Proc. Natl. Acad. Sci. USA
Vol. 95, pp. 14238–14243, November 1998
Evolution

Persistence of neutral polymorphisms in Lake Victoria cichlid fish

(speciation/gene trees)

SANDRA NAGL*, HERBERT TICHY*, WERNER E. MAYER*, NAOYUKI TAKAHATA†, AND JAN KLEIN*‡

*Max-Planck-Institut für Biologie, Abteilung Immungenetik, Corrensstrasse 42, D-72076 Tübingen, Germany; and †Department of Biosystems Science, Graduate University for Advanced Studies, Hayama, Kanagawa 240–01, Japan

Communicated by Masatoshi Nei, Pennsylvania State University, University Park, PA, September 23, 1998 (received for review June 8, 1998)

ABSTRACT Phylogenetic trees for groups of closely related species often have different topologies, depending on the genes used. One explanation for the discordant topologies is the persistence of polymorphisms through the speciation phase, followed by differential fixation of alleles in the resulting species. The existence of transspecies polymorphisms has been documented for alleles maintained by balancing selection but not for neutral alleles. In the present study, transspecific persistence of neutral polymorphisms was tested in the endemic haplochromine species flock of Lake Victoria cichlid fish. Putative noncoding region polymorphisms were identified at four randomly selected nuclear loci and tested on a collection of 12 Lake Victoria species and their putative riverine ancestors. At all loci, the same polymorphism was found to be present in nearly all the tested species, both lacustrine and riverine. Different polymorphisms at these loci were found in cichlids of other East African lakes (Malawi and Tanganyika). The Lake Victoria polymorphisms must have therefore arisen after the flocks now inhabiting the three great lakes diverged from one another, but before the riverine ancestors of the Lake Victoria flock colonized the Lake. Calculations based on the mtDNA clock suggest that the polymorphisms have persisted for about 1.4 million years. To maintain neutral polymorphisms for such a long time, the population size must have remained large throughout the entire period.

Genetic theory predicts that a neutral polymorphism at a nuclear locus of a diploid organism will persist in a population for an average of $4N_e$ generations, where N_e is the effective population size, roughly the number of breeding individuals (1). The theory implies that, under certain circumstances, neutral polymorphisms may persist through the phase of species formation. This persistence may thereby complicate phylogenetic reconstruction, because a phylogenetic tree of a gene may not necessarily reflect the phylogeny of a species accurately (2–4). The circumstances include a relatively short speciation phase and a large founding population size. Although discrepancies between gene and species trees have indeed been observed repeatedly (5, 6), as far as we know, the persistence of neutral polymorphism through the speciation phase has never been documented. An opportunity to test the prediction and to determine the frequency of persisting neutral polymorphisms is offered by the haplochromine species flock of Lake Victoria in East Africa.

Haplochromines are fish of the family Cichlidae (7) that have recently undergone explosive adaptive radiation in the great lakes of the East African Rift Valley and their satellites (8–10). The radiation in Lake Victoria occurred recently. Although the lake apparently dried out completely for a period of several thousand years and did not begin to fill with water until 12,400 years ago (11), it is now inhabited by some 300 species that had originally been assigned to a single genus, *Haplochromis* (8), but more recently they have been divided among 33 genera (12). Most of the species are endemic to Lake Victoria and are therefore believed to have arisen within the period of some 12,000 years since the reconstitution of the lake. The young age of the Lake Victoria haplochromines is supported by molecular studies of mtDNA (13), the major histocompatibility complex (*Mhc*) genes (14), and other genetic markers (15). The mtDNA studies also suggest that Lake Victoria haplochromines may be monophyletic (13). The founders of the flock may have come from the rivers in the Lake Victoria basin, in which their descendants may still live (8). The recency of speciation makes the Lake Victoria haplochromines a suitable model for testing the predicted persistence of neutral polymorphisms through the speciation phase. The present study has been designed to test this prediction.

MATERIALS AND METHODS

Fish. Cichlid fish were caught in East African lakes and rivers (Fig. 1) during expeditions in October and November of 1993, 1995, and 1996. Additional samples were kindly provided by Lothar Seegers (Dinslaken, Germany), who also helped us with the identification of the different species. Voucher specimens of the species used have been deposited at the Musée Royal de l'Afrique Centrale, Tervuren, Belgium (16).

Preparation of Genomic DNA. Pieces of fins were fixed in 70% ethanol. 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), followed by phenol/chloroform extraction.

