Title

Complex histories of repeated gene flow in Cameroon crater lake cichlids cast doubt on one of the clearest examples of sympatric speciation

Journal

Evolution, 69(6)

ISSN

0014-3820

Authors

Martin, Christopher H
Cutler, Joseph S
Friel, John P
et al.

Publication Date

2015-06-01

DOI

10.1111/evo.12674

Peer reviewed

doi:10.1111/evo.12674

Complex histories of repeated gene flow in Cameroon crater lake cichlids cast doubt on one of the clearest examples of sympatric speciation

Christopher H. Martin,[1,2] Joseph S. Cutler,[3] John P. Friel,[4] Cyrille Dening Touokong,[5] Graham Coop,[6] and Peter C. Wainwright[6]

[1]*Department of Biology, University of North Carolina, Chapel Hill, North Carolina*

[2]*E-mail: chmartin@unc.edu*

[3]*Department of Conservation Biology, University of California, Santa Cruz, California*

[4]*Alabama Museum of Natural History, University of Alabama, Tuscaloosa, Alabama*

[5]*Gulf Aquatics, Inc., Douala, Cameroon*

[6]*Center for Population Biology and Department of Evolution & Ecology, University of California, Davis, California*

Received December 11, 2014
Accepted April 20, 2015

One of the most celebrated examples of sympatric speciation in nature are monophyletic radiations of cichlid fishes endemic to Cameroon crater lakes. However, phylogenetic inference of monophyly may not detect complex colonization histories involving some allopatric isolation, such as double invasions obscured by genome-wide gene flow. Population genomic approaches are better suited to test hypotheses of sympatric speciation in these cases. Here, we use comprehensive sampling from all four sympatric crater lake cichlid radiations in Cameroon and outgroups across Africa combined with next-generation sequencing to genotype tens of thousands of SNPs. We find considerable evidence of gene flow between all four radiations and neighboring riverine populations after initial colonization. In a few cases, some sympatric species are more closely related to outgroups than others, consistent with secondary gene flow facilitating their speciation. Our results do not rule out sympatric speciation in Cameroon cichlids, but rather reveal a complex history of speciation with gene flow, including allopatric and sympatric phases, resulting in both reproductively isolated species and incipient species complexes. The best remaining non-cichlid examples of sympatric speciation all involve assortative mating within microhabitats. We speculate that this feature may be necessary to complete the process of sympatric speciation in nature.

KEY WORDS: **Adaptive radiation, admixture, ecological speciation, gene flow, introgression, magic trait, next-generation sequencing, population genomics, RADseq.**

Sympatric speciation, the evolution of reproductive isolation without the benefit of geographic barriers or isolation by distance, has fascinated evolutionary biologists since its initial conception by Darwin as his "principle of divergence" (Turelli et al. 2001). This endpoint on the speciation-with-gene-flow continuum delights the minds of theorists and empiricists alike because it embodies the power of natural and sexual selection to create new species through complex and often counterintuitive interactions (Dieckmann and Doebeli 1999; Kondrashov and Kondrashov 1999; Bolnick and Fitzpatrick 2007; Fitzpatrick et al. 2008; Servedio and Bürger 2014). However, despite substantial fascination, the long absence of convincing examples in nature and restrictive conditions predicted by early models resulted in the dismissal of this process throughout most of the 20th

EVOLUTION
INTERNATIONAL JOURNAL OF ORGANIC EVOLUTION

century (Mayr 1963; Felsenstein 1981; Bolnick and Fitzpatrick 2007).

The discovery of the Cameroon crater lake cichlid radiations almost single-handedly reversed this consensus and reignited interest in the possibility of sympatric speciation in nature (Turelli et al. 2001; Coyne and Orr 2004; Bolnick and Fitzpatrick 2007). Schliewen et al.'s landmark studies outlined a compelling case for sympatric speciation in three different isolated volcanic crater lakes, each containing endemic radiations of cichlid fishes (Schliewen et al. 1994, 2001; Schliewen and Klee 2004). Sympatric radiations were restricted to lakes so small, uniform, and remote that sympatric diversification seemed much more likely than transient phases of geographic isolation. Completely uniform crater basins meant that historical changes in water level could not have subdivided the lakes, eliminating the possibility of within-lake barriers (Schliewen et al. 1994). Furthermore, the radiations were so speciose (up to 11 species) that speciation in situ appeared far more likely than allopatric speciation in a neighboring river, followed by dispersal into the isolated volcanic crater, and subsequent extinction of the original riverine population for each of the endemic species (Schliewen et al. 1994). Finally, mitochondrial and amplified fragment length polymorphism (AFLP) phylogenies supported monophyly: species within the radiations were more closely related to each other than to any riverine outgroups (Schliewen et al. 1994, 2001; Schliewen and Klee 2004). Many authors continue to cite these cichlid radiations as "almost certainly" resulting from sympatric speciation (Kondrashov and Kondrashov 1999; Cristescu et al. 2010) and Coyne and Orr (2004) concluded that "there is no more convincing case in nature" in their comprehensive review of the field.

Sympatric radiations in each crater suggested that the rarity of sympatric speciation in nature may only reflect the rarity of environments in which historical geographic barriers can be ruled out. Indeed, many more putative examples of sympatric speciation in remote crater lakes (Barluenga et al. 2006; Seehausen 2006; Elmer et al. 2010b) and other remote islands (Savolainen et al. 2006) were documented thereafter. For example, in Nicaragua at least three of eight crater lakes appear to contain sympatric Midas cichlid species (Elmer et al. 2010a). On Lord Howe Island as many as 20% of plant species appear to result from sympatric speciation based on the monophyly of sister species endemic to the island (Papadopolus et al. 2011).

However, population genomic analyses now provide the power to distinguish more complex scenarios of colonization and hybridization during adaptive radiation and speciation with gene flow (Heliconius Genome Consortium 2012; Pickrell and Pritchard 2012; Alcaide et al. 2014; Martin and Feinstein 2014; Pease and Hahn 2015). Phylogenetic inference of monophyly for a sympatric radiation is not sufficient to establish that these species shared a single ancestral founding population. For example, the simplest alternative is that two separate waves of colonization by the same source population produced two sympatric species on an island following reinforcement or character displacement after secondary contact (Pfennig and Pfennig 2012). These two species will become apparent sister species after extensive gene flow because their paraphyletic colonization history will be masked if species differences are restricted to small islands of genetic differentiation (Wu 2001; Turner et al. 2005). Mitochondrial introgression among sympatric species is known to obscure phylogenetic relationships (e.g., Shaw 2002), but there are now bountiful examples of genome-wide introgression at nuclear loci in many nascent species that will also obscure population histories in majority-rule phylogenies (Heliconius Genome Consortium 2012; Keller et al. 2013; Martin and Feinstein 2014; Poelstra et al. 2014). Perhaps the best-known examples are stickleback species pairs. Although sympatric species pairs appear to be sister taxa in some lakes, they are believed to have evolved from two waves of colonization of glacial lakes, resulting in speciation of first the benthic ecomorph and then the limnetic ecomorph after the second invasion (Schluter and McPhail 1992). Encouragingly, this alternate double-invasion scenario is supported by geological evidence of repeated flooding and by physiological evidence that the later-arriving limnetic species has higher salinity tolerance (Kassen et al. 1995; Gow et al. 2008).

The distinction between a single colonization and multiple waves is important for identifying the relevant mechanisms driving speciation in nature. Speciation models widely agree that any period of geographic isolation or isolation by distance during the divergence process will facilitate speciation by allowing some initial amount of assortative mating to evolve in allopatry or by restricting gene flow between ecotypes in different habitats (Kirkpatrick and Ravigné 2002; Gavrilets 2004; Doebeli et al. 2005). Indeed, allopatry is no different than perfect assortative mating in these models (Kirkpatrick and Ravigné 2002). Without these barriers, sympatric speciation requires the evolution of assortative mating by ecotype (or more controversially, by sexual signal [Bolnick and Fitzpatrick 2007]) from within a randomly mating population. Models of this process indicate that divergence in sympatry generally requires either much stronger disruptive selection or the presence of traits under divergent ecological selection that also cause some assortative mating through their effects on mating location, mating preferences, or mating cues, known as "magic traits" (Servedio et al. 2011). To begin to test such predictions from the abundance of speciation models and understand their relevance to natural speciation processes (e.g., Martin 2012, 2013), we first need to know whether multiple colonizations were involved in the most compelling examples of sympatric divergence in nature.

It is also important to ask whether secondary colonization and hybridization helped drive additional speciation events by

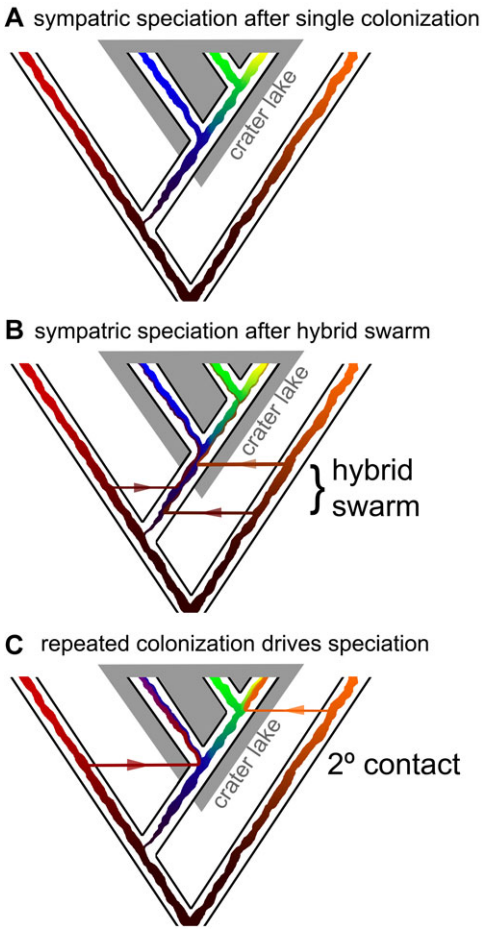

Figure 1. **Three scenarios for sympatric radiation within a crater lake (gray triangle) in which changing colors indicate changing allele frequencies through time and horizontal lines depict migration into the crater. (A) Sympatric speciation following a single colonization event from a neighboring riverine population (brown). (B) Sympatric speciation after a hybrid swarm is formed from repeated colonization by neighboring populations. All resulting species share a similar proportion of their ancestry with outgroups due to this initial period of panmixia. (C) Repeated colonization resulting in speciation with gene flow due to secondary contact between a neighboring population and the initial founder population. Some species within the radiation share more of their ancestry with certain outgroups than others.**

contrasting two hybridization scenarios: adaptive radiation from a hybrid swarm versus adaptive radiation from repeated colonization and character displacement (Seehausen 2004; Pfennig and Pfennig 2012; Fig. 1). Adaptive radiation from a sympatric hybrid swarm would still be an example of sympatric speciation if no geographic barriers separated populations during the divergence process, even if the founding population was composed of multiple colonizing lineages (Fig. 1B). Conversely, if the initial founding population evolves partial reproductive isolation before the second wave of colonists arrives (Fig. 1C), then this initial period of allopatry at least partially contributed to the divergence of these populations. These two double-invasion scenarios are on a continuum: from complete panmixia of a hybrid swarm before sympatric speciation to allopatric speciation with limited gene flow after secondary contact. The latter scenario predicts that some species within the radiation will share more ancestry with the secondary colonizing population than others (Fig. 1C). The hybrid swarm scenario predicts that sympatric species will share a relatively even proportion of ancestry with each of the colonizing populations (Fig. 1B). Distinguishing between these hybridization scenarios is critical to establish whether geographic isolation facilitated speciation within a sympatric radiation.

