First publ. in: Proceedings of the Royal Society B 263 (1996), pp. 1589-1598

Evolutionary conservation of microsatellite flanking regions and their use in resolving the phylogeny of cichlid fishes (Pisces: Perciformes)

RAFAEL ZARDOYA[1], DANA M. VOLLMER[1], CLARK CRADDOCK[2], JEFFREY T. STREELMAN[2], STEVE KARL[2] AND AXEL MEYER[1]*

[1] *Department of Ecology and Evolution, and Program in Genetics, State University of New York, Stony Brook, New York 11794–5245, U.S.A.*
[2] *Department of Biology, University of South Florida, Tampa, Florida 33620–5150, U.S.A.*

SUMMARY

A phylogeny of the principal lineages of cichlid fishes and two other fish families of the suborder Labroidei was based on phylogenetic information from DNA sequences of the flanking region of a $(CA)_n$microsatellite locus. Microsatellite $(CA)_n$ containing clones from a genomic library of an African cichlid fish from Lake Tanganyika, *Tropheus moorii*, were sequenced and primers for the polymerase chain reaction designed. All primers amplified the homologous microsatellite loci in many more than the source species and one microsatellite flanking locus (*TmoM27*) was particularly conserved and amplified in several lineages of perciform fishes that diverged more than 80–100 million years ago. Despite the extensive level of evolutionary conservation of this microsatellite flanking region (MFR), this nuclear region contained reliable phylogenetic information in the form of both point and length mutations. A phylogeny of cichlids based on this MFR agrees with other phylogenetic hypotheses based on morphological, mitochondrial, and anonymous nuclear DNA. Madagascan and Indian cichlids are found to be paraphyletic and the most basal group in the family Cichlidae. African and Neotropical cichlids are both monophyletic and sistergroups. Within African lineages, the East African cichlids are most likely to be monophyletic and the West African cichlids are probably paraphyletic and basal to all African species. The focal microsatellite locus contained much variation in $(CA)_n$repeats in African cichlids and in surfperches (up to 64 repeats), but was short (with only 2–4 repeats) and almost invariant in Neotropical cichlids. The design of phylogenetically highly versatile MFR-primers will be of use not only for phylogeny reconstruction among families of perciform fishes, but also for population-level work in the thousands of species belonging to this highly species-rich suborder of fishes.

1. INTRODUCTION

Microsatellites or simple sequence repeats (SSRS) are DNA sequences that are made up of short (2–5 bp) tandem repeats and are highly variable in copy number even among individuals from the same population (Litt & Luty 1989; Tautz 1989; Weber & May 1989). The fast rates of microsatellite evolution are believed to be caused by replication slippage events (Schlötterer & Tautz 1992). The most common type of motif is a CA–GT di-nucleotide repeat (Tautz *et al.* 1986). This simple repetitive sequence is widely interspersed in eukaryotic genomes, including yeast, insects, fish, amphibians and mammals (Hamada *et al.* 1982) and it is estimated to occur every 20–30 kb on average (Stallings *et al.* 1991). Microsatellites have been traditionally classified according to the nature of the repeat as perfect (without interruptions), imperfect (with one or more interruptions) or compound (adjacent tandem repeats of a different sequence) (Weber 1990).

Microsatellites, because of their extremely high levels of polymorphism, are widely used in a population-level evolutionary context e.g. as indicators of kinship, gene flow, and population structure (reviewed in Queller *et al.* 1993; Bruford & Wayne 1993). Microsatellite fingerprinting approaches have also recently been applied to address questions about paternity in cichlids (Kellog *et al.* 1995). Unfortunately, extensive preliminary molecular work is necessary to develop PCR-primers for individual loci for each particular population study, a circumstance that continues to stand in the way of even more widespread application of single-locus microsatellite markers.

It came as a surprise when it was noted recently that homologous microsatellite loci can persist for about 300 million years (Ma) in turtles (FitzSimmons *et al.* 1995) and even in fish for 470 Ma (Rico *et al.* 1996). It would appear that some microsatellite loci, despite their

* Author to whom correspondence should be addressed: (email: ameyer@life.bio.sunysb.edu).

extremely fast rates of repeat evolution, are quite conservative in their flanking regions and hence can persist for long evolutionary time spans much unchanged. MFRs sometimes contain length mutations across species which might produce identical length variants that could compromise microsatellite population level work (and comparisons of levels of variation across species for homologous loci) and phylogenetic inferences as these length variants in the flanking regions can potentially mimic allelic length variation in the repeat region (see, for example, FitzSimmons *et al.* 1995; Rico *et al.* 1996).

So far, the flanking regions of microsatellites have not been explored for their phylogenetic use. We investigated the evolutionary conservation and phylogenetic information content of a MFR in cichlids and other labroid fishes. Cichlid fishes have received wide attention from evolutionary biologists because they are extremely diverse morphologically, behaviorally, and ecologically (see, for example, Fryer & Iles 1972; Meyer *et al.* 1990; Meyer 1993*b*; Meyer 1994; Meyer *et al.* 1996). With an estimated number of species exceeding 1 500, cichlids are one of the most species-rich families of vertebrates. Cichlids have been studied in particular to understand the formation of 'species flocks' i.e. adaptive radiations that live in single lakes. The geographic region of greatest diversity of cichlids is the East African Rift lakes, where each of the three great lakes Victoria, Malawi and Tanganyika house adaptive radiations with hundreds of endemic species each (reviewed in see, for example, Fryer & Iles 1972; Echelle & Kornfield 1984; Meyer 1993*b*). There the morphological diversification and speciation in some cichlid lineages took place extremely rapidly and many new species arose within only a few thousand generations (Meyer *et al.* 1990; Owen *et al.* 1990; Meyer 1994; Meyer *et al.* 1996). Because of the recent application of molecular phylogenetic methods significant progress has been made in the understanding of phylogenetic relationships of cichlids within and between the three East African lakes (Meyer *et al.* 1990; Meyer *et al.* 1991; Sturmbauer & Meyer 1993; Moran & Kornfield 1993; Kocher *et al.* 1993; Sturmbauer *et al.* 1994; Kocher *et al.* 1995; Meyer *et al.* 1996; reviewed in Meyer 1993*b*).

