Multilocus Phylogeny of Cichlid Fishes (*Pisces: Perciformes*): Evolutionary Comparison of Microsatellite and Single-Copy Nuclear Loci

J. Todd Streelman, Rafael Zardoya,†[1] Axel Meyer,†[2] and Stephen A. Karl**

*Department of Biology, University of South Florida; and †Department of Ecology and Evolution and Program in Genetics, State University of New York at Stony Brook

Among vertebrates, cichlid fishes are the paradigmatic example of adaptive radiation and ecological specialization. In turn, molecular genetic studies of cichlids have focused primarily on more recently diverged groups. Here, we present an evolutionary hypothesis of the major lineages of cichlid fishes based on DNA sequence data from two nuclear loci. One marker, *Tmo-4C4,* is a single-copy locus containing a region of amino acid similarity to the muscle protein TITIN. Flanking sequence from a second, microsatellite, locus, *Tmo-M27,* shows similarity to mammalian RAS guanine nucleotide-releasing factor. We compare and combine data from these loci to evaluate phylogenetic performance. In separate and combined analyses, the sequence data support and clarify previous morphological hypotheses of cichlid major-group relationships. Indian and Malagasy cichlids form a basal, paraphyletic group. Neotropical cichlids are the sister clade to an African assemblage composed of the paraphyletic west and Pan-African lineages and a group of east African rift lake taxa. We use a consensus phylogeny of the Cichlidae to trace evolutionary changes in the microsatellite repeat motif at *Tmo-M27.* Analysis reveals that the repeat region was nearly lost in the ancestor to cichlids and then amplified extensively in African taxa. Results demonstrate that these two new DNA markers could be widely applied in perciform systematics. Furthermore, the comparative approach can unveil mutational dynamics of simple-sequence repeat loci over long periods of fish evolution. Simple-sequence repeat regions are increasingly being found in introns of important regulatory genes. We address issues involving their function and suggest caution in making assumptions of strict neutrality.

Introduction

The freshwater fish family Cichlidae is a multiform mosaic of ancient and recent lineages. The current distribution throughout Africa, India, Madagascar, and the Neotropics suggests a Gondwana-wide ancestral range (Brown and Gibson 1983) and a pattern of divergence matching the splitting of the continental masses (Stiassny 1987, 1991; fig. 1). In this vicariant scenario, cichlids would have roots dating back at least 130 MYA (Storey 1995). In contrast to the relative antiquity of the family, the apparent recent and rapid evolution within some genera has attracted widespread attention (e.g., Fryer and Iles 1972; Meyer et al. 1990; Goldschmidt 1996). For example, in east Africa, the 300-plus species of the haplochromine flock of Lake Victoria are believed to share a common ancestor less than 200,000 years ago (Meyer et al. 1990; Johnson, Scholz, and Talbot 1996; Kaufman, Chapman, and Chapman 1997). Like Galapagos finches and Hawaiian drosophilids, east African cichlids are a familiar example of adaptive radiation and trophic specialization (Futuyma 1986). Evolutionists in general (Charlesworth, Lande, and Slatkin 1982; Raup and Jablonski 1986) and cichlid specialists in particular (Dominey 1984; Ribbink 1990) have long sought connections between micro- and macro-level phenomena that might explain trends in comparative diversity. Central to understanding these patterns and processes of cichlid fish evolution are reliable estimates of the phylogenetic relationships of ancient as well as recent groups within the family.

To date, most molecular genetic studies of cichlids have focused on the more intermediate timescales of divergence of 2–20 MYA (e.g., Nishida 1991; Klein et al. 1993; Sturmbauer and Meyer 1993; Moran, Kornfield, and Reinthal 1994; Sturmbauer, Verheyen, and Meyer 1994; Kocher et al. 1995). Allozyme and mitochondrial (mt)DNA data have provided phylogenetic resolution among and sometimes within the cichlid fish groups of the African rift lakes Tanganyika, Malawi, and Victoria. However, theoretical considerations (Pamilo and Nei 1988; Ball, Neigel, and Avise 1990) and empirical limitations (Kornfield 1978; Moran and Kornfield 1993; Meyer 1994; Kocher et al. 1995) have prompted the hunt for additional markers. To sharpen the resolution of previous studies of cichlids and other taxa, molecular evolutionists have initiated a search of ''alternative gene space'' (Graybeal 1994; Sultmann et al. 1995; Orti and Meyer 1996; Zardoya et al. 1996).

Previously, we reported a molecular phylogenetic analysis of the major groups of cichlid fishes (Africa, India, Madagascar, and the Neotropics) using microsatellite flanking regions (MFRs; Zardoya et al. 1996). Here, we present data from a single-copy nuclear (scn)DNA locus for many of the same taxa included in Zardoya et al. (1996). We compare the evolutionary dynamics and phylogenetic performance of the MFRs with those of the scnDNA locus using the same set of species. In addition, we employ a consensus phylogenetic hypothesis for the Cichlidae to trace patterns of evolution in the microsatellite repeat region. This evolutionary framework is useful in resolving the mutational patterns of simple-sequence repeats (Jin et al. 1996; Rico, Rico,

[1] Present address: Museo Nacional de Ciencias Naturales, Madrid, Spain.

[2] Present address: Department of Biology, University of Konstanz, Konstanz, Germany.

Key words: microsatellite, single-copy nuclear DNA, molecular evolution, cichlid evolution, multilocus phylogeny, TITIN, RAS-GRF.

Address for correspondence and reprints: Stephen A. Karl, Department of Biology, SCA 110, University of South Florida, Tampa, Florida 33620-5150. E-mail: karl@chuma.cas.usf.edu.

Mol. Biol. Evol. 15(7):798–808. 1998

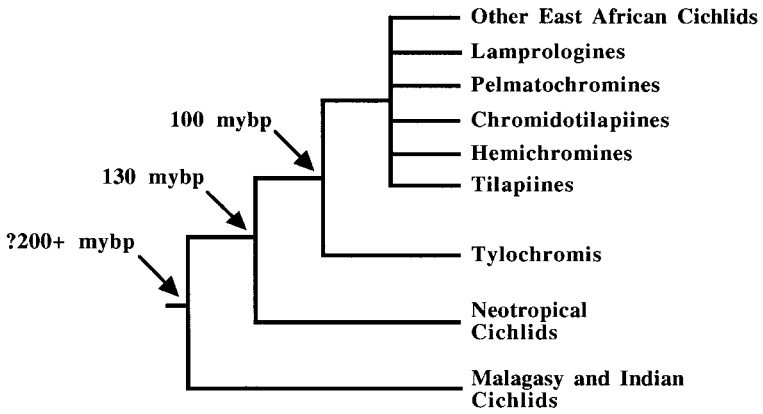

FIG 1.—Intrafamilial phylogenetic hypothesis of cichlid fishes based on morphological characters (Stiassny 1991). Dates at nodes indicate approximate time of separation of land masses based on the geological record (Storey 1995).

and Hewitt 1996; Messier, Li, and Stewart 1996; Orti, Pearse, and Avise 1997).

Materials and Methods

Library Construction and Marker Isolation

The taxa of fishes included in this study are listed in table 1. Isolation of scnDNA loci follows Karl and Avise (1993). Briefly, a genomic DNA library was constructed from a single individual of *Tropheus moorii*(Cichlidae). Cloned insert size was estimated by polymerase chain reaction (PCR) amplification (Saiki et al. 1985; Mullis et al. 1986) of the recombinant clones using universal M13 sequencing primers (Gussow and Clackson 1989). The genomic copy number of 272 clones was determined by dot blot hybridization with radiolabeled total-cell DNA. To identify clones containing microsatellites, a radiolabeled $(GT)_8$ probe was used to screen dot blots, and positive clones were sequenced. The flanking regions of at least one microsatellite clone, *Tmo-M27*, proved useful in cichlid phylogenetic analysis (Zardoya et al. 1996). Here, all sequence position references for *Tmo-M27* are relative to the alignment presented as figure 1 in Zardoya et al. (1996).

