Title
Hybridization and the genomic basis of adaptive radiation in Caribbean pupfish and Cameroonian crater lake cichlids

Author
Richards, Emilie

Publication Date
2022

Peer reviewed|Thesis/dissertation

Hybridization and the genomic basis of adaptive radiation in Caribbean pupfish and Cameroonian crater lake cichlids

by

Emilie J. Richards

A dissertation submitted in partial satisfaction of the

requirements for the degree of

Doctor of Philosophy

in

Integrative Biology

in the

Graduate Division

of the

University of California, Berkeley

Committee in charge:

Professor Christopher H. Martin, Chair
Professor Priya Moorjani
Professor Noah K. Whiteman

Spring 2022

Abstract

Hybridization and the genomic basis of adaptive radiation in Caribbean pupfish and Cameroonian crater lake cichlids

by

Emilie J. Richards

Doctor of Philosophy in Integrative Biology

University of California, Berkeley

Professor Christopher H. Martin, Chair

Adaptive radiations have fascinated evolutionary biologists for centuries for the bursts of phenotypic, ecological and species diversity they contain. The availability of resources in new environments with few competitors has long been seen as the major force driving adaptive radiations, but it is an equally longstanding question why only some lineages rapidly diversify in response to such new ecological opportunities while others do not. Thus, the origins and major features of adaptive radiations are still controversial.

In Chapter 1, I find evidence of introgression following past hybridization events in the genomes of species from a rare radiation of trophic specialist *Cyprinodon* pupfish endemic to San Salvador Island, Bahamas. Extensive histories of hybridization are quickly becoming another common feature that most classic adaptive radiations share, such hybridization with other species is now being viewed as another major force driving adaptive radiations in addition to ecological opportunity. Using whole genome resequencing of 42 individuals, I characterized signatures of genetic divergence and selection between the three species of the San Salvador Island radiation to identify regions of the genome involved in diversification processes on this island and found some of these regions also contain signatures of introgression, supporting a role for hybridization bringing in adaptive variation potentially relevant for the phenotypic and ecological divergence observed in trophic specialists of the radiation.

In Chapter 2, I searched for genomic signatures of secondary contact and introgression in one of the most convincing examples of sympatric speciation in the wild: the Barombi Mbo cichlid radiation. Using whole genome resequencing of 28 individuals, I characterized genetic and functional diversity in regions of the genome that have experienced gene flow differently among the species to determine the role of gene flow in the speciation process. I discovered signatures of introgression that were not shared between all species in the radiation, suggesting that hybridization with allopatric populations did occur after species divergence had already begun within the radiation. Such evidence has previously been used to reject potential case studies of sympatric speciation because it's likely that allopatric divergence contributed to the speciation events observed. However, amongst the sympatric species of the Barombi Mbo radiation, very few regions of the genome appear to have experienced differential introgression of genetic variation from allopatric outgroup populations and there was no clear evidence that introgression

following the initial stages of diversification into different species in lake brought in essential adaptive genetic variation for ecological and morphological diversity. This finding emphasizes the equivocal support that simply documenting hybridization and secondary contact in a system has on its relevancy to speciation processes and suggest that we should not rule out the possibility of sympatric speciation in one of the most celebrated examples in nature quite yet.

In Chapter 3, I reviewed theoretical models of speciation and revisit how we can connect them to the findings from recent genomic re-analyses of classic sympatric speciation case studies in the wild have revealed complex histories of secondary gene flow from outgroups. These documented histories of gene flow post-initial divergence have cast doubt on the status of sympatric speciation since they call in question whether any allopatric divergence in the outgroup populations has contributed to speciation processes within the sympatric speciation case study. I summarize theoretical differences among different types of sympatric speciation and speciation-with-gene-flow models, and propose genomic analyses for distinguishing which models apply to any given empirical case study based on the timing and function of adaptive introgression. Investigating whether secondary gene flow contributed to reproductive isolation in these empirical case studies will aid the field in better determining whether predictions of sympatric speciation theory are ultimately borne out in nature.

In Chapter 4, I expanded on the findings of adaptive introgression in the rare radiation of trophic specialists pupfish on San Salvador Island to test the hybrid origins hypothesis that adaptive radiations originate from hybrid swarms. To do this, I reconstructed the spatial and temporal histories of adaptive alleles underlying major phenotypic axes of diversification from the genomes of 202 Caribbean pupfishes. Using a combination of population genomics, transcriptomics, and genome-wide association mapping, I demonstrate that this microendemic adaptive radiation of trophic specialists on San Salvador Island, Bahamas experienced twice as much adaptive introgression as generalist populations on neighboring islands. Additionally using selective sweep timing analyses, I find evidence that adaptive divergence occurred in stages of diversification across different sources of genetic variation and trait axes once the radiation started. First, standing regulatory variation in genes associated with feeding behavior were swept to fixation by selection, then standing regulatory variation in genes associated with craniofacial and muscular development and finally a de novo non-synonymous substitution in an osteogenic transcription factor swept to fixation most recently. These results support another major hypothesis about adaptive radiations: that they proceed in temporal stages of divergence along different trait axes. The results from this study overall demonstrate how ancient alleles maintained in distinct environmental refugia can be assembled into new adaptive combinations to form adaptive radiations.

Finally, in Chapter 5, I characterized a new intermediate ‘wide-mouth’ scale-eating ecomorph in the sympatric radiation of *Cyprinodon* pupfishes on San Salvador Island. Additionally, I leveraged this ecomorph’s shared ancestry and shared novel ecological niche with scale-eater *C. desquamator* to explore the evolutionary origins of novelty and the genetic divergence that occurs during such major ecological transitions as generalist to scale-eater. Evolutionary novelty is a hallmark of adaptive radiation, yet we still don’t know how such novelties evolve on a microevolutionary scale. This intermediate ecomorph consumes scales in the wild but is genetically diverged from its sister scale-eating species and morphologically

distinct from all the other species in the radiation. I used the timing of selective sweeps on shared and unique adaptive variants in the two scale-eating specialists to characterize the adaptive walk from generalist to scale-eater. Shared adaptive regions swept first in both the specialist *C. desquamator* and the intermediate 'wide-mouth' ecomorph, followed by unique sweeps of introgressed variation in 'wide-mouth' and de novo variation in *C. desquamator*. Selection on the same adaptive alleles may have allowed both scale-eating species access to the same area of the fitness landscape but epistatic interactions with private mutations and introgressed variation in each lineage may have resulted in divergent paths to scale-eating, ultimately contributing to diverse evolutionary outcomes even from a shared starting point.

Table of Contents

Dedication

This dissertation is dedicated to all the fellow bubbles on the tide of empire out there.

Acknowledgements

The most important thing I learned from my PhD research is that even crazy, rare, and unique organisms we think are doing something completely new and hardly ever seen before are actually relying very strongly on subtle, yet crucial supporting systems. This finding has paralleled what getting a PhD was like for me. Me obtaining a PhD has relied so heavily upon my own network of peers, colleagues, friends, and family. They have truly all are the generalists that made me a not-so-special specialists and I wouldn't have it any other way.

Despite being the first in my immediate family to get a STEM PhD, my journey through graduate school has brought me closer to my family in spirit if not physical distance. My family has supported me through the altogether too many years I have spent chasing my goal to be a scientist. I appreciate their expansive support and empathy for what that kind of journey entailed even when they didn't quite always understand what I was doing (mostly because I repeatedly fail to communicate what I do in a relatable way. Note to all you – don't use the word coalescence at the dinner table). My sister Riane has embarked on a parallel journey chasing her own unique dreams and she has been a source of solidarity when my own journey down a path less traveled by has felt lonely. Maybe one day soon I will let her tattoo me, but definitely not with that manatee-mermaid hybrid she offered me last time. My mom's ceaseless generosity and thoughtfulness has really influenced the way I try to show up with my mentees and colleagues – the relationships that have been pivotal to making it through my PhD. She also provided me with so many examples of how to champion yourself and others and I still aspire to be more like her in that regard. My Dad has taught me from being on the receiving end of his endless consideration, just how powerful doing small acts of service for others and the ability to find humor in most situations can be for helping myself and others make our way through the hard times. And I wouldn't be a grateful daughter if I didn't mention that yes, Dad - the name 'Daddy Awesome Fish' is still in the running for the formal species name of the 'wide-mouth' ecomorph I discovered during my PhD. And finally, I want to thank my Papa, who made following my dreams possible by helping put me through college. He is the kind of generous dream chaser I try to remind myself to be when I am feeling particularly pessimistic. I miss you Papa and wish we had more time together, but your generosity has impacted my life in a way I will never forget.

It has been quite literally a coast to coast journey these six years of my PhD, and my advisor Chris Martin has been there through it all. His willingness to share stories about his personal life and family really solidified to me that real people that I can relate to participate in science from the moment I joined the lab. I can't tell you how much it would make me smile to walk into the fish room and see that his daughters had left little doodles on the fish tanks. I came into his lab with a background in statistical phylogenetic and zero experience with live animal research, but his infectious enthusiasm for fish was hard not to catch on to and I thank him for encouraging a deeper love of fish in me – especially a love of ones doing absolute bonkers things out there. Working with fish in the lab gave me much joy during my PhD – so a special thank you to Noodle, Pinecone, Bubble Boy, and Big Boy for giving me something to look forward to every time I stepped in the fish rooms.

The members of the Martin Lab have made lab communities so special to me and I hope my own lab group one day has a similar dynamic to what we had across these six years. Joe

McGirr snatched the official title of first graduate student in the lab away from me by deciding to join the lab a few months before I did, but I appreciate having his support in our efforts to pioneer the way into pupfish genomics territory. I have always admired Joe for his effortless ability to blend and transition between absolute pure silliness and smart/insightful conversation. I especially want to thank him for bringing me along for those small moments of awe in which he shared with me his wonder for all that we were accomplishing in science - even though we were quite often a little tipsy when such moments came to us. Truly all the graduate students, postdocs and lab managers that I overlapped with in the lab (Jackie Galvez, David Tian, Austin Patton, Feña Palominos, Joe Heras, Jelmer Poelstra, Nicole Suren, Kristi Dixon, Logan Buie, Kate Gould) have been a phenomenal support system – especially in a pandemic environment that taxed all our support systems.

One of the biggest surprises of my PhD was getting to join and interact with the community members of the Museum of Vertebrate Zoology. At first, I was very nervous to join such a lauded community full of renowned naturalists, curators, scientists at a time in which I felt like none of those things myself. But they had me hooked and eager for more interactions after the very first museum tour. I am so appreciative of them sharing their passion, wisdom, and kindness with me. They have forever changed the way I view museums and left me so impressed by the truly expansive set of skills and knowledge that they all bring to bear in collections such as theirs. I will miss having such daily access to endless discovery at my fingertips, awaiting me just behind the next specimen cabinet door, but will always look back fondly on what I did manage to discover during my time there. I also appreciate them answering all my borderline insulting questions about the collections over the years, like whether we have any marsupials in the collections.

There were two undergraduates who I directly mentored one-on-one research projects in pupfish with during my PhD, Takao and Anay, and they were absolute delights. I am so impressed with all they accomplished (especially in pandemic times) and I am so proud of them for chasing their dreams. They gave me a glimpse of what it could be like as a professor mentoring researchers one day and reminded me of how powerful a motivator the simple act of getting to share yourself with people can be for wanting to participate in scientific endeavors. I hope they learned even a fraction of things from me as I learned from them. And truly, there were so many undergraduates I connected with while teaching courses as part of my degree. They brought me so many small moments of joy and made me feel a part of the university outside of my research and office space. Thanks for tolerating and even sometimes laughing at all my terrible puns in my lectures.

My PhD has also really revealed to me the necessity of solid science friendships the blend both scientific and personal support. Molly Womack and Mallory Ballinger really made Berkeley and the MVZ feel like home. Additionally, thanks to Mallory for introducing me to Taco Bell Cantina - my lab will tell you I never shut up about it. You two have always been a safe space to talk about life and science and I really do always think it’s been nice seeing you.

Last but definitely not least, this entire PhD dissertation exists in large part to one Michelle St John. Although it’s difficult to see from the outside since she was a co-author on only one of my dissertation chapters, there are traces of Michelle in every single bit of the

science I did during my PhD. I have never met a more whip smart, funny, creative, and generous being. And I don't say this in that way that we can say it when we are idolizing people from a far by because I haven't been able to tell where her desk ends and mine begins for years. Michelle has taught me almost too much about fish, stats, and friendship in our five years together and I am hoping she continues to give me lessons in the future. I enjoy being both supported and pushed by her. I have enjoyed laughing, and stressing out, just doing anything with her these last five years. I didn't know science could be so fun until I started working with Michelle and I don't know where I would be in my pursuit of it today without her support. Michelle is an excellent mentor and friend and I look forward to working with her in the future even when our desks do become isolated by distance. I owe much of my sanity, happiness, and productivity during my PhD to her and I hope she got even a tenth of those same benefits from me in return.

Lastly, I want to acknowledge how proud I am of myself and literally everyone who takes this journey through a PhD – its rewarding yes, but its challenging as all heck and I have never pushed myself so hard professionally or personally before. I have learned much more about myself during this process than I have about evolutionary biology and I hope you do to. For all its many, many flaws – I have never found myself more at grounded or at home than in science communities and there are so many people out there who have contributed to that feeling in big and small ways and it has made all the difference to me.

Dissertation Introduction

Evolutionary radiations have long been central to our understanding of evolution because they generate a wealth of ecological, phenotypic, and species diversity in rapid bursts. Some lineages diversify much more and much faster than others, leaving behind a striking pattern of species unevenness across the tree of life (Simpson 1944a; Alfaro et al. 2009; Glor 2010; Uyeda et al. 2011; Rabosky et al. 2012; Landis and Schraiber 2017). Such bursts are additionally fascinating because they contradict much of our current understanding of speciation processes (Martin and Richards 2019). Mechanistic speciation models predict that diversification should slow with time as available niche space becomes increasingly subdivided and disruptive selection becomes weaker with each recurrent speciation event (e.g. (Dieckmann and Doebeli 1999; Polechová et al. 2005; Bolnick 2006). In opposition to the expected pattern that species and trait divergence should slow down with each subsequent speciation event adaptive radiations often contain an initial phase of rapid, repeated bouts of speciation (Harmon et al. 2010; Stroud and Losos 2016; Martin and Richards 2019). There is little theory out there about how multiple speciation events can occur simultaneously (but see;(Kondrashov and Kondrashov 1999; Gavrilets 2014)) and the origins, underlying mechanisms, and major features of adaptive radiation are still poorly understood.

Ecological opportunity, or access to new environments with abundant resources and few competitors has long been viewed as the major driver of radiations (Schluter 2000; Harmon et al. 2010; Stroud and Losos 2016). However, common sources of ecological opportunities such as evolution of key innovations does not always lead to increased diversification (Seehausen 2006; Harmon and Harrison 2015)(Alfaro et al. 2009a, Harmon & Harrison 2015, McGee et al. 2015, Rabosky 2017, Seehausen 2006) and many evolutionary lineages colonize new environments and have not diversified into adaptive radiations (Roderick and Gillespie 1998; Lovette et al. 2002; Martin and Wainwright 2013a; Martin 2016b). This suggest other microevolutionary forces may be necessary to explain the origins of adaptive radiation and identifying additional explanatory factors for why some lineages radiation and others do not is a major goal in evolutionary biology.

In our search for additional factors, genomic analysis of many classic adaptive radiations over the past two decades has revealed that extensive histories of hybridization are prevalent among most radiations and has led to an additional origins hypothesis. The hybrid swarm hypotheses proposes that hybridization among distinct lineages can introduce genetic diversity and novel allele combinations genome-wide that may trigger rapid diversification in the presence of abundant ecological opportunity (Seehausen 2004). However, it is still unclear how often hybridization is necessary for rapid diversification, as opposed to simply being pervasive throughout the history of any young rapidly diversifying group (Berner and Salzburger 2015; Martin et al. 2015a). One of the only examples with strong evidence of hybridization leading to ecological and species diversification is that of several hybrid species within a radiation of *Helianthus* sunflowers (Schwarzbach and Rieseberg 2002; Welch and Rieseberg 2002; Rieseberg et al. 2003; Gross and Rieseberg 2005; Whitney et al. 2010; Schumer et al. 2014b). Even these convincing cases may simply represent examples of multiple homoploid speciation events within an already radiating lineage rather than a hybrid swarm scenario. So while there is convincing

evidence that hybridization can facilitate diversification among species pairs, it is still unclear whether hybridization is a major factor constraining adaptive radiation in some lineages. Additionally, very few studies look at differences in hybridization and introgression signatures between closely related evolutionary lineages that have and haven't radiation to be able to tell whether hybridization is sufficient to explain what triggers adaptive radiation.

Beyond the origins of adaptive radiations, how adaptive radiations proceed once they begin is an equally as mysterious. For example, whether adaptive divergence during radiation proceeds in temporal stages is still highly debated. One of the predominant arguments that speciation and adaptive radiation proceed in adaptive stages across distinct axes of trait divergence remains highly controversial. For example, one hypothesis proposes three stages of vertebrate adaptive radiation: first shifts in habitat, followed by divergence in trophic morphology, and finally sexual communication signals like color (Streelman and Danley 2003). However, existing evidence for these hypotheses that radiations occur in adaptive stages have largely been based on the phylogenetic distribution of extant traits (Diamond 1986; Losos et al. 1995; Danley and Kocher 2001; Streelman et al. 2002; Streelman and Danley 2003; Ronco et al. 2020) and depend heavily ancestral state reconstructions of rapidly diversifying traits that can be highly unreliable without fossil data (Glor 2010; Sallan and Friedman 2012; Duchêne and Lanfear 2015). The timing of diversification on a phylogeny is also confounded by different rates of diversification across different trait axes (Glor 2010) and rapidly evolving traits can appear to have diverged discretely and only recently even if their divergence occurred continuously across time and conjointly with other traits. Recent population genomic approaches for investigating the timing of selection across multiple trait axes and identifying the sources of genetic variation in recent radiations provide a new opportunity to understand both the temporal and spatial dynamics of adaptation and speciation during adaptive radiation (Nakagome et al. 2016; Smith et al. 2018; Miller et al. 2019; Richards et al. 2019).

Additionally, spectacular ecological transitions, such as blood-drinking (Grant and Grant 2008) or plant carnivory (Givnish et al. 1984, 1997), are a hallmark of adaptive radiations, yet we still know little about how such seemingly discontinuous transitions occur during rapid these bursts of diversification. Recent conceptual frameworks for understanding adaptation to novel fitness peaks suggest that ecological transitions are more likely to occur in stages (Blount et al. 2012; Erwin 2019, 2021). The initial emergence of a novel trait is likely to require further adaptive refinement to become successfully incorporated into the functional ecology of an organism. One of the best examples is the evolution of the novel niche of aerobic citrate metabolism in *E. coli*, in which evolution of *citT* promotor capture in one intermediate lineage actualized the phenotype of aerobic citrate metabolism while further mutations at *dctA* in subsequent lineages refined the efficiency of citrate utilization that allowed it to fully occupy this niche (Blount et al. 2012, 2020; Quandt et al. 2014; Turner et al. 2015). Such experimental studies lend support to the idea that novel ecological transitions also occur in a series of stages at the genetic level in which a complex series of mutations arise that potentiate, actualize, and refine the adaptations to colonize new fitness peaks (Jagdish et al. 2020). However, these studies are limited largely to adaptation in highly experimental settings (Weinreich et al. 2006; Blount et al. 2008, 2012) or along a single selective axis (Fry 2014; Tarvin et al. 2017; Karageorgi et al. 2019) and the universality of these findings to novel ecological transitions in the wild is largely unknown.

Thus much about the microevolutionary processes that lead to the paradoxical burst of diversification observed in adaptive radiations remains largely unknown and poorly tested from a theoretical and empirical standpoint. Despite substantial evidence of adaptive introgression during radiation, very few studies have compared adaptive introgression between closely-related radiating and non-radiating lineages to distinguish introgression as necessary for radiation as predicted by the hybrid swarm hypothesis. Empirical evidence for such predications has been difficult to gather from classic adaptive radiations, because many are restricted within a single unique environment (such Darwin's Finches in the Galapagos archipelago) making it hard to disentangle ecological opportunity from genetic factors or repeatedly radiated across similar unique environments (such as Anolis lizards across the Greater Antilles), making it hard to find systems with radiations nested within many closely related outgroup populations that haven't radiated in similar conditions. However, a nascent adaptive radiation of Caribbean pupfishes provides an excellent opportunity to assess whether hybridization is strongly associated with adaptive radiation and understanding how the rapid evolution of major ecological transitions occur in nature (Turner et al. 2008; Martin and Wainwright 2013a) This radiation contains a wide-spread generalist pupfish species (*Cyprinodon variegatus*) that occurs in sympatry with two previously described trophic specialists that are endemic to many of the hypersaline lakes on the island: a molluscivore (*C. bronotheroides*) with a novel nasal protrusion which is an oral-sheller of gastropods (St. John et al. 2020) and a scale-eating specialist (*C. desquamator*) with two-fold larger oral jaws (Martin and Wainwright 2013a). The evolutionary novelties in this system originated recently; the hypersaline lakes on San Salvador Island were dry during the last glacial maximum 6-20 kya years ago (Hagey and Mylroie 1995; Turner et al. 2008; Clark et al. 2009). Intriguingly, we recently discovered a fourth species of pupfish living in sympatry with the two specialists and generalist on San Salvador Island (Richards and Martin 2017a).

My dissertation research leveraged this rare and young radiation of Caribbean pupfishes that occurs only on a single island, yet is nested within a large network of closely related single species pupfish lineages on similar islands, to test out major hypotheses about how adaptive radiations start and proceed. In chapters 1-4, I explore the evidence across this rare radiation and other similar radiations that hybridization plays a necessary role in adaptive radiation. In chapter 4-5, I additionally explore the microevolutionary forces underlying major features of the adaptive radiation to investigate how the major ecological transitions occurred between generalist to scale-eating specialist and whether adaptive diversification occurred in stages across the broader scale of the entire radiation.

Chapter 1: Adaptive introgression from distant Caribbean islands contributed to the diversification of a microendemic adaptive radiation of trophic specialist pupfishes

This chapter has been previously published and is reproduced here in accordance with the journal's article sharing policy:

Richards EJ, Martin CH. 2017. Adaptive introgression from distant Caribbean islands contributed to the diversification of a microendemic adaptive radiation of trophic specialist pupfishes. *PloS Genetics 13 (8)*: 68-80. DOI: 10.1371/journal.pgen.1006919

1.1. Abstract

Rapid diversification often involves complex histories of gene flow that leave variable and conflicting signatures of evolutionary relatedness across the genome. Identifying the extent and source of variation in these evolutionary relationships can provide insight into the evolutionary mechanisms involved in rapid radiations. Here we compare the discordant evolutionary relationships associated with species phenotypes across 42 whole genomes from a sympatric adaptive radiation of *Cyprinodon* pupfishes endemic to San Salvador Island, Bahamas and several outgroup pupfish species in order to understand the rarity of these trophic specialists within the larger radiation of *Cyprinodon*. 82% of the genome depicts close evolutionary relationships among the San Salvador Island species reflecting their geographic proximity, but the vast majority of variants fixed between specialist species lie in regions with discordant topologies. Top candidate adaptive introgression regions include signatures of selective sweeps and adaptive introgression of genetic variation from a single population in the northwestern Bahamas into each of the specialist species. Hard selective sweeps of genetic variation on San Salvador Island contributed 5 times more to speciation of trophic specialists than adaptive introgression of Caribbean genetic variation; however, four of the 11 introgressed regions came from a single distant island and were associated with the primary axis of oral jaw divergence within the radiation. For example, standing variation in a proto-oncogene (*ski*) known to have effects on jaw size introgressed into one San Salvador Island specialist from an island 300 km away approximately 10 kya. The complex emerging picture of the origins of adaptive radiation on San Salvador Island indicates that multiple sources of genetic variation contributed to the adaptive phenotypes of novel trophic specialists on the island. Our findings suggest that a suite of factors, including rare adaptive introgression, may be necessary for adaptive radiation in addition to ecological opportunity.

1.1.1. Author summary

Groups of closely related species can rapidly evolve to occupy diverse ecological roles, but the ecological and genetic conditions that trigger this diversification are still highly debated. We examined patterns of molecular evolution across the genomes of a recent radiation of pupfishes that includes two trophic specialists. Despite apparently widespread ecological opportunities and gene flow across the Caribbean, this radiation is endemic to a single Bahamian island. Using the whole genomes of 42 pupfish we found evidence of extensive and previously unexpected variation in evolutionary relatedness among Caribbean pupfish. While adaptive introgression appears to be rare across the genomes of the San Salvador Island species, it may have introduced adaptive variants important in the evolution of the complex phenotypes of the specialists. Four of the 11 candidate adaptive introgression regions contain genes with known effects on jaw morphology in zebrafish or associated with pupfish jaw size, the primary axis of phenotypic divergence between species in this system. Our findings that multiple sources of genetic variation contribute to the San Salvador Island radiation suggests that a complex suite of factors, including hybridization with other species, may be necessary for adaptive radiation in addition to ecological opportunity.

1.2. Introduction

Adaptive radiations are central to our understanding of evolution because they generate a wealth of ecological, phenotypic, and species diversity in rapid bursts. However, the mechanisms that trigger rapid bursts of trait divergence, niche evolution, and diversification characteristic of classic adaptive radiations are still debated. The availability of resources in new environments with few competitors has long been seen as the major force driving adaptive radiations (Simpson 1944a; Schluter 2000; Losos and Mahler 2010), but it is a longstanding question why only some lineages rapidly diversify in response to such ecological opportunities while others do not (Roderick and Gillespie 1998; Burns et al. 2002; Thorpe et al. 2008; Seehausen and Wagner 2014; Erwin 2015; Harmon and Harrison 2015).

While gene flow can impede or reverse diversification among incipient species by reducing genetic differentiation and subsequent recombination can break down locally adapted haplotypes (Grant 2002; Lenomand 2002; Coyne and Orr 2004b; Taylor et al. 2006), it can also introduce adaptive genetic variants (Abbott et al. 2013; Seehausen 2013) and/or genetic incompatibilities (Wright et al. 2013; Schumer et al. 2014a, 2015) that initiate or contribute to the process of speciation. A growing number of studies have identified gene flow and genome-wide introgression across a range of adaptive radiations (Garrigan et al. 2012; Hedrick 2013; Martin et al. 2013a, 2015a; Fontaine et al. 2015; Lamichhaney et al. 2015; Malinsky et al. 2015; Kautt et al. 2016a), contributing to the emerging view that gene flow is pervasive throughout the history of many young rapidly diversifying groups and may be necessary for adaptive radiation. Examples of adaptive radiations with histories of extensive hybridization include *Heliconius* butterflies (Mallet et al. 2007; The Heliconius Genome Consortium et al. 2012; Martin et al. 2013b), Darwin's finches (Lamichhaney et al. 2015; Palmer and Kronforst 2015; Almen et al.

2016; Han et al. 2017), *Anopheles* mosquitos (Fontaine et al. 2015; Wen et al. 2016), and cichlids (Brawand et al. 2014; Ford et al. 2015; Malinsky et al. 2015; Martin et al. 2015a; Gante et al. 2016; Kautt et al. 2016a; Meier et al. 2017a,b). The hybrid swarm hypothesis (Seehausen 2004) proposes that hybridization among distinct lineages can introduce genetic diversity and novel allele combinations genome-wide that may trigger rapid diversification in the presence of abundant ecological opportunity. However, it is still unclear how often hybridization is necessary for rapid diversification, as opposed to simply being pervasive throughout the history of any young rapidly diversifying group (Berner and Salzburger 2015; Martin et al. 2015a). One of the only examples with strong evidence of hybridization leading to ecological and species diversification is that of several hybrid species within a radiation of *Helianthus* sunflowers (Schwarzbach and Rieseberg 2002; Welch and Rieseberg 2002; Rieseberg et al. 2003; Gross and Rieseberg 2005; Whitney et al. 2010; Schumer et al. 2014b). However, these may simply represent examples of multiple homoploid speciation events within an already radiating lineage rather than a hybrid swarm scenario. So while there is convincing evidence that hybridization can facilitate diversification among species pairs (but see (Kautt et al. 2016a; Meier et al. 2017a) for a potential multispecies outcome of hybridization), it is still unclear whether gene flow is a major factor constraining adaptive radiation in some lineages or if ecological opportunity is the sole constraint.

The adaptive radiation of San Salvador Island pupfishes provides an outstanding system to compare the contributions of different sources of genetic variation to rapid diversification and the role of gene flow in the evolution of complex phenotypes. Pupfish species of the genus *Cyprinodon* inhabit saline lakes and coastal areas across the Caribbean and Atlantic and nearly all pupfishes are allopatric, dietary generalists consuming algae and small invertebrates (Martin and Wainwright 2011). In contrast, three *Cyprinodon* species live sympatrically in the hypersaline lakes of San Salvador Island and comprise a small radiation that has occurred within the past 10,000 years based on the most recent glacial maximum when these lakes were dry due to lowered sea levels (Hagey and Mylroie 1995; Holtmeier 2000; Turner et al. 2008). This radiation is composed of the widespread generalist algae-eating species *Cyprinodon variegatus* and two endemic specialists that coexist with the generalist in all habitats in some lakes. These specialists have adapted to unique trophic niches using novel morphologies: the molluscivore *Cyprinodon brontotheroides* with a unique nasal protrusion and the scale-eater *Cyprinodon desquamator* with enlarged oral jaws and adductor mandibulae muscles (Martin and Wainwright 2011, 2013a). Surveys of populations living on neighboring islands in the Bahamas and phylogenetic analyses with other *Cyprinodon* species indicate that these specialist species are endemic to the hypersaline lakes of San Salvador Island and that both specialists arose from a generalist common ancestor during this recent radiation (Martin 2016b).

The currently available ecological and genetic data on the group provides little indication as to why this radiation is localized to a single island. Variation in ecological opportunity among hypersaline lake environments in the Caribbean does not appear to explain the rarity of this radiation (Martin 2016b). This finding suggests a potentially important role for sufficient genetic variation to respond to abundant, underutilized resources in these environments. However, a hybrid swarm hypothesis about the origins of the radiation does not appear to explain its rarity either: genetic diversity is comparable among islands and gene flow occurs among all Caribbean islands investigated, not only into San Salvador Island (Martin 2016b). Novel traits and increased rates of diversification associated with them are well documented in this system (Martin and Wainwright 2011, 2013d; Martin 2016b), but understanding the rarity of this

adaptive radiation requires a thorough investigation of the underlying genetic variation that accompanies these rare ecological transitions. A recent study investigating the genetic basis of trophic specialists in this radiation revealed very few regions underlying these phenotypes (McGirr and Martin 2016a). Only thousands of variants out of 12 million were fixed between the scale-eater and molluscivore species. Since genetic divergence is limited to particular regions, localized rather than genome-wide investigations of the genome will be important for understanding how genetic variation, possibly originating outside of San Salvador Island, has contributed to the exceptional phenotypic diversification restricted to this island. Here, we use a machine-learning approach to identify regions of the genome with different evolutionary relationships among 42 pupfish genomes sampled from the San Salvador Island radiation, two distant Caribbean islands, and 3 additional outgroups. We then scan the genome for evidence of localized introgression with pupfish populations outside of San Salvador Island and compare the relative contributions of adaptive introgression from two distant islands and hard selective sweeps to the divergence of each specialist species.

1.3. Results

1.3.1. Extensive variation in patterns of evolutionary relatedness across the genome

To identify localized patterns of population history across the genome, we used the machine-learning approach SAGUARO. SAGUARO combines a hidden Markov model with a self-organizing map to characterize local topologies across the genome among aligned individuals (Zamani et al. 2013). This method does not require any *a priori* hypotheses about the relationships among individuals, but rather infers them directly from the genome by finding regions of consecutive nucleotides with a similar pattern of genetic differentiation, building hypotheses about relationships among individuals from these genetic differences, and then assigning regions of the genome to these hypothesized local topologies. Since smaller segments with fewer informative SNPs are more likely to be incorrectly assigned to a hypothesized topology by chance (pers. comm. M.G. Grabherr), we tested various minimum SNP filters for reducing the amount of short, uninformative segments assigned to topologies by chance and found that increasingly stringent filters over 20 SNPs did not substantially reduce the number of uninformative segments. Using this approach and our 20 SNP filter, we partitioned the genome into a total of 15 unique topologies across 227,248 genomic segments that ranged from 101-324,088 base pairs in length (median: 852 bp) (S1 and S2 Figures; S1 Table). The 15th topology was uninformative about either species or population level relationships, so it was removed from downstream analyses.

The most prevalent history across 64% of the genome featured the expected species phylogeny for this group from previous genome-wide studies (Martin and Wainwright 2011; Martin and Feinstein 2014; Martin 2016b), in which all individuals from San Salvador Island grouped by species into a single clade with distant relationships to outgroup generalist pupfish populations from other islands in the Caribbean, Death Valley in California, and a second radiation in Mexico spanning the most divergent branch of the *Cyprinodon* tree (Figure 1). Unlike previous genome-wide phylogenies (Martin and Feinstein 2014; Martin 2016b), and with the exception of a few individuals that grouped with molluscivores by lake, the generalists on San Salvador Island form a discrete clade from the molluscivores and scale-eaters.

Within this dominant topology, scale-eater*s* from six lakes on San Salvador Island fell into one of two separate clades: small-jawed individuals from Osprey Lake, Great Lake, and Oyster Pond and large-jawed individuals from Crescent Pond, Stout Lake, Osprey Lake, and Little Lake (Figure 1). Molluscivores did not form a single clade as individuals from some lakes (Crescent Pond and Moon Rock) were more closely related to generalists from the same lake than molluscivores from other lakes, similar to previous genome-wide phylogenies (Martin and Feinstein 2014). Another topology covering 10% of the genome was very similar to the dominant one, differing only in the relationships among San Salvador Island generalists (Figure S1). Additional topologies spanning 7.6% of the genome featured a single San Salvador Island clade but also depicted a closer relationship between San Salvador Island and the outgroups as well as groupings of all three San Salvador Island species by lake in Crescent Pond and Moon Rock Pond. When combined with the dominant topology, only 82.6% of the genome supported the expected San Salvador Island clade (S1 Table).

In other regions of the genome, San Salvador Island did not form a single clade (Figures 2A-C and S2, S1 Table). The most frequently observed alternative relationships depicted specialist individuals as a clade outside of the San Salvador Island group and sister to all the outgroup *Cyprinodon* species (Figures 2A,B). The 'large-jawed scale-eater topology' featured large-jawed scale-eaters outside of the San Salvador Island clade, sister to all other outgroups, and was assigned to 4,437 segments covering 3.27% of the genome (Figure 2A). Another topology, the 'molluscivore topology', showed a similar pattern in which the molluscivores formed a single clade outside of the San Salvador Island group and sister to all other outgroups (Figure 2B). This molluscivore topology was assigned to 3,916 segments and covered 3.11% of the genome. Another 2,029 segments covering 1.66% of the genome were assigned to a topology where both the large-jawed and small-jawed scale-eaters formed a combined clade outside of the San Salvador Island group, the 'combined scale-eater topology' (Figure 2C). Other topologies featuring one of the specialists separated from the rest of San Salvador Island covered 0.76%-2.48% of the genome (S1 Table).

Unexpectedly, all 14 informative topologies separated scale-eaters into groups of small- and large-jawed individuals and the relationships between these two groups and other species differed across different regions of the genome. In some regions, the small-jawed scale-eater individuals were sister to the large-jawed scale-eaters (Figures 1,2B-C, S1 and S2). In other regions, the small-jawed scale-eaters were more closely related to the generalists and molluscivores (Figures 2A, S1 and S2). These small-jawed scale-eaters may be a product of ongoing hybridization between species on San Salvador Island or a new 'occasional' scale-eating ecomorph, perhaps representing an intermediate yet viable stage on the evolutionary path towards large-jawed scale-eaters, in which scales form the majority of their diet (Martin and Wainwright 2013d). The presence of homozygous genotypes in all five individuals of small-jawed scale-eaters for variants fixed in both large-jawed scale-eaters and generalists is not consistent with first generation hybrids (S2 Table). They also do not fit the ancestry proportions expected in F_2 hybrids (χ^2=429.6, P=5.16e^{-94}). We might expect increased linkage disequilibrium (LD) in the small-jawed scale-eaters if they represent recent hybridization events between distinct populations. Consistent with this idea, LD decays more slowly in the small-jawed scale-eaters (after approximately 120 kb) than in the three San Salvador Island species (after approximately 50kb: Figure S3). However, strong LD and long haplotype blocks may also result

from other evolutionary phenomena like recent population bottlenecks (e.g. (Reich et al. 2001)). Demographic modeling with a larger sample will be needed to distinguish whether these small-jawed scale-eaters represent hybrids from ongoing or recent gene flow on San Salvador Island or a potential new ecomorph.

1.3.2. Localized introgression into both specialists from across the Caribbean

We examined signals of introgression from two distant pupfish generalist populations in the Caribbean: Lake Cunningham, New Providence Island in the Bahamas (described as the endemic species *Cyprinodon laciniatus* (Hubbs and Miller 1942)) and Etang Saumatre / Lac Azuei in the Dominican Republic (described as the endemic species *Cyprinodon bondi* (Smith 1989)). *C. laciniatus* exhibits morphological variation not observed in other generalist species, including laciniated scales and variation in oral jaw size (Hubbs and Miller 1942), although not the extreme oral jaw morphologies observed in the specialists, and is an interesting candidate for looking at adaptive introgression of variants involved in oral jaw size morphology on San Salvador Island. *C. bondi* is a generalist species of the *variegatus* complex from the south-eastern end of the range of Greater Antillean pupfish and introgression with San Salvador Island populations would suggest that Caribbean-wide gene flow may have contributed to the adaptive radiation on San Salvador Island. We characterized the genomic landscape of introgression in the three San Salvador Island species using f_4 statistics that were initially developed to test for introgression among human populations (Reich et al. 2009; Patterson et al. 2012; Pickrell and Pritchard 2012).

Genome-wide f_4 tests provided evidence of introgression between Caribbean outgroups and San Salvador Island. f_4 values significantly deviated from the null hypothesis of no introgression (f_4 =0) in the scale-eater/molluscivore ($Z = 4.2$, $P = 2.67 \times 10^{-5}$), and scale-eater/generalist combinations ($Z = 4.67$, $P = 3.01 \times 10^{-6}$), but were not significant in the molluscivore/generalist combination ($Z = -1.63$, $P = 0.103$).

When f_4 was calculated in windows, we found that 181 10-kb regions out of 100,260 (0.18%) contained significant evidence of introgression between *C. laciniatus* or *C. bondi* and the San Salvador Island specialists (Figure 3A). Introgressed regions were scattered across the genome in 107 of the 9,259 scaffolds in our dataset. These regions were not typically concentrated in one section of the genome, with the largest cluster within a single scaffold containing 12% of the total (Figure 3A).

The genomic regions with significant evidence of introgression varied between the two specialists (Figure 3B,C): only 15 regions from the 176 and 112 regions with significant evidence of introgression were shared between generalist/scale-eater and generalist/molluscivore comparisons, respectively. This suggests that admixture with other Caribbean populations occurred multiple times and independently for each specialist or that different introgressed regions were used by the two specialists after a single admixture event (see Figure S4-S6 for full Manhattan plots).

We also tested for introgression with the small-jawed scale-eaters excluded to search for potential introgression with the large-jawed scale-eaters alone (Figure S7). Introgressed regions were less variable between the two groups of scale-eaters, with 122 of 209 candidate introgressed regions shared. The 87 introgressed regions unique to the large-jawed scale-eaters suggest that some introgression may have occurred between populations on other Caribbean

islands and the large-jawed scale-eater population independently from the small-jawed scale-eaters.

Regions of low diversity and low recombination may be biased when genome-wide tests of introgression, such as the f_4 statistic, are applied to genomic windows (Martin et al. 2015b). To assess whether our introgressed regions were the result of this bias, we looked at π estimates across the detected regions of introgression in comparison to the genome-wide estimates (mean D_{xy}= *0.007*; mean π scale-eater=*0.0048* ; mean π molluscivore =*0.0054*) and variance in f_4 statistic values. f_4 statistics do appear slightly sensitive to the level of diversity in a region, with variance in f_4 values having a weak negative correlation with mean scaffold π (Pearson's r = -0.18; Figure S8), and a weaker correlation between the value of f_4 and π (Pearson's r = -0.013; Figure S9). However, in selecting our top candidate introgressed regions, we assessed π in all three San Salvador Island species and looked for other signals of introgression to complement the f_4 test. This included pairwise estimates of D_{xy} between each San Salvador Island species and outgroups, TREEMIX analyses used to infer admixture events on population graphs (Pickrell and Pritchard 2012), presence of alternative topologies in the regions, and maximum likelihood trees supporting close relationships between outgroups and either of the specialists. f_4 outliers that appeared in extensive regions of low diversity in all three San Salvador Island species and did not have supporting evidence from other statistics or trees were excluded from the list of candidates as potential false positives in areas of low recombination (n = 2; Figures S10 and S11).

1.3.3. Multiple sources of genetic variation underlie species divergence

The relationships observed in the three alternative topologies (Figure 2) underlie most of the divergence observed between the molluscivores and scale-eaters: 75% and 88% of the fixed SNPs between molluscivores and large-jawed scale-eaters and molluscivores and all scale-eaters, respectively, fall within these topologies that make up less than 5% of the genome in total. Many of these regions contained candidate genes previously associated with variation in *Cyprinodon* jaw size within the San Salvador Island radiation (McGirr and Martin 2016a): 18 of the 31 candidate jaw genes occurred in the combined scale-eater topology and 1 candidate region in the molluscivore topology.

We also assessed the relative contributions of different sources of genetic variation to the divergence between the two specialists (also see Figure S12). Selective sweeps of introgressed variation from our two focal outgroups contributed 5 and 8 times less to species divergence between the scale-eaters and molluscivores, respectively, than sweeps of other sources of genetic variation (Figure 4). Adaptive introgression in regions of high divergence among the specialists appears to be rare, occurring in only 0.006 and 0.016% of the scale-eater and molluscivore genomes, respectively. The higher percentage in the molluscivore genome may be due to stronger bottlenecks in their past than in the scale-eaters, rather than more selective sweeps in this species. Within individual lakes, molluscivores have lower genetic diversity than both scale-eaters and generalists (Martin and Feinstein 2014). When segments are additionally separated based on topology assigned by SAGUARO, the alternative topologies contained a greater proportion of regions with introgressed genetic variation and selective sweeps than those regions assigned to the dominant topology. None of the fixed SNPs in adaptive introgression candidates occurred in a segment assigned to the dominant topology (Figure S12).

1.3.4. Adaptive introgression contributed to localized adaptive radiation

In general, selective sweeps of introgressed genetic variation that contributed to species divergence between the specialists were rare. However, four of the 11 candidate adaptive introgression regions contained genes with known craniofacial effects in model organisms or have been strongly associated with oral jaw size variation in the specialists (McGirr and Martin 2016a), the primary axis of diversification in this system (Table 1 and S3). Only one of these, the proto-oncogene *ski*, has both known craniofacial effects and was associated with jaw size variation in the specialists. *Ski* encodes for a corepressor protein involved in the SMAD-dependent transcription growth factor B pathway (Nagase et al. 1990; Engert et al. 1995; Wotton and Massagué 2001). Mutations in *ski* cause marked reductions in skeletal muscle mass, depressed nasal bridges, and shortened, thick lower jaw bones in mice (Berk et al. 1997; Colmenares et al. 2002) and malformed craniofacial cartilage and shortened lower jaws in zebrafish (Doyle et al. 2012). These phenotypic changes are remarkably similar to the novel craniofacial morphologies in San Salvador Island molluscivore pupfishes, including increased nasal/maxillary protrusion, shortened lower jaw, and thicker dentary and articular bones (Martin and Wainwright 2013a).

The candidate adaptive introgression region spans the start of *ski* and contains three fixed SNPs, one in the 3' untranslated region, one in the 3rd codon position of an exon, and one in an intron. This region contains a signature of high absolute genetic divergence between the two specialists and a selective sweep in the molluscivore (Figure 5). This region also features low nucleotide diversity within scale-eaters and negative estimates of Tajima's D, although this does not appear to be as strong as in the molluscivores. Several lines of evidence point towards the introgression of *ski* variants between molluscivores and *C. laciniatus*. Genetic differentiation is minimal between molluscivores and *C. laciniatus* (D_{xy} = 0.0011) (Figure 5) and higher in all other pairwise comparisons (D_{xy} > 0.013) between the two specialists and two outgroup Caribbean pupfish species (S4 Table), indicating gene flow between the molluscivores on San Salvador Island and the generalist *C. laciniatus* on New Providence Island. Taking a closer look at the genetic variation in this region, we observe that the *ski* SNPs fixed in the San Salvador Island molluscivores are homozygous in *C. laciniatus* and segregating in the generalists (Figure 6A), suggesting that they occur at an appreciable frequency in the generalists. The surrounding molluscivore genetic background of the fixed *ski* SNPs is very similar to *C. laciniatus* (Figure 6B). In this 10-kb region, only 62 SNPs differ between the molluscivores and *C. laciniatus* in our sample. Segments of this region were assigned to the combined scale-eater topology (Figure 2C) and a maximum likelihood tree of the SAGUARO segment containing these three fixed SNPs features *C. laciniatus* in a clade with molluscivores (Figure S13).

In addition to *ski*, one other adaptively introgressed candidate region with known craniofacial effects in fish lies in the RNA-binding protein *rbms3*, a posttranscriptional regulator in the same SMAD-dependent transcription growth factor B pathway. Mutations in this gene cause cartilage and neural crest related abnormalities in zebrafish (Jayasena and Bronner 2012).

This region contains a non-coding SNP fixed in the San Salvador Island scale-eaters that is homozygous in *C. laciniatus* and segregating in the generalist population, a signature of high absolute genetic divergence between the two specialists, and a selective sweep in the scale-eater (Figure 7). Several lines of evidence point towards the introgression of *rbms3* variants between scale-eater and *C. laciniatus*. First, genetic differentiation is minimal between scale-eaters and *C. laciniatus* (D_{xy} = 0.002) and higher in all other pairwise comparisons (D_{xy} > 0.0104) between the two specialists and two outgroup Caribbean pupfish species (S4 Table). Segments of this region were assigned to the combined scale-eater topology (Figure 2C) and a maximum likelihood tree of the segment containing the fixed SNP features *C. laciniatus* in a clade with scale-eaters (Figure S14). Similar to the pattern we find in *rmbs3*, another candidate region previously associated with oral jaw size variation on San Salvador Island spanning *pard3* contained fixed scale-eater variants shared with *C. laciniatus*, strong genetic similarity in the surrounding region between the two and signs of a selective sweep in the scale-eaters (S15 and S16 Figures, Table 1).

In an unannotated candidate adaptive introgression region which has previously been associated with oral jaw size variation on San Salvador Island, we find a slightly different pattern than those mentioned above. The direction of introgression appears to be between *C. laciniatus* and the molluscivores, but is under a selective sweep in the scale-eaters (S17 and S18 Figures, Table 1 and S3). We also see a similar pattern in *nbea*, where the direction of introgression appears to be between *C. laciniatus* and scale-eaters but is under a selective sweep in the molluscivores (S19 and S20 Figures, Table 1 and S3). *Nbea* encodes for a scaffolding protein involved in neurotransmitter release and synaptic functioning and has been identified as a candidate gene for non-syndromic autism disorder (Volders et al. 2011; Cullinane et al. 2013; Nuytens et al. 2013). In zebrafish, mutations disrupt electrical and chemical synapse formation and cause behavioral abnormalities such as decreased startle response (Miller et al. 2015). Introgression in this regions is of interest because behavior is another axis of divergence between specialists in this system alongside craniofacial traits, as the species vary in mate choice (Kodric-Brown and West 2014; West and Kodric-Brown 2015), aggression, and prey capture behavior (Martin and Wainwright 2013d). Both of these candidate regions feature nearly equivalent negative Tajima's D statistics and low nucleotide diversity in the both of the specialists. The regions do not appear to be under strong selection in the generalist populations on San Salvador Island, so the signatures of selective sweeps in both specialists most likely stem from parallel molecular evolution in these regions rather than purifying selection in the ancestral population. Seven of 11 candidate regions show this pattern of equivalent low diversity and negative Tajima's D statistics in both specialists (Table 1 and S3).

The other 6 adaptive introgression candidates contained genes with a variety of functions including angiogenesis, calcium ion binding, embryonic eye morphogenesis, and RNA binding (Table 1) and had similar patterns to those mentioned above. Four of these regions feature low genetic diversity in both specialists. Two of these candidates lie in consecutive regions of the gene *srbd1*, which encodes for an RNA binding protein, and it appears that one has introgressed between the molluscivores and *C. laciniatus* and the other between scale-eaters and *C. laciniatus*. Both of these regions appear to be under a selective sweep in both of the specialists (Table 1 and S3).

Overall, potential adaptive variants contributing to species divergence among the specialists appear to be coming from New Providence Island in the northern Caribbean, rather

than the southern Caribbean (Table 1 and S3). Since it is impossible to infer the directionally of gene flow directly from f_4 values, we used TREEMIX (Pickrell and Pritchard 2012) to visualize gene flow in adaptively introgressed regions. Across the candidate adaptive introgression regions, we found evidence of an admixture event directly from *C. laciniatus* into the molluscivores in *ski* and *ltbp2* and *C. laciniatus* into scale-eaters in *srbd1* (Figure 8 and S5 Table). This suggests that genetic variation found on New Providence Island introgressed into the San Salvador Island radiation. There is no direct evidence from the TREEMIX population graphs of admixture from *C. bondi* into a specialist in the candidate regions (S5 Table), and D_{xy} between *C. bondi* and the specialists in pairwise comparisons is greater than those found between *C. laciniatus* and specialists across these regions (S3 Table). Both lines of evidence suggest that the high f_4 values in these regions stem from gene flow between *C. laciniatus* and the specialists rather than *C. bondi*.

1.4. Discussion

1.4.1. Diverse sources of genetic variation contributed to a highly localized adaptive radiation

Our investigation of genetic variation reveals that multiple sources of genetic variation were important for the assembly of the complex phenotypes associated with the novel ecological transitions seen only on San Salvador Island, Bahamas. While species divergence appears to mostly come from selective sweeps of variation from San Salvador Island (Figure 4), rare adaptive introgression has also played a role in the radiation (Table 1; Figures 5,7,S15,S17, and S19). The adaptive introgression we found in this study has come from large admixture events into San Salvador Island from a generalist pupfish population on another Bahamian island approximately 300 km away. In contrast, we found no evidence of introgression from a generalist population 700 km away in the Dominican Republic in our top candidate regions (Table 1 and S5), although it is impossible to rule out that candidate adaptive variants may also exist in this population at lower frequencies. Importantly, our limited sampling of one individual from each of two distant islands suggests that long-distance adaptive introgression is common and arises from abundant genetic variation found in only some parts of the Caribbean. An intriguing implication of these findings is that adaptive variants within the San Salvador Island radiation may have been partly assembled from the overlap of different pools of standing variation distributed across different parts of the Caribbean.

We found introgressed variants in four genes associated with the primary axis of jaw size variation within the radiation, as well as one in a gene with known behavioral effects in zebrafish. Both specialists appear to have candidate introgressed adaptive variants implicated in jaw morphology. Our best candidate for molluscivores was a region containing three fixed variants previously associated with jaw size variation on San Salvador Island in the proto-oncogene *ski*, which introgressed from *C. laciniatus*, another pupfish species on an island 300 km away (Figures 5,6, and 8A, Table 1). The best candidate for scale-eaters was a region containing a single fixed variant in the gene *rbms3* (Figure 7, Table 1), which is also present in *C. laciniatus*. Other candidate regions contained genes with functions in behavior, angiogenesis, calcium ion binding, embryonic eye morphogenesis, and RNA binding (Table 1).

We rarely know the source of candidate variants involved in diversification or the contributions of multiple sources of genetic variation to rapid diversification. Genomic

investigations of other adaptive radiations have also inferred roles for multiple genetic sources contributing to rapid diversification. For example, in the apple maggot fly, ancient gene flow from Mexican populations introduced an inversion affecting key diapause traits that aided the sympatric host shift to apples in the United States (Feder et al. 2003a). Hybridization within Darwin's finches also appears to play a role in the origin of new lineages through adaptive introgression of functional loci contributing to beak shape differences between species (Lamichhaney et al. 2015). In a *Mimulus* species complex, introgression of a locus affecting flower color appears to have been a driver of adaptation in the early stages of their diversification (Stankowski and Streisfeld 2015). However, even in case studies demonstrating multiple sources of genetic variation, the relative contributions to the diverse ecological traits in these radiations still remain unknown in most cases (but see (Pease et al. 2016)).

1.4.2. The genomic landscape of introgression differs between sympatric trophic specialists

Only 10% of all introgressed regions in either the molluscivore or scale-eater were shared between the two. This minimal overlap may reflect the complexity of different performance demands. Performance in the two specialists involves very different sets of functional traits (i.e. higher mechanical advantage and a novel nasal protrusion in the molluscivores vs. enlarged oral jaws and adductor muscles in the scale-eaters (Martin and Wainwright 2013d)) and divergent selective regimes (narrow and shallow vs. wide and deep fitness valleys (Martin and Wainwright 2013c; Martin 2016b; Martin et al. 2017)). The extensive variability in the genetic variation that introgressed between the two specialists may reflect multidimensional adaptation to two distinct trophic niches in this radiation, rather than variation along a linear axis (e.g. see (Harmon et al. 2005; Nosil and Harmon 2009; Nosil et al. 2009; Doebeli and Ispolatov 2010; Nosil and Hohenlohe 2012; Ispolatov and Doebeli 2013)).

1.4.3. Did introgression trigger adaptive radiation?

Although introgression is rare and localized across the genome, it was likely important for the assembly of the complex phenotypes observed on San Salvador Island (e.g. *ski and rbms3*). Our findings suggest two alternative possibilities. One intriguing possibility is that rare introgression of the necessary adaptive alleles into San Salvador Island may have been required to trigger the radiation in the presence of ecological opportunity. Indeed, a paradox in this system is why generalist populations in hypersaline lakes on neighboring islands with similar levels of ecological opportunity, lake areas, and overall genetic diversity have not radiated (Martin 2016b). Alternatively, adaptive radiation on San Salvador Island may have initiated from standing and *de novo* variation and only later benefited from introgressed alleles to further refine species phenotypes. Of course, these scenarios are not mutually exclusive and may vary across loci. Based on our TREEMIX analysis, introgression from *C. laciniatus* into the molluscivores brought the *ski* variants (Figure 8A), but the candidate adaptive variants in this region are also segregating in the generalist population (Figure 6).

We can roughly estimate the timing of introgression for this *ski* region from the number of variants that have accumulated between the *C. laciniatus* and molluscivore haplotypes ($n = 62$ differences; Figure 6). Assuming neutrality, the observed genetic differences between the two lineages should equal $2\mu t$, the time since their divergence in each lineage and μ, the mutation

rate (Masatoshi 1972). Using mutation rate estimates ranging from $5.37x10^{-7}$ (phylogeny-based estimate of *Cyprinodon* substitution rate (Martin et al. 2016)) to $1.32x10^{-7}$ mutations $site^{-1}$ $year^{-1}$ (estimated from a cichlid pedigree estimate of the per generation mutation rate (Recknagel et al. 2013) using a pupfish generation time of 6 months), introgression of the *ski* adaptive haplotype from *C. laciniatus* into the molluscivore specialist occurred between 5,700 to 23,500 years ago. The 10,000 year age estimate of the San Salvador Island radiation (based on estimates of dry lakes on the island (Hagey and Mylroie 1995; Holtmeier 2000; Turner et al. 2008)) falls within this window. This suggests the intriguing scenario in which widespread introgression during the last glacial maximum may have triggered adaptive radiation within pupfish populations isolated in the saline lakes of San Salvador Island during their initial formation. The 10-fold larger land mass of the Great Bahama Bank during this time could have created the opportunity for larger pupfish populations and greater genetic diversity. These pupfish populations would have been connected more extensively across the region than currently by the increased expanses of coastline habitats on the exposed bank. However, these are only exploratory inferences of the directionality of gene flow and timing of introgression. They should be confirmed with demographic analyses focused on testing different scenarios of admixture into San Salvador Island (e.g. (Pinho and Hey 2010; Filatov et al. 2016; Kautt et al. 2016b,a; Martin et al. 2016; Rougemont et al. 2016)).

While there are rare and convincing examples of hybridization leading to homoploid speciation (reviewed in (Schumer et al. 2014b)), no study, including ours, has yet provided convincing evidence that hybridization was directly involved in triggering an adaptive radiation. For example, while there is strong evidence in Darwin's finches that adaptive introgression of a loci controlling beak shape has contributed to phenotypic diversity of finches in the Galapagos, this hybridization occurred between members within the radiation (Lamichhaney et al. 2015). Similarly, a recent study argued that hybridization between ancestral lineages of the Lake Victoria superflock cichlid radiations and distant riverine cichlid lineages fueled the radiations, based on evidence of equal admixture proportions across the genomes of the Victorian radiations from the riverine lineages and the presence of allelic variation in opsins in the riverine lineages which are also important in the Victoria radiation (Meier et al. 2017a). However, the timing of introgression and necessity of introgressed alleles for initiating adaptive radiations remains unclear in these systems, including our own. Admittedly, hybridization as the necessary and sufficient trigger of adaptive radiation is a difficult prediction to test.

Those examples with more direct evidence linking hybridization to adaptation and reproductive isolation within a radiation are often special cases where a single introgressed adaptive allele automatically results in increased reproductive isolation. Examples include introgressed adaptive loci controlling wing patterns in *Heliconius* butterflies involved in mimicry and mate selection (Jiggins et al. 2001; The Heliconius Genome Consortium et al. 2012), a locus controlling copper tolerance in *Mimulus* that is tightly associated with one causing hybrid lethality (Wright et al. 2013), and loci contributing to differing insecticide resistance in the M/S mosquito mating types (Lee et al. 2013b; Marsden et al. 2014; Norris et al. 2015). While these cases provide convincing evidence that adaptive introgression can facilitate both ecological divergence and reproductive isolation, it is still unclear whether this introgression has actually triggered or simply contributed to the ongoing process of adaptive radiation.

Truly addressing the question of whether adaptive introgression triggered the radiation on San Salvador Island will require a better understanding of the timing of introgression and the necessity of introgressed variation for the speciation process. Although we have candidate alleles

(e.g. in *ski* and *rbms3*) that we think play a role in the evolution of complex specialist phenotypes, it still remains unclear what minimal set of alleles is necessary for the major ecological transitions in this system. Knowledge of the age of variants important for these transitions, and whether these variants are present and adaptive in the other non-radiating lineages of Caribbean generalist populations is needed. Estimation of the age of introgressed variation relative to standing or *de novo* could also shed light on whether adaptive introgression simply contributed to an ongoing diversification process or triggered it on San Salvador Island.

1.4.4. A new small-jawed scale-eating species within the radiation?

We also found evidence of a distinct clade of small-jawed scale-eaters, separate from the large-jawed scale-eaters (Figures 1 and 2). The consistent clustering of this clade across the genome suggests that they may be a distinct, partially reproductively isolated population on San Salvador Island, rather than a product of hybridization between generalists and scale-eaters in the lakes where they exist sympatrically (Figures 1 and 2; S1 and S2 Figures). They have only been observed in six lakes connected to the Great Lake System on San Salvador Island (Great Lake, Mermaid's Pond, Osprey Pond, Oyster Pond, Little Lake, and Stout's Lake), but not in isolated lakes such as Crescent Pond. Consistent with this pattern of occurrence, F_2 hybrid phenotypes resembling the scale-eaters have previously been shown to have extremely low survival and growth rates in these isolated lakes (Martin and Wainwright 2013c).

Small-jawed scale-eaters may represent a viable intermediate ecotype on the evolutionary path toward more specialized scale-eating. Small-jawed scale-eater diets appear to be consistent with intermediate levels of scale-eating. Preliminary gut content analyses revealed that scales were found in the stomachs of 33% of small-jawed scale-eaters (n = 33) compared to 91% of large-jawed scale-eaters (n = 53). The idea that specialization can open the door to further specialization has been seen in other systems, including pollinator syndromes for bees, hummingbirds, and hawkmoths in *Mimulus* (Fenster et al. 2004; Whittall and Hodges 2007; Thomson and Wilson 2008), Darwin's ground finch specializing on blood on two islands in their range (Schluter and Grant 1984), and transitions in mammals between omnivory, carnivory, and herbivory (Price et al. 2012). If small-jawed scale-eaters represent an ecotype stepping stone on the path toward more specialized scale-eating, we might expect regions of the genome to reflect a nested relationship between the large-jawed and small-jawed scale-eaters. We see this predicted pattern in the combined scale-eater topology that underlies most of the fixed variants between the two scale-eating species (Figure 2).

If the small-jawed scale-eaters were instead the result of recent or recurrent hybridization events, we would expect certain patterns of large-jawed scale-eater and generalist ancestry across their genomes. For example, if they represent F_1 hybrids, they should have equal ancestry from the two parental species across their genomes. The lack of fit in all five small-jawed scale-eater individuals to the ancestry proportions excepted if they represent F_1 hybrids, F_2 hybrids, or backcrosses to parental species (S2 Table) suggests that the small-jawed scale-eaters are not the result of such recent hybridization events, although they might have resulted from more complicated scenarios of hybridization that do not follow these simple patterns of ancestry(Fitzpatrick 2012; Gompert and Buerkle 2016). LD does appear to be stronger in the small-jawed scaler-eaters than in the three San Salvador Island species (Figure S3), a pattern expected in recent hybrids of distinct populations. These small-jawed scale-eaters may indeed be

the products of ongoing or recent gene flow on San Salvador Island. A reconstruction of the history of gene flow among San Salvador Island species from demographic modeling with a larger sample, along with estimates of selection and reproductive isolation in the small-jawed scale-eaters, will be needed to assess whether they represent the products of ongoing gene flow on San Salvador Island or a potential new ecomorph.

1.4.5. Conclusion

Here we demonstrate that the complex phenotypes associated with the novel ecological transitions within a nascent adaptive radiation of San Salvador Island pupfishes arose from multiple sources of genetic variation spread across the Caribbean. The variation important to this radiation is localized to small regions across the genome that are obscured by genome-wide summaries of the history of the radiation. Species divergence appears to mostly come from selective sweeps of standing or *de novo* genetic variation on San Salvador Island, but rare adaptive introgression events may also be necessary for the evolution of trophic specialists. This genomic landscape of introgression is variable between the specialists and has come from large admixture events from populations as far as 742 km across the Caribbean, although all top adaptive introgression candidates appear to have introgressed from a population 300 km away in the northwestern Bahamas. Our findings that multiple sources of genetic variation contribute to the San Salvador Island radiation suggests a complex suite of factors, including rare adaptive introgression, may be required to trigger adaptive radiation in the presence of ecological opportunity.

1.5. Methods

1.5.1. Study system and sampling

Individual pupfish were caught in hypersaline lakes on San Salvador Island in the Bahamas with either a hand or seine net in 2011, 2013, and 2015. Samples were collected from eight isolated lakes on this island (Crescent Pond, Great Lake, Little Lake, Mermaid Pond, Moon Rock Pond, Oyster Lake, Osprey Lake, and Stout's Lake) and one estuary (Pigeon Creek). 13 *Cyprinodon variegatus* were sampled from all eight lakes on San Salvador Island; 10 *C. brontotheroides* were sampled from four lakes; and 14 *C. desquamator* were sampled from six lakes. The specialist species occur in sympatry with the generalists in only some of the lakes. Individual pupfish that were collected from other localities outside of San Salvador Island served as outgroups to the San Salvador Island radiation, including *C. laciniatus* from Lake Cunningham, New Providence Island in the Bahamas, *C. bondi* from Etang Saumatre lake in the Dominican Republic*, C. diabolis* from Devil's Hole in California (collected as a dead specimen by National Park Staff in 2012), as well as captive-bred individuals of extinct-in-the-wild species *C. simus* and *C. maya* originating from Laguna Chichancanab, Quintana Roo, Mexico. Fish were euthanized by an overdose of buffered MS-222 (Finquel, Inc.) following approved protocols from University of California, Davis Institutional Animal Care and Use Committee (#17455) and University of California, Berkeley Animal Care and Use Committee (AUP-2015-01-7053) and

stored in 95-100% ethanol. Only degraded tissue was available for *C. diabolis*, as described in (Martin et al. 2016).

1.5.2. Genomic sequencing and bioinformatics

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

2.8 billion raw reads were mapped from 42 individuals to the *Cyprinodon* reference genome (NCBI, *C. variegatus* Annotation Release 100, total sequence length = 1,035,184,475; number of scaffold = 9,259, scaffold N50, = 835,301; contig N50 = 20,803) with the Burrows-Wheeler Alignment Tool (Li and Durbin 2011) (v 0.7.12). 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 (DePristo et al. 2011)(v 3.5) to call and refine our 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) (DePristo et al. 2011; Marsden et al. 2014) because we lacked high-quality known variants for these non-model species. Our final dataset after filtering contained 16 million variants and a mean sequencing coverage of 7.2X per individual (range: 5.2–9.3X).

1.5.3. Characterization of genomic heterogeneity in evolutionary relationships among individuals

We used the machine learning program SAGUARO (Zamani et al. 2013) to identify regions of the genome that contain different signals about the evolutionary relationships across San Salvador Island and outgroup *Cyprinodon* species. Saguaro combines a hidden Markov model with a self-organizing map (SOM) to characterize local phylogenetic relationships among individuals without requiring *a priori* hypotheses about the relationships. When diploid data is used, the SOM selects one allele at random for training. This method infers local relationships among individuals in the form of genetic distance matrices and assigns segments across the genomes to these topologies. These genetic distance matrices can then be transformed into neighborhood joining trees to visualize patterns of evolutionary relatedness across the genome. Three independent runs of SAGUARO were started using the program's default settings and each was allowed to assign 15 different topologies across the genome. To determine how many topologies to estimate, analogous to a scree plot (Cattell 1966; Joliffe 2002), we plotted the proportion of the genome explained by each hypothesized topology and looked for an inflection point (Figure S21). We also looked at the neighborhood joining trees to assess whether additional topologies were informative about the evolutionary relationships among individuals (Figure S21). The 15th topology and additional topologies that we investigated tended to be uninformative about the evolutionary relationships among individuals and represented less than 0.5% of the genome. We excluded the last topology (15th) from downstream analyses due to lack

of genetic distinction at both the level of populations and species included in the proposed genetic distance matrix and the low percentage of the genome assigned to it. The 14 topologies included in downstream analyses and the total percentages of the genome assigned to them were robust across all three independent runs. These topologies also appeared to be fairly robust to the influences of poorly mapped regions in the genome. We generated a mask file to identify poorly mapped regions in our dataset using the program SNPable (http://bit.ly/snpable; k-mer length =50, and 'stringency'=0.5) and removed these segments from downstream analyses of the topologies. Rerunning the SAGUARO analysis on the masked dataset resulted in very similar trees across the 14 different topologies, with the exception of several generalist individuals grouping with molluscivores in the molluscivore topology (Figure S22).

1.5.4. Comparison of linkage disequilibrium among San Salvador Island species

We calculated LD within each of the San Salvador Island species and compared it to estimates for the small-jawed scale-eaters to look for patterns of high linkage consistent with recent hybridization events. Pairwise LD across the largest scaffold in our dataset (4.2 Mb) was calculated for each species using the 'r2 inter-chr' function in PLINK v1.90 (Purcell et al. 2007) for five individuals. These were chosen from a pool of individuals from Great Lake system populations (average genome-wide $F_{st} < 0.05$ across these lakes for each of the species) to balance the effects of small sample sizes and population structure on estimates of LD and more accurately compare LD decay between species. LD may be overestimated for each of the species due to the small number of individuals available to calculate it from in this study, and should be compared to estimates from other studies with caution.

1.5.5. Characterization of introgression patterns across the genome

We characterized the heterogeneity in introgression across the genome using f_4 statistics that were initially developed to test for introgression among human populations (Reich et al. 2009; Patterson et al. 2012; Pickrell and Pritchard 2012). The f_4 statistic tests if branches among a four-taxon tree lack residual genotypic covariance (as expected in the presence of incomplete lineage sorting and no introgression) by comparing allele frequencies among the three possible unrooted trees. A previous study (Martin 2016b) provided evidence of potential admixture with the Caribbean outgroup species used in this study, preventing their use in a D-statistic framework which requires designation of an outgroup with no potential introgression.

To look for evidence of gene flow across the Caribbean, we focused on tests of introgression with the two outgroup clades from our sample that came from other Caribbean islands in the Bahamas and Dominican Republic. Based on the tree ((P1, P2),(*C. laciniatus*, *C. bondi*)), f_4 statistics were calculated for all three possible combinations of P1,P2 among the pooled populations of generalists, scale-eaters, and molluscivores on San Salvador Island. These f_4 statistics were calculated using the population allele frequencies of biallelic SNPs and summarized over windows of 10 kb with a minimum of 50 variant sites using a custom python script (modified from ABBABABA.py created by Simon H. Martin, available on https://github.com/simonhmartin/genomics_general; (Martin et al. 2015b); our modified version is provided in the supplemental material), allowing for up to 10% missing data within a

population per site. All 10 molluscivore and 14 scale-eater individuals from San Salvador Island were used in the tests for comparison to the molluscivore and combined scale-eater topologies, respectively. In another calculation of f_4 statistics across the genome, the 5 small-jawed scale-eater individuals were excluded for the comparison to the large-jawed scale-eater topology. Although only single individuals from New Providence Island, Bahamas and the Dominican Republic were used to represent *C. laciniatus* and *C. bondi* in the f_4 tests, the individuals that were sequenced are a random sample from these populations and should be representative. This resulted in 100,276 f_4 statistics (mean $f_4 = -2x10^{-4}$) calculated across the genome for the test that included all scale-eaters and 100,097 f_4 statistics (mean $f_4 = -9x10^{-5}$) for the test excluding the small-jawed scale-eaters.

We conducted 1,000 permutations of the f_4 test to evaluate the significance of f_4 values in sliding windows across the genome. For each permutation, individuals from the four original populations were randomly assigned without replacement to one of the four populations based on the tree ((P1,P2),(P3,P4)) to assess how likely a given f_4 value would be observed by chance within our empirical dataset. We calculated the 1% tails of this null distribution and used these thresholds for our candidate introgressed regions (i.e. significant at alpha = 0.02). The null distribution illustrating the 1st and 99th quantiles for all combinations of the sliding window f_4 test are provided in the supplementary material (Figure S23). Each candidate introgressed region was assigned a *P*-value by counting the number of permutations that had an f_4 value greater than (or lesser than if the f_4 value was negative) or equal to the observed value.

It is difficult to distinguish between genetic variation that is similar among taxa due to introgression from a hybridization event and that from ancestral population structure, so some of the regions with significant f_4 values may represent the biased assortment of genetic variation into modern lineages from a structured ancestral population [51]. A recent simulation study (Martin et al. 2015b) found that extending the use of genome-wide introgression statistics such as Patterson's D statistic to small genomic regions can result in a bias of detecting statistical outliers mostly in genomic regions of reduced diversity. Although it hasn't been formally tested, f_4 statistics may be subject to the same biases, so we additionally considered the nucleotide diversity present in outlier f_4 regions in downstream analyses by comparing π across the detected regions of introgression in comparison to scaffold- and genome-wide estimates among the three San Salvador Island species.

1.5.6. Comparison of patterns of introgression to patterns of genetic divergence and diversity

We then calculated several population genetic summary statistics in sliding windows across the genome to compare to the f_4 patterns of introgression: F_{st}, between-population nucleotide divergence (D_{xy}), within-population nucleotide diversity (π) for pairwise species comparisons, and Tajima's D estimates of selection in each species. D_{xy} between molluscivores and scale-eaters was calculated over the same 10-kb windows as the f_4 tests using the python script popGenWindows.py created by Simon Martin (available on https://github.com/simonhmartin/genomics_general; (Martin et al. 2015b)). Since our vcf file contained only variant sites and this script does not factor the missing sites into the calculation of D_{xy} by assuming they are invariant, we post-hoc incorporated the missing sites as invariant sites in the calculation of D_{xy}. Missing sites in our dataset may include poorly aligned regions with

lots of variants, so by assuming the missing sites are all invariant, D_{xy} may be underestimated in this study and should be compared to diversity values from other organisms with caution.

The remaining statistics were calculated in non-overlapping sliding windows of 10 kb using 'wier-fst-pop', 'window-pi', and 'TajimaD' functions in VCFtools v.0.1.14 (Danecek et al. 2011a). Negative values of Tajima's D indicate a reduction in nucleotide variation across segregating sites (Tajima 1989), which may result from hard selective sweeps due to positive selection. To determine regions of the genome potentially under positive selection, we created a null distribution of Tajima's D values expected for each of three species under neutral coalescent theory using ms-move (Garrigan and Geneva 2014), a program that adds more flexibility in incorporating introgression events into the coalescent simulator ms (Hudson 2002). Based on the demographic history estimated for the three San Salvador Island species in a previous study (McGirr and Martin 2016a), we incorporated a 100-fold decrease in population size approximately 10,000 years ago (-eN 0.8 0.01) and an introgression event from one population into another to mimic introgression between a San Salvador Island species and an outgroup population at the beginning of the radiation (ex. -ej 0.8 2 1 –ev 0.8 2 1 0.1). We estimated the null distribution of Tajima's D for 100,000 loci for 10-14 individuals with a variable number of segregating sites (ranging from 50 to the maximum observed in a 10-kb window of the genomes of each species). We modeled the timing of introgression from approximately 6,000-23,000 years (based on the rough estimate of the timing of introgression of *ski* in this study) with 10% of population composed of migrants (although the distribution appeared robust to variations in this fraction). Tajima's D values were calculated from the simulated loci using the 'sample stats' feature available in the ms package (Garrigan and Geneva 2014). The simulated introgression event and bottleneck skewed the null distribution towards negative Tajima's D values (Figure S24). Windows from the observed genomes that had Tajima's D values in the lower 2% tail of the null distribution were considered candidate regions for selective sweeps.

We also estimated regions under selective sweeps from the expected neutral folded site frequency spectrum calculated with SweeD (Pavlidis et al. 2013). In this calculation, we included the bottleneck of a 100-fold decrease around 10,000 years ago and the recommended grid size of 1 kb across scaffolds to calculate the composite likelihood ratio (CLR) of a sweep. The values of CLR from 1 kb windows were averaged across 10-kb to compare with the other statistics calculated in windows. Windows with an average CLR estimate above the 98th percentile across the background site frequency spectrum for their respective scaffold were considered candidate regions under a selective sweep.

We also used the function 'wier-fst-pop' to calculate F_{st} across individual SNPs to locate SNPs fixed between species and identify whether candidate adaptive introgression regions potentially contributed to species divergence. We assessed mean coverage across individuals at SNPs fixed between specialists and found that they ranged from 4.8-8.2x. The SNPs fixed in this study may be an overestimate of the variants potentially contributing to diversification in the specialists, as alleles may be missing from our individuals at these sites due to the low coverage. Average coverage and standard deviation across SNPs fixed in candidate regions are reported in the supplementary material (S3 Table). Only regions of overlap between significant f_4 values, strongly negative Tajima's D values, 98th percentile CLR estimates, and fixed SNPs between the two specialists were considered candidate adaptive introgression regions that have contributed to species divergence. For each of these regions, we looked for annotated genes and searched their gene ontology in the phenotype database 'Phenoscape' (Mabee et al. 2012; Midford et al. 2013; Manda et al. 2015; Edmunds et al. 2016) and AmiGO2 (Balsa-Canto et al. 2016) for pertinent

functions, particularly skeletal system effects. Skeletal features, particularly craniofacial morphologies such as jaw length, have extremely high rates of diversification among the species on San Salvador Island (Martin and Wainwright 2011; Martin 2016b) and likely play a key role in the diversification of this group.

1.5.7. Estimation of the direction of gene flow in candidate adaptive introgression regions

While the sign of f_4 hints at the directionality of introgression (e.g. for the tree (P1,P2),(P3,P4), a positive f_4 value indicates gene flow either between P1 and P3 or P2 and P4), the lack of an explicit outgroup in the f_4 statistics makes it difficult to determine the exact direction of gene flow among the included populations and limits our ability to determine if candidate introgressed regions came from admixture with *C. laciniatus* or *C. bondi*. We examined each candidate region for signs of directionality using several methods.

To visualize gene flow among the Caribbean populations included in this study, we used TREEMIX v1.12 (Pickrell and Pritchard 2012) to estimate population graphs with 0-4 admixture events connecting populations. Population graphs were estimated for each region with a significant f_4 value, each with a minimum of 50 SNPs. The number of admixture events was estimated by comparing the rate of change in log likelihood of each additional event, an approach similar to one used in Evanno et al. ((Evanno et al. 2005); also see (Martin 2016b)). However, this analysis should be viewed only as an exploratory tool as the reliability of TREEMIX to detect the number of admixture events has not been tested. This method was designed to be applied on genome-wide allele frequencies and estimates covariance in allele frequencies among populations in branch lengths using a model that assumes allele frequency differences between populations are solely caused by genetic drift (Pickrell and Pritchard 2012). The use of fewer SNPs (≥50) in our window-based approach also makes it harder to reliably distinguish between the different likelihoods for the number of migration events. The reliability of inference under these conditions has not been evaluated, however the migration events inferred in our TREEMIX results were consistent with our findings from our formal f_4 test for gene flow.

We also compared pairwise nucleotide diversity between *C. bondi*, *C. laciniatus*, molluscivores, and scale-eaters to determine which pairs are most genetically similar in the candidate introgression regions. Since our genomic dataset only included single individuals from *C. bondi* and *C.laciniatus* and F_{st} estimates are a relative measure of divergence based on within population diversity, we used calculated D_{xy}, an absolute measure of genetic divergence between-populations. Finally, we generated maximum likelihood phylogenetic trees for the SAGUAROsegment containing the fixed SNPs under a GTR+GAMMA model of sequence evolution using RaxML v.8.2.10 (Stamatakis 2014). Support for nodes was assessed by bootstrapping, allowing the number of bootstraps determined by autoMRE function in RaxML, which ranged from 900-1,000 among regions.

Acknowledgements

We thank J. McGirr, J. Poelstra, and the UNC EEGAS journal club for helpful discussion of this manuscript; K. Gould for performing DNA extractions; J. McGirr for collecting stomach content data; M. Grabherr for discussion of filtering strategies; members of the American Killifish Association for supplying tissues, in particular A. Morales, M. Schneider, J. Cokendolpher, and

A. Kodric-Brown; R. Hanna, K. Guerrero, and L. Simons for assistance obtaining permits; the Gerace Research Centre for accommodation; the governments of the Bahamas, Dominican Republic, and the National Park Service and U.S. Fish and Wildlife Service for permission to collect pupfish samples; the Vincent J. Coates Genomics Sequencing Laboratory and Functional Genomics Laboratory at UC Berkeley, supported by NIH S10 OD018174 Instrumentation Grant, for performing whole-genome library preparation and sequencing; and the UNC ITS Research Computing Services for computational resources.

1.6. Figures

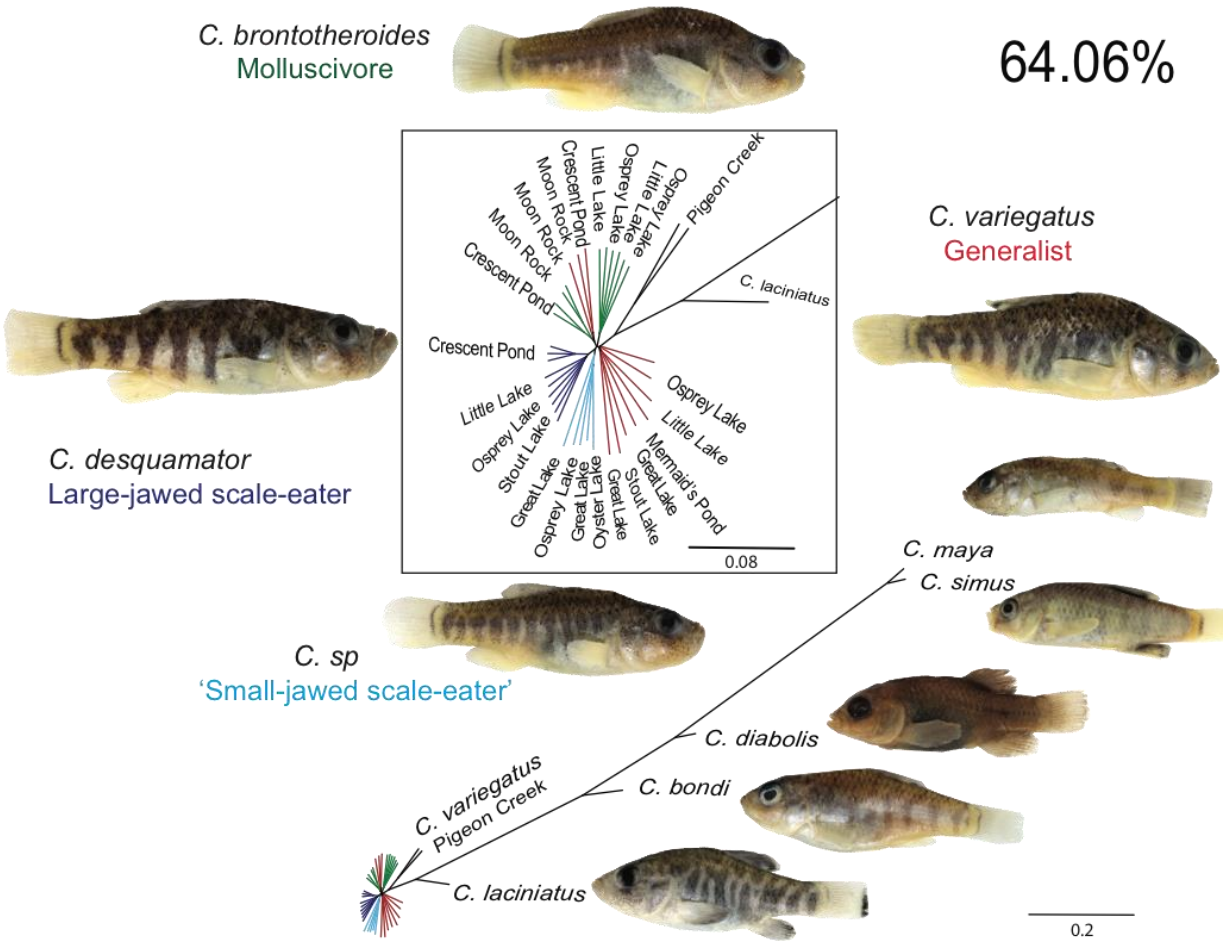


Figure 1. The most common topology showinga monophyletic San Salvador clade covering 64% of the genome. San Salvador generalists (red), molluscivores (green), large-jawed scale-eaters (dark blue), small-jawed scale-eaters (light blue), and outgroup species (black) in the Caribbean, California, and Mexico. Other topologies featuring a monophyletic San Salvador clade are presented in Figure S1.

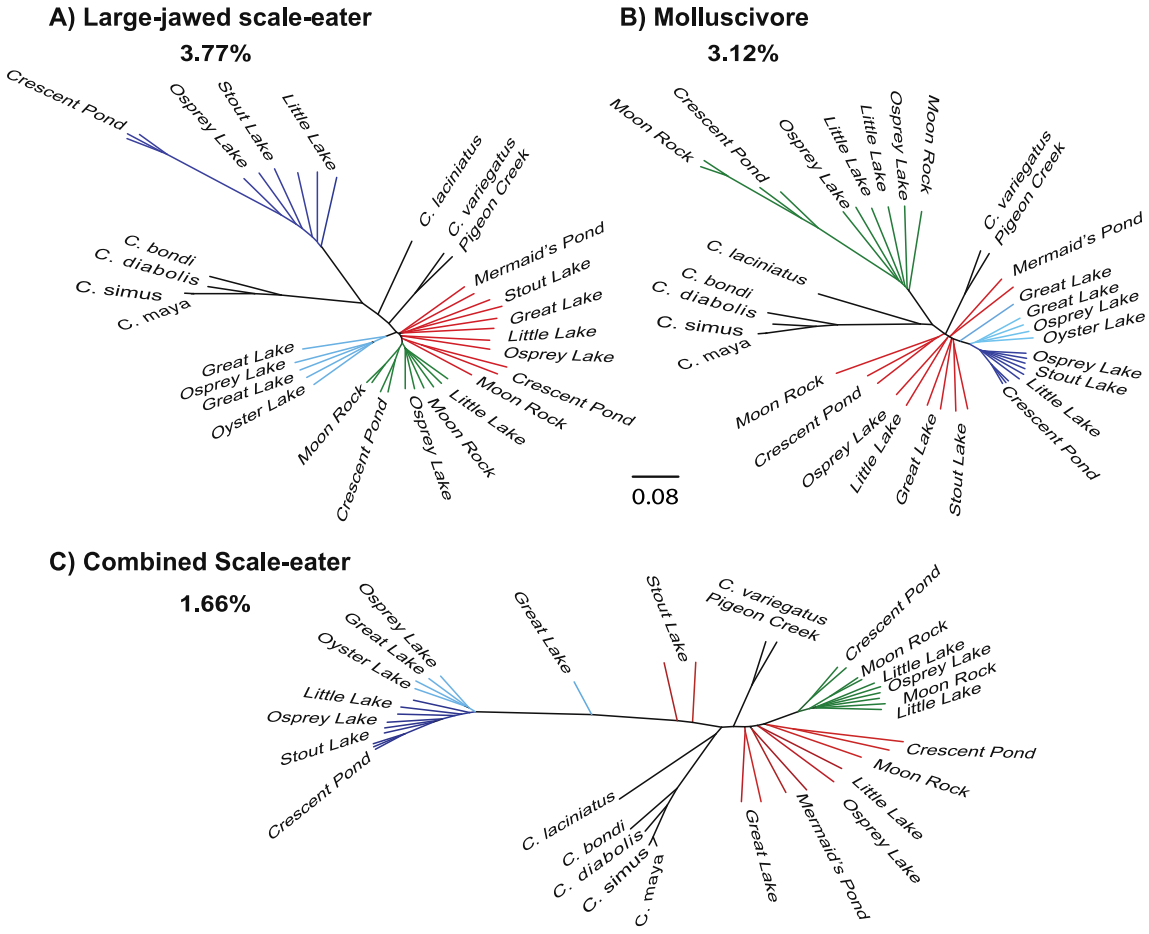


Figure 2. Alternative topologies estimated in SAGUARO showing San Salvador specialists grouping with outgroups. (A) The large-jawed scale-eater topology covered 3.77% of the genome, in which large-jawed scale-eater individuals showed a sister relationship to outgroup pupfish species. (B) The molluscivore topology covered a non-overlapping 3.12% of the genome, in which molluscivores showed a sister relationship to outgroup pupfish species. (C) The combined scale-eater topology covered a non-overlapping 1.66% of the genome, where all scale-eaters (along with two generalists from Stout's Lake) showed a sister relationship to the outgroup pupfish species. Additional alternative topologies are presented in Figure S2.

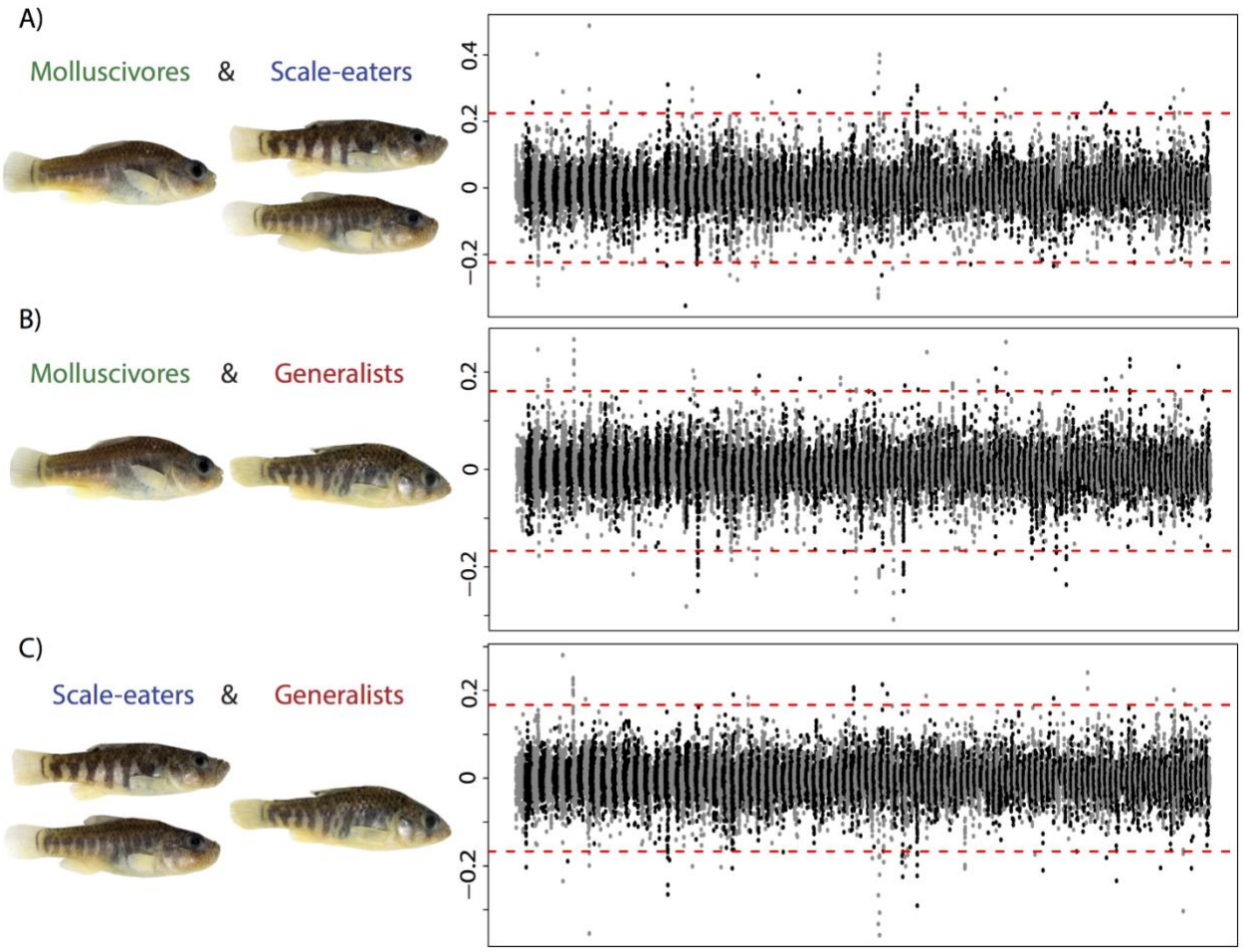


Figure 3. Variable introgression from distant Caribbean islands across the genomes of the San Salvador trophic specialists. Manhattan plot of the f_4 values between the *C. laciniatus* from New Providence Island, Bahamas, *C. bondi* from Etang Saumatre Dominican Republic and (A) molluscivores and scale-eaters on San Salvador, (B) molluscivores and generalists from San Salvador, (C) scale-eaters and generalists on San Salvador. Alternating gray/black colors indicate different scaffolds from the largest 170 scaffolds of the genome. Dotted red lines mark the permutation based significance threshold. Full Manhattan plots for each comparison are presented in Figures S3-S5.

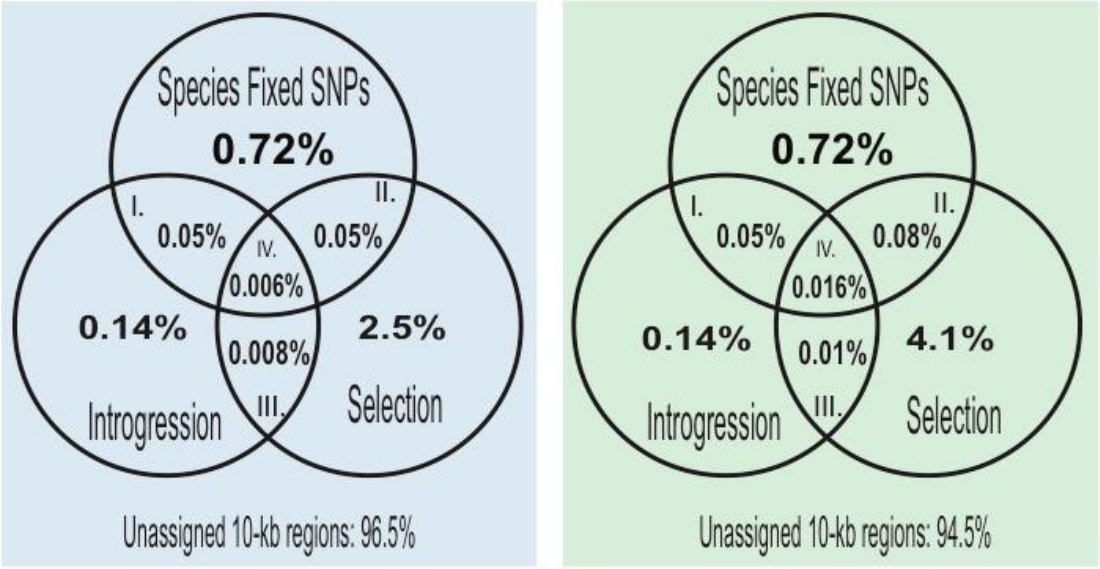


Figure 4. Contributions of selection and introgression to species divergence. Venn diagrams of the contribution of different sources of genetic variation to speciation in this system based on fixed SNPs between the 23molluscivores and combined scale-eaters, significant f_4 values of introgression, hard selective sweeps (the lower 2% of the simulated distribution of Tajima's D values) in (A) combined scale-eaters and (B) molluscivores. We calculated the percentage of I) Introgression: regions that contain introgressed genetic variation from the Caribbean contributing to species divergence but not under a selective sweep, II) Selective Sweeps: regions that have undergone strong selective sweeps from genetic variation on San Salvador not introgressed from our two outgroup populations, III) Adaptive introgression: adaptively introgressed regions not contributing to species divergence , and IV) Adaptive introgression involved in speciation: regions that have undergone selective sweeps of introgressed variation that contribute to species divergence of the two specialist species. The percentage of 10-kb windows not assigned to the above categories is provided below each Venn diagram. Overlap of these categories with the dominant and alternative topologies is provided in Figure S9.

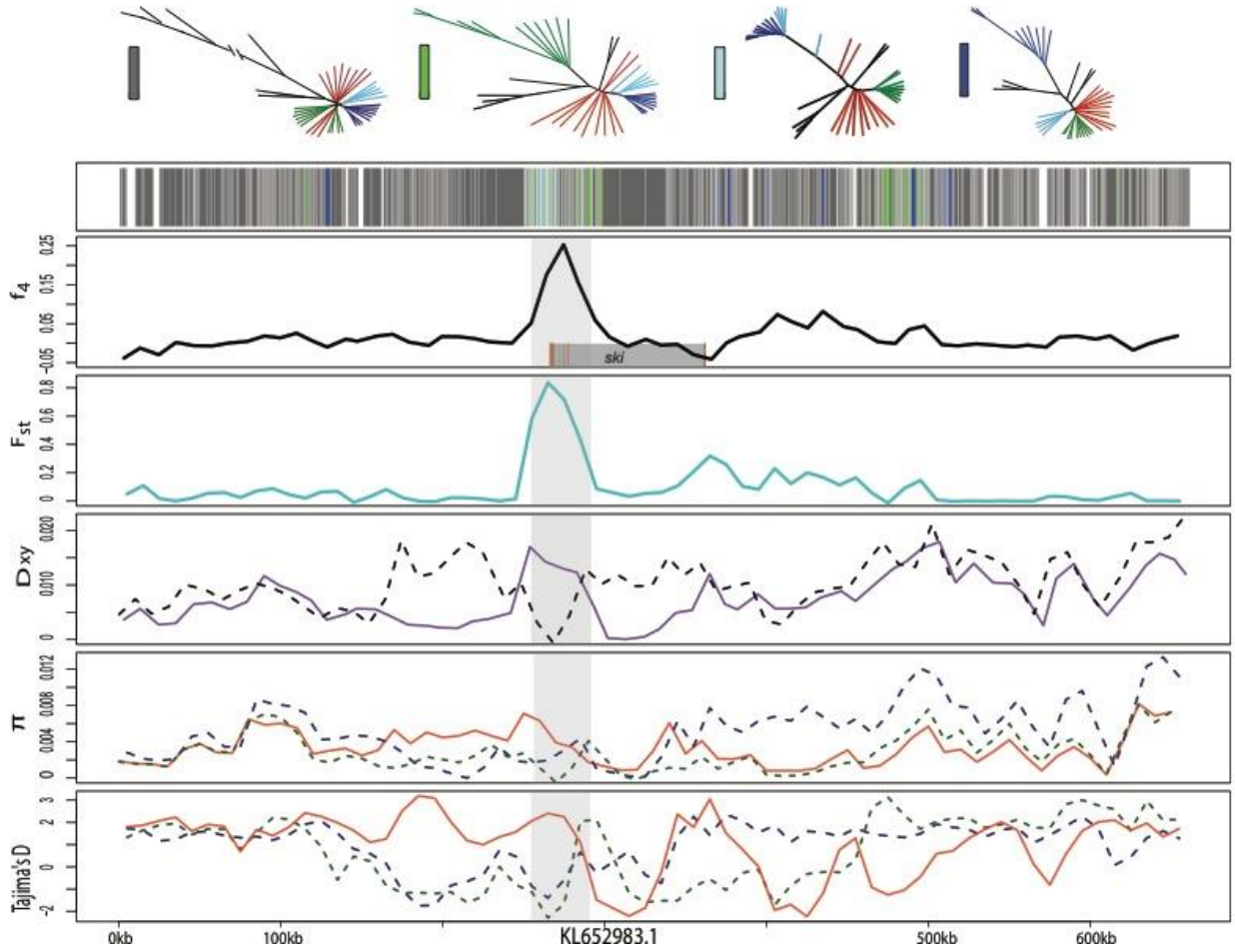


Figure 5. Candidate adaptive introgression region in craniofacial development gene *ski*. Fixed variants in this region were previously associated with pupfish oral jaw size [55]**.** Row 1 shows the history assigned by SAGUARO to segments along a 600-kb scaffold (dark grey: dominant topology; blue: large-jawed scale-eater topology; light blue: combined scale-eater topology**;** green: molluscivore topology; light grey: all other topologies; white: unassigned segments). Row 2 shows average f_4 value across non-overlapping 10-kb windows between mollsucivores/scale-eaters. Shaded grey box shows region annotated for *ski* gene with exons in red. Row 3 shows average F_{st} value across non-overlapping 10-kb windows between molluscivores/scale-eaters (turquoise). Row 4 shows between-population divergence (D_{xy}) across non-overlapping 10-kb windows between molluscivores/scale-eaters (purple) and molluscivores/*C. laciniatus* (grey-dashed). Row 5 shows within-population diversity (π) across non-overlapping 10-kb windows (blue-dashed: scale-eater; green: molluscivore). Row 6 shows Tajima's D across non-overlapping 10-kb windows (blue-dashed: scale-eater; green: molluscivore.

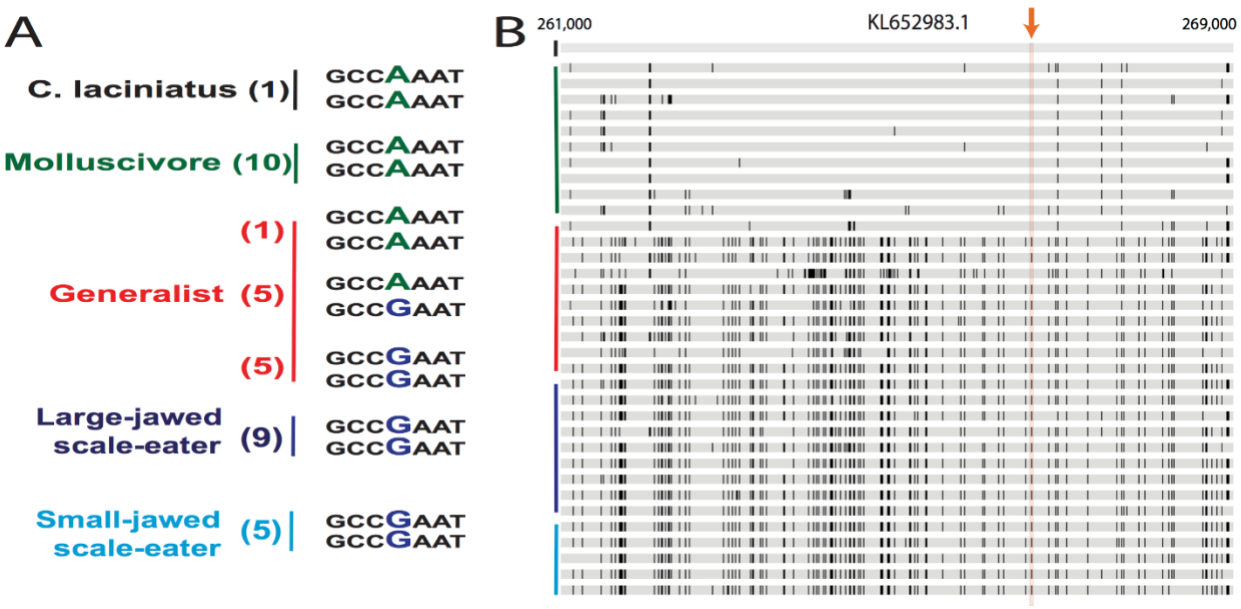


Figure 6. Genetic diversity surrounding the fixed variant in *ski* region assigned to the combined scale-eater topology. (A) The 3' untranslated region variant fixed between the two specialists. The number of individuals with the haplotype(s) are located in parentheses next to species names. The other two fixed SNPs follow the same pattern across species as the SNP shown. (B) A comparison of the San Salvador genotypes (green=molluscivore; red=generalists; blue=scale-eater) with the *C. laciniatus* genotype (black) across an 8-kb window surrounding the fixed variant (orange arrow). The alleles that do not match the alleles of *C. laciniatus* are highlighted with black bars. The arrow points to the conflicting genotypes in the surrounding 8-kb region of the SNPs.

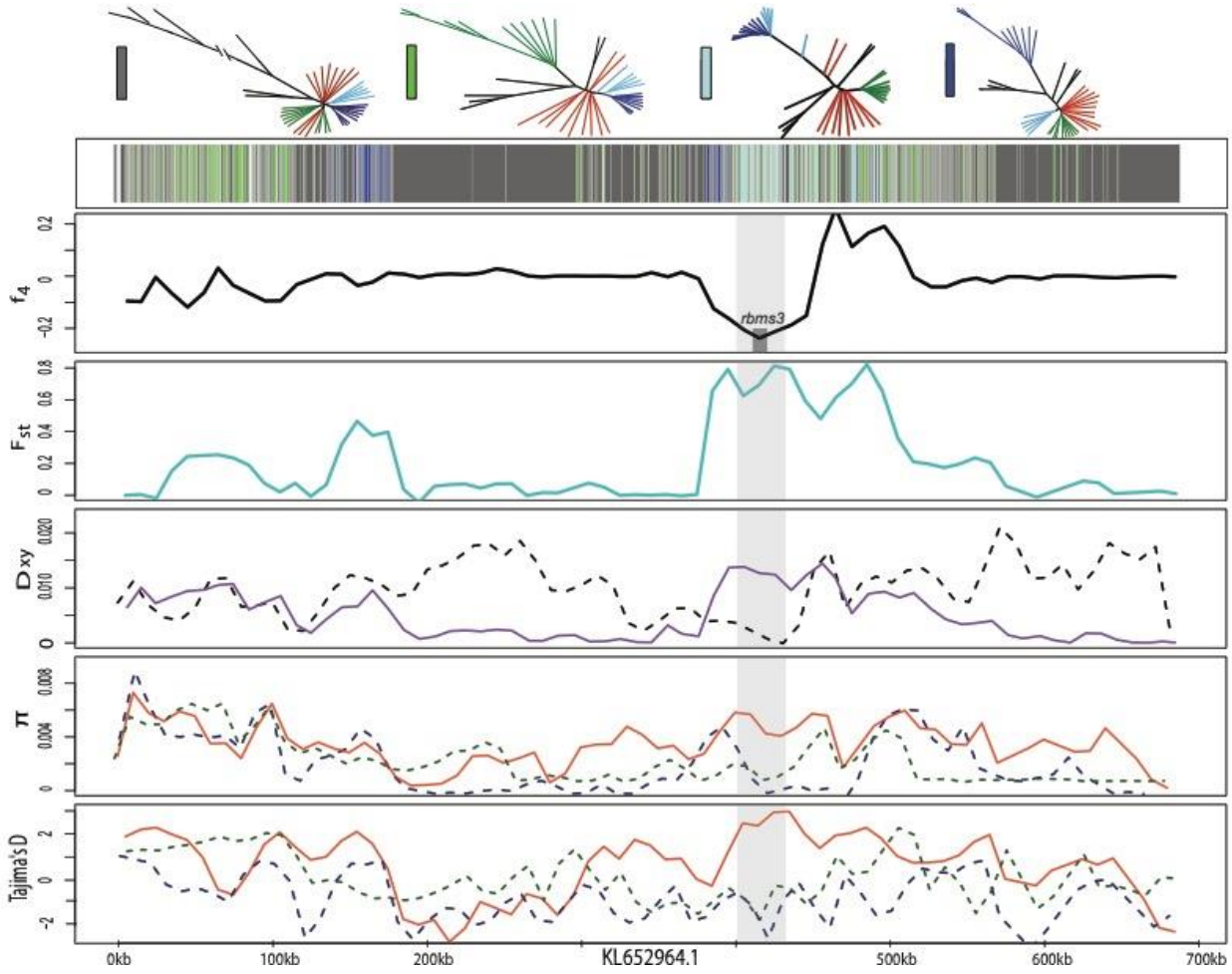


Figure 7. Candidate adaptive introgression region in craniofacial development gene *rbms3*. Row 1 shows the history assigned by SAGUARO to segments along a 700-kb scaffold (dark grey: dominant topology; blue: large-jawed scale-eater topology; light blue: combined scale-eater topology**;** green: molluscivore topology; light grey: all other topologies; white: unassigned segments). Row 2 shows average f_4 value across non-overlapping 10-kb windows between mollsucivores/scale-eaters. Shaded grey box shows region annotated for *rbms3* gene. Row 3 shows average F_{st} value across non-overlapping 10-kb windows between molluscivores/scale-eaters (turquoise). Row 4 shows between-population divergence (D_{xy}) across non-overlapping 10-kb windows between molluscivores/scale-eaters (purple) and scale-eaters/*C. laciniatus* (grey-dashed). Row 5 shows within-population diversity (π) across non-overlapping 10-kb windows (blue-dashed: scale-eater; green: molluscivore). Row 6 shows Tajima's D across non-overlapping 10-kb windows (blue-dashed: scale-eater; green: molluscivore).