Here we offer a critical reappraisal of sympatric speciation in Cameroon crater lake cichlid radiations by estimating the complex colonization history of these craters using population genomics. We used double-digest restriction site associated DNA (RAD) sequencing (Peterson et al. 2012) combined with comprehensive sampling of all three Cameroon crater lakes containing four sympatric cichlid radiations (24 of 25 species), the most closely related outgroup populations in surrounding drainages, and additional outgroup species across Africa (Fig. 2). We analyzed these data with both phylogenetic and population genomic methods to test for (1) monophyly, (2) genetic structure shared with outgroups, and (3) introgression between the sympatric radiations and outgroup populations. We also compared patterns of shared outgroup ancestry among the sympatric species to ask whether multiple waves of colonization may have facilitated additional speciation events within the craters.

Methods

SAMPLING

Fifty-three taxa were collected over six African field expeditions, including 24 of 25 species within the four sympatric radiations and 29 outgroups. We specifically targeted the most closely related populations in drainages surrounding the crater lakes, including four riverine populations of *Sarotherodon galilaeus*, four riverine populations of *Coptodon guineensis*, and two crater lake populations of *Coptodon kottae* (Fig. 2, Appendix S1).

The four Cameroon crater lake radiations fall into two main subclades within haplotilapiine cichlids: the *S. galilaeus* species complex within the Oreochromini and the *C. guineensis* species complex within the Coptodonini (Schwarzer et al. 2009; Dunz and Schliewen 2010). The widespread riverine species *S. galilaeus* (Oreochromini) colonized the 2.3-km-wide volcanic crater lake Barombi Mbo and founded the Barombi Mbo *Sarotherodon* radiation of 11 endemic species (Schliewen et al. 1994; Schliewen and Klee 2004). This radiation shared a most recent common ancestor (MRCA) with riverine species between 1 and 2.5 million years ago (95% credible

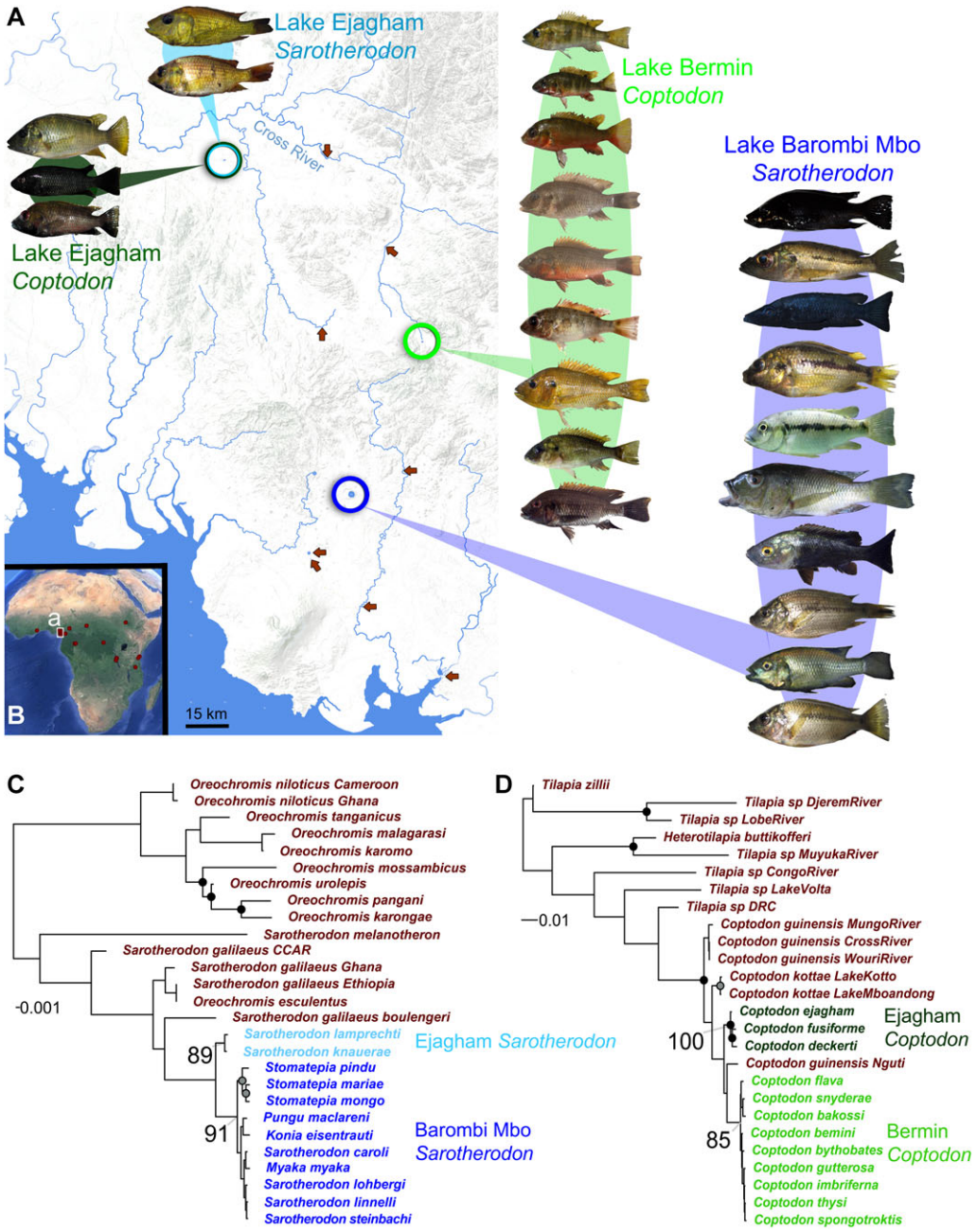

Figure 2. **(A) Sympatric cichlid radiations and outgroup sampling (brown arrows) relative to surrounding river drainages within the volcanic belt of Cameroon (Google Maps Terrain). Representative photos of each sympatric species are shown in the same order as species names listed in (C) and (D). Colors denoting each radiation and outgroups are consistent throughout the manuscript. (B) Outgroup sampling across Africa (brown dots; Google Earth satellite image). (C) Maximum likelihood phylogenies for (C) *Sarotherodon* and (D) *Coptodon* plus outgroups estimated from concatenated datasets of 33,201 and 37,175 SNPs, respectively, genotyped in at least six taxa. Black dots (●) and gray dots (●) indicate nodes with 100% and 95% bootstrap support, respectively. Exact bootstrap support for sympatric radiations is indicated.**

interval: supplement to Friedman et al. 2013), consistent with geological age estimates for Barombi Mbo crater (Cornen et al. 1992). *S. galilaeus* also colonized tiny Lake Ejagham (900 × 600 m) and founded an endemic species pair of *Sarotherodon* (*S. lamprechti/S. knauerae*, formerly *S.* sp. "bighead"/*S.* sp. "mudfeeder": Schliewen et al. 2001; Neumann et al. 2011). Lake Ejagham (a solution basin, not a crater lake) is estimated to be only 10,000 to 100,000 years old (D. Livingston, pers. comm. in Schliewen et al. 2001) and no time-calibrated estimates of divergence time for this species pair are available.

The distantly related riverine species *C. guineensis* also founded two sympatric radiations in Cameroon: a radiation of nine species endemic to 700-m-diameter crater lake Bermin (Beme), the Bermin *Coptodon*, and a radiation of three species endemic to Lake Ejagham, the Ejagham *Coptodon* (*C. fusiforme/deckerti/ejagham*; there is a fourth described species *C. nigrans* unsupported by genetic structure: Dunz and Schliewen 2010). Lake Bermin is estimated to have formed 100,000 to 2 million years ago (G. Kling, pers. comm. in Stiassny et al. 1992).

We analyzed the Coptodonini and Oreochromini separately in all analyses due to their distant relationships to each other (nearly spanning the full history of African haplotilapiines). Within the Oreochromini, we sampled 10 of the 11 species in Barombi Mbo (missing one deep-water specialist, *Konia dikume*, which is closely related to *Konia eisentrauti* in our sample), the two *Sarotherodon* sister species in Lake Ejagham, four riverine outgroup populations of *Sarotherodon galilaeus*, and 11 additional species across West Africa (Appendix S1, Fig. 2). Within the Coptodonini, we sampled all nine *Coptodon* species in Bermin, the three *Coptodon* species in Ejagham, the single species *Coptodon kottae* endemic to the neighboring crater lakes Barombi ba Kotto and Mboandong, four riverine outgroup populations of *Coptodon guineensis*, and eight additional outgroup species across Africa (Appendix S1, Fig. 2). Specimens were collected over the course of six field expeditions to Africa from 2006 to 2013 led by numerous investigators, including four different expeditions each independently led by CHM, JSC, JPF, and CDT (Appendix S1). Tissue samples were preserved in 95% ethanol or RNAlater (Ambion) and stored at –20° C after return to the United States. In some cases, samples collected by CDT were stored for two years at 25°C in ethanol of unknown concentration.

GENOMIC LIBRARY PREPARATION

Genomic DNA was extracted from each sample using Qiagen blood and tissue extraction kits (Qiagen, Inc., Venlo, Limburg, Netherlands). DNA concentration was quantified on a QuBit 2.0 fluorometer (2.0, Life Technologies, Carlsbad, CA, USA) and equalized to 50 ng/uL. DNA quality was assessed using a nanodrop and agarose gel electrophoresis and ranged widely due to varying levels of sample preservation. Three double-digest RAD-seq libraries (each including 96 barcoded samples, not all used for this study) were prepared following the protocol of Peterson et al. (2012) with the following modifications: (1) use of the genotyping-by-sequencing PCR primers and adapters described in Elshire et al. (2011) and used in Martin and Feinstein (2014); (2) eliminating the size-selection and DNA quantification step between digestion and ligation reactions, following Elshire et al. (2011); (3) substituting low-cost Sera-Mag beads for bead size-selection as described in Rohland and Reich (2012); and (4) omitting the optional streptavidin-purification step. We used high-fidelity restriction enzyme SbfI for infrequent cutting and NlaIII for frequent cutting (New England Biolabs, Inc., Ipswich, MA, USA). Ninety-six barcoded adapters with non-nested molecular barcodes between 4 and 8 bp in length separated by at least three mutational steps were calculated using the Deena GBS Barcode Generator (http://www.deenabio.com/services/gbs-adapters) and ordered unmodified with standard purification from Life Technologies. These barcodes were previously successfully applied in Martin and Feinstein (2014).

