MOLECULAR ECOLOGY

Molecular Ecology (2015) **24**, 2686–2701 doi: 10.1111/mec.13205

A genetic demographic analysis of Lake Malawi rock-dwelling cichlids using spatio-temporal sampling

MARTIN HUSEMANN,*† RACHEL NGUYEN,* BAOQING DING* and PATRICK D. DANLEY*

Biology Department, Baylor University, One Bear Place #97388, Waco TX 76798, USA, †Department of General Zoology, Institute of Biology, Martin-Luther University Halle-Wittenberg, Hoher Weg 8, Halle (Saale) D-06120, Germany

Abstract

We estimated the effective population sizes (N_e) and tested for short-term temporal demographic stability of populations of two Lake Malawi cichlids: *Maylandia benetos*, a micro-endemic, and *Maylandia zebra*, a widespread species found across the lake. We sampled a total of 351 individuals, genotyped them at 13 microsatellite loci and sequenced their mitochondrial D-loop to estimate genetic diversity, population structure, demographic history and effective population sizes. At the microsatellite loci, genetic diversity was high in all populations. Yet, genetic diversity was relatively low for the sequence data. Microsatellites yielded mean N_e estimates of 481 individuals (±99 SD) for *M. benetos* and between 597 (±106.3 SD) and 1524 (±483.9 SD) individuals for local populations of *M. zebra*. The microsatellite data indicated no deviations from mutation–drift equilibrium. *Maylandia zebra* was further found to be in migration–drift equilibrium. Temporal fluctuations in allele frequencies were limited across the sampling period for both species. Bayesian Skyline analyses suggested a recent expansion of *M. zebra* populations in line with lake-level fluctuations, whereas the demographic history of *M. benetos* could only be estimated for the very recent past. Divergence time estimates placed the origin of *M. benetos* within the last 100 ka after the refilling of the lake and suggested that it split off the sympatric *M. zebra* population. Overall, our data indicate that micro-endemics and populations in less favourable habitats have smaller N_e, indicating that drift may play an important role driving their divergence. Yet, despite small population sizes, high genetic variation can be maintained.

Keywords: Bayesian Skyline plots, biodiversity conservation, demographic history, drift, East Africa, effective population size, population genetics

Received 17 December 2013; revision received 8 April 2015; accepted 13 April 2015

Introduction

The effective population size (N_e) determines how vulnerable a population is to stochastic forces (Fisher 1930; Frankham 1995; Vucetich *et al.* 1997; Palstra & Ruzzante 2008; Charlesworth 2009), influences the rate of evolution (Lanfear *et al.* 2014) and as such is one of the most important variables in evolutionary biology, and population and conservation genetics (Luikart *et al.* 2010). Therefore, estimating N_e in natural populations is a first step in understanding the forces leading to population differentiation (Charlesworth 2009) and can provide valuable information on the conservation status of a population or species (Luikart *et al.* 2010). For example, N_e can be used to estimate the viability of a population: effective population sizes of 500–5000 are considered necessary for species to maintain their evolutionary potential (Lynch & Lande 1997; Franklin & Frankham 1998; Traill *et al.* 2007; Jamieson & Allendorf 2012; but see Frankham *et al.* 2014 for additional thoughts on these numbers). Populations with a N_e smaller than 500 individuals are more prone to the effects of genetic drift and hence may loose genetic variation more rapidly (Lande & Barrowclough 1987; Newman & Pilsen 1997). The relative importance of drift in the process of speciation

Correspondence: Martin Husemann, Fax: +49 (0)345 55 27428; E-mail: martinhusemann@yahoo.de

remains debated, partly because of a lack of empirical studies (e.g. Orr & Smith 1998). However, theoretical studies have suggested that drift alone may be sufficient to cause speciation under specific circumstances (Uyeda *et al.* 2009), but may take very long if the population size is not very small and the adaptive valley is not shallow (Gavrilets 2003).

Therefore, estimating N_e is not only essential for making informed species management and conservation decisions (Waples 1990, 2002; Ellstrand & Elam 1993; Nunney & Elam 1994), but can also help to understand the potential contribution of drift to speciation (Orr & Smith 1998). However, estimating N_e in natural populations is not trivial as many approaches require genetic samples from populations taken at multiple points in time (Pollak 1983; Waples 1989; Jorde & Ryman 1995; William & Slatkin 1999; Wang & Whitlock 2003; Barker 2011). Such samples can be rare especially in organisms where tissues are not specifically stored for future genetic analysis (Wandeler *et al.* 2007; Nielsen & Hansen 2008; Habel *et al.* 2014). Despite these problems, generating temporal genetic data can be rewarding and help to answer long-posed evolutionary and conservation questions. One such question is the vulnerability of micro-endemic species to stochastic events and the relative impact of drift in their diversification.

The East African cichlids are the most diverse extant vertebrate radiation known. More than 2000 species, most of which are micro-endemics, have evolved in the three East African Great Lakes: Lakes Victoria, Tanganyika and Malawi (Meyer 1993; Seehausen 2000, 2006; Takahashi *et al.* 2001; Kocher 2004; Salzburger & Meyer 2004; Sturmbauer *et al.* 2011; Danley *et al.* 2012). While Lake Tanganyika's cichlids invaded the lake and diversified over the past 6–12 million years, the cichlid species flocks in Lakes Malawi and Victoria have evolved within the 1–2 million years (Salzburger *et al.* 2005). The diversification of the cichlid fauna in the lakes has been strongly influenced by the geographic and climatic history of the region (Genner *et al.* 2010; Aguilée *et al.* 2011, 2013; Danley *et al.* 2012), as well as the shore and basin morphology (e.g. Sturmbauer *et al.* 2001; Genner *et al.* 2010). The patchy distribution of habitat types together with the low dispersal rate of most species has led to high levels of microendemism in the rock-dwelling cichlids (e.g. Markert *et al.* 1999; Allender *et al.* 2003). Most species are found at single, often small, habitat patches in the lake, suggesting that populations may be small and subject to the effects of drift. Therefore, stochastic processes may play a profound role in the evolution of these micro-endemic species and might pose severe threats to their persistence. Few species are found at multiple locations throughout the lake, and previous population genetic studies on these cosmopolitan species have demonstrated that populations of these species are connected by low levels of gene flow (Arnegard *et al.* 1999; Danley *et al.* 2000). Therefore, these populations are expected to have larger effective sizes relative to micro-endemics, unless all populations are already at the carrying capacity of the habitat; further, in widespread species the effects of drift may be mitigated by the influence of migration.

In this study, we sampled the only known population of the micro-endemic species *Maylandia benetos* from Mazinzi Reef and three populations of the widespread species *Maylandia zebra* at three points in time (1996, 2010 and 2012) across 16 years. We genotyped 13 microsatellites and sequenced the mitochondrial D-loop to investigate the spatiotemporal population structures, effective population sizes and demographic histories of the two study species. Specifically, we hypothesize that (i) the Mazinzi Reef population of the widespread species *M. zebra* is larger than that of the sympatric micro-endemic *M. benetos*. We expect these differences in local population sizes because *M. zebra* is expected to receive migrants from neighbouring subpopulations, thereby increasing its effective population size, whereas no such migrant sources are available for *M. benetos*. This, of course, only holds true if the populations are not at their carrying capacity, in which case migration would only help to maintain genetic diversity in the population. (ii) Further, *M. benetos* is suggested to be the evolutionary younger species, as Mazinzi Reef was desiccated in historic time frames, and hence the endemic population may have evolved since recent inundation of Mazinzi Reef. *Maylandia zebra* in turn can have recolonized this location after refilling of the lake. (iii) The potentially younger and smaller population of the micro-endemic *M. benetos* is more strongly influenced by drift than the sympatric population of the widespread *M. zebra*, because drift effects cannot be buffered by migration in the micro-endemic; further, demographic fluctuations may be stronger in the micro-endemic, both in historical time frames and in the short time frame of the study.

Materials and Methods

Study species and locations

The genus *Maylandia* is among the most diverse genera of rock-dwelling cichlids in the lake with 31 described species (Ciccotto *et al.* 2011; Stauffer *et al.* 2013). *Maylandia zebra* is one of the few species that can be found at nearly every rocky habitat throughout the lake. In contrast, most *Maylandia* species are micro-endemics and occur only at a single location in the lake. One of these is *Maylandia benetos* which is endemic to Mazinzi Reef,

where it co-occurs with three other *Maylandia* species, one of which is *M. zebra* (Danley 2011; Husemann *et al.* 2014). Previous studies have indicated that *M. zebra* may not be monophyletic (e.g. Allender *et al.* 2003; Husemann *et al.* 2014). However, in this manuscript, we use the currently accepted taxonomy and treat *M. zebra* and *M. benetos* as separate species (in accordance with premating isolation and lack of natural hybrids (e.g. Ding *et al.* 2014) and treat local occurrences of the barred phenotype as populations of *M. zebra*.

In 1996, 2010 and 2012, we collected *M. zebra* at 3 locations (Mazinzi Reef, Harbour Island and Illala Gap, Fig. 1, Table 1). *Maylandia benetos* was collected at Mazinzi Reef in the same years. Mazinzi Reef (14.1428S, 34.9650E) is a small submerged reef about 10 000 m^2 in size. The sandy habitat around the reef and the relatively large distance to the next stretch of rocky shoreline make it fairly isolated relative to our other sampling locations, Illala Gap (14.0001S, 34.8483E) and Harbour Island (14.0689S, 34.929E). Illala Gap and Harbour Island are well-connected sites at the edge and in the centre of large stretches of rocky habitat (Fig. 1, Danley *et al.* 2000). At Illala Gap, we sampled close to a narrow channel with relatively poor habitat conditions (i.e. large boulders resulting in few caves, high turbidity due to boat traffic and stronger currents at the narrow point). We sampled ~30 adult individuals per population using SCUBA and gill nets (all handled compliant with Baylor IACUC Protocol no. 08-09). Fish were fin-clipped in the field and released on site; fin tissue was either dried, stored in ethanol or preserved in a DMSO storage solution (20% DMSO, 0.25 M EDTA, saturated with NaCl, pH = 7.5) until further processing.

