

Origin and Speciation of Haplochromine Fishes in East African Crater Lakes Investigated by the Analysis of Their mtDNA, *Mhc* Genes, and SINEs

Akie Sato, Naoko Takezaki, Herbert Tichy, Felipe Figueroa, Werner E. Mayer, and Jan Klein

Max-Planck-Institut für Biologie, Abteilung Immunogenetik, Tübingen, Germany

The Western Branch of the East African Great Rift Valley is pocketed with craters of extinct or dormant volcanoes. Many of the craters are filled with water, and the lakes are inhabited by fishes. The objective of the present study was to determine the amount and nature of genetic variation in haplochromine fishes inhabiting two of these crater lakes, Lake Lutoto and Lake Nshere, and to use this information to infer the origin and history of the two populations. To this end, sequences of mitochondrial (mt) DNA control region, exon 2 of major histocompatibility complex (*Mhc*) class II *B* genes, and short interspersed elements (SINEs) were analyzed. The results indicate that the Lake Nshere and Lake Lutoto fishes originated from different but related large founding populations derived from the Kazinga Channel, which connects Lake Edward and Lake George. Some of the genetic polymorphism that existed in the ancestral populations was lost in the populations of the two lakes. The polymorphism that has been retained has persisted for some 50,000 generations (years). During this time, new mutations arose and became fixed in each of the two populations in the mtDNA, giving rise to sets of diagnostic substitutions. Each population evolved in isolation after the colonization of the lakes less than 50,000 years ago. There appears to be no population structure within the crater lake fishes, and their present effective population sizes are in the order of 10^4 to 10^5 individuals. Comparisons with the endemic haplochromine species of Lake Victoria reveal interesting parallels, as well as differences, which may help to understand the nature of the speciation process.

Introduction

The formation of the East African Great Rift Valley, a complex system of faults and escarpments, was accompanied by extensive volcanic activity that left behind many craters along both the eastern and western branches of the Rift (Schlüter 1997). As the volcanoes became extinct or dormant, many of the craters filled with water, and some of the crater lakes came to be inhabited by fish. The circumstances under which the fish reached the crater lakes have been a subject of speculation for many years. Where the lakes established a temporary or permanent connection with nearby river systems, the origin seems obvious. For some of the crater lakes, however, past or present contacts with rivers could not be documented, and more fanciful explanations, such as transfer of fish fry or fertilized eggs by birds or water mammals, have been invoked.

Genetic characterization of a fish population can provide information about the probable size and composition of the founding population, the time of the founding event, and other related questions. This information can then be used to make inferences not only about the circumstances of the colonization but also about the process of speciation and species divergence. Here we use three genetic systems to probe the origin, population dynamics, and speciation of lacustrine fishes—mitochondrial (mt) DNA, genes of the major histocompatibility complex (*Mhc*), and short interspersed repetitive elements

(SINEs). The mtDNA control region with its high substitution rate and maternal transmission is the tool of choice in efforts to identify the most recent common ancestor (MRCA) of a population (Avice et al. 1987). Our survey of mtDNA control region sequences in lacustrine and fluviatile haplochromine cichlids of the Lake Victoria basin revealed the existence of seven major haplogroups, designated I through VII (Nagl et al. 2000). Each haplogroup is characterized by a set of diagnostic substitutions and by the clustering of its members in a single clade on phylogenetic trees. The haplogroups are distributed differentially over the basin, the most widespread being haplogroup V centered on Lake Victoria, the Lake Edward region (including Lake George and Lake Albert), and the rivers around these lakes. A number of additional haplogroups apparently exist among the fishes of Lake Tanganyika and Lake Malawi, which were involved in the survey only tangentially. Haplogroup V is subdivided further into subgroups VA through VD, which are distinguished by a more restrictive set of diagnostic substitutions. Each of the subgroups, however, may be subdivisible even further on typing additional specimens. In particular, subgroups VB and VC show a heterogeneity indicative of future splitting.

Mhc genes with their high polymorphism and long persistence of allelic lineages are well suited for exploring the population dynamics in the history of a species (Klein et al. 1998). Fish *Mhc* genes, like those of other jawed vertebrates, fall into two classes, I and II, and each class into two subclasses, *A* and *B*, encoding the α and β chains of the $\alpha\beta$ heterodimeric protein (Klein et al. 1997; Shand and Dixon 2001). Since most of the work on cichlid *Mhc* genes was done on the class II *B* genes (Klein et al. 1993; Ono et al. 1993a, 1993b; Sato et al. 1997; Málaga-Trillo et al. 1998; Figueroa et al. 2000), we focused on the most variable part of these genes, exon 2 encoding the $\beta 1$ domain of the class II β chain.

Fish SINEs are retroposons that are several hundred base pairs long and contain a region homologous to

Key words: cichlids, haplochromine fishes, *Mhc*, mtDNA control region, SINE, crater lakes.

E-mail: akie.sato@tuebingen.mpg.de.

The nucleotide sequences reported in this publication have been submitted to the GenBank database and assigned the following accession numbers: AY164681 to AY164696 (mitochondrial control region), AY211028 to AY211042 (*Mhc* class II *B*), AY211024 to AY211027 (SINEs).

Mol. Biol. Evol. 20(9):1448–1462. 2003

DOI: 10.1093/molbev/msg151

Molecular Biology and Evolution, Vol. 20, No. 9,

© Society for Molecular Biology and Evolution 2003; all rights reserved.

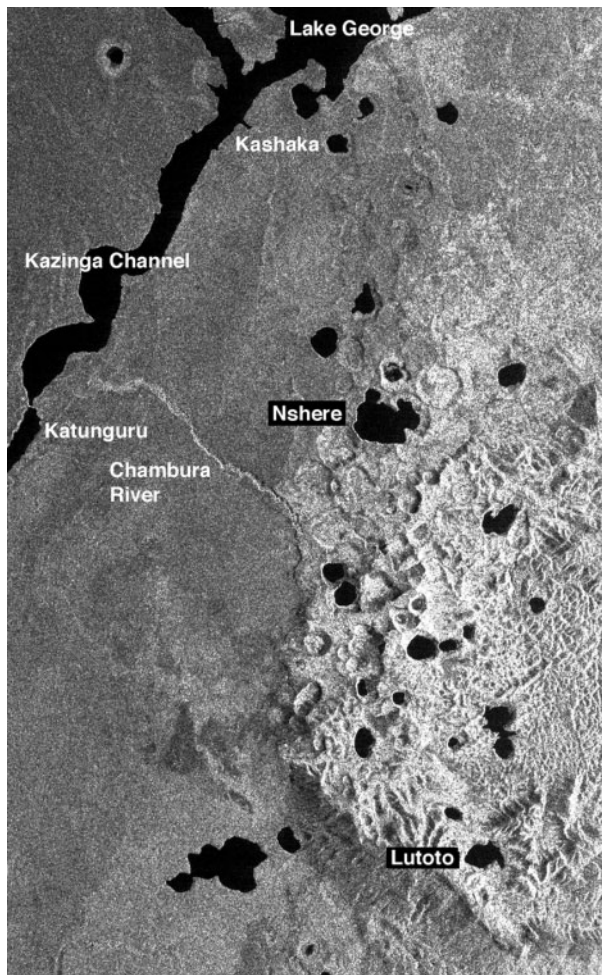


FIG. 1.—Satellite photograph of the Lake Edward region, East Africa, in which Lake Lutoto and Lake Nshere are located. The Kazinga Channel connects Lake Edward and Lake George. (Source: Section of the JPL / NASA SIR C 11 / X-SAR picture Nyiragongo, Zaire, DT058_61Segment7 taken on Nov. 18, 1998.)

a transfer RNA gene in their 5' part and frequently also contain a region corresponding to a segment of a long interspersed repetitive element (LINE) in their 3' part (Okada 1991; Schmid and Maraia 1992; Okada et al. 1997). In a recent study, Terai and colleagues (2003) have shown that haplochromines of the Lake Victoria basin contain two types of SINEs, fixed and polymorphic. The former are found in all individuals of a given species or species assemblage, whereas the latter are present in only some individuals of a population. As there is no indication for selection influencing the evolution of SINEs, these loci provide a nuclear counterpart to the neutrally evolving control region segment of the mtDNA on the one hand and the selection driven evolution of the *Mhc* genes on the other hand. Furthermore, the presumed random distribution of SINEs among the chromosomes enables sampling of different segments of the nuclear genome.

