J Mol Evol (1997) 45:70–83

JOURNAL OF MOLECULAR EVOLUTION

© Springer-Verlag New York Inc. 1997

Evolution of the Mitochondrial DNA Control Region in the *mbuna* (Cichlidae) Species Flock of Lake Malawi, East Africa

Alex Parker, Irv Kornfield

Department of Zoology, University of Maine, Orono, ME 04469, USA

Received: 6 April 1996 / Accepted: 3 March 1997

Abstract. Considerable controversy has surrounded the application of mitochondrial DNA data to reconstruction of evolutionary relationships among the endemic cichlids of Lake Malawi. Central to this debate has been the issue of whether lineage sorting is complete, and thus whether these data actually reflect species phylogeny, or simply gene genealogy. Review of all mtDNA control region sequences available for members of one monophyletic subset of this species flock, the Malawi rockfishes, or *mbuna*, strongly indicates that lineage sorting is incomplete: Character-based analyses of these sequences reconstruct gene, not species, interrelationships. Analysis of the pattern of nucleotide substitutions differentiating these mtDNA alleles suggests that pyrimidine residues undergo transition substitutions more often than do purines. Estimation of the magnitude of derived sequence differentiation in light of the reconstructed gene genealogy suggests that the *mbuna* may be of considerably more recent vintage than previous molecular characterizations have indicated.

Key words: *mbuna* species — Lake Malawi — Mitochondrial DNA control region

Introduction

Lake Malawi cichlid fishes comprise the largest and most diverse known assemblage of vertebrates. Perhaps as many as 1,000 species inhabit the lake, most of them undescribed (Ribbink et al. 1983; Lewis et al. 1986; Reinthal 1990; Konings 1990). These taxa provide a spectacular example of adaptive radiation, spanning trophic guilds that include benthic algae grazers, sand-dwelling planktivores, pelagic piscivores, ambush predators, and specialized paedophages (Fryer and Iles 1972; Futuyma 1986; Eccles and Trewavas 1989). They provide a unique model system with which to investigate the evolutionary process, and in particular mechanisms of speciation (Echelle and Kornfield 1984; Keenlyside 1991; Ricklefs and Schluter 1993; Turner 1994). Of particular interest to speciation theorists are the rockfishes or *mbuna* (Fryer 1959), a monophyletic group of lithophilous, algae-grazing endemics (Moran et al. 1994). They are a diverse and abundant assemblage of more than 200 species in 10 genera that feed and breed primarily on algae-covered rocks and cobbles. Evolutionary aspects of *mbuna* biology have been reviewed in Keenlyside (1991) and by Meyer (1993) and Turner (1994).

A body of comparative molecular evidence convincingly suggests that many East African cichlid species are of extremely recent origin. Analyses of allozymes (Kornfield 1978) and mitochondrial (mt) DNA (Meyer et al. 1990; Moran and Kornfield 1993, 1995; Bowers et al. 1994) have revealed exceeding slight differentiation among large numbers of biological species, with estimates of mtDNA sequence divergence similar to those associated with conspecific comparisons in other fishes (e.g., Avise 1994).

The rapidity of cichlid radiations has suggested to many that some exceptional mechanism(s) may be involved in their diversification; explanations have included habitat complexity (Temple 1969), lake-level

fluctuation (Fryer 1959; Trewavas 1935), mutation (Fryer and Iles 1972), reproductive characteristics (Kosswig 1947; Dominey 1984), and trophic polymorphism (Sage and Selander 1975). Similarly, allopatric (Greenwood 1984; McKaye and Gray 1984), microallopatric (Fryer 1959) and sympatric (Kosswig 1963; McKaye 1991; Turner 1994) modes of speciation have been suggested. Given their rapid evolution, polygynandry (Kellogg et al. 1995; Parker and Kornfield 1996a), and diverse breeding coloration, there has been extensive speculation about the role of sexual selection in speciation of Malawi cichlids (Dominey 1984; McElroy and Kornfield 1990; McKaye 1991; Turner 1994). While there is general consensus that almost all of the Malawi fauna originated within the lake proper (Mayr 1984), little progress has been made in elucidating the actual causes of their flabbergasting diversity. Several critical issues cannot be addressed conclusively in the absence of a reliable phylogenetic context within which to frame and then test the appropriate hypotheses.

Reconstructing the evolutionary history of the *mbuna* by molecular means has proved extremely difficult, owing principally to the recency of their radiation (Kornfield 1991; Kornfield and Parker 1997). Isozyme analyses have provided compelling evidence of reproductive isolation among extremely similar forms (Kornfield 1978; McKaye et al. 1982, 1984), but isozymes are insufficiently variable for reconstruction of evolutionary relationships. The advent of mtDNA analysis in evolutionary studies (reviewed in Avise 1994) offered hope that the basic problems of cichlid phylogenetics would be resolved. Comparative studies of higher-level relationships within the Malawi cichlid fauna by restriction enzyme (Moran et al. 1994) and DNA sequence analysis (Meyer 1993) of mtDNA provided unambiguous definition of portions of the endemic radiation. In particular, the basic dichotomy between *mbuna* and most of the remaining haplochromines was supported, as were the affinities of a few well-differentiated lineages. Despite these successes, however, resolution of relationships within the *mbuna* remained problematic.

Using restriction fragment-length polymorphism (RFLP) analysis, Moran and Kornfield (1993) identified four distinct mtDNA lineages within the *mbuna* (designated α, $\beta1$, $\beta2$, and $\beta3$), differing by an average of 1.5% sequence divergence, but the affinities of the taxa involved seemed aberrant; several species were inferred to be more closely related to taxa in other genera than to congeners. An apparent solution to this problem was obtained when sample sizes were increased: The *mbuna* retain an ancestral mtDNA polymorphism (Moran and Kornfield 1993, 1995). Extensive mtDNA restriction analysis demonstrated that both α and β mtDNA alleles were present in a number of common *mbuna* species. Moran and Kornfield (1993) therefore concluded that relationships among these taxa based on mtDNA data reflect a gene tree (Pamilo and Nei 1988) in which lineage sorting (Avise et al. 1984) is incomplete, rather than a phylogenetic or species tree.

A subsequent study employing sequence analysis of the rapidly evolving mtDNA control region (d-loop), however, argued against this conclusion. Bowers et al. (1994) studied extensive collections of two congeneric species, *Melanochromis auratus* and *M. heterochromis,* from several locations in southern Lake Malawi. They also obtained sequence data for two outgroup taxa: *Pseudotropheus zebra* (another *mbuna*) and *Tramitochromis liturus* (a more distantly related non-*mbuna* Malawi cichlid). These data were analyzed in an explicitly phylogenetic context, using nucleotide substitutions as cladistic characters for inference of hierarchical patterns of relationship within and among species. In their analysis, all *M. auratus* alleles were united to the exclusion of other taxa, and significant hierarchical structure was detected among *M. heterochromis* alleles, structure that was consistent with the spatial distribution of populations (Bowers et al. 1994: their Fig. 4). Particularly compelling was a transversion substitution which they reported to be synapomorphic for all six *M. heterochromis* alleles studied (nucleotide position 323 in their Fig. 3). As preliminary sequence data gathered in our laboratory suggested that these six *M. heterochromis* alleles included members of three of the ancestral lineages defined by Moran and Kornfield (1993), the presence of this unique, unambiguous synapomorphy stood in direct contradiction to the retained polymorphism hypothesis. Additionally, Bowers et al. (1994) observed significantly less sequence variation within and among species of *Melanochromis* than between *Melanochromis* and either outgroup taxon, as would be expected if mtDNA sequences did reflect organismal phylogeny. They suggested that their results probably did reflect species, rather than gene, interrelationships, concluding that ''additional sequencing and the development of nuclear markers will be necessary to resolve the question of retained ancestral polymorphisms.''

In an attempt to bring this issue to a convincing state of closure, we here review all available *mbuna* mtDNA control region sequences with the explicit purpose of evaluating the competing hypotheses that these data reflect either gene genealogy (Moran and Kornfield 1993) or species phylogeny (Bowers et al. 1994).

