MOLECULAR ECOLOGY

Molecular Ecology (2010) **19**, 940–951 doi: 10.1111/j.1365-294X.2010.04529.x

Geography disentangles introgression from ancestral polymorphism in Lake Malawi cichlids

MERYL C. MIMS,*† C. DARRIN HULSEY,‡ BENJAMIN M. FITZPATRICK‡ and J. TODD STREELMAN†

*School of Aquatic and Fishery Sciences, University of Washington, 1122 NE Boat Street, Seattle, WA 98195, USA, †School of Biology, Petit Institute for Bioengineering and Bioscience, Georgia Institute of Technology, 315 Ferst Drive, Atlanta, GA 30332, USA, ‡Department of Ecology and Evolutionary Biology, University of Tennessee, Knoxville, TN 37996, USA

Abstract

Phenotypically diverse Lake Malawi cichlids exhibit similar genomes. The extensive sharing of genetic polymorphism among forms has both intrigued and frustrated biologists trying to understand the nature of diversity in this and other rapidly evolving systems. Shared polymorphism might result from hybridization and/or the retention of ancestrally polymorphic alleles. To examine these alternatives, we used new genomic tools to characterize genetic differentiation in widespread, geographically structured populations of *Labeotropheus fuelleborni* and *Metriaclima zebra*. These phenotypically distinct species share mitochondrial DNA (mtDNA) haplotypes and show greater mtDNA differentiation among localities than between species. However, Bayesian analysis of nuclear single nucleotide polymorphism (SNP) data revealed two distinct genetic clusters corresponding perfectly to morphologically diagnosed *L. fuelleborni* and *M. zebra*. This result is a function of the resolving power of the multi-locus dataset, not a conflict between nuclear and mitochondrial partitions. Locus-by-locus analysis showed that mtDNA differentiation between species (F_{CT}) was nearly identical to the median single-locus SNP F_{CT}. Finally, we asked whether there is evidence for gene flow at sites of co-occurrence. We used simulations to generate a null distribution for the level of differentiation between co-occurring populations of *L. fuelleborni* and *M. zebra* expected if there was no hybridization. The null hypothesis was rejected for the SNP data; populations that co-occur at rock reef sites were slightly more similar than expected by chance, suggesting recent gene flow. The coupling of numerous independent markers with extensive geographic sampling and simulations utilized here provides a framework for assessing the prevalence of gene flow in recently diverged species.

Keywords: evolutionary radiation, hybridization, mitochondrial introgression, reproductive isolation, single nucleotide polymorphism, sympatry

Received 24 August 2009; revision received 16 December 2009; accepted 22 December 2009

Introduction

As populations and species from recently evolved groups diverge, the genomes of individuals diverge as well. Early in this process, individual genomes are mosaics comprised of many genes that show no signature of differentiation and a few that do (Campbell & Bernatchez 2004; Via & West 2008; Nosil *et al*. 2009). If an ancestral population carried genetic polymorphism, then descendant lineages are expected to share polymorphic alleles for some time. Recently diverged groups can also come to share genetic polymorphism in some portions of the genome through hybridization if pre-mating barriers are not sufficient to prevent local gene flow (Gow *et al*. 2006). Both ancestral polymorphism and hybridization contribute to adaptation because they provide standing genetic variation upon

Correspondence: J. Todd Streelman, Fax: (404) 894-0519; E-mail: todd.streelman@biology.gatech.edu

which natural selection can act (Seehausen 2004; Grant *et al*. 2005; Barrett & Schluter 2008). Notably, even low rates of hybridization can sustain the sharing of polymorphism across porous species boundaries, but ancestrally polymorphic alleles are expected to reach fixation over time due to natural selection and/or genetic drift (Avise 1994). Evolutionary biologists would thus like to isolate the source of standing genetic variation available for selection and local adaptation in recently evolved groups (Campbell & Bernatchez 2004; Via 2009; Schluter & Conte 2009).

The cichlid fishes of Lake Malawi represent the most rapidly diverging lineage of vertebrates (Danley & Kocher 2001). Understanding how reproductive boundaries form, are maintained, and contribute to the incredible diversity of this species flock stretches the limits of evolutionary inference drawn from molecular analyses. What complicates our understanding of speciation in this group even more is that many of the over 500 morphologically diagnosable Malawi cichlid species are reproductively compatible in the lab and show minor genetic differentiation in the wild (Kocher *et al.* 1995; Won *et al*. 2005, 2006). Lake Malawi cichlids exhibit mosaic genomes, with alleles shared across the entire species flock. For example, only 2–4% of single nucleotide polymorphisms (SNPs) show the signature of evolutionary differentiation (i) between populations of species, (ii) between species, and (iii) between major Malawi lineages (Loh *et al*. 2008). Allele sharing could be the result of recent hybridization, the retention of ancestral polymorphisms, or both. To quantitatively assess the extent to which these two processes shape patterns of genetic diversity in Lake Malawi cichlid lineages, we examined mitochondrial and nuclear divergence among widespread, geographically structured populations of *Labeotropheus fuelleborni* and *Metriaclima zebra*.

Until recently, hybridization was thought to be rare in Lake Malawi (Albertson *et al.* 1999), echoing the thought among vertebrate biologists that the phenomenon was not 'evolutionarily constructive' (Seehausen 2004). Hypotheses of hybrid origin for some Malawi cichlid species and populations were first inferred from transitional morphologies coupled with intermediate geographic locations relative to parent populations (Stauffer *et al.* 1996). Recent molecular studies have added to the evidence that hybridization occurs and suggested that it might be common. Smith *et al*. (2003) examined a *Metriaclima* sp. population at Makanjila Point using microsatellites and concluded that this population is of hybrid origin. Furthermore, Streelman *et al*. (2004) found support for recent hybridization between an introduced population of *Cynotilapia afra* (Munthali & Ribbink 1998) and the native species *M. zebra* in Lake Malawi National Park. Hybridization occurs in the wild between morphologically diagnosable cichlid lineages endemic to Lake Malawi.

Nevertheless, the degree of hybridization and its contribution to the genetic and phenotypic variation found in the Malawi species flock is difficult to ascertain from surveys of genetic variation alone. Because Malawi cichlids have diversified within a timeframe spanning approximately a million years (Owen *et al*. 1990; Won *et al*. 2005; Genner *et al*. 2007), the observation of shared polymorphism might be equally explained by insufficient time for ancestral alleles to become fixed within lineages (Moran & Kornfield 1993). Therefore, the low-resolution phylogenies (or specific nodes of phylogenies) produced by genetic analyses of Malawi cichlids (e.g. Albertson *et al*. 1999; Kidd *et al*. 2006; Hulsey *et al.* 2007) could be the result of extensive hybridization, ancestral polymorphism, or both. Despite continued suggestions of the importance of hybridization to cichlid diversification (Parnell *et al*. 2008; Stelkens & Seehausen 2009), it remains difficult to parse introgression from ancestral polymorphism when both processes are inferred from observations of allele sharing between populations and species.

