

Comparative Phylogenetic Analyses of the Adaptive Radiation of Lake Tanganyika Cichlid Fish: Nuclear Sequences Are Less Homoplasious But Also Less Informative Than Mitochondrial DNA

Céline Clabaut,¹ Walter Salzburger,^{1,2} Axel Meyer¹

¹ Lehrstuhl für Zoologie und Evolutionsbiologie, Department of Biology, University of Konstanz, 78457 Konstanz, Germany

² Center for Junior Research Fellows, University of Konstanz, 78457 Konstanz, Germany

Received: 15 July 2004 / Accepted: 5 May 2005 [Reviewing Editor: Dr. Rafael Zardoya]

Abstract. Over 200 described endemic species make up the adaptive radiation of cichlids in Lake Tanganyika. This species assemblage has been viewed as both an evolutionary reservoir of old cichlid lineages and an evolutionary hotspot from which the modern cichlid lineages arose, seeding the adaptive radiations in Lakes Victoria and Malawi. Here we report on a phylogenetic analysis of Lake Tanganyika cichlids combining the previously determined sequences of the mitochondrial ND2 gene (1047 bp) with newly derived sequences of the nuclear RAG1 gene (~700 bp of intron 2 and ~1100 bp of exon 3). The nuclear data—in agreement with mitochondrial DNA—suggest that Lake Tanganyika harbors several ancient lineages that did not undergo rampant speciation (e.g., Bathybatini, Trematocarini). We find strong support for the monophyly of the most species-rich Tanganyikan group, the Lamprologini, and we propose a new taxonomic group that we term the C-lineage. The Haplochromini and Tropheini both have an 11-bp deletion in the intron of RAG1, strongly supporting the monophyly of this clade and its derived position. Mapping the phylogenetically informative positions revealed that, for certain branches, there are six times fewer apomorphies in RAG1. However, the consistency index of these positions is higher compared to the mitochondrial ND2 gene. Nuclear data therefore provide, on a per-base pair basis, less but more reliable phylogenetic information. Even if in our case RAG1 has not provided as

much phylogenetic information as we expected, we suggest that this marker might be useful in the resolution of the phylogeny of older groups.

Key words: Adaptive radiation — Cichlid species flocks — Explosive speciation — Nuclear DNA phylogeny — NADH Dehydrogenase Subunit II — RAG1 — C-lineage

Introduction

The Great Lakes of East Africa are among the world's most diverse freshwater ecosystems. In particular, Lakes Victoria, Malawi, and Tanganyika (Fig. 1) contain unique species flocks of cichlid fish including hundreds of endemic species each (Fryer and Iles 1972). The sheer number of species, the degree of ecological and morphological specialization, and the rapidity of lineage formation characterize the East African cichlid assemblages and make them well-known model systems for the study of adaptive radiations and explosive speciation (Fryer and Iles 1972; Stiassny and Meyer 1999; Kornfield and Smith 2000; Kocher et al. 2004; Salzburger and Meyer 2004).

Lake Tanganyika is the oldest of the Great Lakes and—with an estimated age of 9 to 12 million years (Cohen et al. 1993, 1997)—the second oldest lake in the world, surpassed only by Lake Baikal. An estimated 250 endemic species of cichlids occur in Lake

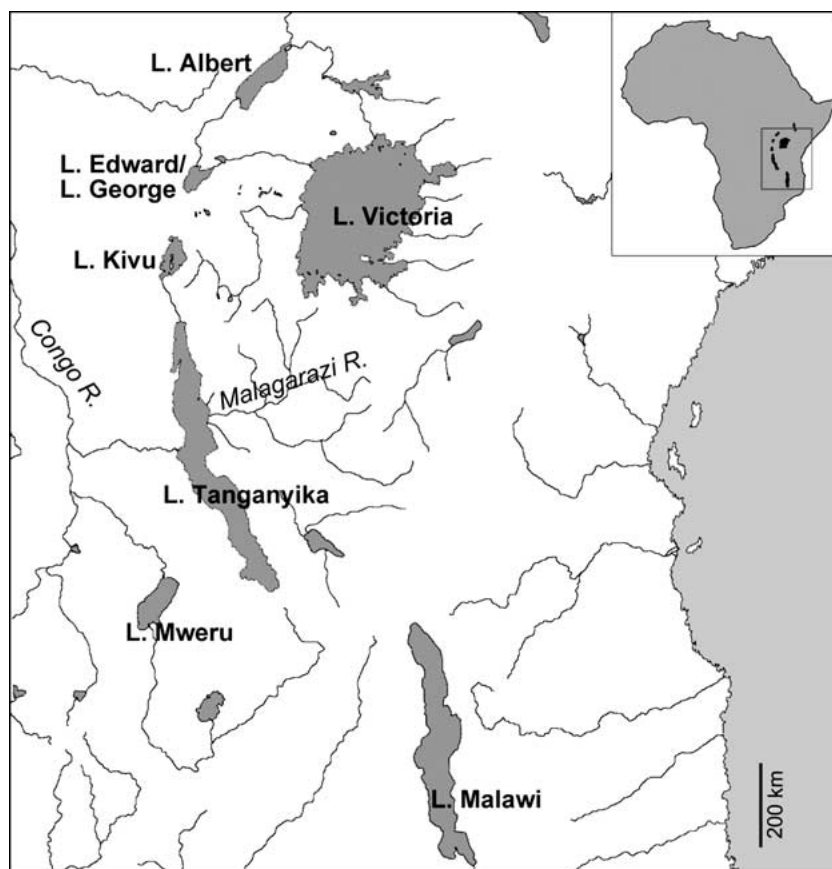


Fig. 1. Map of East Africa showing the lakes and the main river systems of that region.

Tanganyika (Snoeks et al. 1994; Turner et al. 2001). Although this number lies below the estimates for Lake Victoria (~500) and Lake Malawi (~1000), the Lake Tanganyika cichlids are morphologically, ecologically, and behaviorally the most diverse. Poll (1986) recognized 12 distinct tribes based on morphological characters. Eight of these tribes are endemic to the lake and—compared to other cichlid species flocks—relatively species poor: Eretmodini (4 species), Cyprichromini (6), Perissodini (9), Bathybatini (8), Trematocarini (10), Limnochromini (13), Tropheini (24), and Ectodini (32). The remaining four tribes (Haplochromini, Tilapiini, Lamprologini, and Tylochromini) have representatives that occur in other locations in Africa and three of them are the most species-rich tribes of cichlids. Poll (1986) assigned four cichlid genera of Lake Tanganyika and its surrounding rivers to the Haplochromini (~1800 species), which are distributed throughout the African continent but virtually absent from West Africa (Greenwood 1981) and include the species flock of Lake Malawi and the Lake Victoria region superflock (Greenwood 1973, 1980; Verheyen et al. 2003; Salzburger et al. 2005). Also, two species of Tilapiini are found in Lake Tanganyika. The most species-rich tribe of cichlids from Lake Tanganyika, the Lamprologini (79 species), includes at least five represen-

tatives that are found outside the lake in the Congo River system and in the Malagarazi River (Sturmbauer et al. 1994; Schelly and Stiassny 2004; Schelly et al. 2005) (see Fig. 1). Finally, the Tylochromini, which have their center of diversity in West Africa (Stiassny 1990), are represented in Lake Tanganyika by a single species. Recently, Takahashi (2003) suggested the erection of five additional tribes for genera so far assigned to one of Poll's (1986) 12 tribes. Based on internal and external morphological features, Takahashi (2003) elevated the endemic genera *Benthochromis*, *Boulengerochromis*, *Cyphotilapia*, "*Ctenochromis*" *benthicola*, and *Greenwoodochromis* to tribal status.

The reconstruction of the phylogeny of the East African cichlids in general, and that of the species flocks of the three Great Lakes in particular, remains a challenge, despite considerable effort by both morphological as well as molecular phylogeneticists (see, e.g., Poll 1986; Nishida 1991; 1997; Kocher et al. 1995; Albertson et al. 1999; Shaw et al. 2000; Salzburger et al. 2002a, 2005; Takahashi 2003; Allender et al. 2003). Due to the extreme pace of lineage formation and the relatively young age of at least some of these species flocks, the analysis of commonly used markers such as mitochondrial DNA (mtDNA) sequences was still not able to resolve all phylogenetic issues. For example, the

persistence of ancestral polymorphisms as a consequence of incomplete lineage sorting has been suggested to occur, in particular, in the extremely young and therefore closely related cichlid species of Lakes Malawi (Moran and Kornfield 1995) and Victoria (Nagl et al. 1998; Verheyen et al. 2003; Abila et al. 2004). Recent attempts to overcome this problem include the analysis of AFLP markers, which have successfully been applied to a subset of species from Lakes Malawi and Victoria (Albertson et al. 1999; Allender et al. 2003; Seehausen et al. 2003). The phylogenetic problems are different in the older Lake Tanganyika, where shared haplotypes between species are not known (Sturmbauer et al. 2003; Koblmüller et al. 2004)—except as a result of hybridization (Rüber et al. 2001; Salzburger et al. 2002b).

For Lake Tanganyika's species assemblage, morphological characters have, generally, proven to be useful for the assignment of species into tribes (Poll 1986; Takahashi 2003). However, the use of the same characters for phylogenetic purposes may be misleading. On one hand, a large amount of variation exists within tribes, while, on the other hand, convergence in ecomorphological traits and coloration patterns appears to be common between groups that are only distantly related (Kocher et al. 1993; Meyer 1993a). These potential problems and difficulties with morphological data argue for the usage of molecular data to resolve the phylogenetic relationships within Lake Tanganyika's cichlid assemblage.

Several phylogenetic hypotheses exist for the Lake Tanganyika species assemblage based on molecular markers. Nishida (1991) used allozyme data to distinguish several more ancient Tanganyikan tribes (Tylochromini, Tilapiini, Bathybatini, Trematocarini, Lamprologini) from the derived "H-lineage"—a clade combining the tribes Perissodini, Limnochromini, Ectodini, Eretmodini, Cyprichromini, Haplochromini, and Tropheini. These allozyme data (Nishida 1991) already revealed a close relationship between the Tropheini and the Haplochromini, which was confirmed in all later DNA sequence-based studies (see, e.g., Sturmbauer and Meyer 1993; Kocher et al. 1995; Salzburger et al. 2002a, 2005; Verheyen et al. 2003) and combined analyses of allozyme and mtDNA data sets (Nishida 1997). Many studies on the phylogeny of Lake Tanganyika cichlids are based on mitochondrial DNA sequences—e.g., Kocher et al. (1995) focused on the ND2 gene and Salzburger et al. (2002a) on combined DNA sequences of the control region, the cytochrome *b*, and the ND2 gene—confirmed the existence of several ancient groups in Lake Tanganyika with the placement of representatives of the Tylochromini, Tilapiini, Bathybatini and Trematocarini (only in Salzburger et al. 2002a) as sister group to all remaining tribes. In contrast to Nishida (1991), however, the Eretmodini were not resolved

within the "H-lineage" (*sensu* Nishida 1991) but were placed as sister group to the Lamprologini (Kocher et al. 1995) or as sister group to the Lamprologini plus the remaining tribes of Nishida's (1991) original "H-lineage" instead (Salzburger et al. 2002a; see also Takahashi 2003). Kocher et al. (1995), Nishida (1997), and Salzburger et al. (2002a) already found that *Cyphotilapia frontosa* (originally assigned to the Tropheini by Poll [1986]) should be assigned to its own new tribe (see Takahashi 2003). Also, the insertion patterns of short interspersed nuclear elements (SINEs) in Lake Tanganyika cichlids have been used as phylogenetic markers (Takahashi et al. 1998, 2001; Terai et al. 2004), which confirmed the prior phylogenetic hypotheses based on mitochondrial DNA. In an extensive analysis, Takahashi et al. (2001) found incongruent insertion patterns in some of the SINEs, which they interpreted as signature of ancient incomplete lineage sorting in the course of the primary radiation of the Lake Tanganyika cichlids. However, since SINE insertions are evolutionarily rare events in rapidly evolving lineages, the resulting phylogenies often lack sufficient resolution. So far, no phylogenetic study on Lake Tanganyika's cichlid species flock based on nuclear DNA sequences has been done.