Materials and Methods

Twenty-four *mbuna* mtDNA control region sequences were retrieved from GenBank, and ten others were taken directly from the print literature; authorship and accession numbers are given in Table 1. An additional 25 sequences were determined by polymerase chain reaction (PCR) amplification using primers TDK-A and TDK-E (Lee et al. 1995). Nineteen of the samples sequenced were ultrapurified mitochondrial DNA preparations previously assayed by restriction endonuclease digestion; these were explicitly selected to include samples displaying

Table 1. Taxa and collection data for sequences examined; each entry represents a distinct allele X taxon combination

Taxon	Code	*N*	Location	Authors[a]	GenBank #
Pseudotropheus zebra BB 1[b]	PZ1	3	Thumbi Is. W	PK	U90761
P. zebra BB 2[b]	PZ2	1	Mumbo Is.	PK	U90768
P. zebra BB 3[b]	PZ3	1	Nkhata Bay	PK	U90777
P. zebra red dorsal[b]	PR1	1	Nakantenga Is.	PK	U90760
P. zebra gold	PG1	1	Nkhata Bay	PK	U90770
P. tropheops black 1[b]	PT1	2	Nkhata Bay	PK	U90763
P. tropheops black 2[b]	PT2	1	Nkhata Bay	PK	U90771
P. tropheops black 3	PT3	3	Likoma Is.	RM	—
P. tropheops orange chest[b]	PO1	2	Thumbi Is. W	PK	U90778
P. tropheops orange chest	PO2	4	Otter Point	RM	—
P. tropheops orange chest	PO3	3	Zimbawe	RM	—
P. tropheops red cheek	PD1	3	Likoma Is.	RM	—
P. tropheops membe	PE1	4	Likoma Is.	RM	—
P. tropheops yellow chin	PY1	3	Likoma Is.	RM	—
P. tropheops broad mouth	PM1	2	Otter Point	RM	—
P. tropheops chinyamwezi	PH1	3	Chinyamwezi Is.	RM	—
P. cf. *gracilior*	PC1	2	Otter Point	RM	—
P. cf. *microstoma*	PS1	2	Otter Point	RM	—
P. barlowi 1[b]	PB1	1	Maleri Is.	PK	U90764
P. barlowi 2[b]	PB2	1	Maleri Is.	PK	U90776
P. xanstomachus 1[b]	PX1	1	Maleri Is.	PK	U90762
P. xanstomachus 2[b]	PX2	1	Maleri Is.	PK	U90765
P. williamsi 1[b]	PW1	1	Nakantenga Is.	PK	U90772
P. williamsi 2[b]	PW2	1	Nakantenga Is.	PK	U90775
Melanochromis auratus 1	MA1	47	Lake Malawi[c]	BSK	U01926
M. auratus 2	MA2	9	Chidunga Rocks	BSK	U01927
M. auratus 3	MA3	5	Chidunga Rocks	BSK	U01928
M. auratus 4	MA4	5	Nakantenga Is.	BSK	U01929
M. auratus 5	MA5	1	Mitande	BSK	U01930
M. heterochromis 1	MH1	11	Lake Malawi[d]	BSK	U01936
M. heterochromis 2	MH2	13	Mumbo Is.	BSK	U01937
M. heterochromis 3	MH3	1	Mumbo Is.	BSK	U01938
M. heterochromis 4[e]	MH4	16	Thumbi Is. W	BSK/PK	U01939
M. heterochromis 5	MH5	1	Chinyamwezi Is.	BSK	U01940
M. heterochromis 6	MH6	1	Chinyamwezi Is.	BSK	U01941
M. heterochromis 7[f]	MH7	1	Domwe Is.	PK	U90780
M. parallelus 1	MP1	1	Likoma Is.	SK	U01953
M. parallelus 2	MP2	1	Likoma Is.	PK	U90769
M. parallelus 3	MP3	1	Likoma Is.	PK	U90782
M. simulans	MN1	1	Masinje	SK	U01944
M. melanopterus	MT1	1	Mazinzi Reef	SK	U01952
M. johanni	MJ1	1	Masinje	SK	U01942
M. sp. black + white johanni	MW1	1	Likoma Is.	SK	U01935
M. sp. lepidophage	ML1	1	Masinje	SK	U01943
M. sp. blotch 1	MC1	1	Chisumulu Is.	SK	U01932
M. sp. blotch 2	MC2	1	Masinje	SK	U01933
M. sp. maingano	MM1	1	Likoma Is.	SK	U01951
M. sp. blue	MB1	1	Mpanga	SK	U01934
M. sp. slab	MS1	1	Maleri/Mbenji Is.	SK	U01945
Labeotropheus fuelleborni 1[b]	LF1	1	Mbenji Is.	PK	U90766
L. fuelleborni 2[b]	LF2	1	Mbenji Is.	PK	U90773
L. fuelleborni 3	LF3	1	Nkhata Bay	PK	U90774
Iodotropheus sprengerae[b]	IS1	1	Chinyamwezi Is.	PK	U90767
Genyochromis mento[b]	GM1	1	Thumbi Is. W	PK	U90779
Lethrinops gossei[b]	LG1	1	Monkey Bay	PK	U90781
Cyathochromis obliquidens[b]	CO1	1	Thumbi Is. W	PK	U90759
Petrotilapia sp.	PTR	1	Lake Malawi	KEA	U01113
Tramitichromis liturus	TLT	1	Lake Malawi	BSK	—

[a] Sequences were published as follows: KEA, Kocher et al. (1993); BSK, Bowers et al. (1994); SK, J.R. Stauffer and T.D. Kocher (unpublished data submitted to GenBank); RM, Reinthal and Meyer (1997); PK, Parker and Kornfield, this paper

[b] Individuals for which mtDNA alleles were previously defined using restriction analysis (Moran and Kornfield 1993)

[c] Allele found at Domwe Is., Thumbi Is. East, Namalenge Is., Mazinzi Reef, and Chidunga Rocks

[d] Allele found at Domwe Is., Zimbawe Reef, and Mumbo Is

[e] Includes sample mhet575 (see text)

[f] Sequence determined from sample mhet251 (see text)

P. zebra 'red dorsal'	PR1	ccgggctctgccttccatgtaaacgcaatgcatatatgtattaacaccattattttatat
P. zebra 'BB' 3	PZ3	....
P. xanstomachus 1	PX1	....
P. tropheops 'black' 1	PT1	....
M. johanni	MJ1	....
M. parallelus 1	MP1	....
M. heterochromis 6	MH6	....
P. barlowi 1	PB1	....
M. sp. 'black and white'	MW1	....
M. sp. 'lepidophage'	ML1	....
P. tropheops 'red cheek'	PD1	....
P. tropheops 'black' 3	PT3	....
P. tropheops 'membe'	PE1	....
P. xanstomachus 2	PX2	....
L. fuelleborni 1	LF1	....
I. sprengerae	IS1	....
M. sp. 'blotch' 2	MC2	....
P. zebra 'BB' 2	PZ2	....
M. sp. 'blotch' 1	MC1	....
M. parallelus 2	MP2	....
P. zebra 'gold'	PG1	....
M. sp. 'maingano'	MM1	....
C. obliqudens	CO1	..............t...
P. tropheops 'black' 2	PT2	..............t...
L. fuelleborni 2	LF2	..............t...
P.tropheops 'yellow chin'	PY1	..............t...
P. williamsi 2	PW2	..............t...
M. melanopterus	MT1	..............t...
M. heterochromis 4	MH4	..............t...
P. williamsi 1	PW1	....
P. barlowi 2	PB2	....
L. fuelleborni 3	LF3	....
M. auratus 1	MA1	....
P. zebra 'BB' 1	PZ1	....
P. tropheops 'orange chest' 1	PO1	....
P. tropheops 'orange chest' 2	PO2	....
P. tropheops 'broad mouth'	PM1	....
P. cf. *gracilior*	PC1	....
P. cf. *microstoma*	PS1	....
P. tropheops 'orange chest' 3	PO3	....
P. tropheops 'chinyamwezi'	PH1	....
G. mento	GM1	....a.....
M. auratus 2	MA2	....
M. auratus 3	MA3	....
M. auratus 4	MA4	....
M. auratus 5	MA5	....
M. heterochromis 1	MH1	....
M. heterochromis 2	MH2	........................a...................................
M. heterochromis 3	MH3	........................a...................................
M. heterochromis 5	MH5	....
Petrotilapia sp.	PTR	....
M. heterochromis 7	MH7	........................a...................................
M. sp. 'blue'	MB1	..............t...
M. simulans	MN1	....
M. sp. 'slab'	MS1	....
L. gossei	LG1	....
T. liturus	TLT	....

