fhod3	c_bro_v1_0_scaf53	18622119	18644926	2
galnt1	c_bro_v1_0_scaf53	20852048	20872629	17
glipr2	c_bro_v1_0_scaf53	20433230	20435503	3
hdac9b	c_bro_v1_0_scaf53	19008287	19034268	1
mag	c_bro_v1_0_scaf53	17408478	17413240	2
map7d1	c_bro_v1_0_scaf53	29904810	29922183	25
mindy3	c_bro_v1_0_scaf53	20097197	20106215	8
nacad	c_bro_v1_0_scaf53	20437309	20451974	2
pxn1	c_bro_v1_0_scaf53	20366555	20367417	1
rasip1	c_bro_v1_0_scaf53	24769523	24786366	13
slc2a3	c_bro_v1_0_scaf53	24809669	24817209	15
steap4	c_bro_v1_0_scaf53	20313856	20325260	26
tbrg4	c_bro_v1_0_scaf53	20454806	20462512	2
them4	c_bro_v1_0_scaf53	21823050	21830844	5
tnc	c_bro_v1_0_scaf53	18536783	18542213	1
twist1	**c_bro_v1_0_scaf53**	**18968733**	**18969242**	**1**
zhx2	c_bro_v1_0_scaf53	11078442	11084544	6
znf628	c_bro_v1_0_scaf53	24721275	24732863	6

trim25	**c_bro_v1_0_scaf60**	**1610217**	**1614325**	**2**
znf214	c_bro_v1_0_scaf60	1787099	1793538	1
foxo3	c_bro_v1_0_scaf7	12823341	12824321	3
myct1	c_bro_v1_0_scaf7	13100090	13100656	1
otof	c_bro_v1_0_scaf7	12616933	12629352	3
otof	c_bro_v1_0_scaf7	12642391	12658039	5
smek1	c_bro_v1_0_scaf7	12319537	12332574	1
43530	c_bro_v1_0_scaf752	1258	12292	29
nat1	c_bro_v1_0_scaf752	13172	14020	7
zdhhc20	c_bro_v1_0_scaf752	16935	24566	7
galr2	**c_bro_v1_0_scaf8**	**19974117**	**19979248**	**2**
grid2ip	c_bro_v1_0_scaf8	21581872	21603752	4
map2k6	c_bro_v1_0_scaf8	19746299	19760895	3
dcun1d2	c_bro_v1_0_scaf9	28311034	28313774	7
fhl2	c_bro_v1_0_scaf9	25288775	25292382	1
fut9	c_bro_v1_0_scaf9	25262573	25263652	16

Table S3. Candidate adaptive regions for the San Salvador Island molluscivore. Location of the genic regions that contained signatures of a strong selective sweep in the molluscivore (SweeD CLR ≥ 4.47) and at least one divergent variant between the specialists ($F_{st} \geq 0.95$). Full list of variants, including one unannotated candidate regions provided in Data S3. Regions highlighted in Figure S5 are listed in bold.

Gene	Scaffold	Gene Start	Gene Stop	Number of Variants
alox15b	c_bro_v1_0_scaf1	34682742	34695090	1
coq7	c_bro_v1_0_scaf1	28974409	28979038	3
gga1	c_bro_v1_0_scaf1	29195804	29209213	5
gpr83	c_bro_v1_0_scaf1	38351481	38355816	2
klf1	c_bro_v1_0_scaf1	29239984	29242454	13
notum2	c_bro_v1_0_scaf1	28950946	28957848	1
rbm20	c_bro_v1_0_scaf1	15024176	15044016	1
rps15a	c_bro_v1_0_scaf1	28942599	28947456	2
atp8a2	c_bro_v1_0_scaf11	13000335	13035561	92
cd226	c_bro_v1_0_scaf11	10936603	10941232	6
ncoa2	c_bro_v1_0_scaf11	11949666	11977882	7
shisa2	c_bro_v1_0_scaf11	12945178	12953040	18
spice1	c_bro_v1_0_scaf11	12934206	12942196	4
ube2w	c_bro_v1_0_scaf11	11253461	11259709	48
abhd8	c_bro_v1_0_scaf16	13452740	13457468	17
b3gnt3	c_bro_v1_0_scaf16	10003286	10004410	15
b3gnt3	c_bro_v1_0_scaf16	10019232	10020410	1
eef1d	c_bro_v1_0_scaf16	10028318	10042958	64
ptprs	c_bro_v1_0_scaf16	8205473	8246024	20
pycr3	c_bro_v1_0_scaf16	10045452	10047013	8
rfc4	c_bro_v1_0_scaf16	35817866	35832867	31
anks1a	c_bro_v1_0_scaf18	18164811	18167681	1
mybph	c_bro_v1_0_scaf18	26461834	26474649	7
nfasc	c_bro_v1_0_scaf18	17031686	17047770	1
sarg	c_bro_v1_0_scaf18	18185730	18187828	2
trim44	c_bro_v1_0_scaf19	6431393	6435783	14
b3gat1	c_bro_v1_0_scaf21	29988110	29992848	1
cntn5	c_bro_v1_0_scaf21	10012673	10063457	1
tiparp	**c_bro_v1_0_scaf21**	**33709833**	**33728566**	**1**
trarg1	c_bro_v1_0_scaf21	25190856	25191383	1
atad2	c_bro_v1_0_scaf22	7942666	7961336	3
cyp26b1	**c_bro_v1_0_scaf24**	**20457960**	**20473004**	**8**

ext1	c_bro_v1_0_scaf26	271389	272345	8
ext1b	c_bro_v1_0_scaf26	241224	252635	1
sox9	c_bro_v1_0_scaf27	22135691	22136918	2
bri3bp	c_bro_v1_0_scaf33	12638129	12642531	26
gnaq	c_bro_v1_0_scaf33	12884125	12889121	9
wdr31	c_bro_v1_0_scaf33	12650071	12652945	20
cadps	c_bro_v1_0_scaf34	25394387	25411387	2
znf362	c_bro_v1_0_scaf34	27775403	27792854	1
dlx6a	c_bro_v1_0_scaf37	12742190	12744024	1
mylipa	c_bro_v1_0_scaf37	8279827	8292615	1
trps1	c_bro_v1_0_scaf37	5649512	5665892	2
vps9d1	c_bro_v1_0_scaf4	15227575	15257418	1
slc29a3	c_bro_v1_0_scaf43	13679707	13685975	2
ttc33	c_bro_v1_0_scaf47	16128909	16148227	7
esrp2	c_bro_v1_0_scaf5	34229725	34252121	1
fn1	c_bro_v1_0_scaf52	19355212	19387175	1
st6gal2	c_bro_v1_0_scaf52	6730438	6731400	2
vgll3	c_bro_v1_0_scaf52	23953279	23956671	1
cox6b1	c_bro_v1_0_scaf53	24790612	24793003	8
map7d1	c_bro_v1_0_scaf53	29904810	29922183	25
rasip1	c_bro_v1_0_scaf53	24769523	24786366	13
slc2a3	c_bro_v1_0_scaf53	24809669	24817209	15
zhx2	c_bro_v1_0_scaf53	11078442	11084544	5
znf628	c_bro_v1_0_scaf53	24721275	24732863	5
foxo3	c_bro_v1_0_scaf7	12823341	12824321	3
otof	c_bro_v1_0_scaf7	12616933	12629352	11
otof	c_bro_v1_0_scaf7	12642391	12658039	5
smek1	HiC_scaffold_7	12319537	12332574	1

Table S4.
Full list of functional terms associated with genes in candidate regions for the scale-eaters that were significantly enriched (FDR <0.05) in a GO analysis. Focal functional terms related

to key axes of diversification in this system: habitat preference (scale-eating/snail-eating niches), trophic morphology, and/or pigmentation.

