MOLECULAR ECOLOGY

Molecular Ecology (2014) **23**, 1846–1862 doi: 10.1111/mec.12658

# Novel trophic niches drive variable progress towards ecological speciation within an adaptive radiation of pupfishes

CHRISTOPHER H. MARTIN*† and LAURA C. FEINSTEIN†‡

*Department of Environmental Science, Policy, and Management, University of California, Berkeley, CA 94720, USA, †Department of Evolution & Ecology, University of California, Davis, CA 94616, USA, ‡California Council on Science and Technology, 1130 K Street, Sacramento, CA 95814, USA

## Abstract

**Adaptive radiation is recognized by a rapid burst of phenotypic, ecological and species diversification. However, it is unknown whether different species within an adaptive radiation evolve reproductive isolation at different rates. We compared patterns of genetic differentiation between nascent species within an adaptive radiation of *Cyprinodon* pupfishes using genotyping by sequencing. Similar to classic adaptive radiations, this clade exhibits rapid morphological diversification rates and two species are novel trophic specialists, a scale-eater and hard-shelled prey specialist (durophage), yet the radiation is <10 000 years old. Both specialists and an abundant generalist species all coexist in the benthic zone of lakes on San Salvador Island, Bahamas. Based on 13 912 single-nucleotide polymorphisms (SNPs), we found consistent differences in genetic differentiation between each specialist species and the generalist across seven lakes. The scale-eater showed the greatest genetic differentiation and clustered by species across lakes, whereas durophage populations often clustered with sympatric generalist populations, consistent with parallel speciation across lakes. However, we found strong evidence of admixture between durophage populations in different lakes, supporting a single origin of this species and genome-wide introgression with sympatric generalist populations. We conclude that the scale-eater is further along the speciation-with-gene-flow continuum than the durophage and suggest that different adaptive landscapes underlying these two niche environments drive variable progress towards speciation within the same habitat. Our previous measurements of fitness surfaces in these lakes support this conclusion: the scale-eating fitness peak may be more distant than the durophage peak on the complex adaptive landscape driving adaptive radiation.**

*Keywords*: adaptive radiation, ecological novelty, next-generation sequencing, parapatric speciation, population genomics, RADseq, sympatric speciation

*Received 6 May 2013; revision received 9 December 2013; accepted 20 December 2013*

## Introduction

Ecological speciation, the evolution of reproductive isolation due to divergent ecological selection, is generally believed to occur quickly relative to nonecological speciation (McCune 1997; Rundell & Price 2009; Schluter & Conte 2009). Indeed, greater ecological divergence between species pairs is correlated with greater reproductive isolation (Funk *et al.* 2006), and most examples of extremely rapid speciation and phenotypic diversification involve divergent ecology (Hendry & Kinnison 1999; Schluter 2000; Ackerly 2009; Martin & Wainwright 2011; but see Martin & Genner 2009; Arnegard *et al.* 2010).

However, ecological speciation rates vary greatly even among similar taxa or populations of the same

Correspondence: Christopher H. Martin;
E-mail: chmartin@berkeley.edu

species adapting to similar environments (e.g. Lovette *et al.* 2002; Arbogast *et al.* 2006; Seehausen *et al.* 2008; Berner *et al.* 2009, 2010; Hendry *et al.* 2009; Nosil *et al.* 2009; Merrill *et al.* 2011; Rosenblum & Harmon 2011; Martin 2012, 2013). Variable progress towards speciation reflects different positions along the speciation-with-gene-flow continuum, from complete panmixia among populations, to divergent ecotypes, to limited hybridization among phenotypically and ecologically distinct species, and finally cessation of all gene flow due to the build-up of intrinsic reproductive incompatibilities (Coyne & Orr 2004; Mallet 2008; Hendry *et al.* 2009; Nosil *et al.* 2009).

Differences in progress towards ecological speciation can frequently be explained by two categories of ecological mechanisms. First, variable rates of speciation between different taxa adapting to the same environment are often due to intrinsic differences in the spatial scale of gene flow among taxa (Kisel & Barraclough 2010), for example differences in dispersal distance or connectivity among habitat patches (Wagner & McCune 2009; Rosenblum & Harmon 2011). Second, variable rates of speciation among populations of the same species adapting to similar environments are often due to differences between these environments in the steepness of habitat gradients, such as stream-lake ecotones (Berner *et al.* 2009, 2010; Hendry *et al.* 2009), light gradients (Seehausen *et al.* 2008) or the intensity of predation (Nosil *et al.* 2009). These habitat gradients reflect the underlying selection surface: steeper gradients correspond to stronger disruptive selection (Bolnick & Lau 2008; Seehausen *et al.* 2008; Berner *et al.* 2009) which is predicted to drive faster rates of speciation (Dieckmann & Doebeli 1999; Turelli *et al.* 2001; Kirkpatrick & Ravigné 2002; Coyne & Orr 2004). However, variable progress towards ecological speciation driven by divergent dietary niches within the same habitat is rarely studied despite the expectation that these niches correspond to different underlying selection surfaces and the frequent observation that sympatric trophic differentiation and ecological novelty are major features of many classic adaptive radiations (Schluter 2000; Martin & Wainwright 2011, 2013a).

Here, we introduce a new case study of variable progress towards speciation driven by divergent dietary niches within a sympatric adaptive radiation of *Cyprinodon* pupfishes endemic to San Salvador Island, Bahamas. All three nascent species within this radiation coexist within the same benthic habitat in several shallow hypersaline lakes spread across the 20-km island. Differences in habitat gradients cannot explain variable speciation rates because all three species forage, hold territories and breed within the same habitat (Turner *et al.* 2008; Martin & Wainwright 2011). Dispersal distances also do not vary: all three species lay adhesive eggs which produce benthic larvae with no pelagic phase. Moreover, most of the lakes on San Salvador are small (Fig. 1) and shallow [maximum depth: 12 m

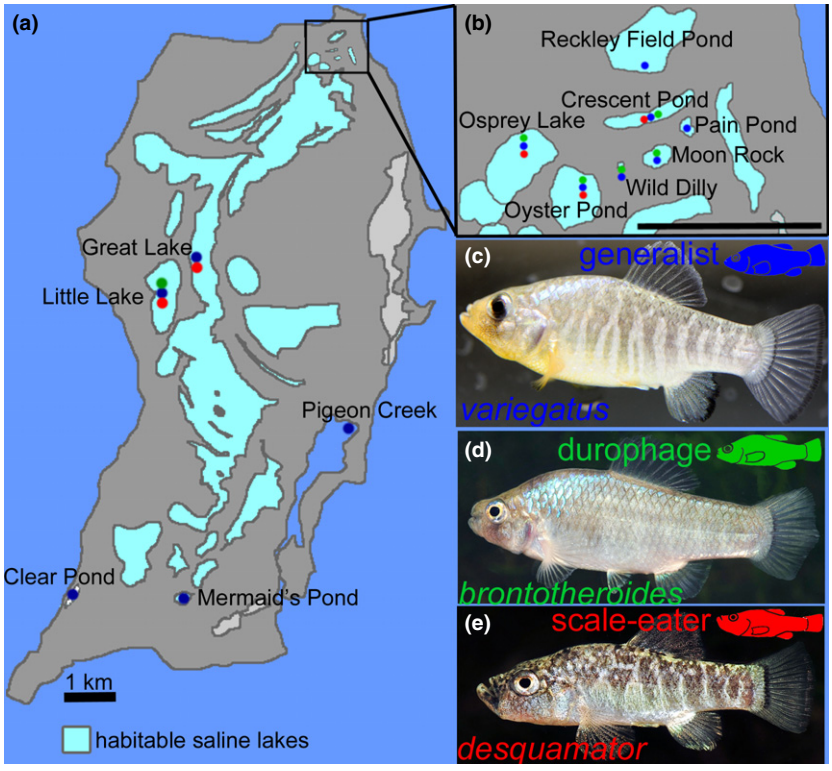


**Fig. 1** (a, b) Collection sites for populations sampled on San Salvador in habitable hypersaline lakes (turquoise) and Pigeon Creek estuary. (c) blue: generalist *variegatus* with typical trophic morphology for the genus; (d) green: durophage *brontotheroides* with unique nasal protrusion; (e) red: scale-eater *desquamator* with enlarged oral jaws. Paired or triplet dots indicate sampled sympatric *Cyprinodon* populations. Colours indicate species throughout the article. (d, e) Photos by Tony Terceira.

(Turner *et al.* 2008)] and do not provide adequate dispersal barriers for mobile fishes.

Instead, these species experience different selective regimes within the same habitat due to differences in their foraging ecology and diet. One newly described species, *Cyprinodon desquamator* (Martin & Wainwright 2013c), is a specialized scale-eating predator (lepidophage), which bites scales from other pupfishes with high-speed lateral strikes and greatly enlarged oral jaws. Multi-lake surveys of stomach contents and stable isotopes indicate that scales comprise at least 50% of its diet (Martin & Wainwright 2013a). A second specialist, *Cyprinodon brontotheroides* (Martin & Wainwright 2013c), frequently feeds on hard-shelled prey (e.g. ostracods and gastropods) and also has a unique species-diagnostic trophic morphology: its upper jaws are encased by a nasal protrusion formed from an enlarged anteriodorsal skeletal extension on the maxilla (Figs 1 and S1, Supporting information; Martin & Wainwright 2011, 2013a,c). This novel wedge-like appendage may facilitate dislodging gastropods from their shells (C. H. Martin, personal observation). The third species in the radiation, *Cyprinodon variegatus*, is a generalist algivore/invertivore, similar in morphology and diet to all other allopatric populations of *variegatus* across the Caribbean and Atlantic coast and likely similar to the common ancestor of the San Salvador radiation (Martin & Wainwright 2011). Populations of all three species in saline lakes on San Salvador form a clade (Martin & Wainwright 2011), and differences in trophic morphology are heritable (Martin & Wainwright 2013b). Geological evidence for the age of these saline lakes suggests this radiation is <10 000 years old (Hagey & Mylroie 1995; Turner *et al.* 2008; Martin & Wainwright 2011).