The two main aims of the present study were (1) to determine the origin, structure, and effective size of the haplochromine populations in Lake Lutoto and Lake Nshere and to ascertain the genetic relationship between

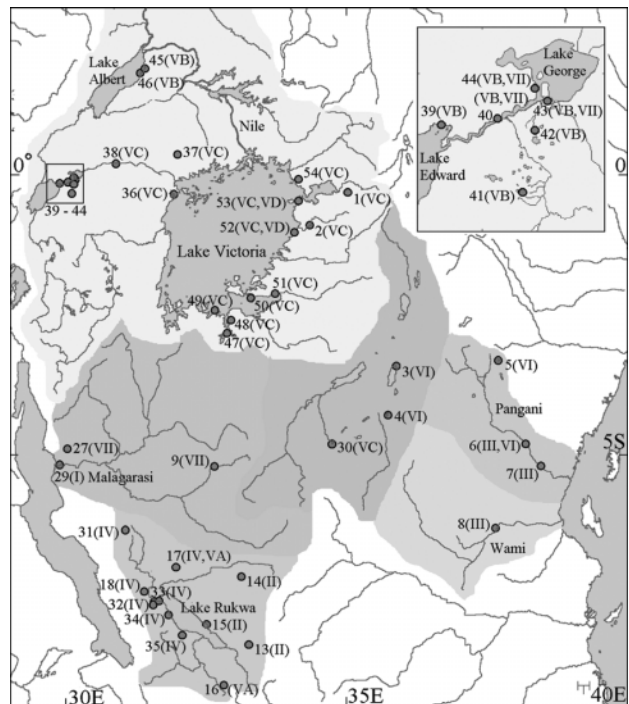


FIG. 2.—The localities in the Lake Victoria and the neighboring region at which samples of haplochromines were collected.

them and (2) to provide a basis for comparing the genetic variability existing in these two isolated populations with that present in the endemic Lake Victoria species and thus pave a way toward a better understanding of the speciation process in these fishes.

Materials and Methods

Localities

Lake Nshere and Lake Lutoto are located in the Bunyaruguru county in southwest Uganda. Judging from their location and shape, and from the geology of the region (Boven et al. 1998), both lakes originated by flooding of craters of extinct volcanoes. They are at a distance of approximately 20 km from each other (fig. 1). They are small (surface areas of $< 1 \text{ km}^2$) and are surrounded by even smaller lakes (fig. 1), the waters of at least some of which have a high salt concentration. Lake Nshere (or Chibwera) is located at an altitude of 970 m in the middle of a steppe-like flatland, the Chambura Game Reserve administered by the Queen Elizabeth National Parks. It is connected via a small river, which dries out periodically, with the Kazinga Channel and so with Lake George and Lake Edward (fig. 2). The distance from Lake Nshere to the Kazinga Channel is approximately 15 km. Lake Lutoto (or Rutoto, also known as Lake Nkugute) is located at an altitude of approximately 1,250 m in a mountainous area (the height of nearby Mt. Kasunju is 1,950 m [fig. 1]). It is connected by a small, rapidly flowing brook to Lake Mugogo at a distance of approximately 3.5 km, the difference in altitude between the two lakes being approximately 180 m. Lake Mugogo drains into the Chambura River system and so into the

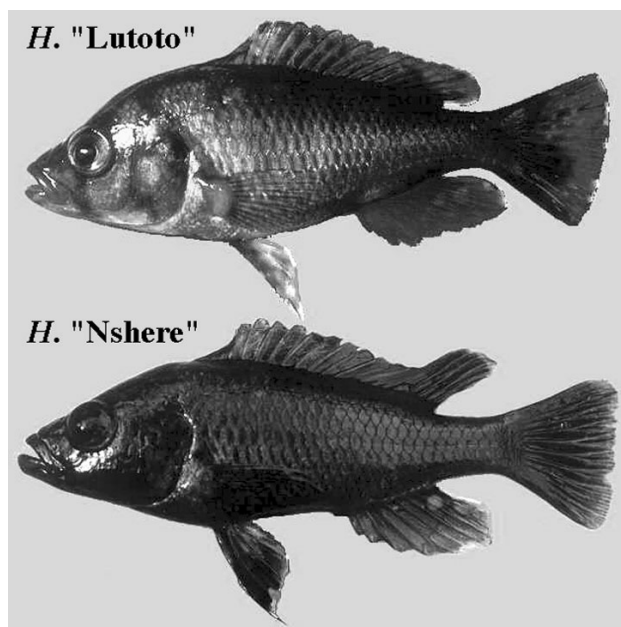


FIG. 3.—The phenotypes of *Haplochromis* “Lutoto” and *H.* “Nshere.”

Kazinga Channel at a distance of approximately 40 km. Although tilapias (*Oreochromis spp.*) were apparently introduced into Lake Lutoto some time ago, the studied species seem to be the only haplochromines in the lake.

Fishes

As far as we could determine, the species of haplochromines living in Lake Nshere and Lake Lutoto have not been described. For reasons that will become apparent later, we regard them as two different species, distinct from those inhabiting other East African lakes and rivers. We refer to them tentatively as *Haplochromis* “Nshere” and *Haplochromis* “Lutoto.” The adults of both species have an average standard body length of 55 mm and hence belong to the category of small haplochromines. Males of *H.* “Nshere” are dark black with prominent orange-yellow egg dumplings. The profile of their heads is straight. Females are a pale greenish color. *H.* “Lutoto” males have dark vertical bars on a dark greenish background. Their head profile forms a smooth concave line (fig. 3). The fishes were caught by angling, approximately 50 individuals in each lake, in less than 30 min using three fishing rods. Apparently, the fishes are abundant in both lakes.

Polymerase Chain Reaction (PCR), Cloning, and Sequencing

Amplifications were performed in the PTC-200 Programmable Thermal Controller (MJ Research, Biozym, Hessisch Oldendorf, Germany) or in the GeneAmp PCR System 9700 (Applied Biosystems, Weiterstadt, Germany). For the amplification of *Mhc* class II *B* exon 2, two Hot Start PCR systems were applied. In one system, 100 ng of genomic DNA were added to a reaction mixture of Mg²⁺-free 1 × PCR buffer, pH 8.5 (Invitrogen,

Table 1
PCR Primers Used

Locus	Orientation	Sequence 5'→3'
Mitochondrial DNA		
CR	F	TAAAATCCTTCCTACTGCTCA
CR	R	TCAAACAAAATATGAATAACAAAACA
<i>Mhc</i> class II <i>B</i>		
	F	TCTGAGTTATRGKTGGRTCGCTGTC
	R	GCAGTACSTCTCCTTCTAGCTYTC
SINEs		
1357	F	CAGCTGGATGAGTAGTGAARSTTCTG
1357	R	CTGTTTCCAACCTTGCACAGCATGG
1303	F	GAAGTGTATGGTCATGGTGAA
1303	R	TTTCTTTATGAAGTGCTGTGTTAGT
1304	F	TACCCTACTATCCAAATCATGTGA
1304	R	GCTCAGTACTTCTGTGACACTTTC
1327	F	AATTCCTCCACCTGTCATAGA
1327	R	CCTAGCTTCGATAGCCAAAGTAGT
1823	F	AAGAGATGTGTAGCAACCTAGCAA
1823	R	CAGATCCTTCACTGCTGTGTCT
1832	F	TGAACTGACAGCTAATTGCATACA
1832	R	CCCTCGCTCTTCCATATCAGA
1802	F	TGTAGTGTAGTGGTAAAGTCCCTAGTGT
1802	R	CAAAGACTGGAATTGGTGTACTGTA

NOTE.—CR indicates control region; F indicates forward; R indicates reverse.

Karlsruhe, Germany), 0.2 mM of each of the four deoxynucleoside triphosphates (Amersham Biosciences, Freiburg, Germany), 1 μM of each of the sense and antisense primers, 2.5 units of *Taq* DNA polymerase (Amersham Biosciences), and 0.4 units of *Pfu* DNA polymerase (Stratagene, Amsterdam, The Netherlands). The reaction mixture was overlaid with 2.5 mM Mg²⁺ HotWax Beads (Invitrogen) to initiate the reaction at high temperature. The amplification consisted of DNA denaturation for 1 min at 94°C, annealing for 15 sec at the appropriate annealing temperature, and 7 min extension at 72°C, followed by 34 cycles of 15 sec denaturation at 94°C, 15 sec annealing at the required temperature, and 2 min extension at 72°C. The reaction was completed by a final primer extension for 7 min at 72°C. Alternatively, 100 ng of genomic DNA were added to a reaction mixture of 1 × PCR buffer (Qiagen, Hilden, Germany), 2.5 mM MgCl₂, 0.2 mM of each of the four deoxynucleoside triphosphates (Amersham Biosciences), 1 μM of each of the sense and antisense primers, 2.5 units of HotStar *Taq* DNA polymerase (Qiagen), and 0.4 units of *Pfu* DNA polymerase (Stratagene). To activate the enzyme, the amplification was initiated by a DNA denaturation step of 15 min at 95°C. The PCR was carried out with HotStar *Taq* DNA polymerase under the conditions given above. PCR products were purified by agarose gel electrophoresis and extraction using the QIA EX II Gel Extraction Kit (Qiagen) and cloned into the pUC18 vector with the help of the Sure Clone Ligation Kit (Amersham Biosciences). The primers used in the various PCRs are listed in table 1. The DNA was sequenced with the help of the Dynamic Direct Cycle Sequencing Kit with 7-deaza dGTP (Amersham Biosciences). Sequencing reactions were processed by the LI-COR Long ReadIR 4200 DNA sequencer (MWG Biotech, Ebersberg, Germany).

Southern DNA Blotting and Hybridization

Five micrograms of genomic DNA were digested with *Eco* RI restriction endonuclease for 18 h under the conditions recommended by the supplier (Roche Diagnostics, Mannheim, Germany), and fragments were separated by agarose gel electrophoresis and blotted onto Hybond-N+ nylon filters (Amersham Biosciences). Pre-hybridization, hybridization, and probe labeling were carried out using the AlkPhos Direct Hybridization and Detection Kit (Amersham Biosciences). One hundred nanograms of DNA were used to label the probe. After the overnight hybridization, the filters were washed according to the AlkPhos Direct protocol. After the application of the chemiluminescent detection reagent CDP-Star of the kit, Hyperfilm ECL (Amersham Biosciences) was exposed to the blot for 6 h and developed.