PCR, Cloning, and Sequencing. Amplifications (17) were carried out in the PTC-100 and PTC-200 Programmable Thermal Controller (MJ Research, Oldendorf, Germany). Genomic DNA (50–100 ng) was added to the reaction mixture of 1× 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 (Pharmacia), 0.2 μM of each of the sense (S) and antisense (A) primers, and 2.5 units of *Taq* DNA polymerase (Pharmacia). The PCR program consisted of a denaturation step for 3 min at 94°C, followed by 35 cycles of 40 sec of denaturation at 94°C, 30 sec of annealing at 48–61°C depending on the primer combination, and 2 min of 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+}

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. §1734 solely to indicate this fact.

© 1998 by The National Academy of Sciences 0027-8424/98/9514238-6$2.00/0
PNAS is available online at www.pnas.org.

Abbreviations: A, antisense; G6P, glucose-6-phosphatase; indel, insertion/deletion site; S, sense; UTR, untranslated region.

Data deposition: The sequences reported in this paper have been deposited in the GenBank database (accession nos. for *HN49* alleles 1–8, AF069077–AF069084; *SN-Y* alleles 1–12, AF069085–AF069096).

‡To whom reprint requests should be addressed. e-mail: jan.klein@tuebingen.mpg.de.

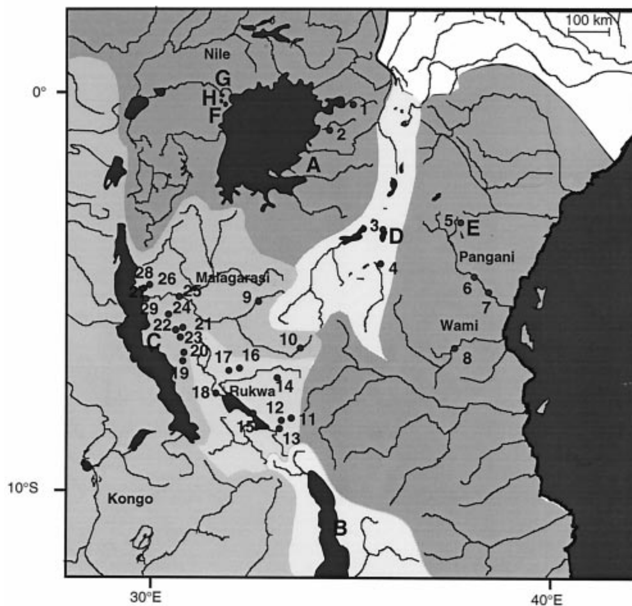

FIG. 1. Geography of East Africa. Drainage systems are indicated by different shadings. A, Lake Victoria; B, Lake Malawi; C, Lake Tanganyika; D, Lake Manyara; E, Lake Chala; F, Lake Nabugabo; G, Lake Kayania; H, Lake Kayugi. Cichlids were caught at localities indicated by numbers.

beads (Invitrogen) were added to the mixture. The following 12 primers were used: P209 [glucose-6-phosphatase (*G6P*), coding region, S], TGCTCACTTCCCACACC; P215 (*G6P*, coding region, S), CAACCAAGATGAGGATTATGAG; P216 [*G6P*, 3′ untranslated region (UTR), A], AGACACT-GAAAAGACAGTTATT; P217, (*G6P*, 3′ UTR, A), AGA-CACTGAAAAGACAGTTCTA; GSP1 (*G6P*, 3′ UTR, A), TGGCTGCGTGTATGTGTAAAAATC; HN43–86 (actin, exon 6, S), TACGCCAACAATGTGCTCTCC; HN43–139 (actin, exon 7, A), GCATGGTTCAGTGGTGGTTTT; HN49–32 (*HN49*, exon, S), TAGAGCAGGAAAGGAG-GAAGG; HN49–297 (*HN49*, S), AGAGGCCGCTTCCCC-GAG; HN49–600 (*HN49*, A), GGCGTCTGTCGTCTCTGT-CC; SN-Y-35 (*SN-Y*, S) CGTCTCTGTCCTCGACACCTG; and SN-Y-246 (*SN-Y*, A) TCTTCTGTGTTGTGCCATGCG.

PCR products were isolated, cloned by using standard methods, and sequenced in an ALF sequencer (Pharmacia). Sequence alignments were prepared with the SEQPUP program, version 0.6e (18).

Single-Strand Conformation Polymorphism Typing. We mixed 6 μl of a PCR product with an equal volume of single-strand conformation polymorphism loading dye (95% formamide/10 mM NaOH/0.1% bromophenol blue/0.1% xylene cyanole). The samples were then denatured at 95°C for 5 min and cooled in ice water for 5 min. A 7.6-μl portion of each sample was loaded on a GeneGel Excel 12.5% polyacrylamide gel (Pharmacia) and subjected to electrophoresis in the GenePhor system (Pharmacia) for 10 min at 200 V, 10 mA, and 5 W and then for 2 h and 30 min at 375 V, 15 mA, and 10 W at 15°C. The DNA was visualized with the DNA Silver Staining Kit (Pharmacia).