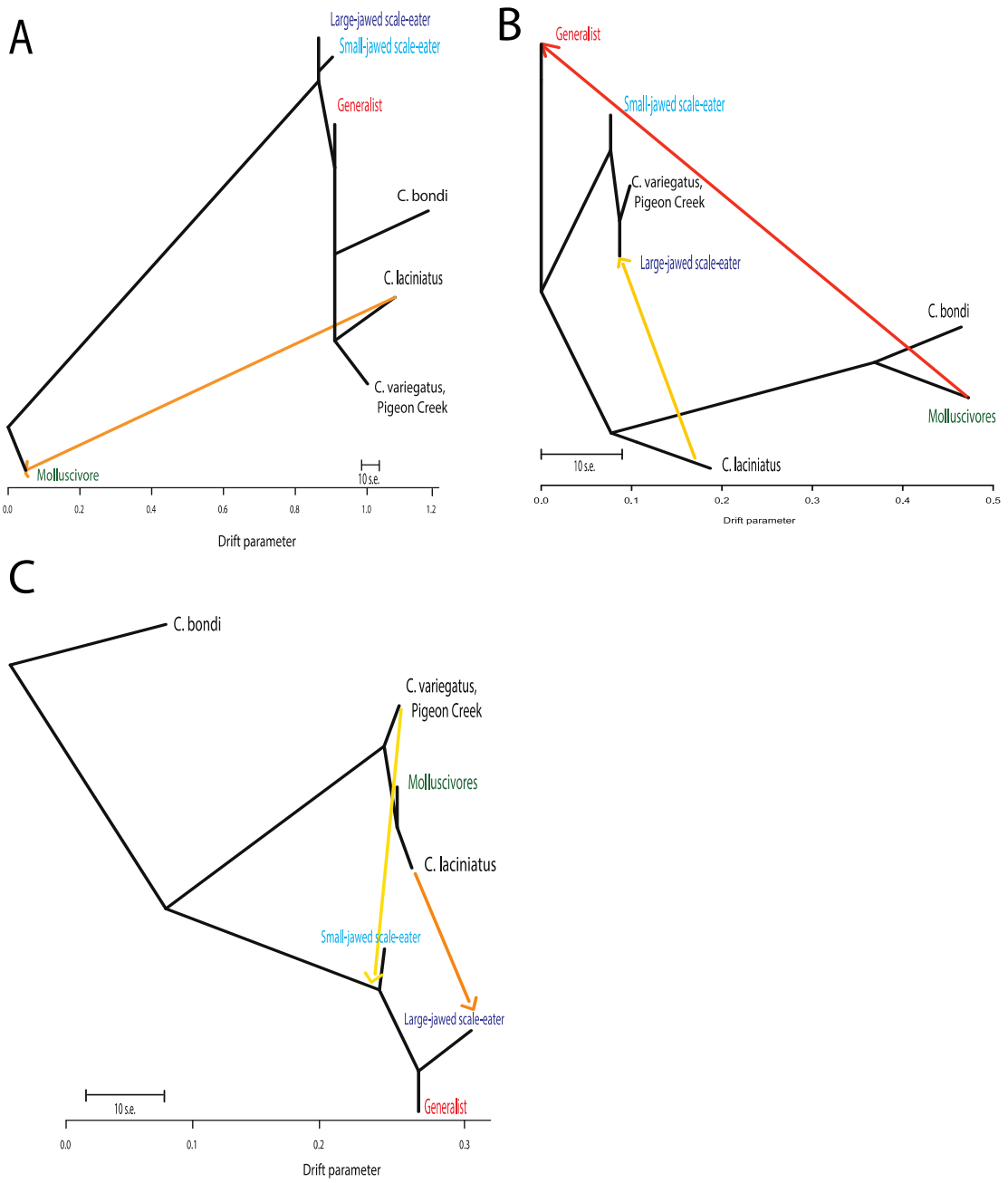


Figure 8. Introgression into the specialists on San Salvador from *C. laciniatus*. TREEMIX population graphs for three candidate adaptive introgression regions showing gene flow from *C. laciniatus* into the molluscivores in the A) *ski* region (change in composite log-likelihood with increase in number of migration events: m=0, LnL: -320; m=1, LnL: 81; m=2 , LnL: 93), (B) *ltpb2* (m=0, LnL: 11; m=1, LnL: 62; m=2, LnL: 137; m=3 , LnL: 62), and (C) showing gene flow from *C. laciniatus* into the molluscivores (m=0, LnL: -17 ; m=1 , LnL: -11; m=2 , LnL: 74; m=3 , LnL: 68).

1.7. Tables

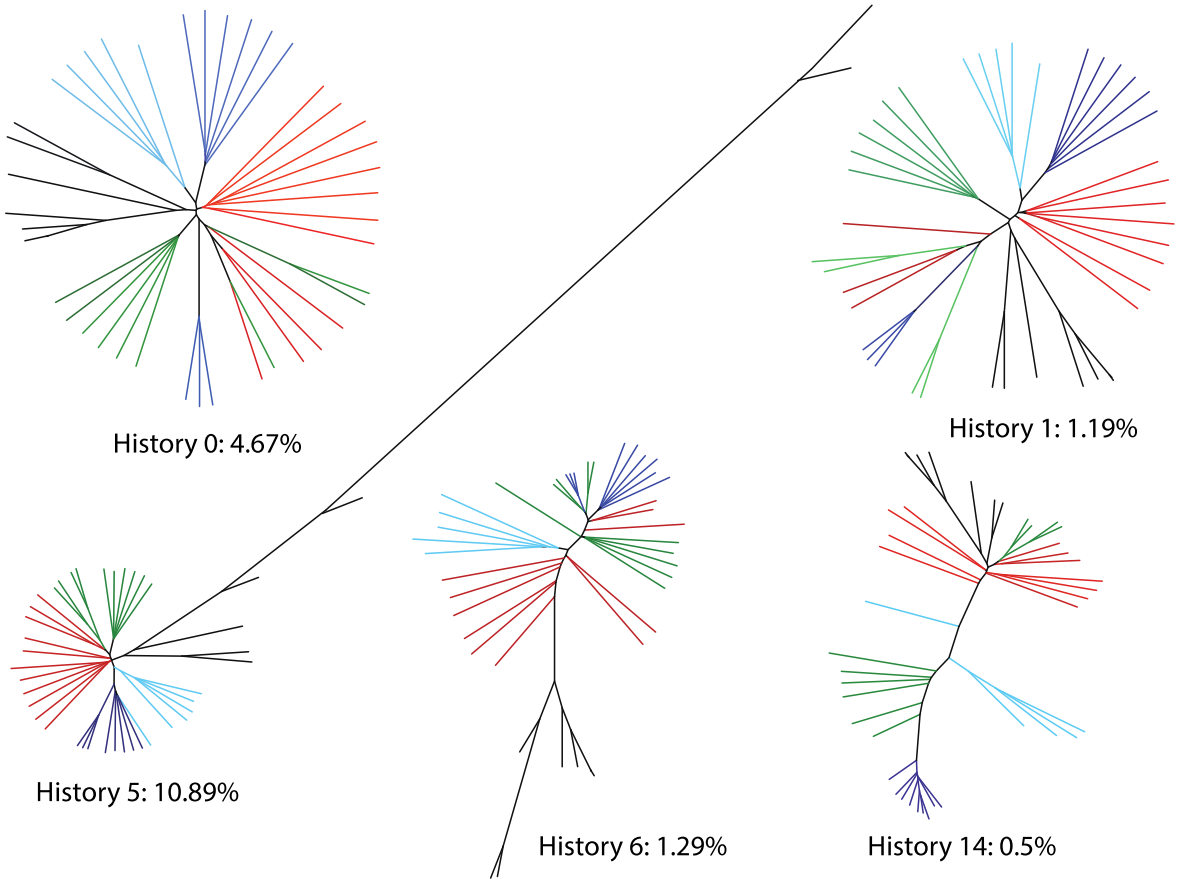
Table 1. Adaptive introgression candidates in San Salvador Island specialists. These 11 candidate regions feature significant f_4 values and signatures of selective sweeps in specialists, SNPs fixed between specialists, and low genetic divergence (D_{xy}) between *C. laciniatus* and one of the specialists. The number of fixed SNPs that were in coding positions of a gene are provided in parentheses after the total number in the region. The specialist(s) with a selective sweep detected in the 98th percentile of the SweeD composite likelihood ratio test and the lowest levels of genetic diversity (π) and Tajima's D estimates within the 2% lower tail of the simulated Tajima's D distribution are listed for each region.

Gene	f_4	Sweep	TREEMIX Directionality	Fixed SNPs	Segregating in generalist?	Low π	Low Tajima's D	GO terms
ski†	0.261**	molluscivores	*C.laciniatus*➡ molluscivores	3(1)	Yes	molluscivore	molluscivore	SMAD binding, cartilage development
rbms3	-0.274*	scale-eaters		1(0)	Yes	scale-eater	scale-eater	RNA binding
pard3†	-0.223**	scale-eaters		57(0)	Yes	scale-eater/ molluscivore	scale-eater	embryonic eye morphogenesis, neuroblast proliferation
NA†	0.255**	scale-eaters		14(-)	Yes	scale-eater/ molluscivore	scale-eater/ molluscivore	-
nbea	-0.28**	molluscivores		40(0)	Yes	scale-eater/ molluscivore	molluscivore	synapse assembly, dendrite development
celf4	0.246**	scale-eaters		27(0)	Yes	scale-eater/ molluscivore	scale-eater	mRNA binding, alternative mRNA splicing
NA	-0.279**	molluscivores/ scale-eaters		1(-)	Yes	molluscivore	molluscivore	-
ltbp2	-0.255**	molluscivores	*C.laciniatus*➡ scale-eaters	2(0)	Yes	molluscivore	molluscivore	microfibril proliferation, calcium ion binding,
srbd1	0.269*	molluscivores/ scale-eaters	*C.laciniatus*➡ molluscivores	19(0)	Yes	scale-eater/ molluscivore	molluscivore	nucleic acid binding, RNA binding
srbd1	-0.267**	molluscivores/ scale-eaters		20(0)	Yes	scale-eater/ molluscivore	molluscivore	nucleic acid binding, RNA binding
mcu	-0.228**	molluscivores		7(0)	Yes	scale-eater/ molluscivore	molluscivore/ scale-eater	mitochondrial calcium homeostasis

P*-value=0.001;P-value<0.001; † gene associated with oral jaw size morphology in San Salvador Island pupfish (McGirr and Martin 2016a); (-) unannotated region

1.8. Supplemental Materials

1.8.1. Supplemental Figures



S1 FigureTopologies featuring a monophyletic San Salvador clade. Black lineages are the *Cyprinodon* outgroups, red lineages are the San Salvador generalists, green lineages are the San Salvador molluscivores, dark blue lineages are the large-jawed scale-eaters and light blue lineages are the small-jawed scale-eaters. Percentages indicate the proportion of the *Cyprinodon* genome assigned to each topology.

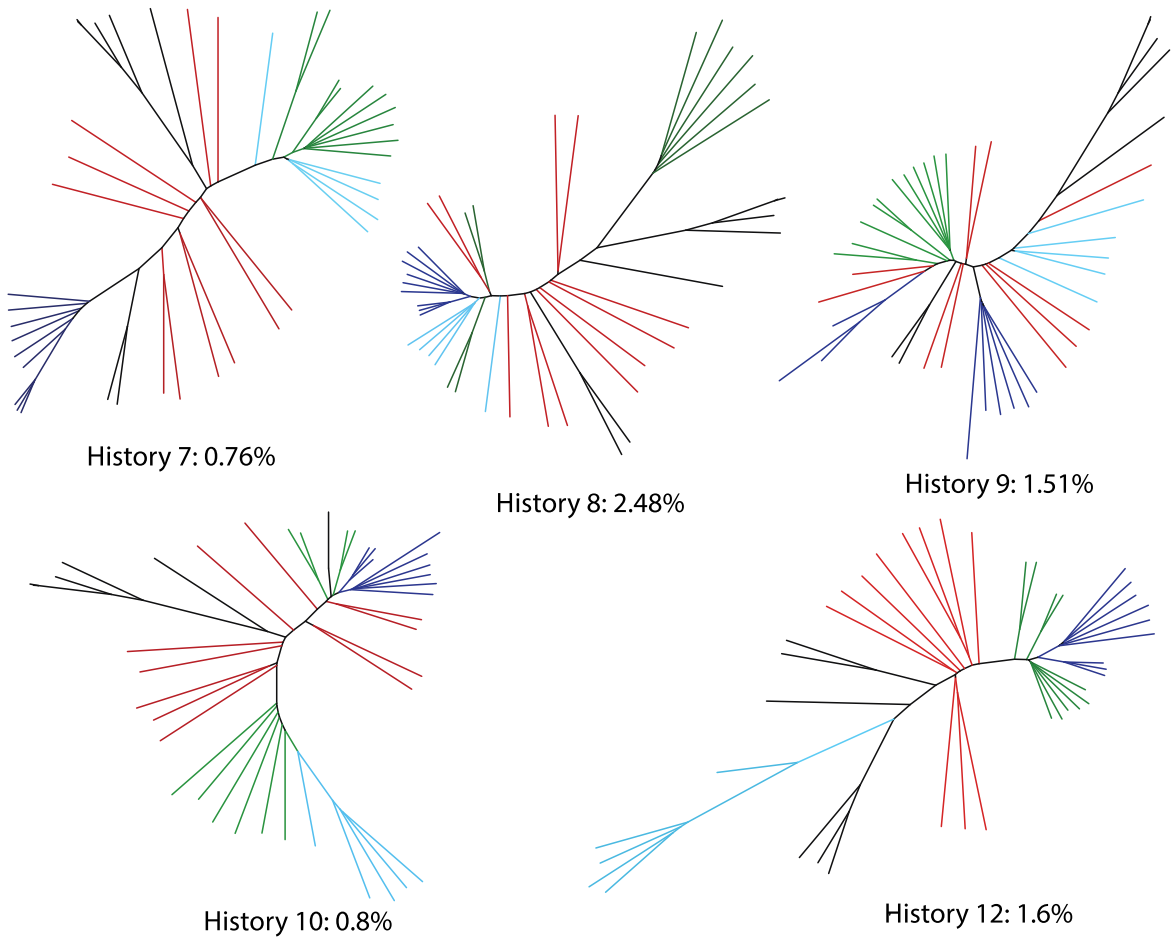
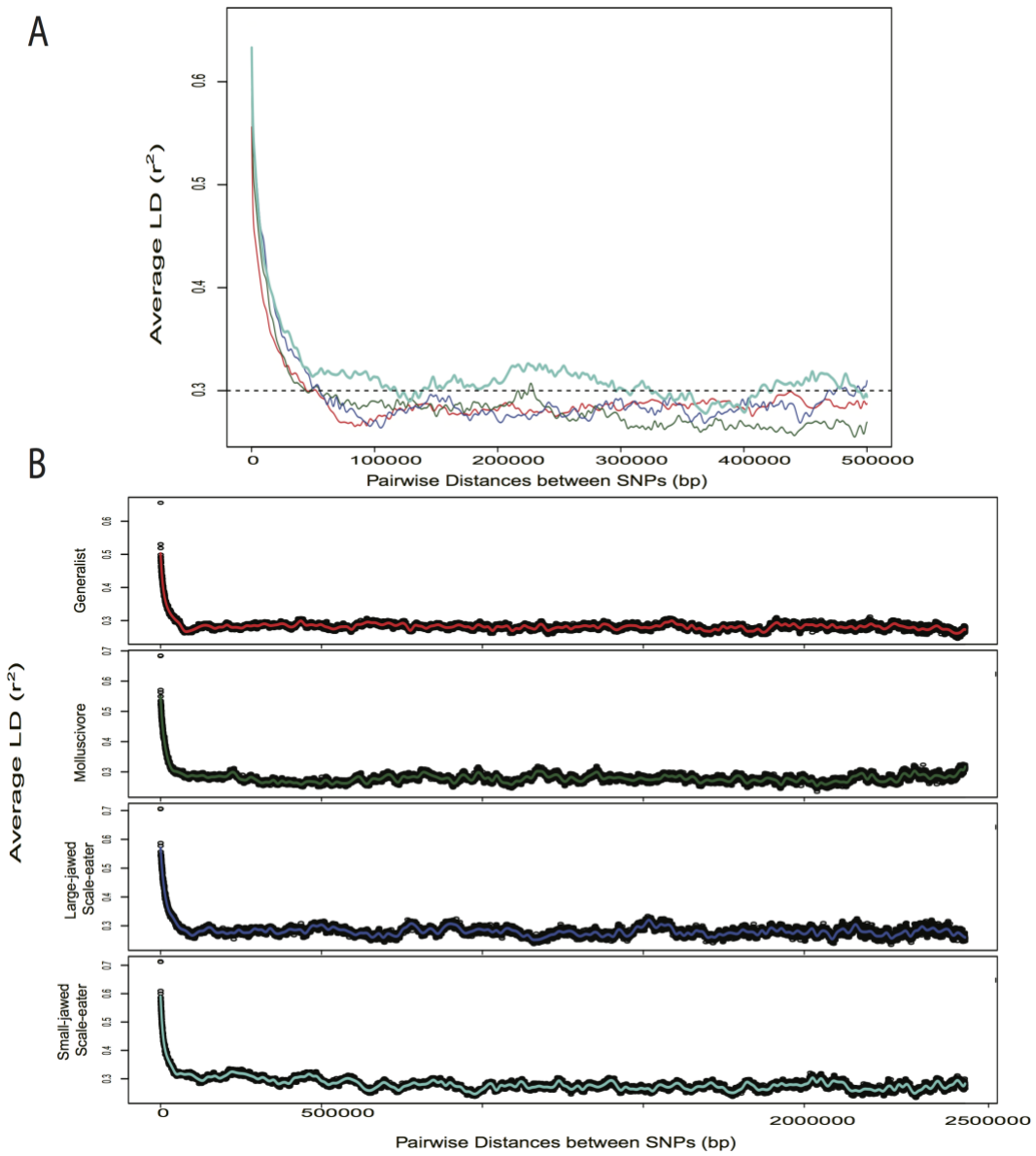
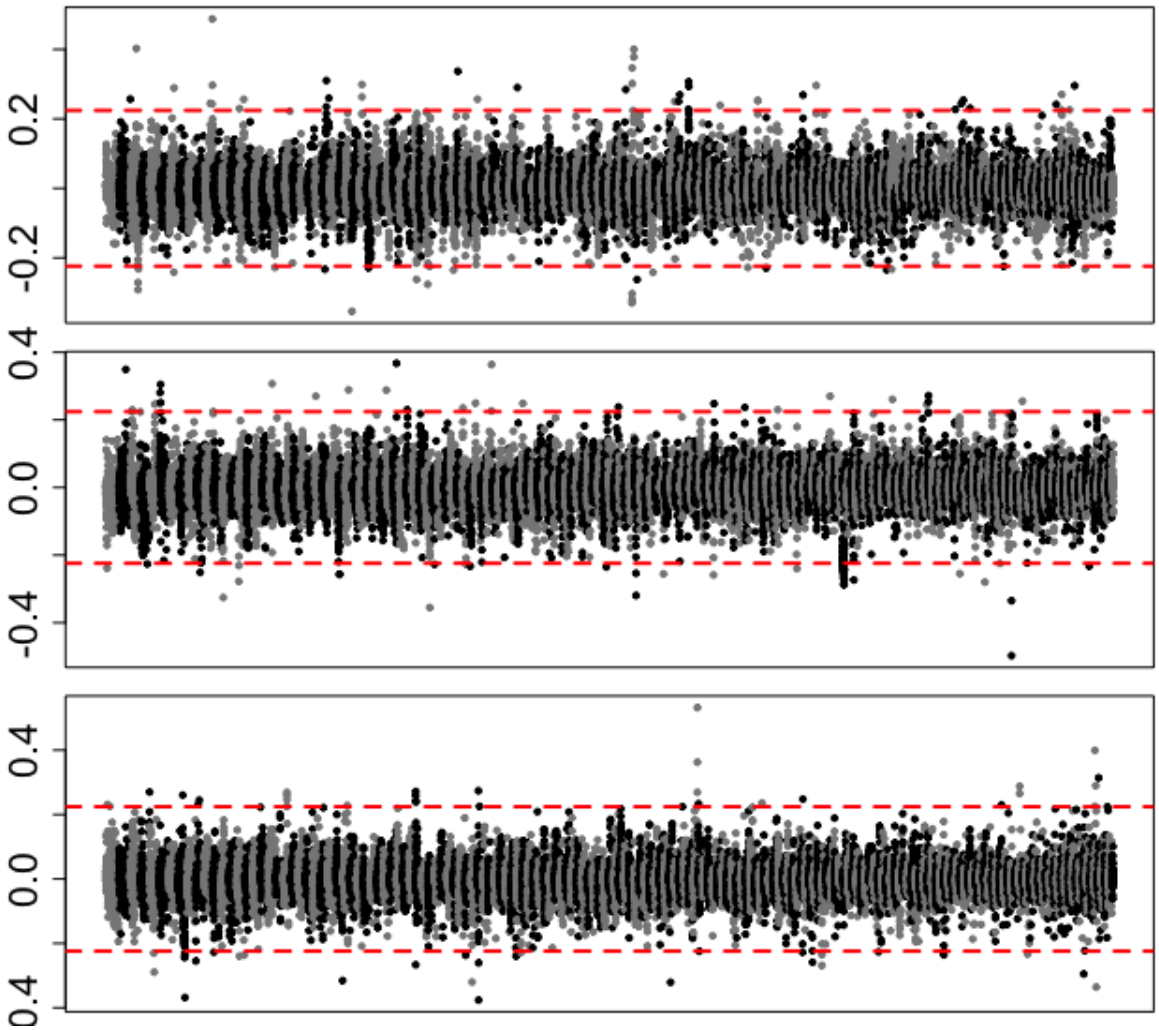


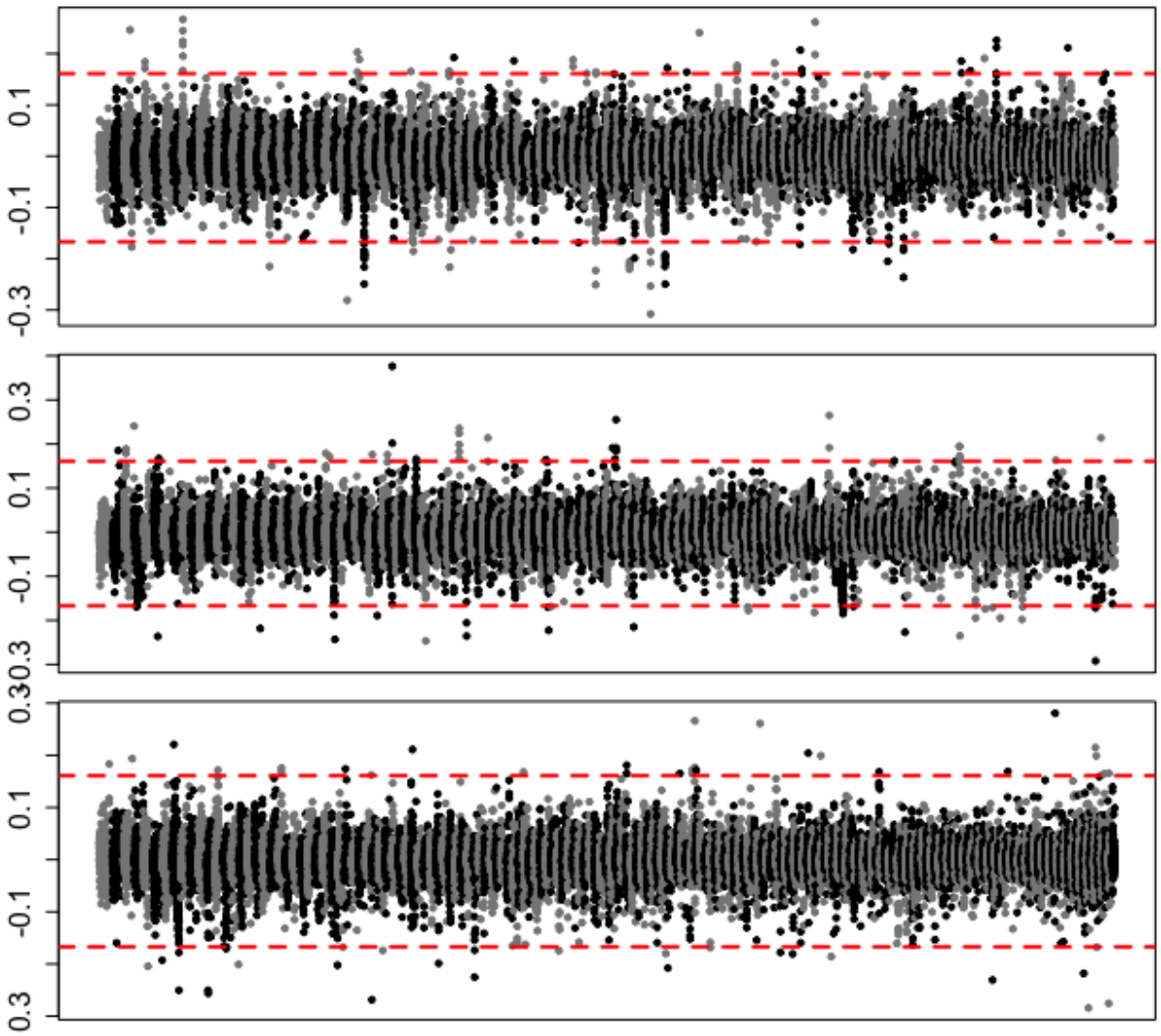
Figure S2. Topologies featuring a non-monophyletic San Salvador clade. Black lineages are the *Cyprinodon* outgroups, red lineages are the San Salvador generalists, green lineages are the San Salvador molluscivores, dark blue lineages are the large-jawed scale-eaters and light blue lineages are the small jawed scale-eater. Percentages indicate the proportion of the *Cyprinodon* genome assigned to each topology.



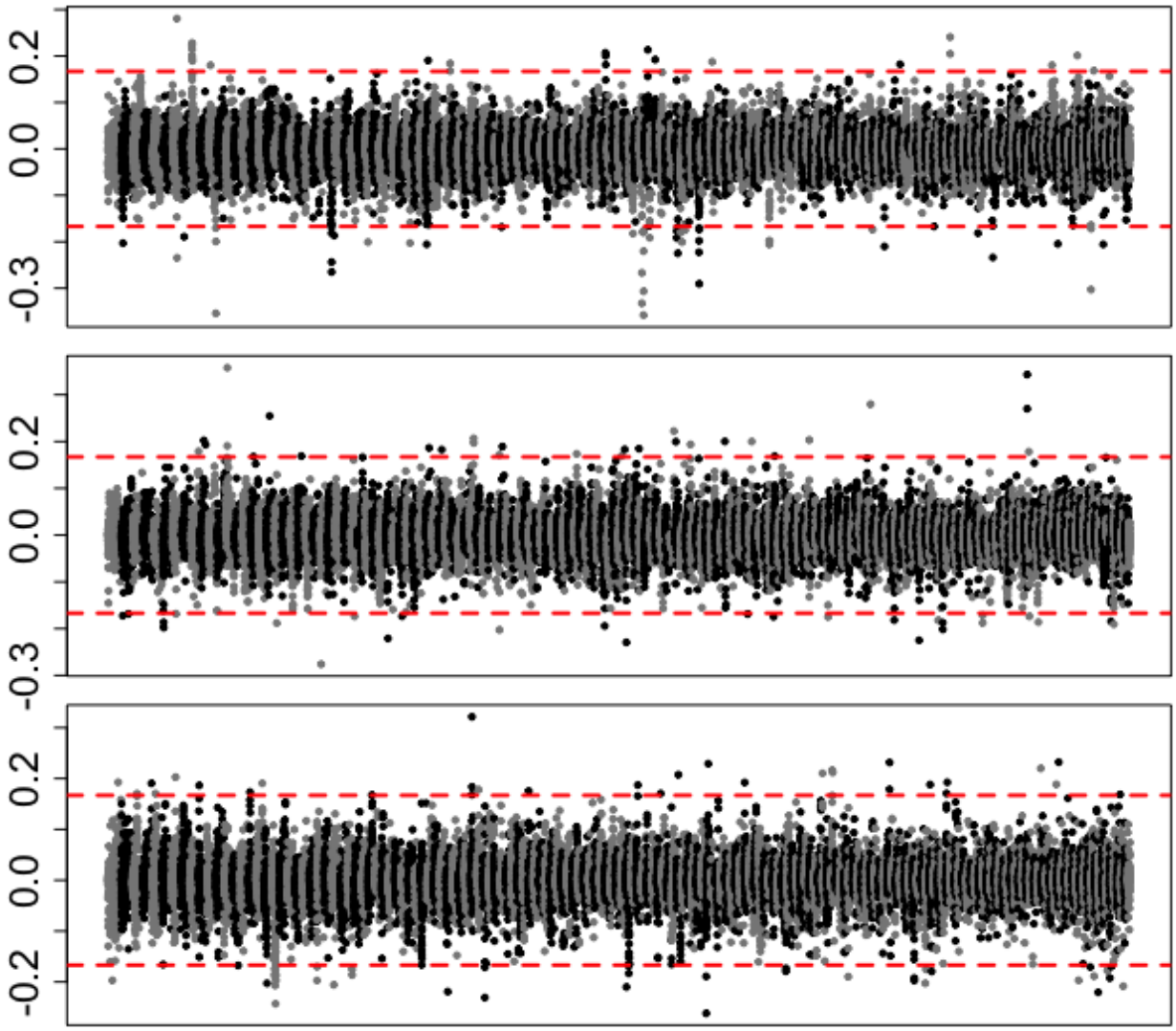
S3 FigureLinkage disequilibrium decay in San Salvador pupfishes. Average r^2 values for pairwise SNPs A) within a distance of 500,000 bp of each other and B) across the entirety of the largest scaffold of the genome (KL652500.1, 4.2 Mb). r^2 was calculated from 5 individuals of each of the San Salvador species: generalists (red), molluscivores (green), large-jawed scale-eaters (dark blue) and small-jawed scale-eaters (light blue). The black horizontal dashed line in panel A is arbitrarily set at r^2=0.3 as a marker for comparing decay between the four groups.



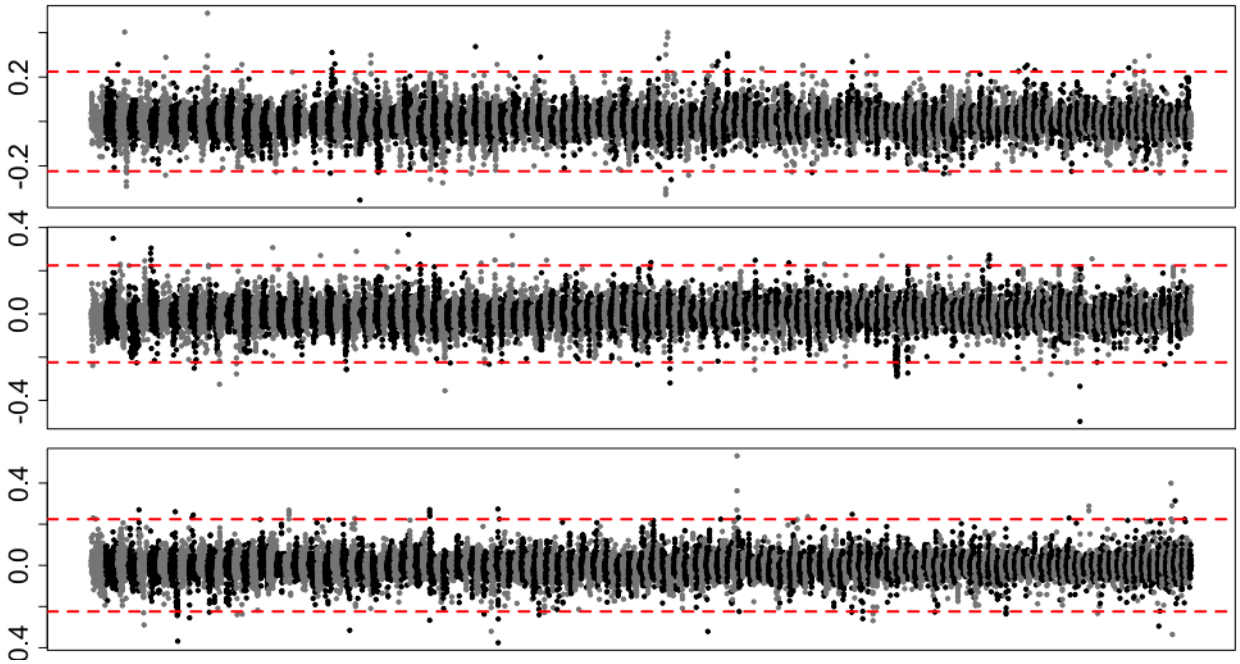
S4 FigureVisualization of introgression across the genomes of molluscivores and scale-eaters. Manhattan plot of the f_4 values between the San Salvador molluscivores, scale-eaters, *C. laciniatus* from New Providence Island, Bahamas and *C. bondi* from Etang Saumatre, Dominican Republic. Alternating gray/black colors indicate different scaffolds, starting with the largest scaffolds in the top row and the smallest scaffolds in the bottom row. Dotted red lines mark the permutation based two-tailed significance level threshold of 0.02.



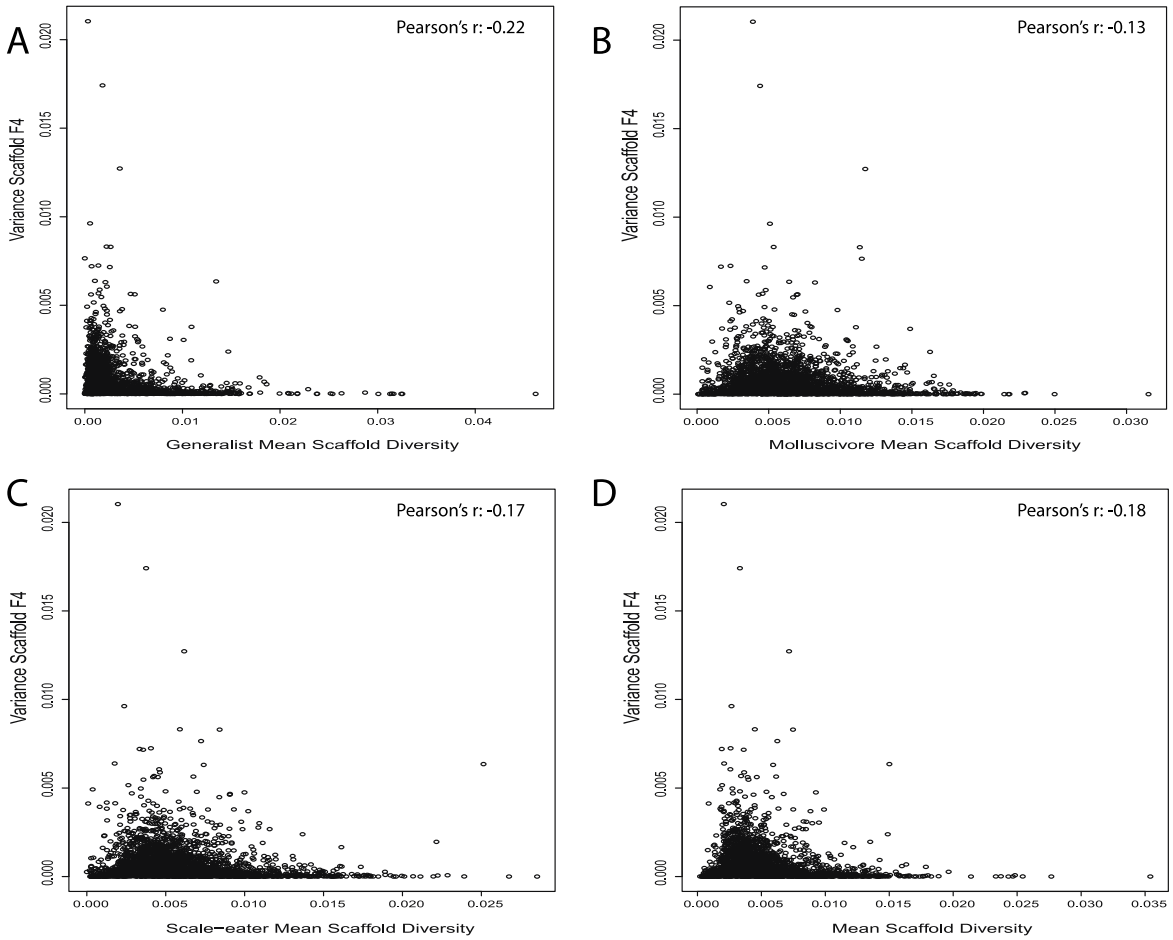
S5 FigureVisualization of introgression across the genomes of molluscivores and generalists. Manhattan plot of the f_4 values between the San Salvador molluscivores, generalists, *C. laciniatus* from New Providence Island, Bahamas and *C. bondi* from Etang Saumatre, Dominican Republic. Alternating gray/black colors indicate different scaffolds, starting with the largest scaffolds in the top row and the smallest scaffolds in the bottom row. Dotted red lines mark the permutation based two-tailed significance level threshold of 0.02.



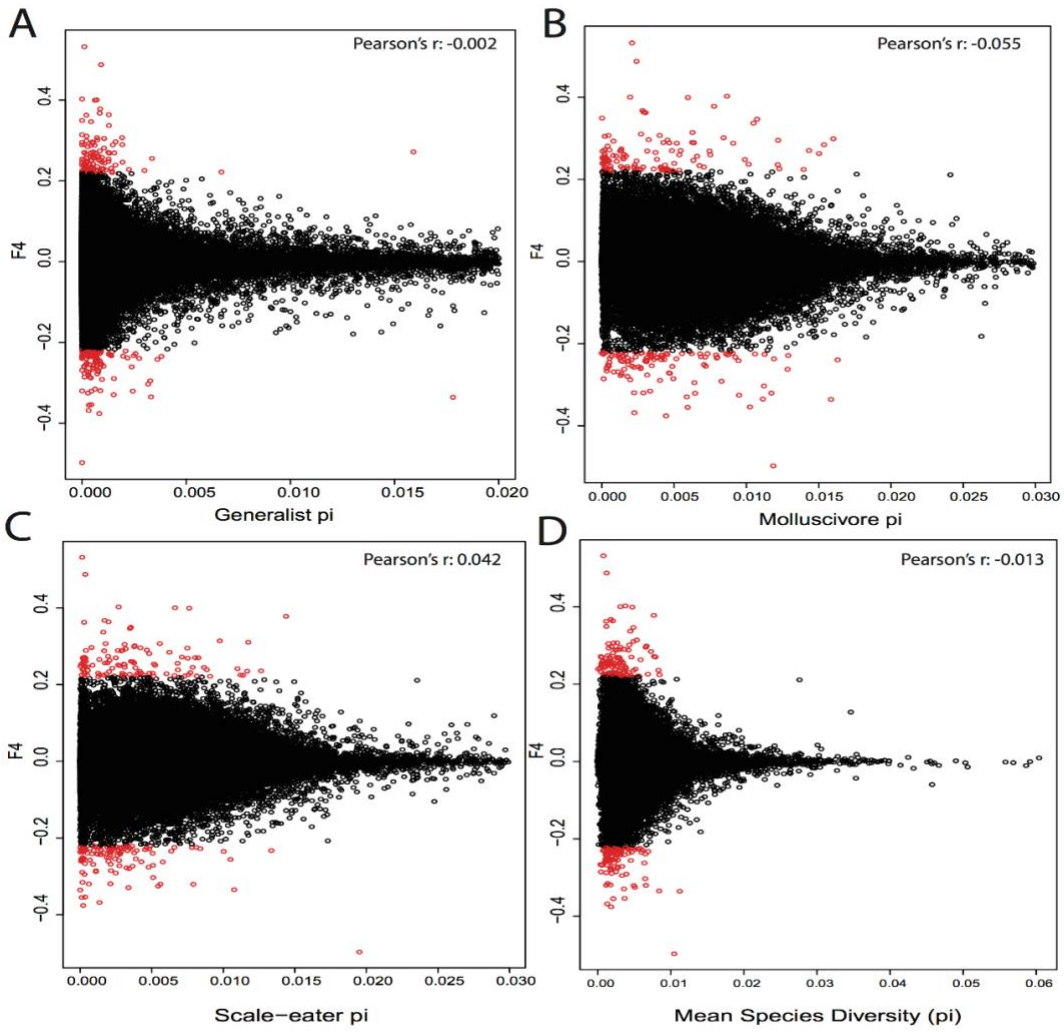
S6 FigureVisualization of introgression across the genomes of scale-eaters and generalists. Manhattan plot of the f_4 values between the San Salvador large-jawed scale-eaters, small-jawed scale-eaters, *C. laciniatus* from New Providence Island, Bahamas and *C. bondi* from Etang Saumatre, Dominican Republic. Alternating gray/black colors indicate different scaffolds, starting with the largest scaffolds in the top row and the smallest scaffolds in the bottom row. Dotted red lines mark the permutation based two-tailed significance level thresholds of 0.02.



S7 FigureVisualization of introgression across the genomes of large-jawed scale-eaters and molluscivores. Manhattan plot of the f_4 values between the San Salvador large-jawed scale-eaters, molluscivores, *C. laciniatus* from New Providence Island, Bahamas and *C. bondi* from Etang Saumautre, Dominican Republic. Alternating gray/black colors indicate different scaffolds, starting with the largest scaffolds in the top row and the smallest scaffolds in the bottom row. Dotted red lines mark the permutation based two-tailed significance level thresholds of 0.02.



S8 FigureComparison of variance in f_4 to genetic diversity statistics over 10-kb non-overlapping windows. The variance in f_4 statistic of a region compared to within-population diversity in A) molluscivores B) scale-eaters, and C) generalists, D) and average within-population diversity in all three species.



S9 FigureComparison of f_4 to genetic diversity statistics over 10-kb non-overlapping windows. Red dots indicate 10-kb regions with signals of introgression above permutations based significance level. The f_4 statistic of a region compared to within-population diversity in A) molluscivores and scale-eaters B) scale-eaters and generalists, and C) molluscivores and generalists D) and average within-population diversity in all three species.

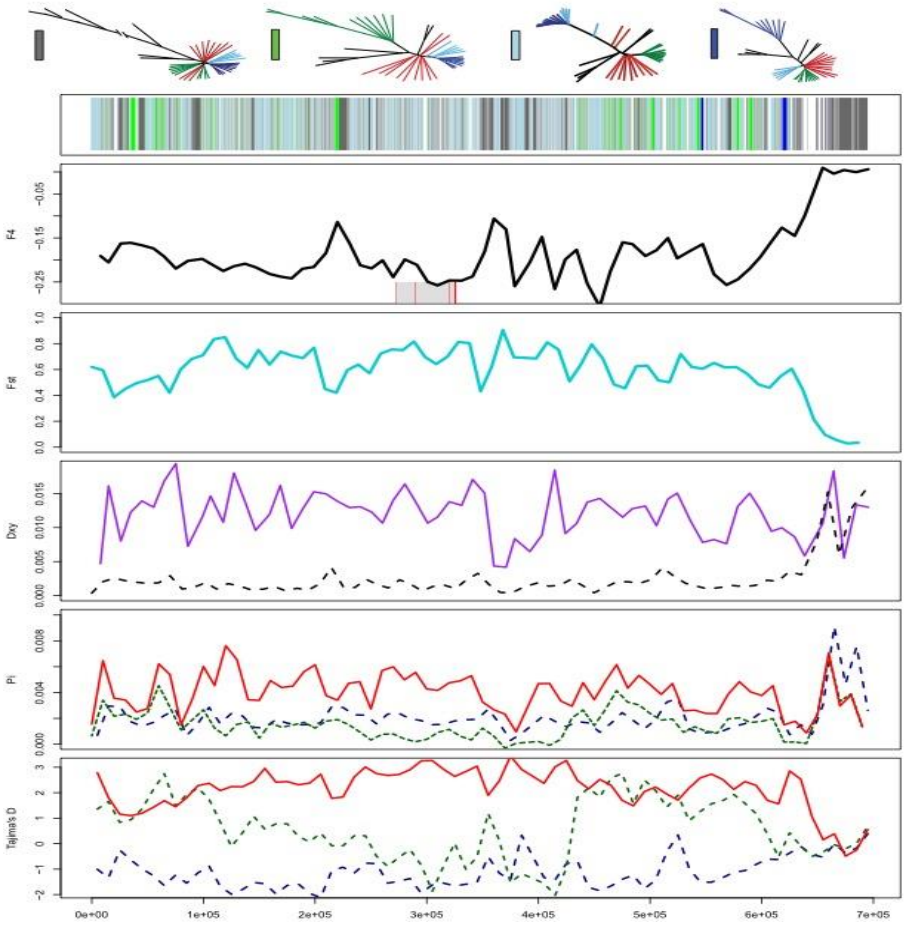


Figure S10. Candidate adaptive introgression regions in gene *wnt7b* with low diversity in all San Salvador species. Row 1 shows the history assigned by SAGUARO to segments along a 700-kb scaffold (dark grey: dominant topology; blue: large-jawed scale-eater topology; light blue: combined scale-eater topology**;** green: molluscivore topology; light grey: all other topologies; white: unassigned segments). Row 2 shows average f_4 value across non-overlapping 10-kb windows between molluscivores/scale-eaters. Shaded grey box shows region annotated for *wnt7b* gene with exons in red. Row 3 shows average F_{st} value across non-overlapping 10-kb windows between molluscivores/scale-eaters (turquoise). Row 4 shows between-population divergence (D_{xy}) across non-overlapping 10-kb windows between molluscivores/scale-eaters and scale-eaters/*C. laciniatus* (grey-dashed). Row 5 shows within-population diversity (π) across non-overlapping 10-kb windows (blue-dashed: scale-eater; green: molluscivore). Row 6 shows Tajima's D across non-overlapping 10-kb windows (blue-dashed: scale-eater; green: molluscivore.

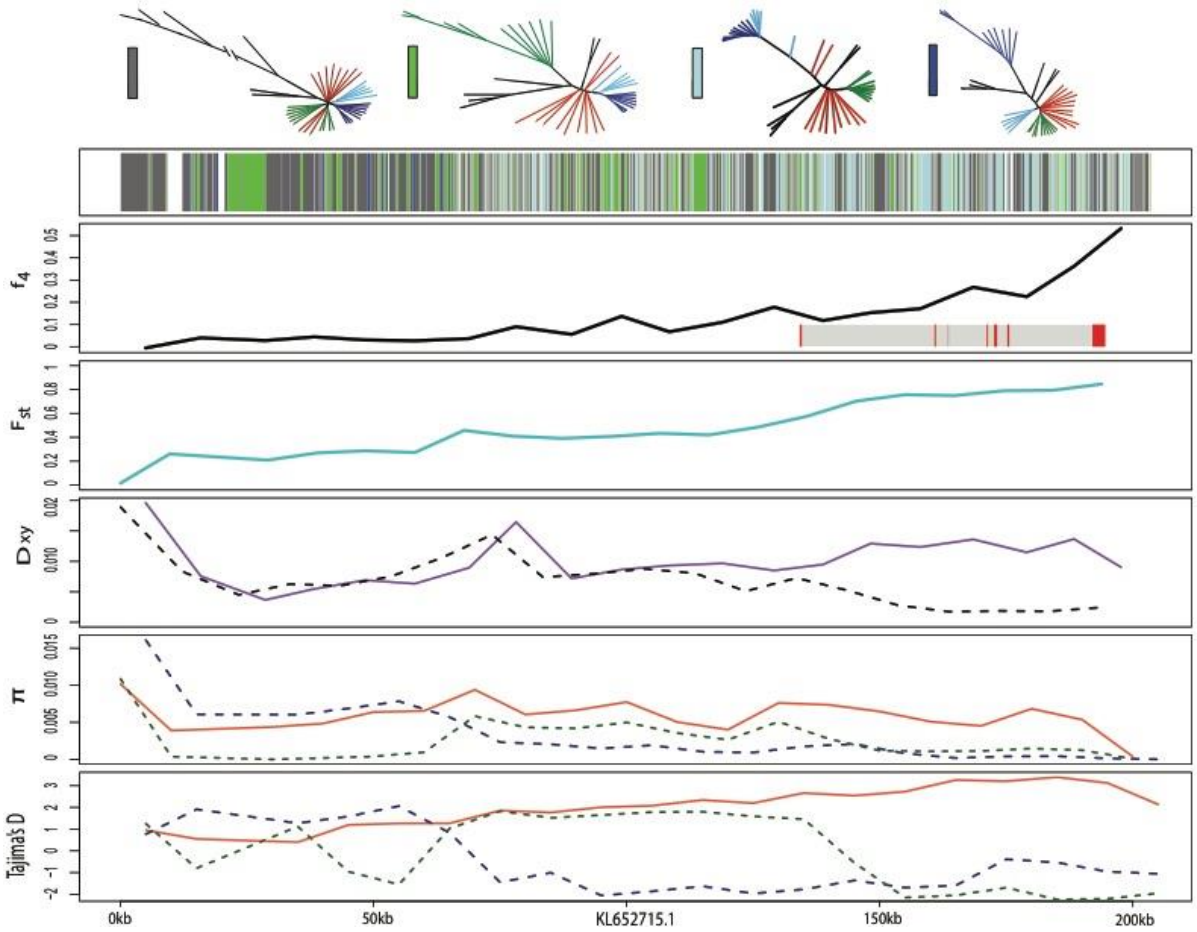


Figure S11. Candidate adaptive introgression region in gene *plekhg* with low diversity in all San Salvador species. Fixed variants in this region were previously associated with pupfish oral jaw size (McGirr and Martin 2016a)**.** Row 1 shows the history assigned by SAGUARO to segments along a 200-kb scaffold (dark grey: dominant topology; blue: large-jawed scale-eater topology; light blue: combined scale-eater topology**;** green: molluscivore topology; light grey: all other topologies; white: unassigned segments). Row 2 shows average f_4 value across non-overlapping 10-kb windows between mollsucivores/scale-eaters. Shaded grey box shows region annotated for plekhg gene with exons in red. Row 3 shows average F_{st} value across non-overlapping 10-kb windows between molluscivores/scale-eaters (turquoise). Row 4 shows between-population divergence (D_{xy}) across non-overlapping 10-kb windows between molluscivores/scale-eaters and molluscivores/*C. laciniatus* (grey-dashed). Row 5 shows within-population diversity (π) across non-overlapping 10-kb windows (blue-dashed: scale-eater; green: molluscivore). Row 6 shows Tajima's D across non-overlapping 10-kb windows (blue-dashed: scale-eater; green: molluscivore.

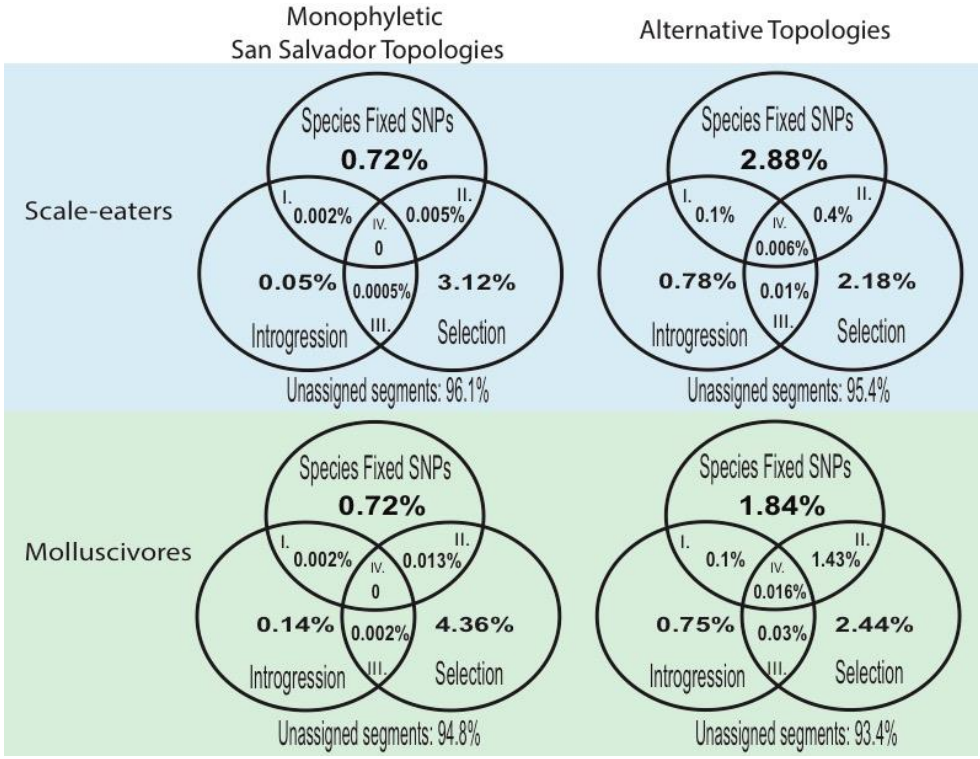


Figure S12. The percentage of segments assigned to the monophyletic San Salvador and alternative topologies that contains signatures of species divergence, selection, and introgression. Venn diagrams of the contribution of different sources of genetic variation to speciation in this system based on the overlap of regions with fixed SNPs between the molluscivore and large-jawed scale-eater, significant f_4 values of introgression, and Tajima's D values below the simulation based lower one-tailed significance level of 0.02. Under each topology, we calculated the percentage of I) regions that contain introgressed genetic variation from the Caribbean contributing to species divergence, II) regions that have undergone strong selective sweeps from non-introgressed genetic variation on San Salvador, III) adaptively introgressed regions not contributing to species divergence , and IV) regions that have undergone selective sweeps of introgressed variation that contributed to species divergence of the two specialists (i.e. contain fixed SNPs between the specialists). The percentage of segments assigned to topologies, but not assigned to any of the above categories, are provided below the Venn diagrams.

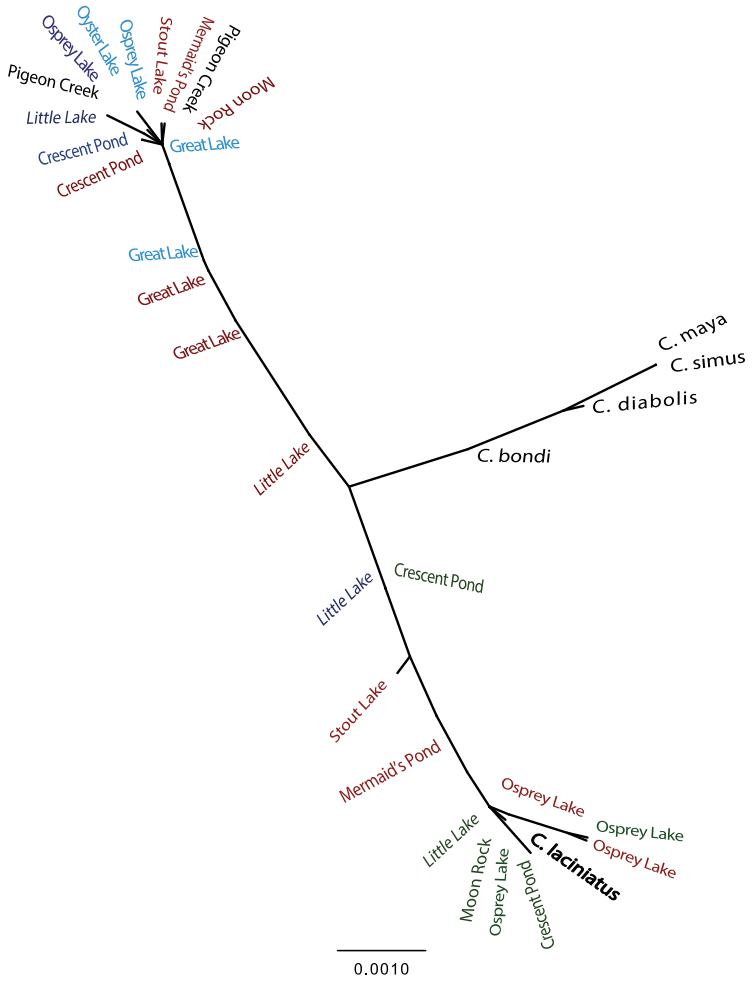


Figure S13. Maximum likelihood tree of the 4,753 bp SAGUARO segment containing the SNPs fixed between specialists in the gene *ski.* The names indicate the pond locality of the individuals (green: molluscivores; dark blue: large-jawed scale-eaters; light blue: small-jawed scale-eaters; black: pupfish outgroups). The scale bar indicates number of substitutions/bp.

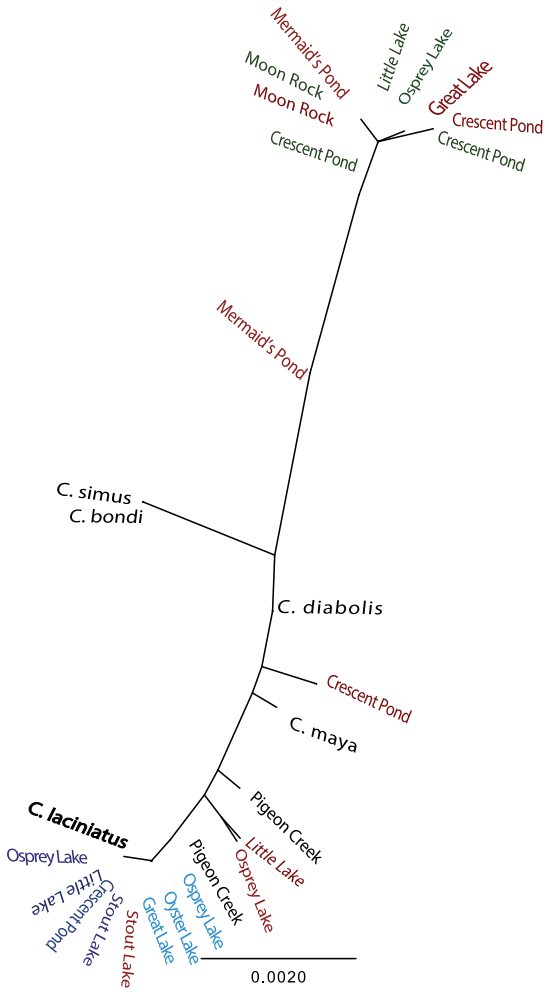


Figure S14. Maximum likelihood tree of the 2,788 bp SAGUARO segment containing the SNPs fixed between specialists in the gene *rbms3*. The names indicate the pond locality of the individuals (green: molluscivores; dark blue: large-jawed scale-eaters; light blue: small-jawed scale-eaters; black: pupfish outgroups). The scale bar indicates number of substitutions/bp.

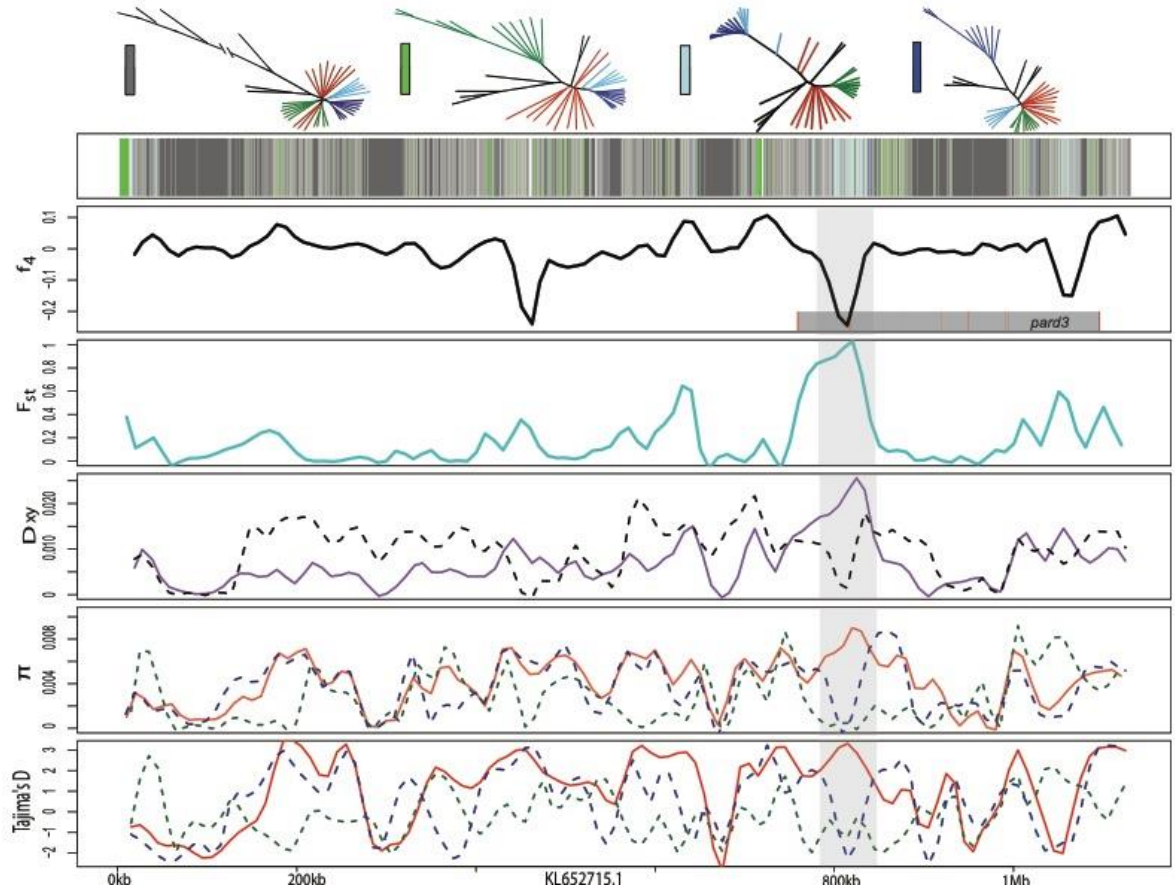


Figure S15. Candidate adaptive introgression region in the gene *pard3*. Fixed variants in this region were previously associated with pupfish oral jaw size (McGirr and Martin 2016a)**.** Row 1 shows the history assigned by SAGUARO to segments along a 1-Mb scaffold (dark grey: dominant topology; blue: large-jawed scale-eater topology; light blue: combined scale-eater topology**;** green: molluscivore topology; light grey: all other topologies; white: unassigned segments). Row 2 shows average f_4 value across non-overlapping 10-kb windows between mollsucivores/scale-eaters. Shaded grey box shows region annotated for *pard3* gene with exons in red. Row 3 shows average F_{st} value across non-overlapping 10-kb windows between molluscivores/scale-eaters (turquoise). Row 4 shows between-population divergence (D_{xy}) across non-overlapping 10-kb windows between molluscivores/scale-eaters (purple) and scale-eaters/*C. laciniatus* (grey-dashed). Row 5 shows within-population diversity (π) across non-overlapping 10-kb windows (blue-dashed: scale-eater; green: molluscivore). Row 6 shows Tajima's D across non-overlapping 10-kb windows (blue-dashed: scale-eater; green: molluscivore.

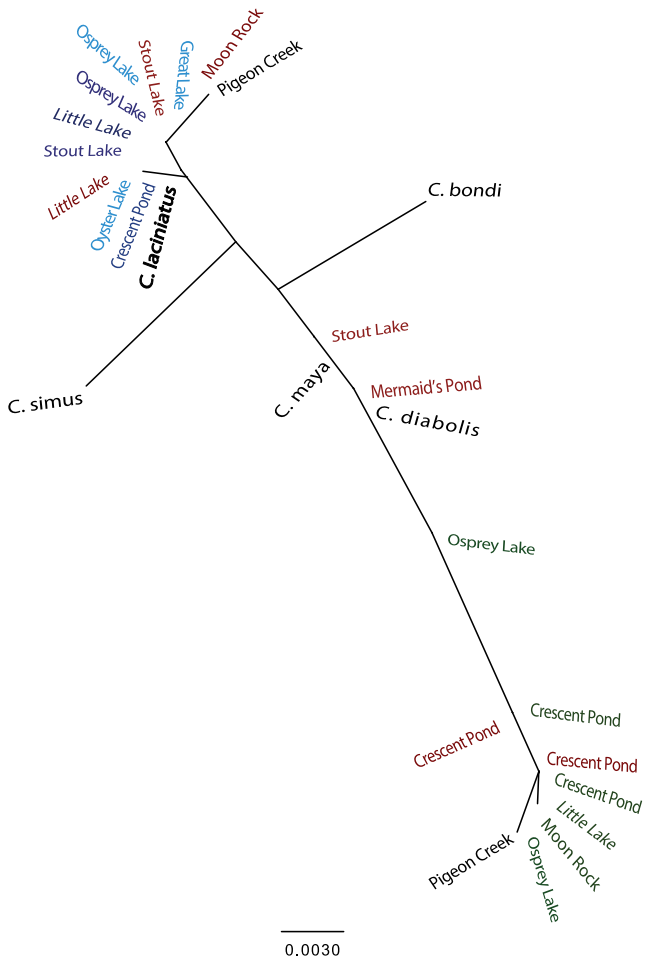


Figure S16. Maximum likelihood tree of the 708 bp SAGUARO segment containing the SNPs fixed between specialists in the gene *pard3*. The names indicate the pond locality of the individuals (green: molluscivores; dark blue: large-jawed scale-eaters; light blue: small-jawed scale-eaters; black: pupfish outgroups). The scale bar indicates number of substitutions/bp.

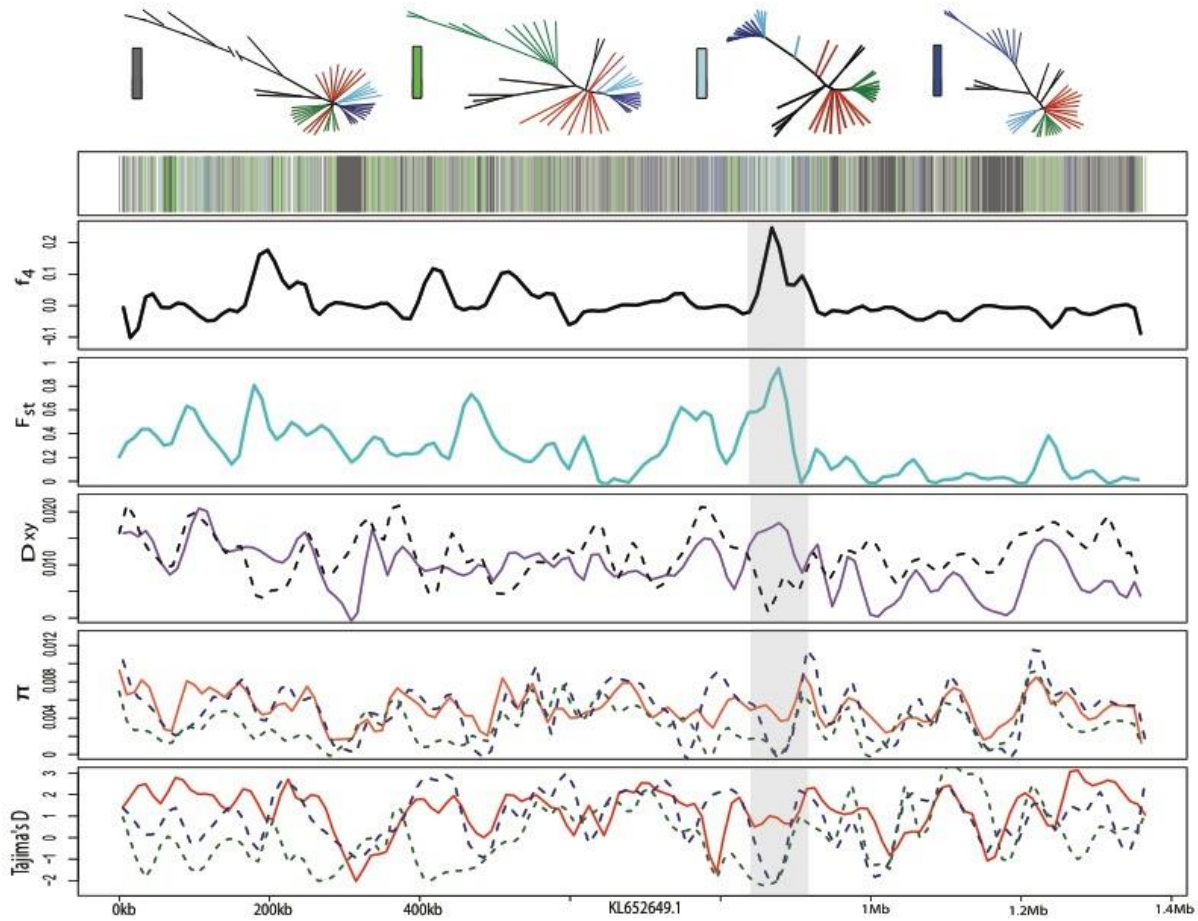


Figure S17. Candidate adaptive introgression region in an unannotated region. Fixed variants in this region were previously associated with pupfish oral jaw size (McGirr and Martin 2016a)**.** Row 1 shows the history assigned by SAGUARO to segments along a 1.4-Mb scaffold (dark grey: dominant topology; blue: large-jawed scale-eater topology; light blue: combined scale-eater topology**;** green: molluscivore topology; light grey: all other topologies; white: unassigned segments). Row 2 shows average f_4 value across non-overlapping 10-kb windows between mollsucivores/scale-eaters. Row 3 shows average F_{st} value across non-overlapping 10-kb windows between molluscivores/scale-eaters (turquoise). Row 4 shows between-population divergence (D_{xy}) across non-overlapping 10-kb windows between molluscivores/scale-eaters (purple) and molluscivores/*C. laciniatus* (grey-dashed). Row 5 shows within-population diversity (π) across non-overlapping 10-kb windows (blue-dashed: scale-eater; green: molluscivore). Row 6 shows Tajima's D across non-overlapping 10-kb windows (blue-dashed: scale-eater; green: molluscivore.

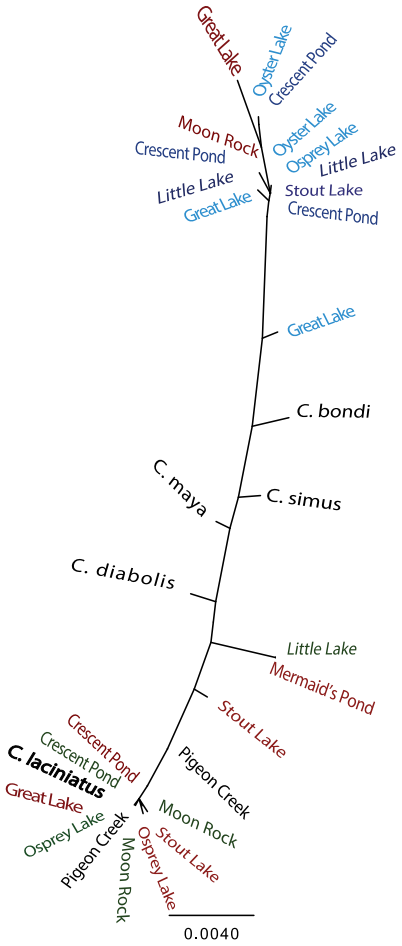


Figure S18. Maximum likelihood tree of the 2,943 bp SAGUARO segment containing the SNPs fixed between specialists in the unannotated region on scaffold KL652649.1. The names indicate the pond locality of the individuals (green: molluscivores; dark blue: large-jawed scale-eaters; light blue: small-jawed scale-eaters; black: pupfish outgroups). The scale bar indicates number of substitutions/bp.

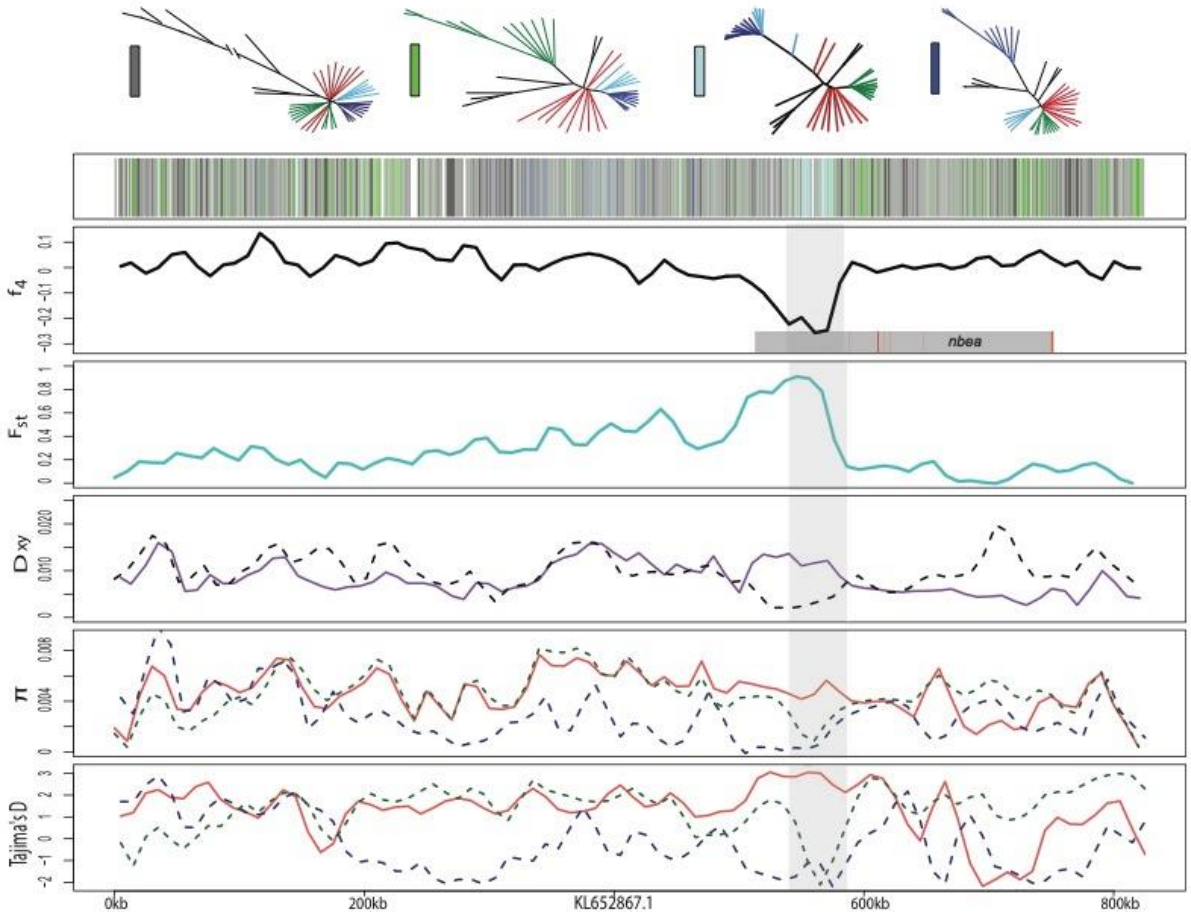


Figure S19. Candidate adaptive introgression region within gene *nbea*. Row 1 shows the history assigned by SAGUARO to segments along a 800-kb scaffold (dark grey: dominant topology; blue: large-jawed scale-eater topology; light blue: combined scale-eater topology**;** green: molluscivore topology; light grey: all other topologies; white: unassigned segments). Row 2 shows average f_4 value across non-overlapping 10-kb windows between mollsucivores/scale-eaters. Shaded grey box shows region annotated for *nbea* gene with exons in red. Row 3 shows average F_{st} value across non-overlapping 10-kb windows between molluscivores/scale-eaters (turquoise). Row 4 shows between-population divergence (D_{xy}) across non-overlapping 10-kb windows between molluscivores/scale-eaters (purple) and scale-eaters/*C. laciniatus* (grey-dashed). Row 5 shows within-population diversity (π) across non-overlapping 10-kb windows (blue-dashed: scale-eater; green: molluscivore). Row 6 shows Tajima's D across non-overlapping 10-kb windows (blue-dashed: scale-eater; green: molluscivore.

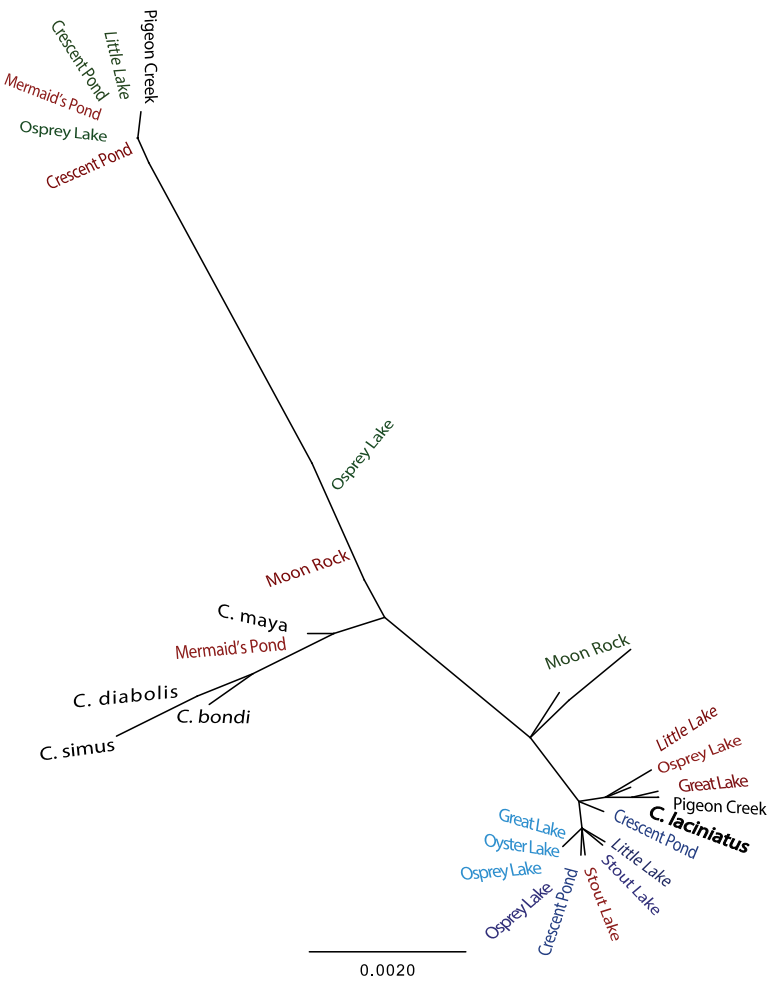


Figure S20. Maximum likelihood tree of the 1,902 bp SAGUARO segment containing the SNPs fixed between specialists in the gene *nbea* The names indicate the pond locality of the individuals (green: molluscivores; dark blue: large-jawed scale-eaters; light blue: small-jawed scale-eaters; black: pupfish outgroups). The scale bar indicates number of substitutions/bp.

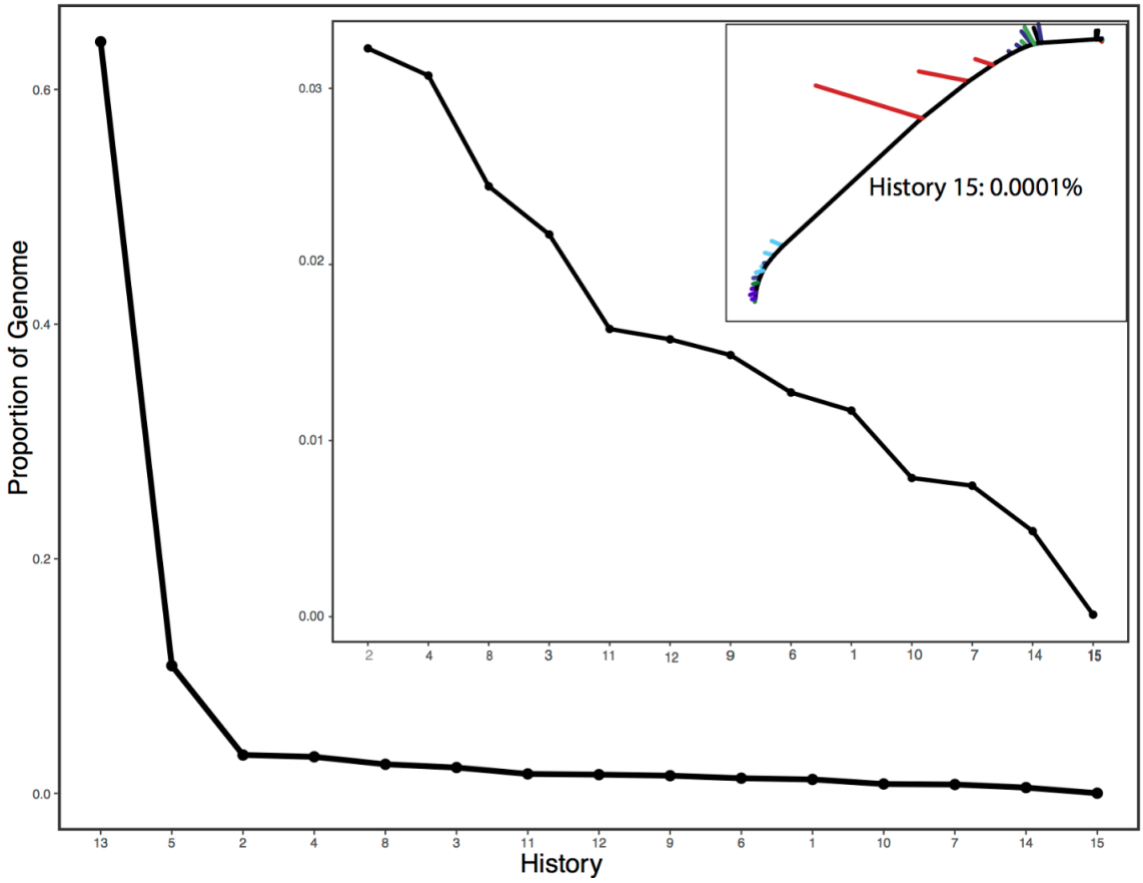


Figure S21. The proportion of the genome assigned to each topology by SAGRUARO. The insert is a closer look at the 13 topologies assigned to the smallest proportion of the genome and the largely uninformative 15th topology. This suggests saturation in the variance explained by topologies at 14.

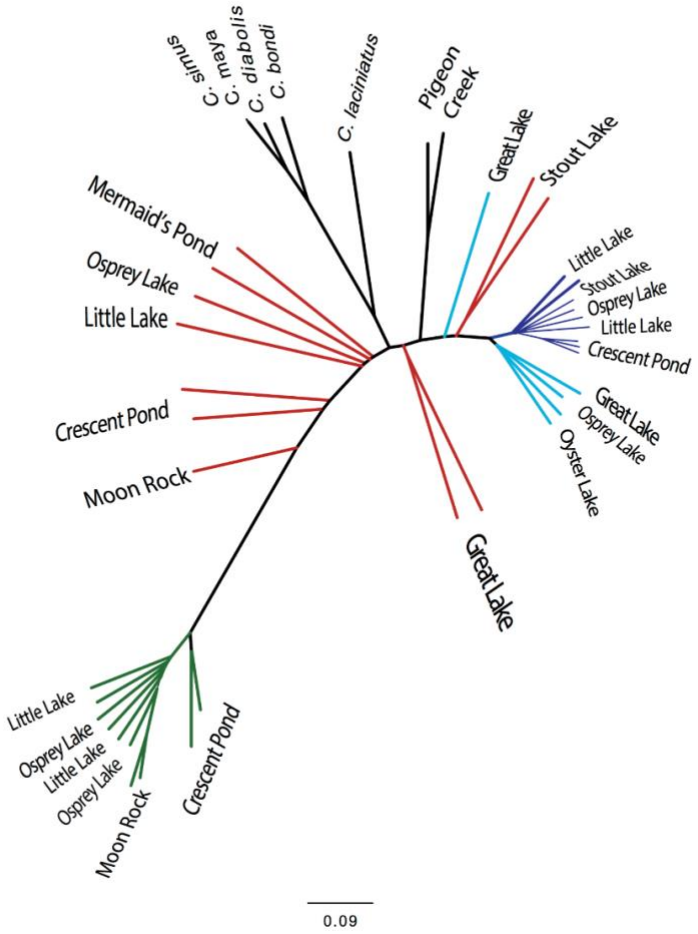


Figure S22. Molluscivore tree at the end of 15 iterations of SAGUARO on the masked genomic dataset. Black lineages are the *Cyprinodon* outgroups, red lineages are the San Salvador generalists, green lineages are the San Salvador molluscivores, dark blue lineages are the large jawed scale-eaters and light blue lineages are the small jawed scale-eater. This topology differs from the molluscivore topology created from unmasked genomic dataset (Figure 2A) in that along with the molluscivores, generalists from Mermaid's Pond, Osprey Lake, Little Lake, Crescent Pond, and Moon Rock Pond appear more closely related to outgroup populations than other San Salvador populations.

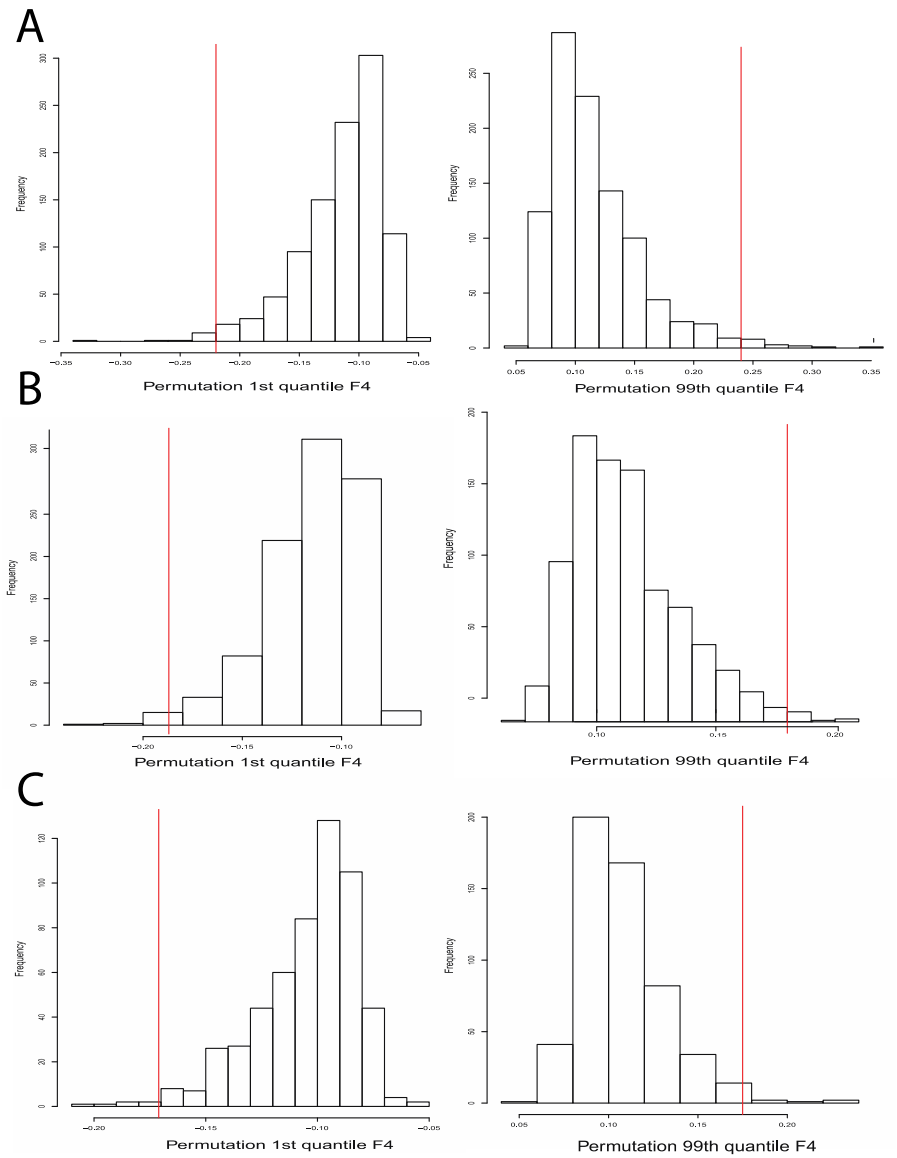


Figure S23. *f*$_4$ 1st and 99th quantiles of null distributions generated from permutations of f4 test across sliding windows of the genome. The red lines represent the 1st quantile (left panels) and 99th (right panels) observed *f*$_4$ values with less than 1% chance of being in the null permutation based distributions of the *f*$_4$ test combinations including a) molluscivores and scale-eaters, b) molluscivores and generalists, and c) scale-eaters and generalists.

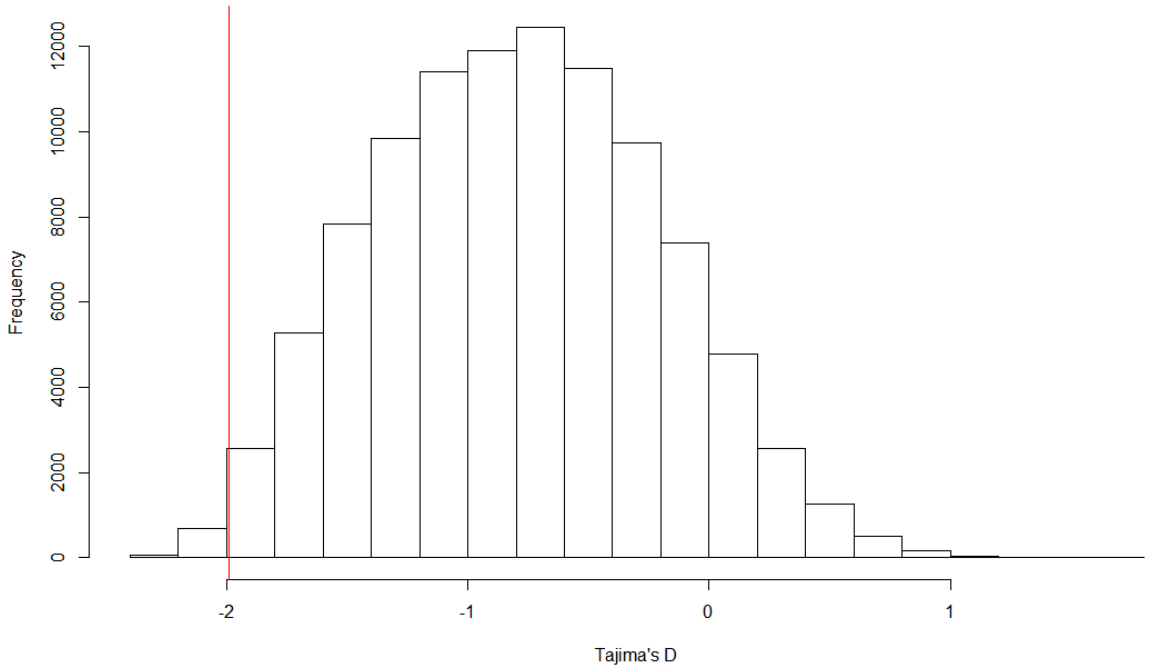


Figure S24. Distribution of Tajima's D values from a coalescence simulation including a bottleneck and introgression. The red line represents the 2nd percentile of the distribution and observed values greater than or equal to this were used to determine regions potentially under selective sweeps.

1.8.2. Supplemental Tables

Table S1. Hypothesized topologies from the SAGUARO analysis.

	Cumulative length (bp)	Percent
Monophyletic		**82.6**
Dominant	580,033,357	64.06
History 5	98,537,488	10.89
History 0	42,238,737	4.67
History 6	11,706,448	1.29
History 1	10,761,277	1.19
History 14	4,479,983	0.5
Non-monophyletic		**17.4**
Large jawed scale-eater	29,638,216	3.27
Molluscivore	28,221,917	3.12
History 8	22,451,018	2.48
History 3	19,942,360	2.2
Scale-eater	15,017,585	1.66
History 12	14,475,837	1.6
History 9	13,655,853	1.51
History 10	7,250,529	0.8
History 7	6,848,171	0.76
Total	**905,258,776**	**100**

Table S2. Ancestry proportions expected of small-jawed scale-eaters if they represent hybrids of the large-jawed scale-eaters and generalists. Ancestry in small-jawed scale-eaters is assigned for SNPs fixed (n=1,887) between the large-jawed scale-eaters and generalists. The genotype count in the 5 small-jawed scale-eater individuals at SNPs fixed between the large-jawed scale-eaters and generalists that are homozygous for one of the parental genotypes or heterozygous between the two (the proportion of loci in is provided in the parentheses). The observed proportion of ancestry in small-jawed scale-eaters does not fit the proportions expected for F_1 hybrids (all heterozygotes), backcross with the one of the parental species (half heterozgyous, half homozygous for parental allele), or F_2 hybrids (half heterzygous, one-fourth homozygous for large-jawed scale-eater, and one-fourth homozygous for small-jawed scale-eater). X^2 value and P-value are provided for the X^2 goodness-of-fit test of the observed proportions of ancestry across the genome to those expected of F_2 hybrids.

Individual	Homozygous Large-jawed Scale-eater	Heterozygous	Homozygous Generalist	F_2 hybrid χ^2	F_2 hybrid χ^2 P-value
GREP1	785 (0.42)	227 (0.15)	825 (0.43)	538.28	1.3×10^{-117}
GREP2	261 (0.13)	514 (0.27)	1129 (0.59)	446.01	1.41×10^{-97}
OSPP2	625 (0.33)	423 (0.22)	839 (0.45)	332.57	9.01×10^{-71}
OYSP1	729 (0.39)	300 (0.16)	858 (0.45)	500.31	2.28×10^{-109}
OYSP3	863 (0.46)	228 (0.12)	796 (0.42)	634.61	1.56×10^{-138}
Average	652.5 (0.35)	348.4 (0.18)	886.9 (0.47)	429.6	5.16×10^{-94}

Table S3. 11 candidate adaptive introgression regions in San Salvador specialists. Adaptively introgressed regions and gene annotations for fixed SNPs between scale-eater and molluscivore species that lie in genomic regions assigned to one of the three alternative topologies. Asterisks (*) indicate SNPs in gene regions associated with San Salvador pupfish oral jaw size variation in a previous study (McGirr and Martin 2016a). Bolded genes have known functional effects on craniofacial traits in a model system. Regions that are not annotated for genes are indicated with a dash (-). *P*-values indicate the number of permutations of the candidate region with f_4 values greater than or equal to the observed f_4 value. The number of fixed SNPs that were in coding positions of a gene are provide in parentheses after the total number of fixed SNPs in the candidate adaptive introgression region. The specialist(s) with a selective sweep detected in the 98th percentile of SweeD composite likelihood ratio test.

								Generalist		Molluscivore		Scale-eater	
Scaffold	Segment	Gene	f_4	*P*-value	Sweep	Fixed SNPs	Avg SNP Coverage	Tajima's D	pi	Tajima's D	pi	Tajima's D	pi
KL652649.1	863668-873661	NA*	0.2536	0	scale-eaters	14(-)	5.57+0.34	0.86	0.006	-2.48	0.0011	-2.04	0.00024
KL652702.1	312277-322263	celf4	0.2461	0.001	scale-eaters	27(0)	6.36+0.53	2.72	0.0042	-2.29	0.00045	-1.54	0.00003 2
KL652715.1	799363-809363	pard3*	-0.223	0	scale-eaters	57(0)	5.55+0.63	2.95	0.0065	-2.18	0.0056	-0.68	0.00022
KL652867.1	545190-575190	nbea	-0.28	0	molluscivores	40(0)	6.02+0.39	2.63	0.0043	-0.86	0.00009 3	-2.18	0.00036
KL652964.1	411177-421153	**rbms3**	-0.2735	0.001	scale-eaters	1(0)	5.21	2.95	0.0041	-1.08	0.0011	-2.19	0.00001
KL652983.1	266059-276054	**ski***	0.2606	0	molluscivores	3(1)	6.19+0.11	2.22	0.0046	-2.24	0.00008	-1.18	0.0016
KL653033.1	403145-413142	NA	-0.2798	0	molluscivores / scale-eaters	1(-)	6.05	2.09	0.0055	-1.96	0.001	-0.81	0.0038
KL653171.1	362672-372487	ltbp2	-0.2546	0	molluscivores	2(0)	4.88+0.17	0.52	0.0055	-2.02	0.0022	0.49	0.0038
KL653356.1	50344-70348	srbd1	0.26	0.001	molluscivores / scale-eaters	19(0)	5.93+0.68	3.25	0.006	-2.11	0.0003	-1.36	0.00018
KL653356.1	70356-80348	srbd1	-0.2666	0	molluscivores / scale-eaters	20(0)	5.92+0.59	2.38	0.0056	-1.98	0.00005	-0.167	0.00027
KL653906.1	10377-20368	mcu	-0.2275	0	molluscivores	7(0)	6.01+0.96	0.46	0.0034	-2.19	0.00087	-1.66	0.0017
Introgressed regions removed from candidates													
KL653706.1	183622-203517	plekhg*	0.43	0.002	molluscivores	41(0)	6.04+0.69	2.13	0.0003 5	-1.96	0.00006 6	-1.05	0.00001 1

KL652959.1	265592-315719	wnt7b	-0.24	0.004	molluscivores / scale-eaters	28(0)	5.75+0.78	2.64	0.001	-1.16	0.0026	-1.94	0.0057

Table S4. Pairwise genetic divergence (Dxy) between molluscivores, scale-eaters, Lake Cunningham, New Providence Island (*C. laciniatus*) and Etang Saumatre, Dominican Republic (*C. bondi*). NA*(2649) is the unannotated candidate adaptive introgression region on scaffold KL652649.1 and NA (3033) is the unannotated candidate adaptive introgression region on scaffold KL653033.1. The two species with the lowest Dxy are bolded for each region.

Comparison	ski	pard3	plekhg	NA*(2649)	rbms3	nbea	wnt7b	celf4	ltbp2	srbd1 (+f4)	srbd1 (-f4)	mcu	NA (3033)
scale-eater v. molluscivore	0.013	0.017	0.014	0.019	0.013	0.012	0.014	0.1	0.01	0.0014	0.015	0.013	0.019
scale-eater v. generalist	0.0075	0.0091	0.0045	0.0009	0.0104	0.0073	0.008	0.008	0.01	0.008	0.01	0.004	0.008
molluscivore v. generalist	0.0073	0.0087	0.014	0.0195	0.0059	0.0069	0.0079	0.0019	0.001	0.005	0.007	0.0099	0.005
scale-eater vs. *C. laciniatus*	0.014	**0.001**	0.015	0.019	**0.002**	**0.0017**	**0.001**	0.011	**0.008**	0.014	**0.003**	**0.0017**	0.012
molluscivore v. *C. laciniatus*	**0.0011**	0.018	**0.0018**	**0.0008**	0.015	0.012	0.015	**0.0004**	0.012	**0.0005**	0.012	0.014	**0.0002**
scale-eater v. *C. bondi*	0.019	0.024	0.016	0.022	0.018	0.015	0.016	0.014	0.018	0.016	0.017	0.012	0.015
molluscivore v. *C. bondi*	0.017	0.022	0.013	0.024	0.012	0.013	0.015	0.013	**0.009**	0.015	0.012	0.016	0.014

S5 Table: Summary of admixture events inferred by TREEMIX for the adaptive introgression regions assigned to the three alternative topologies. San Salvador generalist (A), San Salvador large-jawed scale-eater (L), San Salvador small-jawed scale-eater (S), San Salvador molluscivore (M), *C. laciniatus* from New Providence Island Bahamas (CUN), *C. bondi* from Dominican Republic (ETA), most recent common ancestor of Caribbean pupfish lineages (MRC).

Scaffold	**Segment**	**F4**	**gene**	Admix events	CUN into L	CUN into M	ETA into L	ETA into M	Other			
KL652649.1	863668-873661	0.2536	NA*	3					root M/CUN/PIG/G into S	MRC into A	MRC into PIG	
KL652702.1	312277-322263	0.2461	celf4	2					M into L	L into S	S into L	
KL652715.1	799363-809363	-0.223	pard3*	2					M into MRC	M into G		
KL652867.1	545190-575190	-0.28	nbea	3					CUN into PIG	S into ETA	M into S	
KL652983.1	269059-279054	0.2606	**ski***	1		X						
KL653033.1	403145-413142	-0.279	NA	4					ETA into A	M into A	M into S	M into L
KL653171.1	362672-372487	-0.255	ltbp2	2	X				M into A			
KL653356.1	50344-70348	0.26	srbd1	2					PIG into S	root M,CUN,L into G		
KL653356.1	70356-80348	-0.267	srbd1	2	X				PIG into S			
KL653906.1	10377-20368	-0.228	mcu	2					M into PIG	M into A		
KL652964.1	411177-421153	-0.275	rbms3	3					PIG into ETA	root L,S into PIG	root L,S into A	

Inter-chapter Transition

In Chapter 1, I found evidence of adaptive introgression in a rare radiation of *Cyprinodon* pupfish on San Salvador Island, Bahamas. Such rare radiations are useful for exploring hypothesis about the origins of adaptive radiations because they are often nested within an extensive of closely related outgroup lineages that have not experienced similar diversification. One weakness of such systems however is that they do not contain a series of repeated, independent instances of adaptive radiation and thus it's difficult to tease apart simply correlated from causative factor that have strong associations with radiation in these systems. Stronger support for factors directly involved diversification processes might instead come from cumulative evidence across many similarly rare radiations spanning a broad array of systems. One such similarly rare set of radiations can be found in cichlids radiations from a series of crater lakes in Cameroon. In Chapter 2, I use similar genomic scan approaches as used in Chapter 1 to explore evidence for adaptive introgression contributing to the diversification that occurred in the sympatric species that form the Barombi Mbo cichlid radiation.

Chapter 2: Don't throw out the sympatric speciation with the crater lake water: fine-scale investigation of introgression provides equivocal support for causal role of secondary gene flow in one of the clearest examples of sympatric speciation

This chapter has been previously published and is reproduced here in accordance with the journal's article sharing policy:

Richards EJ, Poelstra, JP, Martin CH. 2017. Don't throw out the sympatric speciation with the crater lake water: fine-scale investigation of introgression provides equivocal support for causal role of secondary gene flow in one of the clearest examples of sympatric speciation. *Evolution Letters 2 (5)*: 520-540. DOI: 10.1002/evl3.78

2.1. Abstract

Genomic data has revealed complex histories of colonization and repeated gene flow previously unrecognized in some of the most celebrated examples of sympatric speciation and radiation. However, much of the evidence for secondary gene flow into these radiations comes from summary statistics calculated from sparse genomic sampling without knowledge of which specific genomic regions introgressed. This tells us little about how gene flow potentially influenced sympatric diversification. Here we investigated whole genomes of Barombi Mbo crater lake cichlids for fine-scale patterns of introgression with neighboring riverine cichlid populations. We found evidence of secondary gene flow into the radiation scattered across < 0.24% of the genome; however, from our analyses it is not clear if the functional diversity in these regions contributed to the ecological, sexual, and morphological diversity found in the lake. Unlike similar studies, we found no obvious candidate genes for adaptive introgression and we cannot rule out that secondary gene flow was predominantly neutral with respect to the diversification process. We also found evidence for differential assortment of ancestral polymorphisms found in riverine populations between sympatric sister species, suggesting the presence of an ancestral hybrid swarm. Although the history of gene flow and colonization is more complicated than previously assumed, the lack of compelling evidence for secondary gene flow's role in species diversification suggests that we should not yet rule out one of the most celebrated examples of sympatric speciation in nature without a more thorough investigation of the timing and functional role of each introgressed region.

2.1.1. Impact statement

Sympatric speciation, the evolution of reproductive isolation without the aid of geographic barriers, is a fascinating process for its illustration of the power of natural and sexual selection alone to create new species. Despite exhaustive searches, only a handful of case studies provide convincing evidence for sympatric speciation in nature. Even in some of the clearest examples of sympatric speciation, the use of genomic data has revealed more complicated histories of gene flow from geographically separated populations than once thought. Evidence for these complicated histories is typically collected as a single point estimate of the entire genome, which can tell us that there was secondary gene flow but nothing about how that gene flow may have contributed to the evolution of new species. Here we exhaustively search whole genomes for signatures of secondary gene flow into the adaptive radiation of Barombi Mbo crater lake cichlids, one of the clearest case studies of sympatric speciation that has recently come under doubt. We characterized genetic and functional diversity in regions of the genome that have experienced gene flow differently among the species to determine the role of gene flow in the speciation process. Very few regions of the genome appear to have experienced differential gene flow and there was no clear evidence that gene flow after initial diversification in lake brought in essential adaptive genetic variation for ecological and morphological diversity. We conclude that multiple colonizations of the lake before diversification began (i.e. a hybrid swarm) may have contributed more to the radiation than secondary gene flow after initial diversification. We still cannot rule out the possibility of sympatric speciation in one of the most celebrated examples in nature.

2.2. Introduction

Sympatric speciation, the endpoint on the speciation-with-gene-flow continuum, is defined as the evolution of reproductive isolation without the aid of geographic barriers under complete panmixia and constant gene flow between diverging populations (Coyne and Orr 2004a; Fitzpatrick et al. 2008; Mallet et al. 2009a). Sympatric speciation has fascinated evolutionary biologists since Darwin for its illustration of the power of complex interactions between natural and sexual selection to create new species. Despite intense searches, very few case studies have been able to meet the rigorous criteria for demonstrating sympatric speciation in nature (Coyne and Orr 2004a; Bolnick and Fitzpatrick 2007). Even in some of the more convincing examples that do meet these criteria, genomic data have revealed more complex evolutionary histories of multiple colonization and repeated gene flow than previously assumed (Papadopolus et al. 2011; The Heliconius Genome Consortium et al. 2012; Geiger et al. 2013; Alcaide et al. 2014; Igea et al. 2015; Malinsky et al. 2015; Martin et al. 2015a; Kautt et al. 2016b).