The sequences of an additional 16 single-copy clone inserts (usually 200–600 nt from each end) were determined using Sequenase T7 DNA polymerase sequencing kits (U.S. Biochemicals). Primers suitable for PCR amplification were designed with the aid of the computer program OLIGO 4.0 (National Biosciences, Inc.) and were synthesized by CyberSyn Inc. (Camden, N.J.). Primers made to one clone, *Tmo-4C4*, amplified genomic DNA from a number of cichlid fish species spanning the global range. Primers from other cloned sequences yielded successful amplifications of select African cichlids (unpublished data). Nucleotide and inferred polypeptide sequences from both *Tmo-4C4* and *Tmo-M27* were used as queries in sequence similarity searches of GenBank using default parameters of BLAST (Altschul et al. 1990; Gish and States 1993). Putative open reading frames (ORFs) in the nucleotide sequences were identified using the program DNA Strider 1.0 (Commissariat a l'Energie Atomique, France).

DNA Amplification and Sequencing of the *Tmo-4C4* Locus

Total-cell DNA was extracted from individual fish following standard phenol/chloroform procedures (Ausubel et al. 1993) and amplified with PCR primers specific for locus *Tmo-4C4*. Fifty-microliter amplification reactions contained 2 μl total-cell DNA, 2.5 mM $MgCl_2$, 1 × reaction buffer (Promega), 5 μg bovine serum albumin, 12.5 pmol of each primer, and 1.25 U *Taq* DNA polymerase (Promega). Cycling parameters were 1 cycle of 2 min at 95°C and 30 cycles of 30 s at 95°C, 30 s at 55°C, and 1 min at 72°C, with a final extension of 7 min at 72°C. After PCR amplifications, 5 μl of the reaction mix was assayed for the amount and fidelity of amplification by agarose gel electrophoresis. Free nucleotides and unused primers from successful amplifications were removed by centrifugal filtration with Millipore Ultrafree-MC (30,000 NMWL) filter units. Both strands of the resulting purified DNAs were sequenced in one of two ways. Sequences were obtained either from direct sequencing of double-stranded PCR products or from clones of amplified DNA (TA cloning kit; Invitrogen). In both cases, sequencing was performed using an ABI automated DNA sequencer. Sequences for all taxa and *Tmo-4C4* primers have been deposited in GenBank under accession numbers U70326–U70333 and U70335–U70361. Sequences for *Tmo-M27* are published in Zardoya et al. (1996) and are available in GenBank (U63654–U63680).

Sequence Alignment and Phylogenetic Analyses

The *Tmo-4C4* sequences were aligned by eye using the computer program SeqEd (Applied Biosystems). Gaps were included to increase sequence similarity, and when appropriate, to maintain putative ORFs. For *Tmo-4C4*, only the *Perca* sequence required gapping that consisted of a single, 3-nt region (i.e., one putative amino acid, positions 325–327; Streelman and Karl 1997). This gap was treated as missing data in all subsequent analyses. *Tmo-M27* alignments were as in Zardoya et al. (1996). Phylogenetic analyses used *Tmo-4C4* scnDNA data alone or in combination with the MFR of *Tmo-M27*. When combined, the microsatellite repeat region

Table 1
Taxa of Fishes Included in this Study and Geographic Information for the Cichlidae

Order Perciformes
- Percidae
  - ***Perca fluviatilis***
- Centropomidae
  - *Lates niloticus*
- Embiotocidae
  - *Amphistichus rhodoterus*
  - ***Micrometrus minimus***
  - ***Damalichthys vacca***
  - *Cymatogaster aggregata*
- Pomacentridae
  - ***Abudefduf saxatilis***
  - ***Dascyllus trimaculatus***
  - ***Dascyllus arnanus***
- Labridae
  - *Halichoeres maculipinna*
  - *Sparisoma chrysopterum* (Haiti)
  - *Sparisoma chrysopterum* (Florida)
  - *Sparisoma radians*
- Cichlidae
  - India
    - ***Etroplus maculatus***
  - Madagascar
    - ***Paretroplus polyactis***
    - *Oxylapia polli*
    - *Ptychochromoides betsileanus*
  - Neotropics
    - ***Astronotus ocellatus***
    - ***Crenicichla saxatilis***
    - *Cichla ocellaris*
    - *Cichlasoma citrinellum*
  - West Africa
    - ***Pelvicachromis pulcher***
    - ***Hemichromis bimaculatus***
  - Pan Africa
    - *Oreochromis leucostictus*
    - *Tilapia zillii*
    - ***Tylochromis polylepis***
    - *Oreochromis niloticus*
  - East Africa
    - ***Boulengerochromis microlepis***
    - *Julidochromis regani*
    - *Lamprologus brichardi*
    - *Neolamprologus compressiceps*
    - ***Astatotilapia calliptera***
    - ***Serranochromis robustus***
    - ***Tropheus moorii***
    - *Cunningtonia longiventralis*
    - *Enantiopus melanogenys*
    - *Cyprichromis leptosoma*
    - ***Labidochromis caereuleus***
    - ***Astatoreochromis alluaudi***
    - *Haplochromis* sp.
    - *Pseudotropheus tropheops*
    - *Pseudocrenilabrus multicolor*
    - *Chalinochromis brichardi*

NOTE.—Taxa in bold were also included in Zardoya et al. (1996). Underlined taxa were only in Zardoya et al. (1996).

of *Tmo-M27* was omitted (as in Zardoya et al. 1996), and gaps were treated as missing data.

Sequence data analysis and bootstrapping were performed using PAUP (heuristic searches with random addition and MULPARS in effect; Swofford 1993). Since previous results suggest that alternative weighting schemes have little effect on tree topology and resolution, transitions, transversions, and codon positions were given equal weight in all parsimony analyses (Zardoya et al. 1996; Streelman and Karl 1997).

Neighbor-joining analyses were performed using PHYLIP. For locus *Tmo-4C4*, we used maximum-likelihood distances with empirical base frequencies (programs DNADIST and NEIGHBOR of PHYLIP). This scheme was chosen because sequences at locus *Tmo-4C4* displayed nucleotide frequency biases (see *Results*), and maximum-likelihood distances are believed to be more accurate than the Kimura two-parameter model under these conditions (Felsenstein 1989). Bootstrapping of neighbor-joining trees was conducted using the SEQBOOT, DNADIST, NEIGHBOR, and CONSENSE programs of PHYLIP. Combined *Tmo-4C4* and *Tmo-M27* data were analyzed using both parsimony and distance approaches as above (see figure legends for details). The program MacClade (Maddison and Maddison 1992) was used to estimate tree-based parameters such as the number of substitutions per nucleotide site and the number of changes at first, second, and third codon positions. All tree-building analyses with *Tmo-4C4* and combined data used *Perca fluviatilis* (Percidae) as an outgroup.

Results

Molecular characterization of *Tmo-4C4*

A single ORF starting at position 1 in the aligned sequences and extending the entire 511 nt of locus *Tmo-4C4* was identified and used to aid in sequence alignment (Streelman and Karl 1997). Nucleotide sequence similarity searches of GenBank resulted in no probable matches. Searches with the inferred polypeptide sequence, however, revealed several genes of high similarity. One third, or approximately 60 amino acids from the 5′ end of *Tmo-4C4*, matched several regions of TITIN (accession number X90568) and TITIN-like proteins likely to be homologs (e.g., UNC-89 [U33058], CONNECTIN [D83008], PERLECAN [M85289], TWITCHIN [X15423]). Notably, each of these proteins has multiple immunoglobulin (IG) domains believed to function in the elasticity of muscle contraction (at least for TITIN; Labeit and Kolmerer 1995; Linke et al. 1996). Individual IG domains within a polypeptide, however, are quite variable at the amino acid level. When *Tmo-4C4* is compared with human TITIN, the region of highest similarity contains 24 (34.8%) identical and 15 (21.7%) charge-conservative amino acids over 69 residues. Other regions in human TITIN, mostly IG domains, display lower levels of similarity.