We conducted a comparative phylogenetic analysis of Lake Tanganyika cichlids based on both mitochondrial (1047 bp of the ND2 gene) and nuclear (~1800 bp of the RAG1 gene including exon and intron sequences) DNA. The analyses were performed on 43 ingroup taxa. At least one representative of each of the 16 tribes (Takahashi 2003) was included in this study, except for the Benthochromini, the Greenwoodochromini, and *Ctenochromis benthicola*, for which no tissue was available. Several riverine species and species of the Lake Malawi species flock have been included (see Table 1).

Recently nuclear DNA (ncDNA) sequences have increasingly been used for phylogenetic studies of more closely related taxa (see, e.g., Koepfli and Wayne 2003). The combination of ncDNA with mtDNA is advantageous, as these two kinds of markers belong to different linkage groups (Brower et al. 1996) and are inherited in different ways. One advantage of ncDNA compared to mtDNA is the generally observed reduced level of homoplasy among more distantly related taxa as a consequence of the slower rate of evolution (see, e.g., Brown et al. 1982; Avise 1994). Due to those differences in mutation rates between ncDNA and mtDNA data sets, we decided to apply two different phylogenetic algorithms that take those differences into account: maximum likelihood (ML) and Bayesian inference (BI). Bayesian phylogenetic inference has recently been shown to deal efficiently with complex models of molecular evolution (Nylander et al. 2004).

Table 1. Characterization of the studied species of Lake Tanganyika cichlids: Species names follow the fishbase Web site nomenclature (www.fishbase.org), and tribe assignments follow the nomenclature of Takahashi (2003)

Taxonomy information		Voucher		GenBank accession No.		
Tribe	Taxon	Sampling locality	No. (institution)	ND2	RAG1 intron 2	RAG1 exon 3
	<i>Heterochromis multidentis</i>	West Africa	n/a	AF398214	—	—
Bathybatini	<i>Bathybates sp.</i>	Lake Tanganyika	T1992-270B (1)	U07239	DQ012183	DQ012225
Cyphotilapiini	<i>Cyphotilapia frontosa</i>	Lake Tanganyika	AM-139 (2)	U07247	DQ012176	DQ012219
Cyprichromini	<i>Cyprichromis leptosoma</i>	Lake Tanganyika	T1995-1822B (1)	AF398224	DQ012192	DQ012223
	<i>Paracyprichromis brieri</i>	Lake Tanganyika	AM-53E (2)	AF398223	—	DQ012212
Ectodini	<i>Aulonocranus dewindti</i>	Lake Tanganyika	AM-8.3 (2)	AY337782	DQ012198	DQ012240
	<i>Cunningtonia longiventralis</i>	Lake Tanganyika	AM-90-16VIII (2)	AY337780	DQ012201	DQ012243
	<i>Cyathopharynx furcifer</i>	Lake Tanganyika	AM-90-16VIII14 (2)	AY337781	DQ012199	DQ012241
	<i>Ectodus descampsi</i>	Lake Tanganyika	AM-90-II7 (2)	AY337790	DQ012190	DQ012232
	<i>Enantiopus melanogenys</i>	Lake Tanganyika	AM-90-IV (2)	AY337770	DQ012196	DQ012238
	<i>Grammatotria lemairii</i>	Lake Tanganyika	AM-71 (2)	AY337787	DQ012200	DQ012242
	<i>Ophthalmotilapia nasuta</i>	Lake Tanganyika	T1992-414B (1)	AY337783	DQ012197	DQ012239
	<i>Xenotilapia ochrogenys</i>	Lake Tanganyika	T2004-4D8 (2)	AY337767	DQ012179	DQ012221
Eretmodini	<i>Eretmodus cyanostictus</i>	Lake Tanganyika	AM-75 (2)	AF398220	DQ012194	DQ012236
	<i>Spathodus erythrodon</i>	Lake Tanganyika	T1992-740 (1)	AF398218	DQ012175	DQ012218
Haplochromini	<i>Astatoreochromis alluaudi</i>	Lake Kanyaboli	R173 (2)	AY930071	—	DQ012217
	<i>Astatotilapia burtoni</i>	Lake Tanganyika	T2004-419 (2)	AF317266	DQ012203	DQ012245
	<i>Haplochromis paludinosus</i>	Nanganga River	AM-T2 (2)	AY930107	DQ012191	DQ012233
	<i>Melanochromis auritus</i>	Lake Malawi	AM-29T (2)	AY930069	DQ012177	DQ012220
	<i>Metriaclima zebra</i>	Lake Malawi	AM-30T (2)	U07263	DQ012207	DQ012249
	<i>Orthochromis malagaraziensis</i> ^a	Malagarazi River	AM-T5 (2)	AF398232	DQ012187	DQ012229
	<i>Orthochromis uvinzae</i> ^a	Malagarazi River	AM-TZ94-112b (2)	AY930048	DQ012172	DQ012214
	<i>Pseudocrenilabrus multicolor</i>	Lake Kanyaboli	R082 (2)	AY930070	DQ012173	DQ012215
Lamprologini	<i>Altolamprologus compressiceps</i>	Lake Tanganyika	AM-90-16VIIIIN (2)	AF398229	DQ012171	DQ012213
	<i>Julidochromis ornatus</i>	Lake Tanganyika	AM-31T (2)	AF398230	DQ012195	DQ012237
	<i>Lamprologus congoensis</i>	Congo River	AM-307 (2)	AF317272	DQ012188	DQ012230
	<i>Lamprologus cylindricus</i>	Lake Tanganyika	AM-Lc (2)	DQ093115	DQ012209	DQ012251
	<i>Lamprologus teugelsi</i> ^b	Congo River	AM-Lt (2)	AF398225	DQ012184	DQ012226
	<i>Neolamprologus calliurus</i>	Lake Tanganyika	AM-Nc (2)	AF398227	DQ012205	DQ012247
	<i>Neolamprologus leleupi</i>	Lake Tanganyika	AM-Nl (2)	DQ093113	DQ012206	DQ012248
Limnochromini	<i>Limnochromis auritus</i>	Lake Tanganyika	CS-2320 (3)	AF398216	DQ012204	DQ012246
Perissodini	<i>Perissodus microlepis</i>	Lake Tanganyika	T1995-1449B (1)	AF398222	DQ012202	DQ012244
	<i>Plecodus straeleni</i>	Lake Tanganyika	T2004-4H9 (2)	AF398221	DQ012178	—
Tilapiini	<i>Boulengerochromis microlepis</i>	Lake Tanganyika	T2004-1H1a (2)	U07240	DQ012193	DQ012235
	<i>Oreochromis tanganicae</i>	Lake Tanganyika	T2004-3F4 (2)	AF317240	DQ012181	DQ012223
Trematocarini	<i>Trematocara unimaculatum</i>	Lake Tanganyika	T2004-3G3 (2)	AF317268	DQ012185	DQ012227
Tropheini	<i>Ctenochromis horei</i>	Lake Tanganyika	T1995-547B (1)	AY930100	DQ012208	DQ012250
	<i>Limnotilapia dardennii</i>	Lake Tanganyika	T2004-1A6 (2)	DQ093109	—	DQ012211
	<i>Lobochilotes labiatus</i>	Lake Tanganyika	T1995-1403B (1)	U07254	—	DQ012210
	<i>Petrochromis polyodon</i>	Lake Tanganyika	T2004-1D8 (2)	AY930068	DQ012174	DQ012216
	<i>Simochromis babaulti</i>	Lake Tanganyika	T1992-1153 (1)	DQ093110	DQ012182	DQ012224
	<i>Simochromis diagramma</i>	Lake Tanganyika	T1992-312 (1)	AY930087	DQ012186	DQ012228
	<i>Tropheus duboisi</i>	Lake Tanganyika	AM-T77 (2)	AY930085	DQ012180	DQ012222
Tylochromini	<i>Tylochromis polylepis</i>	Lake Tanganyika	AM-341 (2)	AF398215	DQ012189	DQ012231

Note. Voucher specimens are available from (1) The Royal Museum of Central Africa, Tervuren, Belgium; (2) Axel Meyer, Lehrstuhl für Zoologie und Evolutionsbiologie, Department of Biology, University of Konstanz, Konstanz, Germany; and (3) Christian Sturmbauer, Institute of Zoology, University Graz, Graz, Austria.

^a The Malagarazi River *Orthochromis* are classified as Haplochromini here, although they should be placed in a new tribe (see this study and Salzburger et al. 2005).

^b Note that the newly described *L. teugelsi* (Schelly and Stiassny 2004) was identified in previous studies (e.g., Salzburger et al. 2002a) as *L. mocquardi*.

We also applied three strategies to quantify and qualify the phylogenetic signal present in the three partitions (ND2, intron 2 of RAG1, and exon 3 of RAG1): (i) the likelihood mapping method (Strimmer and von Haeseler 1996, 1997) implemented in the program TREE-PUZZLE 5.0 (Schmidt et al.

2002); (ii) the mapping of phylogenetic informative positions and autapomorphies on the obtained phylogenies and the determination of the consistency index (Kluge and Farris 1969) for each variable position; and (iii) the calculation of the sequences divergence between pairs of taxa, to



Fig. 2. Organization of the RAG1 gene showing the positions of the primers used in this study. Primer HBF3 was published previously (Brinkmann et al. 2004); primers CF1, CR1, CR5, and Kali F1 were designed for this study.

visualize and quantify the rate of molecular evolution, as well as to establish the relative contribution of the different partitions of the data set to the pairwise distances.

