THE ROYAL SOCIETY

Received 25 March 2002
Accepted 6 August 2002
Published online 9 December 2002

Nuclear markers reveal unexpected genetic variation and a Congolese–Nilotic origin of the Lake Victoria cichlid species flock

Ole Seehausen[1,2*], Egbert Koetsier[2], Maria Victoria Schneider[2], Lauren J. Chapman[3,4], Colin A. Chapman[3,4], Mairi E. Knight[1], George F. Turner[1], Jacques J. M. van Alphen[2] and Roger Bills[5]

[1]*Molecular and Evolutionary Ecology Group, Department of Biological Sciences, University of Hull, Hull HU6 7RX, UK*
[2]*Institute of Evolutionary and Ecological Sciences, University of Leiden, PO Box 9516, 2300 RA Leiden, The Netherlands*
[3]*Department of Zoology, University of Florida, Gainesville, FL 32611, USA*
[4]*Department of Zoology, Makerere University, Kampala, Uganda*
[5]*South African Institute for Aquatic Biodiversity, Somerset Street, Private Bag 1015, Grahamstown 6140, South Africa*

Phylogenetic analyses based on mitochondrial (mt) DNA have indicated that the cichlid species flock of the Lake Victoria region is derived from a single ancestral species found in East African rivers, closely related to the ancestor of the Lake Malawi cichlid species flock. The Lake Victoria flock contains ten times less mtDNA variation than the Lake Malawi radiation, consistent with current estimates of the ages of the lakes. We present results of a phylogenetic investigation using nuclear (amplified fragment length polymorphism) markers and a wider coverage of riverine haplochromines. We demonstrate that the Lake Victoria–Edward flock is derived from the morphologically and ecologically diverse cichlid genus *Thoracochromis* from the Congo and Nile, rather than from the phenotypically conservative East African *Astatotilapia*. This implies that the ability to express much of the morphological diversity found in the species flock may by far pre-date the origin of the flock. Our data indicate that the nuclear diversity of the Lake Victoria–Edward species flock is similar to that of the Lake Malawi flock, indicating that the genetic diversity is considerably older than the 15 000 years that have passed since the lake began to refill. Most of this variation is manifested in trans-species polymorphisms, indicating very recent cladogenesis from a genetically very diverse founder stock. Our data do not confirm strict monophyly of either of the species flocks, but raise the possibility that these flocks have arisen from hybrid swarms.

Keywords: adaptive radiation; amplified fragment length polymorphisms; cichlid fish; explosive speciation; hybridization; Lake Victoria

1. INTRODUCTION

The 'species flocks' of endemic haplochromine cichlid fishes in Lakes Victoria and Malawi are the largest known recent adaptive radiations of animals and are each believed to have evolved from a single founder species in a strikingly short time (Meyer *et al.* 1990; Meyer 1993). Molecular and geological evidence agree that colonization occurred in Lake Malawi between 500 000 and 2 million years before the present (BP) (Meyer *et al.* 1990; Meyer 1993; Johnson *et al.* 1996, 2000; Sturmbauer *et al.* 2001). Geological and palaeolimnological evidence indicates that the Lake Victoria basin is *ca*. 400 000 years old but that the lake was completely dry for several thousand years, and re-filled only 15 000 years BP (Johnson *et al.* 1996, 2000; Talbot & Laerdal 2000). Extremely low neutral genetic variation reported within the Lake Victoria cichlid species flock (Meyer *et al.* 1990; Nagl *et al.* 2000; Farias *et al.* 2001) makes it hard to see how selection can have produced the large functional diversity observed in this flock (Fryer & Iles 1972; Greenwood 1974) from a single ancestral population in a period of time that translates into between 30 000 (for Lake Victoria) and 400 000 (for the entire Lake Victoria–Edward region) cichlid generations (Fryer 2001; Seehausen 2002). Two major criticisms may be levelled at the current picture:

(i) many of the riverine cichlid species that might be good candidates as sister taxa for all or part of the lacustrine radiations are absent from existing molecular phylogenies;
(ii) the evidence for monophyly and lack of genetic variation comes entirely from studies of mitochrondrial (mt) DNA.

Maternally inherited, non-recombining mtDNA is prone to give a false impression of monophyly when sharing of haplotypes is the result of fixation following introgression among independently evolved taxa (Tegelstrom 1987; DeMarais *et al.* 1992; Bernatchez *et al.* 1995; Arnold 1997; Taylor & McPhail 2000; Lu *et al.* 2001;

* Author for correspondence (o.seehausen@hull.ac.uk).

Parson & Shaw 2001; Rueber *et al.* 2001; Yamada *et al.* 2001). Postzygotic isolation between haplochromine species is weak. Intergeneric hybrids are viable and fertile among Lake Malawi species (McElroy & Kornfield 1993; Stauffer *et al.* 1996; M. E. Knight, unpublished data) and among Lake Victoria species (Seehausen *et al.* 1997), as well as between riverine and Lake Victoria species (Crapon de Caprona & Fritzsch 1984). Haplochromines of Lake Victoria hybridize in nature when visual mate choice is constrained by water turbidity (Seehausen *et al.* 1997), as would probably have occurred when rivers backponded to flood nutrient-rich grassland as the lake filled. Mitochondrial DNA haplotype sharing between phenotypically distinct species of Lake Malawi cichlids (Moran & Kornfield 1993) may be due to insufficient time for lineage sorting between speciation events, but hybridization is an alternative explanation. Introgression of mtDNA has now been demonstrated between distinct species and even genera of cichlids from Lake Tanganyika (Rueber *et al.* 2001; Salzburger *et al.* 2002). If introgression had occurred between more than one founding species at the root of an adaptive radiation this might explain the rapid origin of functional genetic variation expressed in the many different adaptive forms in the current species flocks. This challenge to the mtDNA-derived view of the explosive origin of African cichlid species flocks from single ancestors has been largely overlooked.

