First publ. in: Science, Vol. 300 (2003), No. 5617, pp. 325-329

Origin of the Superflock of Cichlid Fishes from Lake Victoria, East Africa

Erik Verheyen,[1] Walter Salzburger,[2] Jos Snoeks,[3] Axel Meyer[2]*

Lake Victoria harbors a unique species-rich flock of more than 500 endemic haplochromine cichlid fishes. The origin, age, and mechanism of diversification of this extraordinary radiation are still debated. Geological evidence suggests that the lake dried out completely about 14,700 years ago. On the basis of phylogenetic analyses of almost 300 DNA sequences of the mitochondrial control region of East African cichlids, we find that the Lake Victoria cichlid flock is derived from the geologically older Lake Kivu. We suggest that the two seeding lineages may have already been lake-adapted when they colonized Lake Victoria. A haplotype analysis further shows that the most recent desiccation of Lake Victoria did not lead to a complete extinction of its endemic cichlid fauna and that the major lineage diversification took place about 100,000 years ago.

The adaptive radiation of more than 500 species of haplochromine cichlids from Lake Victoria (LV) is an astounding example of a morphologically, ecologically, and behaviorally diverse species flock (*1*). This fauna was initially assumed to have evolved in about 750,000 years, the presumed age of the LV basin (*2*). The neighboring Lakes Edward, George, and Kivu (LK) (Fig. 1) harbor less species-rich assemblages of cichlids that, combined with the LV radiation, have been grouped into a "superflock" of closely related species (*3*, *4*).

Molecular phylogenetic studies (*5*–*7*) suggested, contrary to earlier morphology-based studies, that LV's endemic cichlids evolved within the confines of their lake basin, and that they are of monophyletic origin (*5*, *6*). These studies also estimated that the endemic cichlids are an extremely young species flock, with an age of less than 200,000 years (*5*). This low age estimate was reduced further when geological data suggested that LV dried out entirely between 15,600 and 14,700 years ago (*8*, *9*). The hypothesis that the LV superflock evolved within the past 14,700 years (*8*–*10*) led to alternative scenarios, which implied that LV never dried out completely (*11*) or that part of the original superflock survived in river refugia, from which the LV basin was recolonized (*7*, *10*, *12*). Because the LV basin is considerably deeper than other water bodies in its surroundings, residual water bodies situated inside the lake basin are the most likely refugia (*13*, *14*).

[1]Vertebrate Department, Royal Belgian Institute of Natural Sciences, Vautierstraat 29, 1000 Brussels, Belgium. [2]Department of Biology, Lehrstuhl für Zoologie und Evolutionsbiologie, University of Konstanz, Universitätsstrasse 10, 78457, Konstanz, Germany. [3]Vertebrate Section, Royal Museum for Central Africa, Leuvensesteenweg 13, 3080 Tervuren, Belgium.

*To whom correspondence should be addressed. E-mail: axel.meyer@uni-konstanz.de

Competing hypotheses on the origin of the LV superflock have been proposed (*5*–*7*), and the most recent study suggested a Congolese-Nilotic ancestor (*15*). Yet, there is still uncertainty about the timing, the geographic origin of the populations that originally colonized LV, and the refugia from which cichlids recolonized LV after its most recent desiccation. The most direct way to study the geographic origin of faunas is through phylogeographic analyses using a dense and relevant geographic sampling.

Lake Kivu is located in the African Rift Valley and is completely isolated from the much younger LV (*16*). The formation of the Virunga Volcanoes about 14,000 to 11,000 years ago (*17*) or 25,000 to 20,000 years ago (*16*) interrupted the previous connection of proto–lake Kivu to the northern Rift lakes. Lake Kivu harbors 26 fish species, of which 15 are endemic haplochromines (*18*). Though presently connected to the Tanganyika system via the Rusizi, it is part of the East Coast ichthyogeographical region that includes Lakes Victoria, Edward, and George (*18*).