Over all fish sequences, the nucleotide composition of *Tmo-4C4* has a slight bias in cytosine (19.61%) and adenine (28.45%). There are 57 first, 49 second, and 149 third variable putative codon positions. When calculated over the most parsimonious tree, third-position substitutions outnumber first-position substitutions 5:1 and second-position substitutions 6:1. There are 80 variable

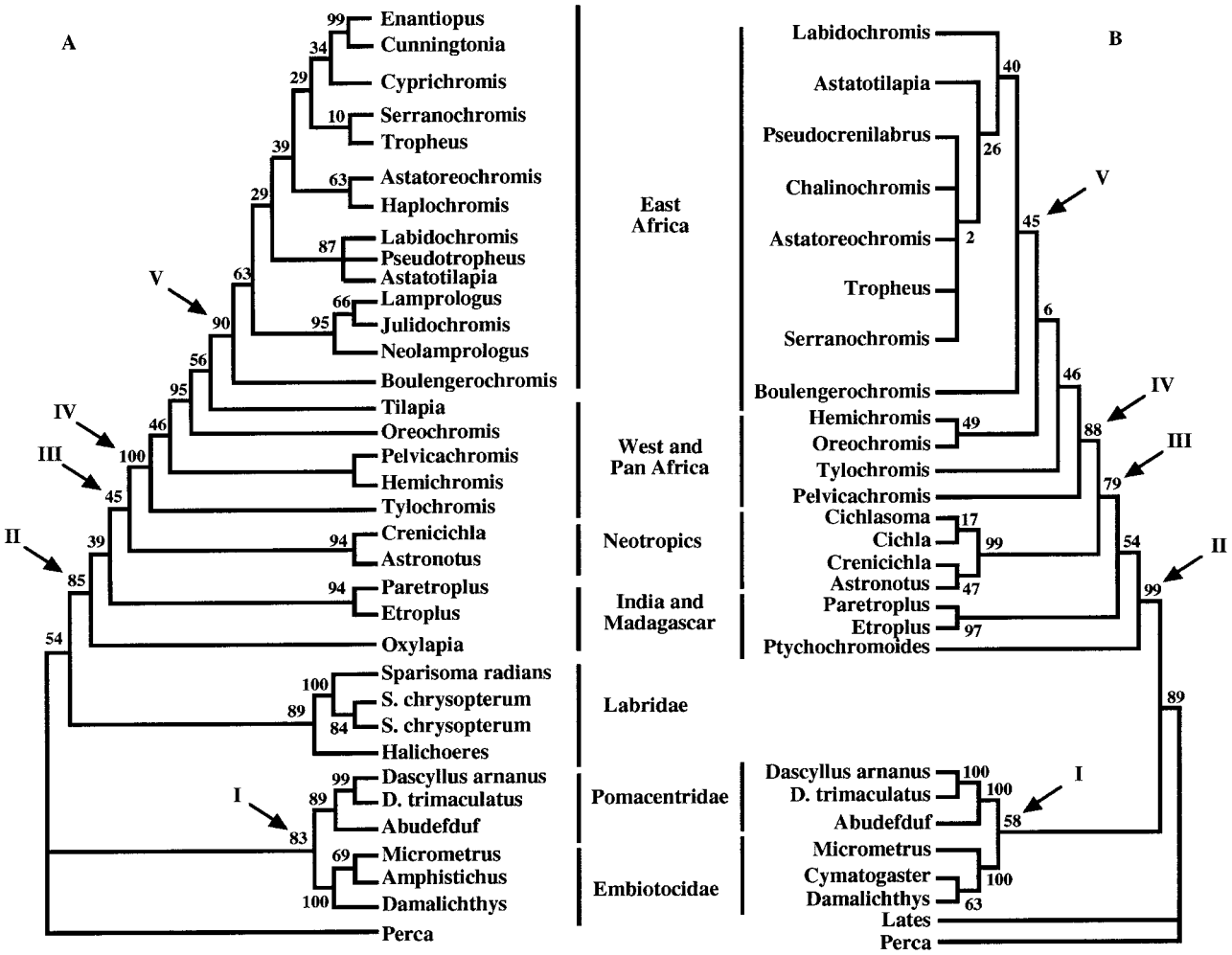

FIG. 2.—Fifty percent majority-rule maximum-parsimony bootstrap consensus tree (and groups compatible with it) for loci (*A*) *Tmo-4C4* and (*B*) *Tmo-M27* (Zardoya et al. 1996). Numbers at nodes represent the support in 200 bootstrap replicates for *Tmo-4C4* and 1,000 replicates for *Tmo-M27*. *Perca* was used as an outgroup in the *Tmo-4C4* analysis, and both *Perca* and *Lates* were used as outgroups at *Tmo-M27*. All taxon names are as in table 1. Roman numerals indicate nodes referred to in text: I—embiotocids and pomacentrids, II—Cichlidae, III—division of Neotropical and African cichlids, IV—African cichlids, V—east African cichlids.

(of 170 total) putative amino acid positions in the 35 taxa at *Tmo-4C4*.

Molecular Characterization of *Tmo-M27*

The *Tmo-M27* sequences contain ORFs starting at position 459 (i.e., −1 frame of the 3′ end) and extending for at least 45 amino acids in the 5′ direction of the aligned sequence shown in Zardoya et al. (1996). All sequences in this region can be translated in both the −1 and −2 reading frames. *Abudefduf* can be translated in all three reverse-compliment reading frames resulting in inferred polypeptides of 64–94 amino acids in length. Notably, each ORF in all taxa can span the microsatellite region. However, a potential 5′ GT intron-splicing junction (Long, deSouza, and Gilbert 1995) is present, suggesting that the microsatellite motif is located in an intron. In support of this premise, no significant ORFs were found in the 5′ end sequence of *Tmo-M27*. Sequence similarity searches using both DNA and inferred polypeptides resulted in several significant matches to the −1 reading frame of the 3′-end 78 nt (26 putative amino acids). Nearly all matches were to RAS-specific guanine nucleotide-releasing factor (RAS-GRF). *Tmo-M27* was 92%–96% identical to mouse (U67326), rat (P28818), and human (S62035) RAS-GRF homologs.

In the DNA sequence of the 3′ ORF, there are 1 first, 4 second, and 12 third variable putative codon positions over all fish sequences. Third-position substitutions outnumber first-position substitutions 16:1 and second-position substitutions 4:1 when calculated over the topology shown in figure 2B. Over 27 fish taxa, there are three variable putative amino acid positions in this region of *Tmo-M27*.

Phylogenetic Analysis of scnDNA and MFR Loci

Phylogenetic analyses of both *Tmo-4C4* and *Tmo-M27* loci are in strong agreement in the placement of most lineages (fig. 2). Embiotocids and pomacentrids are sister groups in separate, as well as in combined, analyses (node I in figs. 2, 3, and 4). Indian and Malagasy cichlids appear as paraphyletic and basal members of a monophyletic Cichlidae (node II). Neotropical cich-