All workers since Boulenger (Boulenger 1898) believe that cichlids are a natural group (see, for example, Pellegrin 1903; Regan 1920; Regan 1922; reviewed in Stiassny 1981; Stiassny 1991). Detailed knowledge about the relationships of the principal lineages of cichlid fishes is largely wanting, and the phylogenetic relationships of the East African lineages to those of the rest of Africa, South America, Madagascar, and India still remain somewhat uncertain (see Stiassny 1991; Sültmann *et al.* 1995). Stiassny (1991) recently summarized and analysed the phylogenetic knowledge of cichlids at the family level (see, for example, Regan 1920; Regan 1922; Cichocki 1976; Stiassny 1981; Olivier 1984; Stiassny 1987) based on morphological data. The main findings are: (i) the Madagascan and Indian cichlids form a (probably paraphyletic) group at the base of the family; (ii) all Neotropical (about 350 species) and all African taxa (more than 1 500 species) (except possibly *Heterochromis*) form monophyletic groups respectively; and (iii) the African *Tylochromis* is the sistergroup to all other Africa cichlids (except possibly *Heterochromis* which might group with the Madagascan + Indian species).

The Cichlidae are one of four families in the suborder Labroidei (Kaufman & Liem 1982; Stiassny & Jensen 1987) which also includes Embiotocidae (surf perches), Pomacentridae (damselfishes) and Labridae (wrasses, odacids and parrotfishes). Interrelationships of labroidei families also remain unclear and debated (Liem & Greenwood 1981; Kaufman & Liem 1982; reviewed in Stiassny 1991). The most recent phylogenetic analysis of the relationships within this suborder (Stiassny & Jensen 1987) placed Cichlidae as sistergroup to the other three families. However, this analysis (Stiassny & Jensen 1987) did not resolve the controversy concerning the interrelationships of the Labroidei unambiguously because the most parsimonious cladogram required independent losses and gains of several complex characters. Because of their biogeographic distribution it is likely that the diversification of cichlids must have taken place before the fragmentation of Gondwanaland (i.e. *ca.* 70–95 Ma).

We developed versatile PCR primers for population level work in many groups of perciform fishes, and in particular, cichlids. Versatile PCR-primers in microsatellite flanking regions can serve two purposes at once: (i) they can be used to retrieve phylogenetic information contained in the flanking regions to establish relationships among distantly related species; and (ii) to amplify individual homologous microsatellite loci for population level questions in a large number of species without having to invest the time and expense to develop species-specific primers for each study anew.

2. MATERIALS AND METHODS

(*a*) *Species and microsatellite loci*

Table 1 lists 27 species, from five perciform families, that were used in this study. Also four species of labrid fish (*Halichoeres maculipinna*, *Halichoeres bivittatus*, *Lachnolaimus* sp., *Clepticus parrai*) were included in this study, but none of the primers were found to amplify the corresponding locus in these species. All specimens, with the exception of the Indian cichlid *Etroplus maculatus*, were collected in the wild. A total of six microsatellite clones were sequenced and the PCR-primers proved to be of various levels of versatility for future population level work in cichlid fishes and other perciforms (tables 1 and 2).

(*b*) *Library construction and clone sequencing*

CA–GT-clones of a specimen of *Tropheus moorii*, a cichlid species endemic to Lake Tanganyika were obtained from a genomic library. *Pst*I DNA fragments were ligated to pBluescript SK + (Stratagene) and transformed into SURE cells (Stratagene). The *Pst*I library was screened with a radiolabelled GT probe. The insert sizes of the positive clones ranged from about 0.5–2.5 kb. Clones were sequenced with M13 universal (−40) and reverse primers with an Applied

Table 1. *Classification of the perciform species analysed in this study and versatility of the MFR PCR primers*

(For each locus the result of the PCR amplification is shown (+) product, (−) no obvious product, (?) product of questionable size, (n) not attempted)

	TmoM5	*TmoM7*	*TmoM11*	*TmoM13*	*TmoM25*	*TmoM27*
Family Cichlidae						
East Africa:						
Boulengerochromis microlepis	+	+	+	?	+	+
Serranochromis robustus	−	−	−	+	+	+
Tropheus moorii	+	+	+	+	+	+
Astatoreochromis alluaudi	+	+	+	+	+	+
Astatotilapia calliptera	+	−	+	+	+	+
Pseudocrenilabrus multicolor	+	+	+	+	+	+
Labidochromis caeruleus	+	−	?	+	+	+
Chalinochromis brichardi	n	n	n	n	n	+
Pan-African:						
Tylochromis polylepis	−	+	−	−	+	+
Oreochromis niloticus	+	−	−	−	?	+
West African:						
Hemichromis bimaculatus	−	+	−	−	+	+
Pelvicachromis pulcher	−	−	−	−	+	+
Neotropics:						
Cichla ocellaris	−	−	−	−	−	+
Crenicichla saxatilis	−	−	−	−	−	+
Astronotus ocellatus	−	−	−	−	−	+
Cichlasoma citrinellum	−	−	−	−	−	+
Madagascar and India:						
Ptychochromoides betsileanus	−	−	−	−	−	+
Paretroplus polyactis	−	−	−	−	−	+
Etroplus maculatus	−	−	−	−	−	+
Family Pomacentridae						
Abudefduf saxiatilis	−	−	−	−	−	+
Dascyllus trimaculatus	−	−	−	−	−	+
Dascyllus aruanus	−	−	−	−	−	+
Family Embiotocidae						
Micrometrus minimus	−	−	−	−	−	+
Damalichthys vacca	−	−	−	−	−	+
Cymatogaster aggregata	−	−	−	−	−	+
Family Centropomidae						
Lates niloticus	−	−	−	−	−	+
Family Percidae						
Perca fluviatilis	−	−	−	−	−	+

Table 2. *Versatility of microsatellite primers in perciform fishes*

locus	primer sequences (5′-3′)	category*	repeat motif (original clone)	expected size (bp)	annealing T (°C)
TmoM5	F gct caa tat tct cag ctg acg ca R aga aca gcg ctg gct atg aaa agg t	compound-perfect	$(GC)_6(AC)_{14}$	300–400	48
TmoM7	F ctg cag cct cgc tca cca cgt at R cac cag ata act gca cag ccc ag	imperfect	$(CA)AA(CA)_{10}$	250–300	50
TmoM11	F att cag gta gag acg aaa tat ta R tag tca cag ttt aca cac aac	perfect	$(AC)_{19}$	150–250	50
TmoM13	F cgc agg gtg ttc ttc agg tgt at R aaa tca cca tat tca tat gtt	imperfect	$(CA)_4CG(CG)_{24}$	250–300	50
TmoM25	F ctg cag tgg cac atc aag aat gag cag cgg t R caa gaa cct ttc aag tca ttt tg	perfect	$(CA)_{23}$	350–400	52
TmoM27	F agg cag gca att acc ttg atg tt R tac taa ctc tga aag aac ctg tga t	imperfect	$(CA)_{13}CC(CA)_5$	350–500	48

* According to Weber (1990).