However, much of the support for complicated histories involving repeated gene flow into sympatric radiations comes from genome-wide tests for gene flow (e.g. (Lamichhaney et al. 2015; Martin et al. 2015a; Meier et al. 2017a)). One prediction of models of speciation with gene flow is that divergence between incipient species should be heterogeneous across the genome (Nosil et al. 2008; Feder et al. 2012; Nosil and Feder 2012). Indeed, high heterogeneity in genomic differentiation has been found across the genomes of many recent or incipient sister species (e.g. (Jones et al. 2012; Martin et al. 2013b; Poelstra et al. 2014; Soria-Carrasco et al. 2014; Malinsky et al. 2015; McGirr and Martin 2016b)), although other processes besides

differential gene flow across the genome can produce similar heterogeneous patterns (Noor and Bennett 2009; Nachman and Payseur 2012; Cutter and Payseur 2013; Cruickshank and Hahn 2014; Guerrero and Hahn 2017; Ravinet et al. 2017). Only a handful of genes may directly contribute to the speciation process whereas the rest of the genome is porous to gene flow while reproductive isolation is incomplete (Wu 2001b; Wu and Ting 2004). Therefore, gene flow detected at the genome-wide level from populations outside the sympatric radiation does not by itself constitute evidence that secondary gene flow was involved in the divergence process among incipient species and shaped the radiation.

The Cameroon crater lake cichlid radiations are some of the most compelling cases for sympatric speciation in the wild (Coyne and Orr 2004a; Bolnick and Fitzpatrick 2007; Martin 2012). The most speciose of these radiations is found in the isolated 2.3 km wide volcanic crater lake Barombi Mbo (Trewavas et al. 1972; Schliewen et al. 1994; Schliewen and Klee 2004). Barombi Mbo hosts a radiation of 11 endemic cichlid species, many of which have clear morphological and ecological separation from other sympatric species (Schliewen et al. 1994). Some endemics have evolved unique specializations, such as the spongivore *Pungu maclareni*, seasonal lekking species *Myaka myaka* (CHM pers. obs.), and the deep-water hypoxia specialist *Konia dikume* (Trewavas et al. 1972). Other endemic species, such as *Stomatepia mariae* and *S. pindu*, appear to be incipient or stalled species complexes with only slight morphological and ecological divergence at the extremes of a unimodal distribution of phenotypes (Martin 2012). However, evidence of differential introgression, weak support for Barombi Mbo monophyly, and differences in levels of shared ancestry with outgroup riverine populations from genome-wide RADseq data all suggest secondary gene flow into the radiation after the initial colonization, casting doubt on one of the best examples of sympatric speciation in the wild (Martin et al. 2015a).

Here we dissect those signals of repeated gene flow to investigate their role in the radiation using whole-genome sequences. We performed exhaustive searches for all genetic patterns consistent with secondary gene flow into the ancestral Barombi Mbo population or into subclades after their initial divergence. We used a machine learning approach to finely dissect and partition different phylogenetic signals across the genome and sliding window genomic scans to test for differential introgression. We found evidence of both shared introgression between sister species and across subclades in the radiation as well as differential introgression among sister species across small regions of the genome. However, functional and genetic diversity in these regions do not paint a clear picture of how introgressed variants may have contributed to speciation in these groups. Our results suggest that either 1) rare introgression of variants in poorly characterized genetic pathways contributed to the morphological and ecological diversity of the radiation (speciation with an allopatric phase), 2) secondary gene flow was predominantly or completely neutral and did not contribute to diversification in Barombi Mbo (sympatric speciation with gene flow), or 3) multiple colonizations of the lake before diversification brought in genetic variation that was then differentially sorted among incipient species (sympatric speciation from a hybrid swarm).

2.3. Methods

2.3.1 Sampling and Genome Sequencing

We sequenced whole genomes of 1-3 individuals from 10 out of the 11 species within the sympatric radiation of Oreochromine cichlids in Cameroon crater lake Barombi Mbo (excluding *Sarotherodon steinbachi* which is morphologically and ecologically similar to the other three *Sarotherodon* species), an endemic *Sarotherodon* species pair from Lake Ejagham, and outgroup *Sarotherodon* individuals from all three river drainages surrounding the lake: Meme, Mungo, and Cross rivers (Figure 1A). Details on the collection, extraction, alignment to the *Oreochromis niloticus* reference genome (Brawand et al. 2014), and variant calling protocols following the standard GATK pipeline (McKenna et al. 2010) are provided in the supplementary methods.

2.3.2. Characterization of introgression patterns across the genome

We exhaustively searched our genomic dataset for patterns of non-monophyletic Barombi Mbo relationships using the machine learning program *Saguaro* (Zamani et al. 2013) to identify regions of the genome that supported phylogenetic topologies consistent with expectations from multiple colonizations and secondary gene flow into the radiation (i.e. paraphyletic/polyphyletic Barombi Mbo radiations). This method infers relationships among individuals in the form of genetic distance matrices and assigns segments across the genomes to different topologies without *a priori* hypotheses about these relationships. We partitioned the genome into a total of 75 unique topologies (well past the inflection point at 30 topologies where the percentage of the genome covered by each additional topology plateaus; FigureS1) to exhaustively search for relationships where subclades or individual Barombi Mbo species were more closely related to riverine populations than other species in the crater lake, suggesting sympatric speciation after a hybrid swarm (i.e. differential sorting of ancestral polymorphism) or secondary gene flow into this subclade (introgression). Because we instructed *Saguaro* to propose trees beyond the normal stopping rules based on the percentage of the genome explained by additional topologies in order to exhaustively search for introgressed regions, the genomic proportions assigned to each tree may not be accurate. These proportions are therefore treated with caution and only used to complement the other results in this study. Details on the *Saguaro* analysis and filtering strategies for calculating proportions are provided in the supplementary methods.

We also characterized differential introgression within subclades of the radiation on both a genome-wide and local level. We tested for differential introgression between Barombi Mbo species and riverine populations at the genome-wide level using f_4 statistics (Reich et al. 2009; Patterson et al. 2012; Pickrell and Pritchard 2012). We focused on tests of introgression with the three surrounding riverine populations of the most closely related outgroup *Sarotherodon galilaeus* in the nearby Mungo and Meme rivers (MM) and *S. galilaeus* from the more distant Cross River (CR). Based on the tree ((P1, P2),(*S. galilaeus* MM, *S. galilaeus* CR)), f_4 statistics were calculated for combinations of species among a) *Stomatepia*, b) the *Konia + Pungu* subclade, and c) *Myaka myaka* with *S. linnelli* as a representative of its sister *Sarotherodon* group. This subset of groupings was chosen to make these analyses more tractable by focusing on species with unique trophic ecologies within the radiation.

Genome-wide f_4 statistics were calculated using the fourpop function in *Treemix* (Pickrell and Pritchard 2012). Standard error was estimated by jackknifing in windows of 1,000 adjacent

SNPs to account for linkage disequilibrium. We also visualized the directionality of genome-wide introgression detected with the f_4 statistics using *TreeMix* (v 1.13) (Pickrell and Pritchard 2012). *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 among populations. We ran *TreeMix* with *S. galilaeus* as root, and with 0 through 20 migration edges. To determine the most likely number of migration events, we performed likelihood-ratio tests comparing each graph to one fewer migration event, starting with 1 versus 0 events, and took as the most likely value the first non-significant comparison.

We investigated whether observed signatures of differential introgression at the genome-wide level contributed variation potentially important to the divergence between species using the f_d statistic, a test statistic similar to f_4, but designed to be calculated across sliding genomic windows (Martin et al. 2015b). 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 (Martin et al. 2015b). We used individuals of *Coptodon kottae* from another Cameroon crater lake (Figure 1A) as our distantly related outgroup population for this test and focused on introgression between Barombi Mbo species and surrounding riverine populations *S. galilaeus* MM and *S. galilaeus* CR. Based on the tree ((P1,P2),(*S. galilaeus*, *C. kottae*)), the f_d statistic was calculated for the same combinations of populations used in the genome-wide tests in 50-kb sliding windows with a minimum of 100 variant sites with no missing data within a population using the ABBABABA.py script (Martin et al. 2015b, available on https://github.com/simonhmartin/genomics_general). Significance of f_d values in sliding windows across the genome were evaluated using simulations with no migration using ms-move (Garrigan and Geneva 2014). We used the multi-species Markovian coalescent approach (MSMC: (Schiffels and Durbin 2014)) to estimate effective population size changes through time in the Barombi Mbo cichlids and riverine outgroups in order to parameterize our simulations. Significant regions were those containing f_d values above the values calculated from simulated windows. For more details on the sliding window calculations of f_d and significance assessment, see supplementary methods.

We divided significant introgressed regions into three categories: 1) regions found only in a single Barombi Mbo species, 2) regions shared with another species within the same subclade (suggestive of introgression before divergence of the subclade), and 3) regions shared between two or more species across the entire clade, suggesting sympatric speciation after a hybrid swarm (i.e. differential sorting of ancestral polymorphism) or secondary gene flow into multiple subclades (introgression). We also looked for overlap between introgressed candidate regions between tests involving *S. galilaeus* MM and *S. galilaeus* CR, another pattern suggestive of a hybrid swarm after initial colonization. For each of these regions, we looked for annotated genes using the well annotated NCBI *Oreochromis* Annotation Release 102 and searched their gene ontology in the phenotype database 'Phenoscape' (Mabee et al. 2012; Midford et al. 2013; Manda et al. 2015; Edmunds et al. 2016) and AmiGO2 (Balsa-Canto et al. 2016) for functions related to the trophic specializations and observed morphological differences among specialist species, such as skeletal system, circulatory system, metabolism, or pigmentation. It is possible that some of the topologies consistent with introgression with outgroups and introgressed regions from f_d tests stem from introgression from unsampled or extinct populations rather than *S. galilaeus* MM or *S.galilaeus* CR directly, however this should not change the overall conclusion

that secondary gene flow events occurred in the history of the radiation or the functional support for the importance of the introgressed regions that we detected.

2.3.3. Comparison of patterns of introgression to patterns of genetic divergence and diversity

Reduced levels of genetic polymorphism in a population may indicate a strong selective sweep. We examined introgressed regions found in only a single Barombi Mbo species for evidence of selection, suggesting that secondary gene flow brought in variation potentially important for speciation. To examine genetic diversity in candidate introgressed regions, we calculated between-population nucleotide divergence (D_{xy}) and within-population nucleotide diversity (π) for pairwise species comparisons among our Barombi Mbo focal specialist species and the riverine outgroups over the same 50-kb windows as the f_d tests (see supplementary methods for more details on these calculations).

2.4. Results

2.4.1. Widespread evidence of polyphyly in Barombi Mbo radiation scattered across small regions of the genome

After conservative filtering of segments to remove uninformative regions (Table S1), the Barombi Mbo cichlid radiation formed a monophyletic group across 53% of the genome whereas only 0.6% was assigned to phylogenies indicating a polyphyletic Barombi Mbo using the machine-learning *Saguaro* approach. These polyphyletic relationships are consistent with many processes, including secondary gene flow, incomplete lineage sorting, strong selection (balancing, divergent, background), and ancestral population structure. The most prevalent phylogeny spanned 38.2% of the genome and featured the expected species phylogeny for this group, in which all Barombi Mbo individuals formed a single clade with distant relationships to outgroup riverine *S. galilaeus* populations (Figure 1B). The second most prevalent topology (spanning 11.8% of the genome) featured identical evolutionary relationships, except for a much shorter branch leading to *S. galilaeus* Mungo and Meme River. Branch lengths produced by *Saguaro* have no direct interpretation as an evolutionary distance (analogous to a neighbor-joining tree), but may be useful for comparison to similar topologies with different branch lengths, e.g. indicating regions with higher divergence rates (Zamani et al. 2013).

In the 0.6% of the genome supporting polyphyletic Barombi Mbo relationships, we found evidence consistent with multiple colonizations of the lake. Since we were looking for patterns consistent with secondary gene flow or a hybrid swarm for subclades of the radiation, we focused on topologies where single species or entire subclades were more closely related to outgroups than other Barombi Mbo species, which represented only 0.24% of the genome (Figure 2; FigureS2-4). Some topologies featured an entire subclade (e.g. *Stomatepia*) as a monophyletic group more closely related to the riverine populations than other Barombi Mbo species, a relationship consistent with a hybrid swarm scenario before the diversification of the *Stomatepia* subclade (Figure 2A-C). Other topologies featured individual species more closely related to outgroup riverine populations than sister species (Figure 2D), a relationship consistent with secondary gene flow into that lineage after the initiation of divergence.

The general pattern of polyphyletic relationships held after maximum likelihood phylogenetic analyses of a sample of polyphyletic regions initially estimated using *Saguaro*. However, nearly all *Saguaro* topologies in which a single species was more closely related to outgroups differed from the maximum likelihood tree of the same region, which instead featured several Barombi Mbo species as more closely related to the outgroups. The topologies that were consistent across the two analyses and had greater than 85% bootstrap support for the polyphyletic relationship featured entire subclades of Barombi Mbo as more closely related to riverine populations (Figure 2). For example, in *Stomatepia* we found topologies that grouped multiple species with riverine populations (Figure 2A). In the *Konia + Pungu* subclade, we saw a similar pattern in a topology where the entire subclade was sister to the riverine outgroup populations (Figure 2B). In the zooplanktivore *M. myaka*, we did find a topology in which *M. myaka* was sister to the riverine populations (Figure 2C), but also a topology where *M. myaka*, along with all the Barombi Mbo *Sarotherodon* species, were sister to the riverine outgroup populations (Figure 2D).

2.4.2. Genome-wide evidence for differential introgression into the radiation

Consistent with evidence of differential introgression from RADseq data (Martin et al. 2015a), genome-wide f_4 tests provided evidence of genome-wide differential gene flow between some Barombi Mbo sister species and the outgroup riverine species (Table 1). There was significant evidence of genome-wide introgression in tests involving both *S. pindu* in the *Stomatepia* species complex and the hypoxia specialist *K. dikume* in the *Konia + Pungu* subclade. Some species pair combinations within these subclades did not show evidence of differential gene flow, suggesting that there may still be sympatric speciation occurring for some species within the radiation. For example, there was no significant secondary gene flow detected genome-wide in the tests involving sister species *S. mariae* and *S. mongo* or *M. myaka* and *S. linnelli* (Table 1). These results are similar to previous RADseq genome-wide f_4 statistics of the radiation, with significant secondary gene flow detected into Barombi Mbo from tests involving *S. pindu* and *S. linnelli* (Martin et al. 2015a). In the previous study, these tests were only significant when more closely related *Sarotherodon* species from Lake Ejagham were used as outgroups rather than more distantly related riverine populations of *S. galilaeus* from Ghana and the Democratic Republic of Congo. It is possible that the introgression detected previously came from an unsampled or extinct riverine population; more likely, there was introgression from the Cameroon *S. galilaeus* riverine populations examined in this study into the Ejagham *Sarotherodon* species examined previously (Martin et al. 2015a).

We also found evidence for widespread gene flow connecting populations across Barombi Mbo and neighboring riverine populations in highly interconnected population graphs using *TreeMix* (Pickrell and Pritchard 2012); the likelihood of each graph did not plateau until reaching 10 admixture events (Figure 3). On the *TreeMix* population graph with 10 admixture events, gene flow from the Mungo/Meme River populations of *S. galilaeus* occurred directly into individual species *S. mongo* and *K. eisentrauti* rather than the ancestral node of their respective subclades (Figure S5). The proportion of admixture inferred for these two events (0.1% into *S. mongo* and 0.4% into *K. eisentrauti*) was similar to the small proportions of the genome assigned to topologies consistent with secondary gene flow in the *Saguaro* analyses. These admixture events pointing to the tips of the graphs suggest secondary gene flow between nearby riverine populations and individual species within the radiation. In all population graphs allowing up to

21 migration events, any admixture from outgroup riverine populations appears to be coming from the Mungo and Meme rivers rather than the Cross River, consistent with the closer geographic proximity of the former drainages.

2.4.3. Very few genomic regions contain signatures of differential introgression into individual Barombi Mbo species

We characterized heterogeneity in introgression across the genome by calculating f_d statistics in 50-kb sliding windows among the same combinations of Barombi Mbo species used in genome-wide tests to investigate whether differential introgression contributed variation potentially important in the divergence between species. Very few regions of the genome introgressed into single species from *S. galilaeus* MM (Figure 3A-C;S6). Among *Stomatepia* species, only a single candidate introgressed region was found to be differentially introgressed from *S. galilaeus* MM into *S. pindu*, *S. mariae*, and *S.mongo* (50, 80, 80-kb respectively), suggesting very little secondary gene flow after initial diversification of *Stomatepia* began (Table 2). Similarly, secondary introgression was also detected for both *Konia* species, *P. maclareni,* and *M. myaka* (Table 2). However, only 0.0054-0.1% of the genome introgressed into a single species of a Barombi Mbo subclade from *S.galilaeus* MM, in blocks ranging from 50 to 95-kb in size. Most of the introgressed regions determined from f_d statistics overlapped with regions assigned to introgression topologies from *Saguaro* (Figure 2), although these topologies made up only a proportion of each of the introgressed regions (0.09-43.7%; Table S2). This incomplete overlap may be driven by the fixed window size of the f_d statistic test, or the exhaustive genome partitioning in our *Saguaro* approach leading to over segmentation of the topologies in these regions.

In contrast, a larger proportion of the genome appears introgressed from *S. galilaeus* CR based on our coalescent simulation cut-offs (Figure S7). Introgression with *S. galilaeus* CR and a single species of Barombi Mbo was detected in 0.7-5% of the genome. These candidate introgressed regions ranged in size from 50 to 220-kb. The size range of introgressed regions was larger in tests with *S. galilaeus* CR than *S. galilaeus* MM (Figure S8). Larger blocks of introgressed regions could indicate that gene flow from Cross River was more recent than Mungo or Meme River. Absolute genetic divergence between Barombi Mbo species and *S. galilaeus* CR is about 7 times smaller across the genome than between Barombi Mbo and *S. galilaeus* MM (Figure S9-11;Table S3). It is difficult to assess whether all these candidate introgressed regions from *S. galilaeus* CR represent long tracks of introgressed material given that 1) our moderate sequencing depth prevents phasing and 2) the percentage and length of introgression tracts may change based on the timing of gene flow and divergence in this system. Such knowledge will require demographic modeling of different scenarios of gene flow and its timing, which is still computationally intractable for 11 species. Given this, we focus on the candidate introgressed regions involving *S. galilaeus* MM population. We also use the *S. galilaeus* CR introgressed regions that overlap with the *S. galilaeus* MM introgressed regions to look for patterns of secondary gene flow resulting in a hybrid swarm.

2.4.4. Evidence for a hybrid swarm

We found shared introgression signals between species both within and among subclades of Barombi Mbo (Figure 3, Table S4-S5) in which multiple species shared introgressed regions

from a riverine population. This suggests that some of this introgression may have occurred in the ancestral stages of the radiation and differentially sorted among species as they diversified. A majority of these (19/27) were also candidate introgression regions from *S. galilaeus* CR, in which one or more Barombi Mbo species are more closely related to *S. galilaeus* MM while other Barombi Mbo species are more closely related to *S. galilaeus* CR (Table 2). This pattern is suggestive of a hybrid swarm scenario from multiple colonizations by different riverine populations before diversification and the sorting of these ancestral polymorphisms among incipient Barombi Mbo species.

2.4.5. Candidate genes for adaptive diversification within introgressed regions

Although we found evidence of differential introgression among sister species scattered across a small proportion of the genome, the types of genes found in these regions do not paint a clear picture of how this variation may have contributed to speciation (Table 2). For example, differential introgression in *Stomatepia* occurred in regions with genes involved in immune response (Table 2), with no obvious links to the divergent morphological, ecological, or color patterning traits observed among these species (Martin 2012) nor to those traits normally associated with adaptive radiation in cichlid fishes such as body shape, pharyngeal jaw morphology, retinal pigments, or male coloration (Kocher 2004; Barluenga et al. 2006; Wagner et al. 2012; Brawand et al. 2014; Malinsky et al. 2015; Meier et al. 2017). Similarly, in both the *Konia* + *Pungu* and *Myaka* + *Sarotherodon* subclades, introgressed regions were near genes involved in a large range of biological processes not directly associated with adaptive ecological traits in these species, such as *K. dikume*'s hypoxia tolerance, *P. maclareni's* spongivory, and *M. myaka's* zooplanktivory.

The clearest candidate for adaptive diversification showing differential introgression into a single species was found between the shallow-water detritivore *K. eisentrauti* and deep-water hypoxia specialist *K. dikume*. One introgressed region in *K. eisentrauti* contains *ephb4*, a gene involved in blood vessel formation (Herbert et al. 2009; Kawasaki et al. 2014). *K. dikume's* deep water specialization includes higher blood volume with higher concentrations of hemoglobin (Green et al. 1973). This region overlapped with a region of the genome assigned to a polyphyletic Barombi Mbo topology that featured both *Konia* species plus *P. maclareni* sister to riverine outgroups by *Saguaro* (Figure 4B). However, this candidate introgressed region is found within a small scaffold that predominantly exhibits f_d values slightly above our simulation cutoff. False positive f_d values can arise from low nucleotide diversity in a region (Martin et al. 2015b), so even this seemingly convincing candidate gene introgression may be driven by lower genetic diversity in the *K. eisentrauti* sequence compared to *K. dikume* rather than true introgression with *S. galileaus* MM.

From the perspective of gene annotations, there is better support for the contributions of secondary gene flow to the radiation in the introgressed regions that were shared across multiple species of the radiation. For example, a region introgressed into *K. eisentrauti* and *P. maclareni* contains *pafah1b3*, a gene involved in platelet activation activity and spermatogenesis in mice (Prescott et al. 2000; Koizumi et al. 2003; Yan et al. 2003), which while not directly associated with *K. dikume's* traits of higher blood volume with higher concentrations of hemoglobin, is still associated with blood. Another region contains an olfactory receptor gene, *or52e8*. The introgressed region containing *or52e8* has signatures of introgression from *S. galilaeus* MM with *K. dikume* and *M. myaka*, as well as from *S. galilaeus* CR with *K. eisentrauti*, *S. mongo*, and *S.*

mariae. Olfactory signaling is a key component of mate choice and species recognition in cichlids (Plenderleith et al. 2005; Blais et al. 2009). For instance, introduction of olfactory sensory variation from riverine populations may have been important for the divergence of *S. mariae* and *S. pindu*, which appear to exhibit greater sexual than ecological divergence (Martin et al. 2012). Similarly, introgression of olfactory alleles may have played a key role in triggering rapid speciation in Lake Ejagham *Coptodon* cichlids after an 8,000 year waiting period (Poelstra et al. in press).

Finally, two shared introgressed regions contained genes with known skeletal affects: *pfn1* and *mdn1*. The first has been shown to effect skeletal development of long bones of limbs in mice (Bottcher et al. 2009) and cell migration during gastrulation in zebrafish (Lai et al. 2008). Mutations in *mdn1* have known craniofacial effects in zebrafish, causing changes in cranial, mandibular arch, and eye size (Busch-Nentwich et al. 2012; Kettleborgouh et al. 2013). The region containing *mdn1* has signatures of introgression from *S. galilaeus* MM with *P. maclareni* and *S. pindu*, as well as introgression from *S. galilaeus* CR with *K. dikume* and *S. mongo*. Craniofacial traits such as jaw length, head depth, and orbit diameter are often divergent among these Cameroon cichlid species (Martin 2012). For instance, *P. maclareni* has smaller and thicker oral jaws, variation in tooth size, and thicker lips compared to other Barombi Mbo species (Trewavas et al. 1972). *Stomatepia mongo* has a highly elongated snout compared to sister species *S. mariae* and *pindu* (Trewavas et al. 1972; Musilová et al. 2014). Signatures of introgression from different riverine populations in regions containing putative adaptive alleles suggest that secondary gene flow from multiple riverine populations before diversification began (i.e. a hybrid swarm scenario) may have contributed variation important for the radiation.

2.4.6. Weak support for selection on candidate introgressed regions

Within the limitations of our sample sizes for each population in this study, the amount of genetic diversity in introgressed regions does not suggest strong divergent selection on introgressed genetic variation due to selective sweeps. In line with the presence of peaks in f_d values in these regions, between-population diversity (D_{xy}) was typically high between one of the species and its sister species (Figure 4; Table S6). However, within-population diversity across many of these regions was often greater or comparable to scaffold and genome-wide averages (Table S6), suggesting these regions may not have experienced hard selective sweeps that would support their role in adaptive divergence among species. Only a few differentially introgressed regions in *S. pindu* (n=1), *S. mongo* (n=1), and *P. maclareni* (n=1) exhibit genetic diversity an order of magnitude lower than the linkage group average (Table S6), consistent with a selective sweep.

In summary, although we found evidence for differential secondary gene flow between sister species in the radiation, we did not find clear evidence from introgressed genes with well-known roles in adaptive divergence nor strong evidence of selection on these regions. However, these alleles may certainly still serve adaptive functions or may be involved in genetic incompatibilities between species. Furthermore, given our sample size limitations, we cannot yet provide a good estimate of selection strength on these regions.

2.5. Discussion

2.5.1. Equivocal evidence that secondary gene flow promoted the diversification of Barombi Mbo cichlids

Our fine-scale investigations of introgression across the genomes of a celebrated putative example of sympatric speciation are consistent with two possible scenarios: 1) sympatric speciation in the presence of continuous neutral secondary gene flow into the radiation, or 2) speciation initiated by secondary gene flow; our data are not consistent with a scenario involving an extensive period of allopatry and secondary contact promoting speciation. We found little support for the latter allopatric scenario using both *Saguaro* machine learning and sliding-window f_d statistics to exhaustively search for differential introgression into single species. From the *Saguaro* analyses, our most conservative estimate of introgression into a single species within the radiation ranged from 0.013 -0.019% of the genome. Estimates are similarly small for the f_d statistics, ranging from 0.0054-0.1% of the genome from *S. galilaeus* populations in the neighboring Mungo and Meme Rivers. Furthermore, even these significant outliers may represent false positives, particularly in the case of introgression with *S. galilaeus* from Cross River, where 0.7-5% of the genome appears to have introgressed from this population. More complex coalescent simulations, larger sample sizes, and knowledge of the timing of diversification and gene flow events are needed to accurately assess the signatures observed. It is also difficult to distinguish signatures of differential introgression from the biased assortment of ancestral polymorphism into modern lineages, e.g. a hybrid swarm scenario that is still consistent with sympatric divergence entirely within the crater lake. Finally, even if our statistical outliers represent differentially introgressed regions, their importance to the speciation process is equivocal. Within the limitations of our sample sizes, we found no evidence of selective sweeps in most of these regions to suggest they aided in divergence between species and most regions contained housekeeping genes that do not clearly suggest how introgressed variation would have contributed to the radiation. Nonetheless, these genes may still serve an adaptive function or could be involved in genetic incompatibilities between diverging sympatric populations, possibly introduced by an initial hybrid swarm in the lake (Seehausen 2004, 2013; Schumer et al. 2015).

2.5.2. Evidence for a hybrid swarm further complicates the role of gene flow in the speciation process

Beyond speciation scenarios involving secondary gene flow aiding the completion of speciation, our findings are suggestive of another scenario for sympatric speciation in this system: sympatric speciation from an initial hybrid swarm resulting from the differential sorting of ancestral polymorphisms among incipient species (e.g. see Figure1 in Martin et al. 2015a). A hybrid swarm is not easily detectable using the f_d statistic because introgressed variation could be shared among diverging sister species, producing an f_d value near zero due to a BBBA rather than ABBA pattern (Reich et al. 2009; Patterson et al. 2012). However, some of the f_d peaks appear to be 1) shared across at least two of the sister species in a subclade, 2) shared between species of different subclades, or 3) contain variation from both riverine populations (Mungo/Meme and Cross Rivers) that appears differentially sorted among sister species. All three of these patterns are consistent with an ancestral hybrid swarm before divergence between sister species occurred.

This pattern of differential sorting of variation from a hybrid swarm from f_d analyses could also result from a lack of power in this statistic to distinguish the directionality of the introgression detected in those regions when using biallelic patterns and four populations (e.g. when two populations share similar allele patterns, the other two populations can share the opposite allele pattern by default). However, we also found evidence that entire subclades (e.g. *Stomatepia*) were more closely related to riverine populations than other Barombi Mbo subclades using our *Saguaro* analyses, a pattern that is also consistent with a hybrid swarm (e.g. Figure2).

It is in these regions that we find the best candidate genes for secondary gene flow contributing to diversification in this system. These candidates are known to affect important traits in other adaptive radiations of cichlids, including olfactory signaling (Nikaido et al. 2013, 2014; Azzouzi et al. 2014; Keller-Costa et al. 2015) and pharyngeal jaw morphology (Muschick et al. 2011; Brawand et al. 2014; Malinsky et al. 2015). These findings are similar to studies on other systems using similar approaches which found compelling cases for adaptive introgression contributing to diversification (e.g. (Abi-Rached et al. 2011; The Heliconius Genome Consortium et al. 2012; Huerta-Sánchez et al. 2014; Lamichhaney et al. 2015; Stankowski and Streisfeld 2015; Arnold et al. 2016; Meier et al. 2017a)), including our own previous work (Richards and Martin 2017a). For example, several studies have found convincing candidate genes/variants in introgressed regions suggesting that adaptive introgression played a role in shaping ecological and morphological diversity. These include the detection of introgressed alleles linked to wing-color patterning involved in mimicry and mate selection in *Heliconius* butterflies (The Heliconius Genome Consortium et al. 2012), flower coloration involved in pollinator preferences for *Mimulus* species (Stankowski and Streisfeld 2015), and oral jaw size variation involved in scale-eating trophic specialization in *Cyprinodon* pupfishes (Richards and Martin 2017a).

2.5.3. The challenges of supporting or rejecting a role for secondary gene flow from tests of differential introgression

There are some caveats to our interpretations of secondary gene flow and its functional role in the ecological and morphological diversity observed within the lake. Recombination rate varies across genomes and determines the scale over which patterns of admixture and differentiation vary (Smukowski and Noor 2011). In our fixed sliding window size of 50-kb, we may have missed important patterns of introgression in regions of recombination hotspots, where such patterns are expected to be very localized (Schumer et al. 2018). Shared variation among species may reflect unsorted polymorphisms from structured ancestral populations rather than hybridization. Introgression events can also be hard to distinguish from ongoing balancing selection of ancestral polymorphism that is sieved between species (Guerrero and Hahn 2017).

Furthermore, while we focused on searching for genetic signatures of hard selective sweeps, introgressed regions with intermediate to high nucleotide diversity may have undergone soft selective sweeps, in which selection drives multiple adaptive haplotypes to fixation. Some of these introgressed regions may have been adaptive and undergone soft selective sweeps, although the relative contributions of hard sweeps versus soft sweeps during adaptation and speciation is still the subject of much debate (Hermisson and Pennings 2005, 2017; Pritchard et al. 2010; Jensen 2014; Schrider et al. 2015).

Although a search for candidate genes with plausible roles in speciation when considering whether introgressed regions contributed to divergence is a good starting point, there

are some caveats to ruling out the importance of secondary gene flow based solely on gene annotations. Not all the effects of a gene may be known, even for model organisms, and there is growing support for an omnigenic model of the link between genotype and phenotype, where all genes expressed in relevant cells can potentially influence a trait (Boyle et al. 2017). Likewise, not all traits involved in species divergence are known and more cryptic traits (e.g. metabolism, physiology) could be as important as the more obviously divergent morphological traits. Understanding the traits and genes involved in speciation is becoming the difficult problem of functional genetic analyses for anything sweeping within a population (see also Richards et al. 2018).

Even if there are no coding regions present in the introgressed region, it may still contain important regulatory elements that affect genes in other regions which underlie speciation traits. It could also house variants that cause genetic incompatibilities and provide a source of postzygotic reproductive isolation. Divergence caused by genetic incompatibilities may not leave signatures of hard selective sweeps in the genome either. One prediction of the hybrid swarm hypothesis is that some of the novel allele combinations introduced by hybridization could include segregating genetic incompatibilities (Seehausen 2004, 2013; Abbott et al. 2013). The presence of DMIs has been shown in other African cichlids (Stelkens et al. 2010) , but little is known about what specific regions of the genome contribute to these patterns. Detecting candidate regions of DMIs in non-model organisms is a currently growing body of work mainly investigated in relatively recent hybrid zones (Schumer et al. 2014a, 2015, 2018), but is a promising avenue of future research, particularly given the support for a potential hybrid swarm in the evolutionary history of Barombi Mbo shown here.

While the limitations of the methods used in this study make it hard to definitively support or reject a role for gene flow in the adaptive radiation of Barombi Mbo, we highlight some of the gaps in knowledge that need to be filled before we can understand what role the observed secondary gene flow played in diversification. A more thorough understanding of the evolutionary history of the group (e.g. the timing and duration of gene flow events, divergence times among species) will be useful in determining the amount of introgression we would expect to see in the contemporary genomes of the species under different speciation scenarios. A thorough investigation of all the regions of the genome that have undergone soft and hard selective sweeps and the timing of those sweeps, alongside a search for genetic incompatibilities among species, will make it easier to understand the functional importance of observed introgressed regions to diversification. The demographic analyses needed for a radiation of 11 species are still largely intractable, but will eventually provide much needed insights on the role of the secondary gene flow observed in this classic putative example of sympatric speciation. However, some recently proposed statistics that infer the direction and relative timing of gene flow to speciation provide a promising future avenue for distinguishing between different scenarios of introgression (e.g. hybrid speciation vs. introgression after initial divergence; (Hibbins and Hahn 2018) without extensive demographic modeling.

2.5.4. Best and worst remaining cases for sympatric speciation within the Barombi Mbo cichlid radiation

While the radiation as a whole may not have entirely arisen in sympatry, some sister species within Barombi Mbo may be better case studies of the process than others. Within the three-species *Stomatepia* subclade, there is little evidence that secondary gene flow played an

important role in speciation. On a genome-wide level, we detected secondary gene flow for *S. pindu* and *S. mariae,* but not *S. mongo* (Table 1). This is intriguing given that a previous study found continuous unimodal morphological variation connecting the *pindu/mariae* species complex (Martin 2012), but did not include the rare and morphologically distinctive *S. mongo* (Musilová et al. 2014). Furthermore, while *pindu/mariae* species complex appear unimodal across ecological phenotypes and trophic axes, they do show bimodality along axes of color/pattern, indicating that disruptive sexual selection on coloration may be the primary driver and initial barrier to reproduction for this group (Martin 2012). This provides a category of functional traits to search for when trying to determine if secondary gene flow played a role in their current level of divergence.

No reproductive isolating barriers or genetic incompatibilities have yet been quantified among any Barombi Mbo species (in contrast to the strong assortative mating by size, color, morphology, and diet documented in Lake Ejagham *Coptodon* species (Martin 2013)). *S. pindu* and *S. mariae* can produce viable F1 hybrids in a no-choice laboratory environment (CHM pers. obs.); however, all Barombi Mbo species appear to avoid hybridizing when housed in mixed-species laboratory aquaria and no hybrid courtship has been observed in the field (CHM pers. obs.). Estimates from scale growth rings suggest only moderate disruptive selection on ecological traits in the *Stomatepia pindu/mariae* species pair. These species were also not significantly different in their dietary source of carbon or relative trophic position from stable isotope analyses (Martin 2012) and their diets are only marginally distinct based on stomach content analysis, containing predominantly insect/shrimp prey in *pindu* and insect/fish prey in *mariae* (Trewavas 1972; CHM unpublished data). Very little is known about reproductive isolating barriers in other Barombi Mbo cichlids, aside from the interesting and unique temporal isolation of *Myaka myaka* due to its seasonal lekking behavior (CHM pers. obs.).

On a fine scale, the few introgressed regions unique to each of the *Stomatepia* species contained only immune response genes. Shared signatures of introgression among two of the three species or with other Barombi Mbo species represented a larger proportion of the genome than differentially introgressed regions within each species, although both types of introgression were rare across the genome. In those regions that introgressed in more than one *Stomatepia* species, but from different riverine populations, we found genes with functions more typically associated with cichlid radiations. Even for the two monotypic specialist species *M. myaka* and *P. maclareni*, introgressed regions found solely in these species contained mainly housekeeping genes, suggesting secondary gene flow may have been neutral in the evolution of their trophic specializations. However, they also share signatures of introgression in the same regions as *Stomatepia*, suggesting that these monotypic specialists may have obtained variation for their specialized traits from sorting of ancestral polymorphisms within a hybrid swarm before the radiation began.

Among all the ecologically divergent species pairs focused on in this study, *K. eisentrauti* and *K. dikume* are the least convincing as a putative example of sympatric speciation between sister species. Differentially introgressed regions between *K. dikume* and *K. eisentrauti* include a region containing *ephb4*, involved in heart and blood vessel development. Given *K. dikume*'s hypoxia specialization, this region is the best potential candidate in this study for secondary gene flow and an allopatric phase of speciation contributing to diversification between species in the radiation. These two species also exist in microallopatry; *K. eisentrauti* is an abundant detritivore found only along the shallow littoral region of the lake whereas *K. dikume* is a deep-water specialist on *Chaoborus* midge larvae which have only been collected in deep-water gill nets

(Trewavas et al. 1972; Schliewen et al. 1994; CHM pers. obs.). Both species are mouthbrooders and likely breed in non-overlapping habitats although nothing is known about the breeding habits of *K. dikume*.

2.5.5. Conclusion

The complex history of colonization in the Barombi Mbo crater lake cichlid radiation found in this and a previous genome-wide study suggests that secondary gene flow may have played a role in the speciation process, which violates one of the strict criteria for demonstrating sympatric speciation in the wild (Coyne and Orr 2004a). Our fine-scale dissection of introgressed regions across the entire genome suggests that supporting or rejecting a role for secondary gene flow in speciation will require an understanding of the functional alleles within each region and their evolutionary history. Nonetheless, we can rule out a scenario in which extensive secondary gene flow after a long allopatric phase, such as reinforcement, contributed to diversification in any Barombi Mbo species. Instead, small and scattered amounts of secondary gene flow that were differentially sorted among these sympatric species may have provided variation with undiscovered functional effects on the divergent ecologies and morphologies seen in the lake or this gene flow was predominantly neutral with respect to its role in the speciation process. We found more convincing evidence that secondary gene flow contributed adaptive variation during an initial hybrid swarm within the lake and later sorting of that variation among species in sympatry. Disentangling the effects of a putative hybrid swarm from secondary contact on the speciation process will require a better understanding of the timing of gene flow events compared to the diversification times of Barombi Mbo species. We found evidence for gene flow into the radiation both before and after initial diversification of subclades within the lake. Even without this information, equivocal support for a functional role of secondary gene flow in the radiation of Barombi Mbo cichlids suggests that we should not rule out the possibility of sympatric speciation in this system just yet.

2.6. Figures

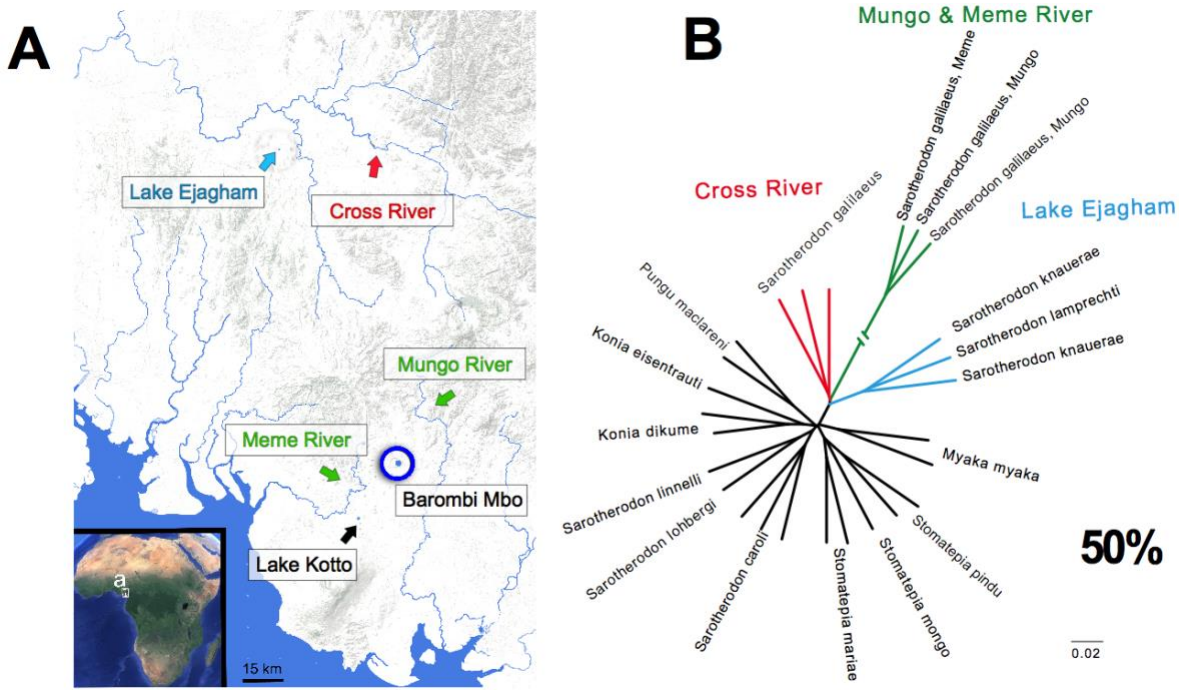


Figure 1. **The predominant phylogenetic relationship among Barombi Mbo radiation and neighboring riverine populations**. A) Map of lakes and riverine drainages within the volcanic belt of Cameroon in the Northwest and Southwest provinces (Ambazonia). Modified from Figure 2 in Martin et al. (2015a). B) The topology assigned to the largest percentage of the genomes across Barombi Mbo radiation and outgroup *Sarotherodon* species. Across most of the genome Barombi Mbo species (black) are more closely related to each other than riverine outgroup populations of *S. galilaeus* Mungo and Meme River (green) and *S. galileaus* Cross River (red), or the Lake Ejagham *Sarotherodon* radiation (blue).

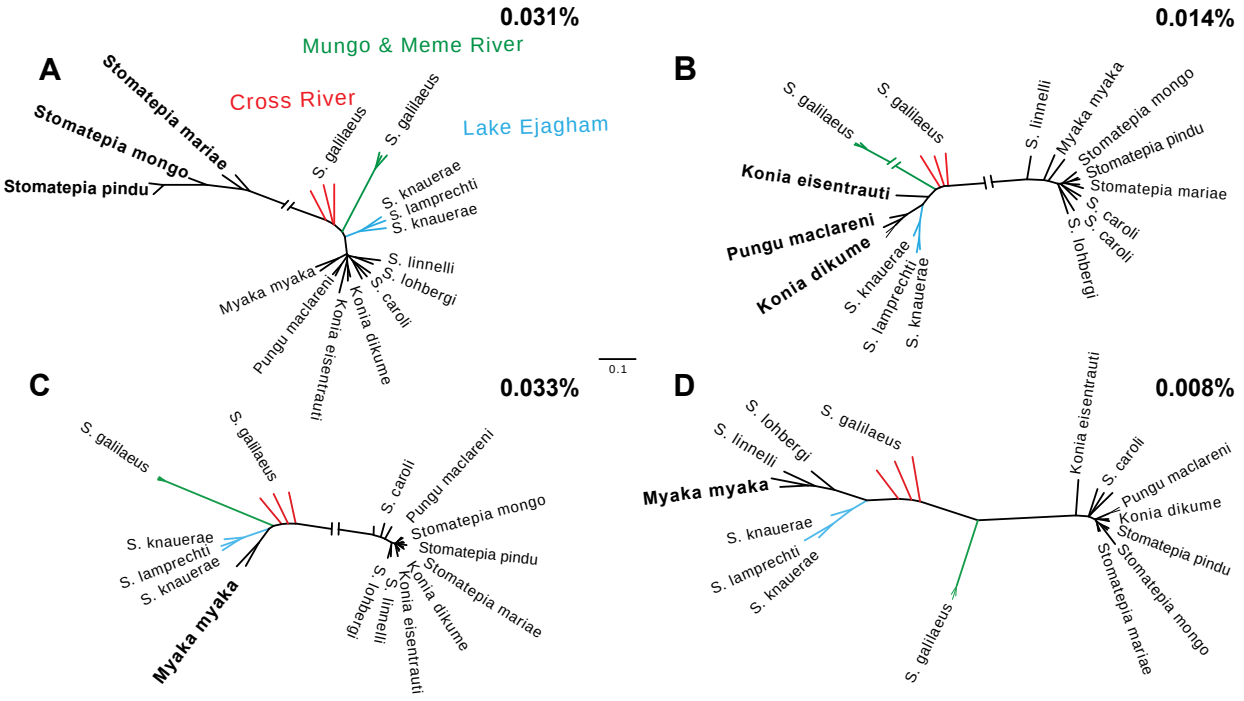


Figure 2. **Topologies featuring Barombi Mbo polyphyly with riverine populations involving the Barombi Mbo species with unique ecologies.** Across small and independent proportions of the genome A) the entire *Stomatepia* clade, B) only *S. pindu*, C) *M. myaka* and *Sarotherodon* species and D) only *M. myaka* were more closely related to outgroups than other Barombi Mbo species. These topologies are consistent with introgression between outgroups and Barombi Mbo and are supported by maximum likelihood analyses. Percentages indicate proportion of the genome assigned to these topologies.

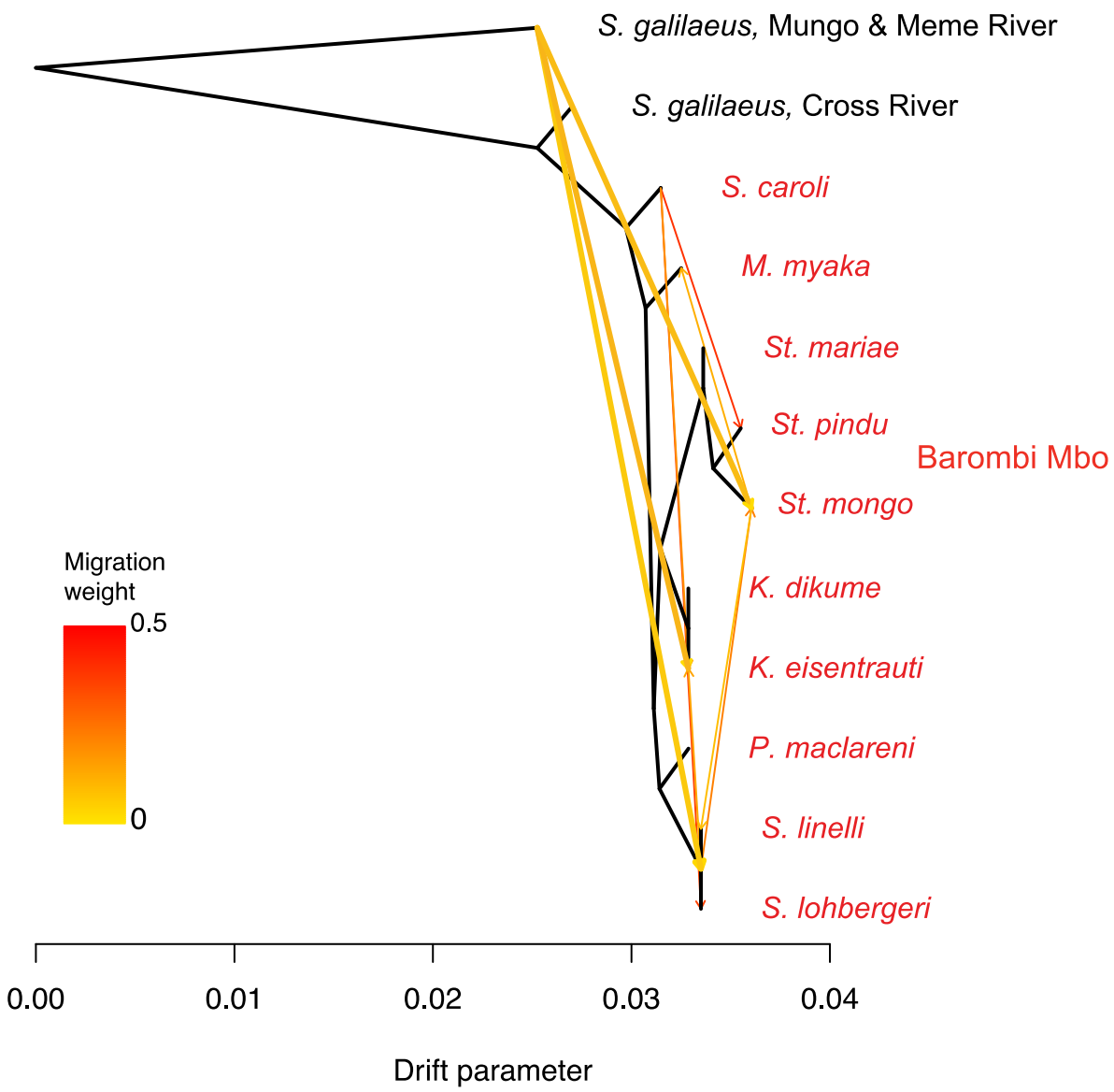


Figure 3. Visualization of genome-wide introgression from riverine *Sarotherodon* populations into Barombi Mbo radiation. *Treemix* graph illustrating 10 admixture events (with heat colors indicating intensity) on a population graph of the radiation. Admixture events from riverine populations into the radiation are indicated with thicker arrows.

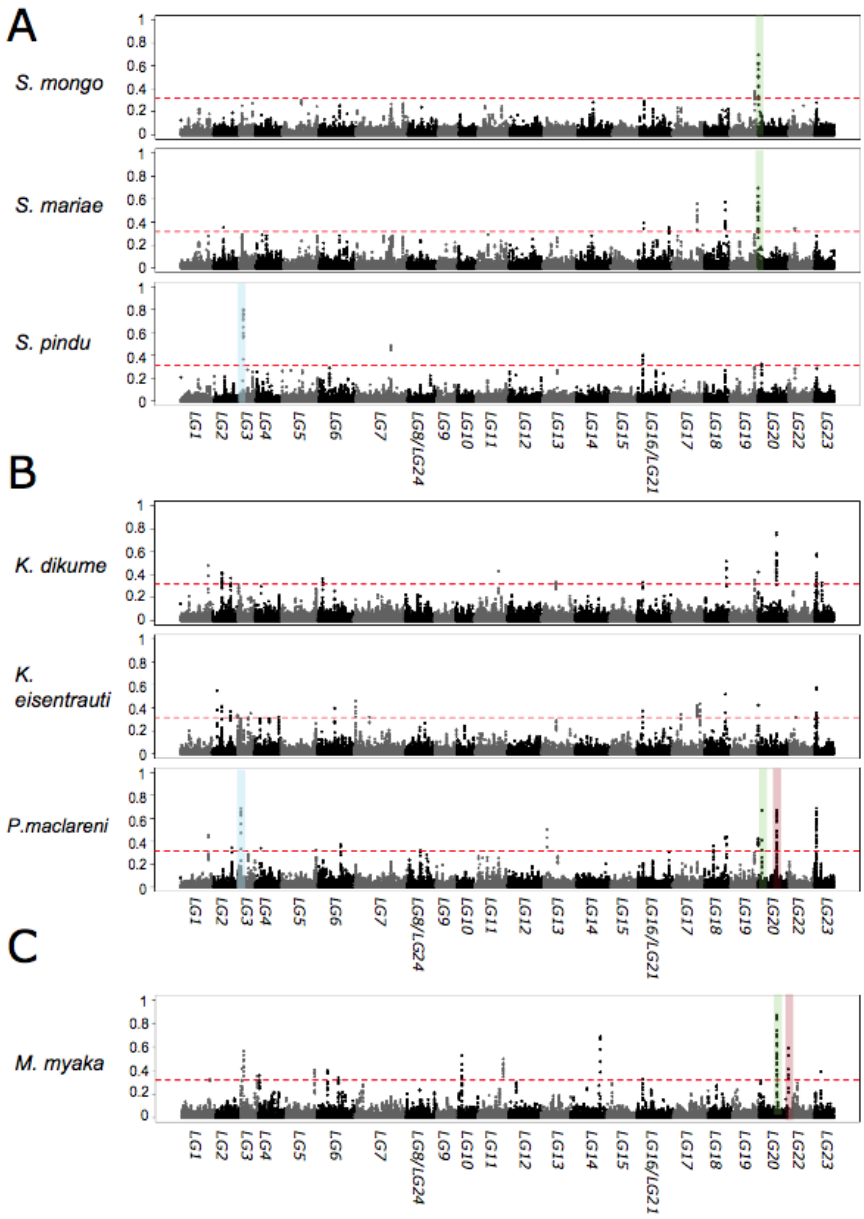


Figure 4. Manhattan plots of f_d values between riverine populations of *S. galilaeus* from Mungo and Meme river and A) the *Stomatepia* species, B) combinations of the three species in the *Konia* and *Pungu* species and C) *Myaka myaka.* Alternating gray/black colors indicate different linkage groups. Dotted red lines mark the coalescent simulation-based significance thresholds for each test (F_d = 0.315). Peaks highlighted in colors represent those signals of introgression shared across different subclades. Manhattan plots for the scaffolds not assigned to the 24 linkage groups are presented in FigureS8.

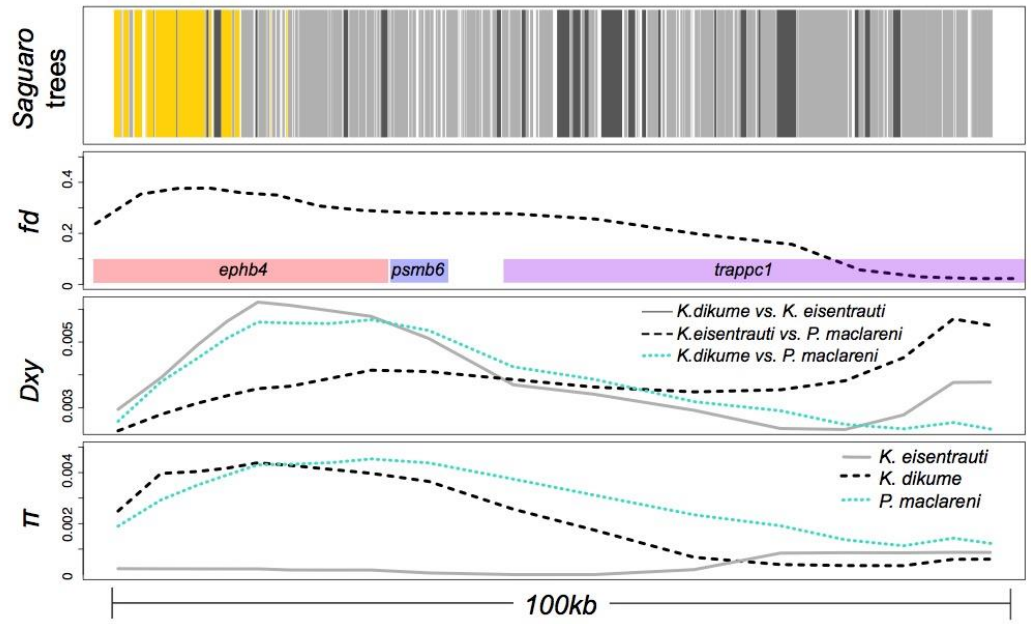


Figure 5. Candidate adaptive introgression region in the *Konia* species pair containing gene *ephb4*. Row 1 shows the topologies assigned by *Saguaro* to the region. Gold blocks were assigned to a topology featuring the *Konia* species pair plus *Pungu maclareni* sister to the riverine *S. galilaeus* populations (Figure 2B), light grey blocks were assigned to the predominant monophyletic topology (Figure 1B), and dark grey blocks were assigned to any other topology. Row 2 shows the peak signal of introgression across scaffold NT_167586.1 detected from the f_d statistic across the three-test combination involving *K. dikume* and *K. eisentrauti* and riverine populations of *S. galilaeus* MM in overlapping 50-kb windows and the genes in this peak (*ephb4*, *psmb6*). Row 3 shows between-population divergence (D_{xy}) among the combinations of *Konia* and *Pungu* species calculated in overlapping 50-kb windows. Row 4 shows within-population diversity (π) in the same non-overlapping 50-kb windows for the species of *Konia* and *Pungu*. The data from Row 2-4 were smoothed using the function smooth.spline in R with a spar of 0.1 for ease of visualization in the figure.

2.7. Tables

Table 1. **Genome-wide f_4 statistics supporting differential introgression within Barombi Mbo radiation.** Tests with significant evidence for differential introgression are highlighted in bold. The f_4 statistic was calculated for pairwise combinations among sister species of Barombi Mbo subclades (A, B) and riverine populations of *S. galilaeus* from the Mungo and Meme Rivers (MM) and Cross River (CR).

Introgression with riverine outgroups: (A,B) ← → (*MM, CR*)	**f_4 statistic**	***Z*-score**	***P*-value**
S. mariae, S. mongo	$-2.04\text{x}10^{-7} \pm -5.15\text{x}10^{-7}$	-0.39	0.69
S. mariae, S. pindu	$\mathbf{-1.92\text{x}10^{-6} \pm -4.48\text{x}10^{-7}}$	**-4.29**	$\mathbf{1.8\text{x}10^{-5}}$
S. mongo, S. pindu	$\mathbf{-1.59\text{x}10^{-6} \pm -4.98\text{x}10^{-7}}$	**-3.19**	**0.0014**
K. dikume, K. eisentrauti	$\mathbf{-2.4\text{x}10^{-6} \pm -6\ \text{x}10^{-7}}$	**-4.01**	$\mathbf{6.3\text{x}10^{-5}}$
K. eisentrauti, P. maclareni	$-2.12\text{x}10^{-7} \pm -6.15\text{x}10^{-7}$	0.35	0.73
K. dikume, P. maclareni	$\mathbf{-2.56\text{x}10^{-6} \pm -5.86\text{x}10^{-7}}$	**-4.37**	$\mathbf{1.2\text{x}10^{-5}}$
M. myaka, S. linnelli	$-4.04\text{x}10^{-7} \pm -7.11\text{x}10^{-7}$	0.56	0.57

Table 2. **Candidate introgressed regions in Barombi Mbo cichlid radiation**. These regions feature significant f_d values between riverine populations of *S. galilaeus* (MM: Mungo and Meme River; CR: Cross River) and the three subclades of the radiation focused on in this study. Unannotated regions with no GO terms are marked with (-).

Linkage Group	Position	Gene(s)	Gene Ontology Terms
		Myaka myaka	
LG3	17560001-17615000	***fcgr2b;fcgr1***	IgG binding,immunoglobin mediated immune response; IgG binding, phagocytosis
LG4	800001-855000	***ifi44***	defense response to virus; GTP binding
LG5	35530001-35595000	***matn4;rbpjl***	growth plate cartilage chondrocyte morphogenesis; transcriptional activitor activity, RNA polymerase II proximal promotor
LG6	10670001-10740000	***prf1;actb***	immunological synapse formation, wide porin chanel activity;dense body,focal adhesion
LG6	20935001-21025000	***jpt2***	cytosol; plasma membrane
LG11	28280001-28375000	***iqgap3;ttc24;gnrh2;igdcc3;polr3gl;hfe***	calmodulin binding, Ras protien signal transduction; biological process; gonadotropin hormone-releasing activity, reproduction;neuromuscular process controlling balance ; nuclear chromatin DNA-directed RNA polymerase;iron ion transport
LG20	1705001-1755000	***atp6ap1;taz***	Rab GTPase binding; positive regulation of bone resorption;O-acyltransferase activity
LG20	19570001-19630000	***sbk1***	protein serine/threonine kinase activity
NT_167475.1	450001-500000	***ier5l;h2dmb1;plgrkt***	biological process;ribonuclease H2 complex; positive regulation of plasminogen activation
NT_167500.1	930001-1000000	***cldn15; il20rb***	cell-cell junction assembly,transforming growth factor beta receptor signaling pathway;interleukin-10 receptor activity
NT_167568.1	10001-60000	***zfp235***	DNA binding, transcription activity
NT_167617.1	320001-400000	***zg57;nebl***	unknown; cardiac muscle thin filament assembly
NT_167716.1	220001-270000	**--**	--
NT_167790.1	255001-270000	**--**	--
NT_167636.1	140001-190000	--	--

Pungu maclareni			
LG3	14525001-14575000	***plac8***	negative regulation of apoptotic process, chromatin binding
LG17	9260001-9315000	***gimap8***	GTPbinding, regulation of T cell apoptotic process
NT_167557.1	1410001-1495000	***pgbd4*;UPC**	DNA binding transcription factor activity
NT_167671.1	265001-330000	***wsc3*; UPC(2)**	Rho protein signal transduction
NT_167747.1	45001-330000	***nlrc3;dclk2***	negative regulation of I-kappaB kinase/NF-kappaB signaling;peptidyl-serine phosphorylation
NT_168010.1	30001-80000	--	--
Konia dikume			
LG2	18660001-18710000	***hmcn1***	basement membrane; fin morphogenesis
NT_168013.1	1-70000	***siglec12;* UPC(2)**	Cell adhesion, carbohydrate binding;--
Konia eisentrauti			
LG6	22460001-22515000	***pitpnc1; vwa7***	phosphatidylinositol transporter activity; extracellular region, biological process
NT_167586.1	220001-290000	***trappc1;psmb6;ephb4****	TRAPP complex;threonine-type endopeptidase activity; ephrin receptor signaling pathway, heart morphogenesis
NT_167663.1	535001-630000	***s100p;s100g; slc29a2*; UPC**	RAGE receptor binding;apical plasma membrane, transition ion binding;nucleoside transmembrane transporter activity;-
Stomatepia mongo			
LG20	150001-230000	***plod3;cyb5d;clec10A***	peptidyl-lysine hydroxylation,procollagen-lysine 5-dioxygenase activity;metal ion binding;carbohydrate binding, adaptive immune response
Stomatepia mariae			
LG16-21	31140001-31220000	***cxcr2;faim; parp9;parp14***	nuetrophil chemotaxis; apoptotic process; positive regulation of interferon-gamma-mediated signaling pathway; negative regulation of interferon-gamma-mediated signaling pathway
Stomatepia pindu			
NT_167675.1	605001-655000	***cmklr1;c5ar1***	G-protein coupled receptor signaling pathway, chemotaxis; complement component C5a receptor activity

*best candidate region for secondary gene flow contributing to diversification; UPC = uncharacterized protein coding gene

2.8. Supplementary Materials

2.8.1. Supplementary Methods

2.8.1.1. *Sampling and Genome Sequencing*

We sequenced whole genomes of 1-3 individuals from 10 out of the 11 species within the sympatric radiation of Oreochromini cichlids in Cameroon crater lake Barombi Mbo (excluding *Sarotherodon steinbachi* which is morphologically and ecologically similar to the other three *Sarotherodon* species), an endemic *Sarotherodon* species pair from Lake Ejagham (*S. lamprechti* and *S. knauerae*), and outgroup *Sarotherodon galilaeus* individuals from all three river drainages flanking the lake: Cross, Meme, and Mungo rivers (e.g. see map in (Schliewen et al. 1994); Figure 1A). Individual cichlids were caught by seine, gill net, or hook-and-line from Barombi Mbo, Lake Ejagham, Cross River, Mungo River, and Meme River in January, 2010 and July, 2016. 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) and University of North Carolina Animal Care and Use Committee (#15-179.0). Whole specimens or tissue samples were stored in 95-100% ethanol or RNAlater (Ambion, Inc.) in the field.

DNA was extracted from muscle tissue using DNeasy Blood and Tissue kits (Qiagen, Inc.) and quantified on a Qubit 3.0 fluorometer (Thermofisher Scientific, Inc.). Genomic libraries were prepared using the automated Apollo 324 system (WaterGen Biosystems, Inc.) at the Vincent J. Coates Genomic Sequencing Center (QB3) at the University of California, Berkeley. Samples were fragmented using Covaris sonication, barcoded with Illumina indices, and quality checked using a Fragment Analyzer (Advanced Analytical Technologies, Inc.). Nine to twelve samples were pooled in four different libraries for 150PE sequencing on four lanes of an Illumina Hiseq4000.

1.9 billion raw reads were mapped from 28 individuals to the *Oreochromis niloticus* reference genome v1.1 (NCBI, total sequence length = 927,679,487; number of scaffold = 5,909, scaffold N50, = 2,766,223; contig N50 = 29,493) with the Burrows-Wheeler Alignment Tool (BWA-MEM v 0.7.15) (Li and Durbin 2009; Li 2013). Duplicate reads were identified using MarkDuplicates and BAM indices were created using BuildBamIndex in Picard Tools (v 2.10.3, http://broadinstitute.github.io/picard). Average coverage across individuals was 10.39 X (range: 6.19-17.67 X). We followed the best practices guide (Van der Auwera et al. 2013) recommended for the Genome Analysis Toolkit (v 3.5) (DePristo et al. 2011) to call and refine our SNP variant dataset using the program HaplotypeCaller. Because we lacked high-quality known variants for these non-model species, we filtered SNPs based on the recommended hard filter criteria (i.e. QD < 2.0; FS < 60; MQRankSum < -12.5; ReadPosRankSum < -8) (DePristo et al. 2011; Marsden et al. 2014) using vcftools "—remove-filtered-all" flag (v 0.1.14) (Danecek et al. 2011a). We further filtered SNPs with a minor allele frequency of less than 0.05, a genotype quality score less than 20, a depth of coverage less than 5, and more than 10% missing data in vcftools.

2.8.1.2. Characterization of introgression patterns across the genome with SAGUARO

First, we exhaustively searched the genomes for patterns of non-monophyletic Barombi Mbo relationships using the machine learning program SAGUARO (Zamani et al. 2013) to identify regions of the genome that contained relationships consistent with expectations from multiple colorizations and secondary gene flow into the radiation (i.e. paraphyletic/polyphyletic Barombi Mbo radiations). Saguaro combines a hidden Markov model with a self-organizing map to characterize variation in phylogenetic relationships among individuals across the genome without requiring *a priori* hypotheses about these relationships or the size of genomic regions. This method infers relationships among individuals in the form of genetic distance matrices and assigns segments across the genomes to different topologies. These genetic distance matrices can then be transformed into neighborhood joining trees to visualize patterns of evolutionary relatedness across the genome. We exhaustively searched the genome for topological variation by partitioning the genome into a total of 75 unique topologies (well past the inflection point at 30 topologies where the percent of genome explained by each additional topology plateaus; Figure S1).

Since smaller segments with fewer informative sites are more likely to be incorrectly assigned to a hypothesized topology by chance, we tested various minimum SNP filters (1, 10, or 20 SNPs) for reducing the amount of short uninformative segments and their effect on the percentage of the genome assigned to topologies. We found that while the percentage of the genome assigned to topologies changes when we apply SNP filters, none of the topologies had all segments entirely removed and the relative proportions of the genome assigned to particular types of topologies were similar across filtering strategies (Table S1). This may be due to uninformative sites being assigned to topologies in a random fashion, such that removing these sites does not change the percentage of the genome disproportionately across topologies. The percentages indicated in the results represent the percentage of base pairs in the genome assigned to topologies after using a 20 SNP minimum filter and represent conservative estimates.

We searched these 75 topologies for evidence of relationships where subclades or individual Barombi Mbo species were more closely related to riverine populations than other species in the crater lake, suggesting sympatric speciation after a hybrid swarm (i.e. differential sorting of ancestral polymorphism) or secondary gene flow into this subclade (introgression). For each of these topologies, we checked if they matched the maximum likelihood topology generated using RaxML from an alignment of the region assigned to the topology by *Saguaro* that contained the most SNPs and a GTRGAMMA model of sequence evolution. Topologies which were not consistent with the maximum likelihood trees with more than 85% bootstrap support were removed from further comparisons to results from other approaches used in this study (e.g. overlap with sliding window tests for introgression).

2.8.1.3. Characterization of introgression patterns across the genome using sliding windows of fd

We characterized heterogeneity in introgression across the genome among these same combinations and investigated whether differential introgression contributed variation potentially important in the divergence between species by calculating f_d statistics in 50-kb sliding windows using a custom python script (modified from ABBABABA.py created by Simon H. Martin, available on https://github.com/simonhmartin/genomics_general; our modified version is provided in the supplementary materials of (Richards and Martin 2017a)). We used the

population allele frequencies of biallelic SNPs, allowing for a minimum of 100 variant sites with no missing data within a population per site. Linkage disequilibrium was calculated pairwise among SNPs across the largest linkage group in the genome (LG 3) using the "--r2" function in plink (Purcell et al. 2007). 50-kb is roughly the point at which linkage disequilibrium among all the groups in this dataset decays below a r^2 value of 0.2 (Figure S12).

Significant regions of introgression were determined by calculating f_d statistic across 150,000 windows containing 250 variants (average number of used variants in empirical windows) simulated under a coalescent model with no gene flow. Empirical windows were considered candidate introgressed regions if the f_d statistic was above the maximum simulated f_d value. Windows of variants were simulated for four populations using ms-move (Garrigan and Geneva 2014) and information about historical changes in population sizes were inferred from MSMC analyses (Schiffels and Durbin 2014). Simulations were run as follows: "msmove 8 1 -I 4 2 1 3 2 -g 1 40 -g 2 40 -g 3 1.59 -g 4 3.9 -eg 0.25 1 0 -eg 0.25 2 0 -eg 1.8 3 0 -eg 1 4 0 -en 2.5 1 0.1 -en 2.5 2 0.1 -en 6 3 0.5 -en 12.5 4 0.5 -T" for tests involving *S.galilaeus* MM and "msmove 8 1 -I 4 2 1 3 2 -g 1 40 -g 2 40 -g 3 1.59 -g 4 3.9 -eg 0.25 1 0 -eg 0.25 2 0 -eg 0.1 3 0 -eg 1 4 0 -en 2.5 1 0.1 -en 2.5 2 0.1 -en 0.4 3 0.167 -en 12.5 4 0.5 -T" for tests involving *S.galilaeus* CR . The simulations were then converted to sequence data using seq-gen and GTR model. These sequences were then inputted into the *fd* statistic pipeline

We ran MSMC on unphased genotypes from the 100 largest scaffolds (size range 0.86 Mb – 51.0 Mbp) for each individual separately. Genotypes were called from bam files using *samtools mpileup* with the following settings: minimum mapping quality of 20 (flag "-q 20"), minimum genotype quality of 20 (flag "-Q 20") and a coefficient for downgrading mapping quality for reads containing excessive mismatches of 50 (flag "-C 50"). As recommended in the MSMC documentation (https://github.com/stschiff/msmc-tools), we used individual-specific masks to mask sites with less than half or more than double the mean coverage for that individual. Nadachowska-brzyska et al. (2016) recommend to only use individuals with a mean coverage of at least 18, yet all our individuals were sequenced at a lower depth. After excluding individuals with depth below 7.5, (the *S. linelli* and *S. mongo* individuals), the genotypes used in this analysis had depths ranging from 7.52 to 16.67 (mean 9.97) across Barombi Mbo individuals, and genotypes for outgroup individuals had depths ranging from 8.10 to 16.97 (mean 11.75). While our MSMC results should therefore be interpreted with caution, the consistency among most individuals of the same species (Figure S13-S14) suggests that the general patterns of the analysis are likely to be robust.

2.8.2. Supplemental Figures

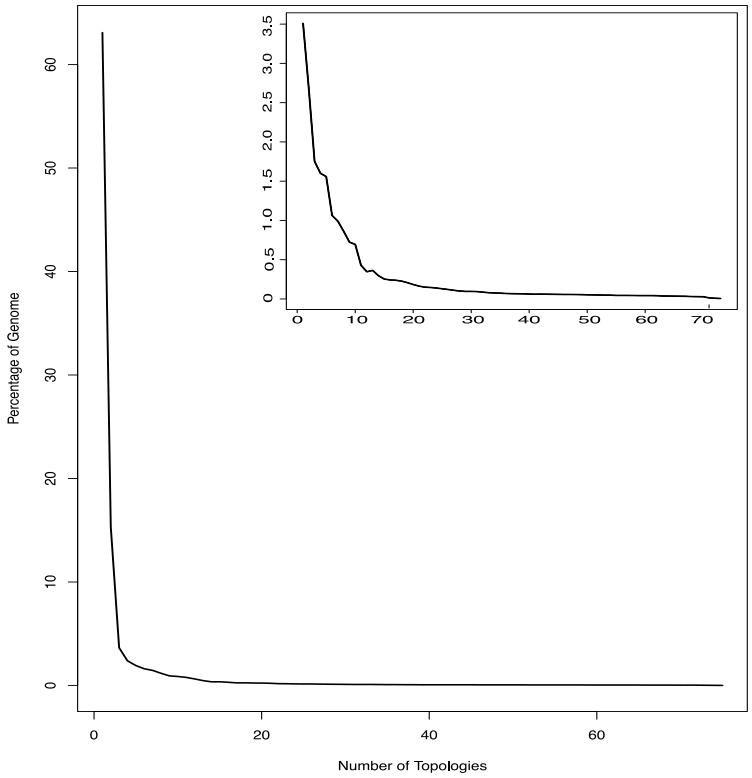


Figure S1. The percentage of the genome assigned to each topology by SAGUARO. The insert is a closer look at all but the two topologies assigned to the largest percentage of the genome (65% and 15%). This suggests saturation in the variance explained by topologies at around 20 proposed topologies.

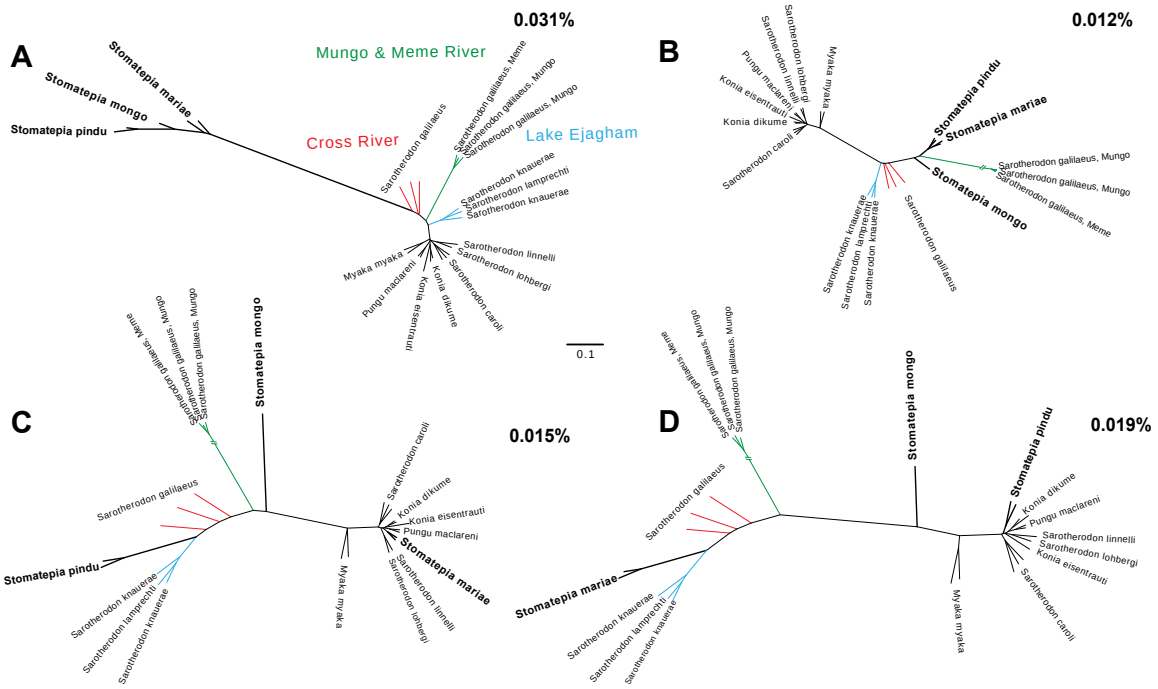


Figure S2. *Saguaro* topologies featuring Barombi Mbo polyphyly with riverine populations involving the *Stomatepia* three-species complex. Across small and independent proportions of the genome A-B) the entire *Stomatepia* clade (cacti 43 and 51), C) only *S. pindu* (cacti 10), and D) only *S. mariae* are more closely related to outgroups than other Barombi Mbo species (cacti 30).

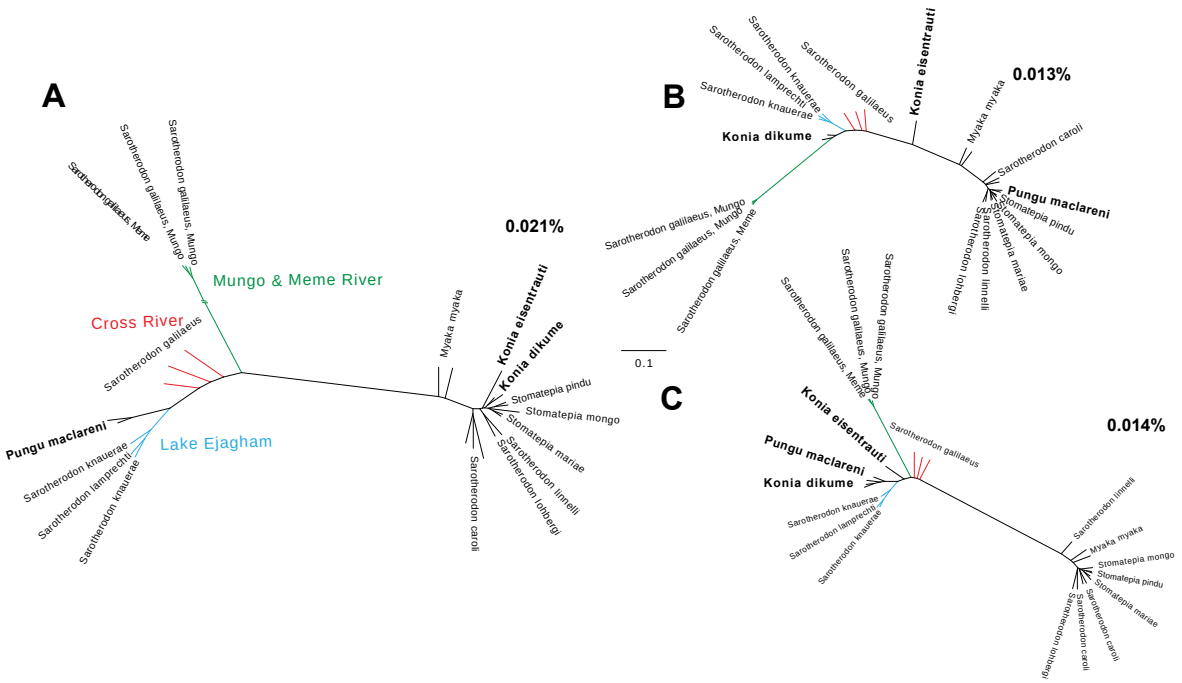


Figure S3. *Saguaro* topologies featuring Barombi Mbo polyphyly with riverine populations involving the *Konia* + *Pungu* subclade. Across small and independent proportions of the genome A) only *P. maclareni* (cacti 13), B) only *K. dikume* (cacti 7), and C) the entire *Konia* + *Pungu* subclade are more closely related to outgroups than other Barombi Mbo species (cacti 39).

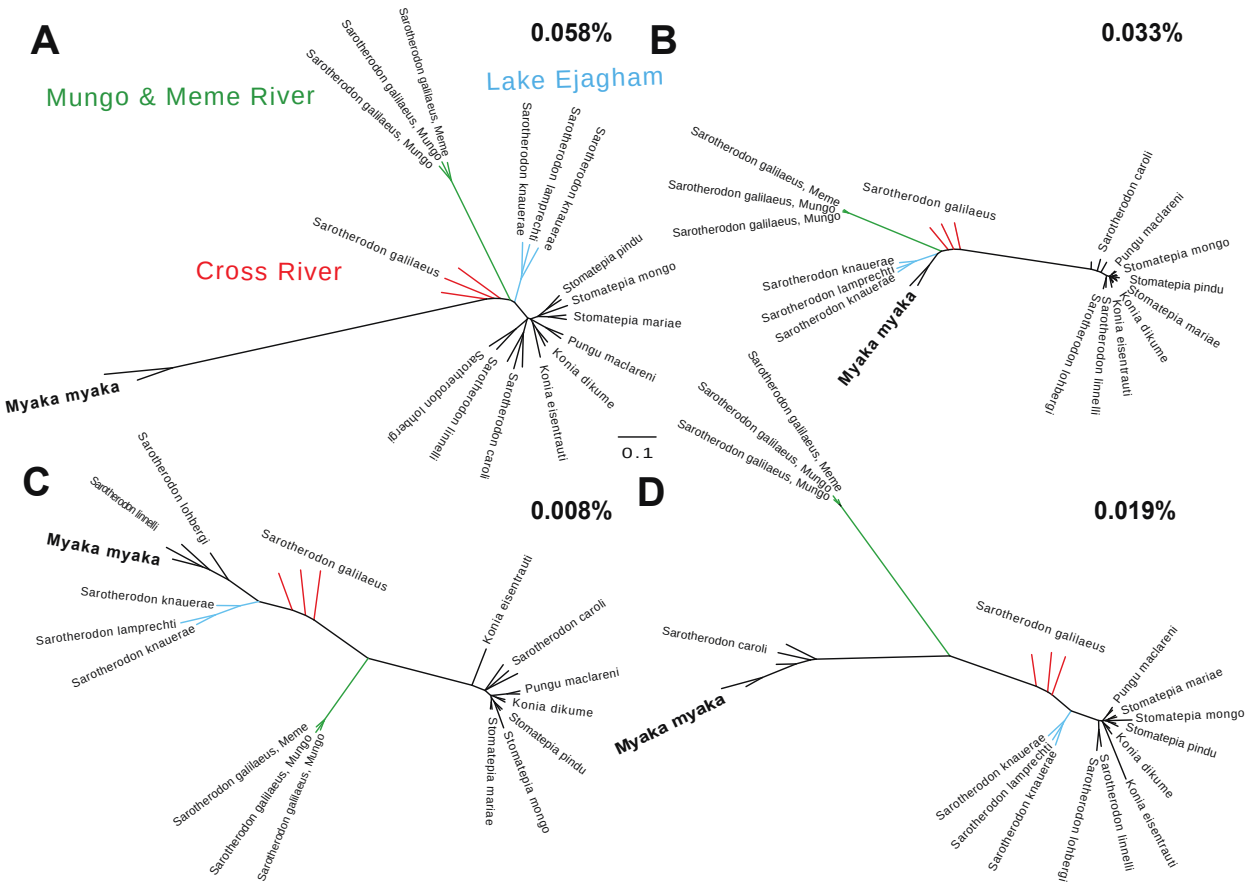


Figure S4. *Saguaro* topologies featuring Barombi Mbo polyphyly with riverine populations involving the *Myaka* + *Sarotherodon* subclade. Across small and independent proportions of the genome A-B) only *M. myaka* (cacti 17 and 3), C) *M. myaka* and two Barombi Mbo *Sarotherdon* species (*S. linelli* and *S. lohbergi* ; cacti 38), and D) *M. myaka* and *S. caroli* are more closely related to outgroups than other Barombi Mbo species (cacti 41).

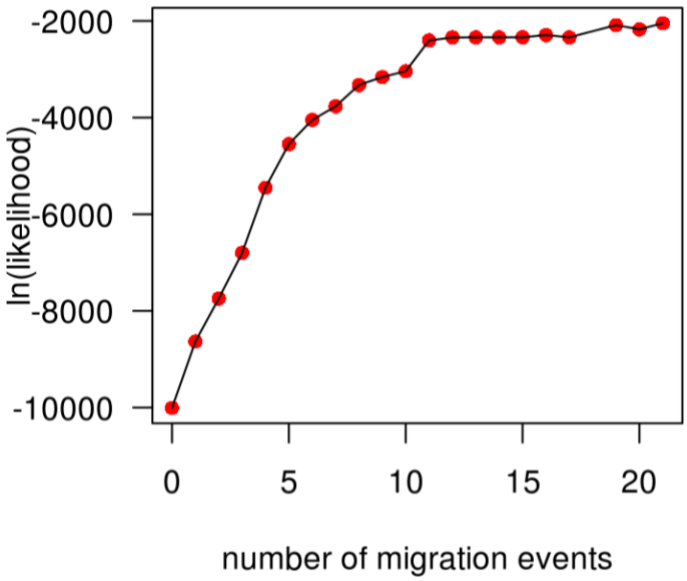


Figure S5. The log-likelihood of Treemix population graphs for Barombi Mbo cichlids as a function of the number of migration events.

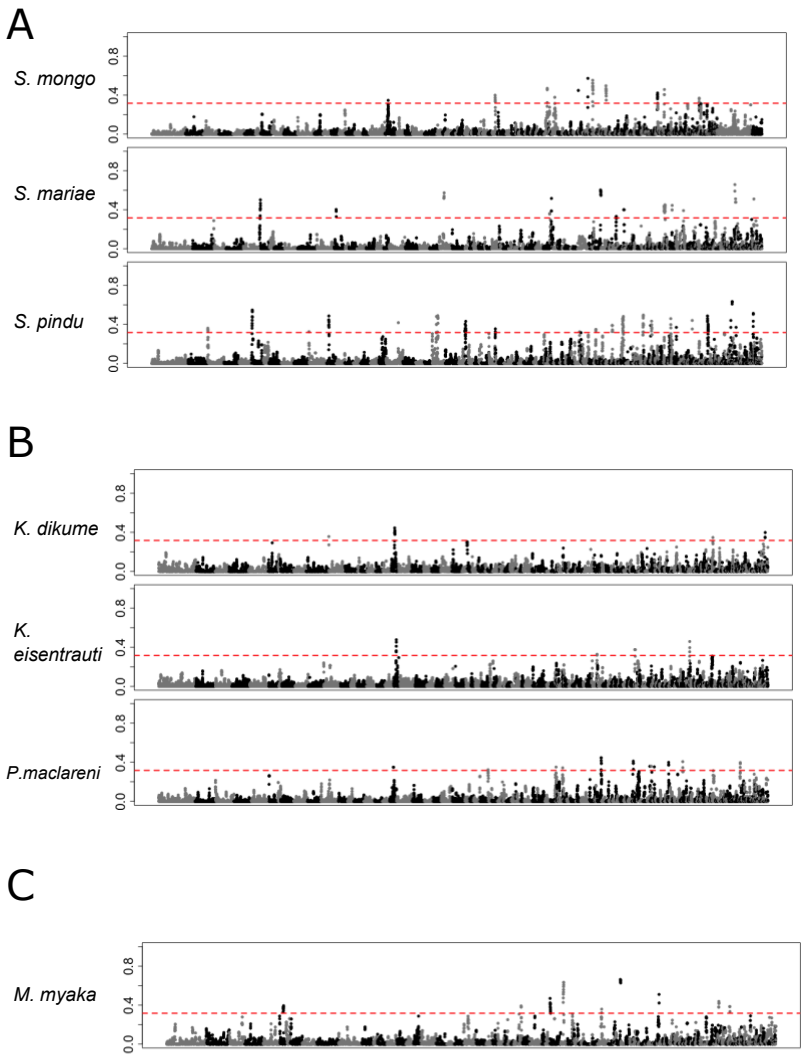


Figure S6. Visualization of introgression with *S. galilaeus* MM across unassigned scaffolds for Barombi Mbo. Alternating gray/black colors indicate different unplaced scaffolds, arranged from largest to smallest in size. Dotted red lines mark the coalescent simulation-based significance thresholds for each test (F_d = 0.315). Peaks highlighted in colors represent those signals of introgression shared across different subclades.

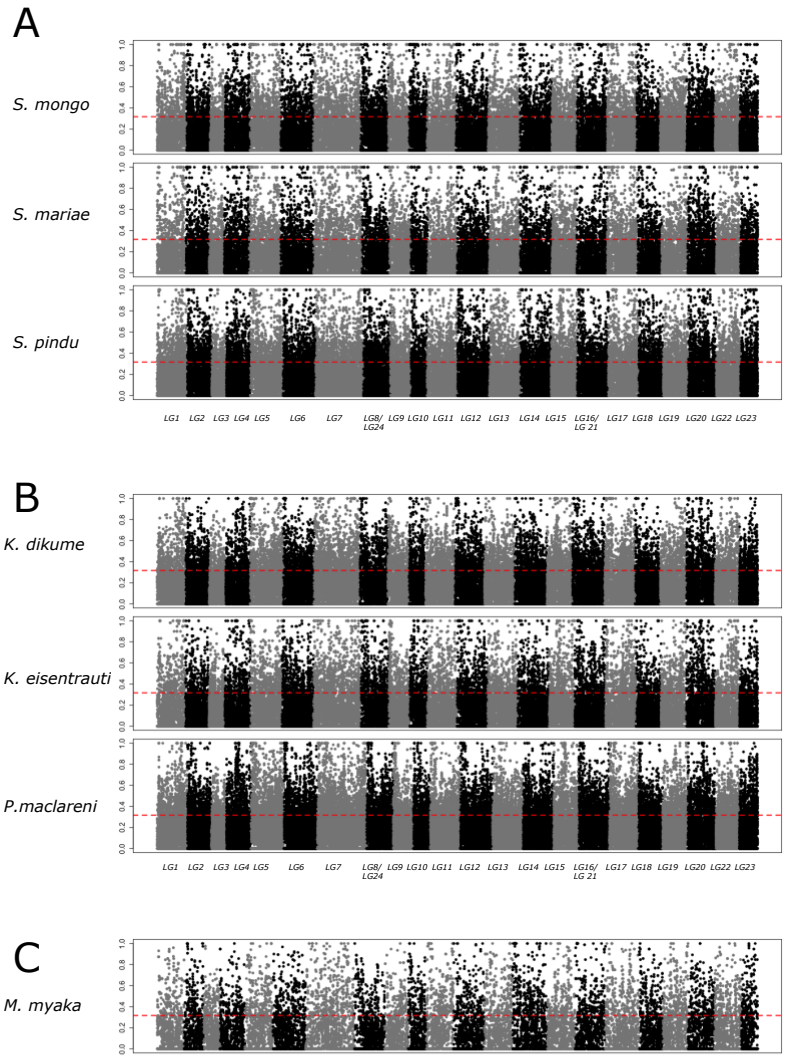


Figure S7. Visualization of introgression with *S. galilaeus* CR across linkage groups for Barombi Mbo. Alternating gray/black colors indicate different linkage groups. Dotted red lines mark the coalescent simulation-based significance thresholds for each test (F_d = 0.325). Peaks highlighted in colors represent those signals of introgression shared across different subclades.

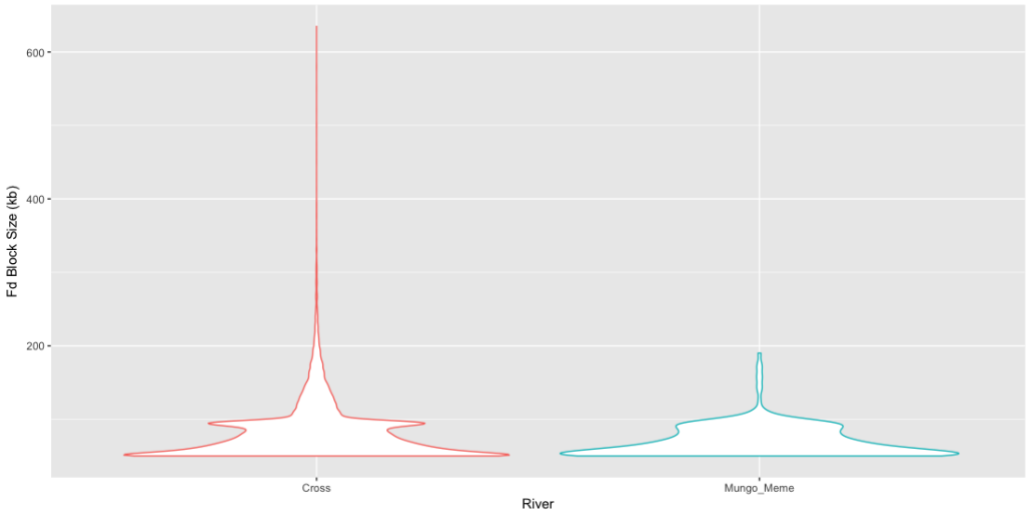


Figure S8. Distribution of introgression block sizes detected from sliding window *fd* statistic tests for Barombi Mbo species. These distributions are for all introgression blocks across all the combinations of tests run between Barombi Mbo with *S.galilaeus* populations from Cross River (red) and Mungo and Meme River (blue).

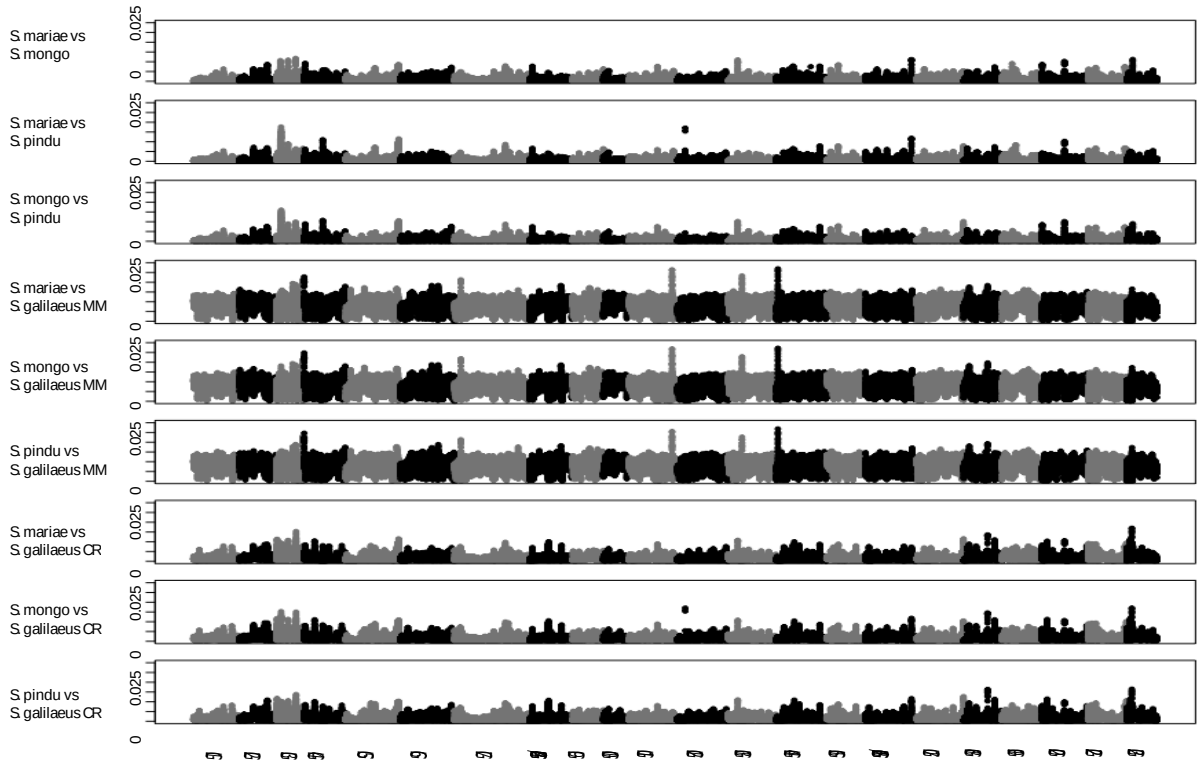


Figure S9. Visualization of divergence across the genome for species of *Stomatepia* and *S. galilaeus* populations from Mungo/Meme River (MM) and Cross River (CR). Alternating gray/black colors indicate different linkage groups.

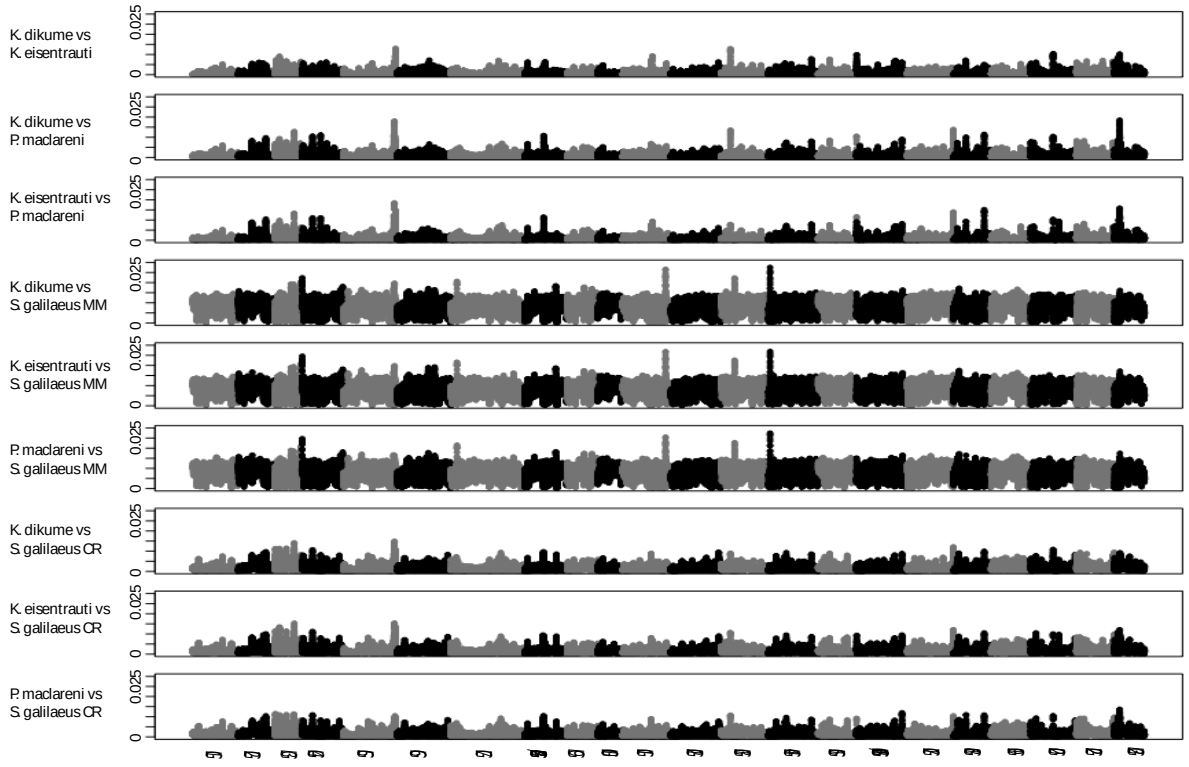


Figure S10. Visualization of divergence across the genome for species of *Konia* and *S. galilaeus* populations from Mungo/Meme River and Cross River. Alternating gray/black colors indicate different linkage groups.

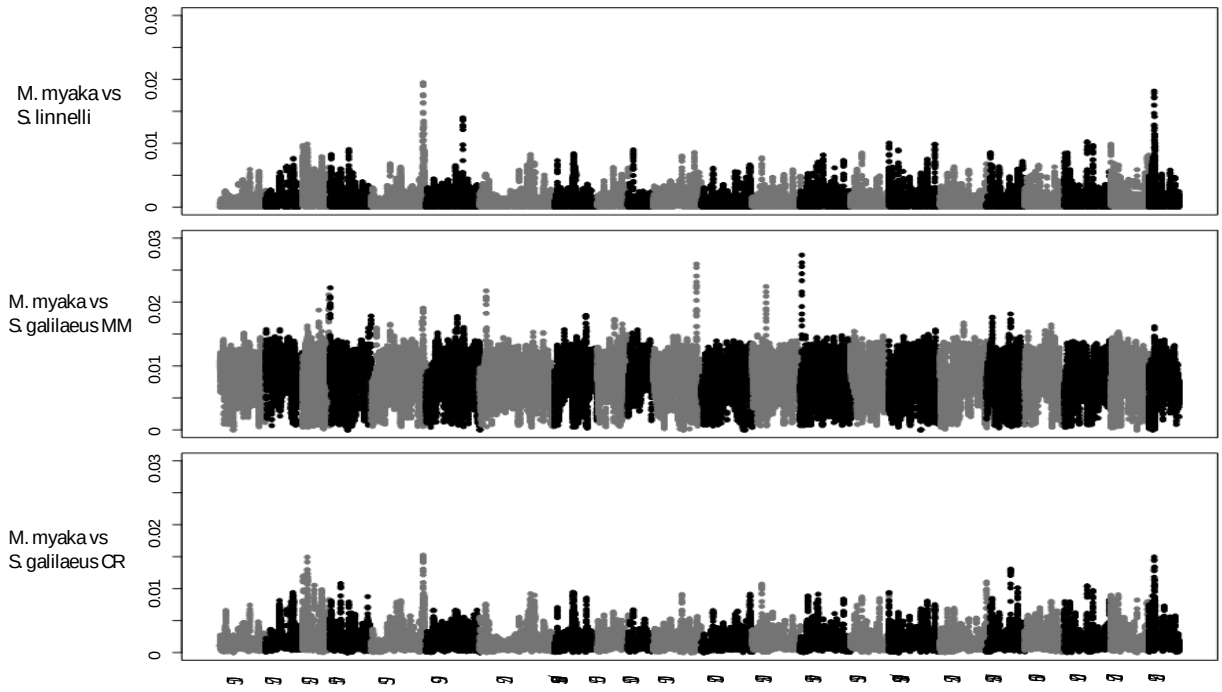


Figure S11. Visualization of divergence across the genome between *M. myaka* and *S.linnelli, S. galilaeus* populations from Mungo/Meme River, and Cross River. Alternating gray/black colors indicate different linkage groups.

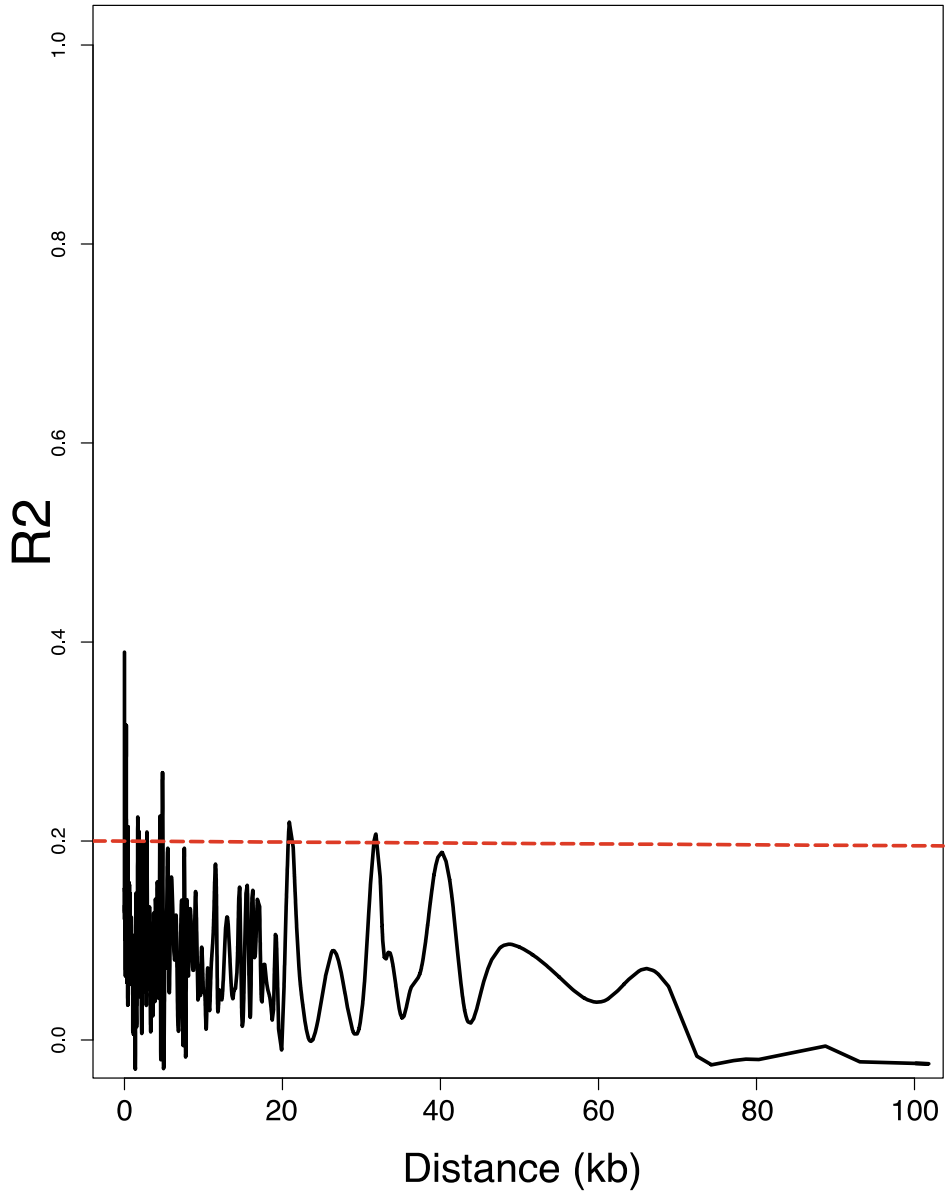


Figure S12. Linkage disequilibrium decay among individuals used in this study. The correlation coefficient (r^2) values between pairwise SNPs on the largest linkage group in the dataset (LG3) against distance between the two SNPs. Arbitrary line set at r^2=0.2 for visualizing decay across distance.

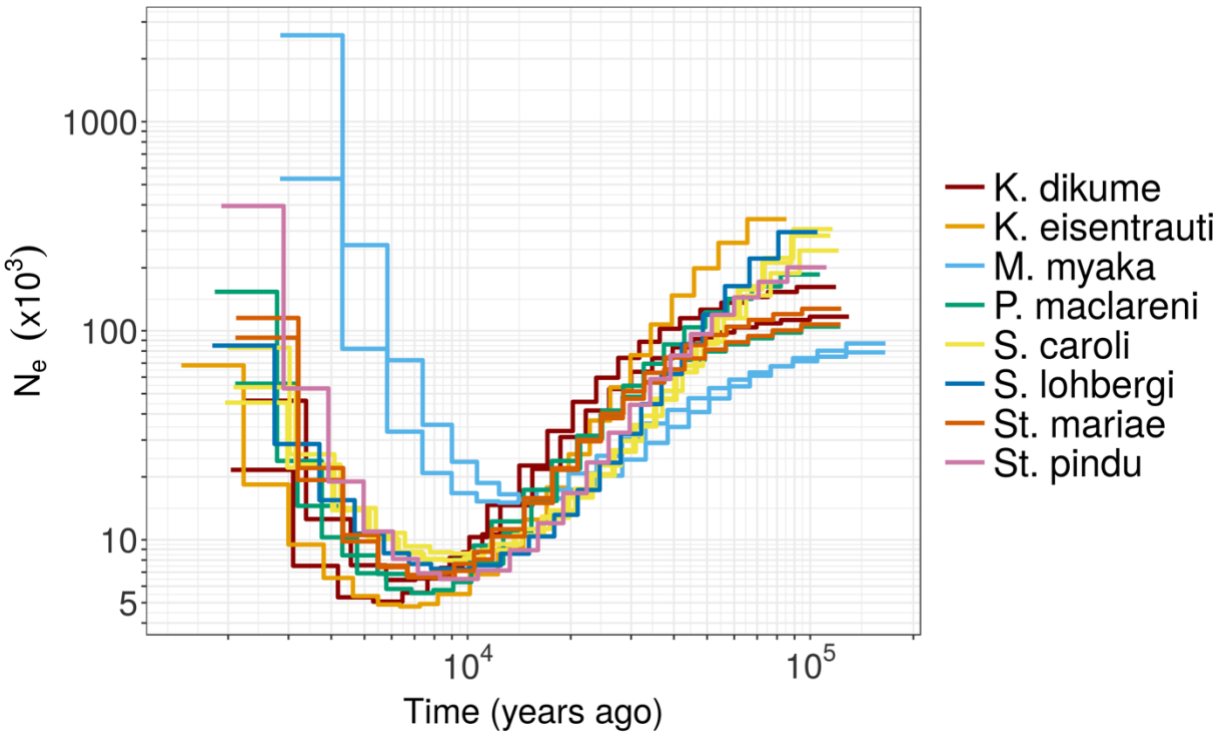


Figure S13. Ancestral Population Size of Barombi Mbo Species. Historical effective population sizes estimated by the Multiple Sequentially Markovian Coalescent (MSMC) for Barombi Mbo species and *Sarotherodon* riverine populations from Mungo and Meme River using a one year generation time and mutation rate estimated for cichlids (7.1×10^{-9} mutations per site per year derived from (Guo et al. 2013)).

