Phylogenetic Relationships and Ancient Incomplete Lineage Sorting Among Cichlid Fishes in Lake Tanganyika as Revealed by Analysis of the Insertion of Retroposons

Kazuhiko Takahashi, Yohey Terai,* Mutsumi Nishida,† and Norihiro Okada**

*Graduate School of Bioscience and Biotechnology, Tokyo Institute of Technology, Yokohama, Japan; and †Division of Molecular Marine Biology, Ocean Research Institute, University of Tokyo, Tokyo, Japan

Lake Tanganyika harbors numerous endemic species of extremely diverse cichlid fish that have been classified into 12 major taxonomic groups known as tribes. Analysis of short interspersed element (SINE) insertion data has been acknowledged to be a powerful tool for the elucidation of phylogenetic relationships, and we applied this method in an attempt to clarify such relationships among these cichlids. We studied insertion patterns of 38 SINEs in total, 24 of which supported the monophyly of three clades. The other 14 loci revealed extensive incongruence in terms of the patterns of SINE insertions. These incongruencies most likely stem from a period of adaptive radiation. One possible explanation for this phenomenon is the extensive incomplete lineage sorting of alleles for the presence or absence of a SINE during successive speciation events which took place about 5–10 MYA. The present study is the first to report the successful application of the SINE method in demonstrating the existence of such possible ''ancient'' incomplete lineage sorting. We discuss the possibility that it might potentially be very difficult to resolve the species phylogeny of a group that radiated explosively, even by resolving the genealogies of more than 10 nuclear loci, as a consequence of incomplete lineage sorting during speciation.

Introduction

The African Great Lakes consist of Lakes Victoria, Tanganyika, and Malawi, and each lake harbors a large number of endemic species of cichlid fish (Fryer and Iles 1972; Greenwood 1984; Coulter 1991). The fish in each lake exhibit remarkable diversity in terms of morphology, ecology, and behavior, and this diversity was acquired through independent and explosive adaptive radiation. Lake Tanganyika is the oldest of these lakes, with an estimated age of 9–12 Myr (Cohen, Soreghan, and Scholz 1993), and cichlids in this lake exhibit the widest morphological diversity (Greenwood 1984). Indeed, it has been suggested that Lake Tanganyika is an evolutionary reservoir of multiple ancient lineages, one of which includes the entire genetic diversity of cichlids in Lakes Victoria and Malawi (Nishida 1991; Kocher et al. 1993, 1995; Sturmbauer and Meyer 1993; Sturmbauer, Verheyen, and Meyer 1994; Nishida 1997; Mayer, Tichy, and Klein 1998).

The cichlid species in Lake Tanganyika have been classified into 12 tribes (Poll 1986). Molecular phylogenetic studies in recent years have resolved some aspects of their phylogeny by exploring various markers, such as allozymes (Nishida 1991, 1997), the control region of mitochondrial DNA (Kocher et al. 1993; Sturmbauer and Meyer 1993; Sturmbauer, Verheyen, and Meyer 1994), the genes for cytochrome *b* (Sturmbauer and Meyer 1993; Sturmbauer, Verheyen, and Meyer 1994) and subunit 2 of NADH dehydrogenase (ND2; Kocher et al. 1995) in the mitochondrial genome, and noncoding regions in the nuclear genome (Sültmann et al. 1995; Mayer, Tichy, and Klein 1998). Nevertheless, some aspects of their phylogenetic relationships remain unresolved. Moreover, even when some relationships are supported by high bootstrap values in analyses using single-locus markers, such as genes in mitochondrial DNA, the possibility of misinterpretation of phylogeny remains, since the tree obtained is only a gene tree, which might differ from the species tree as a result of incomplete lineage sorting of ancestral polymorphisms during successive rounds of speciation (Nei 1987; Pamilo and Nei 1988; Takahata 1989; Avise 2000) and/or interspecific hybridization (Avise 2000). We cannot ignore this possibility, particularly when we attempt to infer the phylogeny of cichlids in the African Great Lakes, since extensive incomplete lineage sorting has already been reported in studies of cichlid flocks in both Lake Malawi (Moran and Kornfield 1993, 1995; Parker and Kornfield 1997; Albertson et al. 1999; Takahashi et al. 2001) and Lake Victoria (Nagl et al. 1998). Thus, the phylogeny of the cichlids in Lake Tanganyika needs to be reevaluated.

In recent years, short interspersed elements (SINEs) have been shown to be powerful phylogenetic markers (Murata et al. 1993, 1996; Shimamura et al. 1997; Hamada et al. 1998*a*; Takahashi et al. 1998, 2001; Nikaido, Rooney, and Okada 1999). These elements are retroposons, and they multiply in genomes via the reverse transcription of RNA that has been transcribed from a parental sequence (Weiner, Deininger, and Efstratiadis 1986). The random choice of the sites of integration of SINEs and the irreversible nature of the integration at each site are extremely useful in efforts to reconstruct phylogeny, since homoplasy is minimized (Okada 1991; Cook and Tristem 1997; Miyamoto 1999; Shedlock,

Abbreviations: AFC family, African cichlid family; MVH clade, the clade formed by cichlids of Lakes Malawi and Victoria and Haplochromini; MVhL clade, the clade formed by cichlids of Lakes Malawi and Victoria, the H-lineage, and Lamprologini; MVHT clade, the clade formed by cichlids of Lakes Malawi and Victoria, Haplochromini, and Tropheini; SINE, short interspersed element.

Key words: adaptive radiation, incomplete lineage sorting, Lake Tanganyika, cichlid, retroposon, SINE.

Address for correspondence and reprints: Norihiro Okada, Graduate School of Bioscience and Biotechnology, Tokyo Institute of Technology, 4259 Nagatsuta-cho, Midori-ku, Yokohama 226-8501, Japan. E-mail: nokada@bio.titech.ac.jp.

Mol. Biol. Evol. 18(11):2057–2066. 2001

Table 1
Fish Species Analyzed by PCR in this Study

Species	Geographic Distribution	Tribe
Cyrtocara moorii	Lake Malawi	—
Haplochromis nyererei	Lake Victoria	—
Astatotilapia burtoni	Lake Tanganyika	Haplochromini
Petrochromis fasciolatus	Lake Tanganyika	Tropheini
Eretmodus cyanostictus	Lake Tanganyika	Eretmodini
Perissodus microlepis	Lake Tanganyika	Perissodini
Paracyprichromis brieni	Lake Tanganyika	Cyprichromini
Limnochromis staneri	Lake Tanganyika	Limnochromini
Cyathopharynx furcifer	Lake Tanganyika	Ectodini
Lamprologus lemairii	Lake Tanganyika	Lamprologini
Boulengerochromis microlepis	Lake Tanganyika	Tilapiini
Trematocara unimaculatum	Lake Tanganyika	Trematocarini
Bathybates fasciatus	Lake Tanganyika	Bathybatini
Tylochromis polylepis	Lake Tanganyika	Tylochromini

NOTE.—The nomenclature of each species follows that of Poll (1986). The *H. nyererei* specimen was purchased from a commercial source in Japan; all the other specimens were collected in the field.