Heteroduplex Analysis. We mixed 5 μl of a PCR product with an equal volume of a reference PCR product from an individual homozygous for the studied marker. The samples were denatured for 5 min at 95°C and annealed for 5 min at room temperature. A 7.6-μl portion of each sample was loaded on a polyacrylamide gel and subjected to electrophoresis as described for single-strand conformation polymorphism typing.

RESULTS

We identified five molecular markers at four loci and used them to survey haplochromine species from Lake Victoria and the rivers in the lake's basin. We obtained one of the markers as a byproduct from another study, and it was identified as being located in the 3′ UTR of the *G6P* gene. The remaining four markers were picked up randomly from a directional cDNA library prepared from *Astatotilapia nubila*. The selected phage clones were PCR-amplified. The 5′ and 3′ ends of the inserts were sequenced. Locus-specific primers, based on the sequences obtained, were used to PCR-amplify genomic DNA from several individuals of the same species. Length differences between the amplification products obtained from cDNA and genomic DNA were taken as an indication of the presence of introns. To detect polymorphism, the intron-containing PCR products from different species were cloned and sequenced. A method was then established for distinguishing the variants and was used in a survey of haplochromine species. A brief description of the markers and the results we obtained by using them follows.

G6P*.** The primer pair P209 and GSP1 was used to amplify a 1.3-kb PCR fragment of this gene. Sequencing of the product showed two alleles differing by a 3-bp (AAT) insertion/deletion site (indel) in the 3′ UTR of the sequence. To distinguish the alleles, an upstream primer, P215, was used in combination with primers spanning the indel site, either P216, specific for the AAT allele, or P217, specific for the "" allele, (where the asterisks indicate the absence of nucleotides corresponding to the AAT sites). In a survey of 78 fish, the two alleles were found in 10 of the 11 tested species representing the major trophic groups of the Lake Victoria haplochromine flock; the *** allele was absent from the nonendemic species *Astatoreochromis alluaudi* and from *Lipochromis melanopterus* (Table 1). The latter is a pedophage (7); whether it truly lacks the *** allele remains uncertain, because only five individuals were available for testing. The two alleles were also found in some of the river species, although fish at some localities seemed to be fixed for one or the other allele (Table 1). In *Astatotilapia burtoni*, the *** allele was absent from six of the seven localities tested. The *** allele is absent in cichlids of Lake Malawi and in the *Tropheus* lineage of Lake Tanganyika, although some of the *Tropheus* species possess another allele characterized by a 304-bp deletion that starts next to the 3-bp deletion site of Lake Victoria cichlids (data not shown).

Actin. The primer pair HN43–86 and HN43–139 amplified intron 6 (161 bp) and parts of exons 6 and 7 (together, 495 bp) of the skeletal actin 1 gene. Polymorphism was detected 37 bp downstream from the exon 6/intron 6 border in the form of a single base-pair substitution that was part of a *Hin*dIII restriction site (AAGCTT or AAGTTT). Digestion by *Hin*dIII of the 656-bp PCR product from the AAGCTT homozygotes yielded two fragments (518 bp and 138 bp). Digestion of the products from AAGCTT heterozygotes yielded three bands (656 bp, 518 bp, and 138 bp). The treatment of the product from AACTTT homozygotes left the 656-bp band undigested. The two alleles were found to be present in all of the 12 tested endemic Lake Victoria species at frequencies ranging from 16% to 50% for the "−" allele (where − indicates the absence of the *Hin*dIII site; Table 1). The two alleles were also present in the riverine species, although, in these, the "+" allele prevailed with the − allele apparently absent from 15 of the 23 river localities. The − allele was absent in the nonendemic species *A. alluaudi* from Lake Victoria and in cichlid species from Lake Malawi and Tanganyika (data not shown).

***HN49*-Indel Polymorphism.** The primer pair HN49–32 and HN49–600 amplified a 1012-bp PCR fragment from an unidentified locus. Comparison of cDNA and genomic DNA sequences indicated the presence in the genomic fragment of two putative introns, X (388 bp) and Y (131 bp). Sequence

Table 1. Frequency (%) of alleles at *G6P*, actin, and *HN49* loci in cichlid fishes

