MOLECULAR ECOLOGY
RESOURCES
Molecular Ecology Resources (2013) **13**, 347–353 doi: 10.1111/1755-0998.12066

# Mitochondrial genome primers for Lake Malawi cichlids

C. DARRIN HULSEY, BENJAMIN P. KECK, HUGO ALAMILLO and BRIAN C. O'MEARA
*Department of Ecology and Evolutionary Biology, University of Tennessee, Knoxville, TN 37996, USA*

## Abstract

**Resolving the evolutionary history of rapidly diversifying lineages like the Lake Malawi Cichlid Flock demands powerful phylogenetic tools. Although this clade of over 500 species of fish likely diversified in less than two million years, the availability of extensive sequence data sets, such as complete mitochondrial genomes, could help resolve evolutionary patterns in this group. Using a large number of newly developed primers, we generated whole mitochondrial genome sequences for 14 Lake Malawi cichlids. We compared sequence divergence across protein-coding regions of the mitochondrial genome and also compared divergence in the mitochondrial loci to divergence at two nuclear protein-coding loci, Mitfb and Dlx2. Despite the widespread sharing of haplotypes of identical sequences at individual loci, the combined use of all protein-coding mitochondrial loci provided a bifurcating phylogenetic hypothesis for the exemplars of major lineages within the Lake Malawi cichlid radiation. The primers presented here could have substantial utility for evolutionary analyses of mitochondrial evolution and hybridization within this diverse clade.**

*Keywords*: adaptive Radiation, East African Rift Lakes, genomes, mbuna, tilapia, utaka

*Received 20 August 2012; revision received 11 December 2012; accepted 12 December 2012*

## Introduction

Reconstructing the evolutionary history of rapidly diversifying lineages is difficult. For instance, attempts to use molecular data to resolve the phylogeny of the Lake Malawi Cichlid Flock (LMCF) have proved frustrating. This is largely due to the short time frame (approximately 2 million years) in which these over 500 species of fish are thought to have diversified (Kocher *et al.* 1995; Genner *et al.* 2007; Hulsey *et al.* 2010). Approximately 99% of the cichlids in Lake Malawi belong to a single clade that apparently diversified so rapidly that species placed within different genera share alleles at both mitochondrial and nuclear loci (Danley & Kocher 2001; Won *et al.* 2006; Hulsey *et al.* 2010). However, resolving relationships within this rapidly radiating group of fishes should now be feasible as genomic resources are becoming increasingly available (Hulsey 2009). Using a host of new primers, we generated mitochondrial genomes for a diversity of Malawi cichlids and used these genomes to examine the patterns of mitochondrial evolution and gene tree evolution for the LMCF.

Components of the mitochondrial genome are frequently the initial marker used to infer phylogenetic relationships among rapidly diversifying animal clades (Kornfield & Smith 2000; Hulsey *et al.* 2007; Fraser *et al.* 2009). Mitochondrial markers are also the primary source of DNA sequences that have been used to parameterize molecular clock estimates of the age of the LMCF (Genner *et al.* 2007; Hulsey *et al.* 2010). Mitochondrial loci have relatively rapid rates of evolution and rapid sorting of polymorphism compared with nuclear encoding loci due to the lower effective population size of mitochondrial vs. nuclear loci (Avise & Walker 1988; Avise *et al.* 1988; Moore 1995). However, particular regions of the mitochondrial genome could exhibit exceptionally slow or rapid rates of change in the LMCF and provide misleading estimates of this group's diversification rate. For instance, most studies of evolutionary relationships within the LMCF have relied on the mitochondrial ND2 gene or the control region (Moran & Kornfield 1993; Parker & Kornfield 1997; Hulsey *et al.* 2007). Because the control region plays a role in mitochondrial replication and its functional properties can vary during times of increased replication rate (Excoffier & Yang 1999; Brown & Clayton 2002), the control region and possibly several of the protein-coding regions might have substantially different rates of evolution compared with the remainder of the mitochondrial genome.

Different species within the LMCF often share identical mitochondrial haplotypes when single mitochondrial loci are examined (Moran & Kornfield 1993; Parker & Kornfield 1997; Hulsey *et al.* 2007). Sharing of mitochondrial

Correspondence: C. Darrin Hulsey, Fax: (001) 65-974-3067;
E-mail: chulsey@utk.edu

alleles could be due to introgression of the entire mitochondria when otherwise reproductively isolated species hybridize (Chan & Levin 2005; Bachtrog *et al.* 2006; Bossu & Near 2009). However, the mitochondrial genome might not show exceptional degrees of introgression compared with nuclear genes within the LMCF (Mims *et al.* 2010). Furthermore, the appearance of allele sharing could simply be a result of inadequate time for molecular divergence in the individual mitochondrial regions previously sequenced (Won *et al.* 2005, 2006). Comparing sequences of all protein-coding loci in the mitochondrial genome of LMCF species will allow us to determine if particular regions simply show limited variability while the entire mitochondrial genome contains sufficient signal to reconstruct mitochondrial gene trees. The mitochondrial protein-coding loci might also be exceptionally informative when compared with nuclear protein-coding loci and provide a set of bifurcating relationships against which to test future evolutionary analyses of the nuclear genome.

