THE ROYAL SOCIETY

The origin and age of haplochromine fishes in Lake Victoria, East Africa

Sandra Nagl[1], Herbert Tichy[1], Werner E. Mayer[1], Naoko Takezaki[2], Naoyuki Takahata[2] and Jan Klein[1*]

[1]*Max-Planck-Institut für Biologie, Abteilung Immungenetik, Corrensstrasse 42, D-72076, Germany*
[2]*Graduate University for Advanced Studies, Department of Biosystems Science, Hayama, Kanagawa 240-0193, Japan*

According to a widely held view, the more than 300 species of haplochromine cichlid fishes in Lake Victoria (LV), East Africa, originated from a single founder species in less than 12 000 years. This view, however, does not follow from the published geological and molecular evidence. The former does indeed suggest that the LV basin dried out less than 15 000 years ago, but it does not provide any information about the species that re-colonized the new lake or that remained in the rivers draining the area. The molecular evidence is inconclusive with respect to the origin of the LV haplochromines because cichlids from critical regions around LV were not adequately sampled; and as far as the age of the LV haplochromines is concerned, it in fact led to an estimate of 250 000–750 000 years old. In the present study, mitochondrial DNA (control region) variation was determined by heteroduplex and sequencing analyses of more than 670 specimens collected at widely distributed East African riverine and lacustrine localities. The analyses revealed the existence of seven haplogroups (I–VII) distinguishable by characteristic substitutions. All endemic LV samples tested fell into one of these haplogroups (V) which, however, was also found to be present at various other localities, both riverine and lacustrine, outside LV. Within this haplogroup, four subgroups (VA through VD) could be distinguished, two of which (VB and VC) were represented in LV and at other localities. The great majority of the LV haplochromine species could be classified as belonging to the VC subgroup, which was found only in LV and in the rivers draining into it. Hence, while the endemic haplochromine species of LV could not have originated from a single founding population, the lake does harbour a large species flock which probably arose *in situ*.

Keywords: adaptive radiation; mitochondrial DNA; speciation; cichlid fishes; species flock

1. INTRODUCTION

Cichlid fishes (family Cichlidae, order Perciformes) are widely distributed in the rivers and lakes of Africa, South America, India and Sri Lanka, but they are particularly numerous and rich in species in the East African Great Lakes (Victoria, Malawi and Tanganyika), where they are undergoing bursts of adaptive radiation (Fryer & Iles 1972; Keenleyside 1991). All cichlid species of the Lake Victoria (LV) basin (i.e. Lakes Victoria, Kioga, Edward, George, Nabugabo and Kivu) are collectively referred to as the 'haplochromines' or the 'haplochromine species flock'; they are related to a similar collection of haplochromine fishes in Lake Malawi and presumably derive from a common ancestor of the tribe Haplochromini in Lake Tanganyika (Eccles & Trewavas 1989). Originally, almost all the LV cichlids were included in the genus *Haplochromis* by at least some authors. Later, five species, *Astatoreochromis alluaudi*, *Hoplotilapia retrodens*, *Platytaeniodus degeni*, *Macropleurodus bicolor* and *Paralabidochromis victoriae*, were assigned each to a different monotypic genus (Greenwood 1956). Later still, Greenwood (1979, 1980) split the genus *Haplochromis* into more than 20 genera and this classification has been accepted by some authors (e.g. Lippitsch 1993, 1995; Seehausen 1996), but not by others (Daget *et al*. 1991; Van Oijen 1996).

Approximately 150 LV haplochromine species have been formally described, but the actual number of existing species may be twice as high (Greenwood 1981; Daget *et al*. 1991; Van Oijen 1996; Seehausen 1996). The individual species have adapted to various ecological niches existing in the lake, the major adaptations being to different food resources. They differ in the size and shape of their bodies, head morphology, male breeding coloration, trophic specialization and breeding behaviour. Although field observations documenting disassortative mating of the different forms are mostly lacking, the low frequency of intermediate forms suggests that few interspecies hybridizations take place. In aquaria, however, many of the species interbreed and produce fertile hybrids, although some are reproductively isolated by post-zygotic mechanisms (Crapon de Caprona & Fritzsch 1984; A. Sato, Z. Zaleska-Rutczynska and J. Klein, unpublished data). Surveys of other regions in East Africa led Greenwood (1981, p. i) to conclude that 'about 99% of the haplochromine species from the lake are strictly endemic, with only a few penetrating the affluent and effluent rivers of the lake, and then for no great distances'. The endemism was explained by assuming that the species flock arose *in situ*, in the lake, from a single founding lineage (Greenwood 1981). Seemingly strong support for this interpretation was provided by mitochondrial DNA (mtDNA) control region sequences and the observation that on phylogenetic trees the sequences of the LV endemics clustered in a single clade (Meyer *et al*. 1990). Haplochromines, however, also inhabit other East African lakes and rivers and since these were not adequately sampled in this

*Author for correspondence (jan.klein@tuebingen.mpg.de).

Received 29 November 1999 *Accepted* 21 February 2000

particular study, the question whether the flock evolved in LV or in some other body of water was not answered. Particularly critical for the study's conclusion was the omission of samples from Lakes Edward, George and Albert (Lake Edward region, LER), which are older than the modern LV (Pickford *et al*. 1993; Schlüter 1997). Since river connections exist and apparently existed in a much extended form in the past between the LER and the LV region, the possibility must be considered that the LV species flock began to radiate elsewhere, before it migrated into LV where the radiation then continued. Such a scenario would resolve the paradox which arose in the meantime: from the intra- and interflock mtDNA divergences, it was estimated that the flock was between 250 000 and 750 000 years old (Meyer *et al*. 1990); yet, subsequently published geological data (Johnson *et al*. 1996) indicated that the ancient LV dried out in the Late Pleistocene. The basin began to fill with water from rivers crossing it only *ca*. 12 400 years ago. Indeed, in a recent study based on ribosomal RNA internal transcribed spacer 1 (ITS-1) sequences, the authors concluded that 'the modern assemblage [i.e. the LV flock] must have been derived from reinvasion by the products of earlier cladogenesis events' (Booton *et al*. 1999, p. 273). This conclusion is based on the observation that out of the 15 sequences compared (one sequence per species), one sequence from Lake Edward and another from Lake Kachira near LV intermingled phylogenetically with the ten LV flock sequences. Since the sample sizes in this study were inadequate to support such a conclusion, the question of the LV flock's origin and age must be regarded as unresolved. The present study was undertaken in an attempt to answer the question by sequencing the mtDNA control region of representative specimens from the LV region (comprising LV, L. Nabugabo, L. Kayugi and L. Kayania), the LER and major East African river systems.

2. MATERIAL AND METHODS

(a) *Source and isolation of DNA*

Cichlid fishes were collected during expeditions in 1993, 1995, 1996 and 1998 (table 1); a few were kindly provided by Dr Lothar Seegers (Dinslaken, Germany). Voucher specimens of the collected species have been deposited at the Nationaal Natuurhistorisch Museum, Leiden, The Netherlands, the Musée Royal de l'Afrique Centrale, Turveren, Belgium, and the Max-Planck-Institut für Biologie, Tübingen, Germany. Since the taxonomy of the riverine haplochromines remains unresolved, no attempt was made to identify them all the way down to species level; these samples are, instead, referred to only by the localities of their origin (figure 1). Pieces of fins from the collected specimens were stored in 70% ethanol and genomic DNA was isolated with the QIAamp Tissue Kit (Qiagen, Hilden, Germany). Contaminating RNA was removed by digestion with RNase A (30 min at 37 °C), which was then followed by phenol–chloroform extraction.

(b) *Polymerase chain reaction, cloning and sequencing*

Amplifications were carried out in PTC-100 and PTC-200 Programmable Thermal Controllers (MJ Research, Oldendorf, Germany). Genomic DNA (50–100 ng) was added to the reaction mixture of 1 × polymerase chain reaction (PCR) buffer (50 mM KCl, 1.5 mM $MgCl_2$, 10 mM Tris HCl, pH 9.0), 0.2 mM of each of the four deoxynucleoside triphosphates (Amersham Pharmacia Biotech, Freiburg, Germany), 0.2 μM of each of the sense (S) and antisense (A) primers and 2.5 units of *Taq* DNA polymerase (Amersham Pharmacia Biotech). The PCR programme consisted of a denaturation step for 3 min at 94 °C, followed by 35 cycles, each cycle consisting of 40 s denaturation at 94 °C, 30 s annealing at 51–54 °C, depending on the primer combination, and a 2 min extension at 72 °C. The reactions were completed by a final extension step for 10 min at 72 °C. Hot-start PCR amplifications were carried out as above, except that a 1 × $MgCl_2$-free PCR buffer was used instead of the standard buffer, and 1.5 mM HotWax Mg^{2+} beads (Invitrogen, Leek, The Netherlands) were added to the mixture. The PCR primers used (SNmt-UP1: 5′-TAAAATCCTTCCTACTGCTTCA-3′ and SNmt-LP1: 5'-TCAAACAAAATATGAATAACAAACA-3′) were specific for the $tRNA^{Pro}$ $tRNA^{Thr}$ gene and the 3′-end of the control region, respectively; the amplification product therefore encompassed almost the entire control region (*ca*. 850 base pairs (bp)). The products were cloned in the pGEM-T Easy Vector (Promega, Mannheim, Germany) and sequenced by using the LI-COR 4200 DNA sequencer (MWG Biotech, Ebersberg, Germany).