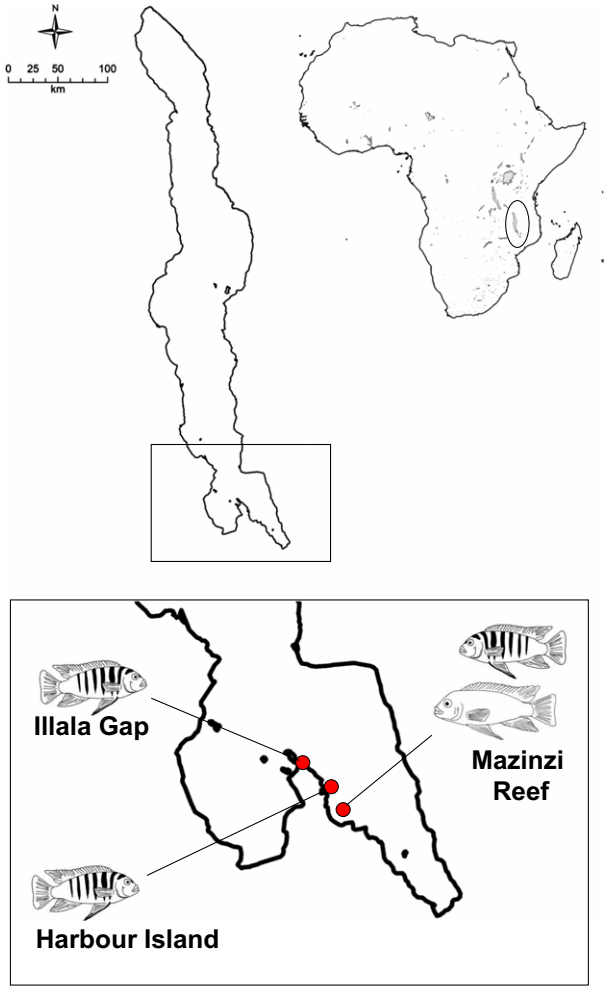

Fig. 1 Sampling map displaying Lake Malawi and showing the locations in the southern part of the lake where populations of the widespread *Maylandia zebra* and the micro-endemic *Maylandia benetos* were sampled in 1996, 2010 and 2012.

Molecular analyses

Microsatellites. DNA was isolated using the Qiagen DNeasy blood and tissue kit (Qiagen, Hilden, Germany) following the manufacturer's protocol for tissue samples. A total of 13 microsatellite loci were genotyped for 351 individuals in this study (Tables 1 and 3, Appendix I). The loci were originally developed for *M. zebra* and had been tested on other Malawi rock-dwelling cichlids (Albertson *et al.* 2003). We chose markers to be on different linkage groups (Albertson *et al.* 2003) to ensure they were physically unlinked. Microsatellites were amplified using standard PCR procedures and labelled using the technique described by Schuelke (2000). We used the fluorescent dyes VIC, 6-FAM and PET for fragment visualization. The master mix for each reaction consisted of 14.7 μL diH_2O, 2.0 μL 10× buffer (Thermo Scientific), 1.6 μL dNTPs (Thermo Scientific, 0.2 μM each), 0.1 μL forward primer + M-13, 0.4 μL reverse primer (0.5 μM, Integrated DNA Technologies, Appendix I), 0.2 μL Taq (1.2 U, DyNAzyme, Thermo Scientific) and 1 μL of template. The following amplification conditions were used: initial denaturation for 5 min at 94 °C, followed by 30 cycles of 30 s at 94 °C, 45 s at 56–61 °C (depending on locus, Appendix I) and 45 s at 72 °C, followed by 8 cycles of 30 s at 94 °C, 45 s at 53 °C and 45 s at 72 °C for M-13 binding, finishing with 10 min elongation at 72 °C. PCR products were visualized on 1% agarose gels stained with gel red (Biotium, Inc. Hayward, CA). Three markers (each 1 μL) with different labels were subsequently pooled and denatured with 7 μL Hi-Di Formamide (Applied Biosystems). Fragment analysis was performed at the Sequencing Facility at Yale University using LIZ-500 as size standard. Genotypes were determined from chromatograms manually using the microsatellite plug-in in GENEIOUS. MICRO-CHECKER (Van Oosterhout *et al.* 2004) was used to test our data for genotyping errors due to stutter bands and null alleles. CREATE v.1.37 (Coombs *et al.*

Table 1 Sampling list providing the sample numbers for each year and each population including the number of individuals sampled for microsatellites, D-loop sequences and the respective GenBank Accession nos

Species	Location	Sampling year	Number of individuals sampled for microsatellites	Number of individuals sampled for D-loop sequences	GenBank Accession nos.
Maylandia zebra	Harbour Island	1996	31	29	KC960349–KC960377
		2010	30	25	KC960277–KC960301
		2012	29	27	KC960250–KC960276
		All	90	81	
	Illala Gap	1996	30	29	KC960320–KC960348
		2010	30	18	KC960302–KC960319
		2012	30	26	KC960172–KC960197
		All	90	73	
	Mazinzi Reef	1996	29	28	KC960407–KC960434
		2010	30	25	KC208879–KC208904
		2012	26	25	KC960225–KC960249
		All	85	78	
	All		265	232	
Maylandia benetos	Mazinzi Reef	1996	30	29	KC960378–KC960406
		2010	29	29	KC208850–KC208878
		2012	27	27	KC960198–KC960224
		All	86	85	
	All		351	317	

2008) was used to generate input files for downstream analyses.

Standard population genetic analyses were performed in ARLEQUIN v.3.5.1.2 (Excoffier & Lischer 2010), GENEPOP on the WEB v.4.2 (Raymond & Rousset 1995) and FSTAT v.2.9.32 (Goudet 1995). We estimated the number of alleles across all loci per population, the mean number of alleles per locus, the observed heterozygosity (H_O), the expected heterozygosity (H_E) and allelic richness (AR). We tested for deviations from Hardy–Weinberg equilibrium (HWE) and linkage disequilibrium (LD) and applied a Bonferroni correction to account for multiple simultaneous tests. Multiple hierarchical analyses of molecular variance (AMOVA, Excoffier *et al.* 1992) were performed to partition the molecular variance. We used three separate set-ups: (i) we used species as groups and locations as populations within groups (temporal samples were pooled) and further analysed within population diversity. The other two AMOVA were performed to partition the variance (ii) between *M. zebra* subpopulations, temporal samples of subpopulations and within these temporal samples and (iii) between temporal samples of *M. benetos* and within them. The genetic divergence among species, local populations and temporal samples was calculated as F_{ST} with ARLEQUIN.

Effective population size. To test for differences in population size between the micro-endemic and the widespread species (hypothesis 1), we estimated last generation N_e for our study populations from microsatellite data using five different approaches. Of the five, one method uses single time points; the approach provides estimates based on gametic disequilibrium between alleles from different neutral loci (Waples & Do 2008); we used the current version of the LD method adjusting for missing data which is implemented as the LD option in N_EESTIMATOR v.2.1 (Do *et al.* 2013). A minimum allele frequency of 0.02 was employed, and 95% confidence intervals were generated using jackknifing. The other four methods use multiple temporal cohorts, but are based on different principles (for all temporal methods, we employed a generation time of 1 year): TM3 uses Bayesian statistics and assumes closed populations (migration is very low between our sampled populations) sampled at two time points close enough together in time (we used the 1996 and 2012 samples) so that mutation has a negligible effect on the observed gene frequencies. MLN_E (Wang & Whitlock 2003) uses a maximum-likelihood approach to estimate drift between temporally spaced populations. TEMPOFS estimates genetic drift between temporally spaced samples using the F_s measure of allele frequency change (Jorde & Ryman 2007). Lastly, we employed the temporal method using moment-based F-statistics implemented in N_EESTIMATOR (Pollak 1983). As all approaches have their advantages and disadvantages (Schwartz *et al.* 1998), we calculated the harmonic means of all point estimates as our final estimate for N_e for each population (Waples 2005; Johnstone *et al.* 2012) with R version 3.1.2 (R Development Team). Standard deviations were derived from the for-

mula: SD = sqrt((mean(1/*x*))^(−4)*var(1/*x*)/length(*x*)), where *x* refers to the array of the five point estimates of N_e for a population. We further tested for pairwise differences between the populations using Welch's two-sample *t*-test.

Equilibrium testing. To test whether populations are in mutation–drift equilibrium, we used microsatellite data from the most recent temporal samples (2012) and the software BOTTLENECK v.1.2.02 (Cornuet & Luikart 1996; Piry *et al.* 1999). As the mutation model is generally not well understood in microsatellites, we tested our data under both the stepwise mutation model (SMM) and the two-phase model (TPM)). The SMM is the simplest model and assumes mutations happen in single repeat steps (Ohta & Kimura 1973). The TPM was developed more recently, allows both single and multistep mutations and assigns different probabilities to each type of mutation (DiRienzo *et al.* 1994). As this model implements stepwise mutations in addition to multistep mutations, it may conform to microsatellite data more appropriately than the other models (Neff *et al.* 1999). When employing TPM, standard settings with a probability of 70 and a variance of 30 (10^4 replications) were used. Wilcoxon's signed-ranks statistics as implemented in BOTTLENECK were used to test whether observed heterozygosity exceeded the expectation under mutation–drift equilibrium.