Materials and Methods

Molecular Methods

We determined DNA sequences of an ~1800-bp segment of the nuclear RAG1 gene of 43 specimens of Lake Tanganyika cichlids representing 14 tribes (according to Takahashi et al. 2003). The RAG1 segment included 692 bp of intron 2 and 1111 bp of exon 3 (Fig. 2). DNA was extracted from ethanol preserved fin-clips or muscle tissue using a high-salt extraction followed by an ethanol precipitation (Brufford et al. 1998). Polymerase chain reaction (PCR) amplification was performed according to standard methods (12 μ l of HPLC water, 2.5 μ l of each primer at 10 μ M, 2.5 μ l of the reaction mix, and 2.5 μ l of dNTP at 10 μ M). Two pairs of primers have been used in order to obtain two overlapping PCR products. For the sequence containing the second RAG1 intron the primers were (forward) HBF3 5'-AARGGGGGACGICCNCHICAGC-3' (Brinkmann et al. 2004) or KaliF1 5'-AAGGGTTTATGTTCAATCAA-3' and (reverse) CR1 5'-AGGGCTGGAATATCTGGCGG-3' (both designed for this study). For the sequence exclusively situated in the exonic region, the primers used were (forward) CF1 5'-GCCGCCAGAT CTTCCAGCCCT-3' and (reverse) CR5 5'-TGCGGGCG TAGTTTCCATTCA-3' (both newly designed). The PCR products were purified with Qiaquick spin columns (Qiagen) following the manufacturer's protocol. If necessary, PCR products were cloned using the TA cloning kit (Invitrogen). Sequencing reactions were performed according to standard methods for all five primers using the Big Dye sequencing chemistry v3.0 (Applied Biosystems). The DNA sequences were determined on an ABI 3100 capillary sequencer (Applied Biosystems). A list of specimens included in this study, their assignment to 1 of the 16 tribes according to Takahashi (2003), and the GenBank accession numbers of the DNA sequences are listed in Table 1. Voucher specimens are deposited in the Royal Museum for Central Africa in Tervuren or are available from the authors.

Phylogenetic Analyses

DNA sequences were aligned with the computer programs Sequencher (GeneCodes) and ClustalX. Seventeen gaps encoded as indels had to be included in the intron region of RAG1 (including an 11-bp-long deletion), and two triplets of gaps had to be included in the exon. We also analyzed 44 sequences of the mitochondrial NADH Dehydrogenase Subunit II gene (ND2; 1047 bp) available from GenBank (see Table 1) (Koblmüller et al. 2004; Salzburger et al. 2002a, 2005). Phylogenetic analyses were performed on several data partitions: ND2, RAG1, and a combined data set. For four specimens it was not possible to amplify the intron of RAG1, and for two specimens we encountered the same problem with the exon of RAG1. Since the resulting missing data represent less than a third of the combined sequence, we decided to include these taxa in the phylogenetic analyses. Also, this decision seems justified based on the observation that char-

acters that are complete (i.e., scored across all taxa, as here for the ND2 sequence) increase accuracy much more effectively in highly incomplete taxa than do characters that are scored in only some taxa (Wiens 2003).

For phylogenetic reconstruction we used ML (with PAUP* 4.0b10 [Swofford 2002]) and BI (with Mr. Bayes [Ronquist and Huelsenbeck 2003]) methods. *Heterochromis multidentis* and *Tylochromis polylepis* were declared as outgroup (according to Stiasny 1990; Lippitsch 1995; Farias et al. 2000; Salzburger et al. 2002a) for the analyses of the combined and the ND2 data set, resulting in a total of 44 taxa; for the RAG1 data set, only *Tylochromis polylepis* was used as single outgroup, resulting in a total of 43 taxa. This was necessary because *Heterochromis* was among the taxa for which no RAG1 sequence could be amplified.

We ran the Modeltest 3.06 routine (Posada and Crandall 1998) to determine, with a hierarchical likelihood ratio test, the appropriate model of molecular evolution for ML analyses. For each data set, a model was chosen from a set of 56 models. We used the GTR+I+G model (Rodriguez et al. 1990) for the ND2 and for the RAG1 data set, and the HKY+I+G model (Hasegawa et al. 1985) for the combined data set (see Table 2 for model parameters). The ML tree of the combined data set was calculated with PAUP* 4.0b10 (Swofford 2002) using an iterative approach. After an initial heuristic search (default parameters) the model parameters were reestimated based on the obtained phylogeny with the "likelihood score" option in PAUP* and used for a new cycle of heuristic searches. This procedure was repeated until the $-\ln$ likelihood did not change anymore in two consecutive runs (see Swofford and Sullivan 2003). We then performed a bootstrap analysis with 100 replicates. BI analyses were performed with MrBayes 3.0 (Huelsenbeck and Ronquist 2001), which uses a Markov chain Monte Carlo (MCMC) approach for sampling the joint posterior probability distributions. We ran four MCMCs in parallel for 10 million generations, sampling trees every 10 generations and excluding 10% of all trees as burn-in. The different partitions were taken into account in the combined data set. For the ND2 and the RAG1 data sets, we performed a single ML heuristic search and the BI was run for 1 million generations (tree sampling every 10 generations; burn in = 10%).

To assess the support for crucial branches in the RAG1, we used the one-tailed Shimodaira-Hasegawa (SH) topology test using RELL bootstraps with 1000 replicates as implemented in PAUP* (Shimodaira and Hasegawa 1999; see also Felsenstein 2004). We constructed a tree in which the monophyly of the *Orthochromis* species was enforced and compared it to the best unconstrained tree.

Finally, we assessed the heterogeneity in phylogenetic signal (i.e., incongruence) among the individual data sets and partitions. We used the level of bootstrap support (BS) in ML trees to measure the degree of conflict among topologies recovered in the separate analyses as well as the posterior probabilities (PP) values estimated by the Bayesian analysis. We also performed a likelihood mapping analysis with TREE-PUZZLE 5.0 (Schmidt et al. 2002), to visualize the phylogenetic signal in the combined data set as well as in the mitochondrial and nuclear partition independently.

Analysis of Character Evolution

For the reconstruction and comparison of character evolution in the nuclear and mitochondrial gene segments, we used the maxi-

Table 2. Models defined by Model Test for the different partitions and the model used for the combined data set after an iterative optimization (see Materials and Methods for further details)

Data set	ND2 alone	RAG1 alone	Combined	Iterative
Best-fitting model/estimates	GTR+I+G	GTR+I+G	HKY+I+G	HKY+I+G
Frequency A	0.297	0.2462	0.2525	0.2517
Frequency C	0.3795	0.2279	0.2945	0.2943
Frequency G	0.0846	0.2412	0.1878	0.1873
Frequency T	0.2389	0.2847	0.2652	0.2667
Transition/transversion ratio			2.9883	3.0086
Gamma shape parameter	0.9192	0.7920	0.5625	0.5689
Proportion of invariable sites	0.3565	0.3225	0.4416	0.4396
Number of substitution types	6	6	2	2
κ				
A–C	0.4049	0.8280	n/a	n/a
A–G	11.3001	1.5762	n/a	n/a
A–T	0.6488	0.7101	n/a	n/a
C–G	0.8348	1.4702	n/a	n/a
C–T	4.2669	2.5227	n/a	n/a
G–T	1	1	n/a	n/a

mum likelihood phylogeny of the combined data set and mapped the parsimonious informative positions and autapomorphies on the respective branches. We mapped those positions by using the “list of apomorphies” command in the “describe tree” menu in PAUP* (Swofford 2002) using the accelerated transformation (ACCTRAN) character-state optimization under default settings, with gaps handled as missing data. We differentiated between the source DNA segment of each informative position (ND2, intron 2 of RAG1, or exon 3 of RAG1), in order to quantify the relative contribution of each DNA segment to the phylogenetic signal in the entire combined data set. We also calculated the consistency index (CI [Kluge and Farris 1969]) for each informative position. We then compared the number of parsimonious informative positions and the CI at the level of deep nodes to those in the Ectodini and the Lamprologini clade, which were chosen as examples for monophyletic groups derived from these deep nodes.

Pairwise Genetic Distances Comparison

We calculated with PAUP* (Swofford 2002) the pairwise distances in the different partitions (ND2, exon 3 of RAG1, intron 2 of RAG1, and the combined RAG1 sequences) as well as for the entire combined data set, using the HKY+G+I model of molecular evolution and the model parameters estimated from the combined data set topology. After excluding those pairs of taxa involving a segment of missing data, we plotted for each taxon pair the corrected pairwise distances of ND2 vs. RAG1 and intron 2 of RAG1 vs. exon 3 of RAG1 in order to test for differences in the rate of molecular evolution between the different partitions. The pairwise distance ratios of RAG1/ND2 and intron 2 of RAG1/exon 3 of RAG1 for each pairwise comparison were then used to calculate the average pairwise distance ratios for both types of comparisons. From these ratios we estimated the length of nuclear DNA segment that would have to be sequenced in order to reach a similar level of divergence that is present in the mtDNA. We also performed such estimations for the two partitions of the ncDNA. We then divided our data set into five taxon clusters according to the results of the phylogenetic analyses (see Figs. 3–5) in order to highlight differences in the ND2 vs. RAG1 pairwise distances. This partition into five taxon clusters was used to compare distances within each of the five clusters (closely related taxa) to those among the five clusters (distantly related taxa). Finally, we grouped taxon clusters I, II, and III (H-lineage *sensu* Salzburger et al. [2002a] or “C-lineage” as defined here) and compared their ND2 vs. RAG1

pairwise distances with those of groups IV and V plus *Tylochromis polylepis* (“remaining taxa”) in order to delineate the relative contribution of mtDNA vs. ncDNA variation at different phylogenetic “depths.”

Results

In the RAG1 sequences, 154 heterozygous sites in the 1803-bp alignment of the 43 species were scored as polymorphisms on the basis of equal peaks heights in electropherograms (27 as Y, 31 as M, 38 as R, 16 as K, 34 as S, and 8 as W). We also noticed that the intron contains a deletion of 11 bp, found in *As-tatotilapia burtoni*, *Metriaclima zebra*, *Haplochromis paludinosus*, *Ctenochromis horei*, *Melanochromis auratus*, and *Pseudocrenilabus multicolor* (Haplochromini) and *Simochromis diagramma*, *Simochromis babaulti*, *Tropheus duboisi*, and *Petrochromis polyodon* (Tropheini). The exon of *M. auratus* shows an additional codon (CAA between position 106 and position 108 in our alignment), whereas *Cyphotilapia frontosa* lacks a codon (AGG between position 544 and position 546). The four-cluster likelihood mapping analyses revealed a percentage fraction of 18.6% unresolved quartet topologies for the RAG1 data set (64.2% fully resolved) and 1% unresolved quartet topologies for the ND2 data set (95.7% fully resolved). In the combined data set, only 0.2% of the quartet topologies were found to be unresolved (97.8% fully resolved).