Functional Category	Enrichment FDR	Genes in list	Total genes	Genes
Neuron differentiation	0.00608452	25	1400	*map1b,atp8a2,tcf12,gpm6a,nr4a2,brinp3,tnc,ptprs,mag,foxo3,med1,rnf6,aldh1a2,gnat2,pdlim5,trim46,nfasc,washc5,zhx2,th,ext1,galr2,anks1a,chrna7,dok6*
Camera-type eye morphogenesis	0.00608452	7	114	*tbc1d20,atp8a2,gnat2,zhx2,th,tfap2a,twist1*
Generation of neurons	0.00608452	26	1553	*map1b,atp8a2,tcf12,gpm6a,nr4a2,brinp3,tnc,ptprs,mag,foxo3,twist1,med1,rnf6,aldh1a2,gnat2,pdlim5,trim46,nfasc,washc5,zhx2,th,ext1,galr2,anks1a,chrna7,dok6*
Muscle tissue development	0.00608452	12	400	*cyp26b1,eya2,kcnk2,smyd1,fhl2,cenpf,twist1,med1,aldh1a2,fhod3,pdlim5,tiparp*
Regulation of biological quality	0.00608452	50	4146	*kcnk2,klf1,dnm1,foxo3,atp8a1,abhd8,atp8a2,gnaq,ptger4,chrna7,gpr20,pde4d,xrn1,cyp26b1,cfap20,ube2k,rasip1,trim44,crocc,eya2,prlh,ptprs,mag,map2k6,otof,med1,rnf6,steap4,aldh1a2,map1b,gnat2,fhod3,dysf,slc16a1,tsc22d3,pdlim5,cadps,tiparp,nxn,rmi1,th,galr2,dgat1,grid2ip,tbc1d20,tbrg4,them4,trim46,rfc4,cyp21a2*
Cell development	0.00708671	32	2196	*map1b,atp8a2,tcf12,gpm6a,brinp3,tnc,ptprs,mag,fhl2,foxo3,twist1,med1,tbc1d20,rnf6,aldh1a2,gnat2,fhod3,dysf,nr4a2,tdrd5,pdlim5,trim46,nfasc,washc5,zhx2,th,ext1,galr2,anks1a,pde4d,chrna7,dok6*
Neural retina development	0.00819009	5	64	*atp8a2,gnat2,gpm6a,zhx2,tfap2a*
Feeding behavior	0.00819009	6	102	*cfap20,prlh,atp8a2,rmi1,th,galr2*
Striated muscle tissue development	0.00819009	11	385	*cyp26b1,eya2,kcnk2,smyd1,fhl2,cenpf,twist1,med1,aldh1a2,fhod3,pdlim5*
Neurogenesis	0.00819009	26	1663	*map1b,atp8a2,tcf12,gpm6a,nr4a2,brinp3,tnc,ptprs,mag,foxo3,twist1,med1,rnf6,aldh1a2,gnat2,pdlim5,trim46,nfasc,washc5,zhx2,th,ext1,galr2,anks1a,chrna7,dok6*
Response to lipid	0.00819009	19	997	*rnf6,brinp3,ptger4,card8,med1,cyp26b1,tnc,pde4d,xrn1,foxo3,trim25,gpr83,aldh1a2,ncoa2,irf8,nr4a2,hnf4g,th,fhl2*
Eating behavior	0.00819009	4	33	*prlh,atp8a2,rmi1,th*
Camera-type eye development	0.00819009	10	317	*med1,tbc1d20,aldh1a2,atp8a2,gnat2,gpm6a,zhx2,th,tfap2a,twist1*
Developmental growth	0.00819009	15	651	*tnc,prlh,kcnk2,ptprs,mag,pde4d,foxo3,med1,rnf6,map1b,atp8a2,dysf,pdlim5,rmi1,trim46*
Eye morphogenesis	0.0084973	7	152	*tbc1d20,atp8a2,gnat2,zhx2,th,tfap2a,twist1*
Embryonic camera-type eye development	0.01187832	4	39	*aldh1a2,th,twist1,tfap2a*
Regulation of phospholipid translocation	0.01187832	2	3	*atp8a1,atp8a2*
Positive regulation of phospholipid translocation	0.01187832	2	3	*atp8a1,atp8a2*
Negative regulation of axon extension	0.01370935	4	41	*ptprs,mag,rnf6,trim46*