To test for migration–drift equilibrium, the program 2MOD v.0.2 was used (Beaumont 2000) with our microsatellite data. The program estimates the relative likelihoods of a 'drift alone' model vs. a 'drift–migration' model to identify populations subjected to genetic drift and migration as opposed to those only affected by drift. We performed two separate runs each with 100 000 iterations. The first run included the three populations of *M. zebra*, whereas the second run was performed for *M. benetos*. The first 10% of iterations were discarded as burn-in.

Mitochondrial DNA. We further amplified a 695-bp fragment of the mitochondrial D-loop for 317 specimens (Tables 1 and 2) using the forward primer HapThr-2: 5′ CCTACTCCCAAAGCTAGGATC 3′ and the reverse primer Fish12s: 5′ TGCGGAGACTTGCATGTGTAAG 3′ (Joyce *et al.* 2005). PCR was performed using the following set-up: 12.2 μL of diH_2O, 2 μL of 10× PCR buffer (reaction concentration 1×), 1.6 μL of dNTP mixture (0.2 μM each, Thermo Scientific, Waltham, Massachusetts, USA), 0.2 μL of DyNAzyme™ DNA Polymerase (1.2 U, Thermo Scientific), 1 μL of each primer (0.5 μM, Integrated DNA Technologies, Coralville, Iowa, USA) and 2 μL of DNA template (either pure extract, 1:10, 1:50 dilution) for a total volume of 20 μL. Amplification conditions were as follows: initial denaturation at 94 °C for 3 min, followed by 30 cycles of 94 °C for 1 min denaturation, 58 °C for 1 min annealing and 72 °C for 2 min elongation, with a final elongation step at 72 °C for 10 min. A total of 10 μL PCR product was purified using 4 μL ExoSAP-IT enzyme mix (Affymetrix, Santa Clara, California, USA). The purified products were sequenced at the Sequencing Facility at Yale University. All sequences were deposited in GenBank (Accession nos. are given in Table 1).

We aligned our sequences using GENEIOUS v.6.0.3 (Drummond *et al.* 2011); base calls were checked by

Table 2 Sequence summary statistics for the D-loop (sequence length 695 bp, gaps considered as 5th site); #s – number of sequences, #h – number of haplotypes, ss – segregating sites, hd – haplotype diversity, K – average number of nucleotide differences, π – nucleotide diversity

Species	Location	Sampling year	#s	#h	ss	hd	K	π
Maylandia zebra	Harbour Island	1996	29	9	13	0.840	4.101	0.0059
		2010	25	8	10	0.667	4.127	0.0060
		2012	27	6	8	0.772	3.410	0.0049
		All	81	18	17	0.805	4.299	0.0055
	Illala Gap	1996	29	4	5	0.254	1.335	0.0019
		2010	18	5	7	0.575	1.497	0.0022
		2012	26	4	3	0.452	0.495	0.0007
		All	73	10	10	0.688	1.284	0.0019
	Mazinzi Reef	1996	28	5	6	0.683	1.902	0.0027
		2010	25	4	4	0.637	1.653	0.0024
		2012	25	5	5	0.470	0.907	0.0013
		All	78	9	10	0.624	1.920	0.0023
Maylandia benetos	Mazinzi Reef	1996	29	3	2	0.246	0.305	0.0005
		2010	29	3	3	0.069	0.207	0.0003
		2012	27	3	2	0.330	0.348	0.0005
		All	85	6	6	0.279	0.341	0.0005

eye. General statistics of sequence variation were calculated with DNASP v.5.10 (Librado & Rozas 2009). Estimates of genetic differentiation among species, populations and temporal samples were calculated as Φ_{ST} with ARLEQUIN v. 3.5.1.2 (Excoffier & Lischer 2010) and were tested for significance using 1000 permutations. AMOVA were performed with similar set-ups as for the microsatellites.

We obtained basic estimates of divergence times using a coalescent model and a relaxed exponential clock as implemented in BEAST. This analysis was performed to test the hypotheses of a younger age of the micro-endemic *M. benetos* due to the relatively recent desiccation of the reef (hypothesis 2). For this, we grouped the sequences by populations, yet did not constrain species to be monophyletic (see Allender *et al.* 2003 and Husemann *et al.* 2014 for evidence of paraphyly of *M. zebra*). We added a sequence from *Labeotropheus trewawase* as out-group (AY911790). The HKY + I substitution model was determined as most suitable for the data with JMODELTEST v.2.1.1 (Posada 2008). We used a published substitution rate of 0.0324 changes per site per million years (SE 0.0139) for the D-loop in cichlids which has been independently estimated by Genner *et al.* (2010) and Koblmüller *et al.* (2011). The simplest coalescent model (constant size) was used. The best clock model was estimated by running all four available clock models and using TRACER to determine the best model for our data set. All analyses were run for 100 million generations sampling every 10 000 generations for a total of 10 000 samples. Convergence of the run was confirmed with TRACER by checking the effective sample size (ESS). The exponential relaxed clock had the best likelihood scores (−1371.613, second best: lognormal relaxed clock −1381.808) and was therefore chosen as best clock model. The tree was summarized with TREEANNOTATOR v.1.7.4 (implemented in the BEAST package) and visualized with FIGTREE v.1.3.1 (Rambaut 2009).

To test for demographic stability in both species (hypothesis 3), we estimated the demographic history of each population independently using Bayesian Skyline analysis (Drummond *et al.* 2005). As strict clocks are considered appropriate for intraspecific data (Hein *et al.* 2005), we used the substitution rate of 0.0324 changes per site per million years (SE 0.0139) and the HKY + I substitution model (see above). The temporal samples for each population were pooled due to limited genetic differentiation (see AMOVA results). Each analysis was run for 100 million steps sampling every 10 000 steps under default settings. The output from BEAST v.1.7.4 (Drummond *et al.* 2012) was subsequently analysed in TRACER v.1.5 (Rambaut & Drummond 2009) from which the data were exported into Excel and displayed.

Results

Microsatellites

We amplified 13 microsatellite loci for a total of 351 individuals. MICRO-CHECKER did not find evidence for allele dropout or genotyping error in any sample; yet, the program suggested the presence of null alleles for some markers at single temporal population samples (details in Table 3). No marker, however, consistently showed a null allele problem across multiple populations. Global analyses of LD with GENEPOP suggested that no significant linkage existed for any of the loci

Table 3 Summary statistics of microsatellite data giving the number of alleles, the average number of alleles per locus, the observed and expected heterozygosity and allelic richness. The loci with the possibility of null alleles as revealed by MICRO-CHECKER are also provided

Species	Location	Sampling year	Number of alleles	Average number of alleles	H_O	H_E	Allelic richness*	Possible null alleles
Maylandia zebra	Harbour Island	1996	277	21.3 (SD 4.8)	0.936	0.937	18.212	UNH2135
		2010	299	23.0 (SD 5.4)	0.863	0.936	19.355	UNH2037, UNH362, UNH2204
		2012	275	21.2 (SD 5.9)	0.924	0.934	18.318	—
	Illala Gap	1996	282	21.7 (SD 4.8)	0.909	0.938	18.468	UNH2135, UNH362
		2010	250	19.2 (SD 3.4)	0.897	0.928	16.619	UNH2155, UNH2139
		2012	263	20.2 (SD 4.0)	0.898	0.931	17.518	UNH2112
	Mazinzi Reef	1996	211	16.2 (SD 3.9)	0.909	0.897	14.229	—
		2010	193	14.8 (SD 4.2)	0.863	0.882	13.150	—
		2012	184	14.2 (SD 4.8)	0.897	0.881	12.999	UNH362
Maylandia benetos	Mazinzi Reef	1996	206	15.8 (SD 3.9)	0.895	0.890	13.855	UNH2037, UNH2190
		2010	186	14.3 (SD 3.9)	0.870	0.884	12.899	UNH2166
		2012	193	14.8 (SD 4.4)	0.917	0.881	13.380	—

*Based on minimal sample size of 20 individuals.

consistent with the choice of markers on different chromosomes. HWE was confirmed for most loci across populations and temporal samples after Bonferroni correction. The only exception was UNH2135 for *Maylandia benetos* and *Maylandia zebra* in 1996 at Mazinzi Reef. The deviation from HWE was caused by an excess of homozygotes at UNH2135 for both species at Mazinzi Reef in 1996. Population subdivision is an unlikely reason for this excess in these populations as they exist at a small submerged reef, rejecting Wahlund effects as an explanation. No evidence for null alleles was found by MICRO-CHECKER for the locus, which, however, might be a false negative.

The number of alleles across all loci found within temporal samples of populations across species varied between 184 (*M. zebra* at Mazinzi Reef, 2012) and 299 (*M. zebra* from Harbour Island, 2010) (Table 3). Accordingly, allelic richness was highest for the Harbour Island population varying between 18.212 and 19.355 across temporal samples. In comparison, *M. zebra* and *M. benetos* from Mazinzi had the lowest allelic richness varying among 12.899 to 14.229 (Table 3). Estimates of observed heterozygosity ranged from 0.863 for *M. zebra* from Harbour Island and Mazinzi Reef in 2010 and 0.936 for *M. zebra* from Harbour Island in 1996.