To exclude the possibility that some of the sequences may have been derived from fragments of mtDNA integrated into nuclear DNA (so-called 'numts'), representative samples of each haplogroup were used as templates for extra-long PCR. The primers SNmt-618LP: 5′-TCAACCTTGTCTGTCATTTCTG TATGCACTCTGAG-3′ and SNmt-868UP: 5′-TGCAAACCC CCCGGAAACAGGAAAAGCTCT-3′ were located in the control region and were chosen so as to amplify nearly the entire 16 kb mtDNA circle. The reaction was carried out in the PTC-100 Programmable Thermal Controller by using the Expand^{TM} Long Template PCR System Kit (Roche Diagnostics, Mannheim, Germany). The PCR product was then used as a template for a nested PCR with primers SNmt-UP1 and SNmt-531LP: 5′-TACCACCATTAACTTATGCAAGC-3′.

(c) *Heteroduplex analysis*

Primers SNmt-131UP: 5′-CAACATTTTAACTAAGGGGTA CATAA-3' and SNmt-280LP: 5′-TGATGGTGGGCTCTTAC TACATTA-3′ were used to amplify an approximately 120 bp-long segment from the 5′ variable part of the control region. Six microlitres of a PCR product were mixed with an equal volume of a reference PCR product. The samples were denatured for 5 min at 95 °C and annealed for 5 min at room temperature. An 8.2-μl aliquot of each sample was loaded on a polyacrylamide gel and subjected to electrophoresis as described previously (Nagl *et al*. 1998).

(d) *Sequence and phylogenetic analysis*

The mtDNA control region sequences were aligned by using the CLUSTALW 1.75 program (Thompson *et al*. 1994) and were then modified by visual inspection. There were 845 sites in total, including insertions and deletions, of which 18 sites were excluded from the analysis because of ambiguities in the alignment. The full alignment can be accessed in the electronic appendix available at The Royal Society Web site. A neighbour-joining tree (Saitou & Nei 1987) was drawn based on the proportion of nucleotide differences (*p*-distances). Gaps were considered as a fifth state and sites with unidentified nucleotide states were excluded from the analysis. Reliability of nodes was evaluated by 1000 bootstrap replications (Felsenstein 1985).

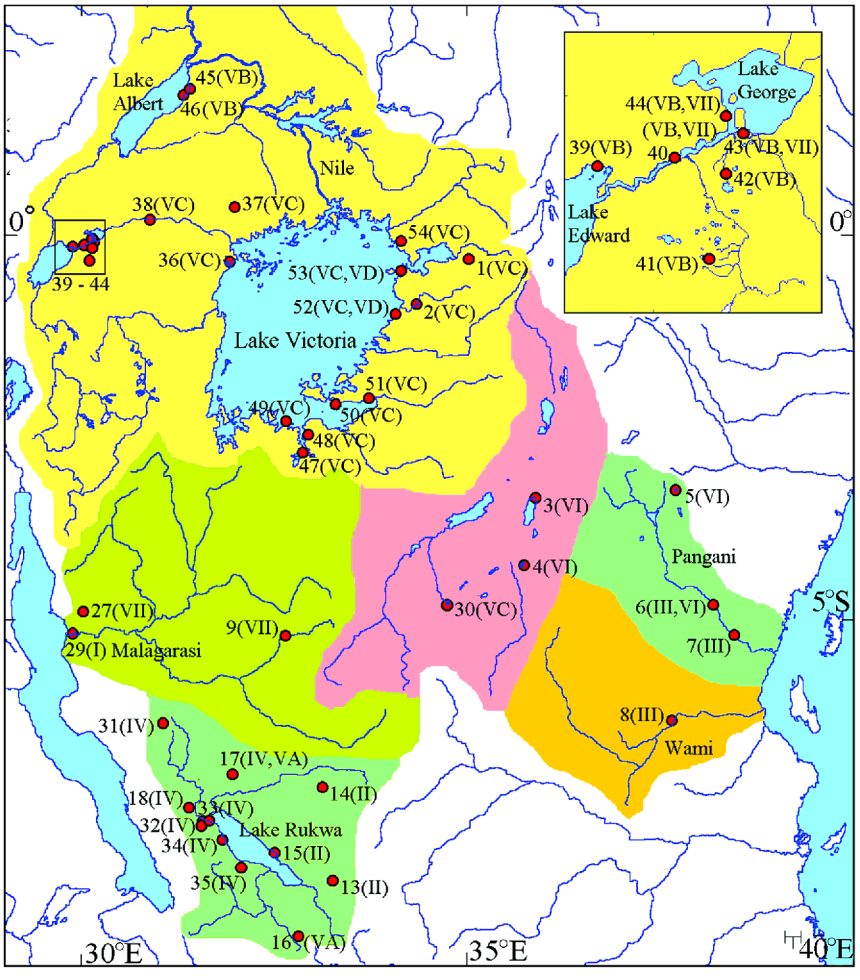


Figure 1. River systems (differentiated by colours) and major lakes of East Africa. Arabic numerals indicate localities at which haplochromines investigated in this study were collected. The mtDNA (control region) haplogroups found at these locations are indicated by Roman numerals in parentheses. Altogether, 670 samples were analysed. The names of the localities (and the number of individuals tested from each locality) are listed. 1, Pond Apida (1); 2, Migori River (1); 3, Lake Manyara (10); 4, Lake Babati (13); 5, Lake Chala (3); 6, Pangani River (3); 7, Pangani River (3); 8, Wami River (3); 9, Wala River (1); 13, Lupa River (2); 14, Piti River (6); 15, Lake Rukwa (1); 16, Myunga River (1); 17, Wogo River (7); 18, Kitilda/Rukwa (19); 27, Malagarasi River (2); 29, Igalala River/Malagarasi (7); 30, Lake Singida (8); 31, Mpanda River (17); 32, Muze River (17); 33, Lake Rukwa (19); 34, Zimba River (35); 35, Nakanga/Lake Rukwa (42); 36, Lake Nabugabo, Lake Kayugi and Lake Kayania (25); 37, Lake Wamala (12); 38, Kabagole/Katonga River (9); 39, Katwe/Lake Edward (27); 40, Katunguru Bridge/Kazinga Channel (33); 41, Lake Lutoto (10); 42, Lake Chibwera (10); 43, Kashaka/Lake George (21); 44, Kasenyi/Lake George (29); 45, Butiaba/Lake Albert (47); 46, Bugoigo, Lake Albert (28); 47, Buzumu Gulf (7); 48, Mwanza Gulf (20); 49, Chamagati Island (2); 50, Zue Island (8); 51, Speke Gulf (12); 52, Muhuru (81); 53, Rusinga (29); 54, Anyanga (39).

Because of the large number of sequences involved, maximum-parsimony trees (Fitch 1971) were drawn by conducting a heuristic search using the PAUP3.20d2 program (Swofford 1993). Ten thousand equally parsimonious trees were saved and a 50% majority consensus tree was constructed.