In this study, we aimed to disentangle the contributions of gene flow and ancestral polymorphism to patterns of mitochondrial and nuclear DNA diversity in two extensively studied Lake Malawi species. Both *L. fuelleborni* and *M. zebra* are members of the rock-dwelling Malawi lineage called 'mbuna' (Genner & Turner 2005). Mbuna species inhabit rock-reef communities discontinuously distributed throughout the lake; these communities are characterized by a high degree of endemism and large variance in species richness (Ribbink *et al*. 1983). Mbuna individuals are highly site-specific as adults, exhibit 'homing' behaviour if experimentally displaced (Hert 1992), and are the products of maternal mouth-brooding as embryos (Genner & Turner 2005). *L. fuelleborni* and *M. zebra* are among the most cosmopolitan of mbuna [they are found locally at most rock reefs (Ribbink *et al*. 1983)], and both species exhibit population genetic structure, morphological differentiation by population, and local colour morphs (Arnegard *et al*. 1999; Danley *et al*. 2000; Streelman *et al*. 2007a).

There are several lines of indirect evidence to suggest that *L. fuelleborni* and *M. zebra* might hybridize in the wild. First, these and other mbuna share mtDNA haplotypes (Moran & Kornfield 1993; Shaw *et al*. 2000; Hulsey *et al*. 2007). Second, *L. fuelleborni* and *M. zebra* produce viable and fertile hybrid offspring in the lab (McElroy & Kornfield 1993; Albertson *et al*. 2003). Third, species in the genus *Tropheops* are morphological intermediates between *L. fuelleborni* and *M. zebra* and might be the

result of hybrid speciation (McElroy & Kornfield 1993; Albertson & Kocher 2005). Reproductive barriers between these and other mbuna species might frequently break down in the wild (Genner & Turner 2005). On the other hand, *L. fuelleborni* and *M. zebra* are morphologically distinct. The square jaw, overhanging snout, and possession of exclusively tricuspid oral jaw teeth categorically distinguish *L. fuelleborni* from *M. zebra* which exhibits a more typical rounded oral jaw coupled with bicuspid and tricuspid teeth (Ribbink *et al*. 1983; Reinthal 1990; Albertson *et al*. 2003). The two species often exhibit divergent coloration in sympatry, and this is thought to be important in cichlid mate recognition (Kornfield & Smith 2000). These morphological differences suggest that prezygotic and/or ecological mechanisms may ensure reproductive isolation between the two species (Kornfield & Smith 2000; Genner & Turner 2005).

Because species of mbuna such as *L. fuelleborni* and *M. zebra* are isolated to rock-reefs throughout the lake (Ribbink *et al*. 1983) and migration of fish between these rocky 'islands' is thought to be low (Danley *et al*. 2000), the dynamics of hybridization at one island could be independent of evolutionary processes at others. Given that co-occurring *L. fuelleborni* and *M. zebra* have the opportunity for gene flow, whereas populations at distinct rock reefs generally do not, this archipelago-like range of each species provides natural replication to test if co-occurring populations share more alleles than expected by chance. If co-occurring populations of the two species are more similar than randomly chosen population pairs, then this would be suggestive of local gene flow. Alternatively, if co-occurring individuals of the two species are no more genetically similar than individuals from random population pairs, it would suggest that recent hybridization between the two species is virtually non-existent, and that the observation of allele sharing is better explained by ancestral polymorphism or ancient hybridization. Combining spatial null models with extensive genotyping of mitochondrial and nuclear DNA markers should provide a robust view of the prevalence of recent and local gene flow between *L. fuelleborni* and *M. zebra*.

Materials and methods

Geographic sampling

Individuals of *L. fuelleborni* and *M. zebra* were collected from 13 locations throughout the southern end of Lake Malawi (Table 1). A total of 76 *L. fuelleborni* and 83 *M. zebra* were included in our analyses of mitochondrial *ND2* variation. No *M. zebra* were included from Chinyamkwazi and Chinyanwezi because they do not occur on these islands (Ribbink *et al*. 1983; JTS personal observation). Populations of *L. fuelleborni* are found at Masinge, Mazinzi Reef and Thumbi West, but these were either not sampled or not included in our analysis. We also included sequences from 29 individuals of other mbuna species collected from various locations reported in Hulsey *et al*. (2007), to examine whether major *L. fuelleborni* and *M. zebra* mitochondrial haplotypes were species-specific. Other mbuna species in the mitochondrial haplotype network included members of the genera *Cynotilapia* (*C. afra*), *Labidochromis* (*L. gigas*), *Melanochromis* (*M. auratus* and *M. vermivorus*), *Metriaclima* (*M. aurora* and *M. callainos*), *Psuedotropheus* (*P. crabro* and *P. elongatus*), and *Tropheops* (*T.* 'orange chest', *T.* 'red cheek', *T. gracilior*, and *T. microstoma*). Nuclear SNP analysis included 83 *M. zebra* individuals and 61 *L. fuelleborni* individuals, following Loh *et al*. (2008). Slight variation in sample size between populations used in the mtDNA and the SNP analyses was due to exclusion of poor-quality SNP data for certain individuals; however, all populations represented in the mitochondrial haplotype network were represented by individuals (usually the same individuals) in the SNP analysis.

Table 1 Geographic locations and the number of individuals per site, per species, used in analysis (see also Fig. 2)

Location	Loc. Code	*L. fuelleborni*	*M. zebra*
Chinyamwezi	CMY	8	-
Chinyankwazi	CNK	8	-
Chiofu Bay	CHB	8	6
Domwe Island	DMW	5	10
Eccles Reef	ECC	10	6
Makanjila Point	MKP	9	9
Masinge	MSJ	-	9
Mazinzi Reef	MZI	-	7
Mumbo	MBO	9	8
Otter Point	OTP	2	8
Thumbi West Island	TWI	-	5
West Reef	WST	8	6
Zimbawe	ZBW	9	9
	Total	76	83

A dash (–) indicates no individuals were sampled from that site.

DNA isolation and sequencing

Total genomic DNA was isolated from fin clips at the University of New Hampshire in the laboratory of TD Kocher. We examined mitochondrial variation in *L. fuelleborni*, *M. zebra*, and several other closely related mbuna species at 1040 bp of the *ND2* locus (Kocher *et al*. 1995). We also sequenced three exons and two introns (949 bp) of the nuclear encoded *dlx2* gene for

M. zebra and *L. fuelleborni* individuals. The *dlx2* gene is important for early development of vertebrate brains, jaws and dentitions (Panganiban & Rubenstein 2002). Variable sites recovered from the *dlx2* sequences were analysed both separately and combined with other nuclear SNP genotypes, as discussed below. All new sequences were deposited in Genbank (GU192064–GU192357). The Genbank numbers of previously reported *ND2* sequences utilized here are available from MCM.