Fig. 1. Aligned control region sequences for 58 mtDNA alleles observed among 180 individuals of 34 species. Nucleotides identical to the top sequence are indicated by *dots;* alignment gaps by *dashes.*

each of the mtDNA lineages described by Moran and Kornfield (1993; see Table 1). Four additional DNA samples were prepared by proteinase K digestion of ethanol-preserved muscle tissue from recently collected specimens, followed by organic extraction and ethanol precipitation. To resolve the issue of an apparently synapomorphic nucleotide substitution in all reported *M. heterochromis* sequences (Bowers et al. 1994; see above), we also sequenced two DNA samples from this taxon, designated mhet251 and mhet575, which were kindly provided by T. D. Kocher. Although these particular samples had not previously been sequenced, based upon collection locality they were predicted to exhibit alleles MH1 and MH4 (Fig. 1), respectively (T. D. Kocher, personal communication). One discrepancy was noted between a published sequence (*M. auratus* allele MA2 (Bowers et al. 1994)) and the corresponding GenBank submission (U01927): Nucleotide position 302 (as numbered by Bowers et al. (1994); see position 219 in Fig. 1) is given as a C in their Fig. 3, but as an A in the GenBank submission. Since this substitution constitutes a single autapomorphic difference from the consensus sequence, we suspect it was inadvertently omitted from the figures of Bowers et al. (1994); in our analyses we use the sequence as reported in GenBank.

PCR amplifications were carried out in 20-μl volumes using standard buffer conditions (AmpliTaq, Perkin Elmer-Cetus) and 35 cycles

```
PR1  caaacatatcctatatattaatacatatcatttataaaaacatagacaaatataccacatatttgtttaaaccattttaactaaagggtacat
PZ3  .............................................................................................
PX1  .............................................................................................
PT1  .............................................................................................
MJ1  .............................................................................................
MP1  .............................................................................................
MH6  .............................................................................................
PB1  .............................................................................................
MW1  ................................................................................t............
ML1  .............................................................................................
PD1  .............................................................................................
PT3  .............................................................................................
PE1  .............................................................................................
PX2  .............................................................................................
LF1  .............................................................................................
IS1  .............................................................................................
MC2  .............................................................................................
PZ2  .............................................................................................
MC1  ..............................................t..............................................
MP2  .............................................................................................
PG1  ...........................-.................................................................
MM1  ..................................c..........................................................
CO1  .............................................................................................
PT2  .............................................................................................
LF2  .................................c...........................................................
PY1  .............................................................................................
PW2  .................................c...........................................................
MT1  .............................................................................................
MH4  .............................................................................................
PW1  ..................................c..........................................................
PB2  ..................................c..........................................................
LF3  ..................................c..........................................................
MA1  ..................................c..........................................................
PZ1  ..................................c..........................................................
PO1  ..................................c..........................................................
PO2  ..................................c..........................................................
PM1  ..................................c..........................................................
PC1  ..................................c..........................................................
PS1  ..................................c..........................................................
PO3  ..................................c..........................................................
PH1  ..................................c..........................................................
GM1  .............................................................................................
MA2  ..................................c..........................................................
MA3  ...........................c......c..........................................................
MA4  ..................................c..........................................................
MA5  ..................................c..........................................................
MH1  ..................................c....t.....................................................
MH2  ..................................c..........................................................
MH3  ..................................c....t.....................................................
MH5  ..................................c..........................................................
PTR  t..........................c................................cc........................a......
MH7  ..................................c..........................................................
MB1  ..................................c..........................................................
MN1  ..................................c..........................................................
MS1  .........................c........c..........................................................
LG1  ..................................c..........................................................
TLT  ..................................c..........................................................
```

Fig. 1. Continued.

of 94°C, 30 s; 50°C, 30 s; 72°C, 60 s. Sequencing templates were prepared via asymmetric PCR, carried out in 50-μl volumes; PCR conditions were identical to the above except for 50-fold dilution of one primer. Template DNA was purified with celullose acetate ultrafiltration columns (Millipore) and sequenced via standard dideoxy chain-termination protocols (Sequenase v 2.0, USB), using α^{35}S-labeled dATP. Sequencing reaction products were electrophoresed for 3 h through 6% polyacrylamide gels; subsequent to electrophoresis, gels were dried onto chromatography paper and exposed to X-ray film for 1 to 3 days. Sequences were read by eye from autoradiographs and entered into ESEE (Cabot and Beckenbach 1989) for manual alignment; all nucleotide positions which varied from the consensus sequence were rechecked against autorads or GenBank submissions. The published control region sequences we obtained, and those we determined, were of varying lengths; subsequent to alignment, sequences were truncated at one or both ends, leaving 337 bp of aligned sequence for all 58 alleles. In addition to equalizing the number of nucleotide positions in all sequences, this procedure removed portions of some sequences which extended into the proline tRNA (5′ end; Meyer et al. (1990)) or the central conserved block (3′ end; Shields and Kocher (1991)); thus all nucleotide positions examined lie in the "hypervariable" segment between these two functionally conserved regions.

```
PR1   aaaccataattgaatattc-tccaataaatacccattaattactaaacgatagtttaagaccgatcacaactctcactagttaagttatacca
PZ3   ...................-.........................................................................
PX1   ...................-.........................................................................
PT1   ...................-.........................................................................
MJ1   ...................-.........................................................................
MP1   ...................-.........................................................................
MH6   ...................-.........................................................................
PB1   ...................-.........................................................................
MW1   ...................-.........................................................................
ML1   ...................-....................c....................................................
PD1   ...................-.........................................................................
PT3   ...................-.....c...................................................................
PE1   ...................-.........................................................................
PX2   .........c.........-.........................................................................
LF1   .........c.........-.........................................................................
IS1   .........c.........-.........................................................................
MC2   .........c.........-.........................................................................
PZ2   .........c.........-.........................................................................
MC1   .........c.........-.........................................................................
MP2   .........c.........-.........................................................................
PG1   .........c.........-.........................................................................
MM1   .........c.........-.........................................................................
CO1   .........c.........-.........................................................................
PT2   .........c.........-.........................................................................
LF2   .........c......c..-.........................................................................
PY1   .........c.........-.........................................................................
PW2   .........c......c..-.........................................................................
MT1   .........c......c..-............tt...........................................................
MH4   .........c......c..-.............t...........................................................
PW1   .........c.........-.............t...........................................................
PB2   .........c.........-.............t...........................................................
LF3   .........c.........-.............t...........................................................
MA1   .........c.........-.............t...........................................................
PZ1   .........c.........-.............t......c....................................................
PO1   .........c.........-.............t...........................................................
PO2   .........c.........-.............t...........................................................
PM1   .........c.........-.............t...........................................................
PC1   .........c.........-.............t...........................................................
PS1   .........c.........-............tt...........................................................
PO3   .........c.........-.............t...........................................................
PH1   .........c.........-.............t...........................................................
GM1   .........c.........-.............t...........................................................
MA2   .........c.........-.............t...............................a...........................
MA3   .........c.........-.............t...........................................................
MA4   .........c.........-.............t...........................................................
MA5   .........c.........-.............t...........................................................
MH1   ...................-.............t...........................................................
MH2   .........c.........-.............t...........................................................
MH3   .........c.........-.............t...........................................................
MH5   .........c.........-...........t.t...........................................................
PTR   .........c.........-...........t.t.c.........................................................
MH7   .........c.........-.............t...........................................................
MB1   .........c.........-...........t.t...........................................................
MN1   .........c.........c.............t...........................................................
MS1   .........c.........-.............t...........................................................
LG1   .........c.........-.............t...........................................................
TLT   .........c.........-.............t...........................................................
```

Fig. 1. Continued.