Results of the Phylogenetic Analysis

RAG1

For this analysis, *Tylochromis polylepis* was chosen as outgroup. The maximum likelihood analysis (Fig. 3A) identified a clade containing *Orthochromis uvinzae* and *Oreochromis tanganicae* as sister group to

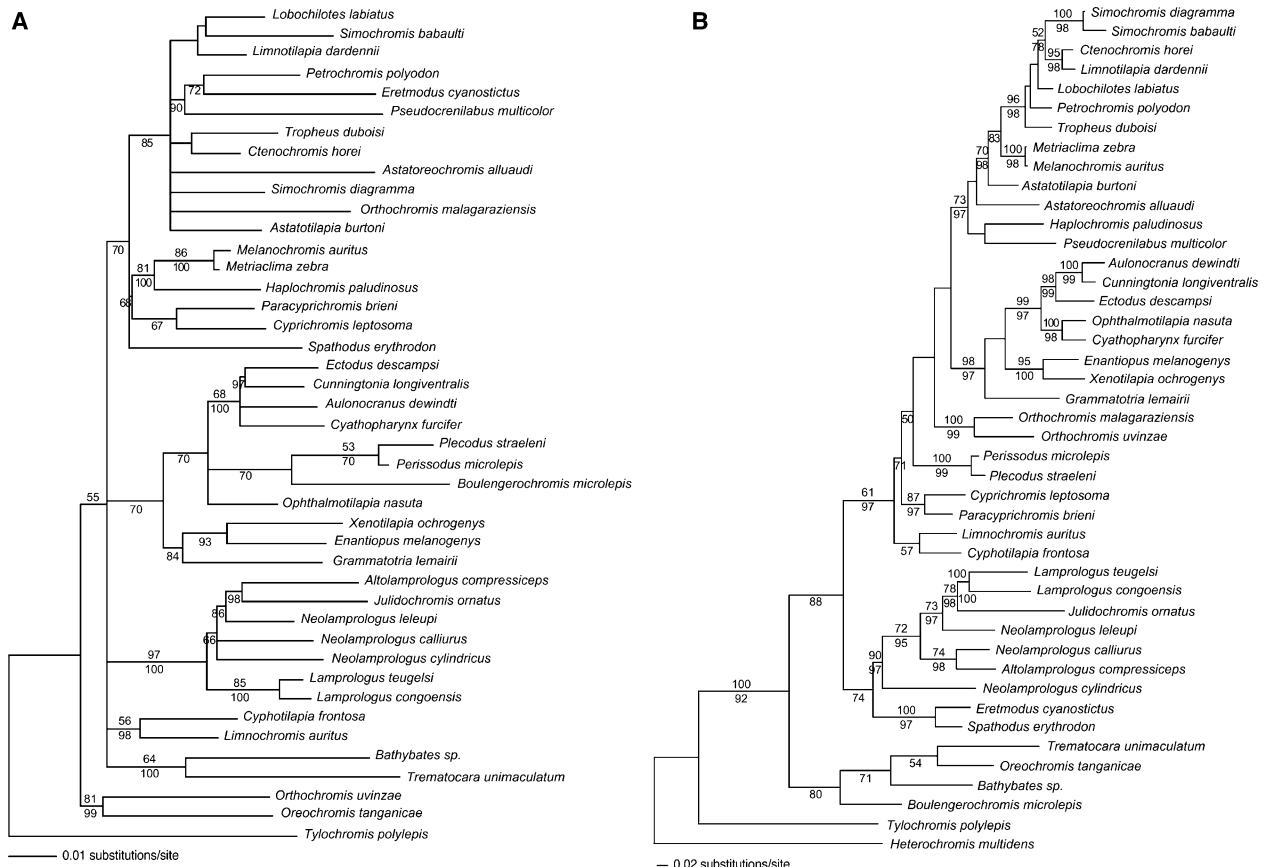


Fig. 3. Maximum likelihood (GTR + I + G) phylogram of Lake Tanganyika cichlids based on RAG1 alone (A) (42 ingroup taxa; *Tylochromis polylepis* as outgroup) and ND2 alone (B) (43 ingroup taxa; *Heterochromis multidens* as outgroup). Numbers above

the branches are ML bootstrap estimates (100 replicates); numbers below the branches are Bayesian posterior probabilities (1 million generations; 10% burn-in). Values < 50% are not shown.

all remaining ingroup taxa. However, the SH test did not reject the monophyly of the two *Orthochromis* representatives when comparing the ML phylogeny to a tree in which these two taxa have been constrained as a clade ($p = 0.097$). The remaining taxa were clustered into five clades forming a polytomy: *Bathybates* sp. plus *Trematocara unimaculatum*, *Limnochromis auritus* plus *Cyphotilapia frontosa*, the Lamprologini, a clade containing the Ectodini plus *Boulengerochromis microlepis* and the Perissodini, and a poorly supported clade including the Haplochromini, the Tropheini, the Eretmodini, the Cyprichromini, and *Orthochromis malagaraziensis* (note that in this analysis, the synapomorphic 11-bp insertion of the Haplochromini was not considered). Only the monophyly of the Lamprologini (BS = 97, PP = 100) was strongly supported; the monophyly of *Bathybates* sp. plus *Trematocara unimaculatum* received moderate support (BS = 64, PP = 100) as well as the monophyly of *Limnochromis auritus* plus *Cyphotilapia frontosa* (BS = 56, PP = 98). The monophyly of the Ectodini/*Boulengerochromis microlepis*/Perissodini clade (PP = 70) and the monophyly of the clade comprised by the Haplochromini the Tropheini, the Eretmodini,

the Cyprichromini, and *Orthochromis malagaraziensis* (PP = 70) were supported only in the Bayesian inference phylogeny.

ND2

For the analysis of the ND2 sequences, *Heterochromis multidens* was used as outgroup. In the ML (Fig. 3B) and BI trees, *Tylochromis polylepis* occupied the most ancestral position, followed by a clade formed by *Boulengerochromis microlepis*, *Bathybates* sp., *Oreochromis tanganyicae*, and *Trematocara unimaculatum*. These form the sister group of two major clades, a clade formed by the Eretmodini plus the Lamprologini and a clade including all remaining taxa. The sister group relationship of Eretmodini and Lamprologini was supported by low BS and PP. In the second clade, *Limnochromis auritus* and *Cyphotilapia frontosa* form the most ancestral lineages, sister group to the Cyprichromini, the Perissodini, the *Orthochromis* species, the Ectodini, and the Haplochromini including the Tropheini. The monophyly of the tribes Eretmodini, Lamprologini, Cyprichromini, Perissodini, Ectodini, Tropheini, and Haplochromini (but excluding *Orth-*

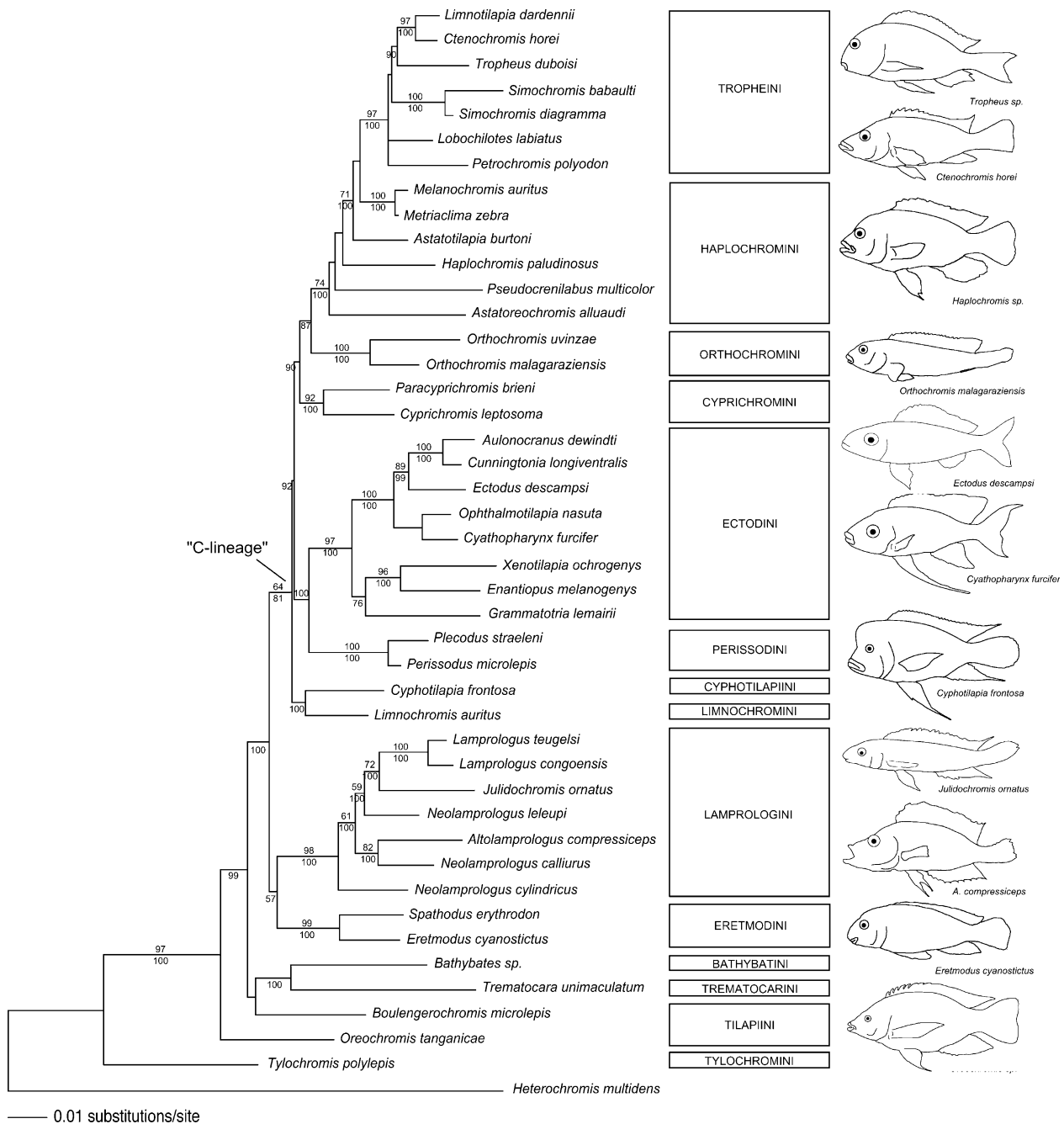


Fig. 4. Maximum likelihood (HKY + I + G) phylogram of Lake Tanganyika cichlids based on ND2 and intron 2 of RAG1 and exon 3 of RAG1 of 44 taxa (declaring *Heterochromis multidens* as outgroup). Numbers above the branches are ML bootstrap esti-

mates (100 replicates); numbers below the branches are Bayesian posterior probabilities (10 million generations; 10% burn-in). Values < 50% are not shown. Names of the corresponding tribes and representatives of some tribes are shown at the right.

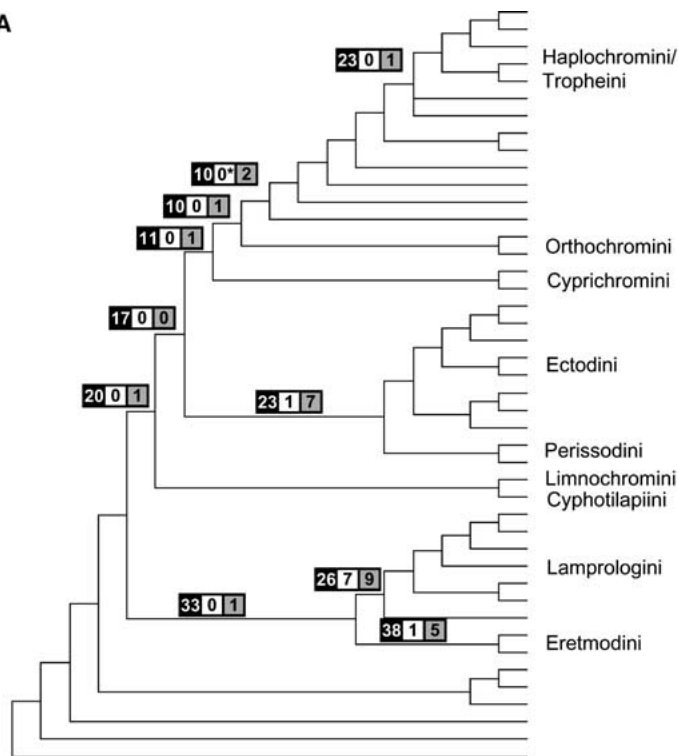
ochromis [see Salzburger et al. 2002a, 2005]) was supported by high BS and PP.