Phylogenetic Analysis

The multiple alignments of sequence data were made by using ClustalW 1.82 (Thompson, Higgins, and Gibson 1994) and checked by eye. The phylogenetic trees were drawn by the neighbor-joining (NJ) method (Saitou and Nei 1987) and evaluated by 1,000 bootstrap replications.

In the analyses of mtDNA control region sequences, the sites 518 to 535 of the alignment in Nagl et al. (2000) were eliminated because they contain many single nucleotide repeats and the sequences are not reliable. Sites with indels were also deleted. NJ trees were drawn using *p*-distances. Unless the substitution rate varies extensively in different lineages, *p*-distance is better in obtaining a correct tree topology than distances based on a complex substitution model, even when the substitution pattern of the nucleotide changes follows a complex model, particularly for closely related sequences (Nei and Kumar 2000). The alignment of group V sequences (fig. 4 [only variable sites are shown]) includes sequences from Nagl et al. (2000). The NJ tree for the group V sequences (fig. 5) was drawn based on the alignment in figure 4 with 810 shared sites.

In the mtDNA analyses, estimations of the mutation rate and population size, as well as Tajima's test for neutrality, were based on an alignment containing sequences from the neighboring riverine and Lake Malawi fishes and a few Lake Tanganyika species as an outgroup with 804 shared sites (data not shown). The major branching patterns among the group V sequences of the NJ tree (not shown) based on the latter alignment were identical to those shown in figure 5. The GenBank accession numbers of the sequences that are not included in the group V NJ tree are as follows. Lake Tanganyika: *Petrochromis orthognathus* (U12549), *Lobochilotes labiatus* (U12550), *Tropheus duboisi* (AF400737); Lake Malawi: *Protomelas fenestratus* (AF213625), *Champschromis spilorhynchus* (U12553), *Mylochromis lateristriga* (AF213624), *Cyrtocara moorii* (U12554), *Lethrinops auritus* (U12551), *Stigmatochromis woodi* (AF213626), *Labeotropheus trewavasae* (AF213623), *Pseudotropheus sp. "msobo"* (AF213622), *Melanochromis melanopterus* (AF213619), *Melanochromis parallelus*

(AF213621), *Petrotilapia sp.* (U12547), *Metriaclima callainos* (AF213620), *Rhamphochromis sp.* (U12548); Haplochromines of the rivers and lakes in the Lake Victoria region: AF213559 to AF213609.

Since in the mtDNA control region the transition-transversion ratio is relatively high and the substitution rate varies extensively across sites (e.g., Tamura and Nei 1993), Kimura's two-parameter distance with gamma correction was used for time estimation. The gamma parameter ($\alpha = 0.12$) was estimated by the maximum-likelihood method using the tree topology of the NJ tree and the baseml in PAML3.12 program (Yang 2002). In the maximum-likelihood analysis, the transition-transversion rate ratio was estimated as 7.1. Observed base frequencies were 0.31, 0.31, 0.16, and 0.23 for A, T, G, and C, respectively.

The average distance (*h*) between the two descendant clusters of interior nodes of the mtDNA control region NJ tree was computed as

$$h = \sum_{i \in A, j \in B} d_{ij} / (2nm)$$

where d_{ij} is the distance between sequences *i* and *j*, A and B are the two descendant clusters of the node, and *n* and *m* are the numbers of the sequences in the clusters A and B, respectively.

The average distance between the mtDNA control region sequences of Lake Malawi and Lake Victoria fishes was 0.0445 ± 0.0073 . Assuming that the separation of these two groups occurred 1 to 2 MYA (see also Sturmbauer et al. 2001), the mutation rate for the mtDNA control region was estimated as $2.2 - 4.5 \times 10^{-8}$ per site per year. This rate is similar to the mutation rate (1.8×10^{-8}) estimated for other fish (snook [Donaldson and Wilson 1999]), but much lower than the rate for human mtDNA control region (0.75×10^{-7} [Tamura and Nei 1993]).

An NJ tree of the cichlid *Mhc* class II *B* exon 2 sequences was drawn using all the available cichlid and closely related sequences in the GenBank database. In the NJ tree shown in figure 5, only representative sequences were included because of space limitations, and medaka fish (*Oryzias latipes*) sequences were used as an outgroup. In the alignment, there were 160 shared nucleotide sites after sites with missing nucleotides or indels were excluded. Jukes-Cantor distances were used.

Detailed information regarding the fish samples and the sequence data obtained for mtDNA control region, *Mhc* class II *B* exon 2, and SINE 1357 are available as Supplementary Material online.

Computer Simulation of Coalescence Process

The coalescence theory (Hudson 1983; Tajima 1983) and the polymorphism found at the SINE loci were used to obtain the lower limit of the effective population size. According to the theory, allelic lineages that exist in a current population have a genealogy that can be traced back to an MRCA. For a neutral nuclear locus of a population with an effective size *N*, the time (t_j) during which exactly *j* allelic lineages exist in the population

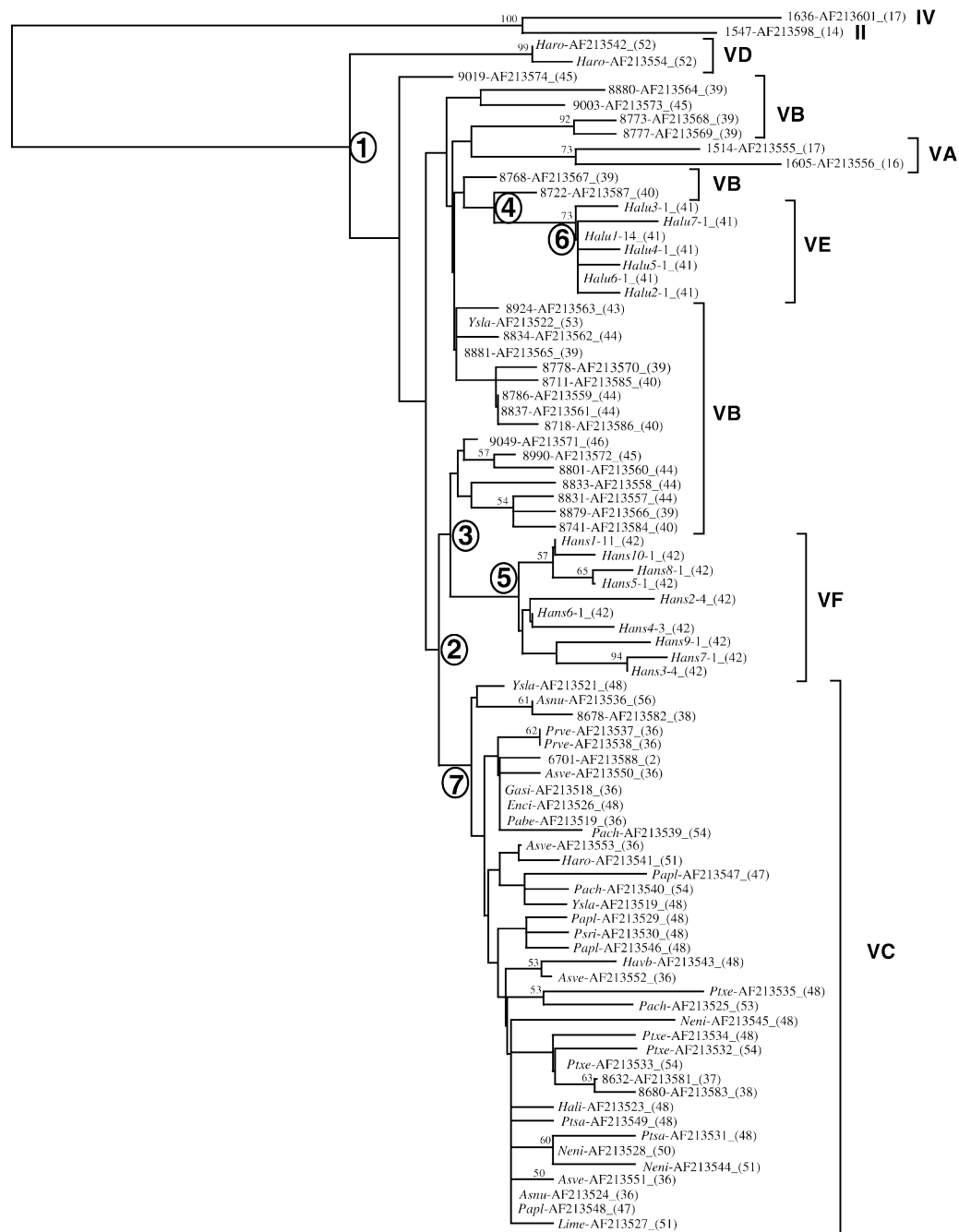


FIG. 5.—Neighbor-joining tree of mtDNA control region haplogroup V sequences. The tree is based on the alignment in figure 4. The sequence designations are as in figure 4. Subgroups of group V are enclosed in brackets. Encircled numbers indicate important nodes; other numbers at nodes are bootstrap values (only those >50% are shown).