We used nuclear DNA to test the mtDNA-based hypotheses of (i) single ancestry for the Lake Victoria and Malawi species flocks, and (ii) lack of genetic variation among Victorian haplochromine cichlids. We performed a genome-wide screening for polymorphic sites with amplified fragment length polymorphisms (AFLPs; Vos *et al.* 1995; Mueller & Wolfenbarger 1999), a rich source of polymorphic markers powerful in resolving phylogenetic relationships both among closely related taxa (Albertson *et al.* 1999; Giannasi *et al.* 2001; Parson & Shaw 2001; Schneider *et al.* 2002) and among lineages that separated several million years ago (Albertson *et al.* 1999; Ogden & Thorpe 2002).

2. MATERIAL AND METHODS

(a) *Taxonomic sampling*

We sampled haplochromine cichlids from each major drainage system adjacent to, or historically believed to have been in contact with the Lake Victoria region: Nile, Congo, eastflowing basins draining into the Indian Ocean and the Lake Tanganyika basin. We also collected samples from Lake Malawi and the nearby Lake Chilwa and from the Zambezi River. The lake species flocks were phylogenetically broadly sampled, such that all major intralacustrine lineages were represented (table 1). Three species of endemic Lake Victoria cichlid species that were no longer found in nature were sampled from aquarium stock maintained at the University of Leiden (table 1).

(b) *Molecular techniques*

Genomic DNA was digested with the restriction enzymes *Mse* I (1 unit) and *Eco* RI (5 units; New England Biolaboratories). Polymerase chain reaction (PCR) adaptors of 16 bp in length specific to the cutting sites were ligated to the ends of the restriction fragments (Vos *et al.* (1995); overnight at a temperature of 37 °C). We performed a preselective amplification with one selective base on each primer (*Mse* I-C and *Eco* RI-A) and five different selective amplifications, using five different combinations of primers with an additional two-base extension (M-CTA + E-AAC, M-CAT + E-ACA, M-CAG + E-ACA, M-CAC + E-AAC, M-CAC + E-ACA). PCR was performed on a BIL-CO PCR thermal cycler. DNA quality and concentration were checked after restriction–ligation, preselective and selective amplification on agarose gels. Visualization of the selectively amplified and fluorescent dye labelled fragments was carried out on an ABI Prism 377 automated sequencer. Internal size standards (GeneScan Rox) were run in every lane. Reproducibility was tested by repeating the process from restriction to fragment scoring for two fishes with each of three primer combinations. Average band sharing between products of the same accession was 95% ± 5%. Products of the same accession were always resolved as sister taxa in neighbour-joining trees produced from matrices containing the full dataset.

(c) *Data preparation*

Gel images were scanned and electropherograms aligned with Applied Biosystems GENESCAN v. 3.1 software, using the internal size standards. Aligned data were imported into Applied Biosystems GENOTYPER v. 2.1 software. Standardized electropherograms were scored for the presence or absence of fragments of between 50 and 500 bases in size. Overlapping fragment size categories were excluded. Three minimum signal intensities (300, 600, 1500 r.f.u.) were used as thresholds in defining fragments and a subsequent presence/absence scoring threshold of 50 r.f.u.

(d) *Genetic and phylogenetic analysis*

The raw data matrix was subjected to a series of tests for adequacy of phylogenetic signal using PAUP* v. 4b8 (Swofford 2001) by plotting the distribution of 1 000 000 random trees, with calculation of the skewness parameter, g_1 (Hillis & Huelsenbeck 1992). Pairwise genetic distances were then calculated from the binary matrices using the algorithm of Link *et al.* (1995) in TREECON v. 1.3b (Van de Peer & De Wachter 1994). This distance is based on shared and unique characters and ignores shared absence. Related distance algorithms that were also used were Nei & Li (1979; implemented in TREECON v. 1.3b) and the methods implemented by the program RESTDIST (Felsenstein 2001).

We applied five tree-building algorithms to the distance matrices. Two of these do not assume a molecular clock: Fitch & Margoliash (1967) with unconstrained branch length (implemented in FITCH in PHYLIP v. 3.6a2; carried out with global rearrangements and 10 different random input orders of taxa); and neighbour-joining (Saitou & Nei (1987); implemented in PHYLIP v. 3.6a2 and TREECON v. 1.3b). The three others assume a molecular clock: Fitch & Margoliash (1967) with constrained branch-length; minimum evolution (Kidd & Sgaramella-Zonta 1971; Rzhetsky & Nei 1992) both implemented in KITSCH in PHYLIP v. 3.6a2 with global rearrangements and carried out with 100 different random input orders; and UPGMA (Sneath & Sokal (1973); implemented in PHYLIP v. 3.6a2 and TREECON v. 1.3b). We tested the assumption of a molecular clock by comparing the ratio of the difference in the mean sum of squares obtained by KITSCH and by FITCH over the sum of squares obtained by FITCH to the *F*-distribution (Felsenstein 1984). The test indicated that the assumption of a molecular clock has to be rejected ($F = 9.24$, 36 and 630 d.f.). Because the assumption of the clock test, that distances are truly independent, is of

Table 1. The samples of cichlids used in the AFLP analysis and their geographical origin.
(Species, collection site, geographical distribution. For some species, two individuals were accessioned.)

East African

Astatoreochromis alluaudi LV, Lake Victoria, Lake Victoria drainage[a]; *A. alluaudi* LE, Lake Saka, Lake Edward–George drainage[a]; *Astatotilapia calliptera*, Lake Malawi, *As. calliptera*, Zambezi River, Lake Malawi–Zambezi–Limpopo; *As.* sp. cf. *calliptera*, Lake Chilwa, Eastflowing Rovuma drainage; *As. bloyeti*, Lake Manyara, eastflowing rivers of Tanzania: from Pangani to Rufiji; *As. burtoni,* Lake Tanganyika, Malagarasi–Lake Tanganyika

Nilotic–Congolese

Thoracochromis brauschii, Lac Fwa, Central Congo; *T. pharyngalis*, Lake Edward, Lake Edward[b]; *T. petronius*, Kazinga Channel, Lake George[b]; *As. flavijosephi*, Lake Kinneret, Israel–Syria and possibly Egypt