Eye development	0.01408643	10	365	*med1,tbc1d20,aldh1a2,atp8a2,gnat2,gpm6a,zhx2,th,tfap2a,twist1*
Growth	0.01408643	18	1018	*sertad2,st7l,tnc,prlh,kcnk2,ptprs,mag,pde4d,foxo3,med1,rnf6,map1b,atp8a2,dysf,irf8,pdlim5,rmi1,trim46*
Cellular response to lipid	0.01408643	14	671	*rnf6,brinp3,ptger4,card8,med1,cyp26b1,tnc,pde4d,foxo3,aldh1a2,irf8,nr4a2,hnf4g,fhl2*
Visual system development	0.01408643	10	366	*med1,tbc1d20,aldh1a2,atp8a2,gnat2,gpm6a,zhx2,th,tfap2a,twist1*
Anatomical structure morphogenesis	0.01669332	34	2702	*map1b,tfap2a,cyp26b1,tnc,esrp2,ptprs,rasip1,mag,fhl2,foxo3,twist1,med1,tbc1d20,rnf6,aldh1a2,atp8a2,gnat2,fhod3,dysf,gpm6a,nr4a2,itga5,pdlim5,trim46,nfasc,tiparp,zhx2,th,ext1,crispld1,chrna7,bcor,eya2,dok6*
Neuron development	0.01669332	19	1140	*map1b,atp8a2,gpm6a,tnc,ptprs,mag,rnf6,gnat2,nr4a2,pdlim5,trim46,nfasc,washc5,th,ext1,galr2,anks1a,chrna7,dok6*
Sensory system development	0.01669332	10	377	*med1,tbc1d20,aldh1a2,atp8a2,gnat2,gpm6a,zhx2,th,tfap2a,twist1*
Cell differentiation	0.01725075	48	4372	*tnc,klf1,map1b,atp8a2,tcf12,gpm6a,nr4a2,brinp3,smyd1,foxo3,glipr2,med1,tfap2a,cyp26b1,prlh,ptprs,rasip1,mag,fhl2,cenpf,twist1,tbc1d20,rnf6,steap4,aldh1a2,gnat2,fhod3,dysf,irf8,tdrd5,pdlim5,trim46,nfasc,tiparp,washc5,nxn,ptger4,zhx2,th,ext1,galr2,itga5,anks1a,trps1,pde4d,chrna7,eya2,dok6*
Behavior	0.01791936	13	619	*cfap20,prlh,kcnk2,atp8a1,atp8a2,ncoa2,nr4a2,slc16a1,itga5,rmi1,th,chrna7,galr2*
Intracellular receptor signaling pathway	0.02008492	9	323	*rnf6,med1,cyp26b1,twist1,aldh1a2,nr4a2,hnf4g,fhl2,map2k6*
Reduction of food intake in response to dietary excess	0.02011061	2	5	*prlh,rmi1*
Nervous system development	0.02011061	31	2439	*cox6b1,map1b,atp8a2,tcf12,gpm6a,nr4a2,brinp3,tnc,prlh,ptprs,mag,foxo3,twist1,med1,rnf6,aldh1a2,gnat2,pdlim5,trim46,nfasc,washc5,fut9,zhx2,th,ext1,galr2,cenpf,anks1a,tfap2a,chrna7,dok6*
Sensory organ development	0.02011061	12	561	*cyp26b1,kcnk2,med1,tbc1d20,aldh1a2,atp8a2,gnat2,gpm6a,zhx2,th,tfap2a,twist1*
Regulation of axon extension	0.02011061	5	94	*ptprs,mag,rnf6,map1b,trim46*
Response to vitamin	0.02011061	5	92	*cyp26b1,tnc,trim25,med1,aldh1a2*
Cellular hormone metabolic process	0.02011061	6	142	*cyp26b1,aldh1a2,tiparp,dgat1,med1,cyp21a2*
Negative regulation of growth	0.02011061	8	267	*sertad2,st7l,kcnk2,ptprs,mag,rnf6,irf8,trim46*
Retina morphogenesis in camera-type eye	0.02011061	4	53	*atp8a2,gnat2,zhx2,tfap2a*
Sensory organ morphogenesis	0.02011061	8	267	*cyp26b1,tbc1d20,atp8a2,gnat2,zhx2,th,tfap2a,twist1*
Negative regulation of	0.02019114	6	147	*atad2,xrn1,twist1,znf451,bcor,cenpf*

chromosome organization				
Regulation of neuron differentiation	0.02113374	13	656	*tcf12,brinp3,ptprs,mag,foxo3,med1,rnf6,map1b,atp8a2,pdlim5,washc5,zhx2,trim46*
Retina development in camera-type eye	0.02113374	6	149	*med1,atp8a2,gnat2,gpm6a,zhx2,tfap2a*
Neuron projection morphogenesis	0.02244972	13	662	*map1b,ptprs,mag,rnf6,atp8a2,gpm6a,nr4a2,pdlim5,trim46,nfasc,ext1,chrna7,dok6*
Response to hormone	0.02411051	17	1031	*foxo3,med1,rnf6,ncoa2,nr4a2,ptger4,chrna7,tnc,prlh,xrn1,gpr83,aldh1a2,hnf4g,th,trarg1,fhl2,gnaq*
Negative regulation of developmental growth	0.02447961	5	105	*kcnk2,ptprs,mag,rnf6,trim46*
Cell morphogenesis involved in neuron differentiation	0.02447961	12	591	*map1b,ptprs,mag,rnf6,atp8a2,nr4a2,pdlim5,trim46,nfasc,ext1,chrna7,dok6*
Cell projection morphogenesis	0.02447961	13	678	*map1b,ptprs,mag,rnf6,atp8a2,gpm6a,nr4a2,pdlim5,trim46,nfasc,ext1,chrna7,dok6*
Axon development	0.02447961	11	510	*map1b,tnc,ptprs,mag,rnf6,atp8a2,nr4a2,trim46,nfasc,ext1,dok6*
Plasma membrane bounded cell projection morphogenesis	0.02447961	13	676	*map1b,ptprs,mag,rnf6,atp8a2,gpm6a,nr4a2,pdlim5,trim46,nfasc,ext1,chrna7,dok6*
Response to organic cyclic compound	0.02545138	16	962	*rnf6,med1,tiparp,tnc,pde4d,xrn1,foxo3,trim25,gpr83,aldh1a2,ncoa2,nr4a2,slc16a1,hnf4g,th,fhl2*
Negative regulation of neuron differentiation	0.02545138	7	224	*ptprs,mag,foxo3,med1,rnf6,zhx2,trim46*
Embryonic camera-type eye morphogenesis	0.02545138	3	28	*th,twist1,tfap2a*
Regulation of extent of cell growth	0.02545138	5	109	*ptprs,mag,rnf6,map1b,trim46*
Protein K48-linked ubiquitination	0.02545138	4	62	*ube2k,march6,trim44,rnf6*
Negative regulation of chromatin organization	0.02545138	4	63	*atad2,twist1,znf451,bcor*
Cellular developmental process	0.02579863	48	4587	*tnc,klf1,map1b,atp8a2,tcf12,gpm6a,nr4a2,brinp3,smyd1,foxo3,glipr2,med1,tfap2a,cyp26b1,prlh,ptprs,rasip1,mag,fhl2,cenpf,twist1,tbc1d20,rnf6,steap4,aldh1a2,gnat2,fhod3,dysf,irf8,tdrd5,pdlim5,trim46,nfasc,tiparp,washc5,nxn,ptger4,zhx2,th,ext1,galr2,itga5,anks1a,trps1,pde4d,chrna7,eya2,dok6*