Species	Locality	*G6P*			Actin			*HN49*		
		n	AAT	***	*n*	+	−	*n*	+	−
Lake Victoria and satellite lakes										
Astatotilapia nubila		4	87.5	12.5	4	75	25	16	93.8	6.2
Astatotilapia velifer		0			6	58.3	41.7	39	98.7	1.3
Enterochromis cinctus		7	71.4	28.6	7	57.1	42.9	8	100	0
Haplochromis pyrrhocephalus		6	83.3	16.7	6	83.3	16.7	36	93.1	6.9
Lipochromis melanopterus		5	100	0	6	50	50	6	100	0
Neochromis nigricans		7	57.1	42.9	8	75	25	8	87.5	12.5
Paralabidochromis chilotes		11	54.5	45.5	11	77.3	22.7	23	93.5	6.5
Paralabidochromis plagiodon		7	57.1	42.9	8	87.5	12.5	8	87.5	12.5
Prognathochromis venator		6	66.7	33.3	8	68.8	31.2	8	93.8	6.2
Psammochromis riponianus		8	56.2	43.8	8	81.2	18.8	8	100	0
Ptyochromis sauvagei		8	37.5	62.5	8	75	25	8	81.3	18.7
Ptyochromis xenognathus		9	61.1	38.9	10	60	40	31	87.1	12.9
Astatoreochromis alluaudi		5	100	0	6	100	0	6	100	0
Rivers										
Astatotilapia sparsidens	3	2	100	0	2	100	0	1	0	100
Astatotilapia bloyeti CHALA	5	1	100	0	1	100	0	1	0	100
Astatotilapia bloyeti	1	19	15.8	84.2	19	57.9	42.1	19	81.6	18.4
	2	15	0	100	15	43.3	56.7	15	100	0
	4	6	100	0	10	100	0	10	0	100
	6	6	50	50	5	60	40	6	50	50
	7	3	50	50	2	0	100	3	100	0
	8	4	100	0	4	25	75	3	100	0
Astatotilapia katavi	11	1	0	100	1	100	0	1	0	100
	13	3	0	100	3	66.7	33.3	4	0	100
	14	11	36.4	63.6	10	100	0	11	0	100
	15	1	100	0	1	100	0	1	0	100
	17	7	14.3	85.7	7	92.9	7.1	7	0	100
	18, 20	3	0	100	3	100	0	3	0	100
	23	2	25	75	2	100	0	2	0	100
Astatotilapia burtoni	9	2	0	100	2	25	75			
	10, 28	2	100	0	2	100	0	2	0	100
	24	1	100	0	1	100	0	1	50	50
	26	1	100	0	1	100	0	1	100	0
	27	1	100	0	1	100	0			
	29	20	100	0	15	100	0	17	100	0

n, number of individuals. Localities are indicated by numbers as in Fig. 1. Allelic designations at locus *G6P*, AAT or *** indicate the presence or absence of the AAT trinucleotide; at locus actin, + or − indicate the presence or absence of *Hin*dIII restriction site; and at locus *HN49*, + or − indicate the presence or absence of 14-bp insert.

analysis of intron Y showed a 14-bp deletion polymorphism. For the species survey, a new primer, HN49–297, was used in combination with HN49–600 to yield a 359-bp intron-Y-containing PCR fragment that, when analyzed on a 12.5% polyacrylamide gel, could be distinguished from the 345-bp product containing the deletion. Both alleles were found in 9 of the 11 endemic Lake Victoria species and 2 of the riverine species *(Astatotilapia bloyeti* and *A. burtoni*; Table 2). The allele with the deletion could not be found in the tested Lake Malawi and Lake Tanganyika cichlids (data not shown).

***HN49*-Substitutional Polymorphism.** In addition to the indel polymorphism, intron Y of the HN49 locus also displayed substitutional polymorphism, which could be surveyed by heteroduplex analysis. A 359-bp PCR product was amplified from intron Y by using the primer pair HN49–297 and HN49–600 on genomic DNA from individuals previously identified as insertion homozygotes. The products were mixed with a similarly amplified product obtained from a single reference individual homozygous for the deletion allele, and the mixture was subsequently denatured. After annealing, the mixture was subjected to gel analysis, and the existence of a number of band patterns was indicated. The identity of the individual bands was established by a combination of DNA-mixing experiments and sequencing. Alleles responsible for the patterns could be thus identified. A similar approach was used to test the variability of the deletion homozygotes. Altogether, eight alleles were identified in the Lake Victoria and the riverine cichlid species. (Allele 8 contains the deletion and differs from allele 7 in one nucleotide pair; it was found in two individuals of *A. nubila* only; Table 2.) Several additional alleles were found in cichlids of Lakes Malawi and Tanganyika (data not shown). Of the eight alleles, six were found in the insertion homozygotes and two in deletion homozygotes; this distribution reflects the difference in the frequency of the insertion versus the deletion alleles. The alleles differed by 1- to 4-bp substitutions scattered over the entire fragment (Fig. 2). The common alleles were found in most of the 12 tested Lake Victoria species, as well as in at least some of the riverine species. The rare alleles had a more restricted distribution, but the sample sizes were not sufficiently large to decide in which species the alleles were truly absent. Allele 1, the most common of the set, was found in Lakes Victoria and Malawi, but not in Lake Tanganyika, the oldest of the three great bodies of water. The higher substitutional variability of the insertion genes compared to the deletion genes might be a reflection of a difference in their age, the former being older than the latter. The reason for the high variability of intron Y, however, remains a mystery.