Astatotilapia nubila; *Asve*, *A. velifer*; *Enci*, *Enterochromis cinctus*; *Gasi*, *Gaurochromis simpsoni*; *Hali*, *Haplochromis lividus*; *Halu*, *H. "Lutoto"*; *Hans*, *H. "Nshere"*; *Haro*, *H. rockkribensis*; *Havb*, *H. sp. "velvet black"*; *Lime*, *Lipochromis melanopterus*; *Neni*, *Neochromis nigricans*; *Pabe*, *Paralabidochromis beadlei*; *Pach*, *P. chilotes*; *Papl*, *P. plagiodon*; *Prve*, *Prognathochromis venator*; *Psri*, *Psammochromis riponianus*; *Ptse*, *Ptyochromis sauvagei*; *Ptse*, *P. xenognathus*; *Ysla*, *Yssichromis lapaogramma*. Substitutions diagnostic for the bracketed subgroups are highlighted. The accession numbers of *H. "Nshere"* (*Hans*) and *H. "Lutoto"* (*Halu*) sequences are AY164690/AY164695/AY164696 (*Hans*1), AF213578 (*Hans*2), AF213580 (*Hans*3), AF213579/AY164695 (*Hans*4), AY164687 (*Hans*5), AY164688 (*Hans*6), AY164689 (*Hans*7), AY164691 (*Hans*8), AY164692 (*Hans*9), AY164693 (*Hans*10), AF213577/AY164682/AY164683 (*Halu*1), AY164681 (*Halu*2), AF213575 (*Halu*3), AF213576 (*Halu*4), AY164684 (*Halu*5), AY164685 (*Halu*6), and AY164686 (*Halu*7). Some of the accession numbers are from Nagl et al. (2000). The samples sequenced in Nagl et al. (2000) were resequenced and confirmed in this study.

corresponds to a random number generated according to the probability density of the exponential distribution with a mean of $2/[j(j-1)]$ in units of $2N$ generations (Simonsen, Churchill, and Aquadro 1995). In the computer simulation starting with n lineages, the coalescence process was repeated until $n_a - 1$ lineages remained in the population, where n is the number of genes sampled in either the *H. "Nshere"* or *H. "Lutoto"* populations, and n_a is the number of alleles shared between *H. "Nshere"* or *H. "Lutoto"* and other lake populations. The time

$$t_a = \sum_{j=n}^{n_a} t_j$$

required to reach this stage was recorded, the expectation of t_a being

$$E(t_a) = 2 \left(\frac{1}{n_a} - \frac{1}{n} \right)$$

in units of $2N$. By applying various values of N , we searched for the case in which the probability that $t_a > 50,000$ years (the geological estimate of the age of Lakes Nshere and Lutoto being $< 50,000$ years) was smaller than 0.05. For this value of N , the hypothesis that n_a allelic lineages persisted for 50,000 years could be rejected at the 5% significance level. This value of N represents the lower limit for the effective population size of a crater lake population.

Computer Simulation for Founding Population Size and the Number of Alleles Retained

To examine the relationship between the minimum number of *Mhc* alleles retained in the population, the founding population size, and the population growth rate after the founding event, another computer simulation was conducted. The simulation was a slight modification of that used by Vincek et al. (1997) who assumed overdominant selection for *Mhc* loci with the selection coefficient $s = 0.01$. The alleles for the founding population of N_b individuals were drawn at random from the parental population. After the founding event, population growth started immediately with rate r and continued until the number of individuals in the population reached 10,000. Then, the number of alleles retained in the population was recorded. This process was repeated 1,000 times for each combination of the values of N_b and r . The ranges of the N_b and r values varied from 5 to 500 and from 0.01 to 0.5, respectively. The allele frequencies in the ancestral population were assumed to be those reported for the *HLA-DRB1* locus in Caucasians (Imanishi, Wakisaka, and Gojobori 1991). In a preliminary study, we also assumed that the frequencies of alleles in the ancestral population are equal. The results for these two situations were essentially the same.

Results

We divide the description of the results into two parts. In the first part we characterize the mtDNA control region, SINE, and *Mhc* class II *B* loci of *H. "Nshere"* and *H.*

"Lutoto." In the second part we use this information to make inferences regarding these two populations. We begin with mtDNA.

mtDNA Control Region

Control regions of mtDNA were sequenced from 28 fish caught in Lake Nshere and 20 specimens from Lake Lutoto. Ten distinct *H. "Nshere"* and seven *H. "Lutoto"* haplotypes were found (fig. 4) among the 48 sequences obtained. The *H. "Nshere"* sequences showed more variability (one to seven differences) than those of *H. "Lutoto"* (one to three differences). Phylogenetic analysis revealed all 48 sequences to belong to haplogroup V characterized by the presence of T, C, C, A, and A at sites 27, 87, 96, 348, and 825, respectively (see Nagl et al. 2000 and fig. 4). Each of the two sets of sequences forms a separate cluster (a clade on a phylogenetic tree [fig. 5]) and each is identified by distinct diagnostic substitutions. All Lake Nshere sequences have A and T at sites 107 and 830, respectively, whereas all Lake Lutoto sequences share T, T, and G at sites 100, 347, and 830, respectively (fig. 4). We denote the control region subgroup of the VB haplogroup present in *H. "Lutoto"* as VE and that of *H. "Nshere"* as VF. The *H. "Lutoto"* sequences appear to be most closely related to a particular subset of VB sequences with which they share an A at site 635 (fig. 4). Similarly, the *H. "Nshere"* sequences are most closely related to another subset of VB sequences that have a G at site 635. Some of the VF sequences also share an A at site 167 with some members of this VB subset, but this nucleotide is also present in some members of the VC subgroup (figs. 4 and 5). These observations lead us to three conclusions. First, the haplochromine fishes inhabiting Lake Lutoto and Lake Nshere originated in an area in which the haplogroup VB is common. Second, the fishes of these two lakes originated from two distinct but related populations, probably occupying different parts of the area of wide VB-haplogroup distribution (see *Discussion*). Third, the fish populations of the two lakes have been isolated from other populations long enough to attain fixation of diagnostic substitutions. A network presentation of mtDNA haplotypes of the VB group and the crater lake fishes is available as Supplementary Material online.

Using the mutation rate of $2.2 - 4.5 \times 10^{-8}$ per site per year (see *Material and Methods*), we estimate the age of the nodes that separate the crater lake fishes from fishes of other localities (nodes 3 and 4 in fig. 5) as 0.18 ± 0.06 to 0.09 ± 0.03 MYA for *H. "Nshere"* and 0.12 ± 0.05 to 0.06 ± 0.03 MYA for *H. "Lutoto"* (table 2). These estimates give the upper limit for the time when the isolation of the crater lake fish occurred. The common ancestral nodes of the Lake Nshere and the Lake Lutoto fishes (nodes 5 and 6 in fig. 5) are estimated to date to 0.13 ± 0.04 to 0.06 ± 0.02 and 0.080 ± 0.04 to 0.04 ± 0.02 MYA, respectively. In contrast, the common ancestral node (node 7 in fig. 5) of the Lake Victoria fishes is dated to 0.23 ± 0.08 to 0.12 ± 0.04 MYA (table 2). These estimates suggest that the isolation of Lake Victoria fishes from other fishes occurred somewhat earlier than that of the crater lake fishes.