The molecular clock of the haplochromine mtDNA control region was calibrated by using the geological age of Lake Malawi as the divergence time of the Lake Malawi and LV flocks (but see § 3). Unfortunately, the geology of Lake Malawi is complex and the age of the lake uncertain, the estimates ranging from two to four million years (Myr) ago (Bromage *et al.* 1995). This range was therefore used for the estimates in table 2, which were obtained as follows: for each pair of sequences i and j, multiple-hit correction of substitutions was made based on the two-parameter model (Kimura 1980), and nucleotide substitutions per site d_{ij} were estimated. The rate parameter was assumed either to be constant (K2P) or to vary according to the γ-distribution (K2PG; Jin & Nei 1990). The α-parameter of the γ-distribution, which is inversely related to the extent of rate variation across sites, was estimated by parsimony-based (Kocher & Wilson 1991) and maximum-likelihood (Yang 1994) methods using the program package PAML (Yang 1998). In addition, estimates were also made based on transversions (Tv) alone and these were considered the most accurate because the transitions included in the other two estimates were partially saturated. To date, a node (c) between a pair of monophyletic clusters of sequences, the average number (d_c) of nucleotide substitutions in all pairwise comparisons was computed by using the formula

$$d_c = \frac{1}{mn}\sum_{i=1}^{m}\sum_{j=1}^{n} d_{ij},$$

Table 1. *List of haplochromine species from Lake Victoria and Lake Edward regions tested in the present study*

(K. Ch., Kazinga Channel; LE, Lake Edward; LG, Lake George; LV, Lake Victoria; LVR, Lakes Victoria, Nabugabo, Kayugi and Kayania. c.f., *con forma*: a species complex; individual species are difficult to differentiate morphologically.)

species name	abbreviation	lake	locality	distribution	mtDNA group
Astatotilapia					
aeneocolor	*Asae*	LG	44	LG, K. Ch.	VB
elegans	*Asel*	LG	44	LE, LG, K. Ch.	VB
nubila	*Asnu*	LVR	36, 48, 53	LVR, L. Kyoga, LE, LG and rivers	VC
schubotziellus c.f.	*Assch*	LG	43, 44	LG, K. Ch., probably LE	VB, VII
velifer	*Asve*	LN	36	LN	VC
Enterochromis					
cinctus	*Enci*	LV	48, 51	LV	VC
nigripinnis	*Enni*	K. Ch.	40	LE, LG, K. Ch.	VB
Gaurochromis					
angustifrons c.f.	*Gaan*	K. Ch., LG	40, 43, 44	LE, LG, K. Ch.	VB, VII
simpsoni	*Gasi*	LN	36	LN	VC
Haplochromis					
lividus	*Hali*	LV	48	LV	VC
petronius	*Hape*	LG	43	LG	VII
velvet black	*Havb*	LV	48, 49, 50	LV	VC
Harpagochromis					
squamipinnis	*Harsq*	LE	39	LE, LG, K. Ch.	VB
Lipochromis					
melanopterus	*Lime*	LV	51, 48	LV	VC
Neochromis					
nigricans	*Neni*	LV	48, 50, 51	LV, Victoria Nile	VC
Prognathochromis					
venator	*Prve*	L. Kayugi, L. Kayania	36	LN, L. Kayugi, L. Kayania	VC
Paralabidochromis					
beadlei	*Pabe*	LN	36	LN	VC
chilotes	*Pach*	LV	52, 53, 54	LV, Victoria Nile	VC
labiatus	*Pala*	LE	39	LE, LG	VB
plagiodon	*Papl*	LV	47, 48	LV	VC
Psammochromis					
riponianus	*Psri*	LV	48	LV, Victoria Nile	VC
schubotzi	*Pssc*	LE	39	LE, LG, K. Ch.	VB
Ptyochromis					
sauvagei	*Ptsa*	LV	48, 53	LV	VC
xenognathus	*Ptxe*	LV	48, 53,	LV	VC
Yssichromis					
laparogramma	*Yila*	LV	47, 48, 53	LV	VC, VB
pappenheimi	*Yipa*	LG	43	LG, LE, K. Ch.	VB
'mbipi'	—	LV	48, 51, 52, 53, 54	LV	VC, VD

where m and n are the numbers of sequences in a monophyletic cluster and the other cluster, respectively. The 'height' in table 1 was defined as $d_c/2$. The sampling error for each d_c or height was estimated by the method of Takezaki *et al.* (1995). One-half of the sampling error of d_c was assigned as the sampling error of the height. The rate (r) of nucleotide substitutions was calibrated as $r = d_c/2T$, where T is the time of divergence (the coalescent point) of two haplogroups from a common ancestral lineage, under the assumption of a correspondence with geological separation. The time of divergence of the two populations was corrected by subtracting the average haplotype diversity within the populations from the haplotype diversity between the populations (Klicka & Zink 1999). This correction factor was about 33% for the divergence of the riverine cichlids from each other and from the Lake Malawi cichlids. Since the age estimate of the latter population, which was used to calibrate the divergence rate, was likewise dependent on the correction factor, the calculated rate was not dependent on the correction. Note that the divergence of subgroup VC from subgroups VA + VB was *ca.* 10% of the LV–Lake Malawi flocks divergence, irrespective of the distance measure used.

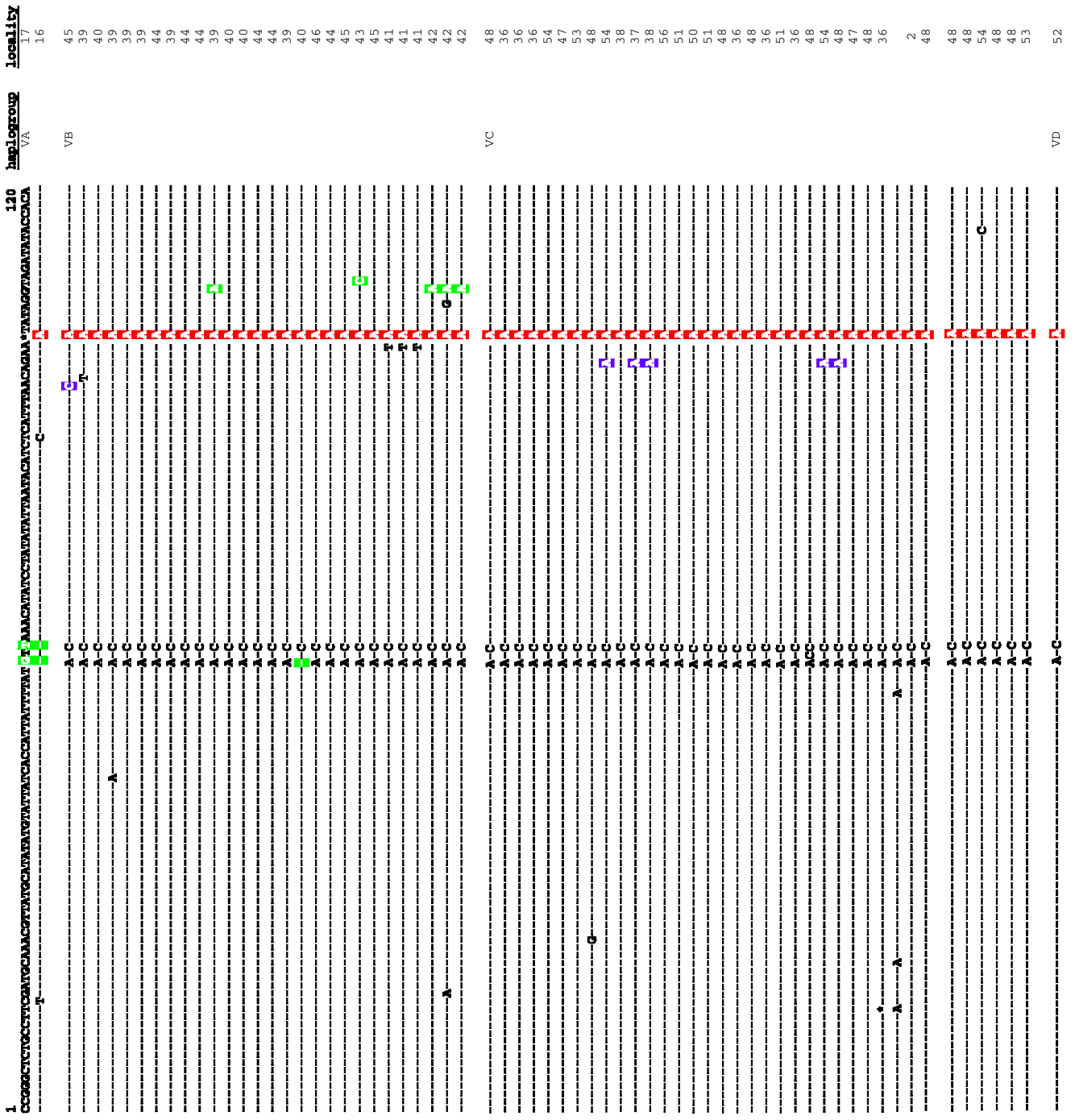
3. RESULTS

Haplochromines were sampled from the lakes and rivers in the area between the western and eastern branches of the East African Rift system (table 1 and figure 1). The control region of the mtDNA extracted from the samples was either sequenced or subjected to heteroduplex analysis. The sequencing established the presence of 114 control region haplotypes among the 164 samples tested. The haplotypes fall into seven haplogroups (I–VII), not counting those found in the Lake Malawi and Lake Tanganyika cichlid fishes. Each of the seven

Table 2. *Average heights of the nodes of the phylogenetic tree in figure 3 and divergence times of different clusters*

(The node numbers correspond to those shown in the neighbour-joining tree (figure 3). Node 0, Lake Tanganyika (LT) and *Astatoreochromis alluaudi*; node 1, LT and Lake Malawi (LM); node 2, LM and the rest; node 3, group V and the rest; node 4, groups II + IV and VI + I + III + VII; node 5, groups II and IV; node 6, groups VI and I + III + VII; node 7, groups I and III + VII; node 8, groups III and VII; node 9, group VD and the rest of group V; node 10, subgroups VA and VB; node 11, subgroups VB and VC. Average heights of the nodes are half the average distances between the two descendant clusters of the nodes. The values are given as percentages. The divergence times are estimated by assuming the node 2 to represent a divergence of 2 Myr or 4 Myr in columns A and B, respectively. The values are given in units of million years. K2P: Kimura two-parameter distance, K2PG: Kimura two-parameter distance with γ-parameters (taking the rate heterogeneity across sites into account) of $\alpha = 0.29$ and 0.11; Tv, transversional distance with Kimura two-parameter correction.)