AMOVA detected the majority of genetic variance within populations (94.38%, Table 4). Little variance was found between species (2.89%) and between populations within species (2.73%). However, the power of AMOVA was low in this case as only a single population (i.e. the only population) of *M. benetos* is included. When analysing molecular variance within species separately, in both species the majority of molecular variance was found within temporal samples of populations (*M. zebra*: 96.70%, *M. benetos*: 98.76%) and little variance was found between populations of *M. zebra* (2.36%) or between temporal samples of either species (*M. zebra*: 0.94, *M. benetos*: 1.24). Estimates of genetic divergence (F_{ST}) were similarly high and significant between species and between populations, but low for comparisons of temporal samples within populations (Table 5).

Table 4 Results from AMOVA for mtDNA data and microsatellite data; for each data set, three different AMOVA were performed: using the whole data set testing for among species, among populations within species and within population diversity, and for each species separately, testing for differences among populations (*Maylandia zebra* only) and for variance components within and between temporal samples

Source of variation	d.f.	Sum of squares	Variance component	Percentage variance	*P*-value
Mitochondrial					
Among species (*M. zebra* vs., *Maylandia benetos*)	1	156.350	0.405	15.94	0.496
Among populations within species	2	197.348	1.266	49.76	<0.001
Within populations	313	273.091	0.872	34.30	<0.001
Total	316	626.091	2.544		
M. zebra					
Among populations	2	197.348	1.232	51.13	0.002
Among temporal samples within populations	6	20.756	0.093	3.85	0.004
Within temporal samples	223	45.02	1.085	45.02	<0.001
Total	231	459.991	2.409		
M. benetos					
Among temporal samples	2	0.549	0.004	2.62	0.078
Within temporal samples	82	12.792	0.156	97.38	
Total	84	13.341	0.160		
Microsatellites					
Among species (*M. zebra* vs., *M. benetos*)	1	61.575	0.135	2.89	0.26
Among populations within species	2	53.911	0.128	2.73	<0.001
Within populations	698	3076.367	4.408	94.38	<0.001
Total	701	3191.853	4.670		
M. zebra					
Among populations	2	57.099	0.119	2.36	0.004
Among temporal samples within populations	6	45.686	0.470	0.94	<0.001
Within temporal samples	521	2527.593	4.851	96.70	<0.001
Total	529	2630.377	5.017		
M. benetos					
Among temporal samples	2	13.318	0.049	1.24	<0.001
Within temporal samples	169	655.403	3.878	98.76	
Total	171	668.721	3.927		

Table 5 Φ_{ST} from D-loop (below diagonal) and F_{ST} from microsatellites (above diagonal) (Mb – *Maylandia benetos* from Mazinzi Reef, MR – *Maylandia zebra* from Mazinzi Reef, HI – *M. zebra* from Harbour Island, IG – *M. zebra* from Illala Gap, 12 – sampling date 2012, 10 – sampling date 2010, 96 – sampling date 1996). Light grey shading indicates differentiation between temporal samples within populations; dark grey shading indicates differentiation between species. Bold values indicate significance at $\alpha = 0.05$

	Mb12	Mb10	Mb96	MR12	MR10	MR96	HI12	HI10	HI96	IG12	IG10	IG96
Mb12	0	0.002	0.003	**0.086**	**0.089**	**0.078**	**0.050**	**0.048**	**0.057**	**0.062**	**0.049**	**0.050**
Mb10	0.025	0	0.002	**0.085**	**0.086**	**0.077**	**0.049**	**0.048**	**0.056**	**0.060**	**0.051**	**0.052**
Mb96	0.009	0.040	0	**0.079**	**0.077**	**0.070**	**0.047**	**0.042**	**0.052**	**0.056**	**0.044**	**0.045**
MR12	**0.764**	**0.785**	**0.780**	0	**0.009**	0.000	**0.041**	**0.042**	**0.041**	**0.053**	**0.052**	**0.041**
MR10	**0.538**	**0.561**	**0.558**	**0.087**	0	**0.006**	**0.039**	**0.042**	**0.040**	**0.054**	**0.057**	**0.043**
MR96	**0.414**	**0.433**	**0.433**	**0.177**	−0.015	0	**0.033**	**0.036**	**0.035**	**0.048**	**0.050**	**0.037**
HI12	**0.556**	**0.576**	**0.567**	**0.427**	**0.344**	**0.348**	0	**0.006**	**0.003**	**0.016**	**0.026**	**0.010**
HI10	**0.552**	**0.576**	**0.560**	**0.487**	**0.406**	**0.397**	**0.110**	0	**0.005**	**0.011**	**0.021**	**0.007**
HI96	**0.541**	**0.560**	**0.552**	**0.399**	**0.324**	**0.333**	−0.024	0.107	0	**0.014**	**0.024**	**0.009**
IG12	**0.909**	**0.923**	**0.916**	**0.825**	**0.768**	**0.754**	**0.572**	**0.335**	**0.539**	0	**0.024**	**0.011**
IG10	**0.844**	**0.864**	**0.854**	**0.770**	**0.698**	**0.679**	**0.489**	**0.208**	**0.465**	**0.254**	0	**0.017**
IG96	**0.827**	**0.845**	**0.835**	**0.751**	**0.686**	**0.669**	**0.488**	**0.194**	**0.461**	**0.130**	0.061	0

Effective population size

The estimates of N_e varied between populations and methods (Table 6), but were in general agreement with the findings of the Skyline analysis (see below, Fig. 2). *Maylandia zebra* from Harbour Island consistently had the largest estimates of effective population size (Table 6, except MLN$_E$). In contrast, the micro-endemic *M. benetos* from Mazinzi Reef had the smallest effective population size for most methods (except TM3). The estimates for *M. zebra* from the Illala Gap were also relatively small and only slightly exceeded the population size estimates for *M. benetos*. The *M. zebra* population from Mazinzi Reef had relatively large population size estimates, which were intermediate between those of the Harbour Island and Illala Gap populations. The harmonic means across all temporal methods were highest for *M. zebra* from Harbour Island (N_e = 1524, ± 483.9 SD), followed by the Mazinzi Reef population of the same species (1103 ± 140), and the population from Illala Gap (597 ± 106.3). The micro-endemic *M. benetos* from Mazinzi Reef had the smallest mean N_e estimate (481 ± 99). The pairwise *t*-tests revealed a significant difference in the effective population sizes of the *M. zebra* population from Mazinzi and the *M. zebra* population from the Illala Gap (P = 0.03). Further, N_e from *M. zebra* from Mazinzi significantly differed from the sympatric population of *M. benetos* (P = 0.02). All other comparisons yielded nonsignificant results ($P > 0.05$).

Equilibrium testing

Our two-tailed Wilcoxon tests for mutation–drift equilibrium revealed significant deviations from mutation–drift equilibrium for *M. zebra* from Mazinzi under the SMM ($P < 0.001$), but not under the TPM (P = 1.000). The Illala Gap population tests were not significant for either model (TPM P = 0.340, SMM P = 0.455). The Harbour Island population tests were significant under the SMM (P = 0.048), but not TPM (P = 0.068). Tests of mutation–drift equilibrium for *M. benetos* suggested deviations from the equilibrium only for the SMM (P = 0.002), but not for the TPM (P = 0.542). As expected for populations that did not experience a bottleneck, all populations showed L-shaped allele frequencies. The test for migration–drift equilibrium with 2MOD indicated that the *M. zebra* populations are in migration–drift equilibrium (100% support), whereas the signal was less clear for *M. benetos*, but suggested that drift explained the data better than the equilibrium model (57.7%).

Mitochondrial DNA

A total of 695 bp were sequenced for 317 individuals (18–29 per population and time point, Table 2). From these 317 individuals, 30 haplotypes were recovered with 25 segregating sites. The haplotype diversity for the whole data set was 0.82. The largest number of haplotypes found in any population and temporal cohort was 9 for *M. zebra* from Harbour Island sampled in 1996. *Maylandia zebra* at Harbour Island was generally the most diverse population and had the highest number of haplotypes in every sampling year (6–9 haplotypes) and a total of 14 different haplotypes across years. *Maylandia benetos* from Mazinzi had a lower diversity than any *M. zebra* population with three haplotypes found in each sampling year and a total number of five haplotypes across all years. The populations of

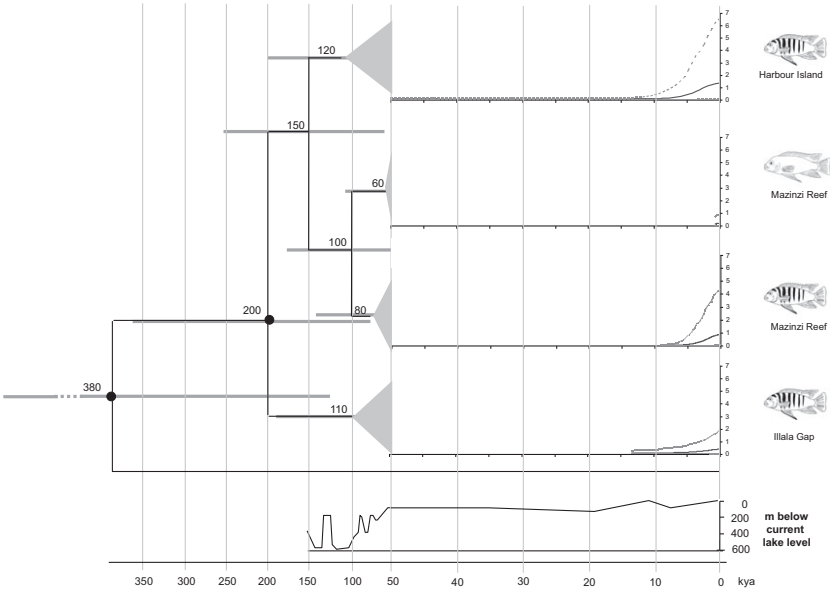

Fig. 2 Chronogram estimating divergence times between the studied populations. Divergence times are given in ka. The grey thick bars indicate 95% HPD intervals. For each population and species, the demographic history derived from Bayesian Skyline analyses is displayed for the last 50 ka. The *y*-axis represents a product of female effective population size (N_e) and generation time (τ, in millions of years); the axis was scaled to a relative size of seven for all plots for comparability; the means (solid dark line) and the upper and lower 95% CI (grey dashed lines) are displayed for all populations. At the bottom of the graph, the lake level according to Scholz *et al.* (2007) is depicted until 150 ka before present.