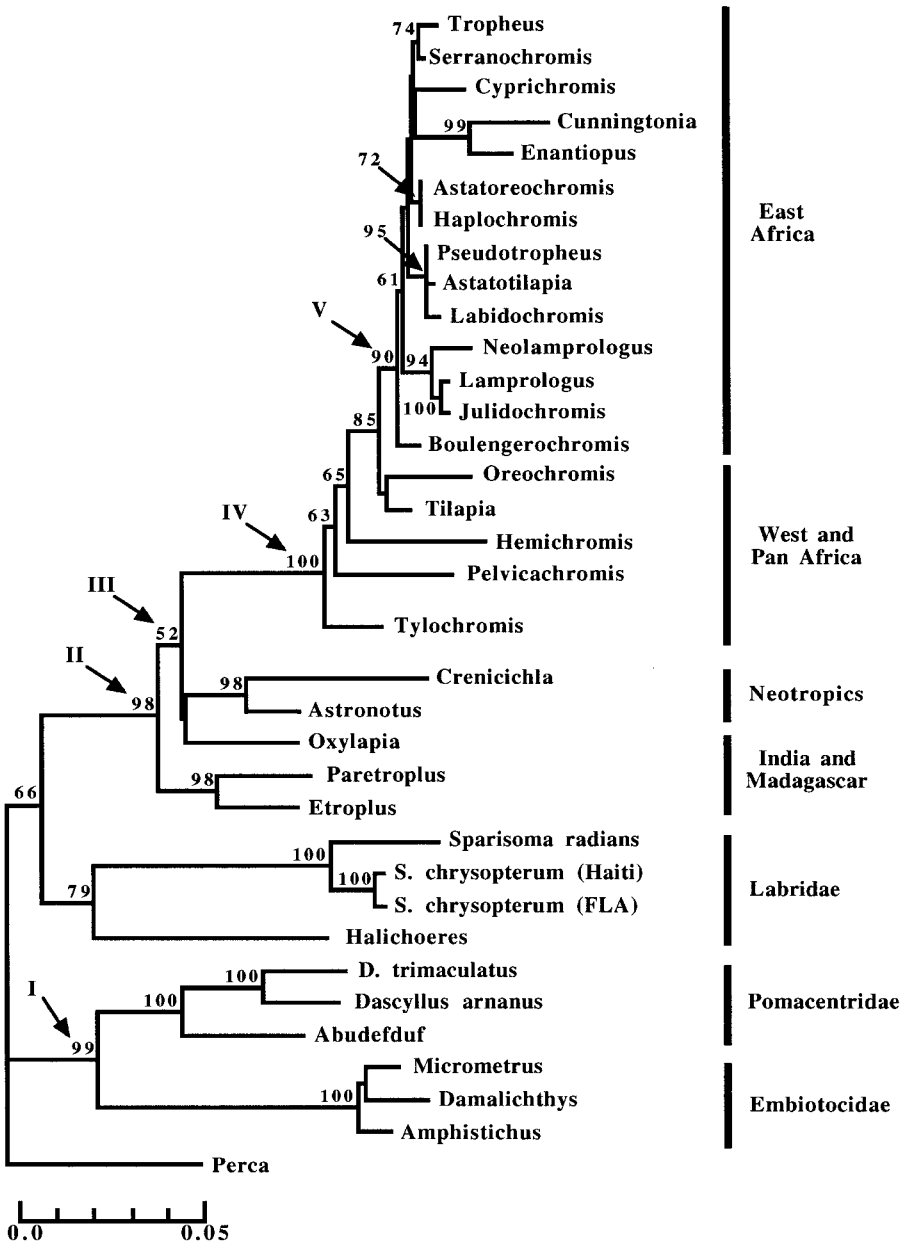

FIG. 3.—Fifty percent majority-rule neighbor-joining bootstrap consensus tree (and groups compatible with it) of maximum-likelihood distances between sequences at locus *Tmo-4C4* with *Perca* used as an outgroup. Distances were calculated using empirical base frequencies with third positions set at a rate category three times that of first and second positions (DNADIST program of PHYLIP [Felsenstein 1989]). Numbers at nodes represent bootstrap support in 200 replicates without rate categories. The topology and support for a neighbor-joining analysis using Kimura two-parameter distances were identical. The scale at the bottom indicates genetic distance. Roman numerals at nodes are as in figure 1.

lids are the sister group (node III) to a monophyletic African clade (node IV). Within Africa, west African cichlids branch off first, followed by Pan-African taxa. The west/Pan-Africans are paraphyletic and basal to an east African assemblage (node V). Relationships among east African cichlids, however, are poorly resolved. There are discrepancies among loci regarding the most basal African taxon. The *Tmo-4C4* data indicate that *Tylochromis* is the basal African cichlid, while data from *Tmo-M27* suggest that *Pelvicachromis* is the first to diverge. The *Tmo-4C4* sequences also confer a basal east African position upon the Tanganyikan tribe Lamprologini (*Lamprologus*, *Julidochromis.* and *Neolamprologus*). Bootstrap support for most of the clades within the east African assemblage is quite low, and exact relationships among taxa are uncertain.

The topology and bootstrap support of the *Tmo-4C4* neighbor-joining tree constructed from maximum-likelihood distances are consistent with those of the parsimony analysis (fig. 3). Neighbor-joining analysis does, however, indicate that there may be variable rates of evolution among taxa at *Tmo-4C4.* Preliminary relative-rate tests suggest that African cichlids as a whole may be evolving more rapidly than the basal cichlid species

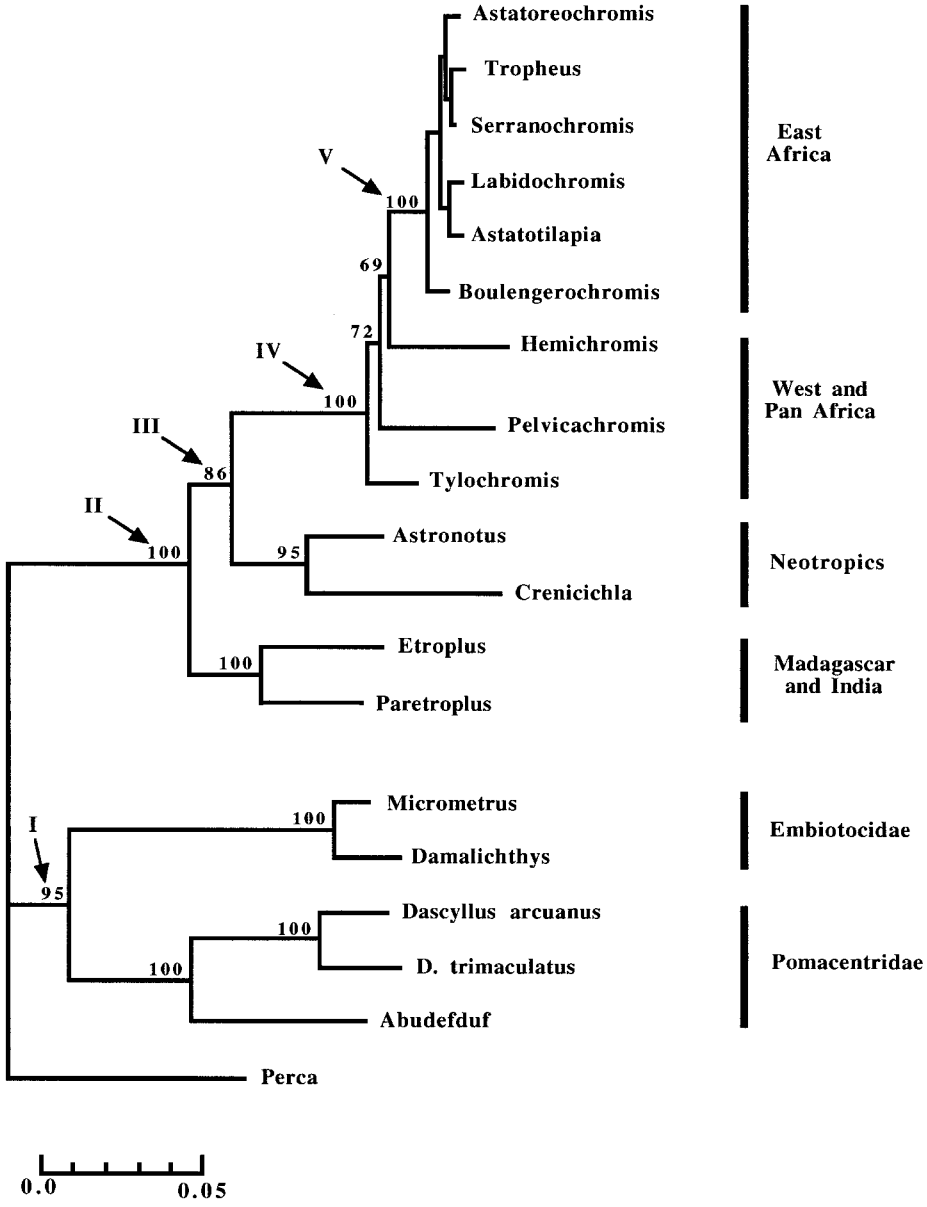

FIG. 4.—Fifty percent majority-rule neighbor-joining bootstrap consensus tree (and groups compatible with it) of maximum-likelihood distances for combined *Tmo-4C4* and *Tmo-M27* data. *Perca* was used as an outgroup. Distances were calculated using empirical base frequencies with codon position rate categories for *Tmo-4C4* set as in figure 3. For *Tmo-M27* sequences, the last 78 nucleotide positions were set with rate categories for codon positions first:second:third = 1:1:3. All other nucleotides at *Tmo-M27* were given the rate category of a third position. Numbers at nodes represent bootstrap support in 200 replicates without rate categories. The topology and support for a neighbor-joining analysis using Kimura two-parameter distances were identical. The scale at the bottom indicates genetic distance. Roman numerals at nodes are as in figure 1.

(i.e., Neotropical, Malagasy, and Indian taxa; unpublished data). The significance of this difference, however, is uncertain given the small number of basal taxa included in this study.

Parsimony and neighbor-joining analyses of combined data from *Tmo-M27* and *Tmo-4C4* result in identical relationships of cichlid major groups and show increased bootstrap support relative to separate analyses (fig. 4). A combined analysis does not, however, resolve the relationships among basal African taxa. A parsimony analysis (not shown) places *Pelvicachromis* while the neighbor-joining analysis places *Tylochromis* as the most basal African taxa.