Milinkovitch, and Okada 2000; Shedlock and Okada 2000). The above-mentioned characteristics of retroposons should also be useful for detection of incomplete lineage sorting (Hamada et al. 1998*b*; Shedlock, Milinkovitch, and Okada 2000; Shedlock and Okada 2000). In the present study, we isolated many members of the AFC family of SINEs, a family that has previously been described in cichlids (Takahashi et al. 1998; Terai, Takahashi, and Okada 1998), and used them to investigate the phylogenetic relationships among the ancient lineages of cichlid fish in Lake Tanganyika. We also investigated the existence and extent (if any) of incomplete lineage sorting of ancestral polymorphisms among these fish. This is the first comprehensive report, to our knowledge, to indicate that the analysis of SINE insertions is effective in attempts to discover possible ''ancient'' incomplete lineage sorting.

Materials and Methods

We examined one representative species from each of the various tribes of cichlids in Lake Tanganyika, as well as from each of Lakes Malawi and Victoria (table 1). Thirty-eight loci at which AFC SINEs had been inserted were isolated randomly from genomic libraries of *Lepidiolamprologus elongatus* (loci 2 and 4), *Ophthalmotilapia ventralis* (loci 213–259), *Tropheus moorii* (loci 305–343), *Haplotaxodon microlepis* (loci 452–455), *Dimidiochromis compressiceps* (loci 1201–1569), *Melanochromis jonjohnsonae* (loci 1654–1666), and *Haplochromis machadoi* (locus 1715). On the basis of these loci, we designed primers that flanked each SINE unit (table 2) and performed polymerase chain reactions (PCRs) with genomic DNA from various cichlids as templates. The phylogenetic tree was constructed on the basis of the presence or absence of a SINE unit at the individual orthologous loci. Experiments were performed by standard techniques as reported previously (Takahashi et al. 1998).

Results

Monophyly of Three Clades

Figure 1 shows representative results of PCR experiments. In the case of locus 1715, longer products containing one unit of the AFC SINE were amplified from the genomes of 10 species that represented lineages in Lakes Malawi and Victoria and the tribes Haplochromini, Tropheini, Eretmodini, Perissodini, Cyprichromini, Limnochromini, Ectodini, and Lamprologini. By contrast, four species from the tribes Tilapiini, Trematocarini, Bathybatini, and Tylochromini yielded shorter products without the SINE unit (fig. 1-I*A*, panel *a*). To confirm these results, we performed Southern hybridization experiments using two kinds of probes. The first hybridization experiment, performed with a probe specific for members of the AFC family, confirmed that the longer PCR products contained a unit of this family (fig. 1-I*A*, panel *b*). In the second hybridization experiment, with a probe specific for the flanking region of the unit at locus 1715, we confirmed that all of the longer and shorter products had faithfully been amplified from the orthologous locus (fig. 1-I*A*, panel *c*). Thus, the unit of the AFC family appeared to be specifically shared among the 10 species that yielded the longer products. A similar pattern of insertion of a sequence of an AFC SINE was also observed for locus 245 (fig. 1-I*B*) except that a shorter product was generated from the genome of *Cyrtocara moorii*. This result was thought to be due to deletion of a non-SINE sequence rather than to the absence of the SINE sequence, since a signal was detected after hybridization with the SINE probe (fig. 1-I*B*, panel *b*). It was further confirmed by sequencing of the shorter product. In addition, the pattern obtained from locus 254 (fig. 1-I*C*) also revealed similar results, with the exception of the absence of a PCR product from the genome of *Perissodus microlepis*. The presence of sequences of the AFC family in the 10 species exclusively was confirmed by the analysis of an additional seven loci (213, 214, 247, 314, 455, 1569, and 1666; table 3), strongly

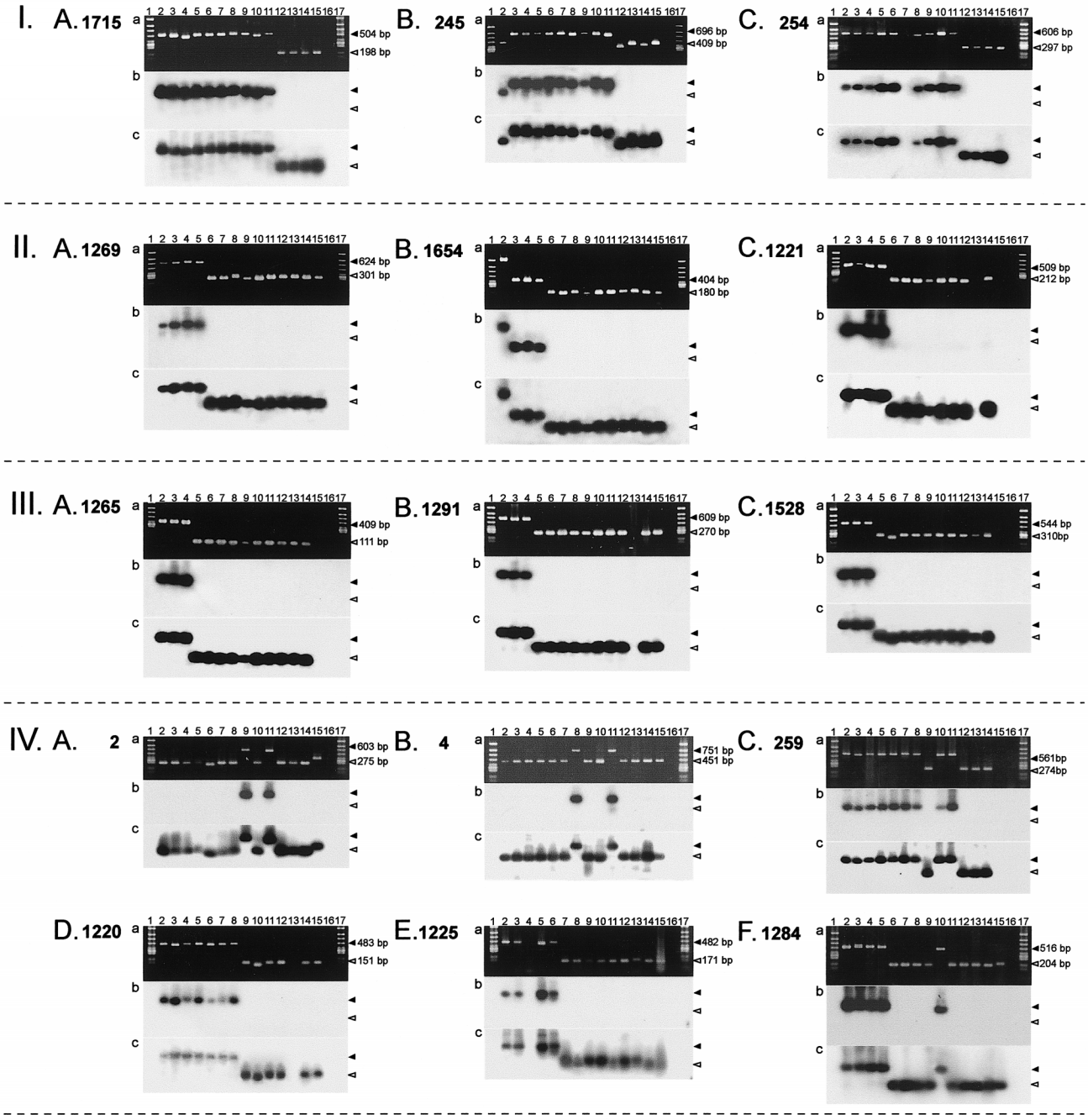
Table 2
Sequences of Primers Used for PCRs