To test the competing hypotheses described above, we compared relationships among mtDNA alleles deduced in two distinct ways. Because some sequence alleles were present in multiple individuals of single taxa, and others in individuals of different taxa, and due to the fact that the number of individuals sampled per taxon varies widely in our data set (Table 1), the OTUs we consider differ between the two analyses. First, a maximum parsimony tree connecting the alleles was constructed using PAUP v3.1.1 (Swofford 1993); this hypothesis of relationship is presented as a spanning tree rather than a more conventional terminally labeled "phylogenetic" tree to emphasize that many alleles are most parsimoniously assigned to internal nodes. In this tree, which relates alleles without reference to the taxa in which they occur, some OTUs include multiple taxa exhibiting identical alleles while other taxa are represented by multiple alleles, and thus multiple OTUs. This hypothesis of relationship will subsequently be referred to as the "gene tree." Second, we considered the hypothesis that these sequence data might reflect evolutionary relationships among *mbuna* species (Bowers et al. 1994) by constructing terminally labeled trees, again using PAUP v 3.1.1 (Swofford 1993). All searches were conducted under a constraint tree which required both species and genera (including the *Pseudotropheus tropheops* species complex) to be monophyletic, assumptions which are well supported by morphological

```
PR1  agtacccaccatcctattcattacccatatttaatgtagtaagagcccaccatcagttgattccttaatgtcaacggttcttgaaggtcaag
PZ3  ............................................................................................
PX1  ............................................................................................
PT1  ............................................................................................
MJ1  ............................................................................................
MP1  ............................................................................................
MH6  ............................................................................................
PB1  ............................................................................................
MW1  ............................................................................................
ML1  ............................................................................................
PD1  ............................................................................................
PT3  ...........c................................................................................
PE1  ............................................................................................
PX2  ............................................................................................
LF1  ............................................................................................
IS1  ............................................................................................
MCI  ............................................................................................
PZ2  ............................................................................................
MC1  .........................t..................................................................
MP2  ............................................................................................
PG1  ............................................................................................
MM1  ............................................................................................
CO1  ............................................................................................
PT2  ............................................................................................
LF2  ............................................................................................
PY1  ............................................................................................
PW2  ............................................................................................
MT1  ............................................................................................
MH4  ............................................................................................
PW1  ............t...............................................................................
PB2  ............t...............................................................................
LF3  ............t...............................................................................
MA1  ............t...............................................................................
PZ1  ............t...............................................................................
PO1  ............t........c......................................................................
PO2  ............t.......c.......................................................................
PM1  ............t.......c.......................................................................
PC1  ............t.......c.......................................................................
PS1  ............t...............................................................................
PO3  ............t...............................................................................
PH1  ............t.......c.......................................................................
GM1  ............t...............................................................................
MA2  ............t...............................................................................
MA3  ............t...............................................................................
MA4  ............t................................................c...........t..................
MA5  ............t.....t.........................................................................
MH1  ............t...............................................................................
MH2  ............t.c..........................................................t..................
MH3  ............t.c.............................................................................
MH5  ............................................................................................
PTR  ..............................................................t.............................
MH7  ............t.......................................................................g.......
MB1  ............t...............................................................................
MN1  ............t...............................................................................
MS1  ............t...............................................................................
LG1  ............................................................................................
TLT  ............................................................................................
```

Fig. 1. Continued.

data (Fryer and Iles 1972; Ribbink et al. 1983; Reinthal 1987); this phylogenetic hypothesis is subsequently referred to as the "species tree." No other assumptions regarding hierarchical relationships of alleles, species, or genera were imposed. Since this analysis considers relationships among taxa, all allele × taxon combinations were treated as distinct OTUs. All variable nucleotide positions were treated as unordered, equally weighted characters. Due to the large number of OTUs exhibiting either identical or autapomorphic alleles, enumeration of all possible minimum-length trees was not feasible. To ensure that we correctly determined the minimum possible treelength for the data set, ten independent heuristic searches were conducted using the TBR branch-swapping algorithm of PAUP v3.1.1, each beginning with a different random number seed. During each search, 1,000 minimum-length trees were allowed to accumulate (MAXTREES = 1,000), and branch-swapping was conducted on all of them; subsequently, one of the 1,000 minimum-length trees from each search was randomly selected for reconstruction of character-state changes (nucleotide substitutions) using the minimum F-value option. In both analyses, the non-*mbuna* Malawi cichlid *Tramitochromis liturus* was employed as an outgroup.

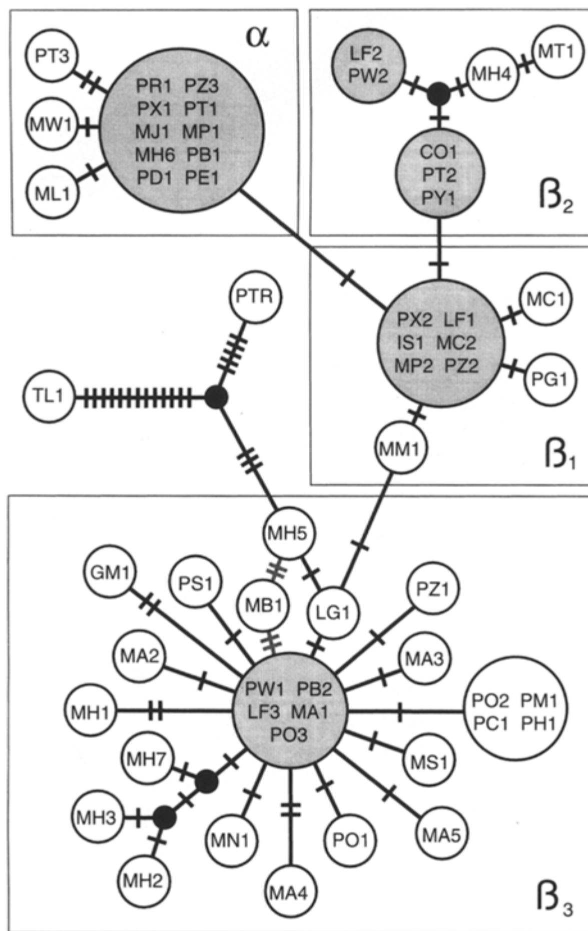

Fig. 2. Maximum parsimony relationships among 58 mtDNA alleles (see Table 1). Each *crossbar* represents one nucleotide substitution. The area of each *open circle* is proportional to the number of species in which that allele occurs. *Small black circles* represent intermediate alleles not observed in the present data set. *Gray circles* denote sequence alleles shared among multiple genera, and therefore inferred to be retained from the common ancestor of the *mbuna*.

Results and Discussion

Gene Tree or Species Tree?

Aligned control region sequences are presented in Fig. 1. Forty-six nucleotide positions (13.61%) are variable, resulting in a total of 58 alleles, 47 of which occur in a single taxon only. Sequences determined from the *Melanochromis heterochromis* samples provided by T. D. Kocher differ from published sequences in one very important way: the T-A transversion reported at position 323 of Bowers et al. (1994; position 239 in Fig. 1) was not present in either of the two individuals examined. Sample mhet 575 is otherwise identical to the published MH4 allele, as predicted, while sample mhet 251 exhibits a novel allele, here designated MH7 (Table 1, Fig. 1). We suspect that the reported synapomorphic transversion constitutes either a data editing error or an artefact of the automated cycle-sequencing procedure used to generate the data presented by Bowers et al. (1994); in our experience, this procedure is somewhat more error-prone than manual, T7 DNA polymerase-based sequencing, especially when sequencing PCR amplimers. Regardless of its origin, the absence of this putatively synapomorphic character compromises the conclusions of Bowers et al. (1994). In light of these observations we have deleted the T-A transversion from all *M. heterochromis* alleles for our analyses (Fig. 1, position 239).

Two maximum parsimony trees were found; they are shown in Fig. 2. These hypotheses differ only in the immediate affinity of allele MB1 (gray lines in Fig. 2); each requires 60 steps, including 49 transition substitutions, nine transversion substitutions, one insertion, and one deletion; this hypothesis requires inference of four intermediate sequence alleles which were not observed in our data set (Fig. 2). All but one of the *mbuna* alleles fell into one of four clusters, each of which is characterized by one or two common alleles that are shared among members of several genera (shaded circles in Fig. 2). Importantly, these four clusters, here identified by analysis of control region sequences, correspond exactly to the four ancestral lineages hypothesized by Moran and Kornfield based on whole-molecule RFLP data: All alleles for which we have both sequence and RFLP data ($N = 17$) fall into the same clusters regardless of which data set is used. We have therefore assigned our sequences to these four ancestral lineages (boxes in Fig. 2), maintaining the terminology established by Moran and Kornfield (1993). The only allele not assigned to one of the four lineages is that found in the sole individual of *Petrotilapia* sequenced (allele PTR). This allele lies opposite all the other sequences, across the root defined by outgroup taxon *T. liturus* (Fig. 2); as such it represents a sister lineage to all other observed *mbuna* mtDNA alleles.

Each heuristic search conducted under the monophyly constraint tree (see Materials and Methods) produced 1,000 unrooted minimum-length trees of 102 steps each; during each of the ten searches, over 500,000 topological rearrangements were performed by the TBR branch-swapping algorithm. While not as conclusive as an exact search (Swofford 1993), this result nonetheless provides compelling evidence that a minimum of 102 steps are required to reconcile these data with a phylogenetic hypothesis requiring monophyletic species and genera. One 102-step tree was selected at random from the 1,000 saved during each search for reconstruction of character-state changes. The number and type of substitutions inferred at each variable site were identical among all ten minimum-length trees (Table 2).