We examined several patterns of molecular evolution of the mitochondrial genome in the LMCF. Using exemplars of the major lineages in the LMCF, we documented the patterns of mitochondrial divergence within the radiation and compared it with two protein-coding loci from the nuclear genome for the same set of individuals. We also determined if members of the LMCF commonly exhibited shared haplotypes at individual loci. Finally, we used the protein-coding regions of the complete mitochondrial genomes to reconstruct a phylogeny for 14 exemplars of major lineages within the LMCF.

## Materials and methods

To generate primers for the mitochondrial genome, we first created alignments of fully sequenced mitochondrial genomes for *Tropheus duboisi* (AP006015), *Neolamprologus brichardi* (AP006014) and *Oreochromis niloticus* (GU238433) from Mabuchi *et al.* (2007) using the program SEQUENCHER version 4.1 (Gene Codes). Using SEQUENCHER, we focused our primer design efforts on regions we identified by eye that ranged from 22 to 27 base pairs (bp) in length and that also showed little divergence between these three African cichlid species. We initially produced a series of forward and reverse primer pairs that spanned aligned regions from 1000 to 2000 bp in length. If PCR reactions of these initial primer pairs failed to produce useful sequences, alternative primer pairs that spanned the regions but represented less conserved regions were generated. All primers were then realigned to the genomes to ensure they did not have multiple priming sites.

Forty-five novel and six previously published primers were used to amplify the entire mitochondrial genomes (Table 1; Fig. 1). For sequencing, total genomic DNA was isolated from caudal fin clips using DNeasy extraction kits (Qiagen). A 1.0-$\mu$L aliquot of extraction was used to provide a DNA template for polymerase chain reaction (PCR). The master mix for all PCRs consisted of combining 13.0 $\mu$L of GoTaq Hot Start Green Master Mix (Promega), 2.5 $\mu$L of both the forward and reverse primers and 6.0 $\mu$L of nuclease-free water (Promega) for a final combined master mix volume of 24.0 $\mu$L. Amplifications of the DNA template and master mix were carried out in a Perkin–Elmer DNA thermocycler using standard methods. Thermal cycling conditions for all regions sequenced consisted of an initial denaturation step of 94 °C (30 s), an annealing step of either 55 °C or 49 °C (30 s) and an extension step of 72 °C (1.5 min). A final incubation of 72 °C for 5 min was added to the end of the PCR reaction to ensure complete extension of amplified products. Subsequently, the PCR products were electrophoretically separated from unincorporated primers and dNTPs using electrophoresis in low-melting-point agarose gel with ethidium bromide (1 mg/L) added and run in Tris-acetate buffer (pH 7.8). Positively amplified DNA was then enzymatically purified. The PCR products were sequenced using Sanger sequencing at the High-Throughput DNA Sequencing Facility at the University of Washington. The amplicon lengths of the primer pairs ranged from 312 to 2641 bp (Table 1). Gene sequences were assembled from individual sequencing reactions using SEQUENCHER version 4.1.

We tested the primers on 14 cichlids that span the morphological diversity of the haplochromine cichlids in the LMCF. Five species of the rock-dwelling mbuna including *Cynotilapia afra* (JN628861), *Genyochromis mento* (JN628858), *Petrotilapia nigra* (JN628852), *Pseudotropheus crabro* (JN628854) and *Tropheops* 'redcheek' (JN628862) were sequenced. Nine species that represent other major lineages in the lake were also sequenced including *Astatotilapia calliptera* (JN628855), *Cheilochromis euchilus* (JN252050), *Dimidiochromis compressiceps* (JN628856), *Diplotaxodon limnothrissa* (JN628851), *Nimbochromis linni* (JN628853), *Pallidochromis tokoloshi* (JN628859), *Rhamphochromis esox* (JN628860), *Taeniolethrinops preaorbitals* (JN628857) and *Trematocranus placodon* (JN628850).