Phylogenetic Pattern of the *Tmo-M27* Repeat Region

The data presented here and in Zardoya et al. (1996) allow us to assess the expansion and contraction patterns of the microsatellite repeat region. Figure 5 traces changes in the number of repeats in the *Tmo-M27* microsatellite region onto a consensus phylogeny of cichlid fishes. In the outgroups to the Cichlidae, the microsatellite region is usually long and often perfect. The

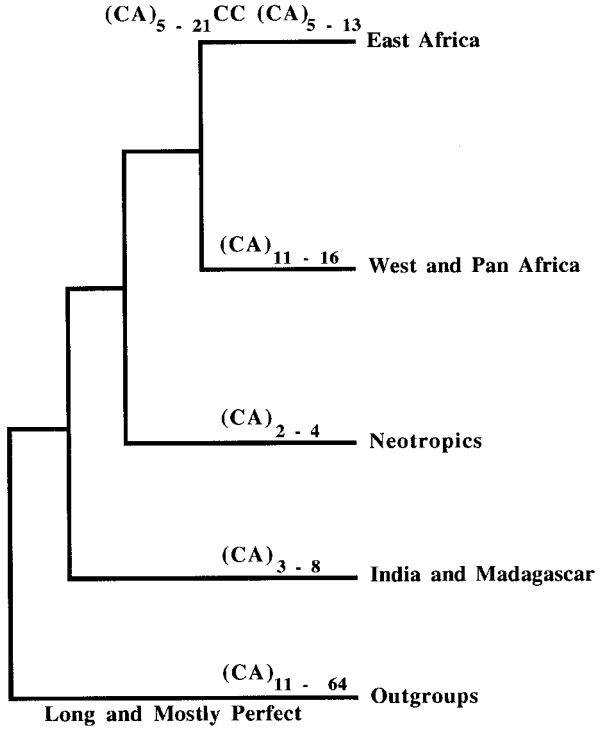

FIG. 5.—Consensus phylogeny of cichlid relationships with the length of the microsatellite repeat superimposed. Taxonomic designations are as in table 1.

Malagasy, Indian, and Neotropical cichlids, however, have truncated (i.e., 2–8) repeat regions that, with the exception of *Etroplus*, also are perfect. In contrast, African cichlids are characterized by long microsatellite regions (up to 34 repeats), with most taxa possessing imperfect runs.

Discussion

Phylogeny of Cichlid Fishes

Alone and combined, scnDNA and MFR sequences corroborate and refine the hypothesized cichlid major-group relationships previously summarized by Stiassny (1991) (fig. 1). Both markers appear to perform equally well in resolving these groups. Left uncertain are the relationships among west, Pan-, and east African cichlids. Morphological data (Oliver 1984; Stiassny 1991) place *Tylochromis* as the sister group to all other African taxa, but this hypothesis has not been corroborated by molecular approaches (Kocher et al. 1995; Sultmann et al. 1995; Zardoya et al. 1996). Here, either *Tylochromis* or *Pelvicachromis* is the basal African taxon depending on which locus (*Tmo-4* supports *Tylochromis, Tmo-M27* supports *Pelvicachromis*) or phylogenetic technique (parsimony analysis of combined data supports *Pelvicachromis*, distance supports *Tylochromis*) is assayed. Data from *Tmo-4C4* support monophyly of the Tanganyikan tribes Lamprologini and Ectodini as well as the Lake Victoria flock (albeit with few taxa). Both *Tmo-4C4* and *Tmo-M27* loci contain too little variation to reliably trace recent evolutionary events within east African cichlids (e.g., *Astatoreochromis* and *Haplochromis* have identical sequences at *Tmo-4C4*). However, these data bolster claims of explosive speciation of lacustrine assemblages within this fertile biogeographic region (Johnson, Scholz, and Talbot 1996).

MFRs performed as well as scnDNA in reconstructing evolutionary relationships, at least among the taxa included in this study. The scnDNA locus *Tmo-4C4* did provide better resolution among more closely related taxa (e.g., relationships between the African species; fig. 2). This, however, may be due to the increased number of characters and differences in the taxa sampled at this locus. Alternatively, the difference in resolution may indicate that MFRs are more conserved than scnDNA loci at shallower (local) versus deeper (global) divergence levels. In addition, MFRs may be influenced strongly by vagaries of evolution at the local level (Jin et al. 1996) which are not observed globally. For example, mutations causing interruptions in the repeat motif have dramatic effects on the rate and pattern of evolution in the microsatellite region (Estoup et al. 1995), and these effects could extend into the flanking sequence.

Evolution of *Tmo-4C4*

Sequence similarity analysis of *Tmo-4C4* indicates that it may code for a TITIN-like protein. The amino acid similarity to known proteins, however, is low. In the best match of *Tmo-4C4* to TITIN polypeptides (human), only 36% of the residues are identical, and an additional 20% are charge-conserved. These measures of amino acid similarity make assignment of homology uncertain. However, if *Tmo-4C4* is an immunoglobulin domain, it is possible that the locus may be under positive selection, favoring amino acid diversity (Tanaka and Nei 1989; Kuma, Iwabe, and Miyata 1995). There are several reasons why we believe that the phylogenetic results observed are not driven by selection (positive, overdominant, or otherwise). First, excluding first and second codon positions from phylogenetic analysis of *Tmo-4C4* does not change the topology of the resulting evolutionary trees (data not shown). Second, synonymous substitutions outnumber nonsynonymous substitutions at least five to one. Under models of positive diversifying selection, nonsynonymous mutations generally exceed synonymous changes (Kreitman and Akashi 1995). Third, diversifying selection on IG sequences has been demonstrated to occur predominantly in IG genes involved in the immune response (Hughes 1997). Nonimmune system IG domains similar to those found in TITIN have not been shown to be influenced by selection. Finally, only approximately one third of *Tmo-4C4* shows similarity to TITIN sequences, and when this region is omitted from nucleotide or polypeptide searches, no significant matches are returned.

Evolution of *Tmo-M27* MFR Regions

The putative polypeptide sequence from the 3′ microsatellite flanking region of locus *Tmo-M27* is remarkably similar across all taxa in this study (only 3 out of 26 putative amino acid positions are variable). In fact, this similarity extends to humans and murine rodents, with no more than five variable positions when

mammals and fishes are combined. Comparison of the human- or rat-inferred polypeptide sequences to *Tmo-M27* reveal 100% similarity when adjusted for charge conservation (i.e., 25 identical and 1 charge-conserved or 24 identical and 2 charge-conserved amino acid changes, respectively). This represents an extreme level of protein conservation over roughly 400 Myr of vertebrate evolution.

Considering only the fish DNA sequences in this region, there are a total of 16 variable nucleotide sites, with third codon positions outnumbering first and second codon positions 12:1 and 4:1, respectively. Again, this pattern is typical of coding regions where most of the nucleotide changes are confined to synonymous third codon positions. Overall, the majority of the phylogenetic information at this locus comes from the 5′ putative noncoding flanking region and putative third-codon-position changes in the 3′ ORF. The remarkable consistency of results between *Tmo-4C4* and *Tmo-M27* as well as with previous morphological (Stiassny 1991) and mitochondrial DNA studies (Kocher et al. 1995; unpublished data) argues against phylogenetic biases due to peculiarities of selection in any of these data sets.

Evolution of *Tmo-M27* Microsatellite Region

Armed with a consensus cichlid phylogeny drawn from morphological and molecular data (fig. 5), we can begin to characterize the evolutionary dynamics of the *Tmo-M27* microsatellite region. This locus appears to have undergone considerable expansion and contraction among the groups included in this study. In all outgroups, the repeat is long (e.g., 64 repeats in *Damalichthys*) and typically not interrupted. Basal cichlids (India, Madagascar, and Neotropics), however, have short repeats that are nearly always perfect (e.g., two repeats in *Crenicichla;* fig. 5). If the consensus phylogeny represents true evolutionary relationships, or at least that the Cichlidae is not ancestral to the outgroup taxa (which include *Perca* and *Lates*), then it would appear that the microsatellite region of *Tmo-M27* was dramatically reduced or nearly lost in the cichlid ancestor. The repeat motif is then expanded in the ancestor to African taxa. Notably, the largest differences in repeat numbers are found within the east African clade; the youngest and most species-rich group.

What might explain the allelic pattern in cichlid fishes? Observed taxonomic distributions in microsatellite repeat number can be explained either by inherent properties of a mutational mechanism or by selection. Additionally, since only a single individual of each species was included in this study, we cannot eliminate the possibility that the observed pattern is due to sampling biases. However, it seems highly unlikely that a blind draw of a single individual from each of 7 non-African taxa would result in only small alleles, whereas a similarly blind draw from each of 12 African taxa results in only large, interrupted alleles. Furthermore, in one instance where we sequenced multiple individuals from a single species (six individuals of *Cichlasoma citrinellum*), all individuals had an identical number of repeats (Zardoya et al. 1996).