Despite its young age, this radiation has features of classic adaptive radiations: (i) rapid morphological diversification rates, (ii) ecological novelty and (iii) diversification on an adaptive landscape with multiple fitness peaks. First, several trophic traits, such as jaw length and adductor muscle mass, are evolving up to 51 times faster than background rates in other pupfishes, one of the fastest rates of trait diversification measured in any radiation (Martin & Wainwright 2011). Second, while scale-eating is known in other fish groups (Sazima 1983), this radiation contains the only specialized scale-eater within over 1500 species of Cyprinodontiformes; the most closely related species with convergent scale-eating ecology is separated by 168 million years of evolution, providing a quantitative measure of the rarity and novelty of this foraging niche (Martin & Wainwright 2013a). Third, measurements of the growth and survival of interspecific hybrids from this radiation placed in field enclosures on San Salvador suggest that competition on a complex adaptive landscape is driving their diversification (Martin & Wainwright 2013b). We found strong evidence for two fitness peaks on the hybrid morphospace corresponding to the phenotypes of two of the three species observed in the radiation (Martin & Wainwright 2013b). Thus, although these species have diverged on microevolutionary timescales placing them solidly in the realm of population genetics, *Cyprinodon* pupfishes exhibit the major features of classic adaptive radiation and we have detailed knowledge of the underlying adaptive landscape driving their diversification.

We evaluated levels of gene flow among the three species in this radiation across seven isolated lakes on San Salvador Island and outgroup populations of *C. variegatus* and related species using a next-generation genotyping-by-sequencing approach (GBS: Elshire *et al.* 2011), resulting in 13 912 SNP markers with at least 50% coverage across individuals. We contrasted levels of population structure between the two trophic specialist and abundant generalist species within and across lakes, scanned for divergent outlier loci and calculated admixture statistics among populations to address three questions: (i) Can divergent trophic niches drive variable progress towards ecological speciation between specialists within the same habitat? (ii) Did specialist ecotypes evolve multiple times independently in different lake populations or once followed by dispersal and genome-wide introgression? and (iii) Do additional habitat variables affect progress towards ecological speciation?

## Methods

### *Sampling*

We sampled 32 populations of *Cyprinodon* from across the Caribbean (Fig. 1, Table S1, Supporting information), including three formally described species within the *variegatus* complex (*variegatus*, *higuey* and *laciniatus*) and two outgroups (*dearborni* and *bondi*; Martin & Wainwright 2011), with a focus on the three sympatric species (*variegatus*, *brontheroides* and *desquamator*) endemic to saline lakes on San Salvador. Although we often sampled few individuals per population, this will not necessarily bias our estimates of population genetic differentiation when using unbiased estimators, such as Cockerham and Weir's $F_{ST}$ (1984) used in this study (Catchen *et al.* 2013), due to the large number of genetic markers sampled per individual (Willing *et al.* 2012). For example, genetic differentiation as small as $F_{ST} = 0.04$ can be detected with samples of only two individuals per population genotyped at 5000 markers (Willing *et al.* 2012: Table S2, Supporting information: 95% confidence interval = [−0.007, 0.033], mean = 0.011,

true $F_{ST}$ = 0.010). Furthermore, increasing the number of markers (without sampling more individuals) continues to reduce the error in estimates of genetic differentiation (Willing *et al.* 2012).

Populations of all three *Cyprinodon* species on San Salvador were collected in July 2008 and July 2011. Populations from the Dominican Republic (*C. bondi* and *C. higuey*) and surrounding Bahamian islands (New Providence, Rum Cay, Crooked Island) were collected in May 2011 and July 2011, respectively. Additional outgroups were acquired from the aquarium trade and private collectors. All specimens were euthanized in an overdose of MS-222 and preserved in 95% ethanol.

Extensive sampling on San Salvador Island revealed only six saline lakes plus the Great Lake system containing *desquamator* and/or *brontotheroides*, always in sympatry with *variegatus* (Fig. 1, Table S1, Supporting information). Four lakes contained all three species (Crescent Pond, Little Lake, Osprey Lake, Oyster Lake), the Great Lake system contained only *variegatus* and *desquamator* (Great Lake plus satellite lakes such as Mermaid Pond; *desquamator* in Mermaid Pond were not sampled for this study), and two lakes contained only *variegatus* and *brontotheroides* (Moon Rock Pond and Wild Dilly Pond). The latter population of *brontotheroides* was extremely small, and thus, only limited sampling was conducted (Wild Dilly Pond: $n$ = 3 in 2008; $n$ = 3 in 2011). Nearly all other permanent, habitable water bodies on the island were explored but found to contain only *variegatus* (Reckley Field Pond, Pain Pond, Clear Pond, Six Pack Pond, Pigeon Creek, Watling's blue hole, three additional blue holes, and two quarries; Fig. 1). Individuals within each population were chosen randomly for sequencing. All three *Cyprinodon* species on San Salvador were distinguished based on discrete differences in jaw morphology (Martin & Wainwright 2011, 2013b,c).

### *Molecular methods*

Genomic DNA was extracted from muscle tissue samples using a standard CTAB/chloroform extraction protocol. DNA concentration in each sample was measured on a fluorometer with PicoGreen (Life Technologies Inc.) and standardized across samples before library preparation. A double-digest RAD-seq library was prepared following the GBS protocol of Elshire *et al.* (2011) with an additional streptavidin-purification step. We used high-fidelity restriction enzyme SbfI for infrequent cutting and NlaIII for frequent cutting (New England Biolabs Inc.). 96 individual barcodes, ranging in length from 4 to 8 base pairs and differing by at least three mutational steps (further controls described in Elshire *et al.* 2011), were calculated using the Deena Bioinformatics GBS Barcode Generator (http://www.deenabio.com/services/gbs-adapters) and synthesized unmodified with standard purification (Life Technologies Inc.). We arranged the 96 individual samples randomly within the 96-well plate. After ligating barcoded adapters to cut fragments, we pooled the 96 individuals into one library (further described in Elshire *et al.* 2011). To attach Illumina adapters, pooled fragments were amplified in 12 independent 50 μL reaction volumes with Phusion high-fidelity DNA polymerase (Thermo Fisher Scientific, Inc.) in a thermocycler at 98 °C for 30 s, followed by 18 cycles of 98 °C for 30 s, 65 °C for 30 s, 72 °C for 30 s and a final extension step of 72 °C for 4 min. We added an additional purification step by biotinylating one of the PCR primers used for amplification. Fragments with common adapters on both ends do not incorporate biotin and were washed from the library after amplification by binding the biotinylated fragments to streptavidin beads (Pierce streptavidin-agarose resin: Thermo Fisher Scientific, Inc.). The distribution of fragment sizes in the final library was checked on an Agilent Bioanalyzer High Sensitivity Chip (Agilent Technologies, Inc.). The library was then sequenced using single-end chemistry to 150 base pairs on one lane of an Illumina HiSeq 2000 at the Vincent J. Coates Genomic Sequencing Laboratory, California Institute for Quantitative Biosciences.

### *Quality filtering and SNP-calling*

300 685 851 150-bp single-end raw sequence reads were visualized with FastQC (Babraham Bioinformatics, Babraham Institute) to assess declining quality scores relative to read length. Based on this assessment, reads were trimmed to 135 base pairs using the FastX-Toolkit (http://hannonlab.cshl.edu/fastx_toolkit/index.html) as subsequent base calls dropped to a median Phred quality score of 2 (50% base call accuracy). The Stacks pipeline (v. 99994 and v. 1.05 for population script; Catchen *et al.* 2011, 2013) was used to filter single-end reads and call SNPs. Reads with a mean Phred quality score below 10 (default option in Stacks pipeline) within a sliding window (20% of read length) or including any uncalled bases (N) were discarded. Reads were then sorted by individual barcode, automatically correcting single errors within barcodes and discarding ambiguous barcodes, before trimming the barcoded adapters from each read and truncating all reads to a length of 108 base pairs to eliminate terminal low-quality base calls. The final quality-filtered data set contained 136 023 370 108-bp reads.

For each individual, reads were first aligned *de novo* into stacks of homologous sequence reads with a minimum depth of 10 identical reads ($m$ = 10; *Stacks pipeline*

*parameters in italics*). Additional reads were then aligned to these stacks allowing for up to two mismatches ($M = 2$). This parameter only constrains pairwise differences, so stacks with more than two polymorphic sites are possible (Catchen *et al.* 2011; Wagner *et al.* 2012). Stacks with excessive numbers of reads (more than two standard deviations above the mean) were discarded as putative paralogous loci ($-t$). Consensus sequences from each stack within each individual were then used to build a catalog of loci across all individuals included in our analyses ($n = 73$; 23 individuals in the sequencing library were not used for this study). Individual SNPs were called based on their likelihood (Catchen *et al.* 2013) relative to this reference catalog of loci, allowing for up to two mismatches between an individual's consensus sequence and the catalog sequence ($n = 2$) and discarding ambiguous matches.

A total of 112 645 loci were identified across 73 individuals. This data set was reduced to 45 334 biallelic SNPs in 30 947 loci (mean: 1.5 SNPs per locus) by filtering to only those loci with at least 20 sequenced reads ($m = 20$). The observed number of well-sampled loci (30 947) matched reasonably well with the predicted 48 828 loci generated by an eight-base cutter in the 1.6 Gb *variegatus* genome ($1.6\ \text{Gb}/4^8$ cut sites $\times$ 2 assuming equal AT:GC ratio; Hinegardner & Rosen 1972). SNP data were exported from the Stacks pipeline in variant call format (.vcf) and converted into raw PLINK format (Purcell *et al.* 2007: http://pngu.mgh.harvard.edu/~purcell/plink/) using vcftools (Danecek *et al.* 2011: http://vcftools.sourceforge.net) for additional quality control. Four individuals were excluded due to low genotyping coverage (>95% missing data). SNPs were excluded if not genotyped in at least 50% of the remaining individuals. This resulted in a final data set of 13 912 SNPs with percentages of missing data for each individual ranging from 6% to 87% and with no apparent bias across populations (Table S2, Supporting information). Varying parameters of minimum stack depth or minimum genotyping coverage did not qualitatively affect our results.

*Genetic clustering analyses.* Hierarchical clustering among individuals and populations was analysed in three ways. First, we used a neighbour-joining tree based on genetic distances among individuals. We caution that this tree is purely a standard genetic clustering method for analysis of allele frequency differences between individuals (e.g. Gautier *et al.* 2010; Andrew *et al.* 2012) and should not be interpreted as a species-level phylogeny, particularly because the assumption of a bifurcating history is known to be violated in this case (see Table 2). A genetic distance matrix was calculated from the 13 912 SNP data set using the dist.gene function from the ape package (Paradis *et al.* 2004) in R (v. 2.15.2; R Core Team 2013). A neighbour-joining tree was then estimated from this distance matrix using the MVR* algorithm (Criscuolo & Gascuel 2008) implemented within the ape package in R. 1000 bootstrap samples from the SNP data matrix were used to assess nodal support. We also counted the number of times *desquamator* and *brontotheroides* individuals formed a single clade within each bootstrap sample using the is.monophyletic function in ape.