To determine the patterns of LMCF divergence at mitochondrial loci, we identified the mean, maximum and minimum uncorrected sequence divergence for all mitochondrial protein-coding genes. To provide a reference for the relative amount of divergence in nuclear gene protein-coding loci compared with mitochondrial protein-coding loci, we also examined 397 bp of the nuclear gene Mitfb (Won *et al.* 2006) and the 795 bp of the Dlx2 gene (Hulsey *et al.* 2010). The same individuals sequenced for the mitochondrial genomes in this study were used to sequence these nuclear gene sequences that

**Table 1** Summary of primers used to amplify sequences of the entire mitochondrial genome sequences. The mitochondrial genomic regions and the primer pairs amplified are listed. The primers were also mapped to the *Dimidiochromis compressiceps* mitochondrial genome and the expected amplicon size from the primer pair is provided. Amplicons were generated at an annealing temperature of either [(a)] 55 °C or [(b)] 49 °C and the temperature used for each pair is noted with the appropriate subscript following the amplicon size. For each primer, the first base pairs sequenced as estimated from the end of the control region in the *D. compressiceps* genome are given in the 'Site' column. A few of the primers were used in more than one primer pair. Three primer sets used to amplify the mitochondria were taken from published studies (* Kocher *et al*. 1995; # Lee *et al*. 1995; † Palumbi 1996), but most were generated from alignments of complete cichlid mitochondrial genomes (Mabuchi *et al*. 2007)

| Regions amplified | Amplicon size (bp) | Site | Primer sets | 3′ to 5′ |
|---|---|---|---|---|
| Phe, Val, 12S | 988[a] | 34 | MitCich1F | AGCTTAACTAAAGCATAACACTGAA |
| | | 1022 | MitCich1R | AAGGGAGATGCTTTKCTGTCTTAG |
| 12S, Val, 16S | 312[b] | 858 | MitCich2F | GAGGATTTAGCAGTAAGCAGA |
| | | 1170 | MitCich2R | CGTCGCCCGTACTTAGGTGGA |
| 12S, Val, 16S | 1161[b] | 992 | MitCich3F | AAGTCGTAACATGGTAAGTGTACC |
| | | 2153 | MitCich3R | AAGGAGACAGTTAAGCCCTCGT |
| 16S, Leu, ND1 | 925[b] | 2014 | MitCich4F | CAYAAGCCTCGCCTGTTTACC |
| | | 2939 | MitCich4R | TGTATGTAGCCRAGRACTTTTCGT |
| 16S | 577[b] | 2022 | 16Sar † | CGCCTGTTTATCAAAAACAT |
| | | 2599 | 16Sbr † | CCGGTCTGAACTCAGATCACGT |
| 16S, Leu, ND1, Ile, Gln, Met, ND2, Trp, Ala, Asn, Cys, Tyr, CoxI | 3382[b] | 2116 5498 | MitCich5F<br>MitCich5R | TGACCGTGCAAAGGTAGCGCAAT<br>GGTGCCGATGTCTTTGTGATTAGTT |
| Leu, ND1, Ile, Gln, Met | 1175[b] | 2834 | MitCich6F | CTAAGCCCTTTYCACGGAGGTTC |
| | | 4009 | MitCich6R | AAGGAAGGAGTTTAACCAACATG |
| Leu, ND1, Ile, Gln, Met, ND2, Trp, Ala, Asn, Cys, Tyr, CoxI | 2664[a] | 2834 5498 | MitCich6F<br>MitCich5R | CTAAGCCCTTTYCACGGAGGTTC<br>GGTGCCGATGTCTTTGTGATTAGTT |
| Gln, Met, ND2, Trp, Ala, Asn | 1316[a] | 3941 | GLN * | CTACCTGAAGAGATCAAAAC |
| | | 5257 | ASN * | CGCGTTTAGCTGTTAACTAA |
| ND2, Trp, Ala, Asn, Cys, Cys, Tyr, CoxI | 539[b] | 5032 5571 | MitCich7F<br>MitCich7R | TCGCCACTTCAACTGCAATAA<br>GTTCTGCCCGAATTAGGAGGCT |
| Tyr, CoxI | 1105[a] | 5487 | MitCich8F | TCYTACCTGTGGCAATCACACG |
| | | 6592 | MitCich8R | GAGGACATAGTGGAAGTGGGCAA |
| CoxI, Ser, Asp, CoxII | 909[b] | 6470 | MitCich9F | ACTCTGCACGGCGGCTCAATTA |
| | | 7379 | MitCich9R | CGATTTCTTGGGAGTCTAGGAT |
| CoxII, Leu, Atp8, Atp6 | 1185[a] | 7292 | MitCich10F | ATAGAAGAACTTCTTCACTTCCACGA |
| | | 8477 | MitCich10R | TTCGTATGCCAATAATKACTGT |
| Atp6, CoxIII | 722[a] | 8183 | MitCich11F | TTGACCAATTYATGAGCCCCACATA |
| | | 8905 | MitCich11R | ATTCAGATTGCTAAACCGGAKGT |
| Atp6, CoxIII | 270[a] | 8784 | MitCich12F | GCCATAATTCAAGCCTATGTC |
| | | 9054 | MitCich12R | AGGCCTCCGATACGGRATGATC |
| Atp6, CoxIII, Gly, ND3, Arg | 1243[a] | 8838 | MitCich13F | ACGTYTAATGGCCCATCAAG |
| | | 10081 | MitCich13R | GGGTCATTAGGYGGTTRTGGACTT |
| ND3, Arg | 333[a] | 9811 | MitCich14F | CCTACGAATGTGGCTTTGATC |
| | | 10144 | MitCich14R | TCGGTGGAATGCCAGGCCTGCTA |
| Arg, ND4L, ND4 | 1080[a] | 10058 | MitCich15F | GTTAGTTTAAGAAAAACCCTTGA |
| | | 11138 | MitCich15R | GTTCTTTGGTGAGWGGTTCCTAG |
| ND4 | 603[a] | 11068 | MitCich16F | GCACACGTTGAAGCCCCAATG |
| | | 11671 | MitCich16R | GTGYTCTCGGGAGTGWGTTGGGT |
| ND4, His, His, Ser, Leu | 388[a] | 11574 | MitCich17F | ACTGRTCYTGATGAACAATTGC |
| | | 11962 | MitCich17R | ATAGCTGCTACTTGGAYTTGCAC |
| Leu, ND5 | 1259[a] | 11942 | MitCich18F | GATAACAGCTCATCCATTGGT |
| | | 13201 | MitCich18R | CTCAGGCGTTTAGGTARGATGT |
| ND5, ND6, Glu, CytB | 2641[a] | 12901 | MitCich19F | ATCGTTGCTTTCTCAACCTCAAG |
| | | 15542 | MitCich19R | TCTAGTCAGGTTTGGTCRACT |
| ND5, ND6 | 1071[b] | 13379 | MitCich20F | AACCCCATCAAACGCCTAGCCCTG |
| | | 14450 | MitCich20R | TCAACTAGTGCGTCGTTTGCAAT |
| ND6, Glu, CytB | 1143[a] | 14399 | MitCich21F | GACTTGAAAAACCACCGTTGTWATTC |