Combined Data Set

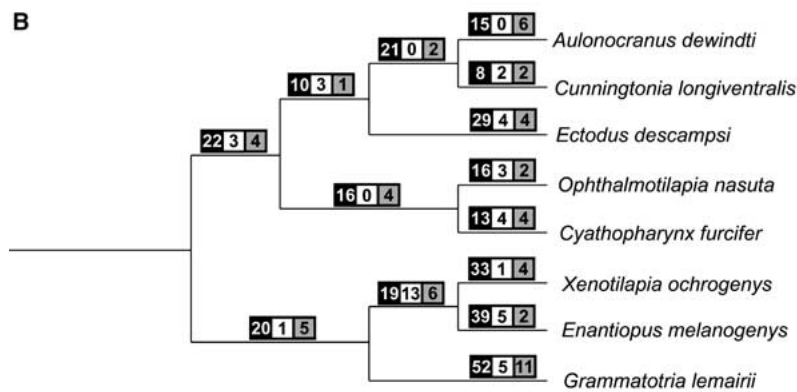
The combined analysis with *Heterochromis multidens* as outgroup led to congruent results in ML and BI analyses (Fig. 4). We obtained *Tylochromis polylepis* as the most ancestral split, placed as sister group to *Oreochromis tanganyicae*, a clade constituted

by *Boulengerochromis microlepis*, *Bathybates sp.*, and *Trematocara unimaculatum* and a group including all remaining taxa. In this group, the Eretmodini and Lamprologini were placed as sister group to a clade formed by *Cyphotilapia frontosa* plus *Limnochromis auritus*, a clade including the Perissodini sister to the Ectodini, a clade formed by the Cyprichromini, the Orthochromis species, and the Haplochromini/Tropheini assemblage (see Fig. 4 for BS and PP).

A



B



C

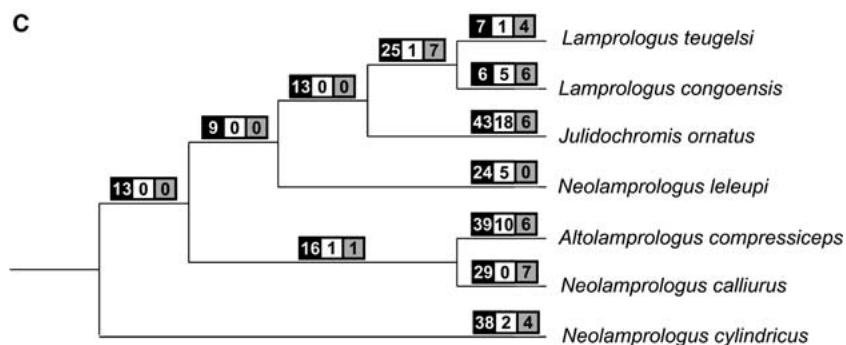


Fig. 5. Mapping of the parsimony informative position contained in the sequences of ND2 (black box), intron 2 of RAG1 (white box), and exon 3 of RAG1 (gray box) on the obtained ML tree (A) (see Fig. 4) as well as on the Ectodini clade (B) and the Lamprologini clade (C). The values on terminal branches in B and C represent autapomorphies of the respective taxon.

Mapping of Molecular Character Evolution

Of the 1047 bp of ND2, 415 characters were parsimony informative (39.64%). In the ncDNA (1803 bp), only 119 characters were parsimony informative (6.60%). Thus, about six times more informative

positions were contained in the mtDNA compared to the ncDNA data set (length corrected). However, the CI of the tree of the combined data set was higher when only the ncDNA positions are taken into account (CI = 0.751) than when the changes of the mtDNA alone are mapped on the combined tree

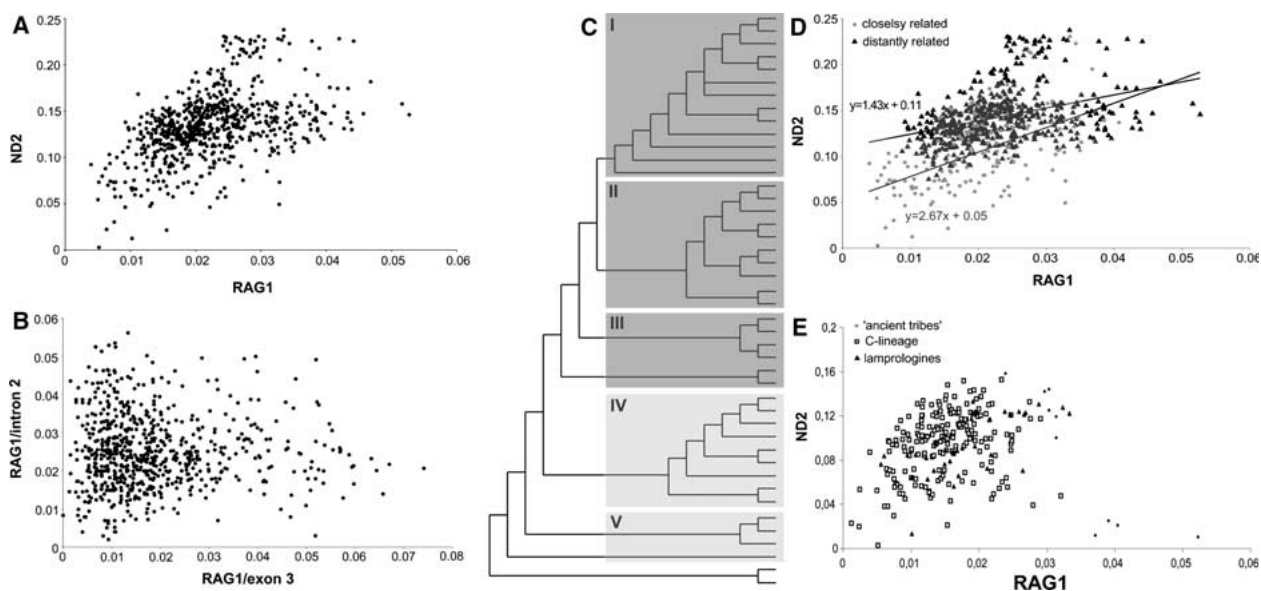


Fig. 6. Pairwise distance comparisons. **A** Pairwise distances of ND2 (y-axis) versus pairwise distances of RAG1 (x-axis) of all possible taxon pairs among 43 cichlid species analyzed (*Heterochromis multidentis* was not used). **B** Pairwise distances of RAG1/intron 2 versus RAG1/exon 3 of all possible taxon pairs among the 43 cichlids. **C** Definition of five clusters of taxa based on the ML topology of the combined data set (see Fig. 4). Clusters I–III are referred to as “C-lineage”; clusters IV and V, as “ancient tribes.” **D**

Pairwise distance comparison (ND2 vs. RAG1) among closely (within groups I–V) and distantly (between groups I–V) related taxa. The trend-lines for the two groups of plots are depicted. **E** Pairwise distance comparison (ND2 vs. RAG1) among “ancient tribes” and “young lineages.” Taxon pairs involving two lamprologine representatives are highlighted. This cloud of plots, although belonging to a more ancient tribe, is similar to those of the C-lineage pairwise comparisons.

(CI = 0.379). The numbers of parsimonious informative positions and autapomorphies are depicted on relevant branches in Fig. 5.

At the level of deeper nodes (Fig. 5A), PAUP* assigned 9.82 times more apomorphic characters in the mtDNA partition (211 of 1047bp) than in the ncDNA partition (37 of 1803bp). However, by calculating the mean of the consistency index of these apomorphies, we found that the nuclear positions (CI = 0.57) are less homoplasious than the mitochondrial ones (CI = 0.29). At the level of terminal nodes, 315 apomorphies were found in the mtDNA partition of the Ectodini, compared to 101 in the ncDNA (Fig. 5B), and 262 apomorphies were found in the mtDNA of the Lamprologini, compared to 84 in the ncDNA (Fig. 5C). Here, the mtDNA provides 5.37 (for the Ectodini and for the Lamprologini) times more informative positions compared to the ncDNA (length corrected). Again, the nuclear positions are less homoplasious (CI = 0.79 for the Ectodini, 0.79 for the Lamprologini) than the mitochondrial ones (CI = 0.38 for the Ectodini, CI = 0.39 for the Lamprologini).

Pairwise Distance Comparisons

The corrected pairwise distances in the ND2 partition were on average 10.17 times higher than those in RAG1. In RAG1, the corrected pairwise distances

were found to be about 1.11 times higher in the exon than in the intron. This contradicts the expectation that intron sequences—because of the presumed absence of evolutionary constraints—would be characterized by higher rates of evolution compared to exon sequences.

The plot of corrected pairwise distances of RAG1 vs. ND2 (Fig. 6A) showed clearly that for each taxon pair the corrected pairwise distance is higher in ND2. Pairwise distances in RAG1 ranged from 0.13% (*Ctenochromis horei*–*Limnotilapia dardennii*) to 4.16% (*Petrochromis polyodon*–*Altolamprologus compressiceps*), while they ranged from 0.29% (*Metriaclima zebra*–*Melanochromis auritus*) to 20.39% (*Tylochromis polylepis*–*Simochromis babaulti*) for the ND2 gene. The plot of corrected pairwise distances of intron 2 of RAG1 vs. exon 3 of RAG1 (Fig. 6B) showed a more equal distribution around the bisecting line. Pairwise distances in intron 2 of RAG1 ranged from 0% (*Cyathopharynx furcifer*–*Aulonocranus dewindti*) to 5.5% (*Julidochromis ornatus*–*Trematocara unimaculatum*), and those in exon 3 of RAG1, from 0.18% (*Metriaclima zebra*–*Melanochromis auritus*) to 5.85% (*Plecodus straeleni*–*Spathodus erythrodon*).

The ND2 vs. RAG1 pairwise distance comparisons between closely and distantly related taxa (as defined in Fig. 6C) revealed that, on the whole, the ncDNA partition contributed to a similar extent to the pairwise genetic distances at both levels. By calculating

the average of the relative contribution of the RAG1 partition to the total genetic distances over all pairwise comparisons, we found that in the group of closely related taxa, RAG1 contributed on average to 14.99% of the total distance, and in the distantly related taxa we found that RAG1 contributed on average to 14.65% of the total distance. When looking at the trend-lines of those two clusters of plots, we noticed that the longer the total distance was, the higher was the contribution of RAG1 to it. This observation is accentuated for the distantly related taxon pairs (see trend-lines in Fig. 6D). In the pairwise comparisons of the ancestral lineages, RAG1 contributes 39.7% to the total distance (mtDNA plus ncDNA) in the Lamprologini 17.1% and only 14.5% in the C-lineage taxa.

Comparing the C-lineage taxa to the remaining ones (Fig. 6E), we observed a similar trend. The distances among taxon-pairs of the C-lineage are generally shorter compared to those among taxon-pairs in the remaining taxa. Also, the pairwise distances within the Lamprologini (highlighted in Fig. 6E)—which, as a tribe, does not belong to the C-lineage—are equivalent to those found within the C-lineage.