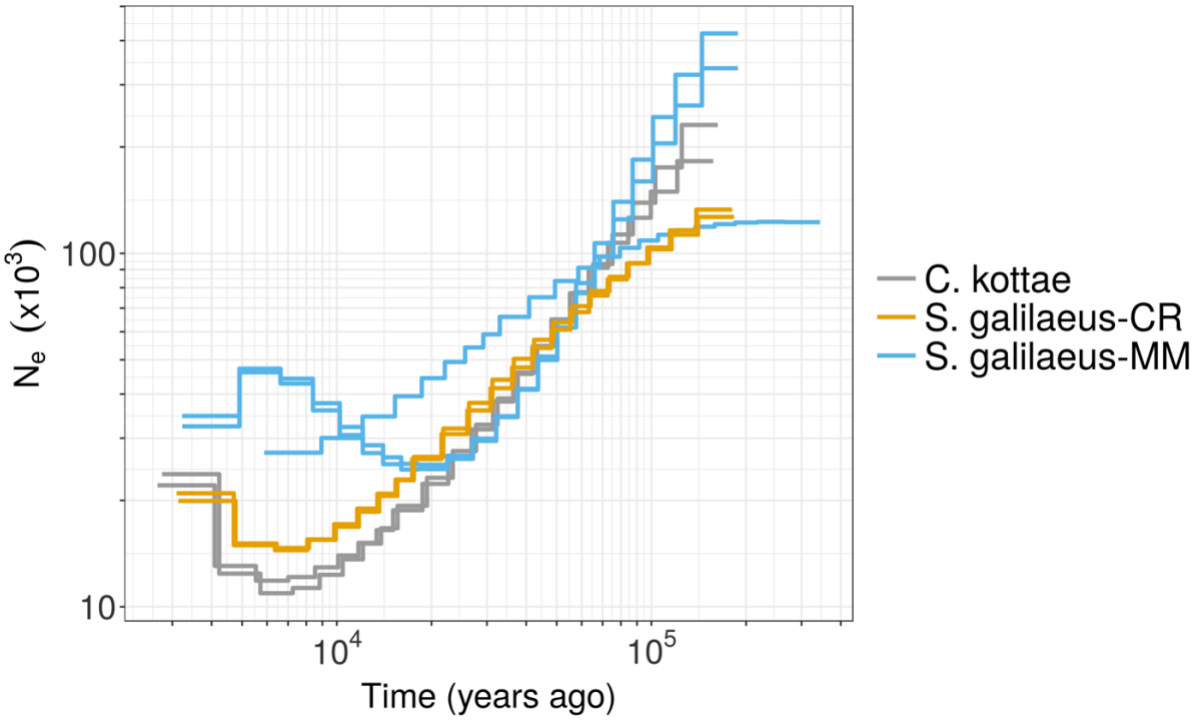


Figure S14. Ancestral population size of outgroups lineages. Historical effective population sizes estimated by the Multiple Sequentially Markovian Coalescent (MSMC) for *C. kottae* and *Sarotherodon* riverine populations from Cross River using a one year generation time and mutation rate estimated for stickleback (7.1×10^{-9} mutations per site per year derived from (Guo et al. 2013)).

2.8.3. Supplemental Tables

Table S1. Percentages of the genome assigned to topologies under various filtering criteria. The relative percentage (out of all segments assigned in SAGUARO analysis) compared to the absolute percentage (out of the entire genome length 927,679,487) for each unique topology featuring either a monophyletic or non-monophyletic Barombi Mbo. We calculated both percentages under different filtering criteria of each segment assigned to a topology having at least one SNP, greater or equal to than 10 SNPs or 20 SNPs.

Monophyletic						
	Relative Percentage			**Absolute Percentage**		
Topology ID	**None**	**10 SNPs**	**20 SNPs**	**None**	**10 SNPs**	**20 SNPs**
1	0.172	0.049	0.023	0.151	0.033	0.012
2	0.045	0.042	0.017	0.039	0.028	0.009
4	0.246	0.189	0.168	0.216	0.128	0.090
6	1.078	0.279	0.135	0.948	0.189	0.072
11	0.344	0.093	0.044	0.303	0.063	0.024
12	0.374	0.125	0.091	0.329	0.085	0.049
15	0.726	0.246	0.170	0.639	0.167	0.091
16	0.994	0.319	0.210	0.875	0.216	0.113
19	0.407	0.153	0.100	0.358	0.104	0.054
20	1.677	1.581	1.572	1.476	1.072	0.844
26	0.703	0.371	0.224	0.619	0.252	0.120
29	0.172	0.096	0.067	0.152	0.065	0.036
35	0.855	0.247	0.147	0.752	0.167	0.079
42	0.061	0.035	0.027	0.054	0.024	0.014
46	0.249	0.283	0.269	0.219	0.192	0.145
54	3.504	1.735	1.003	3.084	1.176	0.538
55	0.070	0.033	0.024	0.061	0.023	0.013
60	0.060	0.044	0.037	0.053	0.030	0.020
62	0.051	0.028	0.017	0.045	0.019	0.009
Figure 1B (65)	**15.195**	**18.624**	**22.089**	**13.375**	**12.622**	**11.855**
67	1.689	0.559	0.120	1.486	0.379	0.064
68	2.700	1.016	0.287	2.377	0.069	0.154
Figure 1B (69)	**63.184**	**71.224**	**71.340**	**55.614**	**48.271**	**38.289**
70	0.045	0.024	0.019	0.039	0.016	0.010
71	1.508	0.762	0.613	1.327	0.517	0.329
72	0.036	0.017	0.010	0.032	0.012	0.006

Total percentage	94.6	98.17	98.23	84.62	65.92	53.04
Non-Monophyletic						
	Relative Percentage			**Absolute Percentage**		
Topology ID	**None**	**10 SNPs**	**20 SNPs**	**None**	**10 SNPs**	**20 SNPs**
0	0.006	0.002	0.001	0.005	0.001	0.001
Figure 2D (3)	**0.224**	**0.087**	**0.061**	**0.197**	**0.059**	**0.033**
5	0.058	0.053	0.027	0.051	0.036	0.014
7	0.147	0.042	0.024	0.130	0.029	0.013
8	0.057	0.024	0.021	0.050	0.016	0.011
9	0.029	0.022	0.022	0.025	0.015	0.012
10	0.121	0.048	0.029	0.107	0.033	0.015
13	0.145	0.065	0.040	0.127	0.044	0.021
14	0.141	0.070	0.055	0.124	0.047	0.029
17	0.305	0.152	0.107	0.269	0.103	0.058
18	0.095	0.042	0.031	0.084	0.028	0.016
21	0.096	0.038	0.021	0.085	0.026	0.011
22	0.044	0.030	0.024	0.039	0.020	0.013
23	0.076	0.037	0.018	0.067	0.025	0.010
24	0.043	0.017	0.010	0.038	0.011	0.005
25	0.128	0.047	0.026	0.113	0.032	0.014
27	0.059	0.031	0.022	0.052	0.021	0.012
28	0.057	0.035	0.025	0.050	0.024	0.013
30	0.096	0.048	0.036	0.084	0.032	0.019
31	0.038	0.017	0.009	0.033	0.012	0.005
32	0.051	0.040	0.016	0.045	0.027	0.008
33	0.065	0.026	0.016	0.058	0.018	0.008
34	0.045	0.032	0.021	0.040	0.022	0.011
36	0.042	0.030	0.026	0.037	0.021	0.014
37	0.053	0.023	0.017	0.047	0.016	0.009
Figure 2C (38)	**0.054**	**0.023**	**0.015**	**0.048**	**0.016**	**0.008**
Figure 2B (39)	**0.101**	**0.041**	**0.026**	**0.089**	**0.028**	**0.014**
40	0.048	0.025	0.019	0.042	0.017	0.010
41	**0.110**	**0.052**	**0.035**	**0.097**	**0.035**	**0.019**

Figure 2A (43)	**0.216**	**0.104**	**0.057**	**0.190**	**0.070**	**0.031**
44	0.081	0.041	0.020	0.072	0.027	0.011
45	0.057	0.029	0.020	0.050	0.020	0.011
47	0.073	0.048	0.031	0.064	0.033	0.017
48	0.029	0.021	0.009	0.025	0.014	0.005
49	0.035	0.019	0.015	0.031	0.013	0.008
50	0.066	0.029	0.023	0.058	0.020	0.013
51	0.236	0.062	0.022	0.207	0.042	0.012
52	0.031	0.017	0.014	0.027	0.011	0.007
53	0.061	0.028	0.017	0.054	0.019	0.009
56	0.067	0.036	0.025	0.059	0.025	0.013
57	0.084	0.043	0.033	0.074	0.029	0.018
58	0.036	0.021	0.012	0.032	0.014	0.006
59	0.041	0.031	0.018	0.036	0.021	0.010
61	0.061	0.023	0.015	0.054	0.015	0.008
63	0.048	0.023	0.016	0.043	0.016	0.009
64	0.034	0.018	0.011	0.030	0.012	0.006
66	0.043	0.020	0.014	0.038	0.014	0.008
73	0.014	0.006	0.003	0.012	0.004	0.002
74	0.009	0.006	0.004	0.008	0.004	0.002
Total Percentage	3.85	1.82	1.18	3.39	1.24	0.632

Table S2. Overlap of *Saguaro* and f_d introgressed regions. The cumulative length of regions assigned by *Saguaro* to the supported polyphyletic topologies (Figure 2) and the percentage of the significant f_d regions that were assigned to these topologies.

Linkage Group	f_d region position	Cumulative length of *Saguaro* regions	Percentage
Regions uniquely introgressed into a single species			
LG 3	14525001-14575000	165	0.3300066
LG 4	800001-855000	697	1.2672958
LG 5	35530001-35595000	103	0.158464
LG 6	10670001-10740000	554	0.7914399
LG 6	20935001-10740000	5351	5.9456216
LG 6	22460001-22515000	1553	2.8236877
LG 11	28280001-28375000	115	0.1210539
LG 16-21	31140001-31220000	2197	2.7462843
LG 17	9260001-9315000	76	0.1381843
LG 20	150001-230000	1389	1.7362717
LG 20	1705001-1755000	6662	13.3242665
LG 20	19570001-19630000	483	0.8050134
NT_167475.1	450001-500000	5530	11.0602212
NT_167500.1	930001-1000000	7074	10.1058587
NT_167557.1	1410001-1495000	8877	10.4436523
NT_167586.1	220001-290000	8286	11.837312
NT_167617.1	320001-400000	34957	43.6967962
NT_167636.1	140001-190000	339	0.6780136
NT_167663.1	535001-630000	4933	5.1926862
NT_167671.1	265001-330000	4257	6.5493315
NT_167675.1	605001-655000	189	0.3780076
NT_167747.1	45001-330000	8788	3.0835196
NT_168010.1	30001-80000	275	0.550011
NT_168013.1	1-70000	3721	5.3157902
Regions of shared introgression within a subclade			
LG 2	9190001-9275000	3730	4.3882869
NT_167557.1	1110001-1170000	8085	13.4752246
NT_167617.1	105001-160000	2256	4.1018928
NT_167653.1	680001-770000	18279	20.3102257
NT_167679.1	395001-445000	1249	2.49805
NT_167702.1	445001-530000	363	0.4270638
NT_167728.1	40001-120000	73	0.0912511
NT_167736.1	160001-220000	484	0.8066801

Regions of shared introgression within Barombi Mbo			
LG 3	1185001-1235000	138	0.2760055
LG 3	3325001-3425000	19766	19.7661977
LG 5	35510001-35575000	334	0.5138541
LG 17	24560001-24620000	812	1.3533559
LG 20	19550001-19600000	2716	5.4321086
LG 20	300001-370000	2455	3.507193
NT_167475.1	2260001-2340000	10186	12.7326592
NT_167535.1	855001-925000	9610	13.7287676
NT_167709.1	135001-190000	1672	3.0400553
NT_167716.1	100001-265000	10662	6.4618573
NT_167744.1	260001-340000	8070	10.0876261
NT_167891.1	55001-125000	454	0.6485807
NT_167891.1	75001-130000	1549	2.8164148

Table S3. Average D_{xy} among Barombi Mbo and riverine *Sarotherodon* species across the genome.

	K. eisentrauti	*K. dikume*	*P. maclareni*	*S. mongo*	*S. mariae*	*S. pindu*	*M. myaka*	*S. linnelli*	*S. galilaeus* MM	*S. galilaeus* CR
K. eisentrauti	--	0.00062	0.00079	0.00083	0.00087	0.00081	0.00089	0.00081	0.00801	0.0015
K. dikume		--	0.00086	0.00088	0.00092	0.00087	0.00094	0.00087	0.00794	0.00154
P. maclareni			--	0.00088	0.00092	0.00086	0.00094	0.00082	0.00788	0.00152
S. mongo				--	0.00071	0.00059	0.00092	0.00088	0.00782	0.00152
S. mariae					--	0.00068	0.00097	0.00091	0.00789	0.00156
S. pindu						--	0.00091	0.00085	0.008	0.00152
M.myaka							--	0.00091	0.0078	0.00152
S. linnelli								--	0.0077	0.00149
S. galilaeus MM									--	0.0076
S. galilaeus CR										--

Table S4. Candidate introgressed regions in Barombi Mbo cichlid radiation that are shared across multiple species in a subclade. These regions feature significant f_d values between riverine populations of *S. galilaeus* MM (Mungo and Meme River) and the three subclades of the radiation focused on in this study. Unannotated regions with no GO terms are marked with (-).

Linkage Group	Position	Gene(s)	Gene Ontology Terms
		Konia	
NC_022200.1	9190001-9275000	UPC(6)	--
NT_167557.1	1110001-1170000	fam111a; rnf175; trim2	fibrillar center, chromatin;ubiquitin protein ligase activity; ubiquitin-protein transferase activity
NT_167679.1	395001-445000	--	--
NT_167702.1	445001-530000	lipe; cd79a; arhgef1	triglyceride catabolic process;B cell receptor signaling pathway;regulation of Rho protein signal transduction
NT_167891.1	155001-225000	cxcr3; pafah1b3*; cnfn; tlr13 ; hmcn1; ceacam20; ceacam5; UPC	CXCR3 chemokine receptor binding;platelet-activating factor acetyltransferase activity, spermatogenesis;cornified envelope,keratinization;toll-like receptor 13 signaling pathway;basement membrane; fin morphogenesis;positive regulation of cytokine production;negative regulation of myotube differentiation
NC_022205.1	1390001-1445000	--	--
NT_167617.1	105001-160000	gastrula zinc finger XICGF26.1	unknown
NT_167653.1	680001-770000	dock3	small GTPase mediated signal transduction
NT_167736.1	160001-220000	btnl2; UPC(3)	negative regulation of T cell receptor signaling pathway
		Stomatepia	
NT_167728.1	40001-120000	H2-Q10; H2-LHX9; tnc	ribonuclease H2 complex;ribonuclease H2 complex;prostate gland epithelium morphogenesis

*best candidate region for secondary gene flow contributing to diversification; UPC = uncharacterized protein coding gene

Table S5. Candidate introgressed regions in Barombi Mbo cichlid radiation that are shared across multiple species across Barombi Mbo. These regions feature significant f_d values between riverine populations of *S. galilaeus* MM (Mungo and Meme River) and the three subclades of the radiation focused on in this study. Unannotated regions with no GO terms are marked with (-).

Linkage Group	Position	Gene(s)	Gene Ontology Terms
LG1	26250001-26315000	adgrg1	heparin binding
LG3	3325001-3425000	UPC(9); UPG (1)	--
LG3	1185001-1240000	sh3tc1	biological process
LG5	35500001-35575000	or52e8*;matn4	olfactory receptor activity;growth plate cartilage chondrocyte morphogenesis,calcium ion binding
LG16-21	5365001-5420000	--	--
LG17	24560001-24620000	prdm4;tmem209;ahcyl2	histone methyltransferase binding*;integral component of membrane;S-adenosylmethionine cycle activity
LG18	20715001-20770000	tep1;znrf1;grina;parp2;UPC	RNA binding; E3 ubiquitin protein ligase activity;negative regulation of endoplasmic reticulum stress-induced intrinsic apoptotic signaling pathway;DNA ligation involved in DNA repair
LG20	19300001-19410000	klhdc8b	ubiquitin-protein transferase activity
LG20	19550001-19615000	sbk; UPC	protein serine/threonine kinase activity
NT_167475.1	2260001-2340000	serp1 ;C-type natriuretic peptide 2;eno3;pfn1*;UPC(2)	endoplasmic reticulum unfolded protein response ;unknown;phosphopyruvate hydratase activity;positive regulation of actin filament bundle assembly*;--
NT_167535.1	855001-930000	dmbt1	defense response to Gram-negative bacterium (Digestive Phnotype
NT_167709.1	135001-230000	hepacam	cell-cell junction; cell adhesion
NT_167716.1	95001-260000	st8sia5	sialylation,sialyltransferase activity
NT_167744.1	260001-340000	siglec1;gpr87;p2ry14	regulation of osteoclast development;G-protein coupled receptor signaling pathway*;G-protein coupled receptor signaling pathway
NT_167891.1	55001-130000	--	--
NT_167927.1	1-55000	mdn1*	rRNA processing

*best candidate region for secondary gene flow contributing to diversification; UPC = uncharacterized protein coding gene

Table S6. Within-population genetic diversity in introgressed regions in Barombi Mbo and *Sarotherodon* riverine populations. Bolded rows indicate introgressed regions where with-in population nucleotide diversity is >10X lower than the scaffold average.

Konia + Pungu											
		K.eisentrauti		***K.dikume***		***P. maclareni***		***S. galilaeus* MM**		***S. galilaeus* CR**	
scaffold	**position**	**f_d region**	**LG avg**	**f_d region**	**LG avg**	**f_d region**	**LG avg**	**f_d region**	**LG avg**	**f_d region**	**LG avg**
K. dikume introgressed regions											
NC_022200.1	18660001 - 18710000	0.00054	0.0004	0.00178	0.0004	0.00042	0.0004	0.00314	0.0004	0.00349	0.001
NT_168013.1	1-70000	0.00068	0.0007	0.00068	0.001	0.00068	0.0007	0.00068	0.002	0.00068	0.001
***K. eisentrauti* introgressed regions**											
NC_022204.1	22460001 - 22515000	0.00145	0.0004	0.00115	0.0004	0.00174	0.0004	0.00116	0.0005	0.00096	0.001
NT_167586.1	220001-290000	0.00089	0.001	0.00121	0.001	0.00095	0.001	0.00166	0.002	0.00171	0.002
NT_167663.1	535001-630000	0.00059	0.0005	0.00056	0.0005	0.00062	0.0005	0.00034	0.00032	0.00565	0.0003
***P. maclareni* introgressed regions**											
NC_022201.1	14525001 - 14575000	0.00263	0.0008	0.0023	0.0008	0.00218	0.0008	0.00218	0.0008	0.00302	0.001
NC_022215.1	9260001-9315000	0.00367	0.0004	0.00308	0.0004	0.00194	0.0004	0.00049	0.0004	0.00263	0.0004
NT_167557.1	1410001-1495000	0.00185	0.002	0.00166	0.001	0.00149	0.001	0.00117	0.001	0.00279	0.002
NT_167671.1	265001-330000	0.00272	0.001	0.00235	0.001	0.00187	0.001	0.00128	0.002	0.00281	0.002
NT_168010.1	**30001-80000**	**6.86E-05**	**0.0001**	**0.00052**	**0.001**	**0.00081**	**0.001**	**0.00061**	**0.0006**	**0.00161**	**0.001**
Stomatepia											
		S. mariae		***S. pindu***		***S. mongo***		***S. galilaeus* MM**		***S. galilaeus* CR**	
scaffold	**position**	**f_d region**	**LG avg**	**f_d region**	**LG avg**	**f_d region**	**LG avg**	**f_d region**	**LG avg**	**f_d region**	**LG avg**
***S. mongo* introgressed regions**											
NC_022218.1	**150001-230000**	**8.5296E-05**	**0.004**	**0.00143**	**0.004**	**0**	**0.004**	**0.00071**	**0.004**	**0.00205**	**0.001**
***S. mariae* introgressed region**											
NC_022214.1	31140001 - 31220000	0.00118	0.004	0.00149	0.004	0.00104	0.004	0.00094	0.004	0.00232	0.001
***S. pindu* introgressed region**											
NT_167675.1	**605001-655000**	**0.00272**	**0.0008**	**0.00056**	**0.005**	**0.00394**	**0.001**	**0.00109**	**0.0008**	**0.00369**	**0.001**
Myaka + Sarotherodon											
		M. myaka		***S. linnelli***		***S. galilaeus* MM**		***S. galilaeus* CR**			
scaffold	**position**	**f_d region**	**LG avg**	**f_d region**	**LG avg**	**f_d region**	**LG avg**	**f_d region**	**LG avg**		
NC_022201.1	17560001 - 17610000	0.00201	0.001	0.00198	0.001	0.00106	0.0007	0.00362	0.001		
NC_022202.1	800001-855000	0.00747	0.0006	0.00924	0.0005	0.00657	0.0005	0.00825	0.001		

NC_022203.1	35530001 - 35595000	0.0049	0.0005	0.01373	0.0005	0.00974	0.0004	0.00778	0.001
NC_022204.1	10670001 - 10740000	0.00747	0.0006	0.00924	0.0005	0.00657	0.0005	0.00825	0.001
NC_022204.1	20935001 - 21025000	0.00253	0.0006	0.00262	0.0005	0.00288	0.0005	0.00306	0.001
NC_022209.1	28280001 - 28375000	0.0035626 5	0.0005	0.00229944	0.0005	0.00138	0.0005	0.00499	0.001
NC_022218.1	1705001-1755000	0.0009556 6	0.0006	0.00149491	0.0005	0.00099	0.0003	0.00207	0.001
NC_022218.1	19570001 - 19630000	0.0006839	0.0006	0.00099304	0.0005	0.00015	0.0003	0.00155	0.001
NT_167475.1	450001-500000	0.0035656 9	0.0024	0.00619587	0.0029	0.00512	0.0024	0.00201	0.003
NT_167500.1	835001-1000000	0.0028197	0.0009	0.00074963	0.0006	0.00111	0.0006	0.00401	0.002
NT_167568.1	10001-60000	0.0029674 2	0.002	0.0031527	0.001	0.00238	0.002	0.00266	0.001
NT_167617.1	320001-400000	0.0011110 5	0.002	0.00013603	0.002	0.00046	0.002	0.00137	0.002
NT_167716.1	220001-270000	0.0050563 3	0.002	0.01043704	0.004	0.00528	0.004	0.00753	0.002
NT_167790.1	255001-305000	0.0045841 3	0.001	0.0037551	0.001	0.00379		0.00401	0.001
NT_167636.1	140001-190000	0.0059642	0.003	0.00871405	0.003	0.00564	0.003	0.00417	0.003

Inter-chapter Transition

In Chapter 1 and 2, I found evidence for histories of hybridization in two different radiations. However, this opposing results of these two chapters in terms of how introgression from these hybridization events contributed to diversification processes highlighted the potentially equivocal impact that hybridization may have in any particular radiation. For example, in Chapter 2, I found very few differentially introgressed regions in the genomes of species from Barombi Mbo crater lake cichlid radiation in Cameroon and these regions did not correspond to regions of the genome that have strongly diverged between species within the radiation. Previous genome-wide inferences of hybridization detected signatures of secondary contact in this radiation, which casted doubt on one of the best examples of sympatric speciation in the wild. However, the findings of Chapter 2 emphasized that on a regional scale across the genome, this hybridization may not have played a strong role in diversification processes in this sympatric radiation. These findings prompted an exploration of how such evidence of complex histories of secondary gene flow that we are detecting fits into our current understanding of speciation mechanisms from theoretical models. These models provide us with more direct understanding of the mechanisms involved in speciation and the ways in which those mechanisms vary in strength across different speciation scenarios that span various geographic scales. In Chapter 3, I discuss how to leverage our understanding of speciation mechanisms from theoretical speciation models to re-evaluate cases of sympatric speciation in the wild in which signatures of secondary gene have been detected and suggest analyses to assess the impact that secondary gene flow has had on speciation processes.

Chapter 3: Searching for sympatric speciation in the genomic era

This chapter has been previously published and is reproduced here in accordance with the journal's article sharing policy:

Richards EJ, Servedio M, and Martin CH. 2019. Searching for sympatric speciation in the genomic era. *BioEssays* 41 (7) DOI: 10.1002/bies.201900047

3.1. Abstract

Sympatric speciation illustrates how natural and sexual selection may create new species in isolation without geographic barriers. However, recent genomic reanalyses of classic examples of sympatric speciation have revealed complex histories of secondary gene flow from outgroups into the radiation. In contrast, the rich theoretical literature on this process distinguishes among a diverse range of models based on simple genetic histories and different types of reproductive isolating barriers. Thus, there is a need to revisit how to connect theoretical models of sympatric speciation and their predictions to empirical case studies in the face of widespread gene flow. We review and summarize theoretical differences among different types of sympatric speciation and speciation-with-gene-flow models, and propose genomic analyses for distinguishing which models apply to case studies based on the timing and function of adaptive introgression. Investigating whether secondary gene flow contributed to reproductive isolation is necessary to test whether predictions of theory are ultimately borne out in nature.

3.2. Introduction: What is sympatric speciation?

As Mayr famously quipped, sympatric speciation is like the Lernean Hydra: "which grew two new heads whenever one of its heads was cut off" (p. 451; (Mayr 1963)). The latest incarnation of this phenomenon has occurred over the past decade: sympatric speciation now means two different things to different research groups. We stress that our goal here is not to offer a new definition of sympatric speciation nor grow a new head on the hydra, but only to clarify existing usage so that we can focus on reconciling diverse theoretical models with existing empirical examples of this process.

Subsequent to Mayr's classic definition based on geography (1947), sympatric speciation was redefined over the past two decades in a population genetic context as the most extreme endpoint on the continuum of divergence with gene flow: panmictic gene flow and no initial divergence at the start of speciation (Via 2001; Coyne and Orr 2004a; Bolnick and Fitzpatrick 2007; Fitzpatrick et al. 2008, 2009). In the context of theoretical speciation models, this type of speciation process is the most difficult because the starting conditions involve no pre-existing divergence among loci involved in reproductive isolation. Instead, linkage disequilibrium must build up through time within a population through the action of disruptive natural selection and strong assortative mating by ecotype, despite the countervailing eroding force of recombination (Felsenstein 1981; Dieckmann and Doebeli 1999; Gavrilets 2004; Doebeli et al. 2005).

Recently, the definition of sympatric speciation has been expanded to focus more on biogeographical context in line with Mayr's original definition (Mallet et al. 2009b), in which the speciation process is defined as sympatric a) as long as diverging populations are within 'cruising range' of each other and b) regardless of whether secondary gene flow provided alleles contributing to reproductive isolating barriers in sympatry. Cruising range provides a practical empirical definition of gene flow between diverging sympatric populations, allowing for some geographic or microallopatric population structure (e.g. (Mallet et al. 2009b; Foote 2018)). Allowing for the secondary gene flow of alleles contributing some or all of the reproductive isolation between sympatric populations also expands the definition of sympatric speciation to include both 'hard' and 'easy' processes under one umbrella. Theoretical models show that speciation is much easier from starting conditions that involve some level of initial divergence and/or restricted gene flow (Otto et al. 2008), for example, if the alleles necessary for reproductive isolation first become physically linked in allopatry (Box 1).

These two definitions, the population genetic and biogeographic, reflect different perspectives on the value of studying sympatric speciation. The biogeographic definition (for clarity, we will refer to this process as 'easy sympatric speciation'), with its broader range of starting conditions that are easier to verify in nature, increases, perhaps vastly, the number of empirical speciation events that could be categorized as examples of sympatric speciation. This definition values the frequency of sympatric diverging populations in nature compared to allopatric speciation, as an estimate of the overall importance of "sympatry" in contributing to biodiversity. The population genetic definition (for clarity, we will refer to this more restrictive process as 'hard sympatric speciation'), with its narrow set of starting conditions that remain challenging to verify in nature, finds value in studying both the easy and hard processes of sympatric speciation defined in theoretical models. Namely it values the theoretical possibility of creating new species solely through the power of divergent selection alone, regardless of whether this process is common in nature. Here we focus on the types of questions that genomic data now allow us to ask to improve the search for examples of both the easy and hard processes of sympatric speciation, and investigate the range of speciation mechanisms found in nature from among those shown to be plausible in theory.

Under the biogeographical definition of sympatric speciation, there is little difference in terms of the speciation mechanisms said to be involved in scenarios that start with initial panmixia (i.e. hard sympatric speciation) versus those that start with some geographic or microallopatric population structure (i.e. easy sympatric speciation (Mallet et al. 2009b; Foote 2018)). However, this contrasts with the theoretical literature, which differentiates models of hard sympatric speciation from other models of speciation with gene flow. Indeed, theory teaches us that the hard process of sympatric speciation (without the aid of secondary gene flow contributing to reproductive isolation) is uniquely and notoriously difficult (Coyne and Orr 1989), in part because quite specific conditions of resource availability (e.g., (Dieckmann and Doebeli 1999; Polechová et al. 2005)), mating traits and preferences (e.g., (Weissing et al. 2011; Norvaišas and Kisdi 2012)), and search costs (e.g., (Kopp and Hermisson 2008)) must be met for it to occur. Some argue that the effort to discern the exact geographic scenario and initial conditions of speciation would be better spent on finding loci involved in reproductive isolation (i.e. 'barrier loci' (Nosil and Schluter 2011; Ravinet et al. 2017)). This is an important first step and we can glean something about the process of speciation from gene annotations of barrier loci and linkage architecture. However, understanding whether any one particular locus or potential mechanism was necessary for speciation often requires placing genomic discoveries in the

context of speciation models that explicitly compare the importance of such factors and mechanisms in driving divergence, models whose outcomes are highly dependent on the initial conditions before sympatric divergence.

3.3. Main Text

3.3.1. Different mechanistic processes underlie divergence in sympatry

Regardless of definition, it is necessary to distinguish among different sympatric divergence processes to understand which classes of speciation models and predictions apply to specific case studies. We here distinguish different scenarios (Figure 1) that will result in two sister species in sympatry based on whether secondary gene flow aided in population divergence: 1) hard sympatric speciation without gene flow; 2) hard sympatric speciation in the presence of a) neutral secondary gene flow or b) after differential sorting of an ancestral hybrid swarm. In the latter case, we also distinguish whether the ancestral hybrid swarm population achieved panmixia before later divergence (i.e. hard sympatric speciation); otherwise, differential sorting of haplotypes within the hybrid swarm is better described by secondary contact speciation-with-gene-flow models rather than sympatric speciation models. 3) Easy sympatric speciation may be aided by secondary gene flow that a) triggers initial sympatric divergence or b) increases divergence after initial divergence in sympatry becomes stalled, an outcome of many sympatric speciation models without sufficiently strong disruptive selection (Matessi et al. 2001; Bolnick and Doebeli 2003; Bolnick 2006; Bürger and Schneider 2006). Finally, 4) secondary contact after a period of allopatry between two populations can result in coexistence or reinforcement, if there is not collapse into a single admixed population (Turelli et al. 2001; Kirkpatrick and Ravigné 2002; Otto et al. 2008; Garner et al. 2018) or extinction of one or both populations.

3.3.1.1. Some scenarios of sympatric divergence are easier than others

We consider scenarios 1 and 2a to be examples of hard sympatric speciation, whereas scenarios 3 and 4 would be examples of speciation aided by secondary gene flow, a much easier process in theory. Interestingly, hybrid swarm scenarios (2b) exist in a gray area, since substantial initial gene flow from multiple sources may increase ecological or preference variation within a population that is sufficient to trigger later sympatric divergence, even without segregating inversions or genetic incompatibilities (van Doorn and Weissing 2001; Seehausen 2004, 2013). So far, we know of no examples of scenario 1 within any case study of sympatric sister species examined using genomic tools; even long diverged species show some evidence of introgression from outgroups in their past (e.g. (Turissini and Matute 2017)). In contrast, sympatric speciation with neutral gene flow (Scenario 2a, and conditionally Scenario 2b) and speciation aided by gene flow (Scenarios 3 and 4) frequently appear to operate concurrently even within a single sympatric adaptive radiation (e.g. (Wagner and McCune 2009; Martin 2012; Meier et al. 2017a)).

3.3.1.2. Why should we distinguish between different sympatric divergence scenarios?

It is important to distinguish these scenarios because theoretical models predict that sympatric divergence unaided by any form of secondary gene flow is substantially more difficult than other

speciation with gene flow scenarios (Box 1). Gene flow throughout the speciation process allows recombination to break down linkage disequilibrium among alleles associated with ecological divergence and assortative mating. There are also three different types of reproductive isolating traits to consider within sympatric speciation models: the most difficult process involves independently segregating loci for ecotype, female preferences, and male traits within the population, whereas sympatric divergence is much easier if any of these three types of traits are combined (i.e. cannot become disassociated by gene flow and recombination), such as assortative mating based on phenotype matching instead of separate loci for preference and traits (Kopp et al. 2017; Servedio and Boughman 2017) or "magic" traits (such as assortative mating based on microhabitat preference; (Dieckmann and Doebeli 1999; Gavrilets 2004; Servedio et al. 2011)). Sympatric speciation by sexual selection alone is also theoretically possible (albeit considered highly unlikely) if there is substantial preference variation either initially within the population or through secondary gene flow (van Doorn and Weissing 2001; Weissing et al. 2011).

Any form of linkage disequilibrium among ecological and mate choice loci formed in allopatry, whether due to physical linkage, selection, or drift, can thus tend to shift the initial starting conditions of panmixia in favor of sympatric divergence (Kirkpatrick and Ravigné 2002). However, linkage disequilibrium without physical linkage subsides within a relatively small number of generations after secondary sympatry and thus may not allow sufficient time for the evolution of assortative mating within the population. In contrast, pre-existing physical linkage among ecological loci has been shown to increase the probability of divergence, especially when it captures already divergent alleles, as is more likely after a period of divergence in allopatry before secondary contact (Kirkpatrick and Barton 2006; Feder and Nosil 2009). Similarly, physical linkage can cause preference and trait alleles to mimic phenotype matching, although even tight linkage can break down over long timescales (shown in a model with population structure: (Servedio and Bürger 2018)). Segregating inversions in the ancestral population are now well-known empirical examples of physical linkage promoting divergence in sympatry (Kirkpatrick and Barton 2006; Fuller et al. 2018). Sympatric divergence is also limited by many other restrictive conditions including the costs of female choosiness and strengths of disruptive selection and assortative mating.

Despite extensive searches for examples of sympatric speciation in the wild, there are few convincing case studies due to the difficulty of ruling out historical allopatric scenarios (see below) and ruling out a role of introgression in speciation. Furthermore, the role of magic traits or matching vs. preference/trait mechanisms is not fully understood in any existing case study. Thus, we still have very limited empirical tests of an extensive theoretical literature and diverse competing models of the notoriously difficult process of sympatric speciation (Gavrilets 2003, 2014; Kopp et al. 2017; Servedio and Boughman 2017).

3.3.2. The classic criteria for sympatric speciation do not distinguish between different sympatric divergence scenarios

There are four traditional criteria for demonstrating hard sympatric speciation without secondary gene flow (e.g. Scenario 1 in Figure 1A): 1) sister species have to be reproductively isolated, 2) form a monophyletic group, 3) largely overlap in ranges, and 4) have biogeographic and evolutionary histories that make periods of allopatric divergence highly unlikely (Coyne and Orr

2004a). Very few case studies have been able to meet these rigorous criteria despite intensive searches (Coyne and Orr 2004a; Bolnick and Fitzpatrick 2007). This has led to the prominent status of crater lake cichlids as some of the best examples of sympatric speciation in the wild due to the uniform shape of isolated volcanic lakes, which convincingly rule out phases of allopatry due to water level changes (Box 2; (Barluenga et al. 2006)).

The monophyly criterion assumes that monophyly arises only when a single ancestral population underlies the present-day daughter species. This is typically met by inferring a single phylogeny from one or more loci. This single point-estimate view of evolutionary history is problematic because it obscures the presence of non-bifurcating relationships among organisms (e.g. sister species that derived ancestry from multiple source populations due to extensive gene flow or hybrid speciation) and the real variation in evolutionary histories among genes across the genome itself (e.g. (Hahn and Nakhleh 2016)). Few regions of the genome may initially contribute to reproductive isolation resulting in a heterogeneous genomic landscape of differentiation among incipient species (Wu 2001a), a pattern now extensively supported across case studies (e.g. (Fontaine et al. 2015; McGirr and Martin 2016b; Campbell et al. 2018)). Therefore, monophyletic relationships are consistent with, but not exclusive to, a scenario of sympatric speciation. Examining heterogeneous evolutionary histories across regions relevant to speciation is thus crucial for understanding the processes and conditions under which sympatric divergence can occur.

3.3.3. The ‘new’ problem of sympatric speciation is to establish or reject a functional role for secondary gene flow

While genomics has increased our ability to resolve evolutionary relationships among organisms, it has also revealed more complex evolutionary histories of multiple colonizations and extensive secondary gene flow in nearly all examples of sympatric speciation that have been examined with genomic data so far (Barluenga and Meyer 2010; Geiger et al. 2013; Elmer et al. 2014; Igea et al. 2015; Machado-Schiaffino et al. 2015; Malinsky et al. 2015; Martin et al. 2015a; Kautt et al. 2016a; Poelstra et al. 2018; Richards et al. 2018); e.g. to our knowledge Lord Howe Island palms and indigobirds have not yet been directly examined for secondary gene flow with an outgroup). Indeed, only a handful of genes may directly contribute to the speciation process whereas the rest of the genome is porous to gene flow while reproductive isolation is incomplete (Wu 2001a; Wu and Ting 2004). Examples of sympatric speciation without secondary gene flow (Scenario 1) are now even rarer after applying modern genomic tools to search for introgression. Instead, it is still possible that even the hard process of sympatric speciation may occur in the face of secondary gene flow in nearly all these examples (Scenario 2a; (Richards et al. 2018)). Importantly, most evidence of secondary gene flow impacting putative examples of sympatric speciation comes from genome-wide tests of introgression from outgroup lineages that do not look at how that secondary gene flow has impacted reproductive isolating barriers between diverging populations in sympatry (e.g. (Martin et al. 2015a; Kautt et al. 2016a)). In case studies of sympatric speciation that involve radiations of species, secondary gene flow may also impact only some of the diverging populations such that some species within a radiation may better represent sympatric speciation scenarios than others. Therefore, introgression detected at the genome-wide level from lineages outside the speciation event tells us that secondary gene flow

has occurred, but little about the divergence process among incipient sympatric species and how that gene flow shaped the process of speciation.

The challenge of understanding the hard process of sympatric speciation in the genomic era is establishing or rejecting a functional role for the secondary gene flow commonly present during the speciation process, in effect ruling out scenarios 3 and 4 in favor of scenario 2 (Figure 1). Even if signatures of secondary gene flow are detected, speciation could still have occurred solely via mechanisms of hard sympatric speciation if that secondary gene flow did not play a causal role in divergence (Scenario 2a and possibly 2b). In contrast, secondary gene flow could play a causal role if it introduced novel genetic variation or physically linked alleles (e.g. a segregating inversion) that promoted divergence, through mechanisms such as inflating variance through the creation of a hybrid swarm **(**Scenario 2b**;** ((Seehausen 2004, 2013; Martin 2016b; Meier et al. 2017a)), adaptive introgression (Scenario 3; (Hedrick 2013; Huerta-Sanchez et al. 2014; Stankowski and Streisfeld 2015; Richards and Martin 2017a)), transgressive segregation (Scenarios 2-3; (Rieseberg et al. 1999; Kagawa and Takimoto 2017)), or hybrid speciation ((Schumer et al. 2014b)). Beyond examples of sympatric speciation, genetic variation brought in through gene flow with divergent lineages has been found in many empirical examples of rapid speciation and recognized as a potential mechanism for rapid speciation and adaptive radiation (reviewed in (Marques et al. 2019)). Here we propose and discuss genomic analyses that may help to establish or reject a functional role of secondary gene flow in the speciation process (Figure 1). This is necessary to identify putative cases of hard sympatric speciation when gene flow appears to be nearly universal in the wild, particularly among sympatric diverging populations.

3.3.4. Genomic analyses can aid in distinguishing between different scenarios of sympatric speciation

Although genome-wide analyses of introgression provide a starting point, ultimately consideration of the time of arrival and functional role of each introgressed region within extant sympatric sister species pairs will be necessary to distinguish between hard sympatric speciation in which incidental gene flow does not contribute to reproductive isolating barriers (Scenario 2a) versus easy sympatric speciation in which divergence is aided by secondary gene flow (Scenario 3; e.g. segregating inversions (Feder et al. 2003b; Fishman et al. 2013) or balancing selection on regions containing multiple barrier loci (Guerrero and Hahn 2017; Nelson and Cresko 2018)). We suggest four major types of genomic analyses to address questions about the role of secondary gene flow and identify sympatric speciation with gene flow: analyses to 1) estimate the timing of introgression into sympatric sister species relative to their divergence time, 2) infer the presence and timing of selective sweeps within sympatric sister species, 3) annotate candidate adaptive introgression regions for functional elements or trait associations that may be relevant to speciation, and 4) if closely related non-speciating outgroups are available, confirm the lack of selective sweeps of these regions in outgroups. Some of these analyses are already being applied to examples of sympatric speciation in the wild, particularly crater lake cichlid systems, taking on the difficult task of distinguishing between sympatric speciation scenarios (Box 2). These analyses are by no means trivial, as evidence for either ‘hard’ or ‘easy’ sympatric speciation scenarios in the wild remain sparse, but recently developed methods have made it possible to start addressing such challenging questions.

3.3.4.1. Is the observed secondary gene flow concurrent with divergence times?

Estimating the duration of gene flow and the timing of introgression into a sister species from an outgroup relative to the timing of divergence between sympatric sister species will help distinguish between scenarios of sympatric speciation, speciation with gene flow, and secondary contact. If populations diverged in sympatry independent of any concurrent secondary gene flow (Scenario 2a), we might expect to see weak concordance of the timing of gene flow with divergence times among species, for example discrete gene flow events that date well before or after divergence times among species. In the case of both discrete gene flow events surrounding divergence time estimates or continuous gene flow from the time of colonization to the present, more information about function and selection on regions introgressed near the time of speciation will be needed. Increasingly sophisticated approaches for detecting fine-scale patterns of introgression and inferring the timing and duration of gene flow from genomic data are becoming available (Box 3).

3.3.4.2. Did any of the introgressed regions experience selective sweeps and did the timing of these sweeps align with species divergence time?

We can use information about selective sweeps of introgressed variation to further characterize the role of secondary gene flow in sympatric divergence. When an allele is selectively favored in a population, positive selection may cause it to increase in frequency and form a localized selective sweep of reduced genetic variation surrounding the adaptive variant (Smith and Haigh 1974). Such regions of high differentiation in recently diverged species are often targeted as candidates for speciation genes, although other processes not directly associated with speciation can lead to similar patterns of high heterogeneity in differentiation across a genome (reviewed in (Nachman and Payseur 2012; Cruickshank and Hahn 2014; Ravinet et al. 2017)). If speciation was recent or ongoing, there may be strong signatures of a selective sweep for particular haplotypes in at least one of the sister species for regions involved in the divergence process (e.g. regions containing selective sweeps overlap regions of strong divergence; Figure1b). If secondary gene flow was neutral with respect to speciation, we may find no signatures of selective sweeps in those introgressed regions. However, care should be taken with any significance thresholds used for calling regions candidates for adaptive introgression, ideally thresholds resulting from genetic simulations. For example, false positives overlapping among all three categories are possible depending on the frequency of regions that are strongly differentiated, experienced a selective sweep, or introgressed, even if secondary gene flow was neutral.

Importantly, a sweep of the same introgressed region in both sympatric sister species may be interpreted as adaptation to the same new environment, which may not contribute to reproductive isolation between the pair (dependent on their respective genetic backgrounds; e.g. (Clarkson et al. 2014; Stölting et al. 2015)). However, this pattern is also consistent with the sweep of a region contributing to a ‘one-allele’ mechanism of mate choice (Felsenstein 1981; Kopp et al. 2017; Servedio and Boughman 2017), such as increased female choosiness in both sympatric sister species, which *would* contribute to reproductive isolation (e.g. (Ortíz-Barrientos

and Noor 2005)). Thus, selective sweeps of an introgressed region in both sympatric sister species do not rule out its role in aiding the speciation process.

Alternatively, if selective sweeps are detected, the timing of selective sweeps can give indirect evidence about their role in speciation. If the timing of introgression predates the timing of the selective sweep, it is challenging to infer the importance of an introgressed region for speciation because linkage disequilibrium among loci relevant to speciation may take time to build up. However, the absence of selective sweeps or introgression until long after species divergence would suggest that introgression was not relevant to speciation.

3.3.4.3. Is there support for a causal role of secondary gene flow based on functional genetic analyses of variants in the region?

Another potential source of evidence for the functional importance of gene flow can come from genome-wide association studies (GWAS) between variants in introgressed regions and traits involved in ecological or sexual isolation between sister species. The conservation of sequences within introgressed regions across taxa may also provide strong evidence of a functional role (e.g. PhastCons (Siepel et al. 2005)). However, many complex traits are driven by a large number of variants of small effect and ruling out a functional role for gene flow from gene annotations is difficult (e.g. see the omnigenic model; (Boyle et al. 2017)). Finally, and most powerfully, genome editing and gene expression reporter systems are increasingly tractable in non-model systems (e.g. (Kratochwil et al. 2017; Cleves et al. 2018)). This is ultimately an asymmetric problem: finding evidence that an introgressed region may have contributed to reproductive isolation is easier than demonstrating that no introgressed regions contributed to reproductive isolation in any way (Richards et al. 2018).

3.3.4.4. Are there similar patterns of selection or divergence in the introgessed regions in closely related outgroup populations?

Thorough investigation of these same regions in outgroups to the sympatric species gives added power to distinguish whether secondary gene flow aided sympatric divergence. If non-diversifying, closely related species exist in similar environments and haven't diversified in a similar manner but share signatures of selective sweeps in the same regions, then the observed introgression may have been neutral relative to speciation, e.g. due to adaptations to shared changes in climate or pathogens or shared regions of reduced recombination or increased background selection. Similarly, several studies comparing genomic landscapes of differentiation across closely related taxa have found that high differentiation observed in the same genomic regions across taxa reflects the action of linked selection across low-recombination regions rather than selection against gene flow at barrier loci (Van Doren et al. 2017; Vijay et al. 2017; Delmore et al. 2018; Ma et al. 2018).

3.4. Conclusions and outlook

Sympatric speciation remains among the most controversial evolutionary processes, beloved by theorists and long sought after by empiricists. While evidence of divergence under the biogeographic definition of sympatry is mounting using traditional genetic criteria of monophyly (Bolnick and Fitzpatrick 2007), genomic data has now revealed the pervasiveness of secondary gene flow and introgression in many of these examples. Future fine-scale investigations of introgression will likely continue to paint a complex picture of the role of secondary gene flow in speciation. Establishing or ruling out a role for secondary gene flow in speciation and discerning which putative cases studies evolved through an 'easy' or 'hard' process of sympatric speciation in the wild will be a formidable task, yet a worthwhile one in its revelation of the sheer power of divergent selection to create species in nature.

Nearly all existing case studies of sympatric speciation involve some form of automatic magic trait, such as assortative mating by habitat (Bush 1975; Feder et al. 2003b; Sorenson et al. 2003), along a depth gradient (Malinsky et al. 2015), or environment-induced phenology shifts (Savolainen et al. 2006). We think that an outstanding remaining question is whether the hard process of sympatric speciation occurs in nature without the aid of some form of magic trait, as originally demonstrated to be possible in theory (Dieckmann and Doebeli 1999). The highly polygenic and multi-dimensional nature of adaptation and mate choice suggests that an 'all-of-the-above' speciation scenario containing a mix of preference/trait, magic trait, and phenotype matching (in which each trait is affected by a wide distribution of allelic effect sizes with varying times of arrival) will be the norm in nature. In contrast, although numerous and diverse, most speciation models continue to address these mechanisms in a piecemeal fashion with an assumption of large effect alleles. It remains unclear how different mechanisms, effect sizes, and times of arrival will interact and compete within a single model.

3.5. Box Text

3.5.1. Box 1. Why do we care whether speciation is sympatric?

Inferences from theoretical models predict that, under a scenario of speciation with gene flow (Scenario 3), introgression can make the process of speciation much easier in three ways. First, by introducing additional variation in ecological traits into the population, introgression could potentially facilitate a branching process due to competition for resources (although we are not aware of a model that assesses this precise situation, it can be inferred from the dynamics of (Dieckmann and Doebeli 1999)). Second, introgression of novel alleles for mating preferences may provide a boost in preference variation that could be an important trigger to aid the evolution of assortative mating under a preference/trait mechanism, which requires preference variation to be large ((van Doorn and Weissing 2001; Weissing et al. 2011)). For example, we found evidence of secondary gene flow of olfactory alleles shortly before the rapid divergence of a Cameroon cichlid radiation in Lake Ejagham, which may have boosted preference variation (Poelstra et al. 2018). Third, secondary gene flow after a period of allopatric isolation may lead to increased linkage disequilibrium between assortative mating and ecological loci or among ecological loci. It seems logical that this might facilitate sympatric speciation as this metric is

often described as progress along the speciation continuum. However, initial linkage disequilibrium has been shown not to matter much in at least some scenarios (Felsenstein 1981) because without physical linkage, linkage disequilibrium breaks down quickly. However, physical linkage may enable these alleles to remain in association for a sufficient time for assortative mating to evolve within the population (e.g., (Servedio and Bürger 2018)). Initial linkage disequilibrium may also increase the probability of allelic capture by an inversion or for selection for new mutations within an inversion that may affect both ecology and assortment (Kirkpatrick and Barton 2006). Finally, higher linkage disequilibrium among ecological loci may in some cases increase the probability of sympatric divergence, but this is in effect similar to varying effect sizes of alleles at ecological loci (e.g. many small effect alleles within a region resemble a large-effect locus (Yeaman et al. 2016)). These predictions could also apply to sympatric radiations. For example, some classic sympatric speciation models (e.g. (Dieckmann and Doebeli 1999)) can yield many more than just two species if left to run for more generations (Polechová et al. 2005; Bolnick 2006).

The fundamental difference between sympatric speciation and speciation with gene flow, including secondary contact scenarios, lies in the fact that very often multiple equilibrium states exist in speciation models, such that loss of divergence and maintenance of divergence in the presence of gene flow are both possible outcomes, depending on the starting conditions of a population (this is nicely illustrated for one measure of divergence by (Kirkpatrick and Ravigné 2002), Figure2). In such cases, speciation is much more easily reached from starting conditions that match those of two populations that have diverged largely in allopatry due to the large amount of allelic variation or pre-existing phenotypic bimodality and assortative mating (e.g., (Otto et al. 2008)). Even for scenarios of speciation with gene flow that are much easier, such as geographic separation between two incipient species that are undergoing gene flow or involving some form of magic trait, differentiation is much more difficult to reach or maintain from an initially homogeneous population than from an initially differentiated one (Cotto and Servedio 2017; Sachdeva and Barton 2017).

3.5.2. *Box 2.* Evidence for sympatric speciation from crater lake cichlid radiations

There are relatively few volcanic chains of crater lakes containing fishes in the tropics, notably found in Cameroon, Nicaragua, Tanzania, Uganda, Madagascar, and Papua New Guinea (Sparks 2004; Seehausen 2006; Malinsky et al. 2015). Although sympatric radiations of endemic fishes are known from other isolated saline, alkali, postglacial, and ancient lakes, only four lineages of cichlids have radiated in the world's crater lakes (Figure 3). The most diverse radiation is Barombi Mbo, Cameroon with eleven endemic cichlid species, followed by Lake Bermin, Cameroon with nine (Schliewen et al. 1994). Nicaraguan crater lakes reach up to five species (Kautt et al. 2016a), the East African craters never exceed two sympatric species (Machado-Schiaffino et al. 2015; Malinsky et al. 2015), and Madagascar's crater lakes contain a single endemic cichlid (Sparks 2004). It remains unknown why regional and lineage diversity varies so greatly because there appears to be no relationship between the occurrence of endemic cichlid radiations and crater lake size or age (up to approximately 5 km diameter and 2 million years old) (Wagner et al. 2014).

The evidence for secondary gene flow is remarkably similar across all crater lake cichlid radiations examined with genomic data so far. Admixture proportions with outgroups are frequently detected within the range of 1-4%: 0.6% in Lake Barombi Mbo *Sarotherodon*

(percentage of polyphyletic trees in *Saguaro*: (Richards et al. 2018)), 1.1% in Lake Massoko *Astatotilapia* (Patterson's *D*: (Malinsky et al. 2015)), 4.3% in Lake Apoyo *Amphilophus* (demographic model:(Kautt et al. 2016a)), and 4.4% in Lake Ejagham *Coptodon* (1,138 f_D outliers: (Poelstra et al. 2018)), although notably these studies all use different metrics of introgression. No case studies have yet found evidence of substantial divergence in allopatry followed by secondary contact (but see Lake Xiloá *Amphilophus* (Kautt et al. 2016a)). Instead, nearly all studies have concluded sympatric divergence with periodic or continuous gene flow, potentially from an initial hybrid swarm population (i.e. introgression from multiple outgroup populations).

Secondary gene flow may have triggered sympatric divergence in a radiation of three *Coptodon* cichlids in Lake Ejagham: demographic analyses of whole genomes suggest that this population did not diversify for 8,000 years despite frequent gene flow until an influx of olfactory receptor alleles 1,000 years ago, coinciding with the first sympatric divergence in the lake (Poelstra et al. 2018). In Lake Victoria, segregating opsin alleles in riverine cichlid populations were differentially sorted among Lake Victorian cichlids and may have triggered their diversification (Meier et al. 2017a).

Strong evidence for hard sympatric speciation in crater lake cichlids without the presence of secondary gene flow remains elusive, but some studies have suggested that introgressed variation may not have played a role in speciation between sympatric species. Malinsky et al. (Malinsky et al. 2015) found that the 1.1% introgression found in Lake Massako cichlids resulting from secondary gene flow from a riverine outgroup postdates the major divergence of a shallow/deep-water sister species pair of cichlids. These timing estimates and the observation of a weak correlation between regions of elevated F_{st} and candidate introgressed regions were used to argue that secondary gene flow may not have played a causal role in divergence in this system. However, this initial introgression may still have aided later sympatric divergence (which admittedly is very difficult to rule out) and nothing is yet known about selective sweeps of introgressed variation.

Very recent sympatric divergence in some crater lakes or the proliferation of many species from a few colonization events may also suggest that divergence occurred in sympatry without the aid of gene flow (Elmer et al. 2010; Kautt et al. 2016a); however, in the former case it remains unclear if incipient divergence will continue to complete reproductive isolation or become stalled as appears to be the case in some species complexes of Cameroon crater lake cichlids (Martin 2012, 2013). Very rare secondary gene flow into the Barombi Mbo cichlid radiation (0.6% introgression) without a clear functional role provides weak evidence of sympatric divergence, but more functional characterization and timing of introgression is needed (Richards et al. 2018). The recent advent of transgenic reporters, CRISPR-Cas9, and *in situ* hybridization genetic tools within Nicaraguan crater lake cichlids provides much promise for future investigations of the role of introgression in sympatric divergence (Kratochwil et al. 2015, 2017).

3.5.3. Box 3. Tools for detecting and timing adaptive introgression

3.5.3.1. Detecting and timing introgression

Although there are a variety of tests to detect gene flow on a local scale or within sliding genomic windows, currently three major types of demographic coalescent modeling approaches

can infer the timing of introgression based on different genomic information: 1) the distribution of allele frequencies from genotype data (site frequency spectrum: e.g; (Gutenkunst et al. 2009; Excoffier et al. 2013)), 2) the distribution of haplotype block lengths from phased genomes: e.g; (Loh et al. 2013; Vernot and Akey 2014), and 3) variation in coalescent patterns among gene trees (Gronau et al. 2011).

Distinguishing introgressed variation from variation that stems from the incomplete sorting of ancestral polymorphisms between diverging species (which doesn't rule out a scenario of hard sympatric speciation) is still challenging because these two processes can often lead to very similar patterns in the genome. However, new population genetic simulation methods enable comparisons of genetic patterns simulated under arbitrarily complex demographic scenarios (Haller and Messer 2019). For example, we may now be able to simulate genomes evolving under complex histories of gene flow and introgression compared to simulated genomes evolving under complex histories of ancestral population structure and biased sorting of ancestral genetic variation, and then use subtle differences between the two simulated datasets to detect introgressed variation in empirical genomes. Improvements in machine-learning methods to distinguish between demographic scenarios that can produce similar genetic patterns are also advancing rapidly (Schrider et al. 2015, 2018; Schrider and Kern 2016; Kern and Schrider 2018).

3.5.3.2. Timing of selective sweeps

Recent methods for estimating the age of a selective sweep exploit different aspects about the pattern of variation surrounding the allele on its haplotype background. These include heuristic approaches that use point estimates of mean haplotype length or the number of derived mutations within a chosen distance of the site (Hudson 2007; Coop et al. 2008), model-based approaches that use demographic information and summary statistics of allele frequencies and linkage disequilibrium to model a distribution of ages that fit the observed data (Beleza et al. 2013; Nakagome et al. 2016; Ormond et al. 2016), and full sequence approaches that leverage the length of ancestral haplotypes surrounding the beneficial allele and the accumulation of derived mutations (Chen et al. 2015; Smith et al. 2018).

3.5.3.3. Functional analyses of introgressed variants

Functional annotation of introgressed regions minimally involves searching an annotated reference genome for genes with relevant functions known from model organisms. Intergenic regions can be searched for evidence of strong sequence conservation across taxa (Siepel et al. 2005) or potential regulatory elements (reviewed in (Chatterjee and Ahituv 2017)). Additionally, genome wide association studies (GWAS) can identify variants in introgressed regions correlated with reproductive isolating barriers. Functional validation of gene and regulatory element variants through genome-editing experiments is also becoming increasingly tractable for non-model organisms (e.g. (Kratochwil et al. 2017)).

3.6. Figures

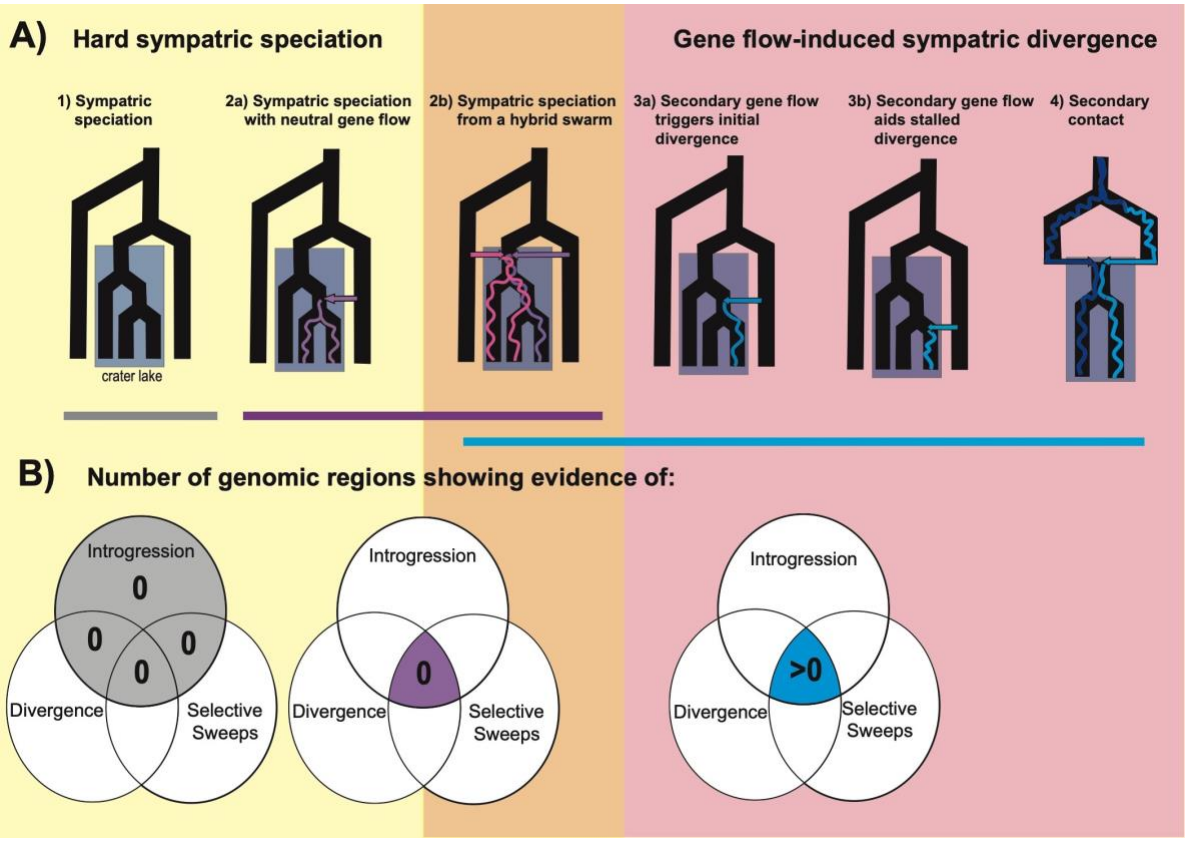


Figure 1: Genomic signatures of sympatric speciation and speciation with gene flow. Speciation scenarios are grouped into hard sympatric speciation scenarios (yellow box; harder process in theoretical models) and other divergence scenarios that can also occur in sympatry (red box; easier processes, which we refer to broadly as sympatric divergence here). Speciation from a hybrid swarm (orange box) can fall under either class of scenarios and additional information is necessary to determine what category of speciation models best describe this process. **A)** The timing of gene flow relative to divergence can be used to distinguish between speciation scenarios. The colored arrows represent gene flow events and the colored lines within the tree are simplified representations of a signature of introgression from that gene flow event into the sympatric species. **B)** Venn diagrams illustrating the number of genomic windows across the entire genome expected to have overlapping signatures of introgression (e.g. f_d outliers), genetic divergence (e.g. F_{st} and D_{xy} outliers), and selective sweeps (e.g. SweeD) for each speciation scenario (e.g. see (Richards and Martin 2017a)). The highlighted sections of the Venn diagrams indicate the key signature that can be used to distinguish between the scenarios. The scenarios that are expected to leave very similar signatures of overlap are grouped by the bars colored with their respective Venn diagram.

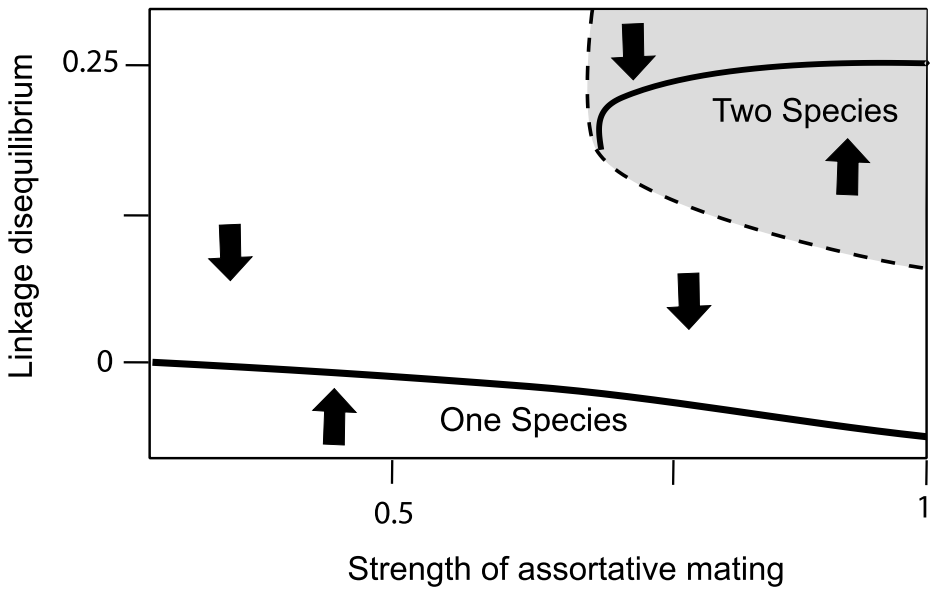


Figure 2. *Two equilibrium cases exist for the linkage disequilibrium (LD), a proxy for differentiation into two distinct "species" in this proof-of-concept model, that can be maintained between two loci that are under disruptive selection and determine assortative mating. With little initial LD, the one-species equilibrium is likely to be reached even when the intensity of assortment is high. When LD in the traits is initially large, as can be the case if there is initially divergence in allopatry, the two-species equilibrium can be reached instead. Modified from* (Kirkpatrick and Ravigné 2002).

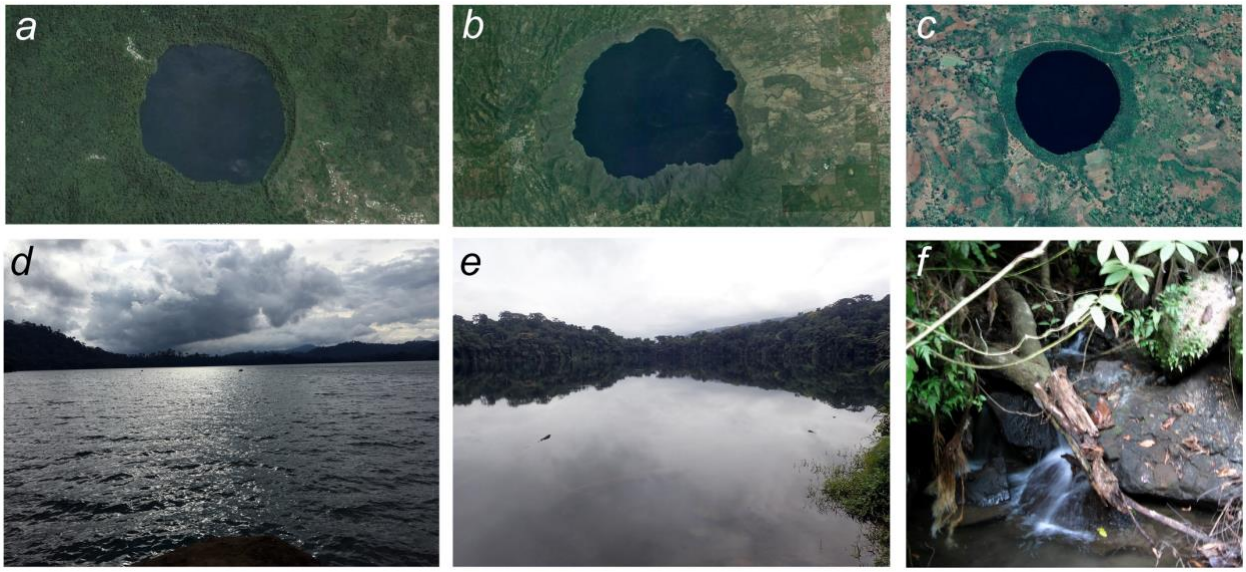


Figure3 Examples of volcanic crater lakes containing endemic cichlid radiations around the globe: *a*,*d*,*f*) Barombi Mbo, Cameroon and its only outlet stream; *b*) Lake Apoyo, Nicaragua, *c*) Lake Massoko, Tanzania, *e*) Lake Bermin, Cameroon. Satellite images (a-c) from Google Earth; (d-f) by CHM.

Inter-chapter Transition

In chapters 1 and 2, I documented extensive evidence for histories of hybridization in two different radiations. Such evidence is the current standard level of support for hybridization playing a role in adaptive radiation. However, these two chapters additionally highlight the potentially equivocal impact those inferred histories of hybridization has had in a particular radiation. In the rare radiation of trophic specialist pupfish on San Salvador Island, I found evidence of introgression with outgroup lineages in regions of the genome that have diverged between species in the radiation, experienced divergent selection in the form of hard selective sweeps and contain genes annotated for relevant craniofacial phenotypes. However, I found opposing results in Barombi Mbo crater lake cichlid radiation in Cameroon, where introgressed regions detected did not correspond to regions of the genome that have strongly diverged between species within the radiation. The multifaceted outcomes of hybridization hinted at by these two systems emphasizes the necessity of additional investigations beyond documenting histories of hybridization. Currently the field interested in the role of hybridization in the origins of adaptive radiations does not know how extensively radiating evolutionary lineages differ from their non-radiating counterparts in their history and capacity for hybridizing. Such comparative studies are hard implement though because many classic adaptive radiations are either restricted within a single unique environment or repeatedly radiated across similar unique environments, making it hard to find systems with radiations nested within many closely related outgroup populations that haven't radiated in similar conditions. However, unique radiations like that of Caribbean pupfishes provide an excellent opportunity to assess major hypotheses about adaptive radiations such as the hybrid swarm hypotheses by asking whether hybridization is strongly associated with adaptive radiation alone in the system. In Chapter 4, I further investigate the signatures of hybridization and adaptive introgression I detected in Chapter 1 by extensively sequencing individual pupfish from across the entire range of *Cyprinodon*. I then use a comparative genomic approach to examine the sources of genetic variation that contribute to the rare radiation of pupfish on San Salvador Island and assess whether the radiation differs in its history of hybridization and introgression from closely related lineages that have not radiated despite experience similar ecological opportunity on nearby islands.

Chapter 4: A vertebrate adaptive radiation is assembled from ancient and disjunct spatiotemporal landscape

This chapter has been previously published and is reproduced here in accordance with the journal's article sharing policy:

Richards, E.J., McGirr, J.A., Wang, J.R., John, M.E.S., Poelstra, J.W., Solano, M.J., O'Connell, D.C., Turner, B.J. and Martin, C.H., 2021. A vertebrate adaptive radiation is assembled from an ancient and disjunct spatiotemporal landscape. *Proceedings of the National Academy of Sciences, 118*(20). DOI: 10.1073/pnas.2011811118.

4.1. Abstract

To investigate the origins and stages of vertebrate adaptive radiation, we reconstructed the spatial and temporal histories of adaptive alleles underlying major phenotypic axes of diversification from the genomes of 202 Caribbean pupfishes. On a single Bahamian island, ancient standing variation from disjunct geographic sources was reassembled into new combinations under strong directional selection for adaptation to novel trophic niches of scale-eating and molluscivory. We found evidence for two longstanding hypotheses of adaptive radiation: hybrid swarm origins and temporal stages of adaptation. Using a combination of population genomics, transcriptomics, and genome-wide association mapping, we demonstrate that this microendemic adaptive radiation of trophic specialists on San Salvador Island, Bahamas experienced twice as much adaptive introgression as generalist populations on neighboring islands and that adaptive divergence occurred in stages. First, standing regulatory variation in genes associated with feeding behavior (*prlh,cfap20,rmi1*) were swept to fixation by selection, then standing regulatory variation in genes associated with craniofacial and muscular development (*itga5,ext1,cyp26b1,galr2*), and finally the only de novo non-synonymous substitution in an osteogenic transcription factor and oncogene (*twist1*) swept to fixation most recently. Our results demonstrate how ancient alleles maintained in distinct environmental refugia can be assembled into new adaptive combinations and provide a new framework for reconstructing the spatiotemporal landscape of adaptation and speciation.

4.1.1. Significance statement

Most biodiversity on earth evolved in rapid bursts of new species, adaptations, and ecological niches. However, this process of adaptive radiation is poorly understood. We used large-scale genomic sequencing across the entire Caribbean range of pupfishes to understand why radiation in this group is restricted to a single Bahamian island. We found that two-fold higher gene flow to this island brought in new combinations of ancient adaptive mutations needed for colonizing novel ecological niches of scale-eating and snail-eating. Adaptation occurred in stages: first selection on feeding behavior, then selection for trophic morphology, and finally refinement through gene coding change. We demonstrate that young, localized radiations can emerge from a vast pool of adaptive genetic variation spread across time and space.

4.2. Introduction

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 (Simpson 1944b; Stroud and Losos 2016). However, the origins and major features of this process are still controversial. For example, ecological opportunity appears necessary but not sufficient to explain why only some lineages radiate and others do not after colonizing similar environments (Wagner et al. 2012; Rabosky 2013; Erwin 2015; Martin 2016b).

One hypothesis about the origins of adaptive radiation is that hybridization between species is necessary to trigger diversification because lineages may be limited in their ability to respond to ecological opportunity by a lack of genetic variation (Seehausen 2004; Marques et al. 2019). Indeed, extensive histories of hybridization have been documented in many classic and spectacular radiations (The Heliconius Genome Consortium et al. 2012; Lamichhaney et al. 2015; Meier et al. 2017a; Richards and Martin 2017b; Richards et al. 2019). Despite substantial evidence of adaptive introgression during radiation, no previous studies have compared adaptive introgression between closely-related radiating and non-radiating lineages to distinguish introgression as necessary for radiation as predicted by the hybrid swarm hypothesis (but see (Meier et al. 2019) for a comparison of genome-wide introgression).

A parallel debate centers on whether adaptive divergence proceeds in temporal stages. Hypotheses that speciation and adaptive radiation proceed in adaptive stages were previously based on the phylogenetic distribution of extant traits (Diamond 1986; Losos et al. 1995; Danley and Kocher 2001; Streelman et al. 2002; Streelman and Danley 2003; Ronco et al. 2020). For example, one model proposed three stages of vertebrate adaptive radiation: first shifts in habitat, followed by divergence in trophic morphology, and finally sexual communication signals like color (Streelman and Danley 2003). Similarly, the behavior-first hypothesis proposed that behavioral changes drive the evolution of other adaptive traits and promote speciation (Mayr 1963; Lande 1981; Huey et al. 2003a; Losos et al. 2004). However, existing evidence for these hypotheses comes from ancestral state reconstructions of rapidly diversifying traits that can be highly unreliable without fossil data (Glor 2010; Sallan and Friedman 2012; Duchêne and Lanfear 2015). The timing of diversification on a phylogeny is also confounded by different rates of diversification across different trait axes (Glor 2010); thus, rapidly evolving traits appear to have diverged most recently even if divergence occurred continuously in multiple traits.

Microevolutionary studies lend additional support to the idea that adaptation occurs in stages, particularly at the genetic level when the order of selection on mutations necessary to successfully adapt to a novel resource is known (Blount 2016; Blount et al. 2018). However, these studies are limited largely to adaptation in highly experimental settings (Weinreich et al. 2006; Blount et al. 2008, 2012) or along a single selective axis such as toxin resistance (Fry 2014; Tarvin et al. 2017; Karageorgi et al. 2019). Whether stages of adaptation occur along multiple ecological and sexual trait axes remains unknown. Recent population genomic approaches for investigating the timing of selection across multiple trait axes and identifying the sources of genetic variation in recent radiations provide a new opportunity to understand both the temporal and spatial dynamics of adaptation and speciation (Nakagome et al. 2016; Smith et al. 2018; Miller et al. 2019; Richards et al. 2019).

Here we use multiple lines of genomic, transcriptomic, and phenotypic evidence from a nascent adaptive radiation of Caribbean pupfishes to test the hypotheses that hybridization was necessary and strongly associated with adaptive radiation and that adaptive divergence occurred in stages. This sympatric radiation contains a widespread generalist algae-eating species, *Cyprinodon variegatus* (G), and two trophic specialists endemic to several hypersaline lakes on San Salvador Island (SSI), Bahamas: a molluscivore, *C. brontotheroides* (M), with a unique nasal protrusion for oral-shelling snails (John et al. 2020) and a scale-eater, *C. desquamator* (S), with two-fold longer oral jaws (Martin and Wainwright 2013b) and adaptive strike kinematics for efficiently removing scales from the sides of prey fish (St. John et al. 2020). This clade exhibits classic hallmarks of adaptive radiation. First, trait diversification rates reach 1,400 times faster than non-radiating generalist populations on neighboring Bahamian islands in nearly identical hypersaline lake environments (Martin 2016b). Second, species divergence is driven by multiple fitness peaks on a surprisingly stable adaptive landscape (Martin and Wainwright 2013c; Martin 2016a; Martin and Gould 2020). Third, craniofacial diversity and ecological novelty within the radiation exceeds all other cyprinodontid species (Martin and Wainwright 2011; Martin 2016b). Fourth, the radiation exhibits striking divergence in other traits, including male reproductive coloration, aggressive behavior, and feeding preferences (Mcgirr and Martin 2018; St. John et al. 2019, 2020; John et al. 2020).

4.3. Results and Discussion

4.3.1. Similar levels of genetic diversity across radiating and non-radiating lineages of Caribbean pupfish

To investigate the spatiotemporal history of adaptive alleles unique to trophic specialists on SSI we constructed the first de novo hybrid assembly for *C. brontotheroides* (1.16 Gb genome 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 aporus* and *Cualac tessellatus* (Figure 1A;Table S1; Data S1). Population structure across the Caribbean was largely explained by geographic distance (Figure 1) and the SSI radiation did not contain higher overall genetic diversity than the rest of the Caribbean (Figure S1). All Caribbean populations experienced similar declines in effective population sizes following the last glacial maximum 19 kya when an order of magnitude more Caribbean coastal habitat was above sea level (Figure 1D).

We scanned 5.5 million single nucleotide polymorphisms (SNPs) across 202 Caribbean-wide pupfish genomes to identify a set of 3,258 scale-eater and 1,477 molluscivore candidate adaptive alleles, respectively. These candidate adaptive alleles showed evidence of both high genetic differentiation between trophic specialists ($F_{st} \geq 0.95$) and significant signatures of a hard selective sweep in both site frequency spectrum (SFS)-based (Pavlidis et al. 2013) and linkage disequilibrium (LD)-based methods ((Alachiotis et al. 2012); Figure2A; FigureS2; Table S2; Data S2 and S3). We hereafter refer to these as “adaptive alleles”. 45% of the selective sweeps

identified in molluscivores were also identified as selective sweeps in scale-eaters, but contained different fixed or nearly fixed alleles (Figure S3), consistent with previously observed patterns of parallel differential gene expression in trophic specialists relative to generalists despite divergent genotypes (Mcgirr and Martin 2018).

We identified 204 genes within 20-kb of adaptive alleles (median=1; range = 0 to 5 genes; 58% of these adaptive alleles were intergenic). 28% were found in cis-regulatory regions (within 20-kb of genes), 12% in intronic regions, and 2% in coding regions, suggesting a substantial role for gene regulatory evolution in the adaptive radiation (Table S3-S4). GO enrichment analysis of adaptive alleles revealed that 12 of the top 15 most significantly enriched terms (FDR<0.008) were related to neurogenesis, behavior, lipid metabolism, or craniofacial development, consistent with the major trophic axis of diversification in this radiation which includes foraging behavior, dietary physiology, and trophic morphology (Figure 2C; Table S5-7). 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) (Mcgirr and Martin 2018).

We then used genome-wide association mapping to identify adaptive alleles significantly associated with major axes of phenotypic diversification in the trophic specialists: caudal fin melanin pigmentation, the main axis of divergence in reproductive coloration (Figure 1); nasal protrusion distance, a unique craniofacial feature of the molluscivore; and oral jaw size, the most divergent morphological trait of the scale-eater (>99th percentile outliers from GEMMA posterior inclusion probabilities (PIP) while controlling for population structure on SSI: Data S7-S9). 136 scale-eater adaptive alleles were associated with oral jaw size (20 genes) and 21 were associated with caudal fin pigmentation (3 genes; Data S9). 152 molluscivore adaptive alleles were associated with oral jaw size (6 genes) and 108 were associated with nasal protrusion distance (3 genes; Data S7-S8). All adaptive alleles associated with pigmentation and nasal protrusion distance were found in cis-regulatory positions within 20-kb of a gene. 89% and 99% of the adaptive alleles associated with oral jaw size in scale-eaters and molluscivores, respectively, were also cis-regulatory.

4.3.2. Adaptive alleles in trophic specialists are broadly distributed across the Caribbean but are only under selection on San Salvador Island

Even though both trophic specialists are endemic to SSI, we found that nearly all their adaptive alleles occurred as standing genetic variation across the Caribbean (molluscivore: 100%; scale-eater: 98%; Figure2A). Furthermore, nearly half the adaptive alleles in SSI trophic specialists were ancient and also found in *Cualac* or *Megupsilon* outgroups to *Cyprinodon* (41% and 55% of scale-eater and molluscivore adaptive alleles, respectively), which diverged over 5 Mya from *Cyprinodon* (Echelle et al. 2005). However, most adaptive alleles in SSI trophic specialists did not show any evidence of selection in other focal generalist populations (only 2% and 6% of scale-eater and molluscivore alleles, respectively; FigureS3) and strong linkage disequilibrium among adaptive alleles in the SSI trophic specialists was not observed in these populations (Figure 2B). Thus, we conclude that novel trophic specialist traits within a microendemic

adaptive radiation were almost entirely assembled from ancient standing genetic variation through strong selection for new combinations of adaptive alleles.

4.3.3. Stronger signatures of hybridization and adaptive introgression in radiating lineages

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 aids adaptive radiation (Figure 3). First, the strongest signal of introgression across the Caribbean was into the root node of the SSI radiation (Figure 3B). Second, trophic specialists on SSI experienced at least twice as much adaptive introgression as other generalist populations across the Caribbean (Mann-Whitney U test: $P = 0.01$ and 0.03 for scale-eaters and molluscivores, respectively; SI Appendix; Figure3D-E). Third, the distribution of tract lengths for adaptive introgression regions was narrower and these tracts were smaller than the overall distribution of introgressed tracts, suggesting that most adaptive introgression resulted from older hybridization events, despite evidence of recent and continuous introgression to the present (Figure 3E-F).

4.3.4. Timing of divergence and selection supports temporal stages of adaptation

SSI pupfish diversified in their specialized foraging behavior on scales or snails, trophic morphology (craniofacial skeleton and musculature), and in their male reproductive coloration, predominantly through contrasting melanin pigmentation throughout the fins and body (Figure 1, 4). We searched all genes in or near adaptive alleles for GO terms relevant to these traits, including feeding behavior, muscle and craniofacial development, and pigmentation (Ensembl 96; Tables S5-7). With the exception of pigmentation, these terms were all highly enriched GO categories (Figure 2; *SI Appendix*). We then tested for stages of adaptation across these multiple trait axes and spatial sources by estimating the relative ages of selective sweeps for all adaptive alleles in or near genes with behavioral or craniofacial annotations and all introgressed and de novo adaptive alleles. We complemented this approach with divergence time estimates between specialists calculated using a measure of absolute divergence (D_{xy}) in the regions surrounding adaptive alleles. For both approaches, time estimates were scaled by a spontaneous mutation rate per generation of 1.56×10^{-8} per base pair, estimated from high coverage (15-69X) sequences of two independent pedigreed crosses of SSI species and 1-2 offspring.

Similar to patterns observed in the vertebrate stages and behavior-first hypotheses (Mayr 1963; Lande 1981; Huey et al. 2003b; Losos et al. 2004), independent adaptive alleles in the *cis*-regulatory regions of three genes associated with feeding behavior (*prlh, cfap20,atp8a1;* (Takayanagi et al. 2008; Suwa et al. 2010; Choquette et al. 2012; Maia et al. 2014; Taye et al. 2017)) were the first three hard sweeps out of all adaptive alleles in our sample associated with behavior, trophic morphology, or pigmentation (Figure 4). These adaptive alleles were much older than the radiation itself and originated as standing genetic variation in the Caribbean (Figure 4A). Feeding behavior was a significantly enriched category in our GO analysis of all genes associated with adaptive alleles (FDR=0.008;Figure2C). Mutations in the prolactin releasing hormone gene (*prlh),* a hormone involved in stimulating milk production in mammals,

also cause overeating in mice (Takayanagi et al. 2008) and may plausibly cause different foraging preferences between SSI specialists. Concurrently, adaptive alleles in the cis-regulatory allele for cilia and flagella associated protein 20 (*cfap20*) were swept to fixation by selection, which affects cilia formation during brain development. *Cfap20* knockouts in *Drosophila* exhibit behavioral defects that affect foraging performance (Maia et al. 2014). *Cfap20* and *prlh* were also differentially expressed between trophic specialists in 2 and 8 dpf whole embryos respectively, consistent with adaptive alleles in the *cis*-regulatory regions of these genes (Data S4-S5;(Mcgirr and Martin 2018)).

We observed significant 'behavior-first' adaptation in which three of the five hard selective sweeps containing genes associated with feeding behavior occurred before any selective sweeps containing genes associated with craniofacial morphology (permutation test: $n = 22$, $P = 0.01$; Figure4A). This pattern was not due to local mutation rate variation and holds across two independent sweep age estimates and across different populations (Figures. S11, S13-15). This initial stage of adaptive divergence in feeding behavior between trophic specialists suggests that behavioral changes might be crucial for initiating trophic specialization. For example, an increased drive to eat due to divergent regulation of *prlh* might be necessary to gain sufficient nutrition from eating scales (approximately 50% of the scale-eater diet) despite the energetic costs of high-speed strikes (Martin and Wainwright 2013b).

However, more recent sweeps of adaptive alleles near genes associated with feeding regulatory behavior in mice (*rmi, slc16a1;* (Hnasko et al. 2004; Suwa et al. 2010)) and behavior in general (*chrna7, nr4a2, ncoa2*) indicated that adaptive divergence of behavior was not restricted entirely to this first stage. Furthermore, at least two genes associated with adaptive alleles in our dataset (*th, atp8a2*) have pleiotropic effects on both behavior and craniofacial morphology based on GO annotations (Table S5-7). The pleiotropic impacts of alleles across different trait axes and stages of adaptation is likely underestimated due to still poorly understood phenotypic effects of gene regulatory networks, particularly for the complex traits of behavior and craniofacial morphology (Boyle et al. 2017).

In a second stage of adaptive divergence in trophic morphology, adaptive alleles in the regulatory regions of genes associated with muscle development (*smyd1, kcnk2)* and craniofacial morphology (*itga5, tiparp, ext1, cyp26b1, galr2*) swept to fixation from Caribbean-wide standing genetic variation and introgressed from three different outgroup populations across the Caribbean (Figure 4E-F). For example, after the initial sweep of behavioral alleles, standing genetic variation in cis-regulatory region of the gene integrin alpha-5 (*itga5*) swept to fixation in the scale-eater (95% HPD: 639-932 ya; Figure4E). *Itga5* is involved in cranial and pharyngeal arch development in zebrafish (Crump et al. 2004; LaMonica et al. 2015). Concurrently, standing variation and a de novo adaptive allele in the cis-regulatory region of the gene *galr2* swept to fixation in the scale-eater (95% HPD: 696-1,008 ya; Figure4E). *Galr2* is another strong craniofacial candidate in this system because the gene produces a transmembrane receptor for galanin, a peptide known to facilitate bone formation (McGowan et al. 2014), was significantly associated with SSI pupfish oral jaw size (99.6th PIP percentile GEMMA; Figure4C; 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 S8; (Martin et al. 2017)).

Similar to the refinement stage of adaptation proposed in the Long Term Evolution Experiment (Blount et al. 2008), we found evidence for a final refinement stage in the radiation. We observed that two adaptive alleles associated with craniofacial morphology swept to fixation

significantly more recently in the scale-eater than any other behavioral or craniofacial adaptive alleles (*tfap2a* and *twist1* (Zhang et al. 1996; Green et al. 2015; Teng et al. 2018); Figure4E). Intriguingly, one of these alleles was the only non-synonymous substitution that was associated with trophic morphology. This substitution occurred in the second exon of *twist1*, a transcription factor known to affect cleft palate and oral jaw size (Teng et al. 2018). Furthermore, this variant was significantly associated with SSI pupfish oral jaw size in a genome-wide association scan (99.5th PIP percentile GEMMA; Figure4C) and is highly conserved across ray-finned fishes (GERP score: 2.19; FigureS4). The second sweep that occurred at the same time as *twist1* involved adaptive alleles in the regulatory position of the gene *tfap2a*. Selection on these alleles may have occurred recently because *tfap2a* appears to be highly pleiotropic, affecting melanocyte, eye, bone, skin, and neuron development (Brewer et al. 2004; Green et al. 2015; Seberg et al. 2017). A final stage of refinement is known from several theoretical and empirical studies of adaptive walks, in which large effect mutations are selected upon only after mutations that minimize their pleiotropic fitness costs arise and fix in the population (Weinreich et al. 2006; Hague et al. 2018; Karageorgi et al. 2019).

We did not observe a strong initial stage of behavioral divergence in the molluscivores. Adaptive alleles for the molluscivore were found in cis-regulatory regions of only a single gene associated with feeding behavior, motor neuron and eye development, *atp8a2* (Bronson et al. 1992; McMillan et al. 2018). These adaptive alleles are among the oldest standing genetic variation, six times older than the radiation itself (Figure 4B,D). However, the molluscivore also experienced more recent sweeps associated with trophic morphology (Figure 4F). Some of the clearest examples were four adaptive alleles in the cis-regulatory region of *ext1*, a gene that causes osteocartilaginous tumors and cranial abnormalities (Mccormick et al. 1998; Sinha et al. 2017), originating from Caribbean-wide standing genetic variation. These alleles were also strongly associated with nasal protrusion distance in our GWAS (99.9th PIP percentile GEMMA; Data S7) and *ext1* was differentially expressed between trophic specialist whole embryos at 8 dpf (Data S4).

We found no evidence for a distinct stage of diversification in sexual signals, despite the striking contrast in male reproductive coloration between SSI trophic specialists in their overall melanin pigmentation exceeding all other *Cyprinodon* species (Martin and Wainwright 2013b). Instead, adaptive alleles associated with melanin pigmentation swept to fixation throughout the process of adaptive radiation. We identified adaptive cis-regulatory alleles in three genes known to affect pigmentation in model organisms (*tfap2a*, *th*, *card8* (Marles et al. 2003; Osualdo and Reed 2011; Seberg et al. 2017)) and two additional adaptive alleles associated with caudal fin melanin pigmentation in SSI pupfishes (>99th PIP percentile GEMMA; Figure4C-E; Data S7). For example, two adaptive alleles in the regulatory region of *card8* originating from standing variation swept to fixation early in the scale-eaters (95% HPD: 866-1,491ya). This gene, which is associated with vitiligo and pigmentation loss in humans (Osualdo and Reed 2011), was significantly associated with pupfish caudal fin pigmentation (99.2th PIP percentile GEMMA; Figure4C; Data S7), and was differentially expressed between trophic specialists in 2 and 8 dpf whole embryos (Data S4-S5,S7). Adaptive alleles in the cis-regulatory regions of *tfap2a* and *th* also swept to fixation in the scale-eaters more recently from standing variation (95% HPD: 292-431 ya; 746-958 ya respectively; Figure4E). Gene knockouts of *tfap2a* cause skin pigmentation defects in mice (Brewer et al. 2004; Seberg et al. 2017) and the gene *th* is associated with skin pigmentation defects in human (Marles et al. 2003) and cuticle pigmentation defects in insects (Qiao et al. 2016). This broad range of sweep ages across candidate pigmentation adaptive

alleles suggests that the distinctive light/dark reproductive coloration associated with trophic specialists diverged throughout the course of adaptive radiation on SSI while repeatedly drawing from existing Caribbean-wide standing genetic variation, rather than during a final stage. This is consistent with the necessary role of pre-mating isolation for adaptation to divergent niches in sympatry (Kirkpatrick and Ravigné 2002; Bolnick and Fitzpatrick 2007; Servedio and Boughman 2017).

4.3.5. Microendemic radiation was assembled from spatially disjunct pools of adaptive alleles

Along with temporal stages of adaptive divergence, we also found disjunct spatial patterns in the sources of adaptive introgression. Introgression from different regions of the Caribbean brought in adaptive alleles for different major axes of phenotypic diversification within the radiation. Adaptive alleles near genes annotated for feeding behavior (*prlh;* Figure4A) and oral jaw size (*cyp26b1* (Laue et al. 2008; Spoorendonk et al. 2008); 99.8th PIP percentile GWAS; Figure4D; Data S8) originated in the northwestern Bahamas (New Providence Island, Exumas, and Cat Island) whereas adaptive introgression of alleles near genes associated with muscle and eye development originated in the Dominican Republic (*cenpf, eya2;* Figure4A). Selective sweeps of adaptive alleles from different sources also occurred at largely distinct times during the radiation (Figure 4G-H, Figure S5). For example, selective sweeps of adaptive alleles from North Carolina were significantly older sweeps of introgressed variation than other populations (ANOVA, $P < 0.03$, SI Appendix, Figure 4G-H). This surprisingly disjointed spatiotemporal patchwork of adaptive introgression across the Caribbean suggests that the extant SSI radiation of trophic specialists was reassembled from distinct pools of genetic variation. Our results are consistent with at least two distinct environmental refugia in other regions of the Caribbean, perhaps due to previous ephemeral adaptive radiations, bridging micro- and macroevolutionary-scale processes (Rosenblum et al. 2012; Martin and Richards 2019).

4.3.6. Microendemic adaptive radiations originate over vast expanses of space and time

We conclude that hybridization substantially contributed to an adaptive radiation of trophic specialist pupfishes endemic to a single island and that the resulting adaptive divergence occurred in stages. The radiation originated from a largely ancient set of alleles maintained within different pools of standing variation in Caribbean and mainland generalist populations. Temporal stages of adaptation observed in this nascent radiation are consistent with selection on behavioral divergence first. Adaptive divergence in trophic morphology occurred next, followed by a final stage of refinement including a non-synonymous substitution in the scale-eaters within a craniofacial transcription factor.

Additionally, our study provides a unique look at the spatial dynamics of alleles involved in adaptive divergence. We found that most adaptive alleles contributing to the major axes of ecological and sexual diversification in this radiation existed in Caribbean generalist populations long before the trophic specialist species on SSI diverged. This genetic variation is distributed across two orders of magnitude larger spatial and temporal scales than the 10 kya radiation endemic to a single 20 km island. 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.

Our understanding of the origins of adaptive radiation in this system remains incomplete. The presence of adaptive introgression in both radiating and non-radiating lineages suggests hybridization was necessary but perhaps not sufficient to trigger adaptive radiation. More subtle and transient factors such as ecosystem productivity or intermediate "stepping-stone" species needed to access novel fitness peaks (see (Richards and Martin 2017a)) might help explain why pupfish adaptive radiations are restricted to SSI and only one other known lake in the Caribbean (Martin and Wainwright 2011). Our study highlights the utility and necessity of including closely related outgroups as controls in testing hypotheses about the mechanisms underlying adaptive radiation. The coincidence of hybridization, ecological opportunity, and sexual selection appear to be the best predictors of adaptive radiation in general (Wagner et al. 2012; Marques et al. 2019; Meier et al. 2019). Most adaptive radiations, including stickleback, African cichlids, Lake Tana barbs, *Anolis* lizards, *Heliconius* butterflies, Hawaiian tetragnathids, and *Brocchinia* bromeliads, share similar patterns of spatial nesting within a widespread clade and intermediate levels of population structure and admixture (Comeault and Matute 2018; Martin and Richards 2019; Gillespie et al. 2020) and we expect similar dynamics to the pupfish system. Research into the broader spatiotemporal landscape of radiations can provide insights about longstanding hypotheses of adaptive radiation and their contributions to global patterns of biodiversity.

4.4. Materials and Methods

4.4.1. Sampling and population genotyping

Cyprinodon pupfishes were collected from across their entire Atlantic and Caribbean range from Massachusetts to Venezuela. Individual DNA samples were resequenced using Illumina Hiseq4000 and Novaseq. We first constructed a hybrid de novo assembly of the *Cyprinodon brontotheroides* genome (1.16 Gb genome size; scaffold N50 = 32 Mb; details in *SI Appendix*). All reads were aligned to this assembly and variants were called and filtered following the best practices guide recommended in the Genome Analysis Toolkit (DePristo et al. 2011). The final dataset used in downstream analyses included 5.5 million SNPs from 202 individuals sampled from 39 localities. This included population level sampling ($n > 8$ individuals) for the three SSI species and five generalist outgroup populations (Figure 1A). To visualize population structure and admixture among Caribbean populations in our dataset, we ran a principle component analysis using the eigenvectors from plink's pca function (v.1.9; (Purcell et al. 2007)) and estimated the proportion of shared ancestry among individuals using ADMIXTURE (v.1.3.0; (Alexander et al. 2009)) on an LD-pruned subset of 2.6 million alleles (*SI Appendix*).

4.4.2. Classification of adaptive alleles in the SSI radiation

We first identified fixed or nearly fixed alleles ($F_{st} \geq 0.95$ between specialists: 4,189) that also showed significant evidence of a hard selective sweep in either specialist population in both the SFS-based and LD-based methods SweeD and OmegaPlus (significance thresholds based on neutral simulations with ms-move: CLR > 4.47 & $\omega > 3.31$; Figure2A; Tables S2-S4, S10). We then characterized the potential function of these adaptive alleles in the specialists in three ways. First, we performed gene ontology (GO) enrichment analyses for genes within 20-kb of adaptive alleles using ShinyGo (Ge and Jung 2018). Additionally, we looked for overlap between genes associated with adaptive alleles and genes previously found to be differentially expressed between the two specialists in whole embryos at 2 and 8 dpf (Mcgirr and Martin 2018). Lastly, we employed a Bayesian sparse linear mixed model (BSLMM) implemented in the GEMMA software package (Zhou et al. 2013) to identify genomic regions associated with variation in lower oral jaw length, caudal fin pigmentation, and nasal protrusion distance across all 78 wild-collected samples of the three SSI species (*SI Appendix*).

4.4.3. Characterizing introgression and spatial distributions of adaptive alleles across the Caribbean

We identified introgression in the specialists in the SSI radiation and generalist outgroup populations on both a genome-wide and local level using Treemix (Pritchard et al. 2010) and the f_d statistic ((Martin et al. 2015b); Figure3). For both specialists, we then looked for introgressed regions that also showed evidence of a hard selective sweep to characterize candidate adaptive introgression regions. For comparison, we search for similar signatures of adaptive introgression in three Caribbean outgroup generalist populations (excluding North Carolina and Venezuela due to lack of equivalent set of outgroup taxa for a four-population test of introgression in them; Table S12; *SI Appendix*).

We surveyed all pupfish populations in our dataset for the scale-eater and molluscivore adaptive alleles. These adaptive alleles were separated into three categories: de novo (observed only on San Salvador Island), standing genetic variation (observed in at least one generalist population outside of San Salvador Island), or introgressed (outlier f_d tests for introgression, significance-based on no-migration simulations with ms-move; $f_d > 0.72$; Table S11-16; SI Appendix). Introgressed adaptive alleles were further separated by geographic source (Fort Fisher, North Carolina, United States (NC); Lake Cunningham, New Providence Island, Bahamas (NP); or Laguna Bavaro, Dominican Republic (DR)).

4.4.4. Detection of stages of adaptation through divergence times and timing of selective sweeps

For all fixed or nearly fixed alleles, we estimated the timing of divergence in the 50-kb region surrounding the variant based on the amount of genetic variation that accumulated between the two specialists. All time estimates were converted to years using a pupfish generation time of one year (Martin 2016b) and a spontaneous mutation rate (1.56×10^{-8}) based on two independent sets of pedigree crosses of SSI pupfishes (Table S9; *SI Appendix*).

We estimated the age of the selective sweeps for all adaptive alleles near or within genes annotated for behavior or trophic morphology GO terms (eye, musculature, mouth, or craniofacial development) from our GO enrichment analyses using coalescent approaches implemented in starTMRCA and McSwan ((Smith et al. 2018; Tournebize et al. 2019); Table S17-19; *SI Appendix*). We also estimated sweep ages for the entire set of de novo and introgressed adaptive alleles. We then compared the 95% highest posterior density (HPD) intervals of age estimates to visualize temporal stages of adaptation across different spatial sources of genetic variation.

4.5. Chapter 4 Figures

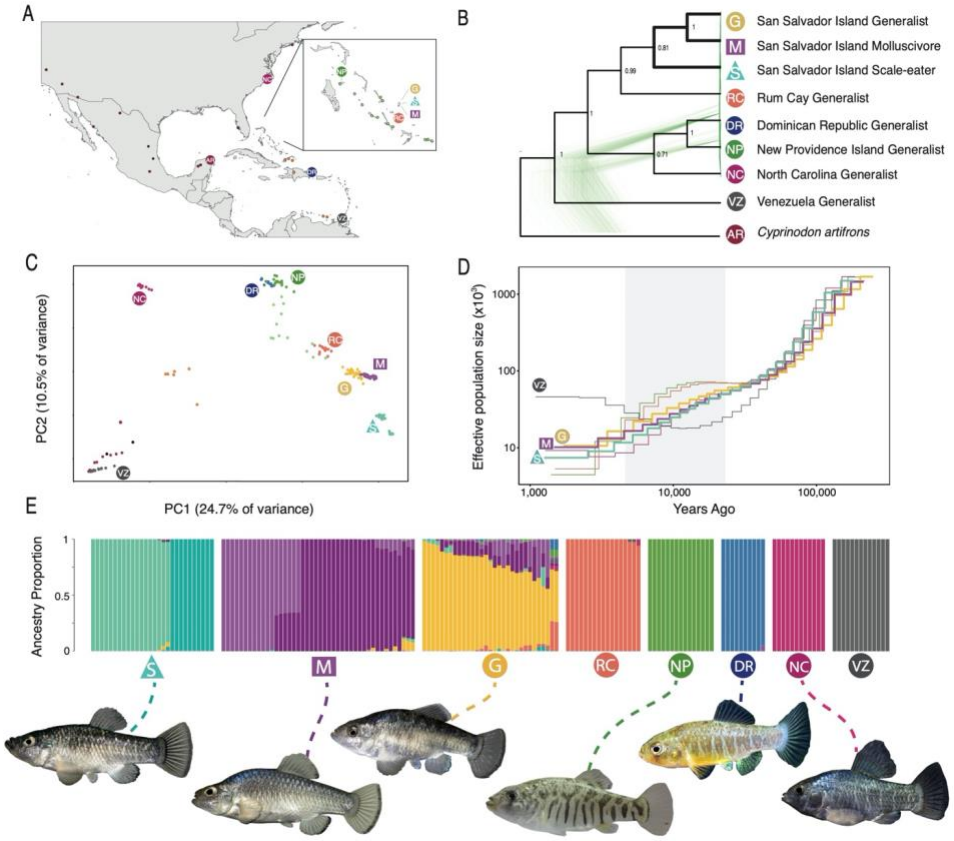


Figure1. Genetic diversity of pupfishes across the Caribbean. A) Sample locations of *Cyprinodon* pupfishes (n = 202) with eight focal populations ($n \geq 10$ per population) marked with large symbols (circles: generalist populations, triangle: scale-eating pupfish, square: molluscivore pupfish) and small dots for individuals from other locations in the Bahamas (light green), other Caribbean locations outside the Bahamas (orange), continental North and South America (maroon), and *Megupsilon* and *Cualac* outgroups to *Cyprinodon* (black). B) Maximum clade credibility cladogram (black) estimated with SNAPP (Bryant et al. 2012) from 10k SNPs for focal populations and the outgroup *Cyprinodon artifrons*. The underlying 500 gene trees (green) randomly sampled from the posterior distribution and visualized with Densitree (Bouckaert 2010) demonstrate the rapid and complex history of divergence in Caribbean pupfishes. C) Principal component analysis of *Cyprinodon* pupfishes. D) Changes in effective population size over time for focal populations in the Caribbean inferred using MSMC (Schiffels and Durbin 2014). The gray shaded area represents the range of estimated ages for the radiation from 6-19 kya based on the filling of hypersaline lakes on SSI after the last glacial maximum (Hagey and Mylroie 1995; Turner et al. 2008; Clark et al. 2009). E) Ancestry proportions across individuals on San Salvador Island, Bahamas (SSI: S, M, G) and 5 other focal Caribbean generalist populations estimated from an LD-pruned dataset of 2.3 million SNPs in ADMIXTURE (Alexander et al. 2009) with k=11.