SN-Y. HN49–600, one of the primers used in the analysis of the *HN49* locus, bound also at two other sites in the genome

Table 2. Frequency (%) of alleles at the *HN49* locus as detected by heteroduplex analysis in cichlid fishes

Species	Locality*	*n*	Insertion alleles				Deletion alleles†
			1	2	3	4	7
Lake Victoria and satellite lakes							
Astatotilapia nubila‡		16	22	66	6		1
Astatotilapia velifer		34	24	75			
Enterochromis cinctus		8	12	88			6
Haplochromis pyrrhocephalus		35	4	90			
Lipochromis melanopertus		3	17	83			13
Neochromis nigricans		8	31	56			7
Paralabidochromis chilotes		23	17	76			12
Paralabidochromis plagiodon		8		88			6
Prognathochromis venator		8	6	88			
Psammochromis riponianus		8		100			19
Ptyochromis sauvagei		8	12	69			14
Ptychromis xenognathus		58	19	67			
Rivers							
Astatotilapia bloyeti	1	19	16	66			
	2	15	37	63			
	6	6	50				
	7, 8	6	100				
Astatotilapia burtoni	9	1		50			
	24	1	50				
	26	1				100	
	27	1	100				

n, number of individuals.
*The localities are indicated by numbers as in Fig. 1.
†Allele 8 was found in two individuals (6%) of *A. nubila* only. Insertion alleles 5 and 6 were found in 16 individuals of *A. burtoni* from the locality 29 at frequencies of 6% and 94%, respectively.
‡Specimens collected at three different localities: Lake Victoria (Mwanza), Lake Nabugabo, and Lake Kayugi.

and amplified a 340-bp polymorphic PCR fragment. This fragment was unrelated to *HN49* but showed a 70% sequence similarity to a region of a 7.3-kb *Fugu rubripes* clone that contained the *Hsp70–2* gene (GenBank accession no. Y08577). The region of similarity is about 1.5 kb downstream of this gene. The extent of the similarity and other indicators suggest that the amplified segment contains part of an exonic sequence. At this anonymous locus, 12 alleles could be identified by single-strand conformation polymorphism typing of the 12 Lake Victoria and 4 riverine cichlid species. The distinctiveness of these alleles could be confirmed by sequencing (Fig. 2). The differences between the alleles are mostly in the putative intron region, and most consist of nucleotide substitutions; two alleles, however, differ from the others by a 15-bp deletion. Allele 1, the most common allele, was found in all the tested species, both from Lake Victoria and from the rivers (Table 3). Other alleles seem to have a more restricted distribution, but several of them are present in both the lacustrine and riverine species. Lake Malawi cichlids, both mbuna and nonmbuna, are characterized by a different set of alleles at this locus (data not shown).

DISCUSSION

In the present study, five polymorphisms at four loci were found in species belonging to the haplochromine flock of Lake Victoria by random testing of either mRNA or clones from a cDNA library. The relative ease with which polymorphisms have been found contradicts earlier conclusions about the invariance of Lake Victoria cichlid fish at the molecular level (19). In fact, molecular polymorphisms may be abundant in these fish.

The five polymorphisms occur in noncoding regions, but, in the vicinity of coding sequences, they occur either in the 3′ UTR or in introns. Because of (*i*) their location in regions believed to be largely exempt from natural selection, (*ii*) their association with genes some of which (e.g., *G6P*) are evolving under purifying selection (20), and (*iii*) the fact that some of them are indel polymorphisms, it seems reasonable to assume that most, if not all, are selectively neutral. The only contraindication to neutrality is that two of the polymorphisms (*HN49* and *SN-Y*) are not only relatively high (8 and 12 alleles, respectively) but also include alleles that differ by multiple substitutions (as many as four in some cases). The heightened polymorphism at the two loci could be explained, for example, by a close linkage to an exon or locus under balancing selection or by increased mutability of the region.