	K2P			K2PG						Tv		
				$\alpha = 0.29$			$\alpha = 0.11$					
	height	time		height	time		height	time		height	time	
node		A	B		A	B		A	B		A	B
0	3.65 ± 0.43	2.27 ± 0.27	4.55 ± 0.544	4.39 ± 0.62	2.30 ± 0.33	4.61 ± 0.65	6.12 ± 1.14	2.36 ± 0.44	4.73 ± 0.88	1.63 ± 0.30	2.83 ± 0.52	5.67 ± 1.04
1	3.40 ± 0.39	2.12 ± 0.24	4.24 ± 0.49	4.08 ± 0.55	2.14 ± 0.29	4.28 ± 0.58	5.65 ± 0.99	2.18 ± 0.38	4.36 ± 0.76	1.22 ± 0.24	2.12 ± 0.42	4.24 ± 0.84
2	3.21 ± 0.37	2.00 ± 0.23	4.00 ± 0.46	3.81 ± 0.52	2.00 ± 0.27	4.00 ± 0.55	5.18 ± 0.91	2.00 ± 0.359	4.00 ± 0.70	1.15 ± 0.23	2.00 ± 0.40	4.00 ± 0.80
3	2.37 ± 0.32	1.48 ± 0.20	2.95 ± 0.40	2.69 ± 0.41	1.41 ± 0.22	2.82 ± 0.43	3.36 ± 0.62	1.30 ± 0.24	2.59 ± 0.48	0.81 ± 0.20	1.42 ± 0.35	2.82 ± 0.70
4	1.67 ± 0.24	1.04 ± 0.15	2.08 ± 0.30	1.85 ± 0.29	0.97 ± 0.15	1.94 ± 0.30	2.19 ± 0.40	0.85 ± 0.15	1.69 ± 0.31	0.40 ± 0.11	0.70 ± 0.19	1.39 ± 0.38
5	0.67 ± 0.19	0.48 ± 0.12	0.83 ± 0.24	0.69 ± 0.21	0.36 ± 0.11	0.72 ± 0.22	7.41 ± 0.24	2.86 ± 0.09	5.72 ± 0.19	0.13 ± 0.08	0.23 ± 0.14	0.45 ± 0.28
6	1.98 ± 0.29	1.23 ± 0.18	2.47 ± 0.36	2.21 ± 0.37	1.16 ± 0.19	2.32 ± 0.39	2.71 ± 0.53	1.05 ± 0.20	2.09 ± 0.41	0.49 ± 0.14	0.85 ± 0.24	1.70 ± 0.49
7	1.79 ± 0.30	1.12 ± 0.19	2.23 ± 0.37	1.98 ± 0.37	1.04 ± 0.19	2.08 ± 0.39	2.35 ± 0.52	0.91 ± 0.20	1.81 ± 0.40	0.54 ± 0.16	0.94 ± 0.28	1.88 ± 0.56
8	0.85 ± 0.19	0.53 ± 0.12	1.06 ± 0.24	0.88 ± 0.21	0.46 ± 0.11	0.92 ± 0.22	0.95 ± 0.24	0.37 ± 0.09	0.73 ± 0.19	0.28 ± 0.10	0.49 ± 0.17	0.97 ± 0.35
9	0.63 ± 0.17	0.39 ± 0.11	0.78 ± 0.21	0.65 ± 0.19	0.34 ± 0.10	0.69 ± 0.20	0.70 ± 0.21	0.27 ± 0.08	0.54 ± 0.16	0.06 ± 0.04	0.11 ± 0.07	0.21 ± 0.13
10	0.63 ± 0.17	0.39 ± 0.11	0.79 ± 0.21	0.66 ± 0.19	0.35 ± 0.10	0.69 ± 0.20	0.72 ± 0.22	0.28 ± 0.08	0.56 ± 0.19	0.03 ± 0.03	0.05 ± 0.05	0.10 ± 0.10
11	0.36 ± 0.11	0.22 ± 0.07	0.45 ± 0.14	0.37 ± 0.11	0.19 ± 0.06	0.39 ± 0.12	0.38 ± 0.12	0.15 ± 0.05	0.29 ± 0.09	0.09 ± 0.05	0.16 ± 0.09	0.31 ± 0.17


(a)
sequence
haplogroup
locality
VA
VB
VC
VD
Ecol
Arve616
Prve687
Prve691
Pach5722
Papl160
Pach
Ptze350
Ptze6864
8678
8632
8680
Amm586
Harc486
Hmz1
Hmz309
Ptze
Amm
Prvab121
Arve619
Lize
Arve605
Hmz1817
Ptze6865
Ptze326
Papl201
Ptze320
Arve663
HG012352
6701
Papl
Parl
Papl73
Pach5721
Yale179
Yale335
Yale6937
6745