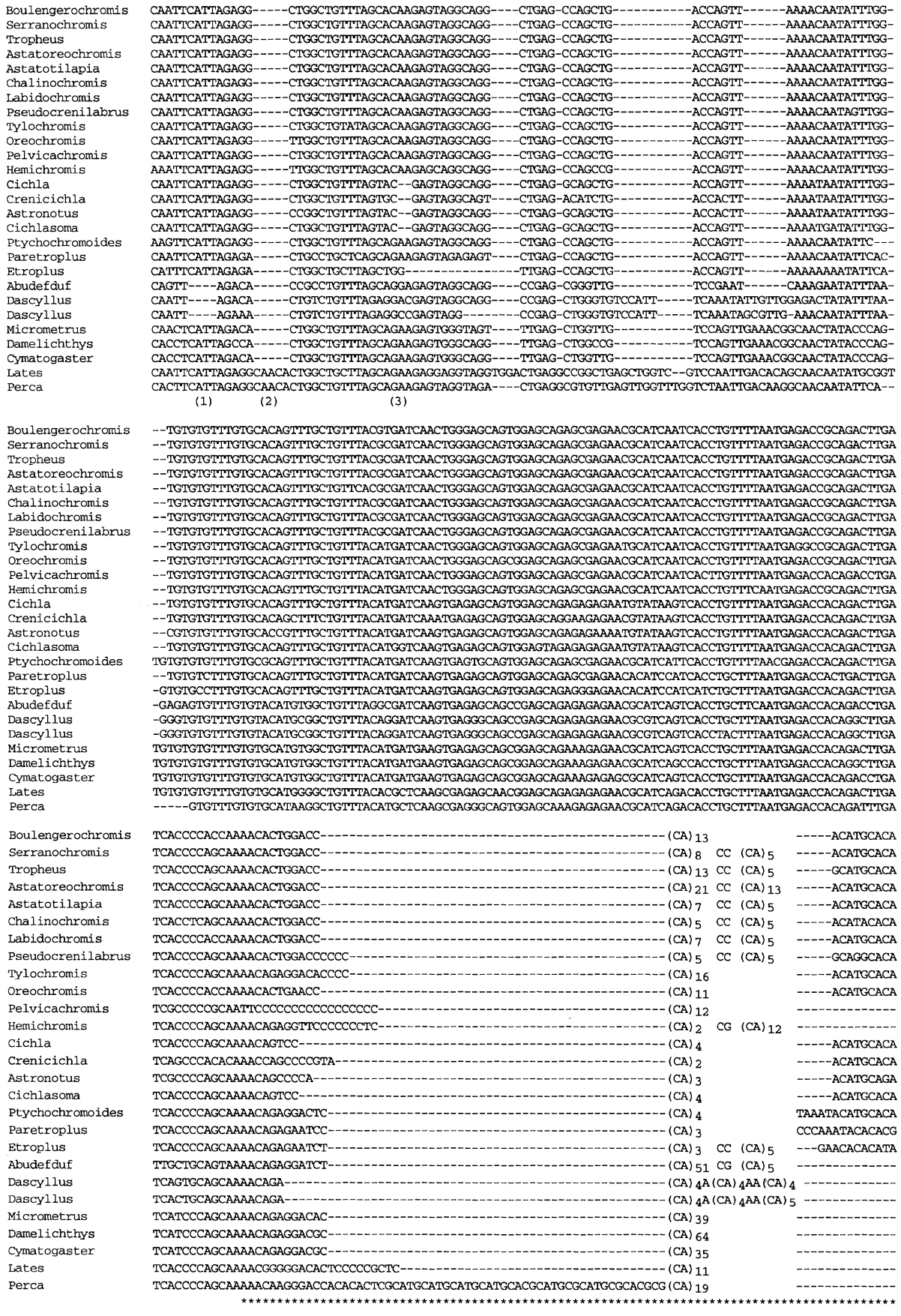

Figure 1. For legend see facing page.

Biosystems 373A *Stretch* DNA sequencer using the Taq Dye Deoxy Terminator Cycle Sequencing Kit (Applied Biosystems) according to manufacturers instructions. Specifically designed oligonucleotides were needed to walk into those positive clones where SSRs were not localized close to the borders of the insert.

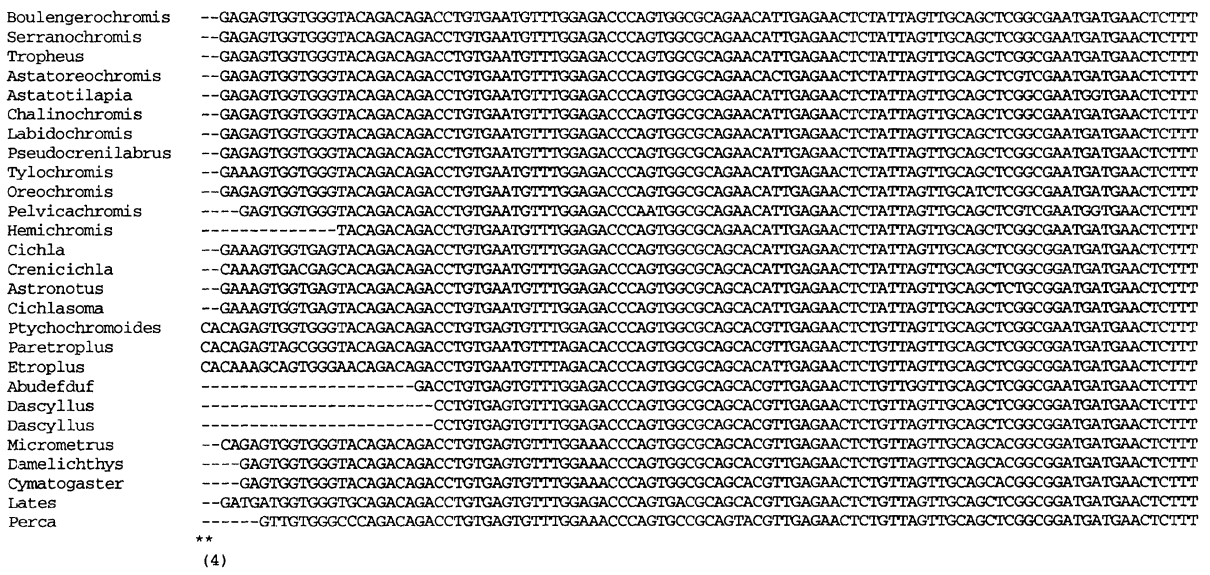

Figure 1. Aligned DNA sequences of the *TmoM27* microsatellite flanking regions from 27 species of perciform fishes. SSRs are indicated by $(CA)_n$ repeats. Asterisks indicate positions that were excluded from the phylogenetic analyses and numbers in parentheses indicate phylogenetically informative length variations (see text).