The comparison of the polymorphisms found in the different East African lakes and rivers allows us to put limits on the interval of time within which the individual variants must have arisen. All five polymorphisms are shared by the various species of the Lake Victoria haplochromines, as well as by these haplochromines and at least some of the related species in the river systems in the Lake Victoria region; however, they seem largely absent from the cichlid fish of Lakes Malawi and Tanganyika, as well as in the nonendemic species *A. alluaudi*. All these alleles must have, therefore, arisen before the Lake Victoria flock began to radiate but after it diverged from the various endemic lineages inhabiting Lakes Malawi and Tanganyika. If one assumes that the riverine species tested in the present study included the ancestors (or their descendants) of the Lake Victoria flock, then the emergence of the polymorphisms must have predated the divergence of the flock from these ancestors. Therefore, most of the polymorphisms must have arisen more than 12,000 and less than 2 million years ago.

With a few exceptions, which may be the result of inadequate sampling, all the polymorphisms were found to be shared by the various species of the endemic Lake Victoria haplochromine flock. The species tested included representatives of the various trophic groups that are believed to have diverged from one another early in the flock's evolution (21). The transspecies polymorphisms must have therefore been passed from

HN49

Allele	57	79	81	82	83	84	85	86	87	88	89	90	91	92	93	94	95	97	100	131	139	177	229	236	249	260	332
	G	A	T	C	A	C	A	A	T	G	T	T	G	A	T	G	C	A	C	C	G	C	C	C	G	C	C
1	-	-	-	-	-	-	G	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
2	-	T	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	T	-	T	-	-	-	-
3	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-		-	-	T	-	-	-	-	-	-	-
4	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	C	-	-	-	T	-	-	T	-	-
5		-	-	-	-	-	-	-	C	-	-	-	-	-	-	-	G	-	-	-	-	-	-	A	-	T	-
6	A	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	T	-	-	-	-	-	-	-	-
7	-	-	*	*	*	*	*	*	*	*	*	*	*	*	*	*	-	-	-	-	-	-	-	-	-	-	-
8	-	-	*	*	*	*	*	*	*	*	*	*	*	*	*	*	-	-	-	A	-	-	-	-	-	-	-

SN-Y

Allele	25	28	46	107	108	109	110	111	112	113	114	115	116	117	118	119	120	121	157	170	184	200
	G	G	C	G	G	C	A	A	A	A	C	G	C	C	A	A	A	A	C	T	T	C
1	T	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
2	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
3	-	-	-	-	A	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	C	-
4	-	-	-	-	A	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
5	-	-	-	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	T	-	C	-
6	T	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	C	-	-
7	T	-	T	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	C	-	-
8	T	-	T	-	-	-	-	-	-	-	T	-	-	-	-	-	-	-	-	C	-	-
9	-	-	-	-	A	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	C	T
10	T	A	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
11	-	-	-	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	T	-	-	-
12	T	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	T	-	-	-

FIG. 2. Informative nucleotide sites of the loci *HN49* and *SN-Y*. The top line shows the position of the nucleotide in the consensus sequence. A dash indicates identity to the consensus; an asterisk indicates a deletion.

ancestral to descendant species in each evolutionary lineage. Explorations of transspecies polymorphisms have often been restricted to loci known to be under balancing selection, such as the *Mhc* loci in jawed vertebrates (22) or the self-incompatibility loci in flowering plants (23). The transspecies polymorphism described here provides examples of passage of presumably neutral variants through numerous speciation phases. That such passages probably occur has been suspected for some time (2–5); our study not only bears out this suspicion but also shows that the phenomenon is widespread and must be taken into consideration in phylogenetic reconstruction. It will complicate attempts to reconstruct the phylogenies of recently radiated species, exemplified by the Lake Victoria haplochromine flock. However, because radiations like those observed in Lake Victoria may have occurred at the time of divergence of most major lineages, ancestral polymorphism might complicate many other phylogenetic reconstructions as well.

The maintenance of neutral polymorphisms requires a certain magnitude and constancy of N_e. The fact that all the polymorphisms found turned out to be transspecific indicates that, generally, speciation is completed before the segregating neutral alleles become fixed. Therefore, it can be expected that very few fixed molecular markers will be available for phylogenetic reconstruction of adaptively radiating species groups. However, the neutral polymorphisms can be used for phylogenetic reconstruction based on genetic distances calculated from gene frequencies. The number of loci reported here is too low to attempt such a reconstruction, but the observed interspecific differences in genetic distances (not shown) provide a promising lead in this direction.