Discussion

Missing Data and Support Values

Support values such as BS and PP are generally considered to be good indicators of the robustness of branches in phylogenetic trees. Unfortunately, real data sets do not always result in phylogenies with excellent support values, and biological phenomena such as the rapid lineage formation in adaptive radiations or explosive speciation events are particularly problematic to resolve due to the pace of cladogenesis within a short period of time (see, e.g., Sturmbauer et al. 2003). When nodes are not supported by high bootstrap values (or when branches are not significantly different from a length of zero), they may have arisen by chance alone (Flynn and Nedbal 1998) or reflect the biological reality of an adaptive radiation. In the absence of real contradiction between different topologies obtained from different data sets (e.g., mtDNA and ncDNA) and different phylogenetic algorithms, the resulting topology may be considered to be robust despite only moderate support values. In our case, the topologies obtained from separate as well as combined data sets, and including mtDNA and ncDNA, are congruent for most relevant branches, although some support values are low (Figs. 3 and 4).

The analyses presented were based on the combined mtDNA and ncDNA sequences data set from 42 ingroup taxa. Six of those were partially incom-

plete for the ncDNA partition (four were missing the intron sequence and two the exon sequence). Including incomplete taxa in an analysis is unlikely to be problematic as long as there is a sufficient amount of phylogenetic information in at least one broadly sampled data set to allow the resolution of these taxa in a phylogeny (Bininda-Emonds and Sanderson 2001). As the ND2 tree is not incongruent with the trees obtained from the combined data set, we suppose that ND2 by itself contains enough informative characters to achieve a strongly supported topology (Salzburger et al. 2002b). Wiens (2003) stressed that the estimated relationships among complete sets of taxa are seemingly unaffected by the inclusion of incompletely sampled taxa. This is again confirmed in our analyses by comparing the tree obtained with ND2 alone to the one constructed from the combined data set. In the combined analysis, the six taxa for which no complete RAG1 sequence was available were placed at the same position (or at least within the same clade) as they were in the ND2 tree. What might be affected by the inclusion of incomplete taxa are support values rather than the overall topology.

The Phylogeny of the Lake Tanganyika Cichlid Assemblage

The assemblage of cichlid fish in Lake Tanganyika is older than the species flocks of Lakes Malawi and Victoria (Fryer and Iles 1972; Salzburger et al. 2005), which is reflected in larger genetic differences facilitating the phylogenetic reconstruction of its cichlid lineages. Lake Tanganyika cichlids have so far been included in several phylogenetic studies using different sets of molecular markers. Nishida (1991) used allozymes and identified several ancient lineages of cichlids in Lake Tanganyika, as well as the derived mouthbrooders, which he termed “H-lineage” (see also Nishida [1997] for an analysis combining allozyme and mtDNA data). Later studies based on mtDNA alone (Kocher et al. 1995; Salzburger et al. 2002a) established a relatively clear scenario for the evolution of the Lake Tanganyika cichlid assemblage. The lake seems to harbor several older lineages that colonized the lake after its formation as well as younger tribes that are likely to have evolved in the course of a primary lacustrine radiation within Lake Tanganyika (Salzburger et al. 2002a). Among these tribes are the Haplochromini, which also include the species flocks of Lakes Malawi and Victoria, and it has been suggested that Lake Tanganyika is the cradle of all haplochromine cichlids (Fryer and Iles 1972; Salzburger et al. 2002a, 2005). Also, ncDNA sequence data suggested that haplochromine cichlids are derived from Lake Tanganyika stocks (Mayer et al. 1998), albeit with low bootstrap support. However, the exact relationships among several tribes re-

mained unclear and only a relatively small number of Tanganyikan cichlid species had been analyzed so far, particularly for ncDNA, although phylogenies based on several loci belonging to different linkage groups appear advantageous (see, e.g., Brower et al. 1996; Takahashi et al. 2001; Seehausen et al. 2004).

In our analyses of the mtDNA data set (Fig. 3B) and combining mtDNA and ncDNA (Fig. 4), we corroborate that *Tylochromis polylepis* represents the most ancestral lineage of the species assemblage in Lake Tanganyika, which might have seeded the lake independently (probably only recently) from its presumed West African center of origin, where most of its congeners still can be found (Stiassny 1990). Consistently, a weakly supported Lake Tanganyika endemic clade consisting of *Boulengerochromis microlepis*, *Oreochromis tanganicae*, *Bathybates* sp., and *Trematocara unimaculatum* forms the most ancestral group in the phylogeny of the combined data set. Despite the apparently ancestral position of *O. tanganicae* in our phylogeny, this species is likely to have colonized the lake only recently, since it shows a high level of sequence identity with its sister species *O. niloticus* (Klett and Meyer 2002). This distance is similar to those found between sister species within the Lamprologini or the Ectodini. It seems interesting that this species was able to adapt to an ecological niche in a lake in which a highly specialized species assemblage of cichlids was already present (see also Salzburger et al. 2005).

When considering only ncDNA (Fig. 3A), *Oreochromis tanganicae* and *Orthochromis uvinzae* were placed as the most basal sister group to all remaining taxa. The SH test, however, could not reject the monophyly of the two *Orthochromis* species. *Trematocara unimaculatum* and *Bathybates* sp. form one of the five clades of the basal polytomy, and *Boulengerochromis microlepis* was resolved in another of these clades in the ncDNA partition, together with the Ectodini and Perissodini. While the *Trematocara unimaculatum*/*Bathybates* sp. clade received moderate bootstrap support, the latter clade was not supported by the maximum likelihood bootstrap analysis. Our results thus corroborate the morphology-based assignment of *Boulengerochromis microlepis* in its own tribe, Boulengerochromini (Takahashi 2003), since it has never been found to form a monophyletic clade with *Oreochromis tanganicae* (Klett and Meyer 2002).

In the combined and in the ND2 data set alone, the Eretmodini were—just as found by Kocher et al. (1995) and Salzburger et al. (2002a)—not resolved within the “H-lineage” (*sensu* Nishida 1991) but were placed as sister group of the Lamprologini instead. In the BI phylogeny of the combined data set, the relative position of the Lamprologini and Eretmodini is poorly supported (PP = 57). In the ML and BI analyses of the ncDNA partition, in which an 11-bp insertion

characteristic to the Haplochromini and Tropheini was not taken into account (see below), the two eretmodine taxa were resolved in a clade containing the Haplochromini, Tropheini, and Cyprichromini (see Nishida 1991; Sturmbauer and Meyer 1993). The monophyly of that clade was weakly supported in the BI phylogeny (PP = 70). Therefore, further data and analyses will be necessary in order to assess the exact phylogenetic placement of the Eretmodini with respect to the Lamprologini and the C-lineage.

In all analyses the Lamprologini were recovered as a monophyletic group (BS = 98 and PP = 100 for the combined data sets, BS = 97 and PP = 100 for RAG1 alone). Interestingly, in the topologies based on ncDNA, the riverine lamprologines (*L. teugelsi* and *L. congoensis*) were placed as sister group to all remaining taxa—unlike in all previous mtDNA analyses (Sturmbauer et al. 1994; Salzburger et al. 2002a) as well as in our trees based on the combined data set and ND2 alone (see Figs. 3 and 4). If this seeming conflict would hold after sequencing of additional ncDNA segments, this finding would suggest that the ancestors of riverine lamprologines presently found in the Congo drainage gave rise to the radiation of the most species-rich tribe in Lake Tanganyika, rendering those Congo taxa as ancestral relict species rather than, as suggested before (Sturmbauer et al. 1994), a lineage that emigrated from Lake Tanganyika.

Our combined analyses revealed strong support for the monophyly of what we propose as the “C-lineage,” the clade containing the Cyphotilapiini, Limnochromini, Perissodini, Ectodini, Cyprichromini, genus *Orthochromis*, Haplochromini, and Tropheini (i.e., the “redefined H-lineage” of Salzburger et al. [2002a]). Within the C-lineage, *Limnochromis auritus* and *Cyphotilapia frontosa*, which were also grouped together in the ncDNA partition alone, occupied the most ancestral position in all analyses. Therefore, *Cyphotilapia frontosa* finds its place at the root of to the C-lineage, and not within the Tropheini, where it was placed originally (Poll 1956) based on its *Tropheus*-like pharyngeal apophysis. These results, again, support the exclusion of *Cyphotilapia* from the Tropheini (see also Nishida 1997; Lippitsch 1998; Salzburger et al. 2002a) and the creation of a new tribe, the Cyphotilapiini (see also Salzburger et al. 2002a; Takahashi 2003). In the ML analysis, as well as in the BI phylogeny of the combined data set, the Perissodini were grouped together with the Ectodini, sister to the Cyprichromini, the two species of *Orthochromis*, and the Haplochromini/Tropheini. In the ND2 partition, the Cyprichromini were placed as sister group to the Perissodini, the two *Orthochromis* species, the Ectodini, and the Haplochromini/Tropheini.

The topology of the Ectodini is similar to that found by Kolbmüller et al. (2004) based on mtDNA only. The monophyly of the Haplochromini/Trophe-

ini was supported by high BS and PP. The results of these analyses support the inclusion of *Ctenochromis horei* within the Tropheini, an assignment already proposed by Takahashi et al. (2003) and Sturmbauer et al. (2003). We furthermore identified an 11-bp synapomorphic deletion in intron 2 of RAG1 of all Haplochromini/Tropheini, which strongly supports the monophyly of this group, and its derived position (the lower support in the ncDNA partition is most likely due to the fact that this 11-bp partition was not taken into account in the ML analyses).

The placement of the two species of *Orthochromis* as sister group to the Haplochromini in the combined tree (Fig. 4) contradicts the topology found for the ND2 alone tree (Fig. 3B) and previous findings based on mtDNA (Salzburger et al. 2002a, 2005). However, the Orthochromini do not contain the haplochromine-specific 11-bp deletion in the intron region RAG1, and their sister group relationship to the Haplochromini is only poorly supported. More data would be necessary to clarify that issue. We note that *Orthochromis polyacanthus* from the Congo drainage was previously shown to be a member of the Haplochromini clade and, hence, not related to *O. malagaraziensis* from the Malagarazi drainage (Salzburger et al. 2002a, 2005; see also De Vos and Seegers 1998). It is thus likely that the genus *Orthochromis* is polyphyletic and in need of a taxonomic revision (see also Salzburger et al. 2005).

In the C-lineage, a large number of species arose within a short period of time and are, thus, separated by only small genetic distances. Clearly, there is no correlation between the degree of molecular and morphological variation among the cichlid adaptive radiations (see also Sturmbauer and Meyer 1992; Verheyen et al. 2003). This is the reason why morphology-based phylogenetic studies are unlikely to resolve the relationships among cichlids. Furthermore, this independence of evolution makes it more likely that the propensity for explosive speciation in this group is likely based on genomic features of cichlids rather than on a simple function of time or mutation rate. Therefore, in the future, comparative evolutionary genomic analyses should be facilitated (Santini et al. 2003). The proposed sequencing of cichlid genomes (Kocher et al. 2004) would certainly contribute to a better understanding of the “cichlid problem.”