Here, we adopt a comprehensive phylogeographic approach to address the question of the origin of the haplochromine superflock, using sequence data published by Nagl *et al*. (*7*) and new sequences obtained in this study. These include: mitochondrial control region sequences (847 base pairs long) from East African haplochromine cichlids of lakes Victoria, Edward, George, Albert, representative East African rivers, and new sequences from 133 LK haplochromines representing 14 out of its 15 known endemic haplochromines, as well as sequences from fishes from small lakes in Uganda and Burundi. Sources of sequences and Genbank accession numbers are given in table S1.

In agreement with earlier studies (*6*, *7*), our phylogenetic analyses (*19*) show that the Lake Victoria, Edward, George, and Albert haplochromines form a monophyletic superflock that includes two specimens (*7*) originating from Tanzanian rivers close to Lake Rukwa (Figs. 1 and 2A). Except for three specimens of all the assayed haplochromines from LK, all Kivu cichlids sampled belong to this superflock (Fig.

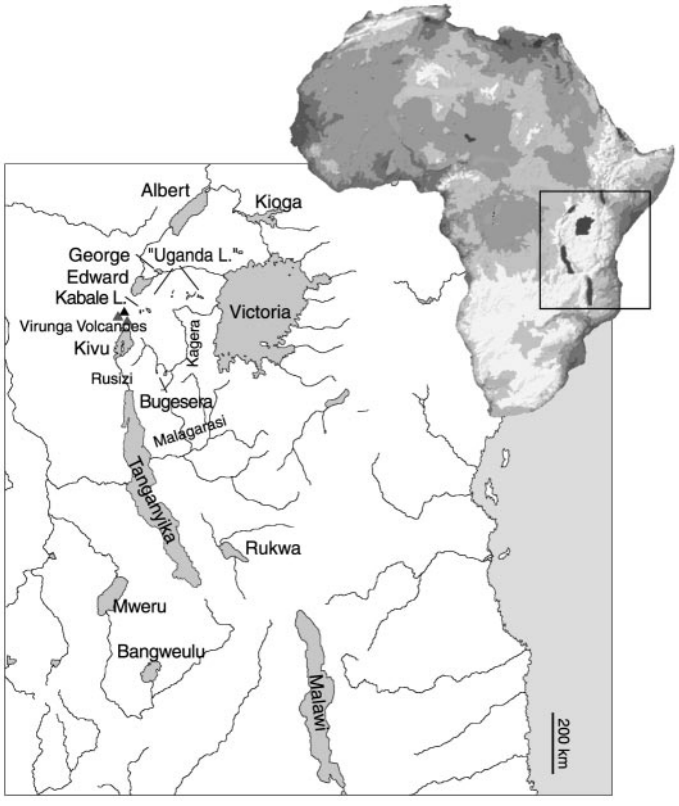

Fig. 1. Map of East Africa showing the lakes and rivers studied.

2). These three specimens have been identified as *Haplochromis gracilior* (Fig. 2C), an endemic LK species that is clearly phylogenetically distinct from the superflock haplochromines (by 30 to 42 mutations), and also occupies a pivotal position because it does not belong to the superflock, but appears to be its most recent sister species.

Our haplotype network approach (*19*, *20*)—permitting a fine-grained reconstruction of the evolutionary histories of these young lacustrine faunas—shows that the cichlids from LK are crucial for the evolutionary history of the LV superflock. The haplotypes of *H. gracilior* have a state "A" in the diagnostic site 630 (Fig. 3C) and are therefore connected to the network through the central Rift Valley haplotype (25 in Fig. 3C). The haplotype network demonstrates the extensive sampling, because almost all possible haplotypes are represented (Fig. 3C). Four additional observations support the crucial role of LK haplochromines in the evolution of all the haplochromines of eastern Africa: (i) the haplotypes of fishes from other lakes are connected by, and therefore derived from, LK haplotypes; (ii) the LK fish are relatively more diverse, although LK currently contains only 15 species as compared to more than 500 in LV; we detected 41 haplotypes in the faunas from both lakes; (iii) even excluding *H. gracilior*, LK haplotypes show an average pairwise distance of 0.6% as compared to 0.5% for LV; (iv) a central haplotype (25 in Fig. 3C) is found in some species from all large lakes, but more than 50% of the fishes with this haplotype are LK endemics.