Lake Victoria–Edward endemics

Platytaeniodus degeni (2)[c]; *Ptyochromis sauvagei*; *Labrochromis ishmaeli* (2)[c]; *Harpagochromis howesi* complex; *Yssichromis pyrrhocephalus*; *Y. piceatus*[c]; *Neochromis omnicaeruleus* (2); Mbipia mbipi (2); *Pundamilia nyererei* (2); all endemic to Lake Victoria; *'Haplochromis' elegans*, Kazinga Channel, Lake Edward–George

Lake Malawi endemics

Rhamphochromis esox; *Copadichromis virginalis*; *Diplotaxodon limnothrissa*; *Taeniolethrinops laticeps*; **Mbuna**: *Pseudotropheus tropheops* 'mauve' Mara Rocks; *P. tropheops* 'mauve' Nkhata Bay; *P. tropheops* 'olive'; *P. (Metriaclima) livingstonii*; *P. (Metriaclima) zebra* Nkhata Bay (2); *P. (Metriaclima) fainzilberi* Mara Rocks; *P. (Metriaclima) fainzilberi* Ruarwe

[a] This species does not belong to the endemic Lake Victoria–Edward species flock but to an older lineage restricted to the Lake Victoria–Edward and Lake Tanganyika regions (Greenwood 1979; Meyer *et al.* 1990).
[b] These species do not belong to the endemic Lake Victoria–Edward species flock but to a lineage widely distributed in the upper Nile (Greenwood 1979).
[c] These species were no longer found in nature; samples comprised laboratory stock maintained at the University of Leiden.

course violated, we performed phylogenetic analysis with and without the clock assumption.

We also carried out a maximum-parsimony analysis on the discrete character matrix (Wagner parsimony (Kluge & Farris 1969; Swofford & Maddison 1987)) by using the MIX program in PHYLIP v. 3.6a2 and MACCLADE v. 3.0 (Maddison & Maddison 1992). To test the observed tree against phylogenetic null hypotheses, we constructed constraint trees in MACCLADE and compared them with the observed tree. We used a parsimony analogue of the Kishino & Hasegawa (1989) test, in which the test statistic was the sum of the differences in the number of steps each site required on the two trees. The null hypothesis was that this sum is zero (Page & Holmes 1998). All trees were rooted on *Astatoreochromis alluaudi*, the outgroup position of that is well established (Meyer *et al.* 1990; Nagl *et al.* 2000).

To compare the genetic variation in the Lake Victoria species flock with the genetic variation in the Lake Malawi species flock, we compared pairwise genetic distances within each flock with a *t*-test, with one mean distance for each sample.

3. RESULTS

(a) *Genetic variation and phylogenetic signal*

Two hundred and forty-six, 192, and 134 polymorphic sites were obtained for the peak height thresholds of 300, 600 and 1500 r.f.u., respectively, and between 33 and 74 polymorphic sites per primer pair at 300 r.f.u. The skewness parameter g_1 revealed significant phylogenetic signal in all parts of the tree (table 2). It was stronger in the dataset scored at a threshold of 300 r.f.u., used for all further analyses.

Genetic distances among the samples from Lake Victoria (mean ± s.d.: 0.40 ± 0.03) and Lake Malawi (0.39 ± 0.03) did not differ significantly ($t = 1.71$, 23 d.f., $p = 0.17$). The ratio of maximum genetic distance within genera (or species) to between genera (or species) was higher for endemic Victorian taxa (genera = 0.82, species = 0.82) than Malawian endemics (genera = 0.70, species = 0.68). This indicates more recent cladogenesis in Lake Victoria, consistent with the short branches in the Lake Victoria phylogeny (figure 1).

(b) *Phylogenetic analyses*

All tree-building methods yielded very similar results both in terms of topology and distribution of internal branch lengths (figure 1). Invariably, there was a deep split into an East African and a Central–North African clade. The East African clade included all East African *Astatotilapia* species and all samples from Lake Malawi but none of the samples from Lake Victoria. The topology within this clade was very similar to that recovered in mtDNA studies (figure 1 (Meyer 1993; Shaw *et al.* 2000); see table 1 for which species constitute the Mbuna). However, whereas mtDNA placed a *Rhamphochromis–Diplotaxodon* clade as basal and *Copadichromis virginalis* closer to phenotypically similar species deeper in the species flock (Shaw *et al.* 2000), *C. virginalis* was resolved by our analyses as basal to the rest of the endemic Lake Malawi cichlids (figure 1*b–d*). The Fitch–Margoliash clock and Wagner parsimony placed *C. virginalis* even basal to a clade that comprised the other Lake Malawi cichlids and some riverine *Astatotilapia* species, breaking up the monophyly of Lake Malawi endemics (figure 1*c*). All methods resolved *Astatotilapia calliptera* as being polyphyletic. The population sampled from within Lake Malawi appeared to be more closely related to endemic Lake Malawi cichlids than to conspecific populations from the Zambezi system and Lake Chilwa (figure 1*b–d*).

The basal position within the Central and North African clade was invariably occupied by the North Nilotic *Astatotilapia flavijosephi*. Other taxa included the Congolese *Thoracochromis brauschii*, the South Nilotic *Thoracochromis pharyngalis* and *Thoracochromis petronius* and all endemic taxa from Lakes Victoria and Edward. Invariably, the genus *Thoracochromis* was resolved as a paraphyletic sister taxon to the haplochromines of Lake Victoria and

Table 2. Skewness parameter g_1 estimated from 1 000 000 random trees.
(300, 600, 1500 refer to the scoring threshold; LM, Lake Malawi endemics; LV, Lake Victoria endemics.)

dataset test	*n* taxa	*n* char[a]	g_1	*p*
entire 300 matrix	38	203/246	–0.35	<0.0001
entire 600 matrix	38	164/192	–0.34	<0.0001
entire 1500 matrix	38	123/134	–0.30	<0.0001
300 without *A. alluaudi*	36	198/239	–0.34	<0.0001
300 without *A. alluaudi* and LM	23	141/205	–0.34	<0.001
300 LV and Mbuna each represented by a single taxon	18	148/209	–0.51	<0.001
300 without *A. alluaudi* and LM; LV represented by single taxon	10	84/163	–0.55	<0.01
300 Lake Malawi only	13	87/155	–0.33	<0.01
300 LV and Lake Edward only	15	88/141	–0.19	0.01
300 Lake Victoria only	14	85/138	–0.19	0.01
300 Mbuna only	8	53/107	–0.37	0.05

[a] Parsimony informative sites/all variable sites.