Locus	5′-Flanking	3′-Flanking
2	5′-AAGTATGGCAGAACAGAGAAGGC-3′	3′-GGAACCACACAACAACTCAATG-5′
4	5′-ACTTGAAACTAGAGACTGAGGCCAC-3′	3′-CTCAGAATCGAAGGATCATAATCGGAGG-5′
213	5′-AATCCCACAATGATATGTCCTA-3′	3′-CTACGTGTGAGGTAAGTGTG-5′
	5′-GAAGCAAGTCCATTTTAAAGTTAACAGAC-3′	3′-ACGAATGTGTCGACAACTACAAAG-5′
214	5′-GACAGCTTTCACTGCTATTATG-3′	3′-AGGACTAATCGTCTCGGTTGA-5′
	5′-ATTAGGCAAAATGGGGATATCCA-3′	3′-ATCGTCTCGGTTGAGTGTTAAA-5′
	5′-TATCCAGGTGACAGGGGAAC-3′	
216	5′-GTTACTCCATCATCTAAGCAAGT-3′	3′-ATCCACTGGGTCTCAAACTC-5′
225	5′-CTGGTTGTCACAGTCCAATGG-3′	3′-CCACACCAGGAACACATTATATA-5′
230	5′-TCAGCTTCATATCTACACAAGGATTG-3′	3′-CGTCAAATACTTACGTAGACACAA-5′
245	5′-CAGCACAGAAGCTGACTTAGTTA-3′	3′-TTGGTAGTACAGTGAAGACGA-5′
247	5′-GCATTATGTTTATGGATCAGGTT-3′	3′-TGGACTAATGTATGCTCGTCAAT-5′
	5′-ACCTTTTGTCAGTCACTTTGC-3′	3′-TCATTTGTCGTGTTGGTCCAA-5′
254	5′-AAGCAAGCTGAAACATCTTTG-3′	3′-TACTAATTACGGACGTGATCC-5′
259	5′-AGACACTGCAATCAGCCTT-3′	3′-GTTATAGAGTCGAGTGTGTCC-5′
	5′-TCCCTCTCTAATTTTCTCAGCCAT-3′	3′-TAGATGTCAGTCCGAAGTTAGTTTA-5′
305	5′-CAGAACTTCCACATCCACCATAG-3′	3′-CCCTGACAACTTCCATTATGT-5′
314	5′-GCAAACTTTCCTTCCTCCTCTC-3′	3′-AAGTGGACTACGTGTTCTGACC-5′
		3′-TACGTCTCCTGGAAGGTAGAAT-5′
328	5′-GCGACTTTGAGCAGAATAATCAG-3′	3′-GGATCTATTGTAGGGACTGA-5′
	5′-CTCAGTTTGGTGGTTCCTACAGA-3′	3′-GTAAAGGGAACGAAGCAGAAC-5′
330	5′-ATCAAAGAATAACCCGCCAA-3′	3′-TACTGAGTGGATTACAATCTCTAT-5′
	5′-CAAATAATCTGCGGTCATATCCATAC-3′	3′-GCTCAGGGACAAACAGAGAAATTT-5′
343	5′-CTGAAACGGAGGAATGGCTTC-3′	3′-ACGGAGAACGACGTAGACACTGTGGCTA-5′
	5′-TGTTAGCTTCACTTATTTCACCATAGT-3′	3′-AGTAGATTATGATGTAAAGTAACAAACAD-5′
452	5′-AGAGCATTTGGCGAGCTGTA-3′	3′-TGTCCCAGATACCTTCGACTTAA-5′
455	5′-GTAGGACTGCTTTCTTCCATTTGTG-3′	3′-GTTGTTGAACCCTTAGGAGACG-5′
1201	5′-GAGTCTCTCAATGAGTTCTCACT-3′	3′-CGTCTCTATTCTACTCTACGAG-5′
1220	5′-CCATAGATGCAGGGAGGGACA-3′	3′-CTTGTTCACCGATACCTTTAACG-5′
1221	5′-GACCTGAATTGGAAAACTATCAC-3′	3′-GTCCGACCATAAGCATGACA-5′
1225	5′-GAGATAGTGACGGGATGAGGG-3′	3′-CACCGTACATCCACCTCACCAC-5′
1233	5′-TGCTTGAGCTGTTGTTGGTC-3′	3′-GTTTCTTCGAGTAACTTCTTGTAAAG-5′
1238	5′-TAACAGTGGTTTGGATATTTGTGC-3′	3′-GATACTAGCGGAAACTGAGGATAC-5′
	5′-TGATGTCATCTCAATTGCTGTCATT-3′	3′-AGTCTAGGTCAAGTGTCAGGAGT-5′
1245	5′-GGAGGAGTAGCAGCTCTAGTCTGAG-3′	3′-CCGTAACCATTATTCTCGCAAAATC-5′
1262	5′-TCTCTTCAGGAGGTGCCACA-3′	3′-TCTAATGAGACATCGGAAATTGGTC-5′
	5′-TGGTCTTCCTCTTTTCATCCTG-3′	3′-GTATCGTCGAAATTACACCTTTTATT-5′
1265	5′-GCTGGGCTTTTCTTTATAGTTACGTTCA-3′	3′-AGACTGTTTTATACGGCCTCACG-5′
1269	5′-GAGAGGGATAAATGGTGCACTG-3′	3′-GTCCTGAATGGGTAAGTAGAACT-5′
1277	5′-CAAATATACCAGGATTGGATTGAG-3′	3′-GCAAGTAAGAGCAAATGCACGG-5′
		3′-AGTACAGGTTCGTAGTGGAAAT-5′
1281	5′-ATGCCCTCACCTGAACTGTTG-3′	3′-ATCGTTCAATCGTACACCGTATGT-5′
1284	5′-ATTCTTCTTCTTCCATCAACTGTCATT-3′	3′-CTCGCATCCGAAACTATTACTAC-5′
1291	5′-ACTATGAAGAAACCACCATGTGC-3′	3′-GTAGTGTGTTCCCTCATAACGAA-5′
1516	5′-ACAGGCAGACCTCCATTGTCA-3′	3′-GACACATAACAACCGCTCAACTC-5′
1528	5′-ATTCAACCAGTCCTATTCAATATTTACTC-3′	3′-GTGTGTCGCTCTTTCTCGTCT-5′
1569	5′-AGTGTTACCAGATGTGCTGATG-3′	3′-CTTTGTAATCTTGTTAGCAGAATTGAT-5′
		3′-AAATCACGTGTAGTGTTGCTTTA-5′
1654	5′-TGAGGCGTAGTGTGGGTAATAGC-3′	3′-TGTCCGTGTCCTTTCTCGAAAC-5′
1666	5′-TTTAAACAGAGTAGCCTGCAAAGT-3′	3′-CATAAGAGTCGAGTAGAAGGTGAC-5′
	5′-CCATCACTAGGCCTTAGACTATGAG-3′	3′-TCAAGTTGGATCGTGTGTCGT-5′
1715	5′-TTACAGAGAACGGAAGGGAAAG-3′	3′-AGAAGAGGGTAAAGTGATTAATCCTT-5′

NOTE.—Loci accompanied by more than two kinds of primers were amplified by PCRs using different combinations of primers, depending on species.

supporting the monophyly of these 10 species. This monophyletic clade included species from Lakes Malawi and Victoria, the seven tribes that belong to the ''H-lineage'' suggested by Nishida (1991; Haplochromini, Tropheini, Eretmodini, Perissodini, Cyprichromini, Limnochromini, and Ectodini), and the tribe Lamprologini. We designated this clade the ''MVhL'' clade for convenience (we use the lowercase letter ''h'' to abbreviate ''H-lineage'' to avoid confusion with the ''H'' that is used later to abbreviate the name of the tribe Haplochromini).