Circulatory system development	0.02618802	17	1064	*th,kcnk2,rasip1,smyd1,fhl2,twist1,med1,aldh1a2,fhod3,dysf,itga5,pdlim5,tiparp,nxn,rbm20,chrna7,bcor*
Cell part morphogenesis	0.0264455	13	697	*map1b,ptprs,mag,rnf6,atp8a2,gpm6a,nr4a2,pdlim5,trim46,nfasc,ext1,chrna7,dok6*
Anatomical structure maturation	0.0264455	6	167	*foxo3,aldh1a2,nr4a2,nfasc,washc5,anks1a*
Response to extracellular stimulus	0.02777357	11	532	*nr4a2,cyp26b1,tnc,prlh,foxo3,trim25,med1,aldh1a2,slc16a1,rmi1,th*
Response to oxygen-containing compound	0.0281392	23	1689	*foxo3,nr4a2,brinp3,ptger4,th,card8,chrna7,cyp26b1,tnc,prlh,klf1,dnm1,map2k6,pde4d,xrn1,trim25,med1,aldh1a2,ncoa2,irf8,rmi1,trarg1,gnaq*
Cellular response to oxygen-containing compound	0.02824704	18	1178	*foxo3,nr4a2,brinp3,ptger4,card8,cyp26b1,tnc,klf1,map2k6,pde4d,xrn1,med1,aldh1a2,irf8,th,trarg1,gnaq,chrna7*
Response to axon injury	0.02862969	4	69	*tnc,kcnk2,ptprs,mag*
Negative regulation of axonogenesis	0.02862969	4	69	*ptprs,mag,rnf6,trim46*
Developmental growth involved in morphogenesis	0.02862969	7	237	*tnc,ptprs,mag,med1,rnf6,map1b,trim46*
Negative regulation of protein polyubiquitination	0.02862969	2	8	*trim44,dysf*
Regulation of phospholipid transport	0.02862969	2	8	*atp8a1,atp8a2*
Positive regulation of phospholipid transport	0.02862969	2	8	*atp8a1,atp8a2*
Neuron projection development	0.02913429	16	997	*map1b,gpm6a,tnc,ptprs,mag,rnf6,atp8a2,nr4a2,pdlim5,trim46,nfasc,washc5,ext1,galr2,chrna7,dok6*
Cell maturation	0.03039891	6	178	*foxo3,nr4a2,tdrd5,nfasc,washc5,anks1a*
Axon extension	0.03039891	5	120	*ptprs,mag,rnf6,map1b,trim46*
Heart development	0.03089855	11	552	*th,kcnk2,smyd1,fhl2,twist1,med1,aldh1a2,fhod3,pdlim5,rbm20,bcor*
Cellular response to chemical stimulus	0.03089855	38	3443	*foxo3,med1,rnf6,ncoa2,nr4a2,brinp3,ptger4,shisa2,cyp26b1,card8,tfap2a,irf8,tiparp,trim44,tnc,kcnk2,klf1,map2k6,pde4d,xrn1,trim25,twist1,aldh1a2,dysf,slc16a1,hnf4g,nxn,th,trarg1,ube2k,znf451,gnaq,chrna7,fhl2,esrp2,itga5,cmbl,nat1*
Axonogenesis	0.0310125	10	471	*map1b,ptprs,mag,rnf6,atp8a2,nr4a2,trim46,nfasc,ext1,dok6*
Embryonic forelimb morphogenesis	0.03241747	3	34	*twist1,aldh1a2,tfap2a*

Cellular response to retinoic acid	0.03315736	4	74	*brinp3,cyp26b1,tnc,aldh1a2*
Homeostatic process	0.03339051	25	1962	*klf1,foxo3,abhd8,ptger4,gpr20,xrn1,ube2k,cyp26b1,crocc,prlh,map2k6,pde4d,med1,steap4,gnat2,slc16a1,tsc22d3,nxn,rmi1,th,galr2,dgat1,tbc1d20,chrna7,rfc4*
Negative regulation of cellular component organization	0.03339051	13	739	*atad2,xrn1,ptger4,ptprs,mag,twist1,rnf6,fhod3,dysf,znf451,bcor,trim46,cenpf*
Response to vitamin D	0.03347122	3	35	*tnc,trim25,med1*
Animal organ morphogenesis	0.03351451	16	1027	*tfap2a,cyp26b1,tnc,esrp2,fhl2,foxo3,twist1,med1,tbc1d20,aldh1a2,atp8a2,gnat2,tiparp,zhx2,th,bcor*
System development	0.03351451	50	4976	*cox6b1,klf1,map1b,atp8a2,tcf12,gpm6a,nr4a2,brinp3,th,foxo3,glipr2,tfap2a,cyp26b1,tnc,prlh,kcnk2,esrp2,ptprs,rasip1,mag,smyd1,fhl2,cenpf,twist1,med1,tbc1d20,rnf6,aldh1a2,gnat2,fhod3,dysf,irf8,itga5,pdlim5,trim46,nfasc,tiparp,washc5,nxn,ptger4,fut9,zhx2,ext1,galr2,rbm20,chrna7,bcor,anks1a,trps1,dok6*
Embryonic eye morphogenesis	0.03498708	3	36	*th,twist1,tfap2a*
Cell morphogenesis involved in differentiation	0.03535026	13	751	*map1b,ptprs,mag,tbc1d20,rnf6,atp8a2,nr4a2,pdlim5,trim46,nfasc,ext1,chrna7,dok6*
Negative regulation of cell growth	0.03535026	6	191	*sertad2,st7l,ptprs,mag,rnf6,trim46*
Embryonic limb morphogenesis	0.03535026	5	130	*cyp26b1,twist1,med1,aldh1a2,tfap2a*
Embryonic appendage morphogenesis	0.03535026	5	130	*cyp26b1,twist1,med1,aldh1a2,tfap2a*
Protein localization to axon	0.03535026	2	10	*trim46,nfasc*
Regulation of developmental growth	0.03607979	8	333	*prlh,kcnk2,ptprs,mag,rnf6,map1b,atp8a2,trim46*
Cellular response to organic cyclic compound	0.03631117	11	579	*rnf6,med1,tiparp,tnc,pde4d,xrn1,foxo3,nr4a2,slc16a1,hnf4g,fhl2*
Response to nutrient levels	0.03921368	10	500	*cyp26b1,tnc,prlh,foxo3,trim25,med1,aldh1a2,slc16a1,rmi1,th*
Response to steroid hormone	0.03921368	9	418	*rnf6,med1,foxo3,gpr83,ncoa2,nr4a2,hnf4g,th,fhl2*
Regulation of tooth mineralization	0.04037623	2	11	*tfap2a,bcor*
Oxidation-reduction process	0.04038443	16	1061	*cyp26b1,steap4,coq7,prlh,tsta3,pycr3,mtrr,cox6b1,aldh1a2,ppp1r3a,nxn,th,cyp21a2,twist1,tbrg4,tstd1*
Response to endogenous stimulus	0.04173373	22	1692	*foxo3,med1,rnf6,ncoa2,nr4a2,ptger4,shisa2,chrna7,tnc,prlh,klf1,pde4d,xrn1,gpr83,aldh1a2,hnf4g,th,trarg1,znf451,fhl2,gnaq,esrp2*