Thus the gene geneology hypothesis is, from the maximum parsimony perspective, strongly favored over the species phylogeny interpretation (60 vs 102 steps). On this basis we assert that the sequence data reported here corroborate the conclusion of Moran and Kornfield (1993) that the *mbuna* retain an ancestral genetic poly-

Table 2. Distribution of number of substitutions per site versus negative binomial expectations, as inferred under competing hypotheses of relationship[a]

	Number of sites			
Number of substitutions	Gene tree	Negative binomial	Species tree	Negative binomial
0	293	293	293	293
1	36	35.67	31	23.6
2	6	7.74	6	9.47
3	4	1.92	2	4.82
4	0		1	2.71
5	0		0	1.61
6	0		0	0.99
7	0	0.70	0	0.62
8	0		2	0.40
9	0		1	0.26
10+	0		2	0.52
	$P = 0.292$ (1 df)		$P = 0.046$ (4 df)	

[a] Expected values were pooled to produce cell values greater than 1 for contingency table tests

Table 3. Nucleotide frequencies (percent) and base compositional biases

Nucleotide	All control region	Variable control region	Fourfold degenerate[a]
G	10.1	10.2	4.5
A	36.8	33.4	25.6
T	31.7	34.0	25.7
C	21.5	22.4	44.2
%GC	31.6	32.6	48.7
G − C skew[b]	−0.361	−0.374	−0.815
A − T skew[b]	0.075	0.009	−0.002

[a] Data from Kocher et al. (1995); these values differ significantly from both control region columns ($\chi^2 = 38.85$, 3 df, $P << 0.001$)
[b] G − C skew calculated as (G − C)/(G + C), A − T skew as (A − T)/(A + T)

morphism that predates evolution of the species flock, which obviates use of mtDNA for cladistic reconstruction of relationships among *mbuna* species. We find the presence of several common alleles shared among multiple genera (shaded circles in Fig. 2), each associated with a number of rarer, derived alleles associated with single taxa, to be particularly indicative of the presence of an ancestral polymorphism. Nevertheless, several alternative explanations might be advanced.

It might be contended that the species phylogeny hypothesis does in fact reflect evolutionary relationships among these sequences, and that these relationships are obscured by the presence of mutational "hot spots"—nucleotides characterized by such accelerated substitution rates that they have undergone multiple, convergent, and back mutations even over the brief time period during which these sequences have diverged. In the present case, the five nucleotide positions inferred to have experienced from eight to ten substitutions on the species tree (Fig. 1; Table 2) are the obvious candidates for such mutational hot spots. Two arguments can be advanced against this interpretation. First, as noted above, the pattern of relationship indicated by cladistic analysis of control region sequences is entirely concordant with that derived from whole-molecule RFLP analysis. Therefore if control region mutational hot spots have engendered a misleading pattern of relationship, similar hot spots throughout the entire *mbuna* mitochondrial genome must have independently mutated to produce a convergent, misleading pattern. We find a coincidence of this magnitude exceedingly improbable. Second, although there clearly is variation in mutation rate among control region nucleotides, as demonstrated by the excellent fit of the number of substitutions per site on the gene tree to the negative binomial distribution (Table 2), the magnitude of variation required to engender the distribution of substitutions per site reconstructed on the species tree is improbable. It has lately been popular to model variation in control region mutation rate using the gamma distribution (e.g., Tamura and Nei 1993; Wakeley 1993; Lyrholm et al. 1996), whereby the magnitude of rate variation is embodied in the gamma distribution's shape parameter, α. Smaller values of α denote greater variation; α = infinity corresponds to no variation, in which case the Poisson and negative binomial distributions are identical. The value of α based on the gene tree hypothesis is 0.416, which is similar to values reported for humans ($\alpha = 0.47$ (Wakeley 1993)) and for other, noncichlid teleosts ($\alpha = 0.31$ (Parker and Kornfield, submitted)). Acceptance of the species tree hypothesis, however, implies that the five most highly variable control region nucleotides (Table 2; above) exhibit a mutation rate more than 70 times greater than that of the remaining nucleotide positions, or, from the continuous distribution perspective, $\alpha = 0.077$; these values are unprecedented in the literature describing control region sequence evolution.

If we do accept the gene tree hypothesis as an appropriate account of the evolutionary history of these sequences, there still remains a potential alternative explanation for their incongruent distribution with regard to species relationships: hybridization. This explanation is less likely than the ancestral polymorphism hypothesis for two reasons. First, the observed distribution of alleles among taxa would imply hybridization not just among closely related species, but among morphologically and ecologically divergent *genera,* some of which inhabit widely disjunct parts of Lake Malawi. (For example, alleles PR1 and PE are identical, but were sampled from species whose ranges are separated by hundreds of kilometers.) Despite the tremendous effort that has been directed at photographic documentation of the Malawi cichlid fauna, the only documented case of hybridization in natural populations of *mbuna* involved a single putative hybrid between two very closely related species of the *P. zebra* complex (Konings 1990). While we cannot

Table 4. Substitution frequencies

Substitution	Number observed	Number expected[a]
G – T	0	5.0
G – C	0	3.3
A – T	5	16.2
A – C	4	10.7
T – C	42	10.9
G – A	7	4.9

[a] Expected number of substitutions corrected for observed base composition of variable sites

Table 5. Estimated mtDNA allele coalescence times[a]

Allele group	N[b]	Coalescence time (years)	Standard deviation
All	11.74	1,884,000	528,000
alpha	2.67	541,000	283,000
beta 1	2.18	517,000	276,000
beta 2	2.00	443,000	256,000
beta 3	2.84	568,000	290,000
P. zebra BB	3.33	640,000	308,000
P. tropheops black	2.67	542,000	283,000
P. tropheops orange chest	1.33	344,000	225,000
M. auratus	2.00	443,000	256,000
M. heterochromis	5.43	950,000	375,000
L. fuelleborni	4.00	739,000	331,000

[a] Values for the four ancestral lineages (α, β1, β2, and β3) estimate the age of the *mbuna* species flock; note that values for alleles segregating in individual species are often larger, emphasizing the widespread absence of lineage sorting. Mutation rate is assumed to be 1×10^{-8} substitutions · base pair^{-1} · year^{-1}

[b] Mean number of substitutions differentiating alleles in each group

absolutely rule out the possibility of hybridization in the distant past, there is no evidence to indicate the level of present-day intergeneric gene flow that would be required to explain our observations. For these reasons, we believe the ancestral polymorphism hypothesis to be the best supported and most reasonable explanation for the phenomenon we document here.

Nucleotide Composition and Compositional Biases

Mean nucleotide frequencies (Table 3) were calculated from all observed alleles; individual sequences differ only slightly, if at all, owing to the small number of substitutions differentiating them; *mbuna* control region sequences are characterized by AT-richness (%GC = 31.46), negative G-C skew (–0.3586), and very minor A-T skew (0.0736), as are control region sequences in other fishes (Parker and Kornfield 1996b) and fourfold degenerate codon positions in mitochondrial protein coding genes of many animal taxa (Perna and Kocher 1995). The most striking difference between *mbuna* control region sequences and fourfold-degenerate codon positions lies in the frequency of guanine residues. These account for 10.1% of nucleotide positions in these sequences vs a mean of 4.8% of fourfold-degenerate sites in several mitochondrial genes of nine vertebrate taxa (Perna and Kocher 1995) and 4.5% of such sites in the NADH dehydrogenase 2 (ND2) gene of a variety of cichlid taxa (Kocher et al. 1995). Similarly, an average of only 2.48% of fourfold-degenerate sites in shark cytochrome *b* genes are occupied by guanine residues (Martin 1995). This relative excess of guanine residues accounts for the less pronounced G-C skew observed in *mbuna* control region sequences (vs, for instance, –0.815 for fourfold-degenerate sites in cichlid ND2 genes (Kocher et al. 1995)).