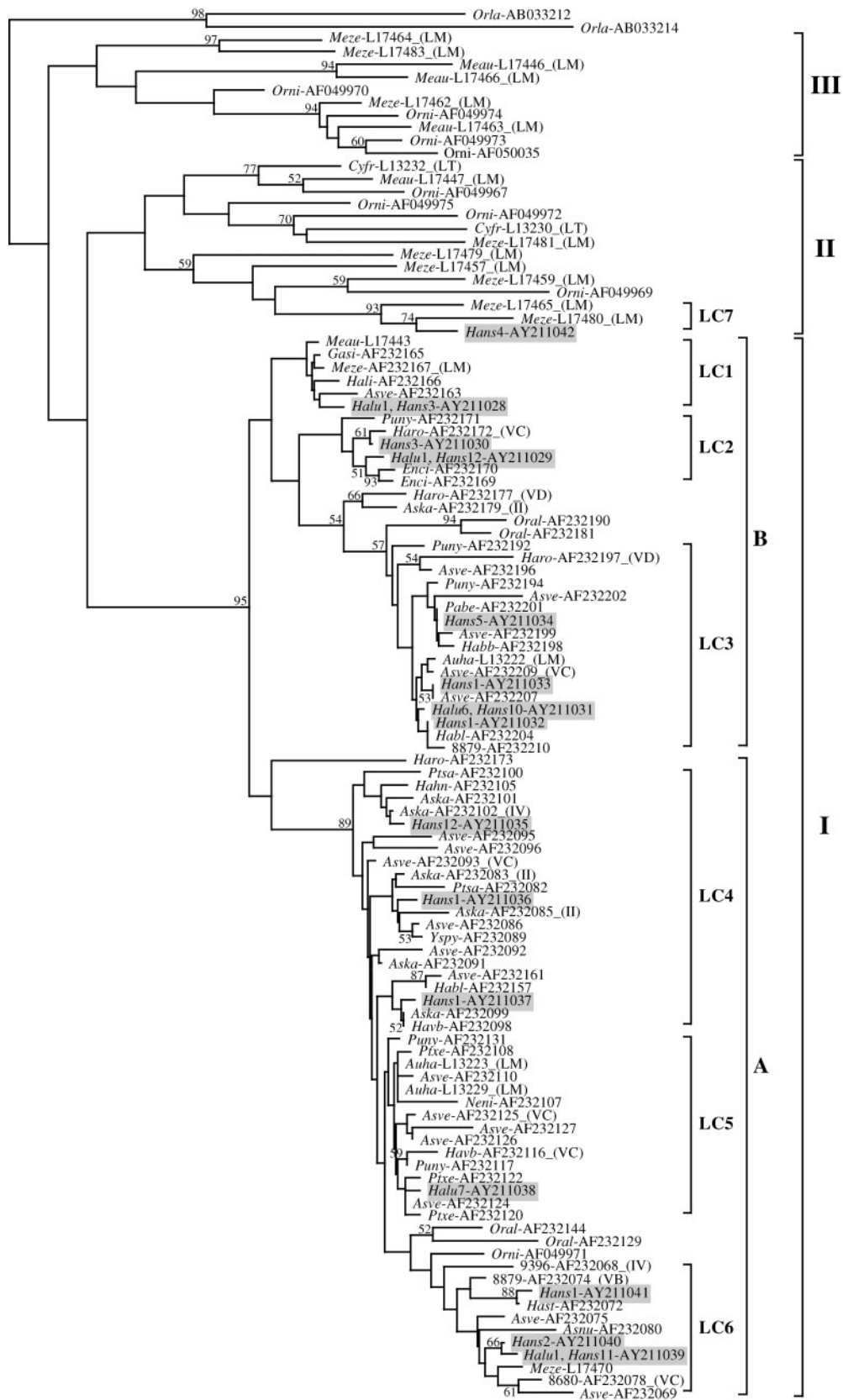


Fig. 7.—Neighbor-joining tree of *Mhc* class II *B* exon 2 sequences of East African cichlid fishes. *H.* “Nshere” and *H.* “Lutoto” sequences produced in the present study are identified as *Hans* or *Halu*, respectively, followed by the number of the individuals and the GenBank accession number, and are shaded. Other sequences are from Figueroa et al. (2000), Málaga-Trillo et al. (1998), and Ono et al. (1993*a*, 1993*b*). I, II, and III indicate lineages, A and B indicate sublineages, and LC1 through LC7 indicate different loci. Individual sequences are identified by species

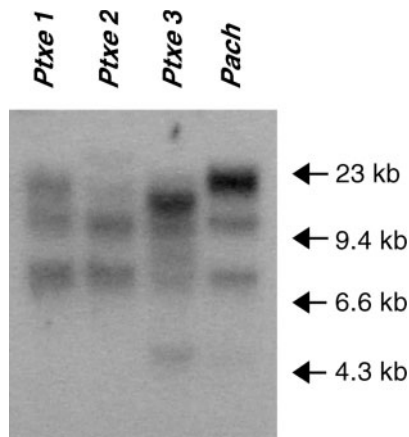


FIG. 8.—Southern blot analysis of genomic DNA isolated from three *Ptyochromis xenognathus* (*Ptxe*) and *Parabidochromis chilotes* (*Pach*) fish. All four samples were digested with *Eco* RI and the blots hybridized with an *Mhc* class II *B* exon 3a probe (Figueroa et al. 1995) (the 5' part of exon 3, which is interrupted by an intron in these fishes). Marker sizes in kilobases (kb) are indicated.

the separation of sequences into these clades is influenced by the choice of primers used for PCR amplification is unclear at present.

One unique sequence of *H.* “Nshere” is included in lineage II; all other sequences of fishes from the two crater lakes are included in lineage I and are approximately equally distributed between the sublineages IA and IB (fig. 7). Thus, a minimum of three class II *B* loci must exist in fishes from crater lakes (IA, IB, and II). Additional loci are defined by distinct clades within lineage I. Altogether, seven possible class II *B* loci are defined in the haplochromines of the two crater lakes (fig. 7). The predicted number of loci is also supported by the results of the Southern blot analysis (fig. 8).

Distribution of SINEs

Terai and his coworkers (2003) identified a set of SINEs that have been inserted into the genomes of endemic Lake Victoria and Lake Edward region haplochromines after the divergence of the Lake Victoria from the Lake Malawi flock and persist in most of the species and populations as presence or absence polymorphisms. To test for the presence or absence of these SINEs in the individual *H.* “Nshere” and *H.* “Lutoto” specimens, we used primers specific for the flanking regions of SINEs 1350, 1840, 1424, 1801, 1807del, 1909, and 1919 in PCR. Of the seven SINEs tested, two were found to be polymorphic in *H.* “Nshere” (1918 and 1801), whereas only one (1918) was polymorphic in *H.* “Lutoto”; all other

SINEs were fixed in the two populations (table 3 [see also Terai et al. 2003]). Comparison of observed and expected frequencies of the genotypes at the polymorphic SINEs revealed both populations to be in Hardy-Weinberg equilibrium (table 3). This result suggests that neither of the two populations is structured.

One of the fixed SINEs (1357) described by Terai et al. (2003) was shown by these authors to display sequence polymorphism, revealing the existence of at least 14 alleles in the haplochromines of the Lake Victoria basin. To obtain additional information about the *H.* “Nshere” and *H.* “Lutoto” populations, we sequenced the 1357 locus of individual fishes from these populations (fig. 9). The sequences obtained in the present study were shorter than those from Terai et al. (2003) in that they lacked 36 bp of the 5' region and 38 bp of the 3' region. For this reason, alleles *D1* and *D3*, which differ at site 31 in the missing region, could not be distinguished in the sequences obtained in this study. Of the 14 alleles, at least four (*D1* or *D3*, *D4*, *D5*, and *E1*) were found in *H.* “Nshere,” but only one (*D1* or *D3*) was detected in *H.* “Lutoto.” No new SINE 1357 alleles were found in either of the two populations.

Present Population Sizes of *H.* “Nshere” and *H.* “Lutoto”

Information about the effective population sizes of *H.* “Nshere” and *H.* “Lutoto” is provided by the mtDNA control region and SINE data. In the case of the mtDNA control region sequences, we computed the nucleotide diversity (average number of nucleotide differences per site, π [see Nei 1987]) as being 3.41×10^{-3} and 7.46×10^{-4} for the *H.* “Nshere” and *H.* “Lutoto” populations, respectively. In a mutation-drift equilibrium, $E(\pi) = 2N_f\mu$, where N_f is the effective population size of females and μ is the mutation rate per generation. By taking $\mu = 2.2 \times 10^{-8}$ to 4.5×10^{-8} per site per year (see *Materials and Methods*) and assuming that 1 generation = 1 year in cichlid fishes, we obtain $N_f = 3.4 \times 10^4$ to 7.7×10^4 for *H.* “Nshere” and $N_f = 0.8 \times 10^4$ to 1.7×10^4 for *H.* “Lutoto.”

The Lower Limit for the Effective Population Size of Crater Lake Fishes

Since polymorphism at some of the SINE loci is shared by the crater lake fishes and fishes of the neighboring region, we used the number of shared alleles and the coalescence theory to obtain the lower limit of the effective population size for *H.* “Nshere” and *H.* “Lutoto.” As described in *Materials and Methods*, we computed the probability of n_a allelic lineages persisting for 50,000 years by the coalescence algorithm (Hudson

← abbreviations, GenBank accession codes (see Ono et al. 1993a, 1993b; Málaga-Trillo et al. 1998; Figueroa et al. 2000), and locality (LM: Lake Malawi, LT: Lake Tanganyika), or mtDNA haplotype. Species abbreviations are as follows: *Aska*, *Astatotilapia katavi*; *Asnu*, *A. nubila*; *Asve*, *A. velifer*; *Auha*, *Aulonocara hansbaenschii*; *Cyfr*, *Cyphotilapia frontosa*; *Enci*, *Enterochromis cinctus*; *Gasi*, *Gaurochromis simpsoni*; *Habb*, *Haplochromis* sp. “big blue”; *Habl*, *H.* sp. “black”; *Hahn*, *H.* sp. “high neck”; *Hali*, *H. lividus*; *Haro*, *H. rockkribensis*; *Hast*, *H.* sp. “stone”; *Havb*, *H.* sp. “velvet black”; *Meau*, *Melanochromis auratus*; *Meze*, *Metriaclima zebra*; *Neni*, *Neochromis nigricans*; *Oral*, *Oreochromis alcalicus*; *Orla*, *Oryzias latipes*; *Orni*, *Oreochromis niloticus*; *Pabe*, *Paralabidochromis beadlei*; *Ptsa*, *Ptyochromis sauvagei*; *Ptxe*, *P. xenognathus*; *Puny*, *Pundamilia nyererei*; *Yspy*, *Yssichromis pyrrhocephalus*. Numbers at nodes are bootstrap values (only those >50% are shown).