Vitamin metabolic process	0.04365266	5	140	*cyp26b1,mtrr,aldh1a2,slc2a3,aasdhppt*
Positive regulation of transcription, DNA-templated	0.04365266	21	1593	*zfhx4,klf1,foxo3,tcf12,ncoa2,atad2,coq7,twist1,med1,tfap2a,irf8,hnf4g,trim44,galr2,zbed1,ppp3r1,rnf6,nr4a2,sertad2,fhl2,cdk8*
Response to ketone	0.04397175	6	204	*ptger4,tnc,xrn1,foxo3,ncoa2,th*
Negative regulation of transcription by RNA polymerase II	0.04451995	14	896	*zfhx4,foxo3,coq7,zhx2,trps1,fhl2,tfap2a,irf8,twist1,med1,ncoa2,bcor,znf451,nr4a2*
Protein polyubiquitination	0.04451995	7	277	*ubox5,ube2k,rnf6,march6,trim44,dysf,fbxl7*
Response to external stimulus	0.04451995	29	2525	*card8,rps15a,nr4a2,trim44,ptger4,cyp26b1,tnc,prlh,kcnk2,ptprs,mag,pde4d,foxo3,trim25,med1,aldh1a2,atp8a2,gnat2,dysf,ifi44,irf8,gpm6a,slc16a1,nfasc,rmi1,th,ext1,gnaq,dok6*
Response to organic substance	0.04451995	37	3461	*foxo3,med1,rnf6,ncoa2,march6,nr4a2,brinp3,ptger4,th,shisa2,card8,aldh1a2,irf8,tiparp,trim44,chrna7,cyp26b1,tnc,prlh,klf1,dnm1,map2k6,pde4d,xrn1,trim25,twist1,gpr83,slc16a1,hnf4g,rmi1,trarg1,ube2k,znf451,fhl2,gnaq,esrp2,itga5*
Negative regulation of macromolecule metabolic process	0.04451995	32	2872	*serpinb1,zfhx4,foxo3,atad2,coq7,zhx2,xrn1,trps1,card8,fhl2,cenpf,twist1,tfap2a,irf8,trim44,bcor,kcnk2,pde4d,smyd1,med1,dysf,ncoa2,nxn,c1d,chrna7,rasip1,znf451,tbrg4,nr4a2,gnaq,tiparp,rps15a*
Positive regulation of gene expression	0.04451995	25	2046	*zfhx4,klf1,foxo3,tcf12,ncoa2,atad2,coq7,twist1,med1,tfap2a,irf8,hnf4g,trim44,galr2,zbed1,ppp3r1,cyp26b1,tnc,rnf6,aldh1a2,nr4a2,sertad2,rbm20,fhl2,cdk8*
Negative regulation of gene expression	0.04451995	24	1952	*zfhx4,foxo3,atad2,coq7,zhx2,trps1,card8,fhl2,cenpf,twist1,tfap2a,irf8,bcor,xrn1,smyd1,med1,dysf,ncoa2,c1d,znf451,tbrg4,nr4a2,tiparp,rps15a*
Developmental maturation	0.04451995	7	282	*foxo3,aldh1a2,nr4a2,tdrd5,nfasc,washc5,anks1a*
Negative regulation of histone modification	0.04451995	3	44	*twist1,znf451,bcor*
Protein modification by small protein conjugation	0.04451995	14	891	*dcun1d2,ubox5,ube2k,znf451,rnf6,march6,trim44,bcor,trim25,med1,dysf,nxn,fbxl7,zbed1*
Regulation of integrin activation	0.04451995	2	13	*ptger4,rasip1*
Forelimb morphogenesis	0.04451995	3	42	*twist1,aldh1a2,tfap2a*
Dopamine biosynthetic process	0.04451995	2	12	*th,nr4a2*
Negative regulation of transcription, DNA-templated	0.04451995	18	1298	*zfhx4,foxo3,atad2,coq7,zhx2,trps1,fhl2,cenpf,twist1,tfap2a,irf8,bcor,smyd1,med1,ncoa2,c1d,znf451,nr4a2*

Embryonic camera-type eye formation	0.04451995	2	12	*twist1,tfap2a*
Eyelid development in camera-type eye	0.04451995	2	13	*twist1,tfap2a*
Cellular response to organic substance	0.04451995	32	2872	*foxo3,med1,rnf6,ncoa2,nr4a2,brinp3,ptger4,shisa2,card8,irf8,tiparp,trim44,cyp26b1,tnc,klf1,map2k6,pde4d,xrn1,trim25,twist1,aldh1a2,slc16a1,hnf4g,th,trarg1,ube2k,znf451,gnaq,chrna7,fhl2,esrp2,itga5*
Cellular response to alcohol	0.04451995	4	90	*ptger4,tnc,xrn1,foxo3*
Response to amyloid-beta	0.04451995	3	43	*dnm1,foxo3,chrna7*
Negative regulation of cellular macromolecule biosynthetic process	0.04451995	20	1509	*zfhx4,foxo3,atad2,coq7,zhx2,xrn1,trps1,fhl2,cenpf,twist1,tfap2a,irf8,bcor,kcnk2,smyd1,med1,ncoa2,c1d,znf451,nr4a2*
Cardiac muscle tissue development	0.04461977	6	213	*kcnk2,fhl2,med1,aldh1a2,fhod3,pdlim5*
Axon regeneration	0.04525881	3	45	*tnc,ptprs,mag*
Neuron maturation	0.04525881	3	45	*nr4a2,nfasc,anks1a*
Regulation of nucleobase-containing compound metabolic process	0.04595578	44	4374	*eya2,zfhx4,klf1,foxo3,tfap2a,tcf12,ncoa2,atad2,coq7,zhx2,xrn1,esrp2,trps1,fhl2,cenpf,twist1,med1,irf8,rfc4,hnf4g,trim44,galr2,bcor,zbed1,ppp3r1,kcnk2,smyd1,znf45,rnf6,znf214,nr4a2,tsc22d3,znf362,sertad2,c1d,znf628,zfp2,rbm20,vgll3,card8,znf451,trim25,tbrg4,cdk8*
Tissue development	0.04604795	25	2079	*glipr2,tfap2a,cyp26b1,tnc,eya2,kcnk2,esrp2,ptprs,rasip1,smyd1,fhl2,cenpf,twist1,med1,tbc1d20,aldh1a2,fhod3,dysf,pdlim5,tiparp,ext1,itga5,bcor,trps1,pde4d*
DNA-templated transcription, initiation	0.04769373	7	293	*twist1,med1,znf451,znf45,cdk8,nr4a2,hnf4g*
Transcription initiation from RNA polymerase II promoter	0.04769373	6	221	*med1,znf451,znf45,cdk8,nr4a2,hnf4g*
Response to xenobiotic stimulus	0.04769373	7	292	*cyp26b1,foxo3,nr4a2,tiparp,th,cmbl,nat1*
Appendage morphogenesis	0.04769373	5	154	*cyp26b1,twist1,med1,aldh1a2,tfap2a*
Limb morphogenesis	0.04769373	5	154	*cyp26b1,twist1,med1,aldh1a2,tfap2a*
Positive regulation of nucleobase-containing compound metabolic process	0.04769373	24	1978	*eya2,zfhx4,klf1,foxo3,tcf12,ncoa2,atad2,coq7,twist1,med1,tfap2a,irf8,rfc4,hnf4g,trim44,galr2,zbed1,ppp3r1,rnf6,nr4a2,sertad2,rbm20,fhl2,cdk8*