In a similar manner, we identified the MVHT and the MVH clades. The MVHT clade consisted of *C. moorii* (Lake Malawi), *Haplochromis nyererei* (Lake Victoria), *Astatotilapia burtoni* (tribe Haplochromini), and *Petrochromis fasciolatus* (tribe Tropheini), all of which exclusively shared SINE units at loci 328, 330, 343, 1221, 1238, 1262, 1269, 1277, and 1654 (fig. 1-II and table 3). The MVH clade consisted of *C. moorii*, *H. nyererei*, and *A. burtoni*, all of which specifically shared SINE units at loci 1233, 1265, 1281, 1291, and 1528 (fig. 1-III and table 3). The putative MVhL,



1 Markers	7 *Perissodus microlepis* (Perissodini)	13 *Trematocara unimaculatum* (Trematocarini)
2 *Cyrtocara moorii* (Lake Malawi)	8 *Paracyprichromis brieni* (Cyprichromini)	14 *Bathybates fasciatus* (Bathybatini)
3 *Haplochromis nyererei* (Lake Victoria)	9 *Limnochromis staneri* (Limnochromini)	15 *Limnochromis staneri* (Limnochromini)
4 *Astatotilapia burtoni* (Haplochromini)	10 *Cyathopharynx furcifer* (Ectodini)	16 Negative control
5 *Petrochromis fasciolatus* (Tropheini)	11 *Lamprologus lemairii* (Lamprologini)	17 Markers
6 *Eretmodus cyanostictus* (Eretmodini)	12 *Boulengerochromis microlepis* (Tilapiini)	

FIG. 1.—Representative patterns obtained during the analysis of PCR products for detection of orthologous loci in cichlid species. Each group of panels designated A–C in rows I–III and A–F in row IV contains photographs of an agarose gel and two autoradiograms: (*a*) electrophoretic profile of products of PCR; (*b*) results of Southern hybridization of these products, designed to detect an AFC short interspersed element (SINE) sequence; (*c*) rehybridization of the same blot for detection of a locus-specific sequence flanking the site at which the SINE unit was inserted. Closed and open arrowheads indicate expected mobilities of amplified fragments with and without insertion of a SINE, respectively.

MVHT, and MVH clades were consistent with one another, since the entire MVH clade was part of the MVHT clade and the entire MVHT clade was included in the MVhL clade. These observations enabled us to construct a single phylogenetic tree that resolved the three clades (fig. 2).

Extensive Incongruence in Patterns of SINE Insertions Among Specific Lineages

Although the patterns of insertion of AFC SINEs observed at the 24 loci analyzed above unambiguously indicated support for the three monophyletic groups of cichlids, our new tree contained three sets of polytomies. One of these polytomies (the red portion in fig. 2) was due to incongruent patterns of insertion of AFC SINEs at 14 loci (loci 2, 4, 216, 225, 230, 259, 305, 452, 1201, 1220, 1225, 1245, 1284, and 1516). Six examples of such patterns are shown in figure 1-IV. In the case of locus 2, insertion of an AFC sequence was observed specifically in the genomes of *Limnochromis staneri* and *Lamprologus lemairii*, representatives of the Limnochromini and Lamprologini tribes, respectively (fig. 1-IV*A*). However, when results for locus 4 were taken into account, these species did not seem to be sister species, since at this locus *L. lemairii* shared insertion of an AFC SINE with *Paracyprichromis brieni* (tribe Cyprichromini), with the genome of *L. staneri* lacking this unit (fig. 1-IV*B*). The pattern obtained from analysis of locus 259 also contradicted the results obtained for locus 2, since at locus 259 the lineage leading to *L. staneri* appeared to have diverged first in the MVhL clade (fig. 1-IV*C*). By contrast, analysis of locus 1220 revealed a conflict with the close relationship between *L. lemairii* and *P. brieni*, which was suggested by the analysis of locus 4, because the latter species shared an AFC SINE with species from the MVHT clade as well as *Eretmodus cyanostictus* (tribe Eretmodini) and *P. microlepis* (tribe Perissodini), whereas the unit was not shared by *L. lemairii* (fig. 1-IV*D*). Another example of such a conflict was observed in the case of patterns obtained for locus 1225 (fig. 1-IV*E*) and locus 1284 (fig. 1-IV*F*), which revealed a discrepancy in terms of the species—either *E. cyanostictus* or *Cyathopharynx furcifer* (tribe Ectodini)—that is closest to the MVHT clade. Further examples of ''incongruent'' patterns of insertion of SINEs in the AFC family are shown in table 3. Insertion of SINE units at such ''incongruent'' loci appears to have occurred specifically at a certain period during the evolution of the various species.

Confirmation of Fixation of Alleles with or Without an AFC SINE

To exclude the possibility that polymorphism was responsible for the presence or absence of a particular SINE in the AFC family in each species, we performed experiments using PCR and additional individuals of three selected species, namely, *A. burtoni, E. cyanostictus,* and *P. microlepis,* with primers specific for amplification of SINEs at three selected loci (1265, 259, and 1220; fig. 3). The results demonstrated the absence of such polymorphism in each case examined. Thus, the loci investigated in the present study were considered fixed, with each allele being either with or without a SINE sequence.