Contribution of Nuclear DNA

Several previous studies on the phylogeny of East African cichlid species flocks were based on the ND2 gene (Kocher et al. 1995; Shaw et al. 2000; Salzburger et al. 2002a, 2005). This mtDNA segment has been chosen because, in mammals, it has been known to evolve more rapidly than other mitochondrial proteins coding segments (Anderson et al. 1982; Kocher

et al. 1995). In cichlids it has been shown to be adequate for resolving genus-level phylogenies (Kocher et al. 1995), and it has a higher rate of molecular evolution than, for example, the widely used cytochrome *b* gene (Cummings et al. 1995; Russo et al. 1996; Zardoya and Meyer 1996; Salzburger et al. 2002a). Here we determined, for the first time, a ncDNA segment (about 1800 bp of the RAG1 gene, including intron 2 and exon 3) in an effort to add the phylogenetic information provided by more slowly evolving DNA fragments in order to gain resolution at the level of deeper nodes of the Lake Tanganyika cichlid phylogeny. This gene has been widely used in phylogeny reconstruction of other groups of animals (e.g., Venkatesh et al. 1999; Brinkmann et al. 2004; Rüber et al. 2004; Van der Meijden et al. 2004), so that primers were available and the gene has been well characterized (e.g., no paralogues have been found so far). Another advantage of using ncDNA in addition to mtDNA is the possibility of combining molecular information provided by independent linkage groups (Brower et al. 1996).

The four-cluster likelihood mapping analysis already indicated that, as expected, the RAG1 segment provided less phylogenetic signal compared to the ND2 sequences. However, the amount of about 81.4% resolved quartet topologies in RAG1 justified its use for phylogenetic reconstructions of these taxa. Also, when combining the mtDNA and ncDNA sequences, the phylogenetic signal improved so that in the combined data set only 0.2% of the quartet topologies remain unresolved. Thus, using the RAG1 partition alone might not lead to the resolution of the phylogeny as accurately as the combined data set can do (see also Rüber et al. 2004).

For phylogenetic analyses in general, fast-evolving (often mtDNA) genes are useful markers of evolutionary history and, particularly, for the resolution of rapid speciation events. Alternatively, very long stretches of slowly evolving DNA have been used as well. Gene segments that evolve relatively rapidly are likely to contain more homoplasious mutations, whereas more slowly evolving genes have fewer informative sites. Because of the accelerated rate of molecular evolution in mtDNA, a discrepancy in the numbers of informative positions was observed between the two partitions (on average there were six times more informative positions in the mtDNA partition). However, this difference is at least partly compensated for by the higher reliability of the phylogenetic informative positions in the ncDNA, which were less homoplasious (ncDNA, CI = 0.75; mtDNA, CI = 0.38). For mtDNA it is known already that only a relatively small number of informative sites provides resolution on the relationships among different tribes of Lake Tanganyika cichlids (see, e.g., Kocher et al. 1995; Salzburger et al. 2002a). A similar pattern was

found in our analysis of the ncDNA partition, where only a few apomorphies at the deeper node level could be identified. However, these showed a relatively high CI and, thus, seem to provide a relatively robust phylogenetic signal. In particular, a relatively rare mutational event, an 11-bp deletion, unambiguously supported the monophyly of the Haplochromini/Tropheini clade (see Salzburger et al. [2005] for a detailed phylogenetic analysis of the haplochromines). Also, within a cluster of the C-lineage (the Ectodini) and in the Lamprologini (not belonging to the C-lineage), relatively more apomorphies could be detected compared to the level of deeper nodes, with even higher consistency indices (Figs. 5B and C). Again, the ncDNA seems to provide more phylogenetic information within the younger lineages compared to the deeper nodes. This can be explained by the fact that apomorphies accumulated after the radiation that gave rise to the Lake Tanganyikan tribes.

The contribution of the ncDNA relative to mtDNA to the pairwise distances tends to be higher for the pairwise distances calculated among distantly related species (Fig. 6D), suggesting that the ncDNA is actually more informative for resolving the relationship among ancient clades. The trend-line in Fig. 6D shows that the longer the distances between two sequences are, the more RAG1 is contributing to the total distance. RAG1 might thus be more efficient for resolving older radiations than the one of the Lake Tanganyika cichlids.

ND2 evolves about 10 times faster than RAG1. This lies in the upper range of the differences in evolutionary rates usually accepted for these two types of DNA markers (Meyer 1993b). It also means that, in order to get the same quantity of phylogenetic information with nuclear DNA as with the ND2 gene, one would have to sequence 10,647 bp of RAG1. In RAG1 of Lake Tanganyikan cichlids, exon 3 evolves equally as rapidly as intron 2. This seems to be inconsistent with the general view that introns might be more informative since they are not as evolutionary constrained as exons and, therefore, would tend to evolve more freely. Usually introns evolve around three times faster than exons since they are expected to evolve approximately as fast as third positions of codon triplets do (Li 1997). In our case, the first and second codon positions in the exon 3 of RAG1 evolved more slowly compared to intron 2, while the third codon positions evolved more rapidly than intron 2.

The pairwise distance comparison in Fig. 6E highlights an interesting position of the pairwise distances within the Lamprologini, which are situated in the range of the C-lineage pairwise comparison. This illustrates that, although the ancestor of the Lamprologini is likely to have evolved before the radiation of the C-lineage, their diversification might

have happened more recently, in parallel to that of the C-lineage tribes.

In general, the ncDNA improved support values in ML and BI, for example, the monophyly of the Lamprologini and the Haplochromini/Tropheini. Future studies with sequences of an intermediate pace of molecular evolution such as unconstrained introns would seem to hold promise. The development of different kinds of markers (such as the 11-bp deletion in the intron of all the Haplochromini and Tropheini included) should also be followed up, since they might contain phylogenetic information that would not be dependent on problems related to the rate of evolution. Thus, even if in the case of the Lake Tanganyika cichlid species flock, the use of other markers might be more appropriate, we believe that RAG1 is suitable for the resolution of phylogenies of older groups.

Acknowledgments. We thank S. Koblmüller, J. Snoeks, C. Sturmbauer, E. Verheyen, and L. De Vos for provision and identification of some of the specimens and R. Gueta, I. Eistetter, T. Mack, and the other members of the Meyer lab for technical assistance. We are grateful to M. Cummings and L. Rüber and the second anonymous reviewer for discussions and valuable comments on the manuscript. This study was supported by the Landesstiftung Baden-Württemberg, the Center for Junior Research Fellows (University Konstanz), and the EU (Marie Curie fellowship) to W.S. and grants from the Deutsche Forschungsgemeinschaft to A.M.

References

- Abila R, Barluenga M, Engelken J, Meyer A, Salzburger W (2004) Population-structure and genetic diversity in a haplochromine cichlid of a satellite lake of Lake Victoria. *Mol Ecol* 13(9):2589–2602
- Albertson RC, Markert JA, Danley PD, Kocher TD (1999) Phylogeny of a rapidly evolving clade: The cichlid fish of Lake Malawi, East Africa. *Proc Natl Acad Sci USA* 96:5107–5110
- Allender CJ, Seehausen O, Knight ME, Turner GF, Maclean N (2003) Divergent selection during speciation of Lake Malawi cichlid fish inferred from parallel radiations in nuptial coloration. *Proc Natl Acad Sci USA* 100:14074–14079
- Anderson S, deBruijn MHL, Coulson AR, Eperon IC, Sanger F, Young I (1982) Complete sequence of bovine mitochondrial DNA. Conserved features of the mammalian mitochondrial genome. *J Mol Biol* 156:683–717
- Avice JC (1984) *Molecular markers, natural history and evolution.* Chapman & Hall, New York
- Bininda-Emonds ORP, Sanderson MJ (2001) Assessment of the accuracy of matrix representation with parsimony analysis supertree reconstruction. *Syst Biol* 50:565–579
- Brinkmann H, Venkatesh B, Brenner S, Meyer A (2004a) Nuclear protein-coding genes support lungfish and not the coelacanth as the closest living relatives of land vertebrates. *Proc Natl Acad Sci* 101:4900–4905
- Brown WM, Prager EM, Wang A, Wilson AC (1982) Mitochondrial DNA sequences of primates: tempo and mode of evolution. *J Mol Evol* 18:225–239
- Brower AVZ, DeSalle R, Vogler A (1996) Gene trees, species tree, and systematics. *Annu Rev Ecol Syst* 27:423–450
- Brufford MW, Hanotte O, Brookfield JFY, Burke T (1998) Multilocus and single-locus DNA fingerprinting. In: Hoelzel A (eds) *Molecular genetic analysis of populations: a prac-*