Edward (figure 1*b–d*). *'Haplochromis' elegans* from Lake Edward was invariably resolved as part of the Lake Victoria radiation, although the Fitch–Margoliash clock algorithm resolved it together with *Harpagochromis* from Lake Victoria as basal in the radiation. Support for monophyly of the combined Lake Victoria–Edward radiation was at best weak. The Fitch–Margoliash clock placed the Lake Victoria endemic *Yssichromis pyrrhocephalus* outside the flock (figure 1*c*) and Minimum Evolution placed the two South-Nilotic *Thoracochromis* species within the flock (figure 1*d*). On all trees the internal branches at the base of the Lake Victoria radiation were much shorter than those at the base of the Lake Malawi radiation.

Distance-based methods without clock resolved the deeper nodes with better bootstrap support (figure 1*b* versus figure 1*c*,*d*). Trees constructed from matrices of other genetic distances (Nei & Li 1979; Felsenstein 2001) were virtually identical to those obtained from the matrix of the Link *et al.* (1995) distances. A Wagner parsimony tree built from the discrete character matrix was identical to the tree obtained by Fitch–Margoliash without clock except that all East African riverine *Astatotilapia* species (*bloyeti, burtoni, calliptera,* sp. 'Lake Chilwa') were resolved between *C. virginalis* and other Lake Malawi endemics (not shown).

We also constructed trees rooted on Lake Malawi cichlids, omitting *Astatoreochromis* from the analysis. It made no difference to the topology and did not affect our finding that the Lake Victoria–Edward haplochromines are nested within the Nilotic–Congolese genus *Thoracochromis*.

Four phylogenetic null hypotheses were formulated and resultant constrained trees compared with the observed most parsimonious tree.

(i) The observed most parsimonious tree required significantly fewer steps per site (Δ) than the most parsimonious constraint trees for the hypothesis that the clade *Astatotilapia bloyeti–Astatotilapia burtoni* is the sister taxon to the endemic cichlids of Lakes Victoria and Edward ($\Delta = -0.017$, s.e. 0.016).
(ii) We were similarly able to reject the hypothesis that *As. calliptera* is a monophyletic or paraphyletic species basal to the Lake Malawi flock ($\Delta = -0.041$, s.e. 0.017). The same test on trees obtained by distance methods (without as well as with clock) also rejected hypotheses (i) and (ii).
(iii) Monophyly of the Lake Malawi species flock was also rejected by the parsimony test ($\Delta = -0.021$, s.e. 0.017), but not by tests on distance-derived trees.
(iv) The hypothesis of monophyly of the Lake Victoria–Edward species flock could not be rejected.

4. DISCUSSION

(a) *Independent origin of lacustrine species flocks*

Our study of 246 polymorphic loci confirms the results of mtDNA sequence analysis that the species flocks in Lakes Victoria and Malawi have evolved independently. Hence, the extensive parallelisms in morphology and ecology (Fryer & Iles 1972; Greenwood 1974) probably reflect true parallel evolution, similar to that between the cichlids of Lake Tanganyika and Lake Malawi (Kocher *et al.* 1993).

(b) *Congolese–Nilotic ancestry of the Lake Victoria species flock*

By contrast to all previous studies of which we are aware, our data indicate that the cichlid species flock of Lake Victoria is derived from the phenotypically diverse Congolese–Nilotic genus *Thoracochromis* and not from the more conservative East African riverine *Astatotilapia.* Sequence variation in mtDNA had indicated a clade including *As. bloyeti* and *As. burtoni* as the sister taxon to the Lake Victoria radiation (Meyer *et al.* 1990; Meyer 1993; Nagl *et al.* 2000). We found that all sampled species of Nilotic or Congolese haplochromines form a paraphyletic group that is ancestral to the species flock of the Lake Victoria region, whereas all the East African *Astatotilapia* species form a paraphyletic group that is ancestral to the species flock of Lake Malawi.

This result is not entirely surprising if the geology of the region is considered. The rivers feeding Lake Victoria were tributaries of the Congo until the region to the West of it became elevated *ca.* 400 000 years BP (Cooke 1958; Fryer & Iles 1972; Johnson *et al.* 2000). It is possible that the shared ancestry with the Congolese branch of *Thoracochromis* predates this event. Subsequent to the