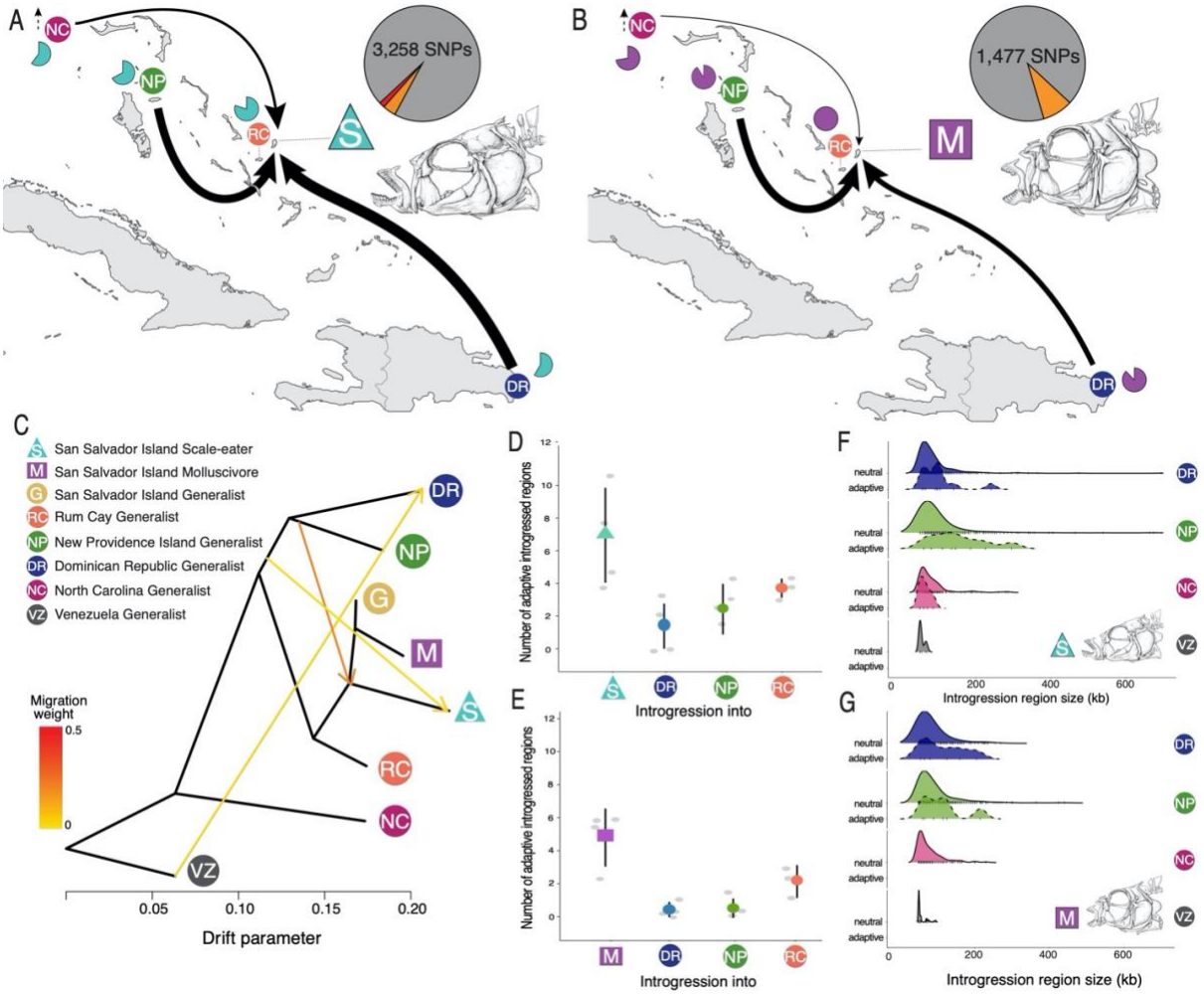


Figure3. History of hybridization across the Caribbean. A-B) Summary map of adaptive introgression into SSI trophic specialists from focal generalist populations across the Caribbean. Arrow thickness is proportional to the number of adaptive introgression regions (outliers based on ms-move simulations with no migration; $f_d > 0.71$; Table S10). Pie charts represent the percentage of all scale-eater (A; teal) and molluscivore (B; purple) adaptive alleles segregating in outgroup generalist populations across the Caribbean. C) Genome-wide population graph inferred from *Treemix* with the three strongest signals of introgression. Note that the strongest signal (in orange) is into the root node of the SSI radiation. D-E) Number of adaptive introgression regions in each focal population (bootstrapped mean and 95% confidence interval in black). Grey dots represent the number of adaptive introgression regions in each focal population from each of the four source populations (Table S12). Note that the total number of adaptive introgression regions is often larger in specialists (S, M) than in other outgroup generalist populations. F-G) Density plots of the tract lengths of adaptive introgression regions (f_d and selective sweep outliers: dashed line) and all introgression regions (f_d outliers only: solid line). Small tick marks below density plots indicate observed introgression regions.

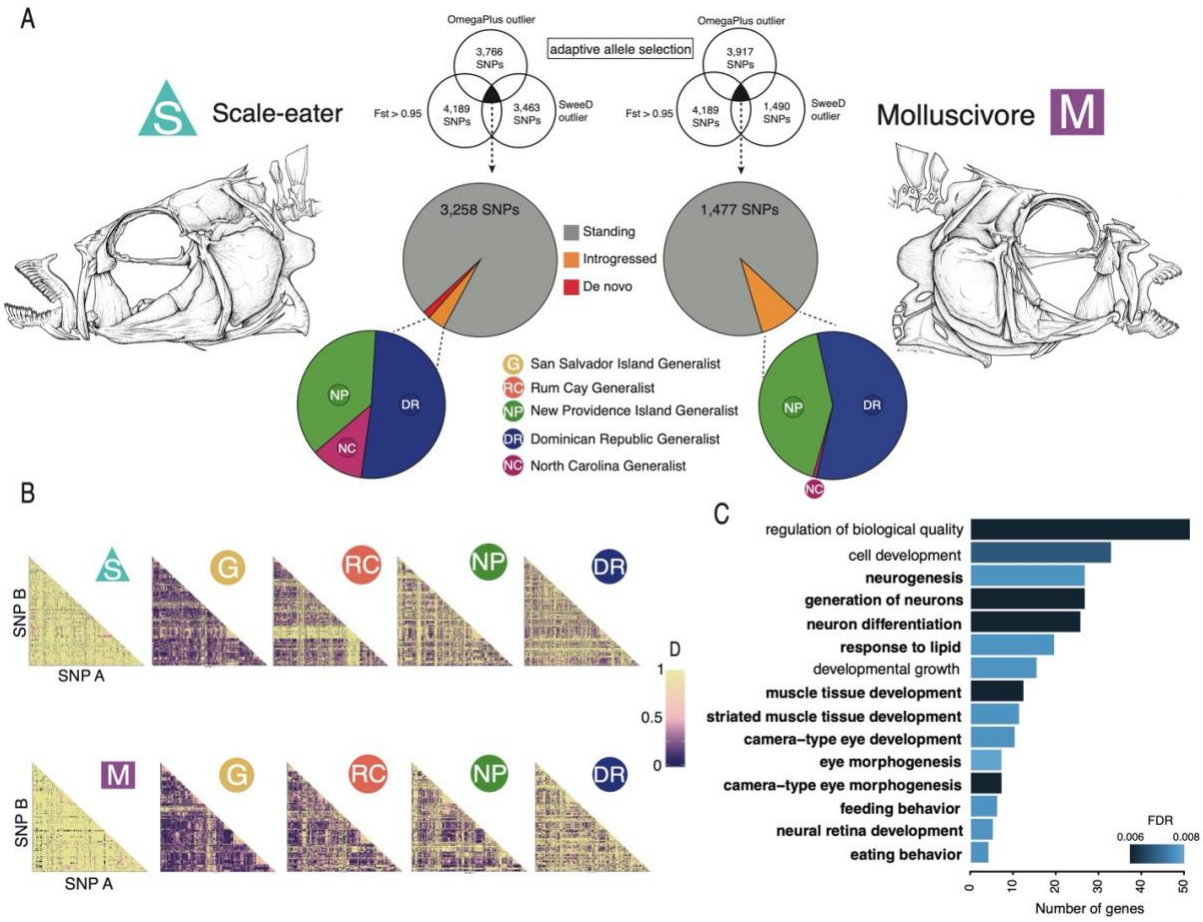


Figure2. All adaptive alleles differentiating trophic specialist pupfishes. A) All fixed and nearly-fixed adaptive alleles in each trophic specialist on SSI (filtered by $F_{st} \geq 0.95$ between specialists and evidence of a hard selective sweep in both an SFS- and LD-based approach (SweeD CLR≥ 4.47; OmegaPlus: ω >3.31)). Adaptive alleles are divided into the three major sources of genetic variation: de novo variation found on SSI only (red), introgression from a specific source population (orange), or standing genetic variation (grey). Introgressed variation is further broken down by source population: New Providence Island, Bahamas (green), Dominican Republic (blue), or North Carolina (magenta; Tables S13-S16). B) Heatmaps of linkage disequilibrium among all pairwise combinations (designated SNP A vs SNP B) of adaptive alleles for scale-eaters (S; top row) and molluscivores (M; bottom row) on SSI in comparison to linkage among these SNPs in generalists on SSI (G) and three other focal generalist populations across the Caribbean. Note the breakdown in linkage disequilibrium among adaptive alleles for trophic specialists in all generalist populations examined. C) Top 15 GO categories in which scale-eater adaptive alleles were significantly enriched, with relevant terms corresponding to the major axis of trophic divergence in this radiation highlighted in bold (FDR < 0.01; full list of terms with FDR < 0.05 in Table S5). Skull illustrations by Sean V. Edgerton.

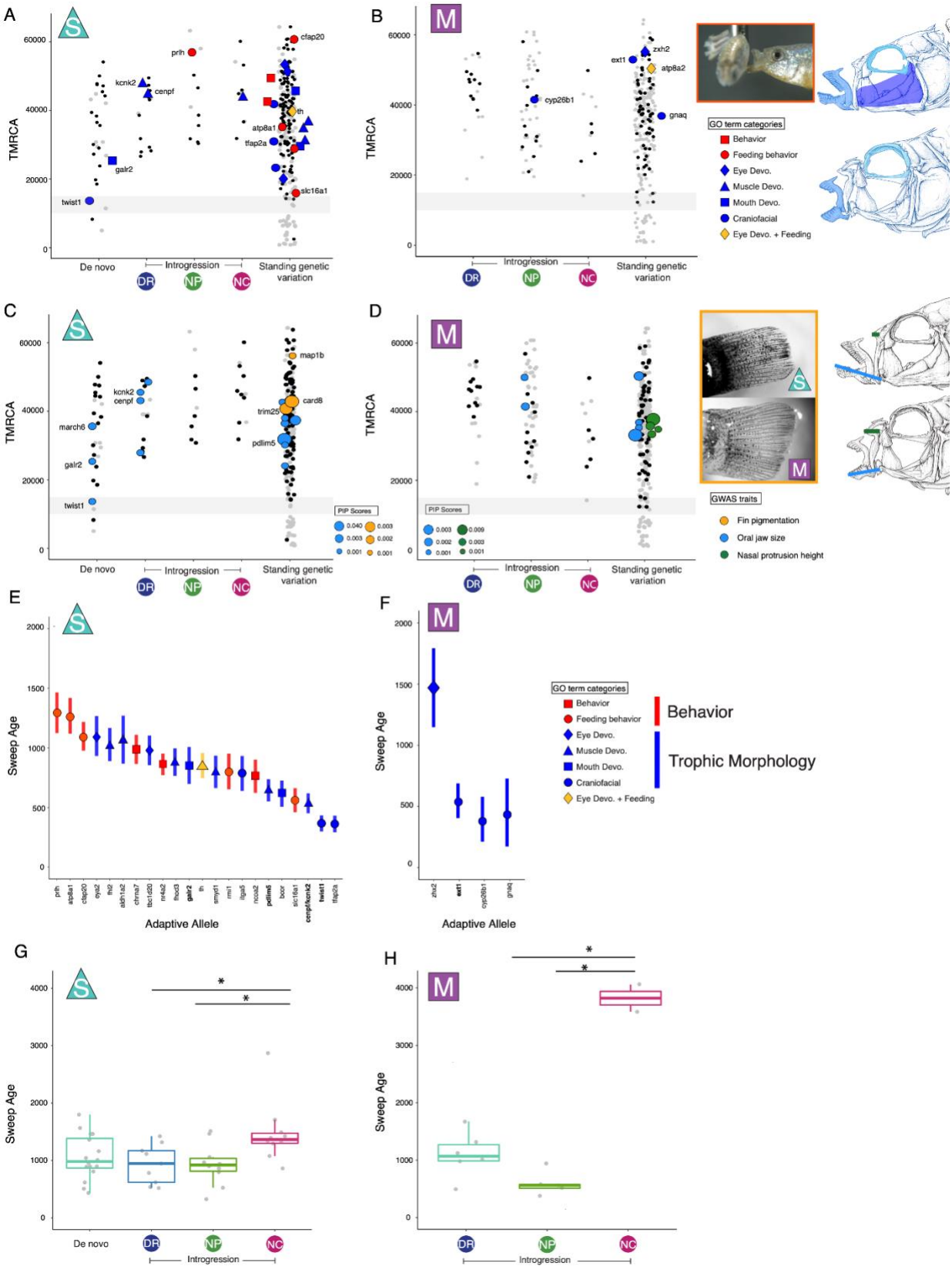

A
S
TMRCA
60000
40000
20000
0
prlh
cfap20
kcnk2
cenpf
th
atp8a1
tfap2a
galr2
slc16a1
twist1
De novo
Introgression
Standing genetic variation
DR
NP
NC
B
M
ext1
zxh2
atp8a2
cyp26b1
gnaq
GO term categories
Behavior
Feeding behavior
Eye Devo.
Muscle Devo.
Mouth Devo.
Craniofacial
Eye Devo. + Feeding
C
map1b
kcnk2
cenpf
card8
trim25
march6
pdlim5
galr2
twist1
PIP Scores
0.040
0.003
0.003
0.002
0.001
0.001
D
S
M
GWAS traits
Fin pigmentation
Oral jaw size
Nasal protrusion height
E
Sweep Age
2000
1500
1000
500
0
Adaptive Allele
F
GO term categories
Behavior
Feeding behavior
Eye Devo.
Muscle Devo.
Mouth Devo.
Craniofacial
Eye Devo. + Feeding
Behavior
Trophic Morphology
G
H
4000
3000
2000
1000
0
*
*
De novo
Introgression

Figure4. The spatiotemporal landscape of adaptive radiation. A-B) Time to most recent common ancestor (TMRCA) of adaptive alleles based on D_{xy} in the 50-kb windows containing adaptive alleles. Each column separates adaptive alleles by their spatial distribution: de novo (SSI only), adaptive introgression from one of three outgroup populations (DR: Dominican Republic, NP: New Providence, NC: North Carolina), and standing genetic variation. Gray bars highlight the approximate origins of the microendemic radiation on SSI at approximately 6-19 kya (based on geological age estimates for filling of hypersaline lakes on SSI (Hagey and Mylroie 1995; Turner et al. 2008) since the last glacial maximum (Clark et al. 2009)). All adaptive alleles associated with genes for behavior (red) and craniofacial morphology (blue) are illustrated by a colored point. Black points show adaptive alleles for non-focal GO terms or unannotated; gray points show all fixed or nearly fixed alleles between specialists ($F_{st} \geq 0.95$) with no signal of a hard selective sweep; and triangles show adaptive alleles associated with pigmentation. Genes discussed in the text are labeled by their associated adaptive allele. C-D) Adaptive alleles are colored by significant association (>99th PIP percentile GEMMA) with oral jaw size, caudal fin pigmentation, and nasal protrusion distance in scale-eaters (C) and molluscivores (D). Dot sizes scale with PIP score. E-F) 95% HPD intervals for selective sweep ages in the scale-eaters and molluscivores. Adaptive alleles within 20-kb of genes are colored by their GO annotations. Bolded genes were significantly associated with oral jaw size in (E) scale-eaters or nasal protrusion distance in (F) molluscivores in the GWAS analysis (>99th percentile PIP GEMMA; Data S7-9). G-H) Boxplots of selective sweep ages across de novo and introgressed adaptive alleles from the three focal outgroup generalist populations. Asterisks indicate significant differences in sweep ages between different source populations (ANOVA, $P < 0.03$).

4.6. Supplemental Materials

4.6.1. Materials and Methods

4.6.1.1. Sampling.

Pupfishes were collected from across the complete Atlantic and Caribbean range of *Cyprinodon* from Massachusetts to Venezuela. For the three species in the SSI radiation, individual pupfish were collected from 15 isolated hypersaline lakes on SSI (Table S1; Data S1) and one estuary (Pigeon Creek) using hand and seine nets between 2011 and 2018. We sequenced 36 *Cyprinodon variegatus*, 47 *C. brontotheroides*, and 39 *C. desquamator* across these lakes, including six lakes in which one or two specialist species occur in sympatry with the generalist (Crescent Pond, Storr's Lake, Little Lake, Oyster Pond, Osprey Lake, Moon Rock Pond). We also sequenced outgroup high-coverage focal populations of generalist pupfish including 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 sequenced 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* ((Echelle et al. 2005; Valdés González et al. 2020), Figure1A; Table S1; Data S1).

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.

4.6.1.2 Genomic Library Prep.

DNA was extracted from muscle tissue using DNeasy Blood and Tissue kits (Qiagen, Inc.) and quantified on a Qubit 3.0 fluorometer (Thermofisher Scientific, Inc.). Genomic libraries were prepared using the automated Apollo 324 system (WaterGen Biosystems, Inc.) at the Vincent J. Coates Genomic Sequencing Center (QB3). Samples were fragmented using Covaris sonication, barcoded with Illumina indices, and quality checked using a Fragment Analyzer (Advanced Analytical Technologies, Inc.). Nine to ten samples were pooled per lane for 150PE sequencing on four lanes of an Illumina Hiseq4000 and an additional 96 individuals were sequenced on one 150PE lane of Illumina Novaseq with S4 chemistry. This included 42 individuals from a previous genomic study (Richards and Martin 2017b).

4.6.1.3 De novo genome assembly and annotation.

We constructed a hybrid de novo assembly from an inbred lab-raised individual of *C. brontotheroides* using three different sequencing technologies: Oxford Nanopore sequencing was performed at UNC's High Throughput Sequencing Facility, a 10X Genomics synthetic long-read library was prepared and sequenced by Hudson Alpha, and Chicago and HiC libraries were prepared and sequenced by Dovetail Genomics. Genomic DNA was extracted from an inbred F4 male *C. brontotheroides* individual, an offspring from three generations of full-sib mating in the lab, starting with an F0 pair collected from Crescent Pond, SSI (the type locality:(Martin and Wainwright 2013b)). 10X sequencing was performed on this individual according to 10X Genomics' recommended protocol and sequenced on an Illumina HiSeq4000, resulting in 460 million 2x150 bp reads. DNA was extracted from this same molluscivore individual for Nanopore sequencing using a modified phenol:chloroform extraction protocol (Green and Sambrook 2017). Two libraries were sequenced on R9.4 flow cells on Nanopore's GridION desktop sequencer – 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, (Weisenfeld et al. 2018)) to produce a preliminary "pseudohap" assembly. Nanopore reads were corrected using FMLRC (Wang et al. 2018). The Supernova assembly was scaffolded with corrected nanopore reads using LINKS (Warren et al. 2015) with 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; (Durand et al. 2016)), followed by scaffolding with 3D-DNA (v180922) (Dudchenko et al. 2017). We performed a final polishing with four rounds of Racon (v1.3.1; (Vaser et al. 2017)) using the corrected Nanopore reads. The final assembly consisted of 1.16 Gbp in 15,698 scaffolds with an N50 of 32,013,756 bp (32 Mb).

To validate our assembly, we ran BUSCO (v3.0.1) (Simão et al. 2015) to identify known single-copy conserved genes. We found 86.4% of BUSCOs in the Actinopterygii class assembled completely, and 83.4% into single copy orthologs. We annotated this assembly using the Maker pipeline (v3.01.02)(Cantarel et al. 2008), providing alternate ESTs and protein evidence for ab-initio gene prediction from *C. variegatus* (Lencer et al. 2017), which is closely related and expected to have very similar genic structure and codon usage. Predicted genes were assigned putative function by aligning (BLASTp) to the UniProt database (Bateman 2019).

4.6.1.4 Population genotyping.

Raw reads were mapped from 222 individuals to our de novo assembly of the *Cyprinodon brontotheroides* reference genome (v 1.0; total sequence length = 1,162,855,435 bp; number of scaffolds = 15,698, scaffold N50 = 32 Mb) with bwa-mem (v 0.7.12; (Li and Durbin 2011)). 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; (DePristo et al. 2011)) 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; (DePristo et al. 2011; Marsden et al. 2014)) because we lacked high-quality known alleles for these non-model species. Poorly mapped regions were removed using a mask file generated from the program SNPable (http://bit.ly/snpable; k-mer length = 50, and 'stringency' = 0.5). SNPs for SSI individuals were additionally filtered to remove those 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; (Danecek et al. 2011b)). This set of 9.3 million SNPs was then further filtered for alleles that had minor allele frequencies above 0.05 and less than 50% missing data across all Caribbean outgroup individuals with population level sampling. The resulting dataset that we used for all downstream analyses, unless otherwise noted, contained 5.5 million SNPs. The MAF threshold we used as a quality filter (excluding minor allele frequencies below 5%) will bias any search for rare alleles in this system. However, our main objective in this study was to characterize candidate adaptive alleles that have swept within specialist populations on SSI, alleles that would not be influenced by this MAF filter because they are not expected to be rare alleles within our specialist populations of interest. For some calculations that are heavily influenced by the presence/absence of minor alleles, such as *Dxy*, π, and allele frequency distributions across Caribbean populations we used a version of the genetic dataset without the minor allele frequency filter and note when we have done so.

4.6.1.5 Population genetic analyses.

The filtered genomic dataset was first pruned to SNPs in linkage disequilibrium using the LD pruning function (--indep-pairwise 50 5 0.5) in plink (v1.9)(Purcell et al. 2007), leaving 2.6 million SNPs. To visualize population structure in our dataset, we ran a principal 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)(Alexander et al. 2009). The number of populations (K) was chosen

using ADMIXTURE's cross-validation method (--cv) across 1-20 values of K. K = 11 populations was then chosen using the broken-stick method, following (Evanno et al. 2005). 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 SSI, 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 SSI, 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 (Figure 1E). None of the 37 single individuals from other locations were removed. The final dataset used in downstream analyses included 202 individuals.

For analyses of genetic variation within sliding windows, we used a window size of 50-kb based on the extent of linkage disequilibrium (LD) along a scaffold estimated by LD decay along the largest scaffold in our genome. We calculated LD decay from pairwise calculations of LD between all SNPs within 100-kb of each other along the largest scaffold using PLINK's LD function (--r^2). Linkage disequilibrium decayed to background rates after 50-kb at a threshold of $r^2 \geq 0.1$ (Figure S6).

Within-population nucleotide diversity (π) was calculated in 50-kb windows across the genome for each of eight focal populations (>10 individuals resequenced) using the python script popGenWindows.py available from https://github.com/simonhmartin/genomics_general (Martin et al. 2015b). Since this calculation can be heavily influenced by minor alleles, we calculated π without the 5% minor allele frequency filter. Instead, we filtered all minor alleles with a read depth less than 5 in order to remove any rare variants that may be the result of sequencing error rather than a true minor allele, resulting in 10.8 million variants. We then calculated *Dxy* and π in sliding windows. The number of nonvariant sites in each window was also factored into these calculations. To ensure equal sample sizes among populations, we downsampled individuals from each population to the number of individuals in the focal population with the lowest sampling (n = 10). We randomly selected 10 individuals from each population before calculating π in sliding windows. We repeated this 100 times and averaged π across the replicates (Figure 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 were small in all comparisons except in the case of two comparisons. The SSI generalist population had a significantly greater genome-wide genetic diversity of an appreciable effect size compared to North Carolina (*Cohen's d*=0.87) and Venezuela generalist populations (*Cohen's d*=1.38). The significantly lower within-population genetic diversity in Venezuela than other generalist populations may be due to a recent population bottleneck that was not observed in any other populations (Figure 1C and S1).

Finally, allowing for some admixture, we calculated highly differentiated SNPs between trophic specialists based on $F_{st} \geq 0.95$ (Figure S2; Table S2-S4; Data S2-S3). F_{st} between the two specialist populations was calculated per variant site using –weir-pop-fist function in vcftools (v.0.1.15; (Danecek et al. 2011b)) on the 5.5 million variant dataset.

4.6.1.6 Mutation rate estimation

The spontaneous mutation rate for Caribbean pupfishes was estimated from moderate to high-coverage sequencing (15-69x) of parents and offspring from two independent pedigreed crosses of SSI species: one cross between a second generation inbred lab-reared generalist and third-generation inbred lab-reared molluscivore individual from Little Lake (*C. variegatus x C. brontotheroides*) and another between a second-generation lab-reared generalist and second-generation lab-reared scale-eater from Little Lake (*C. variegatus x C. desquamator*). Using the same pipeline for alignment to the *C. brontotheroides* reference genome and variant calling as above, we obtained 9 million SNPs 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). Following the mutation rate estimation protocol outlined in (Feng et al. 2017), we independently called alleles 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 alleles (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 by identifying sites where an offspring was heterozygous for an allele not found in either of the parents. We first looked for alleles which were heterozygous in the offspring and alternately homozygous in the parents (i.e. known heterozygous sites). Ten 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 following similar pipelines and filters from several previous studies (Feng et al. 2017; Malinsky et al. 2018; Wu et al. 2020). This included filtering by 1) genotype quality, 2) mapping quality, 3) base quality rank sum, 4) mapping quality rank sum, 5) quality by depth, 6) site depth, 7) allele depth, 8) read position rank sum, 9) strands odds ratios, and 10) fisher strand scores. Sites were filtered to those greater than or equal to the mean score for known heterozygous sites in the offspring for filters 1 and 2 and scores within 2 standard deviations of the mean score for filters 3-10. 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 specific values used for thresholds reported in Table S9). 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 the range of known heterozygous sites (Table S9) and an allele balance score between 0.3 and 0.7. This set of alleles was then filtered for those independently called in both GATK and samtools following (Feng et al. 2017).

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 the same variant quality filters as for the 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). Genomic regions were excluded if 1) read map depth for a variant was not within two standard deviations of the average read map

depth (varies by sample; Table S9), 2) mapping quality scores were less than 50, or 3) base quality scores were less than 30.

Since the de novo mutations observed could have originated on either chromosome, the point estimate of the per site mutation rate is the number of new mutations observed divided by two times the size of the accessible genome, following (Wu et al. 2020). The mutation rates were then averaged across individual offspring for each cross (Table S9) 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 (Guo et al. 2013; Kautt et al. 2016a; Malinsky et al. 2018); however, short-lived smaller species with higher metabolism rates like pupfishes are expected to exhibit faster mutation rates (Martin et al. 2016). We estimated generation times in the field to be approximately one year based on laboratory and field (Martin et al. 2019) longevity studies.

4.6.1.7 Demographic Inferences

Various demographic histories can shift the distribution of low- and high-frequency derived alleles 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 our seven focal populations. We ran MSMC on unphased GATK-called genotypes separately for a high-coverage individual in each of seven focal populations (excluding generalist *C. higuey* due to poor sequencing quality of our single high-coverage individual; 17-28x mean coverage across individuals; Figure 1D;Table S10). As recommended in the MSMC documentation, we masked 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 less than 10 reads as recommended by Nadachowska-Brzyska et al. (Nadachowska-Brzyska et al. 2016). To scale the output of MSMC to real time and effective population sizes, we used a one-year generation time (Martin et al. 2016) and the estimated spontaneous mutation rate of 1.56×10^{-8} per generation per base pair for Caribbean pupfishes (see previous section).

4.6.1.8 Introgression in SSI specialists

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; (Pickrell and Pritchard 2012)). *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 with 0 through 20 migration edges. The most likely number of migration events was chosen using the broken-stick approach (Figure 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 (Martin et al. 2015b). 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 (Martin et al. 2015b). We used 2 individuals of *C. artifrons* from Cancun, Mexico as our outgroup population for this test, which forms the deepest divergence event with *C. variegatus* within the *Cyprinodon* clade (Echelle et al. 2005), and focused on introgression between SSI specialists and outgroup Caribbean generalist populations. Based on the tree (((P1,P2),P3), *C. artifrons*), the f_d statistic was calculated for the combinations of populations in which the focal population (P2) was either the scale-eater or the molluscivore, the other specialist population was the sister group (P1), and P3 was one of the Caribbean outgroup populations (Table S11 and S12). f_d statistics were calculated from 50-kb sliding windows with a minimum of 100 variant sites and no missing data in a population using the ABBABABA.py script (available on https://github.com/simonhmartin/genomics_general;(Martin et al. 2015b)). 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 for focal generalist populations (whenever we had sister groups to fit the relationships necessary for the test (Table S12B and S12D)).

Significance of f_d values in sliding windows across the genome was evaluated using simulations with no migration using ms-move (Garrigan and Geneva 2014). We used estimates of changes in effective population size for each population from our MSMC analyses. We set the divergence time between the two specialists to 10,000 years based on the age of the hypersaline lakes on SSI . The threshold for significant introgression regions was determined by simulating f_d statistics across the genome under a coalescent model with no gene flow, consisting of 150,000 50-kb windows each containing the mean number of alleles observed in our dataset. Empirical windows were considered candidates for introgression if the f_d statistic was above the maximum simulated f_d value (Table S11). We merged consecutive 50-kb f_d outlier windows to estimate the sizes of introgressed regions and approximate the age of introgression events (Figure 3E-F).

4.6.1.9 Search for candidate adaptive alleles in SSI specialists

4.6.1.9.1 Selective sweep detection.

We searched for hard selective sweeps in the trophic specialist populations using two different approaches. The first method is based on the site frequency spectrum (SFS) calculated with SweeD (v.3.3.4;(Pavlidis et al. 2013)). This method calculates 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 (Garrigan and Geneva 2014), controlling for the impact of the inferred population size decreases over time for each population from MSMC runs mentioned above (Figure 1D; Table S7). The CLR ratios for the simulated datasets were then used to assess outlier CLR ratios from the empirical dataset. We considered regions with CLR ratios above the 95th percentile value of CLR from the neutral simulated dataset as candidate hard selective sweep regions (scale-eater: CLR > 5.28; molluscivore: CLR > 4.47; Table S7). We also inferred candidate hard selective sweep

regions for the five focal Caribbean generalist populations (sample size ≥ 10) following the same method outlined above for the specialists (Table S10).

To complement our SweeD selection analyses, we also used an LD-based approach for detecting hard selective sweeps implemented in OmegaPlus (Alachiotis et al. 2012). OmegaPlus implements the ω-statistic introduced in (Kim and Neilsen 2004) that looks for strong patterns of elevated LD in regions that are associated with selective sweeps. We estimated ω-statistic values in similar 50-kb windows across the scaffolds and across the same simulated datasets used in the SweeD analysis to assess outlier selective sweep regions in the specialist genomes. There was strong overlap in the candidate adaptive alleles between OmegaPlus and SweeD for 93% of candidate adaptive alleles in the scale-eater and 99% of candidate adaptive alleles in the molluscivore (Table S2). OmegaPlus detected many more outlier regions than SweeD (Table S2). LD-based estimates are ideally suited for use with haplotype data rather than genotype data and might be more susceptible to high false positive rates in cases where the demographic model is overly simplistic (Pavlidis and Alachiotis 2017). To be conservative, we only analyzed candidate adaptive alleles detected by both methods.

We chose to focus on detecting hard selective sweeps for our candidate adaptive variants because a) their stronger pattern is easier to discern from neutral processes with our moderate population-level sampling and coverage, and b) theoretical and experimental work suggest that soft sweeps of multiple copies of an allele are unlikely for groups with smaller population sizes (Jensen 2014). However, we acknowledge that we may have missed some candidate adaptive variation in the specialists in the form of partial or soft selective sweeps.

4.6.1.9.2 Selection of candidate adaptive allele for both specialists

To identify candidate adaptive alleles underlying trophic specialists species divergence on SSI, we looked for strongly divergent SNPs between the two specialist species in regions of the genome that showed evidence of hard selective sweeps. We considered divergent SNPs to be those that were nearly fixed ($F_{st} \geq 0.95$) between the specialists to accommodate the small amounts of admixture that can occur between these nascent species (Figure S2; Table S3-S4; Data S2-S3). For the rest of this study, we considered the 3,258 and 1,477 alleles that were nearly fixed between the species on San Salvador ($F_{st} \geq 0.95$) and located in a candidate selective sweep (empirical CLR > demographic simulations CLR; empirical ω > demographic simulations; Table S2) as the adaptive alleles for the scale-eater and molluscivore, respectively (Table S3-S4; Data S2-S3).

4.6.1.9.3 Categorization of the spatial distribution of adaptive alleles.

We then surveyed all pupfish individuals sampled from outside these populations for this set of adaptive alleles. Alleles were separated into three categories of genetic variation: de novo (the specialist allele was found only on SSI), introgressed (the specialist allele fell in a candidate introgression region determined in the Introgression section) or standing genetic variation (the specialist allele was also found in at least one generalist population sampled outside of SSI). Introgressed variation was further parsed by geographic region of the outgroup source generalist population: North Carolina (NC), New Providence Island (NP), or Dominican Republic (DR).

Given that the majority of the adaptive alleles for both specialists (98 and 100% the scale-eater and molluscivores, respectively) exist as standing genetic variation across the Caribbean (Figure 2A), we looked for how many of these adaptive alleles in the specialists also showed evidence of hard selective sweeps in focal generalist populations outside of SSI. Only 2% of the scale-eater adaptive alleles and 6% of the molluscivore adaptive alleles occurred in regions that similarly exhibited signatures of a hard selective sweep in generalist populations outside of SSI (Figure S3).

4.6.1.10 Introgression in outgroup generalist populations

We were interested in whether San Salvador Island specialist genomes exhibited more introgression in regions undergoing hard selective sweeps than other generalist populations. In the absence of a clear null expectation for the number of introgressed regions, we calculated the number of these adaptive introgression regions for the specialists that were also outlier f_d regions in other combinations of populations across the Caribbean (Table S11), to determine if those adaptive introgression regions observed in the specialists had also introgressed in other populations. Since several outgroup generalist populations had multiple values for the number of adaptive introgression regions (due to several different combinations of sister lineages (P1) available for testing against: Table S11), only the mean number of adaptive introgression regions per generalist population was shown for ease of visualization (Table S11; Figure3E-F). North Carolina and Venezuela were excluded as focal populations for these introgression tests because we lacked suitable outgroup taxa for them. Since these counts were not normally distributed, we used the non-parametric Mann-Whitney U test to determine if the mean number of adaptive introgression regions in each specialist was greater than the mean in the rest of the Caribbean (Table S12A v. S12B and Table S12C v. S12D) and calculated 95% confidence intervals around these means using the boot.ci function in the R package boot (v1.3; Figure3C). Since neither of the SSI specialists appear to have 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.

4.6.1.11. Functional characterization of adaptive alleles through GO analysis

We performed gene ontology (GO) enrichment analyses for genes near candidate adaptive alleles using ShinyGo (v.0.51;(Ge and Jung 2018)). 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; Figure2C; Figure4; Table S5), we also 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 (Balsa-Canto et al. 2016). All genes had consistent annotations across databases, except *galr2*. *Galr2* was annotated for feeding behavior in the Biological Processes database (Ensemble 92), but recent studies indicate that it does not play a role in feeding behavior (Wang

et al. 1998; Anderson et al. 2013). Thus, we removed its annotation as a candidate gene for feeding behavior, but kept it as a candidate for trophic morphology (Table S5-S6).

4.6.1.12. Functional characterization of adaptive alleles through genome-wide association mapping

4.6.1.12.1. 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 (Figure S8A; Data S6). Each specimen was also measured for standard length using digital calipers 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 (Figure 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 males remain paler throughout their body and fins (Figure 4). This pair of sympatric specialists exceeds the lightness contrast in male reproductive breeding coloration observed across all other *Cyprinodon* pupfishes. Females of each species show the same general pattern of lightness/darkness. We detected no difference in the total number of melanocytes on the caudal, anal, or pectoral fins among the SSI species (data not shown). Instead, we found that scale-eater individuals were significantly darker overall on their caudal fins (two-tailed *t*-test, t=5.25, df=45.5, *P* -value= 3.8×10^{-6}; Figure4B; Data S6), perhaps due to larger melanocyte areas relative to molluscivores. We found similar patterns for anal and pectoral fins and used only caudal fin lightness values for genome-wide association mapping (data not shown). A Meiji EMZ-8TR stereomicroscope with standardized external illumination and an OMAX 18 Mp digital microscope camera was used to take lateral photographs of the caudal fin of each individual against the same white reference background in each image (Figure 4B;Data S6). Adobe Photoshop (Creative Cloud) was used to select a rectangular area from inside the caudal fin, not including the caudal peduncle region or terminal marginal band, and measure the mean overall lightness of this region relative to a control region selected from the illuminated white background (following (Martin and Johnsen 2007)). Standardized caudal fin pigmentation was then calculated as the proportion of the caudal fin lightness value relative to the control background lightness value for downstream analyses.

4.6.1.12.2. Genome-wide association mapping analyses

We employed a Bayesian sparse linear mixed model (BSLMM) implemented in the GEMMA software package (v. 0.94.1; (Zhou et al. 2013)) to identify genomic regions associated with variation in lower oral jaw length, caudal fin pigmentation, and nasal protrusion distance across the three species on SSI. We only included individuals from SSI given extensive Caribbean-wide population structure (Figure 1C). We specifically performed genome-wide association mapping with GEMMA because of its demonstrated effectiveness in accounting for relatedness among samples and in controlling for population stratification by internally calculating a genetic relatedness matrix and incorporating it as a covariate in the BSLMM. The BSLMM uses Markov Chain Monte Carlo (MCMC) sampling to estimate the proportion of phenotypic variation explained by all SNPs included in the analysis (proportion of phenotypic variance explained (PVE); FigureS10A-C), only SNPs of large effect (proportion of genetic variance explained by sparse effects (PGE); FigureS10D-F), and the number of large-effect SNPs needed to explain PGE (nSNPs; FigureS10G-I). GEMMA also estimates a posterior inclusion probability (PIP) for each SNP. We used PIP, the proportion of steps in MCMC chain 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 MCMC runs of the BSLMM with 100 million steps and 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 (Figure 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 following (McGirr and Martin 2016b). All 10 independent runs were consistent in reporting the strongest associations for the same 20-kb windows. We identified regions strongly associated with our traits of interests by a PIP score in the 99th percentile 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, FigureS10H). 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 (Martin et al. 2017).

4.6.1.13. Functional characterization of adaptive alleles through differential gene expression and QTL analysis from previous studies

4.6.1.13.1. Differential gene expression

Additionally, we looked for overlap between genes associated with our set of adaptive alleles 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 previous studies (Mcgirr and Martin 2018; McGirr and Martin 2020). Tables with significantly differentially expressed genes at 2 and 8 dpf from these studies are provided in Data S4 and S5.

4.6.1.13.2. QTL analysis for jaw size

We also investigated our set of adaptive alleles for effects on craniofacial morphology by overlapping scaffolds with a previously published linkage map and QTL analysis of an F2 intercross between specialist species (Martin et al. 2017). We overlapped markers from this study that spanned the 95% Bayesian credible interval for a significant QTL for lower jaw length (LG15; taken from Figure S2 in (Martin et al. 2017)). 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+ (Camacho et al. 2009) and we selected the result with the highest percent identity and lowest e-value (Table S8). We then looked at all the genic regions within the interval between these two markers to investigate overlap between the QTL region and the alleles 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 an 18 Mb region on scaffold c_bro_v1_0_ scaf8 (Table S8). However, this large region contained only a few adaptive alleles associated with the genes *map2k6* (3 alleles), *galr2* (2 alleles), and *grid2ip* (4 alleles).

4.6.1.1.14. Timing of divergence for adaptive alleles

If adaptive diversification in this radiation of pupfishes occurred in temporal stages as proposed in other systems (e.g. 'behavior-first evolution'; (Mayr 1963; Huey et al. 2003a; Losos et al. 2004)), we predicted that there would be an ordering of divergence times among sweeps containing genes annotated for traits related to different trait axes in this system (Table S6-S7). In order to determine if there have been stages of adaptation in this adaptive radiation of pupfishes, we first estimated divergence times between molluscivores and scale-eaters for each adaptive allele. 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 the 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 (Hudson 2007; Coop et al. 2008). We estimated sequence divergence in regions surrounding alleles using D_{xy}, an absolute measure of genetic divergence. We calculated D_{xy} in 50-kb windows between the genomes for the SSI specialists (scale-eater vs. snail-eater) using the python script popGenWindows.py available from https://github.com/simonhmartin/genomics_general (Martin et al. 2015b).

To get a heuristic estimate of divergence time between specialists at these adaptive alleles, we used this D_{xy} count of the number of alleles 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 (Masatoshi 1972). Using the per generation mutation rate estimated above (1.56×10^{-8}), we calculated the time since divergence for adaptive alleles and compared that time to the estimated 6-19 kya age of the radiation (based on estimates of the last period of drying of hypersaline lake basins on SSI (Hagey and Mylroie 1995; Turner et al. 2008) and the last glacial maximum (Clark et al. 2009)).

To look for stages of diversification along different trait axes using these divergence time estimates, we matched adaptive alleles to potential phenotypes in two ways: 1) from our GO enrichment analyses for genes relevant to the major axes of adaptive radiation in this system (e.g. craniofacial morphology and behavior), and 2) regions strongly associated with either lower jaw size, nasal protrusion distance, or caudal fin pigmentation in the GWAS for SSI pupfish species. We found 31 regions containing adaptive alleles in or near genes with relevant GO terms and 24 regions containing adaptive alleles significantly associated with traits in the GWAS (Figure 4).

Six significantly enriched GO terms from the GO enrichment analysis of all the adaptive alleles reflect major axes of trait diversification in the radiation: divergent behavior or feeding behavior (GO terms: behavior and feeding behavior) and divergent craniofacial morphology (GO terms: eye, muscle tissue, skeletal and mouth development). There is strong morphological divergence in oral jaw size, eye orbit diameter, and adductor muscle mass among the SSI species. We therefore focused our comparison of divergence time estimates on alleles associated with genes annotated for these traits and 6 GO terms in downstream analyses of stages of adaptation across different trait axes. Melanin pigmentation is another divergent trait in this system, but it was not a significantly enriched GO term in our analyses. We include descriptions of alleles potentially relevant to pigmentation in the main text.

We then plotted the divergence time estimates for all adaptive alleles based on their spatial origins (de novo on SSI, introgression, or standing genetic variation). We also plotted all neutral regions that contained a fixed or nearly fixed allele, but no signature of a hard selective sweep (Figure 4, S11 and S12). We pruned alleles by randomly selecting one from the group of alleles that fell within the same 50-kb window so that each plotted point was independent. Some windows had multiple alleles with different spatial distributions (e.g. de novo vs. standing genetic variation), so we made an alternative plot for alternative spatial distributions of alleles that occurred within a single 50-kb window (the smaller vs larger spatial distribution; Figure4 and FigureS12). This applied to several adaptive alleles that were characterized as either introgressed or standing genetic variation in two regions containing genes with relevant adaptive annotations (*galr2* and *kcnk2*). In Figure 4 we plotted these alleles in the introgression and de novo columns. In Figure S12 we plotted these alleles in the standing genetic variation column.

We also explored the impact that the choice of pairwise species used in D_{xy} calculations had on the estimates of divergence times and relative ordering of those times among adaptive alleles. We measured D_{xy} between each of the specialists and *C. artifrons*, the outgroup used in the f_d statistic to estimate divergence times. The ordering of divergence times among genes and across phenotypic axes in this new calculation was similar to the ordering found for divergence times estimated with D_{xy} between the specialists (Figure 4, Figure S12). This indicates that the older divergence times among some regions is probably not due to 3 in mutation rate between the specialists on SSI that isn't observed in other outgroup generalist populations.

4.6.1.15. Timing of selective sweeps on adaptive alleles

4.6.1.15.1. Estimating posterior distribution of sweep ages for adaptive alleles

We also looked for evidence that adaptation occurred in stages by estimating the ages of selective sweeps of adaptive alleles. We used a coalescent-based approach implemented in the R

package starTMRCA (v0.6.1; (Smith et al. 2018)) to get sweep age estimates for adaptive alleles. Estimating sweep ages for all 1,477 molluscivore adaptive alleles and 3,258 scale-eater adaptive alleles was computationally infeasible using this Bayesian approach, so we chose to estimate sweep ages for two subsets of these adaptive alleles (Table S17-18). For the first subset, we estimated sweep ages for all alleles in or near (within 20-kb of) genes annotated for significantly enriched GO terms from our GO enrichment analysis that were relevant to behavior and trophic morphology. This subset included all 12 genes assigned to the behavior GO term, all 10 genes assigned to eye development GO term, all 12 genes assigned to the muscle tissue development GO terms, and all 4 genes assigned to the mouth development GO term (Table S5-S7). Several genes were annotated for multiple GO terms, so we ended up estimating sweep ages for a set of adaptive alleles associated with 25 different genes with relevant GO terms for the scale-eater and 6 for the molluscivore. For the second subset of adaptive alleles, we estimated sweep ages for all de novo and introgressed alleles regardless of annotation. This left a large pool of adaptive alleles distributed as standing genetic variation (illustrated in Figure4) that we could not estimate selective sweeps for. Therefore in a third subset of alleles, we selected all alleles with equivalently old and young divergence age estimates to our adaptive alleles from the first subset (those annotated for focal GO terms). In these alleles, we investigated what the genes they were in or near are annotated for to determine if they had any relevance to behavior or trophic morphology we may have missed. If the regions surrounding the adaptive alleles were unannotated, we aligned the 100-kb region surrounding the allele to the references genomes of *C. variegatus*, zebrafish and medaka available on Ensembl 96 (Yates et al. 2020) using the same protocol in Section 1.14.2 to look for potentially relevant gene annotations we may have missed in annotating the *C. bronotheriodes* reference genome in this study.

For each candidate adaptive allele from the two subsets mentioned above, a 1-Mb window surrounding the variant was extracted into a separate vcf for both specialist populations and the SSI generalists. We removed 2 generalist and 1 molluscivore individuals from this analysis that had with more than 10% missing data because starTMRCA requires complete genotype data. For all remaining individuals, we then used the LD KKNI command in Tassel5 (Bradbury et al. 2007) to infer missing sites based on LD if possible. After this imputation step, we then removed the small number of sites with any missing data across individuals within each population.

We then input this dataset with no missing allele information into starTMRCA. We used the mutation rate estimate of 1.56×10^{-8} substitutions per base pair estimated in this study and a recombination rate of 3.11×10^{-8} (from genome-wide recombination rate estimate for stickleback; (Roesti et al. 2013)) in order to estimate the age of selective sweeps for adaptive alleles. For the cases in which we had more than one adaptive allele in a selective sweep region, the variant with the highest F_{st} was chosen as the location of the beneficial allele for the sweep age estimate. We thus estimated sweep ages for 86 sets of adaptive alleles across scale-eater and molluscivores. We calculated posterior distributions of sweep age estimates using three independent runs of 10,000 steps. All runs were checked for convergence of age estimates between and within runs.

We then ran permutation tests to determine how likely the ordering of selective sweep ages by trait axes (i.e. feeding behavior, trophic morphology) was to occur by chance alone. To do this we randomly reassigned the ordering of the ages we estimated across the 22 sets of adaptive

alleles 10,000 times without replacement. Then we estimated the probability of seeing the observed number of times the oldest sweep ages were all associated with a particular trait axes by counting the number of random permutations which matched or exceeded the observe pattern. For example, 5 out of these 22 adaptive allele sets were associated with feeding behavior. We then counted how many random permutations had an ordering in which the first three (the observed pattern), four, or five oldest sweeps were associated with feeding behavior to calculate an empirical *P*-value.

4.6.1.15.2. The robustness of sweep age estimates across genealogical assumptions

Additionally, we explored how robust these sweep age estimates were to the assumption made by starTMRCA that the sweep left a star-shaped genealogy pattern. This pattern is expected for sweeps that arose from a single copy of an allele in which many alleles in one generation coalesce back to a single ancestor in the previous generation. We wanted to explore how robust our age estimates were particularly because we are comparing alleles with very different spatial distributions (de novo, introgressed, and standing). If the underlying allelic genealogy does not follow the star-shaped pattern of coalescence expected by selective sweeps from a single allele copy and instead swept from multiple copies in a soft sweep, using different subsets of individuals from a population or species could result in vastly different sweep age estimates (Smith et al. 2018) and indicate that they do not fit the star-shaped pattern assumed by starTMRCA .

Therefore, we re-estimated our sweep ages solely using the Osprey lake populations of scale-eaters and molluscivores and compared these age estimates to those from the entire population of scale-eaters on SSI. The age estimates for Osprey Lake were very similar to the entire SSI population and the relative ordering of age estimates across adaptive alleles was nearly identical (Figure S14). This indicates that the sweep ages estimates, particularly their relative ordering, were robust to differences in spatial distribution and potential differences in genealogical patterns among alleles.

4.6.1.15.3. The robustness of sweep age estimates across different methods

We also explored the robustness of selective sweep ages estimated by starTMRCA by additionally estimating sweep ages using an independent R package called McSwan (v1.1.1;; https://github.com/sunyatin/McSwan; (Tournebize et al. 2019)). 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 (Tournebize et al. 2019). 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 (Tournebize et al. 2019).

With a mutation rate estimate, neutral demographic model (effective population size changes and divergence events), and variant file, McSwan generates simulated and observed SFSs and a prior of sweep ages, whose upper bound is determined 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 the SFSs required by McSwan to estimate sweep ages, we used our estimated mutation rate (1.56×10^{-8}), the same demographic models of changes in effective population sizes used in our SweeD runs for the generalists and scale-eater populations (Table S10), and a divergence time estimate between SSI generalist and scale-eater of 10,000 years. 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 22 sets of adaptive alleles highlighted in Figure 4. To precisely determine the boundaries of hard selective sweeps, McSwan iterates its genomic scans over adjacent windows of various lengths and offsets and compares the empirical SFS to the simulated SFS under selection to assign regions as selective sweeps. We set up the iterative scans across these 500-kb regions in sliding windows that ranged from 1000 bp to 200-kb in length and a minimum of 50 alleles required per window. Each sliding scan of the 500-kb region used 100 overlapping steps (default setting). We then looked for overlap between the regions detected as hard selective sweeps by McSwan with adaptive allelles previously detected with SweeD and F_{st} (Table S2-S3).

For these 11 regions, we filtered the distribution of sweep ages for estimates that had a stability value (a parameter that represents the strength of support for a selective sweep model over a neutral model) in the 95th percentile. To get a likely range of selective sweep age estimates for each region, we calculated the 95% high posterior density (HPD) region with the R package HDIntervals (v0.2; https://cran.r-project.org/web/packages/HDInterval/index.html) from their respective posterior distributions. We repeated this process for the 6 sets of adaptive alleles found in the molluscivore, only three of which were also detected as being under a selective sweep in McSwan. The 95% HPD of these age estimates for the scale-eater and molluscivore populations are presented in Figure 4C, S15 and Table S19 and the full posteriors are shown in Figure S16 and S17. We then assessed the probability of observing the same ordering of sweep ages across alleles from different trait axes (i.e. feeding behavior and trophic morphology) using the same permutation approach described in Section 1.13.1.

4.6.2. Supplemental Results and Discussion

4.6.2.1. Spatiotemporal stages of adaption based on timing of divergence among adaptive alleles

4.6.2.2.1 Evidence of stages of adaptation across different axes of trait diversification from divergence time estimates

Based on relevant GO terms, we found that several adaptive alleles in or near genes annotated for feeding behavior exhibited the oldest divergence times (Figure 4A and S12) while adaptive alleles in or near genes annotated for craniofacial morphology and pigmentation showed younger divergence times (Figure 4A and S12). Similarly, we found younger divergence times among regions with genes annotated for traits related to trophic morphology based on GWAS annotations (Figure 4B).

When we compare divergence estimates from across all adaptive alleles and not just those with relevant GO annotations, there are three sets of alleles with similarly old divergence time estimates to our oldest feeding behavior candidates (*prlh* and *cfap20*; Figure 4A) in the scale-eater and six sets of alleles with similarly old divergence time estimates to our oldest eye morphology candidate in the molluscivores (*zhx2*; Figure 4B). We investigated the genomic regions surrounding these adaptive alleles for any annotations relevant to behavior or craniofacial morphology that we may have missed from the GO enrichment analysis. If the regions were unannotated in our *C. brontotheroides* genome, we blasted the regions to the *C. variegatus* and model organism medaka and zebrafish references genomes on Ensembl (96;(Yates et al. 2020)) to check for additional gene annotations.

From this additional search, we found three sets of adaptive alleles with similar divergence times to the oldest feeding behavior alleles (*prlh* and *cfap20; Figure* 4A) but the single gene (*gpr20)* these alleles were near did not appear to have any relevant annotations for behavior or craniofacial morphology and the additional two unannotated regions were also unannotated in the other reference genomes (*Cyprinodon variegatus*, Medaka, and zebrafish). Similarly in molluscivores, two sets of adaptive alleles with older divergence estimates (Figure 4B) were not near any genes annotated for feeding behavior or craniofacial traits (*shisa2* and *gga1*) and the four unannotated regions were unannotated in other reference genomes as well. We also searched all adaptive alleles comparable in age to the youngest adaptive alleles from our stages of adaptation analysis (*twist1* and *slc16a1*). The genes associated with these two sets of alleles (*tstd1* and *slc35e1*) with younger ages than the *twist1* allele similarly did not have relevant annotations for feeding behavior or craniofacial morphology.

4.6.2.1.3. The ordering of divergence times among adaptive alleles not driven by variation in mutation rate among regions of the genome

"Differences in mutation rate across the genome could confound our estimates of divergence times. For example, regions with the oldest divergence time estimates might only appear old because they are located in regions with higher mutation rates than other regions in the genome. To explore this possibility, we found that the scaffolds containing feeding behavior genes do not appear to have higher counts of de novo mutations in our controlled laboratory crosses (Figure S13A-C) nor more called variants than other scaffolds in the larger genomic dataset of wild individuals from across the Caribbean. Thus, we did not find any evidence of elevated mutation rates on the three scaffolds containing the oldest divergence times for feeding behavior genes (Figure S13D)."

4.6.2.2. Spatiotemporal stages of adaption based on timing of selection on adaptive alleles

4.6.2.2.1. Evidence of stages of adaptation across different axes of trait diversification from starTMRCA

Although we ran starTMRCA on all scale-eater and molluscivores adaptive alleles that were in or near all genes annotated for behavior or craniofacial morphology from our GO enrichment analysis, we were unable to get estimates for twelve sets of adaptive alleles due to poor convergence within 10,000 steps across the 3 independent runs in starTMRCA. These alleles were discarded from sweep age comparisons. The lack of power to estimate sweep ages with certainty for these alleles may be due to weaker selection on these adaptive alleles or greater variability in the strength of selection across populations in different lakes.

Therefore, our downstream analyses included sweep age estimates from 26 of the 31 sets of sweep age estimates from alleles associated with genes that have behavior or craniofacial morphology GO term annotations (22 of 25 for scale-eater, and 4 of 6 in molluscivores; Figure 4E-F). For the molluscivore, we are missing sweep age estimates for the adaptive alleles near the gene *atp8a2* (annotated for eye development and feeding behavior) and *tiparp* (annotated for craniofacial morphology). For the scale-eater, we are missing sweep ages for adaptive alleles in or near two genes annotated for eye development (*gnat2, zhx2*) and one annotated for muscle tissue development (*med1*). However, we did have sweep age estimates for all adaptive alleles in or near genes relevant to behavior and mouth morphology for the scale-eater. We therefore believe that the 'behavior-first' stage of adaptation we see is fairly robust in comparison to a second stage of adaptive divergence in trophic morphology.

We observed a notably 'behavior-first' stage of adaptive diversification, largely driven by the fact that the three oldest selective sweeps occurred in adaptive alleles in or near genes annotated for feeding behavior among the scale-eater alleles. We further investigated the probability that this 'behavior first' pattern could occur by chance using a permutation test. The probability that the first three or more of the oldest selective sweeps would all be associated with feeding behavior by chance alone is small (permutation test, *P-value* =0.01).

4.6.2.2.2. Evidence of stages of adaptation across different axes of trait diversification from McSwan

For the scale-eater population, only 8 of the 25 sets of alleles detected as hard selective sweeps using SweeD were also detected as hard selective sweeps using McSwan and given age estimates. In Tournebize et al. (Tournebize et al. 2019), they noted low power to detect selective sweeps when selection was relatively weak ($s \leq 0.05$) and recent (Supplemental information Section 2 of (Tournebize et al. 2019)). In one case, the alleles surrounding the adjacent genes *cenpf* and *kcnk2* were detected within the same large selective sweep in McSwan and thus have the same age estimates (Figure S15B). However, the twelve additional adaptive alleles

undetected by McSwan may be under weaker selection or more recent. Due to the very recent timing of selection in this system and the much larger set of sweep age estimates obtained from starTMRCA, we present only the starTMRCA sweep ages in the main text.

We also found a similar 'behavior-first' stage of adaptive diversification with this smaller subset of sweep age estimates from McSwan. The two oldest sweeps in the scale-eater were both associated with feeding behavior (prlh and cfap20). The probability of observing this pattern by chance alone is small (permutation test; *P-value* = 0.033).

4.6.2.2.3. Spatiotemporal stages of adaptive introgression from different source populations

We estimated selective sweep ages across all de novo and introgressed variants in the scale-eater and molluscivores regardless of gene annotations as well. We find evidence that introgressed adaptive alleles swept before any de novo adaptive alleles (Figure S5) and selection on introgressed variation occurred throughout the process of radiation. Introgressed alleles sweeping before de novo alleles further supports a role for hybridization being necessary for radiation in this system.

We also assessed whether there were significant differences in the timing of selection across de novo and introgressed alleles coming from different source populations using ANOVA. We found that alleles originating in North Carolina swept significantly earlier than introgressed alleles from New Providence Island and the Dominican Republic (*P*=0.03 and *P*=0.02 respectively; Figure 4G-H). Sweeps of adaptive alleles introgressed from North Carolina also trended older than sweeps of de novo adaptive alleles, although this was not a significant difference (*P*=0.06). Sweeps of de novo adaptive alleles occurred concurrently with sweeps of introgressed alleles from New Providence Island and the Dominican Republic (P=0.61).

4.6.3. Supplemental Figures

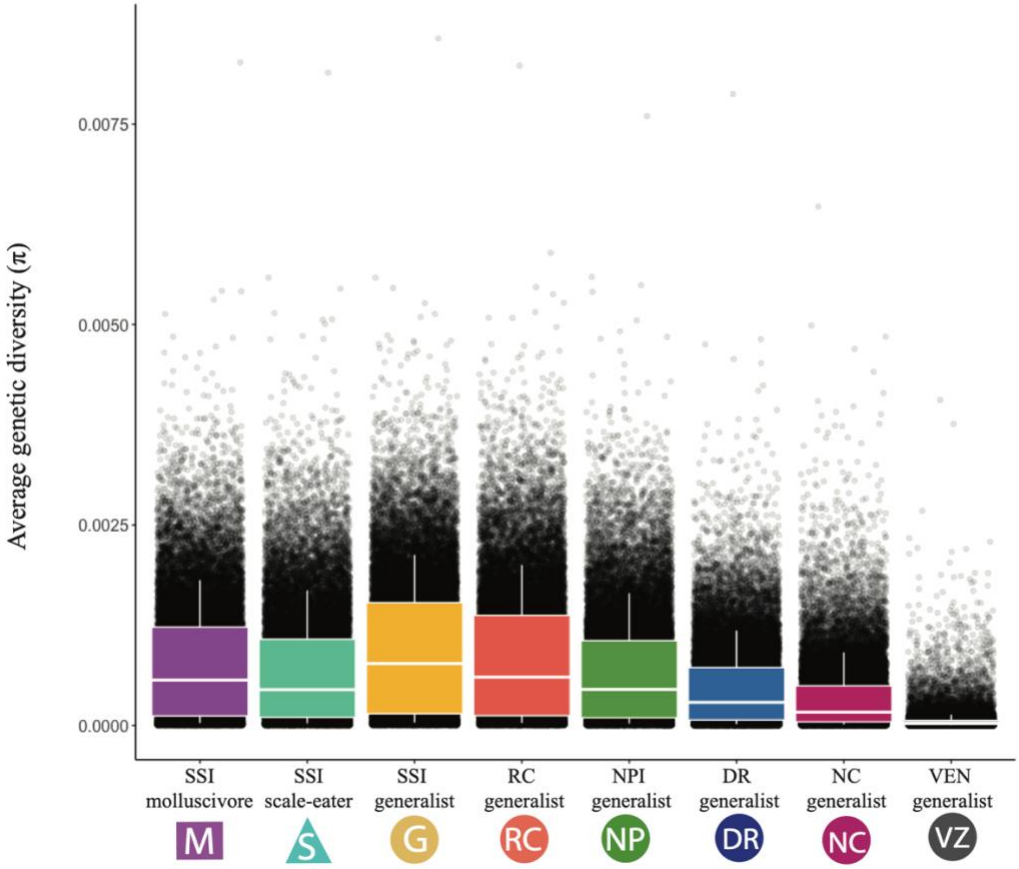


Figure S1. Similar genome-wide level genetic diversity across Caribbean pupfish populations. Within population (π) nucleotide diversity in 50-kb sliding windows across the genomes of the SSI (SSI) species and generalist species on Rum Cay (RC), New Providence Island (NPI), Dominican Republic (DR), North 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.

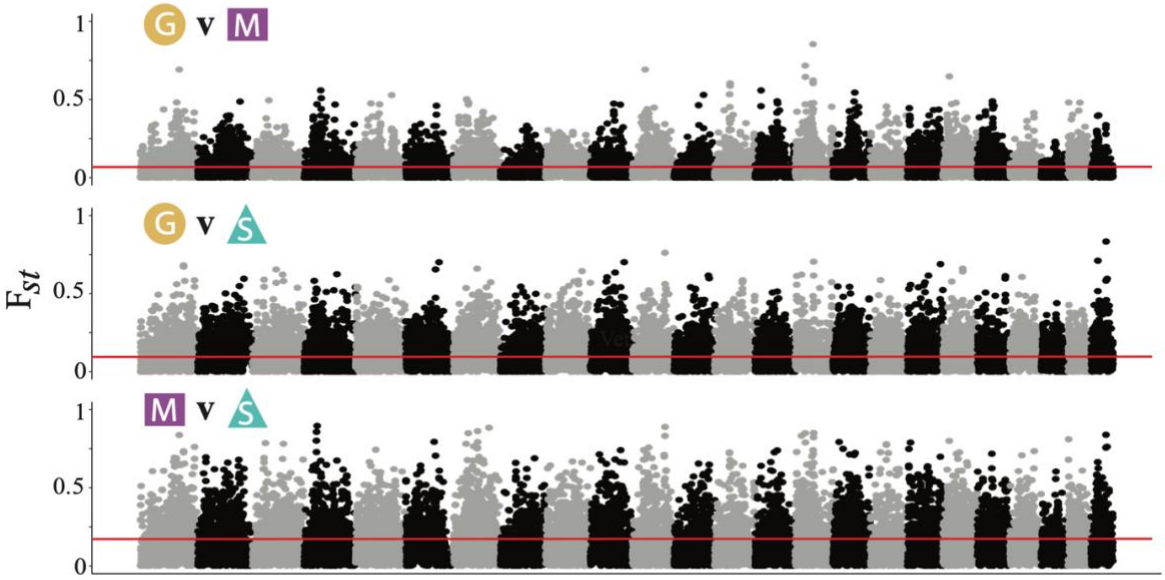


Figure S2. Genetic divergence among SSI species. Manhattan plot of Fst in 50-kb windows across the genome for the three SSI species on the largest 24 scaffolds in the molluscivore (C. brontotheroides) genome corresponding to the 24 chromosomes in Cyprinodon (Stevenson 1981). Solid red line represents the average Fst values for each comparison (generalist vs. molluscivore; 0.07; generalist vs. scale-eater: 0.11; molluscivore vs. scale-eater: 0.15).

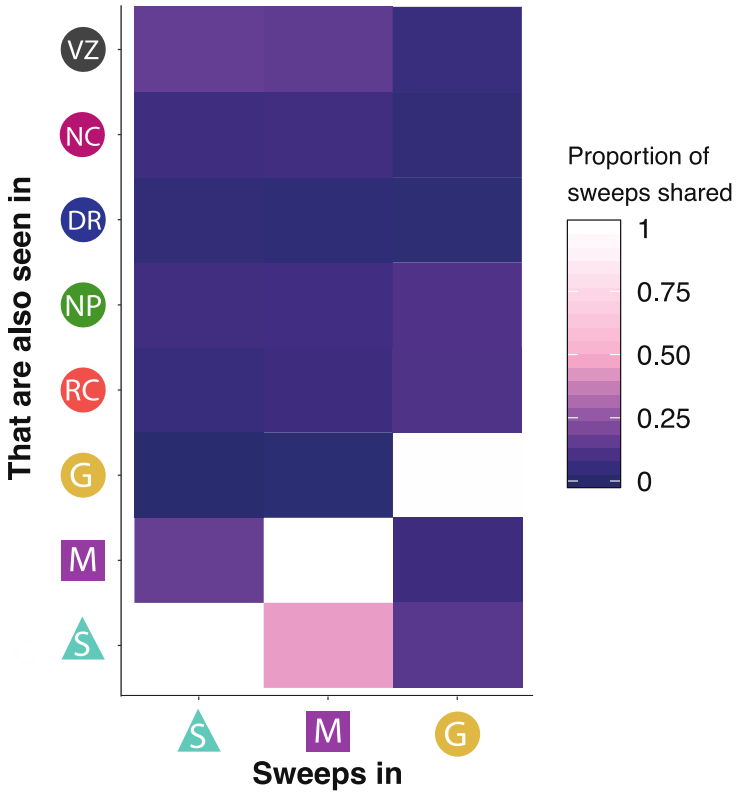


Figure S3. Selective sweeps in SSI population shared with other Caribbean populations. The proportion of hard selective sweeps in the SSI species that are also found sweeping in other Caribbean populations. Regions under hard selective sweep were identified as those with a SweeD CLR estimate greater than those calculated from demographic simulations of a similar sized population evolving neutrally (e.g. SweeD CLR > 5.28 for scale-eaters and SweeD CLR > 4.43 for molluscivores, see Table S8 for threshold values for all populations). Note that 42% of hard selective sweeps in the molluscivore population also showed signs of a sweep in the scale-eater population.

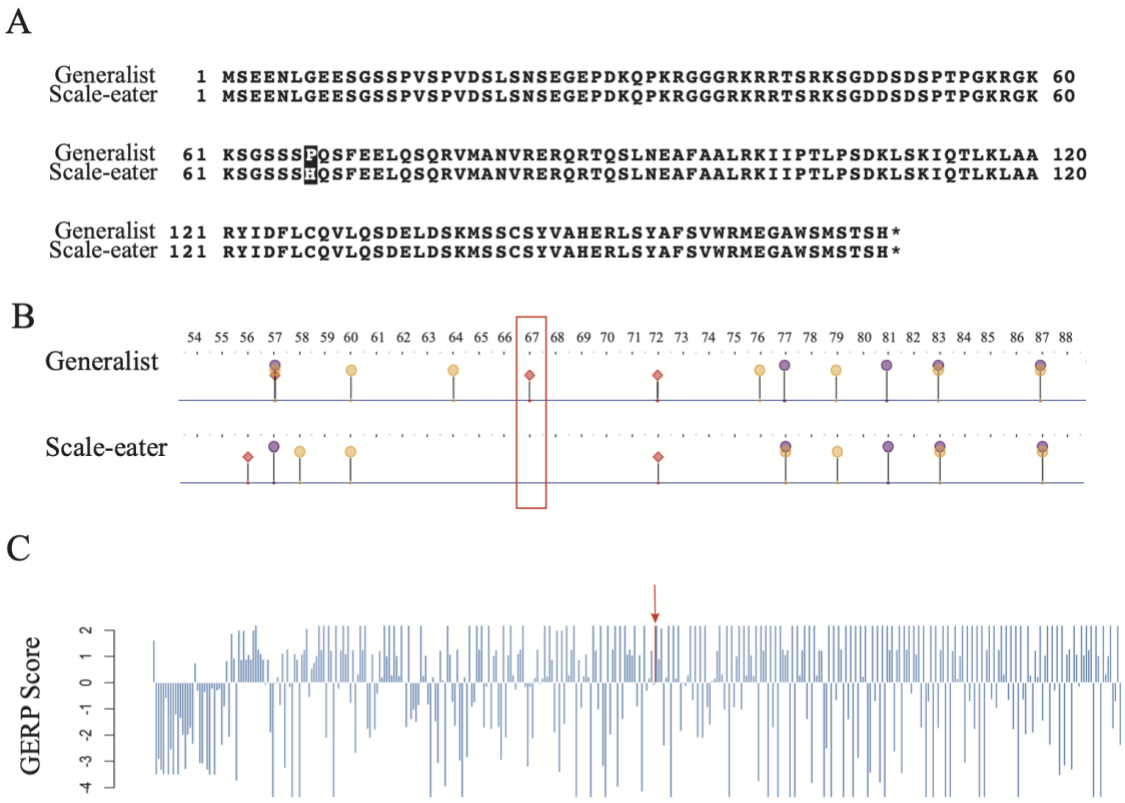


Figure S4. Sequence conservation among fishes around candidate gene *twist1*.
A) Amino acid sequence of *twist1* protein for SSI generalists and scale-eaters. The non-synonymous substitution that is nearly fixed between the two species changes the amino acid from a proline to histidine (highlighted in 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 (Ofran and Rost 2007). 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). A conservation score above 2 is considered highly conserved (Davydov et al. 2010).

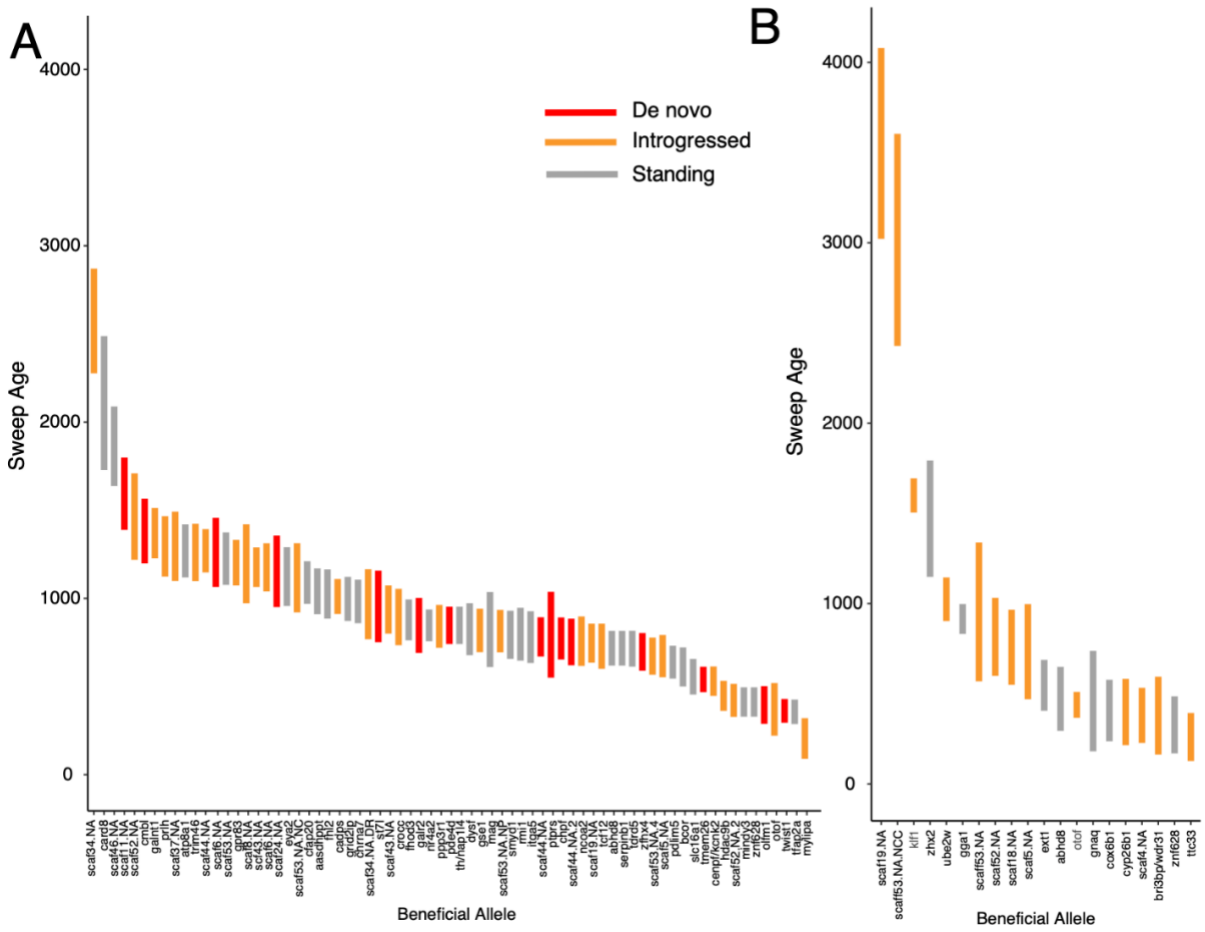


Figure S5. Selective sweep ages across spatial source of genetic variation.
95% HPD interval of the posterior distribution for selective sweep ages estimates calculated from starTMRCA for all introgressed and de novo of the specialists adaptive alleles, as well as all adaptive alleles in or near (within 20-kb) of genes annotated for behavior and craniofacial GO terms in our GO enrichment analysis (Figure 4). Selective sweep ages in the scale-eaters (A) and molluscivores (B) are colored by spatial distribution of the adaptive genetic variation (standing, introgressed or de novo alleles). Adaptive alleles are labeled by the gene region they are associated with. Alleles that are in unannotated regions are labeled by the scaffold they are found on. The exact position of the variant on that scaffold is listed in Table S16).

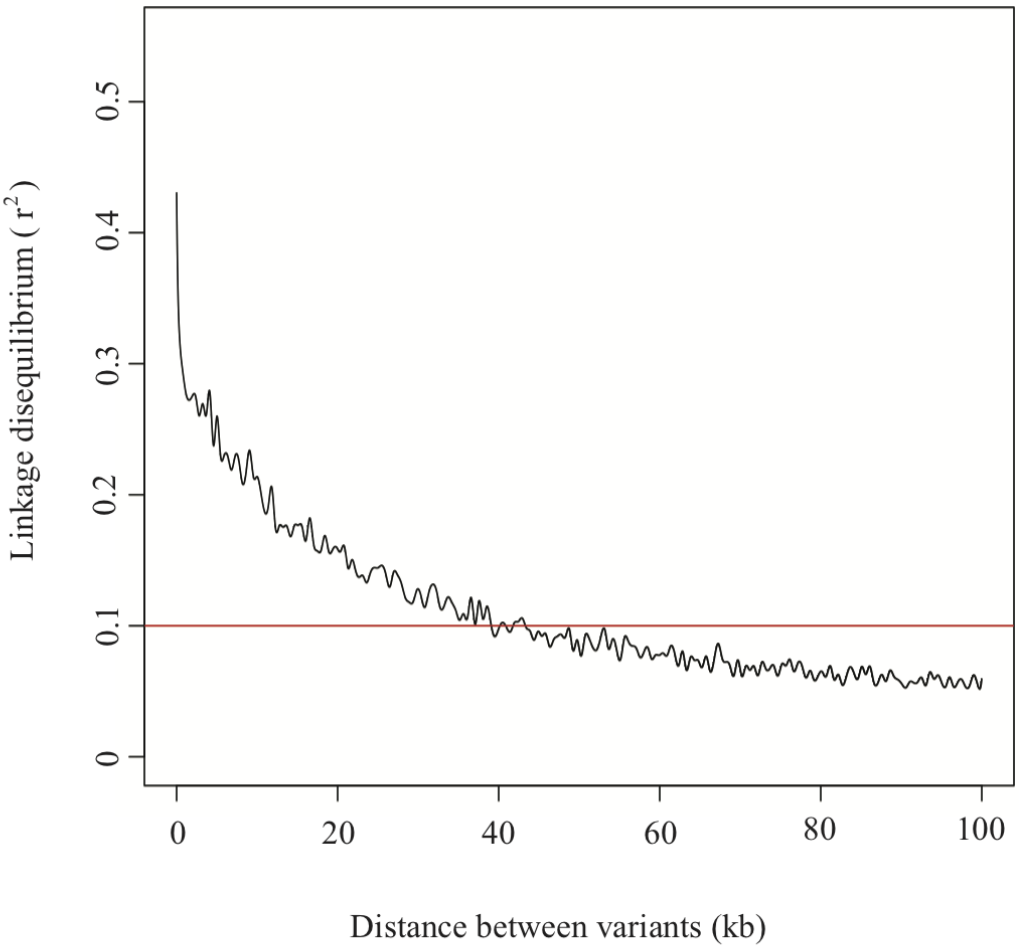


Figure S6. Linkage disequilibrium decay along the genome. LD decay over pairwise combinations of alleles within 100 kb of each other on the longest scaffold in the genome (49,059,223 bp), with r^2=0.1 marked for reference. From this pattern of decay, we chose a window size of 50-kb for sliding windows analyses used in this study.

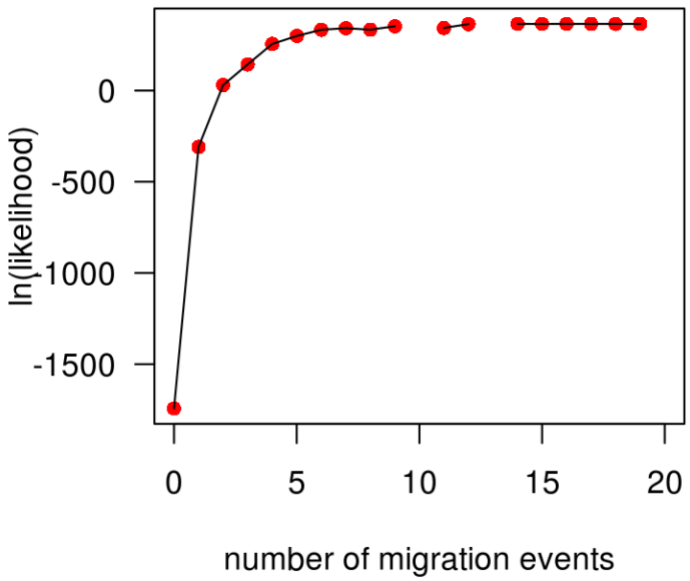


Figure S7. The likelihood of migration events on the TREEMIX population graph of admixture events across Caribbean populations. The log likelihood of different population graphs with 0-20 migration events model on them using TREEMIX (Pickrell and Pritchard 2012) and an LD-pruned set of 2.3 million SNPs across the SSI species, the 5 focal outgroup generalist populations (>8 individuals) and *C.artifrons*. 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 Figure4B.

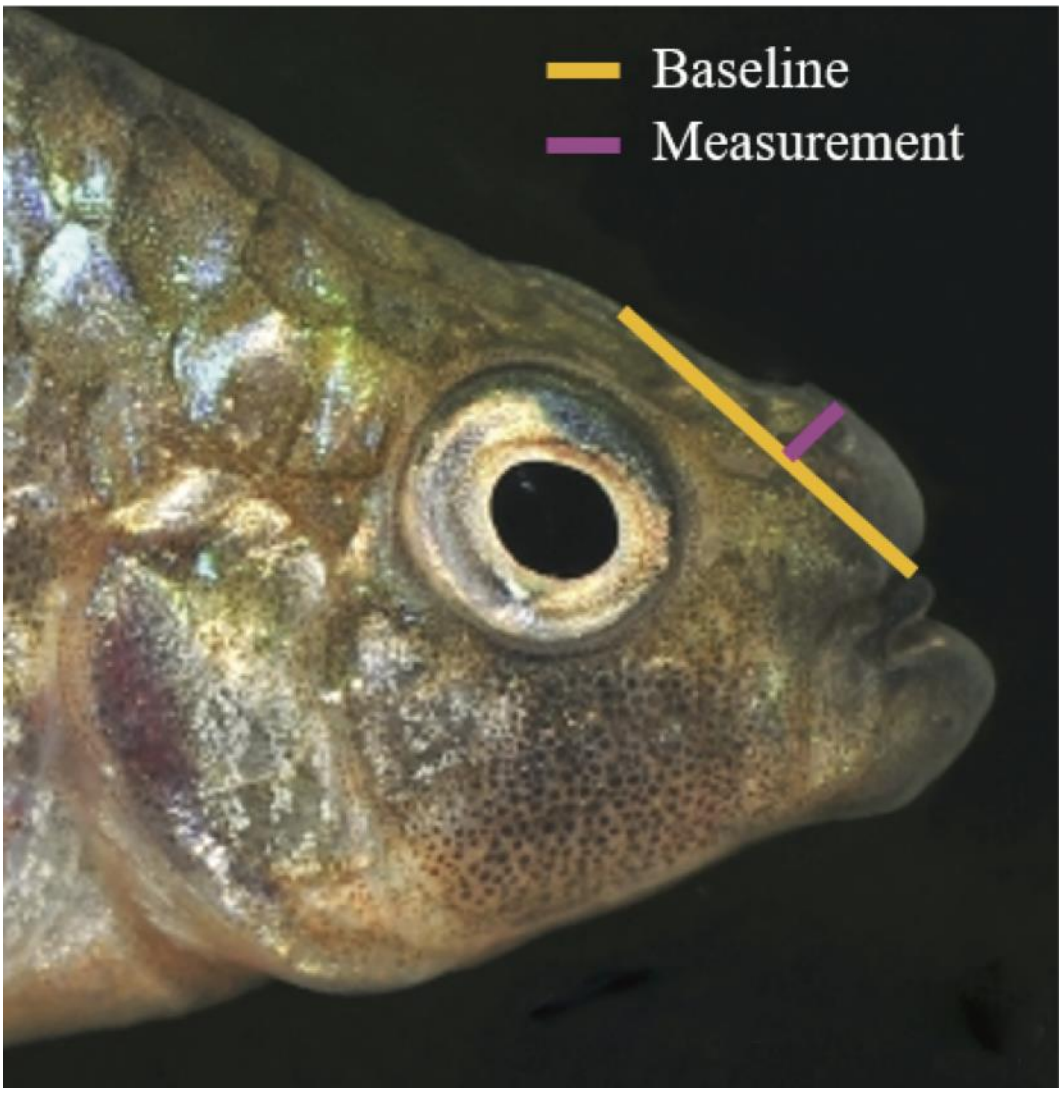


Figure S8. Example image of nasal protrusion distance measurement for GWAS. The purple line represents the nasal protrusion distance on a *C. brontotheroides* specimen. The yellow line represents a baseline tangent line from the dorsal surface of the neurocranium to the tip of the premaxilla used for reference. Photo by Tony Terceira.

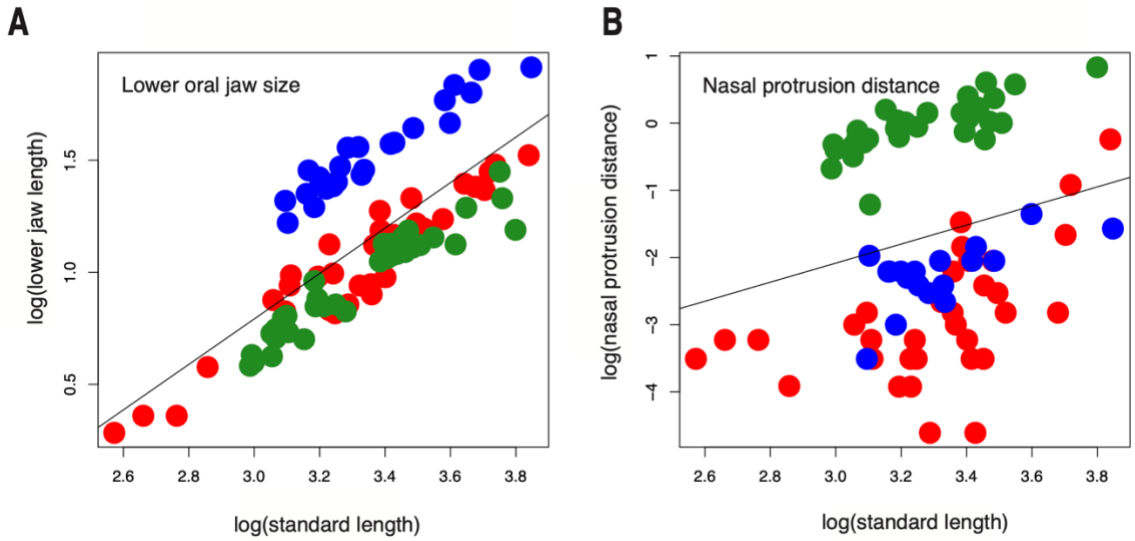


Figure S9. Standardized craniofacial trait measurements in SSI species. Log-transformed A) lower oral jaw length (mm) and B) nasal protrusion distance (mm) standardized by log-transformed standard length (mm) for SSI generalist (red), molluscivore (green), and scale-eater (blue).

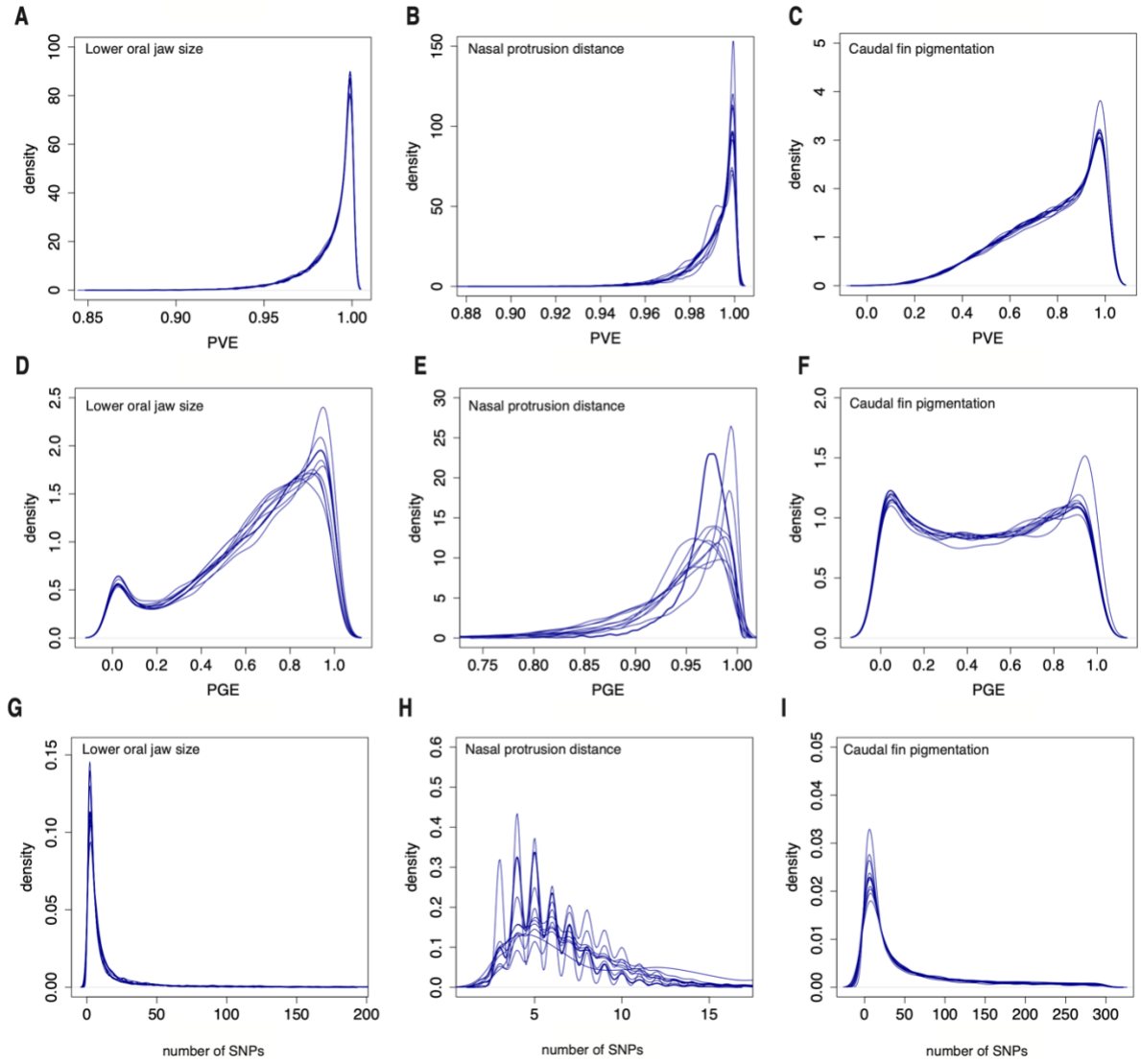


Figure S10. Posterior density distributions for hyper-parameters describing the proportion of variance in phenotypes for the three focal traits. The variance in lower jaw size, nasal protrusion distance, and caudal fin pigmentation explained by A-C) every SNP (proportion of phenotype variance explained (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.

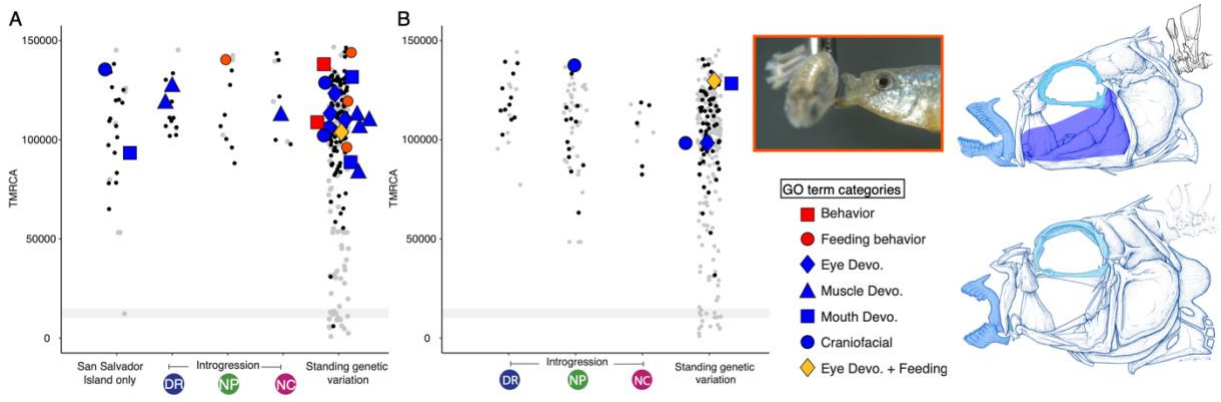


Figure S11. The spatiotemporal landscape of adaptive radiation based on divergence time from an outgroup generalist population. Time to most recent common ancestor (TMRCA) of adaptive alleles based on D_{xy} in their 50-kb window. TMRCA estimates based on genetic divergence (*Dxy*) between outgroup *C. artifrons* and A) scale-eaters or B) molluscivores. Each column separates adaptive alleles by their spatial distribution: de novo (SSI only), adaptive introgression from one of three outgroup populations (DR: Dominican Republic, NP: New Providence, NC: North Carolina), and standing genetic variation. Gray bars highlight the approximate origins of the microendemic radiation on SSI at approximately 6-19 kya (based on range of geological age estimates for filling of hypersaline lakes on SSI (Hagey and Mylroie 1995; Turner et al. 2008) since the last glacial maximum (Clark et al. 2009)). All adaptive alleles associated with genes for behavior (red) or craniofacial morphology (blue) are illustrated by a colored point. Black points show adaptive alleles for non-focal GO terms or unannotated; gray points show all nearly fixed alleles between specialists ($F_{st} \geq 0.95$) with no signal of a hard selective sweep.

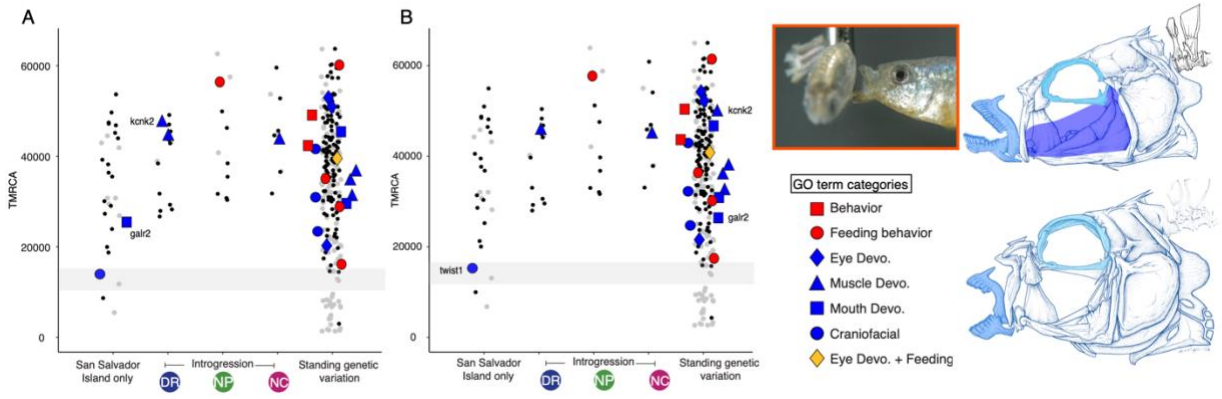


Figure S12. The alternative spatiotemporal landscape of adaptive radiation in scale-eaters. Divergence time plot in which the two labelled alleles that were plotted in the introgression and de novo column (A;) are plotted as their alternative spatial distribution in standing variation column (B). Time to most recent common ancestor (TMRCA) of adaptive alleles based on D_{xy} in their 50-kb window based on the larger spatial distrubtion of adaptive alleles. TMRCA estimates based on genetic divergence (*Dxy*) between the two specialists across alternative distributions for scale-eater adaptive alleles. Points labeled with gene names indicate the two alleles in which there are two or more adaptive alleles in the same linkage block that have different spatial distributions: A) alleles with smaller spatial scales (de novo or introgressed) and B) and alleles with larger spatial scales (standing genetic variation). Each column separates adaptive alleles by their spatial distribution: de novo (SSI only), adaptive introgression from one of three outgroup populations (DR: Dominican Republic, NP: New Providence, NC: North Carolina), and standing genetic variation. Gray bars highlight the approximate origins of the microendemic radiation on SSI at approximately 6-19 kya (based on range of geological age estimates for filling of hypersaline lakes on SSI (Hagey and Mylroie 1995; Turner et al. 2008) since the last glacial maximum (Clark et al. 2009)). All adaptive alleles associated with genes for behavior (red) or craniofacial morphology (blue) are illustrated by a colored point. Black points show adaptive alleles for non-focal GO terms or unannotated; gray points show all nearly fixed alleles between specialists ($F_{st} \geq 0.95$) with no signal of a hard selective sweep.
Time to most recent common ancestor (TMRCA) of the region surrounding candidate adaptive alleles (LD-pruned so that each point is independent) based on relative genetic divergence metric *Dxy* (Nei and Li 1979) which captures only the amount divergence that has accumulated since the two populations diverged for A) scale-eaters and B) molluscivores. Each column separates adaptive alleles by their spatial distribution: de novo (SSI only), adaptive introgression from one of three outgroup populations (DR: Dominican Republic, NP: New Providence, NC: North Carolina), and standing genetic variation. Gray bars highlight the approximate origins of the microendemic radiation on SSI: from the last glacial maximum (approximately 6-19 kya; ranging from to the youngest age estimate for filling of hypersaline lakes on SSI (Clark et al. 2009)) to the last glacial maximum before which lakes on SSI were completely dry (Turner et al. 2008)).Alleles are colored by evidence of hard selective sweeps: black for fixed or nearly fixed

($F_{st} \geq 0.95$) adaptive alleles annotated for non-focal GO terms or unannotated; gray for fixed or nearly fixed alleles between specialists with no signal of hard selective sweep; and triangles represent alleles additionally associated with pigmentation. All alleles annotated for the GO categories of behavior (red shades) and craniofacial morphology (blue shades) are included. Genes highlighted in the text are labeled by their associated variant. Yellow shade indicates genes annotated for feeding behavior and eye development. Triangle shape indicates gens also annotated for pigmentation.

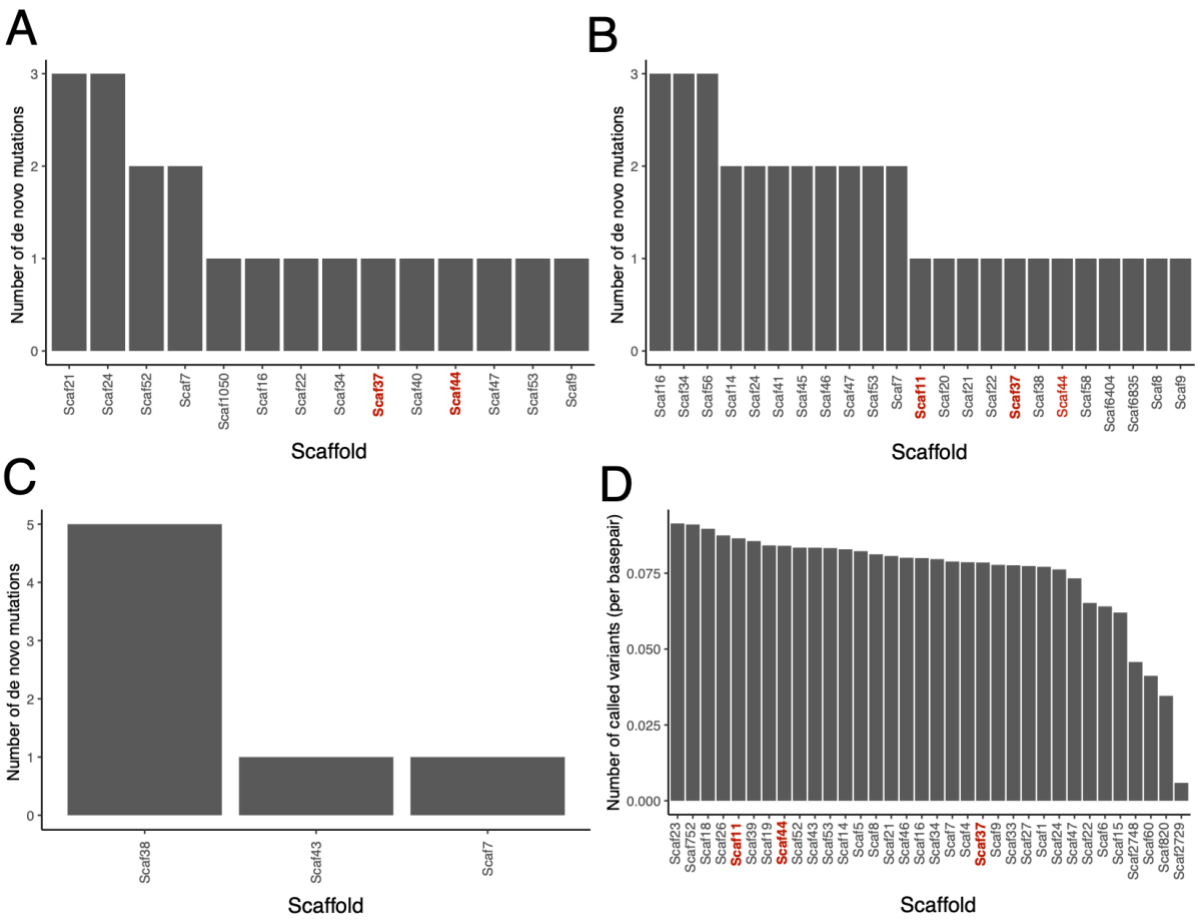


Figure S13. Raw counts of alleles found across scaffolds. The count of de novo mutations in genome of A-B) two hybrids from molluscivores x generalist cross and C) single hybrid from scale-eater x generalist sequenced to high coverage (15-69x) that were used to estimate average mutations rate for pupfish. D) The relative number of alleles per scaffold (absolute count divided by number of base pairs in the scaffold) that candidate adaptive alleles were found on. Scaffolds highlighted in red are three scaffolds that contain the feeding behavior genes with the oldest divergence time and selective sweep age estimates.

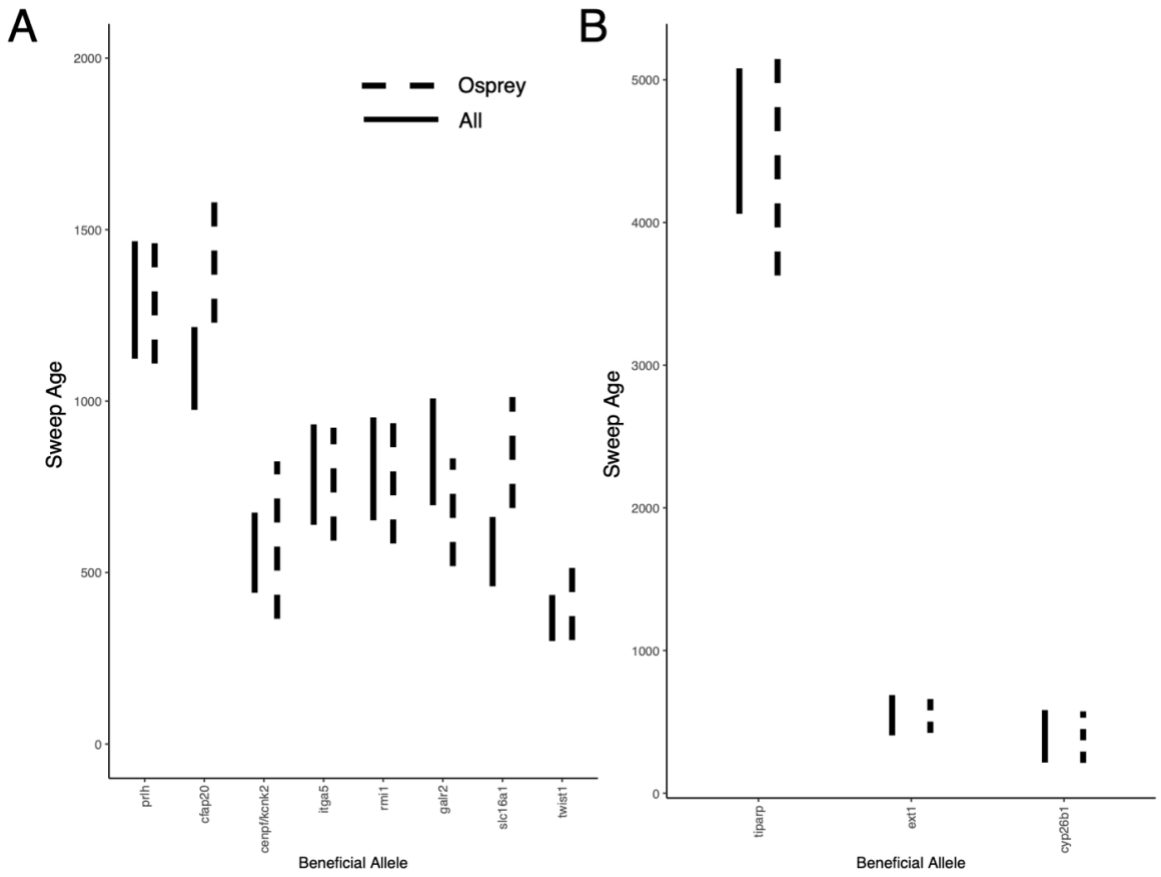


Figure S14. Allele age estimates from single population of specialists compared to estimates from all individuals. A) 95% HPD interval from the posterior distribution of allele age estimates calculated with starTMRCA on all scale-eater individuals (*N*=26) compared to just individuals from the Osprey Lake population (*N*=11). B) 95% HPD interval from the posterior distribution of allele age estimates calculated with starTMRCA on all molluscivore individuals (*N*=43) compared to just individuals from the Osprey Lake population (*N*=10).

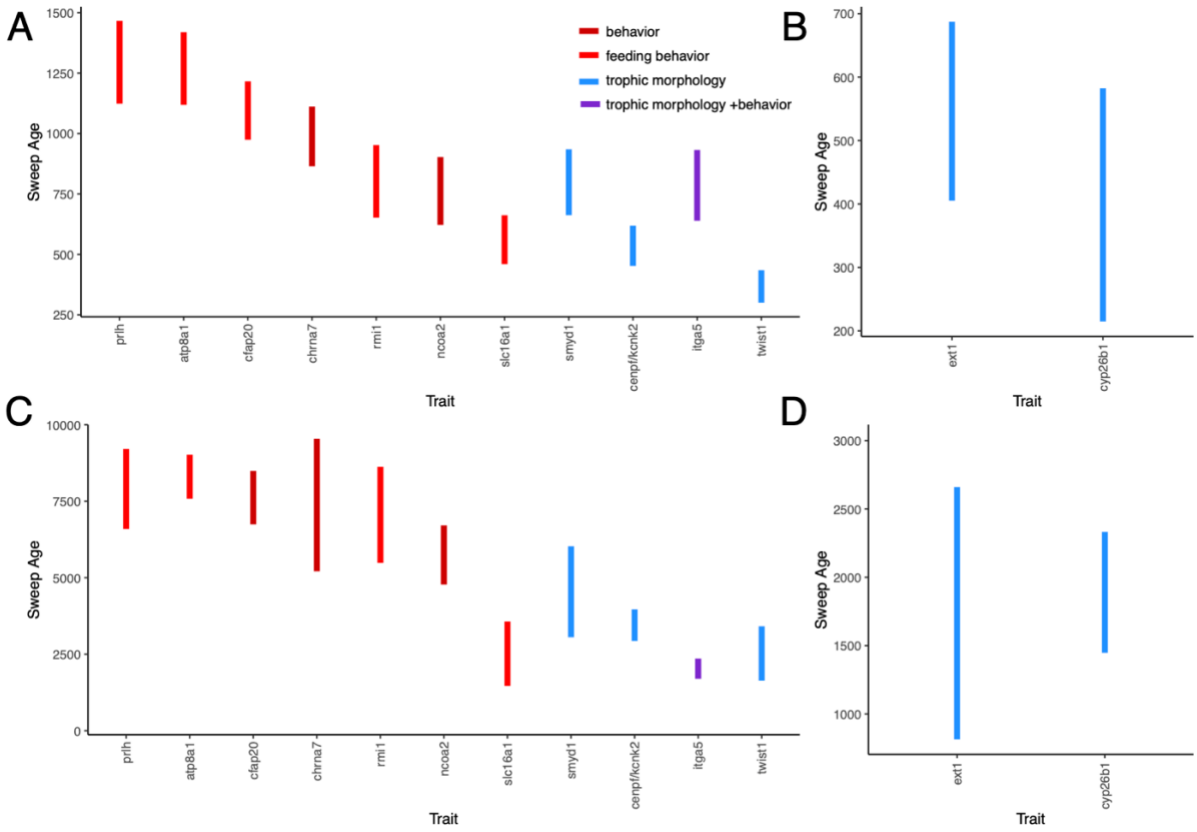


Figure S15. Overlapping hard selective sweep age estimates from starTMRCA and McSwan.

95% HPD interval for selective sweep ages for overlapping set of adaptive alleles across starTMRCA for scale-eaters (A) and molluscivores (B) compared to the 95% HPD interval estimate from McSwan for scale-eaters (C) and molluscivores (D). Selective sweep ages are colored by GO annotations relevant to two major stages of adaptation: behavior (behavior and feeding behavior), trophic morphology (craniofacial, muscle development) and both.