(*c*) PCR *primers design and* PCR *amplification*

Based on the DNA sequences of the *Tropheus* microsatellite-containing clones, six primer pairs for the polymerase chain reaction (PCR) were designed (table 2). The placement of primers was such that they would be about 150 bp up- and downstream of the $(CA)_n$ repeat, allowing the unambiguous verification of homology of the PCR-product. By placing the primer sites at such an unusually large distance from the microsatellite it was possible to explore the phylogenetic use of these flanking regions, the evolutionary conservation of these flanking regions and to simultaneously use these PCR-primers for future population-level studies that will use microsatellite length variation as population genetic marker. Most of the primers worked only in cichlid species from East Africa (tables 1, 2) as expected because the source species is an endemic cichlid of Lake Tanganyika. As East Africa contains more than 1000 species of cichlid fishes, for most of which these primer pairs are expected to work, they might be of use for workers interested in population level studies of East African cichlids. The principal MFR studied here was *TmoM27* as it appeared to have the highest degree of phylogenetic versatility and amplified the homologous locus in 27 of 31 species for which it was tested (tables 1 and 2). DNA was amplified (94 °C 60 s, 48–52 °C 60 s, 72 °C 60 s) for 35 cycles on a Perkin Elmer 480 PCR machine in 25 µl volumes (Tris 67 mM, pH 8.8; 1.5 mM $MgCl_2$, 0.4 mM of each dNTP, 75 ng of each primer, and 1 unit of AmpliTaq DNA polymerase (Perkin-Elmer-Cetus) (see table 2 for exact annealing temperatures). Aliquots of the PCR products were ligated into pGEM-T plasmid vectors (Promega). Clones for each PCR-product were sequenced for both strands as described above. Sequences determined here are available in GenBank (accession numbers: U63654–U63680).

(*d*) *Phylogenetic analyses*

DNA sequences of *TmoM27* were aligned using CLUSTAL W (Thompson *et al.* 1994) followed by refinement by eye (see figure 1). Ambiguously aligned nucleotides (142 positions indicated with asterisks in figure 1), exclusively located in the microsatellite repeat region, were excluded from the phylogenetic analyses and gaps resulting from the alignment were treated as missing data. Sequences were subjected to the maximum-parsimony (MP) method (PAUP* version 4.0d46, Swofford 1993) using heuristic searches with ten random stepwise additions of taxa to find the most parsimonious trees. Transitions and transversions were given equal weight; if weighted 1:2 the effects on the resulting topologies were negligible. Also maximum likelihood (ML) (two parameter default model in PAUP* using empirical base frequencies) and neighbor-joining (NJ) (Saitou & Nei 1987) (based on Kimura two parameter corrected distance matrices) analyses were done with PAUP* (Swofford 1993). Robustness of the phylogenetic results was tested by bootstrapping (Felsenstein 1985) (as implemented in PAUP* with 1000 pseudoreplications each). A spectral analysis (Hendy *et al.* 1994) was done with Spectrum (version 1.01) (Charleston 1995) using the average spectrum option to calculate the support of each split and selecting for splits with estimated length greater than 0.001.

3. RESULTS

(*a*) *Versatility of the* PCR *primers*

The versatility of the six PCR-primer pairs for MFRs was tested in 31 species from six families of perciform fishes (tables 1, 2). They had various levels of conservation and therefore phylogenetic versatility (tables 1, 2). Some PCR primer pairs only amplified homologous loci of microsatellite containing DNA in other closely related African but not, for example, in Neotropical cichlids or other families of percoid fishes (table 1). The primer pair designed from clone *TmoM27* proved to be the most versatile. This set of primers was able to amplify the corresponding locus from three (Cichlidae, Embiotocidae and Pomacentridae) families of fish from the suborder Labroidei as well as from the outgroups (*Perca fluviatilis*, the perch (Family Percidae) and *Lates niloticus*, the nile perch (Family Centropomidae)). Somewhat surprisingly, these PCR primers did not however, amplify this locus

in four species of fish from the family Labridae, the fourth family in the suborder Labroidei, in which they were tested. None of the MFR sequences have significant similarity to sequences available in GenBank.

(*b*) *Microsatellite flanking regions*

Figure 1 shows the aligned sequences (459 positions) for the set of 27 species (table 1) in which *TmoM27* was amplified, cloned, and sequenced. The *TmoM27* locus is characterized by two stretches of conserved flanking regions about 100–150 pb on either side of the SSR across the taxa studied.

Interestingly, these regions appear to be more variable in close proximity to the SSR. Point mutations dominate numerically in the MFR although several, sometimes sizeable, length mutations were observed as well (figure 1). Several phylogenetically informative insertion–deletion events can be observed (figure 1), for example, those that synapomorphically differentiate: (i) Pomacentrids from other taxa; (ii) ingroup from outgroup species in this data set; (iii) Neotropical cichlids from other cichlids; and (iv) cichlids from all other families of labroids (see corresponding numbers in figure 2). MFR length variation has the potential to mislead in the interpretation of microsatellite repeat length variation (FitzSimmons *et al.* 1995; Rico *et al.* 1996). However, we did not detect any length variation in those cases in which more than one individual was sequenced in a particular species and only in one case was a single base pair substitution in the MFR found between the two alleles from one individual (data not shown).

Base composition in the microsatellite flanking region was quite even among all taxa; only a slight bias favouring A (26.7%) and G (28.1%), a normal percentage of T (25.0%) and an anti-C bias (20.2%) were detected. The average transition–transversion ratio for the MFR of the Tmo 27 locus was 1.76.