The data allow us to address questions concerning the sizes of the populations in the lineages of individual species and of the founding stock of the haplochromines in Lake Victoria. The first question has been approached with the help of the coalescence theory, specifically the formula for the probability $g_{ij}(t)$ that i $(= 2n)$ genes sampled in an extant population descended from j ancestral genes t time units ago (i.e., formulas 6.1 and 6.2 in ref. 24). The time units are expressed conveniently as $t = T/(2N_e)$, where T is the number of generations separating the extant population N_e from the ancestral population. Because, at each of the four tested loci, certain alleles are shared between the lacustrine and riverine species, j is no smaller than the number (k) of shared alleles. For conservative estimates of N_e, we set $j = k$. We also set T as time of divergence of the lacustrine and riverine species, estimated from mtDNA

Table 3. Frequency (%) of alleles at the *SN-Y* locus in cichlid fishes

Species	*n*	Alleles*							
		1	2	3	4	5	6	7	8
Lake Victoria and satellite lakes									
Astatotilapia nubila	16	25	28	31	13	3			
Astatotilapia velifer	37	24	57	5	8				
Enterochromis cinctus	7	14	36	29	14				
Haplochromis pyrrhocephalus	33	14	8	5	48	3			
Lipochromis melanopterus	6	25		42					
Neochromis nigricans	8	13	19	19					
Paralabidochromis chilotes	27	48	4	43		2			
Paralabidochromis plagiodon	8	6	63	13	19				
Prognathochromis venator	8	44	50						
Psammochromis riponianus	8	12	88						
Ptyochromis sauvagei	8	25	38	25		6			
Ptyochromis xenognathus	30	10	25	55					
Rivers									
Astatotilapia sparsidens	2						100		
Astatotilapia bloyeti	54	33		7		18	20	2	2
Astatotilapia katavi	29	38	2		2			43	12
Astatotilapia burtoni	7	21					29	14	14

n, number of individuals.

*Only alleles shared between species are shown. The following alleles were present in only two species: 9 (*N. nigricans*, *P. chilotes*), 10 (*H. pyrrhocephalus*, *P. venator*), 11 (*N. nigricans*, *A. bloyeti*), and 12 (*P. chilotes*, *P. sauvagei*).

Table 4. Estimates of t and π for 12 Lake Victoria haplochromine species

	Species											
	1	2	3	4	5	6	7	8	9	10	11	12
t	0.64	—	0.80	0.64	1.31	0.71	0.77	1.16	0.96	1.40	0.72	0.73
π (%) *HN49*	0.36	0.33	0.19	0.16	0.25	0.43	0.32	0.19	0.18	0	0.25	0.27
π (%) *SN-Y*	0.41	0.20	0.27	0.20	0.19	0.28	0.53	0.16	0.28	0.10	0.49	0.25

Species 1 is *A. nubila*; 2, *A. velifer*; 3, *E. cinctus*; 4, *H. pyrrhocephalus*; 5, *L. melanopterus*; 6, *N. nigricans*; 7, *P. chilotes*; 8, *P. plagiodon*; 9, *P. venator*; 10, *P. riponianus*; 11, *P. sauvagei*; and 12, *P. xenognathus*.

sequences. The nucleotide substitution rate was estimated to be 2% per site per 1 million years based on full-length control region sequences from Lake Malawi and Lake Victoria cichlids (S.N., W.E.M., H.T., and J.K., unpublished data) and on the age of Lake Malawi (2 million years; ref. 25). With these values, we estimate that the haplochromines of Lake Victoria and those of the Lake Victoria river system diverged 1.4 million years ago. Assuming the generation time of cichlid fish to be $\approx$3 years (26), we obtain $T = 470{,}000$ generations. Because the coalescence-based estimate of t is $\approx$1 for all 12 Lake Victoria species (Table 4), N_e of the individual species lineages must have been of the order of 10^5.

To estimate the size of the population that founded the Lake Victoria haplochromine flock we use the nucleotide diversity (π) computed for the *HN49* and *SN-Y* loci from pairwise comparisons of the nucleotide sequences (Table 4). We infer the lower boundary for π from the number and frequencies of the alleles shared between the lacustrine and riverine species under the assumption that the average frequencies did not change in time so that they can be given as averages over all extant species. We then obtain $\pi = 0.36\%$ for the *HN49* locus and $\pi = 0.48\%$ for the *SN-Y* locus, with an average of $\pi = 0.42\%$ for the two loci. This average does not differ much from that of the individual lacustrine species, indicating that the nucleotide diversity has not changed during the interval that separates the founding stock from the descendant species. Assuming the mutation rate (μ) of the loci to be of the order of 10^{-8} per site per generation (27), we estimate, from the formula $\pi = 4N_e\mu$, that N_e is of the order of 10^5.

Both estimates are based on several assumptions, of which the two most important ones are the values of the substitution rate of the mtDNA control region and the mutation rates of the nuclear genes. However, even if these rates were one order of magnitude higher than computed or estimated, the two principal implications of the present study would remain relevant, namely that presumably neutral polymorphisms can persist in populations for a long enough time to be passed from species to species along an evolutionary lineage and that N_e in the Lake Victoria haplochromine lineages have been large over long periods of time.