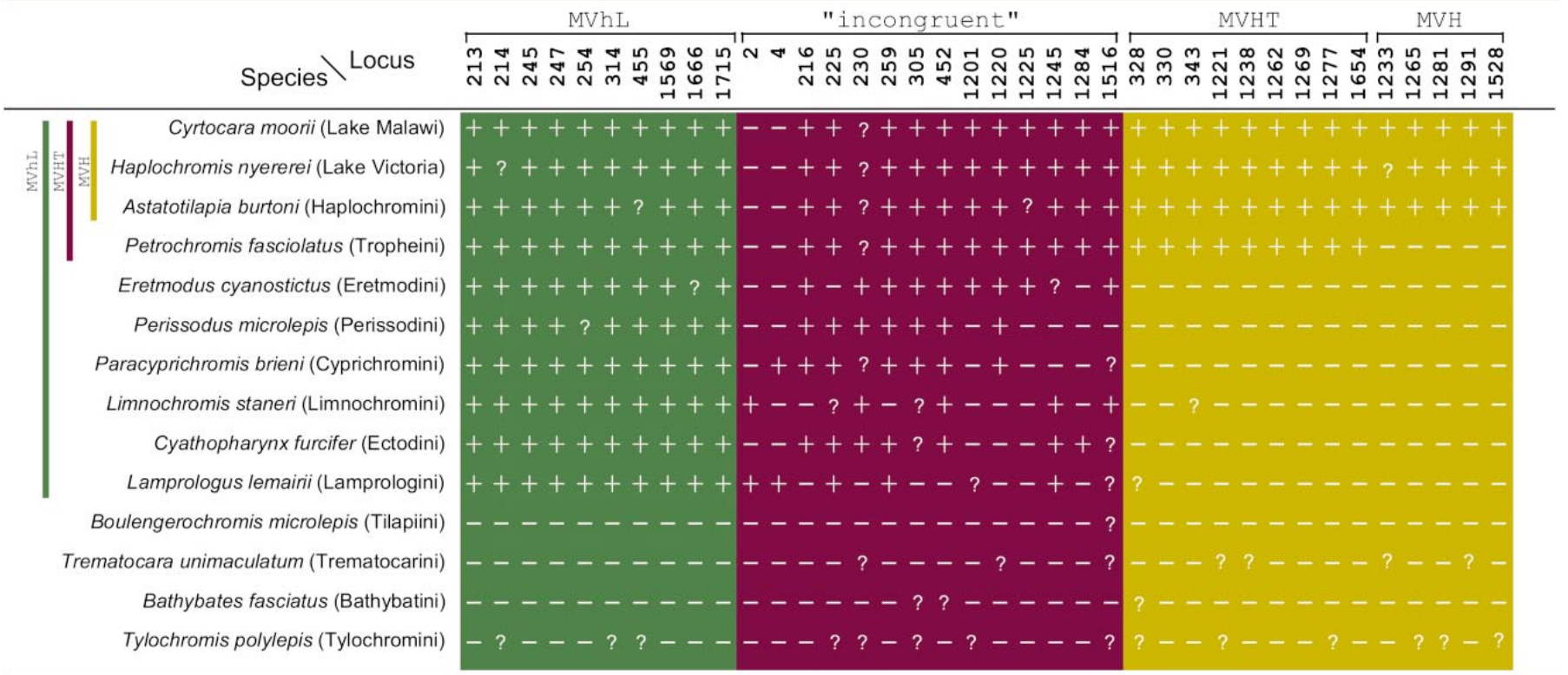
Discussion

In the present analysis, we investigated the patterns of SINE insertion in the AFC family at 38 independent loci. We were able to divide these patterns into two groups: a ''congruent'' group, and an ''incongruent'' group. The former group of patterns supported the existence of the three monophyletic groups, MVhL, MVHT, and MVH. The three clades were strongly supported by SINE insertions at 10, 9, and 5 independent loci, respectively. Furthermore, the three clades also seem to be in generally good agreement with the results of early studies of molecular phylogeny using various markers (Meyer et al. 1990; Meyer, Kocher, and Wilson 1991; Nishida 1991, 1997; Sturmbauer and Meyer 1993; Sturmbauer, Verheyen, and Meyer 1994; Kocher et al. 1995; Mayer, Tichy, and Klein 1998) even though representative species from each lake and tribe were not necessarily the same.

The phylogenetic tree shown in figure 2 might provide a basis for the evolution of phenotypes such as breeding behavior during the adaptive radiation of cichlids in Lake Tanganyika. The breeding behavior of cichlids in Lake Tanganyika is characterized by two primary types, mouthbrooding and substrate spawning (Fryer and Iles 1972; Barlow 1991). Mouthbrooders incubate their eggs in the buccal cavity, whereas substrate spawners take care of their eggs on a substratum. In Lake Tanganyika, *Tilapia, Boulengerochromis,* and all of the species in the Lamprologini tribe are substrate spawners, whereas all the other cichlids are mouthbrooders. The tree obtained in the present study clearly indicates that mouthbrooders and substrate spawners are polyphyletic (fig. 2). This result is consistent with the notion that cichlids have evolved breeding behavior multiple times in different lineages during their adaptive radiation (Barlow 1991; Sturmbauer and Meyer 1993).

Possible explanations for the incongruent patterns of insertion of SINE sequences are interspecific hybridization and/or the incomplete lineage sorting of ancestral polymorphism of SINEs. Both of these possibilities should remain in consideration because of the difficulty in distinguishing between them based on the present data. However, if we follow the observation that interspecific hybridization seems rare in the extant faunas of cichlids in East Africa (see discussions in Moran and Kornfield [1993, 1995] and Parker and Kornfield [1997]) and postulate that a similar situation also existed when the major lineages in Lake Tanganyika radiated, incomplete lineage sorting may be the more likely explanation. The MVhL clade was strongly supported, as shown in table 3. Therefore, the length of the internode basal to MVhL (from time X to time Y in fig. 4) appears to have been sufficient for some SINEs to become fixed in the population. Other SINEs that were amplified more recently within this internode (indicated by the yellow

Table 3
A Presence/Absence Matrix for an AFC SINE at Each Locus Analyzed



NOTE. +, Presence of the indicated AFC SINE; -, absence of the indicated AFC SINE; ?, "no data" due to failure of amplification of a product by PCR. Bars on the left side of the names of species indicate the three momophyletic groups that were unambiguously identified in this study. Loci were divided into four groups by reference to patterns of insertion of AFC SINEs.

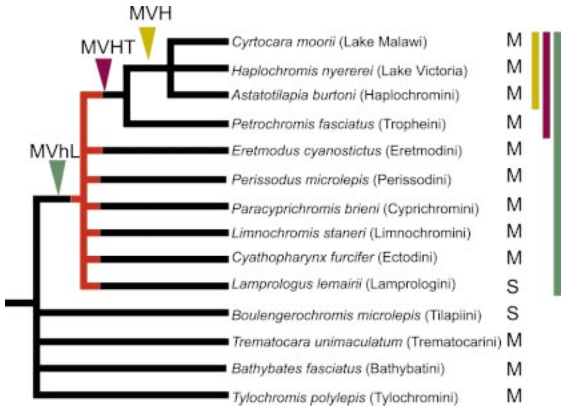


FIG. 2.—A phylogenetic tree for cichlid species in the 12 tribes in Lake Tanganyika, as well as representative species from Lakes Malawi and Victoria, based on the present analysis of the insertion of AFC short interspersed elements (SINEs). Names of the tribes and lakes are shown in parentheses after the names of the corresponding species. Arrowheads indicate internodes deduced from insertion of a SINE unit at each of 24 loci analyzed, and the names above the arrowheads are the designations of the identified clades. These clades were supported by the patterns of insertion of a SINE unit at loci 213, 214, 245, 247, 254, 314, 455, 1569, 1666, and 1715 (the MVhL clade); at loci 328, 330, 343, 1221, 1238, 1262, 1269, 1277, and 1654 (the MVHT clade); and at loci 1233, 1265, 1281, 1291, and 1528 (the MVH clade). Vertical bars on the right side of the tree also indicate the three clades. The red portion of the tree indicates the period during which putative incomplete lineage sorting of ancestral polymorphisms occurred extensively. M = mouthbrooder; S = substrate spawner.