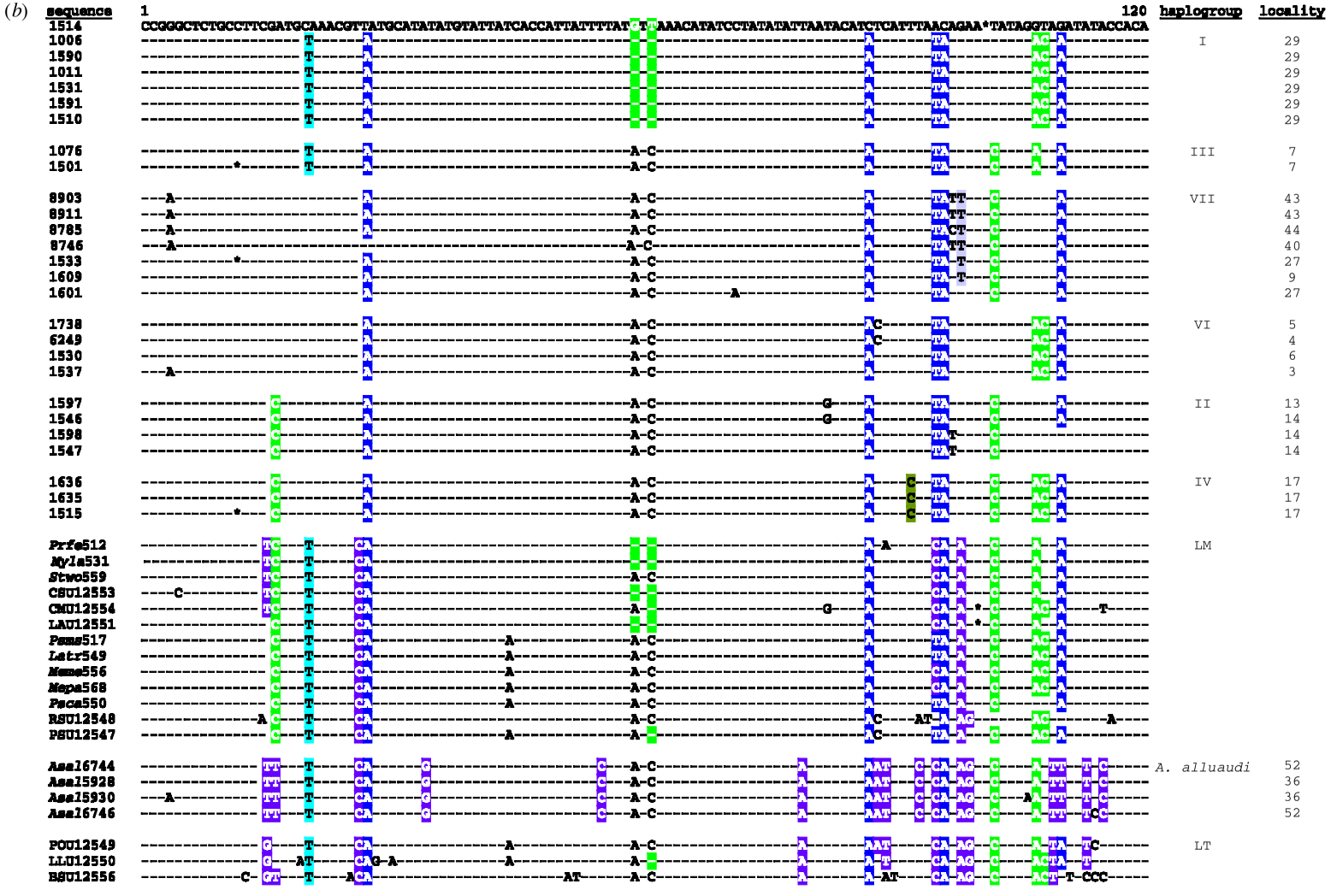


Figure 2(*a*, *b*). Alignment of mtDNA control region nucleotide sequences of East African haplochromine fishes (the first 120 sites; the full alignment can be viewed in the electronic appendix). Dashes indicate identity with the reference sequence at the top and asterisks indicate indels introduced for optimal alignment. Haplogroups are separated by gaps and are indicated by numbers and letters on the right-hand side. Diagnostic substitutions are highlighted. For species abbreviations see table 1. HOU12552, *Haplochromis obesus* (Lee *et al.* 1995). The Lake Malawi (LM) and Lake Tanganyika (LT) species were *Latr*, *Labeotropheus trewavasae*; *Meme*, *Melanochromis melanopterus*; *Mepa*, *Melanochromis parallelus*; *Myla*, *Mylochromis lateristriga*; *Prfe*, *Protomelas fenestratus*; *Psca*, *Pseudotropheus callainos*; *Psms*, *Pseudotropheus msobo*; *Stwo*, *Stigmatochromis woodi*. CSU12553, *Champsochromis spilorhynchus*; CMU12554, *Cyrtocara moorii*; LAU12551, *Lethrinops auritus*; PSU12547, *Petrotilapia* sp.; RSU12548, *Rhamphochromis* sp.; POU12549, *Petrochromis orthognathus*; LLU12550, *Lobochilotes labiatus*; BSU12556, *Bathybates* sp. (Lee *et al.* 1995). Specimens from the non-endemic species *Astatoreochromis alluaudi* (*Asal*) were caught in LV and Lake Kajaboli. Sequences identified by numbers only are from rivers.

Table 3. *Diagnostic sites of group V subgroups of haplochromine mtDNA control region sequences*

	subgroup			
site	VA	VB	VC	VD
59	G	A	A	A
61	T	C	C	C
128	T	C	C	C
271	T	T	C	T
423	T	T	T	C
438	C	C	C	T
635	A	A/G	G(A)	G

haplogroups is defined by 7–41 diagnostic substitutions (figure 2). A diagnostic substitution is shared by all or nearly all members of one group but is absent in most other groups. Haplotypes of the same group differ from one another by scattered substitutions at other sites. All groups are also distinguishable by their characteristic banding patterns on heteroduplex analysis and many more individuals (a total of 670) could therefore be assigned to them than were sequenced. Group V was highly heterogeneous in that it was represented by numerous heteroduplex patterns.

All samples of the presumably endemic haplochromine species from LV and its satellite, Lake Nabugabo (46 samples typed by sequencing and 209 by heteroduplex analysis) fell into group V. The sample set consisted of 15 species from nine different localities, eight near the eastern and one near the western coast (figure 1 and table 1). The species represented 12 genera, defined according to Greenwood's (1979, 1980) revision of the genus *Haplochromis*; they embodied eight different trophic groups. Out of the two non-endemic species tested, one (*Astatotilapia nubila*, formerly *Haplochromis nubilus*) also bore group V, whereas the other (*Astatoreochromis alluaudi*) did not. The former species has been reported from lakes and rivers in the entire LV basin, including various waters in the LER (Trewavas 1933; Greenwood 1965). *A. alluaudi* may have been limited originally to the LV basin, but more recently has also been introduced into other regions of East Africa (Greenwood 1956). Its mtDNA control region sequence places it outside the seven haplogroups in a position between Lake Tanganyika and Lake Malawi cichlids (figures 2 and 3). Group-V-bearing fishes were also found in the rivers draining into LV, in the lakes and rivers of the LER, in the Lake Rukwa region (localities 16 and 17) and in Lake Singida (locality 30, figure 1). The other haplogroups were distributed as follows: group I, Malagarasi River (locality 29); group II, Lake Rukwa region (localities 13, 14 and 15); group III, Wami (locality 8) and Pangani (localities 6 and 7); group IV, Lake Rukwa region (localities 17, 18, 31, 32, 33, 34 and 35); group VI, Lake Babati (localities 3 and 4) and the Pangani region (localities 5 and 6); group VII, the Malagarasi region (localities 9 and 27) and LER (localities 40, 43 and 44). Group V is thus the most widely distributed of the seven haplogroups, the remaining six showing a tendency toward restriction to one drainage system (figure 1).

The group V sequences are differentiated into four subgroups denoted VA through VD. All four subgroups share the substitutions diagnostic of group V (27T, 87C, 95A, 96C, 101A, 169G, 172C, 348A, 368T and 825A); they are distinguished by one or more substitutions characteristic of each of them (table 3). Subgroup VA was found at two localities (16 and 17) in the Lake Rukwa region. Subgroup VB was identified in haplochromine fishes collected in the LER (eight localities; see figure 1), and in one species from LV. The exceptional LV individual was caught at locality 53 and identified as representing the species *Yssichromis laparogramma*. The other seven individuals of this species from localities 47, 48 and 53 were typed as bearing the VC subgroup. Subgroup VC was found to be present in all tested specimens of the endemic LV haplochromine flock, with the exception of the VB- and VD-bearing individuals; it was also present in the rivers draining into LV (localities 1, 2, 36, 37 and 38). Among the genera tested were seven which are, according to Greenwood (1980), represented by different species in LV and in the LER: *Astatotilapia velifer* (LV region), *Astatotilapia schubotziellus* (LER), *Astatotilapia aeneocolor* (LER), *Astatotilapia elegans* (LER); *Enterochromis cinctus* (LV), *Enterochromis nigripinnis* (LER); *Gaurochromis simpsoni* (LV region), *Gaurochromis angustifrons* (LER); *Haplochromis lividus* (LV), *Haplochromis petronius* (LER); *Paralabdochromis chilotes* (LV), *Paralabdochromis beadlei* (LV region), *Paralabdochromis plagiodon* (LV), *Paralabdochromis labiatus* (LER); *Psammochromis riponianus* (LV), *Psammochromis schubotzi* (LER); *Yssichromis laparogramma* (LV), *Yssichromis pappenheimi* (LER). In each case (with the exception of the one *Y. laparogramma* individual mentioned above), all the LV region individuals were typed as belonging to the VC subgroup, while all the LER fishes could be classified as bearing VB-subgroup haplotypes. Finally, subgroup VD occurred in 41 specimens obtained from rocky-shore localities, one (locality 52) near the estuary of the Migori River (locality 2) and the other (locality 53) *ca.* 100 km north. The specimens had the phenotype of fishes belonging to the *Haplochromis rockkribensis* complex of rock-dwelling cichlids ('mbipi') described recently by Seehausen (1996). Other fishes from these two localities had a similar appearance, but typed as belonging to the VC subgroup, which otherwise encompasses the great majority of the haplochromine flock.

Differentiation of riverine control region sequences into seven haplogroups is also indicated by phylogenetic analyses, based on either genetic distances (figure 3) or maximum-parsimony considerations (figure 4). (Both methods yielded largely congruent results; because of the large number of sequences involved, the application of maximum-likelihood methods was judged as impractical.) The existence of the individual haplogroups is strongly supported by high bootstrap values, but the support for the order in which the groups diverged from each other is only modest or low. Apparently, the populations in which the different haplogroups arose diverged in close succession and subsequently evolved without significant gene flow between them. The seven haplogroups were rooted by a collection of control-region sequences obtained from Lake Malawi and Lake Tanganyika cichlids (listed in figure 2), as well as the non-endemic LV species *A. alluaudi*.

To date the divergences of the haplogroups, we used the geological record of the East African Great Lakes (Schlüter 1997). The record is, however, an unreliable source of information for divergence time estimates for two main reasons. First, since different parts of the lakes have different geological histories, the age estimates vary considerably according to the particular region analysed. And second, since a lake need not have been colonized immediately following its formation and may have undergone periods of desiccation throughout its history, its age is not necessarily the age of the fishes now living in it. We have dealt with these uncertainties in two ways. First, instead of taking a single value, we provide a range of divergence estimates based on the range of a lake's age estimates. Specifically, the age of Lake Malawi was taken to be in the range 2–4 Myr (Bromage *et al*. 1995; Schlüter 1997), and these values were then used to estimate the divergence time of the Lake Malawi and LV cichlids and so to calibrate the molecular clock of the mtDNA control region (table 2). Second, we applied to the cichlids the rate estimated for sharks on the basis of a reasonably complete fossil record (Martin *et al*. 1992). Since, however, the shark rate was obtained for the mtDNA cytochrome *b* gene, we used it first to estimate the divergence time of the LV and Lake Malawi cichlids from cytochrome *b* sequences deposited in the databases (Sturmbauer & Meyer 1993; Cantatore *et al*. 1994; Mayer *et al*. 1998) and then the value thus obtained to calibrate the control region clock. The LV–Lake Malawi divergence time estimate was 2 Myr, which corresponds to the lowest value of the geological range estimate. The estimated control region divergence rate of the cichlid fishes (5.6% per site per 10^6 years) is approximately three times slower than the corresponding human (mammalian) rate (Horai *et al*. 1995). That the teleost mtDNA evolves more slowly than mammalian mtDNA is also indicated by data calibrated by the emergence of the Panama Isthmus (Donaldson & Wilson 1999). Below we therefore use the lower values of the divergence time ranges given in table 2.