This difference in base composition presents a challenge to the assumption that all sites in the nonfunctional portion of the mitochondrial control region are unconstrained by selection and thus free to vary in response to the fundamental mutational matrix, as is inferred to be the case for fourfold degenerate codon positions (Perna and Kocher 1995). Since nucleotide positions for which there are obvious reasons to posit stabilizing selection (e.g., first and second codon positions (Kocher et al. 1995; Martin 1995) and structurally conserved elements of mitochondrial ribosomal RNA genes (Parker and Kornfield 1996b)) exhibit higher percentages of guanine, one possible interpretation is that some control region guanine residues are subject to selective constraint. If this is the case, base composition of *variable* control region nucleotides should more closely resemble that of fourfold-degenerate sites. Because the sequences we report here represent a relatively small sample of variable nucleotides (N = 40), we instead used a published alignment of more divergent African cichlid mtDNA control region sequences (Kocher et al. 1993) to estimate the composition of variable positions (N = 117 nucleotides). This estimate is essentially identical to the overall composition of these sequences and is significantly more G-rich than are fourfold-degenerate positions (Table 3).

Patterns of Nucleotide Substitution

Different types of nucleotide substitutions occurred with very different frequencies. Transitions were on the whole 5.45 times as common as transversions (49 TS: 9 TV), in agreement with numerous studies of the evolutionary dynamics of mitochondrial DNA (e.g., Brown et al. 1986). The four transversion types occurred in proportions roughly concordant with observed base composition (Table 4): only A-T and A-C substitutions were observed, reflecting the numerical dominance of A over G residues. Frequencies of the two types of transition substitution, however, differ substantially from frequency-adjusted null expectations derived using the "revised EOR" method of Collins et al. (1994). Six times as many substitutions interchanged pyrimidines as interchanged purines, although pyrimidines are only slightly more

common (53.12%), representing a frequency-adjusted 2.7:1 excess of pyrimidine over purine substitutions. This phenomenon has not previously been reported for animal mtDNA control regions, although it has for protein coding regions; for instance, at all codon positions of the cytochrome *b* gene of carcharhiniform sharks, pyrimidine transitions accounted for over 80% of all nucleotide substitutions (Martin 1995). In some instances, apparent biases in substitution frequency may be due to inappropriate reconstruction of substitutions on hypothesized phylogenetic trees. Perna and Kocher (1995) critiqued the work of Tamura and Nei (1993), suggesting that if different substitution types occur with very different frequencies, maximum parsimony reconstructions may underrepresent the most rapid substitution type by reconstructing a single, synapomorphic substitution on an internal branch rather than two convergent substitutions on terminal branches. In the present case, it is possible to modify the phylogenetic tree of Fig. 2 to alter the G-A transition at position 25 from a synapomorphy to two separate, independent substitutions. This rearrangement eliminates a convergent pyrimidine transition at position 101, but at the cost of an additional T-C transition at position 259, so the overall bias in transition frequency is essentially unaffected.

Relative Levels of Sequence Divergence

Although independent temporal calibration of a "molecular clock" is unlikely in the absence of a reliable fossil record (as in the case of East African cichlids), comparison of levels of divergence in homologous sequence regions is still of utility for understanding the relative ages of related groups of organisms. Previous estimates of intra-*mbuna* sequence divergence (Moran and Kornfield 1993; Bowers et al. 1994) have employed global comparison of all observed sequence or restriction alleles. In both cases, alleles belonging to multiple mtDNA lineages were present in the data from which these estimates were derived. These estimates thus confound derived sequence differentiation with that which, under the ancestral polymorphism hypothesis, predates the common ancestor of this fauna. We estimated global coalescence times for all of our control region sequences and then estimated the age of the common ancestor of the modern *mbuna* species flock for each ancestral allele group independently (Table 5). Coalescence times were calculated as described by Templeton (1993), assuming a mutation rate of 1×10^{-8} substitutions · base pair^{-1} · year^{-1}. Mutation rates of this magnitude have been estimated for primate mtDNA using fossil evidence (Ruvolo et al. 1996) and for control region sequences in other teleost groups (Parker and Kornfield, submitted). Thus, although we have no cichlid-specific evidence to support the choice of this particular mutation rate, its generality across other taxa suggests that it may not be too far off base. Because coalescence times are a linear function of assumed mutation rate, the values given in Table 5 can be adjusted for any desired mutation rate via multiplication by the appropriate correction factor.

Coalescence time estimates for each of the four hypothesized ancestral mtDNA lineages are satisfyingly similar (Table 5), suggesting an age of approximately 500,000 years for the ancestral *mbuna* species, while the ancestor of all alleles still segregating among members of the species flock is substantially older, perhaps as old as Lake Malawi itself. Thus in Malawi cichlid species there obtains a situation diametrically opposed to that seen in many studies of human mtDNA diversity (e.g., Templeton 1993; Ruvolo 1996): In humans, the common ancestor of all extant mtDNA alleles is thought to be much more recent than the origin of modern *Homo sapiens,* while in the *mbuna,* alleles that arose long before the origin of any extant species are still segregating in those species. There are at least two competing explanations for our observation that relatively ancient alleles are still segregating in many *mbuna* species. First, diversifying selection may be acting on *mbuna* mtDNA, as has been suggested to be the general case for histocompatibility loci (Hughes and Nei 1989). Although natural selection on mtDNA variants has been demonstrated in several taxa (e.g., Rand et al. 1994; Templeton 1996), we are unaware of any evidence supporting diversifying selection. Another possibility is that the observed mtDNA variation is essentially neutral, and thus that the presence of highly divergent alleles segregating within species is indicative of large effective population sizes. The mean time to fixation of a neutral allele (which is the same process as allele coalescence, viewed from the opposite temporal perspective) is $4N_e$ for diploid loci (Kimura and Ohta 1969); for haploid, maternally inherited mtDNA this value becomes $2N_f$, where N_f is the effective number of females. Assuming a generation time of 2 years, estimates of N_f range from 110,000 to 237,000 for the six species in which we have sampled three or more alleles (Table 5). These values are quite large, although perhaps not implausible for 10-cm fishes inhabiting a very large lake, especially in light of their polygynandrous mating system (Kellogg et al. 1995; Parker and Kornfield 1996a), which tends to elevate effective population sizes (Sugg and Chesser 1994). More importantly, since we have concluded that most of the genetic variation we observe in the *mbuna* species flock predates the origin of present-day *mbuna* species, this means that large population sizes were maintained *through* the speciation events giving rise to the species we observe today. This implies that several previously hypothesized speciation mechanisms that depend on transiently small population sizes, including founder effects (Carson and Templton 1984; Moran and Kornfield 1995) and runaway sexual selection (Dominey 1984; Turner 1994), are probably not of general utility for understanding the genesis of Malawi cichlid diversity.