We generated DNA sequence data from PCR products. PCR amplifications were carried out in a Perkin–Elmer DNA thermocycler. The reaction volume was 25 μL [16.0 μL of H_2O, 2.5 μL 10 × $MgCl_2$ PCR buffer, 1.25 μL $MgCl_2$, 2.0 μL dNTPs (10 mM), 1.0 μL of each primer (10 μM), 0.25 μL of TAQ, and 1 μL DNA (~15–20 ng)]. Thermal cycling conditions consisted of an initial denaturation step of 94 °C (45 min), 52 °C (45 s), and 72 °C (2 min 30 s). A final incubation of 72 °C for 4 min was added to ensure complete extension of amplified products. The *ND2* and *dlx2* PCR products were separated from unincorporated primers and dNTPs using electrophoresis in low melting agarose gel run in Tris-acetate buffer (pH 7.8). Ethidium bromide (1.5 mg/μL) was added to the gels for visualization. Positively amplified DNA was then purified using an enzymatic combination of 1 μL of Exonuclease I (10.0 U/μL) and 1 μL shrimp alkaline phosphatase (2.0 U/μL) per 10 μL of PCR. Treated PCR products were used as templates for sequencing reactions (Applied Biosystems terminator cycle sequencing reactions). Sequences were generated at the Automated DNA Sequencing Facility at the University of Washington. Sequencher version 4.1 (Gene Codes, Ann Arbor, MI, USA) was then used to assemble the gene sequences.

Nuclear SNPs

We analysed data from 52 haphazardly chosen SNPs from the nuclear genome, genotyped in *L. fuelleborni* and *M. zebra* individuals (Loh *et al*. 2008). These SNPs were identified by comparative analysis of genomic survey sequences from five Lake Malawi cichlid species; summary statistics for these loci are reported in Loh *et al*. (2008) and the associated online files. The SNP loci were genotyped using the GenomeLab SNPstream Genotyping System Software Suite v2.3 (Beckman Coulter, Inc.) at the Center for Medical Genomics at Emory University, as described (Loh *et al*. 2008).

Mitochondrial haplotype network

A mitochondrial haplotype network of *ND2* sequences from *L. fuelleborni*, *M. zebra* and other mbuna individuals was constructed using TCS 1.21 software (Clement *et al*. 2000). Sequences were aligned for analysis using Clustal X (Thompson *et al*. 1999), and codon positions were defined using MacClade 4.0 (Maddison & Maddison 2000). The haplotypes, once imported, were analysed using parsimony to generate the network, and the frequencies of haplotypes in each haplogroup were tallied. Additionally, haplotype frequencies for *L. fuelleborni* and *M. zebra* were depicted on a map of rock-reef sampling locations (Fig. 2).

Bayesian clustering

Twenty-three SNPs identified from *dlx2* gene sequences (above) of *L. fuelleborni* and *M. zebra* individuals were combined with 52 SNPs from throughout the genomes of both species (Loh *et al*. 2008). SNPs were treated in two ways. First, we employed a model (Falush *et al*. 2003) that assumed complete linkage among the 23 *dlx2* SNPs and analysed these linked markers in conjunction with the 52 additional independent markers (Table 2, Complete linkage). Second, the 52 other SNPs were assumed unlinked and analysed alone, excluding data from *dlx2* (Table 2, Excluded).

To describe overall structure in the SNP data and to identify admixed genotypes (Gompert *et al*. 2006), we used STRUCTURE 2.2 (Pritchard *et al*. 2000). Probable *K* (i.e. the number of genetic clusters supported by the data) values for the combined dataset of all *L. fuelleborni* and *M. zebra* were expected to fall between $K = 2$ (number of species) and $K = 21$ (number of individual samplings). The optimal number of genetic clusters was determined using the ad hoc statistic ΔK, based on the rate of change in the log probability of data between successive *K* values (Evanno *et al*. 2005). For each *K* (from 1 through 20), we ran ten Markov-Chain Monte Carlo (MCMC) replicates with burn in equal to 10 000 generations and post-burn in equal to 100 000 generations.

AMOVA

We used nested analysis of molecular variance (AMOVA) in Arlequin 3.1 (Excoffier *et al*. 2005) to (i) evaluate genetic differentiation within and between species and (ii) to compare mitochondrial and nuclear data. Nucleotide divergence between individual mtDNA haplotypes (Kimura two-parameter distance) and *dlx2* genotypes (number of differences between individuals) was used to calculate Φ_{ST} (correlation of haplotypes/genotypes within samples relative to the entire data set), Φ_{SC} (correlation within sites relative to the species), and Φ_{CT} (correlation within species relative to entire data set). These values cannot be directly compared to the SNP

data, for which only the traditional (allele frequency-based) F_{ST}, F_{SC}, and F_{CT} can be calculated. Therefore, we also used mtDNA haplotype frequencies to calculate F_{ST}, F_{SC}, and F_{CT}. A similar analysis based on allele frequencies of *dlx2* could not be performed because the SNP data cannot be reliably converted to phased haplotype data.

Multilocus AMOVA of SNPs is not directly comparable to single locus results from mtDNA and *dlx2* because the multilocus results estimate average levels of differentiation and are expected to have lower sampling variance, but also might be strongly influenced by a few highly differentiated markers. Therefore, we performed locus-by-locus AMOVA with the SNP data to evaluate the distribution of *F*-statistics among markers. No appropriate method exists for testing whether the distribution of F_{CT} among markers deviates from a neutral model (Beaumont 2005; Foll & Gaggiotti 2008), but markers in the tails of the distribution might be considered candidates for further analysis (Loh *et al*. 2008). Since mitochondrial DNA might more readily introgress after hybridization (Chan & Levin 2005; Bachtrog *et al*. 2006; Bossu & Near 2009), we were interested in whether mtDNA is an outlier relative to the distribution of F_{CT} for nuclear markers. We therefore compared the F_{CT} of *ND2* to that for each individual SNP locus.

Gene flow between species

To address whether co-occurring populations of *L. fuelleborni* and *M. zebra* are more similar than random interspecific pairs of these two species, we constructed a null distribution by randomly sampling eight *L. fuelleborni* and eight *M. zebra* populations to form eight random pairs (the actual number of sites where the species co-occur in our data set). We then calculated the average pairwise F_{ST} for each of 10 000 such replicates. This procedure estimates the distribution of average F_{ST} between *L. fuelleborni* and *M. zebra* populations assuming no relationship between F_{ST} and whether or not the populations co-occur at sites. We then compared the average pairwise F_{ST} from the eight co-occurring comparisons to this simulated distribution (estimated separately for each marker subset of the data). The fraction of randomizations with mean F_{ST} less than or equal to the observed mean for co-occurring populations is an estimate of the one-tailed *P*-value for the null hypothesis of no relationship between genetic differentiation and co-occurrence. Rejection of this hypothesis leads to acceptance of the alternative that co-occurring populations are more likely to share allelic variation than are random population pairs. This would support the notion that recent gene flow had occurred between the two species where they co-occur. Randomizations were performed using custom-written code in R (http://www.r-project.org).