Negative regulation of neurogenesis	0.04769373	7	292	*ptprs,mag,foxo3,med1,rnf6,zhx2,trim46*
Negative regulation of nitrogen compound metabolic process	0.04769373	29	2551	*serpinb1,zfhx4,foxo3,atad2,coq7,zhx2,xrn1,trps1,card8,fhl2,cenpf,twist1,tfap2a,irf8,trim44,bcor,kcnk2,pde4d,smyd1,med1,dysf,ncoa2,nxn,c1d,chrna7,rasip1,znf451,nr4a2,gnaq*
Roof of mouth development	0.04769373	4	94	*twist1,tiparp,tfap2a,bcor*
Cellular response to endogenous stimulus	0.04769373	19	1431	*foxo3,med1,rnf6,ncoa2,nr4a2,ptger4,shisa2,tnc,klf1,pde4d,xrn1,hnf4g,th,trarg1,znf451,gnaq,chrna7,fhl2,esrp2*
Response to nutrient	0.04807838	6	222	*cyp26b1,tnc,trim25,med1,aldh1a2,slc16a1*
Regulation of gene expression	0.04807838	47	4798	*zfhx4,klf1,znf451,foxo3,tfap2a,tcf12,ncoa2,atad2,coq7,zhx2,esrp2,trps1,card8,fhl2,cenpf,twist1,med1,irf8,hnf4g,trim44,galr2,bcor,zbed1,ppp3r1,cyp26b1,tnc,xrn1,smyd1,znf45,rnf6,aldh1a2,dysf,znf214,nr4a2,tsc22d3,znf362,sertad2,c1d,znf628,zfp2,rbm20,vgll3,trim25,tbrg4,tiparp,cdk8,rps15a*
Developmental cell growth	0.04807838	6	223	*ptprs,mag,rnf6,map1b,pdlim5,trim46*
Oocyte development	0.04807838	3	48	*foxo3,tdrd5,washc5*
Regulation of neurogenesis	0.04807838	13	824	*tcf12,brinp3,ptprs,mag,foxo3,med1,rnf6,map1b,atp8a2,pdlim5,washc5,zhx2,trim46*

Table S5. Top 5 BLAST hits for LG15 QTL. Bolded values indicate the top hit that was used to determine the region the significant oral jaw size QTL aligned to an 18-Mb region on scaffold c_bro_v1_0_scaf8 (8840660-27314762**)** in the *C. brontotheroides* reference genome that contained 3 genes (*map2k6*, *galr2*, and *grid2ip*).

LG15 marker	Scaffold	% identity	Length (bp)	Mismatch	Start	End	E-value	Bitscore
10999	**c_bro_v1_0_scaf8**	**97.917**	**96**	**2**	**8840660**	**8840755**	**1.95E-42**	**174**
10999	c_bro_v1_0_scaf8	100	17	0	17544438	17544422	4.6	34.2
10999	c_bro_v1_0_scaf36	100	20	0	201747	201728	0.074	40.1
10999	c_bro_v1_0_scaf7	100	19	0	13795738	13795756	0.29	38.2
10999	c_bro_v1_0_scaf52	100	18	0	23185857	23185840	1.2	36.2
10999	c_bro_v1_0_scaf38	100	18	0	1880370	1880387	1.2	36.2
33382	**c_bro_v1_0_scaf8**	**100**	**93**	**0**	**27314670**	**27314762**	**1.97E-45**	**184**
33382	c_bro_v1_0_scaf8	93.617	47	3	26627380	26627426	8.05E-11	69.9
33382	c_bro_v1_0_scaf8	93.617	47	3	27916662	27916616	8.05E-11	69.9
33382	c_bro_v1_0_scaf8	95.238	42	2	1464518	1464477	3.18E-10	67.9
33382	c_bro_v1_0_scaf8	95.238	42	2	11224060	11224019	3.18E-10	67.9

Table S6.
Per generation mutation rate estimation from high coverage sequencing of parents and F1 from two crosses of San Salvador Island species.

Cross	*C. variegatus x C. brontotheroides*		*C. variegatus x C. desquamator*
Offspring	F1.A	F1.B	F1.A
Avg. coverage	67.5X	45.1X	32.7X
Known heterozygous sites genotype quality (GQ)	X>99	X>99	X>99
Known heterozygous sites base quality rank sum (BaseQRankSum)	1.4<x<2.6	1.4<x<2.6	1.4<x<2.7
Known heterozygous sites mapping quality (MQ)	x>54	x>54	x>54
Known heterozygous sites mapping quality rank sum (MQRankSum)	1.6<x<1.9	1.6<x<1.9	1.4<x<2
Known heterozygous sites quality by depth (QD)	24<x<36	24<x<36	24<x<36
Known heterozygous sites depth (DP)	27 <x<77	15 < x< 54	12< x < 39
Known heterozygous sites allele depth (AD)	10<x<42	5<x<30	4< x < 21
Known heterozygous sites read position rank sum (ReadPosRankSum)	-1.8<x<2.3	1.8<x<2.3	1.4<x<2.34
Known heterozygous sites StrandOddsRatio (SOR)	0.17<x<1.4	0.14<x<1.4	0.19<x<1.3
Known heterozygous sites FisherStrand (FS)	4.6<x<7.5	4.6<x<7.3	45<x<7.5
GATK new mutation sites (bp)	9114	8936	331
mpileup new mutation sites (bp)	14772	14182	7206
Shared variants (bp)	20	37	9
Accessible genome (bp)	698887016	712364816	695995433
Mutation rate estimate	1.43×10^{-8}	2.59×10^{-8}	6.46×10^{-9}

Table S7.

Parameters for selective sweep analyses. The average coverage, composite likelihood ratio threshold based on neutral simulations, and the population size change parameters and individual used for each species.