Table 3
Observed and Expected Genotype Frequencies (%) at SINE Loci in Crater Lake Cichlids

Locus	<i>H.</i> “Nshere” (<i>n</i> = 28)					<i>H.</i> “Lutoto” (<i>n</i> = 20)						
		+/+	+/-	-/-	na	χ^2		+/+	+/-	-/-	na	χ^2
1918	obs	27	54	19	2	0.18	obs	63	26	11	1	1.41
	exp	29	50	21			exp	58	36	6		
1801	obs	22	59	19	1	0.94	obs	100	0	0	3	—
	exp	27	50	23			exp	100	0	0		
1424	obs	100	0	0	1	—	obs	100	0	0	0	—
	exp	100	0	0			exp	100	0	0		
1919	obs	0	0	100	1	—	obs	0	0	100	0	—
	exp	0	0	100			exp	0	0	100		
1350	obs	100	0	0	0	—	obs	100	0	0	1	—
	exp	100	0	0			exp	100	0	0		
1840	obs	100	0	0	0	—	obs	100	0	0	4	—
	exp	100	0	0			exp	100	0	0		
1807		+/+	+/+d	+d/+d				+/+	+/+d	+d/+d		
	obs	100	0	0	2	—	obs	100	0	0	4	—
	exp	100	0	0			exp	100	0	0		

NOTE.—*n* is the number of individuals tested; na is the number of individuals for which data were not obtained; obs indicates observed genotype frequencies; exp is the expected genotype frequencies; +/+, +/-, and -/- indicate homozygote for the presence of SINE insertion, heterozygote, and homozygote for the absence of SINE insertion, respectively; +d indicates SINE insertion variant with 216-bp deletion. None of the χ^2 values indicated significant deviation from Hardy-Weinberg equilibrium.

1983; Tajima 1983) for various population sizes (*N*) and searched for the value of *N* for which *n_a* lineages persist for more than 50,000 years in 5% of the iterations of the computer simulation.

We can assume *n_a* = 2 in the case of the presence/absence polymorphism of SINEs 1918 and 1801 in *H.* “Nshere” and 1918 in *H.* “Lutoto”; and *n_a* = 4 for the alleles determined by sequencing SINE 1357 and found to be shared between *H.* “Nshere” and other lake populations. Using the *n_a* = 2 value, we can reject the hypothesis that *N* was smaller than 6,000 individuals at the 5% significance level for the past 50,000 years of existence of the *H.* “Nshere” and *H.* “Lutoto” populations. The 6,000

individuals thus represent the lower limit for the effective sizes of these two populations. Using the *n_a* = 4 value, the results of the computer simulation indicate that *N* = 24,000 is the lower limit of the effective *H.* “Nshere” population size that assures the persistence of at least four allelic lineages for the last 50,000 years.

Population History

Tajima (1989) designed a test that examines

$$d = k - S / \left(\sum_i^{n-1} 1/i \right) = 0$$

by using the statistic

$$D = d \sqrt{\hat{V}(d)}$$

where *S* is the number of segregating sites in nucleotide sequences, *n* is the number of sequences obtained, and *k* is the average number of nucleotide differences per sequence. This test is based on relationships existing under the mutation-drift equilibrium

$$E(S) = \left(\sum_i^{n-1} 1/i \right) M$$

and *E(k)* = *M*, where *M* = 2*N_f**v* for the mtDNA sequences and *v* is the mutation rate per sequence in this case. It was developed for testing the neutrality of a locus under the assumption that a population is in the mutation-drift equilibrium. However, if we assume that neutrality holds at the locus, the test can also be used for examining a population size change in the past. A negative value of *D* indicates an excess of *S* relative to *k*, which suggests a recent rapid accumulation of mutations by population expansion. The results of the Tajima test applied to the *H.* “Nshere” and *H.* “Lutoto” populations are shown in table 4. Tajima’s *D* statistic is negative for both *H.*

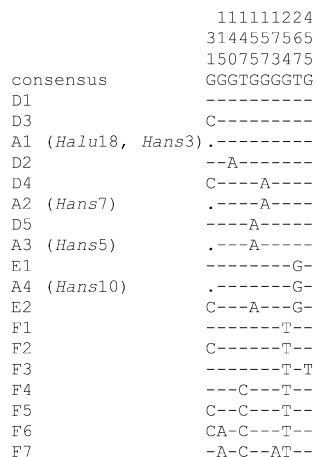


FIG. 9.—Nucleotide sequences of SINE 1357 alleles identified in Lake Victoria basin haplochromines (only variable sites are shown). Alleles found in *H.* “Nshere” and *H.* “Lutoto” are identified in parentheses by the respective numbers of the individuals from which the allele sequences were obtained. Sequences A1 through A4 are from the present study and start at site 37 and end at site 455; all other sequences are from Terai et al. (2003) and begin at site 31. Other symbols are as in figure 4. Dots (.) indicate that no data are available. The GenBank accession numbers for A1 to A4 are AY211024 to AY211027.

Table 4
The Nucleotide Differences in mtDNA Control Region and Tajima's Test of Neutrality

	<i>H.</i> "Nshere"	<i>H.</i> "Lutoto"
<i>n</i>	28	20
π	3.41×10^{-3}	7.46×10^{-4}
<i>k</i>	3.12	0.60
<i>S</i>	14.00	6.00
<i>D</i>	-0.45	-2.06 ^a

NOTE.—*n* is the number of samples from which sequences are obtained; π is the average nucleotide difference per site; *k* is the average nucleotide difference per sequence; *S* is the number of segregating sites; *D* is Tajima's *D* statistic.

^a *D* is significant at 5% level.

"Nshere" and *H.* "Lutoto," but the result is significant only for the latter, suggesting that the population of *H.* "Lutoto" expanded recently. The population size estimates shown in the earlier section were obtained by π , which is equivalent to *k*, because $\pi = k/m$ where *m* is the number of nucleotide sites examined. If we use the relationship

$$E(S) = \left(\sum_{i=1}^{n-1} 1/i \right) (2N_f v)$$

N_f is estimated as 2.4×10^4 to 4.7×10^4 for *H.* "Lutoto." Although a large error is associated with the estimate using *S*, it is likely that the female population size of *H.* "Lutoto" is currently larger than 0.8×10^4 to 1.7×10^4 .

Founding Population Size

Among the exon 2 *Mhc* class II *B* gene sequences we obtained, we found 14 different ones in *H.* "Nshere" and five different ones in *H.* "Lutoto." All these sequences, with one exception, belong to lineage I; the exceptional sequence found in *H.* "Nshere" belongs to the lineage II sequences of *Metriaclima zebra* (fig. 5). The lineage I sequences can be divided into six loci (see fig. 5), or they can be treated as one locus. By considering these two extreme possibilities, the numbers of alleles that were passed onto the founding populations of the crater lakes from the ancestral population are one to four for *H.* "Nshere" and one for *H.* "Lutoto" in the former case and 12 for *H.* "Nshere" and five for *H.* "Lutoto" in the latter case.

In a computer simulation similar to that described by Vincek et al. (1997), we examined the number of alleles retained for various founding population sizes ($N_b = 5$ to 500) and growth rates ($r = 0.01$ to 0.5) (see *Materials and Methods*). Only representative values ($r = 0.01, 0.05,$ and 0.5) are shown in figure 10. In figure 10A, the average numbers of alleles (n_{ave}) retained when the number of individuals reached 10,000 in 1,000 replications of computer simulation are shown in relation to the founding population size (N_b). In figure 10B, instead of the average number of alleles (n_{ave}), the 95% upper limit of the number of alleles (n_m) retained in the population is shown. That is, in at least 95% of the replications, the number of alleles retained in the population was smaller than n_m . For each N_b value, we can reject that the number of alleles retained

in the population is smaller than n_m at the 5% significance level. Thus, N_b values in figure 10B can give a lower limit of the founding population size for the observed number of alleles n_m . The result of the simulation suggests that four or five alleles can be retained even in the case of a small founding population (5 to 10 individuals) if the growth rate of the population after the founding event is not small (< 0.1 to 0.2). The founding population must contain at least 20 breeding individuals to retain 12 alleles, even for a growth rate as high as 0.5. For a relatively low growth rate (0.01), the founding population must be at least 50 to 70 to retain four to five alleles and at least 300 individuals to retain 12 alleles.

Discussion

Mitochondrial DNA haplotyping assigns the *H.* "Nshere" and *H.* "Lutoto" populations to the VB group, which is widely distributed throughout the Lake Edward region. On the phylogenetic tree (fig. 5), the set of VB sequences closest to those of *H.* "Nshere" stemmed from localities 39 (Kazinga Channel), 40 (Lake Edward), 44 (Lake George), and 45 and 46 (Lake Albert [see fig. 2]). Similarly, the set of VB sequences closest to those of *H.* "Lutoto" is derived from the same area (localities 39 and 40), although here related sequences were also found in Lake Albert. We suggest, therefore, that the founders of the two crater lake populations originated in the Kazinga Channel but that haplotypes closely related to those they carried were distributed across much of the entire region. Since the nearest neighbors of *H.* "Nshere" and *H.* "Lutoto" on the tree are different subsets of the VB subgroup, we suggest further that the founders came from different parts of the Kazinga Channel or from different populations and that the colonizations of the two lakes occurred independently. There is certainly no evidence that *H.* "Nshere" originated from *H.* "Lutoto," or vice versa.