Mutational processes are a more likely explanation of the observed pattern. Although still controversial, it appears that the primary molecular mechanism responsible for expansion and contraction of microsatellite loci involves the number of repeat units and DNA slippage-repair (Levinson and Gutman 1987; Schug, Mackay, and Aquadro 1997). When microsatellite motifs approach their lower limit (i.e., two), slippage-repair may be unable to create new, larger alleles. Once this point is reached, species may tend to be fixed for low repeat number over extended evolutionary periods (Gray and Jeffreys 1991). The observation of few repeats in basal cichlids may simply reflect conservation of allele size due to constraints on mutation. Hence, the reduced number of repeats at this locus may have been maintained independently for over 80 Myr in the Neotropical, Indian, and Malagasy lineages.

According to this scenario, the ancestor to African cichlids evolved an increased number of repeats, restoring the capacity of slippage mechanisms to produce larger allele sizes and larger variance in repeat number. Variation in repeat number in east African cichlids may have accompanied stints of gene diversification during multiple founder-flush events in the last 5 Myr (Scholz and Rosendahl 1988; Owen et al. 1990; Johnson, Scholz, and Talbot 1996; Verheyen et al. 1996).

Repeat number itself may be constrained by selection. Although much of the population genetic and evolutionary biology literature assumes implicitly or explicitly that microsatellite loci are evolving neutrally, there is a growing body of evidence suggesting otherwise. For instance, some dinucleotide and many trinucleotide repeats are associated with genetic instability and disease (Thibodeau, Bren, and Schaid 1993; de-Graaff et al. 1995; Loesch et al. 1995; Zhong et al. 1995; Wierdl et al. 1996; Wierdl, Dominska, and Petes 1997). Other researchers have proposed a functional role for microsatellites. Recent evidence indicates that the CAG trinucleotide repeat may compete for cytoplasmic RNA binding proteins (McLaughlin, Spencer, and Eberwine 1996), that purine/pyrimidine dinucleotide repeats are tightly bound by an RNA editor (Herbert et al. 1995), and that losses or gains of repeats can affect gene expression (Thornton et al. 1997). Microsatellites have been described as recombinational hot spots (Harding et al. 1993) and have been implicated in transposition and gene conversion events (Thompson-Stewart, Karpen, and Spradling 1994). Furthermore, repeat loci can show directionality in repeat copy number (Rubinsztein et al. 1995; Amos et al. 1996; Primmer et al. 1996; Wierdl, Dominska, and Petes 1997; but see Ellegren et al. 1997).

Our results from *Tmo-M27* point to a potential alternative explanation for the variation in $(CA)_n$ copy number. The 3′ flanking region of *Tmo-M27* is nearly 100% identical at the amino acid level to mammalian RAS-GRFs. Sequence similarity is generally lost near the microsatellite region, and the consensus GT immediately following the putative RAS-GRF ORF may function as an intron/exon splice junction. The microsatellite region and the 5′ flanking region of *Tmo-M27,*

therefore, are probably in an intron region and will be spliced out of the functional mRNA transcript.

What are the possible consequences of dinucleotide expansion near a RAS-GRF exon? Alternating polypurine/pyrimidine repeats are believed to form a left-handed DNA conformation (Z-DNA; Hamada, Petrino, and Kakunaga 1982; Gross, Huang, and Garrard 1985) which can alter the expression of nearby genes (Nordheim and Rich 1983; Hamada et al. 1984; Naylor and Clark 1990). RAS-GRF activates RAS proteins which control cell proliferation and development (Katz and McCormick 1997). Differential expression of *ras-grf* may have important ramifications for the regulation of RAS (Quilliam et al. 1995). If number or composition of repeats at *Tmo-M27* affects the expression of a RAS-GRF-like protein, then repeat length may be under selection. This may be a factor contributing to the complex pattern of microsatellite evolution at *Tmo-M27.* This possibility reinforces the suggestions of others (Moore et al. 1991; Rico, Rico, and Hewitt 1996) that microsatellite repeat evolution is strongly affected by genomic location.

Although difficult to demonstrate conclusively, these interpretations should provide critical cautionary information to population geneticists and evolutionary biologists who treat microsatellite loci as strictly neutral markers. The occurrence of a microsatellite (i.e., a potential regulatory region) imbedded within an important developmental cistron (in this case, RAS-GRF) implies that such assertions of neutrality may be invalid. It is known that dinucleotide repeats are dispersed throughout eukaryotic genomes (Stallings et al. 1991) and that this distribution may not be random (Brahmachari et al. 1995). Microsatellites are known from many well-characterized genes (e.g., prolactin, insulin-like growth factor, globins, actins, vertebrate growth hormone), and a regulatory role may be more widespread than previously presumed. Furthermore, few investigators report the flanking regions surrounding repeat motifs (but see Moore et al. 1991). Unfortunately, this information is largely unavailable due to the fact that standard techniques employ just enough sequence data at a microsatellite locus to identify PCR priming sites. Information on the MFRs is a minimal requirement for complete understanding of the evolutionary dynamics of simple sequence repeats.

Applications of *Tmo-4C4* and *Tmo-M27* in Fish Phylogeny

It has been difficult to find informative molecular markers for taxa believed to have diverged 50–200 MYA (Cantatore et al. 1994; Graybeal 1994; Meyer 1994). Mitochondrial genes, clearly the workhorses of modern molecular systematics, generally contain too many differences at the DNA level and too few at the amino acid level. Moreover, there is a growing appreciation that mitochondrial DNA represents only a single-gene genealogy and that the results from this single locus may not be representative of organismal evolution (Pamilo and Nei 1988; Ball, Neigel, and Avise 1990). Both nuclear loci, *Tmo-4C4* and *Tmo-M27,* perform well at reconstructing relationships among recently diverged cichlid fishes as well as at deeper taxonomic levels (Streelman and Karl 1997).

Acknowledgments

We thank A. L. (''Sam'') Bass, B. W. Bowen, R. C. Vrijenhoek, and two anonymous reviewers for valuable comments on the manuscript and C. Craddock and D. Vollmer for laboratory and technical assistance. R.Z. is sponsored by a postdoctoral grant from the Ministerio de Education y Ciencia of Spain. We are grateful to the National Science Foundation, U.S.A (grant DEB9615178), and the Max-Planck-Society, Germany for financial support to A.M. and to the University of South Florida, Sigma Xi, The American Cichlid Association, and The American Museum of Natural History for financial support to J.T.S and S.A.K.

LITERATURE CITED

ALTSCHUL S. F., W. GISH, W. MILLER, E. W. MYERS, and D. J. LIPMAN. 1990. Basic local alignment search tool. J. Mol. Biol. **215**: 403–410.

AMOS, W., S. J. SAWCER, R. W. FEAKES, and D. C. RUBINSZTEIN. 1996. Microsatellites show mutational bias and heterozygote instability. Nat. Genet. **13**:390–391.

AUSUBEL, F. M., R. BRENT, R. E. KINGSTON, D. D. MOORE, J. G. SEIDMAN, J. A. SMITH, K. STRUHL, P. WANG-IVERSON, and S. G. BONITZ. 1993. Current protocols in molecular biology. Greene Publishing Associates and Wiley-Interscience, New York.

BALL, R. M. JR., J. E. NEIGEL, and J. C. AVISE. 1990. Gene genealogies within the organismal pedigrees of random-mating populations. Evolution **44**:360–370.

BRAHMACHARI, S. K., G. MEERA, P. S. SARKAR, P. BALAGURUMOORTHY, J. TRIPATHI, S. RAGHAVAN, U. SHALIGRAM, and S. PATASKAR. 1995. Simple repetitive sequences in the genome: structure and functional significance. Electrophoresis **16**:1705–1714.

BROWN, J. H., and A. C. GIBSON. 1983. Biogeography. C. V. Mosley, St. Louis.

CANTATORE, P., M. ROBERTI, G. PESOLE, A. LUDOVICO, F. MILLELA, M. N. GADALETA, and C. SACCONE. 1994. Evolutionary analysis of cytochrome *b* sequences in some Perciformes: evidence for a slower rate of evolution than in mammals. J. Mol. Evol. **39**:589–597.

CHARLESWORTH, B., R. LANDE, and M. SLATKIN. 1982. A neo-Darwinian commentary on macroevolution. Evolution **36**: 474–498.