**Table 1** (Continued)

| Regions amplified | Amplicon size (bp) | Site | Primer sets | 3′ to 5′ |
|---|---|---|---|---|
| | | 15542 | MitCich19R | TCTAGTGCAGGTTTGGTCRACT |
| Glu, CytB, Thr, Pro | 1257[a] | 14399 | MitCich21F | GACTTGAAAAACCACCGTTGTWATTC |
| | | 15656 | MitCich21R | AGRATCCTAGCTTTGGGAGYTAG |
| Tyr, Pro, Control Region, 12S | 1032[a] | 15628 | MitCich22F | CGAAGGTTAAGTCCTTCCTACYGCT |
| | | 74 | MitCich22R | TGACAGTAAAGTCAGGACCAAGC |
| Control Region, Phe | 592[a] | 15999 | CRLeeL # | AGTAAGAGCCCCACCATCAGT |
| | | 5 | CRLeeM # | TATGCTTTAGTTAAGGCTACG |
| Control Region, Phe, Val, 12S | 1608[a] | 16000 | MitCich23F | TGTAGTAAGAGCCCACCATCAG |
| | | 1022 | MitCich1R | AAGGGAGATGCTTTKCTGTCTTAG |
| Control Region, Phe | 488[b] | 16172 | MitCich24F | ATGGTGGTAATACATACTCCTCG |
| | | 74 | MitCich22R | TGACAGTAAAGTCAGGACCAAGC |

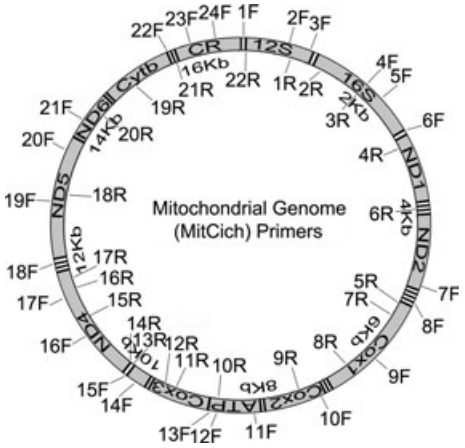


**Fig. 1** The annealing location of the new mitochondrial primers presented in Table 1. Primers are shown with respect to the location of protein-coding loci in the Lake Malawi cichlid flock (LMCF) mitochondrial genome. The mitochondrial tRNAs are not labelled but are noted with double black lines. The annealing locations of all forward primers are depicted outside the circular genome and all reverse primers are placed inside the representation of the mitochondrial genome. Lengths of 2000 bp starting with the end of the control region are demarcated to highlight the amplicon sizes and annealing sites of the mitochondrial genome primers.