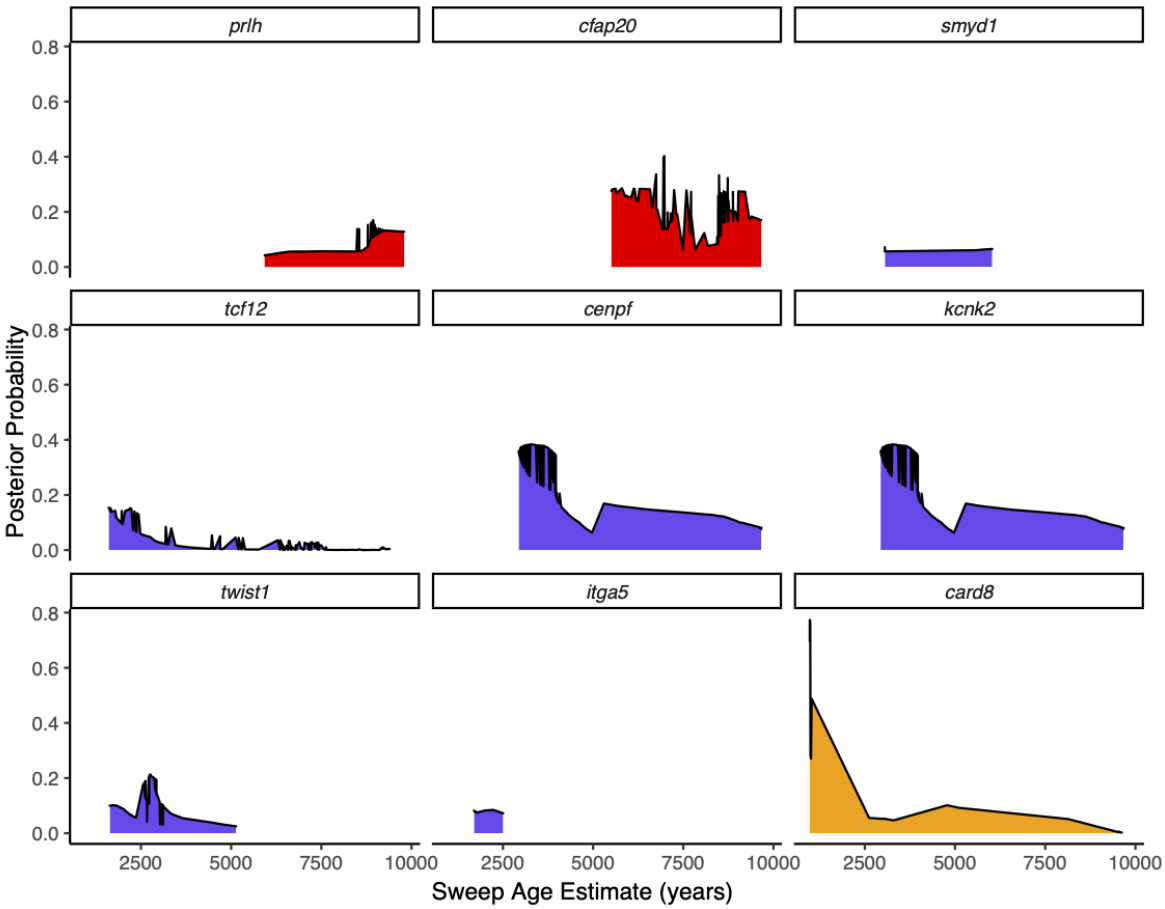


Figure S16. Full posterior distributions for scale-eater sweeps. The posterior distributions of sweep ages estimated from focal adaptive alleles (Table S13) calculated from McSwan. These nine regions contained fixed or nearly fixed variants ($F_{st} \geq 0.95$) between specialists that were estimated to be hard selective sweeps using both SweeD and McSwan. Sweep ages are colored based on GO and GWAS annotations relevant to the stages proposed in the stages of adaptation: feeding behavior (red), trophic morphology (craniofacial and muscle: blue-violet), and sexual communication (pigmentation: orange).

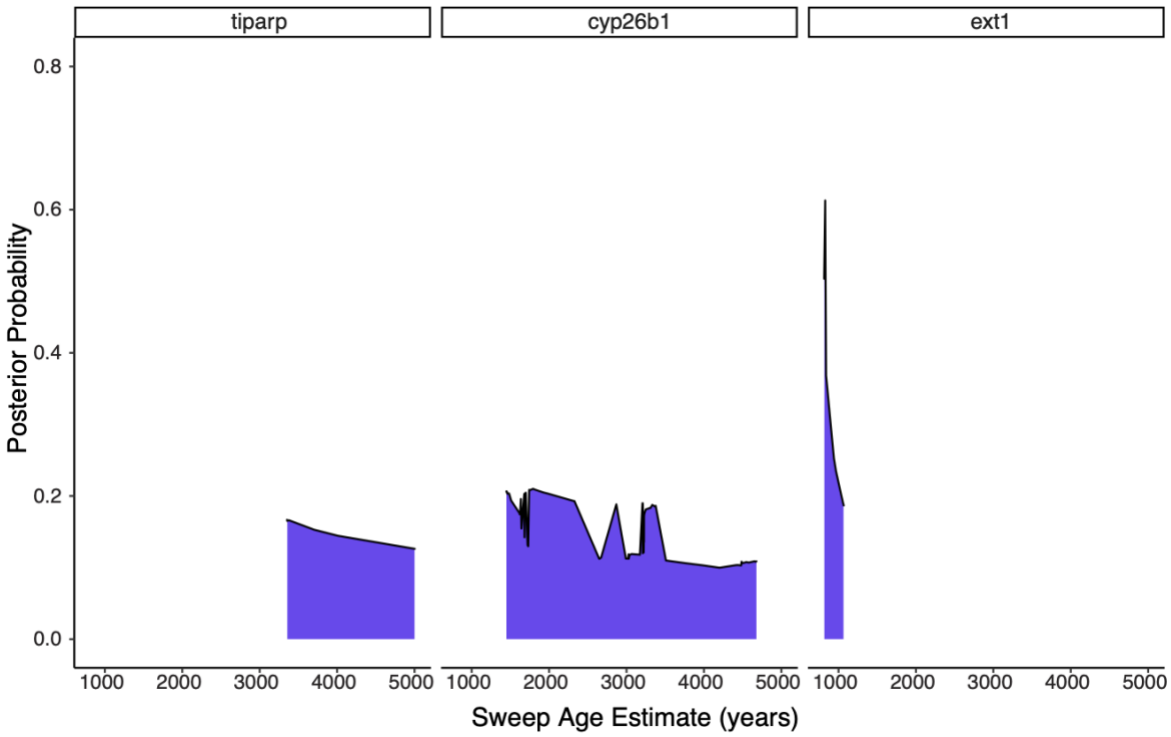


Figure S17. Full posterior distributions for molluscivore sweeps. The posterior distributions of sweep ages estimated for focal adaptive alleles (Table S13) calculated from McSwan. These three regions contained fixed or nearly fixed variants ($F_{st} \geq 0.95$) between specialists that were estimated to be hard selective sweeps using both SweeD and McSwan. Sweep ages 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).

4.6.4. Supplemental Tables

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

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

SSI molluscivore	*Cyprinodon brontotheroides*	Osprey Lake	SSI, Bahamas	12
SSI molluscivore	*Cyprinodon brontotheroides*	Oyster Pond	SSI, Bahamas	8
SSI scale-eater	*Cyprinodon desquamator*	Crescent Pond	SSI, Bahamas	10
SSI scale-eater	*Cyprinodon desquamator*	Little Lake	SSI, Bahamas	5
SSI scale-eater	*Cyprinodon desquamator*	Osprey Lake	SSI, Bahamas	10
SSI scale-eater	*Cyprinodon desquamator*	Oyster Lake	SSI, 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	1
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

Table S2. The number of selective sweeps found in specialist genomes. The number of selective sweeps detected in total across the specialist genomes using and SFS-based approach SweeD and LD-based approach OmegaPlus. Hard selective sweeps were determined based on demographic simulation-based thresholds (SweeD CLR > 5.28; OmegaPlus ω > 3.31 for scale-eaters and SweeD CLR > 4.47; OmegaPlus ω > 4.23 for molluscivores). The alleles that overlapped with nearly fixed (F_{st} ≥0.95) SNP(s) between the specialists with hard selective sweeps detecting jointly in both sweep programs were then used the total number of candidate adaptive alleles in this study.

	Molluscivore		Scale-eater	
	SweeD	OmegaPlus	SweeD	OmegaPlus
Number of selective sweeps detected	8269	12060	14729	18387
Number of windows tested	52744	49822	52696	51561
Number of alleles that overlap with sweep	1490	3917	3463	3766
Number of alleles with uniquely detected selective sweep	13	2427	230	303
Number of alleles with jointly detected selective sweep	1477		3233	

Table S3. Candidate adaptive alleles for the San Salvador Island (SSI) scale-eater. Location of the genic regions that contained signatures of a strong selective sweep in the scale-eater (above demographic simulation based thresholds SweeD CLR > 5.28;OmegaPlus ω > 3.31)and at least one divergent variant between the specialists ($F_{st} \geq 0.95$). Full list of alleles, including unannotated candidate regions provided in Data S2. Adaptive alleles highlighted in Figure 4 are listed in bold.

Gene	Scaffold	Gene Start	Gene End	Number of Alleles
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 S4. Adaptive alleles for the San Salvador Island (SSI) molluscivore.
Location of the genic regions that contained signatures of a strong selective sweep in the molluscivore (SweeD CLR ≥ 4.47; OmegaPlus ω > 4.23) and at least one divergent variant between the specialists (F_{st} ≥ 0.95). Full list of alleles, including one unannotated candidate regions provided in Data S3. Adaptive alleles highlighted in Figure S5 are listed in bold.

Gene	Scaffold	Gene Start	Gene Stop	Number of Alleles
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 S5. Full list of functional terms associated with genes in adaptive alleles 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 chromosome organization	0.02019114	6	147	*atad2,xrn1,twist1,znf451,bcor,cenpf*
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	0.04769373	29	2551	*serpinb1,zfhx4,foxo3,atad2,coq7,zhx2,xrn1,trps1,card8,fhl2,cenpf,twist1,tfap2a,irf8,trim44,bcor,kcnk2,pde4d,s*

compound metabolic process				*myd1,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 S6. Scale-eater adaptive alleles used for assessing stages of adaptation.
We estimated ages for all adaptive alleles that were in or near (within 20-kb) of a gene associated with a GO term for behavior or craniofacial traits on the Ensemble 96 annotation database and were significantly enriched in our GO enrichment analysis (Table S5). Sweep ages, stage category assignment, any additional annotations we found for genes and their references are provided. Also included is a partial list of other significantly enriched GO terms for each gene. For visual clarity in the table, the broader GO terms (terms that > 1000 genes listed in database) are not included. See Table S5 for full list. Sweep ages are listed as the 95% HPD range (X indicates missing age estimates because estimates across starTMRCA runs did not converge for that sweep).

Gene	Sweep Age	Stages Category	GO enrichment annotations	GWAS annotations	Other annotations	References for other annotations	Other GO enrichment annotations (Partial list)
galr2	696-1008	craniofacial	behavior; feeding behavior	oral jaw size	bone tissue development	(Wang et al. 1998; McDonald et al. 2007; Anderson et al. 2013)	Behavior, Feeding behavior
cfap20	974-1215	feeding behavior	behavior; feeding behavior	--	--	(Maia et al. 2014)	Behavior, Feeding behavior
atp8a1	118-1419	behavior	behavior	--	--	(Taye et al. 2017)	Regulation of phospholipid transport, Positive regulation of phospholipid transport, Regulation of phospholipid translocation, Positive regulation of phospholipid translocation
rmi1	652-952	feeding behavior	behavior; feeding behavior	--	--	(Suwa et al. 2010)	Developmental growth, Response to extracellular stimulus, Response to nutrient levels, Eating behavior, Reduction of food intake in response to dietary excess
th	746-958	feeding behavior/ craniofacial	eye development; behavior; feeding behavior	--	--	(Marles et al. 2003; Qiao et al. 2016)	Camera-type eye development, Response to xenobiotic stimulus, Sensory organ morphogenesis, Response to ketone, Eye morphogenesis, Camera-type eye morphogenesis, Embryonic camera-type eye development, Embryonic eye morphogenesis, Eating behavior, Embryonic camera-type eye

							morphogenesis, Dopamine biosynthetic process
ncoa2	622-902	behavior	behavior	--	--	--	Response to steroid hormone, Response to ketone
nr4a2	762-942	behavior	behavior	--	--		Cell morphogenesis involved in neuron differentiation, Cellular response to organic cyclic compound, Response to extracellular stimulus, Axon development, Axonogenesis, Response to steroid hormone, Intracellular receptor signaling pathway, DNA-templated transcription,Response to xenobiotic stimulus, Developmental maturation, Transcription initiation from RNA polymerase II promoter, Cell maturation, Anatomical structure maturation, Neuron maturation, Dopamine biosynthetic process
kcnk2	452-618	craniofacial	Muscle tissue development; behavior	oral jaw size	--		Developmental growth, Behavior, Sensory organ development, Heart development, Striated muscle tissue development, Regulation of developmental growth, Negative regulation of growth, Cardiac muscle tissue development, Negative regulation of developmental growth, Response to axon injury
slc16a1	459-661	behavior	Behavior	--	feeding behavior	(Hnasko et al. 2004)	Response to organic cyclic compound,Cellular response to organic cyclic compound, Response to extracellular stimulus, Response to nutrient levels, Response to nutrient
itga5	639-932	craniofacial	behavior	--	eye development; pharengeal arch development	(Crump et al. 2004; LaMonica et al. 2015)	Anatomical structure morphogenesis, Tissue development, Circulatory system development
chrna7	864-1111	behavior	behavior	--		--	Neuron projection development, Cell morphogenesis involved in differentiation, Cell part morphogenesis, Cell projection

							morphogenesis, Plasma membrane bounded cell projection morphogenesis, Neuron projection morphogenesis , Cell morphogenesis involved in neuron differentiation, Response to amyloid-beta
med1	X	craniofacial	eye development; muscle tissue development	--		--	Intracellular receptor signaling pathway, Camera-type eye development, "DNA-templated transcription,Negative regulation of neurogenesis, Developmental growth involved in morphogenesis, Negative regulation of neuron differentiation, Response to nutrient, Transcription initiation from RNA polymerase II promoter, Cardiac muscle tissue development, Appendage morphogenesis, Limb morphogenesis, Retina development in camera-type eye, Cellular hormone metabolic process, Embryonic limb morphogenesis, Embryonic appendage morphogenesis, Response to vitamin, Response to vitamin D
gnat2	X	craniofacial	eye development	--		--	Camera-type eye development, Sensory organ morphogenesis, Eye morphogenesis, Retina development in camera-type eye, Camera-type eye morphogenesis, Neural retina development, Retina morphogenesis in camera-type eye
eya2	962-1295	muscle	muscle tissue development	--	--	--	Striated muscle tissue development
tfap2a	292-431	craniofacial	eye devolopment; mouth development;		pigmentation; embryonic cranial skeleton morphogenesis	(Brewer et al. 2004; Green et al. 2015; Seberg et al. 2017)	Camera-type eye development, Sensory organ morphogenesis, Appendage morphogenesis, Limb morphogenesis, Eye morphogenesis, Retina development in camera-type eye, Embryonic limb morphogenesis, Embryonic appendage morphogenesis,

							Camera-type eye morphogenesis, Roof of mouth development, Neural retina development, Retina morphogenesis in camera-type eye, Forelimb morphogenesis, Embryonic camera-type eye development, Embryonic eye morphogenesis, Embryonic forelimb morphogenesis, Embryonic camera-type eye morphogenesis, Eyelid development in camera-type eye, Embryonic camera-type eye formation, Regulation of tooth mineralization
tbc1d20	854-1103	craniofacial	eye development	--	--	--	Cell morphogenesis involved in differentiation, Sensory organ development, Sensory system development, Visual system development, Eye development, Camera-type eye development, Sensory organ morphogenesis, Eye morphogenesis, Camera-type eye morphogenesis
smyd1	662-934	muscle	muscle tissue development	--	--	--	Circulatory system development, Heart development, Striated muscle tissue development
cenpf	452-618	muscle	muscle tissue development	oral jaw size	--	--	Negative regulation of cellular component organization, Striated muscle tissue development, Negative regulation of chromosome organization
pdlim5	550-736	muscle	muscle tissue development	oral jaw size	behavior	(Horiuchi et al. 2013)	Neuron projection development, Regulation of neurogenesis, Cell morphogenesis involved in differentiation, Cell part morphogenesis, Cell projection morphogenesis, Plasma membrane bounded cell projection morphogenesis, Neuron projection morphogenesis, Regulation of neuron differentiation, Developmental growth, Cell morphogenesis involved in neuron differentiation, Heart development,

							Striated muscle tissue development, Developmental cell growth, Cardiac muscle tissue development
bcor	505-727	craniofacial	mouth development	--	retina development	(Lee et al. 2013a)	Negative regulation of transcription by RNA polymerase II, Protein modification by small protein conjugation, Negative regulation of cellular component organization, Heart development, Negative regulation of chromosome organization, Negative regulation of chromatin organization, Negative regulation of histone modification, Regulation of tooth mineralization
fhod3	767-999	muscle	muscle tissue development	--	ear development	IMPC: https://www.mousephenotype.org/data/genes/MGI:1925847#phenotypesTab	Circulatory system development, Negative regulation of cellular component organization, Heart development, Striated muscle tissue development, Cardiac muscle tissue development
twist1	300-434	craniofacial	muscle tissue development;eye development	oral jaw size	mandibular arch skeleton	(Das and Crump 2012; Teng et al. 2018)	Camera-type eye development, DNA-templated transcription,Sensory organ morphogenesis, Appendage morphogenesis, Limb morphogenesis, Eye morphogenesis, Negative regulation of chromosome organization, Embryonic limb morphogenesis, Embryonic appendage morphogenesis, Camera-type eye morphogenesis, , Negative regulation of chromatin organization, Negative regulation of histone modification, Forelimb morphogenesis, Embryonic camera-type eye development, Embryonic eye morphogenesis, Embryonic forelimb morphogenesis, Embryonic camera-type eye morphogenesis, Eyelid development in camera-type eye, Embryonic camera-type eye formation

	zhx2	X	craniofacial	eye development	--	--	--	Camera-type eye development, Negative regulation of neurogenesis, Sensory organ morphogenesis, Negative regulation of neuron differentiation, Eye morphogenesis, Retina development in camera-type eye, Camera-type eye morphogenesis, Neural retina development, Retina morphogenesis in camera-type eye
	fhl2	890-1169	craniofacial	muscle tissue development	--	--	--	Response to lipid, Response to organic cyclic compound, Negative regulation of transcription by RNA polymerase II, Cellular response to lipid, Cellular response to organic cyclic compound, Heart development, Response to steroid hormone, Striated muscle tissue development, Intracellular receptor signaling pathway, Cardiac muscle tissue development
	prlh	1123-1466	feeding behavior	behavior; feeding behavior	--	--	(Takayanagi et al. 2008)	Developmental growth, Response to extracellular stimulus, Response to nutrient levels, Regulation of developmental growth, Eating behavior, Reduction of food intake in response to dietary excess
	ald1ha2	878-1279	craniofacial	muscle tissue development, eye development	--	limb morphogenesis	(Zhao et al. 2010)	Camera-type eye development, Developmental maturation, Response to nutrient, Cardiac muscle tissue development, Anatomical structure maturation, Appendage morphogenesis, Limb morphogenesis, Cellular hormone metabolic process, Vitamin metabolic process, Embryonic limb morphogenesis, Embryonic appendage morphogenesis, Response to vitamin, Cellular response to retinoic acid, Forelimb morphogenesis, Embryonic camera-type eye development, Embryonic forelimb morphogenesis

Table S7. Molluscivore adaptive alleles used for assessing stages of adaptation. We estimated ages for all adaptive alleles that were in or near (within 20-kb) of a gene associated with a GO term for behavior or craniofacial traits on the Ensemble 96 annotation database and were significantly enriched in our GO enrichment analysis (Table S5). Sweep ages, stage category assignment, any additional annotations we found for genes and their references are provided. Also included is a partial list of other significantly enriched GO terms for each gene. For visual clarity in the table, the broader GO terms (terms that > 1000 genes listed in database) are not included. See Table S5 for full list. Sweep ages are listed as the 95% HPD range (X indicates missing age estimates because estimates across starTMRCA runs did not converge for that sweep).

Gene	Sweep Age	Stages Category	GO enrichment annotations	GWAS annotations	Other researched relevant annotations	References for other annotations	Other GO enrichment annotations
cyp26b1	214-582	craniofacial	muscle tissue development	--	craniofacial development	(Laue et al. 2008; Spoorendonk et al. 2008)	Striated muscle tissue development, Intracellular receptor signaling pathway, Response to xenobiotic stimulus, Sensory organ morphogenesis, Response to nutrient, Appendage morphogenesis, Limb morphogenesis, Cellular hormone metabolic process, Vitamin metabolic process, Embryonic limb morphogenesis, Embryonic appendage morphogenesis, Response to vitamin, Cellular response to retinoic acid
ext1	405-687	craniofacial	cranial skeletal system development	nose height	--	(Mccormick et al. 1998)	Neuron projection development, Cell morphogenesis involved in differentiation,Cell part morphogenesis,Cell projection morphogenesis, Plasma membrane bounded cell projection morphogenesis, Neuron projection morphogenesis, Cell morphogenesis involved in neuron differentiation, Axon development, Axonogenesis
gnaq	180-737	craniofacial	skeletal system development	--	pigmentation; jaw size	(Fitch et al. 2003; Shirley et al. 2013)	Regulation of biological quality, Response to organic substance, Cellular response to chemical stimulus, Negative regulation of macromolecule metabolic process, Cellular response to organic substance, Negative regulation of nitrogen compound metabolic process, Response to external stimulus, Response to endogenous stimulus, Response to oxygen-containing compound, Cellular response to endogenous stimulus, Cellular response to oxygen-containing compound, Response to hormone

zhx2	1147-1793	craniofacial	eye development	--	--		Regulation of neurogenesis, Regulation of neuron differentiation, Sensory organ development, Sensory system development, Visual system development, Eye development, Camera-type eye development, Negative regulation of neurogenesis, Sensory organ morphogenesis, Negative regulation of neuron differentiation, Eye morphogenesis, Retina development in camera-type eye, Camera-type eye morphogenesis, Neural retina development, Retina morphogenesis in camera-type eye
tiparp	x	craniofacial	mouth development; muscle tissue development	--		--	Response to organic cyclic compound, Cellular response to organic cyclic compound , Response to xenobiotic stimulus, Cellular hormone metabolic process, Roof of mouth development
atp8a2	x	feeding behavior	behavior; feeding behavior	--		(McMillan et al. 2018)	Developmental growth, Response to extracellular stimulus, Response to nutrient levels, Eating behavior, Reduction of food intake in response to dietary excess

Table S8. 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 S9. Per generation mutation rate estimation from high coverage sequencing of parents and F1 from two crosses of San Salvador Island (SSI) species. Details about the average coverage of genome sequences in three offspring across two crosses, the number of de novo variants at steps in the filtering pipeline, and the specific filter thresholds used for each individual to filter down to high quality de novo variants in each (shared alleles).

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 alleles (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 S10. Parameters for selective sweep analyses in SweeD.
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 S11. The number of introgression regions in the SSI specialists. We determined introgressed regions of the genome as a 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 (alleles with $F_{st} \geq 0.95$) and signatures of a hard selective sweep (above demographic simulation based thresholds SweeD CLR > 5.28; OmegaPlus ω > 3.31 for scale-eaters and SweeD CLR > 4.47; OmegaPlus ω > 4.23 for molluscivores) to determine the number of adaptive introgression regions. These adaptive introgression regions range in size from 50-kb to 110-kb in length. For each introgression test, *C. artifrons* was used as the outgroup population (e.g. O) while the other specialist was used as the sister species (e.g. P1).

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 S12. Caribbean pupfish populations used to detect signatures of introgression in San Salvador Island (SSI) specialists and generalist lineages on other islands. The f_d 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 S13. Candidate adaptive introgression regions from Rum Cay generalists (*C. variegatus*) and San Salvador Island (SSI) specialists. We determined introgressed regions of the genome as regions 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 Rum Cay were then overlapped with regions of the genome with strong genetic divergence (alleles with $F_{st} \geq 0.95$) and signatures of a hard selective sweep (above demographic simulation-based thresholds SweeD CLR > 5.28; OmegaPlus ω > 3.31 for scale-eaters and SweeD CLR > 4.47; OmegaPlus ω > 4.23 for molluscivores) to determine the number of adaptive introgression regions. For each introgression test, *C. artifrons* was used as the outgroup population (e.g. O) while the other specialist was used as the sister species (e.g. P1).

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 S14. Candidate adaptive introgression regions from Dominican Republic generalists (*C. higuey*) and San Salvador Island (SSI) specialists. We determined introgressed regions of the genome as regions 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 Dominican Republic population were then overlapped with regions of the genome with strong genetic divergence (alleles with $F_{st} \geq 0.95$) and signatures of a hard selective sweep (above demographic simulation based thresholds SweeD CLR > 5.28;OmegaPlus ω > 3.31 for scale-eaters and SweeD CLR > 4.47; OmegaPlus ω > 4.23 for molluscivores) to determine the number of adaptive introgression regions. For each introgression test, C. artifrons was used as the outgroup population (e.g. O) while the other specialist was used as the sister species (e.g. P1).

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 S15. Candidate adaptive introgression regions from New Providence Island generalists (*C. laciniatus*) and San Salvador Island (SSI) specialists. We determined introgressed regions of the genome as regions 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 New Providence Island population were then overlapped with regions of the genome with strong genetic divergence (alleles with $F_{st} \geq 0.95$) and signatures of a hard selective sweep (above demographic simulation based thresholds SweeD CLR > 5.28;OmegaPlus ω> 3.31 for scale-eaters and SweeD CLR > 4.47; OmegaPlus ω > 4.23 for molluscivores) to determine the number of adaptive introgression regions. For each introgression test, *C. artifrons* was used as the outgroup population (e.g. O) while the other specialist was used as the sister species (e.g. P1).

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 S16. Candidate adaptive introgression regions from North Carolina Coast generalists (*C. variegatus*) and San Salvador Island (SSI) specialists.
We determined introgressed regions of the genome as regions 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 North Carolina population were then overlapped with regions of the genome with strong genetic divergence (alleles with $F_{st} \geq 0.95$) and signatures of a hard selective sweep (above demographic simulation based thresholds SweeD CLR > 5.28; OmegaPlus ω> 3.31 for scale-eaters and SweeD CLR > 4.47; OmegaPlus ω> 4.23 for molluscivores) to determine the number of adaptive introgression regions. For each introgression test, *C. artifrons* was used as the outgroup population (e.g. O) while the other specialist was used as the sister species (e.g. P1).

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 S17. Selective sweep ages on San Salvador Island using coalescent-based starTMRCA approach. The 95% high posterior density region of the posterior distribution of sweep ages for all denovo and introgressed adaptive alleles in scale-eater estimated using starTMRCA. A selection of standing variants that were calculated for the stages of adaptation analyses (GO terms related to behavior and craniofacial morphology) included as well. Introgressed adaptive alleles are labeled by the population introgressed from: New Providence Island (INTRO.NP), Dominican Republic (INTRO.DR), and North Carolina (INTRO.NC).

Gene	Spatial Distribution	Scaffold	Position	Mean Age	95 % HPD Lower	95% HPD Upper
scaf34.NA	INTRO.NC	c_bro_v1_0_scaf34	17475008	2583	2277	2871
card8	SGV	c_bro_v1_0_scaf46	1311093	2095	1728	2488
scaf46.NA	SGV	c_bro_v1_0_scaf46	13200234	1867	1637	2088
scaf11.NA	de novo	c_bro_v1_0_scaf11	21634014	1585	1388	1799
scaf52.NA	INTRO.NC	c_bro_v1_0_scaf52	4987013	1463	1219	1709
cmbl	de novo	c_bro_v1_0_scaf11	9924142	1375	1199	1566
galnt1	INTRO.NP	c_bro_v1_0_scaf53	20864827	1365	1227	1513
prlh	INTRO.NP	c_bro_v1_0_scaf11	9496004	1289	1124	1466
scaf37.NA	INTRO.NC	c_bro_v1_0_scaf37	14881950	1284	1099	1492
atp8a1	SGV	c_bro_v1_0_scaf44	14973114	1277	1119	1419
trim46	INTRO.DR	c_bro_v1_0_scaf37	3700741	1268	1098	1423
scaf44.NA	INTRO.NC	c_bro_v1_0_scaf44	28137436	1268	1147	1393
scaf6.NA	de novo	c_bro_v1_0_scaf6	923414	1259	1070	1462
scaf53.NA	SGV	c_bro_v1_0_scaf53	4776006	1222	1082	1380
gpr83	INTRO.NC	c_bro_v1_0_scaf1	38363517	1203	1079	1337
scaf8.NA	INTRO	c_bro_v1_0_scaf8	16314185	1193	978	1425
scaf43.NA	INTRO.NC	c_bro_v1_0_scaf43	27190362	1185	1070	1295
scaf6.NA	INTRO	c_bro_v1_0_scaf6	955941	1174	1045	1318
scaf24.NA	DENOVO	c_bro_v1_0_scaf24	20383519	1159	956	1361
eya2	SGV	c_bro_v1_0_scaf34	32255078	1131	962	1296
scaf53.NA.NC	INTRO.NC	c_bro_v1_0_scaf53	10409675	1096	926	1318
cfap20	SGV	c_bro_v1_0_scaf37	5095975	1093	974	1216
aldh1a2	SGV	c_bro_v1_0_scaf5	27704112	1063	878	1279
aasdhppt	SGV	c_bro_v1_0_scaf21	26917283	1046	916	1175
fhl2	SGV	c_bro_v1_0_scaf9	25305758	1020	891	1170
cadps	INTRO.DR	c_bro_v1_0_scaf34	25417185	1012	917	1115
grid2ip	SGV	c_bro_v1_0_scaf8	21601776	998	877	1127
chrna7	SGV	c_bro_v1_0_scaf37	3593615	986	864	1111
scaf34.NA.DR	INTRO.DR	c_bro_v1_0_scaf34	22649365	984	773	1171

st7l	de novo	c_bro_v1_0_scaf34	31258254	940	757	1163
scaf43.NA	INTRO.NC	c_bro_v1_0_scaf43	18320970	936	805	1079
fhod3	SGV	c_bro_v1_0_scaf53	18640776	888	768	999
crocc	INTRO.NP	c_bro_v1_0_scaf16	32982520	888	740	1060
galr2	de novo	c_bro_v1_0_scaf8	19961303	861	696	1008
nr4a2	SGV	c_bro_v1_0_scaf52	13841760	853	762	942
ppp3r1	INTRO.NP	c_bro_v1_0_scaf26	30297160	851	725	968
pde4d	DENOVO	c_bro_v1_0_scaf21	32304491	847	747	958
th/nap1l4	SGV	c_bro_v1_0_scaf19	7822448	847	747	958
dysf	SGV	c_bro_v1_0_scaf24	20221166	830	683	978
gse1	INTRO.DR	c_bro_v1_0_scaf5	28411973	829	700	947
mag	SGV	c_bro_v1_0_scaf53	17420175	824	615	1042
scaf53.NA.NP	INTRO.NP	c_bro_v1_0_scaf53	12368389	821	700	940
smyd1	SGV	c_bro_v1_0_scaf39	1662237	802	662	935
itga5	SGV	c_bro_v1_0_scaf18	28962001	792	639	932
rmi1	SGV	c_bro_v1_0_scaf39	4281152	789	652	952
ptprs	de novo	c_bro_v1_0_scaf16	8251751	789	555	1043
scaf44.NA	de novo	c_bro_v1_0_scaf44	10558794	782	676	898
chpf	de novo	c_bro_v1_0_scaf52	21897888	776	659	897
scaff44.NA.2	de novo	c_bro_v1_0_scaf44	16942340	763	626	890
ncoa2	INTRO.NP	c_bro_v1_0_scaf11	11975827	760	622	903
scaf19.NA	INTRO.NP	c_bro_v1_0_scaf19	6605756	756	641	862
tcf12	INTRO.NP	c_bro_v1_0_scaf5	27887771	729	606	862
abhd8	SGV	c_bro_v1_0_scaf16	13454820	729	624	821
serpinb1	SGV	c_bro_v1_0_scaf16	10637011	729	624	821
tdrd5	SGV	c_bro_v1_0_scaf16	12833025	720	618	822
zfhx4	de novo	c_bro_v1_0_scaf11	8072317	694	595	809
scaf53.NA.4	INTRO.DR	c_bro_v1_0_scaf53	32457769	675	573	783
scaf5.NA	INTRO.NP	c_bro_v1_0_scaf5	28307404	675	558	798
pdlim5	SGV	c_bro_v1_0_scaf47	24141970	645	550	737
bcor	SGV	c_bro_v1_0_scaf52	5558993	613	506	727
slc16a1	SGV	c_bro_v1_0_scaf18	29613954	556	460	662
tmem26	de novo	c_bro_v1_0_scaf43	26585181	546	473	617
cenpf/kcnk2	INTRO.DR	c_bro_v1_0_scaf44	12538313	533	452	619
hdac9b	INTRO.DR	c_bro_v1_0_scaf53	18998120	445	367	538
scaf52.NA.2	INTRO.DR	c_bro_v1_0_scaf52	13758756	424	333	521
mindy3	SGV	c_bro_v1_0_scaf53	20112997	420	334	501
znf628	SGV	c_bro_v1_0_scaf53	24744443	420	334	501
olfm1	de novo	c_bro_v1_0_scaf47	14782939	398	294	507

otof	INTRO.NP	c_bro_v1_0_scaf7	12603683	371	227	525
twist1	de novo	c_bro_v1_0_scaf53	18968932	367	300	435
tfap2a	SGV	c_bro_v1_0_scaf34	32255078	359	293	431
mylipa	INTRO.NP	c_bro_v1_0_scaf37	8265887	206	95	326

Table S18. .Selective sweep ages on San Salvador Island using coalescent-based starTMRCA approach. The 95% high posterior density region of the posterior distribution of sweep ages for all introgressed candidate alelles in molluscivore estimated using starTMRCA. A selection of standing variants (SGV) that were calculated for the stages of adaptation analyses (GO terms related to behavior and craniofacial morphology) included as well. Introgressed adaptive alleles are labeled by the population introgressed from: New Providence Island (INTRO.NP), Dominican Republic (INTRO.DR), and North Carolina (INTRO.NC).

Gene	Spatial Distribution	Scaffold	Position	Mean Age	95 HPD Lower	95 HPD Upper
abhd8	SGV	c_bro_v1_0_scaf16	13455352	471	294	649
cox6b1	SGV	c_bro_v1_0_scaf53	24790621	402	236	577
cyp26b1	INTRO.NP	c_bro_v1_0_scaf24	20486531	396	215	582
ext1	SGV	c_bro_v1_0_scaf26	264812	546	405	687
gga1	SGV	c_bro_v1_0_scaf1	29209563	914	831	997
gnaq	SGV	c_bro_v1_0_scaf33	12883992	439	181	737
zhx2	SGV	c_bro_v1_0_scaf53	11080970	1490	1148	1793
znf628	SGV	c_bro_v1_0_scaf53	24744443	325	169	486
18.NA	INTRO.NP	c_bro_v1_0_scaf18	2258923	763	550	965
19.NA	INTRO.NC	c_bro_v1_0_scaf19	7642081	3560	3021	4080
4.NA	INTRO.NP	c_bro_v1_0_scaf4	16217615	369	227	532
43.NA.NCC	INTRO.NC	c_bro_v1_0_scaf43	27435224	408	249	576
5.NA	INTRO.DR	c_bro_v1_0_scaf5	27117941	726	469	997
52.NA	INTRO.DR	c_bro_v1_0_scaf52	4982714	820	599	1032
53.NA	INTRO.DR	c_bro_v1_0_scaf53	10434469	903	569	1339
53.NA.NCC	INTRO.NC	c_bro_v1_0_scaf53	10904586	3131	2428	3604
bri3bp.wdr31	INTRO.NP	c_bro_v1_0_scaf33	12644789	376	162	594
klf1	INTRO.DR	c_bro_v1_0_scaf1	29253566	1603	1504	1694
otof	INTRO.DR	c_bro_v1_0_scaf7	12606143	444	367	510
trim44	INTRO.DR	c_bro_v1_0_scaf19	6441342	15922	14639	17127
ttc33	INTRO.NP	c_bro_v1_0_scaf47	16146578	262	127	393
ube2w	INTRO.DR	c_bro_v1_0_scaf11	11268935	1022	903	1144

Table S19. Selective sweep ages on San Salvador Island (SSI) using coalescent-based McSwan approach. 95% high posterior density region of the posterior distribution of sweep ages of adaptive alleles in scale-eater and molluscivore genomes estimated using McSwan (Tournebize et al. 2019).

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	8490
prlh	habitat preference	c_bro_v1_0_scaf11	9200146	9276987	76841	6594	9210
card8	pigmentation	c_bro_v1_0_scaf46	1451011	1663431	212420	973	5097
kcnk2, cenpf	trophic morphology	c_bro_v1_0_scaf44	12227155	12305895	78740	2936	3966
smyd1	trophic morphology	c_bro_v1_0_scaf39	1643098	1647708	4610	3054	6030
tcf12	trophic morphology	c_bro_v1_0_scaf5	27975725	28016276	40551	1607	5119
twist1	trophic morphology	c_bro_v1_0_scaf53	18953132	19092361	139229	1636	3413
itga5	trophic morphology	c_bro_v1_0_scaf18	28040450	28049258	8808	1697	2357
Molluscivore							
ext1	trophic morphology	c_bro_v1_0_scaf26	162903	230930	68027	814	1060
tiparp	trophic morphology	c_bro_v1_0_scaf21	33602383	33606685	4302	3353	5003
cyp26b1	trophic morphology	c_bro_v1_0_scaf24	20527588	20602663	75075	1447	4567

Inter-chapter transition

In Chapter 4, I explored the evidence for introgression in the genomes of both radiating and non-radiating lineages of *Cyprinodon* pupfish. I found that the trophic specialist pupfish in the San Salvador Island radiation contained twice as much adaptive introgression in their genomes than found in the generalist-only outgroup pupfish lineages on other islands that don't contain a radiation**.** While these results helped strengthen the case for hybridization playing an important role in pupfish adaptive radiation, I also found evidence that non-radiating populations of pupfish experienced some level of introgression from the exact same sources as well. These results represent some of the first findings that histories of hybridization may be quite similar between radiating and non-radiating lineages and thus indicate that hybridization may be necessary but not sufficient to produce adaptive radiations in the presence of ecological opportunity. My first four chapters highlight the need to explore genomic patterns of selection, introgression and mutation in more depth to determine if other microevolutionary forces interact with hybridization to create adaptive radiations. One place to start in this effort is to conduct fine-scale investigations characterizing the genetic transitions that occur during particular diversification events, to better understand the impact that introgression following hybridization has on adaptive radiation. In Chapter 5, I characterize the timing of selection and genetic sources of candidate adaptive alleles involved in ecological transition to specializing on a diet of scales by doing a focused comparative genomic analysis of the generalist and two scale-eating specialist species in the radiation of pupfish on San Salvador Island.

Chapter 5: We get by with a little help from our friends: shared adaptive variation provides a bridge to novel ecological specialists during adaptive radiation

This chapter is currently in press with Proceedings of Royal Society B. A publicly available version of this chapter has been published as a biorxiv preprint and is reproduced here in accordance with the preprint server's article sharing policy:

Richards, E.J. and Martin, C.H. 2022. We get by with a little help from our friends: shared adaptive variation provides a bridge to novel ecological specialists during adaptive radiation. *Biorxiv*. DOI: 10.1101/2021.07.01.450755.

5.1. Abstract

Adaptive radiations involve astounding bursts of phenotypic, ecological, and species diversity. However, the microevolutionary processes that underlie the origins of these bursts are still poorly understood. We report the discovery of an intermediate 'wide-mouth' scale-eating ecomorph in a sympatric radiation of *Cyprinodon* pupfishes, illuminating the transition from a widespread algae-eating generalist to a novel microendemic scale-eating specialist. We first show that this ecomorph occurs in sympatry with generalist *C. variegatus* and scale-eating specialist *C. desquamator* on San Salvador Island, Bahamas, but is genetically differentiated, morphologically distinct, and often consumes scales. We then compared the timing of selective sweeps on shared and unique adaptive variants in trophic specialists to characterize their adaptive walk. Shared adaptive regions swept first in both the specialist *desquamator* and the intermediate 'wide-mouth' ecomorph, followed by unique sweeps of introgressed variation in 'wide-mouth' and de novo variation in *desquamator*. The two scale-eating populations additionally shared 9% of their hard selective sweeps with molluscivores *C. brontotheroides*, despite no single common ancestor among specialists. Our work provides a new microevolutionary framework for investigating how major ecological transitions occur and illustrates how both shared and unique genetic variation can provide a bridge for multiple species to access novel ecological niches.

5.2. Introduction

Rapid bursts of diversification and repeated bouts of speciation like those seen in adaptive radiations contradict current mechanistic speciation models that predict diversification should slow with time as available niche space becomes increasingly subdivided and disruptive selection becomes weaker with each recurrent speciation event (e.g. (Dieckmann and Doebeli 1999; Polechová et al. 2005; Bolnick 2006)). Diversification on complex adaptive landscapes

with multiple empty fitness peaks corresponding to different niches provides an alternative mechanism to niche subdivision (Kondrashov and Kondrashov 1999; Gavrilets 2014; Martin and Richards 2019). However, these landscapes present a new problem to our mechanistic understanding of adaptive radiations: How do populations manage to escape local optima, cross fitness valleys, and access new fitness peaks (Arnold et al. 2001; Svennson and Calsbeek 2012; Martin and Wainwright 2013c; Martin and Gould 2020)? Colonizing new fitness peaks on the adaptive landscape presents challenges because it requires transitions in behaviors, morphological traits, or a combination of the two that allow organisms to adapt to new ecological niches (Calsbeek and Irschick 2007). Spectacular ecological transitions do often occur during adaptive radiation, such as blood-drinking (Grant and Grant 2008) or plant carnivory (Givnish et al. 1984, 1997), yet it is still poorly understood how such seemingly discontinuous transitions occur.

Recent conceptual frameworks for understanding adaptation to novel fitness peaks suggest that these major ecological transitions likely occur in stages of potentiation, actualization and refinement (Blount et al. 2012; Erwin 2021). The initial emergence of a novel trait likely requires further refinement to become successfully incorporated into the functional ecology of an organism. Several experimental lab studies suggest that novel ecological transitions are highly contingent on accruing a series of mutations that incrementally refine adaptations to colonize new fitness peaks (Blount et al. 2012; Quandt et al. 2014). This idea that genetic background is important in setting the stage for adaptation also underlies many hypotheses for adaptive radiation, such as the hybrid swarm and syngameon hypotheses – in which radiations are driven by acquiring novel combinations of alleles through the exchange of genetic variation either from distinct lineages outside the radiation or within the radiation itself (Seehausen 2004). However, we are only just beginning to explore how gene flow and shared genetic variation gives recipient lineages access to new fitness peaks in the wild and generates adaptive radiations (Martin and Richards 2019).

An adaptive radiation of trophic specialist pupfishes on San Salvador Island (SSI) in the Bahamas is an excellent system for understanding how the rapid evolution of major ecological transitions occurs in nature. This radiation contains a widespread generalist pupfish species (*Cyprinodon variegatus*) that occurs in sympatry with two previously described trophic specialists that are endemic to the hypersaline lakes on the island: a molluscivore (*C. bronotheroides*) with a novel nasal protrusion which is an oral-sheller of gastropods (John et al. 2020) and a scale-eating specialist (*C. desquamator*) with two-fold larger oral jaws (Martin and Wainwright 2013a). The evolutionary novelties in this system originated recently; the lakes on SSI were dry during the last glacial maximum 6-20 kya years ago (Turner et al. 2008; Clark et al. 2009). Intriguingly, we recently discovered a fourth species of pupfish living in sympatry with the two specialists and generalist on SSI (Richards and Martin 2017b). This species exhibits intermediate jaw morphology between *C. desquamator* and *C. variegatus* (figure 1). Here we refer to this new ecomorph as the 'wide-mouth' because its mouth is wider than any other species in the radiation. The multi-peak fitness landscape driving this radiation suggests that *C. desquamator* is isolated by a large fitness valley from *C. variegatus* and *C. brontotheroides* (Martin and Wainwright 2013c) and this intermediate 'wide-mouth' may provide clues about the microevolutionary processes underlying how the observed novel fitness peaks are traversed in the wild.

Here we first investigate the position of the 'wide-mouth' on the ecological spectrum from generalist to scale-eating specialist using a combination of morphological, behavioral,

dietary, and genomic data. We then estimated the demographic history of the 'wide-mouth' and explored the spatial origins and timing of selection on shared and unique genetic variation involved in adaptation to scale-eating to better understand this ecological transition. Our results suggest that while intermediate in jaw length known to be relevant for the highly specialized scale-eater *C. desquamator, C sp*. 'wide-mouth' demonstrates transgressive morphology and a distinct genetic background. Our investigation of the timing of selection and genetic origins of the adaptive alleles shared and unique between the two scale-eating species indicates divergent adaptive walks that are highly dependent on their genetic background. Despite shared origins, access to unique genetic variation in each of the two scale-eating sister species likely resulted in distinct adaptive walks and ultimately contributed to the diversity of ecological specialists observed in this radiation.

5.3. Methods

5.3.1. Ecological and morphological characterization of 'wide-mouth' scale-eater

C. variegatus, C. desquamator, and *C.* sp. 'wide-mouth' individuals from 3 lake populations (Osprey Lake, Great Lake, and Oyster Pond) in which we had sufficient specimens (n=84; *C. brontotheroides* not shown) were measured for 9 external morphological traits using digital calipers. Traits were selected for specific connections to foraging performance which differed across the three species in a previous study (Martin and Wainwright 2013c). We also characterized diet for *C. variegatus, C. desquamator,* and 'wide-mouth' in Osprey Lake from stomach content analyses (*n*=10 per species) and stable isotope analyses of muscle tissue from wild-collected samples (*n*=75). Dietary overlap was characterized by comparison of population mean scale count from gut contents using ANOVA and ellipse areas and bivariate means on isotope biplots using SIBER (Jackson et al. 2011). See supplemental methods for more details on sample sizes and analyses.

5.3.2. Genomic library preparation and variant filtration

To explore the evolutionary history of *C.* sp. 'wide-mouth', we sequenced whole genomes of 24 individuals following protocols used in a previous study (Richards et al. 2021) that included genomes from *C. variegatus, C. desquamator,* and *C. brontotheroides*. Our final genetic dataset after filtering contained 6.4 million variants across 110 individuals from the four species (7.9x median coverage). See supplemental methods for the full sequencing and genotyping protocol.

5.3.3. Genomic origins of the 'wide-mouth' scale-eater

We first tested whether these *C.* sp. 'wide-mouth' individuals represented recent (e.g. F1/F2) hybrids of *C. variegatus* and *C. desquamator* in the wild using principal component and ADMIXTURE analyses to look for the genome-wide pattern expected in PCAs when recent hybrids between two populations are included. We also used formal tests for introgression and

admixed populations, f_3 and f_4-statistics (Peter 2016), to assess whether 'wide-mouth' are the byproduct of recent admixture. Finally, we used *fastsimcoal2* (v2.6.0.3;(Excoffier et al. 2013)), a demographic modeling approach based on the folded minor allele frequency spectrum (mSFS), to discriminate among alternative evolutionary scenarios for the origin of 'wide-mouth' and estimated divergence times among all four species based on the best model fit from an AIC test (see supplementary methods for more detail).

5.3.4. Characterization of unique and shared adaptive alleles among specialists

Across all four populations in Osprey Lake, we looked for regions that showed evidence of a hard selective sweep using SweeD (v.3.3.4;(Pavlidis et al. 2013)). The composite likelihood ratio (CLR) for a hard selective sweep was calculated in 50-kb windows across scaffolds that were at least 100-kb in length (99 scaffolds; 85.6% of the genome). Significance thresholds were determined using CLR values from neutral sequences simulated under MSMC inferred demographic scenarios of historical effective population size changes (Supplemental methods; figure S1; table S3).

Next, we searched for candidate adaptive alleles associated with species divergence by overlapping selective sweep regions with regions of high genetic divergence based on fixed or nearly fixed SNPs between species. We chose to also look at regions with nearly fixed SNPs ($F_{st} \geq 0.95$) to accommodate ongoing gene flow between these young species. F_{st} between the populations and species was calculated per variant site using –weir-pop-fist function in vcftools (v.0.1.15;(Danecek et al. 2011b)).

5.3.5. Timing of selection on candidate adaptive alleles

We also determined the relative age of candidate adaptive alleles by generating estimates of coalescent times using starTMRCA (v0.6.1;(Smith et al. 2018)). For each candidate adaptive allele that was unique to the three specialists and the 16 shared alleles between *C. desquamator* and 'wide-mouth', a 1-Mb window surrounding the variant was extracted into separate vcfs for each species. These sets of variants were then analyzed in starTMRCA with a mutation rate of 1.56×10^{-8} substitutions per base pair (from Caribbean pupfishes (Richards et al. 2021)) and a recombination rate of 3.11×10^{-8} (from stickleback; (Roesti et al. 2013)). Each analysis was run three times per focal adaptive allele and all runs were checked for convergence between and within runs. Most runs rapidly converged within the initial 6000 steps, but 5 runs did not converge after an additional 4000 steps and were discarded from further analysis. See supplementary methods for more details on timing analyses.

5.3.6. Characterization of adaptive introgression adaptive alleles in *C.* sp 'wide-mouth'

Lastly, we investigated the spatial origins of adaptive alleles shared and unique to the two scale-eating specialists by searching in our previous study spanning Caribbean-wide outgroup populations for these alleles (Richards et al. 2021). Adaptive alleles were assigned as standing genetic variation if observed in any population outside SSI or de novo if they were only observed within populations on SSI. Additionally, we investigated signatures of introgression across the

genome of *C.* sp 'wide-mouth' and *C. desquamator* to determine if they showed evidence of adaptive introgression from outgroup generalist populations as observed previously (Richards et al. 2021). See supplementary methods for more details on introgression analyses.

5.4. Results

5.4.1. 'Wide-mouth' ecomorph is ecologically intermediate and morphologically distinct

We found that the 'wide-mouth' ecomorph is morphologically distinct from *C. desquamator* and *C. variegatus* across a suite of craniofacial traits (figure 2A-B). The lower jaw length of 'wide-mouth' was intermediate between *C. desquamator* and *C. variegatus* (figure 2C), while the buccal width and adductor mandibulae height were 8% larger in 'wide-mouth' than *C. desquamator* (figure 2E-F). These morphological differences were consistent across multiple lakes (figure S2). Small modifications in craniofacial morphology among these species have major impacts on scale-eating performance in this system by altering kinematic traits such as peak gap size which is partially controlled by the length of the lower jaw, jaw protrusion distance, and the angle of the lower jaw relative to the suspensorium (St. John et al. 2020).

C. sp 'wide-mouth' also did not show morphological divergence comparable to that observed in the molluscivore *C. brontotheroides.* The molluscivore specialist presents an opposing axis of morphological divergence to the scale-eating specialists, with shorter oral jaw length and larger eye diameter than even the generalist *C. variegatus*, in addition to a novel nasal protrusion of the maxilla not observed in any other Cyprinodontidae species (Martin and Wainwright 2013d).

Morphological traits were heritable in a common garden laboratory environment after one generation: lab-reared *C.* sp. 'wide-mouth' displayed significantly larger buccal width than *C. desquamator* (t-test; $P = 0.003$) and maintained their characteristic intermediate jaw lengths (ANOVA; $P = 0.03$, figure S3). There was also some evidence of phenotypic plasticity in both *C. desquamator* and *C. sp* 'wide-mouth' compared to wild individuals likely caused by the common lab diet. See supplementary results for more details.

5.4.2. 'Wide-mouth' occupies a distinct intermediate scale-eating niche

We found that 'wide-mouth' ingested scales, but at a significantly lower frequency than C. *desquamator* (Wilcoxon Rank Sum test, $P = 0.004$; figure 3A). We did not detect any scales in *C. variegatus* guts (figure 3A). Detritus made up the rest of the *C. sp.* 'wide-mouth' and *C. desquamator* diets and was the dominant component of *C. variegatus* gut contents, except for a single individual with one mollusc shell. A previous study that characterized contents of *C. variegatus*, *C. brontotheroides*, and *C. desquamator* populations across several ponds also found detritus to be the dominant component of each species' diet (49-71%) and nearly zero scales in the gut contents of *C. variegatus* and *C. brontotheroides* (Martin and Wainwright 2013d).

The intermediate scale-eating dietary niche of the wide-mouth ecomorph is complemented by our stable isotope analyses, which provide long-term snapshots of the carbon

sources and relative trophic levels in these species. Osprey Lake individuals collected on the same day from the same site differed in δ15N levels across species (ANOVA, $P = 4.55x10^{-6}$; figure 3B and S4); 'wide-mouth' δ15N was intermediate between *C. variegatus* and *C. desquamator* (Tukey HSD; $P= 1.34 x10^{-5}$ & $1.11 x10^{-4}$ respectively), supporting its intermediate trophic position. SIBER analyses of trophic niche position indicate distinct positioning of the wide-mouth ecomorph based on the lack of extensive overlap in niche space measured by standard ellipse areas and bivariate means with 95% confidence intervals of isotope values among the species (figure 3B).

5.4.3. C. sp. 'wide-mouth' did not result from hybridization between *C. variegatus* and *C. desquamator*

Several lines of genomic evidence from PCA, ADMIXTURE, and *f-statistic* analyses support the 'wide-mouth' ecomorph as a genetically distinct species rather than a recent hybrid between *C. desquamator* and *C. variegatus* on SSI as their intermediate jaw morphology might suggest (figure 4A-C & S5-6; see Supplementary material for more details). Demographic modeling of divergence and gene flow on SSI places *C. sp* 'wide-mouth' as sister to *C. desquamator*, supporting previous phylogenetic inference (Richards and Martin 2017a). In the best supported model of 28 demographic models tested (table S2), 'wide-mouth' and *C. desquamator* diverged 11,658 years ago (95 CI: 8,257-20,113 years; figure S7B; table S2) with ongoing gene flow. Additionally, the amount of genetic divergence between populations indicates that *C. desquamator* and *C.* sp. 'wide-mouth' are more genetically diverged from each other than to the generalist *C. variegatus* (e.g. F_{st} in figure 4C).

5.4.4. C. sp 'wide-mouth' is comprised of both shared and unique adaptive alleles

Next we looked at regions of the genome in both *C. desquamator* and *C.* sp. wide-mouth that showed strong evidence of hard selective sweeps. We found 6 shared hard selective sweeps in both species containing a total of 15 SNPS that were fixed or nearly-fixed compared to the sympatric generalist *C. variegatus* (figure 4E, 5): 10 SNPs were in unannotated regions, two were in the introns of the gene *daam2,* and three were in putative regulatory regions (with 20-kb) of the genes *usp50, atp8a1,* and *znf214* (one variant each). Shared adaptive alleles in the gene *daam2,* a wnt signaling regulator, are intriguing because knockdown of this gene causes abnormal snout morphology in mice (Dickinson et al. 2016) and abnormal cranial and skeletal development in zebrafish (Kida et al. 2007).

We also found unique sets of adaptive alleles in *C.* sp 'wide-mouth' and *C. desquamator* (figure 4E;figure 5). None of the adaptive alleles unique to *C.* sp. 'wide-mouth' were in or near genes annotated for craniofacial phenotypes in model organisms, despite its distinctive craniofacial morphologies. In *C. desquamator,* three of 12 unique adaptive alleles were in or near genes associated with or known to affect craniofacial phenotypes: a *de novo* non-synonymous coding substitution in the gene *twist1*, several putative regulatory variants near the gene *gnaq,* and 8 variants in or near the gene *bri3bp,* which is located inside a QTL region for cranial height in pupfish (St. John, Michelle E. et al. 2021). *C. brontotheriodes* also contained at least one unique candidate craniofacial adaptive allele: a non-synonymous coding substitution in the gene *kat6b* (figure 5), which is associated with abnormal craniofacial morphologies, including shorter mandibles, in mice (Thomas et al. 2000). This pattern of unique alleles relevant to craniofacial phenotypes in specialists *C. brontotheriodes* and *C. desquamator*, but not *C.* sp 'wide-mouth',

holds even if we lower the threshold to the top 1 percentile of F_{st} outliers between specialists and generalist (see supplemental results; figure S10).

5.4.5. The origins of adaptive alleles in C. sp 'wide-mouth' and desquamator

The adaptive alleles shared by *C. desquamator* and *C.* sp. 'wide-mouth' occurred as low frequency standing genetic variation in the Caribbean, with the exception of a single de novo allele on SSI located in an unannotated region on scaffold 6 (figure 5). The adaptive alleles unique to *C. desquamator* and *C.* sp 'wide-mouth' also predominantly came from standing genetic variation (84% and 81%, respectively). 14% of adaptive alleles unique to *C. desquamator* were de novo mutations to SSI and 2% occurred in candidate introgression regions (table S7). We found the opposite in *C. sp* 'wide-mouth': only 4% of their unique adaptive alleles were de novo mutations whereas 15% occurred in candidate introgression regions (table S8). This adaptive introgression was detected for generalist populations sampled from North Carolina and Laguna Bavaro in the Dominican Republic (table S8; figure S11). Using the Relative Node Depth (RND) statistic, we also discovered that 5 of the 6 shared adaptive alleles (all except for the unannotated region on scaffold 43;table S6) appear introgressed between *C. desquamator* and *C.* sp. 'wide-mouth', suggesting a substantial contribution of introgression to the adaptive alleles observed in *C.* sp. 'wide-mouth'.

5.4.6. Timing of selection on adaptive alleles reveals features of the adaptive walk to scale-eating

Selective sweeps occurred much more recently in both populations than their inferred divergence times (figure 5). Intriguingly, selection on 4 of the 6 adaptive alleles occurred significantly earlier in *C. desquamator* than 'wide-mouth'. Only a single adaptive allele had an older median age estimate in 'wide-mouth' than *C. desquamator*, although the 95% HPD intervals overlapped between the species (figure 5). Additionally, overall we found a significant difference in timing of selection between shared and unique adaptive alleles in the two scale-eater populations (ANOVA *P*-value = 0.00478). In *C. desquamator*, shared adaptive alleles swept before any unique adaptive alleles (Tukey HSD *P*-value = 0.003217; figure 5). For the 'wide-mouth', shared adaptive alleles with *C. desquamator* also generally swept before those unique to the species, despite these unique alleles being standing and introgressed variation from the Caribbean (figure 5). However, this difference was not significant due to one unique adaptive allele (*slitrk5)* that swept first (figure 5; ANOVA, Tukey HSD; *P* = 0.8367). This adaptive allele resides in a region that appears to be introgressed with the Laguna Bavaro generalist population in the Dominican Republic where this allele also show signs of a hard selective sweep (Richards et al. 2021). The older age estimate of this sweep in *C.* sp 'wide-mouth' might be due to older shared selection for the alleles in other Caribbean populations before introgression with *C.* sp 'wide-mouth'. All other introgressed adaptive alleles in *C.* sp. 'wide-mouth' swept more recently than shared sweeps with *desquamator*, including the shared de novo allele, and were not under selection in outgroup generalist populations.

Intriguingly, all but one of the de novo adaptive alleles in *desquamator* swept at the same time in the recent past (figure 5). Only one of these adaptive alleles in *olfm1* region has 95% HPD age range that overlaps with the next oldest selective sweep on standing genetic variation (*gnaq*; figure 5), suggesting a discrete stage of selection on new mutations in *C. desquamator*.

5.4.7. Shared signatures of selection across the three specialists in the radiation

Lastly, we compared the genetic divergence and selection patterns observed in the two scale-eating specialists to the divergent molluscivores specialist *C. brontotheroides* to investigate the extent of allele sharing among all three trophic specialists in this adaptive radiation. We found that no fixed or nearly-fixed alleles relative to the generalist *C. variegatus* were shared across all three specialists (figure S9-S10; supplementary results). However, we did find evidence of 44 shared selective sweeps across all three specialist populations that were not shared with *C. variegatus* populations (figure S12C). These shared regions were significantly enriched for genes annotated for metabolic processes (figure S12D), suggesting shared selection for metabolizing the more protein-rich diet across all three trophic specialists (also see (Mcgirr and Martin 2018)).

5.5. Discussion

5.5.1. Discovery of a new cryptic intermediate scale-eater highlights the power of reusing genetic variation to access novel niches

The hallmark of adaptive radiation is a rapid burst of diversification which is predicted by theory to slow down over time as niche subdivision increases (Martin and Richards 2019). An alternative possibility is that radiations can be self-propagating and that the diversity generated within the first stages of radiation helps beget further diversity (Whittaker 1977). This could happen through exploitation of new trophic levels created by new species or physical alterations of the environment by new species that may create additional opportunities for speciation (reviewed in (Stroud and Losos 2016; Martin and Richards 2019)). The diversity begets diversity hypothesis can also be visualized as the exploration of a complex multi-peaked fitness landscape; as species in the radiation colonize new peaks, this provides access to additional neighboring fitness peaks to fuel rapid radiation. Our discovery of a cryptic new scale-eating species through morphological, dietary, and genomic analyses revealed shared nearly-fixed or fixed adaptive alleles in both scale-eating species relative to the generalist *C. variegatus*. While *C.* sp. 'wide-mouth' is ecologically intermediate in its scale-eating behavior, our estimates of the relative timing of selective sweeps suggest that these shared alleles were first selected upon in the more specialized scale-eater *C. desquamator*, although unaccounted for demographic differences may also be contributing to differences in estimated timing between species.

Intriguingly, the shared adaptive alleles between *C. desquamator* and *C.* sp. 'wide-mouth' have potentially introgressed more recently rather than selected upon in their shared common ancestor. Five of the six regions surrounding these shared adaptive alleles showed patterns of high genetic similarity consistent with introgression (table S10). Alternatively, this genetic similarity may also be caused by strong background selection on shared ancestral variation. Effective population sizes are not drastically different between the two species and exon density is not in the upper tail of the genome wide-distribution (figure S1;table S10), two factors found in other studies where background selection tends to confound adaptive introgression inferences (Kim et al. 2018; Zhang et al. 2020). However, we do not have extensive knowledge of recombination breakpoints in this non-model system to distinguish between strong background selection on shared ancestral variation and adaptive introgression scenarios for each candidate adaptative introgression region.

We also found strong signatures of introgression in *C.* sp. 'wide-mouth' genomes from outgroup generalist populations that were not shared with *C. desquamator* (figure S11; table S11). Craniofacial morphology is a major axis of diversification between trophic specialists in this system (Martin and Wainwright 2011), yet *C.* sp. 'wide-mouth' appears to have little unique genetic variation relevant for craniofacial traits compared to the other two specialists (figure S10). Despite this, they do exhibit transgressive craniofacial phenotypes not seen in the other specialists. An intriguing implication of these findings is that hybridization may allow different species to share many of the same adaptive alleles to occupy distinct but similar niches, in line with the syngameon and 'diversity begets diversity' hypotheses of adaptive radiation (Whittaker 1977; Seehausen 2004).

5.5.2 An adaptive walk underlies the major ecological transition from generalist to scale-eating specialist

The foundational model of adaptation is that it proceeds in 'adaptive walks' towards fitness optima that involve the sequential fixation of adaptive alleles that move a population in the phenotypic direction of the local optimum (Orr 2005). The distinct timing of selection across different adaptive alleles in both *C. desquamator* and *C.* sp 'wide-mouth' suggests that the ecological transition from generalist to novel scale-eating specialist involved an adaptive walk in which selection on a beneficial allele was contingent on prior fixation of other adaptive alleles in each specialists' genetic background (see supplemental materials for further discussion). This is best highlighted by the pattern observed in *C. desquamator* in which nearly all de novo mutations swept at the same time in a distinct selective stage from other adaptive variants rather than being selected upon as they originated (figure 5).

5.5.3. The (un)predictability of adaptive walks to novel ecological niches

A recent study characterizing genotypic fitness landscapes underlying the transition from *C. variegatus* and *C. desquamator* revealed a rugged landscape with many local fitness peaks, likely due to epistatic interactions among alleles (Patton et al. 2021). The staggered timing of selection on alleles lends support to this finding. Epistasis can reduce the number of adaptive walks selection will promote (Weinreich et al. 2006), and might explain why the same adaptive alleles were the first to undergo hard selective sweeps in both 'wide-mouth' and *desquamator.*

We also found evidence for shared selective sweeps across all three specialists in regions that are enriched for genes annotated for metabolic processes such as short chain fatty acid and propionate metabolism (figure S12D). The lack of fixed alleles in these regions relevant to dietary specialization suggests polygenic selection (see supplemental for more discussion). Subtle shifts of allele frequencies across the genome can lead to divergent genomic backgrounds that give populations access to different ecological niches (e.g. (Hansen 2013; Otte et al. 2021)).

While both shared sweeps among all specialists and shared adaptive alleles among the two scale-eating species suggest constrained adaptive walks along overlapping genotypic pathways, we still see most selective sweeps are unique to each species in this radiation (figure 4; figure S11). Curiously, some adaptive standing genetic variation rose to high frequency in *C. desquamator* but did not similarly undergo selection in *C.* sp 'wide-mouth', despite its adaptation to a similar ecological niche and the presence of these alleles segregating at low frequency in the 'wide-mouth' population. This highlights the dual influence of epistatic interactions on adaptive

walks in rugged landscapes – epistasis reduces number of available paths but increases the number of local fitness peaks populations can get stuck on (Fragata et al. 2019). Selection on the same adaptive alleles may have allowed both scale-eating species access to the same area of the fitness landscape but epistatic interactions with private mutations and introgressed variation in each lineage may have resulted in divergent paths to scale-eating, ultimately contributing to diverse evolutionary outcomes even from a shared starting point.

The use of adaptive alleles from distinct spatial sources, the distinct morphologies and divergent genomic backgrounds, and potential introgression of adaptive alleles from the more specialized scale-eater *C. desquamator* into *C.* sp 'wide-mouth' reveals a tangled path for novel ecological transitions in nature. The complex epistatic interactions at microevolutionary scales implicated in this study make it all the more fascinating that novel ecological transitions are a common macroevolutionary feature of adaptive radiation.

5.6. Chapter 5 Figures

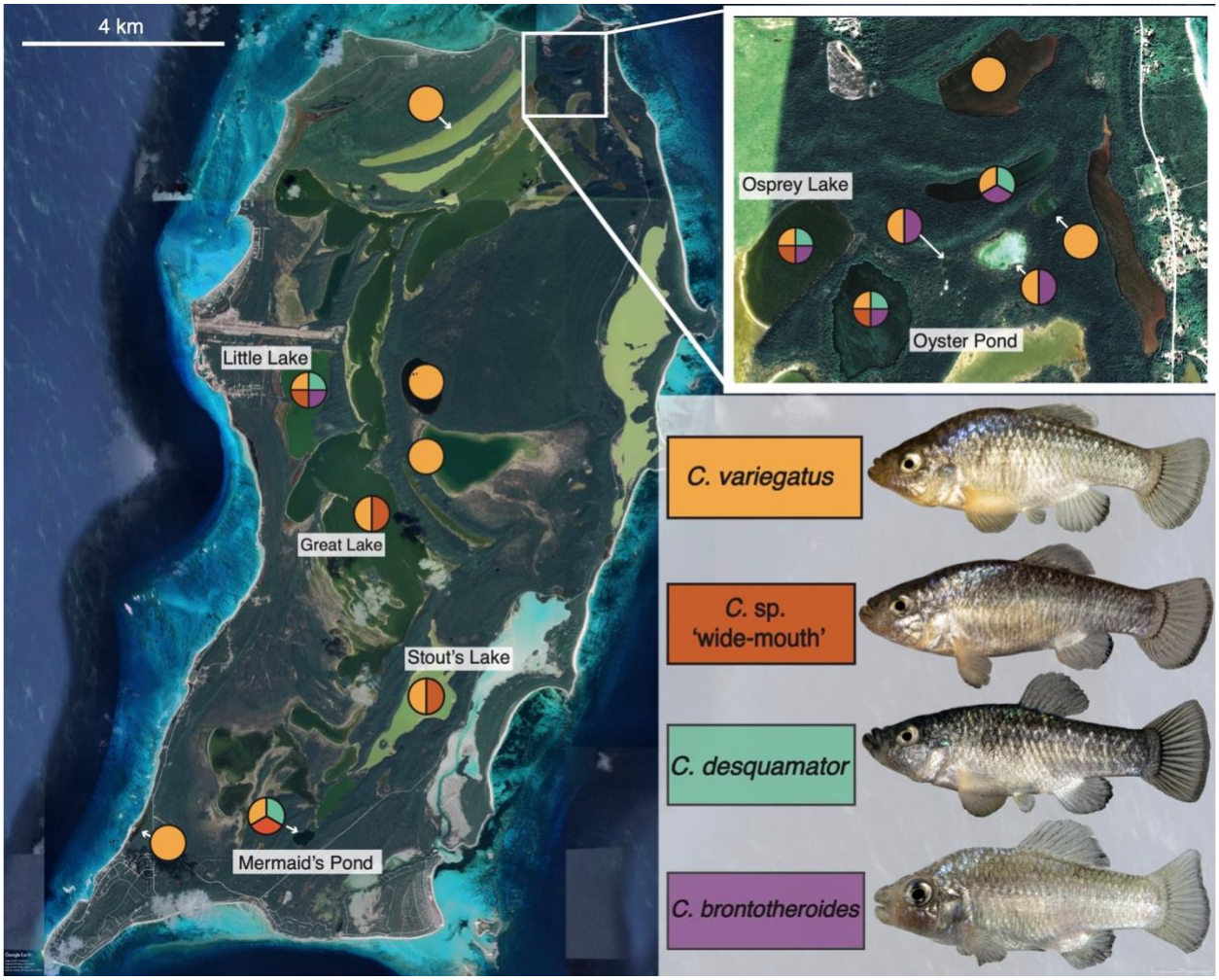


Figure 1. The San Salvador Island radiation of pupfish**.** Pie charts indicate the presence of sympatric *Cyprinodon* species in each lake and are color-coded with representative pictures of generalist *C. variegatus* (gold), recently discovered *C.* sp. 'wide-mouth' ecomorph (red-orange) with intermediate jaws, scale-eater *C. desquamator* (teal) with the largest oral jaws, and molluscivore *C. brontotheroides* (purple) with characteristic nasal protrusion. Labeled lakes contain all known *C.* sp. 'wide-mouth' populations sampled for this study. Satellite image from Google Earth.

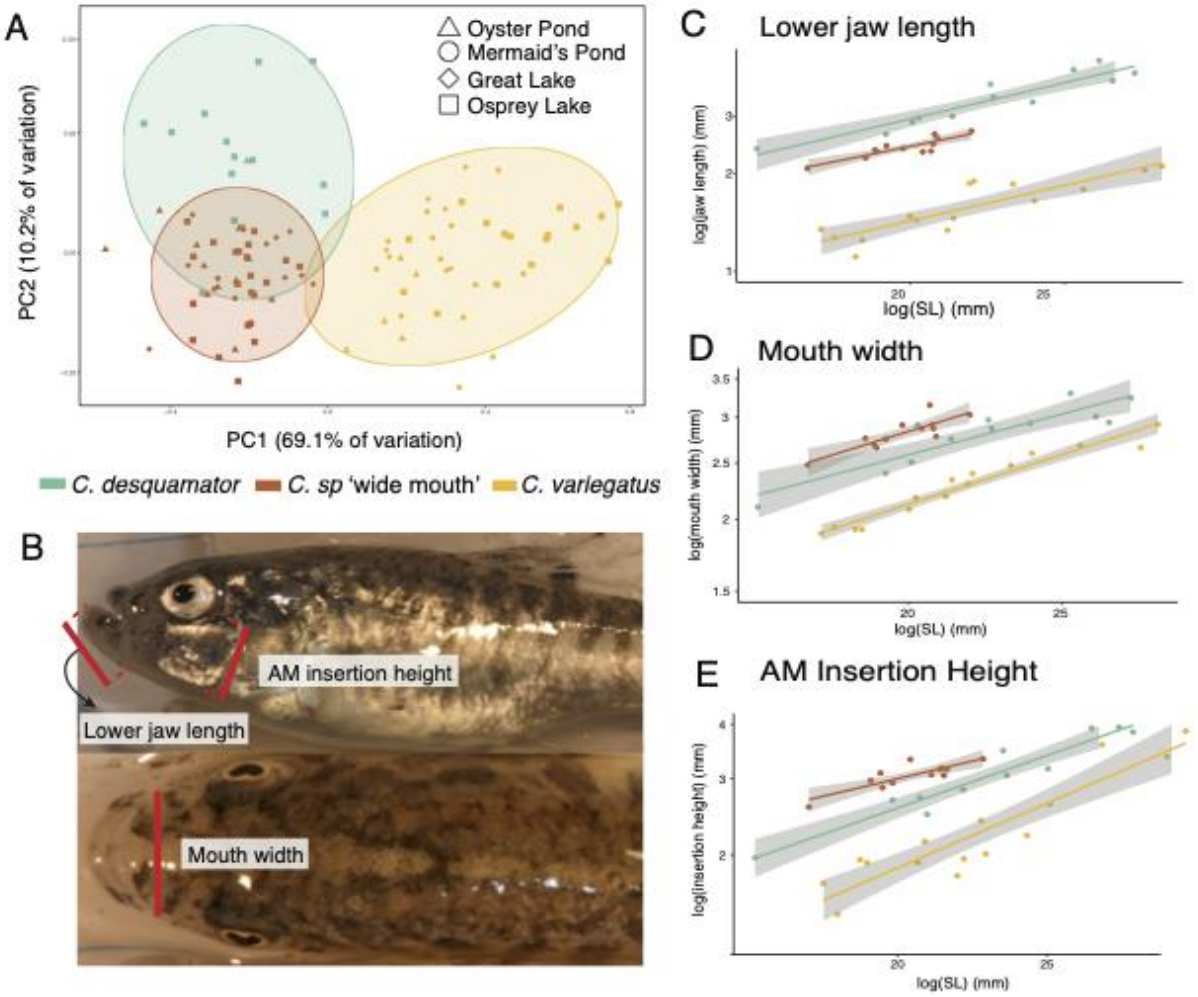


Figure 2. *C.* sp. 'wide-mouth' has distinct morphology within the San Salvador Island adaptive radiation. A) First two principal components of morphological diversity for 8 size-corrected traits and 95% confidence ellipses by species (*C. variegatus* : gold; *C.* sp. 'wide-mouth': red-orange; *C. desquamator*: teal; *C. brontotheriodes* not shown). PC1 is mainly described by lower jaw length and PC2 by adductor mandibulae insertion height, buccal width, and neurocranium width. B) Depictions of the three external measurements that best distinguished *C.* sp. 'wide-mouth' from both *C. desquamator* and *C. variegatus*, measured using digital calipers. C-E) The relationship between standard length (mm) of individuals and their C) lower jaw length, D) buccal cavity width, and E) adductor mandibulae insertion height (AM insertion) across individuals of the three species in Osprey Lake. 95% confidence bands for linear models in gray.

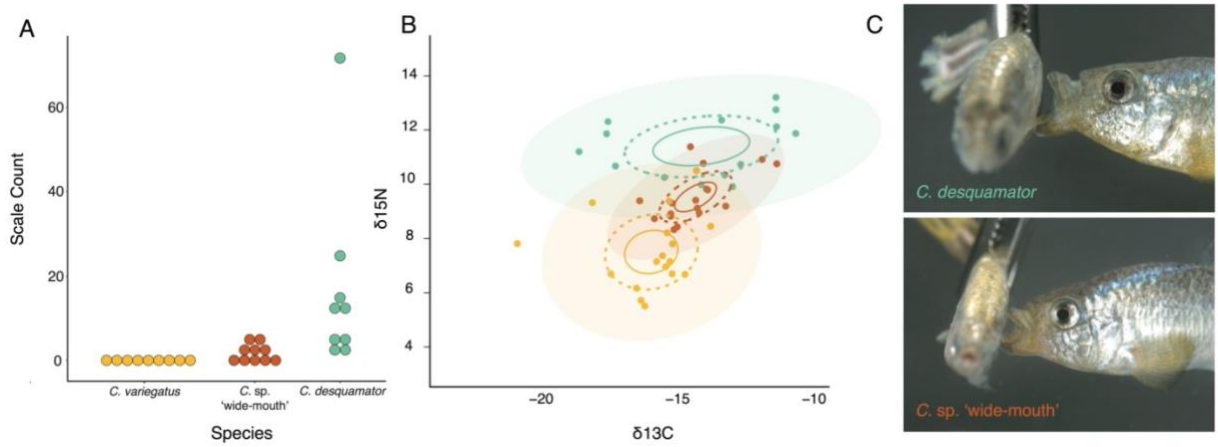


Figure 3. *C.* sp. 'wide-mouth' ingests scales. A) Scale counts from gut content analysis of the hindgut of Osprey pupfish populations (10 individuals per species). B) Relative trophic position (δ15N stable isotope ratio) and dietary carbon source (δ13C stable isotope ratio) with 95% confidence ellipses for generalist and scale-eating species. Solid lines represent 95% confidence intervals around bivariate mean, dotted lines represent standard ellipse areas corrected for sample size (contain 40% of data; SEAc), shaded circles represent ellipse area that contain 95% of the data calculated using the R package SIBER. C) Still images of scale-eating strikes in *C. desquamator* and *C.* sp. 'wide-mouth' filmed at 1100 fps on a Phantom VEO 440S camera.

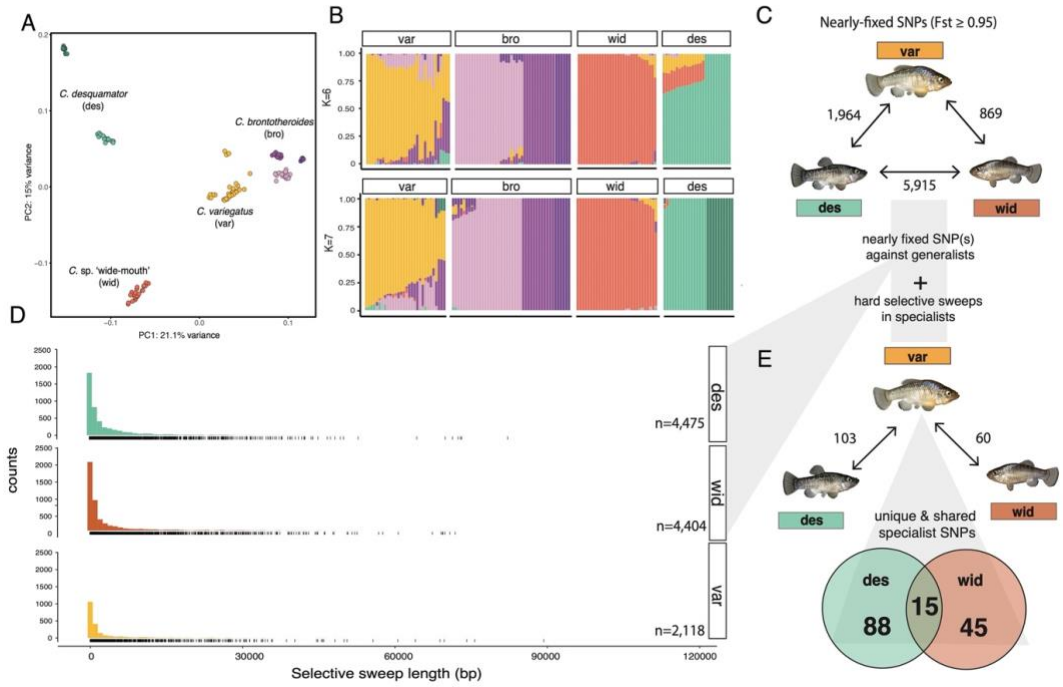


Figure 4. Patterns of selection and genetic divergence in specialist genomes**.** A) Principal components analysis of the four focal groups on San Salvador Island based on an LD-pruned subset of genetic variants (78,840 SNPs). B) Ancestry proportions across individuals of the four focal groups. Proportions were inferred from ADMIXTURE analyses with 2 values of K with the highest likelihood on the same LD-pruned dataset in A. C) Selective sweep length distributions across generalist and scale-eating species. Rug plot below each histogram represents the counts of selective sweeps in different length bins. D) The total number of fixed or nearly-fixed SNPs ($F_{st} \geq 0.95$) between each group in Osprey Pond. E) The number of adaptive alleles (fixed or nearly-fixed SNPS [$F_{st} \geq 0.95$] relative to *C. variegatus* and under selection in each population of specialists in Osprey Lake. Venn diagram highlights those adaptive alleles that are unique to each specialist and shared with other specialists.

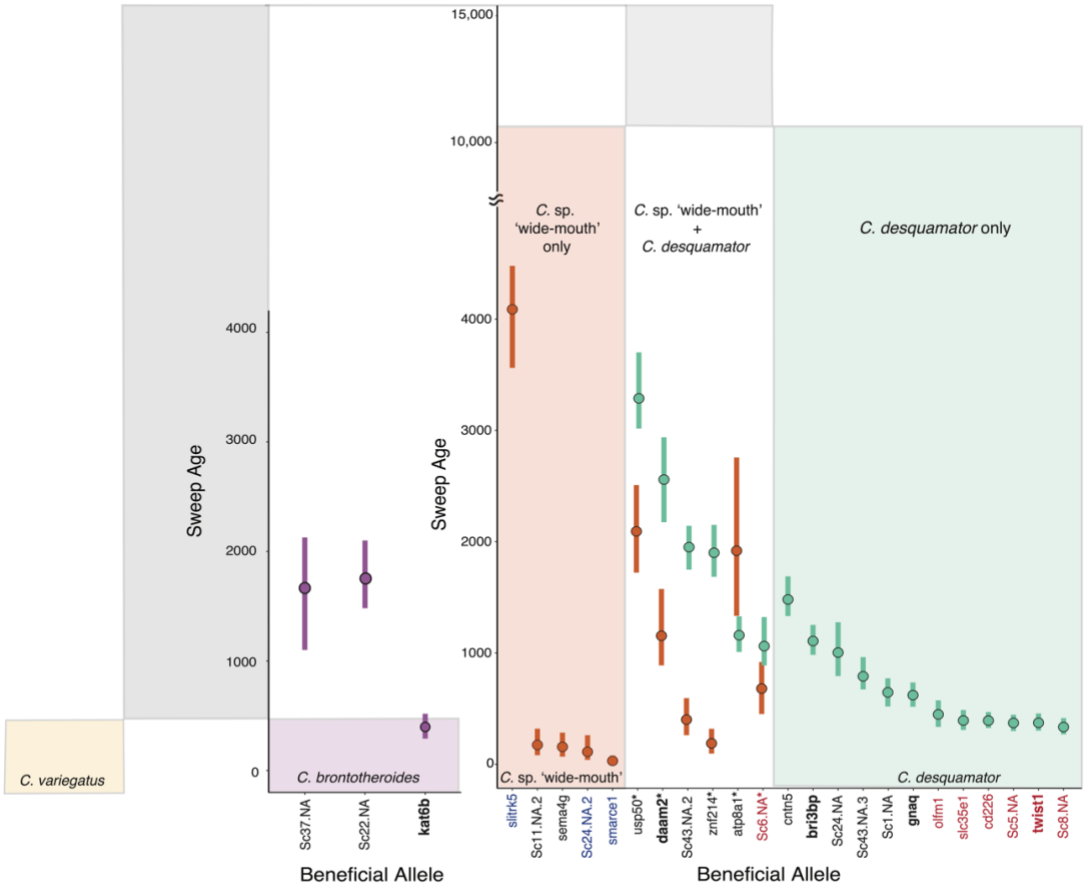


Figure 5. Timing of selection on adaptive alleles in trophic specialists nested within the demographic history of the radiation. The mean and 95% HPD estimates for the timing of selection on sets of fixed or nearly fixed SNPs (named by the gene they are in or within 20-kb of) for the three specialist populations found in sympatry in Osprey Lake (sweeps in *C. variegatus* not shown). The age of each beneficial allele is color coded by the species and the inferred demographic history is displayed in the background for comparison. Gene names highlighted in bold are associated with oral jaw size. Gene names are colored by source of genetic variation (de novo: red; introgressed with outgroup: blue; standing genetic variation: black). Gene names with asterisk indicate those inferred as introgressed between *C. desquamator* and *C.* sp 'wide-mouth'.

5.7. Supplemental Materials

5.6.1. Supplemental Methods

5.6.1.1. Sampling

C. sp. 'wide-mouth' individuals were collected on San Salvador Island in the Bahamas using hand and seine nets between 2017 and 2018. 'Wide-mouth' individuals were collected from 5 lakes in which taxa occurs in sympatry with just *C. variegatus* (Great Lake, and Stout's Lake) or with generalist and *C. desquamator* (Oyster Pond, Osprey Lake, and Mermaid's Pond,; Figure1). In Oyster Pond and Little Lake, *C. desquamator* is extremely rare (only a single Oyster Pond individual grouped genetically and morphologically with *C. desquamator*), so Osprey Lake appears to be the only lake where all three species and the recently discovered 'wide-mouth' coexist at appreciable frequencies in sympatry. 'Wide-mouth' individuals have also been collected in Little Lake in the past (Bruce Turner, pers. obs. 1990s).

Fishes were euthanized in an overdose of buffered MS-222 (Finquel, Inc.) following approved protocols from 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.

5.6.1.2. Ecological and morphological characterization of 'wide-mouth' scale-eater

Multiple individuals from 3 lake populations (Osprey Lake, Great Lake, and Oyster Pond) in which we had measurable specimens of the species (n=84) were measured for 9 traits using digital calipers. Traits were selected for specific connections to foraging performance (Wainwright et al. 2004; Patricia Hernandez et al. 2009) and that differed across the three groups in a previous study when externally measured (Martin and Wainwright 2013c). For each specimen we measured (1) standard length from the anterior tip of the premaxilla to the posterior margin on the midline of the caudal peduncle, (2) lower jaw length from the medial point of the dentary on the lower jaw to the distal point of rotation on the quadrate-articular joint, (3) the width of the buccal from the distance between the proximal sides of the buccal cavity in dorsal view, (4) the height of the adductor mandibulae insertion from the vertex between the vertical and horizontal arms of the preopercle to the dorsal margin of the hyomandibula, (5) the distance from the insertion of the first ray of the dorsal fin to the insertion of the first ray of the anal fin, (6) the height of the caudal peduncle from the dorsal and ventral procurrent rays on the caudal fin, (7) neurocranium width from the narrowest distance between the eyes from the dorsal view, (8) body width from just behind the posterior margins of each operculum in dorsal view, and (9) diameter of the orbit. Measurements were made by a single observer (EJR). Repeatability of measurements was assessed by remeasuring a random subset of 10 individuals per species twice. Measurements on the full dataset were not taken until repeatability was above 85% per trait (r^2 in a linear regression of both sets of measurements). To remove the effects of body size on trait size, we used the residuals from a linear regression of the log-transformed trait on log transformed standard length. We also compared these morphological measurements between lab-reared F2 and wild caught specimens to rule out strong plasticity in traits of interest.

For a subset of individuals collected from each population (*C. variegatus, C. desquamator,* and 'wide-mouth') in Osprey Lake, diet was characterized from stomach content

analyses (*n*=10 each) and stable isotope analyses (*n*=75; 42 *C. variegatus*, 16 *C. desquamator*, 17 'wide-mouth' individuals). These are two complementary approaches that reflect short-term and long-term snapshots of the dietary differences among these groups. For stomach content analysis, a single observer (EJR) dissected out a 1 cm portion of the gut from the posterior hindgut region and counted the number of scales found. For stable isotope analyses, muscle tissue samples were taken from the caudal peduncle region of each individual immediately after euthanasia in an overdose of MS-222. Muscle samples were individually labeled and dried at 60° C for at least 24 hours. Foil-wrapped tissue samples were then sent to the UC Davis Stable isotope facility and analyzed on a PDZ Europa ANCA-GSL elemental analyzer interfaced to a PDZ Europa 20-20 isotope ratio mass spectrometer to characterize the natural abundances of δ13C and δ15N among individuals. δ13C and δ15N levels were summarized across each of the three species to look at overlap among species by 40% and 95% Standard Area Ellipses and 95% confidence ellipses around bivariate means on isotope biplots (figure 3) of the data using R package SIBER (Jackson et al. 2011). Additionally we assessed whether there were significant differences in δ15N values between species through an ANOVA using aov() function from the base statistics package in R (v 4.1.0).

5.6.1.3. Genomic library preparation

We sequenced 24 'wide-mouth' individuals following protocols used in a previous study (Richards et al. 2020) in which we sequenced genomes from *Cyprinodon variegatus*, *C. desquamator*, and *C. brontotheroides*. Raw reads were mapped from these 24 individuals to a de-novo assembly of *Cyprinodon brontotheroides* reference genome (v 1.0; total sequence length = 1,162,855,435 bp; number of scaffolds = 15,698, scaffold N50 = 32 Mbp) with bwa-mem (v 0.7.12;(Li and Durbin 2011)). 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;(DePristo et al. 2011)) to call and refine our single nucleotide polymorphism (SNP) variant dataset using the program HaplotypeCaller. Variants in 'wide-mouth' individuals were called jointly with the 202 individuals sequenced from a previous study (Richards et al. 2020). We filtered SNPs based on the recommended hard filter criteria (i.e. QD < 2.0; FS > 60; MQRankSum < -12.5; ReadPosRankSum < -8;(DePristo et al. 2011; Marsden et al. 2014)) because we lacked high-quality known variants for these non-model species. After selecting for only individuals from San Salvador Island, variants were additionally filtered to remove SNPs with a minor allele frequency below 0.05, genotype quality below 20, or containing more than 10% missing data across all individuals at the site using vcftools (v.0.1.15;(Danecek et al. 2011b)). 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 6.4 million variants with 7.9x median coverage per individual.

5.6.1.4. Testing for signatures of recent hybrids

We first tested whether these 'wide-mouth' individuals represented recent (e.g. F1/F2) hybrids of *C. variegatus* and *C. desquamator* in the wild. These two young species do reproduce with each other in the lab and rare hybrid spawning events have been observed in the wild (unpublished

data ME St. John). As a first visual assessment, we conducted principal component analysis on *C. desquamator*, generalist, and 'wide-mouth' individuals to look for the genome-wide pattern expected in PCAs in which recent hybrids between two populations are included. To perform the PCA, the genetic 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;(Purcell et al. 2007)), leaving 2.6 million variants. We then ran a principal component analysis using the eigenvectors output by plink's pca function (--pca). The first two principal components were plotted in R (R Core Team 2018 v3.5.0).

As a complementary assessment, we used an admixture model-based approach to estimate the proportion of shared ancestry among individuals in our dataset using ADMIXTURE (v.1.3.0)(Alexander et al. 2009). The number of populations (K) was decided upon using ADMIXTURE's cross-validation method (--cv) across 1-10 population values of K. K = 4 was then chosen using the broken-stick method. Ancestry proportions estimated by ADMIXTURE were plotted in R for the K value with the highest likelihood and the two K values surrounding to explore whether the strong signatures of population structure in Crescent Pond *C. desquamator* individuals was masking hybridization signatures in any of the 'wide-mouth' populations.

Lastly, to discriminate among alternative evolutionary scenarios for the origin of 'wide-mouth' and to estimate divergence time among all four species, we used demographic modeling based on the folded minor allele frequency spectrum (mSFS). The observed mSFS was computed using the SFStools script from D. Marques available on github (https://github.com/marqueda/SFS-scripts). For these demographic model comparisons, we used only individuals from Osprey Lake (*C. desquamator*: n=10; *C. variegatus* : n=12; *C.* sp. 'wide-mouth': n=11; *C. brontotheriodes*: n=13) to avoid additional complex population structure across ponds. We contrasted 28 demographic models of different topologies and gene flow scenarios across the four groups (Figure S5;Table 1 and S2). For each model, the fit to the observed multidimensional mSFS was maximized using the composite-likelihood method fastsimcoal (v2.6.0.3;(Excoffier et al. 2013) with 100,000 coalescent simulations, 40 expectation-maximization cycles, and pooling all entries with less than 20 SNPs). For parameter estimates, we used a wide search range with log-uniform distributions with the range of some priors informed by previous estimates of effective population size from MSMC and the age of the last glacial maximum (~20kya) prior to which the lakes would have been dry on SSI. These ranges were not upper bounded by these specified priors and so the simulations were free to explore parameter space that exceeded the priors.

For each demographic model, we ran 100 independent *fastsimcoal2* runs to determine the parameter estimates with the maximum likelihood. The best fitting demographic model was identified using Akaike information criteria (AIC). To get confidence intervals for the parameter estimates from the best fitting model, we simulated 100 mSFS based on the maximum likelihood parameter estimates from our best fitting model and ran 50 independent runs with this model on each simulated SFS. The parameter point estimates from the run with the highest likelihood (of the 50 independent runs) from each simulated SFS were then used to compute 95 percentile confidence intervals as a measure of uncertainty in the parameter estimates from the observed SFS.

5.6.1.5. Testing for shared adaptive alleles

5.6.1.5.1. Patterns of directional selection

Various demographic histories can shift the distribution of low- and high-frequency derived variants to falsely resemble signatures of hard selective sweeps. In a previous study, we used MSMC analyses to infer histories of population size changes in *C. variegatus*, *C. desquamator* and *C. brontotheriodes* (Richards et al. 2021). In this study, we repeated the same analysis for the 'wide-mouth'. In order to account for demography of the 'wide-mouth' population in downstream analyses, we used the MSMC (v. 1.0.1;(Schiffels and Durbin 2014) to infer historical effective population size (N_e) changes in the 'wide-mouth'. We ran MSMC on unphased GATK-called genotypes separately for a single individual 'wide-mouth' from Osprey Lake with 16x mean coverage across its genome (Figure S4). As recommended in the MSMC documentation, we masked 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. (Nadachowska-Brzyska et al. 2016). To scale the output of MSMC to real time and effective population sizes, we used a one-year generation time (Martin et al. 2016) and the estimated spontaneous mutation rate of 1.56×10^{-8} estimated from high coverage sequencing of two independent pedigreed parent-offspring crosses of San Salvador Island species from a previous study (Richards et al. 2021).

Across all four populations in Osprey Lake, we looked for regions that appeared to be under strong divergent selection in the form of a hard selective sweep from the site frequency spectrum using SweeD (v.3.3.4;(Pavlidis et al. 2013)). 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 (Garrigan and Geneva 2014), controlling for the impact of the inferred population size decreases over time for each population from MSMC runs mentioned above (Figure S4; Table S3). 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 (Table S3). We compared selective sweeps across *C. variegatus, C. desquamator, C. brontotheriodes,* and 'wide-mouth' to look for shared and unique selective sweeps among the four groups.

5.6.1.5.2. Characterization of candidate adaptive alleles

We searched for candidate adaptive alleles underlying divergent traits among the species by overlapping selective sweeps regions with regions of high genetic divergence based on fixed or nearly fixed SNPs between groups. We chose to look at nearly fixed SNPs over only fixed variation to accommodate the ongoing geneflow occurring between these young species. We took two approaches to finding fixed or nearly-fixed variants: 1) a fixed threshold of $F_{st} \geq 0.95$ across all comparisons and 2) and threshold of the 99.9th percentile of F_{st}, which varied among comparisons (range of F_{st}: 0.73-0.83).

We made the following pairwise comparisons for F_{st} calculations including a) between *C. variegatus* and each of the specialists, b) each specialist against all other groups and c) shared between two or three specialists against *C. variegatus*. For nearly fixed variants with the $F_{st} \geq$

0.95 threshold, we looked for putative function of the candidate adaptive alleles by looking at gene annotations of any gene the variant was in or near (within 20-kb of the gene, which is within the 50-kb LD decay estimate). We used available gene annotations from model organisms of mice and zebrafish from MGI, ZFIN, and 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 (Balsa-Canto et al. 2016) .

For the 99.9th percentile F_{st} variants: We performed gene ontology (GO) enrichment analyses for genes near candidate adaptive variants using ShinyGo (v.0.51;(Ge and Jung 2018)). In the *C. brontotheroides* reference genome annotations (described in *de novo* genome assembly and annotation section), gene symbols largely match human gene symbols. Thus, the best matching species when we searched for enrichment across biological process ontologies curated mostly human gene functions.

5.6.1.5.3. Timing of selection on candidate adaptive alleles

Lastly, we also determined the relative age of candidate adaptive alleles by generating estimates of coalescent times using the R package starTMRCA (v0.6-1;(Smith et al. 2018)). For each candidate adaptive allele that was unique to the three specialists and the 16 shared alleles between *C. desquamator* and 'wide-mouth', a 1-Mb window surrounding the variant was extracted into separate vcfs for individuals within each species. The program requires complete genotype data so we first filtered out any individuals with more than 10% missing data (1 *C. brontotheriodes* and 2 'wide-mouth' individuals). With Tassel5 (Bradbury et al. 2007), we then used the LD KKNI command to infer missing sites based on LD for remaining individuals with less than 10% missing data. Subsequently we removed the small number of sites with any missing data across individuals within each population. Since several adaptive alleles fall within the same 1-Mb window, we chose a single adaptive allele at random within each 1-Mb window as the focal allele to estimate timing for. This lead to 6 time estimates for the 16 shared adaptive alleles between *C. desquamator* and *C.* sp. 'wide-mouth', 5 time estimates for the unique adaptive alleles to *C.* sp. 'wide-mouth', X time estimates for the unique adaptive alleles to *C. desquamator* and 3 time estimates for the unique adaptive alleles to *C. brontotheroides*.

These sets of variants were then analyzed in starTMRCA with the mutation rate of $1.56x10^{-8}$ substitutions per base pair, and a recombination rate of $3.11x10^{-8}$ (from genome-wide recombination estimate for stickleback; (Roesti et al. 2013)). Each analysis was run three times per focal adaptive allele and all runs were checked for convergence between and within runs. Most runs rapidly converged within the initial 6000 steps, but 5 runs did not converge after an additional 4000 steps and were discarded from further analysis.

5.6.1.6. Characterization of introgression in C. sp 'wide-mouth' genome

5.6.1.6.1. Introgression with C. desquamator in candidate adaptive allele regions

To assess whether shared adaptive alleles between *C.* sp 'wide-mouth' and *C. desquamator* may be due to introgession between the two populations, we calculated Relative Node Depth (RND) statistics (Feder et al. 2005) in 10-kb sliding windows across the genome. RND statistics measure the relative node depth of two taxa compared to an outgroup and is calculated as the quotient of D_{xy} between two species to the average distance from each to an outgroup. This relative amount of divergence compared to an outgroup allows the statistic to be robust to mutation rate variation across the genome, in which low neutral mutation rate regions might be mistaken for a locus that has experienced a recent introgression event. RND was calculated using a custom script that used R package PopGenome (v2.7.5;(Pfeifer et al. 2014)) first calculate D_{xy} values calculated for *C. desquamator*, *C.* sp. 'wide-mouth' and *C. artifrons* in 10-kb windows across the genome and then RND for each scaffold that contained a candidate adaptive variant that was unique or shared between *C. desquamator* and *C.* sp wide-mouth. Significance of RND values was evaluated using simulations with no migration using ms-move (Garrigan and Geneva 2014). We used estimates of changes in effective population size and divergence times for each population from our *fastsimcoal2* analyses and ms-move to simulate neutral genetic divergence between two isolated populations. The threshold for significant introgression regions was determined by simulating genetic sequences under a coalescent model with no gene flow, consisting of 150,000 10-kb windows each containing the mean number of alleles observed in our dataset and running these simulated sequences through the same custom PopGenome script (RND_PopGenome_Artout_null.R) to calculate RND values. Empirical windows were considered candidates for introgression if the RND statistic was in the bottom 5th percentile of simulated values (RND < 0.22; Table S10). To assess whether the sympatric coexistence of four incompletely reproductively isolated species in Osprey Lake has lead to introgression across specialist and generalist population boundaries as well, we assessed whether RND test statistics calculated between *C. variegatus* and *C. desquamator* in Osprey Lake were similarly detected as introgressed in these regions. None of these regions appeared to be introgressed with the generalist *C. variegatus* (Table S10), suggesting the introgressed signature between *C. desquamator* and *C.* sp. 'wide-mouth' is not likely due to neutral introgression likely occurring between all species in the lake due to their incomplete reproductive isolation.

5.6.1.6.2. Characterizing introgression with outgroup Caribbean generalist populations in candidate adaptive allele regions

We also investigated whether any of these candidate adaptive alleles in *C. desquamator* and *C.* sp 'wide-mouth' fell in regions that appear to be introgressed from an outgroup generalist population (similar to analyses done in our previous study (Richards et al. 2021)). First, we extracted the genomes of individuals assigned to our 5 focal outgroup generalist populations in which we had 8 or more individuals sequenced in that previous study. We filtered variants across these 5 populations and our San Salvador Island samples down to those that has a quality of 20, and no more than 10% missing in any population for a dataset of 10,876,882 SNPs across 10 populations (including the four SSI species and the outgroup *C. artifrons*) and the 26 scaffolds that candidate adaptive alleles were found on. To detect introgression involving outgroup populations we used the differential test of introgression, *df*-statistic (Pfeifer and Kapan 2019),

which is designed to look for signatures of introgression across sliding genomic windows (Martin et al. 2015b). The *df*-statistic is 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 (Martin et al. 2015b). We used 2 individuals of *C. artifrons* from Cancun, Mexico as our outgroup population for this test, which forms the deepest divergence event with *C. variegatus* within the *Cyprinodon* clade (Echelle et al. 2005), and focused on introgression between *C. desquamator* or *C. sp.* 'wide-mouth' and outgroup Caribbean generalist populations that was not shared with *C. brontotheroides*. Based on the tree (((P1,P2),P3), *C. artifrons*), the *df* statistic was calculated for the combinations of populations in which the focal population (P2) was either the scale-eater or the molluscivore, the other specialist population was the sister group (P1), and P3 was one of the Caribbean outgroup populations sequenced in a previous study (Fort Fisher, North Carolina Coast;Isla Margarita, Venezuela; Lagunas Bavoros, Dominican Republic, and New Providence Island, Bahamas. *Fd*-statistics were calculated from 10-kb sliding windows using R package PopGenome. Empirical windows were considered candidates for introgression if the *df*-statistic was in the top 95th percentile of simulated values (*df* > 0.55).

5.6.1.6.3. Characterization of admixture history on San Salvador Island with outgroup generalist populations

The results from sliding window tests for introgression surrounding candidate adaptive alleles for *C.* sp. 'wide-mouth' indicated that this cryptic species may be the product of admixture between SSI species and outgroup generalist lineage related to our North Carolina coast population (Table S8 and S10). To further explore this, we created genome-wide admixture graphs using the program qpgraph (Patterson et al. 2012) in R package admixtools2 (v2.0.0; https://uqrmaie1.github.io/admixtools/). Qpgraph can accommodate many populations and admixture events and creates an admixture graph that summarizes all pairwise f_2- and f_3-statistics across included populations and estimates admixture proportions and drift weights between populations. Admixture graphs model the number of admixture events specified by the user and evaluates the fit of the observed *f*-statistic values between populations to those expected from this model to produce a negative log likelihood score that measures the total amount by which the modeled f-statistics failed to match the observed values (where larger likelihood values indicate poorer fit).

Normally in qpgraph analyses, the user manually specifies the topology of the model and the program then solves for the optimal values of the parameters. The recent R package admixtools2 also includes the function find_graphs() that generates and evaluates admixture graphs across user-defined number of iterations to find the a range of feasible and best fitting graph topologies for a set of *f*-statistics. Using the four Osprey Lake populations on San Salvador Island (*C. variegatus, C. brontotheroides, C. desquamator* and *C.* sp. 'wide-mouth', and the 5 Caribbean generalist populations and *C. artifrons* outrgroup mentioned in the above two sections, we ran find_graphs() 10 times to model 1-10 admixture events.

The input data consisted of 20,381 SNPs that remained after filtering the dataset that included the outgroup generalist populations used in the sliding-window tests above and an additional filters for heterozygous sites that had an allele balance between 0.3 and 0.7, minimum 3x depth of coverage, and quality filter of 30. The final dataset used as input for qpgraph included 20,381 SNPs across 10 populations and each find_graphs() run was allowed to explore

topology space for 1000 iterations. The topology with the best fit was extracted from these runs and assigned to be the admixture graph for that admixture event. The model and admixture graph with the lowest likelihood of all admixture events tested (LnL: range 2.16-16.23 across all 10 models), included 7 admixture events. One of these admixture events depicted admixture between *C. desquamator* and a lineage related to the North Carolina coast generalist population resulting in the C. sp 'wide-mouth' population. Other admixture events in this admixture graph, such as an admixture event at the base of the SSI radiation and a separate admixture event involving *C. desquamator* and an outgroup generalist, supported signatures inferred using independent program TreeMix from a previous study(Richards et al. 2021).

The inference of *C.* sp. 'wide-mouth' as admixed between *C. desquamator* and a distantly related generalist species appeared in the 8 admixture graphs modeling 3-10 admixtures events, suggesting it's a strong signature in the dataset. However, the outgroup Caribbean generalist population that acted as one of the donors varied across different runs from being sister to North Carolina coast population to falling somewhere along the branch that divides the Venezuela population from the rest of the Caribbean populations. This may suggest that we are missing the true donor population in our sample. To further support this signature of *C.* sp. 'wide-mouth' experiencing introgression following a potential secondary contact event between San Salvador Island and an outgroup generalist population, we used f_4-statistics to look for genome-wide signatures of differential introgression *between C. desquamator* and *C.* sp 'wide-mouth'.

Differential signatures of introgression between the two would likely indicate introgression that *C.* sp. 'wide-mouth' contains in its genome but is not shared with *C. desquamator* and therefore likely not a signature of shared introgression in the ancestor of the two scale-eating populations on SSI. Genome-wide f_4 statistics were calculated using the f4() function in admixtools2 which calculates f_4-statistics, Z scores and p-values for the specified populations. Following the tree ((P1,P2),(P3,P4)), *C. desquamator* was used as sister species P1, *C.* sp. 'wide-mouth' as focal species P2, the 5 generalist populations from Fort Fisher, North Carolina Coast; Isla Margarita, Venezuela; Lagunas Bavoros, Dominican Republic, Rum Cay, Bahamams and New Providence Island, Bahamas were rotated in as the potential donor population P3 and *C. artifrons* used as population P4 (Table S11). The f_4-statistic that included North Carolina coast population had highest Z-score and indicated a significant amount of gene flow (p-value=0.004;Table S11). However, tests that included Rum Cay and New Providence Island also resulted in f_4-statistics with marginally significant p-values (p-value=0.056 and 0.067 respectively; Table S11), perhaps mirroring that admixture occurred with an unsampled generalist population somewhere North of SSI and most closely related to North Carolina population.