DEGRAAFF, E., R. WILLEMSEN, N. ZHONG, C. E. M. DE DIE-SMULDERS, W. T. BROWN, G. FRELING, and B. OOSTRA. 1995. Instability of the CGG repeat and expression of the FMR1 protein in a male fragile X patient with a lung tumor. Am. J. Hum. Genet. **57**:609–618.

DOMINEY, W. J. 1984. Effects of sexual selection and life history on speciation: species flocks in African cichlids and Hawaiian *Drosophila.* Pp. 231–250 *in* A. A. ECHELLE and I. KORNFIELD, eds. Evolution of fish species flocks. University of Maine Press, Orono.

ELLEGREN, H., S. MOORE, N. ROBINSON, K. BYRNE, W. WARD, and B. C. SHELDON. 1997. Microsatellite evolution—a reciprocal study of repeat lengths at homologous loci in cattle and sheep. Mol. Biol. Evol. **14**:854–860.

ESTOUP, A., C. TAILLIEZ, J.-M. CORNUET, and M. SOLIGNAC. 1995. Size homoplasy and mutational processes of inter-

rupted microsatellites in two bee species, *Apis mellifera* and *Bombus terrestris* (Apidae). Mol. Biol. Evol. **12**:1074–1084.

FELSENSTEIN, J. 1989. PHYLIP—phylogeny inference package (version 3.2). Cladistics **5**:164–166.

FRYER, G., and T. D. ILES. 1972. The cichlid fishes of the great lakes of Africa: their biology and evolution. Oliver and Boyd, Edinburgh.

FUTUYMA, D. J. 1986. Evolutionary biology. Sinauer, Sunderland Mass.

GISH, W., and D. J. STATES. 1993. Identification of protein coding regions by database similarity search. Nat. Genet. **3**: 266–72.

GOLDSCHMIDT, T. 1996. Darwin's dreampond: drama in Lake Victoria. MIT Press, Cambridge, Mass.

GRAY, I. G., and A. J. JEFFREYS. 1991. Evolutionary transience of hypervariable minisatellites in man and the primates. Proc. R. Soc. Lond. Biol. Sci. **243**:241–253.

GRAYBEAL, A. 1994. Evaluating the phylogenetic utility of genes: a search for genes informative about deep divergence among vertebrates. Syst. Biol. **43**:174–193.

GROSS, D. S., S. Y. HUANG, and W. T. GARRARD. 1985. Chromatin structure of the potential Z-forming sequence (dT-dG)$_n$–(dC-dA)$_n$. J. Mol. Biol. **183**:251–265.

GUSSOW, D., and T. CLACKSON. 1989. Direct clone characterization from plaques and colonies by the polymerase chain reaction. Nucleic. Acids. Res. **17**:4000–4001.

HAMADA, H., M. PETRINO, and T. KAKUNAGA. 1982. A novel repeated element with Z-DNA-forming potential is widely found in evolutionary diverse eukaryotic genomes. Proc. Natl. Acad. Sci. USA **79**:6465–6469.

HAMADA, H., M. SEIDMAN, B. H. HOWARD, and C. M. GORMAN. 1984. Enhanced gene expression by the poly(dT-dG)–poly(dC-dA) sequence. Mol. Cell. Biol. **4**:2622–2630.

HARDING, R. M., A. J. BOYCE, J. J. MARTINSON, J. FLINT, and J. B. CLEGG. 1993. A computer simulation study of VNTR population genetics: constrained recombination rules out the infinite alleles model. Genetics **135**:911–922.

HERBERT, A., K. LOWENHAUPT, J. SPITZNER, and A. RICH. 1995. Chicken double-stranded RNA adenosine deaminase has apparent specificity for Z-DNA. Proc. Natl. Acad. Sci. USA **92**:7550–7554.

HUGHES, A. L. 1997. Rapid evolution of immunoglobulin superfamily C2 domains expressed in immune system cells. Mol. Biol. Evol. **14**:1–5.

JIN, L., C. MACAUBAS, J. HALLMAYER, A. KIMURA, and E. MIGNOT. 1996. Mutation-rate varies among alleles at a microsatellite locus: phylogenetic evidence. Proc. Natl. Acad. Sci. USA **93**:15285–15288.

JOHNSON, T. C., C. A. SCHOLZ, and M. R. TALBOT. 1996. Late Pleistocene desiccation of Lake Victoria and rapid evolution of cichlid fishes. Science **273**:1091–1093.

KARL, S. A., and J. C. AVISE. 1993. PCR-based assays of Mendelian polymorphisms from anonymous single copy nuclear DNA: techniques and applications for population genetics. Mol. Biol. Evol. **10**:342–361.

KATZ, M. E., and F. MCCORMICK. 1997. Signal-transduction from multiple ras effectors. Curr. Opin. Genet. Dev. **7**:75–79.

KAUFMAN, L. S., L. J. CHAPMAN, and C. A. CHAPMAN. 1997. Evolution in fast forward: haplochromine fishes in the Lake Victoria region. Endeavour **21**:23–30.

KLEIN, D., H. ONO, C. O'HUIGIN, V. VINCEK, T. GOLDSCHMIDT, and J. KLEIN. 1993. Extensive MHC variability in cichlid fishes of Lake Malawi. Nature **364**:330–334.

KOCHER, T. D., J. A. CONROY, K. R. MCKAYE, J. R. STAUFFER, and S. F. LOCKWOOD. 1995. Evolution of NADH dehydrogenase subunit 2 in east African cichlid fish. Mol. Phylogenet. Evol. **4**:420–432.

KORNFIELD, I. 1978. Evidence for rapid speciation in African cichlid fishes. Experientia **34**:335–336.

KREITMAN, M., and H. AKASHI. 1995. Molecular evidence for natural selection. Annu. Rev. Ecol. Syst. **26**:403–422.

KUMA, K., N. IWABE, and T. MIYATA. 1995. Functional constraints against variations on molecules from the tissue level: slowly evolving brain-specific genes demonstrated by protein kinase and immunoglobulin supergene families. Mol. Biol. Evol. **12**:123–130.

LABEIT, S., and B. KOLMERER. 1995. Titins: giant proteins in charge of muscle ultrastructure and elasticity. Science **270**: 293–296.

LEVINSON, G., and G. A. GUTMAN. 1987. Slipped-strand mispairing: a major mechanism for DNA sequence evolution. Mol. Biol. Evol. **4**:203–221.

LINKE, W. A., M. IVEMEYER, N. OLIVIERI, B. KOLMERER, J. C. RÜEGG, and S. LABEIT. 1996. Towards a molecular understanding of the elasticity of TITIN. J. Mol. Biol. **261**:62–71.

LOESCH D. Z., R. HUGGINS, V. PETROVIC, and H. SLATER. 1995. Expansion of the CGG repeat in fragile X in the FMR1 gene depends on the sex of the offspring. Am. J. Hum. Genet. **57**:1408–1413.

LONG, M., S. J. DESOUZA, and W. GILBERT. 1995. Evolution of the intron-exon structure of eukaryotic genes. Curr. Opin. Genet. Dev. **5**:774–778.

MCLAUGHLIN, B. A., C. SPENCER, and J. EBERWINE. 1996. CAG trinucleotide RNA repeats interact with RNA-binding proteins. Am. J. Hum. Genet. **59**:561–569.

MADDISON, W. P., and D. R. MADDISON. 1992. MacClade: analysis of phylogeny and character evolution. Version 3.0. Sinauer, Sunderland, Mass.

MESSIER, W., S.-H. LI, and C.-B. STEWART. 1996. The birth of microsatellites. Nature **381**:483.

MEYER, A. 1994. Shortcomings of the cytochrome *b* gene as a molecular marker. TREE **9**:278–280.

MEYER, A., T. D. KOCHER, P. BASASIBWAKI, and A. C. WILSON. 1990 Monophyletic origin of Lake Victoria cichlid fishes suggested by mitochondrial DNA sequences. Nature **347**: 550–553.

MOORE, S. S., L. L. SARGEANT, T. J. KING, J. S. MATTICK, M. GEORGES, and D. J. S. HETZEL. 1991. The conservation of dinucleotide microsatellites among mammalian genomes allows the use of heterologous PCR primer pairs in closely related species. Genomics **10**:654–660.

MORAN, P., and I. KORNFIELD. 1993. Retention of an ancestral polymorphism in the mbuna species flock of Lake Malawi. Mol. Biol. Evol. **10**:1015–1029.