are presented in Hulsey *et al.* (2010). Both of these loci contained intronic and exonic regions. We identified the exonic regions based on comparisons to cDNA sequences from the African cichlids *Maylandia zebra* (AY196318) for Mitfb and *Oreochromis niloticus* (AF534538) for Dlx2. Within the *Oreochromis niloticus* genome, Mitfb is located on scaffold GL831152.1: 2,347,962-2,375,409 and Dlx2 is located on scaffold GL831187.1: 1,335,057-1,337,518. To provide relevant comparisons to the protein-coding mitochondrial loci, we compared only the divergence in the exonic regions of the two nuclear genes.

To infer the mitochondrial gene tree of the LMCF species, we used only the protein-coding regions of the 14 Lake Malawi cichlids. We did not use the tRNAs or the control region for this analysis because of potential ambiguities in alignment. The cichlids *Tropheus duboisi* (AP006015), *Neolamprologus brichardi* (AP006014) and *Oreochromis niloticus* (GU28433) were used as outgroups. After aligning each locus with MUSCLE v3.8 (Edgar 2004), we used AIC in JMODELTEST v0.1.1 (Guindon & Gascuel 2003; Posada 2008) to identify the best model of molecular evolution for the protein-coding genes. Bayesian analyses were run employing MRBAYES v.3.1.2 (Ronquist & Huelsenbeck 2003). We partitioned the data by gene, with the previously identified models of molecular evolution specified for the respective partitions, and rates were allowed to vary across partitions (Rannala & Yang 2008). We ran the analysis in MRBAYES v.3.1.2 (Ronquist & Huelsenbeck 2003) for 20 000 000 generations in two simultaneous runs with four Markov chains in each run. Trees and parameters were sampled from the MCMC chain every 1000 generations. We assessed convergence by examining trends in the likelihood vs. generation plots, values for the average standard deviation of the split frequencies, and effective sample sizes in TRACER v1.5 (Rambaut & Drummond 2007). We used the sumt command in MRBAYES to remove the first 20% of trees as burn-in and estimated the maximum clade credibility tree with the postburn-in set of trees.

## Results

The mitochondrial genomes of the LMCF ranged from 16578 to 16585 base pairs (bp) in length. The alignment of all protein-coding loci totalled 11470 bp. Uncorrected sequence divergence among the protein-coding regions of the LMCF mitochondria varied extensively (Table 2). The gene ND6 showed the largest maximum uncorrected genetic distance (4.2%) between members of the LMCF,

**Table 2** The mean, minimum and maximum observed uncorrected sequence divergence at mitochondrial loci was calculated for the sequenced Malawi species

| Gene | Mean% | Min% | Max% |
|---|---|---|---|
| ND1 | 2.4 | 0.0 | 3.6 |
| ND2 | 2.0 | 0.1 | 3.5 |
| Cox1 | 2.2 | 0.0 | 3.6 |
| Cox2 | 1.7 | 0.0 | 2.7 |
| ATP8 | 1.4 | 0.0 | 3.0 |
| ATP6 | 1.8 | 0.0 | 3.4 |
| Cox3 | 1.8 | 0.0 | 2.9 |
| ND3 | 2.4 | 0.0 | 4.0 |
| ND4L | 1.3 | 0.0 | 3.7 |
| ND4 | 2.5 | 0.1 | 3.9 |
| ND5 | 2.4 | 0.1 | 3.5 |
| ND6 | 2.4 | 0.2 | 4.2 |
| CytB | 2.7 | 0.1 | 4.1 |

and Cytb showed the next greatest amount of sequence divergence (4.1%). The control region had a higher maximum per cent divergence (5.2%) than any of the mitochondrial protein-coding loci examined. The gene ND4L showed the smallest average divergence (1.3%) across the LMCF.

The nuclear gene sequencing resulted in 96 bp of Mitfb exonic sequence and 172 bp of Dlx2 exon. The Mitfb sequence was invariable across the region sequenced and the Dlx2 gene exhibited single mutations in both *Cheilochromis euchilus* and *Diplotaxodon limnothrissa*. The two mutations observed were at different locations in Dlx2. The maximum divergence between these two nuclear exons was 0% and 0.6% respectively. The species showing alternative allele differences in Dlx2 are noted in Fig. 2.