The haplotype network also captures the colonization events between LK and the other lakes. First, Lakes Edward, George, and Albert must have been seeded by at least four lineages (starting from haplotypes 25 and 56), and second, the sequence of haplotypes derived from these two central haplotypes suggests that the colonization of the Rift Valley cichlids may have occurred in a stepwise manner, starting from LK, through the Edward-George region, and lastly to Lake Albert (Figs. 3C and 4). It is also apparent that at least two lineages that are derived from haplotype 56 (considered to be ancestral by the network tree approach) seeded LV, thus making the LV species flock diphyletic. The few mutations that separate the Victoria and Kivu haplotypes, and the distribution of the central Rift Valley haplotype 25 over a large range of the sampling area (Fig. 3, A and C), indicate that faunal connections must have existed between these waterbodies until recently. This agrees with the geological evidence for the recent cessation of connections between Lakes Kivu and Edward (*16*, *17*). In addition, there is geological evidence that supports a recent connection between LV and LK (*21*). Such a connection has been traced between the Kibuye and Bugesera areas that probably encompassed what is currently the upper Kagera River basin (Fig. 1). Indeed, seismic patterns and field observations indicate that the western Kivu border-fault segment has served as a master fault for crustal

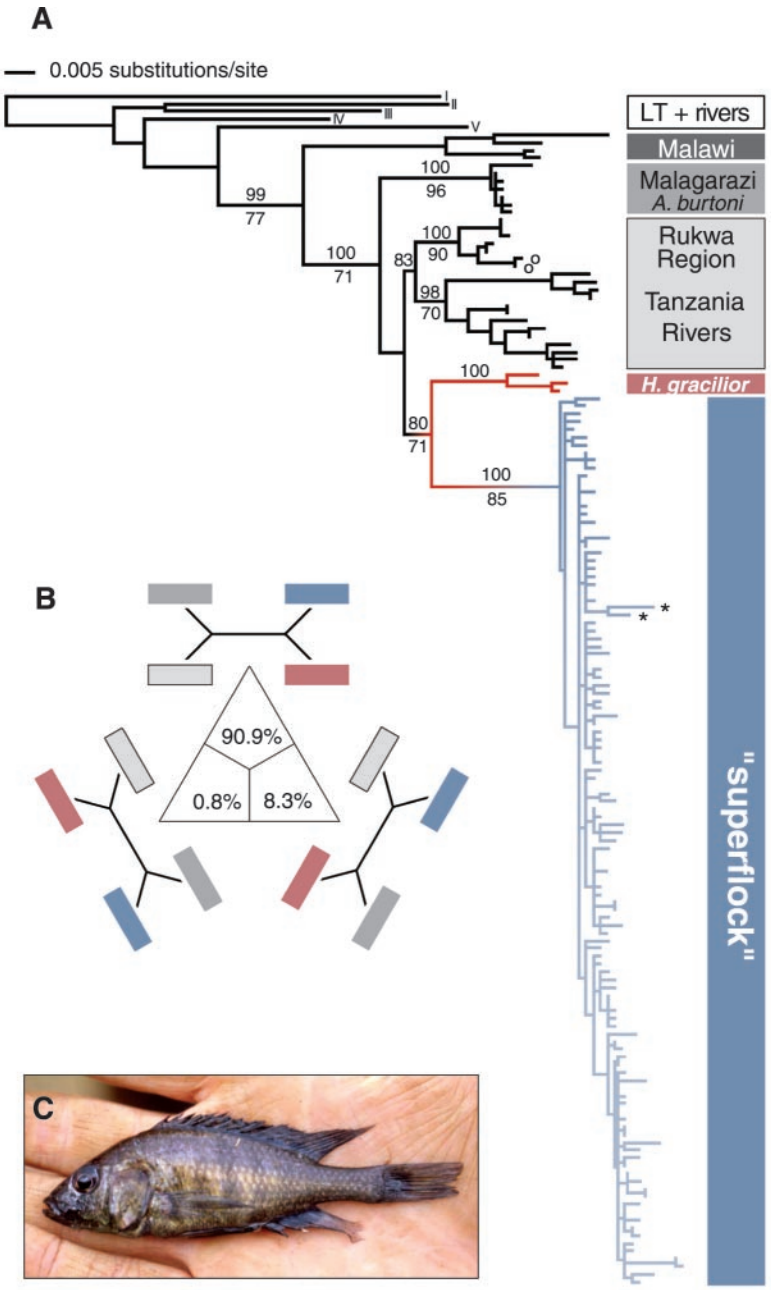