Samples were randomly arranged within each 96-well plate and populations with multiple samples were split across the three plates. We aimed for at least six individuals per population if sufficient samples were available (Appendix S1). Five hundred nanograms of DNA per sample were digested with 0.12 μL SbfI and 0.25 μL NlaIII in 20 μL reaction volumes for 3 h at 37°C. Annealed barcoded and common adapters were then ligated to the digested samples with 0.5 μL T4 DNA ligase (New England Biolabs) in 50 μL reaction volumes for 1 h at 16°C. Ligation reactions were size-selected with SeraMag bead solution at a ratio of 1.5 beads:sample and groups of 48 samples were pooled in 35 μL volumes. The Functional Genomics Laboratory at UC Berkeley then size-selected each pooled library for fragment sizes between 300 and 500 bp (library 1) or between 300 and 400 bp (libraries 2–3) using a Blue Pippin (Sage Science, Beverly, MA, USA). Agilent High-Sensitivity Bioanalyzer chips (Agilent Technologies, Inc., Santa Clara, CA, USA) indicated that, in practice, fragment size ranges ranged from 300 to over 500 bp in all three libraries. Pooled and size-selected libraries of 48 samples each were then amplified in a 50 μL reaction volume with Phusion high-fidelity DNA polymerase (ThermoFisher Scientific, Petaluma, CA, USA) and Illumina primers at 98° C for 30 sec, followed by 12 cycles of 98°C for 10 sec, 65°C for 30 sec, 72°C for 30 sec, and a final extension at 72°C for 7 min before cool-down. PCR reactions were size-selected with a 1.8 bead:sample ratio of SeraMag and reactions from each set of 96 samples were pooled and re-suspended in 18 μL of 1x TE buffer. Final library quality was checked on an Agilent High-Sensitivity Bioanalyzer chip and by qPCR at the Functional Genomics Laboratory. The three libraries were each sequenced single-end to 100 bp on a single lane of an Illumina HiSeq 2500 machine (standard mode); two libraries were sequenced at the Vincent J. Coates Genomic Sequencing Library, California Institute for Quantitative Biosciences, UC Berkeley, and one library at the UC Davis Genome Sequencing Facility.

SEQUENCING

The two genomic libraries sequenced at Berkeley produced 132.8 and 133.3 million single-end reads. The library sequenced at UC Davis produced 90.4 million single-end reads. We used the Stacks

pipeline (version 1.18, Catchen et al. 2013) to filter reads and call SNPs. First, raw reads were sorted by barcode and any reads with low-quality scores (mean Phred score < 10 along a sliding window of 20 bp), uncalled bases, or missing restriction sites or barcodes were discarded using the default settings of process_radtags in Stacks (version 1.18). Barcodes with errors were not rescued. The quality-filtered libraries sequenced at UC Berkeley retained 81% and 83% of total reads, respectively. The quality-filtered library sequenced at UC Davis retained 49.7% of total reads despite identical library preparation, revealing wide variation in the quality of these sequencing centers.

BIOINFORMATICS PIPELINE

SNPs were called from filtered datasets by aligning to the most closely related reference genomes to *Sarotherodon* and *Coptodon*. We used Bowtie 2 (version 2.2.3, Langmead and Salzberg 2012) to align filtered reads from each individual to available reference genomes. For the Coptodonini, we used the *Astatotilapia burtoni* genome assembly (version 1) downloaded from http://bouillabase.org. For the Oreochromini, we used the *Oreochromis niloticus* genome assembly (version 1.1) downloaded from GenBank (Brawand et al. 2014). We then used the pstacks command in Stacks to align stacks of homologous sequences within each individual into loci with a minimum of three sequenced reads per locus (stacks flags in italics: *-m* 3). Loci were then cataloged across all individuals and merged into homologous loci based on genomic position (*-g*). SNPs were called within individuals at loci with at least eight reads (*-m* 8) by their maximum likelihood relative to all individuals genotyped at that locus (Catchen et al. 2013). This pipeline resulted in 51,638 loci and 41,428 loci with at least eight sequenced reads within the Oreochromini and Coptodonini, respectively. These results agree well with the number of RAD loci identified in other African cichlid RADseq studies surveying SbfI restriction sites: for example, 66,500 loci identified in an *Aulonocara baenschi/Tramitichromis intermedius* cross (O'Quin et al. 2013); 89,927–136,000 loci identified in Lake Victoria rock cichlids (Wagner et al. 2012; Keller et al. 2013); and 69,889 loci within two families of *O. niloticus* (Palaiokostas et al. 2013). Our numbers of detected loci are slightly lower due to filtering by a minimum read depth of eight reads.

The MRCA of *O. niloticus* + Barombi Mbo *Sarotherodon* + Ejagham *Sarotherodon* is estimated at 2.5 to 7 million years ago (Friedman et al. 2013). The MRCA of *A. burtoni* + Bermin *Coptodon* + Ejagham *Coptodon* is estimated at 7.5 to 16 million years ago (95% credible intervals from supplemental chronogram in Friedman et al. 2013). It is unknown whether alignment to such divergent genomes biased our inferences by restricting analyses to more conserved regions of the genome. One well-known bias is that allele dropout due to polymorphic restriction sites underestimates heterozygosity in any RADseq study; however, this bias does not appear to affect estimates of relative genetic differentiation among populations (Arnold et al. 2013).

We excluded all individuals genotyped in less than 5% of total loci and exported SNP data in .phylip or .plink formats (version 1.7, Purcell et al. 2007) for downstream analyses. Due to the varying tolerances of different analyses for missing data, we then tailored minimum genotyping thresholds for each subsequent analysis. We used less-stringent minimum genotyping thresholds for phylogenetic analyses, STRUCTURE, and principal components analysis (PCA) to take advantage of larger datasets. We used more-stringent genotyping thresholds (loci genotyped in every population) for Treemix graphs and f_4 statistics.

PHYLOGENETIC ESTIMATION

We used maximum likelihood to estimate a phylogeny for all species and populations based on the concatenated SNP datasets for the Oreochromini and Coptodonini. We exported only those SNPs fixed within species and variable among them (.phylip format) and genotyped in at least six species or populations (of 28 taxa for the Oreochromini and 26 taxa for the Coptodonini). This resulted in a dataset of 33,201 SNPs for the Oreochromini phylogeny and 37,175 SNPs for the Coptodonini. We also explored analyses using only high-coverage loci by filtering for SNPs genotyped in at least half of all species ($\geq$14), resulting in 10,024 SNPs for the Oreochromini and 13,732 SNPs for the Coptodonini.

We used RaxML (version 8.0.2, Stamatakis 2014a) to estimate the maximum likelihood topology for 1000 bootstrap samples using the rapid bootstrapping algorithm under the GTR + Γ model of nucleotide substitution. We included a correction to the likelihood approximation for the ascertainment bias of using only variable sites and estimated empirical base frequencies (ASC_GTRGAMMAX: Stamatakis 2014b). We caution that lack of invariant sites prevents accurate inference of branch lengths in our phylogeny (Lemmon and Lemmon 2013); however, this does not affect our goal to assess evidence for monophyly in the four sympatric cichlid radiations. Nonetheless, we stress that our maximum likelihood phylogenies only estimate an average topology across a concatenated SNP dataset and assume a bifurcating tree (Edwards 2009). Such phylogenetic evidence for monophyly can only provide a point estimate of the majority-rule branching pattern and does not address subsequent gene flow.

GENETIC CLUSTERING ANALYSES

For all further population genetic analyses, we focused only on the *Sarotherodon* and *Coptodon* species complexes, rather than the full phylogenies with *Oreochromis* and *Tilapia* outgroups. We used two methods to visualize genetic structure among our populations and species: (1) PCA of genetic variance (Price

et al. 2006) and (2) Bayesian hierarchical clustering using STRUCTURE (version 2.3.4, Pritchard et al. 2000). For PCA analyses, we used PLINK (version 1.7, Purcell et al. 2007) to filter each dataset for loci genotyped in at least 50% of individuals and excluded individuals with less than 5% of loci genotyped overall. We then visualized principal components of variance in these datasets using probabilistic PCA, implemented in the pcaMethods package in R (Stacklies et al. 2007). This algorithm is designed to be robust to large amounts of missing data (Stacklies et al. 2007).

We used STRUCTURE (version 2.3.4; Pritchard et al. 2000; Falush et al. 2003) to estimate proportions of admixture for every individual in the dataset. To limit the effects of linkage disequilibrium, we exported only the first SNP per locus and filtered by SNPs genotyped in at least half of all individuals. We then evaluated levels of genetic structure for $k = 3$–8 in both the *Coptodon* and *Sarotherodon* species complexes (Table S1). For each run, we used the admixture model with correlated allele frequencies and ran MCMC chains for 100,000 generations, removing the first 50% as burn-in. Posterior probabilities from at least six independent runs were evaluated for each level of k. We aggregated independent runs of STRUCTURE using STRUCTURE Harvester (version 0.6.94, Earl 2012) and CLUMPP (Jakobsson and Rosenberg 2007). We then inferred the estimated number of genetic clusters using the log likelihood of the data and Evanno's method based on the rate of change of the log likelihood (Evanno et al. 2005).

TESTS OF INTROGRESSION

We calculated f_4 statistics to provide formal tests of secondary gene flow between each sympatric radiation and outgroups. Essentially, f_4 statistics test if residual genotypic covariance among branches in a four-taxon tree is zero (as expected due to incomplete lineage sorting) or significantly different from zero (indicating more recent gene flow between branches) and were developed to test for introgression among human populations (Reich et al. 2009; Pickrell and Pritchard 2012). Given a set of four populations, f_4 statistics compare allele frequencies among the three possible unrooted trees to test which pairs form clades (supplement to Reich et al. 2009, p. 25). Unlike D-statistics, also known as "ABBA/BABA tests" (Green et al. 2010; Durand et al. 2011; Heliconius Genome Consortium 2012), which require that an unadmixed outgroup be used to root the tree and localize which of two possible branches is admixed, neither the f_4 or f_3 statistics require rooted trees in their calculation (supplement to Reich et al. 2009, p. 21, 25). The significance of the f_4 statistics is assessed by calculating the standard error of f_4 statistics through a jackknifing procedure (Pickrell and Pritchard 2012).