The oldest of the seven control region sequence assemblages is group V, which diverged from the ancestor of all the other haplogroups an estimated 1.4 Myr ago (table 2). The next split may have been between the ancestors of groups II + IV and the ancestors of groups III + VI + VII *ca*. 1.3 Myr ago, and the other divergences then followed in relatively short intervals (figure 3 and table 2). Within group V, the oldest split appears to have been of the subgroup VD from the ancestor of all the other subgroups an estimated 110 000 years ago. Whether subgroup VD represents a single species or species complex remains to be determined. The subgroup has thus far been found only in LV. It may encompass descendants of the species that survived the desiccation of the palaeolake either in rivers crossing the basin or in persisting pools.

The order in which the remaining three subgroups separated from one another is not resolved by either of the two phylogenetic trees. Apparently, the separation occurred within too short an interval to be unambiguously reflected in the available sequences. While subgroup VC is clearly separated from the other subgroups on both the distance and maximum-parsimony trees, subgroup VA appears to emerge from the midst of subgroup VB (see § 4). Unfortunately, the divergence time of subgroup VC from the other subgroups cannot be determined with a degree of precision one might have wished to achieve. The estimate in table 2 provides a value of 160 000 years which is, paradoxically, higher than the divergence time of the presumably oldest subgroup (VD). This discrepancy is undoubtedly a manifestation of the paucity of substitutions differentiating the subgroups as indicated by the high standard error values of the estimates. All one can conclude at present is that subgroup VC probably diverged from the other subgroups less than 160 000 years ago; it is highly improbable that the divergence took place 12 000 years ago or less. This latter conclusion, however, does not preclude the possibility that the morphological divergence of the species comprising subgroup VC occurred during the 12 000-year interval. In fact, we argue in § 4 that it very likely did.

4. DISCUSSION

Our study reveals a striking discrepancy between molecular and morphological variation in the East African cichlid fishes. On the one hand, in LV the haplochromines have diverged morphologically to the degree that they have been classified by some authors (Greenwood 1980; Lippitsch 1993, 1995; Seehausen 1996) into not only distinct species, but also different genera. At the molecular level, by contrast, the members of the flock display considerable trans-specifically shared polymorphism (Nagl *et al*. 1998) but no fixed substitutions differentiating the distinct morphological forms (see also Meyer *et al*. 1990). On the other hand, in the East African rivers, the haplochromines have been regarded by morphologists as a rather homogeneous group (Greenwood 1981), yet at the molecular level they fall into at least seven clearly distinguishable groups with genetic distances between groups in their control region sequences ranging from 0.0109 to 0.0592. If one were to follow the recent proposal by Avise & Johns (1999) to base classifications of living forms on molecularly determined genetic distances, all the members of the endemic LV flock would fall into a single species, whereas the riverine forms would constitute at least seven different genera.

There are at least three molecularly distinguishable types of haplochromines in LV. Most abundant are the fishes classified as belonging to subgroup VC. The species classically described as comprising the endemic LV flock—as far as they can be found and identified in the collections tested by the present study—all belong to this subgroup, with the single exception of the one individual of *Y. laparogramma*. One might be tempted to dismiss this exception as being the result of, for example, a mix-up, but a careful check provided no support for such an explanation. More importantly, a re-evaluation of sequences published by Meyer *et al*. (1990) reveals that the species *Prognathochromis paraguarti* reported by them as being caught in the vicinity of Jinja on the Ugandan side of LV, has in fact a subgroup VB haplotype. Meyer and colleagues also list in their study a sequence of *Platytaeniodus degeni* which we would classify as belonging to subgroup VB. *P. degeni*, is one of the genera denoted by Greenwood (1956) as monotypic and hence potentially perhaps outside of the main flock. Taken together, there is

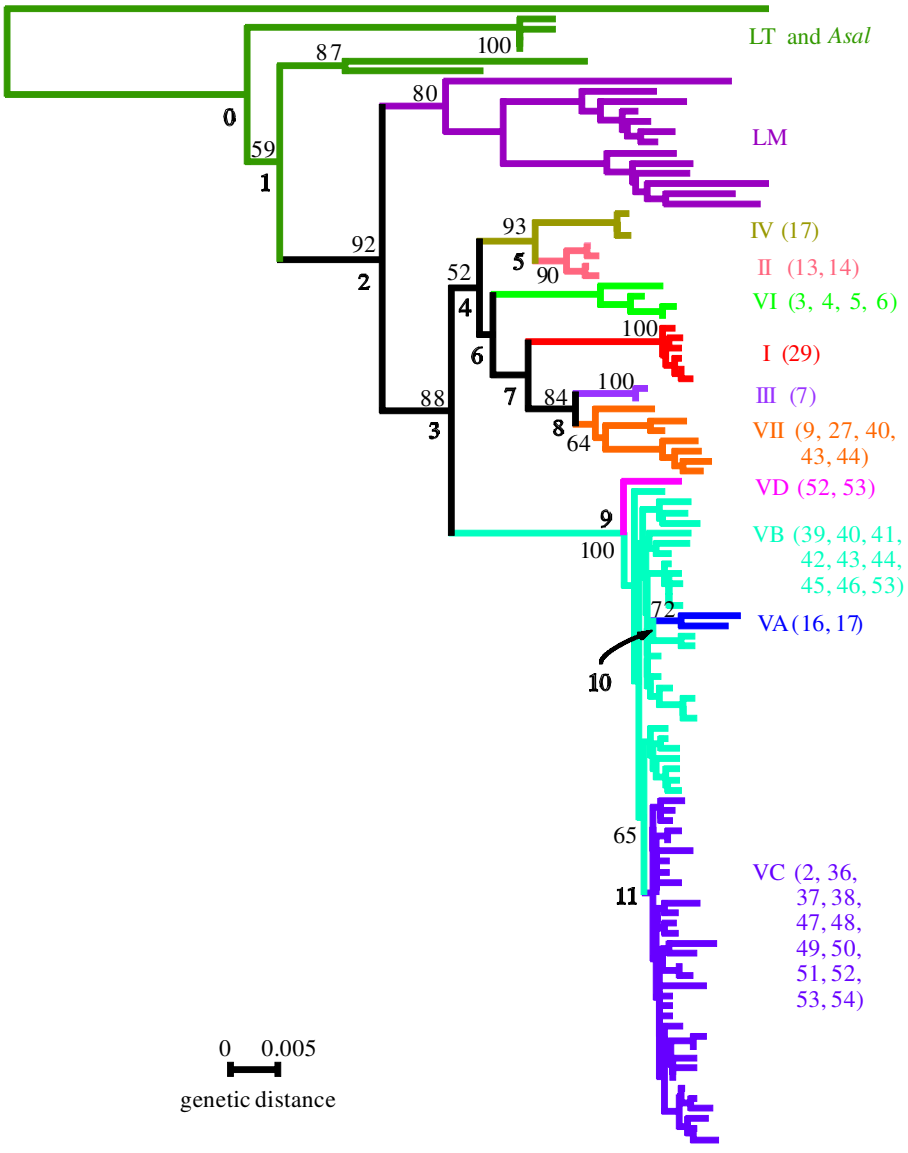


Figure 3. Neighbour-joining tree of East African haplochromine fishes based on mtDNA control region sequences. Haplogroups are designated I–VII; *Asal*, *Astatoreochromis alluaudi*; LM, Lake Malawi; LT, Lake Tanganyika. The localities in parentheses correspond to those in figure 1. Important branch nodes are numbered 1–11 by outlined numbers; solid numbers indicate the percentage of recovery in 1000 bootstrap replications.

an indication for the presence in LV of haplochromines bearing control region haplotypes that otherwise occur mostly in the LER. Another indication that LV haplochromines are not molecularly as uniform as originally believed is the presence of the subgroup VD sequences among the 'mbipi'-like fishes.