Second, we estimated a maximum-likelihood tree for all individuals from the concatenated 13 912 SNP data matrix using RAXML (v. 7.7.5; Stamatakis *et al.* 2005), similar to the approach of Jones *et al.* (2013). We conducted 1000 nonparametric bootstraps using the rapid-bootstrapping algorithm and GTR+CAT model of nucleotide substitution (Stamatakis 2006). However, we caution that no likelihood models in RAXML yet account for samples of only variable sites and ignore recombination among loci, which is known to bias phylogenetic inference. Furthermore, without invariant sites, branch lengths should be ignored (Lemmon and Lemmon 2013).

Third, we estimated a maximum-likelihood population tree for sympatric populations of *variegatus* and *brontotheroides* using TREEMIX (v. 1.12; Pickrell & Pritchard 2012). This tree provides a visualization of the majority–topology relationships later tested using Patterson's D-statistics (see below; Table 2). We only included lakes in which both species were sampled and in which these sympatric interspecific populations exhibited a sister relationship relative to allopatric populations (Crescent Pond, Little Lake, Moon Rock Lake) for later formal testing with the D-statistic. We also included the neighbouring *variegatus* population on Rum Cay as a closely related outgroup. To limit the effects of linkage disequilibrium, we exported only the first SNP from each locus for analysis. The maximum-likelihood population tree was then estimated from the allele frequencies of 1290 SNP's present in all focal populations.

Nonhierarchical clustering among individuals within saline lakes on San Salvador was visualized with principal components analysis (PCA; Patterson *et al.* 2006). To handle missing data in our SNP data set, we used probabilistic PCA implemented within the pcaMethods package (Stacklies *et al.* 2007) in R, which is designed to be robust to large amounts of missing data (Stacklies *et al.* 2007).

Finally, we used unsupervised Bayesian clustering (STRUCTURE v. 2.3.4; Pritchard *et al.* 2000; Falush *et al.* 2003) to estimate the posterior probability of population membership in each of $k$ genetic clusters for each individual. To limit the effects of linkage disequilibrium on

these analyses, we exported only the first SNP from each locus for analysis, resulting in 4202 SNPs with 50% coverage across all 56 individuals sampled from saline lakes on San Salvador. We evaluated the log likelihood of the SNP data set and $\Delta K$, the rate of change of the log likelihood (Evanno *et al.* 2005), for levels of population substructure from 1 to 12 (Fig. S2, Supporting information). For each value of $k$, we used the admixture model with correlated allele frequencies (Falush *et al.* 2003) and ran MCMC chains for 50 000 steps before calculating posterior probabilities of group membership from the next 50 000 steps. Likelihood values from at least two independent runs were averaged for each value of $k$ (Fig. S2, Supporting information). The largest value of $\Delta K$ was at $k = 2$, so we visualized population substructure for $k = 2$–4 from ten independent runs each ($k = 5$–7 is also presented in Fig. S3, Supporting information). Posterior samples from each run were aggregated using CLUMPP (v. 1.1.2; Jakobsson & Rosenberg 2007) and plotted in R.

*$F_{ST}$ outlier genomic scans.* We estimated the number of loci putatively under selection in each specialist species by comparing allele frequency differences at each SNP. We used a Bayesian approach implemented in BAYESCAN v. 2.1 (Foll & Gaggiotti 2008) which compares allele frequencies between subpopulations connected by a common migrant gene pool and computes the posterior probability of a neutral or selective model at each locus. This approach is robust to complex demographic scenarios, such as the variable subpopulation sizes examined here (Foll & Gaggiotti 2008). Although not accounting for hierarchical subpopulation structure is known to elevate the rate of false positives in outlier detection (Excoffier *et al.* 2009), we pooled individuals of each species for analyses due to our limited and unequal sample sizes within lakes and to avoid inflating false positives in recently bottlenecked populations (Foll & Gaggiotti 2008). We conducted independent outlier analyses for each trophic specialist, comparing all pooled populations of *desquamator* or *brontotheroides* to all *variegatus* populations on San Salvador Island. To further reduce false positives, we only used SNPs from loci with at least 10 sequenced reads, present in at least 50% of individuals in each species pool, and with a minor allele frequency of at least 25%, resulting in 5079 SNPs for our outlier analyses. We used prior odds of 10:1 for the neutral model relative to selective model at each SNP and the default options for running the MCMC search in Bayescan, which we conducted three times independently.

*Admixture analyses.* To test whether the *brontotheroides* ecotype evolved repeatedly in parallel across different lake populations, as suggested by the genetic distance and maximum-likelihood trees (Fig. 2), we looked for admixture between these populations that would be ignored by tree-based hierarchical clustering methods. We used Patterson's D-statistic to formally test for recent admixture between *brontotheroides* populations in different lakes (Reich *et al.* 2009; Green *et al.* 2010; Durand *et al.* 2011; Eaton & Ree 2013). The D-statistic is a measure of admixture between three closely related populations which is robust to random sequencing error and most other demographic assumptions (Durand *et al.* 2011). Consider a four-population pectinate tree with three populations of interest, A, B, C and an outgroup O [i.e. (((A,B),C),O)], indicating the hierarchical relationship between these populations based on the majority of markers; allele patterns such as aabb and bbaa on this tree support a sister relationship between populations A and B. The D-statistic compares the ratio of incongruent allele patterns which do not support the majority topology: alleles shared between A and C (baba) and alleles shared between B and C (abba). If no introgression has occurred between populations A and C or B and C since initial population divergence, then stochastic lineage sorting should result in equal ratios of incongruent allele patterns abba and baba. If introgression has occurred, the D-statistic measures how much the abba/baba ratio is skewed away from zero, implicating admixture between populations which share more incongruent alleles than expected by chance relative to the population tree (also see discussion in Eaton & Ree 2013; Pickrell & Pritchard 2012).

To visualize and direct our choice of focal populations for these tests, we first used TREEMIX (v. 1.12) to identify maximum-likelihood admixture events connecting populations on the tree of *variegatus* and *brontotheroides* populations (Pickrell & Pritchard 2012). This visualization is based on the same principle as the D-statistic and fits admixture events to those populations which share more incongruent alleles than expected by chance, relative to the maximum-likelihood tree estimated without admixture (Pickrell & Pritchard 2012). We compared the likelihood and AIC of 0-3 admixture events on the population tree (Table S4, Supporting information) and illustrated the position of the two most strongly supported admixture events with orange lines on the neighbour-joining and maximum-likelihood trees for comparison between panels in Fig. 2.

We calculated D-statistics for sympatric populations of *variegatus* and *brontotheroides* (A and B), an allopatric population of *variegatus* or *brontotheroides* in a different lake (C), and a closely related outgroup population (O: *variegatus* on Rum Cay). For clarity, we only included lakes in which both species were sampled and exhibited

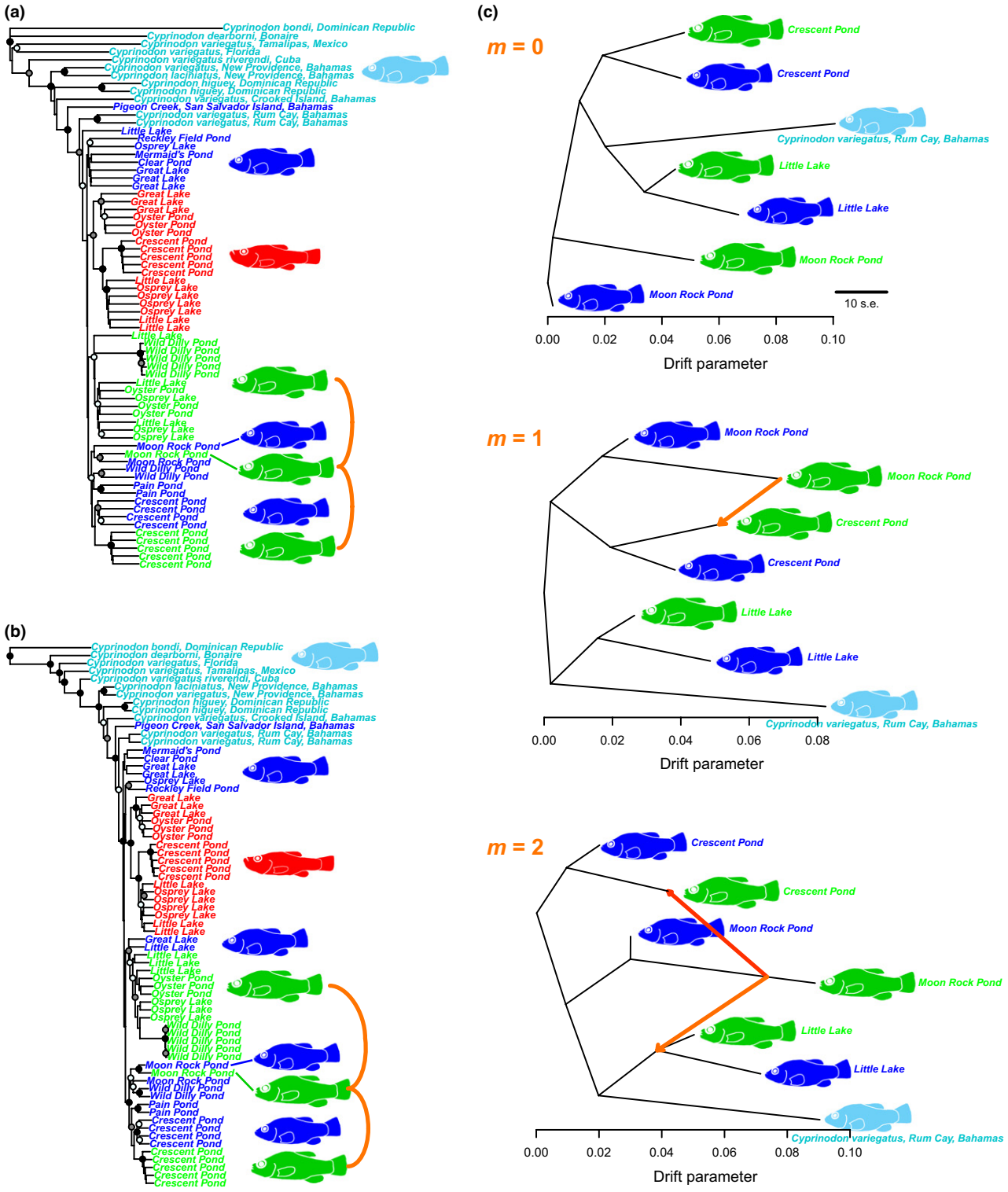


**Fig. 2** (a) Neighbour-joining tree based on genetic distances from 13 912 single-nucleotide polymorphisms with 1000 bootstrap replicates (bootstrap support: ●100%, ◉95%, ○80%) indicating hierarchical clustering relationships between the individuals and populations sampled (blue: *variegatus* on San Salvador Island, green: *brontotheroides*, red: *desquamator*, light blue: *variegatus* outgroups throughout Caribbean). (b) Maximum-likelihood tree based on 1000 bootstrap replicates (bootstrap support: ●100%, ◉95%, ○80%). (c) Maximum-likelihood trees for focal populations of *brontotheroides* and *variegatus* with 0, 1 and 2 admixture events (m: highlighted in orange; ≥2 admixture events are strongly supported by the data: Tables 2 and S4, Supporting information). Significant admixture events between populations (Tables 2 and S4, Supporting information) are indicated by the orange lines on each tree in a, b.