- tical approach. Oxford University Press, Oxford, pp 287–336
- Cohen A, Lezzar K, Tiercelin J, Soreghan M (1997) New paleontologic and lake level reconstructions of Lake Tanganyika: implication for tectonic, climatic and biological evolution in a rift lake. *Basin Res* 7:107–132
- Cohen A, Soreghan M, Scholz C (1993) Estimating the age of formation of lakes: An example from Lake Tanganyika, East African Rift system. *Geology* 21:511–514
- Cummings MP, Otto SP, Wakeley J (1995) Sampling properties of DNA sequence data in phylogenetic analysis. *Mol Biol Evol* 12:814–822
- De Vos L, Seegers L (1998) Seven new *Orthochromis* species (Teleostei: Cichlidae) from the Malagarazi, Luiche und Rugufu basins (Lake Tanganyika drainage), with notes on their reproductive biology. *Ichthyol Explor Freshwaters* 9:371–420
- Farias IP, Orti G, Meyer A (2000) Total evidences: molecules, morphology, and the phylogenetics of cichlids fish. *Mol Dev Evol* 288:76–92
- Felsenstein J (2004) *Inferring phylogenies*. Sinauer Associate, Sunderland, MA
- Flynn JJ, Nedbal MA (1998) Phylogeny of the Carnivora (Mammalia): Congruence vs. incompatibility among multiple data sets. *Mol Phylogenet Evol* 9:414–426
- Fryer G, Iles TD (1972) The cichlid fish of the Great Lakes of Africa. T.H.F., Neptune City, NJ
- Greenwood PH (1973) A revision of *Haplochromis* and related species (pisces, Cichlidae) from Lake George, Uganda. *Bull Br Mus Nat Hist (Zool)* 25:141–242
- Greenwood PH (1980) Towards a phyletic classification of the “genus” *Haplochromis* (Pisces, Cichlidae) and related taxa. Part II. The species from Lakes Victoria, Nabugabo, Edward, Georg and Kivu. *Bull Br Mus Nat Hist (Zool)* 39:1–101
- Greenwood PH (1981) The haplochromine fish of the East African lakes. Cornell University Press, Ithaca, NY
- Hasegawa M, Kishino H, Yano T (1985) Dating of the human–ape splitting by a molecular clock of mitochondrial DNA. *J Mol Evol* 22:160–174
- Huelsenbeck JP, Ronquist F (2001) MrBayes: Bayesian inference of phylogeny. *Bioinformatics* 17:754–755
- Klett V, Meyer A (2002) What, if anything, is a Tilapia?—Mitochondrial ND2 phylogeny of Tilapiines and the evolution of parental care systems in the african cichlid fish. *Mol Biol Evol* 19:865–883
- Kluge AG, Farris JS (1969) Quantitative phyletics and the evolution of anurans. *Syst Zool* 18:1–32
- Koblmüller S, Salzburger W, Stumbauer C (2004) Evolutionary relationships in the sand-dwelling cichlid lineage of Lake Tanganyika suggest multiple colonization of rocky habitats and convergent origin of biparental mouthbrooding. *J Mol Evol* 58:79–96
- Kocher TD (2004) Adaptive evolution and explosive speciation: the cichlid fish model. *Nat Rev Genet* 5:288–298
- Kocher TD, Conroy JA, McKaye KR, Stauffer JR (1993) Similar morphologies of cichlid fish in Lakes Tanganyika and Malawi are due to convergence. *Mol Phyl Evol* 4:420–432
- Kocher TD, Conroy JA, McKaye KR, Stauffer JR, Lockwood SF (1995) Evolution of NADH Dehydrogenase Subunit 2 in East African cichlid fish. *Mol Phylogenet Evol* 4:420–432
- Kocher TD, Fernald R, Hofmann H, Meyer A, Okada N, Penman D, Seehausen O (2004) Genome sequence of a cichlid fish: the Nile tilapia (*Oreochromis niloticus*). Proposal submitted to the JGI Community Sequencing Program by the Cichlid Genome Consortium. Available at: <http://www.hcgs.unh.edu/cichlid/>
- Koepfli K-P, Wayne RK (2003) Type I STS markers are more informative than cytochrome b in phylogenetic reconstruction of the Mustilidae (Mammalia:Carnivora). *Syst Biol* 52:571–593
- Kornfield I, Smith PF (2000) African cichlid fish: model systems for evolutionary biology. *Annu Rev Ecol Syst* 31:163–196
- Li W-H (1997) *Molecular evolution*. Sinauer Associates, Sunderland, MA
- Lippitsch E (1995) Scale and squamation character polarity and phyletic assessment in the family Cichlidae. *J Fish Biol* 47:91–106
- Lippitsch E (1998) Phylogenetic study of cichlid fish in Lake Tanganyika: a lepidological approach. *J Fish Biol* 53:752–766
- Mayer WE, H. T, Klein J (1998) Phylogeny of African cichlid fish as revealed by molecular markers. *Heredity* 80:702–714
- Meyer A (1993a) Phylogenetic relationships and evolutionary processes in East African cichlid fish. *TREE* 8:279–284
- Meyer A (1993b) Evolution of mitochondrial DNA in fish. In: Hochachka PW, Mommsen TP (eds) *Molecular biology frontiers, biochemistry and molecular biology of fish*. Elsevier Science, Amsterdam, pp 1–38
- Moran P, Kornfield I (1995) Were population bottlenecks associated with radiation of the mbuna species flock (Teleostei: Cichlidae) of Lake Malawi? *Mol Biol Evol* 12:1085–1093
- Nagl S, Tichy H, Mayer WE, Takahata N, Klein J (1998) Persistence of neutral polymorphisms in Lake Victoria cichlid fish. *Proc Natl Acad Sci USA* 95:14238–14243
- Nishida M (1991) Lake Tanganyika as an evolutionary reservoir of old lineages of East African fish: inferring form allozymes data. *Experientia* 47:974–979
- Nishida M (1997) Phylogenetic relationships and evolution of Lake Tanganyika cichlids: a molecular perspective. In: Kawanabe H, Hori M, Nagoshi M (eds) *Fish communities in Lake Tanganyika*. Kyoto University Press, Kyoto, pp 1–24
- Nylander JAA, Ronquist F, Huelsenbeck JP, Nieves-Aldrey JL (2004) Bayesian phylogenetic analysis of combined data. *Syst Biol* 53:47–67
- Poll M (1956) Resultats scientifique. Exploration hydrobiologique belge au lac Tanganyika (1946–1947). Poissons Cichlidae. Institut Royal des Sciences Naturelles de Belgique, Brussels
- Poll M (1986) Classification des Cichlidae du lac Tanganyika—Tribus, genres et especes. Académie royale de Belgique, Memoires de la Classe des Sciences, coll. in 8°, 2^{ème} série, T. XVI, fascicule Z
- Posada D, Crandall KA (1998) ModelTest: Testing the model of DNA substitution. *Bioinformatics* 14:817–818
- Rodriguez F, Oliver JF, Marin A, Medina JR (1990) The general stochastic model of nucleotide substitutions. *J Theor Biol* 142:485–501
- Ronquist F, Huelsenbeck JP (2003) MrBayes 3: Bayesian phylogenetic inference under mixed models. *Bioinformatics* 19:1572–1574
- Rüber L, Meyer A, Sturmbauer C, Verheyen E (2001) Population structure in two sympatric species of Lake Tanganyika cichlid tribe Eretmodini: evidence for introgression. *Mol Ecol* 10:1207–1225
- Rüber L, Britz R, Tan HH, Ng PKL, Zardoya R (2004) Evolution of mouthbrooding and life-history correlates in the fighting fish genus *Betta*. *Evolution* 58:799–813
- Russo CA, Takezaki N, Nei M (1996) Efficiencies of different genes and different tree-building methods in recovering a known vertebrate phylogeny. *Mol Biol Evol* 13:525–536
- Salzburger W, Meyer A (2004) The species flocks of East African cichlid fish: recent advances in molecular phylogenetics and population genetics. *Naturwissenschaften* 91: 277–290
- Salzburger W, Meyer A, Baric S, Verheyen E, Stumbauer C (2002a) Phylogeny of the Lake Tanganyika Cichlid species flock and its relationship to the central and East African Haplochromine cichlid fish faunas. *Syst Biol* 51:113–135
- Salzburger W, Baric S, Sturmbauer C (2002b) Speciation via introgressive hybridization in East African cichlids? *Mol Ecol* 11:619

- Salzburger W, Mack T, Verheyen E, Meyer A (2005) Out of Tanganyika: Genesis, explosive speciation, key innovations and phylogeography of the haplochromine cichlid fish. *BMC Evol Biol* 5:17
- Santini S, Boore JL, Meyer A (2003) Evolutionary conservation of regulatory elements in vertebrate *Hox* gene clusters. *Genet Res* 13:1111–1122
- Schelly RC, Stiassny MLJ (2004) Revision of the congo river *Lamprologus* Schilthuis, 1891 (Teleostei: Cichlidae), with descriptions of two new species. *Am Mus Nov* 3451:1–40
- Schelly R, Salzburger W, Koblmüller S, Duftner N, Sturmbauer C (2005) Phylogenetic relationships of the lamprologine cichlid genus *Lepidiolamprologus* (Teleostei: Perciformes) based on mitochondrial and nuclear sequences, suggesting introgressive hybridization. *MPE* (in press)
- Schmidt HA, Strimmer K, Vingron M, von Haeseler A (2002) TREE-PUZZLE: maximum likelihood phylogenetic analysis using quartets and parallel computing. *Bioinformatics* 18:502–504
- Seehausen O (2004) Hybridization and adaptive radiation. *TREE* 19:198–207
- Seehausen O, Koetsier E, Schneider MV, Chapman LJ, Chapman CA, Knight ME, Turner GF, van Alphen JJM, Bills R (2003) Nuclear markers reveal unexpected genetic variation and a Congolese/Nilotic origin of the Lake Victoria cichlid species flock. *Proc Roy Soc Lond* 270:129–137
- Shaw PW, Turner GF, Iddid MR, Robinson RL, Carvalho GR (2000) Genetic population structure indicates sympatric speciation of Lake Malawi pelagic cichlids. *Proc Roy Soc Lond B* 267:2273–2280
- Shimodaira H, Hasegawa M (1999) Multiple comparisons of log-likelihoods with applications to phylogenetic inference. *Mol Biol Evol* 16:1114–1116
- Snoeks J, Rüber L, Verheyen E (1994) The Tanganyika problem: comments on the taxonomy and distribution patterns of its cichlid fauna. *Adv Limnol* 44:355–372
- Stiassny MLJ (1990) *Tylochromis*, relationships and the phylogenetic status of the African Cichlidae. *Am Mus Novit* 1993:1–14
- Stiassny MLJ, Meyer A (1999) Cichlids of the African rift lakes. *Sci Am* February:44–49
- Strimmer K, von Haeseler A (1996) Quartet puzzling: A quartet maximum-likelihood method for reconstructing tree topologies. *Mol Biol Evol* 13:964–969
- Strimmer K, von Haeseler A (1997) Likelihood-mapping: A simple method to visualize phylogenetic content of a sequence alignment. *Proc Natl Acad Sci USA* 94:6815–6819
- Sturmbauer C, Meyer A (1992) Genetic divergence, speciation and morphological stasis in a lineage of African cichlid fish. *Nature* 358:578–581
- Sturmbauer C, Meyer A (1993) Mitochondrial phylogeny of the endemic mouthbrooding lineages of cichlid fish of Lake Tanganyika, East Africa. *Mol Biol Evol* 10:751–768
- Sturmbauer CE, Verheyen E, Meyer A (1994) Mitochondrial phylogeny of the Lamprologini, the major substrate spawning lineage of cichlid fish from Lake Tanganyika in Eastern Africa. *Mol Biol Evol* 10:751–768
- Sturmbauer C, Hainz U, Baric S, Verheyen E, Salzburger W (2003) Evolution of the tribe Tropheini from Lake Tanganyika: synchronized explosive speciation producing multiple evolutionary parallelism. *Hydrobiologia* 500:51–64
- Swofford DL (2002) PAUP*. Phylogenetic analysis using parsimony (*and other methods), version 4.0. Sinauer Associates, Sunderland, MA
- Swofford DL, Sullivan J (2003) Phylogeny inference based on parsimony and other methods: Practice. In: Salemi M, Vandamme A-M (eds) *The phylogenetic handbook, a practical approach to DNA and protein phylogeny*. Cambridge University Press, Cambridge, pp 182–206
- Takahashi K, Terai Y, Nishida M, Okada N (1998) A novel family of short interspersed repetitive elements (SINEs) from cichlids: the patterns of insertion of SINEs at orthologous loci support the proposed monophyly of four major groups of cichlid fish in Lake Tanganyika. *Mol Biol Evol* 15:391–407
- Takahashi K, Terai Y, Nishida M, Okada N (2001) Phylogenetic relationships and ancient incomplete lineage sorting among cichlid fish in Lake Tanganyika as revealed by analysis of the insertion of retrotransposons. *Mol Biol Evol* 18:2057–2066
- Takahashi T (2003) Systematics of Tanganyikan cichlid fish (Teleostei: Perciformes). *Ichthyological Res* 50:367–382
- Terai Y, Takezaki N, Mayer WE, Tichy H, Takahata N, Klein J, Okada N (2004) Phylogenetic relationships among east african Haplochromine fish as revealed by short interspersed elements (SINEs). *J Mol Evol* 58:64–78
- Turner GF, Seehausen O, Knight ME, Allender CJ, Robinson RL (2001) How many species of cichlid fish are there in African lakes? *Mol Ecol* 10:793–806
- Van der Meijden A, Vences M, Meyer A (2004) Novel Phylogenetic relationships of the enigmatic brevicipitine and scaphiophrynine toads as revealed by sequences from the nuclear RAG1 gene. *Proc Roy Soc B (Suppl)* 271:S378–S381
- Venkatesh B, Ning Y, Brenner S (1999) Late changes in spliceosomal introns define clades in vertebrate evolution. *Proc Natl Acad Sci USA* 96:10267–10271
- Verheyen E, Salzburger W, Snoeks J, Meyer A (2003) Origin of the superflock of cichlid fish from Lake Victoria East Africa. *Science* 300:325–329
- Wiens JJ (2003) Missing data, incomplete taxa, and phylogenetic accuracy. *Syst Biol* 52:528–538
- Zardoya R, Meyer A (1996) Phylogenetic performance of mitochondrial protein-coding genes in resolving relationships among vertebrates. *Mol Biol Evol* 13:933–942