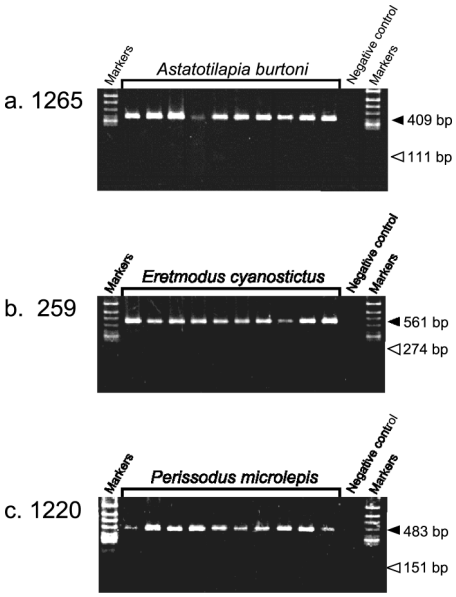


FIG. 3.—Confirmation of the fixation of particular AFC short interspersed elements (SINEs) by electrophoretic analysis of products of PCR. PCR was performed with DNA from each of 10 individual specimens of *Astatotilapia burtoni* (*a*), *Eretmodus cyanostictus* (*b*), and *Perissodus microlepis* (*c*) using primers for amplification of three selected loci, namely, 1265, 259 and 1220, respectively. Closed and open arrowheads indicate the expected mobilities of amplified fragments with and without insertion of an AFC SINE sequence.

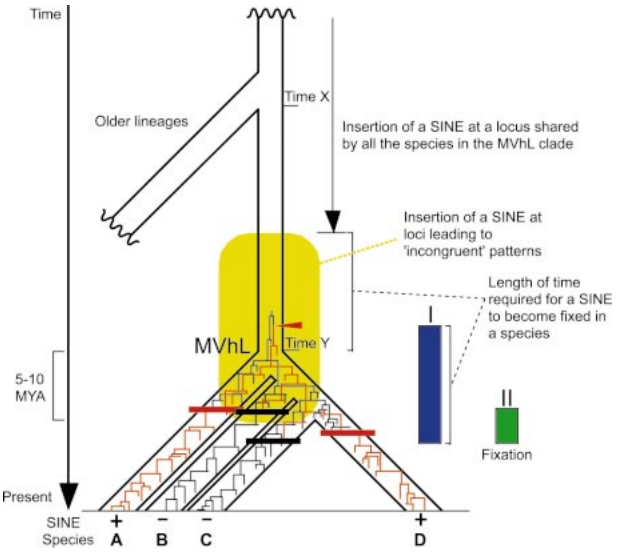


FIG. 4.—A scheme showing a hypothetical example of putative incomplete lineage sorting among lineages in the MVhL clade and the subsequent fixation of an allele with or without a short interspersed element (SINE) sequence. The large tree indicates the phylogenetic relationship among the four representative species (A–D) from different major lineages in the MVhL clade. Monophyly of the MVhL clade was itself clear, since lineage sorting of loci 213, 214, 245, 247, 254, 314, 455, 1569, 1666, and 1715 might have been completed before the divergence of lineages A–D. The region between time X and time Y represents the internode basal to the MVhL clade. Within the species tree, the gene tree of a certain orthologous locus is shown. A red arrowhead indicates insertion of the SINE sequence at this locus. Accordingly, at this locus, the SINE sequence is absent in the black portion of the gene tree and present in the red portion. Polymorphism with respect to the presence or absence of the SINE sequence existed during period I, and explosive speciation occurred 5–10 MYA (Nishida 1997). After speciation, the allele with the SINE sequence was fixed (+) in each of the lineages that led to species A and D, whereas the allele without this SINE sequence was fixed (−) in each of the lineages that led to species B and C (period II). The red and black thick horizontal bars denote fixation of alleles with and without the SINE sequence, respectively. Thus, the present pattern of SINE insertions at this locus in the four species is in conflict with the species phylogeny and also with the gene phylogeny of some other loci. The figure shows just four species and one locus for simplicity. We found that speciation of at least seven lineages was related to such incomplete lineage sorting, and 14 loci were identified that were incongruent with one another in terms of phylogeny in the present study.

rectangle) might, however, be polymorphic. When a SINE was polymorphic (i.e., presence/absence) at the beginning of the divergence of lineages of the MVhL clade (time Y), this SINE became a possible source of an ''incongruent locus.'' After the divergence of each lineage, an allele with or without a SINE unit at a particular locus might have become fixed stochastically (Pamilo and Nei 1988). This process might cause incongruence among the gene trees of the investigated loci if one locus became fixed for presence while another was fixed for absence. As a hypothetical example, let us consider a certain individual locus at which a SINE unit was inserted at the time indicated by a red arrowhead in figure 4. The SINE at this locus had been polymorphic in most lineages during period I (blue rectangle) but was fixed or lost independently in each lineage during period II (green rectangle). Then, the SINE inserted at this locus was shared by species A and D, while it

was lost in species B and C, even though species D is more distantly related to species A than to species B and C. Since the process of fixation and loss of a SINE at a locus is stochastic, patterns of presence and absence of SINEs can vary among different loci if multiple speciation events overlap in their time courses or occur during short time intervals. This hypothesis, which assumes the fixation of ancestrally polymorphic alleles, is consistent with the absence of a heterozygote of alleles with and without a sequence of the AFC family in the present matrix (table 3), which can be detected by coexistence of long and short PCR products from a single individual, and with our failure to detect polymorphisms in extant species (fig. 3).

An effective population size and intervals between speciations have been proposed to be critical factors in the differential lineage sorting of ancestral polymorphisms (Nei 1987; Pamilo and Nei 1988; Takahata 1989). When effective population sizes are large and constant (or expanding) and intervals between speciations are short, ancestral polymorphisms are likely to be retained in multiple successive lineages. Our present data suggest that the population of the ancestor of the MVhL lineage might have been relatively large and might have diverged rapidly into the different species that formed the ancestral lineages that led, in turn, to the present independent tribes.

Extensive incomplete lineage sorting has also been reported in cichlids in Lake Malawi. Moran and Kornfield (1993, 1995) investigated the restriction fragment length polymorphisms (RFLPs) of the mitochondrial DNA of Mbuna (rock-dwelling) species and suggested the existence of ancestral polymorphisms that had been retained by multiple species. Parker and Kornfield (1997) obtained a similar result when they investigated sequences of the control region of mitochondrial DNA in various species of Mbuna. Albertson et al. (1999) observed amplified fragment length polymorphisms (AFLPs) in several species of Mbuna in Lake Malawi, and their results supported the persistence of ancestral polymorphisms among extant populations. With regard to non-Mbuna species in Lake Malawi, Takahashi et al. (2001) reported a transspecies polymorphism of a SINE insertion at a specific locus and suggested that it might be due to incomplete lineage sorting and/or interspecific hybridization among these species. A similar phenomenon has also been reported in cichlids of Lake Victoria. An analysis of DNA sequence variation at four randomly selected loci in the nuclear genome revealed sharing of multiple alleles among nearly all 12 of the tested species, demonstrating that neutral polymorphisms have persisted beyond species boundaries (Nagl et al. 1998).