Fig. 2. (**A**) Maximum likelihood analyses [general time-reversible model with gamma correction (GTR+G+I)], based on 161 sequences, support the monophyly of the *Haplochromis* superflock and suggest that the Kivu-endemic *H. gracilior* is its sister species. Numbers above the branches represent values from Bayesian inference obtained with the MrBayes program (*32*); numbers below represent quartet puzzling values obtained with PAUP* (*33*). This analysis combines sequence data from East African riverine and lacustrine haplochromines (*7*) (table S1); cichlids from Lake Malawi (*Pseudotropheus* sp. *msobo*, *Labeotropheus trewavasae*, *Lethrinops auritus*, and *Cyrtocara moorii*) (*34*); and other relevant taxa such as the nonendemic *Astatoreochromis alluaudi* (IV, LV region), *Serranochromis* sp. (III, Lake Mweru), *Thoracochromis brauschi* (II, Congo Basin), and *Petrochromis orthognathus* (V, Lake Tanganyika). LT, Lake Tanganyika; *A. burtoni*, *Astatotilapia burtoni*. The tree was rooted with a representative cichlid lineage from Lake Tanganyika [*Limnochromis auritus* (I)] (*30*). Not all members of the superflock are lacustrine. Some occur in the Lake Rukwa region (marked by asterisks), whereas others of predominantly riverine clades (marked by circles) occur in Lakes Edward and George [according to (*7*)]. (**B**) Results of the four-cluster likelihood mapping analysis (*35*), represented as a triangle showing the likelihood support for three alternative topologies. The topology with *H. gracilior* as a sister group to the superflock is strongly supported. (**C**) *H. gracilior*.