5.6.2. Results

5.6.2.1. 'Wide-mouth' ecomorph is ecologically intermediate and morphologically distinct

Morphological traits were heritable in a common garden laboratory environment after one generation: lab-reared *C.* sp. 'wide-mouth' displayed significantly larger buccal width than *C. desquamator* (t-test; $P = 0.003$) and maintained their characteristic intermediate jaw lengths (ANOVA; $P = 0.03$, figure S3). There was also some evidence of phenotypic plasticity in both *C.*

desquamator and *C.* sp 'wide-mouth' compared to wild individuals. Two of the focal three traits that distinguished wild populations of *C.* sp. 'wide-mouth' from the other pupfish species, lower jaw length and buccal width, remain significantly distinct in the lab-reared populations (based on non-overlapping 95% confidence intervals around mean residuals of each species). However, one trait that distinguishes *C.* sp. 'wide-mouth' in the wild appears to be quite plastic under laboratory rearing conditions; there were no significant differences in mean adductor mandibulae height residuals among species in laboratory environments. This is interesting because adductor mandibulae height is a proxy measurement for the cross-sectional width of the adductor muscle directly beneath the skin. The plasticity observed in this proxy may be due to muscle plasticity. In the lab-reared colonies, all species are reared on the same diet, such that the scale-eating populations do not scale-feed as often as they would in the wild and thus lab-reared populations of scale-eaters may not develop the notably increased adductor muscle mass. Interestingly, the distinct differences in adductor mandibulae insertion height among wild populations disappeared in the lab-reared populations (figure S3), indicating phenotypic plasticity associated with the common diet of pellet foods and lack of scale-eating attacks in the lab.

We found that 'wide-mouth' ingest scales, but at a significantly lower frequency than *C. desquamator* (Wilcoxon Rank Sum test, $P = 0.004$; figure 3A). We did not detect any scales in *C. variegatus* guts (figure 3A). Detritus made up the only other gut contents present besides scales in all the specimen of *C.* sp 'wide-mouth' from Osprey Lake that we dissected. This was similar for the other scale-eater *C. desquamator* individuals from Osprey Lake, in which we also found only scales and detritus in their gut. For the generalist species *C. variegatus*, their guts contained only detritus except for a single individual in which we found a snail shell. The low diversity in gut contents among species may be due to the relatively small sample size of individuals we had that we could perform gut content analysis on for this study (n=10 for each species). However, the predominance of detritus in the gut contents of Osprey Lake populations of all three species is not surprising given the thick (<1m) bottom layer of mud and flocculent in this lake(Edwards 2001). Additionally, detritus made the largest percentage (49-71%) of San Salvador Island pupfish diet in other lakes across the island as well. A previous study that conducted gut content analyses with 2-4X larger sample sizes also found that the majority of *C. variegatus, C. brontotheroides,* and *C. desquamator*'s diet is detritus. Except for one population of *C. desquamator* in Crescent Pond where scales made up 50% of their gut contents and detritus only made up 30%. Therefore, the ecological intermediacy of *C.* sp 'wide-mouth' is supported by their intermediate ratio of scales to detritus, which is the predominant axis of dietary divergence between *C. variegatus* and *C. desquamator* in this lake. The lack of other dietary items in the gut contents of *C.* sp 'wide-mouth' further supports our arguments that their ecological divergence is along the same specialist axis of scale-eater *C. desquamator* and that they share little ecological overlap with the other molluscivore specialist *C. brontotheroides.*

5.6.2.3. 'Wide-mouth' did not result from hybridization between C. variegatus and C. desquamator

Several lines of genomic evidence support the 'wide-mouth' ecomorph as a distinct species. First, 'wide-mouth' individuals do not occupy an intermediate position between *C. desquamator* and *C. variegatus* along either of the first two principal components that represent the major axes

of genetic variation in the radiation (figure S4A). In the ADMIXTURE analysis, 'wide-mouth' individuals share more ancestry with each other than with either *C. desquamator* or *C. variegatus* under the two most likely K values (K=6 and 7; figure S4B&5). Lastly, significant positive f_3-statistics and the non-significant f_4-statistics support our inference that the 'wide-mouth' is not a recent admixed population of any pairwise combinations of the *C. brontotheriodes, C. variegatus,* or *C. desquamator* populations (table S1).

5.6.2.4. C. sp. 'wide-mouth' did not result from hybridization between C. variegatus and C. desquamator

We compared 28 different demographic models of divergence and gene flow on San Salvador Island (table S2) that represented all possible topologies among the four species and explored zero, early, and current gene flow scenarios. In the best supported model, 'wide-mouth' was sister to *C. desquamator* with current gene flow and a divergence time estimate of 11,658 years ago (95 CI: 8,257-20,113 years; figure 4; table 1 and S2). This model indicates that the ancestor to 'wide-mouth' and *C. desquamator* populations first diverged from *C. variegatus* 15,390 years ago (95% CI: 10,722-23,927 years; figure 4B). This estimate of the origin of the radiation overlaps well with geological age estimates based on the filling of hypersaline lakes on San Salvador Island after the end of the last glacial maximum period (~6-19K years ago; (Turner et al. 2008; Clark et al. 2009)). *C. variegatus* and *C. brontotheriodes* populations in Osprey Lake diverged recently about 462 years ago (95% CI: 411-1,121 years; figure 4B).

5.6.2.5. Shared and unique adaptive alleles in C. sp. 'wide-mouth' and desquamator

We discovered that the two scale-eating populations shared a set of the same adaptive alleles not found in *C. brontotheriodes* or *C. variegatus* (Figure 5D). This consisted of 15 alleles in 6 shared hard selective sweeps in *C. desquamator* and 'wide-mouth': 10 SNPs were in unannotated regions, two were in the introns of the gene *daam2,* and three were in regulatory regions of the genes *usp50, atp8a1,* and *znf214* (one variant each).

Shared adaptive alleles in the gene *daam2,* a wnt signaling regulator, are intriguing because knockdown of this gene causes abnormal snout morphology, osteoporosis, and changes in insulin and alkaline phosphate levels in mice (Dickinson et al. 2016), and abnormal cranial and skeletal development in zebrafish (Kida et al. 2007). Craniofacial morphology is one of the rapidly diversifying traits in this system, suggesting that divergence in *daam2* may play a role in the shared craniofacial divergence of the two scale-eating species. Similarly, *usp50* functions in protein metabolism and deubiquination (Lee et al. 2017) and may play a role in shared metabolic adaptations to the higher-protein content of a scale-eating or snail-eating diet (also note microbiome divergence of *C. desquamator* with increased prevalence of collagenase-digesting bacteria when reared in a common garden:(Heras and Martin 2021)). *atp8a1* is an ion transporter (Soupene 2008), *znf214* is a transcription binding factor, and the ten unannotated variants were not associated with lower jaw size variation in a previous genome-wide association study of the radiation (Richards et al. 2021).

Despite the divergent craniofacial features of the 'wide-mouth', none of the adaptive alleles unique to the 'wide-mouth' appear to be in or near genes annotated for craniofacial phenotypes in model organisms. In *C. desquamator,* three of the 13 sets of unique adaptive alleles are in or near genes annotated for craniofacial phenotypes: a *de novo* non-synonymous

coding substitution in the gene *twist1*, several putative regulatory variants near the gene *gnaq*, and 8 variants in and near the gene *bri3bp*, which is located inside a QTL region for cranial height in pupfish (St. John et al. 2021). In *C. brontotheriodes*, there is also a candidate craniofacial adaptive allele: a non-synonymous coding substitution in the gene *kat6b*, which is associated with abnormal craniofacial morphologies, including shorter mandibles in mice (Thomas et al. 2000; Kraft et al. 2011) and Ohdo Syndrome and bulbous noses in humans (Clayton-Smith et al. 2011).

This pattern of unique alleles relevant to craniofacial phenotypes in *C. brontotheriodes* and *C. desquamator*, but not 'wide-mouth', holds even if we lower the threshold to the top 1 percentile of F_{st} between specialists and generalist (see supplemental for more details). 'Roof of mouth development', 'bone development' and 'skeletal system development' are in the top 20 most significantly enriched GO terms for *desquamator* candidate adaptive alleles, with 6 significantly enriched terms relevant to cranial and skeletal development (out of 165 enriched terms total). Similarly, 'embryonic skeletal system development' and 'skeletal system development' were significantly enriched terms for *brontotheriodes* candidate adaptive SNP alleles (out of 8 enriched terms total). However, 'wide-mouth' adaptive alleles at this lower F_{st} threshold were not significantly enriched for any GO terms related to craniofacial or skeletal morphology (n=52 terms; Figure S7).

5.6.2.6. Substantial history of admixture in C. sp 'wide-mouth'

To further investigate the signatures of adaptive introgression in this species, we modeled admixture across San Salvador Island and several key generalist populations in the Caribbean with f_3- and f_4-statistics (table S11;figure S10). Previous signatures of admixture involving the base of SSI radiation and outgroup generalist populations (Richards et al. 2021) were detected again in our admixture graph, alongside a new signature of admixture that appears to be secondary contact between a pupfish lineage most closely related to North Carolina in our dataset and *C.* sp 'wide-mouth' (table S11; figure 5 & S10).

5.6.2.7. Shared signatures of selection across the three specialists in the radiation

We found that a higher percentage of the genome was under positive selection in all specialist populations compared to the *C. variegatus* population (figure 5A). This pattern might be expected given the divergent selection pressures the specialists face to adapt to different trophic niches. Of all the specialists, *C. brontotheriodes* exhibited both the most selective sweeps and the longest selective sweeps (figure 5A). This suggests that *C. brontotheriodes* have undergone selection most recently among all the species in Osprey Lake and is supported by the recent divergence time between the *C. brontotheriodes* and *C. variegatus* populations (figure 4D). The shared selective sweeps were shorter than any of the selective sweeps unique to each of the specialists, suggesting that selection for these shared regions was not the most recent or strongest in any of the specialists (figure S11C).

Additionally, all specialists, including the two scale-eating species, were more genetically diverged from each other than to *C. variegatus* species based on the number of SNPs fixed or nearly-fixed between them (figure 5B). This pattern of stronger genetic divergence between specialists also held across fixed SNPs, the top 1% of F_{st}, and genome-wide average F_{st} (table

S4). Next, we identified a set of candidate adaptive alleles for each specialist by filtering for SNPs that were fixed or nearly fixed (F_{st}, ≥ 0.95) relative to *C. variegatus* and found within a hard selective sweep in the focal specialist. In this set of candidate adaptive alleles, *C. brontotheriodes* had the fewest adaptive alleles while *C. desquamator* had the most (figure 5B). All three specialists, including the two scale-eating species, were more genetically diverged from each other than to *C. variegatus* based on the number of SNPs fixed or nearly-fixed between them across multiple F_{st} thresholds (figure S11B; table S4).

5.7.2. Supplemental Discussion

The hallmark of adaptive radiation is a rapid burst of diversification which is predicted by theory to slow down over time as niche subdivision increases (Martin and Richards 2019). An alternative possibility is that radiations can be self-propagating and that the diversity generated within the first stages of radiation helps beget further diversity (Whittaker 1977). This could happen through exploitation of new trophic levels created by new species or physical alterations of the environment by new species that may create additional opportunities for speciation (reviewed in (Stroud and Losos 2016; Martin and Richards 2019)). The diversity begets diversity hypothesis can also be visualized as the exploration of a complex multi-peaked fitness landscape; as species in the radiation colonize new peaks, this provides access to additional neighboring fitness peaks to fuel rapid radiation. At the microevolutionary level, shared adaptive variants can also help other populations colonize new or ecologically similar areas of the fitness landscape. However, the genetic basis and microevolutionary processes underlying major ecological transitions are still poorly understood in nature.

5.7.2.1. An adaptive walk underlies the major ecological transition from generalist to scale-eating specialist

One of the foundational models of adaptation is that it proceeds in 'adaptive walks' towards fitness optima that involve the sequential fixation of adaptive alleles that move a population in the phenotypic direction of the local optimum (Maynard Smith 1970; Gillespie 1984; Kauffman and Levin 1987; Orr 1998, 2005). The distinct timing of selection across different adaptive alleles suggests that the ecological transition from generalist to novel scale-eating specialist involved such an adaptive walk rather than a sudden burst of concurrent selection events after some major environmental shift. These distinct, multiple bouts of selection could be caused by mutation-limited (i.e. waiting for new beneficial alleles; (Bell 2013; Lindsey et al. 2013; Rousselle et al. 2020)) or mutation-order processes (i.e. epistatic interactions; (Mani and Clarke 1990; Schluter 2009; Good et al. 2017).

We found that all of the shared adaptive alleles between the scale-eating populations occurred at low frequency in generalist populations on other Caribbean islands in a previous study ((Richards et al. 2021); Table S7). For example, three copies of the shared *usp50* adaptive allele were found outside of San Salvador Island (Table S7). This indicates that initial bouts of selection occurred on available standing genetic variation and that staggered timing of hard selective sweeps in each trophic specialist most likely reflects mutation-order processes in which selection on a beneficial allele was contingent on prior fixation of other adaptive alleles in each specialists' genetic background. However, several adaptive alleles originated from introgression

or de novo mutations found only on San Salvador Island (Table S7), so part of the adaptive walk may have also been mutation-limited.

5.7.2.1. Cryptic scale-eating species reveals features of the adaptive walk towards scale-eating specialization

Although the two scale-eating species shared a set of adaptive alleles and a recent common ancestor, 5 of 6 shared adaptive regions swept at significantly different times between the two species. This difference in timing may result from several different scenarios: 1) independent adaptive walks to the same scale-eating niche, 2) independent adaptive walks to different scale-eating niches or 3) the adaptive walk of one population depends on the adaptive walk of the other population. We explore each of these scenarios in turn.

First, the difference in timing of selection on the same shared adaptive alleles could indicate independent adaptive walks to the same scale-eating peak that occurred at different times and/or by slightly different routes. *C. desquamator* and 'wide-mouth' populations have predominantly abutting ranges, with only a small amount of geographic overlap on San Salvador Island in four lakes, Osprey, Oyster, Little Lake, and Mermaid's Pond (Figure 1). If this current distribution is representative of their historical ranges and the two lineages began diverging about 11,000 years ago, it is possible that the adaptive walks took place in different lakes and were largely independent of one another. The differences in timing on the shared adaptive alleles and the presence of unique alleles observed in each population are reminiscent of mutation-order speciation (Schluter 2009; Schluter and Conte 2009). In mutation-order speciation the same alleles are favored in populations that are adapting to similar environments, yet by chance and/or epistatic interactions with different genetic backgrounds, similar adaptive alleles fix in just one population. At least one set of alleles near the gene *slitrk5* are unique to the 'wide-mouth' scale-eater and fixed well before selection occurred on the shared adaptive variants between the two scale-eater populations (Figure 5). Epistatic interactions of these *slitrk5* alleles may have prevented the 'wide-mouth' from fixing additional segregating alleles that are uniquely fixed in *C. desquamator*.

Second, the difference in timing of selection may have occurred because the two scale-eating populations are adapting to two different scale-eating niches. The intermediate diet, distinctly sized morphological traits, and smaller body size in the 'wide-mouth' scale-eater compared to *C. desquamator* may indicate the 'wide-mouth' scale-eater is adapting to a different scale-eating niche. While the two scale-eating populations do share some overlap in adaptive alleles and selective sweeps, the majority of selective sweeps are unique to each species (Figure 5 & 6), including neurogenesis, brain, and nervous system development (*slitrk5, sema4g, and smarce1*), whereas unique adaptive alleles in *desquamator* include craniofacial development annotations (*olfm1, gnaq, twist1*) as well. This difference in gene annotations can also be seen at the broader level of regions of the genome under selection (Figure S7). Therefore, the differences in timing at shared adaptive alleles might reflect differences in relative strength of selection due to different selective regimes experienced by the two scale-eating populations. While starTMRCA is fairly robust in its timing estimates to different strengths of the selection (Smith et al. 2018), we cannot rule out that differences in relative timing of selection might partially reflect differences in strength of selection occurring in the two scale-eating populations.

Third, another scenario that could have led to the older timing of selection across shared adaptive alleles in *desquamator* than 'wide mouth' is one in which the adaptive walk of one population depended on adaptive alleles from another. The significantly older timing of selection in *desquamator* on shared adaptive alleles and the low absolute genetic divergence between these two scale-eating population in these regions (Figure 6) suggests that these alleles may have introgressed between the two populations. Given that these shared adaptive alleles appear as standing genetic variation across Caribbean populations, albeit at very low frequency in our sampling, it is likely that the shared alleles were present in the ancestor and were segregating in both populations after they diverged. There are several reasons why adaptive introgression may have been necessary for adaptive divergence of 'wide-mouth' despite the alleles initially segregating in the ancestral population. In line with mutation-order processes (e.g. Mani and Clarke 1990; Schluter 2009; Good et al. 2017), these alleles might not have been adaptive until the right genetic background was present in the 'wide-mouth'. Introgression from *C. desquamator* in which the allele was already swept to high frequency could have raised the frequency of these alleles in 'wide-mouth' and increased the likelihood of fixation for the beneficial alleles (reviewed in (Patwa and Wahl 2008)). This is consistent with previously proposed hypotheses for how diversity begets further diversity in adaptive radiations (Whittaker 1977; Stroud and Losos 2016; Martin and Richards 2019).

5.7.2.3. Shared aspects of genomic landscape of selection among all specialists additionally supports specialization promoting further specialization

Intriguingly we also find evidence of selective sweeps shared across all three specialists despite their divergent adaptations and lack of a shared specialist ancestor. However, there are no commonly shared adaptive alleles fixed against *C. variegatus* in these shared selective sweep regions. Hard selective sweeps indicate a selection scenario in which a single beneficial allele of large effect on a trait is swept to high frequency (reviewed in (Pavlidis and Alachiotis 2017)). The lack of highly divergent alleles in these regions might indicate polygenic selection events underlie the shared signatures across specialists. The broad-scale transition from dietary generalist to dietary specialist may involve polygenic selection with small shifts in the frequency of many alleles. Despite very different expectations about the genetic basis of adaptation, polygenic selection events may have been detected as a hard selective sweep in this study as the two selection types can be challenging to distinguish based solely on patterns of genetic variation (Chevin and Hospital 2008; Höllinger et al. 2019; Thornton 2019; Barghi et al. 2020). Recently developed frameworks provide additional criteria to help distinguish between the two selective regimes (Barghi et al. 2020) but are beyond the scope of this current study.

These shared selective sweeps across all specialists are not the longest selective sweeps in any of the specialist genomes (> 50-kb; Figure 5C), indicating they are not the most recent or strongest selection events. However, these shared selective sweeps appear to be relevant to dietary specialization as these regions are enriched for genes annotated for metabolic processes such as short chain fatty acid and propionate metabolism (Figure 5). Propionates and other short chain fatty acids are common microbiome metabolites (Flint et al. 2012). A recent microbiome study in lab-reared populations found that *C. desquamator* microbiomes appear enriched for *Burkholderiaceae* bacteria that can digest collagen, further supporting an adaptive role for microbiomes in the radiation (Heras and Martin 2021).

We also see pairwise shared selective sweeps between all combinations of two specialists. In contrast to the sweeps shared between all specialists, we see stronger genetic divergence among species in these regions and a proportion of the selective sweeps shared between two specialists do contain fixed or nearly-fixed alleles. However, the specialists that share selective sweeps in these regions are fixed for different adaptive alleles, consistent with previously observed patterns of parallel differential gene expression in specialists despite divergent genotypes (Mcgirr and Martin 2018). These alleles were possibly under balancing selection in the ancestor (most of these alleles are segregating at intermediate frequency in *C. variegatus* currently) and the alternate alleles were driven to fixation between the specialists during their respective divergences. While the species have discrete phenotypes across several morphological traits important to divergence in this system, these traits often lie on a continuous axis (i.e. shorter oral jaw lengths in *C. brontotheriodes*, intermediate jaw lengths in *C. variegatus*, and longer oral jaw lengths in 'wide-mouth' and *C. desquamator*). Some of the alleles that are alternatively fixed between specialists at the same position may have incomplete dominance effects on such traits.

In line with adaptation to divergent trophic niches across the species, we do still see many unique signatures of selection among the specialists. The *C. variegatus* population had the fewest selective sweep signatures. All three specialists had more selective sweeps potentially resulting from strong directional selection for adapting to new fitness peaks. However, while we might expect more selective sweeps in *C. desquamator* genomes given the deeper valley on the fitness landscape isolating the scale-eater phenotypes (Martin and Wainwright 2013c; Martin and Gould 2020), C. *brontotheriodes* had the most selective sweeps detected in their genome. They also have the longest sweeps detected (e.g. 100-kb), suggesting that selection may have occurred most recently in the *C. brontotheriodes*. This matches with the most recent divergence event being between generalist and *C. brontotheriodes* in Osprey Lake. An intriguing implication of this result is the potentially long wait time for speciation of *C. brontotheriodes* species compared *C. desquamator* despite the shallower fitness valley isolating *C. brontotheroides* from generalists. However, we cannot yet rule out that the recent divergence time estimated in Osprey Lake reflects a more recent arrival of *C. brontotheriodes* to that lake after initially diverging long ago in another lake not incorporated in our demographic model.

5.7.3. Supplemental Figures

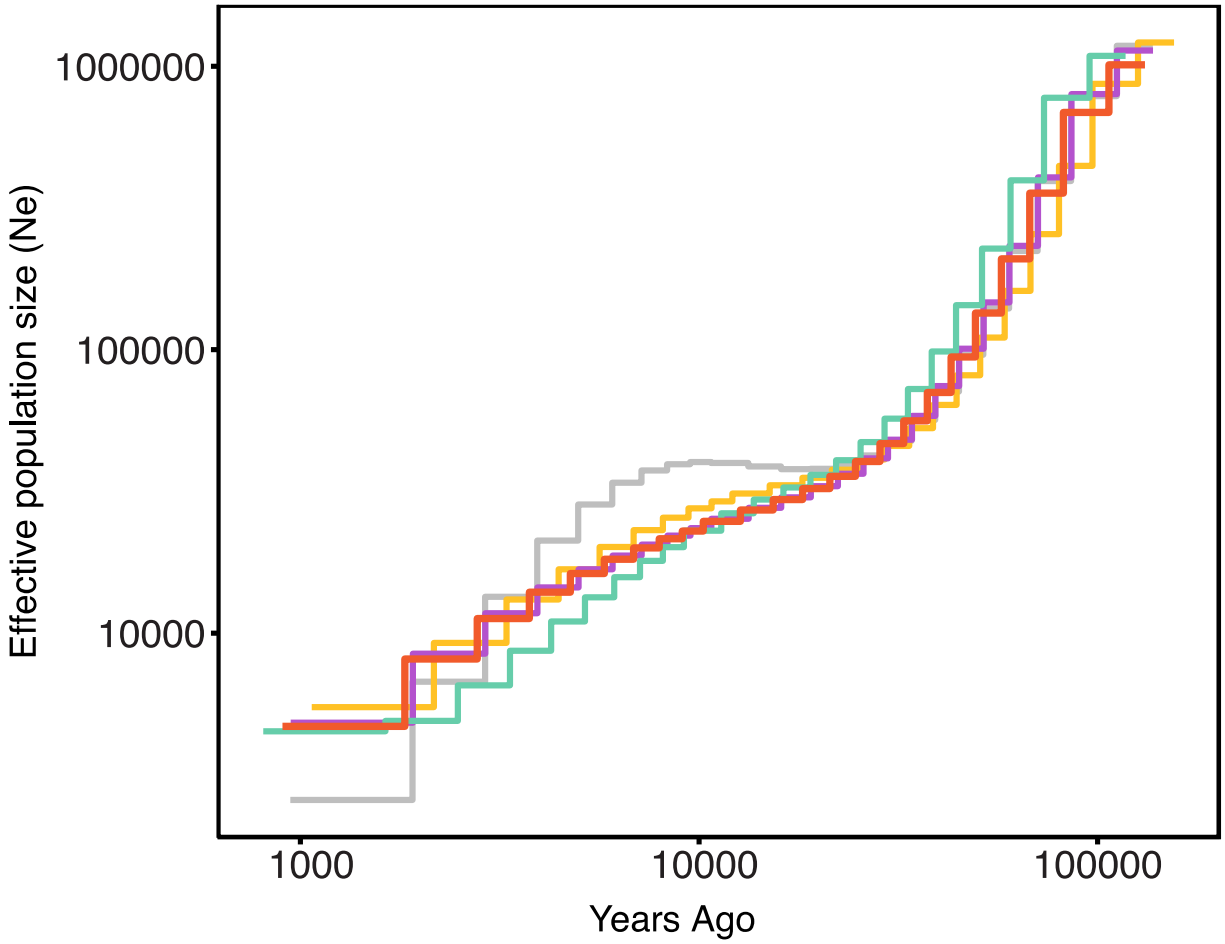


Figure S1. Changes in effective population size over time for Osprey populations. Inferred using MSMC (Schiffels and Durbin 2014) on high-coverage (18-36X) genomes from single individuals of each of the four species in Osprey Lake, San Salvador Island Bahamas: *C. variegatus* (gold), *C. brontotheroides* (purple), *C. desquamator* (teal), and C. sp. 'wide-mouth' (red-orange). Time was scaled using a mutation rate of 1.56x10-8 mutations/basepair/generation and a generation time of 1 year.

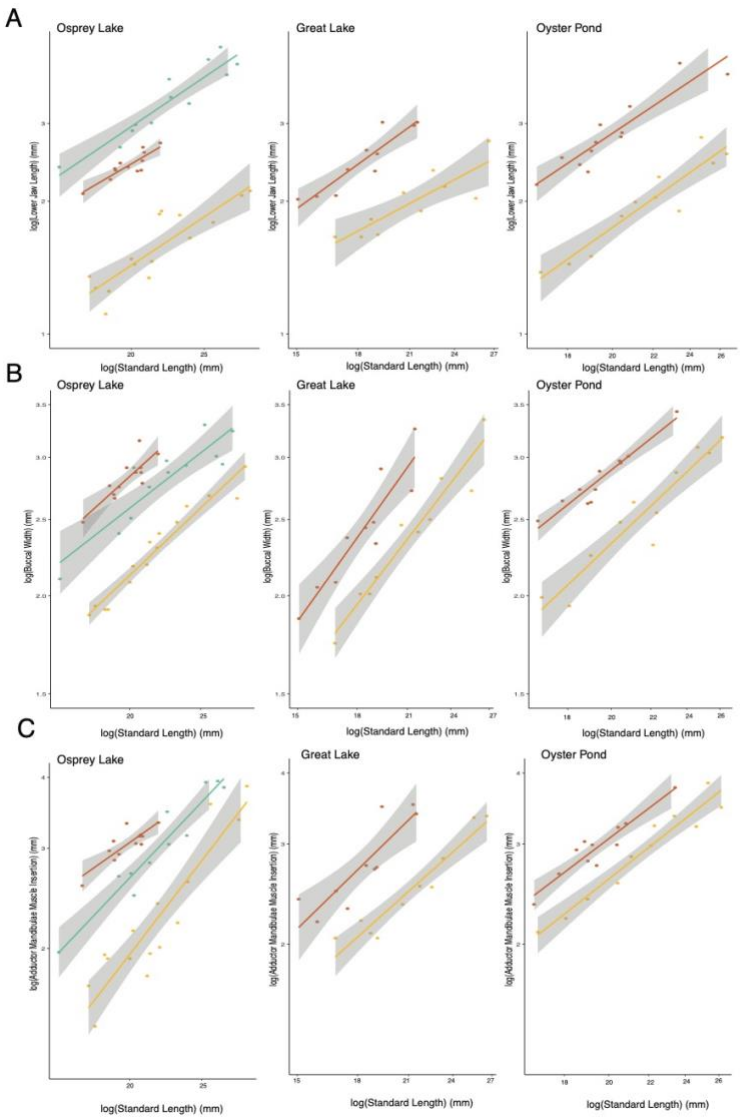


Figure S2. Relationships between three divergent traits and standard length across ponds. The relationship between standard length (mm) of individuals and their D) lower jaw length, E) buccal cavity width, and F) adductor mandibulae muscle insertion height across individuals the three species *C. desquamator* (teal), *C. variegatus* (gold), and *C.* sp. 'wide-mouth' (red-orange) across different ponds (Osprey, Oyster, and Great Lake). Colored lines represent linear model of these relationships for each species with their 95% confidence bands in gray.

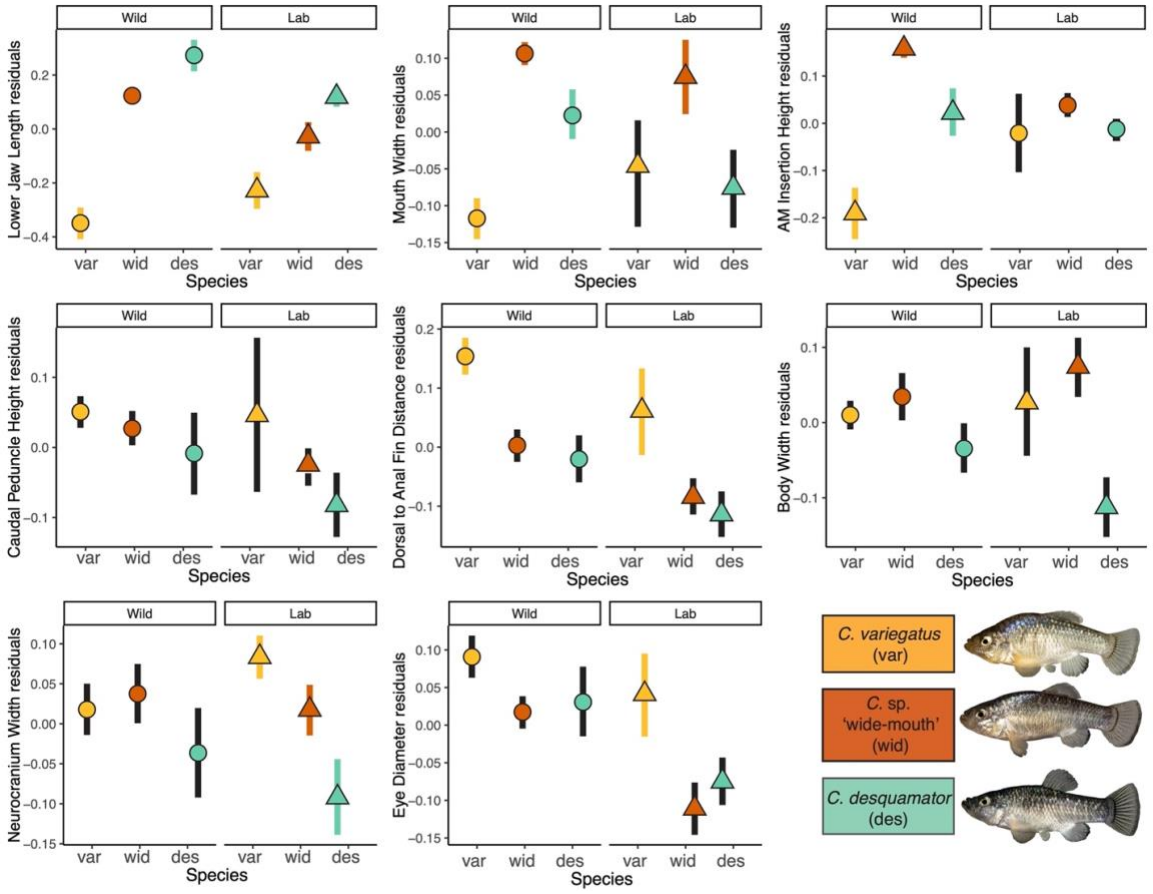


Figure S3. Comparison of the three focal divergent traits in lab and wild populations. A. 95% CI of the standardized sizes of 8 traits measured (see y-axis label on each panel for particular trait) for *C. variegatus* (gold;n=5), *C.* sp. 'wide-mouth' (red-orange;n=20) and *C. desquamator* (teal;n=20). Each plot is broken into two panels of wild caught and lab reared individuals (wild: circle; lab: triangle) represent the mean values of traits from each population. Confidence intervals that overlap between species are highlight in black. Lab raised individuals include F1, F2 and F3 generations. AM insertion height is a proxy for cross-sectional area of the adductor mandibulae

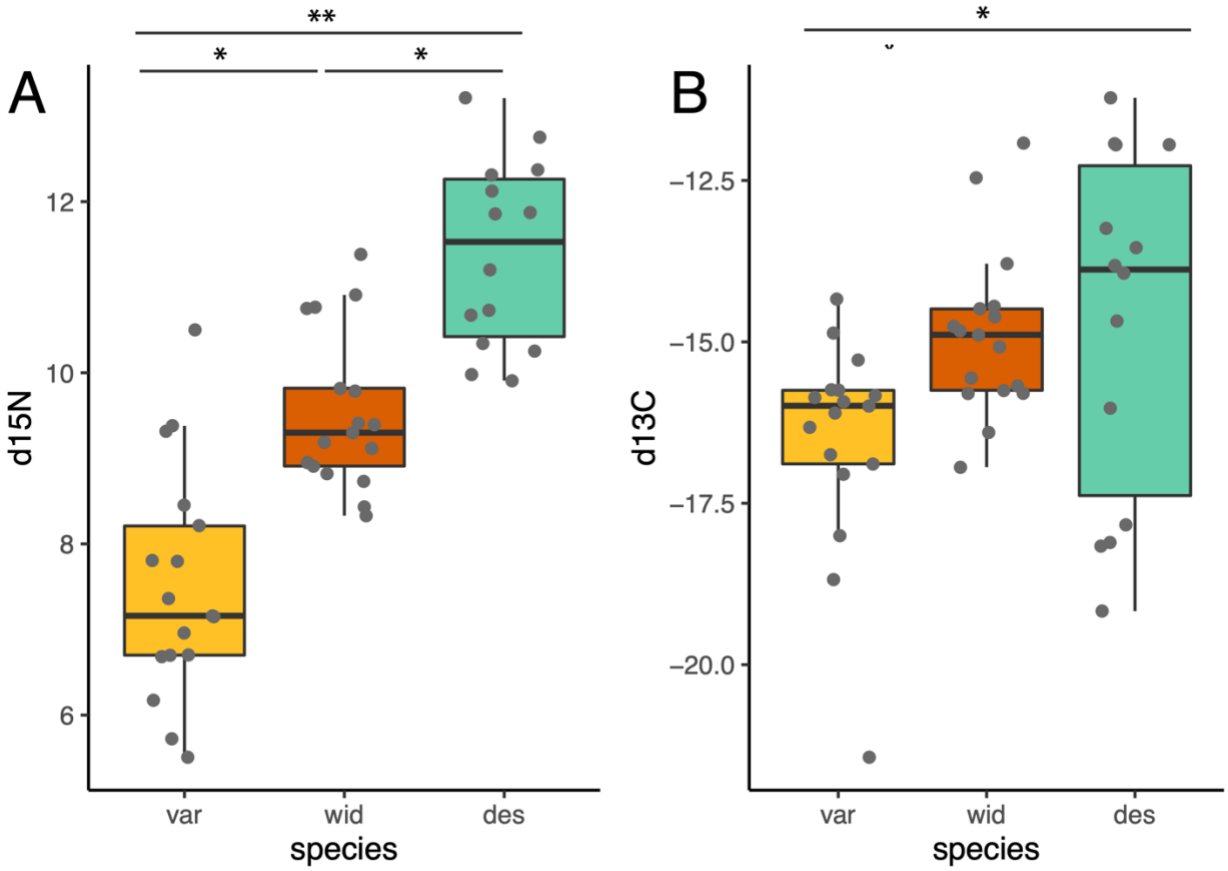


Figure S4. Trophic level comparison among populations in Osprey Lake. d15N values from isotope analysis of muscle tissue from 48 individuals across the three species that span the ecological spectrum from generalist to specialized scale-eater: generalist *C. variegatus* (var;gold), *C. sp.* 'wide-mouth' (wid;red-orange), and *C. desquamator* (des; teal). ANOVA results indicate significant differences in d15N levels across populations indicating that each occupies a distinct trophic level.

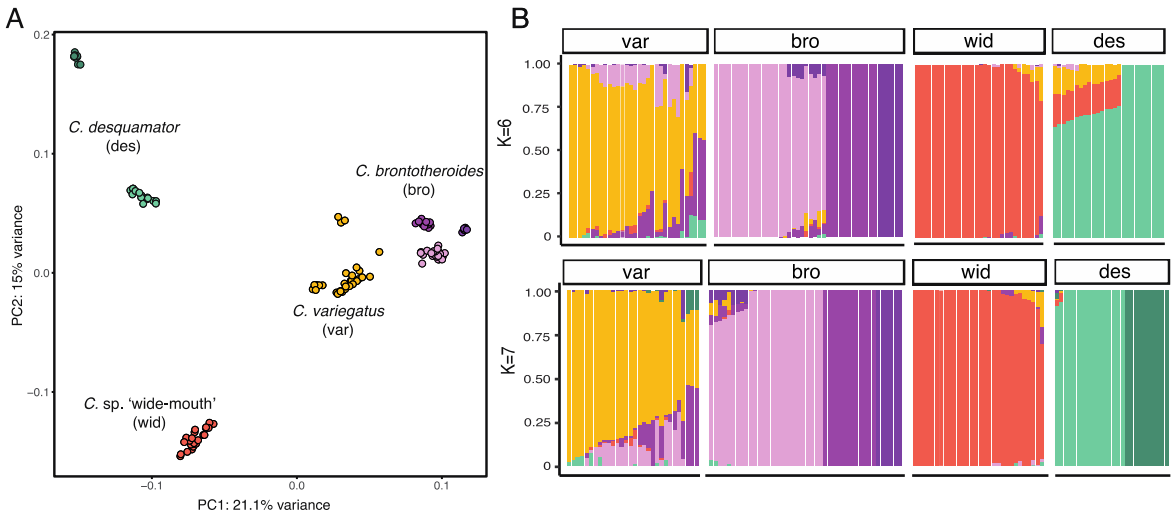


Figure S5. *C.* sp. 'wide-mouth' population did not result from recent hybridization. A) Principal components analysis of the four focal groups on San Salvador Island based on an LD-pruned subset of genetic variants (78,840 SNPs). B) Ancestry proportions across individuals of the four focal groups. Proportions were inferred from ADMIXTURE analyses with 2 values of K with the highest likelihood on the same LD-pruned dataset in A.

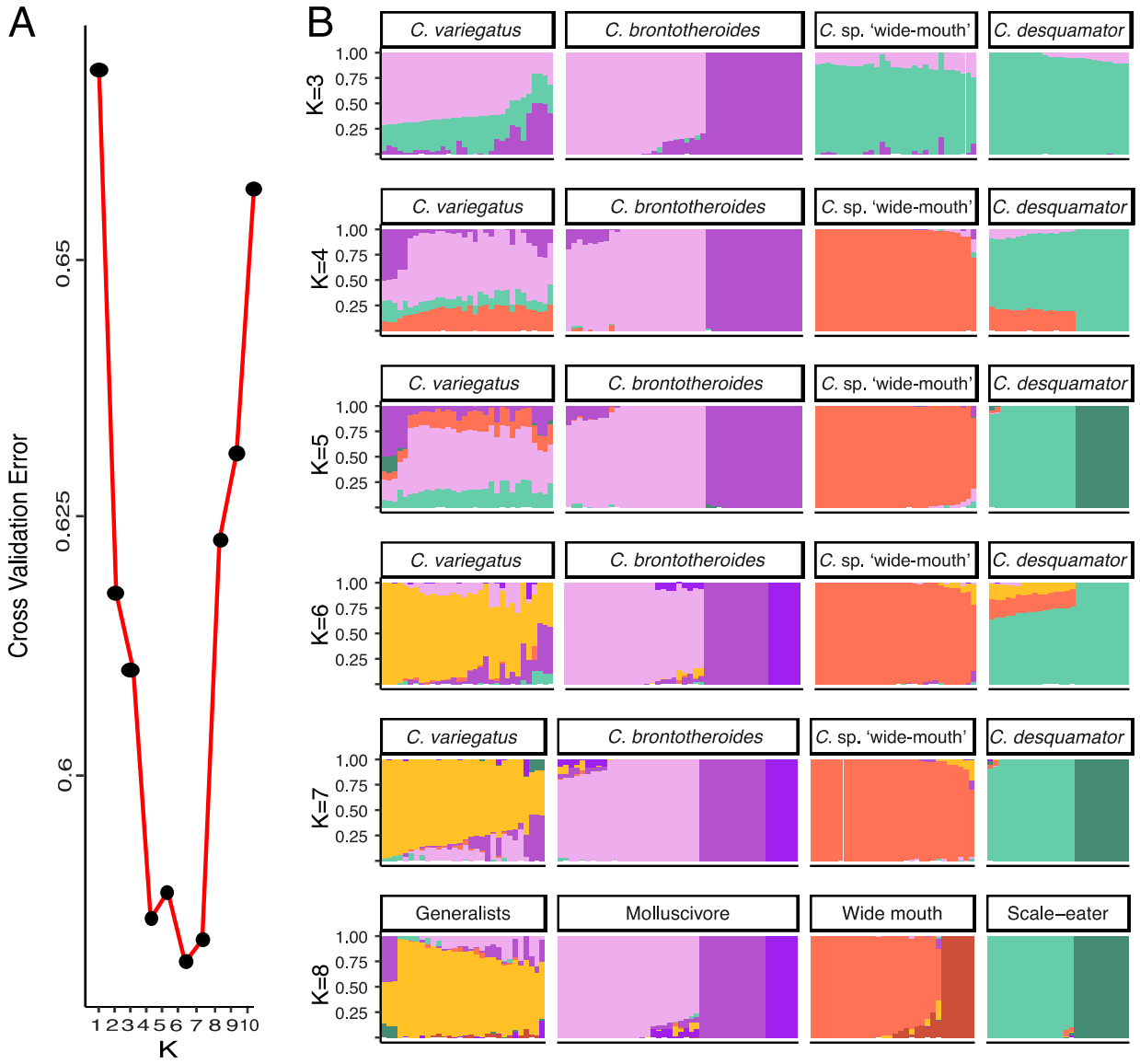


Figure S6. *C.* sp. 'wide-mouth' is not an admixed population. ADMIXTURE analysis of San Salvador Island populations using an LD-pruned subset of genetic variants from 109 individuals (78,840 variants). A) Cross validation error estimates from ADMIXTURE that indicate a model of K=6-7 was the best fit. B) Ancestry proportions across individuals of the three focal groups. Proportions were inferred from ADMIXTURE analyses with 2 values of K with the highest likelihood on the same LD-pruned dataset in A.

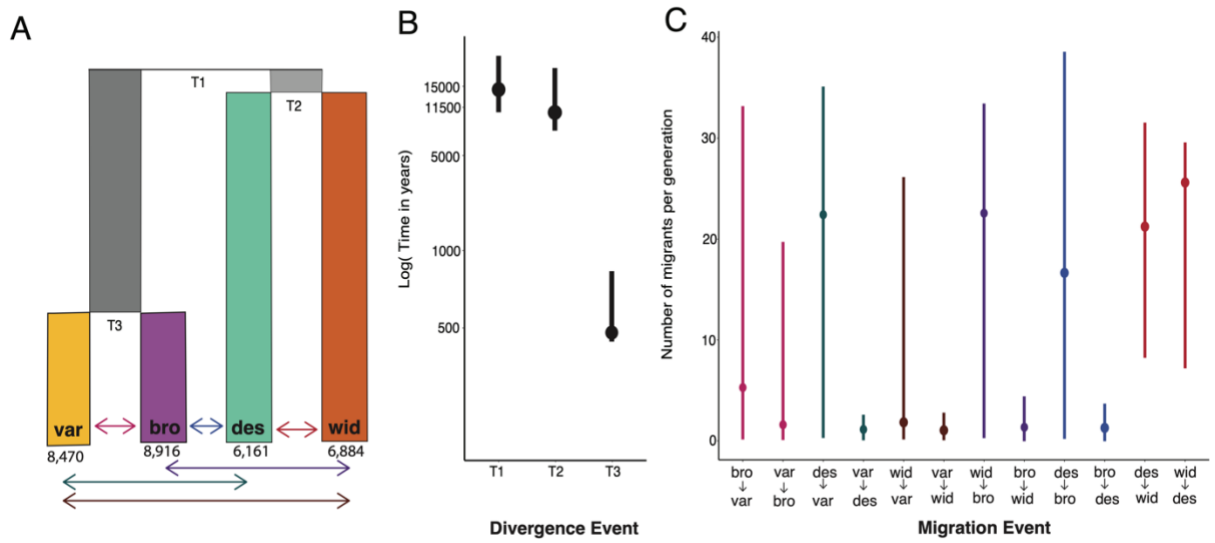


Figure S7. Demographic history of San Salvador Island radiation. A) Best supported demographic model in which *C. desquamator* diverged from the *C.* sp. 'wide-mouth' based on an LD-pruned dataset with no missing information across all Osprey Lake individuals (67,400 SNPs). Divergence time is shown with the maximum likelihood point estimate from the run with the best fit and the 95% confidence interval for that parameter estimate based on 100 bootstrap replicates. B) Maximum likelihood point estimate and 95% confidence intervals for divergence time based on 100 bootstrap replicates from this model. C) Maximum likelihood point estimate and 95% confidence intervals for migration rate parameters involved in the best fitting model depicted in units of the number of migrants per generation.

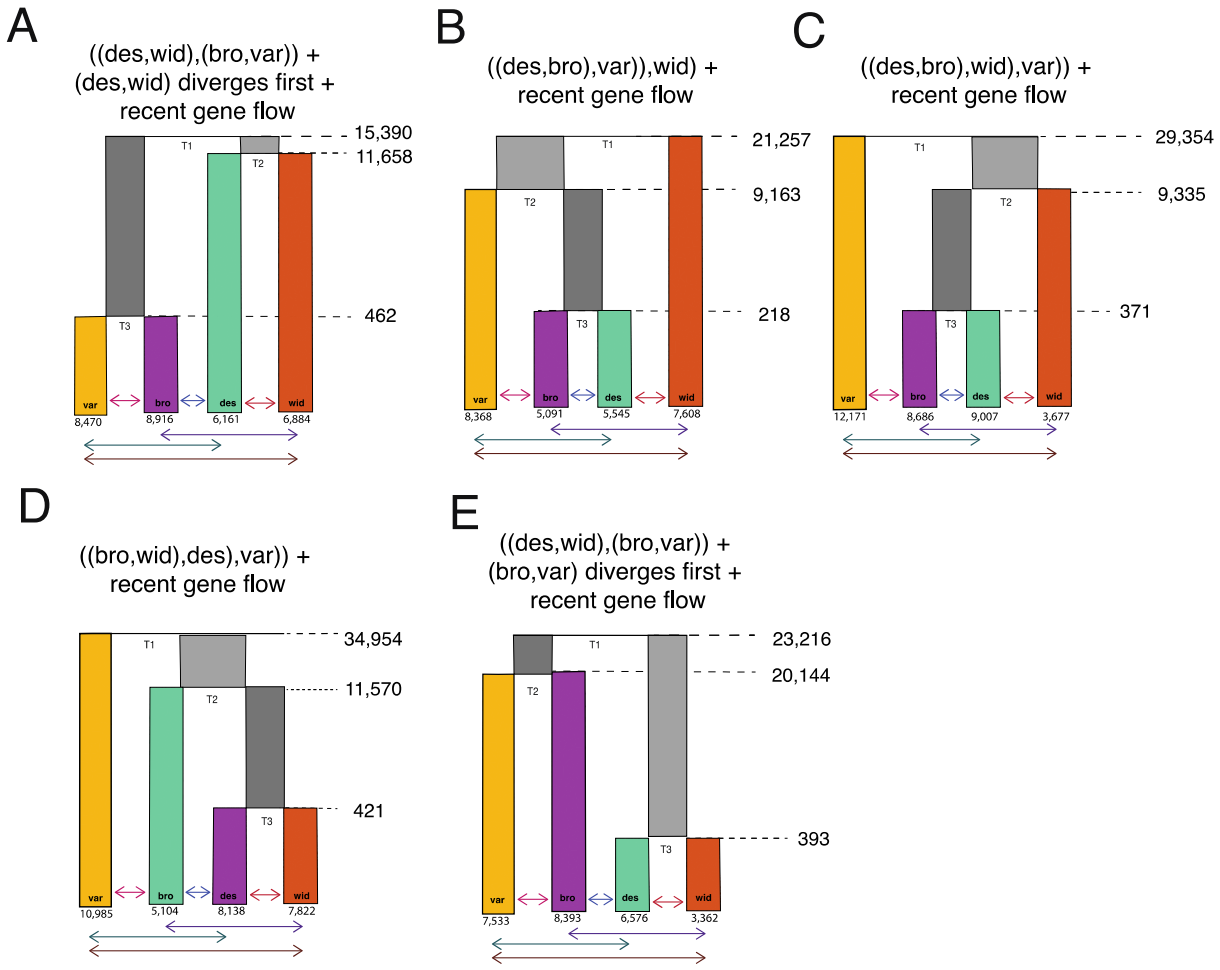


Figure S8. Top five best fitting demographic models for Osprey Lake. From a genetic dataset that was LD-pruned and had no missing information across individuals comprised of 67,400 SNPs. Listed in order of lowest AIC scores (Table 1 and S2) with best supported model in A. All top models were of recent bidirectional gene flow but varied in topology among the four species.

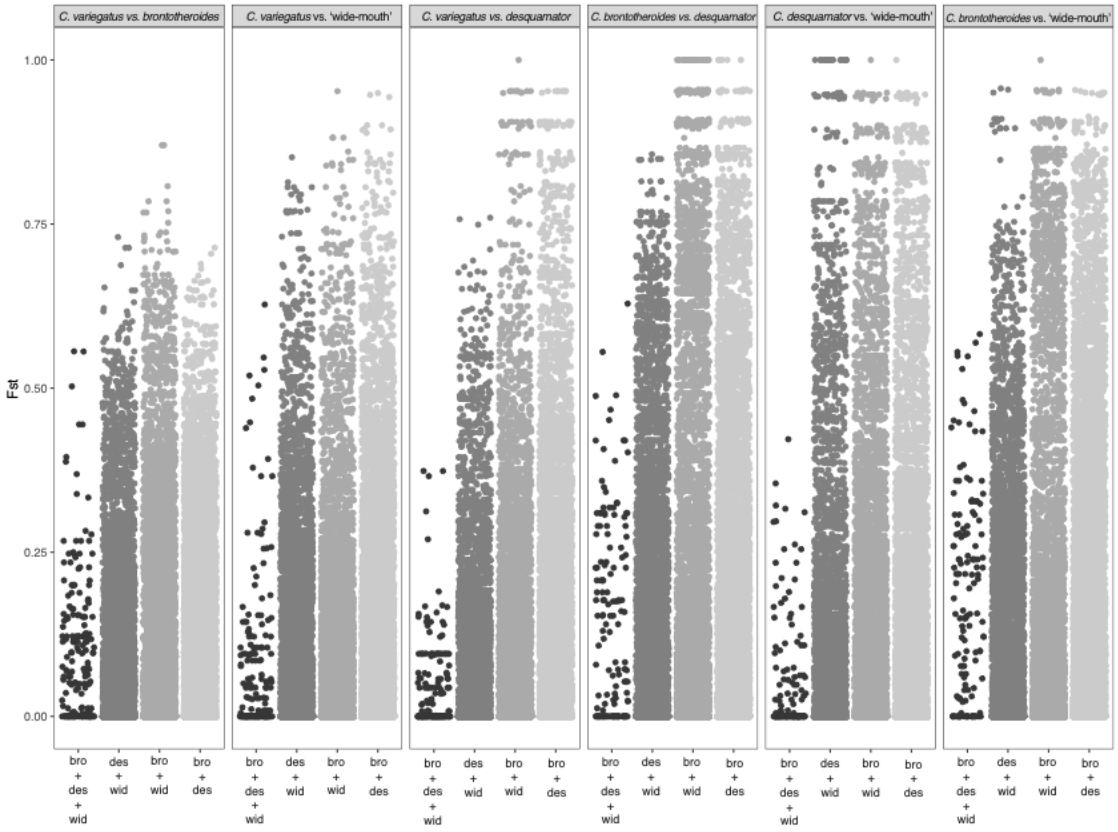


Figure S9. Genetic divergence among populations in Osprey Lake in regions of shared selection across two or more specialists. Each panel represents pairwise F_{st} comparison between two populations for every SNP found in regions under shared selection across two or more specialist species on San Salvador Island: C. brontotheroides (bro); C. desquamator (des), C. sp. 'wide-mouth' (wid). Regions shared across all three specialists (bro+des+wid) do not contain highly divergent SNPs between any of the species in the radiation. Highly divergent SNPs (Fst > 0.75) are only found in regions shared between one or two specialists.

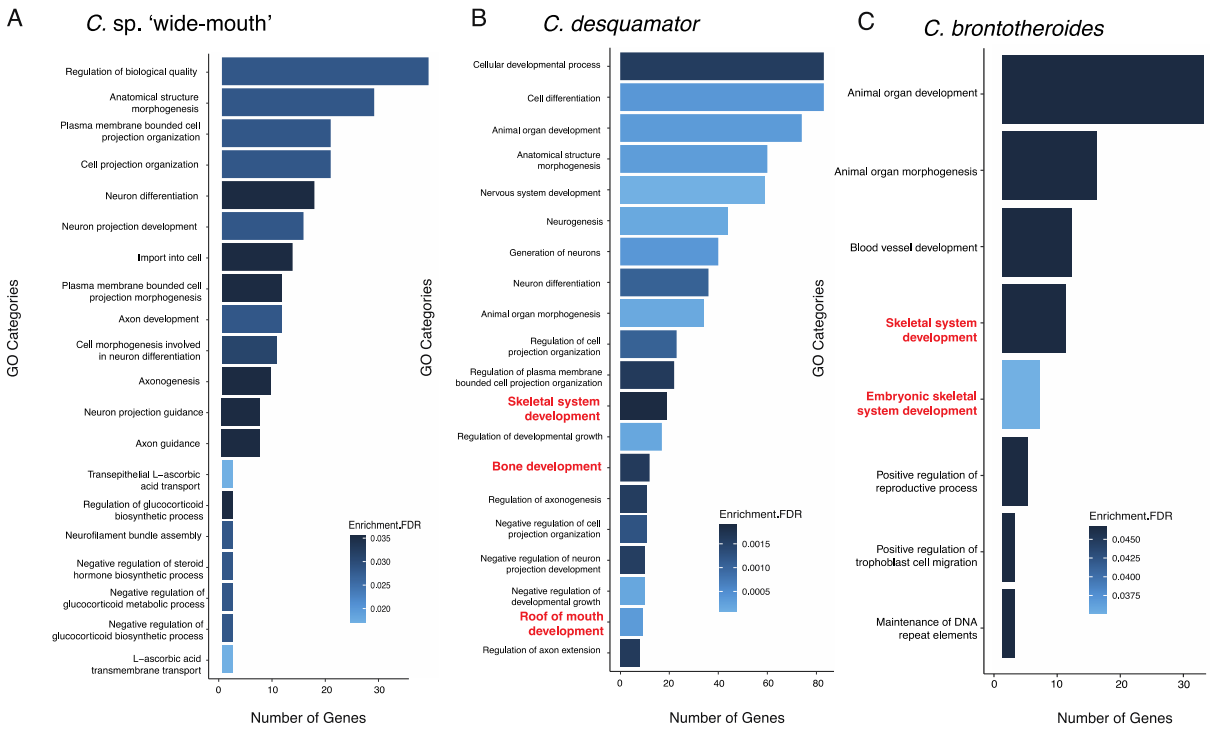


Figure S10. Top 20 enriched GO categories for divergent alleles in each of the three specialists. GO enrichment analyses were performed on genes in or near (within 20-kb) of a SNP that was under a hard selective sweep and strongly diverged from generalist species (top 1% of Fst values across genome) for a) 'wide-mouth' b) *C. desquamator*, and C) *C. brontotheriodes*. GO categories that were significantly enriched for relevant terms corresponding to craniofacial development, the major axes of morphological divergence in this radiation, are highlighted in red. All terms included were significant at an FDR < 0.05 and full list of terms in Data S1-3.

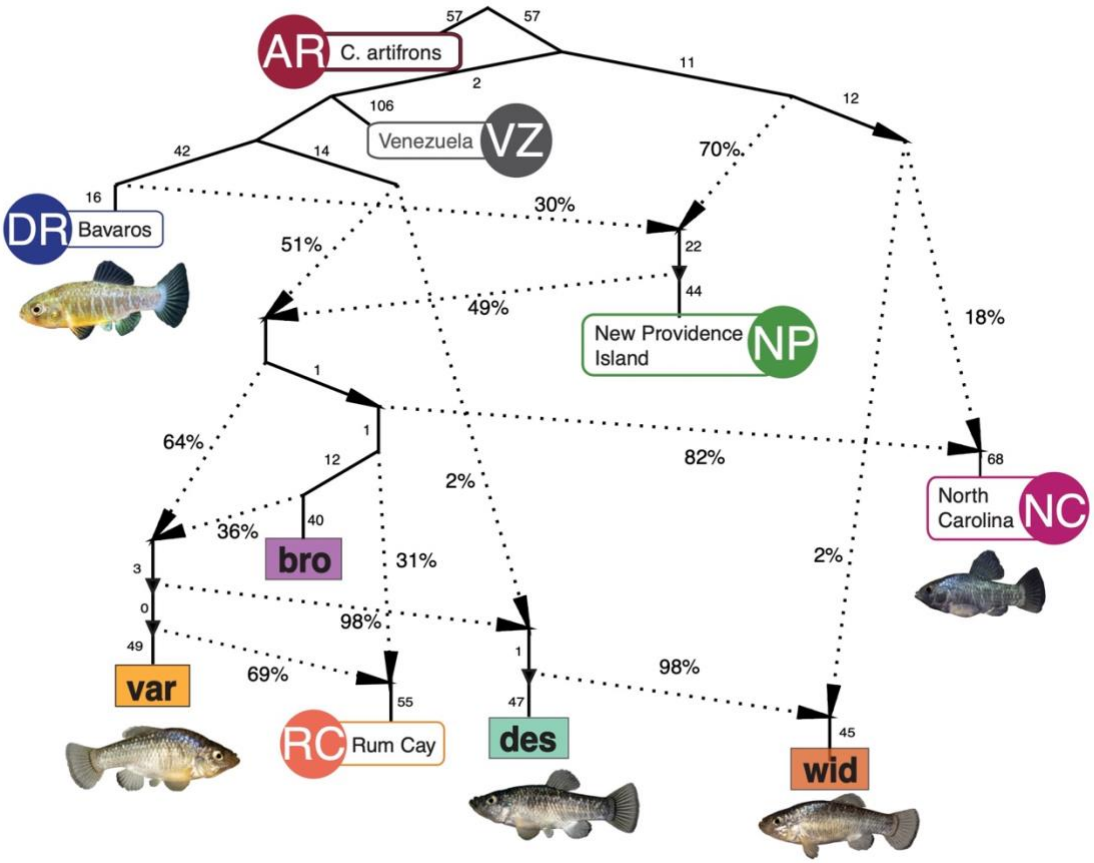


Figure S11. Caribbean pupfish admixture graph. Best-fiting admixture graph from *qpGraph* approach that estimated f_3-statistics across Caribbean populations. This admixture model features 7 admixture events represented by admixture edges and proportions (percentages on dotted branches) and drift edges (numbers on solid branches) that minimize the difference between fitted and estimated f_3-statistics.

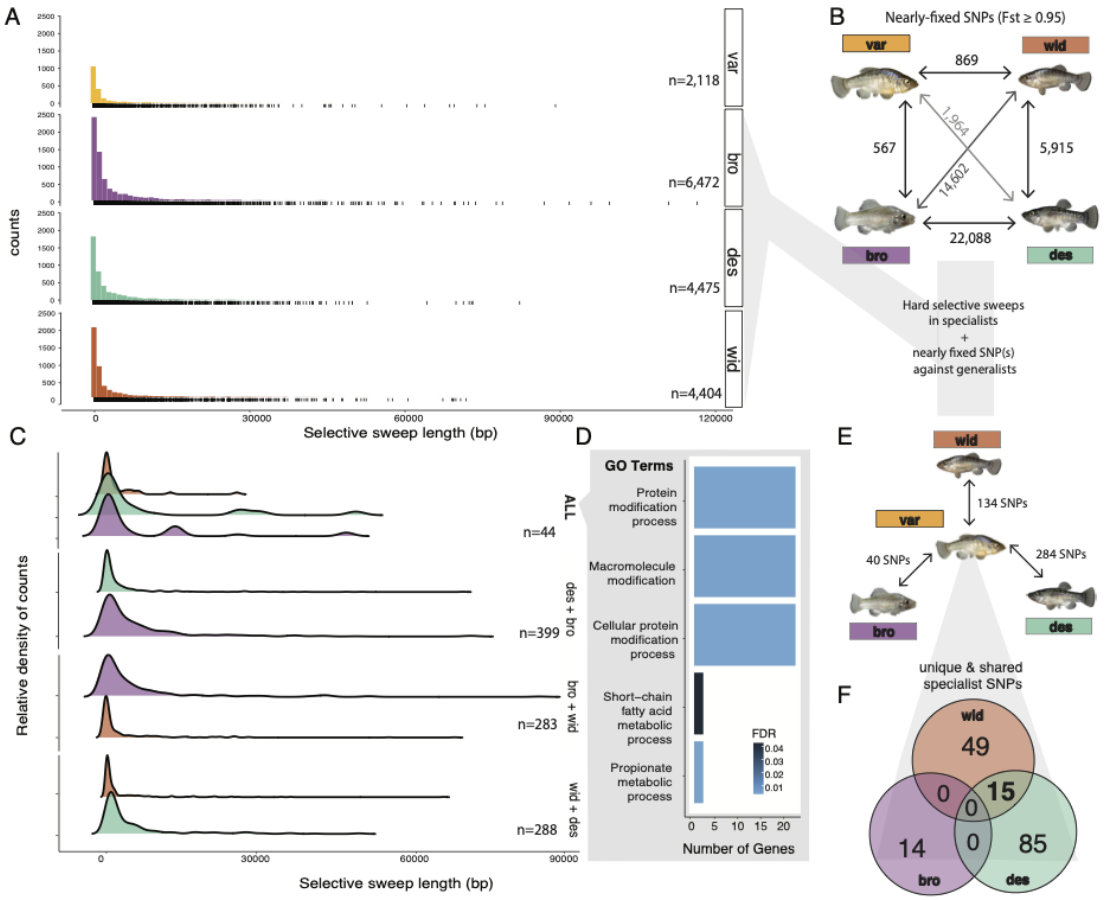


Figure S12. Patterns of selection and genetic divergence in specialist genomes**.** A) Selective sweep length distributions across all four San Salvador Island species. Rug plot below each histogram represents the counts of selective sweeps in different length bins. B) The total number of fixed or nearly-fixed SNPs (F_{st}≥ 0.95) between each group in Osprey Pond. C) Ridgeline plots for length distributions of selective sweeps shared between different combinations of specialists (*C. brontotheriodes*: bro, *C. desquamator*: des, *C.* sp. 'wide-mouth': wid). D) GO enrichment analysis of the 57 genes contained in the shared selective sweep regions across all specialists. E) The number of adaptive alleles (fixed or nearly-fixed SNPS [F_{st}≥ 0.95] relative to *C. variegatus* and under selection in each population of specialists in Osprey Lake. F) Venn diagram highlights those adaptive alleles that are unique to each specialist and shared with other specialists.

5.7.4. Supplemental Tables

Table S1. No significant genome-wide signature of admixture among populations on San Salvador Islands. Results from three formal tests for introgression to assess whether 'wide-mouth' populations are hybrids between generalist *C. variegatus* and scale-eater *C. desquamator*: D-, f_3- and f_4-statistics across all possible combinations of Osprey Lake populations of generalist species. Significant f_3-statistic had Z-scores > -2; significant f_4 and D-statistic Z-scores were smaller than -2 and greater than 2. The only significant signature of admixture (bolded Z-scores) comes from D- and f_4-statistics based on relationships that violate the expected tree (((*C. desquamator*; 'wide-mouth'),Generalist),Outgroup) and therefore should not be interpreted as evidence of 'wide-mouth' being an admixed population. *C. variegatus* population from Rum Cay, the nearest neighbor island in the Bahamas to San Salvador Island was used as an outgroup population for D- and f_4-statistics.*C. var = C. variegatus*; *C. bro = C. brontotheroides*; *C. des = C. desquamator*

f_3-statistic						
A	B	C	--	f_3	stderr	*Z*
C. var	*C. des*	C. sp. 'wide-mouth'		0.0046	0.0008	5.25
f_4- statistic						
P1	P2	P3	P4	f_4	stderr	*Z*
C.sp 'wide-mouth'	*C. des*	*C. var*	Rum Cay	-0.0002	0.00016	-1.2
C.sp 'wide-mouth'	*C. va*	*C. bro*	Rum Cay	0.00018	0.0002	0.92
D-statistic						
P1	P2	P3	Outgroup	D	stderr	*Z*
C.sp.'wide-mouth'	*C. des*	*C. var*	Rum Cay	-0.0018	0.00149	-1.19
C. sp. 'wide-mouth'	*C. var*	*C. bro*	Rum Cay	0.0017	0.00179	0.93

Table S2. Support for the 28 demographic models for the evolution of the 'wide-mouths' from the site frequency spectrum. The likelihood and AIC scores all demographic models estimated in *fastsimcoal2* for the Osprey populations of *C. variegatus* var), *C. brontotheroides* (bro), *C. desquamator* (des), and *C.* sp. 'wide-mouth' (wid) are presented here with a complete list of all models tested reported in Table S2. Change in likelihood (ΔLnL) represents the difference in likelihood from that of a simulated SFS expected by the demographic model tested. Change in AIC (ΔAIC) represents the difference in AIC scores from that of the model with the smallest ΔLnL. All models presented here represent different divergence scenarios with recent gene flow allowed, which had better support from the data than models with no gene flow or early gene flow. Visual representations of the top five models are depicted in Figure S5.

Divergence Model	LnL	ΔLnL	AIC	ΔAIC
(1) ((des,wid),(bro,var)) + (des,wid) diverges first + recent gene flow	-11701	1077	54133	0
(2) ((des,bro),var)) ,wid)) + recent gene flow	-11780	1112	54296	35
(3) ((des,bro),wid),var)) + recent gene flow	-11806	1138	54418	61
(4) ((bro,wid),des),var)) + recent gene flow	-11808	1140	54423	63
(5) ((des,wid),(bro,var)) + (bro,var) diverges first + recent gene flow	-11846	1178	54602	101
(6) ((des,wid), bro),var)) + recent gene flow	-11855	1187	54642	110
(7) ((des,bro),var)) ,wid))+ early gene flow	-12016	1348	55364	271
(8) ((des,var),bro), wid)) + recent gene flow	-12024	1356	55426	279
(9) ((wid,bro),des),var)) + early gene flow	-12086	1418	55688	341
(10) ((wid,des),bro),var)) + early gene flow	-11855	1419	55692	342
(11) ((des,wid),(bro,var)) + early gene flow	-12139	1471	55929	394
(12) ((des,wid), bro),var)) + no gene flow	-12315	1648	56738	570
(13) ((wid,bro),des),var)) + no gene flow	-12326	1658	56787	581
(14) ((des,bro),var)) ,wid))+ no gene flow	-12334	1667	56826	589
(15) ((des,wid),(bro,var)) + recent sister gene flow only	-12394	1726	57106	649
(16) ((des,wid),(bro,var))+ no gene flow	-12396	1728	57109	651
(17) ((wid,var)),bro),des)) + recent gene flow	-12399	1731	57148	654
(18) ((bro,var)) ,wid),des)) + recent gene flow	-12401	1733	57157	656
(19) ((wid,bro),var)),des)) + recent gene flow	-12401	1734	57163	657
(20) ((des,wid),var),bro)) + recent gene flow	-12420	1752	57244	675
(21) ((wid,var)),des),bro)) + recent gene flow	-12425	1757	57265	680
(22) ((des,var),wid),bro)) + recent gene flow	-12425	1757	57267	680
(23) ((wid,var)),(bro,des)) + no gene flow	-23245	12577	107070	11500
(24) ((wid,var)),(bro,des)) + early gene flow	-23658	12990	108976	11913
(25) ((wid,var)),(bro,des)) + recent gene flow	-23659	12991	109003	11914
(26) ((des,var),(bro,wid)) + early gene flow	-34983	24315	161129	23238
(27) ((des,var),(bro,wid)) + no gene flow	-34983	24315	161125	23238
(28) ((des,var),(bro,wid)) + recent gene flow	-34983	24315	161149	23238

Divergence Model	LnL	ΔLnL	AIC	ΔAIC
(1) ((des,wid),(bro,var)) + (des,wid) diverges first + recent gene flow	-11701	1077	54133	0
(2) ((des,bro),var)) ,wid)) + recent gene flow	-11780	1112	54296	35
(3) ((des,bro),wid),var)) + recent gene flow	-11806	1138	54418	61
(4) ((bro,wid),des),var)) + recent gene flow	-11808	1140	54423	63
(5) ((des,wid),(bro,var)) + (bro,var) diverges first + recent gene flow	-11846	1178	54602	101

Table S3. Parameters for selective sweep analyses in SweeD.
The 95th percentile of composite likelihood ratio threshold based on neutral simulations under the demographic scenario of decreasing population size through time inferred with MSMC (Figure S4), and the population size change and haplotype number information parameters required by SweeD for each species.

Species	CLR threshold	SweeD Commands
C. variegatus	0.40320	-folded -strictPolymorphic -G 0.407 -eN 5.45 181.8 -s 24
C. brontotheroides	0.43473	-folded -strictPolymorphic -G 0.389 -eN 5.88 196 -s 32
C. desquamator	0.40537	-folded -strictPolymorphic -G 0.218 -eN 8.11 270 -s 16
C. sp. 'wide-mouth'	0.39262	-folded -strictPolymorphic -G 0.203 -eN 8.57 276 -s 22

Table S4. Genetic divergence among the four populations in Osprey Lake, San Salvador Island across different thresholds of F_{st}. Relative measure of genetic divergence was calculated in pairwise combinations of the different species as the number of fixed SNPs between them, the number of fixed or nearly fixed SNPs between them, the top 1% of F_{st} between them from the distribution of F_{st} between SNPs and the genome-wide average F_{st} across all SNPs in the genome.

F_{st} Comparison	Number of fixed SNPs (F_{st} =1)	Number of nearly fixed SNPS ($F_{st} \geq 0.95$)	Top 1%	Genome-wide average
C. sp. 'wide-mouth' vs *C. desquamator*	5,212	5,915	0.84	0.131
C. sp. 'wide-mouth' vs *C. brontotheroides*	5238	14,602	0.85	0.15
C. sp. 'wide-mouth' vs *C. variegatus*	247	869	0.68	0.104
C. desquamator vs *C. brontotheroides*	6941	22,088	0.86	0.17
C. desquamator vs *C. variegatus*	414	1,964	0.71	0.116
C. variegatus vs *C. brontotheroides*	110	567	0.637	0.09

Table S5. Comparison of the average absolute genetic divergence (D_{xy}) between the four species in Osprey Lake 'wide-mouth', *C. variegatus, C. desquamator and C. brontotheriodes* and the shared adaptive alleles between scale-eater populations.

Comparison	Mean D_{xy}
C. variegatus vs *C. brontotheroides*	0.167
C. variegatus vs *C. desquamator*	0.169
C. variegatus vs *C.* sp. 'wide-mouth'	0.165
C. brontotheroides vs *C. desquamator*	0.162
C. brontotheroides vs *C.* sp. 'wide-mouth'	0.156
C. desquamator vs *C.* sp. 'wide-mouth'	0.157
Selective sweeps in *C.* sp. 'wide-mouth'	0.123
Shared genetic variants with *C. desquamator*	0.0814
Unique genetic variants in *C.* sp. 'wide-mouth'	0.24

Table S6. The distribution of shared adaptive alleles between the two scale-eating species across the Caribbean. Numbers represent the copies of the *C. desquamator* and 'wide-mouth' allele present in outgroup populations collected from San Salvador Island (SS), Rum Cay (RC), Cat Island (CT), Exumas Islands (EX), Long Island (LG), Mayaguana (MG), New Providence Island (NPI) in the Bahamas. Lagunas Bavaros (BAV) and Etang Saumautre (ETA) in Dominican Republic, Sarasota Estuary in Florida (FL), Curacoa (CUR), Caicos Island (CAI) and Isla Margarita (MAR) in Venezuela.

			C. variegatus								*C. laciniatus*	*C. higuey*	*C. bondi*	*C. dearborni*		
Allele	Gene	Distribution	SS	RC	CT	EX	LG	MG	FL	NC	NPI	BAV	ETA	CUR	CAI	MAR
HiC_scaffold_6:907101	NA	SGV	0	0	1	0	0	0	0	12	0	0	0	0	0	0
HiC_scaffold_6:907425	NA	SGV	1	0	0	0	0	0	0	7	0	0	0	0	0	0
HiC_scaffold_6:907665	NA	SGV	0	0	0	0	0	0	0	7	0	0	0	0	0	0
HiC_scaffold_6:908229	NA	SGV	1	0	1	0	0	0	0	6	3	0	0	0	0	0
HiC_scaffold_6:908889	NA	SGV	1	0	0	0	0	0	0	6	0	0	0	0	0	0
HiC_scaffold_6:923182	NA	SGV	0	5	0	1	0	0	1	0	0	0	0	0	0	0
HiC_scaffold_6:923590	NA	SGV	0	5	0	2	0	0	0	0	0	0	0	0	0	0
HiC_scaffold_6:924203	NA	de novo	0	0	0	0	0	0	0	0	0	0	0	0	0	0
HiC_scaffold_6:929621	NA	SGV	1	2	1	0	0	0	0	0	0	0	0	0	0	0
HiC_scaffold_9:31763392	daam2	SGV	6	3	0	1	0	1	1	1	6	1	0	0	0	0
HiC_scaffold_9:31763400	daam2	SGV	8	5	0	1	1	0	1	8	10	8	0	0	0	0
HiC_scaffold_24:5390450	usp50	SGV	1	1	0	0	0	0	0	0	1	0	0	0	1	0
HiC_scaffold_43:27385265	NA	SGV	3	1	0	2	1	1	0	5	1	0	1	2	0	11
HiC_scaffold_43:27389200	NA	SGV	5	1	0	0	1	1	1	12	12	2	1	3	1	12
HiC_scaffold_44:14963656	atp8a1	SGV	3	1	0	1	0	1	1	12	12	2	1	3	1	12
HiC_scaffold_60:1798737	znf214	SGV	3	0	0	0	0	1	0	0	2	0	0	0	0	0

Table S7. The distribution of unique adaptive alleles to the highly specialized scale-eater *C. desquamator* across the Caribbean. Numbers represent the copies of the *C. desquamator* allele present in outgroup populations collected from San Salvador Island (SS), Rum Cay (RC), Cat Island (CT), Exumas Islands (EX), Long Island (LG), Mayaguana (MG), New Providence Island (NPI) in the Bahamas. Lagunas Bavaros (BAV) and Etang Saumautre (ETA) in Dominican Republic, Sarasota Estuary in Florida (FL), Curacoa (CUR), Caicos Island (CAI) and Isla Margarita (MAR) in Venezuela. The spatial distribution of each allele is summarized into three categories: standing genetic variation (SGV); de novo to San Salvador Island populations (de novo) and alleles found in introgressed regions (Intro).

			C. variegatus								*C. laciniatus*	*C. higuey*	*C. bondi*	*C. dearborni*		
Allele	Gene	Distribution	SS	RC	CT	EX	LG	MG	FL	NC	NP	BAV	ETA	CUR	CAI	MAR
HiC_scaffold_1:29498971	NA	SGV	2	1	0	1	1	2	2	0	0	0	2	2	0	22
HiC_scaffold_1:29505055	NA	SGV	6	0	0	0	0	0	0	24	0	0	0	0	0	0
HiC_scaffold_1:29518877	NA	de novo	4	0	0	0	0	0	0	0	0	0	0	0	0	0
HiC_scaffold_1:36701070	NA	SGV	2	0	0	0	0	0	0	24	7	9	0	0	0	0
HiC_scaffold_1:36701336	NA	SGV	2	3	0	0	0	0	0	0	0	0	0	0	0	0
HiC_scaffold_1:36704196	NA	SGV	4	1	0	0	0	0	0	24	5	10	0	0	0	0
HiC_scaffold_1:36705430	NA	SGV	3	1	0	0	0	0	0	24	5	9	0	0	0	0
HiC_scaffold_1:36705978	NA	SGV	1	1	0	0	0	0	0	22	7	10	0	0	0	0
HiC_scaffold_1:36707870	NA	SGV	4	4	2	2	0	2	0	0	0	0	0	0	2	0
HiC_scaffold_1:36713418	NA	SGV	1	0	0	0	0	0	0	24	7	12	2	4	0	0
HiC_scaffold_1:36715711	NA	Intro NC	4	6	3	4	0	2	0	23	9	11	2	4	2	0
HiC_scaffold_1:36715712	NA	Intro NC	4	6	3	4	0	2	2	23	9	11	0	4	2	0
HiC_scaffold_1:36740602	NA	SGV	5	5	2	4	1	1	2	20	4	20	2	4	0	22
HiC_scaffold_1:36741004	NA	SGV	2	4	2	2	1	0	0	22	7	7	0	4	0	24
HiC_scaffold_1:36742540	NA	SGV	2	0	0	0	0	0	0	0	0	1	0	0	0	0
HiC_scaffold_1:36743115	NA	SGV	4	6	2	0	1	0	2	24	6	7	0	4	0	24
HiC_scaffold_1:36744639	NA	SGV	0	1	0	0	0	0	0	0	0	2	0	0	1	0
HiC_scaffold_1:36744717	NA	SGV	1	2	0	2	1	2	2	24	5	5	2	4	2	23
HiC_scaffold_1:36744821	NA	SGV	1	1	0	1	0	0	1	0	0	2	0	0	1	0

HiC_scaffold_1:36745174	NA	SGV	2	6	2	6	1	4	2	22	5	18	2	4	2	24
HiC_scaffold_1:36745293	NA	SGV	3	4	4	6	1	2	2	24	6	20	2	4	2	24
HiC_scaffold_1:36748002	NA	SGV	0	0	0	0	0	0	0	24	0	5	0	0	1	0
HiC_scaffold_5:5681542	NA	de novo	1	0	0	0	0	0	0	0	0	0	0	0	0	0
HiC_scaffold_5:6082906	NA	SGV	3	9	0	2	0	0	0	0	0	0	0	0	0	0
HiC_scaffold_5:6082926	NA	SGV	3	9	0	2	1	0	1	24	0	6	2	0	0	24
HiC_scaffold_5:8296047	NA	de novo	1	0	0	0	0	0	0	0	0	0	0	0	0	0
HiC_scaffold_5:8465173	NA	de novo	2	0	0	0	0	0	0	0	0	0	0	0	0	0
HiC_scaffold_5:21455838	NA	SGV	6	0	0	2	0	0	0	8	0	6	0	2	1	0
HiC_scaffold_8:14526358	NA	de novo	0	0	0	0	0	0	0	0	0	0	0	0	0	0
HiC_scaffold_11:10913509	NA	de novo	3	0	0	0	0	0	0	0	0	0	0	0	0	0
HiC_scaffold_11:10948683	cd226	de novo	5	0	0	0	0	0	0	0	0	0	0	0	0	0
HiC_scaffold_11:12963188	NA	SGV	1	0	0	4	2	1	2	6	1	3	0	0	0	0
HiC_scaffold_21:10041965	cntn5	SGV	7	0	0	0	0	2	0	0	17	0	0	0	1	0
HiC_scaffold_21:31993865	slc35e1	de novo	1	0	0	0	0	0	0	0	0	0	0	0	0	0
HiC_scaffold_24:7400912	NA	SGV	0	4	0	4	0	2	0	0	0	0	0	0	0	0
HiC_scaffold_24:7402072	NA	SGV	1	2	0	2	1	0	1	0	0	0	2	0	0	0
HiC_scaffold_33:12638684	bri3bp	SGV	11	15	1	0	0	2	0	0	4	2	0	4	0	0
HiC_scaffold_33:12644904	bri3bp	SGV	2	9	0	0	0	1	0	0	4	0	0	0	0	0
HiC_scaffold_33:12645055	bri3bp	SGV	3	7	0	2	0	1	1	11	8	15	2	4	2	0
HiC_scaffold_33:12645107	bri3bp	SGV	1	7	0	0	0	0	0	0	4	0	0	0	0	0
HiC_scaffold_33:12645423	bri3bp	SGV	5	8	0	0	0	0	0	0	4	0	0	0	0	0
HiC_scaffold_33:12645669	bri3bp	SGV	1	8	0	0	0	0	0	10	5	0	0	0	0	0
HiC_scaffold_33:12646217	bri3bp	SGV	2	9	0	0	0	0	0	0	6	0	0	0	0	0
HiC_scaffold_33:12651722	wdr31	SGV	4	0	0	0	0	0	0	0	0	1	0	0	0	0
HiC_scaffold_33:12656779	wdr31	SGV	7	12	1	0	0	1	0	3	16	9	2	2	0	0
HiC_scaffold_33:12656781	wdr31	SGV	7	12	1	0	0	1	0	3	16	9	2	2	0	0
HiC_scaffold_33:12656782	wdr31	SGV	7	12	1	0	0	1	0	3	16	9	2	2	0	0
HiC_scaffold_33:12663470	wdr31	SGV	4	9	0	0	1	0	2	8	16	6	2	2	0	21

HiC_scaffold_33:12663497	wdr31	SGV	4	8	0	0	0	0	0	0	13	0	0	0	0	0
HiC_scaffold_33:12663527	wdr31	SGV	4	8	0	0	0	0	0	0	12	0	0	0	0	0
HiC_scaffold_33:12664211	wdr31	SGV	3	11	0	0	0	0	0	0	16	0	0	0	0	0
HiC_scaffold_33:12665844	wdr31	SGV	2	12	0	0	0	0	0	0	18	0	0	0	0	0
HiC_scaffold_33:12666096	wdr31	SGV	7	14	1	4	1	1	1	24	23	2	0	4	0	24
HiC_scaffold_33:12666381	wdr31	SGV	4	11	0	4	0	1	0	0	14	5	0	2	0	0
HiC_scaffold_33:12666394	wdr31	SGV	4	11	0	0	0	0	0	0	17	0	0	0	0	0
HiC_scaffold_33:12670109	wdr31	SGV	5	11	0	0	0	0	0	0	14	0	0	0	0	0
HiC_scaffold_33:12670176	wdr31	SGV	5	10	0	0	2	0	0	6	15	6	2	0	0	18
HiC_scaffold_33:12892230	gnaq	SGV	2	5	0	0	0	0	0	0	19	0	0	0	0	0
HiC_scaffold_33:12893080	gnaq	SGV	1	5	0	0	0	0	0	0	22	1	0	0	0	0
HiC_scaffold_43:26601424	NA	SGV	2	0	0	0	1	0	0	0	0	0	0	0	0	0
HiC_scaffold_43:26602031	NA	SGV	1	1	0	0	1	0	2	21	0	0	0	0	0	1
HiC_scaffold_43:26602094	NA	SGV	0	0	0	0	2	0	0	13	0	0	0	4	0	0
HiC_scaffold_43:26603077	NA	SGV	2	0	0	0	0	2	0	12	0	0	2	0	0	0
HiC_scaffold_43:26603083	NA	SGV	2	0	0	0	0	2	0	12	0	0	2	0	0	0
HiC_scaffold_43:26606120	NA	de novo	1	0	0	0	0	0	0	0	0	0	0	0	0	0
HiC_scaffold_43:26606125	NA	de novo	1	0	0	0	0	0	0	0	0	0	0	0	0	0
HiC_scaffold_44:37672920	NA	SGV	4	0	0	3	0	0	0	6	0	0	0	0	0	0
HiC_scaffold_44:37672992	NA	SGV	6	0	0	3	0	0	2	0	0	0	2	4	0	0
HiC_scaffold_44:37673070	NA	SGV	2	0	0	0	0	0	0	0	0	0	0	0	0	0
HiC_scaffold_44:37673124	NA	SGV	2	0	0	3	0	0	1	0	0	0	0	0	0	0
HiC_scaffold_44:37674761	NA	SGV	3	0	0	3	0	0	2	0	0	0	0	0	0	0
HiC_scaffold_44:37674976	NA	SGV	3	0	0	3	0	0	0	0	0	0	0	0	0	0
HiC_scaffold_44:37675863	NA	SGV	5	0	0	3	0	0	0	0	0	0	0	0	0	0
HiC_scaffold_44:37693328	NA	SGV	4	0	0	0	0	0	0	0	0	0	0	0	0	0
HiC_scaffold_47:14782939	olfm1	SGV	32	17	2	3	1	2	1	12	14	10	1	2	1	12
HiC_scaffold_52:13758756	NA	SGV	0	0	0	0	0	0	0	0	0	0	1	0	0	0
HiC_scaffold_52:16733488	NA	SGV	14	7	0	1	1	0	0	0	0	11	0	0	0	0

HiC_scaffold_52:16737799	NA	SGV	14	7	0	1	1	3	0	15	13	10	0	4	0	20
HiC_scaffold_52:16737807	NA	SGV	14	6	0	0	1	2	0	0	0	8	0	0	0	0
HiC_scaffold_52:16737812	NA	SGV	14	7	0	1	1	3	0	15	13	8	0	4	0	0
HiC_scaffold_52:16737828	NA	SGV	14	6	0	0	1	0	0	0	0	8	0	0	0	0
HiC_scaffold_52:21897888	NA	de novo	1	0	0	0	0	0	0	0	0	0	0	0	0	0
HiC_scaffold_53:18968932	twist1	de novo	1	0	0	0	0	0	0	0	0	0	0	0	0	0
HiC_scaffold_8238:1360	NA	de novo	0	0	0	0	0	0	0	0	0	0	0	0	0	0
HiC_scaffold_9898:5097	NA	SGV	6	1	0	0	0	0	0	0	0	0	0	0	0	0
HiC_scaffold_9898:5770	NA	SGV	5	0	0	4	1	4	0	0	20	0	2	0	0	22
HiC_scaffold_10137:3210	NA	SGV	4	0	0	0	0	0	0	10	1	0	0	0	0	2
HiC_scaffold_10137:5068	NA	SGV	5	0	0	0	0	0	0	9	3	0	2	0	0	0

Table S8. The distribution of unique adaptive alleles to the intermediate scale-eater *C. sp.* 'wide-mouth' across the Caribbean. Numbers represent the copies of the *C. desquamator* allele present in outgroup populations collected from San Salvador Island (SS), Rum Cay (RC), Cat Island (CT), Exumas Islands (EX), Long Island (LG), Mayaguana (MG), New Providence Island (NPI) in the Bahamas. Lagunas Bavaros (BAV) and Etang Saumautre (ETA) in Dominican Republic, Sarasota Estuary in Florida (FL), Curacoa (CUR), Caicos Island (CAI) and Isla Margarita (MAR) in Venezuela. The spatial distribution of each allele is summarized into three categories: standing genetic variation (SGV); de novo to San Salvador Island populations (de novo) and alleles found in introgressed regions (Intro). Putative donor populations for introgressed regions are indicated by the same location codes.

			C. variegatus								*C. laciniatus*	*C. higuey*	*C. bondi*	*C. dearborni*		
Allele	Gene	Distribution	SS	RC	CT	EX	LG	MG	FL	NC	NP	BAV	ETA	CUR	CAI	MAR
HiC_scaffold_1:34344664	smarce1	Intro NC	1	0	0	4	1	2	2	24	2	11	2	4	0	20
HiC_scaffold_1:34346071	smarce1	Intro NC	1	0	0	0	0	1	0	24	0	0	0	0	0	0
HiC_scaffold_1:34346073	smarce1	Intro NC	1	0	0	0	0	1	0	24	0	0	0	0	0	0
HiC_scaffold_5:20660128	NA	SGV	1	0	0	0	0	0	0	0	0	0	0	0	0	0
HiC_scaffold_5:20941055	NA	SGV	1	1	0	3	0	2	0	0	5	6	0	0	1	0
HiC_scaffold_5:20941063	NA	SGV	1	1	0	3	0	2	0	0	5	6	0	0	1	0
HiC_scaffold_5:20945163	NA	SGV	3	1	0	3	0	2	0	0	8	11	0	0	1	2
HiC_scaffold_5:20945417	NA	SGV	1	1	0	4	0	2	0	0	5	7	0	0	1	0
HiC_scaffold_5:20945569	NA	SGV	1	1	0	3	0	3	0	0	4	9	0	4	2	0
HiC_scaffold_5:20964282	NA	SGV	1	1	0	2	1	2	0	15	5	9	0	0	0	0
HiC_scaffold_5:20964283	NA	SGV	1	1	0	2	1	2	0	15	5	9	0	0	0	0
HiC_scaffold_5:20968573	NA	SGV	0	1	0	4	1	3	0	0	7	7	0	4	0	0
HiC_scaffold_5:20969017	NA	SGV	1	1	0	4	1	4	0	0	7	7	0	0	0	0
HiC_scaffold_5:20969195	NA	SGV	2	1	0	2	1	2	0	0	7	6	0	0	0	0
HiC_scaffold_5:20969509	NA	SGV	0	1	0	3	1	2	0	16	6	6	0	4	0	0
HiC_scaffold_5:20969704	NA	SGV	1	2	0	3	1	4	0	0	6	6	0	0	0	0
HiC_scaffold_5:20969759	NA	SGV	1	1	0	3	1	2	0	10	5	5	0	4	0	0
HiC_scaffold_5:20970496	NA	SGV	0	2	0	3	0	2	0	14	6	6	0	4	1	0

HiC_scaffold_5:20971701	NA	SGV	1	1	0	3	1	4	0	0	8	5	0	0	0	1
HiC_scaffold_5:20973584	NA	SGV	1	2	0	4	1	4	0	0	9	6	0	0	0	0
HiC_scaffold_5:20974017	NA	SGV	0	1	0	1	0	4	2	3	8	6	0	0	0	0
HiC_scaffold_5:20977516	NA	SGV	1	2	0	2	1	4	0	0	4	7	0	0	0	0
HiC_scaffold_5:20979898	NA	SGV	0	2	2	3	1	2	1	0	6	5	0	0	1	0
HiC_scaffold_5:20986384	NA	SGV	0	2	0	2	1	2	0	0	8	0	0	0	2	0
HiC_scaffold_5:20988210	NA	SGV	0	3	0	2	1	2	0	0	7	0	0	0	0	0
HiC_scaffold_5:20988292	NA	SGV	0	4	0	3	1	4	0	0	6	1	0	0	2	0
HiC_scaffold_5:20988346	NA	SGV	0	5	0	3	0	2	0	0	8	0	0	0	1	0
HiC_scaffold_11:21559293	NA	SGV	1	22	0	4	0	2	1	0	13	2	0	0	0	0
HiC_scaffold_11:21571679	NA	SGV	1	17	0	1	1	2	1	5	13	2	0	0	1	0
HiC_scaffold_11:21571720	NA	SGV	1	22	0	1	1	0	0	5	0	3	0	0	0	0
HiC_scaffold_11:21571742	NA	SGV	1	22	0	1	1	0	0	5	0	3	0	0	1	0
HiC_scaffold_24:15255378	NA	Intro NC	0	0	0	0	1	0	1	23	0	0	0	0	0	24
HiC_scaffold_43:19331844	semag4	SGV	7	0	2	0	0	0	0	0	0	0	0	0	0	0
HiC_scaffold_44:14922408	NA	SGV	4	11	1	0	0	0	0	0	0	1	1	0	0	0
HiC_scaffold_44:14922640	NA	SGV	2	9	0	0	0	0	0	0	0	0	2	0	2	0
HiC_scaffold_44:14922976	NA	SGV	2	11	0	0	0	0	0	0	3	1	0	4	2	0
HiC_scaffold_52:9589311	slitrk5	Intro BAV	1	0	0	2	1	0	0	21	7	20	0	0	0	20
HiC_scaffold_52:9589319	slitrk5	Intro BAV	1	0	0	2	1	0	0	21	7	20	0	0	0	20
HiC_scaffold_52:9592382	slitrk5	Intro BAV	0	1	0	0	2	0	0	2	1	1	2	0	0	26
HiC_scaffold_5743:1358	NA	SGV	0	4	0	2	1	2	0	0	9	0	0	0	2	0
HiC_scaffold_5789:9634	NA	SGV	0	3	0	2	0	2	0	0	7	0	0	0	0	0
HiC_scaffold_5789:10010	NA	de novo	0	0	0	0	0	0	0	0	0	0	0	0	0	0
HiC_scaffold_5789:11490	NA	SGV	0	4	0	2	2	2	0	0	5	0	0	0	0	0
HiC_scaffold_5789:12093	NA	SGV	0	4	0	2	0	2	0	0	7	0	0	0	1	0
HiC_scaffold_13976:3987	NA	SGV	0	0	0	2	1	2	0	0	5	0	0	0	2	0

Table S10. Assessment of introgression signatures between population in Osprey Lake, San Salvador Island in regions that contain putative adaptive alleles unique to *C. sp*. wide-mouth, unique to *C. desquamator* and shared between both scale-eating populations. Introgressed regions are highlighted in bold. Introgressed regions were those that had a relative node depth (RND) value between two populations. RND statistics were calculated in 10-kb windows between *C. desquamator* and *C.* sp. 'wide-mouth' (des vs wid) and *C. desquamator* and *C.* sp. 'wide-mouth' (des vs var), with the latter test serving as a control for recent and ongoing gene exchange due to sympatric overlap in the lake that is not relevant for scale-eating adaptation. Adaptive allele, gene names, and RND values that are bolded represent candidate introgressed regions with adaptive alleles in them. Candidate introgressed regions were regions that fell below a significance threshold value (RND < 0.28) based on the lower 5th percentile of RND values calculated from simulations of two populations that experience no gene flow during divergence. The number of exons present in each 10-kb window is also included in this table.

Adaptive allele	Gene	RND des vs wid	RND des v var	Introgressed Region	Number of exons
Unique *C. sp.* 'wide-mouth' adaptive alleles					
HiC_scaffold_5:20941063	NA	1.48663438	1.11402561	20940001-20950000	0
HiC_scaffold_5:20964282	NA	1.53346729	1.28660246	20960001-20970000	0
HiC_scaffold_5:20970496	NA	1.52647461	1.29854143	20970001-20980000	0
HiC_scaffold_5:20986384	NA	1.34806367	0.90675156	20980001-20990000	0
HiC_scaffold_11:21559293	NA	1.24490382	0.78312989	21550001-21560000	0
HiC_scaffold_11:21571679	NA	1.74507532	0.87626263	21570001-21580000	0
HiC_scaffold_24:15255378	NA	0.60467498	0.54123148	15250001-15260000	0
HiC_scaffold_43:19331844	Sema4g	1.18018089	0.94111453	19330001-19340000	0
HiC_scaffold_44:14922408	atp8a1	1.60947983	0.97866515	14920001-14930000	0
HiC_scaffold_52:9589311	Slitrk5	1.47013193	1.14495019	9580001-9590000	0
C. desquamator and *C. sp.* 'wide-mouth' adaptive alleles					
HiC_scaffold_6:907101	**NA**	**0.01764957**	0.304609	900001-910000	0
HiC_scaffold_6:923182	**NA**	**0.03560806**	0.41432253	920001-930000	0
HiC_scaffold_6:934123	**NA**	**0.01180417**	0.29362216	930001-940000	0
HiC_scaffold_6:941221	**NA**	**0.03304484**	0.52001902	940001-950000	0
HiC_scaffold_6:951028	NA	0.33613727	0.34973512	950001-960000	**0**
HiC_scaffold_9:31763392	**DAAM2**	**0.0241461**	0.07809267	31750001-31760000	4
HiC_scaffold_9:31763392	**DAAM2**	**0.06961605**	0.83262139	31760001-31770000	5
HiC_scaffold_24:5390450	**Usp50**	**0.0833701**	0.90291628	5390001-5400000	0
HiC_scaffold_43:27385265	NA	1.30551148	2.06180406	27380001-27390000	0
HiC_scaffold_44:14963656	**ATP8A1**	**0.13299985**	1.26725109	14960001-14970000	2
HiC_scaffold_60:1798737	**ZNF214**	**0.08551992**	1.60256069	1790001-1800000	1
Unique *C. desquamator* adaptive alleles					
HiC_scaffold_5:6082906	NA	1.61515973	1.39739811	6080001-6090000	0

HiC_scaffold_11:10913509	cd226	0.25420561	1.04030645	10910001-10920000	2
HiC_scaffold_11:10948683	cd226	1.11058873	0.75472269	10940001-10950000	2
HiC_scaffold_21:10041965	cntn5	1.51795767	1.5514369	10040001-10050000	3
HiC_scaffold_21:31993865	SLC35E1	4.35	4.81944444	31990001-32000000	3
HiC_scaffold_24:7401348	NA	0.62028279	0.62001004	7400001-7410000	0
HiC_scaffold_33:12638684	bri3bp	1.87610126	1.80369995	12630001-12640000	1
HiC_scaffold_33:12641481	bri3bp	2.05234267	1.93074547	12640001-12650000	2
HiC_scaffold_33:12656933	wdr31	2.08801498	1.91525971	12650001-12660000	8
HiC_scaffold_33:12664211	wdr31	2.04327131	1.88029401	12660001-12670000	0
HiC_scaffold_33:12670176	wdr31	2	1.87301587	12670001-12680000	0
HiC_scaffold_33:12891903	gnaq	1.19586694	1.21312932	12890001-12900000	0
HiC_scaffold_43:26606125	NA	1.7651081	1.60827345	26600001-26610000	0
HiC_scaffold_44:37673070	NA	1.3785441	1.29438333	37670001-37680000	0
HiC_scaffold_44:37693328	NA	1.2303647	1.09117949	37690001-37700000	0
HiC_scaffold_47:14782939	**olfm1**	**0.19804878**	0.7647762	14780001-14790000	1
HiC_scaffold_52:16737828	NA	1.19630072	1.25855174	16730001-16740000	0
HiC_scaffold_53:18968932	**twist1**	**0.05856256**	0.54868019	18960001-18970000	2

Table S11.Genome-wide signatures of differential introgression between *C.* sp. 'wide-mouth' and *C. desquamator*. f_4-statistics, standard error, z-scores and p-values for genome-wide test with 5 outgroup generalist populations from across the Caribbean. *C. artifrons* used as outgroup species in which we expect minimal gene flow.
–

P1	P2	P3	Outgroup	f4	Standard Error	Z-score	P-value
wid	des	Dominican Republic	*Artifrons*	-0.0003	0.00017	-1.713	0.082
wid	des	Rum Cay	*Artifrons*	0.0004	0.00018	1.904	0.05
wid	des	North Carolina	*Artifrons*	0.0005	0.00018	2.85	0.0044
wid	des	Venezuela	*Artifrons*	0.0001	0.00014	0.788	0.43
wid	des	New Providence Island	*Artifrons*	0.0003	0.00015	1.813	0.069

Dissertation Conclusion

Together the chapters of my dissertation highlight the complexity of hybridization's impact on the speciation and ecological diversification that occurs during adaptive radiation. The similar historical signatures of hybridization detected in two different radiations of fish, yet the quite different outcomes hybridization had on the local genomic scale in each suggests that the field's current standard of simply detecting hybridization across radiations is not enough to draw strong conclusions about its universal role in triggering adaptive radiation. My research has taken initial steps beyond documenting evidence of hybridization in both *Cyprinodon* pupfish and Barombi Mbo cichlid radiations by exploring the role hybridization has played in diversification processes in each.

In Barombi Mbo cichlid radiation, evidence of historical hybridization events from secondary contact between the radiation and outgroup riverine populations raised doubt about its status as one of the best case studies of sympatric speciation in nature. However, genome-wide signatures of hybridization tells us little about how gene flow potentially influenced sympatric diversification in this system. From my investigations of the localized genomic impact of these detected hybridization events, I discovered that introgressed genomic regions were not strongly divergence between species of the radiation and it is not clear that the introgressed genetic diversity in these regions contributed functionally to the ecological, sexual, and morphological diversity found in the radiation. Although more in depth functional studies will be necessary to truly determine the impact of introgressed variation in this system, the uncertainty about its impact raised here highlights the necessity such investigations of how detected hybridization has impacted genomic divergence in any system.

In the *Cyprinodon* pupfish system, hybridization likely brought together new combinations of largely ancient set of alleles that have been maintained within different pools of standing variation in Caribbean and mainland generalist populations to San Salvador Island. Some of this genetic variation that appears introgressed into the trophic specialist background appears relevant for the notable craniofacial divergence in this radiation based on genome-wide association studies, quantitative trait loci analyses and gene annotations in model organisms. Much of this introgressed variation also appears to be under divergent selection between species within the radiation. The radiation also contains stronger signatures of introgression than found in populations on other islands that did not radiate, lending support to the hybrid origins hypothesis. However, these other populations did experience introgression from some of the same sources and in some cases share 100% of the same genetic variation, highlighting that our understanding of the origins of adaptive radiation in this system remains incomplete. Future research endeavors across multiple systems will be necessary to determine why some lineages respond to hybridization with adaptive radiations and others do not.

The *Cyprinodon* pupfish radiation of trophic specialists on San Salvador Island has been additionally useful for testing out other hypotheses about adaptive radiation and evolutionary novelty because of its young age. At ~10-20 thousand years old, we can detect selection and genetic divergence signatures likely related to the beginning of the radiation. However, I did find added nuance to this stages hypothesis in that stages aren't as temporally distinct across different

axes of traits as proposed. Such temporally overlapping patterns were previously undetectable from standard phylogenetic approaches to this question but were readily detected from the population genomic approach I utilized to estimate the timing of selection. Additionally, my final chapter highlights that the major ecological transitions that occur during the occupation of a novel ecological niche is highly contingent on the genomic background present. The two scale-eating specialist of the San Salvador Island radiation are genetically, morphologically, and ecologically diverged from each other despite occupying a similar novel niche of scale-eating and being sister species. While they have utilized some of the same adaptive alleles in their adaptation to scale-eating, they predominantly use different pools of adaptive variation from each other, often coming from entirely different spatial sources (e.g. de novo vs. introgressed). Despite having access to the same genetic variation, as these populations live in sympatry and interbreed, the intermediate scale-eater in this system did not utilize the suite of de novo alleles that the more specialized scale-eater did. This indicates that complex interactions at the genetic level underly some of these ecological diversification outcomes and that such interactions can lead to very diverse outcomes even among closely related species. This contingency discovered from dissecting the genetic divergence between generalists and scale-eaters in this radiation makes it easier to understand why hybridization that brought in new combinations of genetic variation may not have played a similar role across outgroup pupfish populations as it did in the radiation. The results from my chapters provide clear steps forward for pinpointing the conditions in which hybridization may generate adaptive radiations that include exploring the epistatic interactions among alleles and other factors that contribute to genome evolution variation across radiating and non-radiating lineages.

In Chapter 4, I also provide evidence to support the three stages hypothesis of adaptive radiation. Temporal stages of adaptation observed in this nascent radiation are consistent with selection on behavioral divergence first. Adaptive divergence in trophic morphology occurred next, followed by a final stage of refinement including a non-synonymous substitution in the scale-eaters within a craniofacial transcription factor. Additionally, knowledge of the source of each adaptive allele provided me a unique look at the spatial dynamics of alleles involved in adaptive divergence. We found that most adaptive alleles contributing to the major axes of ecological and sexual diversification in this radiation existed in Caribbean generalist populations long before the trophic specialist species on San Salvador Island diverged. This genetic variation is distributed across two orders of magnitude larger spatial and temporal scales than the 10 kya radiation endemic to a single 20 km island. 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.

In conclusion, my chapters highlight the utility and necessity of including closely related outgroups as controls in testing hypotheses about the mechanisms underlying adaptive radiation. Rare radiations that are constrained to a small range but are spatially nested with a widespread clade of lineages that haven't diversified provide a powerful tool for empirically assessing hypotheses about adaptive radiations. The coincidence of hybridization, ecological opportunity, and sexual selection appear to be the best predictors of adaptive radiation so far. Most adaptive radiations, including stickleback, African cichlids, Lake Tana barbs, *Anolis* lizards, *Heliconius* butterflies, Hawaiian tetragnathids, and *Brocchinia* bromeliads, share similar patterns of spatial nesting within a widespread clade and intermediate levels of population structure and admixture

and one could expect similar dynamics occurring in these radiations to what we observed in the pupfish and crater lake cichlid systems. Research into the broader spatiotemporal landscape of radiations can provide insights about longstanding hypotheses of adaptive radiation and their contributions to global patterns of biodiversity.

References

Abbott, R., D. Albach, S. Ansell, J. W. Arntzen, S. J. E. Baird, N. Bierne, J. Boughman, A. Brelsford, C. A. Buerkle, R. Buggs, R. K. Butlin, U. Dieckmann, F. Eroukhmanoff, A. Grill, S. H. Cahan, J. S. Hermansen, G. Hewitt, A. G. Hudson, C. Jiggins, J. Jones, B. Keller, T. Marczewski, J. Mallet, P. Martinez-Rodriguez, M. Möst, S. Mullen, R. Nichols, A. W. Nolte, C. Parisod, K. Pfennig, A. M. Rice, M. G. Ritchie, B. Seifert, C. M. Smadja, R. Stelkens, J. M. Szymura, R. Väinölä, J. B. W. Wolf, and D. Zinner. 2013. Hybridization and speciation. J. Evol. Biol. 26:229–246.

Abi-Rached, L., M. J. Jobin, S. Kulkarni, A. McWhinnie, K. Dalva, L. Gragert, F. Babrzadeh, B. Gharizadeh, M. Luo, F. A. Plummer, J. Kimani, M. Carrington, D. Middleton, R. Rajalingam, M. Beksac, S. G. E. Marsh, M. Maiers, L. A. Guethlein, S. Tavoularis, A.-M. Little, R. E. Green, P. J. Norman, and P. Parham. 2011. The Shaping of Modern Human Immune Systems by Multiregional Admixture with Archaic Humans. Science (80-.). 334:89–94.