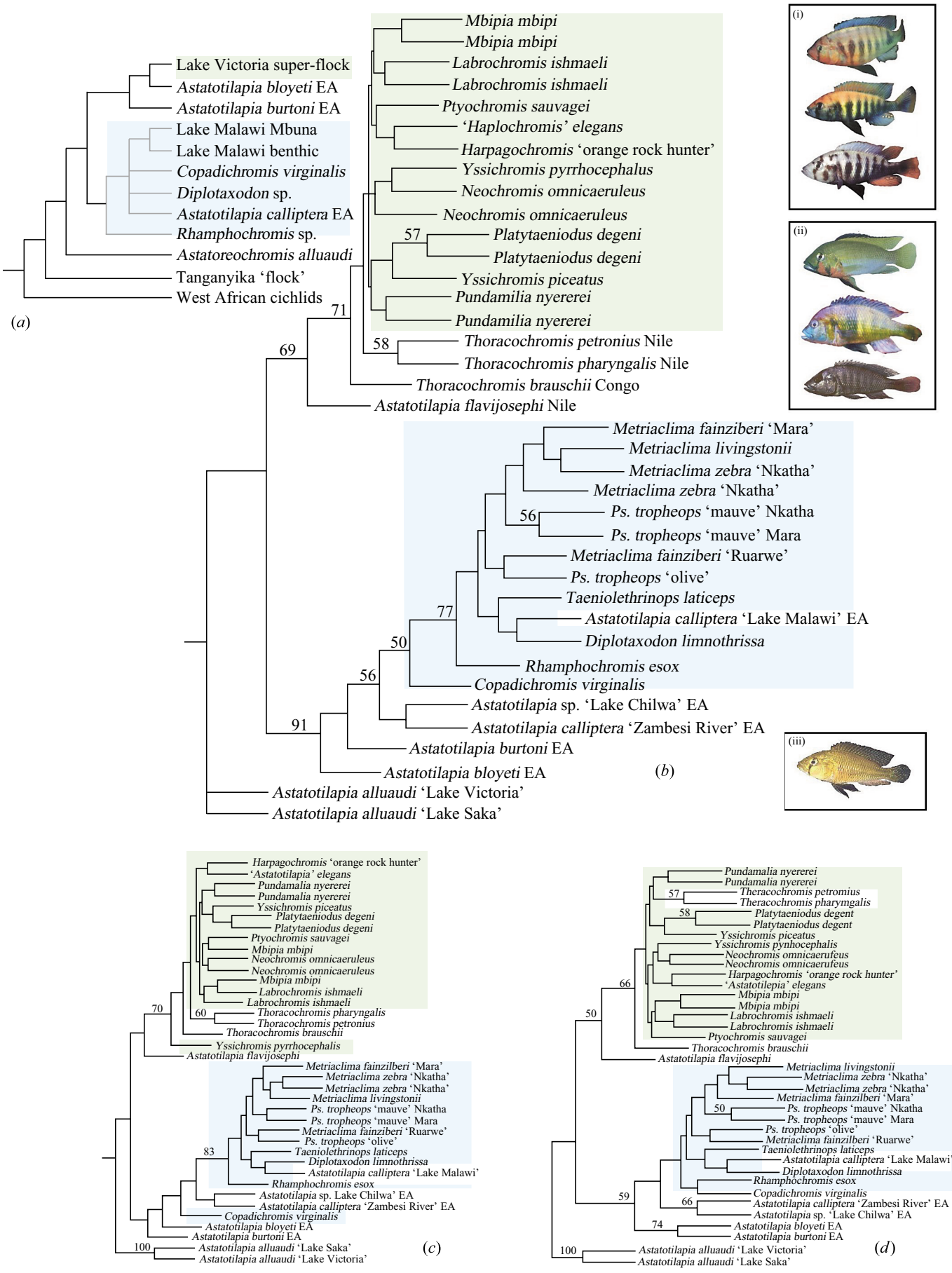

Figure 1. (*Caption overleaf.*)

Figure 1. (*a*) Phylogram for the haplochromine cichlid species flocks in Lake Victoria and Lake Malawi and species in African rivers as resolved by variation in mtDNA (after Meyer (1993); *Copadichromis mloto* is now called *C. virginalis*). All lineages present on this tree were sampled for our AFLP analysis. *Taeniolethrinops* is the representative in our dataset of the clade 'Lake Malawi benthic' A. Lake Malawi flock group B are the Mbuna (table 1). In addition, we sampled Nilotic and Congolese taxa. (*b*) Phylogram resolved by the FITCH program without clock assumption from a pairwise distance matrix based on variation at 246 polymorphic AFLP loci. The figures on the branches are percentage recovery in 100 bootstrap resamplings, each run ten times with different random input orders and global and local rearrangement. The three male nuptial colour types that dominate among the endemic cichlids of Lake Victoria are shown (*i*; three species of the Lake Victoria endemic genus *Pundamilia* representing the types 'red chest', 'red dorsum', 'blue') and compared with male nuptial coloration in *Thoracochromis* (*ii*, *T. brauschii* (photograph, P. Schupke), *T. pharyngalis*, *T. petronius* (photograph, M. Smith) and in *As. bloyeti* (*iii*)). A maximum-parsimony tree resembled this tree but resolved *As. burtoni*, *As. bloyeti* and *As. calliptera* Zambezi–*A*. spec. 'Lake Chilwa' and between *C. virginalis*. (*c*) Like (*b*) but resolved using the program KITSCH (assuming a molecular clock). Each of 100 bootstrap resamplings was run 100 times with different random input orders and global and local rearrangement. Note the positions of *Y. pyrrhocephalus* and *C. virginalis*. (*d*) Like (*b*) but resolved using the Minimum Evolution algorithm in KITSCH (assuming a molecular clock). Each of 100 bootstrap resamplings was run 100 times with different random input orders. Green boxes, endemic members of the Lake Victoria–Edward flock; blue boxes, endemic members of the Lake Malawi flock; EA, inhabitants of East African rivers; Nile/Congo, members of Nilotic–Congolese lineages.

uplift, formerly westflowing rivers reversed flow and formed lakes in the Lake Victoria–Edward region, which ultimately overflowed northwards into the Nile. The current Lake Victoria ichthyofauna, other than haplochromine cichlids, is a mosaic of Nilotic and Congolese taxa without significant contribution of the eastflowing rivers (Seehausen 2002). Introgression involving *Thoracochromis* at the base of the Lake Victoria–Edward radiation would be an alternative explanation for the genetic similarity between Lake Victoria haplochromines and the Nilotic–Congolese genus *Thoracochromis*.

Thoracochromis is anatomically, ecologically, and in coloration much more diverse than the East African *Astatotilapia*. Among the species that we sampled are a rock-living epibios eater (*T. petronius*), an omnivore (*T. brauschii*) and a mollusc crusher (*T. pharyngalis*) (Greenwood 1979). The North-Nilotic *As. flavijosephi*, which our data indicate is more closely related to *Thoracochromis* than to East African *Astatotilapia*, displays intraspecific polymorphism where females have papilliform pharyngeal teeth that are suited to feed on insects, whereas males possess molariform teeth on hypertrophied bones so as to crush molluscs (Werner 1976). Very similar morphological variation characterizes species differences in the Lake Victoria–Edward flock (Greenwood 1974) but is not known in East African *Astatotilapia* species even where they inhabit lakes. *Thoracochromis* also has diverse male nuptial coloration, resembling the three patterns that dominate the Lake Victoria–Edward species flock (blue-grey, red dorsum, red chest; figure 1) where they can be associated with disruptive selection and speciation (Seehausen & Van Alphen 1999). None of these colour patterns occur in *As. bloyeti*, the previously indicated sister taxon, and only blue and yellow patterns occur in *As. burtoni*. The identification of *Thoracochromis* as the likely paraphyletic ancestor of the Lake Victoria–Edward radiation indicates that the ability to express much of the functional genetic diversity found in the species flock in terms of morphology and coloration may by far pre-date the origin of the Lake Victoria–Edward species flock.