a sister relationship relative to populations in other lakes (Crescent Pond, Little Lake, Moon Rock Lake) based on our initial exploratory Treemix analyses. We also simplified these tests by only focusing on sympatric populations within a single lake, rather than pooling neighbouring lake populations. To limit linkage disequilibrium between markers, we sampled only one SNP per locus and only used SNPs genotyped in all populations examined (100% population coverage, number of sequenced reads per locus ≥10), resulting in a data set of 1290 SNPs. Patterson's D-statistics were calculated using the *fourpop* function within TREEMIX (v. 1.12; Pickrell & Pritchard 2012). SNP markers were not aligned to a reference genome; thus, we calculated standard error and significance by jackknifing each SNP marker independently ($k = 1$) rather than in blocks across a sliding window (Pickrell & Pritchard 2012).

*Effects of lake area and density on genetic differentiation.* We first calculated mean pairwise $F_{ST}$ and genetic diversity (average number of pairwise differences per site) for all lake populations on San Salvador (Table S3, Supporting information) using STACKS (v. 1.05; Catchen *et al.* 2013). We measured lake area and geographical distances between lakes using Google Earth and the Area Calculator Tool (http://www.daftlogic.com/projects-google-maps-area-calculator-tool.htm) and then used Mantel tests with 10 000 permutations to test the correlation between geographical distance and genetic differentiation ($F_{ST}$) between populations of each species. We also conducted visual censuses to estimate the relative abundance of each species by counting all individuals along 30 m × 0.3 m transects in the littoral zone (0.3 – 1 m). Censuses were conducted in July and averaged across three field seasons in 2008, 2011, and 2013 (Crescent Pond $n = 7$ transects, Little Lake $n = 6$, Moon Rock $n = 3$, Osprey Lake $n = 4$, Wild Dilly $n = 6$; censuses not available for Great Lake due to turbidity; total individuals counted = 7314).

## Results

### *Clustering analyses: variable genetic differentiation between trophic specialists within lakes*

All lake populations on San Salvador formed a clade on the neighbour-joining and maximum-likelihood trees (Fig. 2a,b). Intraspecific populations within each lake clustered together, supporting the genetic distinctiveness of each species in sympatry (Fig. 2a,b).

The most closely related *variegatus* population was in a lake on the nearest island Rum Cay (35 km distant), instead of the estuarine *variegatus* population on San Salvador (Fig. 1: Pigeon Creek), suggesting that interisland lake populations are more closely related than intraisland lake and estuarine populations, possibly reflecting more recent colonization of Pigeon Creek estuary.

All scale-eating *desquamator* individuals ($n = 18$) across five lakes on San Salvador formed a single clade with 100% bootstrap support in both the neighbour-joining and maximum-likelihood trees (Fig. 2a,b), strongly supporting a single origin of this species followed by dispersal across lakes and limited within-lake introgression. In contrast, all durophage *brontotheroides* individuals ($n = 20$) across six lake populations never formed a clade in any bootstrap sample. In two of six lakes, *brontotheroides* were more closely related to sympatric populations of *variegatus* than allopatric populations of *brontotheroides* (Fig. 2a,b: Crescent Pond and Moon Rock Pond). A sister relationship between sympatric populations of these two species was also supported by the maximum-likelihood population tree (Fig. 2c). This repeated pattern of interspecific clustering within lakes suggests either multiple independent origins of the *brontotheroides* ecotype within each lake or a single origin of *brontotheroides* followed by dispersal and extensive introgression with sympatric *variegatus* populations.

Likewise, PCA indicated that neighbouring genetic clusters of *desquamator* were more divergent from multiple and sometimes overlapping clusters of *variegatus* and *brontotheroides* (Fig. 3), except for the Wild Dilly Pond population of *brontotheroides* which was highly isolated from all other populations. This is the smallest saline lake on San Salvador (635 $m^2$; Fig. 1), and the durophage population occurs at very low frequency within this lake ($n = 2$ out of 2463 fish observed in six transects covering a total area of 42 $m^2$, suggesting a census size of 30 fish). The genetic distinctiveness of this population likely reflects a bottleneck following a recent colonization event, which is known to inflate genetic differentiation due to low within-population genetic diversity (Charlesworth 1998). Three distinct clusters of sympatric *brontotheroides* and *variegatus* were more evident after excluding this population from the PCA (Fig. 3b).

STRUCTURE analyses supported two genetic clusters in the SNP data for all individuals from lakes on San Salvador based on the rate of change in the log probability of the data, $\Delta K$ (Fig. S2, Supporting information); however, a model with three genetic clusters substantially improved the probability of the data (Fig. S2, Supporting information). The largest amount of population structure ($k = 2$) was between *brontotheroides* in Wild Dilly Pond and all other populations on San Salvador, consistent with the results from PCA (Fig. 4a). *Cyprinodon variegatus* and *brontotheroides* individuals also showed more evidence of introgression than *desquamator*

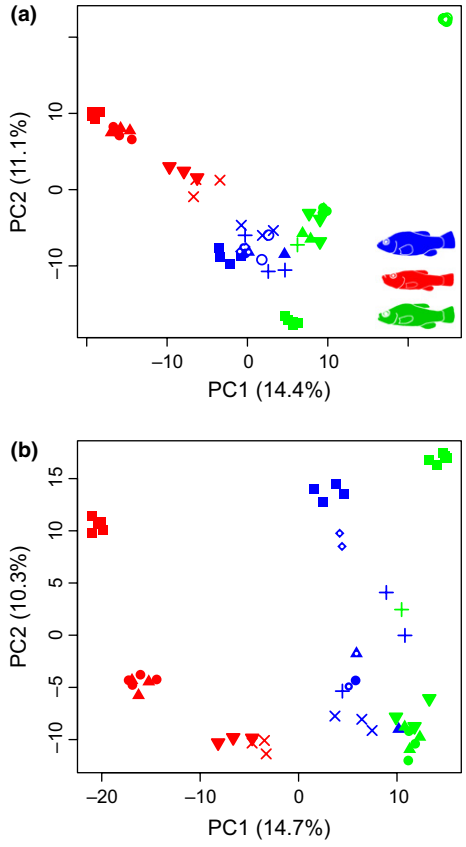


**Fig. 3** First two principal components of variation in 13 912 single-nucleotide polymorphisms across populations of all three *Cyprinodon* species on San Salvador (blue: *variegatus*, green: *brontotheroides*, red: *desquamator*). (a) All saline lake populations on San Salvador; (b) Wild Dilly Pond *brontotheroides* outlier population removed. Population symbols: ■ Crescent Pond, ▲ Little Lake, ● Osprey Lake, ▼Oyster Lake, x Great Lake, + Moon Rock Pond, ○Wild Dilly Pond; with white dot: ♦ Pain Pond, ●Clear Pond, ▲Reckley Field Pond, ■ Mermaid's Pond.

(Fig. 4a). Adding an additional genetic cluster ($k = 3$) partially separated *desquamator* individuals from other species and indicated very little introgression for both *desquamator* and *brontotheroides* in Crescent Pond. Four genetic clusters ($k = 4$) indicated further separation between *desquamator* and other species (Fig. 4c; additional genetic clustering among populations is visualized in Fig. S3, Supporting information).

### *Genomic outlier scans: differences in number of selected loci between trophic specialists*

We consistently identified a greater of number of loci putatively under disruptive selection in comparisons between *desquamator* and *variegatus* than between *brontotheroides* and *variegatus* (Table 1). In *desquamator-variegatus* comparisons, 14 SNPs (of 5079 total) supported a model of diversifying selection at a false discovery rate = 0.05 (Table 1); in *brontotheroides-variegatus* comparisons, only eight SNPs supported a model of diversifying selection at the same false discovery rate (Table 1). None of these selected loci were shared between the two comparisons, although the range of $F_{ST}$ values for outlier loci was similar (Table 1).

### *Admixture analyses: parallel speciation or admixture across lakes*

We found strong evidence of admixture between *brontotheroides* populations in different lakes based on formal tests of admixture using Patterson's D-statistics (Table 2) and visualized on the maximum-likelihood population tree (Fig. 2c; Table S4, Supporting information). Notably, eight of the ten possible four-population subtrees contrasting allopatric *brontotheroides* populations indicated highly significant evidence of admixture (Cb/Cb comparisons in Table 2: $P < 0.01$). In contrast, none of the four-population subtrees contrasting allopatric *variegatus* populations (Cv/Cv) or allopatric *variegatus* and *brontotheroides* populations (Cv/Cb) found significant evidence of admixture (Table 2), indicating that these populations do not share more incongruent alleles than expected by chance. Treemix analyses supported these conclusions: the maximum-likelihood tree with two admixture events connecting allopatric *brontotheroides* populations (Fig. 2c) was strongly supported over models with zero or one admixture event (Table S4, Supporting information).

Despite frequent clustering between sympatric *brontotheroides* and *variegatus* populations at the majority of SNP markers (Figs 2 and 3), *brontotheroides* populations in different lakes share more alleles than expected by chance. Therefore, parallel speciation of *brontotheroides* in each lake can be rejected in favour of a single origin of *brontotheroides* (supported by significant allele-sharing across lakes), followed by widespread introgression between sympatric populations of *variegatus* and *brontotheroides* (which results in genetic clustering by lake at the majority of loci).