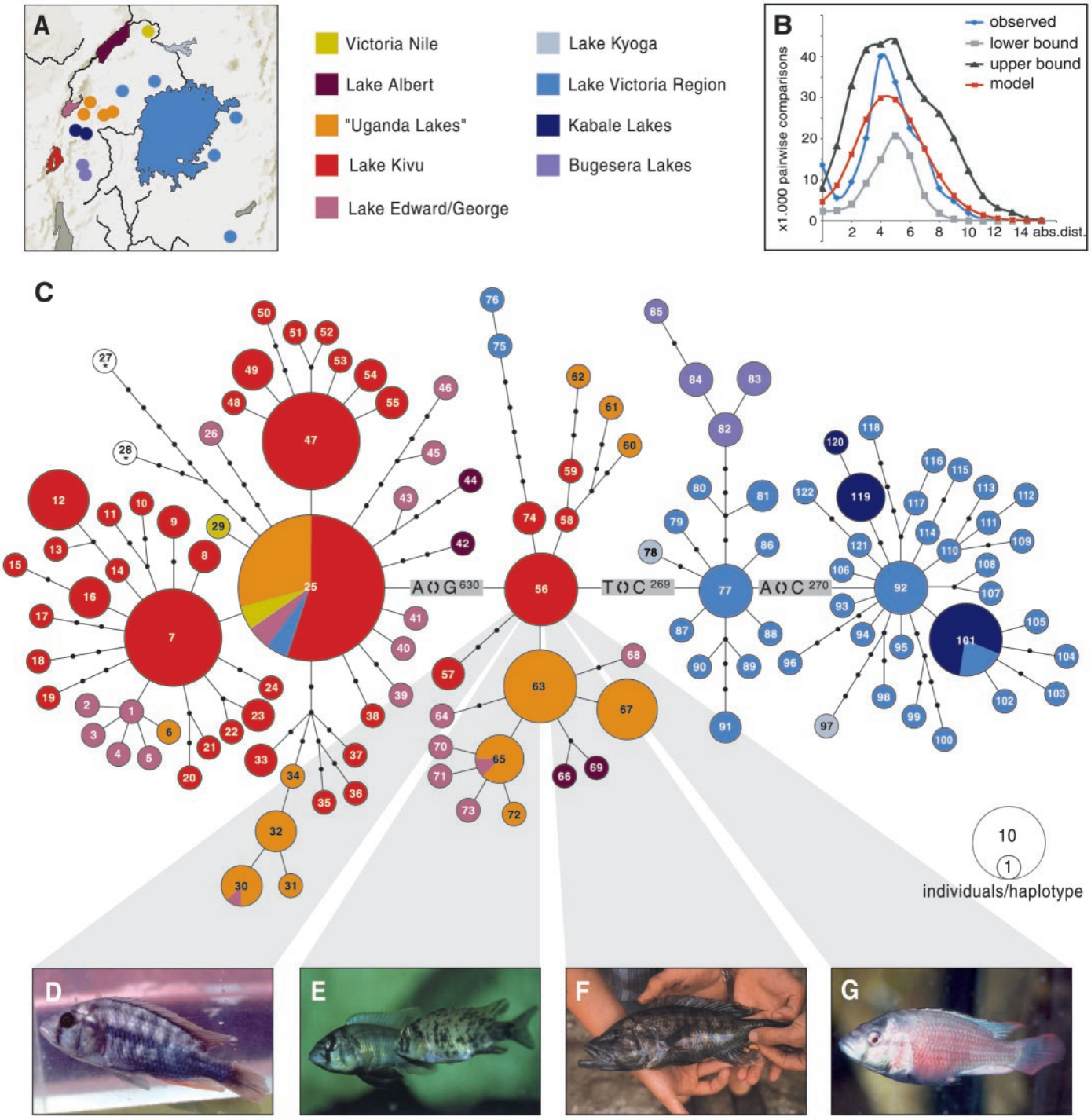

Fig. 3. (**A**) Map showing colored images of the lakes discussed. (**B**) The results obtained using Arlequin 2.1 (*36*) demonstrate that the entire superflock is considerably older than 14,700 years. The major demographic extension occurred 4.2 mutations ago for the entire mitochondrial control region (3.15 mutations for the first 365 base pairs). The latter translates into 0.863% sequence divergence (*24*), which suggests that the maximum population expansion occurred about 98,000 to 132,700 years ago. The upper- and lower-bound curves are 2.5 and 97.5 percentile values of 5000 simulations. (**C**) Unrooted haplotype network of the haplochromine superflock (the specimens indicated with a blue bar in Fig. 2A). Haplotypes are colored according to the respective lakes (specimens are listed in table S1). The sizes of the haplotypes reflect the number of specimens sharing the same haplotype (see scale in the lower right corner). Each of the 122 haplotypes is numbered (for example, haplotype 25 is the central haplotype that is found in LK and other lakes). Mutations characterizing transitions between the four central haplotypes are shown. Haplochromines from the small Uganda lakes were introduced with tilapias from Lake Edward (*37*). (**D** to **G**) Haplotype 56 evolved into two lineages that colonized LV. It is shared by six species that represent a large part of the ecomorphological diversity of the endemic LK flock. Four of those species are shown: (D) *H. crebridens* (epilithic algae grazer), (E) *H. paucidens* (insect eater), (F) *H. vittatus* (piscivore), and (G) *H. rubescens* (epilithic algae grazer). [Photos copyright Africa Museum Tervuren, Belgium]