Geologists date the earliest volcanic activity in the Lake Lutoto and Lake Nshere region to approximately 50,000 years ago (Boven et al. 1998). This date places an upper limit on the origin of the two lakes and hence also on the time of their colonization by fishes. The date is younger than the estimated time required for the mtDNA control region haplotypes now present in *H.* "Nshere" and *H.* "Lutoto" to coalesce to the MRCA. The difficulty with coalescence time estimates, however, is that the calibration of the mtDNA molecular clock is unreliable. For the East African cichlid fishes, the clock is commonly calibrated by the geological age of Lake Malawi (Meyer et al. 1990; Nagl et al. 2000; Sturmbauer et al. 2001), which, however, is uncertain (Schlüter 1997). The estimates range from 0.5 to 4 Myr, depending on the part of the lake examined and the dating method. In an earlier publication (Nagl et al. 2000), we calibrated the clock on both the 2 Myr and 4 Myr age of Lake Malawi. Here, we also take the younger age estimate (1 Myr) into account. Thus calibrated, the molecular clock dates the MRCAs of the mtDNA control region of *H.* "Nshere" to 60,000 or 130,000 years ago and that of *H.* "Lutoto" to 40,000 or 80,000 years ago (the two alternatives represent datings based on the 1 or 2 Myr age of Lake Malawi). Both estimates for *H.* "Nshere" exceed

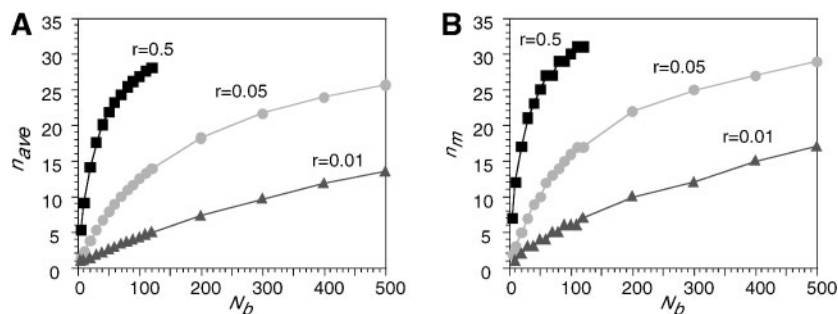


FIG. 10.—The results of computer simulations showing the relationship between the founding population size and the number of alleles retained in the population. N_b is the number of individuals in the founding size; r is the growth rate. (A) n_{ave} is the average number of alleles retained in the population. (B) n_m is the 95% upper limit of the number of alleles. N_b gives the lower limit of the founding population size for retaining n_m alleles. The number of alleles in the population exceeds n_m in less than 5% of iterations.

the upper limit set by the geological dates, but the *H. “Lutoto”* estimates come reasonably close to it. Of course, large errors are associated with these time estimates (table 2), and the geological dating of the crater lakes is subject to uncertainty (Boven et al. 1998). It is therefore difficult to tell whether or not the age of the MRCAs of the crater lake fishes really exceeds the geological date of the lakes. In principle, there can be two reasons for the discrepancy between molecularly and geologically estimated dates. First, the coalescence time to the MRCA need not coincide with the time of lake colonization and emergence of a species because polymorphism could have passed through the speciation phase. In the case of *H. “Nshere”* and *H. “Lutoto,”* however, there is no indication that any of the mtDNA control region haplotypes now present in the population existed before the emergence of these species. Second, Lake Malawi could in fact be younger than 1 Myr or the divergence of the Lake Victoria basin cichlids from the Lake Malawi fishes might have significantly postdated the emergence of the lake (Sturmbauer et al. 2001). Be this as it may, a large standard error of the molecular clock estimate leaves the door open for a younger age of the MRCA. In relative terms, the ages of *H. “Nshere”* and *H. “Lutoto”* may be comparable to those of the endemic Lake Victoria species.

Both the SINE and the mtDNA data indicate that the populations of *H. “Nshere”* and *H. “Lutoto”* are quite large. The lower limit of N (effective size of a whole population) from the SINE data is 6,000 for *H. “Lutoto”* and 24,000 for *H. “Nshere,”* and the N_f (female effective population size) estimate from the mtDNA control region data is 38,000 to 77,000 for *H. “Nshere”* and 8,000 to 17,000 for *H. “Lutoto.”* For lakes that have each a surface area of less than 1 km², these are indeed large effective population sizes. In a large population, selection—where applicable—gains the upper hand over random genetic drift. It can lead to the rapid fixation of mutations introducing adaptive modifications. This potential does not seem to have been exploited by fishes of Lake Nshere and Lake Lutoto, but has apparently been capitalized on by fishes in some other crater lakes and by haplochromines of Lake Victoria. In West Africa, the volcanic crater lakes Barombi Mbo and Bermin in Cameroon harbor 11 and nine tilapiine cichlid species, respectively (Trewavas, Green, and Corbet 1972; Stiassny, Schlieven, and

Dominey 1992), which apparently arose in situ from a single ancestor species by sympatric speciation (Schlieven, Tautz, and Pääbo 1994). Similarly, Lake Tana in Ethiopia provides a home for 14 morphotypes of large barbels (*Barbus intermedius* complex), all adapted to different ecological niches, exploiting different food resources, and derived from a common ancestral species within the lake (Nagelkerke, Sibbing, and Osse 1995; Dixon et al. 1996). The tremendous morphological diversity of the endemic Lake Victoria haplochromines is well documented (Greenwood 1981; Seehausen 1996). The difference between Lake Victoria on the one hand and Lake Nshere and Lake Lutoto on the other is, of course, that Lake Victoria, being the third largest lake in the world, offers many different ecological niches for adaptive radiation that the crater lakes do not. By contrast, Lake Barombi Mbo and Lake Bermin on the one hand and Lake Nshere and Lake Lutoto on the other are comparable in size. Whether they are also comparable in the number of ecological niches they potentially provide for fishes is not known to us. The reasons why cichlids adaptively radiated in the West African but not in the two East African crater lakes are unclear.

The lower limits of N (6,000 for *H. “Lutoto”* and 24,000 for *H. “Nshere”*) obtained in this study were based on the assumption that the size remained constant over 50,000 years. If this was the case, the founding populations of these two species must have been quite large. Although a bottleneck phase is not excluded by these considerations, if it did occur it must have been of short duration and followed by very rapid expansion of the population to its present size. The lower limit of the bottleneck is indicated by the *Mhc* data. It is 10 to approximately 300 breeding individuals, depending on how the data are interpreted. These numbers make the hypothesis involving the “seeding” of the lakes by fish eggs or fry transmitted by animals an unlikely proposition. Even if birds were able to deliver up to 300 founders into each of the two lakes, one would not expect the eggs or fry to be of the same origin in terms of species and locality, whereas all the available evidence suggests a monophyletic origin of both *H. “Nshere”* and *H. “Lutoto”* in a single founding event. Most likely, the founders of *H. “Nshere”* and *H. “Lutoto”* reached the lakes by way of rivers, and perhaps the way to Lake Lutoto was longer or more difficult than the way to

Lake Nshere, so fewer fish reached the former than the latter body of water. In both cases, however, the genetic diversity of the founding populations might have been greater than was ultimately retained. This conclusion is indicated by a comparison of the genetic variability currently found in the crater lake populations and that present among the Kazinga Channel fishes. Some loss of diversity apparently occurred by random genetic drift during the expansion phase after the founding event. The isolation of the crater lake population might have been another factor contributing to the loss of variability.

All the genetic systems we tested consistently indicate that compared with *H. "Nshere," H. "Lutoto"* is genetically a more homogeneous population. Whereas in *H. "Nshere"* we detected 10 different mtDNA control region haplotypes, found four of the six *Mhc* class II *B* loci to be polymorphic with two to four alleles per locus, could establish the presence or absence of polymorphism at two of the five SINE loci, and found four alleles at the SINE 1357 locus, in *H. "Lutoto"* we found seven mtDNA alleles, all the *Mhc* class II *B* loci to be monomorphic, one of the SINE loci to be polymorphic, and only one SINE 1357 allele. The reason for this difference between *H. "Nshere"* and *H. "Lutoto"* populations is not immediately apparent. The founding population of *H. "Lutoto"* might have been smaller than that of *H. "Nshere,"* as mentioned earlier, or, as the Tajima test indicates, the *H. "Lutoto"* population might have experienced a recent bottleneck and currently be in a recovering phase as a consequence of some event such as a change in water level.

In contrast to molecular diversification, morphological diversification reveals just the opposite trend when the populations of the crater lakes and Lake Victoria are compared. The two crater lake species are rather similar in appearance, the differences between them in body shape and coloration being inextensive. Similarly, the morphological variation within each crater lake population is relatively minor (H. Tichy and E. Schraml, unpublished data). By contrast, the endemic Lake Victoria species have diverged much more extensively from each other morphologically and each shows considerable intraspecific variation (Greenwood 1981; Seehausen 1996). These disparate trends can be attributed to the differences in opportunities for adaptation to distinct environmental niches on the one hand and to opportunities to mix diverging populations in the large and small lakes on the other hand.