We thank Dr. Lothar Seegers for specimens and his help in species identification, Ms. Ljubica Sanader for technical assistance, Drs. Yoko Satta, (Graduate University for Advanced Studies, Hayama) and Brent Murray (Max-Planck-Institut für Biologie, Tübingen) for advice, and Ms. Niamh Ní Bhleithín for editorial assistance.

1. Kimura, M. (1955) *Proc. Natl. Acad. Sci. USA* **41,** 144–150.
2. Nei, M. (1972) *Am. Nat.* **106,** 283–292.
3. Nei, M. & Li, W. H. (1979) *Proc. Natl. Acad. Sci. USA* **76,** 5269–5273.
4. Takahata, N. & Nei, M. (1985) *Genetics* **110,** 325–344.
5. Pamilo, P. & Nei, M. (1988) *Mol. Biol. Evol.* **5,** 568–583.
6. Ball, R. M., Neigel, J. E. & Avise, J. C. (1990) *Evolution* **44,** 360–370.
7. Greenwood, P. H. (1981) *The Haplochromine Fishes of the East African Lakes* (Cornell Univ. Press, Ithaca, NY).
8. Greenwood, P. H. (1974) *Bull. Br. Mus. Nat. Hist. Zool. Suppl.* **6,** 1–134.
9. Meyer, A. (1993) *Trends Ecol. Evol.* **8,** 279–284.
10. Keenleyside, M. H. A. (1991) in *Cichlid Fishes: Behavior, Ecology and Evolution* (Chapman & Hall, London), pp. 103–128.
11. Johnson, T. C., Scholz, C. A., Talbot, M. R., Kelts, K., Ricketts, R. D., Ngobi, G., Beuning, K., Ssemmanda, I. & McGill, J. W. (1996) *Science* **273,** 1090–1093.
12. Greenwood, P. H. (1979) *Bull. Br. Mus. Nat. Hist. Zool.* **35,** 265–322.
13. Meyer, A., Kocher, T. D., Basasibwaki, P. & Wilson, A. C. (1990) *Nature (London)* **347,** 550–553.
14. Klein, J., Klein, D., Figueroa, F., Sato, A. & O'hUigin, C. (1997) in *Molecular Systematics of Fishes*, eds. Kocher, T. D. & Stepien, C. A. (Academic, New York), pp. 271–283.
15. Sültmann, H. & Mayer, W. E. (1997) in *Molecular Systematics of Fishes*, eds. Kocher, T. D. & Stepien, C. A. (Academic, New York), pp. 39–52.
16. Seegers, L. (1996) *The Fishes of the Lake Rukwa Drainage*, Annales Sciences Zoologiques (Musée Royal de L'Afrique Centrale, Tervuren, Belgium), Vol. 278.
17. Saiki, R. K., Gelfand, D. H., Stoffel, S., Sharf, S. J., Higuchi, R., Horn, G. T., Mullis, K. B. & Erlich, H. A. (1988) *Science* **239,** 487–491.
18. Gilbert, D. G. (1996) SEQPUP, A Biosequence Editor and Analysis Application (Indiana Univ., Bloomington, Indiana), Version 0.6e.
19. Sage, R. D., Loiselle, P. V., Basasibwaki, P. & Wilson, A. C. (1984) in *Molecular Versus Morphological Change Among Cichlid Fishes of Lake Victoria*, eds. Echelle, A. A. & Kornfield, I. (Univ. of Maine Press, Orono, ME), pp. 185–201.
20. Nagl, S., Mayer, W. E. & Klein, J. (1998) *DNA Res*. **2,** 1–5.
21. Sage, R. D. & Selander, R. K. (1975) *Proc. Natl. Acad. Sci. USA* **72,** 4669–4673.
22. Takahata, N., Satta, Y. & Klein, J. (1992) *Genetics* **130,** 925–938.
23. Ioerger, T. R., Clark, A. G. & Kao, T. H. (1990) *Proc. Natl. Acad. Sci. USA* **87,** 9732–9735.
24. Taveré, S. (1984) *Theor. Popul. Biol.* **26,** 119–164.
25. Fryer, G. & Iles, T. D. (1972) *The Cichlid Fishes of the Great Lakes of Africa* (TFH Publications, Neptune City, NJ).
26. Charlesworth, B. (1980) *Evolution in Age-Structured Populations* (Cambridge Univ. Press, Cambridge, U.K.).
27. Nei, M. (1987) *Molecular Evolutionary Genetics* (Columbia Univ. Press, New York).