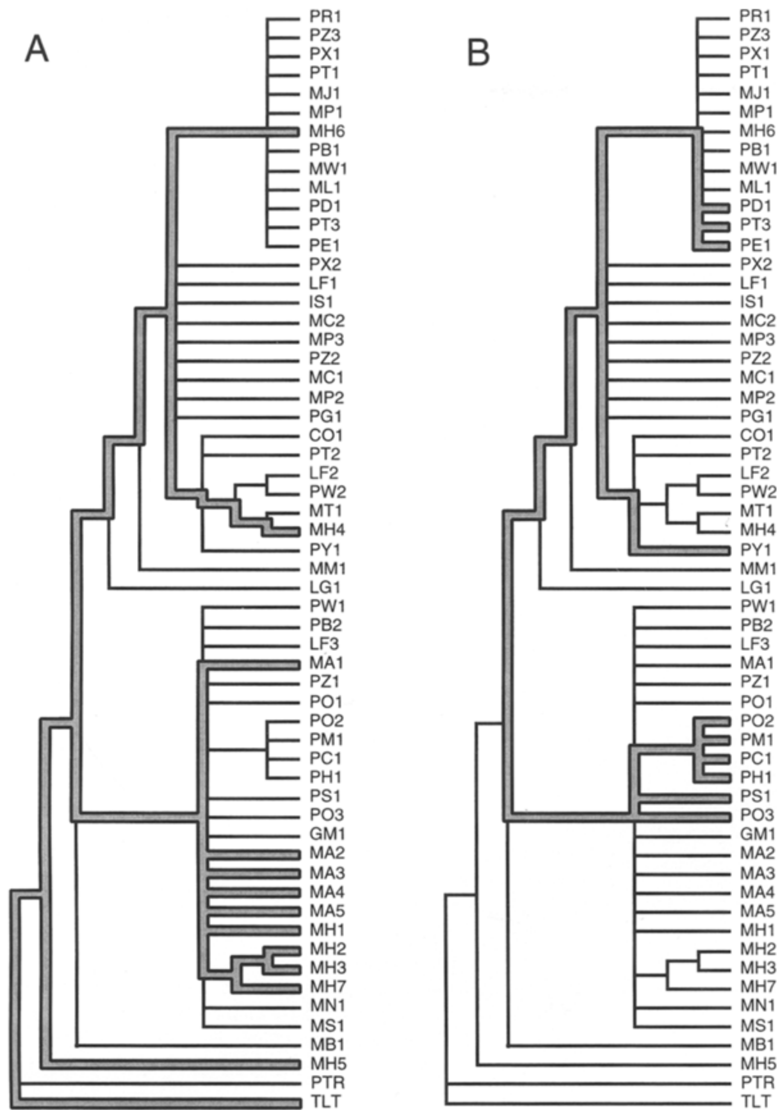

Fig. 3. Ancestral polymorphisms present a substantial hazard when interpretating data sets of limited scope. **A** The results of Bowers et al. (1994) superimposed on a consensus cladogram representing the two MP trees derived from our combined data (see Fig. 2). From this topology *(shaded)* they concluded that *M. auratus (MA1–MA5)* was most likely descended from a *M. heterochromis*–like ancestor *(MH1–MH7)*. In light of the additional data assembled here, this line of reasoning would imply that essentially the entire *mbuna* species flock must be descended from one or another population of *M. heterochromis*. **B** The results of Reinthal and Meyer (1997); they hypothesized that members of the *P. tropheops* species complex sampled in the southern part of Lake Malawi *(PD1, PT3, PE1, PY1)* must be derived from a single colonist species, as they form a monophyletic sister group to species sampled in the northern part of the lake *(PO2, PM1, PC1, PS1)*. This interpretation implies that various members of *Melanochromis*, *Labeotropheus*, *Cyathochromis*, and *Iodotropheus* are also descended from that colonizing *P. tropheops* species.

Conclusions

Our results allow convincing rejection of the hypothesis that relationships among *mbuna* mtDNA sequences reflect organismal phylogeny and confirm the hypothesis that the distribution of mtDNA alleles amongst these species is a consequence of retention of multiple alleles present in a polymorphic ancestral taxon (Moran and Kornfield 1993). Although this situation precludes the application of character-based methods of phylogenetic reconstruction to nucleotide substitutions in *mbuna* mtDNA, the fact that many taxa retain multiple alleles (above; Moran and Kornfield 1995) argues that these polymorphisms, when analyzed appropriately, may be of considerable utility in understanding the population biology, and perhaps even systematics, of the *mbuna* species flock.

More generally, the observation of almost universally incomplete mtDNA lineage sorting in a taxon consisting of at least 200 nominal species underscores the extreme importance of large sample sizes in phylogenetic studies of very closely related organisms; that this dictum applies to outgroup as well as ingroup taxa is clearly illustrated by our reinterpretation of the data of Bowers et al. (1994). This is especially true in cases like the one considered here, where speciation is suspected to have been

extremely rapid, and the potential exists for maintenance of large effective population sizes, which are expected to reduce the effects of genetic drift and retard lineage sorting. The potential pitfalls associated with attempts to interpret such data in a strictly phylogenetic context are clearly illustrated by superimposition of the phylogenetic hypotheses of Bowers et al. (1994) and Reinthal and Meyer (1997) on our more comprehensive data set (Fig. 3A,B, respectively). It is apparent that these authors' conclusions, which seemed perfectly reasonable in light of their restricted data sets, are highly unlikely when considered in this larger context. Based on the topology shown in Fig. 3A, Bowers et al. (1994) concluded that *M. auratus* (MA1–MA5) was most likely descended from a *M. heterochromis*-like ancestor (MH1–MH7). In light of our additional data, this line of reasoning would imply that essentially the entire *mbuna* species flock must be descended from one or another population of *M. heterochromis.* Figure 3B shows the results of Reinthal and Meyer (1997); they hypothesized that members of the *P. tropheops* species complex sampled in the southern part of Lake Malawi (PD1, PT3, PE1, PY1) must be derived from a single colonist species, as they form a monophyletic sister group to species sampled in the northern part of the lake (PO2, PM1, PC1, PS1). Again, this interpretation is unlikely, implying that various members of *Melanochromis, Labeotropheus, Cyathochromis,* and *Iodotropheus* are also descended from the hypothetical colonizing *P. tropheops* species.

Our findings regarding the relative rates of purine vs pyrimidine transitions suggest that the standard practice of simply weighting transversions over transitions by some integer factor (e.g., Lehman et al. (1995); Parker and Kornfield (1995), and numerous other studies) for purposes of phylogenetic analysis does not accurately reflect the evolutionary dynamics of these sequences. Again, comparisons involving control region sequence alignments from several other higher animal taxa should be carried out to test the generality of these observations, preferably in concert with analysis of other mtDNA gene regions from the same taxa. This will facilitate understanding of the differences in evolutionary dynamics of fourfold-degenerate codon positions and control region nucleotide residues, both of which are heavily relied upon for purposes of phylogenetic inference. Coalescence times for alleles derived subsequent to the origin of the *mbuna* species flock place the estimated age of this evolutionary radiation at approximately one-half million years, although the large variances associated with these estimates do not preclude a much more recent origin for some taxa, as is supported by geophysical evidence (Owen et al. 1990).

Acknowledgments. We thank T. D. Kocher for donation of DNA samples and P. Moran for critical reading of an earlier draft of this manuscript. We are particularly indebted to T.D. Kocher for his insightful review. This work was supported by the University of Maine/University of New Hampshire Sea Grant (NA36RG0110), the National Science Foundation (DEB 952183), and a University of Maine Center for Marine Studies fellowship to A.P.

References

Avise JC (1994) Molecular markers, natural history and evolution. Chapman and Hall, New York

Bowers N, Stauffer JR, Kocher TD (1994) Intra- and interspecific mitochondrial DNA sequence variation within two species of rock-dwelling cichlids (Teleostei: Cichlidae) from Lake Malawi, Africa. Mol Phyl Evol 3:75–82

Brown GG, Prager EM, Wang A, Wilson AC (1986) Structural conservation and variation in the d-loop containing region of vertebrate mitochondrial DNA. J Mol Biol 192:503–511

Cabot EL, Beckenbach AT (1989) Simultaneous editing of multiple nucleic acid and protein sequences with ESEE. Comput Appl Biosci 5:233–234

Carson HL, Templeton AR (1984) Genetic revolutions in relation to speciation phenomena: the founding of new populations. Annu Rev Ecol Syst 15:97–131

Collins TM, Kraus F, Estabrook G (1994) Compositional effects and weighting of nucleotide sequences for phylogenetic analysis. Syst Biol 43:449–459

Dominey WJ (1984) Effects of sexual selection and life history on speciation: Species flocks in African cichlids and Hawaiian *Drosophila.* In: Echelle AA, Kornfield I (eds) Evolution of fish species flocks. University of Maine Press, Orono, pp 231–250

Eccles DH, Trewavas E (1989) Malawian cichlid fishes: the classification of some haplochromine genera Lake Fish. Movies Herten, West Germany

Echelle AA, Kornfield I (eds) (1984) Evolution of fish species flocks. University of Maine Press, Orono

Fryer G (1959) The trophic interrelationships and ecology of some littoral communities in Lake Nyasa with special references to the fishes, and a discussion of the evolution of a group of rock-frequenting Cichlidae. Proc. Zool Soc Lond 132:153–281

Fryer G, Iles TD (1972) The cichlid fishes of the great lakes of Africa. Oliver Boyd, Edinborough

Futuyma D (1986) Evolutionary biology. Sinauer, Sunderland, MA

Greenwood PH (1984) African cichlids and evolutionary theories. In: Echelle AA, Kornfield IL (eds) Evolution of fish species flocks. University of Maine Press, Orono, pp 141–154

Hughes AL, Nei M (1989) Nucleotide substitution at major histocompatibility complex class II loci: evidence for overdominant selection. Proc Natl Acad Sci USA 86:958–962

Keenlyside M (ed) (1991) Cichlid fishes: behavior, ecology, and evolution. Chapman and Hall, London

Kellogg KA, Markert JA, Stauffer JR, Kocher TD (1995) Microsatellite variation demonstrates multiple paternity in lekking cichlid fishes from Lake Malawi, Africa. Proc R Soc Lond [Biol] 260B:79–84

Kimura M, Ohta T (1969) The average number of generations until fixation of an individual mutant gene in a finite population. Genetics 61:763–771