extension during the Quaternary and that an uplifting along the central part of the border-fault segment elevated a terrace of late Pleistocene lacustrine sequences to about 500 m above the present lake level (*22*). The breakdown of the eastern rift shoulder caused rivers in the south and in central Rwanda to drain into the proto–lake Kivu. After the eastern rift shoulder was elevated, Rwandan rivers were drained to the north, but after the Virungas volcanic event, the northward drainage of LK was interrupted and the direction of flow of many rivers was reversed, creating the modern river systems east of LK, including a series of lesser lakes in the Bugesera region and in Uganda, after which the connection between the Kagera and LV originated (*23*). Thus, it appears that the tectonic activities leading to the uplift of the present barrier between LK and the Kagera and Bugesera river systems may be of sufficiently recent origin to support our scenario.

A molecular clock has been calibrated for the cichlid control region (*24*) and was used as described in (*24*). It suggests that the vicariance event that split the LV and LK cichlid faunas must have occurred less than 41,500 to 30,500 years ago. This age estimate also falls within the range of the highest geological estimates for the eruption of the Virunga Volcanoes that separate LK from the northern Rift Valley lakes (*17*).

The pattern of genetic variation within the major LV clade differs from the pattern observed in LK and the Rift Valley lakes. There are intermediate haplotypes missing along the long branches derived from haplotypes 77 and 92, which might be the result of a massive extinction, possibly related to the most recent dessication of LV between 15,600 and 14,700 years ago (*8*, *9*). This might be explained by the fact that LV is shallower and would therefore be much more affected by climatological changes than the other large, generally deeper, Rift Vally lakes. The most recent dessication of LV should have eradicated its entire fish fauna and genetic diversity. The estimated timing of the demographic expansion within the entire superflock suggests that the entire superflock is considerably older than the 14,700 years since LV refilled (Fig. 3B) (*8*, *9*). Also, the haplotypes situated at the end of these long branches derived from the central LV haplotypes 77 and 92 must belong to an older radiation, which, according to our data, occurred between 98,000 to 132,700 years ago (Fig. 3B) (*19*). The presence of more distinct and therefore older haplotypes (Fig. 3C) that originated long before this event strongly argues against the view that LV dried out completely. But, clearly, after refilling during the last 14,700 years, LV experienced a vast increase in the number of individuals, but not the origination of many new haplotypes.

The fast radiation of the ecomorphological diversity in LV haplochromines may be explained by their descent from lacustrine, possibly already diversified, Kivu ancestors, and by the finding that all extant LV haplochromines evolved from lineages that survived the most recent low water stand in LV. It is interesting that the "source haplotype" [haplotype 56, which links the lineages of all lakes (Fig. 3C)] is shared by six LK species that display a considerable amount of the ecomorphological diversity that is found in haplochromines (Fig. 3, D to G). In view of this, it is possible that similar morphologies of haplochromines in Lakes Victoria, Edward, George, and Kivu evolved only once (*4*), although the implied monophyly for the entire superflock is not supported by unambiguous morphological evidence (*4*, *25*). Alternatively, evolutionary mechanisms, such as atavisms and the retention of ancestral genetic programs (*1*, *26*, *27*), might explain the rapid origin of morphological novelties and repeated phenotypic diversification. It is noteworthy that a genetically relatively homogeneous species flock (LV) contains a higher degree of morphological divergence and vastly higher number of species (more than 500) than the much smaller cichlid species flocks (15 species in LK, 60 in lakes Edward and George, and 6 in Lake Albert), some of which are genetically more diverse and hence older. The lack of correlation between the genetic divergence and therefore age on one hand, and the morphological diversification and species-richness on the other hand, has been observed before (*28*). Closely related lineages in different lakes will not necessarily have similar speciation rates, and it appears that the youngest and largest basin (LV) provided more opportunities that facilitated speciation.