We filtered our datasets for those loci genotyped in every population and in at least 50% of individuals. Individuals with less than 5% coverage were excluded. f_4 statistics were calculated using the fourpop function in Treemix (version 1.12, Pickrell and Pritchard 2012). To account for linkage disequilibrium, standard error was estimated by jackknifing in windows of 10 adjacent SNPs.

VISUALIZATION OF INTROGRESSION

We used Treemix (version 1.12, Pickrell and Pritchard 2012) to visualize introgression between branches of the phylogeny. This method uses the allele frequencies and a Gaussian approximation for genetic drift among populations to estimate a maximum likelihood population tree. Introgression between branches is then evaluated in a stepwise likelihood procedure, searching the tree for an optimal placement of each subsequent admixture event (Pickrell and Pritchard 2012). An edge on the graph is assigned as a branch on the tree if it contributed the majority of alleles to the descendent population, otherwise it is the "migration" edge (Pickrell and Pritchard 2012, p. 3). Information about the directionality of introgression comes from the asymmetries in the relationships among populations given the tree. For example, if we imagine a tree ((A,B)(C,D)) with subsequent introgression from B into C, population C would show unusually high covariance with A, but B would not with D. In practice this information is contained in the likelihood, a fact that Treemix exploits by, after adding each migration edge, locally maximizing the likelihood by iteratively changing the source and destination of the migration event along with additional local branch swapping (Pickrell and Pritchard 2012, p. 14). We fit three to four admixture events to both the *Coptodon* and *Sarotherodon* trees to visualize the largest gene flow events and estimate their proportion and direction.

Results

PHYLOGENETIC ANALYSES INDICATE WEAK SUPPORT FOR MONOPHYLY

We first tested for monophyly in each of the four sympatric radiations based on concatenated datasets of 33,201 SNPs in *Sarotherodon* and 37,175 SNPs in *Coptodon*. Maximum likelihood phylogenies strongly supported monophyly in only one of the four radiations (bootstrap support = 100% in Ejagham *Coptodon*: Fig. 2D). Bootstrap support for the other three radiations ranged from 85% to 91% (Fig. 2C, D). Paraphyletic topologies included an origin of *C. guineensis* (Cross River, Nguti sample site) within the Bermin *Coptodon* and an origin of *O. esculentus* within the Barombi Mbo *Sarotherodon* or the Ejagham *Sarotherodon*. Phylogenetic analyses based on smaller but more complete SNP datasets found similar weak support for monophyly, including paraphyly of Bermin *Coptodon* with Cross River fish in 79% of bootstrap samples (Fig. S1d).

GENETIC CLUSTERING ANALYSES ARE CONSISTENT WITH INTROGRESSION

We next estimated principal components of genetic variance (PCA) and Bayesian genetic clustering (STRUCTURE) to assess whether sympatric radiations shared ancestry with outgroup populations and visualize how this ancestry was distributed among the species within each radiation. A single ancestral founding event in each crater would result in divergence along independent principal component axes for populations within the crater that would also be equidistant to the river populations in PCA space as they share equal levels of ancestry with outgroups (McVean 2009). Equal levels of ancestry shared with outgroups should also be apparent in STRUCTURE plots, as the crater populations should draw the same amount of ancestry from ancestry clusters shared with outgroup populations. In contrast, three species groups of Barombi *Sarotherodon* shared varying levels of ancestry with outgroup Ejagham *Sarotherodon* (Figs. 3B, 4A). This pattern suggests that introgression (or ancestral population structure) contributed to the sympatric divergence of these three species groups. We filtered to 904 SNPs genotyped completely in 14 high-coverage individuals and found the same pattern, indicating that missing data are not the cause (Fig. S3). The younger Ejagham *Sarotherodon* (0.01–0.1 million years ago) showed mixed ancestry and intermediate placement between Barombi *Sarotherodon* and riverine *Sarotherodon* outgroups (Figs. 3A, 4A, S2), consistent with a history of admixture (McVean 2009).

The younger radiations of Bermin *Coptodon* and Ejagham *Coptodon* (approximately 0.01–0.1 million years ago) also did not diverge along independent principal component axes (Figs. 3C, S2) and shared varying levels of ancestry with closely related Cross River populations and crater lake *C. kottae* populations (Fig. 4B). This pattern suggests additional gene flow between sympatric radiations and neighboring riverine populations. In particular, within the Ejagham *Coptodon* all six admixed individuals came from a single species *C. fusiforme* (Fig. 3C: dark-green circles) and shared larger proportions of their ancestry with outgroup populations in the Cross River (Fig. 4B), suggesting that *C. fusiforme* may have speciated after additional gene flow. Admixed individuals within the Bermin *Coptodon* did not show the same pattern, belonging to four of the nine species (Fig. 3C: light-green assorted shapes); however, there was minimal evidence of genetic differentiation among these nine species (Fig. 4B).

ADMIXTURE STATISTICS SUPPORT INTROGRESSION WITH SYMPATRIC RADIATIONS

Because patterns of genetic clustering can be consistent with multiple processes, we used f_4 statistics (Reich et al. 2009; Patterson et al. 2012; Pickrell and Pritchard 2012) to formally test for the presence of introgression with sympatric radiations after their initial founding. The f_4 statistics supported significant introgression

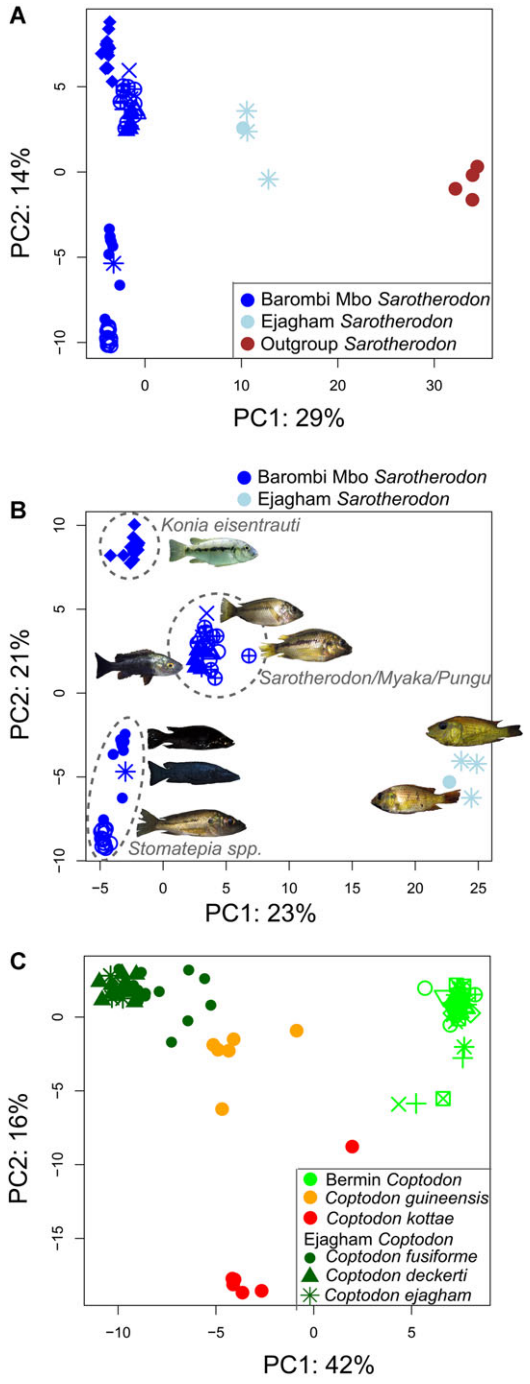

Figure 3. Principal components analysis of genetic variance within (A) the Barombi *Sarotherodon* (dark blue, 10 species) and Ejagham *Sarotherodon* (light blue, two species) relative to riverine *Sarotherodon galilaeus* populations across Africa (4419 SNPs), (B) between only the Barombi *Sarotherodon* and Ejagham *Sarotherodon* with representative photographs of the three distinct clusters of Barombi *Sarotherodon*, and (C) within Bermin *Coptodon* (light green, nine species) and Ejagham *Coptodon* (dark green, three species) relative to closely related outgroups *C. guineensis* Nguti and *C. kottae* (1658 SNPs; analyses with all outgroups presented in Fig. S2). Species within each cichlid radiation are indicated by different shapes.