M. zebra from Mazinzi Reef and the Illala Gap were intermediate with 4–5 haplotypes in each year and a total of 9 and 10 haplotypes across years, respectively.

AMOVA across species showed that most of the variation is found among populations within species (49.76%, Table 4). Most of the remaining variance was found within populations (34.30%), whereas the differentiation between species was rather low and not significant (15.94%, $P = 0.496$). However, the power of AMOVA was low for the same reasons as for the microsatellites. When performing AMOVA within species, 51.13% of the variation was attributed to differences between populations. The amount of variation explaining temporal differences was low in both species (*M. zebra*: 3.85%, *M. benetos*: 2.62), whereas the remaining variance was attributed to variation within temporal samples (*M. zebra*: 45.02%, *M. benetos*: 97.38%). Φ_{ST} estimates suggested a strong population differentiation similar to levels of divergence between the species. Φ_{ST} estimates between temporal samples within populations were low and mostly nonsignificant (Table 5).

The molecular clock analyses (Fig. 2) suggested a divergence between *Maylandia* and *Labeotropeus* (outgroup) at ~380 ka. The divergence between the *M. zebra* population from the Illala Gap and all other locations was dated at about 200 ka. The *M. zebra* population from Harbour Island branched off at around 150 ka. *Maylandia benetos* and *M. zebra* from Mazinzi Reef had a sister group relationship with their divergence time being estimated at ~100 ka, corresponding well with the refilling of the lake after the major low stand at around 120 ka (Fig. 2).

The Bayesian Skyline analyses suggested subtle differences in the demographic histories of the populations (Fig. 2). The *M. zebra* population from Harbour Island started to expand about 6 kya and has the largest size. Similarly, the *M. zebra* population from Mazinzi Reef expanded during that time, yet is slightly smaller. The expansion of the Illala Gap population started more recently (4 kya); the population of *M. benetos* from Mazinzi Reef could not be traced back far in time, but is the smallest and appears to be slightly expanding. The relative sizes of the populations are similar to the N_e estimation from microsatellite data (see below) with *M. zebra* from Harbour Island having the largest and *M. benetos* from Mazinzi Reef having the smallest population size.

Discussion

In this study, we examined the spatial and temporal genetic diversity of two East African cichlids with different geographic ranges (widespread vs. micro-endemic) to examine a number of hypotheses related to these species' current and historical demography. We used microsatellite data from temporally sampled populations and found that mean estimates of N_e ranged between 597 and 1524 individuals for local populations of the widespread *Maylandia zebra* and 481 individuals for the micro-endemic *Maylandia benetos* (Table 6). The comparison of harmonic means yielded significant differences between the sympatric populations of *M. zebra* and *M. benetos* at Mazinzi Reef supporting our first hypothesis that the micro-endemic has a smaller effective population size. The differences in size of local populations estimated from microsatellite data were also reflected in the mitochondrial data (Fig. 2). The splitting of *M. zebra* populations was dated between 200

Table 6 Effective population sizes based on microsatellite data for all studied populations estimated using different approaches. For single time point methods, the most recent year (2012) was used, given are N_e with 95% CI. For the LD method, a minimum allele frequency of 0.02 was used. The pairwise *t*-tests showed significant differences in the mean N_e of the *Maylandia zebra* populations from Mazinzi Reef and the Illala Gap ($P = 0.03$), and for *M. zebra* and *Maylandia benetos* from Mazinzi Reef ($P = 0.02$). All other comparisons were nonsignificant ($P > 0.05$)

Population	N_e (LD, N_EESTIMATOR)	N_e (Moments based, N_EESTIMATOR)	N_e (TM3)	N_e (TEMPOFS)	N_e (MLN_E)	Average of all temporal methods
Sampling	Single time point, most recent	Multiple time points	Multiple time points	Multiple time points	Multiple time points	
Method	Linkage disequilibrium	Moments based	Bayesian approach	Genetic drift among samples	Maximum likelihood	Harmonic means (±SD)
Illala *M. zebra*	834 (244–∞)	631.4 (386.3–1276.7)	795.3 (372.0–1255.7)	353 (196–1617)	665.0 (462.3–1087.3)	597 (±106.3)
Harbour *M. zebra*	6371 (305–∞)	1966.2 (749.8–∞)	2966.8 (0–100 000)	981 (442–∞)	793.8 (524.3–1452.9)	1524 (±483.9)
Mazinzi *M. zebra*	801 (161–∞)	1093.2 (481.6–34319.8)	1794.0 (0–31787.9)	965 (346–∞)	1268.3 (661.7–6511.2)	1103 (±140)
Mazinzi *M. benetos*	392 (137–∞)	554 (324–1217.8)	996.7 (420–1936.7)	289 (111–∞)	638.9 (422.7–1147.1)	481 (±99)

and 150 ka during times of major lake-level fluctuations. The split between the Mazinzi Reef *M. zebra* and *M. benetos* occurred most recently and was dated at 100 ka, supporting our second hypothesis: *M. benetos* is the younger of the two species. The demographic histories of all *M. zebra* populations are characterized by recent expansions since the last low lake-level stands (*c.* 8 kya) and have been found to be in migration–drift equilibrium. The *M. benetos* population could not be traced back far in time, yet seems to have slightly expanded. However, no bottleneck or strong demographic fluctuations could be detected for either species, neither on historical, nor at ecological timescales. Therefore, we did not find support for our third hypothesis of stronger demographic fluctuations in the micro-endemic; yet, 2mod analyses suggested that drift may influence *M. benetos* more strongly than *M. zebra*. Diversity estimates, AMOVA, F_{ST} and Φ_{ST} estimates (Tables 4 and 5) all revealed little temporal genetic changes over the studied 16 years, suggesting that the studied time frame is too short for strong fluctuations in allele frequencies, even in relatively small populations. In the following, we discuss these findings in detail and suggest implications for cichlid evolution and conservation.

Local population sizes of the widespread M. zebra and the micro-endemic M. benetos

Many studies have investigated the population structure and demographic histories of Lake Malawi cichlids (e.g. Markert *et al.* 1999; Danley *et al.* 2000; Smith & Kornfield 2002; Won *et al.* 2005; Genner *et al.* 2010). However, very little is known about the sizes of local populations with only a single study explicitly estimating N_e: Won *et al.* (2005) calculated effective and ancestral population sizes from composite genetic loci in species of the genus *Tropheops*; estimates ranged from 1500 to 47 800 (Won *et al.* 2005). We hypothesized that local effective population sizes of micro-endemics are smaller than those of more widespread species (hypothesis 1). Our estimates of N_e for the widespread *Maylandia zebra* ranged from 597 (±106.3) for the Illala Gap population to 1524 (±483.9) for the Harbour Island population. The N_e of the micro-endemic *M. benetos* from Mazinzi Reef was significantly smaller with only 481 individuals (±99 SD). The differences in the N_e of the Mazinzi Reef populations are also reflected in the census sizes of these populations estimated from transect data (*M. benetos* – 5340 individuals, *M. zebra* – 8800 individuals, P. D. Danley, unpublished data), translating to N_e/N_c ratios of 0.09 and 0.13 for *M. benetos* and *M. zebra*, respectively. The ratios fall well within the range of estimates for vertebrate taxa (Frankham 1995; Palstra & Ruzzante 2008). The differences in local population sizes are similarly reflected in the mitochondrial data, which showed the same relative contemporary population sizes (Fig. 2). The observed differences in N_e between the micro-endemic and the widespread species may have several nonexclusive reasons: (i) inward migration from other close populations can replenish the population of *M. zebra* at Mazinzi Reef; yet, no such source populations are available for *M. benetos* (e.g. Danley 2011). This of course plays only a role, if the populations are not at the carrying capacity of the

habitat. However, in cases of sudden population declines (e.g. due to fishing), migration may play an important balancing role. (ii) The territory size differs between the two species, with *M. benetos* requiring larger territories, leading to smaller local populations (Genner *et al.* 1999; Maruyama *et al.* 2010). This seems unlikely as previous studies have suggested that *M. zebra* actually has larger territories (Danley 2011). (iii) The N_e may be lower in *M. benetos* as mate choice is stronger and fewer individuals reproduce (Danley 2011). While this may be a possible explanation, it seems unlikely, as the census size for *M. benetos* is also lower and the N_e/N_c ratios of both populations are rather similar. Finally, (iv) the fishing pressure for aquarium trade may be higher for the micro-endemic, as there is only a single source population, whereas *M. zebra* can be harvested at many locations in the lake. While we cannot convincingly decide which of the possible explanations for different N_e applies, it seems likely that a combination of stronger fishing pressure for pet trade and lack of other populations as source for inward migration results in lower N_e and N_c for *M. benetos*. These factors probably also apply to other micro-endemics and raise concerns about their conservation status.