Alachiotis, N., A. Stamatakis, and P. Pavlidis. 2012. OmegaPlus: a scalable tool for rapid detection of selective sweeps in whole-genome datasets. Bioinformatics 28:2274–2275.

Alcaide, M., E. S. C. Scordato, T. D. Price, and D. E. Irwin. 2014. Genomic divergence in a ring species complex. Nature 511:83–85. Nature Publishing Group.

Alexander, D. H., J. Novembre, and K. Lange. 2009. Fast model-based estimation of ancestry in unrelated individuals. Genome Res. 19:1655–1664. Cold Spring Harbor Lab.

Alfaro, M. E., F. Santini, C. Brock, H. Alamillo, A. Dornburg, D. L. Rabosky, G. Carnevale, and L. J. Harmon. 2009. Nine exceptional radiations plus high turnover explain species diversity in jawed vertebrates. Proc. Natl. Acad. Sci. U. S. A. 106:13410–4.

Almen, M. S., S. Lamichhaney, J. Berglund, B. R. Grant, P. R. Grant, M. T. Webster, and L. Andersson. 2016. Adaptive radiation of Darwin's finches revisited using whole genome sequencing. BioEssays 38:14–20.

Anderson, M. E., J. Runesson, I. Saar, Ü. Langel, and J. K. Robinson. 2013. Galanin, through GalR1 but not GalR2 receptors, decreases motivation at times of high appetitive behavior. Behav. Brain Res. 239:90–93. Elsevier B.V.

Arnold, B. J., B. Lahner, J. M. DaCosta, C. M. Weisman, J. D. Hollister, D. E. Salt, K. Bomblies, and L. Yant. 2016. Borrowed alleles and convergence in serpentine adaptation. Proc. Natl. Acad. Sci. 113:8320–8325.

Arnold, S. J., M. E. Pfrender, and J. A. G. 2001. The Apaptive Landscape as a conceptual bridge between mico- and macroevolution. Genetica 112–113:9–32.

Azzouzi, N., F. Barloy-Hubler, and F. Galibert. 2014. Inventory of the cichlid olfactory receptor gene repertoires: identification of olfactory genes with more than one coding exon. BMC Genomics 15:586.

Balsa-Canto, E., D. Henriques, A. Gabor, and J. R. Banga. 2016. AMIGO2, a toolbox for dynamic modeling, optimization and control in systems biology. Bioinformatics 32:1–2.

Barghi, N., J. Hermisson, and C. Schlötterer. 2020. Polygenic adaptation: a unifying framework to understand positive selection. Nat. Rev. Genet. 21:769–781. Springer US.

Barluenga, M., and A. Meyer. 2010. Phylogeography, colonization and population history of the Midas cichlid species complex (Amphilophus spp.) in the Nicaraguan crater lakes. BMC Evol. Biol. 10:326. BioMed Central Ltd.

Barluenga, M., K. N. Stölting, W. Salzburger, M. Muschick, and A. Meyer. 2006. Sympatric

speciation in Nicaraguan crater lake cichlid fish. Nature 439:719–723.
Bateman, A. 2019. UniProt: A worldwide hub of protein knowledge. Nucleic Acids Res. 47:D506–D515. Oxford University Press.
Beleza, S., A. M. Santos, B. McEvoy, I. Alves, C. Martinho, E. Cameron, M. D. Shriver, E. J. Parra, and J. Rocha. 2013. The Timing of Pigmentation Lightening in Europeans. Mol. Biol. Evol. 30:24–35.
Bell, G. 2013. Evolutionary rescue and the limits of adaptation. Philos. Trans. R. Soc. B Biol. Sci. 368:20120080. Royal Society.
Berk, M., S. Y. Desai, H. C. Heyman, and C. Colmenares. 1997. Mice lacking the ski proto-oncogene have defects in neurulation, craniofacial patterning, and skeletal muscle development. Genes Dev. 11:2029–2039.
Berner, D., and W. Salzburger. 2015. The genomics of organismal diversification illuminated by adaptive radiations. Trends Genet. 31:491–499. Elsevier Ltd.
Blais, J., M. Plenderleith, C. Rico, M. I. Taylor, O. Seehausen, C. van Oosterhout, and G. F. Turner. 2009. Assortative mating among Lake Malawi cichlid fish populations is not simply predictable from male nuptial colour. BMC Evol. Biol. 9:53.
Blount, Z. D. 2016. A case study in evolutionary contingency. Stud. Hist. Philos. Sci. Part C Stud. Hist. Philos. Biol. Biomed. Sci. 58:82–92.
Blount, Z. D., J. E. Barrick, C. J. Davidson, and R. E. Lenski. 2012. Genomic analysis of a key innovation in an experimental *Escherichia coli* population. Nature 489:513–518.
Blount, Z. D., C. Z. Borland, and R. E. Lenski. 2008. Historical contingency and the evolution of a key innovation in an experimental population of *Escherichia coli*. Proc. Natl. Acad. Sci. U. S. A. 105:7899–906.
Blount, Z. D., R. E. Lenski, and J. B. Losos. 2018. Contingency and determinism in evolution: Replaying life's tape. Science (80-.). 362.
Blount, Z. D., R. Maddamsetti, N. A. Grant, S. T. Ahmed, T. Jagdish, J. A. Baxter, B. A. Sommerfeld, A. Tillman, J. Moore, J. L. Slonczewski, J. E. Barrick, and R. E. Lenski. 2020. Genomic and phenotypic evolution of escherichia coli in a novel citrate-only resource environment. Elife 9:1–64.
Bolnick, D. I. 2006. Multi-species outcomes in a common model of sympatric speciation. J. Theor. Biol. 241:734–744.
Bolnick, D. I., and M. Doebeli. 2003. Sexual dimorphism: two sides of the same ecological coin. Evolution (N. Y). 57:2433–2449.
Bolnick, D. I., and B. M. Fitzpatrick. 2007. Sympatric speciation: models and empirical evidence. Annu. Rev. Ecol. Evol. Syst. 38:459–487.
Bouckaert, R. R. 2010. DensiTree: making sense of sets of phylogenetic trees. Bioinformatics 26:1372–1373.
Boyle, E. A., Y. I. Li, and J. K. Pritchard. 2017. An Expanded View of Complex Traits: From Polygenic to Omnigenic. Cell 169:1177–1186. Elsevier.
Bradbury, P. J., Z. Zhang, D. E. Kroon, T. M. Casstevens, Y. Ramdoss, and E. S. Buckler. 2007. TASSEL: Software for association mapping of complex traits in diverse samples. Bioinformatics 23:2633–2635.
Brawand, D., C. E. Wagner, Y. I. Li, M. Malinsky, I. Keller, S. Fan, O. Simakov, A. Y. Ng, Z. W. Lim, E. Bezault, Turner-Maier, J. J. Johnson, R. Alcazar, H. J. Noh, P. Russell, B. Aken, J. Alföldi, C. Amemiya, N. Azzouzi, J.-F. Baroiller, F. Barloy-Hubler, A. Berlin, R. Bloomquist, K. L. Carleton, M. a. Conte, H. D'Cotta, O. Eshel, L. Gaffney, F. Galibert, H.

F. Gante, S. Gnerre, L. Greuter, R. Guyon, N. S. Haddad, W. Haerty, R. M. Harris, H. a. Hofmann, T. Hourlier, G. Hulata, D. B. Jaffe, M. Lara, L. A.P., I. MacCallum, S. Mwaiko, M. Nikaido, H. Nishihara, C. Ozouf-Costaz, D. J. Penman, D. Przybylski, M. Rakotomanga, S. C. P. Renn, F. J. Ribeiro, M. Ron, W. Salzburger, L. Sanchez-Pulido, M. E. Santos, S. Searle, T. Sharpe, R. Swofford, F. J. Tan, L. Williams, S. Young, S. Yin, N. Okada, T. D. Kocher, E. a. Miska, E. S. Lander, B. Venkatesh, R. D. Fernald, A. Meyer, C. P. Ponting, J. T. Streelman, K. Lindblad-Toh, O. Seehausen, and F. Di Palma. 2014. The genomic substrate for adaptive radiation in African cichlid fish. Nature 513:375–381.

Brewer, S., W. Feng, J. Huang, S. Sullivan, and T. Williams. 2004. Wnt1-Cre-mediated deletion of AP-2α causes multiple neural crest-related defects. Dev. Biol. 267:135–152.

Bronson, R. T., H. Sweet, C. A. Spencer, and M. T. Davisson. 1992. Genetic and age related models of neurodegeneration in mice: dystrophic axons. 71–83.

Bryant, D., R. Bouckaert, J. Felsenstein, N. A. Rosenberg, and A. RoyChoudhury. 2012. Inferring Species Trees Directly from Biallelic Genetic Markers: Bypassing Gene Trees in a Full Coalescent Analysis. Mol. Biol. Evol. 29:1917–1932.

Bürger, R., and K. A. Schneider. 2006. Intraspecific Competitive Divergence and Convergence under Assortative Mating. Am. Nat. 167:190–205. The University of Chicago Press.

Burns, K. J., S. J. Hackett, and N. K. Klein. 2002. Phylogenetic relationships and morphological diversity in Darwin's finches and their relatives. Evolution 56:1240–1252.

Bush, G. L. 1975. Modes of Animal Speciation. Annu. Rev. Ecol. Syst. 6:339–364.

Calsbeek, R., and D. J. Irschick. 2007. The quick and the dead: Correlational selection on morphology, performance, and habitat use in island lizards. Evolution (N. Y). 61:2493–2503.

Camacho, C., G. Coulouris, V. Avagyan, N. Ma, J. Papadopoulos, K. Bealer, and T. L. Madden. 2009. BLAST+: Architecture and applications. BMC Bioinformatics 10:1–9.

Campbell, C. R., J. W. Poelstra, and A. D. Yoder. 2018. What is Speciation Genomics? The roles of ecology, gene flow, and genomic architecture in the formation of species. Biol. J. Linn. Soc. bly063–bly063.

Cantarel, B. L., I. Korf, S. M. C. Robb, G. Parra, E. Ross, B. Moore, C. Holt, A. S. Alvarado, and M. Yandell. 2008. MAKER: An easy-to-use annotation pipeline designed for emerging model organism genomes. Genome Res. 18:188–196.

Cattell, R. B. 1966. The scree test for the number of factors. Multivariate Behav. Res. 1:245–276.

Chatterjee, S., and N. Ahituv. 2017. Gene Regulatory Elements, Major Drivers of Human Disease. Annu. Rev. Genomics Hum. Genet. 18:45–63. Annual Reviews.

Chen, H., J. Hey, and M. Slatkin. 2015. A hidden Markov model for investigating recent positive selection through haplotype structure. Theor. Popul. Biol. 99:18–30. Elsevier Inc.

Chevin, L. M., and F. Hospital. 2008. Selective sweep at a quantitative trait locus in the presence of background genetic variation. Genetics 180:1645–1660.

Choquette, A. C., L. Bouchard, V. Drapeau, S. Lemieux, A. Tremblay, C. Bouchard, M.-C. Vohl, and L. Pérusse. 2012. Association between olfactory receptor genes, eating behavior traits and adiposity: Results from the Quebec Family Study. Physiol. Behav. 105:772–776.

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

Clarkson, C. S., D. Weetman, J. Essandoh, A. E. Yawson, G. Maslen, M. Manske, S. G. Field,

M. Webster, T. Antão, B. MacInnis, D. Kwiatkowski, and M. J. Donnelly. 2014. Adaptive introgression between Anopheles sibling species eliminates a major genomic island but not reproductive isolation. Nat. Commun. 5:4248. Nature Publishing Group.

Clayton-Smith, J., J. O'Sullivan, S. Daly, S. Bhaskar, R. Day, B. Anderson, A. K. Voss, T. Thomas, L. G. Biesecker, P. Smith, A. Fryer, K. E. Chandler, B. Kerr, M. Tassabehji, S. A. Lynch, M. Krajewska-Walasek, S. McKee, J. Smith, E. Sweeney, S. Mansour, S. Mohammed, D. Donnai, and G. Black. 2011. Whole-exome-sequencing identifies mutations in histone acetyltransferase gene KAT6B in individuals with the say-barber-biesecker variant of Ohdo syndrome. Am. J. Hum. Genet. 89:675–681. The American Society of Human Genetics.

Cleves, P. A., J. C. Hart, R. M. Agoglia, M. T. Jimenez, P. A. Erickson, L. Gai, and C. T. Miller. 2018. An intronic enhancer of Bmp6 underlies evolved tooth gain in sticklebacks. PLOS Genet. 14:e1007449. Public Library of Science.

Colmenares, C., H. a Heilstedt, L. G. Shaffer, S. Schwartz, M. Berk, J. C. Murray, and E. Stavnezer. 2002. Loss of the SKI proto-oncogene in individuals affected with 1p36 deletion syndrome is predicted by strain-dependent defects in Ski-/- mice. Nat. Genet. 30:106–9.

Comeault, A. A., and D. R. Matute. 2018. Genetic divergence and the number of hybridizing species affect the path to homoploid hybrid speciation. Proc. Natl. Acad. Sci. 115:9761–9766.

Coop, G., K. Bullaughey, F. Luca, and M. Przeworski. 2008. The Timing of Selection at the Human FOXP2 Gene. Mol. Biol. Evol. 25:1257–1259.

Cotto, O., and M. R. Servedio. 2017. The Roles of Sexual and Viability Selection in the Evolution of Incomplete Reproductive Isolation: From Allopatry to Sympatry. Am. Nat. 190:680–693. The University of Chicago Press.

Coyne, J. A., and A. Orr. 2004a. Speciation. Sutherland, MA, Sunderland, MA.

Coyne, J. A., and A. H. Orr. 1989. Patterns of speciation in Drosophila. Evolution (N. Y). 43:362–381. Wiley/Blackwell (10.1111).

Coyne, J. A., and H. A. Orr. 2004b. Speciation. Sutherland, MA.

Cruickshank, T. E., and M. W. Hahn. 2014. Reanalysis suggests that genomic islands of speciation are due to reduced diversity, not reduced gene flow. Mol. Ecol. 23:3133–3157.

Crump, J. G., M. E. Swartz, and C. B. Kimmel. 2004. An Integrin-Dependent Role of Pouch Endoderm in Hyoid Cartilage Development. 2.

Cullinane, A. R., A. A. Schäffer, and M. Huizing. 2013. The BEACH Is Hot: A LYST of Emerging Roles for BEACH-Domain Containing Proteins in Human Disease. Traffic 14:749–766.

Cutter, A. D., and B. A. Payseur. 2013. Genomic signatures of selection at linked sites: unifying the disparity among species. Nat. Rev. Genet. 14:262–274. Nature Publishing Group.

Danecek, P., 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, and R. Durbin. 2011a. The variant call format and VCFtools. Bioinformatics 27:2156–2158.

Danecek, P., 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, and 1000 Genomes Project Analysis Group. 2011b. The variant call format and VCFtools. Bioinformatics 27:2156–2158.

Danley, P. D., and T. D. Kocher. 2001. Speciation in rapidly diverging systems: lessons from Lake Malawi. Mol. Ecol. 10:1075–1086. John Wiley & Sons, Ltd.

Das, A., and J. G. Crump. 2012. Bmps and Id2a act upstream of twist1 to restrict ectomesenchyme potential of the cranial neural crest. PLoS Genet. 8.

Davydov, E. V., D. L. Goode, M. Sirota, G. M. Cooper, A. Sidow, and S. Batzoglou. 2010. Identifying a high fraction of the human genome to be under selective constraint using GERP++. PLoS Comput. Biol. 6.

Delmore, K. E., J. S. Lugo Ramos, B. M. Van Doren, M. Lundberg, S. Bensch, D. E. Irwin, and M. Liedvogel. 2018. Comparative analysis examining patterns of genomic differentiation across multiple episodes of population divergence in birds. Evol. Lett. 2:76–87. Wiley-Blackwell.

DePristo, M. A., 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, and M. J. Daly. 2011. A framework for variation discovery and genotyping using next-generation DNA sequencing data. Nat. Genet. 43:491–8.

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

Dickinson, M. E., A. M. Flenniken, X. Ji, L. Teboul, M. D. Wong, J. K. White, T. F. Meehan, W. J. Weninger, H. Westerberg, H. Adissu, C. N. Baker, L. Bower, J. M. Brown, L. Brianna Caddle, F. Chiani, D. Clary, J. Cleak, M. J. Daly, J. M. Denegre, B. Doe, M. E. Dolan, S. M. Edie, H. Fuchs, V. Gailus-Durner, A. Galli, A. Gambadoro, J. Gallegos, S. Guo, N. R. Horner, C. wei Hsu, S. J. Johnson, S. Kalaga, L. C. Keith, L. Lanoue, T. N. Lawson, M. Lek, M. Mark, S. Marschall, J. Mason, M. L. McElwee, S. Newbigging, L. M. J. Nutter, K. A. Peterson, R. Ramirez-Solis, D. J. Rowland, E. Ryder, K. E. Samocha, J. R. Seavitt, M. Selloum, Z. Szoke-Kovacs, M. Tamura, A. G. Trainor, I. Tudose, S. Wakana, J. Warren, O. Wendling, D. B. West, L. Wong, A. Yoshiki, D. G. MacArthur, G. P. Tocchini-Valentini, X. Gao, P. Flicek, A. Bradley, W. C. Skarnes, M. J. Justice, H. E. Parkinson, M. Moore, S. Wells, R. E. Braun, K. L. Svenson, M. Hrabe de Angelis, Y. Herault, T. Mohun, A. M. Mallon, R. Mark Henkelman, S. D. M. Brown, D. J. Adams, K. C. Kent Lloyd, C. McKerlie, A. L. Beaudet, M. Bucan, S. A. Murray, M. McKay, B. Urban, C. Lund, E. Froeter, T. LaCasse, A. Mehalow, E. Gordon, L. R. Donahue, R. Taft, P. Kutney, S. Dion, L. Goodwin, S. Kales, R. Urban, K. Palmer, F. Pertuy, D. Bitz, B. Weber, P. Goetz-Reiner, H. Jacobs, E. Le Marchand, A. El Amri, L. El Fertak, H. Ennah, D. Ali-Hadji, A. Ayadi, M. Wattenhofer-Donze, S. Jacquot, P. André, M. C. Birling, G. Pavlovic, T. Sorg, I. Morse, F. Benso, M. E. Stewart, C. Copley, J. Harrison, S. Joynson, R. Guo, D. Qu, S. Spring, L. Yu, J. Ellegood, L. Morikawa, X. Shang, P. Feugas, A. Creighton, P. C. Penton, O. Danisment, N. Griggs, C. L. Tudor, A. L. Green, C. Icoresi Mazzeo, E. Siragher, C. Lillistone, E. Tuck, D. Gleeson, D. Sethi, T. Bayzetinova, J. Burvill, B. Habib, L. Weavers, R. Maswood, E. Miklejewska, M. Woods, E. Grau, S. Newman, C. Sinclair, E. Brown, S. Ayabe, M. Iwama, and A. Murakami. 2016. High-throughput discovery of novel developmental phenotypes. Nature 537:508–514. Nature Publishing Group.

Dieckmann, U., and M. Doebeli. 1999. On the origin of species by sympatric speciation. Nature 400:354. Macmillan Magazines Ltd.

Doebeli, M., U. Dieckmann, J. A. Metz, and D. Tautz. 2005. What we have also learned: adaptive speciation is theoretically plausible. Evolution (N. Y). 59:691–699.

Doebeli, M., and I. Ispolatov. 2010. Complexity and diversity. Science 328:494–7.

Doyle, A. J., J. J. Doyle, S. L. Bessling, S. Maragh, M. E. Lindsay, D. Schepers, E. Gillis, G.

Mortier, T. Homfray, K. Sauls, R. A. Norris, N. D. Huso, D. Leahy, D. W. Mohr, M. J. Caulfield, A. F. Scott, A. Destrée, R. C. Hennekam, P. H. Arn, C. J. Curry, L. Van Laer, A. S. McCallion, B. L. Loeys, and H. C. Dietz. 2012. Mutations in the TGF-β repressor SKI cause Shprintzen-Goldberg syndrome with aortic aneurysm. Nat. Genet. 44:1249–1254. Nature Publishing Group.

Duchêne, S., and R. Lanfear. 2015. 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.

Dudchenko, O., 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, and E. L. Aiden. 2017. De novo assembly of the Aedes aegypti genome using Hi-C yields chromosome-length scaffolds. Science (80-.). 356:92 LP – 95.

Durand, N. C., M. S. Shamim, I. Machol, S. S. P. Rao, M. H. Huntley, E. S. Lander, and E. L. Aiden. 2016. Juicer Provides a One-Click System for Analyzing Loop-Resolution Hi-C Experiments. Cell Syst. 3:95–98.

Echelle, A. A., E. W. Carson, A. F. Echelle, D. Bussche, and T. E. Dowling. 2005. Historical Biogeography of the New-World Pupfish Genus Cyprinodon (Teleostei: Cyprinodontidae). Copeia 2005:320–339.

Edmunds, R. C., B. Su, J. P. Balhoff, B. F. Eames, W. M. Dahdul, H. Lapp, J. G. Lundberg, T. J. Vision, R. A. Dunham, P. M. Mabee, and M. Westerfield. 2016. Phenoscape: Identifying candidate genes for evolutionary phenotypes. Mol. Biol. Evol. 33:13–24.

Edwards, C. D. 2001. Effect of Salinity on the Ecology of Molluscs in the Inland Saline Waters of San Salvador Island: A Experiment in Progress.

Elmer, K. R., S. Fan, H. Kusche, M. Luise Spreitzer, A. F. Kautt, P. Franchini, and A. Meyer. 2014. Parallel evolution of Nicaraguan crater lake cichlid fishes via non-parallel routes. Nat. Commun. 5:1–8. Nature Publishing Group.

Elmer, K. R., T. K. Lehtonen, A. F. Kautt, C. Harrod, and A. Meyer. 2010. Rapid sympatric ecological differentiation of crater lake cichlid fishes within historic times. BMC Biol. 8.

Engert, J. C., S. Servaes, P. Sutrave, S. H. Hughes, and A. Rosenthal. 1995. Activation of a muscle-specific enhancer by the ski proto-oncogene. Nucleic Acids Res. 23:2988–2994.

Erwin, D. H. 2021. A conceptual framework of evolutionary novelty and innovation. Biol. Rev. 96:1–15.

Erwin, D. H. 2015. Novelty and innovation in the history of life. Curr. Biol. 25:R930–R940. Elsevier Ltd.

Erwin, D. H. 2019. Prospects for a General Theory of Evolutionary Novelty. J. Comput. Biol. 26:735–744.

Evanno, G., S. Regnaut, and J. Goudet. 2005. Detecting the number of clusters of individuals using the software STRUCTURE: A simulation study. Mol. Ecol. 14:2611–2620.

Excoffier, L., I. Dupanloup, E. Huerta-Sánchez, V. C. Sousa, and M. Foll. 2013. Robust Demographic Inference from Genomic and SNP Data. PLOS Genet. 9:e1003905. Public Library of Science.

Feder, J. L., S. H. Berlocher, J. B. Roethele, H. Dambroski, J. J. Smith, W. L. Perry, V. Gavrilovic, K. E. Filchak, J. Rull, and M. Aluja. 2003a. Allopatric genetic origins for sympatric host-plant shifts and race formation in Rhagoletis. Proc. Natl. Acad. Sci. U. S. A. 100:10314–10319.

Feder, J. L., S. H. Berlocher, J. B. Roethele, H. Dambroski, J. J. Smith, W. L. Perry, V.

Gavrilovic, K. E. Filchak, J. Rull, and M. Aluja. 2003b. Allopatric genetic origins for sympatric host-plant shifts and race formation in Rhagoletis. Proc. Natl. Acad. Sci. 100:10314–10319.
Feder, J. L., S. P. Egan, and P. Nosil. 2012. The genomics of speciation-with-gene-flow. Trends Genet. 28:342–350. Elsevier Ltd.
Feder, J. L., and P. Nosil. 2009. Chromosomal inversions and species differences: when are genes affecting adaptive divergence and reproductive isolation expected to reside within inversions? Evolution (N. Y). 63:3061–3075. Wiley/Blackwell (10.1111).
Feder, J. L., X. Xie, J. Rull, S. Velez, A. Forbes, B. Leung, H. Dambroski, K. E. Filchak, and M. Aluja. 2005. Mayr, Dobzhansky, and Bush and the complexities of sympatric speciation in Rhagoletis. Proc. Natl. Acad. Sci. U. S. A. 102 Suppl:6573–80.
Felsenstein, J. 1981. Skepticism towards Santa Rosalia, or why are there so few kinds of animals? Evolution (N. Y). 35:124–138. Wiley/Blackwell (10.1111).
Feng, C., M. Pettersson, S. Lamichhaney, C. J. Rubin, N. Rafati, M. Casini, A. Folkvord, and L. Andersson. 2017. Moderate nucleotide diversity in the Atlantic herring is associated with a low mutation rate. Elife 6:1–14.
Fenster, C. B., W. S. Armbruster, P. Wilson, M. R. Dudash, and J. D. Thomson. 2004. Pollination Syndromes and Floral Specialization. Annu. Rev. Ecol. Evol. Syst. 35:375–403.
Filatov, D. A., O. G. Osborne, and A. S. T. Papadopulos. 2016. Demographic history of speciation in a Senecio altitudinal hybrid zone on Mt. Etna. Mol. Ecol. 25:2467–2481.
Fishman, L., A. Stathos, P. M. Beardsley, C. F. Williams, and J. P. Hill. 2013. Chromosomal rearrangements and the genetics of reproductive barriers in mimulus (monkey flowers). Evolution (N. Y). 67:2547–2560.
Fitch, K. R., K. A. McGowan, C. D. Van Raamsdonk, H. Fuchs, D. Lee, A. Puech, Y. Hérault, D. W. Threadgill, M. H. De Angelis, and G. S. Barsh. 2003. Genetics of dark skin in mice. Genes Dev. 17:214–228.
Fitzpatrick, B. M. 2012. Estimating ancestry and heterozygosity of hybrids using molecular markers. BMC Evol. Biol. 12:131.
Fitzpatrick, B. M., J. A. Fordyce, and S. Gavrilets. 2009. Pattern, process and geographic modes of speciation. J. Evol. Biol. 22:2342–2347.
Fitzpatrick, B. M., J. A. Fordyce, and S. Gavrilets. 2008. What, if anything, is sympatric speciation? J. Evol. Biol. 21:1452–9.
Flint, H. J., K. P. Scott, P. Louis, and S. H. Duncan. 2012. The role of the gut microbiota in nutrition and health. Nat. Rev. Gastroenterol. Hepatol. 9:577–589.
Fontaine, M. C., J. B. Pease, A. Steele, R. M. Waterhouse, D. E. Neafsey, I. V. Sharakhov, X. Jiang, A. B. Hall, F. Catteruccia, E. Kakani, S. N. Mitchell, Yi-ChiehWu, H. A. Smith, R. R. Love, and N. J. B. Mara K. La. 2015. Extensive introgression in a malaria vector species complex revealed by phylogenomics. Science (80-.). 347:1258522–1258522.
Foote, A. D. 2018. Sympatric Speciation in the Genomic Era. Trends Ecol. Evol. 33:85–95.
Ford, A. G. P., K. K. Dasmahapatra, L. R??ber, K. Gharbi, T. Cezard, and J. J. Day. 2015. High levels of interspecific gene flow in an endemic cichlid fish adaptive radiation from an extreme lake environment. Mol. Ecol. 24:3421–3440.
Fragata, I., A. Blanckaert, M. A. Dias Louro, D. A. Liberles, and C. Bank. 2019. Evolution in the light of fitness landscape theory. Trends Ecol. Evol. 34:69–82.
Fry, J. D. 2014. Mechanisms of naturally evolved ethanol resistance in Drosophila melanogaster. J. Exp. Biol. 217:3996–4003. Company of Biologists.

Fuller, Z. L., C. J. Leonard, R. E. Young, S. W. Schaeffer, and N. Phadnis. 2018. Ancestral polymorphisms explain the role of chromosomal inversions in speciation. PLOS Genet. 14:e1007526. Public Library of Science.

Gante, H. F., M. Matschiner, M. Malmstr??m, K. S. Jakobsen, S. Jentoft, and W. Salzburger. 2016. Genomics of speciation and introgression in Princess cichlid fishes from Lake Tanganyika. Mol. Ecol., doi: 10.1111/mec.13767.

Garner, A. G., B. E. Goulet, M. C. Farnitano, Y. F. Molina-Henao, and R. Hopkins. 2018. Genomic signatures of reinforcement. Genes (Basel). 9.

Garrigan, D., and A. Geneva. 2014. msmove: A modified version of Hudson's coalescent simulator ms allowing for finer control and tracking of migrant genealogies.

Garrigan, D., S. B. Kingan, A. J. Geneva, P. Andolfatto, A. G. Clark, K. R. Thornton, and D. C. Presgraves. 2012. Genome sequencing reveals complex speciation in the *Drosophila simulans* clade. Genome Res. 22:1499–1511.

Gavrilets, S. 2004. Fitness landscape and origin of species. Princeton University Press, Princeton, NJ.

Gavrilets, S. 2014. Models of speciation: where are we now? J. Hered. 105:743–755.

Gavrilets, S. 2003. PERSPECTIVE: MODELS OF SPECIATION: WHAT HAVE WE LEARNED IN 40 YEARS? Evolution (N. Y). 57:2197–2215. The Society for the Study of Evolution.

Ge, S. X., and D. Jung. 2018. ShinyGO: a graphical enrichment tool for animals and plants. bioRxiv 2.

Geiger, M. F., J. K. McCrary, and U. K. Schliewen. 2013. Crater Lake Apoyo revisited: population genetics of an emerging species flock. PLoS One 8:1–17.

Gillespie, J. H. 1984. Molecular Evolution Over the Mutational Landscape. Evolution (N. Y). 38:1116–1129. [Society for the Study of Evolution, Wiley].

Gillespie, R. G., G. M. Bennett, L. De Meester, J. L. Feder, R. C. Fleischer, L. J. Harmon, A. P. Hendry, M. L. Knope, J. Mallet, C. Martin, C. E. Parent, A. H. Patton, K. S. Pfennig, D. Rubinoff, D. Schluter, O. Seehausen, K. L. Shaw, E. Stacy, M. Stervander, J. T. Stroud, C. Wagner, and G. O. U. Wogan. 2020. Comparing Adaptive Radiations Across Space, Time, and Taxa. J. Hered. 111:1–20.

Givnish, T. J., E. L. Burkhardt, R. E. Happel, and J. D. Weintraub. 1984. Carnivory in the Bromeliad Brocchinia reducta , with a Cost / Benefit Model for the General Restriction of Carnivorous Plants to Sunny , Moist , Nutrient-Poor Habitats. Am. Nat. 124:479–497.

Givnish, T. J., K. J. Sytsma, J. Smith, W. Hahn, B. DH, and E. Burkhardt. 1997. Molecular evolution and adaptive radiation in *Brocchinia* (Bromeliaceae: Pitcairnioideae) atop tepuis of the Guyana Shield. Pp. 259–311 *in* Molecular evolution and adaptive radiation. Cambridge University Press, Cambridge.

Glor, R. E. 2010. Phylogenetic Insights on Adaptive Radiation. Annu. Rev. Ecol. Evol. Syst. 41:251–270. Annual Reviews.

Gompert, Z., and C. A. Buerkle. 2016. What, if anything, are hybrids: enduring truths and challenges associated with population structure and gene flow. Evol. Appl. 9:909–923.

Good, B. H., M. J. McDonald, J. E. Barrick, R. E. Lenski, and M. M. Desai. 2017. The dynamics of molecular evolution over 60,000 generations. Nature 551:45–50.

Grant, P. R. 2002. Unpredictable Evolution in a 30-Year Study of Darwin's Finches. Science (80-.). 296:707–711.

Grant, P. R., and R. B. Grant. 2008. How and Why Species Multiply. Princeton University Press.

Green, J., S. A. Corbet, and E. Betney. 1973. Ecological studies on crater lakes in West Cameroon The blood of endemic cichlids in Barombi Mbo in relation to stratification and their feeding habits. J. Zool. 170:299–308. Blackwell Publishing Ltd.
Green, M. R., and J. Sambrook. 2017. Isolation of high-molecular-weight DNA using organic solvents. Cold Spring Harb. Protoc. 2017:356–359.
Green, R. M., W. Feng, T. Phang, J. L. Fish, H. Li, R. A. Spritz, R. S. Marcucio, J. Hooper, H. Jamniczky, B. Hallgrímsson, and T. Williams. 2015. Tfap2a-dependent changes in mouse facial morphology result in clefting that can be ameliorated by a reduction in Fgf8 gene dosage. DMM Dis. Model. Mech. 8:31–43.
Gronau, I., M. J. Hubisz, B. Gulko, C. G. Danko, and A. Siepel. 2011. Bayesian inference of ancient human demography from individual genome sequences. Nat. Genet. 43:1031. Nature Publishing Group, a division of Macmillan Publishers Limited. All Rights Reserved.
Gross, B. L., and L. H. Rieseberg. 2005. The ecological genetics of homoploid hybrid speciation. J. Hered. 96:241–252.
Guerrero, R. F., and M. W. Hahn. 2017. Speciation as a sieve for ancestral polymorphism. Mol. Ecol. 26:5362–5368.
Guo, B., F. J. J. Chain, E. Bornberg-Bauer, E. H. Leder, and J. Merilä. 2013. Genomic divergence between nine- and three-spined sticklebacks. BMC Genomics 14:756.
Gutenkunst, R. N., R. D. Hernandez, S. H. Williamson, and C. D. Bustamante. 2009. Inferring the Joint Demographic History of Multiple Populations from Multidimensional SNP Frequency Data. PLOS Genet. 5:e1000695. Public Library of Science.
Hagey, F. M., and J. . Mylroie. 1995. Pleistocene lake and lagoon deposits, San Salvador Island, Bahamas. Geol. Soc. Am. 30:77–90.
Hague, M. T. J., G. Toledo, S. L. Geffeney, C. T. Hanifin, E. D. Brodie, and E. D. Brodie. 2018. Large-effect mutations generate trade-off between predatory and locomotor ability during arms race coevolution with deadly prey. Evol. Lett. 2:406–416.
Hahn, M. W., and L. Nakhleh. 2016. Irrational exuberance for resolved species trees. Evolution (N. Y). 70:7–17.
Haller, B. C., and P. W. Messer. 2019. SLiM 3: Forward Genetic Simulations Beyond the Wright–Fisher Model. Mol. Biol. Evol. 36:632–637.
Han, F., S. Lamichhaney, B. R. Grant, P. R. Grant, L. Andersson, and M. T. Webster. 2017. Gene flow, ancient polymorphism, and ecological adaptation shape the genomic landscape of divergence among Darwin's finches. Genome Res. 1–12.
Hansen, T. F. 2013. Why epistasis is important for selection and adaptation. Evolution (N. Y). 67:3501–3511.
Harmon, L. J., and S. Harrison. 2015. Species diversity is dynamic and unbounded at local and continental scales. Am. Nat. 185:584–593.
Harmon, L. J., J. J. Kolbe, J. M. Cheverud, and J. B. Losos. 2005. Convergence and the multidimensional niche. Evolution (N. Y). 59:409–421.
Harmon, L. J., J. B. Losos, T. J. Davies, R. G. Gillespie, J. L. Gittleman, W. B. Jennings, K. H. Kozak, M. A. McPeek, F. Moreno-Roark, T. J. Near, A. Purvis, R. E. Ricklefs, D. Schluter, J. A. Schulte, O. Seehausen, B. L. Sidlauskas, O. Torres-Carvajal, J. T. Weir, and A. Ø. Mooers. 2010. Early bursts of body size and shape evolution are rare in comparative data. Evolution 64:2385–2396.
Hedrick, P. W. 2013. Adaptive introgression in animals: Examples and comparison to new mutation and standing variation as sources of adaptive variation. Mol. Ecol. 22:4606–4618.

Heras, J., and C. Martin. 2021. Nonadaptive radiation of the gut microbiome in an adaptive radiation of Cyprinodon pupfishes with minor shifts for scale-eating. bioRxiv. Cold Spring Harbor Laboratory.

Herbert, S. P., J. Huisken, T. N. Kim, M. E. Feldman, B. T. Houseman, R. A. Wang, K. M. Shokat, and D. Y. R. Stainier. 2009. Arterial-Venous Segregation by Selective Cell Sprouting: An Alternative Mode of Blood Vessel Formation. Science (80-.). 326:294–298.

Hermisson, J., and P. S. Pennings. 2005. Soft Sweeps. Genetics 169:2335–2352.

Hermisson, J., and P. S. Pennings. 2017. Soft sweeps and beyond: understanding the patterns and probabilities of selection footprints under rapid adaptation. Methods Ecol. Evol. 8:700–716.

Hnasko, T. S., M. S. Szczypka, W. A. Alaynick, M. J. During, and R. D. Palmiter. 2004. A role for dopamine in feeding responses produced by orexigenic agents. 1023:309–318.

Höllinger, I., P. S. Pennings, and J. Hermisson. 2019. Polygenic adaptation: From sweeps to subtle frequency shifts. PLOS Genet. 15:e1008035. Public Library of Science.

Holtmeier, C. L. 2000. Heterochrony, maternal effects, and phenotypic variation among sympatric pupfishes. Evolution (N. Y). 2:330–338.

Horiuchi, Y., M. Ishikawa, N. Kaito, Y. Iijima, Y. Tanabe, H. Ishiguro, and T. Arinami. 2013. Experimental Evidence for the Involvement of PDLIM5 in Mood Disorders in Hetero Knockout Mice. PLoS One 8:1–10.

Hubbs, C. L., and R. R. Miller. 1942. Studies of the fishes of the order Cyprinodontes. XVIII. Cyprinodon laciniatus, new species, from the Bahamas.

Hudson, R. R. 2002. Generating samples under a Wright-Fisher neutral model of genetic variation. Bioinformatics 18:337–338.

Hudson, R. R. 2007. The Variance of Coalescent Time Estimates from DNA Sequences. J. Mol. Evol. 64:702–705.

Huerta-Sanchez, E., X. Jin, Asan, Z. Bianba, B. M. Peter, N. Vinckenbosch, Y. Liang, X. Yi, M. He, M. Somel, P. Ni, B. Wang, X. Ou, Huasang, J. Luosang, Z. X. P. Cuo, K. Li, G. Gao, Y. Yin, W. Wang, X. Zhang, X. Xu, H. Yang, Y. Li, J. J. Wang, J. J. Wang, R. Nielsen, E. Huerta-Sánchez, X. Jin, Asan, Z. Bianba, B. M. Peter, N. Vinckenbosch, Y. Liang, X. Yi, M. He, M. Somel, P. Ni, B. Wang, X. Ou, Huasang, J. Luosang, Z. X. P. Cuo, K. Li, G. Gao, Y. Yin, W. Wang, X. Zhang, X. Xu, H. Yang, Y. Li, J. J. Wang, J. J. Wang, and R. Nielsen. 2014. Altitude adaptation in Tibetans caused by introgression of Denisovan-like DNA. Nature 512:194–197. Nature Publishing Group, a division of Macmillan Publishers Limited. All Rights Reserved.

Huerta-Sánchez, E., X. Jin, Asan, Z. Bianba, B. M. Peter, N. Vinckenbosch, Y. Liang, X. Yi, M. He, M. Somel, P. Ni, B. Wang, X. Ou, Huasang, J. Luosang, Z. X. P. Cuo, K. Li, G. Gao, Y. Yin, W. Wang, X. Zhang, X. Xu, H. Yang, Y. Li, J. Wang, J. Wang, and R. Nielsen. 2014. Altitude adaptation in Tibetans caused by introgression of Denisovan-like DNA. Nature 512:194–197.

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

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

Igea, J., D. Bogarín, A. S. T. Papadopulos, and V. Savolainen. 2015. A comparative analysis of island floras challenges taxonomy-based biogeographical models of speciation. Evolution (N. Y). 69:482–491.

Ispolatov, I., and M. Doebeli. 2013. Chaos and Unpredictability in Evolution. Evolution (N. Y).

68:1365–1373.
Jackson, A. L., R. Inger, A. C. Parnell, and S. Bearhop. 2011. Comparing isotopic niche widths among and within communities: SIBER - Stable Isotope Bayesian Ellipses in R. J. Anim. Ecol. 80:595–602.
Jagdish, T., J. J. Morris, B. D. Wade, and Z. D. Blount. 2020. Probing the Deep Genetic Basis of a Novel Trait in Escherichia coli* BT - Evolution in Action: Past, Present and Future: A Festschrift in Honor of Erik D. Goodman. Pp. 107–122 *in* W. Banzhaf, B. H. C. Cheng, K. Deb, K. E. Holekamp, R. E. Lenski, C. Ofria, R. T. Pennock, W. F. Punch, and D. J. Whittaker, eds. Springer International Publishing, Cham.
Jayasena, C. S., and M. E. Bronner. 2012. Rbms3 functions in craniofacial development by posttranscriptionally modulating TGF-?? signaling. J. Cell Biol. 199:453–466.
Jensen, J. D. 2014. On the unfounded enthusiasm for soft selective sweeps. Nat. Commun. 5:5281.
Jiggins, C. D., R. E. Naisbit, R. L. Coe, and J. Mallet. 2001. Reproductive isolation caused by colour pattern mimicry. Nature 411:302–305.
John, M. E. S., K. E. Dixon, and C. H. Martin. 2020. Oral shelling within an adaptive radiation of pupfishes : Testing the adaptive function of a novel nasal protrusion and behavioural preference. 1–9.
Joliffe, I. 2002. Principal component analysis. John Wiley & Sons, Ltd.
Jones, F. C., M. G. Grabherr, Y. F. Chan, P. Russell, E. Mauceli, J. Johnson, R. Swofford, M. Pirun, M. C. Zody, S. White, E. Birney, S. Searle, J. Schmutz, J. Grimwood, M. C. Dickson, R. M. Myers, C. T. Miller, B. R. Summers, A. K. Knecht, S. D. Brady, H. Zhang, A. A. Pollen, T. Howes, C. Amemiya, Broad Institute Genome Sequencing Platform & Whole Genome Assembly Team, E. S. Lander, F. Di Palma, K. Lindblad-Toh, and D. M. Kingsley. 2012. The genomic basis of adaptive evolution in threespine sticklebacks. Nature 484:55–61.
Kagawa, K., and G. Takimoto. 2017. Hybridization can promote adaptive radiation by means of transgressive segregation. Ecol. Lett. 21:264–274. Wiley/Blackwell (10.1111).
Karageorgi, M., S. C. Groen, F. Sumbul, J. N. Pelaez, K. I. Verster, J. M. Aguilar, A. P. Hastings, S. L. Bernstein, T. Matsunaga, M. Astourian, G. Guerra, F. Rico, S. Dobler, A. A. Agrawal, and N. K. Whiteman. 2019. Genome editing retraces the evolution of toxin resistance in the monarch butterfly. Nature 574:409–412. Springer US.
Kauffman, S., and S. Levin. 1987. Towards a general theory of adaptive walks on rugged landscapes. J. Theor. Biol. 128:11–45.
Kautt, A. F., G. Machado-Schiaffino, and A. Meyer. 2016a. Multispecies Outcomes of Sympatric Speciation after Admixture with the Source Population in Two Radiations of Nicaraguan Crater Lake Cichlids. PLoS Genet. 12:1–33.
Kautt, A. F., G. Machado-Schiaffino, J. Torres-Dowdall, and A. Meyer. 2016b. Incipient sympatric speciation in Midas cichlid fish from the youngest and one of the smallest crater lakes in Nicaragua due to differential use of the benthic and limnetic habitats? Ecol. Evol. 6:5342–5357.
Kawasaki, J., S. Aegerter, R. D. Fevurly, A. Mammoto, T. Mammoto, M. Sahin, J. D. Mably, S. J. Fishman, and J. Chan. 2014. RASA1 functions in EPHB4 signaling pathway to suppress endothelial mTORC1 activity. J. Clin. Invest. 124:2774–2784. The American Society for Clinical Investigation.
Keller-Costa, T., A. V. M. Canário, and P. C. Hubbard. 2015. Chemical communication in

cichlids: A mini-review. Gen. Comp. Endocrinol. 221:64–74.
Kern, A. D., and D. R. Schrider. 2018. diploS/HIC: An Updated Approach to Classifying Selective Sweeps. G3 Genes|Genomes|Genetics.
Kida, Y. S., T. Sato, K. Y. Miyasaka, A. Suto, and T. Ogura. 2007. Daam1 regulates the endocytosis of EphB during the convergent extension of the zebrafish notochord. Proc. Natl. Acad. Sci. U. S. A. 104:6708–6713.
Kim, B. Y., C. D. Huber, and K. E. Lohmueller. 2018. Deleterious variation shapes the genomic landscape of introgression. PLoS Genet. 14:1–30.
Kim, Y., and R. Neilsen. 2004. Linkage disequilibrium as a signature of selective sweeps. Genetics 167:1513–1524.
Kirkpatrick, M., and N. H. Barton. 2006. Chromosome inversions, local adaptation and speciation. Genetics 173:419–434.
Kirkpatrick, M., and V. Ravigné. 2002. Speciation by Natural and Sexual Selection: Models and Experiments. . Am. Nat. 159:S22–S35. The University of Chicago Press.
Kocher, T. D. 2004. Adaptive evolution and explosive speciation: the cichlid fish model. Nat. Rev. Genet. 5:288–298.
Kodric-Brown, A., and R. J. D. West. 2014. Asymmetries in premating isolating mechanisms in a sympatric species flock of pupfish (Cyprinodon). Behav. Ecol. 25:69–75.
Koizumi, H., N. Yamaguchi, M. Hattori, T. O. Ishikawa, J. Aoki, M. M. Taketo, K. Inoue, and H. Arai. 2003. Targeted disruption of intracellular type I platelet activating factor-acetylhydrolase catalytic subunits causes severe impairment in spermatogenesis. J. Biol. Chem. 278:12489–12494.
Kondrashov, A. S., and F. A. Kondrashov. 1999. Interactions among quantitative traits in the course of sympatric speciation. Nature 400:351–354. Macmillan Magazines Ltd.
Kopp, M., and J. Hermisson. 2008. Competitive speciation and costs of choosiness. J. Evol. Biol. 21:1005–1023. Wiley/Blackwell (10.1111).
Kopp, M., M. R. Servedio, T. C. Mendelson, R. J. Safran, R. L. Rodríguez, M. E. Hauber, E. C. Scordato, L. B. Symes, C. N. Balakrishnan, D. M. Zonana, and G. S. van Doorn. 2017. Mechanisms of Assortative Mating in Speciation with Gene Flow: Connecting Theory and Empirical Research. Am. Nat. 191:1–20. The University of Chicago Press.
Kraft, M., I. C. Cirstea, A. K. Voss, T. Thomas, I. Goehring, B. N. Sheikh, L. Gordon, H. Scott, G. K. Smyth, M. R. Ahmadian, U. Trautmann, M. Zenker, M. Tartaglia, A. Ekici, A. Reis, H. G. Dörr, A. Rauch, and C. T. Thiel. 2011. Disruption of the histone acetyltransferase MYST4 leads to a noonan syndrome-like phenotype and hyperactivated MAPK signaling in humans and mice. J. Clin. Invest. 121:3479–3491.
Kratochwil, C. F., M. M. Sefton, Y. Liang, and A. Meyer. 2017. Tol2 transposon-mediated transgenesis in the Midas cichlid (Amphilophus citrinellus) — towards understanding gene function and regulatory evolution in an ecological model system for rapid phenotypic diversification. BMC Dev. Biol. 17:15.
Kratochwil, C. F., M. M. Sefton, and A. Meyer. 2015. Embryonic and larval development in the Midas cichlid fish species flock (Amphilophus spp.): A new evo-devo model for the investigation of adaptive novelties and species differences Evolutionary developmental biology. BMC Dev. Biol. 15:1–15.
Lamichhaney, S., 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, and L. Andersson. 2015. Evolution of Darwin's finches and their beaks revealed by genome

sequencing. Nature 518:371–375.
LaMonica, K., H. Ding, and K. B. Artinger. 2015. prdm1a functions upstream of itga5 in zebrafish craniofacial development. Genesis 53:270–277. John Wiley & Sons, Ltd.
Lande, R. 1981. Models of speciation by sexual selection on polygenic traits. Proc. Natl. Acad. Sci. 78:3721 LP – 3725.
Landis, M. J., and J. G. Schraiber. 2017. Pulsed evolution shaped modern vertebrate body sizes.
Laue, K., M. Jänicke, N. Plaster, C. Sonntag, and M. Hammerschmidt. 2008. Restriction of retinoic acid activity by Cyp26b1 is required for proper timing and patterning of osteogenesis during zebrafish development. 3787:3775–3787.
Lee, J., B.-K. Lee, and J. M. Gross. 2013a. Bcl6a function is required during optic cup formation to prevent p53-dependent apoptosis and colobomata. Hum. Mol. Genet. 22:3568–3582.
Lee, J. Y., D. Seo, J. You, S. Chung, J. S. Park, J. H. Lee, S. M. Jung, Y. S. Lee, and S. H. Park. 2017. The deubiquitinating enzyme, ubiquitin-specific peptidase 50, regulates inflammasome activation by targeting the ASC adaptor protein. FEBS Lett. 591:479–490.
Lee, Y., C. D. Marsden, L. C. Norris, T. C. Collier, B. J. Main, A. Fofana, A. J. Cornel, and G. C. Lanzaro. 2013b. Spatiotemporal dynamics of gene flow and hybrid fitness between the M and S forms of the malaria mosquito, Anopheles gambiae. Proc. Natl. Acad. Sci. U. S. A. 110:19854–9.
Lencer, E. S., W. C. Warren, R. Harrison, and A. R. McCune. 2017. The Cyprinodon variegatus genome reveals gene expression changes underlying differences in skull morphology among closely related species. BMC Genomics 18:424. BioMed Central.
Lenomand, T. 2002. Gene flow and the limits to natural selection. Trends Ecol. Evol. 17:183–189.
Li, H. 2013. Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. arXiv 1303:1–3.
Li, H., and R. Durbin. 2009. Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics 25:1754–1760.
Li, H., and R. Durbin. 2011. Inference of human population history from individual whole-genome sequences. Nature 475:493–496. Nature Publishing Group.
Lindsey, H. A., J. Gallie, S. Taylor, and B. Kerr. 2013. Evolutionary rescue from extinction is contingent on a lower rate of environmental change. Nature 494:463–467.
Loh, P.-R., M. Lipson, N. Patterson, P. Moorjani, J. K. Pickrell, D. Reich, and B. Berger. 2013. Inferring Admixture Histories of Human Populations Using Linkage Disequilibrium. Genetics 193:1233 LP – 1254.
Losos, J. B., P. H. Harvey, A. J. Leigh Brown, and J. M. Smith. 1995. Community evolution in Greater Antillean Anolis lizards: phylogenetic patterns and experimental tests. Philos. Trans. R. Soc. London. Ser. B Biol. Sci. 349:69–75. Royal Society.
Losos, J. B., and D. L. Mahler. 2010. Adaptive Radiation: The Interaction of Ecological Opportunity, Adaptation, and Speciation. Pp. 381–420 *in* Evolution since Darwin: the first 150 years. Sinauer Associates, Sunderland, MA.
Losos, J. B., T. W. Schoener, and D. A. Spiller. 2004. Predator-induced behaviour shifts and natural selection in field-experimental lizard populations. Nature 432:505–508.
Lovette, I. J., E. Bermingham, and R. E. Ricklefs. 2002. Clade-specific morphological diversification and adaptive radiation in Hawaiian songbirds. Proc. R. Soc. London. Ser. B Biol. Sci. 269:37–42. Royal Society.
Ma, T., K. Wang, Q. Hu, Z. Xi, D. Wan, Q. Wang, J. Feng, D. Jiang, H. Ahani, R. J. Abbott, M.

Lascoux, E. Nevo, and J. Liu. 2018. Ancient polymorphisms and divergence hitchhiking contribute to genomic islands of divergence within a poplar species complex. Proc. Natl. Acad. Sci. 115:E236 LP-E243.
Mabee, P., J. P. Balhoff, W. M. Dahdul, H. Lapp, P. E. Midford, T. J. Vision, and M. Westerfield. 2012. 500,000 fish phenotypes: The new informatics landscape for evolutionary and developmental biology of the vertebrate skeleton. J. Appl. Ichthyol. 28:300–305.
Machado-Schiaffino, G., A. F. Kautt, H. Kusche, and A. Meyer. 2015. Parallel evolution in Ugandan crater lakes: Repeated evolution of limnetic body shapes in haplochromine cichlid fish. BMC Evol. Biol. 15.
Maia, T. M., D. Gogendeau, C. Pennetier, C. Janke, and R. Basto. 2014. Bug22 influences cilium morphology and the posttranslational modification of ciliary microtubules. Biol. Open 3:138–151.
Malinsky, M., R. J. Challis, A. M. Tyers, S. Schiffels, Y. Terai, B. P. Ngatunga, E. A. Miska, R. Durbin, M. J. Genner, and G. F. Turner. 2015. Genomic islands of speciation separate cichlid ecomorphs in an East African crater lake. Science (80-.). 350:1493–1498.
Malinsky, M., H. Svardal, A. M. Tyers, E. A. Miska, M. J. Genner, G. F. Turner, and R. Durbin. 2018. Whole-genome sequences of Malawi cichlids reveal multiple radiations interconnected by gene flow. Nat. Ecol. Evol. 2:1940–1955.
Mallet, J., M. Beltrán, W. Neukirchen, and M. Linares. 2007. Natural hybridization in heliconiine butterflies: the species boundary as a continuum. BMC Evol. Biol. 7:28.
Mallet, J., A. Meyer, P. Nosil, and J. L. Feder. 2009a. Space, sympatry and speciation. J. Evol. Biol. 22:2332–41.
Mallet, J., A. Meyer, P. Nosil, and J. L. Feder. 2009b. Space, sympatry and speciation. J. Evol. Biol. 22:2332–41.
Manda, P., J. P. Balhoff, H. Lapp, P. Mabee, and T. J. Vision. 2015. Using the phenoscape knowledgebase to relate genetic perturbations to phenotypic evolution. Genesis 53:561–571.
Mani, G. S., and B. C. Clarke. 1990. Mutational order: a major stochastic process in evolution. Proc. R. Soc. London. B. Biol. Sci. 240:29–37. Royal Society.
Marles, L. K., E. M. Peters, D. J. Tobin, N. A. Hibberts, and K. U. Schallreuter. 2003. Tyrosine hydroxylase isoenzyme I is present in human melanosomes: a possible novel function in pigmentation. Exp. Dermatol. 12:61–70. John Wiley & Sons, Ltd.
Marques, D. A., J. I. Meier, and O. Seehausen. 2019. A Combinatorial View on Speciation and Adaptive Radiation. Trends Ecol. Evol., doi: https://doi.org/10.1016/j.tree.2019.02.008.
Marsden, C. D., Y. Lee, K. Kreppel, A. Weakley, A. Cornel, H. M. Ferguson, E. Eskin, and G. C. Lanzaro. 2014. Diversity, differentiation, and linkage disequilibrium: prospects for association mapping in the malaria vector Anopheles arabiensis. G3 (Bethesda). 4:121–31.
Martin, C. H. 2016a. Context dependence in complex adaptive landscapes: frequency and trait-dependent selection surfaces within an adaptive radiation of Caribbean pupfishes. Evolution (N. Y). 70:1265–1282.
Martin, C. H. 2013. Strong assortative mating by diet, color, size, and morphology but limited progress toward sympatric speciation in a classic example: Cameroon crater lake cichlids. Evolution (N. Y). 67:2114–23.
Martin, C. H. 2016b. The cryptic origins of evolutionary novelty: 1000-fold faster trophic diversification rates without increased ecological opportunity or hybrid swarm. Evolution

(N. Y). 70:2504–2519.
Martin, C. H. 2012. Weak Disruptive Selection and Incomplete Phenotypic Divergence in Two Classic Examples of Sympatric Speciation: Cameroon Crater Lake Cichlids. Am. Nat. 180:E90–E109.
Martin, C. H., J. E. Crawford, B. J. Turner, and L. H. Simons. 2016. 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.
Martin, C. H., J. S. Cutler, J. P. Friel, C. Dening Touokong, G. Coop, and P. C. Wainwright. 2015a. Complex histories of repeated gene flow in Cameroon crater lake cichlids cast doubt on one of the clearest examples of sympatric speciation. Evolution (N. Y). 69:1406–1422.
Martin, C. H., P. A. Erickson, and C. T. Miller. 2017. 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.
Martin, C. H., and L. C. Feinstein. 2014. Novel trophic niches drive variable progress towards ecological speciation within an adaptive radiation of pupfishes. Mol. Ecol. 23:1846–1862.
Martin, C. H., K. Gould, and C. Bocklage. 2019. Surprising spatiotemporal stability and frequency-independence across multiple fitness peaks driving adaptive radiation in the wild.
Martin, C. H., and K. J. Gould. 2020. Surprising spatiotemporal stability of a multi-peak fitness landscape revealed by independent field experiments measuring hybrid fitness. Evol. Lett. 530–544.
Martin, C. H., and S. Johnsen. 2007. A field test of the Hamilton-Zuk hypothesis in the Trinidadian guppy (Poecilia reticulata). Behav. Ecol. Sociobiol. 61:1897–1909.
Martin, C. H., and E. J. Richards. 2019. 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.
Martin, C. H., and P. C. Wainwright. 2013a. A remarkable species flock of *Cyprinodon* pupfishes endemic to San Salvador Island, Bahamas. Bull. Peabody Museum Nat. Hist. 54:231–240.
Martin, C. H., and P. C. Wainwright. 2013b. A remarkable species flock of Cyprinodon pupfishes endemic to San Salvador Island, Bahamas. Bull. Peabody Museum Nat. Hist. 54:231–241.
Martin, C. H., and P. C. Wainwright. 2013c. Multiple fitness peaks on the adaptive landscape drive adaptive radiation in the wild. Science 339:208–11.
Martin, C. H., and P. C. Wainwright. 2013d. On the measurement of ecological novelty: scale-eating pupfish are separated by 168 my from other scale-eating fishes. PLoS One 8:e71164.
Martin, C. H., and P. C. Wainwright. 2011. Trophic novelty is linked to exceptional rates of morphological diversification in two adaptive radiations of cyprinodon pupfish. Evolution (N. Y). 65:2197–2212.
Martin, S. H., K. K. Dasmahapatra, N. J. Nadeau, C. Slazar, J. R. Walters, F. Simpson, M. Blaxter, A. Manica, J. Mallet, and C. D. Jiggins. 2013a. Genome-wide evidence for speciation with gene flow in Heliconius butterflies. Genome Res. 23:1817–1828.
Martin, S. H., K. K. Dasmahapatra, N. J. Nadeau, C. Slazar, J. R. Walters, F. Simpson, M. Blaxter, A. Manica, J. Mallet, and C. D. Jiggins. 2013b. Heliconius and sympatric speciation. Genome Res. 23:1817–1828.
Martin, S. H., J. W. Davey, and C. D. Jiggins. 2015b. Evaluating the use of ABBA-BABA statistics to locate introgressed loci. Mol. Biol. Evol. 32:244–257.

Masatoshi, N. 1972. Genetic distance between populations. Am. Nat. 106:283–292.
Matessi, C., A. Gimelfarb, and S. Gavrilets. 2001. Long-term Buildup of Reproductive Isolation Promoted by Disruptive Selection: How Far Does it Go? Selection 2:41–64. Akadémiai Kiadó.
Maynard Smith, J. 1970. Natural Selection and the Concept of a Protein Space. Nature 225:563–564.
Mayr, E. 1963. Animal species and evolution. Belknap, Cambridge, MA.
Mccormick, C., Y. Leduc, D. Martindale, K. Mattison, L. E. Esford, A. P. Dyer, and F. Tufaro. 1998. The putative tumour suppressor EXT1 alters the expression of cell-surface heparan sulfate. 19:158–161.
McDonald, A. C., J. A. Schuijers, A. L. Gundlach, and B. L. Grills. 2007. Galanin treatment offsets the inhibition of bone formation and downregulates the increase in mouse calvarial expression of TNFα and GalR2 mRNA induced by chronic daily injections of an injurious vehicle. Bone 40:895–903.
Mcgirr, J. A., and C. H. Martin. 2018. Parallel evolution of gene expression between trophic specialists despite divergent genotypes and morphologies. Evol. Lett. 2:62–75.
McGirr, J. A., and C. H. Martin. 2020. Ecological divergence in sympatry causes gene misexpression in hybrids. Mol. Ecol. 29:2707–2721.
McGirr, J. A., and C. H. Martin. 2016a. Novel candidate genes underlying extreme trophic specialization in Caribbean pupfishes. Mol. Biol. Evol. msw286.
McGirr, J. A., and C. H. Martin. 2016b. Novel candidate genes underlying extreme trophic specialization in Caribbean pupfishes. Mol. Biol. Evol. 34:873–888.
McGowan, H. W., J. A. Schuijers, B. L. Grills, S. J. McDonald, and A. C. McDonald. 2014. 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.
McKenna, A., M. Hanna, E. Banks, A. Sivachenko, K. Cibulskis, A. Kernytsky, K. Garimella, D. Altshuler, S. Gabriel, M. Daly, and M. A. DePristo. 2010. The genome analysis toolkit: A MapReduce framework for analyzing next-generation DNA sequencing data. Genome Res. 20:1297–1303.
McMillan, H. J., A. Telegrafi, A. Singleton, M. T. Cho, D. Lelli, F. C. Lynn, J. Griffin, A. Asamoah, T. Rinne, C. E. Erasmus, D. A. Koolen, C. A. Haaxma, B. Keren, D. Doummar, C. Mignot, I. Thompson, L. Velsher, M. Dehghani, M. Y. Vahidi Mehrjardi, R. Maroofian, M. Tchan, C. Simons, J. Christodoulou, E. Martín-Hernández, M. J. Guillen Sacoto, L. B. Henderson, H. McLaughlin, L. L. Molday, R. S. Molday, and G. Yoon. 2018. Recessive mutations in ATP8A2 cause severe hypotonia, cognitive impairment, hyperkinetic movement disorders and progressive optic atrophy. Orphanet J. Rare Dis. 13:1–10. Orphanet Journal of Rare Diseases.
Meier, J. I., D. A. Marques, S. Mwaiko, C. E. Wagner, L. Excoffier, and O. Seehausen. 2017a. Ancient hybridization fuels rapid cichlid fish adaptive radiations. Nat. Commun. 8:14363. Nature Publishing Group.
Meier, J. I., V. C. Sousa, D. A. Marques, O. M. Selz, C. E. Wagner, L. Excoffier, and O. Seehausen. 2017b. Demographic modelling with whole-genome data reveals parallel origin of similar Pundamilia cichlid species after hybridization. Mol. Ecol. 26:123–141.
Meier, J. I., R. B. Stelkens, D. A. Joyce, U. K. Schliewen, O. M. Selz, C. E. Wagner, S. Mwaiko, N. Phiri, C. Katongo, and O. Seehausen. 2019. Hybridization explains rapid adaptive

radiation in Lake Mweru cichlid fi shes. Nat. Commun., doi: 10.1038/s41467-019-13278-z. Springer US.
Midford, P. E., T. Dececchi, J. P. Balhoff, W. M. Dahdul, N. Ibrahim, H. Lapp, J. G. Lundberg, P. M. Mabee, P. C. Sereno, M. Westerfield, T. J. Vision, D. C. Blackburn, S. Federhen, P. Mabee, J. Balhoff, W. Dahdul, H. Lapp, P. Midford, T. Vision, M. Westerfield, W. Dahdul, J. Balhoff, J. Engeman, T. Grande, E. Hilton, C. Kothari, H. Lapp, J. Lundberg, P. Midford, T. Vision, M. Westerfield, P. Mabee, F. Fang, M. Toledo-Piza, M. Uhen, A. Barnosky, B. Bills, J. Blois, M. Carrano, M. Carrasco, G. Erickson, J. Eronen, M. Fortelius, R. Graham, E. Grimm, M. O'Leary, A. Mast, W. Piel, P. Polly, L. Saila, J. Balhoff, P. Midford, H. Lapp, M. Ghiselin, S. Schulz, H. Stenzhorn, M. Boeker, D. Thau, N. Franz, B. Smith, M. Ashburner, C. Rosse, J. Bard, W. Bug, W. Ceusters, L. Goldberg, K. Eilbeck, A. Ireland, C. Mungall, O. Consortium, N. Leontis, P. Rocca-Seraa, A. Ruttenberg, S. Sansone, R. Scheuermann, N. Shah, P. Whetzel, S. Lewis, J. Beaulieu, R. Ree, J. Cavender-Bares, G. Weiblen, and M. Donoghue. 2013. The vertebrate taxonomy ontology: a framework for reasoning across model organism and species phenotypes. J. Biomed. Semantics 4:34.
Miller, A. C., L. H. Voelker, A. N. Shah, and C. B. Moens. 2015. Neurobeachin is required postsynaptically for electrical and chemical synapse formation. Curr. Biol. 25:16–28. Elsevier Ltd.
Miller, S. E., A. W. Legan, M. T. Henshaw, K. L. Ostevik, K. Samuk, and F. M. K. Uy. 2019. Evolutionary dynamics of recent selection on cognitive abilities. 1–8.
Muschick, M., M. Barluenga, W. Salzburger, and A. Meyer. 2011. Adaptive phenotypic plasticity in the Midas cichlid fish pharyngeal jaw and its relevance in adaptive radiation. BMC Evol. Biol. 11:116.
Musilová, Z., A. Indermaur, A. R. Bitja Nyom, R. Tropek, C. Martin, and U. K. Schliewen. 2014. Persistence of *Stomatepia mongo*, an endemic cichlid fish of the Barombi Mbo Crater Lake, Southwestern Cameroon, with notes on its life history and behavior. Copeia 14:556–560.
Nachman, M. W., and B. A. Payseur. 2012. Recombination rate variation and speciation: theoretical predictions and empirical results from rabbits and mice. Philos. Trans. R. Soc. B Biol. Sci. 367:409–421.
Nadachowska-Brzyska, K., R. Burri, L. Smeds, and H. Ellegren. 2016. PSMC analysis of effective population sizes in molecular ecology and its application to black-and-white Ficedula flycatchers. Mol. Ecol. 25:1058–1072.
Nagase, T., G. Mizugachi, N. Nomura, R. Ishizaki, Y. Ueno, and S. Ishii. 1990. Requirement of protein co-factor for the DNA-binding function of the human ski proto-oncogene product. Nucleic Acids Res. 18:337–343.
Nakagome, S., G. Alkorta-Aranburu, R. Amato, B. Howie, B. M. Peter, R. R. Hudson, and A. Di Rienzo. 2016. Estimating the Ages of Selection Signals from Different Epochs in Human History. Mol. Biol. Evol. 33:657–669.
Nei, M., and W. H. Li. 1979. Mathematical model for studying genetic variation in terms of restriction endonucleases. Proc. Natl. Acad. Sci. 76:5269 LP – 5273.
Nelson, T. C., and W. A. Cresko. 2018. Ancient genomic variation underlies repeated ecological adaptation in young stickleback populations. Evol. Lett. 2:9–21.
Nikaido, M., T. Ota, T. Hirata, H. Suzuki, Y. Satta, M. Aibara, S. I. Mzighani, C. Sturmbauer, K. Hagino-Yamagishi, and N. Okada. 2014. Multiple Episodic Evolution Events in V1R Receptor Genes of East-African Cichlids. Genome Biol. Evol. 6:1135–1144.

Nikaido, M., H. Suzuki, A. Toyoda, A. Fujiyama, K. Hagino-Yamagishi, T. D. Kocher, K. Carleton, and N. Okada. 2013. Lineage-Specific Expansion of Vomeronasal Type 2 Receptor-Like (OlfC) Genes in Cichlids May Contribute to Diversification of Amino Acid Detection Systems. Genome Biol. Evol. 5:711–722.
Noor, M. a F., and S. M. Bennett. 2009. Islands of speciation or mirages in the desert? Examining the role of restricted recombination in maintaining species. Heredity (Edinb). 103:439–444. Nature Publishing Group.
Norris, L. C., B. J. Main, Y. Lee, T. C. Collier, A. Fofana, A. J. Cornel, and G. C. Lanzaro. 2015. Adaptive introgression in an African malaria mosquito coincident with the increased usage of insecticide-treated bed nets. Proc. Natl. Acad. Sci. 112:815–820.
Norvaišas, P., and E. Kisdi. 2012. Revisiting Santa Rosalia to Unfold a Degeneracy of Classic Models of Speciation. Am. Nat. 180:388–393. The University of Chicago Press.
Nosil, P., S. P. Egan, and D. J. Funk. 2008. Heterogeneous genomic differentiation between walking-stick ecotypes: "isolation by adaptation" and multiple roles for divergent selection. Evolution 62:316–36.
Nosil, P., and J. L. Feder. 2012. Widespread yet heterogeneous genomic divergence. Mol. Ecol. 21:2829–2832.
Nosil, P., and L. J. Harmon. 2009. Niche dimensionality and ecologocial speciation. Pp. 127–154 *in* R. Butlin, ed. Speciation and Patterns of Diversity. Cambridge University Press, Cambridge.
Nosil, P., L. J. Harmon, and O. Seehausen. 2009. Ecological explanations for (incomplete) speciation. Trends Ecol. Evol. 24:145–156.
Nosil, P., and P. Hohenlohe. 2012. Dimensionality of sexual isolation during reinforcement and ecological speciation in. Evol. Ecol. Res. 14:467–485.
Nosil, P., and D. Schluter. 2011. The genes underlying the process of speciation. Trends Ecol. Evol. 26:160–167.
Nuytens, K., I. Gantois, P. Stijnen, E. Iscru, A. Laeremans, L. Serneels, L. Van Eylen, S. A. Liebhaber, K. Devriendt, D. Balschun, L. Arckens, J. W. M. Creemers, and R. D'Hooge. 2013. Haploinsufficiency of the autism candidate gene Neurobeachin induces autism-like behaviors and affects cellular and molecular processes of synaptic plasticity in mice. Neurobiol. Dis. 51:144–151. Elsevier Inc.
Ofran, Y., and B. Rost. 2007. ISIS: Interaction sites identified from sequence. Bioinformatics 23:13–16.
Ormond, L., M. Foll, G. B. Ewing, S. P. Pfeifer, and J. D. Jensen. 2016. Inferring the age of a fixed beneficial allele. Mol. Ecol. 25:157–169.
Orr, H. A. 2005. The genetic theory of adaptation: A brief history. Nat. Rev. Genet. 6:119–127.
Orr, H. A. 1998. The Population Genetics of Adaptation: The Distribution of Factors Fixed during Adaptive Evolution. Evolution (N. Y). 52:935.
Ortíz-Barrientos, D., and M. A. F. Noor. 2005. Evidence for a One-Allele Assortative Mating Locus. Science (80-.). 310:1467 LP – 1467.
Osualdo, A. D., and J. C. Reed. 2011. NLRP1, a regulator of innate immunity associated with vitiligo. Pigment Cell Melanoma Res. 5–8.
Otte, K. A., V. Nolte, F. Mallard, and C. Schlötterer. 2021. The genetic architecture of temperature adaptation is shaped by population ancestry and not by selection regime. Genome Biol. 22:1–25. Genome Biology.
Otto, S. P., M. R. Servedio, and S. L. Nuismer. 2008. Frequency-dependent selection and the

evolution of assortative mating. Genetics 179:2091–2112.
Palmer, D. H., and M. R. Kronforst. 2015. Divergence and gene flow among Darwin's finches: A genome-wide view of adaptive radiation driven by interspecies allele sharing. BioEssays 37:968–974.
Papadopolus, Savolainen, A. S. T. Papadopulos, W. J. Baker, D. Crayn, R. K. Butlin, R. G. Kynast, I. Hutton, and V. Savolainen. 2011. Speciation with gene flow on Lord Howe Island. Proc. Natl. Acad. Sci. U. S. A. 108:13188–93.
Patricia Hernandez, L., A. C. Gibb, and L. Ferry-Graham. 2009. Trophic apparatus in cyprinodontiform fishes: Functional specializations for picking and scraping behaviors. J. Morphol. 270:645–661.
Patterson, N., P. Moorjani, Y. Luo, S. Mallick, N. Rohland, Y. Zhan, T. Genschoreck, T. Webster, and D. Reich. 2012. Ancient admixture in human history. Genetics 192:1069–1093.
Patton, A. H., E. J. Richards, K. J. Gould, L. K. Buie, and H. Christopher. 2021. Adaptive introgression and de novo mutations increase access to novel fitness peaks on the fitness landscape during a vertebrate adaptive radiation.
Patwa, Z., and L. M. Wahl. 2008. The fixation probability of beneficial mutations. J. R. Soc. Interface 5:1279–1289.
Pavlidis, P., and N. Alachiotis. 2017. A survey of methods and tools to detect recent and strong positive selection. J. Biol. Res. 24:7.
Pavlidis, P., D. Živković, A. Stamatakis, and N. Alachiotis. 2013. SweeD: Likelihood-based detection of selective sweeps in thousands of genomes. Mol. Biol. Evol. 30:2224–2234.
Pease, J. B., D. C. Haak, M. W. Hahn, and L. C. Moyle. 2016. Phylogenomics Reveals Three Sources of Adaptive Variation during a Rapid Radiation. PLoS Biol. 14:1–24.
Peter, B. M. 2016. Admixture, population structure, and f-statistics. Genetics 202:1485–1501.
Pfeifer, B., and D. D. Kapan. 2019. Estimates of introgression as a function of pairwise distances. BMC Bioinformatics 20:1–11. BMC Bioinformatics.
Pfeifer, B., U. Wittelsbürger, S. E. Ramos-Onsins, and M. J. Lercher. 2014. PopGenome: An efficient swiss army knife for population genomic analyses in R. Mol. Biol. Evol. 31:1929–1936.
Pickrell, J. K., and J. K. Pritchard. 2012. Inference of Population Splits and Mixtures from Genome-Wide Allele Frequency Data. PLoS Genet. 8:e1002967.
Pinho, C., and J. Hey. 2010. Divergence with gene flow: models and data. Annu. Rev. Ecol. Evol. … 215–232.
Plenderleith, M., C. van Oosterhout, R. Robinson, and G. Turner. 2005. Female preference for conspecific males based on olfactory preference in a Lake Malawi cichlid fish. Biol. Lett. 1:1–8.
Poelstra, J. W., E. J. Richards, and C. H. Martin. 2018. Speciation in sympatry with ongoing secondary gene flow and a potential olfactory trigger in a radiation of Cameroon cichlids. Mol. Ecol. 27:1–19.
Poelstra, J. W., N. Vijay, C. M. Bossu, H. Lantz, B. Ryll, I. Müller, V. Baglione, P. Unneberg, M. Wikelski, M. G. Grabherr, and J. B. W. Wolf. 2014. The genomic landscape underlying phenotypic integrity in the face of gene flow in crows. Science (80-.). 344:1410–4.
Polechová, J., N. H. Barton, and S. Gavrilefs. 2005. Speciation thrrough competition: a critical review. Evolution (N. Y). 59:1194–1210. The Society for the Study of Evolution.
Prescott, S. M., G. a Zimmerman, D. M. Stafforini, and T. M. Mcintyre. 2000. Platelet activating

factor. Annu. Rev. Biochem. 69:419–445.
Price, S. A., S. S. B. Hopkins, K. K. Smith, and V. L. Roth. 2012. Tempo of trophic evolution and its impact on mammalian diversification. Proc. Natl. Acad. Sci. 109:7008–7012.
Pritchard, J. K., J. K. Pickrell, and G. Coop. 2010. The Genetics of Human Adaptation: Hard Sweeps, Soft Sweeps, and Polygenic Adaptation. Curr. Biol. 20:R208–R215.
Purcell, S., 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, and P. C. Sham. 2007. PLINK: A Tool Set for Whole-Genome Association and Population-Based Linkage Analyses. Am. J. Hum. Genet. 81:559–575.
Qiao, L., M. Du, X. Liang, Y. Hao, X. He, F. Si, T. Mei, and B. Chen. 2016. Tyrosine Hydroxylase is crucial for maintaining pupal tanning and immunity in Anopheles sinensis. Sci. Rep. 6:1–11. Nature Publishing Group.
Quandt, E. M., D. E. Deatherage, A. D. Ellington, G. Georgiou, and J. E. Barrick. 2014. Recursive genomewide recombination and sequencing reveals a key refinement step in the evolution of a metabolic innovation in Escherichia coli. Proc. Natl. Acad. Sci. U. S. A. 111:2217–2222.
Rabosky, D. L. 2013. Diversity-Dependence, Ecological Speciation, and the Role of Competition in Macroevolution. Annu. Rev. Ecol. Evol. Syst. 44:481–502.
Rabosky, D. L., G. J. Slater, and M. E. Alfaro. 2012. Clade Age and Species Richness Are Decoupled Across the Eukaryotic Tree of Life. PLOS Biol. 10:e1001381. Public Library of Science.
Ravinet, M., R. Faria, R. K. Butlin, J. Galindo, N. Bierne, M. Rafajlović, M. A. F. Noor, B. Mehlig, and A. M. Westram. 2017. Interpreting the genomic landscape of speciation: a road map for finding barriers to gene flow. J. Evol. Biol. 30:1450–1477.
Recknagel, H., K. R. Elmer, and A. Meyer. 2013. A hybrid genetic linkage map of two ecologically and morphologically divergent Midas cichlid fishes (Amphilophus spp.) obtained by massively parallel DNA sequencing (ddRADSeq). G3 (Bethesda). 3:65–74.
Reich, D. E., M. Cargill, S. Bolk, J. Ireland, P. C. Sabeti, D. J. Richter, T. Lavery, R. Kouyoumjian, S. F. Farhadian, R. Ward, and E. S. Lander. 2001. Linkage disequilibrium in the human genome. 9:199–204.
Reich, D., K. Thangaraj, N. Patterson, A. L. Price, and L. Singh. 2009. Reconstructing Indian population history. Nature 461:489–94.
Richards, E. J., and C. H. Martin. 2017a. Adaptive introgression from distant Caribbean islands contributed to the diversification of a microendemic adaptive radiation of trophic specialist pupfishes. PLOS Genet. 13:e1006919. Public Library of Science.
Richards, E. J., and C. H. Martin. 2017b. Adaptive introgression from distant Caribbean islands contributed to the diversification of a microendemic adaptive radiation of trophic specialist pupfishes. PLoS Genet. 13:1–35.
Richards, E. J., J. A. McGirr, J. R. Wang, M. E. St. John, J. W. Poelstra, M. J. Solano, D. C. O'Connell, B. J. Turner, and C. H. Martin. 2021. A vertebrate adaptive radiation is assembled from an ancient and disjunct spatiotemporal landscape. Proc. Natl. Acad. Sci. U. S. A. 118.
Richards, E. J., J. W. Poelstra, and C. H. Martin. 2018. Don't throw out the sympatric speciation with the crater lake water: fine-scale investigation of introgression provides equivocal support for causal role of secondary gene flow in one of the clearest examples of sympatric speciation. Evol. Lett. 2:524–540.

Richards, E. J., M. R. Servedio, and C. H. Martin. 2019. Searching for Sympatric Speciation in the Genomic Era. BioEssays 41:1900047. John Wiley & Sons, Ltd.
Richards, E., J. McGirr, J. Wang, M. St. John, J. Poelstra, M. Solano, D. O'Connell, B. Turner, and C. Martin. 2020. Major stages of vertebrate adaptive radiation are assembled from a disparate spatiotemporal landscape. , doi: 10.1101/2020.03.12.988774.
Rieseberg, L. H., M. A. Archer, and R. K. Wayne. 1999. Transgressive segregation, adaptation and speciation. Heredity (Edinb). 83:363. The Genetical Society of Great Britain.
Rieseberg, L. H., O. Raymond, D. M. Rosenthal, Z. Lai, K. Livingstone, T. Nakazato, J. L. Durphy, A. E. Schwarzbach, L. A. Donovan, and C. Lexer. 2003. Major Ecological Transitions in Wild Sunflowers Facilitated by Hybridization. Science (80-.). 301:1211–1216.
Roderick, G. K., and R. G. Gillespie. 1998. Speciation and phylogeography of Hawaiian terrestrial arthropods. Mol. Ecol. 7:519–531.
Roesti, M., D. Moser, and D. Berner. 2013. Recombination in the threespine stickleback genome - Patterns and consequences. Mol. Ecol. 22:3014–3027.
Ronco, F., M. Matschiner, A. Böhne, A. Boila, H. H. Büscher, A. El Taher, A. Indermaur, M. Malinsky, V. Ricci, A. Kahmen, S. Jentoft, and W. Salzburger. 2020. Drivers and dynamics of a massive adaptive radiation in cichlid fishes. Nature, doi: 10.1038/s41586-020-2930-4.
Rosenblum, E. B., B. A. J. Sarver, J. W. Brown, S. Des Roches, K. M. Hardwick, T. D. Hether, J. M. Eastman, M. W. Pennell, and L. J. Harmon. 2012. Goldilocks Meets Santa Rosalia: An Ephemeral Speciation Model Explains Patterns of Diversification Across Time Scales. Evol. Biol. 39:255–261.
Rougemont, Q., P.-A. Gagnaire, C. Perrier, C. Genthon, A.-L. Besnard, S. Launey, and G. Evanno. 2016. Inferring the demographic history underlying parallel genomic divergence among pairs of parasitic and non-parasitic lamprey ecotypes. Mol. Ecol. n/a-n/a.
Rousselle, M., P. Simion, M.-K. Tilak, E. Figuet, B. Nabholz, and N. Galtier. 2020. Is adaptation limited by mutation? A timescale-dependent effect of genetic diversity on the adaptive substitution rate in animals. PLOS Genet. 16:e1008668. Public Library of Science.
Sachdeva, H., and N. H. Barton. 2017. Divergence and evolution of assortative mating in a polygenic trait model of speciation with gene flow. Evolution (N. Y). 71:1478–1493. Wiley/Blackwell (10.1111).
Sallan, L. C., and M. Friedman. 2012. Heads or tails: Staged diversification in vertebrate evolutionary radiations. Proc. R. Soc. B Biol. Sci. 279:2025–2032.
Savolainen, V., M.-C. Anstett, C. Lexer, I. Hutton, J. J. Clarkson, M. V Norup, M. P. Powell, D. Springate, N. Salamin, and W. J. Baker. 2006. Sympatric speciation in palms on an oceanic island. Nature 441:210–213.
Schiffels, S., and R. Durbin. 2014. Inferring human population size and separation history from multiple genome sequences. Nat. Genet. 46:919. Nature Publishing Group, a division of Macmillan Publishers Limited. All Rights Reserved.
Schliewen, U. K., and B. Klee. 2004. Reticulate sympatric speciation in Cameroonian crater lake cichlids. Front. Zool. 1:5.
Schliewen, U. K., D. Tautz, and S. Paabo. 1994. Sympatric speciation suggested by monophyly of crater lake cichlids. Nature 368:629–631.
Schluter, D. 2009. Evidence for Ecological Speciation and Its Alternative. Science (80-.). 323:737 LP – 741.
Schluter, D. 2000. The ecology of adaptive radiation. OUP Oxford.

Schluter, D., and G. L. Conte. 2009. Genetics and ecological speciation. Proc. Natl. Acad. Sci. 106:9955–9962.
Schluter, D., and P. R. Grant. 1984. Ecological Correlates of Morphological Evolution in a Darwin ' s Finch , Geospiza difficilis. Evolution (N. Y). 38:856–869.
Schrider, D. R., J. Ayroles, D. R. Matute, and A. D. Kern. 2018. Supervised machine learning reveals introgressed loci in the genomes of Drosophila simulans and D. sechellia. PLOS Genet. 14:e1007341. Public Library of Science.
Schrider, D. R., and A. D. Kern. 2016. S/HIC: Robust Identification of Soft and Hard Sweeps Using Machine Learning. PLOS Genet. 12:e1005928. Public Library of Science.
Schrider, D. R., F. K. Mendes, M. W. Hahn, and A. D. Kern. 2015. Soft Shoulders Ahead: Spurious Signatures of Soft and Partial Selective Sweeps Result from Linked Hard Sweeps. Genetics 200:267–284.
Schumer, M., R. Cui, D. L. Powell, R. Dresner, G. G. Rosenthal, and P. Andolfatto. 2014a. High-resolution mapping reveals hundreds of genetic incompatibilities in hybridizing fish species. Elife 2014:1–21.
Schumer, M., R. Cui, G. G. Rosenthal, and P. Andolfatto. 2015. Reproductive Isolation of Hybrid Populations Driven by Genetic Incompatibilities. PLoS Genet. 11:1–21.
Schumer, M., G. G. Rosenthal, and P. Andolfatto. 2014b. How common is homoploid hybrid speciation? Evolution (N. Y). 68:1553–1560.
Schumer, M., C. Xu, D. L. Powell, A. Durvasula, L. Skov, C. Holland, J. C. Blazier, S. Sankararaman, P. Andolfatto, G. G. Rosenthal, and M. Przeworski. 2018. Natural selection interacts with recombination to shape the evolution of hybrid genomes. Science (80-.).
Schwarzbach, A. E., and L. H. Rieseberg. 2002. Likely multiple origins of a diploid hybrid sunflower species. Mol. Ecol. 11:1703–1715.
Seberg, H. E., E. Van Otterloo, S. K. Loftus, H. Liu, G. Bonde, R. Sompallae, D. E. Gildea, J. F. Santana, J. R. Manak, J. Pavan, T. Williams, and R. A. Cornell. 2017. TFAP2 paralogs regulate melanocyte differentiation in parallel with MITF.
Seehausen, O. 2006. African cichlid fish: a model system in adaptive radiation research. Proc. R. Soc. Biol. Sci. 273:1987–98.
Seehausen, O. 2013. Conditions when hybridization might predispose populations for adaptive radiation. J. Evol. Biol. 26:279–281.
Seehausen, O. 2004. Hybridization and adaptive radiation. Trends Ecol. Evol. 19:198–207.
Seehausen, O. and, and C. E. Wagner. 2014. Speciation in Freshwater Fishes. Annu. Rev. Ecol. Evol. Syst. 45:621–651.
Servedio, M., and R. Bürger. 2018. The Effects on Parapatric Divergence of Linkage between Preference and Trait Loci versus Pleiotropy.
Servedio, M. R., and J. W. Boughman. 2017. The Role of Sexual Selection in Local Adaptation and Speciation. Annu. Rev. Ecol. Evol. Syst. 48:annurev-ecolsys-110316-022905.
Servedio, M. R., G. S. Van Doorn, M. Kopp, A. M. Frame, and P. Nosil. 2011. Magic traits in speciation: 'magic' but not rare? Trends Ecol. Evol. 26:389–397.
Shirley, M. D., H. Tang, C. J. Gallione, J. D. Baugher, L. P. Frelin, B. Cohen, P. E. North, D. A. Marchuk, A. M. Comi, and J. Pevsner. 2013. Sturge–Weber Syndrome and Port-Wine Stains Caused by Somatic Mutation in GNAQ. N. Engl. J. Med. 368:1971–1979. Massachusetts Medical Society.
Siepel, A., G. Bejerano, J. S. Pedersen, A. S. Hinrichs, M. Hou, K. Rosenbloom, H. Clawson, J. Spieth, L. D. W. Hillier, S. Richards, G. M. Weinstock, R. K. Wilson, R. A. Gibbs, W. J.

Kent, W. Miller, and D. Haussler. 2005. Evolutionarily conserved elements in vertebrate, insect, worm, and yeast genomes. Genome Res. 15:1034–1050.
Simão, F. A., R. M. Waterhouse, P. Ioannidis, E. V Kriventseva, and E. M. Zdobnov. 2015. BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. Bioinformatics 31:3210–3212.
Simpson, G. G. 1944a. Tempo and mode in evolution. Columbia University Press.
Simpson, G. G. 1944b. Tempo and mode of evolution. 15th ed. Columbia University Press.
Sinha, S., C. Mundy, T. Bechtold, F. Sgariglia, M. Ibrahim, P. C. Billings, K. Carroll, E. Koyama, K. B. Jones, and M. Pacifici. 2017. Unsuspected osteochondroma-like outgrowths in the cranial base of Hereditary Multiple Exostoses patients and modeling and treatment with a BMP antagonist in mice. 1–26.
Smith, J., G. Coop, M. Stephens, and J. Novembre. 2018. Estimating Time to the Common Ancestor for a Beneficial Allele. Mol. Biol. Evol. 35:1003–1017.
Smith, J. M., and J. Haigh. 1974. The hitch-hiking effect of a favourable gene. Genet. Res. 23:23–35. Cambridge University Press.
Smith, M. L. 1989. Cyprinodon nicholsi, a new pupfish from Hispaniola, and species characteristics of C. bondi Myers (Teleostei: Cyprinodontiformes). Am. Museum Novit. 1–10.
Smukowski, C. S., and M. A. F. Noor. 2011. Recombination rate variation in closely related species. Heredity (Edinb). 107:496–508. The Genetics Society.
Sorenson, M. D., K. M. Sefc, and R. B. Payne. 2003. Speciation by host switch in brood parasitic indigobirds. Nature 424:928–931. Macmillian Magazines Ltd.
Soria-Carrasco, V., Z. Gompert, A. A. Comeault, T. E. Farkas, T. L. Parchman, J. S. Johnston, C. A. Buerkle, J. L. Feder, J. Bast, T. Schwander, S. P. Egan, B. J. Crespi, and P. Nosil. 2014. Stick insect genomes reveal natural selection's role in parallel speciation. Science 344:738–42.
Soupene, E. 2008. ATP8A1 activity and phosphatidylserine transbilayer movement. J. Receptor. Ligand Channel Res. Volume 1:1–10.
Sparks, J. S. 2004. Molecular phylogeny and biogeography of the Malagasy and South Asian cichlids (Teleostei: Perciformes: Cichlidae). Mol. Phylogenet. Evol. 30:599–614.
Spoorendonk, K. M., J. Peterson-maduro, J. Renn, T. Trowe, S. Kranenbarg, C. Winkler, and S. Schulte-merker. 2008. Retinoic acid and Cyp26b1 are critical regulators of osteogenesis in the axial skeleton. 3774:3765–3774.
St. John, Michelle E., J. C. Dunker, Richards, Emilie J., S. Romero, and C. H. Martin. 2021. Parallel genetic changes underlie integrated craniofacial traits in an adaptive radiation of trophic specialist pupfishes. 1–46.
St. John, M. E., R. Holzman, and C. H. Martin. 2020. Rapid adaptive evolution of scale-eating kinematics to a novel ecological niche. J. Exp. Biol. 223:jeb217570.
St. John, M. E., J. A. McGirr, and C. H. Martin. 2019. The behavioral origins of novelty: did increased aggression lead to scale-eating in pupfishes? Behav. Ecol. 30:557–569.
Stamatakis, A. 2014. RAxML version 8: A tool for phylogenetic analysis and post-analysis of large phylogenies. Bioinformatics 30:1312–1313.
Stankowski, S., and M. A. Streisfeld. 2015. Introgressive hybridization facilitates adaptive divergence in a recent radiation of monkeyflowers. Proc. R. Soc. London B 282:20151666.
Stelkens, R. B., K. a Young, and O. Seehausen. 2010. The accumulation of reproductive incompatibilities in African cichlid fish. Evolution 64:617–33.

Stevenson, M. M. 1981. Karyomorphology of Several Species of Cyprinodon. Copeia 494–498.
Stölting, K. N., M. Paris, C. Meier, B. Heinze, S. Castiglione, D. Bartha, and C. Lexer. 2015. Genome-wide patterns of differentiation and spatially varying selection between postglacial recolonization lineages of *Populus alba* (Salicaceae), a widespread forest tree. New Phytol. 207:723–734. Wiley/Blackwell (10.1111).
Streelman, J. T., M. Alfaro, M. W. Westneat, D. R. Bellwood, and S. A. Karl. 2002. Evolutionary history of the parrotfishes: biogeography, ecomorphology, and comparative diversity. Evolution (N. Y). 56:961–971. John Wiley & Sons, Ltd.
Streelman, T. J., and P. D. Danley. 2003. The stages of vertebrate evolutionary radiation. Trends Ecol. Evol. 18:126–131.
Stroud, J. T., and J. B. Losos. 2016. Ecological Opportunity and Adaptive Radiation. Annu. Rev. Ecol. Evol. Syst. 47:507–532.
Suwa, A., M. Yoshino, C. Yamazaki, M. Naitou, and R. Fujikawa. 2010. RMI1 deficiency in mice protects from diet and genetic-induced obesity. 277:677–686.
Svennson, E., and R. Calsbeek. 2012. The adaptive landscape in evolutionary biology. Oxford University Press.
Tajima, F. 1989. Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. Genetics 123:585–595.
Takayanagi, Y., H. Matsumoto, M. Nakata, T. Mera, S. Fukusumi, S. Hinuma, Y. Ueta, T. Yada, G. Leng, and T. Onaka. 2008. Endogenous prolactin-releasing peptide regulates food intake in rodents. J. Clin. Invest. 118:4014–4024.
Tarvin, R. D., C. M. Borghese, W. Sachs, J. C. Santos, Y. Lu, L. A. O'Connell, D. C. Cannatella, R. A. Harris, and H. H. Zakon. 2017. Interacting amino acid replacements allow poison frogs to evolve epibatidine resistance. Science (80-.). 357:1261–1266.
Taye, M., J. Kim, S. H. Yoon, W. Lee, O. Hanotte, T. Dessie, S. Kemp, O. A. Mwai, K. Caetano-Anolles, S. Cho, S. J. Oh, H. K. Lee, and H. Kim. 2017. Whole genome scan reveals the genetic signature of African Ankole cattle breed and potential for higher quality beef. BMC Genet. 18:1–14. BMC Genetics.
Taylor, E. B., J. W. Boughman, M. Groenenboom, M. Sniatynski, D. Schluter, and J. L. Gow. 2006. Speciation in reverse: Morphological and genetic evidence of the collapse of a three-spined stickleback (Gasterosteus aculeatus) species pair. Mol. Ecol. 15:343–355.
Teng, C. S., M. C. Ting, D. T. Farmer, M. Brockop, R. E. Maxson, and J. G. Crump. 2018. Altered bone growth dynamics prefigure craniosynostosis in a zebrafish model of Saethre-Chotzen syndrome. Elife 7:1–23.
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, and C. D. Jiggins. 2012. Butterfly genome reveals promiscuous exchange of mimicry adaptations among species. Nature 487:94–98. Nature Publishing Group.
Thomas, T., A. K. Voss, K. Chowdhury, and P. Gruss. 2000. Querkopf, a MYST family histone acetyltransferase, is required for normal cerebral cortex development. Development 127:2537–2548. England.
Thomson, J. D., and P. Wilson. 2008. Explaining evolutionary shifts between bee and hummingbird pollination: convergence, divergence, and directionality. Int. J. Plant Sci. 169:23–38.
Thornton, K. R. 2019. Polygenic adaptation to an environmental shift: Temporal dynamics of variation under Gaussian stabilizing selection and additive effects on a single trait. Genetics 213:1513–1530.
Thorpe, R. S., Y. Surget-Groba, and H. Johansson. 2008. The relative importance of ecology and geographic isolation for speciation in anoles. Philos. Trans. R. Soc. B 363:3071–81.
Tournebize, R., V. Poncet, M. Jakobsson, Y. Vigouroux, and S. Manel. 2019. McSwan: A joint site frequency spectrum method to detect and date selective sweeps across multiple population genomes. Mol. Ecol. Resour. 19:283–295.
Trewavas, E., J. Green, and S. A. Corbet. 1972. Ecological studies on crater lakes in West Cameroon fishes of Barombi mbo. J. Zool. 167:41–95.
Turelli, M., N. H. Barton, and J. A. Coyne. 2001. Theory and speciation. Trends Ecol. Evol. 16:330–343.
Turissini, D. A., and D. R. Matute. 2017. Fine scale mapping of genomic introgressions within the Drosophila yakuba clade. PLOS Genet. 13:e1006971. Public Library of Science.
Turner, B. J., D. D. Duvernell, T. M. Bunt, and M. G. Barton. 2008. Reproductive isolation among endemic pupfishes (Cyprinodon) on San Salvador Island, Bahamas: Microsatellite evidence. Biol. J. Linn. Soc. 95:566–582.
Turner, C., Z. Blount, D. Mitchell, and R. Lenski. 2015. Evolution and coexistence in response to a key innovation in a long-term evolution experiment with Escherichia coli. bioRxiv 020958.
Uyeda, J. C., T. F. Hansen, S. J. Arnold, and J. Pienaar. 2011. The million-year wait for macroevolutionary bursts. Proc. Natl. Acad. Sci. U. S. A. 108:15908–15913.
Valdés González, A., L. Martínez Estévez, M. E. Ángeles Villeda, and G. Ceballos. 2020. The extinction of the Catarina pupfish Megupsilon aporus and the implications for the conservation of freshwater fish in Mexico. Oryx 54:154–160. Cambridge University Press.
Van der Auwera, G. A., M. 0. Carneiro, C. Hartl, R. Poplin, G. del Angel, A. Levy-Moonshine, T. Jordan, K. Shakir, D. Roazen, J. Thibault, E. Banks, G. V. Kiran, D. Altshuler, S. Gabriel, and M. A. DePristo. 2013. From FastQ Data to High-Confidence Variant Calls: The Genome Analysis Toolkit Best Practices Pipeline. Curr. Protoc. Bioinforma. 43:11.10.1-11.10.33. Wiley-Blackwell.
van Doorn, G. S., and F. J. Weissing. 2001. Ecological versus Sexual Selection Models of Sympatric Speciation: A Synthesis. Selection 2:17–40. Akadémiai Kiadó.
Van Doren, B. M., L. Campagna, B. Helm, C. J. Llera, I. J. Lovette, and M. Liedvogel. 2017. Correlated patterns of genetic diversity and differentiation across an avian family. Mol. Ecol. 26:3982–3997. Wiley/Blackwell (10.1111).
Vaser, R., I. Sovic, N. Nagarajan, and Š. Mile. 2017. Fast and accurate de novo genome assembly from long uncorrected reads. Genome Res. 1–10.

Vernot, B., and J. M. Akey. 2014. Resurrecting Surviving Neandertal Lineages from Modern Human Genomes. Science (80-.). 1245938.
Via, S. 2001. Sympatric speciation in animals: The ugly duckling grows up. Trends Ecol. Evol. 16:381–390.
Vijay, N., M. Weissensteiner, R. Burri, T. Kawakami, H. Ellegren, and J. B. W. Wolf. 2017. Genomewide patterns of variation in genetic diversity are shared among populations, species and higher-order taxa. Mol. Ecol. 26:4284–4295. Wiley/Blackwell (10.1111).
Volders, K., N. K, and V. M. Creemers J. 2011. The Autism candidate gene nuerobeachin encodes a scaffolding protein implicated in membrane trafficking and signaling. Curr. Mol. Med. 11:204–217.
Wagner, C. E., L. J. Harmon, and O. Seehausen. 2014. Cichlid species-area relationships are shaped by adaptive radiations that scale with area. Ecol. Lett. 17:583–592.
Wagner, C. E., L. J. Harmon, and O. Seehausen. 2012. Ecological opportunity and sexual selection together predict adaptive radiation. Nature 487:366–369. Nature Publishing Group.
Wagner, C. E., and A. R. McCune. 2009. Contrasting patterns of spatial genetic structure in sympatric rock-dwelling cichlid fishes. Evolution (N. Y). 63:1312–1326.
Wainwright, P. C., D. R. Bellwood, M. W. Westneat, J. R. Grubich, and A. S. Hoey. 2004. A functional morphospace for the skull of labrid fishes: Patterns of diversity in a complex biomechanical system. Biol. J. Linn. Soc. 82:1–25.
Wang, J. R., J. Holt, L. McMillan, and C. D. Jones. 2018. FMLRC: Hybrid long read error correction using an FM-index. BMC Bioinformatics 19:1–11. BMC Bioinformatics.
Wang, S., L. Ghibaudi, T. Hashemi, C. He, C. Strader, M. Bayne, H. Davis, and J. J. Hwa. 1998. 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.
Warren, R. L., C. Yang, B. P. Vandervalk, B. Behsaz, A. Lagman, S. J. M. Jones, and I. Birol. 2015. LINKS: Scalable, alignment-free scaffolding of draft genomes with long reads. Gigascience 4. GigaScience.
Weinreich, D. M., N. F. Delaney, M. A. DePristo, and D. L. Hartl. 2006. Darwinian Evolution Can Follow Only Very Few Mutational Paths to Fitter Proteins. Science (80-.). 312:111 LP – 114.
Weisenfeld, N. I., V. Kumar, P. Shah, D. M. Church, and D. B. Jaffe. 2018. Corrigendum: Direct determination of diploid genome sequences. Genome Res. 28:757–767.
Weissing, F. J., P. Edelaar, and G. S. van Doorn. 2011. Adaptive speciation theory: a conceptual review. Behav. Ecol. Sociobiol. 65:461–480.
Welch, M. E., and L. H. Rieseberg. 2002. Habitat divergence between a homoploid hybrid sunflower species, Helianthus paradoxus (Asteraceae), and its progenitors. Am. J. Bot. 89:472–478.
Wen, D., Y. Yu, M. W. Hahn, and L. Nakhleh. 2016. Reticulate evolutionary history and extensive introgression in mosquito species revealed by phylogenetic network analysis. Mol. Ecol. 25:2361–2372.
West, R. J. D., and A. Kodric-Brown. 2015. Mate Choice by Both Sexes Maintains Reproductive Isolation in a Species Flock of Pupfish (Cyprinodon spp) in the Bahamas. Ethology 121:793–800.
Whitney, K. D., R. A. Randell, and L. H. Rieseberg. 2010. Adaptive introgression of abiotic

tolerance traits in the sunflower Helianthus annuus. New Phytol. 187:230–239.
Whittaker, R. 1977. Evolution of species diversity in land communities. Evol. Biol.
Whittall, J. B., and S. A. Hodges. 2007. Pollinator shifts drive increasingly long nectar spurs in columbine flowers. Nature 447:706–709.
Wotton, D., and J. Massagué. 2001. Smad Transcriptional Corepressors in TGF$β$ Family Signaling. Pp. 145–164 *in* M. L. Privalsky, ed. Transcriptional Corepressors: Mediators of Eukaryotic Gene Repression. Springer Berlin Heidelberg, Berlin, Heidelberg.
Wright, K. M., D. Lloyd, D. B. Lowry, M. R. Macnair, and J. H. Willis. 2013. Indirect Evolution of Hybrid Lethality Due to Linkage with Selected Locus in Mimulus guttatus. PLoS Biol. 11.
Wu, C.-I. I. 2001a. The genic view of the process of speciation. J. Evol. Biol. 14:851–865.
Wu, C.-I., and C.-T. Ting. 2004. Genes and speciation. Nat. Rev. Genet. 5:114–122.
Wu, C. I. 2001b. The genic view of the process of speciation. J. Evol. Biol. 14:851–865.
Wu, F. L., M. Przeworski, P. Moorjani, M. Przeworski, A. I. Strand, L. A. Cox, L. A. Cox, C. Ober, J. D. Wall, A. I. Strand, and P. Moorjani. 2020. A comparison of humans and baboons suggests germline mutation rates do not track cell divisions.
Yan, W., A. H. Assadi, A. Wynshaw-Boris, G. Eichele, M. M. Matzuk, and G. D. Clark. 2003. Previously uncharacterized roles of platelet-activating factor acetylhydrolase 1b complex in mouse spermatogenesis. Proc. Natl. Acad. Sci. U. S. A. 100:7189–94.
Yates, A. D., P. Achuthan, W. Akanni, J. Allen, J. Allen, J. Alvarez-Jarreta, M. R. Amode, I. M. Armean, A. G. Azov, R. Bennett, J. Bhai, K. Billis, S. Boddu, J. C. Marugán, C. Cummins, C. Davidson, K. Dodiya, R. Fatima, A. Gall, C. G. Giron, L. Gil, T. Grego, L. Haggerty, E. Haskell, T. Hourlier, O. G. Izuogu, S. H. Janacek, T. Juettemann, M. Kay, I. Lavidas, T. Le, D. Lemos, J. G. Martinez, T. Maurel, M. McDowall, A. McMahon, S. Mohanan, B. Moore, M. Nuhn, D. N. Oheh, A. Parker, A. Parton, M. Patricio, M. P. Sakthivel, A. I. Abdul Salam, B. M. Schmitt, H. Schuilenburg, D. Sheppard, M. Sycheva, M. Szuba, K. Taylor, A. Thormann, G. Threadgold, A. Vullo, B. Walts, A. Winterbottom, A. Zadissa, M. Chakiachvili, B. Flint, A. Frankish, S. E. Hunt, G. Iisley, M. Kostadima, N. Langridge, J. E. Loveland, F. J. Martin, J. Morales, J. M. Mudge, M. Muffato, E. Perry, M. Ruffier, S. J. Trevanion, F. Cunningham, K. L. Howe, D. R. Zerbino, and P. Flicek. 2020. Ensembl 2020. Nucleic Acids Res. 48:D682–D688.
Yeaman, S., S. Aeschbacher, and R. Bürger. 2016. The evolution of genomic islands by increased establishment probability of linked alleles. Mol. Ecol. 25:2542–2558.
Zamani, N., P. Russell, H. Lantz, M. P. Hoeppner, J. R. Meadows, N. Vijay, E. Mauceli, F. Di Palma, K. Lindblad-Toh, P. Jern, and M. G. Grabherr. 2013. Unsupervised genome-wide recognition of local relationship patterns. BMC Genomics 14:347.
Zhang, J., S. Hagopian-Donaldson, G. Serbedzija, J. Elsemore, D. Plehn-Dujowich, A. P. McMahon, R. A. Flavell, and T. Williams. 1996. Neural tube, skeletal and body wall defects in mice lacking transcription factor AP-2. Nature 381:238–241.
Zhang, X., B. Kim, K. E. Lohmueller, and E. Huerta-Sánchez. 2020. The impact of recessive deleterious variation on signals of adaptive introgression in human populations. Genetics 215:799–812.
Zhao, X., T. Brade, T. J. Cunningham, and G. Duester. 2010. Retinoic acid controls expression of tissue remodeling genes Hmgn1 and Fgf18 at the digit–interdigit junction. Dev. Dyn. 239:665–671. John Wiley & Sons, Ltd.
Zhou, X., P. Carbonetto, and M. Stephens. 2013. Polygenic modeling with bayesian sparse linear

mixed models. PLoS Genet. 9:e1003264. Public Library of Science.