The discovery that descendents of Kivu haplochromines have colonized LV parallels the finding that descendants of Lake Tanganyika cichlids appear to have colonized river systems (*29*, *30*) and other lakes (*5*, *30*). Therefore, both lakes appear to have acted as evolutionary reservoirs that, because of their greater depth and resulting increased relative stability, conserved lineages that seeded the neighboring rivers and lakes at a later time. The basal position of the LK and Lake Tanganyika haplochromines in the major radiations (LV and Lake Malawi), enhances their evolutionary significance. Based on phylogenetic criteria (*31*), these faunas should be given a high priority in conservation programs.

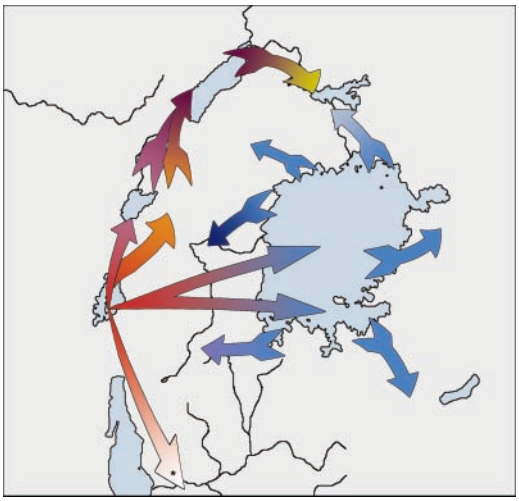

Fig. 4. Scenario of proposed colonization events between the lakes investigated. The asterisk in the arrow on the bottom left side refers to two haplochromines from the Lake Rukwa region (Fig. 2A) (*7*).

References and Notes

1. M. L. J. Stiassny, A. Meyer, *Sci. Am.* **280**, 64 (1999).
2. P. H. Greenwood, *The Haplochromine Fishes of the East African Lakes: Collected Papers on Their Taxonomy, Biology* (Cornell Univ. Press, Ithaca, NY, 1981).
3. _____, *Bull. Br. Mus. Nat. Hist. (Zool.)* **25**, 139 (1973).
4. _____, *Bull. Br. Mus. Nat. Hist. (Zool.)* **39**, 1 (1980).
5. A. Meyer, T. D. Kocher, P. Basasibwaki, A. C. Wilson, *Nature* **347**, 550 (1990).
6. G. C. Booton *et al.*, *Mol. Phylogenet. Evol.* **11**, 273 (1999).
7. S. Nagl *et al., Proc. R. Soc. London Ser. B* **267**, 1049 (2000).
8. T. C. Johnson *et al.*, *Science* **273**, 1091 (1996).
9. T. C. Johnson, K. Kelts, E. Odada, *Ambio* **29**, 2 (2000).
10. O. Seehausen, *Proc. R. Soc. London Ser. B* **269**, 491 (2002).
11. G. Fryer, *Proc. R. Soc. London Ser. B* **268**, 1147 (2001).
12. L. S. Kaufman, L. J. Chapman, C. A. Chapman, *Endeavour* **21**, 23 (1997).
13. G. Fryer, *Hydrobiologia* **354**, 177 (1997).
14. J. Snoeks, *J. Aquaricult. Aq. Sci.* **9**, 150 (2001).
15. O. Seehausen *et al.*, *Proc. R. Soc. London Ser. B* **270**, 129 (2003).
16. L. C. Beadle, *The Inland Waters of Tropical Africa. An Introduction to Tropical Limnology* (Longman, London, 1974).
17. A. Pouclet, *Mus. R. Afr. Centr. Dépt. Géol. Min. Rapp. Ann.* **1977**, 145 (1978).
18. J. Snoeks, L. De Vos, D. Thys van den Audenaerde, *S. Afr. J. Sci.* **93**, 579 (1997).
19. The specimens and their species identifications, geographic origin, accession numbers, and sources are provided in table S1. Materials and methods are available as supporting material on *Science* Online.
20. M. Clement, D. Posada, K. A. Crandall, *Mol. Ecol.* **9**, 1657 (2000).
21. J. Moeyersons, *Mus. R. Afr. Centr. Tervuren (Belgium) Dépt. Géol. Min. Rapp. Ann.* **1978**, 127 (1979).
22. C. Ebinger, *Geol. Soc. Am. Bull.* **101**, 885 (1989).
23. U. Schmidt, in *Palaeoecology of Africa and the Surrounding Islands*, J. Runge, Ed. (Balkema, Lisse, Tokyo, 2001), pp. 51–62.
24. C. Sturmbauer, S. Baric, W. Salzburger, L. Rüber, E. Verheyen, *Mol. Biol. Evol.* **18**, 144, (2001).
25. E. Lippitsch, *J. Fish Biol.* **51**, 284 (1997).
26. M. L. J. Stiassny, *Neth. J. Zool.* **42**, 260 (1992).
27. A. Meyer, in *Homology*, G. R. Bock, G. Cardew, Eds. (Wiley, New York, 1999), pp. 141–157.