(*c*) *Phylogenetic relationships among the principal lineages of cichlids and labroid fishes*

Evolutionary relationships of cichlids and other perciform fishes based on the *TmoM27* microsatellite flanking regions were corroborated by maximum parsimony (Swofford 1993), neighbor-joining (Saitou & Nei 1987), and maximum likelihood phylogenetic reconstruction methods (figure 2). The *TmoM27* MFR data suggest: (i) that the Madagacan + Indian cichlids form a paraphyletic group and are basal in the monophyletic family Cichlidae; (ii) the Neotropical cichlids (about 350 species) and the African species (over 1000 species) are monophyletic, respectively (except possibly *Heterochromis*, which could not be included in this study); (iii) the West African lineages are likely to be paraphyletic and basal to the other African lineages. The inferred phylogeny for labroid families remains somewhat uncertain as indicated by both the conflict in topologies between MP (ML supports the MP topology) and NJ analyses (figure 2*a*, *b*) and by the fact that only moderately high bootstrap values

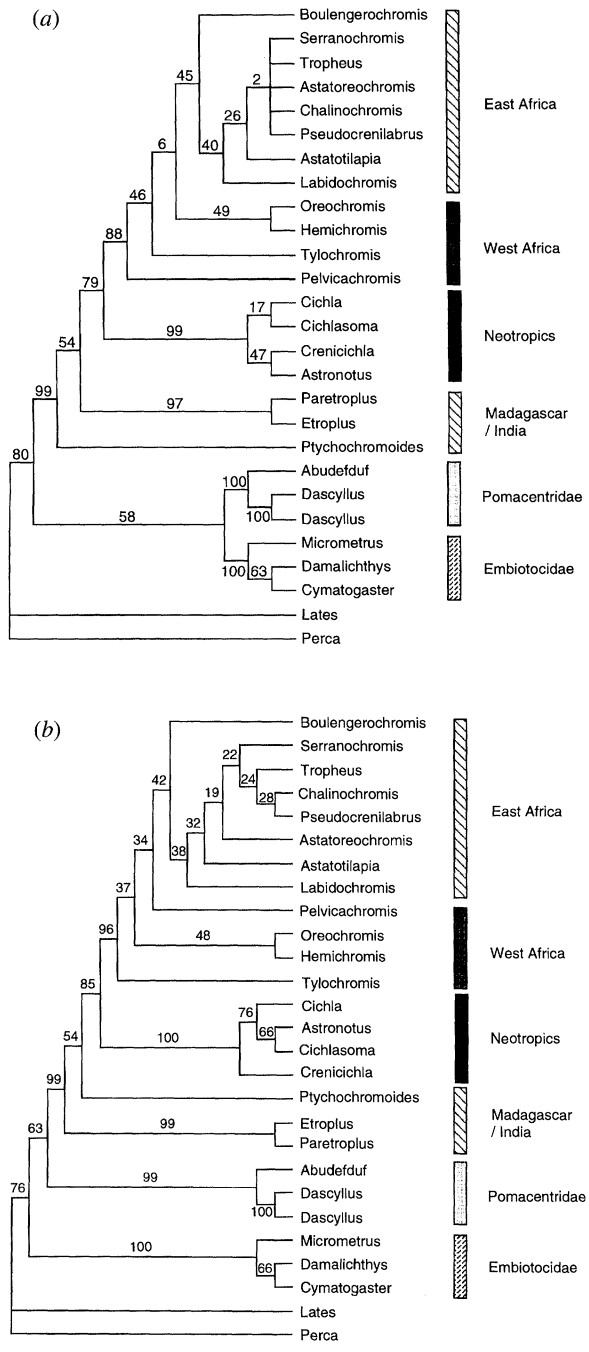

Figure 2. Molecular phylogenetic relationships among the principal lineages of cichlid fishes and two other families in the suborder Labroidei, the surfperches (Embiotocidae) and the damselfishes (Pomacentridae). The Nile perch (*Lates niloticus*) and the perch (*Perca fluviatilis*) were used as outgroups for the analyses. (*a*) 50% majority-rule maximum parsimony bootstrap consensus tree (and other groups compatible with it). Numbers above the branches indicate bootstrap support in 1000 replicates. (PAUP* version 4.0d46, D. Swofford, personal communication). (*b*) 50% majority-rule bootstrap neighbor-joining (Saitou & Nei 1987) consensus tree (and other groups compatible with it). Numbers above the branches indicate bootstrap support in 1000 pseudoreplicates (PAUP* version 4.0d46).

support the relationships among cichlids, surfperches and damselfishes (Felsenstein 1985) (figure 2).

The relationships among the East African cichlids and their relationships to other principal African

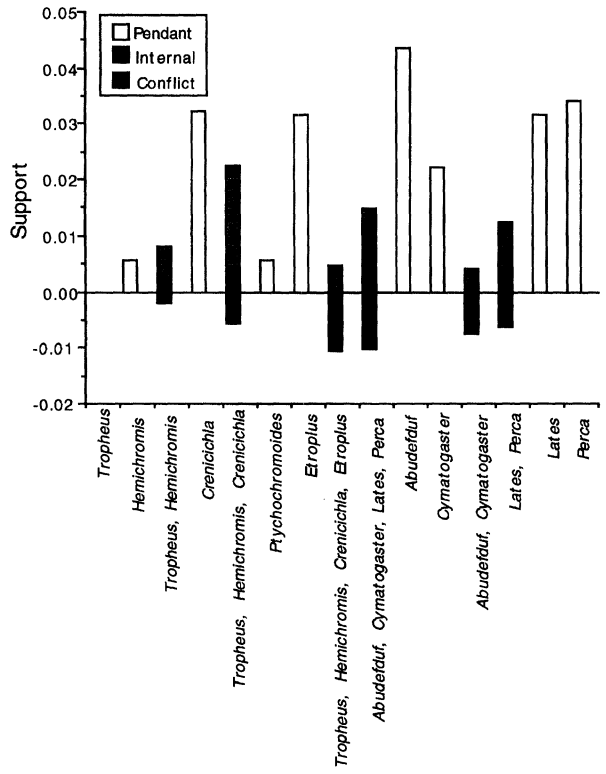

Figure 3. Hadamard log transform of the average bipartition spectrum for locus *TmoM*27 microsatellite DNA sequences for a subset of nine taxa (*Perca*, *Lates*, *Cymatogaster*, *Abudefduf*, *Ptychochromoides*, *Etroplus*, *Crenicichla*, *Hemichromis* and *Tropheus*). The support of internal (shaded) and pendant (white) splits is plotted above the zero point of the *y*-axis. The relative amounts of conflict (black) that each split has with the rest of the data are plotted below the zero point of the *y*-axis. Only those splits with estimated length (expected number of character-state changes per site) greater than 0.001 are shown.

cichlid lineages remain unsure based on *TmoM*27 MFR sequences because of the high level of evolutionary conservation at this locus. The East African lineages differed by typically less than 1.5% uncorrected sequence divergence (figure 1). Maximally 3.5% divergence was observed among all African, up to 7.7% among Neotropical, and up to 11% among the India and Madagascar cichlids (figure 1). These data support the notion from mtDNA data that East African lineages are exceedingly young and that the diversification of these lineages and adaptive radiation took place at an explosive evolutionary pace. It is unclear why African lineages appear to contain less than half the amount of genetic variation than was detected among Neotropical cichlid lineages. This observation needs to be substantiated with larger comparative MFR data sets, but it might suggest a younger age for African cichlids and/or slower rates of molecular evolution. At the family-level up to 15% sequence divergence was observed. Interestingly, among the Embiotocidae and Pomacentridae the amout of uncorrected sequence divergence among (the phyogenetically more limited) set of three species in each family was with maximally 3% much lower than that of the family Cichlidae. The rates of sequence evolution in MFRs of cichlids are 10–20 times slower than mitochondrial DNA sequence evolution (see, for example, Meyer 1993*a*) and around 0.05–0.1% per million years.