The importance of incomplete lineage sorting effects on species phylogeny inference has been suggested from both practical (Nagl et al. 1998; Streelman et al. 1998; Albertson et al. 1999) and theoretical (Nei 1987; Pamilo and Nei 1988; Takahata 1989; Wu 1991; Tachida and Iizuka 1993; Lyons-Weiler and Milinkovitch 1997; Avise 2000) standpoints. Our results indicated the existence of extensive incomplete lineage sorting at 14 loci among the lineages leading to the tribes Eretmodini, Perissodini, Cyprichromini, Limnochromini, Ectodini, and Lamprologini, as well as the MVHT clade. Therefore, we must be careful when attempting to infer phylogenetic relationships among these lineages, particularly when a single-locus marker, including mitochondrial DNA, is used. Let us consider, for example, the monophyly of the H-lineage (tribes in the MVhL clade excluding the Lamprologini). This monophyly was first proposed from the results of allozyme analysis (Nishida 1991, 1997) and was supported by studies of sequences of the control region and the gene for cytochrome *b* of mitochondrial DNA (Sturmbauer and Meyer 1993; Sturmbauer, Verheyen, and Meyer 1994) with bootstrap values of 43%–89%. The H-lineage was, however, suggested to be polyphyletic from the results of a study of mitochondrial DNA (the gene for subunit 2 of NADH dehydrogenase) by Kocher et al. (1995), who placed *Tanganicodus irsacae* (Eretmodini tribe) in the position of a sister group to the Lamprologini tribe. The present study suggests that monophyly of the H-lineage should be reexamined, since we found that the patterns obtained for five (loci 2, 4, 225, 259, and 1245) of the 14 incongruent loci failed to support monophyly.

Multilocus incongruence might be a general phenomenon that has accompanied rapid speciation during the adaptive radiation of various organisms. Further examples of such a phenomenon might be found among the Galapagos finches, Hawaiian Drosophila, and honey creepers (Mayr 1984), as well as among fish in certain other lakes (Schliewen, Tautz, and Pääbo 1994; Strecker et al. 1996), since these organisms are known to form ''species flocks,'' which are seemingly monophyletic groups of closely related species that coexist in the same areas, as do the cichlids in the African Great Lakes. Moreover, adaptive radiation is not necessarily restricted to organisms that form species flocks, and multilocus incongruence might potentially have been a much more general phenomenon than previously expected. Indeed, it has been proposed that discordance between gene phylogeny and species phylogeny (or between gene phylogenies based on studies of different loci) might be due to incomplete lineage sorting in a wide variety of animals and plants, such as species of mouse (Ohtsuka et al. 1996), char (Hamada et al. 1998*b*), rockfish (Alesandrini and Bernardi 1999), and *Brassica* (Tatout et al. 1999). Another possible example of incomplete lineage sorting has been reported in Felidae (Slattery, Murphy, and O'Brien 2000). In this report, two distantly related species in this group were suggested to exclusively share an insertion of a SINE at a specific locus. Although Slattery, Murphy, and O'Brien (2000) proposed that this result was due to parallel insertion of a SINE sequence at the same site in the genomes of these species, we will need many more data to reject the incomplete lineage sorting hypothesis.

In most studies, including the above examples, evidence for incomplete lineage sorting has come from observation of multiple alleles shared among recently radiated species (transspecies polymorphisms). However, this phenomenon may not be specific to groups that have recently radiated. When radiation of species was

ancient enough for the alleles responsible for such transspecies polymorphisms to become fixed, its evidence comes only from discordances among genealogies of different loci. The SINE method appears to be advantageous for detection of such an ancient phenomenon, since, unlike the ordinary methods using sequence data, this method can eliminate the possibility of parallelism, which often cannot be easily distinguished from incomplete lineage sorting as a cause of incongruence among estimated gene trees.

Acknowledgments

The authors thank the Lake Tanganyika Research Group of Kyoto University for the collection and identification of specimens of cichlid fish from Lake Tanganyika, Dr. T. Sato and Dr. M. Yuma for their help with specimens from Lake Malawi, and Dr. M. Nei and Dr. A. P. Rooney for their discussion on data and the manuscript. This work was supported by a Grant-in-Aid for Specially Promoted Research and for Overseas Scientific Surveys from the Ministry of Education, Culture, Sports and Science of Japan to N.O.

LITERATURE CITED

ALBERTSON, R. C., J. A. MARKERT, P. D. DANLEY, and T. D. KOCHER. 1999. Phylogeny of a rapidly evolving clade: the cichlid fishes of Lake Malawi, East Africa. Proc. Natl. Acad. Sci. USA **96**:5107–5110.

ALESANDRINI, S., and G. BERNARDI. 1999. Ancient species flocks and recent speciation events: what can rockfish teach us about cichlids (and vice versa)? J. Mol. Evol. **49**:814–818.

AVISE, J. C. 2000. Phylogeography: the history and formation of species. Harvard University Press, Cambridge, Mass.

BARLOW, G. W. 1991. Mating systems among cichlid fishes. Pp. 173–190 *in* M. H. A. KEENLEYSIDE, ed. Cichlid fishes: behaviour, ecology and evolution. Chapman and Hall, London.

COHEN, A. S., M. G. SOREGHAN, and C. A. SCHOLZ. 1993. Estimating the age of formation of lakes: an example from Lake Tanganyika, East-African Rift system. Geology **21**: 511–514.

COOK, J. M., and M. TRISTEM. 1997. 'SINEs of the times'—transposable elements as clade markers for their hosts. Trends Ecol. Evol. **12**:295–297.

COULTER, G. W. 1991. The benthic fish community. Pp. 151–199 *in* G. W. COULTER, ed. Lake Tanganyika and its life. Oxford University Press, London, Oxford, and New York.

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

GREENWOOD, P. H. 1984. African cichlids and evolutionary theories. Pp. 141–154 *in* A. A. ECHELLE and I. KORNFIELD, eds. Evolution of fish species flocks. University of Maine at Orono Press, Orono, Maine.

HAMADA, M., M. HIMBERG, R. A. BODALY, J. D. REIST, and N. OKADA. 1998*a*. Monophyletic origin of the genera Stenodus and Coregonus, as inferred from an analysis of the insertion of SINEs (short interspersed repetitive elements). Arch. Hydrobiol. Spec. Issues Adv. Limnol. **50**:383–389.

HAMADA, M., N. TAKASAKI, J. D. REIST, A. L. DECICCO, A. GOTO, and N. OKADA. 1998*b*. Detection of the ongoing sorting of ancestrally polymorphic SINEs toward fixation or loss in populations of two species of charr during speciation. Genetics **150**:301–311.

KOCHER, T. D., J. A. CONROY, K. R. MCKAYE, and J. R. STAUFFER. 1993. Similar morphologies of cichlid fish in Lakes Tanganyika and Malawi are due to convergence. Mol. Phylogenet. Evol. **2**:158–165.