### *Effect of ecological variables on genetic differentiation across lakes*

Genetic differentiation between sympatric *brontotheroides* and *variegatus* populations in different lakes varied widely ($F_{ST} = 0.12–0.49$: Table S3, Supporting information) and was marginally correlated with increased lake area (Fig. 5a; $r = 0.824$, $n = 5$, $P = 0.086$) and significantly negatively correlated with the census abundance of *brontotheroides* relative to *variegatus* (Fig. 5b; $r = 0.911$, $n = 5$, $P = 0.032$; with Wild Dilly outlier population

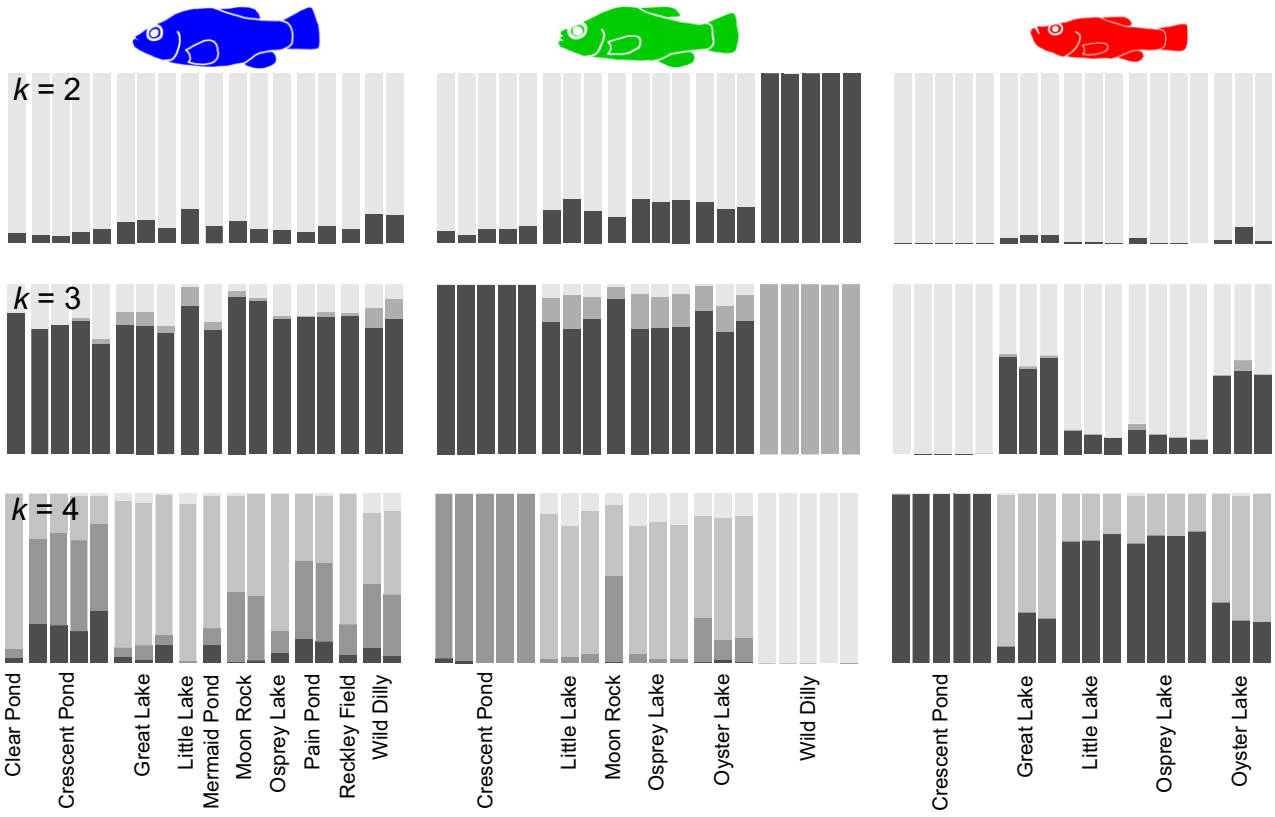


**Fig. 4** STRUCTURE analyses of 4202 single-nucleotide polymorphisms across saline lake populations on San Salvador, Bahamas. Large blocks group individuals by species (blue: *variegatus*, green: *brontotheroides*, red: *desquamator*). Genetic structure was visualized from 2 to 4 genetic clusters based on log likelihood and $\Delta K$ values (Fig. S2, Supporting information; $k$ = 5–7 genetic clusters: Fig. S3, Supporting information).

**Table 1** Bayesian $F_{ST}$ outlier scans identifying loci putatively under diversifying selection for comparisons between *desquamator* or *brontotheroides* and *variegatus* populations in lakes on San Salvador across 5079 SNPs at false discovery rates (FDR) of 1%, 5% and 10%. All outliers identified indicated diversifying selection (alpha > 0), rather than balancing selection (Foll & Gaggiotti 2008). $F_{ST}$ range is presented for all SNPs identified at FDR = 0.10

| | | Outliers under diversifying selection | | | |
|---|---|---|---|---|---|
| Comparison | Run | FDR = 0.10 | FDR = 0.05 | FDR = 0.01 | $F_{ST}$ range |
| *desquamator-variegatus* | 1 | 19 | 14 | 6 | 0.29–0.41 |
| | 2 | 19 | 14 | 6 | |
| | 3 | 19 | 14 | 6 | |
| *brontotheroides-variegatus* | 1 | 14 | 8 | 4 | 0.28–0.47 |
| | 2 | 14 | 8 | 3 | |
| | 3 | 14 | 8 | 3 | |

SNPs, single-nucleotide polymorphisms.

removed: $r$ = 0.867, $n$ = 4, $P$ = 0.133). These ecological variables appear to have no major effect on genetic differentiation between sympatric *desquamator* and *variegatus* populations (Fig. 5a,b), although we were limited by very low sample size ($n$ = 3). Larger lake areas were significantly correlated with increased genetic diversity (average number of pairwise differences per site) in populations of all three species (Fig. 5c; two-way ANOVA, $n$ = 22, log-area effect: $P$ = 0.015, species effect: $P$ = 0.153) and marginally associated with increased relative abundance of *brontotheroides* populations ($r$ = 0.681, $n$ = 5, $P$ = 0.205). The geographical distance between lakes on

**Table 2** Patterson's D-statistics provide evidence against parallel evolution of *brontotheroides* in different lakes on San Salvador due to recent admixture among these populations. Four-population pectinate subtrees (((A,B),C),O), test for admixture between a population of *variegatus* or *brontotheroides* within a single lake (A and B), and an allopatric population of *variegatus* or *brontotheroides* in a different lake (C), with an additional outgroup population (O: *Cyprinodon variegatus* on Rum Cay). Relationships between these populations supporting this topology can be visualized in Fig. 2c

| Admixture test | 4-population tree | D-statistic | SE | *z*-score | *P*-value |
|---|---|---|---|---|---|
| Cvar/Cbro | 1,**2**;**3**,O | −0.007 | 0.003 | −2.05 | 0.05 |
| **Cbro/Cbro** | 1,**2**;**4**,O | **−0.028** | 0.004 | **−7.02** | **$7e^{-12}$** |
| Cbro/Cvar | 1,**2**;**5**,O | −0.002 | 0.004 | −0.55 | 0.34 |
| **Cbro/Cbro** | 1,**2**;**6**,O | **−0.013** | 0.003 | **−3.99** | **$1e^{-4}$** |
| Cvar/Cvar | 1,**3**;**5**,O | −0.002 | 0.004 | −0.45 | 0.36 |
| Cvar/Cbro | 1,**3**;**6**,O | −0.005 | 0.003 | −1.47 | 0.14 |
| **Cbro/Cbro** | 1,**4**;**2**,O | **−0.014** | **0.005** | **−2.94** | **0.01** |
| **Cbro/Cbro** | 1,**4**;**6**,O | **−0.020** | **0.005** | **−4.41** | **$2e^{-5}$** |
| Cvar/Cvar | **1**,5;**3**,O | 0.004 | 0.004 | 0.90 | 0.23 |
| **Cbro/Cbro** | 1,**6**;**4**,O | **−0.013** | **0.005** | **−2.68** | **0.01** |
| Cbro/Cvar | 2,**4**;**5**,O | −0.008 | 0.004 | −1.83 | 0.07 |
| Cbro/Cbro | 2,**4**;**6**,O | −0.007 | 0.004 | −1.82 | 0.08 |
| **Cbro/Cbro** | **2**,5;**4**,O | **0.022** | **0.006** | **3.82** | **$2e^{-4}$** |
| **Cbro/Cbro** | 3,**4**;**2**,O | **−0.020** | 0.004 | **−4.46** | **$2e^{-5}$** |
| Cbro/Cvar | 3,**4**;**5**,O | −0.009 | 0.004 | −1.89 | 0.07 |
| **Cbro/Cbro** | 3,**4**;**6**,O | **−0.015** | 0.004 | **−4.04** | **$1e^{-4}$** |
| Cbro/Cbro | 5,**6**;**4**,O | −0.006 | 0.005 | −1.40 | 0.15 |

Cvar, *C. variegatus*; Cbro, *C. brontotheroides*. Population key: 1 = Crescent Pond *variegatus*, 2 = Crescent Pond *brontotheroides*, 3 = Moon Rock *variegatus*, 4 = Moon Rock *brontotheroides*, 5 = Little Lake *variegatus*, 6 = Little Lake *brontotheroides*, O: Rum Cay *variegatus*.
D-statistics supporting significant recent admixture are highlighted in bold. Negative D-statistics indicate admixture between populations B and C, positive D-statistics indicate admixture between populations A and C (Reich *et al.* 2009; Pickrell & Pritchard 2012); populations tested for admixture within each 4-population tree are highlighted in bold. All Cbro/Cbro tests should be highlighted in grey.

San Salvador was highly correlated with genetic differentiation between populations of *variegatus* (Mantel test, $r = 0.437$, $P = 0.011$), but not between populations of *brontotheroides* ($r = -0.558$, $P = 0.884$) or *desquamator* ($r = -0.577$, $P = 0.870$).