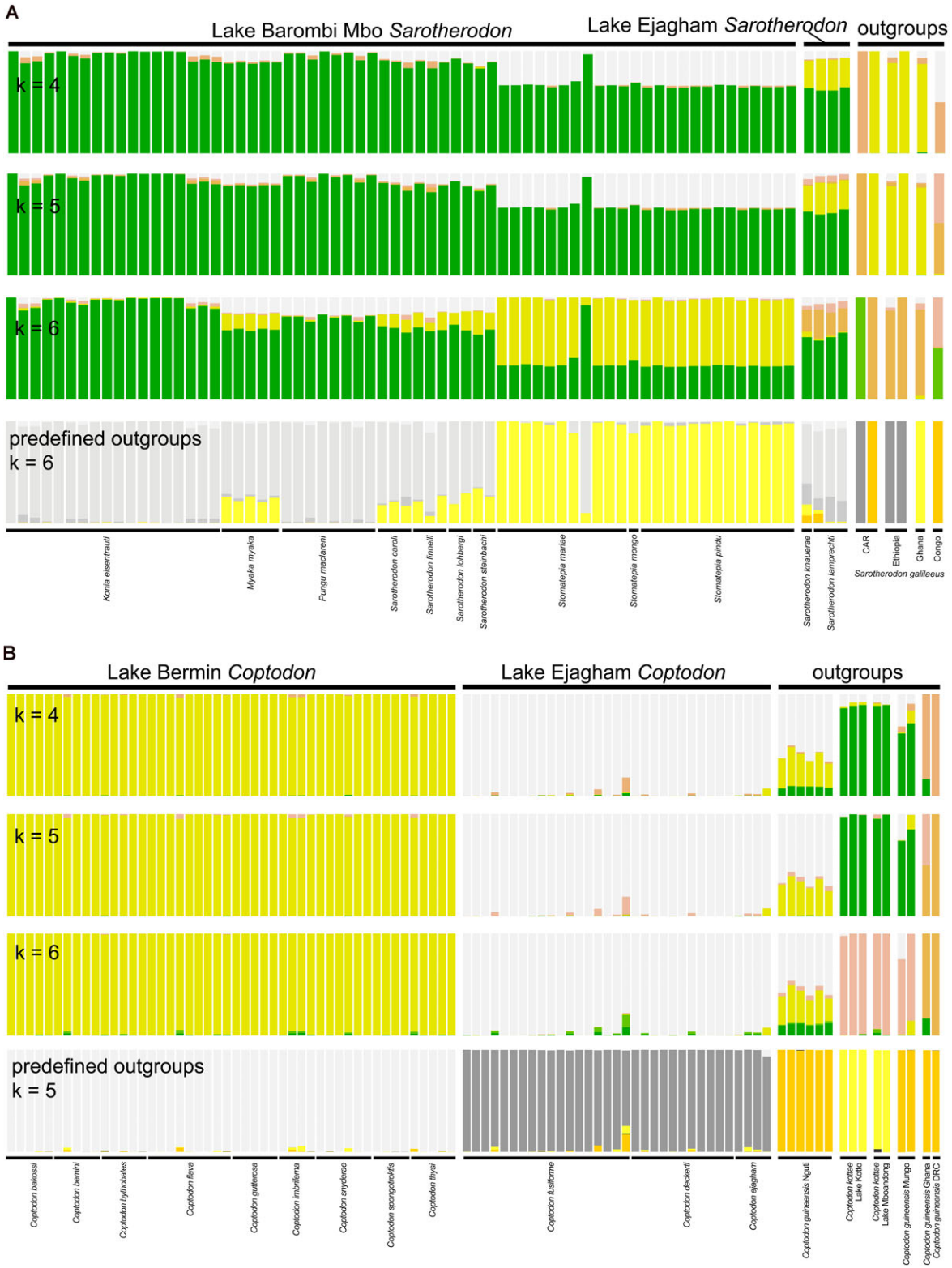

Figure 4. Bayesian clustering analyses showing $k = 4$–6 genetic clusters within the (A) *Sarotherodon* (8271 SNPs pruned to one per locus) and (B) *Coptodon* (6674 SNPs pruned to one per locus) relative to outgroups. The likelihood of the *Sarotherodon* data was maximized at six genetic clusters and Evanno's ad hoc method (Evanno 2005) supported seven genetic clusters; the likelihood of the *Coptodon* data was maximized at five genetic clusters and Evanno's method supported five clusters (Table S1). Population information for outgroups was also used to test for admixture with sympatric radiations (predefined outgroups). Genetic structure within sympatric radiations shared with outgroup populations is highlighted in yellow and orange relative to genetic structure only within sympatric radiations in shades of gray.

Table 1. f_4 statistics supporting introgression with sympatric radiations.

Introgression with:		Four-taxon Tree			
Barombi Mbo ***Sarotherodon***	Ejagham *Sarotherodon*	((A,B);(C,D)) Introgression: (A,B) ← → (C,D)	f_4 Statistic	z-Score	P-value
×	×	*S. lamprechti, S. knauerae*; ***S. lohbergi, S. caroli***	–0.0006 ± 0.0003	–2.282	0.023
×	×	*S. lamprechti, S. knauerae*; ***S. lohbergi, S. linnelli***	–0.0010 ± 0.0005	–2.109	0.035
×	×	*S. lamprechti, S. knauerae*; ***St. pindu, St. mongo***	0.0005 ± 0.0002	2.151	0.032
×	×	*S. lamprechti, S. knauerae*; ***S. lohbergi, M. myaka***	–0.0009 ± 0.0005	–2.063	0.039
×	×	*S. lamprechti, S. knauerae*; ***St. mongo, S. linnelli***	–0.0012 ± 0.0006	–1.948	0.051
	-	*S. lamprechti, S. knauerae*; *S. galilaeus boulengeri*, ***St. mongo***	0.0016 ± 0.0011	1.481	0.139
	–	*S. lamprechti, S. knauerae*; *S. melanotheron Ghana*, ***St. mongo***	0.0016 ± 0.0010	1.605	0.109
–		***S. lohbergi, S. caroli***; *S. galilaeus boulengeri, S. galilaeus Ghana*	0.0008 ± 0.0005	1.627	0.104
–		***S. lohbergi, S. linnelli***; *S. melanotheron Ghana*, *S. knauerae*	–0.0015 ± 0.0009	–1.604	0.109
Bermin ***Coptodon***	Ejagham *Coptodon*				
×	×	*C. deckerti, C. ejagham*; ***C. bemini, C. snyderae***	–0.0012 ± 0.0004	–2.767	0.006
×	×	*C. deckerti, C. ejagham*; ***C. bemini, C. bakossi***	–0.0009 ± 0.0004	–2.201	0.028
×	×	*C. deckerti, C. ejagham*; ***C. bakossi, C. snyderae***	–0.0002 ± 0.0002	–2.120	0.034
×	×	*C. deckerti, C. ejagham*; ***C. flava, C. snyderae***	–0.0010 ± 0.0004	–2.110	0.035
×	×	*C. deckerti, C. ejagham*; ***C. bemini, C. thysi***	–0.0006 ± 0.0003	–2.077	0.038
×		*C. deckerti*, *C.guinensis Ghana*; ***C. bemini, C. snyderae***	–0.0023 ± 0.0008	–2.770	0.006
×		***C. bakossi, C. bythobates***; *C. guinensis Mungo*, *C. fusiforme*	–0.0016 ± 0.0007	–2.190	0.029
	×	*C. deckerti, C. ejagham*; *C. kottae Kotto, C. guinensis DRC*	0.0037 ± 0.0014	2.575	0.010
	×	*C. deckerti, C. fusiforme*; *C. guinensis Ghana*, ***C. bakossi***	0.0032 ± 0.0014	2.312	0.021
	×	*C. deckerti, C. fusiforme*; *C. guinensis Ghana*, ***C. spongotroktis***	0.0032 ± 0.0014	2.354	0.019
	×	*C. deckerti, C. fusiforme*; *C. guinensis Ghana*, ***C. bemini***	0.0032 ± 0.0013	2.411	0.016
	×	*C. deckerti, C. fusiforme*; *C. guinensis Ghana*, ***C. flava***	0.0034 ± 0.0014	2.530	0.011

The most significant four-taxon trees supporting introgression with each sympatric radiation are shown. Some additional nonsignificant four-taxon trees with outgroups are shown for comparison. Species within sympatric radiations are bolded in gray or black. Two-tailed *P*-values are reported for each *z*-score. Note that statistical tests are not independent of each other but should be viewed as the strength of support for introgression, or deviations from a tree-like model of population branching, across various four-taxon subsets.

in all four sympatric radiations (Table 1). Within *Sarotherodon*, introgression between Barombi Mbo *Sarotherodon* and Ejagham *Sarotherodon* was supported by a few subsets of species from these two clades that did not fit a tree-like branching model (Table 1: $P = 0.02$). This introgression may have occurred through an extinct or extant riverine population not sampled in this study because all comparisons with riverine outgroups were nonsignificant (Table 1). In the Bermin and Ejagham *Coptodon* radiations, the strongest evidence for introgression was with neighboring Cross River populations, indicating repeated gene flow ($P = 0.004$).

We found variable support for introgression across four-taxon trees containing different subsets of the species within each radiation. For example, the magnitude of f_4 statistics was twofold higher in Barombi Mbo trees with two *Sarotherodon* species (*S. lohbergi*, *S. linnelli*: –0.001, Table 1) than with two *Stomatepia* species (*St. pindu, St. mongo*: 0.0005). Similarly, f_4 statistics with alternate subsets of Bermin *Coptodon* ranged more than fivefold in magnitude (Table 1). Such variable support for introgression among sympatric species provides evidence that some species groups share more ancestry with outgroups than others, consistent with multiple colonization events that may have facilitated further species divergence within the craters.

We used Treemix (Pickrell and Pritchard 2012) to visualize introgression across maximum likelihood population graphs (trees fit with additional migration events) and estimate the proportion and directionality of introgression based on the genetic covariance of allele frequencies among populations (Fig. 5). Genetic covariance relationships between sympatric cichlid radiations and outgroup taxa recovered a complex history of repeated migration among neighboring rivers and crater lakes for both *Sarotherodon* and *Coptodon* (Fig. 5).

Discussion

REPEATED COLONIZATION OF CAMEROON CRATER LAKES

Despite intense interest, only a handful of case studies have withstood the rigorous criteria for demonstrating sympatric speciation in the wild (Coyne and Orr 2004; Bolnick and Fitzpatrick 2007). This likely reflects both the rarity of isolated and topographically simple environments in which allopatric and parapatric scenarios can be ruled out (often considered the null hypothesis: Coyne and Orr 2004) and strong theoretical predictions that the stringent conditions necessary for this process to proceed will rarely be met in nature (Turelli et al. 2001; Kirkpatrick and Ravigné 2002; Gavrilets 2004). One of the most celebrated examples of sympatric speciation, Cameroon crater lake cichlids, previously rested on a phylogenetic argument for single colonization based on majority-rule support for monophyly from mtDNA sequences and AFLP markers (Schliewen et al. 1994, 2001; Schliewen and Klee 2004). We revisited these examples and examined the evidence for single or multiple colonization based on genome-wide analyses of introgression.

Significant support for introgression from f_4 statistics (Table 1) suggests additional gene flow between Cameroon cichlid radiations and riverine populations after the initial colonization of these three craters. These formal tests were complemented by weak bootstrap support for monophyly in majority-rule phylogenies (Fig. 2) and patterns of shared ancestry with outgroups along major axes of genotypic variance (Fig. 3) and in Bayesian clustering analyses (Fig. 4). These analyses cannot rule out gene flow in the opposite direction from craters to riverine populations. However, Treemix analyses do not support this conclusion and indicate introgression into three of the four sympatric radiations within a complex history of gene flow among taxa (Fig. 5). Analyses of introgression are sensitive to ancestral population structure within the founding population (Durand et al. 2011), for example due to a hybrid swarm (Seehausen 2004). However, if this ancestral structure played no role in species divergence, we would expect similar levels of outgroup ancestry among species within each sympatric radiation (Fig. 1B) in contrast to the observed uneven patterns of outgroup ancestry (Figs. 3, 4). Overall, we have far too little resolution to argue that every species in these craters evolved from hybridization with a secondary colonizing population. Instead, our analyses open a window of allopatry between initial founding and subsequent gene flow that may have played a role in facilitating some speciation events during the course of adaptive radiation.