Results

Mitochondrial haplotypes

The *ND2* network showed three most common haplotypes (labeled 1, 22, 25, Fig. 1); all other haplotypes are likely descendants of these. As expected, all haplotypes

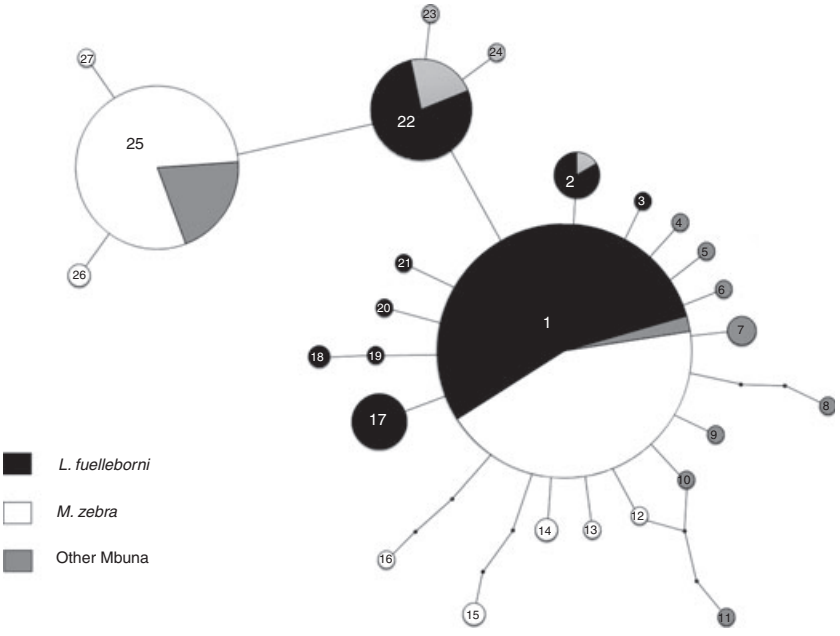

Fig. 1 Parsimony haplotype network of mitochondrial *ND2* sequences for *L. fuelleborni* (black), *M. zebra* (white), and other mbuna species (grey). Each branch (link) represents one base pair difference. Circle size is proportional to the number of individuals represented in the haplotype. Haplotype numbers were arbitrarily chosen.

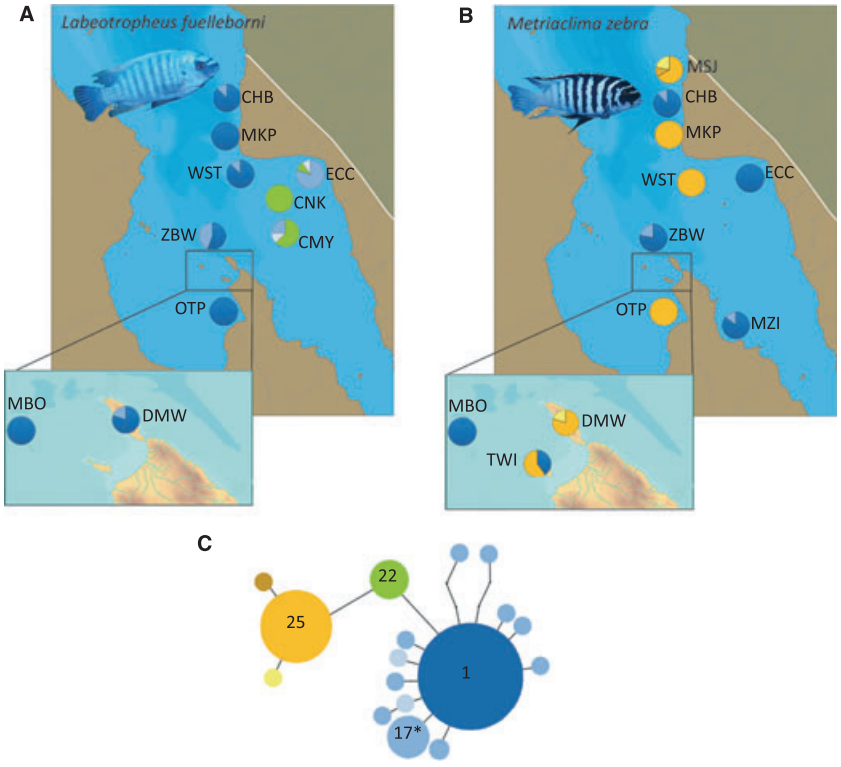

Fig. 2 Geospatial reference maps of *ND2* haplotypes (Fig. 1) for *L. fuelleborni* (a) and *M. zebra* (b). Pie graphs represent haplotype frequencies by location (for location codes, see Table 1), and haplotypes are colour-coded based on the key (c). Haplotype 17 (see also Fig. 1) is denoted by the star (*). Note that other mbuna species (from Fig. 1) are not included here. Lake Malawi maps are modified, with permission, from Cichlid Press (http://www.cichlidpress.com).

Table 2 ΔK values for a given K (number of genetic clusters) for *L. fuelleborni* and *M. zebra* individuals, analysed using all SNPs, where *dlx2* SNPs are treated as a single haplotype (Complete linkage) or excluded (Excluded)

K	Complete linkage	Excluded
2	1361.76	1457.61
3	2.44	2.45
4	1.33	1.96
5	3.47	3.15
6	2.49	1.88
7	0.76	2.23
8	0.82	4.76
9	1.97	0.83
10	1.97	0.86
11	1.29	9.62
12	13.72	0.83
13	1.64	30.5
14	3.05	1.22
15	1.33	1.37
16	2.67	2.45
17	1.97	1.8
18	2.29	1.32
19	2.9	1.32

were quite similar, differing by no more than 8 bases (0.76%). Haplotype 1 includes the majority of all individuals and is found in *L. fuelleborni*, *M. zebra*, and other mbuna species including three of the four *Tropheops* individuals in the analysis (the fourth *Tropheops* individual is included in singleton haplotype 11). In addition to haplotype 1, haplotype 22 and 25 also include a large percentage of total individuals. Haplotypes 1 and 25 together account for 88% of *M. zebra* individuals. While haplotype 25 includes other mbuna species, it does not include any *L. fuelleborni* individuals. Haplotypes 1 and 22 account for 75% of *L. fuelleborni* individuals. Haplotype 22 includes both *L. fuelleborni* individuals and other mbuna species, but no *M. zebra* individuals. None of the three major haplotypes are specific to either *L. fuelleborni* or *M. zebra*, as other mbuna species group within all three haplotypes.

ND2 haplotypes were geographically structured (Fig. 2). Of the 11 locations sampled for *M. zebra* (Fig. 2b), only 1 location (Thumbi West Island) had individuals with both haplotypes 1 and 25. The other ten locations contained only one of the two major *M. zebra* haplotypes. If locations included individuals that fell into peripheral haplotypes (shown in light blue or light yellow), those peripheral haplotypes always grouped with the major haplotype to which they were most closely linked.

The *L. fuelleborni* mitochondrial haplotype map (Fig. 2a) lacks the geographic checkerboard distribution of *M. zebra*. Seven of the ten locations sampled for

L. fuelleborni are exclusively haplotype 1 and its peripheral haplotypes. The three locations that are not exclusively haplotype 1 are clustered in the eastern arm of south Lake Malawi. Haplotype 22 is present only in *L. fuelleborni* individuals from Chinyankwazi (exclusively haplotype 22), Chinyamwezi (haplotype 22 and haplotype 1 peripheral), and in one individual from Eccles Reef. Halpotype 17 (starred) is peripheral to haplotype 1 and is interesting because it is exclusive to Eccles Reef and includes eight of the ten individuals sampled from this location. Haplotype 17 is the only peripheral haplotype which contains a majority of the individuals from a specific location for either species. Haplotype 2 (Fig. 1) also shows evidence for some geographic sorting. While haplotype 2 does appear in some other mbuna species, all *L. fuelleborni* individuals with haplotype 2 are from Zimbawe and Domwe, two geographically adjacent sampling sites.