The population sizes, especially that of *M. benetos*, are relatively small and are at the lower bound of what is considered necessary to maintain the evolutionary potential of populations (Lynch & Lande 1997; Franklin & Frankham 1998; Traill *et al.* 2007). However, as especially for micro-endemics, small effective population sizes represent the natural situation, it may not represent a problem (Habel & Schmitt 2012), unless populations experience unnatural disturbances. This is reflected in the relatively high genetic diversity found in the micro-endemic. Yet, in the presence of strong anthropogenic pressure, that is overharvesting for food or aquarium trade, pollution and increasing boating, small population sizes may become a problem in the future. Hence, conservation plans may be needed. One of the few specific conservation actions suggested for cichlids is the creation of microscale protected areas (Sturmbauer 2008). Our study supports this idea and suggests that specifically micro-endemics require special attention.

In addition to the differences in effective population size between the widespread and the micro-endemic species, we observed differences in local population sizes of *M. zebra*. The population from the Illala Gap is about half the size of the Mazinzi Reef or Harbour Island populations (Table 6). This might be the result of an edge effect, as this population is located at the end of a longer stretch of continuous suitable habitat. It is well established that populations at the edge of a distribution often have smaller population sizes and suffer from reduced genetic diversity as a result of the loss of alleles from the centre of the distribution (Eckert *et al.* 2008; Ray *et al.* 2015). Alternatively, the habitat at Illala Gap might be suboptimal for the species and therefore has a lower capacity. At this location, large boulders and slabs of bedrock are prominent and disturbance due to boat traffic and stronger currents are common, whereas at locations with higher N_e estimates (Harbour Island, Mazinzi Reef), the preferred cobble-rich habitat dominates (P. D. Danley, unpublished data) and disturbance may be rarer. Probably, both factors play a role for the smaller size of the Illala Gap population.

It, however, is important to note that the ranges of point estimates of N_e are relatively large and overlap between all populations. The lack of precision is partially a problem of low sample size. However, in order to generate sample sizes large enough to significantly reduce confidence intervals, very high numbers of samples, on the order of several hundreds to thousands, are necessary (Palstra *et al.* 2009), which are difficult to assemble in a study system like ours. Another problem potentially biasing N_e estimates is migration. While this generally can be excluded as a factor for *M. benetos*, migration may play a role for the local N_e estimates of *M. zebra*. The temporal estimates may be slightly downward-biased in *M. zebra* due to the potential of migration from other locations (Ryman *et al.* 2014), whereas the LD estimates should be relatively stable even in the face of low rates of migration (Waples & England 2011). The calculation of harmonic means of several estimates should help to overcome the biases of single methods and provide the best possible estimates of N_e (e.g. Waples 2005; Barker 2011).

Demographic histories reflect lake-level fluctuations

Lake-level fluctuations are often thought to have strongly influenced the evolution of the East African cichlid radiations (Owen *et al.* 1990; Sturmbauer *et al.* 2001; Genner *et al.* 2010). We dated the divergence of the different study populations using a molecular clock approach which revealed divergence times between 100 and 200 ka (Fig. 2), a time of major fluctuations in the level of Lake Malawi (Scholz *et al.* 2007). The divergence of *M. benetos* from the *M. zebra* population from Mazinzi Reef was dated at 100 ka and therefore coincides with the refilling after a major low stand about 120 ka; the analysis further suggests a younger age of the micro-endemic in comparison with the widespread species supporting our second hypothesis. Bayesian Skyline analyses indicated expansions of all populations of *M. zebra* within the last 10 000 years (Fig. 2), suggesting demographic expansions after the refilling of the lake subsequent to a more recent desiccation event

about 8 ka, which emptied the southern arms of the lake, including all study locations (Cohen *et al.* 2007; Lyons *et al.* 2011; Scholz *et al.* 2011). However, no population bottlenecks were detected for any population based on the analysis of microsatellite data, and the *M. zebra* populations are found to be in equilibrium, suggesting that they are currently stable, despite their relatively small sizes. The *M. benetos* population could not be traced far back in time with the Skyline analysis; yet, it appears that the population also has slightly expanded in the last few hundred years (Fig. 2). However, we cannot find any clear differences in demographic stability of *M. benetos* and *M. zebra*. Yet, our analyses showed that *M. benetos* is not in migration–drift equilibrium, suggesting that it is more strongly affected by drift compared to *M. zebra*. These results need to be interpreted with caution as inference of demographic histories using mitochondrial data is known to have several problems associated with them (Ballard & Kreitman 1995; Balloux 2010). For example, the low diversity at the D-loop and the resulting low coalescent times could have resulted from a selective sweep leading to the fixation of few haplotypes in the population (Ballard & Whitlock 2004). While such confounding effects cannot be excluded, the observed patterns correspond well with paleogeographic events and are in line with findings by other authors (e.g. Genner *et al.* 2010).

Lastly, we tested for allele frequency fluctuations within the sampling period. We observed only little temporal signal in the data, and the populations of both species showed temporal stability in the numbers of alleles and in heterozygosity (*t*-test $P > 0.05$). Likewise, measures of population genetic divergence (F_{ST} and Φ_{ST}) between time points were small and mostly non-significant (Table 5). AMOVA revealed that temporal samples only explained 3.85% of the variance for mitochondrial data in *M. zebra* and 2.62% in *M. benetos* and 0.94% for the microsatellite data in *M. zebra* and 1.24% in *M. benetos* (Table 4). This indicates that drift is not strong enough to cause significant population fluctuations over the ecological timescales sampled in this study, even in micro-endemic species with relatively low effective population sizes and lack of inward migration.

Conclusion

The sympatric populations of the micro-endemic *M. benetos* and the widespread *M. zebra* from Mazinzi Reef differed significantly in their effective population sizes, possibly as a result of differences in anthropogenic fishing pressure and the lack of inward migration for the micro-endemic. Further, local populations of *M. zebra* significantly differed in N_e, probably as a result of connectivity and habitat quality. The evolution and past demography of all populations were strongly influenced by lake-level fluctuations. We did not detect any bottlenecks, but the data indicated that drift may play an important role for the micro-endemic. Yet, across the 16-year time period, allele frequency fluctuations were negligible, suggesting that the time needed for drift to significantly affect populations, even of micro-endemics, exceeds the studied time frame. Our results demonstrate that micro-endemics may be influenced by drift more strongly than more widespread species, but can maintain high genetic diversity, if the populations are relatively undisturbed. Therefore, micro-endemics may require specific conservation strategies involving the establishment of microscale protected areas to protect them against anthropogenic stochastic events that may threaten their persistence.

Acknowledgement

We want to thank R. Zatha, A. Howe, J. Curole, T. Kocher, J. Markert, M. Arnegard and the officers of Lake Malawi National Park, members of Malawi Fisheries Department and faculty at the University of Malawi for field assistance. We are grateful to A. Tellier, C. Sturmbauer, J. C. Habel and three anonymous reviewers for comments on previous versions of the manuscript. J. Wang, C. Vangestel, M. Slotman and G. Athrey provided useful suggestions for the analyses. The study was funded by a Jordan Grant from the American Cichlid Association to MH, Folmar Grants to MH & BD, the C. Gus Glasscock, Jr. Endowed Fund for Excellence in Environmental Sciences to MH and funds to PDD from the Faculty Research Investment Program and the Vice Provost for Research at Baylor University.

References

Aguilée R, Lambert A, Claessen D (2011) Ecological speciation in dynamic landscapes. *Journal of Evolutionary Biology*, **24**, 2663–2677.

Aguilée R, Claessen D, Lambert A (2013) Adaptive radiation driven by the interplay of eco-evolutionary and landscape dynamics. *Evolution*, **67**, 1291–1306.

Albertson RC, Streelman JT, Kocher TD (2003) Directional selection has shaped the oral jaws of Lake Malawi cichlid fishes. *Proceedings of the National Academy of Sciences USA*, **100**, 5252–5257.

Allender CJ, Seehausen O, Knight ME, Turner GF, Madean N (2003) Divergent selection during speciation of Lake Malawi cichlid fishes inferred from parallel radiations in nuptial coloration. *Proceedings of the National Academy of Sciences USA*, **100**, 14074–14079.

Arnegard ME, Markert JA, Danley PD, Stauffer JR 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 B*, **266**, 119–130.

Ballard JWO, Kreitman M (1995) Is mitochondrial DNA a strictly neutral marker? *Trends in Ecology and Evolution*, **10**, 485–488.

Ballard JWO, Whitlock MC (2004) The incomplete natural history of mitochondria. *Molecular Ecology*, **13**, 729–744.

Balloux F (2010) The worm in the fruit of the mitochondrial DNA tree. *Heredity*, **104**, 419–420.

Barker JSF (2011) Effective population size of natural populations of *Drosophila buzzatii*, with a comparative evaluation of nine methods of estimation. *Molecular Ecology*, **20**, 4452–4471.

Beaumont M (2000) Readme_2mod [Documentation file]. Available with the program at http://www.rubic.rdg.ac.uk/~mab/software.html.

Charlesworth B (2009) Effective population size and patterns of molecular evolution and variation. *Nature Reviews Genetics*, **10**, 195–205.

Ciccotto PJ, Konings A, Stauffer JR Jr (2011) Descriptions of five new species in the genus *Metriaclima* (Teleostei: Cichlidae) from Lake Malawi, Africa. *Zootaxa*, **2738**, 1–25.

Cohen AS, Stone JR, Beuning KRM *et al.* (2007) Ecological consequences of early Late Pleistocene megadroughts in tropical Africa. *Proceedings of the National Academy of Sciences USA*, **104**, 16422–16427.

Coombs JA, Letcher BH, Nislow KH (2008) CREATE: a software to create input files from diploid genotypic data for 52 genetic software programs. *Molecular Ecology Resources*, **8**, 578–580.