The fit of the MFR sequence data and the internal conflict of the data with the phylogeny inferred for labroid species was also examined by doing a spectral analysis (Hendy *et al.* 1994). Figure 3 shows the Hadamard transformation (Hendy *et al.* 1994) of the bipartition spectrum found from a simplified data set which included nine genera (*Perca*, *Lates*, *Cymatogaster*, *Abudefduf*, *Ptychochromoides*, *Etroplus*, *Crenicichla*, *Hemichromis* and *Tropheus*) (table 1). The support for internal and pendant splits is moderately high suggesting that *TmoM*27 MFR sequences are good indicators of the labroidei phylogeny. The splits show almost no conflict with the exception of the internal edge for the monophyly of the Cichlidae (figure 3).

(*d*) *Evolution of the microsatellite locus*

Interestingly, the highest repeat numbers of the microsatellite were observed among the embiotocids and some of the pomacentrids (figure 1). The most common type of imperfect CA repeat observed involved a transversion mutation to a CC (figure 1). In *Tropheus*, the East African species from which the microsatellite was originally isolated, it was found to be polymorphic among three individuals sequenced (interestingly, the microsatellite in these specimens is perfect with 20–25 repeats). Remarkably, the repeat numbers seem to be reduced to maximally four in all Neotropical cichlids in which it has been sequenced. *TmoM*27 was sequenced in six specimens from the Neotropical species *Cichlasoma citrinellum* and all were found to have an identical number ($n = 3$) of repeats.

4. DISCUSSION

Our results demonstrate for the first time that MFRs contain reliable phylogenetic information and are able to recover with considerable confidence the phylogenetic relationships within Cichlidae and other families of the suborder Labroidei. We tested six sets of PCR primers, of which only one, *TmoM*27, showed remarkable evolutionary conservation. The other five sets of PCR primers failed to amplify the homologous loci in all species of cichlid fishes that are more distantly related to *Tropheus* (table 1). The *TmoM*27 locus was amplified in several families of perciform fish including Cichlidae, Embiotocidae, Pomacentridae, Percidae and Centropomidae which shared a common ancestor that likely lived more than 100 Ma BP (the oldest known perciform fossil is from the late cretaceous, *ca.* 80–100 Ma (Patterson 1993*a*, *b*). The fact that this set of primers did not amplify in the Labridae does not necessarily imply that they are phylogenetically more distantly related to the three other families of fish from the suborder Labroidei as a single, phylogenetically spurious and autapomorphic mutation in the priming sequences of *TmoM*27 primers could be responsible. Rico *et al.* (1996) have previously found that there is no strong correlation between age of divergence (and hence taxonomic clustering) and the number of MFRs that show sequence conservation.

No DNA studies have, so far, attempted to relate the principal lineage of cichlids, including those from East Africa, West Africa, the Neotropics, Madagascar, and India. A randomly amplified polymorphic DNA based approach to isolate anonymous nuclear DNA (which did not contain microsatellites) from (mostly East) African cichlids was done to identify anonymous nuclear regions that were then used to reconstruct phylogenetic relationships among some lineages of African cichlids with Neotropical outgroups (Sültmann *et al.* 1995). That study supported the monophyly of tilapiine cichlids, African cichlids (*Heterochromis* was not included), and the sistergroup relationship among the Lake Malawi and Lake Victoria species flocks. It also suggested a basal position of *Tylochromis* among the African cichlids but focused mainly on the relationships among East African cichlids and not their relationships to other lineages of cichlids or other families of the suborder Labroidei (Sültmann *et al.* 1995).

The molecular phylogeny among the principal cichlid lineages here obtained based on the *TmoM*27 MFR overwhelmingly agrees with morphology-based cladistic hypotheses (Stiassny 1991). The absence of high levels of conflict in the splits of the bipartition spectrum and the results from the bootstrap analyses highlight that the *TmoM*27 MFR is evolving at appropriate levels for several of the phylogenetic questions posed here. Also unpublished preliminary analyses based on DNA sequence data from one partial (16S ribosomal RNA, about 450 bp) and three complete (cytochrome *b*, ATPase 6, ATPase 8, about 1900 bp) mitochondrial genes on the phylogeny of the principal lineages in the family Cichlidae (A. Meyer *et al.*, unpublished data) agree with Stiassny's phylogeny (Stiassny 1991) and the nuclear DNA phylogeny reported here. The phylogenetic relationships inferred for the family Cichlidae are consistent with the biogeographical distribution of the principal lineages of species and suggests that the cichlid radiation started prior to the splitting of the Gondwana supercontinent (160–150 Ma) and is in agreement with a vicariance hypothesis for the distribution of these freshwater fishes (for example, African and South American cichlids form monophyletic groups) (figure 2). The sistergroup relationship of the Madagascan genus *Paretroplus* with the Indian genus *Etroplus* implies that Madagascan and Indian cichlids are older than the separation of Madagascar from the Indian subcontinent, roughly 85–90 Ma (Storey *et al.* 1995). Because of uncertainties and the missing information from labrids the contested monophyly of and relationships among families in the suborder Labroidei (Kaufman & Liem 1982; Stiassny & Jensen 1987) still remains uncertain based on molecular data. The relationships among Neotropical, West and East African cichlid lineages respectively also remain tentative (figure 2) because of the high conservation of the *TmoM*27 locus and their resolution has to await the outcome of ongoing work on MFRs in cichlids (A. Meyer *et al.*, unpublished data). Whether the low level of sequence variation among African cichlids oberved for *TmoM*27 is typical is currently being investigated (A. Meyer *et al.*, unpublished data). Although the fossil evidence for divergence times among cichlid lineages is not strong enough to estimate rates of sequence evolution with confidence, our estimate compares favourably with other MFRs analysed (e.g. mammals, Schlötterer *et al.* 1991) and is only somewhat higher and probably not significantly different than that of other fishes (Rico *et al.* 1996).