The divergence times of these three subgroups (VB, VC, VD; table 2) are clearly in excess of 12 400 years, so the subgroups must have existed as distinct lineages well before the colonization of the modern LV. In this context, an important point to emphasize is the absence of subgroup VC fishes in the LER. The reason why this is important is that the lakes in this region have often been considered as having been the place in which many species arose before moving to LV following its reappearance 12 400 years ago (Booton *et al.* 1999). While some species now found in LV (those of the subgroup VB) may have indeed come from the LER (although the reverse movement cannot be excluded), the absence of subgroup VC haplotypes in the LER and their predominance in LV renders this scenario for the origin of the LV haplochromine flock highly improbable. In fact, it can be argued that the ancestors of the LV subgroup VC flock did not pass through the LER. All the LER haplochromines tested have a T at site 271 of the control region sequence, whereas all the LV flock individuals have a C at this site. Since the Lake Rukwa region haplochromines also have a T at this site, the substitution apparently became fixed before the LER and Lake Rukwa haplochromines diverged from each other and before the former began to radiate (once reproductive species barriers were established, fixation of the same mutation in all of the numerous species of the flock became highly unlikely). For the same reason, the C at site 271 must have been fixed before the ancestors of the LV flock began to radiate. The radiations of the LER and the LV flocks must have therefore occurred independently. Similar arguments can be put forward for the polymorphism at site 635, at which most LER haplochromines have an A, whereas the overwhelming majority of the LV flock species have a G. Consistent with this interpretation is the considerably shorter separation time (50 000–100 000 years ago) of the Lake Rukwa region from the LER haplochromines in comparison with the longer separation time (160 000–310 000 years ago) of the LER from the LV flock haplochromines (table 2). These data do not

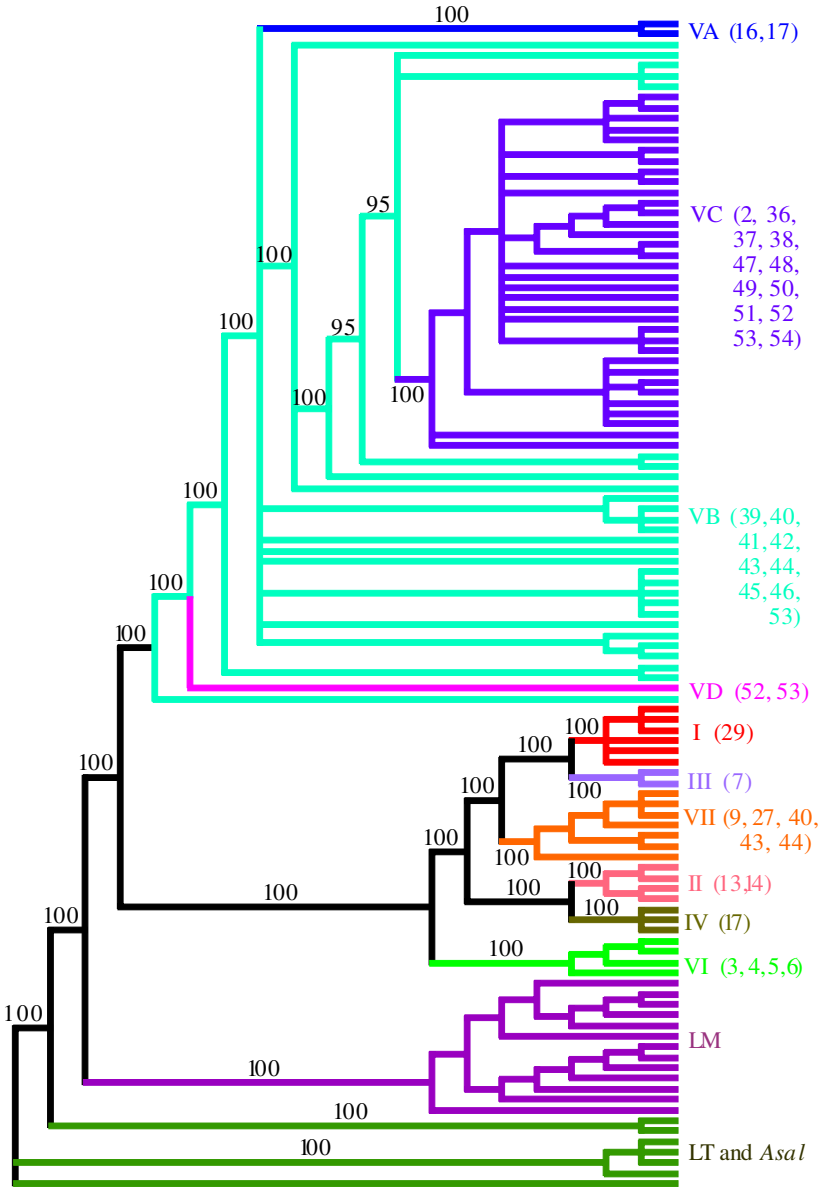


Figure 4. Majority consensus maximum-parsimony tree of East African haplochromine fishes based on mtDNA control region sequences. Haplogroups are designated I–VII; *Asal*, *Astatoreochromis alluaudi*; LM, Lake Malawi; LT, Lake Tanganyika. The localities in parentheses correspond to those in figure 1. The frequency of particular branching topologies is indicated by numbers.

indicate how the separation of the Lake Rukwa region and LER haplochromines occurred, whether the ancestors of the former migrated out of the LER to establish the Lake Rukwa population or the ancestors of the latter migrated north to found the LER flock; the divergence may have also taken place in some other region. Whatever the case might have been, it is important that at the time of the divergence the substitutional differences between the subgroups VA and VB on the one hand and VC on the other were already established.

There is, however, an apparent contradiction in the above interpretation of the data. If the VC-bearing LV species flock was really founded by a single lineage, then one would not expect to find the VC-type species of a given genus in LV and the VB-type species of the same genus in the LER (table 1). The contradiction can be resolved in one of two ways. Either one or the other of the two species has been misidentified, or the two species do not really belong to the same genus and the similarities between them are the result of evolutionary convergence. If the former explanation were to be invoked, one would still have to explain why molecularly distinct species, apparently derived from distinct lineages, resemble each other morphologically to the extent that they can be held for belonging to the same genus. It is therefore more parsimonious to invoke the second explanation. Morphological convergence of forms has been shown to be a common occurrence among Lake Tanganyika lineages and between Lake Tanganyika and Lake Malawi cichlids (Kocher *et al*. 1993; Rüber *et al*. 1999).

If the radiation of the LV flock did not begin in the LER and if, as the geological evidence indicates (Johnson *et al*. 1996), the desiccation of the palaeolake in the LV basin was complete 14 000 years ago, then the origin of

the LV flock founding population must be sought in the rivers that persisted in this region or in other East African river systems. Indeed, group V appears to be the most widely distributed of the seven riverine haplogroups. Since, however, no explosive morphological radiation of haplochromines has been reported for any of the East African river systems, such radiation being apparently conditional upon the colonization of a large lake, it must be concluded that the LV flock originated *in situ* from a non-specialized riverine founding population that colonized the forming LV some 12 400 years ago. The 300 or so species which apparently comprise the LV flock must have therefore arisen within this short period of time.

It is inconceivable, however, that all the mutations responsible for the morphological adaptations differentiating the flock's individual species arose within a mere 12 400 years. Three important conclusions must, therefore, be drawn. First, the mutations must have already been present in the founding population as polymorphisms, but in combinations which allowed the maintenance of the generalist (non-specialist) phenotype characteristic of riverine haplochromines (Greenwood 1979). Only in the lake, under the selection pressure to produce phenotypes adapted to specific ecological niches, were the mutations sorted out into new combinations, allowing the appearance of the various specialized phenotypes in the different niches. Second, to furnish all the necessary polymorphisms (it can reasonably be assumed that most of the morphological characters differentiating the flock's species are polygenetically controlled), the flock's founding population must have been large. These two conclusions are supported by the demonstration that nuclear gene polymorphisms are widely shared not only among the various species of the LV flock but also between the LV flock and the riverine species (Nagl *et al.* 1998). Third, because of the shortness of the time interval since the beginning of the radiation, the species of the LV flock cannot be expected to be fixed for any mutations apart from the ones responsible for the morphological differences among them. Unless the latter are identified, it may therefore be futile to search for molecular markers that could decipher the phylogeny of the flock.

In summary, it is proposed that the ancestors of the LV flock were trophic generalists which lived in the East African river systems in reproductive isolation from other haplochromine populations for at least 1.4 Myr. From this population, a series of subpopulations separated in close succession 100 000–200 000 years ago. One of these subpopulations founded the VD lineage of LV haplochromines; the other the VC subgroup of the LV flock; and the third most of the LER species. The founders of the LV flock entered the forming lake *ca.* 12 400 years ago and began to radiate by adapting to the various ecological niches arising in the lake. The rapid radiation was possible because most of the mutations necessary for the morphological and behavioural adaptations to these niches were already present as polymorphisms in the gene pool of the large founding population. The mutations were sorted out into distinct combinations and fixed by natural selection. Although the specifics of this adaptive radiation may have been characteristic of the haplochromines, the principles may apply to adaptive radiation in general.