28. C. Sturmbauer, A. Meyer, *Nature* **358**, 578, (1992).
29. C. Sturmbauer, E. Verheyen, A. Meyer, *Mol. Biol. Evol.* **11**, 691 (1994).
30. W. Salzburger, A. Meyer, S. Baric, E. Verheyen, C. Sturmbauer. *Syst. Biol.* **51**, 113 (2002).
31. M. L. J. Stiassny, M. C. C. de Pinna, in *Systematics and Conservation Evaluation*, P. L. Forey, C. J. Humphries, R. I. Vane-Wright, Eds. (Clarendon, Oxford, 1994), pp. 235– 249.
32. J. Huelsenbeck, F. Ronquist, MrBayes program, available at http://morphbank.ebc.uu.se/mrbayes (2001).
33. D. Swofford, *PAUP*. Phylogenetic Analysis Using Parsimony (*and other methods)* (Sinauer, Sunderland, MA, 2000), ed. 4.04a.
34. W. J. Lee, J. Conroy, W. H. Howell, T. D. Kocher, *J. Mol. Evol.* **41**, 54 (1995).
35. K. Strimmer, A. van Haeseler, *Proc. Natl. Acad. Sci. U.S.A.* **94**, 6815 (1997).
36. S. Schneider, D. Roessli, L. Excoffier, *ARLEQUIN, A Software for Population Genetics Data Analysis* (Genetics and Biometry Laboratory, Department of Anthropology, Univ. of Geneva, Switzerland, 2000).
37. Anonymous, *Annual Report of the East African Fisheries Research Organisation. Appendix C.* (East African Fisheries Research Organisation, 1953), vol. 41.
38. We thank the Royal Museum for Central Africa–Tervuren and the Institut National de Recherches Scientifiques–Butare for their support of the fieldwork of E.V. in Rwanda in 1984. We acknowledge J. Engelken from the University of Konstanz for lab work; L. De Vos and E. Schraml for specimens; J. Moeyersons, L. Tack, and C. Ebinger for discussions on the geological history of the region; the "Cichlid group" from the Meyer lab and L. Bernatchez for discussions; and J. Day for comments on the manuscript. This work was funded by a visiting professor fellowship from the University of Konstanz; a grant from The Belgian Federal Office for Scientific, Technical, and Cultural Affairs and support through the Ecological Genetics Research Network of the Fund for Scientific Research–Flanders to E.V.; and grants from the Deutsche Forschungsgemeinschaft, the Ministry of Science, and Art of Baden-Württemberg and the University of Konstanz to A.M. and W.S.

Supporting Online Material
www.sciencemag.org/cgi/content/full/1080699/DC1
Materials and Methods
Table S1

21 November 2002; accepted 17 February 2003
Published online 20 March 2003;
10.1126/science.1080699
Include this information when citing this paper.