More broadly, our results raise the possibility of complex colonization histories in other celebrated examples of sympatric speciation (e.g., Barluenga et al. 2006; Elmer et al. 2010b; Papadopolus et al. 2011) and putative examples of convergent evolution across replicate adaptive radiations (Gillespie 2004; Kocher 2004; Losos 2011). In all these examples, singular colonization histories have been inferred from phylogenetic estimates of monophyly. These case studies should be reexamined with population genomic methods that can detect and test for complex colonization histories (e.g., Consortium 2012; Alcaide et al. 2014; Pease and Hahn 2015). For example, reexamination of sympatric fish radiations in other isolated lakes now suggests a role for repeated colonization (Herder et al. 2006; Joyce et al. 2011; Geiger et al. 2013; Martin and Feinstein 2014), with the exception of a putative incipient species pair in a Nicaraguan crater lake (in which no genetic differentiation has been detected so far: Elmer et al. 2010b). Future work incorporating haplotype information to estimate the timing of introgression events from tract lengths will be an important next step toward unraveling the complexity of these processes in nature (Pool and Nielsen 2009; Harris and Nielsen 2013).

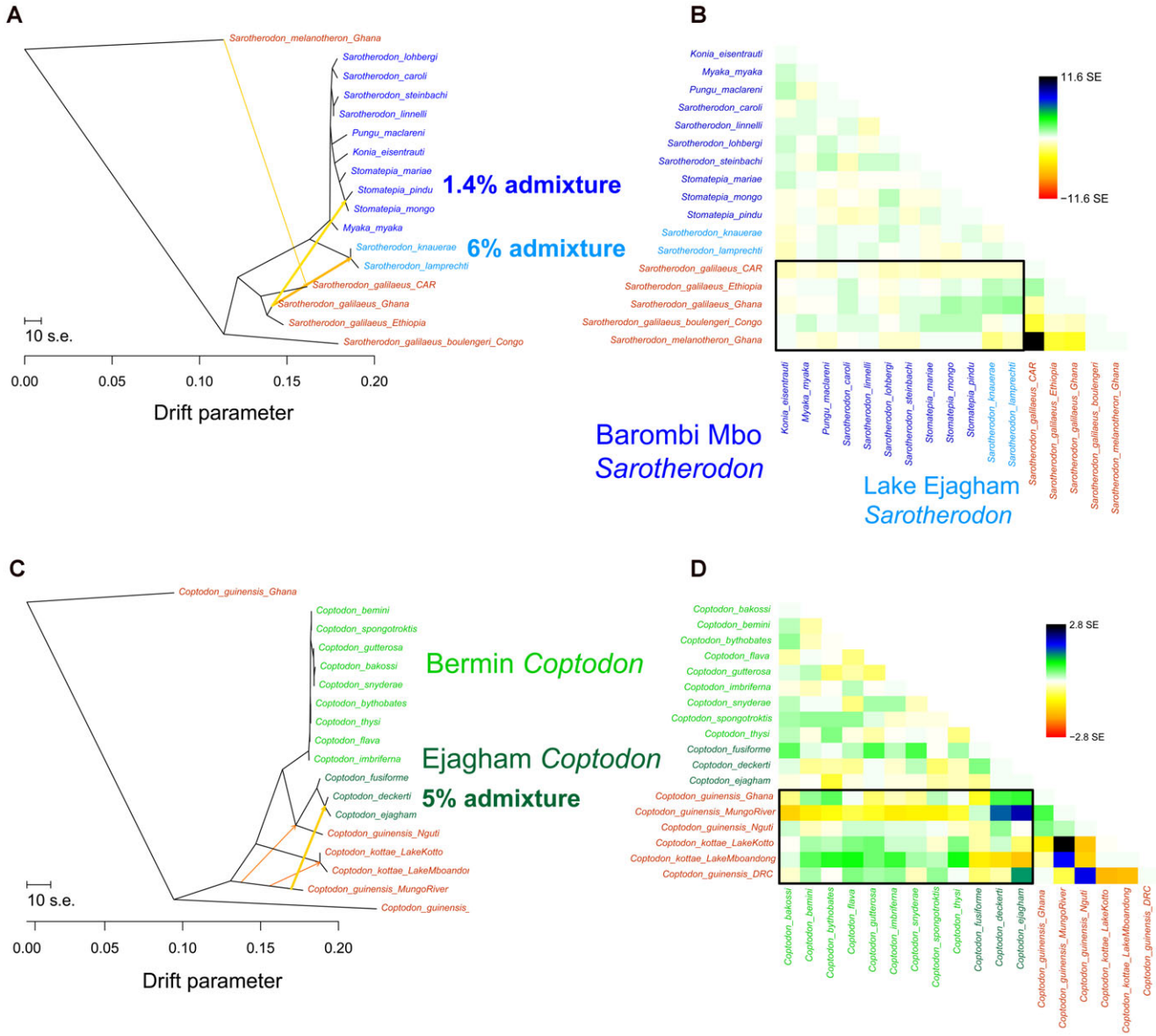

Figure 5. Visualization of introgression into sympatric (A and B) *Sarotherodon* and (C and D) *Coptodon* radiations. Treemix graphs illustrate the three to four strongest introgression events (with heat colors indicating their intensity) for the maximum likelihood phylogenies of (A) *Sarotherodon* estimated from 4441 SNPs genotyped in all populations and (C) *Coptodon* estimated from 1658 SNPs genotyped in all populations. Gene flow between the sympatric radiations and outgroups is annotated with the inferred proportion of admixture. Directionality of gene flow (indicated by the small arrow on each migration edge) was inferred by an iterative maximum likelihood procedure. The scale bar indicates 10 times the average standard error of population relatedness from the genetic variance–covariance matrix of allele frequencies. (B and D) Heat colors indicate residual covariance between each pair of populations (relative to the population tree with zero migration edges) divided by the average standard error across all pairs. Bluer colors indicate populations more closely related to each other than expected under the maximum likelihood tree, suggestive of more recent gene flow (Pickrell and Pritchard 2012, p. 9). The black boxes highlight covariance between sympatric radiations and riverine outgroup species.

EVIDENCE OF SECONDARY GENE FLOW FOR EACH SYMPATRIC RADIATION

We found considerable evidence for secondary gene flow in the youngest radiation of three Ejagham *Coptodon* species. The most significant four-population tests supported secondary gene flow from a distant population of *C. guineensis* in Ghana (Table 1: $P = 0.01$); however, STRUCTURE analyses identified this shared riverine ancestry, illustrated in green, in both the Ghana population and the nearby Cross River population of *C. guineensis* at Nguti (Fig. 4B). Lake Ejagham is within the Cross River drainage, so migration most likely occurred from this river that may also have received migrants from distant riverine populations in Ghana

(Fig. 2). Furthermore, six individuals that appear admixed between Ejagham *Coptodon* and Cross River (Fig. 3B) all come from a single species, *C. fusiforme*. Similarly, the majority of admixed individuals (with "green" ancestry) in the STRUCTURE analysis come from *C. fusiforme* (Fig. 4B) and this species shows different patterns of genetic covariance with outgroup populations than the other two Ejagham species in Treemix analyses (Fig. 5D). This differential pattern of shared ancestry with outgroup species within a sympatric radiation is exactly what we would expect to see if secondary colonization facilitated additional speciation in this lake, as depicted in Figure 1C.

We also found some evidence for secondary gene flow between the young radiation of nine Bermin *Coptodon* species and the neighboring Cross River population of *C. guineensis*, although we cannot infer in which direction (due to no signal in our Treemix analyses) or if gene flow facilitated speciation. The most significant four-population tests supported introgression between *C. snyderae* in Lake Bermin and *C. deckerti* in Lake Ejagham (Table 1: $P = 0.006$). This migration between isolated lakes likely occurred through an intermediary riverine population and both of these lakes are within the Cross River drainage (Fig. 2). Consistent with this migration route, nearby Cross River fish frequently grouped within the Bermin radiation (in 15% and 79% of bootstrap samples: Figs. 2, S1) and shared ancestry with several individuals from Bermin in STRUCTURE (Fig. 4B) and principal components analyses (Fig. 3B). However, these admixed individuals did not consistently come from one species as observed for Ejagham *Coptodon* (Figs. 2–4).

We found patterns consistent with secondary gene flow in both Barombi *Sarotherodon* and Ejagham *Sarotherodon* radiations. The oldest sympatric radiation, Barombi *Sarotherodon*, showed a clear signal of differential relatedness among three main species groups and Ejagham *Sarotherodon* in both PCA (Figs. 3B, S3) and STRUCTURE analyses (Fig. 4A). This signal cannot be explained by a single founding event, but instead suggests either differential introgression or ancestral population structure among these three species groups, perhaps facilitating their early divergence. Four-population tests of a few subsets of the Barombi species with the two Ejagham species also suggest that these Barombi species do not form true clades (e.g., f_4 [Ejagham sp.1, sp.2; Barombi sp.1, sp.2] = –0.0006, P = 0.023: Table 1), providing additional support for differential gene flow into these species groups. However, four-population tests including outgroup riverine samples were not significant (Table 1), possibly due to our lack of sampling a closely related riverine outgroup population in Cameroon (Appendix S1). Finally, the species pair of *Sarotherodon* in Lake Ejagham clearly appears admixed between Barombi *Sarotherodon* and outgroup populations of riverine *Sarotherodon* in other parts of West Africa (Figs. 3A, 4A). This pattern is inconsistent with a single founder event and instead suggests either introgression or speciation from a hybrid swarm (Fig. 1B, C).

Specific patterns of introgression were not always consistent across different analyses. This should be expected given the complex colonization histories inferred, sparse genomic sampling, large amount of missing data inherent to RADseq (Arnold et al. 2013; Davey et al. 2013), absence of closely related genomes for alignment, limited power to detect introgression after substantial drift in these isolated populations (Patterson et al. 2012), and probable absence of the actual colonizing populations in our sample (which may be long extinct). However, inference of introgression with f_4 statistics is robust to missing data and does not require sampling the exact source of introgressed alleles, only closely related populations (Reich et al. 2009; Pickrell and Pritchard 2012). Specific introgression events were also variable in similar studies estimating complex population histories (Brandvain et al. 2014).

ADDITIONAL CRITERIA FOR SYMPATRIC SPECIATION AND THE NECESSITY OF AUTOMATIC MAGIC TRAITS

It is also worth revisiting three additional arguments in support of sympatric speciation in Cameroon cichlids. (1) Are these craters so isolated that repeated colonization is implausible? To the contrary, all four craters have been colonized by multiple fish lineages (six in Barombi Mbo [Trewavas et al. 1972], three in Bermin [Stiassny et al. 1992], and five in Ejagham [Schliewen et al. 2001]), suggesting that repeated colonization by any one population, such as cichlids, is quite possible. (2) Do all sympatric species coexist and breed within the same habitat? This is true for littoral species: for example, all three Ejagham *Coptodon* species guarded breeding territories less than 1 m apart (Martin 2013). However, there are also deep-water and open-water habitat specialists in all four cichlid radiations that may exclusively breed within their microhabitats (Martin 2012). (3) Are sympatric species "good" species with substantial levels of reproductive isolation and distinct phenotypes (Coyne and Orr 2004)? There are ecologically and phenotypically distinct species within Cameroon cichlid radiations (e.g., sponge-eating specialist *Pungu maclareni* [Schliewen et al. 1994]); however, there are also species complexes in all three lakes that exhibit a unimodal distribution of morphologies and appear to be stalled in the earliest phases of incipient speciation, only exhibiting partial bimodality in breeding coloration (Martin 2012, 2013; J. S. Cutler, C. H. Martin, pers. obs.). We found minimal genetic differentiation within these species complexes (Figs. 3, 4).