The remarkable conservation of locus *TmoM*27 allowed us to document the evolution of the microsatellite contained in this locus across different species of perciform fishes. The *TmoM*27 microsatellite varies from only two repeats in some African and Neotropical cichlids to 64 repeats in one Embiotocidae. The propensity of fish microsatellites to be of greater length in terms of number of repeats than those of mammals has been previously postulated based on the analysis of several microsatellite loci in the Atlantic cod (*Gadus morhua*) and two salmonids (*Oncorhynchus mykiss* and *Salmo salar*) (Brooker *et al.* 1994). However, our results show that this propensity may not apply to all fish because there is much length variation between different families (figure 1). Interestingly, the analysis of these data in a phylogenetic context also reveals that there is no apparent phylogenetic trend in the transition from a perfect to an imperfect microsatellite (see figure 1). The *TmoM*27 microsatellite is reduced to only 2–4 repeat units in Neotropical cichlids. This result is surprising given that the four lineages of Neotropical cichlids are rather distantly related to each other and is in agreement with the idea that below a critical number of repeat units, expansion of short repeat sequences occurs relatively slowly (Messier *et al.* 1996).

Whether the tempo and mode of evolution observed for *TmoM*27 will turn out to be typical for MFRs in cichlids, other poikilotherms or even more inclusive taxonomic groups will require work on more MFRs in this and other sets of species of differing levels of phylogenetic inclusiveness. The congruence between this new nuclear DNA marker with previous findings from other nuclear DNA markers, mitochondrial DNA and morphological data are encouraging. The DNA sequences from *TmoM*27 are the first molecular data set to establish with confidence the relationships among the principal lineages of cichlid fishes.

We thank Ciro Rico and Thomas Kocher who provided valuable comments on the manuscript. Eric Schultz, Melanie Stiassny, Peter Reinthal, Andreas Spreinat, Paul Loiselle, Lothar Seegers, Luc de Vos, Irv Kornfield, Frietson Galis, and Christian Sturmbauer generously donated samples. Michael A. Charleston kindly helped with the spectral analysis. Katharina Noack from Stony Brook kindly provided unpublished Victoria cichlid and Nile perch microsatellite flanking region DNA sequences. David Swofford granted permission to publish results based on the test version of his PAUP* program. R. Z. is sponsored by a postdoctoral grant of the Ministerio de Educacion y Ciencia of Spain. We are grateful to the National Science Foundation, U.S.A. (grants DEB- 8918027, BSR-9107838, and BSR-9119867) and the Max-Planck-Society, Gemany for financial support to A. M. This publication was prepared during A. M.'s tenure as Guggenheim Fellow and Miller Visiting Research Professor at the University of California at Berkeley. The support and the generous hospitality of the Miller Institut and the Museum of Vertebrate Zoology is gratefully acknowledged.

REFERENCES

Boulenger, G. A. 1898 A revision of the African and Syrian fishes of the family Cichlidae. Part 2. *Proc. Zool. Soc. Lond.* 132–152.

Brooker, A. L., Cook, D., Bentzen, P., Wright, J. M. & Doyle, R. W. 1994 Organization of microsatellites differs between mammals and cold-water teleost fishes. *Can. J. Fish. Aquat. Sci.* **51**, 1959–1966.

Bruford, M. W. & Wayne, R. K. 1993 Microsatellites and their application to population genetic studies. *Curr. Opin. Genet. Dev.* **3**, 939–943.

Charleston, M. A. 1995 *Spectrum.* (Version 1.01 for Macintosh.)

Cichocki, F. 1976 Cladistic history of cichlid fishes and reproductive categories of the American genera *Acarichthys*, *Biotodoma*, and *Geophagus*. Ann Arbor: University of Michigan.

Echelle, A. A. & Kornfield, I. (eds.) 1984 *Evolution of fish species flocks*. Orono, Maine: University of Maine Press.

Felsenstein, J. 1985 Confidence limits on phylogenies: an approach using the bootstrap. *Evolution* **39**, 783–791.

FitzSimmons, N. N., Moritz, C. & Moore, S. S. 1995 Conservation and dynamics of microsatellite loci over 300 million years of marine turtle evolution. *Molec. Biol. Evol.* **12**, 432–440.

Fryer, G. & Iles, T. D. 1972 *The cichlid fishes of the East Aftican Lakes*. T. D.F.: TFH Publications.

Hamada, H., Petrino, M. G. & Kakunaga, T. 1982 A novel repeated element with Z-DNA-forming potential is widely found in evolutionarily diverse eukaryotic genomes. *Proc. natn. Acad. Sci. U.S.A.* **79**, 6465–6469.

Hendy, M. D., Penny, D. & Steel, M. A. 1994 A discrete fourier analysis for evolutionary trees. *Proc. natn. Acad. Sci. U.S.A.* **91**, 3339–3343.

Kaufman, L. & Liem, K. F. 1982 Fishes of the suborder Labroidei (Pisces: Perciformes): phylogeny, ecology, and evolutionary significance. *Brevoria* **472**, 1–19.

Kellog, K. A., Markert, J. A., Stauffer, J. R. & Kocher, T. D. 1995 Microsatellite variation demonstrates multiple paternity in lekking cichlid fishes from Lake Malawi, Africa. *Proc. R. Soc. Lond.* B **260**, 79–84.

Kocher, T. D., Conroy, J. A., McKaye, K. R. & Stauffer, J. R. 1993 Similar morphologies of cichlid fish in Lakes Tanganyika and Malawi are due to convergence. *Molec. Phyl. Evol.* **2**, 158–165.

Kocher, T. D., Conroy, J. A., McKaye, K. R., Stauffer, J. R. & Lockwood, S. F. 1995 Evolution of NADH dehydrogenase subunit 2 in East African cichlid fish. *Molec. Phyl. Evol.* **4**, 420–432.

Liem, K. F. & Greenwood, P. H. 1981 A functional approach to the phylogeny of pharyngognath teleosts. *Am. Zool.* **21**, 83–101.

Litt, M. & Luty, J. A. 1989 A hypervariable microsatellite revealed by in vitro amplification of a dinucleotide repeat within the cardiac muscle actin gene. *Am. J. hum. Genet.* **44**, 397–401.

Messier, W., Li, S. H. & Stewart, C. B. 1996 The birth of microsatellites. *Nature, Lond.* **381**, 483.

Meyer, A. 1993*a* Evolution of mitochondrial DNA in fishes. In *Molecular biology frontiers, biochemistry and molecular biology of fishes*, vol. 2 (ed. P. W. Hochachka & T. P. Mommsen), pp. 1–38. Oxford: Elsevier Science Publishers.