Cornuet JM, Luikart G (1996) Description and power analysis of two tests for detecting recent population bottlenecks from allele frequency data. *Genetics*, **144**, 2001–2014.

Danley PD (2011) Aggression in closely related Malawi cichlids varies inversely with habitat complexity. *Environmental Biology of Fishes*, **92**, 275–284.

Danley PD, Markert JA, Arnegard ME, Kocher TD (2000) Divergence with gene flow in the rock-dwelling cichlids of Lake Malawi. *Evolution*, **54**, 1725–1737.

Danley PD, Husemann M, Ding B, DiPietro LM, Beverly EJ, Peppe DJ (2012) The impact of the geologic history and paleoclimate on the diversification of East African cichlids. *International Journal of Evolutionary Biology*, **2012**, 574851.

Ding B, Daugherty DW, Husemann M, Chen M, Howe AE, Danley PD (2014) Quantitative genetic analyses of male color pattern and female mate choice in a pair of cichlid fishes of Lake Malawi, East Africa. *PLoS ONE*, **9**, e114798.

DiRienzo A, Peterson AC, Garza JC, Valdes AM, Slatkin M, Freimer NB (1994) Mutational processes of simple-sequence repeat loci in human populations. *Proceedings of the National Academy of Sciences USA*, **91**, 3166–3170.

Do C, Waples RS, Peel D, Macbeth GM, Tillett BJ, Ovenden JR (2013) NeEstimator V2: reimplementation of software for the estimation of contemporary effective population size (*Ne*) from genetic data. *Molecular Ecology Resources*, **1**, 209–214.

Drummond AJ, Rambaut A, Shapiro B, Pybus OG (2005) Bayesian coalescent inference of past population dynamics from molecular sequences. *Molecular Biology and Evolution*, **22**, 1185–1192.

Drummond AJ, Ashton B, Buxton S *et al.* (2011) Geneious v5.4. Available at http://www.geneious.com/.

Drummond AJ, Suchard MA, Xie D, Rambaut A (2012) Bayesian phylogenetics with BEAUti and the BEAST 1.7. *Molecular Biology and Evolution*, **29**, 1969–1973.

Eckert CG, Samis KE, Lougheed SC (2008) Genetic variation across species' geographical ranges: the central–marginal hypothesis and beyond. *Molecular Ecology*, **17**, 1170–1188.

Ellstrand NC, Elam DR (1993) Population genetic consequences of small population size: Implications for plant conservation. *Annual Reviews in Ecology and Systematics*, **24**, 217–242.

Excoffier L, Lischer HE (2010) Arlequin suite ver 3.5: a new series of programs to perform population genetics analyses under Linux and Windows. *Molecular Ecology Resources*, **10**, 564–567.

Excoffier L, Smouse PE, Quattro JM (1992) Analysis of molecular variance inferred from metric distances among DNA haplotypes: application to human mitochondrial DNA restriction data. *Genetics*, **131**, 479–491.

Fisher RA (1930) *The Genetical Theory of Natural Selection*. Clarendon, Oxford.

Frankham R (1995) Effective population-size adult-population size ratios in wildlife – a review. *Genetical Research*, **66**, 95–107.

Frankham R, Bradshaw CJA, Brook BW (2014) Genetics in conservation management: revised recommendations for the 50/500 rules, Red List criteria and population viability analyses. *Biological Conservation*, **170**, 56–63.

Franklin IF, Frankham R (1998) How large must populations be to retain evolutionary potential? *Animal Conservation*, **1**, 69–73.

Gavrilets S (2003) Models of speciation: what have we learned in 40 years? *Evolution*, **57**, 2197–2215.

Genner M, Turner GF, Hawkins SJ (1999) Resource control by territorial male cichlid fish in Lake Malawi. *Journal of Animal Ecology*, **68**, 522–529.

Genner MJ, Knight ME, Haesler MP, Turner GF (2010) Establishment and expansion of Lake Malawi rock fish populations after a dramatic late Pleistocene lake level rise. *Molecular Ecology*, **19**, 170–182.

Goudet J (1995) FSTAT (Version 1.2): a computer program to calculate F-statistics. *Journal of Heredity*, **86**, 485–486.

Habel JC, Schmitt T (2012) The burden of genetic diversity. *Biological Conservation*, **147**, 270–274.

Habel JC, Husemann M, Finger A, Danley PD, Zachos FE (2014) The relevance of time series in molecular ecology and conservation biology. *Biological Reviews*, **89**, 484–492.

Hein J, Schierup MH, Wiuf C (2005) *Gene Genealogies, Variation and Evolution: A Primer in Coalescent Theory*. Oxford University Press, Oxford, UK.

Husemann M, Tobler M, McCauley C, Ding B, Danley PD (2014) The correlation of reproductive and ecological traits may facilitate the coexistence of congeners in Lake Malawi's rock-dwelling cichlids. *Journal of Evolutionary Biology*, **27**, 826–839.

Jamieson IG, Allendorf FW (2012) How does the 50/500 rule apply to MVPs? *Trends in Ecology and Evolution*, **27**, 578–584.

Johnstone DL, O'Connell MF, Palstra FP, Ruzzante DE (2012) Mature male pall contribution to the effective size of an anadromous Atlantic salmon (*Salmo salar*) population over 30 years. *Molecular Ecology*, **22**, 2394–2407.

Jorde PE, Ryman N (1995) Temporal allele frequency change and estimation of effective size in populations with overlapping generations. *Genetics*, **139**, 1077–1090.

Jorde PE, Ryman N (2007) Unbiased estimator for genetic drift and effective population size. *Genetics*, **177**, 927–935.

Joyce DA, Lunt DH, Bills R *et al.* (2005) An extinct cichlid fish radiation emerged in an extinct Pleistocene lake. *Nature*, **435**, 90–95.

Koblmüller S, Salzburger W, Obermüller B, Eigner E, Sturmbauer C, Sefc KM (2011) Separated by sand, fused by dropping water: Habitat barriers and fluctuating water levels steer the evolution of rock-dwelling cichlid populations in Lake Tanganyika. *Molecular Ecology*, **20**, 2272–2290.

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

Lande R, Barrowclough GF (1987) Effective population size, genetic variation, and their use in population management. In: *Viable Populations for Conservation* (ed. Soulé ME), pp. 87–124. Cambridge University Press, Cambridge, UK.

Lanfear R, Kokko H, Eyre-Walker A (2014) Population size and the rate of evolution. *Trends in Ecology & Evolution*, **29**, 33–41.

Librado P, Rozas J (2009) DnaSP v5: a software for comprehensive analysis of DNA polymorphism data. *Bioinformatics*, **25**, 1451–1452.

Luikart G, Ryman N, Tallmon DA, Schwartz MK, Allendorf FW (2010) Estimation of census and effective population sizes: the increasing usefulness of DNA-based approaches. *Conservation Genetics*, **11**, 355–373.

Lynch M, Lande R (1997) The critical effective size for a genetically secure population. *Animal Conservation*, **1**, 70–72.

Lyons RP, Scholz CA, Buoniconti MR, Martin MR (2011) Late Quaternary stratigraphic analysis of the Lake Malawi Rift, East Africa: an integration of drill-core and seismic-reflection data. *Palaeogeography, Plaeoclimatology, Palaeoecology*, **303**, 20–37.

Markert JA, Arnegard ME, Danley PD, Kocher TD (1999) Biogeography and population genetics of the Lake Malawi cichlid *Melanochromis auratus*: habitat transience, philopatry and speciation. *Molecular Ecology*, **8**, 1013–1026.

Maruyama A, Rusuwa B, Yuma M (2010) Asymmetric interspecific territorial competition over food resources amongst Lake Malawi cichlid fishes. *African Zoology*, **45**, 24–31.

Meyer A (1993) Phylogenetic relationships and evolutionary processes in East African cichlid fishes. *Trends in Ecology and Evolution*, **8**, 279–284.

Neff BD, Fu P, Gross MR (1999) Microsatellite evolution in sunfish. *Canadian Journal of Fisheries and Aquatic Sciences*, **56**, 1198–1205.

Newman D, Pilsen D (1997) Increased probability of extinction due to decreased genetic effective population size: experimental populations of *Clarkia pulchella*. *Evolution*, **51**, 354–362.

Nielsen EE, Hansen MM (2008) Waking the dead: the value of population genetic analyses of historical samples. *Fish and Fisheries*, **9**, 450–461.

Nunney L, Elam DR (1994) Estimating the effective population size of conserved populations. *Conservation Biology*, **8**, 175–184.

Ohta T, Kimura M (1973) A model of mutation appropriate to estimate the number of electrophoretically detectable alleles in a finite population. *Genetics Research*, **22**, 201–204.

Orr MR, Smith TB (1998) Ecology and speciation. *Trends in Ecology and Evolution*, **13**, 502–506.

Owen RB, Crossley R, Johnson TC *et al.* (1990) Major low levels of Lake Malawi and their implications for speciation rates in cichlid fishes. *Proceedings of the Royal Society London B*, **240**, 519–553.

Palstra FP, Ruzzante DE (2008) Genetic estimates of contemporary effective population size: what can they tell us about the importance of genetic stochasticity for wild population persistence. *Molecular Ecology*, **17**, 3428–3447.

Palstra FP, O'Connell MF, Ruzzante DE (2009) Age structure, changing demography and effective population size in Atlantic salmon (*Salmo salar*). *Genetics*, **182**, 1233–1249.

Piry S, Luikart G, Cornuet JM (1999) Bottleneck: a computer program for detecting recent reductions in the effective population size using allele frequency data. *Journal of Heredity*, **90**, 502–503.