MORAN, P., I. KORNFIELD, and P. C. REINTHAL. 1994. Molecular systematics and radiation of the haplochromine cichlids of Lake Malawi. Copeia **1994**:274–288.

MULLIS, K., F. FALOONA, S. SCHARF, R. SAIKI, G. HORN, and H. ERLICH. 1986. Specific enzymatic amplification of DNA in vitro: the polymerase chain reaction. Cold Spring Harb. Symp. Quant. Biol. **51**:263–273.

NAYLOR, L. H., and E. M. CLARK. 1990. d(TG)$_n$–d(CA)$_n$ sequences upstream of the rat prolactin gene form Z-DNA and inhibit gene transcription. Nucleic. Acids. Res. **18**: 1595–1601.

NISHIDA, M. 1991. Lake Tanganyika as an evolutionary reservoir of old lineages of cichlid fishes: inferences from allozyme data. Experientia **47**:974–979.

NORDHEIM, A., and A. RICH. 1983. The sequence (dC-dA)$_n$–(dG-dT)$_n$ forms left-handed Z-DNA in negatively supercolied plasmids. Proc. Natl. Acad. Sci. USA **80**:1821–1825.

OLIVER, M. K. 1984. Systematics of African cichlid fishes: determination of the most primitive taxon, and studies on the haplochromines of Lake Malawi (Teleostei, Cichlidae). Ph.D. thesis, Yale University, New Haven, Conn.

ORTI, G., and A. MEYER. 1996. Molecular evolution of ependymin and the phylogenetic resolution of early divergences among euteleost fishes. Mol. Biol. Evol. **13**:556–573.

ORTI, G., D. E. PEARSE, and J. C. AVISE. 1997. Phylogenetic assessment of length variation at a microsatellite locus. Proc. Natl. Acad. Sci. USA **94**:10745–10749.

OWEN, R. B., R. CROSSLEY, T. C. JOHNSON, D. TWEDDLE, I. KORNFIELD, S. DAVISON, D. H. ECCLES, and D. E. ENGSTROM. 1990. Major low levels of Lake Malawi and their implications for speciation rates in cichlid fishes. Proc. R. Soc. Lond. B Biol. Sci. **240**:519–553.

PAMILO, P., and M. NEI. 1988. Relationships between gene trees and species trees. Mol. Biol. Evol. **5**:568–583.

PRIMMER, C. R., H. ELLEGREN, N. SAINO, and A. P. MOLLER. 1996. Directional evolution in germline microsatellite mutations. Nat. Genet. **13**:391–393.

QUILLIAM, L. A., R. KHOSRAVI-FAR, S. Y. HUFF, and C. J. DER. 1995. Guanine nucleotide exchange factors: activators of the Ras superfamily of proteins. BioEssays **17**:395–404.

RAUP, D. M., and D. JABLONSKI. 1986. Patterns and processes in the history of life. Springer-Verlag, Berlin.

RIBBINK, A. J. 1990. Alternative life-history styles of some African cichlid fish. Environ. Biol. Fish. **28**:87–100.

RICO, C., I. RICO, and G. HEWITT. 1996. 470 million years of conservation of microsatellite loci among fish species. Proc. R. Soc. Lond. Biol. Sci. **263**:549–557.

RUBINSZTEIN, D. C., W. AMOS, J. LEGGO, S. GOODBURN, S. JAIN, S-H. LI, R. L. MARGULIS, C. A. ROSS, and M. A. FERGUSON-SMITH. 1995. Microsatellite evolution–evidence for directionality and variation in rate between species. Nat. Genet. **10**:337–343.

SAIKI, R., S. SCHARF, F. FALOONA, K. B. MULLIS, G. T. HORN, and H. A. ERLICH. 1985. Enzymatic amplification of β-globin genomic sequences and restriction site analysis for diagnosis of sickle cell anemia. Science **230**:1350–1354.

SCHOLZ, C. A., and B. R. ROSENDAHL. 1988. Low lake stands in lakes Malawi and Tanganyika, East Africa, delineated with multifold seismic data. Science **240**:1645–1648.

SCHUG, M. D., T. F. C. MACKAY, and C. F. AQUADRO. 1997. Low mutation rates of microsatellite loci in *Drosophila melanogaster*. Nat. Genet. **15**:99–102.

STALLINGS, R. L., R. F. FORD, D. NELSON, D. C. TORNEY, C. E. HILDEBRAND, and R. K. MOYZIS. 1991. Evolution and distribution of $(GT)_n$ repetitive sequences in mammalian genomes. Genomics **10**:807–815.

STIASSNY, M. L. J. 1987. Cichlid familial intrarelationships and the placement of the neotropical genus *Cichla* (Perciformes, Labroidei). J. Nat. Hist. **21**:1311–1331.

———. 1991. Phylogenetic intrarelationships of the family Cichlidae: an overview. Pp. 1–31 *in* M. H. A. KEENLEYSIDE, ed. Cichlid fishes—behavior, ecology and evolution. Croom Helm, London.

STOREY, B. C. 1995. The role of mantle plumes in continental breakup: case histories from Gondwanaland. Nature **377**: 301–308.

STREELMAN, J. T., and S. A. KARL. 1997. Reconstructing labroid evolution using single-copy nuclear DNA. Proc. R. Soc. Lond. Biol. Sci. **264**:1011–1020.

STURMBAUER, C., and A. MEYER. 1993. Mitochondrial phylogeny of the endemic mouthbrooding lineages of cichlid fishes from Lake Tanganyika in Eastern Africa. Mol. Biol. Evol. **10**:751–768.

STURMBAUER, C., E. VERHEYEN, and A. MEYER. 1994. Mitochondrial phylogeny of the Lamprologini, the major substrate spawning lineage of cichlid fishes from Lake Tanganyika in eastern Africa. Mol. Biol. Evol. **11**:691–703.

SULTMANN, H., W. E. MAYER, F. FIGUEROA, H. TICHY, and J. KLEIN. 1995. Phylogenetic analysis of cichlid fishes using nuclear DNA markers. Mol. Biol. Evol. **12**:1033–1047.

SWOFFORD, D. L. 1993. PAUP: phylogenetic anlysis using parsimony. Version 3.1.1. Illinois Natural History Survey, Champaigne.

TANAKA, T., and M. NEI. 1989. Positive Darwinian selection observed at the variable region genes of immunoglobulins. Mol. Biol. Evol. **6**:447–459.

THIBODEAU, S. N., G. BREN, and D. SCHAID. 1993. Microsatellite instability in cancer of the proximal colon. Science **260**:816–819.

THOMPSON-STEWART, D., G. H. KARPEN, and A. C. SRADLING. 1994. A transposable element can drive the concerted evolution of tandemly repetitious DNA. Proc. Natl. Acad. Sci. USA **91**:9042–9046.

THORNTON, C. A., J. P. WYMER, Z. SIMMONS, C. MCCLAIN, and R. T. MOXLEY III. 1997. Expansion of the myotonic dystrophy CTG repeat reduces expression of the flanking *DMAHP* gene. Nat. Genet. **16**:407–409.

VERHEYEN, E., L. RUBER, J. SNOEKS, and A. MEYER. 1996. Mitochondrial phylogeography of rock-dwelling cichlid fishes reveals evolutionary influence of historical lake level fluctuations of Lake Tanganiyka, Africa. Philos. Trans. R. Soc. Lond. Biol. Sci. **351**:797–805.

WIERDL, M., M. DOMINSKA, and T. D. PETES. 1997. Microsatellite instability in yeast: dependence on the length of the microsatellite. Genetics **146**:769–779.

WIERDL, M., C. N. GREENE, A. DATTA, S. JINKS-ROBERTSON, and T. D. PETES. 1996. Destabilization of simple repetitive DNA sequences by transcription in yeast. Genetics **143**: 713–721.

ZARDOYA, R., D. M. VOLLMER, C. CRADDOCK, J. T. STREELMAN, S. KARL, and A. MEYER. 1996. Evolutionary conservation of microsatellite flanking regions and their use in resolving the phylogeny of cichlid fishes (Pisces: Perciformes). Proc. R. Soc. Lond. Biol. Sci. **263**:1589–1598.

ZHONG, N., W. YANG, C. DOBKIN, and W. T. BROWN. 1995. Fragile X gene instability: anchoring AGG's and linked microsatellites. Am. J. Hum. Genet. **57**:351–361.

SHOZO YOKOYAMA, reviewing editor

Accepted March 9, 1998