Meyer, A. 1993*b* Phylogenetic relationships and evolutionary processes in East African cichlid fishes. *Trends Ecol. Evol.* **8**, 279–284.

Meyer, A. 1994 DNA technology and phylogeny of fish: molecular phylogenetic studies of fish. In *Genetics and evolution of aquatic organisms* (ed. A. R. Beaumont), pp. 219–249. London: Chapman and Hall.

Meyer, A., Kocher, T. D., Basasibwaki & Wilson, A. C. 1990 Monophyletic origin of Lake Victoria cichlid fishes suggested by mitochondrial DNA sequences. *Nature, Lond.* **347**, 550–553.

Meyer, A., Kocher, T. D. & Wilson, A. C. 1991 African fishes. *Nature, Lond.* **350**, 467–468.

Meyer, A., Montero, C. M. & Spreinat, A. 1996 Molecular phylogenetic inferences about the evolutionary history of the East African cichlid fish radiations. In *IDEAL (International Decade of East African Lakes): the limnology, climatology and paleoclimatology of the East African lakes* (ed. T. Johnson & E. Odada), pp. 303–323. Gordon and Breach Scientific Publishers.

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

Olivier, M. K. 1984 Systematics of African cichlid fishes: determination of the most primitive taxon, and studies on the haplochromines of Lake Malawi (Teleostei, Cichlidae). New Haven, Connecticut: Yale University Press.

Owen, R. B., Crossley, R., Johnson, C. *et al.* 1990 Major low levels of Lake Malawi and implication for speciation rates in cichlid fishes. *Proc. R. Acad. Sci.* B **240**, 513–553.

Patterson, C. 1993*a* Osteichthyes: Teleostei. In *The fossil record 2* (ed. M. J. Benton), pp. 621–656. London: Chapman & Hall.

Patterson, C. 1993*b* An overview of the early fossil record of acanthomorphs. *Bull. Mar. Sci.* **52**, 29–59.

Pellegrin, J. 1903 Contribution a l'etude anatomique, biologique, et taxonomique des poissons de la familes es cichlides. *Mem. Soc. Zool. France* **16**, 41–402.

Queller, D. C., Strassmann, J. E. & Hughes, C. R. 1993 Microsatellites and kinship. *Trends Ecol. Evol.* **8**, 285–288.

Regan, C. T. 1920 The classification of the fishes of the family Cichlidae. I. The Tanganyikan genera. *A. Mag. nat. Hist.* **5**, 33–53.

Regan, C. T. 1922 The classification of the fishes of the family Cichlidae. II. On African and Syrian genera not restricted to the Great Lakes. *A. Mag. nat. Hist.* **10**, 249–264.

Rico, C., Rico, I. & Hewitt, G. 1996 470 million years of conservation of microsatellite loci among fish species. *Proc. R. Soc. Lond.* B **263**, 549–557.

Saitou, N. & Nei, M. 1987 The neighbor-joining method: a new method for reconstructing phylogenetic trees. *Molec. Biol. Evol.* **4**, 406–425.

Schlötterer, C., Amos, B. & Tautz, D. 1991 Conservation of polymorphic simple sequence loci in cetacean species. *Nature, Lond.* **354**, 63–65.

Schlötterer, C. & Tautz, D. 1992 Slippage synthesis of simple sequence DNA. *Nucleic Acids Res.* **20**, 211–215.

Stallings, R. L., Ford, A. F., Nelson, D., Torney, D. C., Hildebrand, C. E. & Moyzis, R. K. 1991 Evolution and distribution of (GT)n repetitive sequences in mammalian genomes. *Genomics* **10**, 807–815.

Stiassny, M. L. J. 1981 The phyletic status of the family Cichlidae (Pisces: Perciformes): a comparative anatomical investigation. *Neth. J. Zool.* **31**, 275–314.

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

Stiassny, M. L. J. 1991 Phylogenetic intrarelationships of the family Cichlidae: an overview. In *Cichlid fishes: behavior, ecology and evolution* (ed. M. H. A. Keenleyside), pp. 1–31. London: Croom Helm.

Stiassny, M. L. J. & Jensen, J. S. 1987 Labroid inter-

relationships revisited: morphological complexity, 'key innovations', and the study of comparative diversity. *Bull. Mus. comp. Zool.* **151**, 269–319.

Storey, M., Mahoney, J. J., Saunders, A. D., Duncan, R. A., Kelley, S. P. & Coffin, M. F. 1995 Timing of hot spot-related volcanism and the breakup of Madagacar and India. *Science, Wash.* **267**, 852–855.

Sturmbauer, C. & Meyer, A. 1993 Mitochondrial phylogeny of the endemic mouthbrooding lineages of cichlid fishes from Lake Tanganyika in Eastern Africa. *Molec. Biol. Evol.* **10**, 751–768.

Sturmbauer, C., Verheyen, E. & Meyer, A. 1994 Mitochondrial phylogeny of the lamprologini, the major substrate spawning lineage of cichlid fishes from Lake Tanganyika in Eastern Africa. *Molec. Biol. Evol.* **11**, 691–703.

Sültmann, H., Maer, W. E., Figueroa, F., Tichy, H. & Klein, J. 1995 Phylogenetic analysis of cichlid fishes using nuclear DNA markers. *Molec. Biol. Evol.* **12**, 1033–1047.

Swofford, D. L. 1993 *PAUP: Phylogenetic Analysis Using Parsimony*. Champagne, Illinois: Illinois Natural History Survey.

Tautz, D. 1989 Hypervariability of simple sequences as a general source for polymorphic DNA markers. *Nucleic Acids Res.* **17**, 6463–6471.

Tautz, D., Trick, M. & Dover, G. A. 1986 Cryptic simplicity in DNA is a major source of genetic variation. *Nature, Lond.* **322**, 652–656.

Thompson, J. D., Higgins, D. G. & Gibson, T. J. 1994 CLUSTAL W: improving the sensitivity of pregressive multiple sequence alignment through sequence weighting, position specific gap penalties and weight matrix choice. *Nucleic Acids Res.* **22**, 4673–4680.

Weber, J. L. 1990 Informativeness of human (dC-dA)n dot (dG-dT)n polymorphisms. *Genomics* **7**, 524–530.

Weber, J. L. & May, P. E. 1989 Abundant class of human DNA polymophisms which can be typed using the polymerase chain reaction. *Am. J. hum. Genet.* **44**, 388–396.

Received 5 August 1996; accepted 29 August 1996