Species	Average Coverage	CLR threshold	SweeD Commands
SSI generalist	28.87X	4.89	-folded -strictPolymorphic -G 0.4068 -eN 5.45 181.8 -s 64
SSI molluscivore	17.37X	4.47	-folded -strictPolymorphic -G 0.389 -eN 5.88 196 -s 88
SSI scale-eater	18.21X	5.28	-folded -strictPolymorphic -G 0.218 -eN 8.11 270 -s 52
RC	21.04X	4.41	-folded -strictPolymorphic -G 0.23 -eN 11.15 269.1 -s 34
NP	22.67X	2.28	-folded -strictPolymorphic -G 0.198 -eN 13.35 445.07 -s 30
DR	NA	5.37	-folded -strictPolymorphic -G 0.236 -eN 10.83 362.8 -s 20
NCC	27.62X	5.09	-folded -strictPolymorphic -G 0.29 -eN 8.01 374.4 -s 24
VEN	17.21X	18.05	-folded -strictPolymorphic -G 8.87 -eN 0.086 0.345 -eN 1.077 38.78 -s 22

Table S8. The number of introgression regions in the SSI specialists. We determined introgressed regions of the genome as region with a *fd* statistic (ranges from 0 to 1) value above the threshold found in neutral simulations with no gene flow. These introgressed regions from each donor population were then overlapped with regions of the genome with strong genetic divergence (variants with Fst >= 0.95) and signatures of a hard selective sweep (CLR > 5.28 and > 4.47 for scale-eaters and molluscivores respectively) to determine the number of adaptive introgression regions. These adaptive introgression regions range in size from 50-kb to 110-kb in length.

Donor population (P3)	f_d threshold	Number of candidate introgression regions	Number of candidate adaptive introgression regions
Introgression with Molluscivore			
Rum Cay	0.81	536	5
New Providence	0.72	660	7
Dominican Republic	0.81	375	8
North Carolina	0.69	138	0
Venezuela	0.69	54	0
Introgression with Scale-eater			
Rum Cay	0.81	385	5
New Providence	0.72	645	9
Dominican Republic	0.81	426	11
North Carolina	0.71	163	3
Venezuela	0.69	15	0

Table S9.

Combinations of Caribbean pupfish populations used to detect signatures of introgression in San Salvador Island specialists and generalist lineages on other islands. The *fd* statistic was used to detect introgression between combinations of P2 and P3 populations, given the tree (((P1,P2),P3),O). For this series of tests we used *C. artifrons* as the outgroup in which limited gene flow is expected to have occurred with the others.

-	Sister group (P1)	Introgression into (P2)	Introgression from (P3)	Adaptive introgression regions
		Focal introgression regions in scale-eater		
A.	*C. brontotheroides*	*C. desquamator*	*C. laciniatus* NP	11
	C. brontotheroides	*C. desquamator*	*C. higuey* DR	8
	C. brontotheroides	*C. desquamator*	*C. variegatus* NC	4
	C. brontotheroides	*C. desquamator*	*C. dearborni* VZ	0
B.	*C. variegatus* SSI	*C. higuey* DR	*C. laciniatus* NP	2
	C. variegatus SSI	*C. higuey* DR	*C. variegatus* NC	3
	C. variegatus RC	*C. higuey* DR	*C. laciniatus* NP	0
	C. variegatus RC	*C. higuey* DR	*C. variegatus* NC	0
	C. variegatus SSI	*C. laciniatus* NP	*C. variegatus* NC	4
	C. variegatus RC	*C. laciniatus* NP	*C. variegatus* NC	1
	C. variegatus RC	*C. laciniatus* NP	*C. variegatus* NC	2
	C. variegatus SSI	*C. variegatus* RC	*C. higuey* DR	3
	C. variegatus SSI	*C. variegatus* RC	*C. laciniatus* NP	4
	C. variegatus SSI	*C. variegatus* RC	*C. variegatus* NC	4
		Focal introgression regions in molluscivore		
C.	*C. desquamator*	*C. brontotheroides*	*C. laciniatus* NP	5
	C. desquamator	*C. brontotheroides*	*C. higuey* DR	6
	C. desquamator	*C. brontotheroides*	*C. variegatus* NC	2
	C. desquamator	*C. brontotheroides*	*C. dearborni VZ*	0
D.	*C. variegatus* SSI	*C. higuey* DR	*C. laciniatus* NP	0
	C. variegatus SSI	*C. higuey* DR	*C. variegatus* NC	1
	C. variegatus RC	*C. higuey* DR	*C. laciniatus* NP	0
	C. variegatus RC	*C. higuey* DR	*C. variegatus* NC	0
	C. variegatus SSI	*C. laciniatus* NP	*C. variegatus* NC	1
	C. variegatus RC	*C. laciniatus* NP	*C. variegatus* NC	0
	C. variegatus RC	*C. laciniatus* NP	*C. variegatus* NC	0
	C. variegatus SSI	*C. variegatus* RC	*C. higuey* DR	2
	C. variegatus SSI	*C. variegatus* RC	*C. laciniatus* NP	1
	C. variegatus SSI	*C. variegatus* RC	*C. variegatus* NC	3

Table S10.

Candidate adaptive introgression regions from Rum Cay generalists (*C. variegatus*) and San Salvador Island specialists.

Scaffold	Variant Position	Start	End	Gene
Introgression with Molluscivore				
c_bro_v1_0_scaf11	12962909	12965001	13010000	*shisa2, atp8a2*
c_bro_v1_0_scaf16	35813565	35765001	35875000	*rfc4*
c_bro_v1_0_scaf18	18167642	18150001	18215000	*anks1a*
c_bro_v1_0_scaf18	18177499	18150001	18225000	*sarg*
c_bro_v1_0_scaf52	19358574	19345001	19395000	*fn1*
Introgression with Scale-eater				
c_bro_v1_0_scaf1	15017907	14995001	15065000	*rbm20*
c_bro_v1_0_scaf5	28411973	28365001	28455000	*gse1*
c_bro_v1_0_scaf37	3586373	3585001	3650000	*chrna7*
c_bro_v1_0_scaf43	30358142	30355001	30405000	*c1d*
c_bro_v1_0_scaf53	11080970	11080001	11130000	*zhx2*

Table S11.

Candidate adaptive introgression regions from Dominican Republic generalists (*C. higuey*) and San Salvador Island specialists.

Scaffold	Variant Position	Start	End	Gene
Introgression with Molluscivore				
c_bro_v1_0_scaf1	28938769	28935001	28985000	*rps15a*
c_bro_v1_0_scaf1	28962108	28935001	28995000	*notum2*
c_bro_v1_0_scaf1	28969771	28935001	28995000	*coq7*
c_bro_v1_0_scaf7	12326193	12305001	12375000	*smek1*
c_bro_v1_0_scaf7	12606143	12605001	12685000	*otof*
c_bro_v1_0_scaf11	11256440	11210001	11295000	*ube2w*
c_bro_v1_0_scaf18	18167642	18135001	18225000	*anks1a,sarg*
c_bro_v1_0_scaf19	6430544	6410001	6465000	*trim44*
Introgression with Scale-eater				
c_bro_v1_0_scaf5	28411973	28385001	28450000	*gse1*
c_bro_v1_0_scaf8	19759133	19735001	19790000	*map2k6*
c_bro_v1_0_scaf18	28961523	28915001	29010000	*itga5*
c_bro_v1_0_scaf19	7822448	7815001	7870000	*nap1l4*
c_bro_v1_0_scaf34	25414453	25400001	25460000	*cadps*
c_bro_v1_0_scaf34	26069290	26020001	26115000	*srgap3*
c_bro_v1_0_scaf37	3700741	3685001	3750000	*trim46*
c_bro_v1_0_scaf44	12541185	12540001	12620000	*kcnk2, cenpf*
c_bro_v1_0_scaf44	24564920	24540001	24620000	*gpm6a*
c_bro_v1_0_scaf53	18998120	18990001	19045000	*hdac9b*
c_bro_v1_0_scaf53	20294941	20245001	20330000	*steap4*

Table S12. Candidate adaptive introgression regions from New Providence Island generalists (*C. laciniatus*) and San Salvador Island specialists.