(c) *Genetic diversity older than species diversity*

Our data reveal similar levels of nuclear genetic variability within the Lake Victoria cichlids and the Lake Malawi cichlids, despite the fact that Lake Victoria and its species flock are estimated to be ten times younger than the Lake Malawi flock, and the flocks differ tenfold in their variability at mtDNA (Meyer *et al.* 1990; Nagl *et al.* 2000). Little is known about linearity or deviation from linearity of AFLP marker divergence over time. However, the power of AFLPs in resolving shallow (intraspecific populations (Giannasi *et al.* 2001; Parson & Shaw 2001; Schneider *et al.* 2002)) as well as deep (several million years old lineages (Albertson *et al.* 1999; Ogden & Thorpe 2002)) phylogenetic history is documented, and recent studies that directly compared genetic distances obtained from AFLPs and from sequence information found good congruency (Giannasi *et al.* 2001; Maguire *et al.* 2002; Ribeiro *et al.* 2002). Published information on nuclear gene sequences (Mayer *et al.* 1998) is consistent with our finding that the difference in genetic diversity between Lake Victoria and Lake Malawi cichlids is much less apparent at nuclear than at mitochondrial loci. It appears that the Victorian flock has arisen from a founding population that was highly variable at nuclear loci but not at mtDNA. It is difficult to conceive that a single species would have contained sufficient genetic variation to account for the observed nuclear gene diversity after 15 000 years (for the current Lake Victoria flock) or even after 100 000–400 000 years (for the Lake Victoria–Edward radiation).

Our data do not provide evidence for strict monophyly of the Lake Victoria flock. All our analyses place *'Haplochromis' elegans* from Lake Edward among the species from Lake Victoria. That the Lake Victoria radiation may not be monophyletic with regard to the cichlids of Lake Edward has been indicated before (Booton *et al.* 1999; Nagl *et al.* 2000). However, our data do not provide clear support for the monophyly of the wider Lake Victoria–Edward radiation either. Some tree-building algorithms place Nilotic and Congolese species of *Thoracochromis* among the basal branches of the radiation, which would indicate either polyphyly, recent radiation from a *Thoracochromis* ancestor the phenotype of which persisted in some of the endemic lacustrine lineages of Lake Edward, or hybridization. The contrast between muted mitochondrial and relatively large nuclear genetic variation and the weak support for monophyly from analysis of nuclear markers are consistent with the origin of the

present Victoria flock, and possibly also much of the larger Lake Victoria–Edward flock, from a hybrid swarm.

Applying a molecular clock to the mtDNA data indicates that the Lake Victoria–Edward cichlid species flock began to form between 100 000 and 400 000 years BP, shortly after the geological uplift in the west (Nagl *et al.* 2000). The region subsequently underwent three cycles of drier and moister periods, in the last of which Lake Victoria desiccated entirely 17 000 years BP (Johnson *et al.* 2000). The deeper Lake Edward may have persisted, albeit with much reduced water volume. At 14 600 years BP, Lake Victoria filled up rapidly and began to reconnect to Lake Edward via the Katonga river divide, and to overflow once again into the Nile. It is probable that haplochromine cichlids, like other fishes, re-colonized Lake Victoria in the Holocene from the neighbouring Lake Edward and possibly Lake Kivu (Kaufman *et al.* 1997; Seehausen 2002). Lake Edward contains haplochromines of divergent mtDNA haplotypes (Nagl *et al.* 2000), including *Thoracochromis* (Greenwood 1979). We propose that extensive introgression occurred in the wake of the recolonization of Lake Victoria. Lake Victoria cichlids hybridize in turbid water (Seehausen *et al.* 1997), and flooding of the grasslands would probably have been associated with high water turbidity. One of the parental mtDNA haplotypes could have become fixed in the hybrid swarm, as has been documented in cyprinids (DeMarais *et al.* 1992) and stickleback radiations (Taylor & McPhail 2000), either by chance or due to unidirectional geneflow. Crapon de Caprona & Fritzsch (1984) have shown that hybrids between female Lake Victoria cichlids and male *As. burtoni* suffer complete larval mortality, whereas the reciprocal cross yields viable and fertile F1, F2 and backcrosses. As the lake grew and water became clearer, disruptive sexual selection on male coloration in conjunction with increased nuclear genetic variability, transgressive segregation (as also shown by Crapon de Caprona & Fritzsch 1984) and diversifying ecological selection would have permitted rapid speciation and ecological diversification. Given the dramatic climatic fluctuations that the entire region experienced during the Pleistocene (Johnson *et al.* 1996), it is conceivable that even the wider Lake Victoria–Edward species flock has experienced cycles of introgression and speciation.

The identification of the morphologically variable genus *Thoracochromis* as a paraphyletic taxon ancestral to the Lake Victoria–Edward radiation, in conjunction with the contrast of high diversity at nuclear gene loci and extremely little at mtDNA, which might be explained by hybridization that preceded the radiation, perhaps offer a partial solution to the problem of the almost instantaneous origin of adaptive diversity during the radiation of cichlid fish in Lake Victoria (Fryer 2001).

(d) *Lake Malawi*

Our data also provide no clear evidence for the monophyly of the Lake Malawi flock. By contrast to mtDNA phylogenies, several tree-building algorithms that we have used place the endemic *C. virginalis* as the sister group to a clade in which species of the riverine genus *Astatotilapia* lie basal with respect to the other endemic Lake Malawi species. This differs from the relationship estimated by mtDNA. We found the lacustrine population of *As. calliptera* to lie well within a clade otherwise comprised only of Malawian endemics (as indeed is also indicated by mtDNA (Meyer 1993; Shaw *et al.* 2000)). The apparently anomalous positions of these two taxa might be explained by introgressive hybridization. Lake Malawi has witnessed dramatic changes in climate too and was almost dry until 0.5–1 MYBP (Sturmbauer *et al.* 2001).