Thus, the most ecologically and phenotypically distinct species within the Cameroon cichlid radiations are restricted to specialized habitats and breed seasonally, suggesting that speciation was facilitated by an environmental gradient (Doebeli and Dieckmann 2003; Seehausen et al. 2008). In contrast, species complexes that do coexist and breed in the same littoral habitat are not genetically or phenotypically distinct. This suggests that

without the benefit of some spatial isolation, species complexes may become stalled in the earliest phases of the divergence process, as predicted by many theoretical models (e.g., Matessi et al. 2002; Bürger and Schneider 2006; Otto et al. 2008). Consistent with this prediction, although assortative mating was very strong in Ejagham *Coptodon* (Martin 2013), the strength of disruptive selection on trophic morphology in both Barombi Mbo *Stomatepia* and Ejagham *Coptodon* was weak and may not be sufficient to complete the speciation process (Martin 2012).

Available evidence suggests that all crater lake cichlid radiations speciated with the help of double invasions (Schliewen et al. 2006; Geiger et al. 2013) or remain stalled as incipient species complexes (Elmer et al. 2010b; Martin 2012, 2013). To our knowledge, all compelling examples of sympatric speciation besides crater lake cichlid radiations involve some form of automatic linkage between ecological divergence and mating time or location, known as "automatic magic traits" (see review in Servedio et al. 2011). For example, pea aphids mate on their respective host plants in sympatry (Via 1999), a widespread mechanism in which assortative mating automatically results from divergent host preference in phytophagous insects (Berlocher and Feder 2002) and their parasitoids (Forbes et al. 2009). Similar mate-where-you-eat preferences explain sympatric divergence in mole rats (Hadid et al. 2013), spiny mice (Hadid et al. 2014), and even Lorde Howe Island palms, in which growth on different soil types induces different flowering times (Savolainen et al. 2006). Automatic magic traits may have played a role in the speciation of deep-water or open-water habitat specialists within sympatric Cameroon cichlid radiations if these species mate within their microhabitats. In contrast, incipient species complexes in these lakes do not segregate in different mating habitats or breed seasonally (Martin 2013; C. H. Martin, J. S. Cutler, pers. obs.). Although there are many possibilities for classic magic traits in cichlids (e.g., Seehausen et al. 2008; Martin 2010, 2013; Maan and Seehausen 2011), the conspicuous absence of automatic magic traits from sympatric incipient species complexes of cichlids suggests this feature may be needed to complete the process of sympatric speciation in nature.

Conclusion

We found evidence for additional gene flow with neighboring rivers in all four sympatric radiations of Cameroon cichlids following initial crater colonization. This complex colonization history calls into question the celebrated status of Cameroon cichlids as compelling examples of sympatric speciation in nature: any period of allopatry allows for the buildup of partial reproductive isolation with later waves of colonists and violates the strict criterion of sympatric coexistence throughout the speciation process (Coyne and Orr 2004). Indeed, we found evidence that some sympatric species were more closely related to outgroups than others, suggesting they may have speciated due to introgression from later waves of colonists. Nonetheless, we cannot pinpoint a role for secondary colonization in every speciation event and we certainly cannot rule out sympatric speciation confined to subclades within these radiations.

Comparing variable progress toward speciation within these radiations suggests an interesting constraint on sympatric speciation. The most ecologically and phenotypically distinct crater lake cichlid species may benefit from assortative mating within microhabitats. In contrast, species that breed in sympatry show minimal genetic and phenotypic divergence and may be examples of stalled incipient speciation. This pattern is paralleled in the best remaining examples of sympatric speciation in nature: all involve assortative mating by microhabitat. This suggests that existing sympatric speciation models may overestimate the plausibility of this process without automatic linkage between divergent ecology and mating location. Alternatively, the strong and persistent disruptive ecological selection needed to complete sympatric speciation without automatic linkage may be rare in nature.

ACKNOWLEDGMENTS

This study was funded by the Miller Institute for Basic Research in the Sciences, a Young Explorer's Grant from the National Geographic Society, a Lewis and Clark Field Research Grant from the American Philosophical Society, an NSF Graduate Research Fellowship, and a UC Davis Genome Center Pilot Grant to CHM, and by a United States Peace Corps grant to JSC. We gratefully acknowledge the governments of Cameroon, Ghana, DRC, CAR, Tanzania, and Ethiopia for permission to conduct this research. We also thank the local chiefs of Barombi Mbo, Barombi ba Kotto, Eyumojock, Bermin and many others throughout Africa for permission to collect in their lakes and rivers and live in their villages. We thank P. Enyang for field assistance, E. Bezault for providing tissue samples from Ghana, A. Lamboj for permission to use his photo of Stomatepia mongo, and N. Gonwuou for assistance with permits.

DATA ARCHIVING

SNP datasets are deposited in the Dryad Digital Repository, doi 10.5061/dryad.b28p1. Raw sequenced Illumina reads from each individual are deposited in the NCBI Short Read Archive, Bioproject PRJNA282170 and accession numbers SAMN03568274-SAMN03568549.

LITERATURE CITED

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.

Arnold, B., R. B. Corbett-Detig, D. Hartl, and K. Bomblies. 2013. RADseq underestimates diversity and introduces genealogical biases due to nonrandom haplotype sampling. Mol. Ecol. 22:3179–3190.

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.

Berlocher, S., and J. Feder. 2002. Sympatric speciation in phytophagous insects: moving beyond controversy? Annu. Rev. Entomol. 47:773–815.

Bolnick, D. I., and B. M. Fitzpatrick. 2007. Sympatric speciation: models and empirical evidence. Annu. Rev. Ecol. Evol. Syst. 38:459–487.

Brandvain, Y., A. M. Kenney, L. Flagel, G. Coop, and A. L. Sweigart. 2014. Speciation and introgression between *Mimulus nasutus* and *Mimulus guttatus*. PLoS Genet. 10:e1004410.

Brawand, D., C. E. Wagner, Y. I. Li, M. Malinsky, I. Keller, S. Fan, O. Simakov, A. Y. Ng, Z. W. Lim, E. Bezault, et al. 2014. The genomic substrate for adaptive radiation in African cichlid fish. Nature 513:375–381.

Bürger, R., and K. A. Schneider. 2006. Intraspecific competitive divergence and convergence under assortative mating. Am. Nat. 167:190–205.

Catchen, J., P. A. Hohenlohe, S. Bassham, A. Amores, and W. A. Cresko. 2013. Stacks: an analysis tool set for population genomics. Mol. Ecol. 22:3124–3140.

Cornen, G., Y. Bande, P. Giresse, and J. Maley. 1992. The nature and chronostratigraphy of quaternary pyroclastic accumulations from Lake Barombi Mbo (West-Cameroon). J. Volcanol. Geotherm. Res. 51:357–374.

Coyne, J. A., and H. A. Orr. 2004. Speciation. Sinauer Associates, Sunderland, MA.

Cristescu, M. E., S. J. Adamowicz, J. J. Vaillant, and D. G. Haffner. 2010. Ancient lakes revisited: from the ecology to the genetics of speciation. Mol. Ecol. 19:4837–4851.

Davey, J. W., T. Cezard, P. Fuentes-Utrilla, C. Eland, K. Gharbi, and M. L. Blaxter. 2013. Special features of RAD sequencing data: implications for genotyping. Mol. Ecol. 22:3151–3164.

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

Doebeli, M., and U. Dieckmann. 2003. Speciation along environmental gradients. Nature 421:259–264.

Doebeli, M., U. Dieckmann, J. A. Metz, and D. Tautz. 2005. What we have also learned: adaptive speciation is theoretically plausible. Evolution 59:691–695.

Dunz, A. R., and U. K. Schliewen. 2010. Description of a *Tilapia* (*Coptodon*) species flock of Lake Ejagham (Cameroon), including a redescription of *Tilapia deckerti* Thys van den Audenaerde, 1967. Spixiana 33:251–280.

Durand, E. Y., N. Patterson, D. Reich, and M. Slatkin. 2011. Testing for ancient admixture between closely related populations. Mol. Biol. Evol. 28:2239–2252.

Earl, D. A. 2012. STRUCTURE HARVESTER: a webite and program for visualizing STRUCTURE output and implementing the Evanno method. Conserv. Genet. Resour. 4:359–361.

Edwards, S. V. 2009. Is a new and general theory of molecular systematics emerging? Evolution 63:1–19.

Elmer, K. R., H. Kusche, T. K. Lehtonen, and A. Meyer. 2010a. Local variation and parallel evolution: morphological and genetic diversity across a species complex of neotropical crater lake cichlid fishes. Philos. Trans. R. Soc. Lond. B. Biol. Sci. 365:1763–1782.

Elmer, K. R., T. K. Lehtonen, A. F. Kautt, C. Harrod, and A. Meyer. 2010b. Rapid sympatric ecological differentiation of crater lake cichlid fishes within historic times. BMC Biol. 8:60.

Elshire, R., J. Glaubitz, Q. Sun, J. Poland, K. Kawamoto, E. Buckler, and S. Mitchell. 2011. A robust, simple genotyping-by-sequencing (GBS) approach for high diversity species. PLoS One 6:e19379.

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.

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

Felsenstein, J. 1981. Skepticism towards Santa Rosalia, or why are there so few kinds of animals? Evolution 1981:124–138.

Fitzpatrick, B. M., J. A. Fordyce, and S. Gavrilets. 2008. What, if anything, is sympatric speciation? J. Evol. Biol. 21:1452–1459.

Forbes, A. A., T. H. Q. Powell, L. L. Stelinski, J. J. Smith, and J. L. Feder. 2009. Sequential sympatric speciation across trophic levels. Science 323:776–779.

Friedman, M., B. P. Keck, A. Dornburg, R. I. Eytan, C. H. Martin, C. D. Hulsey, P. C. Wainwright, and T. J. Near. 2013. Molecular and fossil evidence place the origin of cichlid fishes long after Gondwanan rifting. Proc. R. Soc. B Biol. Sci. 280:1733.

Gavrilets, S. 2004. Fitness landscapes and the origin of species. Princeton Univ. Press, Princeton, NJ.

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:e74901.