We thank Ms Waltraud Sinz and Ms Ljubica Sanader for technical assistance; Ms Jane Kraushaar for editorial assistance; Mr Erwin Schraml and Dr Iva Dyková for help in sample collecting during the expeditions; members of Haplochromis Ecology Survey Team, Department of Morphology, University of Leiden, The Netherlands, for help in identification of specimens; and the Uganda National Council for Science and Technology, the Tanzania Commission for Science and Technology, Ministry of Tourism, Natural Resources and Environment, Tanzania, the Kenya Ministry of Tourism and Wildlife, Nairobi, and Kenya Fisheries Research Institute, Kisumu, and its crew for help and support. Special thanks for help to Mr Sigi Engelhardt, Kikambala.

REFERENCES

Avise, J. C. & Johns, G. C. 1999 Proposal for a standardized temporal scheme of biological classification for extant species. *Proc. Natl Acad. Sci. USA* **96**, 7358–7363.

Booton, G. C., Kaufman, L., Chandler, M., Oguto-Ohwayo, R., Duan, W. & Fuerst, P. A. 1999 Evolution of the ribosomal RNA internal transcribed spacer one (ITS-1) in cichlid fishes of the Lake Victoria region. *Mol. Phylogenet. Evol.* **11**, 273–282.

Bromage, T. G., Schrenk, F. & Juwayeyi, Y. M. 1995 Paleobiogeography of the Malawi Rift: age and vertebrate paleontology of the Chiwondo Beds, Northern Malawi. *J. Hum. Evol.* **28**, 37–57.

Cantatore, P., Roberti, M., Pesole, G., Ludovico, A., Milella, F., Gadaleta, M. N. & Saccone, C. 1994 Evolutionary analysis of cytochrome *b* sequences in some Perciformes: evidence for a slower rate of evolution than in mammals. *J. Mol. Evol.* **39**, 589–597.

Crapon de Caprona, M.-D. & Fritzsch, B. 1984 Interspecific fertile hybrids of haplochromine Cichlidae (Teleostei) and their possible importance for speciation. *Neth. J. Zool.* **34**, 503–538.

Daget, J., Gosse, J.-P., Teugels, G. G. & Thys van den Audenaerde, D. F. E. 1991 *Check-list of the freshwater fishes of Africa*, vol. 4. Brussels: Institut Royal des Sciences Naturelles de Belgique.

Donaldson, K. A. & Wilson, R. R. 1999 Amphi-panamic 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.

Eccles, D. H. & Trewavas, E. 1989 *Malawian cichlid fishes. The classification of some haplochromine genera.* Herten, Germany: Lake Fish Movies.

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

Fitch, W. M. 1971 Towards defining the course of evolution: minimum change for a specific tree topology. *Syst. Zool.* **20**, 406–416.

Fryer, G. & Iles, T. D. 1972 *The cichlid fishes of the Great Lakes of Africa. Their biology and evolution.* Edinburgh, UK: Oliver Boyd.

Greenwood, P. H. 1956 The monotypic genera of cichlid fishes in Lake Victoria. *Bull. Br. Mus. Nat. Hist.* (*Zool.*) **3**, 295–333.

Greenwood, P. H. 1965 The cichlid fishes of Lake Nabugabo, Uganda. *Bull. Br. Mus. Nat. Hist.* (*Zool.*) **12**, 315–357.

Greenwood, P. H. 1979 Towards a phyletic classification of the 'genus' *Haplochromis* (Pisces, Cichlidae) and related taxa. Part I. *Bull. Br. Mus. Nat. Hist.* (*Zool.*) **35**, 265–322.

Greenwood, P. H. 1980 Towards a phyletic classification of the 'genus' *Haplochromis* (Pisces, Cichlidae) and related taxa. Part II. *Bull. Br. Mus. Nat. Hist.* (*Zool.*) **39**, 1–101.

Greenwood, P. H. 1981 *The haplochromine fishes of the East African Lakes.* Ithaca, NY: Cornell University Press.

Horai, S., Hayasaka, K., Kondo, R., Tsugane, K. & Takahata, N. 1995 Recent African origin of modern humans revealed by complete sequences of hominoid mitochondrial DNAs. *Proc. Natl Acad. Sci. USA* **92**, 532–536.

Jin, L. & Nei, M. 1990 Limitations of the evolutionary parsimony method of phylogenetic analysis. *Mol. Biol. Evol.* **7**, 82–102.

Johnson, T. C., Scholz, C. A., Talbot, M. R., Kelts, K., Ricketts, R. D., Ngobi, G., Beuning, K., Ssemmanda, I. & McGill, J. W. 1996 Late Pleistocene desiccation of Lake Victoria and rapid evolution of cichlid fishes. *Science* **273**, 1091–1093.

Keenleyside, M. H. A. (ed.) 1991 *Cichlid fishes. Behavior, ecology and evolution*. London: Chapman & Hall.

Kimura, M. 1980 A simple method for estimating evolutionary rates of base substitutions through comparative studies of nucleotide sequences. *J. Mol. Evol.* **16**, 111–120.

Klicka, J. & Zink, R. M. 1999 Pleistocene effects on North American songbird evolution. *Proc. R. Soc. Lond.* B **266**, 695–700.

Kocher, T. D. & Wilson, A. C. 1991 Sequence evolution of mitochondrial DNA in humans and chimpanzees: control region and a protein-coding region. In *Evolution of life* (ed. S. Osawa & T. Honjo), pp. 391–413. New York: Springer.

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

Lee, W. J., Conroy, J., Howell, W. H. & Kocher, T. D. 1995 Structure and evolution of teleost mitochondrial control regions. *J. Mol. Evol.* **41**, 54–66.

Lippitsch, E. 1993 A phyletic study on lacustrine haplochromine fishes (Perciformes, Cichlidae) of East Africa, based on scale and squamation characters. *J. Fish Biol.* **42**, 903–946.

Lippitsch, E. 1995 Scale and squamation character polarity and phyletic assessment in the family Cichlidae. *J. Fish Biol.* **47**, 91–106.

Martin, A. P., Naylor, G. J. P. & Palumbi, S. R. 1992 Rates of mitochondrial DNA in sharks are slow compared with mammals. *Nature* **357**, 153–155.

Mayer, W. E., Tichy, H. & Klein, J. 1998 Phylogeny of African cichlid fishes as revealed by molecular markers. *Heredity* **80**, 702–714.

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

Nagl, S., Tichy, H., Mayer, W. E., Takahata, N. & Klein, J. 1998 Persistence of neutral polymorphisms in Lake Victoria cichlid fish. *Proc. Natl Acad. Sci. USA* **95**, 14 238–14 243.

Pickford, M., Senut, B. & Hadoto, D. 1993 *Geology and palaeobiology of the Albertine Rift Valley Uganda–Zaire. 1. Geology*. Orléans, France: Centre International pour la Formation et les Echanges Geologiques.

Rüber, L., Verheyen, E. & Meyer, A. 1999 Replicated evolution of trophic specializations in an endemic cichlid fish lineage from Lake Tanganyika. *Proc. Natl Acad. Sci. USA* **96**, 10 230–10 235.

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

Schlüter, T. 1997 *Geology of East Africa*. Berlin: Gebrüder Borntraeger.

Seehausen, O. 1996 *Lake Victoria rock cichlids. Taxonomy, ecology, and distribution*. Zevenhuizen, The Netherlands: Verduyn Cichlids.

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

Swofford, D. L. 1993 *PAUP: phylogenetic analysis using parsimony*. Champaign, IL: Illinois Natural History Survey.

Takezaki, N., Rzhetsky, A. & Nei, M. 1995 Phylogenetic test of the molecular clock and linearized trees. *Mol. Biol. Evol.* **12**, 823–833.

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

Trewavas, E. 1933 Scientific results of the Cambridge expedition to the East Africa lakes, 1930–1. II. The cichlid fishes. *J. Linn. Soc. (Zool.)* **38**, 309–341.

Van Oijen, M. J. P. 1996 The generic classification of the haplochromine cichlids of Lake Victoria, East Africa. *Zool. Verh. Leiden* **302**, 57–110.

Yang, Z. 1994 Maximum likelihood phylogenetic estimation from DNA sequences with variable rates over sites: approximate methods. *J. Mol. Evol.* **39**, 306–314.

Yang, Z. 1998 PAML, v. 1.3c. London: University College London.

As this paper exceeds the maximum length normally permitted, the authors have agreed to contribute to production costs.

An electronic appendix to this paper can be found at (http://www.pubs.royalsoc.ac.uk/publish/pro_bs/rpb1447.htm).