Bayesian clustering

The STRUCTURE analyses of nuclear SNP data offered little evidence of hybridization between the two species. STRUCTURE was performed with *dlx2* SNPs included as a linked haplotype (Fig. 3) and with the *dlx2* SNPs excluded from the 52 unlinked markers from Loh *et al*. (not shown). In both cases, *L. fuelleborni* and *M. zebra* were assigned to two separate genetic clusters with no misclassified individuals (Fig. 3; Table 2). The most intermediate individual was a *M. zebra* estimated to have 20–25% *L. fuelleborni* ancestry (Fig. 3), however, the 95% confidence interval for this individual included 0% *L. fuelleborni* ancestry. Both analytical treatments (*dlx2* SNPs included as a haplotype, or excluded) of SNP markers yielded statistical support for 2 genetic clusters (Table 2). No congruence was found between *ND2* haplogroups and genetic clusters apparent in the nuclear SNPs (Fig. 3).

AMOVA

The AMOVA demonstrates that for all marker types, the majority of genetic variation is observed within and not between *L. fuelleborni* and *M. zebra* (Table 3). It also confirms the results of other studies that these species are genetically structured in space (Arnegard *et al*. 1999; Danley *et al*. 2000; Streelman *et al*. 2007a). AMOVA results indicate that mtDNA and *dlx2* exhibit similar levels of differentiation between *L. fuelleborni* and *M. zebra* when considering nucleotide divergence. Mitochondrial DNA shows the most differentiation among populations (as expected, given its smaller effective population size). In order to compare mtDNA to the non-*dlx2* SNPs, we examined *F*-statistics rather than sequence divergence. We observed less differentiation between the two species with mtDNA than with the multi-locus SNP data (Table 3). However, we found that the between-species F_{CT} for mtDNA (0.1266) is nearly identical to the median SNP F_{CT} (0.1246) (Fig. 4).

Gene flow between species

Mitochondrial DNA F_{ST} and Φ_{ST} for co-occurring *L. fuelleborni* and *M. zebra* pairs were not significantly different from random population pairs (one tailed $P = 0.583$ and 0.680, respectively; Fig. 5a,b). Likewise, Φ_{ST} calculated from the *dlx2* data for co-occurring pairs was not different than random ($P = 0.738$, Fig. 5C). However, the multilocus F_{ST} calculated from the SNP data was significantly smaller for co-occurring pairs ($P = 0.002$) compared to the null distribution (Fig. 5d). There is a subtle signal of greater average similarity between *L. fuelleborni* and *M. zebra* where they co-occur ($F_{ST} = 0.529$ vs. 0.569 expected), suggesting a small amount of recent gene flow between species at sympatric sites.

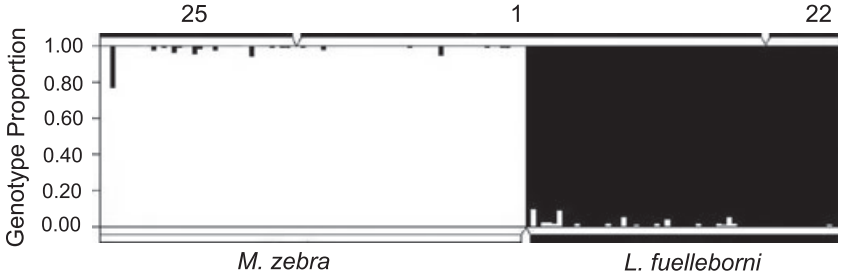

Fig. 3 STRUCTURE plot of genetic assignment for *L. fuelleborni* and *M. zebra*, where $K = 2$ (see ΔK calculations, Table 2), assembled using nuclear SNP data where *dlx2* SNPs are included as a haplotype. Each bar of the plot represents one individual, with estimated likelihood assignment (or genotypic proportion) on the *y*-axis. Across the top we show each individual's *ND2* haplogroup number (see Fig. 1) and across the bottom we plot each individual's morphological species designation.

Table 3 AMOVA results

Source of variation	d.f.	Sum of squares	Variance components	Percentage variation	Phi Statistics	*P*-value
mtDNA (seq-based)						
Among species	1	15.176	0.15278 (Va)	20.45	PhiCT = 0.20449	0.0246
Among pops, within species	19	54.357	0.34766 (Vb)	46.53	PhiSC = 0.58493	<0.0001
Within pops	138	34.045	0.2467 (Vc)	33.02		
Total	158	103.578	0.74714			
mtDNA (hap-based)						
Among species	1	5.592	0.0486 (Va)	12.66	Fct = .12659	0.048
Among pops, within species	18	31.379	0.21175 (Vb)	55.16	Fsc = .63155	<0.0001
Within pops	134	16.554	0.12354 (Vc)	32.18		
Total	153	53.526	0.38389			
DLX2 (seq-based)						
Among species	1	15.775	0.09751 (Va)	19.3	PhiCT = 0.19298	0.003
Among pops, within species	18	36.013	0.11958 (Vb)	23.67	PhiSC = 0.29327	<0.0001
Within pops	268	77.23	0.28817 (Vc)	57.03		
Total	287	129.017	0.50526			
SNPs (allele-based)						
Among species	1	443.605	2.9958 (Va)	39.64	Fct = .39641	0.0001
Among pops, within species	18	389.297	1.28112 (Vb)	16.95	Fsc = .28085	<0.0001
Within pops	268	879.167	3.28048 (Vc)	43.41		
Total	287	1712.069	7.55739			

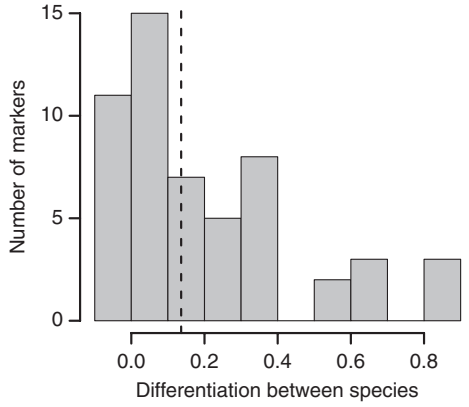

Fig. 4 The genetic differentiation of mitochondrial DNA is similar to that of the median nuclear SNP locus. We plot the F_{CT} of *ND2* (0.1266, dashed line) against the frequency distribution of F_{CT} for individual SNPs (median = 0.1246).

Discussion

The Lake Malawi cichlid flock is an evolutionary enigma because tremendous phenotypic diversity has evolved in a short period of time (Kocher 2004; Streelman *et al*. 2007b; Salzburger 2009) in populations and species with similar genomes and broad sharing of polymorphism (Loh *et al*. 2008). Evolutionary biologists are interested in the sharing of alleles across recently diverged evolutionary lineages because standing genetic variation is an important source for local adaptation (Barrett & Schluter 2008). In this study, we used a large number of independent molecular markers and the replicate spatial distribution of Lake Malawi mbuna populations to evaluate two explanations for allele sharing among groups, ancestral polymorphism and recent gene flow. We observed an overwhelming genetic signal from allele frequencies that *L. fuelleborni* and *M. zebra* represent distinct evolutionary clusters. However, gene flow between species occurs at low levels at sites of co-occurrence.