## Discussion

We found strong evidence of variable progress towards ecological speciation between trophic specialists within a nascent adaptive radiation of *Cyprinodon* pupfishes endemic to saline lakes on San Salvador Island, Bahamas. Two specialized species within this radiation, *Cyprinodon desquamator* and *Cyprinodon brontotheroides*, have recently adapted to highly unusual ecological niches, scale-eating and durophagy, respectively. In contrast to many other case studies of variable progress towards ecological speciation (Nosil & Crespi 2006; Seehausen *et al.* 2008; Berner *et al.* 2009, 2010; Merrill *et al.* 2011;

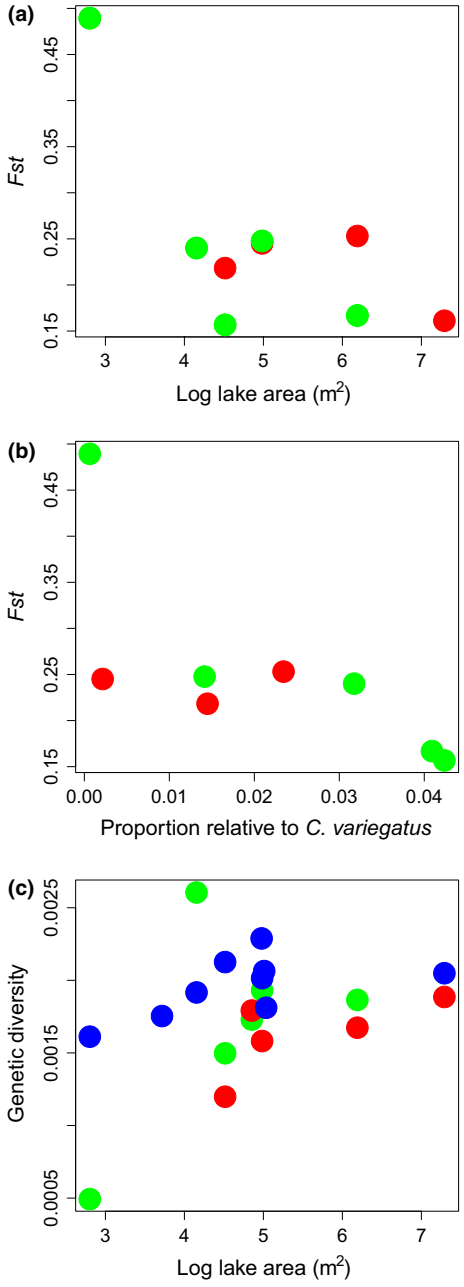


**Fig. 5** Genetic differentiation (genome-wide $F_{ST}$ from 13 912 single-nucleotide polymorphisms: Table S3, Supporting information) between sympatric populations of *variegatus* and *brontotheroides* (green) or *variegatus* and *desquamator* (red) relative to (a) lake area, (b) census abundance of each specialist relative to *variegatus* and (c) genetic diversity (average pairwise nucleotide divergence per site, π, for all populations (blue: *variegatus*, green: *brontotheroides*, red: *desquamator*) relative to log lake area.

Rosenblum & Harmon 2011), these species completely overlap in their microhabitat distribution within lakes with no barriers to gene flow due to habitat gradients or intrinsic differences in dispersal. Instead, their respective niche environments appear to drive the evolution of reproductive isolation at different rates within the same habitat.

We contrasted genetic differentiation between these trophic specialists and the abundant generalist species *Cyprinodon variegatus* within multiple lakes where these species coexist. The scale-eater *desquamator* was strongly isolated from other *Cyprinodon* species across all lakes surveyed and clustered together by species (Figs 2–5). This supports a single origin of *desquamator* with limited introgression following dispersal among lakes. In contrast, durophagous *brontotheroides* populations repeatedly clustered with sympatric *variegatus* populations (Figs 2–5). This interspecific clustering pattern is consistent with either multiple independent origins of the *brontotheroides* ecotype in different lake populations (as often concluded in previous studies, for example Gillespie 2004; Turner *et al.* 2008; also see Shaw 2002) or a single origin of *brontotheroides* followed by widespread introgression with sympatric *variegatus*. To distinguish these two scenarios, tests for admixture between populations in different lakes are needed (Reich *et al.* 2009; Durand *et al.* 2011; Pickrell & Pritchard 2012). We used Treemix and Patterson's D-statistics to visualize and test for admixture between *brontotheroides* populations in different lakes and found strong supporting evidence (Fig. 2c, Tables 2 and S4, Supporting information). Thus, at the majority of markers, these populations clustered with sympatric *variegatus* populations (Figs 2a,b–4), while excess allele-sharing at a minority of sites indicated that *brontheroides* populations in different lakes are more closely related than expected by chance (Fig. 2c, Table 2; also see discussion in Pickrell & Pritchard 2012; Eaton & Ree 2013). These results strongly reject the hypothesis of multiple independent origins of *brontotheroides* in different lakes. Instead, there appears to be widespread introgression between sympatric *variegatus* and *brontotheroides* populations. Species-diagnostic traits of *brontotheroides* may be limited to small regions of the genome, 'genomic islands of speciation' (Turner *et al.* 2005), which are shared across lakes. This conclusion was supported by genome-wide outlier scans, which found a greater number of loci putatively under disruptive selection in comparisons between *desquamator* and *variegatus* than between *brontotheroides* and *variegatus* (Table 1). Overall, this suggests that more regions of the genome are under selection during adaptation to scale-eating than to durophagy.

We conclude that *desquamator* is further along the speciation-with-gene-flow continuum than *brontotheroides*, despite their coexistence within the same habitat and origins within the same micro-endemic adaptive radiation. Why? Perhaps the simplest explanation is that *desquamator* is older than *brontotheroides* (e.g. Lovette *et al.* 2002; Berner *et al.* 2009; Elmer *et al.* 2010). However, we think this explanation is unlikely. First, the neighbour-joining and maximum-likelihood trees (Fig. 2a,b) suggest near simultaneous divergence of the three species after initial colonization of San Salvador by an ancestral population of *variegatus* within the past 10 000 years. Second, *brontotheroides* is distributed more broadly across saline lakes than *desquamator* (Fig. 1); rare dispersal among lakes (as indicated by high $F_{ST}$ values between lake populations: Fig. S3, Supporting information, Fig. 5) suggests that this pattern would be unlikely if *brontotheroides* were younger than *desquamator*. Third, we observed consistent differences in the relative amount of introgression between trophic specialists and sympatric *variegatus* populations across all lakes surveyed: where all three species occurred in sympatry, *brontotheroides* always exhibited more introgression than *desquamator* (Figs 4 and 5; Table S3, Supporting information). No post-zygotic intrinsic incompatibilities have been observed that would limit introgression among these species (Martin & Wainwright 2013b). Instead, this pattern suggests that extrinsic reproductive isolating barriers (such as selection against hybrids) are consistently stronger for *desquamator* than for *brontotheroides* across lake environments. Without such consistent extrinsic reproductive isolating barriers in each lake limiting introgression, sympatric populations would collapse into a hybrid swarm regardless of the times of origin of the two specialist species. This suggests that consistent niche environments across lakes, not time of origin, are driving variable rates of introgression between these trophic specialists.

### *Divergent trophic niches drive variable progress towards speciation*

We suggest that variable progress towards speciation between trophic specialists within the same habitat is driven by the divergent topography of the adaptive landscape between distinct niche environments. These two specialist niches exert highly contrasting performance demands which affect the underlying shape of the adaptive landscape driving adaptive radiation. Predatory scale-removal from evasive, fast-moving fishes requires a high-speed, accurate biting strike and must be balanced against a low net energy gain per successful strike, unlike piscivory (Sazima 1983; Martin & Wainwright 2013a). In contrast, foraging on attached, immobile prey, such as snails, may only require increased mechanical advantage of the lower jaw simple

lever system for increased crushing force (Holzman *et al.* 2012; although note that *brontotheroides* has a novel nasal appendage that may facilitate snail-crushing: Fig. S1, Supporting information). Furthermore, scales are an entirely novel resource throughout the evolutionary history of Cyprinodontiform fishes: the most closely related specialized scale-eater is separated by 168 million years of evolution from *C. desquamator*, providing a quantitative measure of the novelty of this niche (Martin & Wainwright 2013a). In contrast, hard-shelled prey such as gastropods and ostracods are more prevalent in the diet of *brontotheroides*, but are also common components of the diet of *variegatus* and many other *Cyprinodon* species (Martin & Wainwright 2013a).

These divergent performance requirements and ecologies between specialist niches are reflected by different topographies on the phenotypic adaptive landscape. For example, scale-eating may correspond to a steeper fitness peak if this niche requires a very specific, specialized trophic morphology and body shape for successful strikes (Martin & Wainwright 2013a). Steeper fitness peaks could drive more rapid evolution of reproductive isolation than shallow fitness peaks due to steeper disruptive selection gradients (Dieckmann & Doebeli 1999; Kirkpatrick & Ravigné 2002). Second, if novel ecological niches, such as scale-eating, correspond to more divergent phenotypes, this could also contribute to greater reproductive isolation due to a larger fitness valley isolating this niche from others (i.e. stronger ecological selection against scale-eater hybrids). Third, the dimensionality of selection may be higher (more 'multifarious') for complex prey capture behaviours such as scale-eating relative to durophagy (Nosil *et al.* 2009). All three of these attributes of fitness peaks on the adaptive landscape (steepness, distance, and dimensionality) affect progress towards ecological speciation, but are rarely measured (but see Bolnick & Lau 2008; Nosil *et al.* 2009; Martin 2012, 2013; Martin & Wainwright 2013b).

Field measurements of the adaptive landscape on San Salvador support these ideas. Survival and growth of F2 hybrids placed in field enclosures revealed multiple fitness peaks on the pupfish adaptive landscape (Martin & Wainwright 2013b). Hybrids most closely resembling the phenotypes of *variegatus* and *brontotheroides*, respectively, corresponded to neighbouring fitness peaks separated by a small fitness valley; in turn, these two fitness peaks were separated by a much larger fitness valley from hybrids resembling *desquamator*. This large fitness valley provides strong evidence that selection against scale-eating hybrids is greater than selection against durophage hybrids. This inference holds despite our lack of evidence for a third fitness peak corresponding to *desquamator*, which may lie outside the hybrid morphospace examined in our study (Martin & Wainwright 2013b).