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

Gow, J. L., S. M. Rogers, M. Jackson, and D. Schluter. 2008. Ecological predictions lead to the discovery of a benthic–limnetic sympatric species pair of threespine stickleback in Little Quarry Lake, British Columbia. Can. J. Zool. 86:564–571.

Green, R. E., J. Krause, A. W. Briggs, T. Maricic, U. Stenzel, M. Kircher, N. Patterson, H. Li, W. Zhai, M. H.-Y. Fritz, N. F. Hansen, et al. 2010. A draft sequence of the Neanderthal genome. Science 328:710–722.

Hadid, Y., S. Tzur, T. Pavlícek, R. Šumbera, J. Šklíba, M. Lövy, O. Fragman-Sapir, A. Beiles, R. Arieli, S. Raz, et al. 2013. Possible incipient sympatric ecological speciation in blind mole rats (*Spalax*). Proc. Natl. Acad. Sci. USA 110:2587–2592.

Hadid, Y., T. Pavlícek, A. Beiles, R. Ianovici, S. Raz, and E. Nevo. 2014. Sympatric incipient speciation of spiny mice *Acomys* at "Evolution Canyon," Israel. Proc. Natl. Acad. Sci. USA 111:1043–1048.

Harris, K., and R. Nielsen. 2013. Inferring demographic history from a spectrum of shared haplotype lengths. PLoS Genet. 9:e1003521.

Heliconius Genome Consortium. 2012. Butterfly genome reveals promiscuous exchange of mimicry adaptations among species. Nature 487:94–98.

Herder, F., A. W. Nolte, J. Pfaender, J. Schwarzer, R. K. Hadiaty, and U. K. Schliewen. 2006. Adaptive radiation and hybridization in Wallace's Dreamponds: evidence from sailfin silversides in the Malili Lakes of Sulawesi. Proc. R. Soc. B Biol. Sci. 273:2209–2217.

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

Joyce, D. A., D. H. Lunt, M. J. Genner, G. F. Turner, R. Bills, and O. Seehausen. 2011. Repeated colonization and hybridization in Lake Malawi cichlids. Curr. Biol. 21:R108–R109.

Kassen, R., D. Schluter, and J. D. Mcphail. 1995. Evolutionary history of threespine sticklebacks (*Gasterosteus* spp.) in British Columbia: insights from a physiological clock. Can. J. Zool. 73:2154–2158.

Keller, I., C. E. Wagner, L. Greuter, S. Mwaiko, O. M. Selz, A. Sivasundar, S. Wittwer, and O. Seehausen. 2013. Population genomic signatures of divergent adaptation, gene flow and hybrid speciation in the rapid radiation of Lake Victoria cichlid fishes. Mol. Ecol. 22:2848–2863.

Kirkpatrick, M., and V. Ravigné. 2002. Speciation by natural and sexual selection: models and experiments. Am. Nat. 159(Suppl):S22–S35.

Kocher, T. D. 2004. Adaptive evolution and explosive speciation: the cichlid fish model. Nat. Rev. Genet. 5:288–298.

Kondrashov, A., and F. Kondrashov. 1999. Interactions among quantitative traits in the course of sympatric speciation. Nature 400:351–354.

Langmead, B., and S. Salzberg. 2012. Fast gapped-read alignment with Bowtie 2. Nat. Methods 9:357–359.

Lemmon, E. M., and A. R. Lemmon. 2013. High-throughput genomic data in systematics and phylogenetics. Annu. Rev. Ecol. Evol. Syst. 44:99–121.

Losos, J. B. 2011. Lizards in an evolutionary tree: ecology and adaptive radiation of anoles. Univ. of California Press, Berkeley, CA.

Maan, M. E., and O. Seehausen. 2011. Ecology, sexual selection and speciation. Ecol. Lett. 14:591–602.

Martin, C. H. 2010. Unexploited females and unreliable signals of male quality in a Malawi cichlid bower polymorphism. Behav. Ecol. 21:1195–1202.

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. 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, 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.

Matessi, C., A. Gimelfarb, and S. Gavrilets. 2002. Long-term buildup of reproductive isolation promoted by disruptive selection: how far does it go? Selection 2:41–64.

Mayr, E. 1963. Animal species and evolution. Harvard Univ. Press, Cambridge, MA.

McVean, G. 2009. A genealogical interpretation of principal components analysis. PLoS Genet. 5:e1000686.

Neumann, D., M. L. J. Stiassny, and U. K. Schliewen. 2011. Two new sympatric *Sarotherodon* species (Pisces: Cichlidae) endemic to Lake Ejagham, Cameroon, west-central Africa, with comments on the *Sarotherodon galilaeus* species complex. Zootaxa 2765:1–20.

O'Quin, C. T., A. C. Drilea, M. A. Conte, and T. D. Kocher. 2013. Mapping of pigmentation QTL on an anchored genome assembly of the cichlid fish, *Metriaclima zebra*. BMC Genomics 14:287.

Otto, S. P., M. R. Servedio, and S. L. Nuismer. 2008. Frequency-dependent selection and the evolution of assortative mating. Genetics 179:2091–2112.

Palaiokostas, C., M. Bekaert, M. G. Q. Khan, J. B. Taggart, K. Gharbi, B. J. McAndrew, and D. J. Penman. 2013. Mapping and validation of the major sex-determining region in Nile tilapia (*Oreochromis niloticus*) using RAD sequencing. PLoS One 8:e68389.

Papadopolus, S., 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. USA 108:13188–13193.

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 192:1065–1093.

Pease, J., and M. Hahn. 2015. Detection and polarization of introgression in a five-taxon phylogeny. Syst. Biol. In press.

Peterson, B. B. K., J. N. J. Weber, E. H. E. Kay, H. S. Fisher, and H. E. Hoekstra. 2012. Double digest radseq: an inexpensive method for de novo SNP discovery and genotyping in model and non-model species. PLoS One 7:e37135.

Pfennig, D. W., and K. S. Pfennig. 2012. Evolution's wedge: competition and the origins of diversity. Univ. of California Press, Berkeley, CA.

Pickrell, J. K., and J. K. Pritchard. 2012. Inference of population splits and mixtures from genome-wide allele frequency data. PLoS Genet. 8: e1002967.

Poelstra, J. W., N. Vijay, C. M. Bossu, H. Lantz, B. Ryll, I. Müller, V. Baglione, P. Unneberg, M. Wikelski, M. G. Grabherr, et al. 2014. The genomic landscape underlying phenotypic integrity in the face of gene flow in crows. Science 344:1410–1414.

Pool, J. E., and R. Nielsen. 2009. Inference of historical changes in migration rate from the lengths of migrant tracts. Genetics 181:711–719.

Price, A., N. Patterson, and R. Plenge. 2006. Principal components analysis corrects for stratification in genome-wide association studies. Nat. Genet. 38:904–909.

Pritchard, J. K., M. Stephens, and P. Donnelly. 2000. Inference of population structure using multilocus genotype data. Genetics 155:945–959.

Purcell, S., B. Neale, and K. Todd-Brown. 2007. PLINK: a tool set for whole-genome association and population-based linkage analyses. Am. J. Hum. Genet. 81:559–575.

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

Rohland, N., and D. Reich. 2012. Cost-effective, high-throughput DNA sequencing libraries for multiplexed target capture. Genome Res. 22:939–946.

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.

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. Pääbo. 1994. Sympatric speciation suggested by monophyly of crater lake cichlids. Nature 6472:629–632.

Schliewen, U. K., K. Rassmann, M. Markmann, J. Markert, T. Kocher, and D. Tautz. 2001. Genetic and ecological divergence of a monophyletic cichlid species pair under fully sympatric conditions in Lake Ejagham, Cameroon. Mol. Ecol. 10:1471–1488.

Schliewen, U. K., T. D. Kocher, K. R. McKaye, O. Seehausen, and D. Tautz. 2006. Evolutionary biology: evidence for sympatric speciation? Nature 444:E12–E13.

Schluter, D., and J. McPhail. 1992. Ecological character displacement and speciation in sticklebacks. Am. Nat. 1992:85–108.

Schwarzer, J., B. Misof, D. Tautz, and U. K. Schliewen. 2009. The root of the East African cichlid radiations. BMC Evol. Biol. 9:186.

Seehausen, O. 2004. Hybridization and adaptive radiation. Trends Ecol. Evol. 19:198–207.

———. 2006. African cichlid fish: a model system in adaptive radiation research. Proc. Biol. Sci. 273:1987–1998.

Seehausen, O., Y. Terai, I. S. Magalhaes, K. L. Carleton, H. D. J. Mrosso, R. Miyagi, I. van der Sluijs, M. V Schneider, M. E. Maan, H. Tachida, et al. 2008. Speciation through sensory drive in cichlid fish. Nature 455:620–626.

Servedio, M. R., and R. Bürger. 2014. The counterintuitive role of sexual selection in species maintenance and speciation. Proc. Natl. Acad. Sci. USA 111:8113–8118.

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.

Shaw, K. L. 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. Proc. Natl. Acad. Sci. USA 99:16122–16127.

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

Stamatakis, A. 2014a. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. Bioinformatics 30:1312–1313.

———. 2014b. The RAxML v8.0. X Manual. Pp. 1–54.

Stiassny, M. L. J., U. K. Schliewen, and W. J. Dominey. 1992. A new species flock of cichlid fishes from Lake Bermin, Cameroon with a description

of eight new species of *Tilapia* (Labroidei: Cichlidae). Ichthyol. Explor. Freshwaters 3.4:311–346.

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.

Turner, T. L., M. W. Hahn, and S. V Nuzhdin. 2005. Genomic islands of speciation in *Anopheles gambiae*. PLoS Biol. 3:e285.

Via, S. 1999. Reproductive isolation between sympatric races of pea aphids. I. Gene flow restriction and habitat choice. Evolution 53:1446–1457.

Wagner, C. E., L. J. Harmon, and O. Seehausen. 2012. Ecological opportunity and sexual selection together predict adaptive radiation. Nature 487:366–369.

Wu, C.-I. 2001. The genic view of the process of speciation. J. Evol. Biol. 14:851–865.

Associate Editor: M. Hahn
Handling Editor: R. Shaw

Supporting Information

Additional Supporting Information may be found in the online version of this article at the publisher's website:

Figure S1. Maximum likelihood phylogenies for Oreochromini (*a, b*) and Coptodonini (*c, d*).
Figure S2. Principal component analysis of genetic variance.
Figure S3. Principal component analysis of genetic variance between the Barombi *Sarotherodon* and Ejagham *Sarotherodon* sympatric radiations.
Table S1. Summary of STRUCTURE runs and statistics used for calculating Evanno's Δ K (35).
Appendix S1. Species, location, collector, museum voucher numbers for cataloged specimens, and sample size per population for all individuals genotyped at more than 5% of total loci and used for analyses.