Kocher TD, Conroy JA, McKaye KR, Stauffer JR (1993) Similar morphologies of cichlid fish in Lakes Tanganyika and Malawi are due to convergence. Mol Phyl Evol 2:158–165

Kocher TD, Conroy JA, McKaye KR, Stauffer JR, Lockwood SF (1995) Evolution of NADH dehydrogenase subunit 2 in East African cichlid fish. Mol Phyl Evol 4:420–432

Konings A (1990) Konings' book of cichlids and all the other fishes of Lake Malawi. TFH, Neptune City, NJ

Kornfield I (1978) Evidence for rapid speciation in African cichlid fishes. Experientia 34:335–336

Kornfield I (1991) Genetics. In: Keenlyside MHA (ed) Cichlid fishes:

behavior, ecology and evolution. Chapman and Hall, London, pp 103–128

Kornfield I, Parker A (1997) Molecular systematics of a rapidly evolving species flock: the *mbuna* of Lake Malawi and the search for phylogenetic signal. In: Kocher TD, Stepien C (eds) Molecular systematics of fishes. Academic Press, New York, pp 25–37

Kosswig C (1947) Selective mating as a factor for speciation in cichlid fish of East African lakes. Nature 159:604–605

Kosswig C (1963) Ways of speciation in fishes. Copeia 1963:238–244

Lee WJ, Conroy J, Kocher TD (1995) Structure and evolution of teleost mitochondrial control regions. J Mol Evol 41:54–65

Lehman N, Pfrender ME, Morin PA, Crease TJ, Lynch M (1995) A hierarchical molecular phylogeny within the genus *Daphnia.* Mol Phyl Evol 4:395–407

Lewis D, Reinthal P, Trendall J (1986) A guide to the fishes of Lake Malawi National Park. World Wildlife Fund Publications, Gland, Switzerland

Lyrholm T, Leimar O, Gyllensten U (1996) Low diversity and biased substitution patterns in the mitochondrial DNA control region of sperm whales: implications for estimates of time since common ancestry. Mol Biol Evol 13:1318–1326

Martin A (1995) Mitochondrial DNA sequence evolution in sharks: rates, patterns, and phylogenetic inferences. Mol Biol Evol 12: 1114–1123

Mayr E (1984) Evolution of fish species flocks: a commentary. In: Echelle AA, Kornfield I (eds) Evolution of fish species flocks. University of Maine Press, Orono, pp 3–12

McElroy DM, Kornfield I (1990) Sexual selection, reproductive behavior, and speciation in the mbuna species flock of Lake Malawi (Pisces: Cichlidae). Env Biol Fishes 28:273–284

McKaye KR (1991) Sexual selection and the evolution of the cichlid fishes of Lake Malawi, Africa. In: Keenlyside MHA (ed) Cichlid fishes: behavior, ecology and evolution. Chapman and Hall, London, pp 241–257

McKaye KR, Gray WN (1984) Extrinsic barriers to gene flow in rock dwelling cichlids of Lake Malawi: microhabitat heterogeneity and reef colonization. In: Echelle AA, Kornfield I (eds) Evolution of fish species flocks. University of Maine Press, Orono, pp 169–184

McKaye KR, Kocher T, Reinthal P, Kornfield I (1982) Sympatric sibling species complex of *Petrotilapia* Trewavas analysed by enzyme electrophoresis (Pisces: Cichlidae). J Linn Soc 76:91–96

McKaye KR, Kocher TD, Reinthal P, Harrison R, Kornfield I (1984) Genetic evidence for allopatric and sympatric differentiation among morphs of a Lake Malawi cichlid fish. Evolution 38:215–219

Meyer A (1993) Phylogenetic relationships and evolutionary processes in East African cichlid fishes. Trends Ecol Evol 8:279–284

Meyer A, Kocher TD, Bassaslowski P, Wilson AC (1990) Monophyletic origin of Lake Victoria cichlid fishes suggested by mitochondrial DNA sequences. Nature 347:550–553

Moran P, Kornfield I (1993) Retention of an ancestral polymorphism in the mbuna species flock (Pisces: Cichlidae) of Lake Malawi. Mol Biol Evol 10:1015–1029

Moran P, Kornfield I (1995) Evidence of a population bottleneck and extreme philopatry in the mbuna species flock (Teleostei: Cichlidae) of Lake Malawi. Mol Biol Evol 12:1085–1093

Moran P, Kornfield I, Reinthal P (1994) Molecular systematics and radiation of the haplochromine cichlids (Teleostei: Perciformes) of Lake Malawi. Copeia 1994:274–288

Nei M (1987) Molecular evolutionary genetics. Columbia University Press, New York

Owen RB, Crossley R, Johnson TC, Tweddle D, Kornfield I, Davison S, Eccles DH, Engstrom DE (1990) Major low levels of Lake Malawi and implication for speciation rates in cichlid fishes. Proc R Soc Lond [Biol] 240B:519–553

Pamilo P, Nei M (1988) Relationships between gene trees and species trees. Mol Biol Evol 5:568–583

Parker A, Kornfield I (1995) A molecular perspective on evolution and zoogeography of cyprinodontid killifishes. Copeia 1995:8–21.

Parker A, Kornfield I (1996a) Polygynandry in *Pseudotropheus zebra,* a cichlid fish from Lake Malawi, Africa. Env Biol Fishes 47:345–352

Parker A, Kornfield I (1996b) An improved amplification and sequencing strategy for phylogenetic studies using the mitochondrial large subunit ribosomal RNA. Genome 39:793–797

Patterson HE (1980) A comment of mate recognition systems. Evolution 34:330–331

Perna NT, Kocher TD (1995) Patterns of nucleotide composition at fourfold degenerate sites of animal mitochondrial genomes. J Mol Evol 41:353–358

Rand DM, Dorfsman M, Kan LM (1994) Neutral and non-neutral evolution of *Drosophila* mitochondrial DNA. Genetics 138:741–756

Reinthal PN (1987) Morphology, ecology, and behavior of a group of the rock-dwelling fishes (Cichlidae: Perciformes) from Lake Malawi, Africa. Unpublished PhD dissertation, Duke University, Durham

Reinthal PN (1990) The living jewels of Lake Malawi. Natl Geogr 177:42–51

Reinthal PN, Meyer A (1997) Genetic tests of speciation models in African cichlid fishes. In: Givnish TJ, Sytsma KJ (eds) Molecular evolution and adaptive radiations. Cambridge University Press, Cambridge (in press)

Ribbink AJ, Marsh BA, Marsh AC, Ribbink AC, Sharp BJ (1983) A preliminary survey of the cichlid fishes of the rocky habitats in Lake Malawi. S Afr J Sci 18:149–310

Ricklefs RE, Schluter D (1993) Species diversity: regional and historical influences. In: Ricklefs RE, Schluter D (eds) Species diversity in ecological communities. University of Chicago Press, Chicago, pp 350–363

Ruvolo M (1996) A new approach to studying modern human origins: hypothesis testing with coalescence time distributions. Mol Phyl Evol 5:202–219

Sage D, Selander RK (1975) Trophic radiation through polymorphism in cichlid fishes. Proc Natl Acad Sci 72:4669–4673

Shields GF, Kocher TD (1991) Phylogenetic relationships of North American ursids based on analysis of mitochondrial DNA. Evolution 45:218–221

Sugg DW, Chesser RK (1994) Effective population sizes with multiple paternity. Genetics 137:1147–1155

Swofford DL (1993) PAUP: phylogenetic analysis using parsimony, ver 3.1.1. Computer program distributed by the Illinois Natural History Survey, Champaign, IL

Tamura K, Nei M (1993) Estimation of the number of nucleotide substitutions in the control region of mitochondrial DNA in humans and chimpanzees. Mol Biol Evol 10:512–526

Templeton AR (1993) The "Eve" hypothesis: a genetic critique and reanalysis. Am Anthropol 95:51–72

Templeton AR (1996) Contingency tests of neutrality using intra/interspecific gene trees: the rejection of neutrality for the evolution of the mitochondrial cytochrome oxidase II gene in the hominid primates. Genetics 144:1263–1270

Trewavas E (1935) A synopsis of the cichlid fishes of Lake Nyasa. Ann Mag Nat Hist 10:65–118

Turner GF (1994) Speciation mechanisms in Lake Malawi cichlids: a critical review. Arch Hydrobiol 44:139–160

Wakeley J (1993) Substitution rate variation among sites in hypervariable region I of human mitochondrial DNA. J Mol Evol 37:613–623