### *Variable progress towards speciation across different lake environments*

Despite the small size of the island, hypersaline lakes on San Salvador Island differ dramatically in size, depth, turbidity and faunal/floral diversity (Godfrey 1994). Indeed, *brontotheroides* populations display distinct trophic morphologies in different lakes (Fig. S1, Supporting information), perhaps reflecting divergent diets of ostracods or gastropods (Martin & Wainwright 2013a). Biotic and abiotic differences between lakes may be driven by differences in connectivity of the lakes to each other and subterranean tidal exchanges with the sea. We found a significant correlation between geographical distance among lakes and genetic differentiation between *variegatus* populations, but this isolation-by-distance pattern did not explain divergence among *brontotheroides* or *desquamator* populations. Instead, we found a significant negative correlation between the abundance of *brontotheroides* relative to *variegatus* and the extent of genetic differentiation between these species (Fig. 5b). This pattern appears to suggest that as *brontotheroides* become rarer, they become more reproductively isolated from *variegatus*. However, the relative abundance of *brontotheroides* was correlated with lake area and genetic diversity, a rough estimate of effective population size, which suggests that *brontotheroides* populations with low relative abundance are also smaller (the extreme example being Wild Dilly Pond with an estimated census size of 30 individuals). Smaller populations are more likely to go locally extinct which implies that smaller *brontotheroides* populations are more likely to be recent colonists. Large population bottlenecks due to recent colonization are known to inflate estimates of genetic differentiation between populations (e.g. Charlesworth 1998). Thus, we propose that genetic differentiation is inflated in *brontotheroides* populations with low relative abundance due to an artifact of recent population bottlenecks in small lakes.

### *Additional reproductive isolating barriers driving variable progress towards speciation*

We cannot rule out additional nonecological differences between these specialists species which may explain their variable progress towards speciation, such as different cues used in premating isolation (Kodric-Brown & Strecker 2001) or differences in genetic architecture such as inversions or magic traits (Servedio *et al.* 2011). Indeed, our preliminary mate preference trials and field observations suggest that *desquamator* females show stronger preferences for conspecific males than *brontotheroides* females (C. H. Martin, unpublished data). Thus, *desquamator* is more strongly isolated due to both

stronger selection against hybrids (Martin & Wainwright 2013b) and premating isolation, in agreement with the lower levels of gene flow inferred from this study.

## Conclusion

Ecological speciation is not only driven at variable rates between different habitats or when additional selective agents are in play. Here, we show that different niche environments within the same habitat can also drive variable rates of speciation within a sympatric adaptive radiation. If we view speciation as a positive feedback loop (Servedio & Saetre 2003), different niches may drive this loop at different rates. From this top-down perspective, different ecologies exert varying constraints on performance which affect the strength of selection against intermediate ecotypes. This variable selection against hybrids should drive reinforcement of premating isolation at different rates and ultimately result in different rates of speciation and the evolution of postzygotic intrinsic incompatibilities (but see Rabosky & Matute 2013).

This case study also suggests an interesting contrast between ecological rarity and speciation rate. The rarity of scale-eating within Cyprinodontiformes across similar environments suggests that this niche may be extremely hard to colonize in contrast to other specialist niches such as durophagy or piscivory (Martin & Wainwright 2013a). However, once colonized, the evolution of reproductive isolating barriers appears to occur more rapidly within the scale-eating niche than in other specialized niches. Reproductive isolation may evolve faster within niches corresponding to more distant peaks on the adaptive landscape due to stronger selection against hybrids. However, distant fitness peaks may also be more difficult for a population to initially colonize. Thus, rare niches may drive speciation faster. Furthermore, if most nascent species are 'ephemeral' and more vulnerable to extinction (see Uyeda *et al.* 2011; Rosenblum *et al.* 2012) until evolving intrinsic reproductive incompatibilities, we predict that species in rare niches are also more likely to persist due to faster speciation rates. Overall, our case study of variable progress towards ecological speciation driven by different trophic specializations demonstrates an important new ecological dimension in this emerging field: the niche environment in which selection occurs.

## Acknowledgements

This study was funded by an NSF Graduate Research Fellowship, an NSF Dissertation Improvement Grant, an ARCS foundation award, a UCDavis dissertation year fellowship, the Center for Population Biology, the Miller Institute for Basic Sciences, and Gerace Research Centre student grants to CHM. We thank the governments of the Bahamas and the Dominican Republic for permission to conduct this research and the Gerace Research Centre and Kelvin Guerrero for help obtaining permits. Erica Rosenblum, Thomas Schoener, Sarah Longo, Alisa Sedghifar and Sean Rogers provided thoughtful comments on the manuscript. Rick Grosberg, Brenda Cameron, Noelia Carrasquilla, Doug Cook, Don Strong and Sergey Nuzhdin generously provided laboratory space, equipment or assistance developing molecular protocols. Graham Coop and Jeffrey Ross-Ibarra provided helpful advice on library design and statistical analyses. Minyong Chun at the VCGSL kindly advised on sequencing. Al Morales, Chris Butcher and Joseph Scanlan donated additional specimens used for this study. Finally, we thank the Bahamian and Dominican people for their hospitality and logistical support during fieldwork.

## References

Ackerly D (2009) Conservatism and diversification of plant functional traits: evolutionary rates versus phylogenetic signal. *Proceedings of the National Academy of Sciences USA*, **106**, 19699–19706.

Andrew RL, Kane NC, Baute GJ *et al.* (2012) Recent nonhybrid origin of sunflower ecotypes in a novel habitat. *Molecular Ecology*, **22**, 799–813.

Arbogast BS, Drovetski SV, Curry RL *et al.* (2006) The origin and diversification of Galapagos mockingbirds. *Evolution*, **60**, 370–382.

Arnegard ME, McIntyre PB, Harmon LJ *et al.* (2010) Sexual signal evolution outpaces ecological divergence during electric fish species radiation. *The American Naturalist*, **176**, 335–356.

Berner D, Grandchamp A, Hendry AP (2009) Variable progress toward ecological speciation in parapatry: stickleback across eight lake-stream transitions. *Evolution*, **63**, 1740–1753.

Berner D, Roesti M, Hendry AP, Salzburger W (2010) Constraints on speciation suggested by comparing lake-stream stickleback divergence across two continents. *Molecular Ecology*, **19**, 4963–4978.

Bolnick DI, Lau OL (2008) Predictable patterns of disruptive selection in stickleback in postglacial lakes. *The American Naturalist*, **172**, 1–11.

Catchen JM, Amores A, Hohenlohe P, Cresko W, Postlethwait JH (2011) Stacks: building and genotyping loci de novo from short-read sequences. *G3: Genes, Genomes, Genetics*, **1**, 171–182.

Catchen J, Hohenlohe PA, Bassham S *et al.* (2013) Stacks: an analysis tool set for population genomics. *Molecular Ecology*, **22**, 3124–3140.

Charlesworth B (1998) Measures of divergence between populations and the effect of forces that reduce variability. *Molecular Biology and Evolution*, **15**, 538–543.

Core Team R (2013) *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria.

Coyne JA, Orr HA (2004) *Speciation*. Sinauer Associates Inc, Sunderland, Massachussetts.

Criscuolo A, Gascuel O (2008) Fast NJ-like algorithms to deal with incomplete distance matrices. *BMC Bioinformatics*, **9**, 166.

Danecek P, Auton A, Abecasis G *et al.* (2011) The variant call format and VCFtools. *Bioinformatics*, **27**, 2156–2158.

Dieckmann U, Doebeli M (1999) On the origin of species by sympatric speciation. *Nature*, **400**, 354–357.

Durand EY, Patterson N, Reich D, Slatkin M (2011) Testing for ancient admixture between closely related populations. *Molecular Biology and Evolution*, **28**, 2239–2252.

Eaton DAR, Ree RH (2013) Inferring phylogeny and introgression using RADseq data: an example from flowering plants (Pedicularis: Orobanchaceae). *Systematic Biology*, **62**, 689–706.

Elmer KR, Lehtonen TK, Kautt AF, Harrod C, Meyer A (2010) Rapid sympatric ecological differentiation of crater lake cichlid fishes within historic times. *BMC Biology*, **8**, e60.

Elshire R, Glaubitz J, Sun Q (2011) A robust, simple genotyping-by-sequencing (GBS) approach for high diversity species. *PLoS ONE*, **6**, e19379.

Evanno G, Regnaut S, Goudet J (2005) Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study. *Molecular Ecology*, **14**, 2611–2620.

Excoffier L, Hofer T, Foll M (2009) Detecting loci under selection in a hierarchically structure population. *Heredity*, **103**, 285–298.

Falush D, Stephens M, Pritchard JK (2003) Inference of population structure using multilocus genotype data: linked loci and correlated allele frequencies. *Genetics*, **164**, 1567–1587.

Foll M, Gaggiotti O (2008) A genome-scan method to identify selected loci appropriate for both dominant and codominant markers: a Bayesian perspective. *Genetics*, **180**, 977–993.

Funk DJ, Nosil P, Etges WJ (2006) Ecological divergence exhibits consistently positive associations with reproductive isolation across disparate taxa. *Proceedings of the National Academy of Sciences USA*, **103**, 3209–3213.

Gautier M, Laloe D, Moazami-Goudarzi K (2010) Insights into the genetic history of French cattle from dense SNP data on 47 worldwide breeds. *PLoS ONE*, **5**, e13038.

Gillespie R (2004) Community assembly through adaptive radiation in Hawaiian spiders. *Science*, **303**, 356–359.

Godfrey PJ (1994) Natural history of northeastern San Salvador Island: a "New World" where the New World began. Bahamian Field Station, 1994.

Green RE, Krause J, Briggs AW *et al.* (2010) A draft sequence of the Neandertal genome. *Science*, **328**, 710–722.

Hagey FM, Mylroie JE (1995) Pleistocene history of the Bahamas. In: *Terrestrial and Shallow Marine Geology of the Bahamas and Bermuda* (eds Curran HA, White B). Geological Society of America Special Papers, 300, 77–90

Hendry AP, Kinnison MT (1999) Perspective: the pace of modern life: measuring rates of contemporary microevolution. *Evolution*, **53**, 1637–1653.

Hendry AP, Bolnick DI, Berner D, Peichel CL (2009) Along the speciation continuum in sticklebacks. *Journal of Fish Biology*, **75**, 2000–2036.

Hinegardner R, Rosen DE (1972) Cellular DNA content and the evolution of teleostean fishes. *American Naturalist*, **1972**, 621–644.

Holzman R, Collar DC, Price SA, Hulsey CD, Thomson RC, Wainwright PC (2012) Biomechanical trade-offs bias rates of evolution in the feeding apparatus of fishes. *Proceedings of the Royal Society of London B: Biological Sciences*, **279**, 1287–1292.