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.

LYONS-WEILER, J., and M. C. MILINKOVITCH. 1997. A phylogenetic approach to the problem of differential lineage sorting. Mol. Biol. Evol. **14**:968–975.

MAYER, W. E., H. TICHY, and J. KLEIN. 1998. Phylogeny of African cichlid fishes as revealed by molecular markers. Heredity **80**:702–714.

MAYR, E. 1984. Evolution of fish species flocks: a commentary. Pp. 3–11 *in* A. A. ECHELLE and I. KORNFIELD, eds. Evolution of fish species flocks. University of Maine at Orono Press, Orono, Maine.

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.

MEYER, A., T. D. KOCHER, and A. C. WILSON. 1991. African fishes. Nature **350**:467–468.

MIYAMOTO, M. M. 1999. Perfect SINEs of evolutionary history? Curr. Biol. **9**:R816–R819.

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

———. 1995. Were population bottlenecks associated with the radiation of the Mbuna species flock (Teleostei: Cichlidae) of Lake Malawi? Mol. Biol. Evol. **12**:1085–1093.

MURATA, S., N. TAKASAKI, M. SAITOH, and N. OKADA. 1993. Determination of the phylogenetic relationships among Pacific salmonids by using short interspersed elements (SINEs) as temporal landmarks of evolution. Proc. Natl. Acad. Sci. USA **90**:6995–6999.

MURATA, S., N. TAKASAKI, M. SAITOH, H. TACHIDA, and N. OKADA. 1996. Details of retropositional genome dynamics that provide a rationale for a generic division: the distinct branching of all the Pacific salmon and trout (Oncorhynchus) from the Atlantic salmon and trout (Salmo). Genetics **142**:915–926.

NAGL, S., H. TICHY, W. E. MAYER, N. TAKAHATA, and J. KLEIN. 1998. Persistence of neutral polymorphisms in Lake Victoria cichlid fish. Proc. Natl. Acad. Sci. USA **95**:14238–14243.

NEI, M. 1987. Molecular evolutionary genetics. Columbia University Press, New York.

NIKAIDO, M., A. P. ROONEY, and N. OKADA. 1999. Phylogenetic relationships among cetartiodactyls based on insertions of short and long interspersed elements: hippopotamuses are the closest extant relatives of whales. Proc. Natl. Acad. Sci. USA **96**:10261–10266.

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

———. 1997. Phylogenetic relationships and evolution of Tanganyikan cichlids: a molecular perspective. Pp. 1–24 *in* H. KAWANABE, M. HORI, and M. NAGOSHI, eds. Fish communities in Lake Tanganyika. Kyoto University Press, Kyoto, Japan.

OHTSUKA, H., M. OYANAGI, Y. MAFUNE, N. MIYASHITA, T. SHIROISHI, K. MORIWAKI, R. KOMINAMI, and N. SAITO. 1996. The presence/absence polymorphism and evolution of

the p53 pseudogene in the genus Mus. Mol. Phylogenet. Evol. **5**:548–556.

OKADA, N. 1991. SINEs: short interspersed repeated elements of the eukaryotic genome. Trends Ecol. Evol. **6**:358–361.

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

PARKER, A., and I. KORNFIELD. 1997. Evolution of the mitochondrial DNA control region in the Mbuna (Cichlidae) species flock of Lake Malawi, East Africa. J. Mol. Evol. **45**:70–83.

POLL, M. 1986. Classification des Cichlidae du lac Tanganyika: tribus, genres et espèces. Mem. Acad. R. Belg. Cl. Sci. **45**: 5–163.

SCHLIEWEN, U. K., D. TAUTZ, and S. PÄÄBO. 1994. Sympatric speciation suggested by monophyly of crater lake cichlids. Nature **368**:629–632.

SHEDLOCK, A. M., M. C. MILINKOVITCH, and N. OKADA. 2000. SINE evolution, missing data, and the origin of whales. Syst. Biol. **49**:808–817.

SHEDLOCK, A. M., and N. OKADA. 2000. SINE insertions: powerful tools for molecular systematics. BioEssays **22**:148–160.

SHIMAMURA, M., H. YASUE, K. OHSHIMA, H. ABE, H. KATO, T. KISHIRO, M. GOTO, I. MUNECHIKA, and N. OKADA. 1997. Molecular evidence from retroposons that whales form a clade within even-toed ungulates. Nature **388**:666–670.

SLATTERY, J. P., W. J. MURPHY, and S. J. O'BRIEN. 2000. Patterns of diversity among SINE elements isolated from three Y-chromosome genes in carnivores. Mol. Biol. Evol. **17**: 825–829.

STRECKER, U., C. G. MEYER, C. STURMBAUER, and H. WILKENS. 1996. Genetic divergence and Speciation in an extremely young species flock in Mexico formed by the genus Cyprinodon (Cyprinodontidae, Teleostei). Mol. Phylogenet. Evol. **6**:143–149.

STREELMAN, J. T., R. ZARDOYA, A. MEYER, and S. A. KARL. 1998. Multilocus phylogeny of cichlid fishes (Pisces: Perciformes): evolutionary comparison of microsatellite and single-copy nuclear loci. Mol. Biol. Evol. **15**:798–808.

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.

SÜLTMANN, 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.

TACHIDA, H., and M. IIZUKA. 1993. A population genetic study of the evolution of SINEs. I. Polymorphism with regard to the presence or absence of an element. Genetics **133**:1023–1030.

TAKAHASHI, K., M. NISHIDA, M. YUMA, and N. OKADA. 2001. Retroposition of the AFC family of SINEs (short interspersed repetitive elements) before and during the adaptive radiation of cichlid fishes in Lake Malawi and related inferences about phylogeny. J. Mol. Evol. (in press).

TAKAHASHI, K., Y. TERAI, M. NISHIDA, and N. OKADA. 1998. A novel family of short interspersed repetitive elements (SINEs) from cichlids: the patterns of insertion of SINEs at orthologous loci support the proposed monophyly of four major groups of cichlid fishes in Lake Tanganyika. Mol. Biol. Evol. **15**:391–407.

TAKAHATA, N. 1989. Gene genealogy in three related populations: consistency probability between gene and population trees. Genetics **122**:957–966.

TATOUT, C., S. WARWICK, A. LENOIR, and J.-M. DERAGON. 1999. SINE insertions as clade markers for wild crucifer species. Mol. Biol. Evol. **16**:1614–1621.

TERAI, Y., K. TAKAHASHI, and N. OKADA. 1998. SINE cousins: the 3′-end tails of the two oldest and distantly related families of SINEs are descended from the 3′ ends of LINEs with the same genealogical origin. Mol. Biol. Evol. **15**: 1460–1471.

WEINER, A. M., P. L. DEININGER, and A. EFSTRATIADIS. 1986. Nonviral retroposons: genes, pseudogenes, and transposable elements generated by the reverse flow of genetic information. 1986. Annu. Rev. Biochem. **55**:631–661.

WU, C.-I. 1991. Inferences of species phylogeny in relation to segregation of ancient polymorphisms. Genetics **127**:429–435.

NARUYA SAITOU, reviewing editor

Accepted July 13, 2001