A complex history of mitochondrial DNA

The within-island monomorphism and checkerboard-like distribution of mtDNA haplotypes for many of the sampled locations supports the idea that gene flow is low among island populations within these species (Figs 1 and 2). The distribution of the three major haplotypes recovered for *ND2* is also suggestive that spatial isolation plays a substantial role in the genetic divergence of both *L. fuelleborni* and *M. zebra*. Other mbuna species included in the *ND2* analysis largely sorted into the three major haplogroups (Fig. 1), indicating that haplotypes were not species-specific for either *L. fuelleborni* or *M. zebra*. Sharing of mtDNA haplotypes is not unusual for these recently evolved Malawi mbuna species (Hulsey *et al*. 2007). Geographically widespread haplotypes 1 (present in *L. fuelleborni* and *M. zebra*) and 25 (absent in *L. fuelleborni*) were linked by haplotype 22, absent in *M. zebra* and localized in *L. fuelleborni* to the

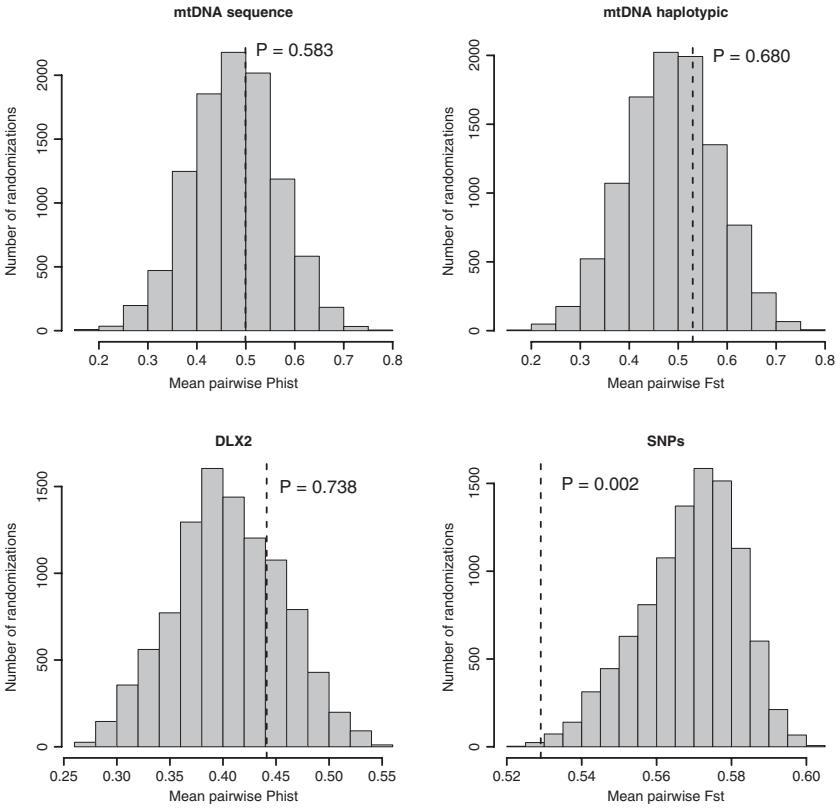

Fig. 5 Comparison of genetic differentiation between *L. fuelleborni* and *M. zebra* when they co-occur at sites (dashed lines) against the frequency distribution of simulated random population pairs for mtDNA (sequence-based and haplotypic), *dlx2*, and multilocus SNP data partitions. Please note differences in scale of x-axes.

eastern sites Chinyankwazi and Chinyamwezi (Figs 1, 2). The most likely explanation for this feature of the haplotype network is extinction of *M. zebra*, carrying haplotype 22, from the isolated eastern locales Chinyankwazi and/or Chinyamwezi (Ribbink *et al*. 1983). In sum, mitochondrial haplotypes are often fixed locally, but mbuna species generally maintain common ancestral alleles.

Evidence from the nuclear genome

Examining numerous independent loci from the nuclear genome provides a robust framework to evaluate the historical processes occurring within and between biological species. The nuclear SNPs support two highly distinct groups that correspond exactly to individuals morphologically identifiable as *L. fuelleborni* and *M. zebra*. No individuals were misclassified by STRUCTURE (Fig. 3). If extensive hybridization were ongoing between *L. fuelleborni* and *M. zebra*, we would have expected admixed genotypes in the STRUCTURE analysis. That the two species were so clearly grouped into two distinct genetic clusters suggests that hybridization is likely to be rare (globally) between the two species.

Although the STRUCTURE analyses appear definitive, the significance of genetic subdivision for either species or populations is difficult to quantify using such heuristic tools alone. The AMOVAs show that the mitochondrial and nuclear genomes of these two cichlid species exhibit significant genetic structure at the population level, as well as between species (Table 3). Coupled with the recovery of single mtDNA haplotypes from most populations (Fig. 2), these results confirm the importance of local demographic processes. While the mtDNA data provide less interspecific discriminatory ability than the SNPs as a whole, the differentiation of mtDNA between *L. fuelleborni* and *M. zebra* is very near the median differentiation of individual nuclear SNPs (Fig. 4). This indicates that the mtDNA is no more or less diagnostic than one would expect by chance for a single locus. These results are not consistent with biased introgression of mitochondria relative to the nuclear genome, as has been reported in other systems (Powell 1983; Bossu & Near 2009).

The subtle signature of gene flow between species

The overwhelming inference that can be drawn from our analyses is that *L. fuelleborni* and *M. zebra* are genetically distinct lineages. If gene flow between the two species occurs, it occurs rarely. For both *ND2* and *dlx2* markers, differentiation at sites of co-occurrence is not significantly different from differentiation among random pairs of interspecific populations (Fig. 5). However,

co-occurring populations showed significantly lower differentiation than expected by chance when we examined the full complement of SNP markers. Therefore, when a large number of independent loci from throughout the genome are examined, there is support for low levels of introgression between *L. fuelleborni* and *M. zebra*. The complete SNP dataset provides higher resolution than the individual mtDNA or *dlx2* analyses, and in this case, the SNPs showed a small signal of hybridization undetectable with single loci examined in isolation. The species *L. fuelleborni* and *M. zebra* are globally genetically distinct, segregate ancestral polymorphism, and exhibit a subtle signature of hybridization where they co-occur.

Hybridization may have substantial impacts on adaptively radiating groups like the Lake Malawi cichlid fishes (Dowling & Secor 1997; Gerber *et al.* 2001; Seehausen 2004; Nolte *et al.* 2005). Gene flow can influence evolution by elevating levels of genetic variation (Lewontin & Birch 1966), and because the island populations of Lake Malawi mbuna are frequently small, additional genetic variation might be critical to population viability. Gene flow impacts the extent of genetic co-variation among traits (Schluter & Conte 2009, the 'transporter' hypothesis), and phenotypic integration is thought to contribute to the evolutionary success of cichlids (Albertson *et al*. 2003, 2005; Hulsey *et al*. 2006; Fraser *et al*. 2009). Hybridization has also been suggested to produce novel Malawi cichlid colour and trophic phenotypes (Streelman *et al*. 2004; Parnell *et al*. 2008). The small amounts of gene flow detected between the two species examined here, if generally characteristic of the mbuna rock-dwellers, might influence community-specific patterns of adaptation in these fishes. For instance, isolated rock reefs harbouring many mbuna species (i.e. Thumbi West houses 36 mbuna species, some of which are introduced and hybridize) may experience different evolutionary trajectories than sites with fewer species (i.e. Zimbawe Rock, ~3 km from Thumbi West, houses nine mbuna species). Coupled with the retention of ancestral polymorphism shared across the species flock (Moran & Kornfield 1993; Loh *et al*. 2008), even low levels of hybridization might allow Malawi cichlids to recycle ancestral variation as local communities of genetic consortia.