Scaffold	Variant Position	Start	End	Gene
Introgression with Molluscivore				
c_bro_v1_0_scaf1	29209555	29160001	29250000	*gga1*
c_bro_v1_0_scaf1	29241942	29195001	29250000	*klf1*
c_bro_v1_0_scaf7	12326193	12300001	12375000	*smek1*
c_bro_v1_0_scaf7	12628199	12610001	12670000	*otof*
c_bro_v1_0_scaf24	20486354	20470001	20540000	*cyp26b1*
c_bro_v1_0_scaf33	12634285	12590001	12655000	*bri3bp, wdr31*
c_bro_v1_0_scaf47	16145704	16110001	16195000	*ttc33*
Introgression with Scale-eater				
c_bro_v1_0_scaf5	27882801	27845001	27900000	*tcf12*
c_bro_v1_0_scaf7	12604722	12555001	12620000	*otof*
c_bro_v1_0_scaf11	9503186	9500001	9550000	*prlh*
c_bro_v1_0_scaf11	11975348	11930001	12010000	*ncoa2*
c_bro_v1_0_scaf16	32982520	32950001	33030000	*crocc*
c_bro_v1_0_scaf18	28961523	28915001	28970000	*itga5*
c_bro_v1_0_scaf37	8265887	8220001	8315000	*mylipa*
c_bro_v1_0_scaf43	30297117	30250001	30325000	*ppp3r1*
c_bro_v1_0_scaf53	20832687	20830001	20880000	*galnt1*

Table S13. Candidate adaptive introgression regions from North Carolina Coast generalists (*C. variegatus*) and San Salvador Island specialists.

Scaffold	Variant Position	Start	End	Gene
Introgression with Scale-eater				
c_bro_v1_0_scaf1	28962108	28945001	28995000	*notum2*,*coq7*
c_bro_v1_0_scaf1	38350857	38330001	38400000	*gpr83*
c_bro_v1_0_scaf34	32388612	32380001	32440000	eya2

Table S14.

Selective sweep ages on San Salvador Island. 95% high posterior density region of the posterior distribution of sweep ages of focal regions in scale-eater and molluscivore genomes estimated using McSwan (*48*).

Gene	Trait	Scaffold	Position Start	Position Stop	Region Size	95% HPD Lower	95% HPD Upper
			Scale-eater				
cfap20	habitat preference	c_bro_v1_0_scaf37	5000841	5017240	16399	6747.04	8490.23
prlh	habitat preference	c_bro_v1_0_scaf11	9200146	9276987	76841	6594.56	9210.36
card8	pigmentation	c_bro_v1_0_scaf46	1451011	1663431	212420	973.64	5097.36
kcnk2, cenpf	trophic morphology	c_bro_v1_0_scaf44	12227155	12305895	78740	2936.44	3966.91
smyd1	trophic morphology	c_bro_v1_0_scaf39	1643098	1647708	4610	3054.91	6030.54
tcf12	trophic morphology	c_bro_v1_0_scaf5	27975725	28016276	40551	1607.22	5119.88
twist1	trophic morphology	c_bro_v1_0_scaf53	18953132	19092361	139229	1636.34	3413.14
itga5	trophic morphology	c_bro_v1_0_scaf18	28040450	28049258	8808	1697.39	2357.06
			Molluscivore				
ext1	trophic morphology	c_bro_v1_0_scaf26	162903	230930	68027	814.01	1060.49
tiparp	trophic morphology	c_bro_v1_0_scaf21	33602383	33606685	4302	3353.84	5003.16
cyp26b1	trophic morphology	c_bro_v1_0_scaf24	20527588	20602663	75075	1447.58	4567.30

Data S1.

***Cyprinodon* pupfish sampling information.** The pond/lake names, localities, island, country, and species names, and individual codes of the pupfish individuals used in this study.

Data S2.

San Salvador Island scale-eater candidate variants. The candidate scale-eater variants that were nearly-fixed (Fst > 0.95) and in a region with a signature of a hard selective sweep and the genes within 20-kb of them.

Data S3.

San Salvador Island molluscivore candidate variants. The candidate molluscivore variants that were nearly-fixed (Fst > 0.95) and in a region with a signature of a hard selective sweep and the genes within 20-kb of them.

Data S4.

Differentially expressed genes at 2 dpf. The gene names and *P*-values of genes that were found to be significantly differential expressed (FDR> 0.05) between scale-eaters and mollsucivores at the 2 days post-fertilization larval stage in a previous study (*36*).

Data S5.

Differentially expressed genes at 8 dpf. The gene names and *P*-values of genes that were found to be significantly differential expressed (FDR> 0.05) between scale-eaters and mollsucivores at the 8 days post-fertilization larval stage in a previous study (*36*).

Data S6.

SSI GWAS trait measurements. Trait values of standard length, lower oral jaw size, nasal protrusion distance, and caudal fin pigmentation measured across the three species of San Salvador Island radiation to include in a GWAS for candidate variants underlying these traits.

Data S7

Top genomic regions associated with lower oral jaw size in a GWAS of SSI species. Regions in which all the variants within a 20-kb windows had a summed PIP score that was in the 99th percentile of all summed PIP scores for association with lower oral jaw size across 10 independent runs of Bayesian linear mixed model implemented in GEMMA (*37*).

Data S8.

Top genomic regions associated with lower caudal fin pigmentation in a GWAS of SSI species. Regions in which all the variants within a 20-kb windows had a summed PIP score that was in the 99th percentile of all summed PIP scores for association with caudal fin pigmentation across 10 independent runs of Bayesian linear mixed model implemented in GEMMA (*37*).

Data S9.

Top genomic regions associated with lower maxillary nasal protrusion in a GWAS of SSI species. Regions in which all the variants within a 20-kb windows had a summed PIP score that

1395 was in the 99th percentile of all summed PIP scores for association with maxillary nasal
1396 protrusion across 10 independent runs of Bayesian linear mixed model implemented in
1397 GEMMA.
1398
1399