The loci ND2, ND4, ND5, ND6, Cytb and the control region showed at least some genetic divergence among all of the species sequenced (Table 2). However, all the other protein-coding loci showed identical sequences (0.0% sequence divergence) between at least two members of the LMCF species sequenced. The mbuna species like *P. nigra*, *G. mento*, *T.* 'redcheek', *P. crabro* and *C. afra* commonly shared identical haplotypes at several of the mitochondrial loci. However, the non-mbuna species *C. euchilus*, *N. linni*, *T. praeorbitalis*, *T. placodon* and *D. compressiceps* also exhibited identical haplotypes at several of the sequenced loci.

Despite the overall low amount of sequence divergence, the mitochondrial genome protein-coding sequences provided a well-supported topology of relationships among the members of the LMCF examined (Fig. 2). The entire mitochondrial genome data set contained 1138 parsimony informative sites. The mitochondrial protein-coding loci in the 14 LMCF

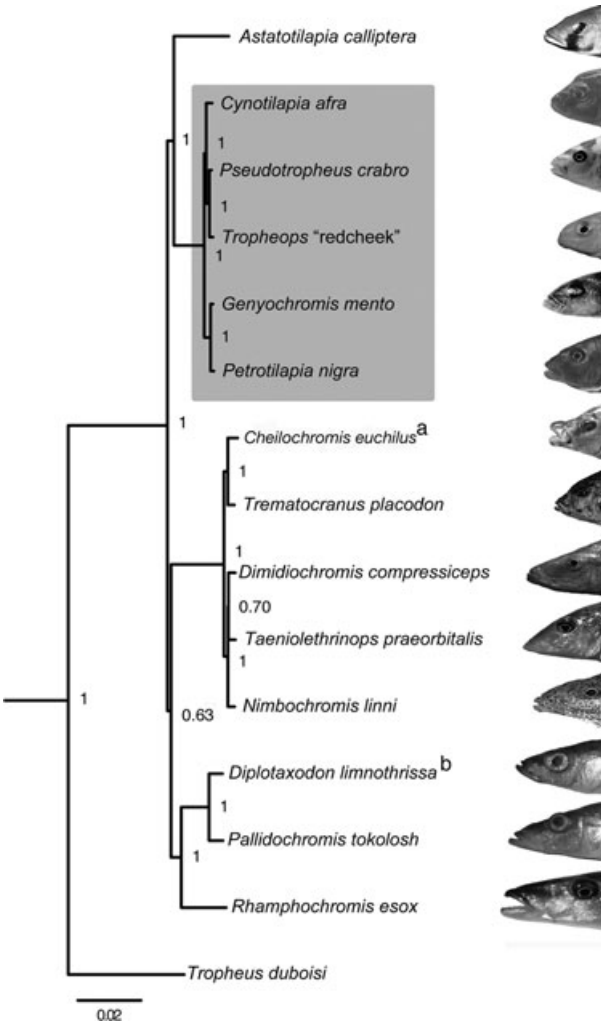


**Fig. 2** Phylogeny of 14 species of the LMCF based on complete mitochondrial genomes. The phylogeny was analysed with *Oreochromis niloticus*, *Neolamprologus brichardi* and *Tropheus duboisi* (only *T. duboisi* is depicted) as outgroups to polarize the LMCF species. Nodes in the phylogeny were generally supported with posterior probabilities of 1.00 (1), and only two nodes within the LMCF had lower support. The rapidly evolving mbuna clade is highlighted with grey. The overall low level of divergence in this mitochondrial phylogeny is depicted with a per cent divergence scale bar. The two species that show alternative alleles at Dlx2 are noted (a,b). Images of the heads of each Malawi species sequenced are shown to the right of the species name.

species contained a total of 846 variable characters, with the locus ND5 having the most (192) and Atp8 (11) having the fewest. We identified convergence in the results of the Bayesian analysis well before the 20% burn-in and the effective sample sizes for all parameters were above 200, indicative of adequate mixing in the MCMC (Rambaut & Drummond 2007). The mbuna *P. nigra*, *G. mento*, *T.* 'redcheek', *P. crabro* and *C. afra* formed a well-supported monophyletic group with strictly

bifurcating relationships. The species *A. calliptera* was strongly supported as sister group to the mbuna clade. The species *T. praeorbitalis* showed slightly closer affinities (0.70 posterior probability) to *D. compressiceps* than either did to *N. linni*. The species *C. euchilus*, *N. linni*, *T. praeorbitalis*, *T. placodon* and *D. compressiceps* formed a strongly supported clade that was placed with moderate support (0.63 posterior probability) as sister group to the deep-water dwelling species *P. tokolosh*, *D. limnothrissa*, and pelagic *R. esox*.