The use of SNPs to address future population-level questions for Lake Malawi cichlids and other recently evolved groups is promising given the ability of these markers to so clearly detail patterns of isolation and gene flow. Furthermore, the coupling of numerous independent markers with extensive geographic sampling and simulations that take advantage of population subdivision should provide a robust framework for assessing the prevalence of gene flow among the genomes of recently diverged species.

Acknowledgements

We thank Tom Kocher, Karen Carleton, Craig Albertson, Michael Kidd, Celeste Kidd, Tyrone Spady, Reade Roberts, Pat Danley, Aimee Howe and Richard Zatha for help collecting fishes and the Kocher lab for cooperation in obtaining DNA. The Malawi Parks department provided permits to obtain fish in Lake Malawi National Park. Nick Parnell, Phillip Hollingsworth and Matt Niemiller provided valuable comments on an earlier draft of this manuscript. This research was funded by grants from the National Science Foundation (IOS 0546423) and the AP Sloan Foundation (BR-4499) to JTS. MCM was a GIT President's Undergraduate Research Scholar.

References

Albertson RC, Kocher TD (2005) Genetic architecture sets limits on transgressive segregation in hybrid cichlid fishes. *Evolution*, **59**, 686–690.

Albertson RC, Markert JA, Danley PD, Kocher TD (1999) Phylogeny of a rapidly evolving clade: The cichlid fishes of Lake Malawi, East Africa. *Proceedings of the National Academy of Sciences of the United States of America*, **96**, 5107–5110.

Albertson RC, Streelman JT, Kocher TD (2003) Genetic basis of adaptive shape differences in the cichlid head. *Journal of Heredity*, **94**, 291–301.

Albertson RC, Streelman JT, Kocher TD, Yelick PC (2005) Integration and evolution of the cichlid mandible: the molecular basis of alternative feeding strategies. *Proceedings of the National Academy of Sciences of the United States of America*, **102**, 16287–16292.

Arnegard ME, Markert JA, Danley PD, Stauffer JR, Ambali AJ, Kocher TD (1999) Population structure and colour variation of the cichlid fish *Labeotropheus fuelleborni* Ahl along a recently formed archipelago of rocky habitat patches in southern Lake Malawi. *Proceedings of the Royal Society of London Series B Biological Sciences*, **226**, 119–130.

Avise JC (1994) *Molecular Markers, Natural History and Evolution*. Chapman & Hall, New York, NY.

Bachtrog D, Thornton K, Clark A, Andolfatto P (2006) Extensive introgression of mitochondrial DNA relative to nuclear genes in the Drosophila yakuba species group. *Evolution*, **60**, 292–302.

Barrett RD, Schluter D (2008) Adaptation from standing genetic variation. *Trends in Ecology & Evolution*, **23**, 38–44.

Beaumont MA (2005) Adaptation and speciation: what can F−st tell us? *Trends in Ecology & Evolution*, **20**, 435–440.

Bossu CM, Near TJ (2009) Gene trees reveal repeated instances of mitochondrial introgression in Orangethroat darters (Percidae: Etheostoma). *Systematic Biology*, **58**, 114–129.

Campbell D, Bernatchez L (2004) Generic scan using AFLP markers as a means to assess the role of directional selection in the divergence of sympatric whitefish ecotypes. *Molecular Biology & Evolution*, **21**, 945–956.

Chan KMA, Levin SA (2005) Leaky prezygotic isolation and porous genomes: rapid introgression of maternally inherited DNA. *Evolution*, **59**, 720–729.

Clement M, Posada D, Crandall KA (2000) TCS: a computer program to estimate gene genealogies. *Molecular Ecology*, **9**, 1657–1660.

Danley PD, Markert JA, Arnegard ME, Kocher TD (2000) Divergence with gene flow in the rock-dwelling cichlids of Lake Malawi. *Evolution*, **54**, 7125–1737.

Danley PD, Kocher TD (2001) Speciation in rapidly diverging systems: lessons from Lake Malawi. *Molecular Ecology*, **10**, 1075–1086.

Dowling T, Secor CL (1997) The role of hybridization and introgression in the diversification of animals. *Annual Review of Ecology and Systematics*, **28**, 593–619.

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, Laval G, Schneider S (2005) Arlequin ver. 3.0: An integrated software package for population genetics data analysis. *Evolutionary Bioinformatics Online*, **1**, 47–50.

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.

Fraser GJ, Hulsey CD, Bloomquist R, Uyesugi K, Manley NR, Streelman JT (2009) An ancient gene network is co-opted for teeth on old and new jaws. *PLoS Biology*, **7**, e1000031.

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.

Gerber AS, Tibbets CA, Dowling TE (2001) The role of introgressive hybridization in the evolution of the *Gila robusta* complex (Teliostei: Cyprinidae). *Evolution*, **55**, 2028–2039.

Genner MJ, Seehausen O, Lunt DH, Joyce DA, Shaw PW, Carvalho GR, Turner GF (2007) Age of cichlids: new dates for ancient lake fish radiations. *Molecular Biology and Evolution*, **24**, 1269–1282.

Genner MJ, Turner GF (2005) The mbuna cichlids of Lake Malawi: a model for rapid speciation and adaptive radiation. *Fish and Fisheries*, **6**, 1–34.

Gompert Z, Fordyce JA, Forister ML, Shapiro AM, Nice CC (2006) Homoploid hybrid speciation in an extreme habitat. *Science*, **314**, 1923–1925.

Gow JL, Peichel CL, Taylor EB (2006) Contrasting hybridization rates between sympatric three-spined sticklebacks highlights the fragility of reproductive barriers between evolutionarily young species. *Molecular Ecology*, **15**, 739–752.

Grant PR, Grant BR, Petren K (2005) Hybridization in the recent past. *American Naturalist*, **166**, 56–67.

Hert E (1992) Homing and home-site fidelity in rock-dwelling cichlids of Lake Malawi, Africa. *Environmental Biology of Fishes*, **33**, 229–237.

Hulsey CD, Garcia de Leon FJ, Rodiles-Hernandez R (2006) Micro- and macroevolutionary decoupling of cichlid jaws: a test of Liem's Key Innovation Hypothesis. *Evolution*, **60**, 2096–2109.

Hulsey CD, Mims MC, Streelman JT (2007) Do constructional constraints influence cichlid craniofacial diversification? *Proceedings of the Royal Society of London Series B Biological Sciences*, **274**, 1867–1875.

Kidd MR, Kidd CE, Kocher TD (2006) Axes of differentiation in the bower-building cichlids of Lake Malawi. *Molecular Ecology*, **15**, 459–478.