Pollak E (1983) A new method for estimating the effective population size from allele frequency changes. *Genetics*, **104**, 531–548.

Posada D (2008) jModeltest: Phylogenetic model averaging. *Molecular Biology and Evolution*, **25**, 1253–1256.

Rambaut A, 2009. FigTree version 1.3.1. Available at http://tree.bio.ed.ac.uk/software/figtree/.

Rambaut A, Drummond AJ (2009) Tracer v1.5. Available at http://beast.bio.ed.ac.uk/Tracer.

Ray JW, Husemann M, Lutz-Carrillo DJ, King RS, Danley PD (2015) Life at the leading edge: genetic impoverishment of the spotted bass, *Micropterus punctulatus*, at its Western edge. *Environmental Biology of Fishes*, DOI:10.1007/s10641-015-0400-x.

Raymond M, Rousset F (1995) GENEPOP (version 1.2): population genetics software for exact tests and ecumenicism. *Journal of Heredity*, **86**, 248–249.

Ryman N, Allendorf FW, Jorde PE, Laikre L, Hössjer O (2014) Samples from subdivided populations yield biased estimates of effective size that overestimate the rate of loss of genetic variation. *Molecular Ecology Resources*, **14**, 87–99.

Salzburger W, Meyer A (2004) The species flocks of East African \cichlid fishes: recent advances in molecular phylogenetics and population genetics. *Naturwissenschaften*, **91**, 277–290.

Salzburger W, Mack T, Verheyen E, Meyer A (2005) Out of Tanganyika: genesis, explosive speciation, key-innovations and phylogeography of the haplochromine cichlid fishes. *BMC Evolutionary Biology*, **5**, 17.

Scholz CA, Johnson TC, Cohen AS *et al.* (2007) East African megadroughts between 135 and 75 thousand years ago and bearing on early-modern human origins. *Proceedings of the National Academy of Sciences USA*, **104**, 16416–16421.

Scholz CA, Cohen AS, Johnson TC, King J, Talbot MR, Brown ET (2011) Scientific drilling in the Great Rift Valley: the 2005 Lake Malawi Scientific Drilling Project – an overview of the past 145,000 years of climate variability in Southern Hemisphere East Africa. *Palaeogeography, Palaeoclimatology, Palaeoecology*, **303**, 3–19.

Schuelke M (2000) An economic method for the fluorescent labeling of PCR fragments. *Nature Biotechnology*, **18**, 233–234.

Schwartz MK, Tallmon DA, Luikart G (1998) Review of DNA-based census and effective population size estimators. *Animal Conservation*, **1**, 293–299.

Seehausen O (2000) Explosive speciation rates and unusual species richness in Haplochromine cichlid fishes: effects of sexual selection. *Advances in Ecological Research*, **31**, 237–274.

Seehausen O (2006) African cichlid fish: a model system in adaptive radiation research. *Proceedings of the Royal Society B*, **273**, 1987–1998.

Smith PF, Kornfield I (2002) Phylogeography of Lake Malawi cichlids of the genus *Pseudotropheus*: significance of allopatric colour variation. *Proceedings of the Royal Society of London B*, **269**, 2495–2502.

Stauffer JR, Black K, Konings AF (2013) Descriptions of five new species of *Metriaclima* (Teleostei: Cichlidae) from Lake Malawi. *Zootaxa*, **3647**, 101–136.

Sturmbauer C (2008) The Great Lakes in East Africa: biological conservation considerations for species flocks. *Hydrobiologia*, **615**, 95–101.

Sturmbauer C, Baric S, Salzburger W, Rüber L, Verheyen E (2001) Lake level fluctuations synchronize genetic divergences of cichlid fishes in African lakes. *Molecular Biology and Evolution*, **18**, 144–154.

Sturmbauer C, Husemann M, Danley PD (2011) New insights on explosive speciation and adaptive radiation from African cichlid fishes. In: *Hotspots* (eds Habel JC, Zachos FE), pp. 333–362. Springer, Berlin, Germany.

Takahashi K, Terai Y, Nishida M, Okada N (2001) Phylogenetic relationships and ancient incomplete lineage sorting among cichlid fishes in Lake Tanganyika as revealed by analysis of the insertion of retroposons. *Molecular Biology and Evolution*, **18**, 2057–2066.

Traill LW, Bradshaw CJA, Brook BW (2007) Minimum viable population size: a meta-analysis of 30 years of published estimates. *Biological Conservation*, **139**, 159–166.

Uyeda JC, Arnold SJ, Hohenlohe PA, Mead LS (2009) Drift promotes speciation by sexual selection. *Evolution*, **63**, 583–594.

Van Oosterhout C, Hutchinson WF, Wills DPM, Shipley P (2004) Micro-checker: software for identifying and correcting genotyping errors in microsatellite data. *Molecular Ecology Notes*, **4**, 535–538.

Vucetich JA, Waite TA, Nunney L (1997) Fluctuating population size and the ratio of effective to census population size. *Evolution*, **51**, 2017–2021.

Wandeler P, Hoeck PEA, Keller LF (2007) Back to the future: museum specimens in population genetics. *Trends in Ecology and Evolution*, **22**, 634–642.

Wang J, Whitlock MC (2003) Estimating effective population size and migration rates from genetic samples over space and time. *Genetics*, **163**, 429–446.

Waples RS (1989) A generalized approach for estimating effective population size from temporal changes in allele frequency. *Genetics*, **121**, 379–391.

Waples RS (1990) Conservation genetics of Pacific salmon. III. Estimating effective population size. *Journal of Heredity*, **81**, 277–289.

Waples RS (2002) Definition and estimation of effective population size in the conservation of endangered species. In: *Population Viability Analysis* (eds Beissinger SR, McCullough DR), pp. 147–168. The University of Chicago Press, Chicago, Illinois.

Waples RS (2005) Genetic estimates of contemporary effective population size: to what time periods do estimates apply? *Molecular Ecology*, **14**, 3335–3352.

Waples RS, Do C (2008) LDN_e: a program for estimating effective population size from data on linkage disequilibrium. *Molecular Ecology Resources*, **8**, 753–756.

Waples RS, England PR (2011) Estimating contemporary effective population size on the basis of linkage disequilibrium in the face of migration. *Genetics*, **189**, 633–644.

William EG, Slatkin M (1999) Using maximum likelihood to estimate population size from temporal changes in allele frequencies. *Genetics*, **152**, 755–761.

Won Y-J, 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 USA*, **102**, 6581–6586.

M.H. and P.D.D. designed the experiment. M.H. and R.N. performed the lab work. M.H. and B.D. analysed the data. M.H. lead the writing, all authors read and approved the manuscript..

Data accessibility

All sequence data were deposited in GenBank (Table 1). The microsatellite data, the mitochondrial alignment and the dated tree file are available from Dryad (doi:10.5061/dryad.53686).

Appendix 1 Microsatellites used in this study: given is their repeat pattern and motive, the range of allele sizes, the annealing temperature, the linkage group on which the locus is found, the primers used and the reference where primers were taken from

ID	Bp repeat	Repeat pattern	Allele sizes	Annealing temp., °C	Genome location	Primer f	Primer r	References
UNH2037	2	CA	126–236	58	LG 2	GGGATTCACTGGCACCTACT	ATGTGGTTCCCAGTGATGGT	Albertson *et al.* (2003)
UNH2086	2	CA	134–282	59	LG 1	AACAGGCCGAGCAGAAAGT	CGATAGGCTGTTTTCCTGGAG	Albertson *et al.* (2003)
UNH2139	2	GT	195–259	56	LG 12	GCAGTGCACATGCGACTTAT	ACAGCCAGCTACTGTGCAAC	Albertson *et al.* (2003)
UNH2204	2	GT	116–194	56	LG 8	CACATCATGTCAATCAGACATCC	GGAGACGGTTCAAAGTCCTG	Albertson *et al.* (2003)
UNH2190	2	GT	120–194	56	LG 10	GTTCGGCTGTGATGGTGATT	AGCGAGGACGGAGCTTTAAC	Albertson *et al.* (2003)
UNH2169	2	GT	115–203	56	LG 11	CCAGTGGGTCCTCCTACAGA	CCCAGTGACTTTGAGGTGTG	Albertson *et al.* (2003)
UNH2112	2	GT	121–227	56	LG 13	CTCGGTGGTCAGAATGAAGG	TTACAGCACTTCACGGTTGC	Albertson *et al.* (2003)
UNH362	2	CA	110–188	56	LG 17	GAACAGCTTTCAGACGGAGG	ACTGAGGCCAGGTGAAGAAA	T. D. Kocher, unpublished
UNH2166	2	CA	119–233	56	LG 16	ACTGGCCCAAAACTGTCAAA	TGTGTGCCAAGGATAGCAAA	Albertson *et al.* (2003)
UNH2065	2	CA	109–221	56	LG 19	CCGGGATGATTTTCTCACTG	CAGCACACGACAGGAGGTT	Albertson *et al.* (2003)
UNH2152	2	GT	117–261	56	LG 3	TGACTGCTGCACATTATAACTCC	CAGCATGAAACTCACTGGAAA	Albertson *et al.* (2003)
UNH2135	2	GT	105–243	56	LG 4	CCTGACAAAGCTGATTGTTCC	GTGAATGCTGAGGCAAGTCA	Albertson *et al.* (2003)
UNH231	2	CA	150–282	56	LG 6	GCCTATTAGTCAAAGCGT	ATTTCTGCAAAAGTTTTCC	B. Y. Lee & T. D. Kocher, unpublished