## Discussion

The patterns of mitochondrial genome evolution within the LMCF largely support the results of previous studies based on single mitochondrial loci. The ND2 gene, which has been frequently used to reconstruct initial relationships among members of the LMCF (Fraser *et al.* 2009; Hulsey *et al.* 2010), showed a similar amount of sequence divergence compared with other loci. The widespread use of this locus to make inferences about the evolution of the LMCF has likely not biased our view of mitochondrial divergence among these cichlids. However, future studies of population level divergence or phylogenetic divergence among LMCF species might consider sequencing gene regions like ND6 or Cytb that evolve at the fastest rate of any of the mitochondrial genes.

None of the mitochondrial genes shows extraordinarily different rates of evolution within the LMCF (Table 2), and all of the regions of the mitochondrial genome examined in the LMCF exhibited less than 5% maximum uncorrected sequence divergence. Likewise, the phenomena of loci showing no sequence divergence between members of the LMCF are not limited to genes that have often been used to examine evolutionary history in the LMCF, like the control region (Parker & Kornfield 1997) or ND2 (Mims *et al.* 2010). The lack of divergence at particular mitochondrial loci is also not a phenomenon limited to the microsympatric mbuna, as several non-mbuna species share identical haplotypes at single loci.

The allele sharing is even more extreme in the two nuclear protein-coding loci examined (Fig. 2). All of the LMCF species examined had identical exons at the Mitfb locus. There were only two mutations in the Dlx2 exon among all of the species examined, and these mutations were not parsimony informative. Extrapolations to the entire proteome of Malawi cichlids from only these two nuclear protein-coding loci could be misleading. However, the total of only two variable sites in the 288 bp of Mitfb and Dlx2 exon examined would indicate a variable site approximately every 150 bp of nuclear genome protein-coding sequence. This low nuclear genome sequence divergence indicates that to obtain the 846 variable protein-coding characters found in the relatively compact mitochondrial genome, one would need to generate more than 11 times more nuclear protein-coding sequence (approximately 127 000 bp) for each of the LMCF species examined. The mitochondrial genome provides a well-supported gene tree for the LMCF, and because the mitochondrial protein-coding genes are physically linked and not subject to recombination, this locus could provide some advantages compared with loci from the nuclear genome (Moore 1995).

Although species shared identical sequences at several mitochondrial loci, the full set of mitochondrial protein-coding genes supported a largely nested set of bifurcating relationships for the Lake Malawi cichlid species (Fig. 2). This type of strictly bifurcating topology is unusual for the LMCF as most phylogenies, even those using multiple loci, are not well resolved (Albertson *et al.* 1999; Hulsey *et al.* 2007, 2010). Also, if each mitochondrial protein-coding locus were examined in isolation, this sharing of a haplotype at a particular locus might be interpreted as evidence of recent introgression wherein different species share the same mitochondrial genome (Parker & Kornfield 1997; Mims *et al.* 2010). However, many morphologically distinct members of the LMCF that have no sequence divergence at some mitochondrial loci do exhibit phylogenetically informative divergence at other loci. The linked nature of the mitochondrial genome indicates mitochondrial introgression during hybridization is not the only process responsible for sharing of mitochondrial haplotypes. The existence of gene tree structure, while nevertheless sharing alleles at particular loci, supports the idea that divergence in the LMCF is so recent that even many rapidly evolving mitochondrial genes have not had time to accumulate phylogenetically informative mutations (Hulsey *et al.* 2010).

It is becoming increasingly clear that genetic incompatibilities between organelle and nuclear genomes could contribute to speciation (Roca *et al.* 2005). Therefore, understanding patterns of divergence in the LMCF mitochondrial genome could provide not only a relatively unambiguous mitochondrial gene tree but also a more thorough understanding of the mechanisms of speciation in this rapidly diversifying clade. Furthermore, discordance between mitochondrial markers and nuclear markers have been used to infer instances of hybridization in the LMCF (Joyce *et al.* 2011), and the full battery of primers generated here could help further resolve patterns of hybridization in the LMCF. However, to fully resolve the phylogeny of this clade, novel phylogenetic methods will likely need to be implemented (Mims *et al.* 2010). The mitochondrial genome will continue to provide a tool for examining and understanding the evolutionary patterns and processes of rapidly diversifying clades like the LMCF.