Kocher TD, Conroy JA, McKaye KR, Stauffer JR, Lockwood SF (1995) Evolution of NADH dehydrogenase subunit 2 in east African cichlid fish. *Molecular Phylogenetics and Evolution*, **4**, 420–432.

Kocher TD (2004) Adaptive evolution and explosive speciation: the cichlid fish model. *Nature Reviews Genetics*, **5**, 288–298.

Kornfield I, Smith PF (2000) African cichlid fishes: Model systems for evolutionary biology. *Annual Review of Ecology and Systematics*, **31**, 163–196.

Lewontin RC, Birch LC (1966) Hybridization as a source of variation for adaptation to a new environment. *Evolution*, **20**, 315–336.

Loh YHE, Katz LS, Mims MC, Kocher TD, Yi SV, Streelman JT (2008) Comparative analysis reveals signatures of differentiation amid genomic polymorphism in Lake Malawi cichlids. *Genome Biology*, **9**, R113.

Maddison DR, Maddison WP (2000) *MacClade 4.0*. Sinauer, Sunderland, MA.

McElroy DM, Kornfield I (1993) Novel jaw morphology in hybrids between Pseudotropheus zebra and Labeotropheus feulleborni (Teleostei, Cichlidae) from Lake Malawi, Africa. *Copeia*, **1993**, 933–945.

Moran P, Kornfield I (1993) Retention of an ancestral polymorphism in the mbuna species flock (Teleostei, Cichlidae) of Lake Malawi. *Molecular Biology and Evolution*, **10**, 1015–1029.

Munthali SM, Ribbink AJ (1998) Condition and fecundity of translocated rock-dwelling cichlid fish in Lake Malawi. *Journal of Zoology (London)*, **244**, 347–355.

Nolte AW, Freyhof J, Stemshorn KC, Tautz D (2005) An invasive lineage of sculpins, *Cottus* sp. (Pisces, Teleostei) in the Rhine with new habitat adaptations has originated from hybridization between old phylogeographic groups. *Proceedings of the Royal Society of London Series B Biological Sciences*, **272**, 2379–2387.

Nosil P, Funk DJ, Ortiz-Barrientos D (2009) Divergent selection and heterogeneous genomic divergence. *Molecular Ecology*, **18**, 375–402.

Owen RB, Crossley R, Johnson TC, Tweddle D, Kornfield I, Davison S, Eccles DH, Engstrom D (1990) Major low levels of Lake Malawi Africa and their implications for speciation rates in cichlid fishes. *Proceedings of the Royal Society of London Series B Biological Sciences*, **240**, 519–553.

Parnell NF, Hulsey CD, Streelman JT (2008) Hybridization produces novelty when the mapping of form to function is many to one. *BMC Evolutionary Biology*, **8**, 122.

Powell JR (1983) Interspecific cytoplasmic gene flow in the absence of nuclear gene flow—evidence from *Drosophila*. *Proceedings of the National Academy of Sciences of the United States of America*, **80**, 492–495.

Panganiban G, Rubenstein JLR (2002) Developmental functions of the Distal–less/Dlx homeobox genes. *Development*, **129**, 4371–4386.

Pritchard JK, Stephens M, Donnelly P (2000) Inference of Population Structure Using Multilocus Genotype Data. *Genetics*, **155**, 945–959.

Reinthal PN (1990) The feeding habits of a group of herbivourous rock–dwelling cichlid fishes cichlidae

perciformes from Lake Malawi Africa. *Environmental Biology of Fishes*, **27**, 215–233.

Ribbink AJ, Marsh BA, Marsh AC, Ribbink AC, Sharp BJ (1983) A preliminary survey of the cichlid fishes of rocky habitats in Lake Malawi Africa. *South African Journal of Zoology*, **188**, 149–310.

Salzburger W (2009) The interaction of sexually and naturally selected traits in the adaptive radiation of cichlid fishes. *Molecular Ecology*, **18**, 169–185.

Schluter D, Conte GL (2009) Genetics and ecological speciation. *Proceedings of the National Academy of Sciences of the United States of America*, **106**(Suppl 1), 9955–9962.

Seehausen O (2004) Hybridization and adaptive radiation. *Trends in Ecology & Evolution*, **19**, 198–207.

Shaw PW, Turner GF, Idid MR, Robinson RL, Carvalho GR (2000) Genetic population structure indicates sympatric speciation of Lake Malawi pelagic cichlids. *Proceedings of the Royal Society Biological Sciences Series B*, **267**, 2273–2280.

Smith PF, Konings A, Kornfield I (2003) Hybrid origin of a cichlid population in Lake Malawi: Implications for genetic variation and species diversity. *Molecular Ecology*, **12**, 2497–2504.

Stauffer Jr JR, Bowers NJ, Kocher TD, McKaye KR (1996) Evidence of hybridization between *Cynotilapia afra* and *Pseudotropheus zebra* (Teleostei: Cichlidae) following an intralacustrine translocation in Lake Malawi. *Copeia*, **1996**, 203–208.

Stelkens R, Seehausen O (2009) Genetic distance between species predicts novel trait expression in their hybrids. *Evolution*, **63**, 884–897.

Streelman JT, Albertson RC, Kocher TD (2007a) Variation in body size and trophic morphology within and among genetically differentiated populations of the cichlid fish, *Metriaclima zebra*, from Lake Malawi. *Freshwater Biology*, **52**, 525–538.

Streelman JT, Peichel CL, Parichy DM (2007b) Developmental genetics of adaptation in fishes: the case for novelty. *Annual Review of Ecology, Evolution and Systematics*, **38**, 655–681.

Streelman JT, Gmyrek SL, Kidd MR, Kidd C, Robinson RL, Hert E, Ambali AJ, Kocher TD (2004) Hybridization and contemporary evolution in an introduced cichlid fish from Lake Malawi National Park. *Molecular Ecology*, **13**, 2471–2479.

Thompson JD, Plewniak F, Poch O (1999) A comprehensive comparison of multiple sequence alignment programs. *Nucleic Acids Research*, **27**, 2682–2690.

Via S (2009) Natural selection in action during speciation. *Proceedings of the National Academy of Sciences of the United States of America*, **106**(Suppl 1), 9939–9946.

Via S, West J (2008) The genetic mosaic suggests a new role for hitchhiking in ecological speciation. *Molecular Ecology*, **17**, 4334–4345.

Won YJ, Sivasundar A, Wang Y, Hey J (2005) On the origin of Lake Malawi cichlid species: A population genetic analysis of divergence. *Proceedings of the National Academy of Sciences of the United States of America*, **102**, 6581–6586.

Won YJ, Wang Y, Sivasundar A, Raincrow J, Hey J (2006) Nuclear gene variation and molecular dating of the cichlid species flock of Lake Malawi. *Molecular Biology and Evolution*, **23**, 828–837.

M.M. is interested in relationships between environmental disturbance, population connectivity, and the evolution of life history strategies in freshwater ecosystems. D.H. studies adaptation and the evolution of complex phenotypes. B.F.'s interests include the genetics and biogeography of speciation, hybridization, and invasion biology. J.T.S.'s present focus is population structure of *Yeti* species.