Jakobsson M, Rosenberg NA (2007) CLUMPP: a cluster matching and permutation program for dealing with label switching and multimodality in analysis of population structure. *Bioinformatics*, **23**, 1801–1806.

Jones JC, Fan S, Franchini P, Schartl M, Meyer A (2013) The evolutionary history of Xiphophorus fish and their sexually selected sword: a genome-wide approach using restriction site-associated DNA sequencing. *Molecular Ecology*, **22**, 2986–3001.

Kirkpatrick M, Ravigné V (2002) Speciation by natural and sexual selection: models and experiments. *The American Naturalist*, **159**, S22–S35.

Kisel Y, Barraclough T (2010) Speciation has a spatial scale that depends on levels of gene flow. *The American Naturalist*, **175**, 316–334.

Kodric-Brown A, Strecker U (2001) Responses of *Cyprinodon maya* and *C. labiosus* females to visual and olfactory cues of conspecific and heterospecific males. *Biological Journal of the Linnean Society*, **74**, 541–548.

Lemmon EM, Lemmon AR (2013) High-throughput genomic data in systematics and phylogenetics. *Annual Review of Ecology, Evolution, and Systematics*, **44**, 727–744.

Lovette IJ, Bermingham E, Ricklefs RE (2002) Clade-specific morphological diversification and adaptive radiation in Hawaiian songbirds. *Proceedings of the Royal Society of London B: Biological Sciences*, **269**, 37–42.

Mallet J (2008) Hybridization, ecological races and the nature of species: empirical evidence for the ease of speciation. *Philosophical Transactions of the Royal Society B: Biological Sciences*, **363**, 2971–2986.

Martin CH (2012) Weak disruptive selection and incomplete phenotypic divergence in two classic examples of sympatric speciation: Cameroon crater lake cichlids. *American Naturalist*, **180**, E90–E109.

Martin CH (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*, **67**, 2114–2123.

Martin CH, Genner MJ (2009) High niche overlap between two successfully coexisting pairs of Lake Malawi cichlids. *Canadian Journal of Fisheries and Aquatic Sciences*, **66**, 579–588.

Martin CH, Wainwright PC (2011) Trophic novelty is linked to exceptional rates of morphological diversification in two adaptive radiations of *Cyprinodon* pupfishes. *Evolution*, **65**, 2197–2212.

Martin CH, Wainwright PC (2013a) On the measurement of ecological novelty: scale-eating pupfish are separated by 168 my from other scale-eating fishes. *PLoS ONE*, **8**, e71164.

Martin CH, Wainwright PC (2013b) Multiple fitness peaks on the adaptive landscape drive adaptive radiation in the wild. *Science*, **339**, 208–211.

Martin CH, Wainwright PC (2013c) A remarkable species flock of *Cyprinodon* pupfishes endemic to San Salvador Island, Bahamas. *Bulletin of the Peabody Museum of Natural History*, **54**, 231–240.

McCune AR (1997) How fast is speciation? Molecular, geological, and phylogenetic evidence from adaptive radiations of fishes. In: *Molecular Evolution and Adaptive Radiation* (eds Givnish T, Sytsma T), pp. 585–610. Cambridge University Press, Cambridge.

Merrill R, Gompert Z, Dembeck L (2011) Mate preference across the speciation continuum in a clade of mimetic butterflies. *Evolution*, **65**, 1489–1500.

Nosil P, Crespi BJ (2006) Experimental evidence that predation promotes divergence in adaptive radiation. *Proceedings of the National Academy of Sciences USA*, **103**, 9090–9095.

Nosil P, Harmon LJ, Seehausen O (2009) Ecological explanations for (incomplete) speciation. *Trends in Ecology and Evolution*, **24**, 145–156.

Paradis AE, Claude J, Strimmer K (2004) APE: analysis of phylogenetics and evolution in R language. *Bioinformatics*, **20**, 289–290.

Patterson N, Price AL, Reich D (2006) Population structure and eigenanalysis. *PLoS Genetics*, **2**, e190.

Pickrell JK, Pritchard JK (2012) Inference of population splits and mixtures from genome-wide allele frequency data. *PLoS Genetics*, **8**, e1002967.

Pritchard JK, Stephens M, Donnelly P (2000) Inference of population structure using multilocus genotype data. *Genetics*, **155**, 945–959.

Purcell S, Neale B, Todd-Brown K (2007) PLINK: a tool set for whole-genome association and population-based linkage analyses. *The American Journal of Human Genetics*, **81**, 559–575.

Rabosky DL, Matute DR (2013) Macroevolutionary speciation rates are decoupled from the evolution of intrinsic reproductive isolation in *Drosophila* and birds. *Proceedings of the National Academy of Sciences USA*, **110**, 15354–15359.

Reich D, Thangaraj K, Patterson N, Price A, Singh L (2009) Reconstructing Indian population history. *Nature*, **461**, 489–494.

Rosenblum EB, Harmon LJ (2011) "Same same but different": replicated ecological speciation at White Sands. *Evolution*, **65**, 946–960.

Rosenblum EB, Sarver BA, Brown JW *et al.* (2012) Goldilocks meets Santa Rosalia: an ephemeral speciation model explains patterns of diversification across time scales. *Evolutionary Biology*, **39**, 255–261.

Rundell RJ, Price TD (2009) Adaptive radiation, nonadaptive radiation, ecological speciation and nonecological speciation. *Trends in Ecology & Evolution*, **24**, 394–399.

Sazima I (1983) Scale-eating in characoids and other fishes. *Environmental Biology of Fishes*, **9**, 87–101.

Schluter D (2000) *Ecology of Adaptive Radiation*. Oxford University Press, New York City, New York.

Schluter D, Conte GL (2009) Genetics and ecological speciation. *Proceedings of the National Academy of Sciences USA*, **106**, 9955–9962.

Seehausen O, Terai Y, Magalhaes IS *et al.* (2008) Speciation through sensory drive in cichlid fish. *Nature*, **455**, 620–626.

Servedio MR, Saetre G-P (2003) Spciation as a positive feedback loop between postzygotic and prezygotic barriers to gene flow. *Proceedings of the Royal Society of London B: Biological Sciences*, **270**, 1473–1479.

Servedio MR, Van Doorn GS, Kopp M, Frame AM, Nosil P (2011) Magic traits in speciation: "magic" but not rare? *Trends in Ecology & Evolution*, **26**, 389–397.

Shaw KL (2002) Conflict between nuclear and mitochondrial DNA phylogenies of a recent species radiation: what mtDNA reveals and conceals about modes of speciation in Hawaiian crickets. *Proceedings of the National Academy of Sciences USA*, **99**, 16122–16127.

Stacklies W, Redestig H, Scholz M, Walther D, Selbig J (2007) pcaMethods–a bioconductor package providing PCA methods for incomplete data. *Bioinformatics*, **23**, 1164–1167.

Stamatakis A (2006) RaxML-VI-HPC: maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. *Bioinformatics*, **22**, 2688–2690.

Stamatakis A, Ludwig T, Meier H (2005) RaxML-III: a fast program for maximum likelihood-based inference of large phylogenetic trees. *Bioinformatics*, **21**, 456–463.

Turelli M, Barton NH, Coyne JA (2001) Theory and speciation. *Trends in Ecology & Evolution*, **16**, 330–343.

Turner TL, Hahn MW, Nuzhdin SV (2005) Genomic islands of speciation in *Anopheles gambiae*. *PLoS Biology*, **3**, e285.

Turner BJ, Duvernell DD, Bunt TM, Barton MG (2008) Reproductive isolation among endemic pupfishes (*Cyprinodon*) on San Salvador Island, Bahamas: microsatellite evidence. *Biological Journal of the Linnaean Society*, **95**, 566–582.

Uyeda JC, Hansen TF, Arnold SJ, Pienaar J (2011) The million-year wait for macroevolutionary bursts. *Proceedings of the National Academy of Sciences USA*, **108**, 15908–15913.

Wagner CE, McCune AR (2009) Contrasting patterns of spatial genetic structure in sympatric rock-dwelling cichlid fishes. *Evolution*, **63**, 1312–1326.

Wagner CE, Keller I, Wittwer S *et al.* (2012) Genome-wide RAD sequence data provide unprecedented resolution of species boundaries and relationships in the Lake Victoria cichlid adaptive radiation. *Molecular Ecology*, **22**, 787–798.

Willing E-M, Dreyer C, van Oosterhout C (2012) Estimates of genetic differentiation measured by $F_{st}$ do not necessarily require large sample sizes when using many SNP markers. *PLoS ONE*, **7**, e42649.

---

C.H.M. designed the study, collected samples, prepared the sequencing library, performed analyses and wrote the article. L.C.F. streamlined and modified molecular protocols, guided library preparation and commented on the article and analyses.

---

## Data accessibility

Raw sequencing reads are available on the NCBI Short Reads Archive (BioProject ID PRJNA231549). SNP matrices and all other data sets used for analyses are deposited in Dryad: doi:10.5061/dryad.5k4k4.

## Supporting information

Additional supporting information may be found in the online version of this article.

**Fig. S1** Representative photographs of *brontotheroides* individuals from two different populations on San Salvador: Little Lake (top row) and Crescent Pond (bottom row).

**Fig. S2** (a) Log likelihood of SNP data (filled circles) across $k = 1$–12 levels of population substructure and $\Delta K$ (open circles) across $k = 2$–12 levels of substructure, where the modal value suggests the true number of subpopulations (Evanno *et al.* 2005). (b) Variance in log likelihood of SNP data among runs

across $k$ = 1–12 levels of population substructure. Log likelihoods were averaged from at least two independent MCMC runs of 50 000 steps after discarding 50 000 steps as burnin.

**Fig. S3** STRUCTURE analyses of 4202 SNPs across saline lake populations on San Salvador Island, Bahamas. Large blocks group individuals by species (blue: *variegatus*, green: *brontotheroides*, red: *desquamator*).

**Table S1** Collection localities, source, and sample size for each population sampled.

**Table S2** Species, location, and proportion of missing loci out of the 13 912 SNPs for each individual sampled.

**Table S3** Pairwise $F_{ST}$ values between all populations in saline lakes on San Salvador based on 23 173 SNPs genotyped in at least 50% of individuals.

**Table S4** Model comparison of population trees presented in Fig. 2c with different numbers of admixture events ($m$ = 0–3).