We are currently analysing nuclear and mtDNA sequence variation of riverine, Lake Victoria and Lake Malawi haplochromines that may shed more light onto the possibility of ancestral hybridization and to further investigate relative levels of genetic variation in different markers.

The authors thank S. Balshine-Earn, F. Witte, P. Shaw and C. Rico for samples; L. Beukeboom, M. Maan and H. J. Mrosso for help in the molecular laboratory; E. Lippitsch for discussion; and R. Lande, D. Lunt, D. Schluter and two anonymous referees for comments on the manuscript. O.S. and J.v.A. were supported during part of this work by the Life Sciences section of the Dutch Science Foundation (NWO grant no. 805-36.025).

REFERENCES

Albertson, R. C., Markert, J. A., Danley, P. D. & Kocher, T. D. 1999 Phylogeny of a rapidly evolving clade: the cichlid fishes of Lake Malawi, East Africa. *Proc. Natl Acad. Sci. USA* **96**, 5107–5110.

Arnold, M. L. 1997 *Natural hybridisation and evolution*. Oxford Series in Ecology and Evolution. Oxford University Press.

Bernatchez, L., Glemet, H., Wilson, C. C. & Danzmann, R. G. 1995 Introgression and fixation of Arctic charr (*Salvelinus alpinus*) mitochondrial genome in an allopatric population of brook trout (*Salvelinus fontinalis*). *Can. J. Fish Aquat. Sci.* **52**, 179–185.

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. Phyl. Evol.* **11**, 273–282.

Cooke, H. B. S. 1958 Observations relating to quarternary environments in East and South Africa. *Trans. Proc. Geol. Soc. S. Afr.* **60(Annexure 5)**, 1–73.

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.

DeMarais, B. D., Dowling, T. E., Douglas, M. E., Minckley, W. L. & Marsh, P. C. 1992 Origin of *Gila seminuda* (Teleostei: Cyprinidae) through introgressive hybridization: implications for evolution and conservation. *Proc. Natl Acad. Sci. USA* **89**, 2747–2751.

Farias, I. P., Orti, G., Sampaio, I., Schneider, H. & Meyer, A. 2001 The cytochrome b gene as a phylogenetic marker: the limits of resolution for analyzing relationships among cichlid fishes. *J. Mol. Evol.* **53**, 89–103.

Felsenstein, J. 1984 Distance methods for inferring phylogenies: a justification. *Evolution* **38**, 16–24.

Felsenstein, W.M. 2001 PHYLIP v. 3.6a2 alpha. Department of Genetics, University of Washington, Seattle.

Fitch, W. M. & Margoliash, E. 1967 Construction of phylogenetic trees. *Science* **155**, 279–284.

Fryer, G. 2001 On the age and origin of the species flock of haplochromine cichlid fishes of Lake Victoria. *Proc. R. Soc. Lond.* B **268**, 1147–1152. (DOI 10.1098/rspb.2001.1601.)

Fryer, G. & Iles, T. D. 1972 *The Cichlid fishes of the great lakes of Africa: their biology and evolution*. London: Oliver & Boyd.

Giannasi, N., Thorpe, R. S. & Malhotra, A. 2001 The use of

amplified fragment length polymorphism in determining species trees at fine taxonomic levels: analysis of a medically important snake, *Trimesurus albolabris*. *Mol. Ecol.* **10**, 419–426.

Greenwood, P. H. 1974 The cichlid fishes of Lake Victoria, East Africa: the biology and evolution of a species flock. *Bull. Br. Mus. Nat. Hist., Zool.* **Suppl. 6**, 1–134.

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.

Hillis, D. M. & Huelsenbeck, J. P. 1992 Signal, noise, and reliability in molecular phylogenetic analyses. *J. Heredity* **83**, 189–195.

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.

Johnson, T. C., Kelts, K. & Odada, E. 2000 The holocene history of Lake Victoria. *Ambio* **29**, 2–11.

Kaufman, L. S., Chapman, L. J. & Chapman, C. A. 1997 Evolution in fast forward: haplochromine fishes of the Lake Victoria region. *Endeavour* **21**, 23–30.

Kidd, K. K. & Sgaramella-Zonta, L. A. 1971 Phylogenetic analysis: concepts and methods. *Am. J. Hum. Gen.* **23**, 235–252.

Kishino, H. & Hasegawa, M. 1989 Evaluation of the maximum likelihood estimate of the evolutionary tree topologies from DNA sequence data, and the branching order in Hominoidea. *J. Mol. Evol.* **29**, 170–179.

Kluge, A. G. & Farris, J. S. 1969 Quantitative phyletics and the evolution of the anurans. *Syst. Zool.* **18**, 1–32.

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.

Link, W., Dixkens, C., Singh, M., Schwall, M. & Melchinger, A. E. 1995 Genetic diversity in European and Mediterranean faba bean germ plasm revealed by RAPD markers. *Theor. Appl. Genet.* **90**, 27–32.

Lu, G., Basley, D. J. & Bernatchez, L. 2001 Contrasting patterns of mitochondrial DNA and microsateliite introgressive hybridization between lineages of lake whitefish (*Coregonus clupeaformis*); relevance for speciation. *Molec. Ecol.* **10**, 965–985.

McElroy, D. M. & Kornfield, I. 1993 Novel jaw morphology in hybrids between *Pseudotropheus zebra* and *Labeotropheus fuelleborni* (Teleostei, Cichlidae) from Lake Malawi, Africa. *Copeia* **4**, 933–945.

Maddison, W. P. & Maddison, D. R. 1992 *MacClade. Analysis of phylogeny and character evolution*. Sunderland, MA: Sinauer.

Maguire, T. L., Peakall, R. & Saenger, P. 2002 Comparative analysis of genetic diversity in the mangrove species *Avicennia marina* (Forsk.) Vierh. (Avicenniaceae) detected by AFLPs and SSRs. *Theor. Appl. Genet.* **104**, 388–398.

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

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

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.

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

Mueller, U. G. & Wolfenbarger, L. L. 1999 AFLP genotyping and fingerprinting. *Trends Ecol. Evol.* **14**, 389–394.