## Acknowledgements

We thank TD Kocher, PD Danley, AE Howe, R Roberts and R Zatha for help collecting fishes in Malawi. AJ Ambali, the University of Malawi, Malawi Parks Service and the Malawi government for logistics and permissions to collect fish in Malawi. Domino Joyce and Michael Oliver provided images of cichlids. We also thank the NSF (IOB 0546423) for funding support to J. Todd Streelman and the University of Tennessee for support (to CDH).

## References

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.

Avise JC, Walker D (1988) Pleistocene phylogeographic effects on avian populations and the speciation process. *Proceedings of the Royal Society Series B: Biological Sciences*, **265**, 457–463.

Avise JC, Ball RM, Arnold J (1988) Current versus historical population sizes in vertebrate species with high gene flow: a comparison based on mitochondrial DNA lineages and inbreeding theory for neutral mutations. *Molecular Biology and Evolution*, **5**, l–344.

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.

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

Brown TA, Clayton DA (2002) Release of replication termination controls mitochondrial DNA copy number after depletion with 2′,3′-dideoxycytidine. *Nucleic Acids Research*, **30**, 2004–2010.

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

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

Edgar RC (2004) MUSCLE: multiple alignment with high accuracy and high throughput. *Nucleic Acids Research*, **32**, 1792–1797.

Excoffier L, Yang ZH (1999) Substitution rate variation among sites in mitochondrial hypervariable region 1 of humans and chimpanzees. *Molecular Biology and Evolution*, **16**, 1357–1368.

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

Genner MJ, Seehausen O, Lunt DH *et al*. (2007) Age of cichlids: new dates for ancient lake fish radiations. *Molecular Biology and Evolution*, **24**, 1269–1282.

Guindon S, Gascuel O (2003) A simple, fast, and accurate method to estimate large phylogenies by maximum-likelihood. *Systematic Biology*, **52**, 696–704.

Hulsey CD (2009) Cichlid genomics and phenotypic diversity in a comparative context. *Integrative and Comparative Biology*, **49**, 618–629.

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

Hulsey CD, Mims MC, Parnell NF, Streelman JT (2010) Comparative rates of lower jaw diversification in cichlid adaptive radiations. *Journal of Evolutionary Biology*, **23**, 1456–1467.

Joyce DA, Lunt DH, Genner MJ, Turner GF, Bills R, Seehausen O (2011) Repeated colonization and hybridization in Lake Malawi cichlids. *Current Biology*, **21**, R108.

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.

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

Lee WJ, Conroy J, Howell WH, Kocher TD (1995) Structure and evolution of teleost mitochondrial control regions. *Journal of Molecular Evolution*, **41**, 54–66.

Mabuchi K, Miya M, Azuma Y, Nishida M (2007) Independent evolution of the specialized pharyngeal jaw apparatus in cichlid and labrid fishes. *BMC Evolutionary Biology*, **7**, 10.

Mims MC, Hulsey CD, Fitzpatrick BM, Streelman JT (2010) Geography disentangles introgression from ancestral polymorphism in Lake Malawi cichlids. *Molecular Ecology*, **19**, 940–951.

Moore WS (1995) Inferring phylogenies from mtDNA variation: mitochondrial-gene trees versus nuclear-gene trees. *Evolution*, **49**, 718–726.

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.

Palumbi SR (1996) Nucleic acids II: the polymerase chain reaction. In: *Molecular Systematics* (eds Hillis DM, Moritz C & Mable BK), pp. 205–247. Sinauer & Associates Inc., Sunderland, Massachusetts.

Parker A, Kornfield I (1997) Evolution of the mitochondrial DNA control region in the mbuna (Cichlidae) species flock of Lake Malawi, East Africa. *Journal of Molecular Evolution*, **45**, 70–83.

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

Rambaut A, Drummond AJ (2007) *Tracer v1.4*, Available from http://beast.bio.ed.ac.uk/Tracer

Rannala B, Yang Z (2008) Phylogenetic inference using whole genomes. *Annual Review of Genomics and Human Genetics*, **9**, 217–31.

Roca AL, Georgiadis N, O'Brien SJ (2005) Cytonuclear genomic dissociation in African elephant species. *Nature Genetics*, **37**, 96–100.

Ronquist F, Huelsenbeck JP (2003) MrBayes 3: Bayesian phylogenetic inference under mixed models. *Bioinformatics*, **19**, 1572–1574.

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.

---

C.D.H. designed the primers, generated the sequences and wrote the manuscript. B.P.K. and H.A. performed the phylogenetic analyses. B.C.O. provided help with analyses and writing the paper.

---

## Data Accessibility

DNA sequences: GenBank accessions JN628850-JN628862, and JN252050.

Protein-coding alignments for phylogeny reconstruction and tables: DRYAD entry doi:10.5061/dryad.57vq0.