At the molecular level, the most striking difference between speciation of haplochromines in the crater lakes and in Lake Victoria is in the fixation of substitutions. Although, as stated earlier, the endemic haplochromines in the crater lakes and in Lake Victoria had approximately the same length of time to evolve, the crater lakes species each accumulated two or three diagnostic substitutions in the control region of their mtDNAs since the time of their origin from the founding populations, but the endemic Lake Victoria species tested thus far have not accrued any such fixed changes (Nagl et al. 1998, 2000). This difference is most likely attributable to the different degrees of isolation of the populations. Whereas the crater lake populations apparently evolve in complete physical iso-

lation with only one species per lake, in Lake Victoria the reproductive barriers between the emerging species may be leaky. The barriers may be sufficiently strong to segregate characters under selection, differentiating the populations morphologically and behaviorally for longer periods of time, but not strong enough to assure fixation of neutral mutations by random genetic drift. Occasional breaches of the barrier may keep these mutations in a polymorphic state. An important implication of these deductions is the tentativeness of the early phases of speciation in the endemic Lake Victoria haplochromines. We suggest that as long as the environmental conditions in the lake remain stable, the different species may evolve in isolation from one another. Dramatic changes in the conditions may, however, usher in a phase of a limited gene flow between some of the species and so prevent fixation of neutral mutations. The ease with which the Lake Victoria haplochromines produce interspecies hybrids in aquaria attests to the weakness of the reproductive barriers between the species.

Acknowledgments

We thank Erwin Schraml for help in sample collecting during the expeditions, as well as Sabine Rosner for technical and Jane Kraushaar for editorial assistance.

Literature Cited

- Awise, J. C., J. Arnold, R. M. Ball, E. Bermingham, T. Lamb, J. E. Neigel, C. A. Reeb, and N. C. Saunders. 1987. Intraspecific phylogeography: the mitochondrial DNA bridge between population genetics and systematics. *Annu. Rev. Ecol. Syst.* **18**:489–522.
- Boven, A., P. Pastels, L. E. Punzalan, T. K. Yamba, and J. H. Musisi. 1998. Quaternary perpotassic magmatism in Uganda (Toro-Ankole Volcanic Province): age assessment and significance for magmatic evolution along the East African Rift. *J. Afr. Earth Sci.* **26**:463–476.
- Dixon, B., L. A. J. Nagelkerke, F. A. Sibbing, E. Egberts, and R. J. M. Stet. 1996. Evolution of *MHC* class II b chain-encoding genes in the Lake Tana barbel species flock (*Barbus intermedius* complex). *Immunogenetics* **44**:419–431.
- Donaldson, K. A., and R. R. Wilson. 1999. Amphipanamic geminates of snook (Percoidei: Centropomidae) provide a calibration of the divergence rate in the mitochondrial DNA control region of fishes. *Mol. Phylogenet. Evol.* **13**:208–213.
- Figuerola, F., W. E. Mayer, H. Sülmann, C. O'hUigin, H. Tichy, Y. Satta, N. Takezaki, N. Takahata, and J. Klein. 2000. *Mhc* class II *B* gene evolution in East African cichlid fishes. *Immunogenetics* **51**:556–575.
- Figuerola, F., H. Ono, H. Tichy, C. O'hUigin, and J. Klein. 1995. Evidence for insertion of a new intron into an *Mhc* gene of perch-like fish. *Proc. R. Soc. Lond. B Biol. Sci.* **257**:325–330.
- Greenwood, P. H. 1981. *The Haplochromine fishes of the East African lakes*. Cornell University Press, Ithaca, New York.
- Hudson, R. R. 1983. Testing the constant-rate neutral allele model with protein sequence data. *Evolution* **37**:203–217.
- Imanishi, T., A. Wakisaka, and T. Gojobori. 1991. Genetic relationships among various human populations indicated by *MHC* polymorphisms. Pp. 627–632 in K. Tsuji, M. Aizawa, and T. Sasazuki, eds. *HLA 1991. Proceedings of the eleventh international histocompatibility workshop and conference*. Oxford University Press, Oxford.

- 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.
- Klein, J., D. Klein, F. Figueroa, A. Sato, and C. O'hUigin. 1997. Major histocompatibility complex genes in the study of fish phylogeny. Pp. 271–283 in T. D. Kocher and C. A. Stepien, eds. *Molecular systematics of fishes*. Academic Press, New York.
- Klein, J., A. Sato, S. Nagl, and C. O'hUigin. 1998. Molecular trans-species polymorphism. *Annu. Rev. Ecol. Syst.* **29**:1–21.
- Málaga-Trillo, E., Z. Zaleska-Rutczynska, B. McAndrew, V. Vincek, F. Figueroa, H. Sülmann, and J. Klein. 1998. Linkage relationships and haplotype polymorphism among cichlid *Mhc* class II *B* loci. *Genetics* **149**:1527–1537.
- 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.
- Nagelkerke, L. A. J., F. A. Sibbing, and J. W. M. Osse. 1995. Morphological divergence in the large barbs (*Barbus* spp.) of Lake Tana, Ethiopia. *Neth. J. Zool.* **45**:431–454.
- 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.
- Nagl, S., H. Tichy, W. E. Mayer, N. Takezaki, N. Takahata, and J. Klein. 2000. The origin and age of the haplochromine species flock in Lake Victoria, East Africa. *Proc. R. Soc. Lond. B Biol. Sci.* **267**:1049–1061.
- Nei, M. 1987. *Molecular evolutionary genetics*. Columbia University Press, New York.
- Nei, M., and S. Kumar. 2000. *Molecular evolution and phylogenetics*. Oxford University Press, Oxford.
- Okada, N. 1991. SINES. *Curr. Opin. Genet. Dev.* **1**:498–504.
- Okada, N., M. Hamada, I. Ogiwara, and K. Ohshima. 1997. SINES and LINES share common 3' sequences: a review. *Gene* **205**:229–243.
- Ono, H., C. O'hUigin, H. Tichy, and J. Klein. 1993b. Major-histocompatibility-complex variation in two species of cichlid fishes from Lake Malawi. *Mol. Biol. Evol.* **10**:1060–1072.
- Ono, H., C. O'hUigin, V. Vincek, and J. Klein. 1993a. Exon-intron organization of fish major histocompatibility complex class II *B* genes. *Immunogenetics* **38**:223–234.
- Saitou, N., and M. Nei. 1987. The neighbor-joining method: a new method for reconstructing phylogenetic trees. *Mol. Biol. Evol.* **4**:406–425.
- Sato, A., D. Klein, H. Sülmann, F. Figueroa, C. O'hUigin, and J. Klein. 1997. Class I *Mhc* genes of cichlid fishes: identification, expression, and polymorphism. *Immunogenetics* **46**:63–72.
- Schliwen, U. K., D. Tautz, and S. Pääbo. 1994. Sympatric speciation suggested by monophyly of crater lake cichlids. *Nature* **368**:629–632.
- Schlüter, T. 1997. *Geology of East Africa*. Gebrüder Borntraeger, Berlin.
- Schmid, C., and R. Maraia. 1992. Transcriptional regulation and transpositional selection of active SINE sequences. *Curr. Opin. Genet. Dev.* **2**:874–882.
- Seehausen, O. 1996. *Lake Victoria rock cichlids: Taxonomy, ecology, and distribution*. Verduijn Cichlids, Zevenhuizen, Netherlands.
- Shand, R., and B. Dixon. 2001. Teleost major histocompatibility genes: diverse but not complex. *Mod. Asp. Immunobiol.* **2**:66–72.
- Simonsen, K. L., G. A. Churchill, and C. F. Aquadro. 1995. Properties of statistical tests of neutrality for DNA polymorphism data. *Genetics* **141**:413–429.
- Stiassny, M. L. J., U. K. Schliwen, and W. J. Dominey. 1992. A new species flock of cichlid fishes from Lake Bermin, Cameroon with a description of eight new species of *Tilapia* (Labroidei; Cichlidae). *Ichthyol. Explor. Freshwaters* **3**:311–346.
- Sturmbauer, C., S. Baric, W. Salzburger, L. Rüber, and E. Verheyen. 2001. Lake level fluctuations synchronize genetic divergences of cichlid fishes in African lakes. *Mol. Biol. Evol.* **18**:144–154.
- Tajima, F. 1983. Evolutionary relationship of DNA sequences in finite populations. *Genetics* **105**:437–460.
- . 1989. Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. *Genetics* **123**:585–595.
- Tamura, K., and M. Nei. 1993. Estimation of the number of nucleotide substitutions in the control region of mitochondrial DNA in humans and chimpanzees. *Mol. Biol. Evol.* **10**:512–526.
- Terai, Y., N. Takezaki, W. E. Mayer, H. Tichy, N. Takahata, J. Klein, and N. Okada. 2003. Phylogenetic relationships among East African haplochromine fishes as revealed by short interspersed elements (SINES). *J. Mol. Evol.* (in press).
- Thompson, J. D., D. G. Higgins, and T. J. Gibson. 1994. Clustal W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, positions-specific gap penalties and weight matrix choice. *Nucleic Acids Res.* **22**:4673–4680.
- Trewavas, E., J. Green, and S. A. Corbet. 1972. Ecological studies on crater lakes in West Cameroon fishes of Barombi Mbo. *J. Zool.* **167**:41–95.
- Vincek, V., C. O'hUigin, Y. Satta, N. Takahata, P. T. Boag, P. R. Grant, B. R. Grant, and J. Klein. 1997. How large was the founding population of Darwin's finches? *Proc. R. Soc. Lond. B Biol. Sci.* **264**:111–118.
- Yang, Z. 2002. PAML A program package for phylogenetic analysis by maximum likelihood. Version 3.12. University College London.

Naruya Saitou, Associate Editor

Accepted April 20, 2003