Nagl, S., Tichy, H., Mayer, W. E., Klein, J., Takezaki, N. & Takahata, N. 2000 The origin and age of haplochromine fishes in Lake Victoria, East Africa. *Proc. R. Soc. Lond.* B **267**, 1049–1061. (DOI 10.1098/rspb.2000.1109.)

Nei, M. & Li, W.-H. 1979 Mathematical model for studying genetic variation in terms of restriction endonucleases. *Proc. Natl Acad. Sci. USA* **76**, 5269–5273.

Ogden, R. & Thorpe, R. S. 2002 The usefulness of amplified fragment length polymorphism markers for taxon discrimination across graduated fine evolutionary levels in Caribbean *Anolis* lizards. *Mol. Ecol.* **11**, 437–445.

Page, R. D. M. & Holmes, E. C. 1998 *Molecular evolution: a phylogenetic approach*. Oxford: Blackwell Science.

Parson, Y. M. & Shaw, K. L. 2001 Species boundaries and genetic diversity among Hawaiian crickets of the genus *Laupala* identified using amplified fragment length polymorphism. *Mol. Ecol.* **10**, 1765–1772.

Ribeiro, M. M., Mariette, S., Vendramin, G. G., Szmidt, A. E., Plomion, C. & Kremer, A. 2002 Comparison of genetic diversity estimates within and among populations of maritime pine using chloroplast simple-sequence repeat and amplified fragment length polymorphism data. *Mol. Ecol.* **11**, 869–877.

Rueber, L., Meyer, A., Sturmbauer, C. & Verheyen, E. 2001 Population structure in two sympatric species of the Lake Tanganyika cichlid tribe Eretmodini: evidence for introgression. *Mol. Ecol.* **10**, 1207–1225.

Rzhetsky, A. & Nei, M. 1992 Statistical properties of the ordinary least-squares, generalized least-squares, and minimum-evolution methods of phylogenetic inference. *J. Mol. Evol.* **35**, 367–375.

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

Salzburger, W., Baric, S. & Sturmbauer, C. 2002 Speciation via introgressive hybridization in East African cichlids? *Mol. Ecol.* **11**, 609.

Schneider, M. V., Beukeboom, L. W., Driessen, G., Lapchin, L., Bernstein, C. & Van Alphen, J. J. M. 2002 Geographical distribution and genetic relatedness of sympatrical thelytokous and arrhenotokous populations of the parasitoid *Venturia canescens* (Hymenoptera). *J. Evol. Biol.* **15**, 191–200.

Seehausen, O. 2002 Patterns in fish radiation are compatible with Pleistocene desiccation of Lake Victoria and 14 600 year history for its cichlid species flock. *Proc. R. Soc. Lond.* B **269**, 491–497. (DOI 10.1098/rspb.2001.1906.)

Seehausen, O. & Van Alphen, J. J. M. 1999 Can sympatric speciation by disruptive sexual selection explain rapid evolution of cichlid diversity in Lake Victoria? *Ecol. Lett.* **2**, 262–271.

Seehausen, O., Van Alphen, J. J. M. & Witte, F. 1997 Cichlid fish diversity threatened by eutrophication that curbs sexual selection. *Science* **277**, 1808–1811.

Shaw, P. W., Turner, G. F., Idid, M. R., Robinson, R. L. & Carvalho, G. R. 2000 Genetic population structure indicates sympatric speciation of Lake Malawi pelagic cichlids. *Proc. R. Soc. Lond.* B **267**, 2273–2280. (DOI 10.1098/rspb.2000.1279.)

Sneath, P. H. A. & Sokal, R. R. 1973 *Numerical taxonomy*. San Francisco, CA: Freeman.

Stauffer, J. R. Jr, Bowers, N. J., Kocher, T. D. & McKaye, K. R. 1996 Evidence of hybridization between *Cynotilapia afra* and *Pseudotropheus zebra* (Teleostei: Cichlidae) following an intralacustrine translocation in Lake Malawi. *Copeia* **1996**, 203–208.

Sturmbauer, C., Baric, S., Salzburger, W., Ruber, L. & Verheyen, E. 2001 Lake level fluctuations synchronize gen-

etic divergences of cichlid fishes in African lakes. *Mol. Biol. Evol.* **18**, 144–154.

Swofford, D. L. 2001 *PAUP* 4b8*. Sunderland, MA: Sinauer.

Swofford, D. L. & Maddison, W. P. 1987 Reconstructing ancestral character states under Wagner parsimony. *Math. Biosci.* **87**, 199–229.

Talbot, M. R. & Laerdal, T. 2000 The late Pleistocene–Holocene palaeolimnology of Lake Victoria, East Africa, based upon elemental and isotopic amalyses of sedimentary organic matter. *J. Palaeolimn.* **23**, 141–164.

Taylor, E. B. & McPhail, J. D. 2000 Historical contingency and ecological determinism interact to prime speciation in sticklebacks, *Gasterosteus*. *Proc. R. Soc. Lond.* B **267**, 2375–2384. (DOI 10.1098/rspb.2000.1294.)

Tegelstrom, M. 1987 Transfer of mitochondrial DNA from the northern red-backed vole (*Clethrionomys rutilus*) to the bank vole (*C. glareolus*). *J. Mol. Evol.* **24**, 218–227.

Van de Peer, Y. & De Wachter, R. 1994 TREECON for Windows: a software package for the construction and drawing of evolutionary trees for the Microsoft Windows environment. *Comput. Applic. Biosci.* **10**, 569–570.

Vos, P., Hogers, R., Bleeker, M., Reijans, M., Van de Lee, T., Hornes, M., Frijters, A., Pot, J., Peleman, J., Kuiper, M. & Zabeau, M. 1995 AFLP: a new technique for DNA fingerprinting. *Nucl. Acid Res.* **23**, 4407–4414.

Werner, Y. L. 1976 Notes on reproduction in the mouthbrooding fish *Haplochromis flavijosephi* (Teleostei: Cichlidae) in the aquarium. *J. Nat. Hist.* **10**, 669–680.

Yamada, M., Higuchi, M. & Goto, A. 2001 Extensive introgression of mitochondrial DNA found between two genetically divergent forms of Threespine Stickleback, *Gasterosteus aculeatus*, around Japan. *Environ. Biol. Fish.* **61**, 269–284.

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