Cichlids of the lower Congo River - a new model system in speciation research?

Dissertation

zur Erlangung des Doktorgrades (Dr. rer. nat.)

der

Mathematisch-Naturwissenschaftlichen Fakultät

der

Rheinischen Friedrich-Wilhelms-Universität Bonn

vorgelegt von

Julia Schwarzer

aus

Höltinghausen

August 2011

Gedruckt mit Genehmigung der Mathematisch-Naturwissenschaftlichen Fakultät

der Rheinischen Friedrich-Wilhelms-Universität Bonn.

Die Arbeit wurde am Zoologischen Forschungsmuseum Alexander

Koenig in Bonn durchgeführt.

Erstgutachter: Prof. Dr. B. Misof

Zweitgutachter: Prof. Dr. T.C.M. Bakker

Tag der mündlichen Prüfung: 18.01.2012

Erscheinungsjahr: 2012

The Cataracts of the Congo

« Allí o mui grande reino está de Congo,
Por nós ja convertido à fé de Christo,
Por onde o Zaire passa claro e longo,
Rio pelos antiguos nunca visto.”

“Here lies the Congo kingdom, great and strong,
Already led by us to Christian ways;
Where flows Zaïre, the river clear and long,
A stream unseen by men of olden days.”

The Lusiada, V. 13.

Preface

Taxonomic remarks

This following section gives a preliminary update on recent taxonomic results concerning Cichlidae (Teleostei, Perciformes) with a focus on Congolian riverine cichlids. Tribus names ending with "-ini" in cichlids are only partly taxonomically available according to to the ICZN-rules, although some are used indiscriminantly. These formal and informal Tribus-names were nevertheless used in the present study as they considerably facilitate the verbal description of cichlid relationships, however, taxonomically unavailable tribus names publishes with "-ini" were consistently renamed as ending with "-ines" to avoid confusion.

Several undescribed species exist in all studied cichlid genera, for which designations corresponding to their sampling location (e.g. *Nanochromis* sp. "Ndongo", *Steatocranus* sp. "Nki") or, if present, according to conspicuous morphological characteristics (*S*. sp. "red eye" or *S*. sp. "bulky head") are used.

For the genera *Steatocranus* Boulenger, 1899 and *Nanochromis,* Pellegrin 1904, a taxonomic revision will follow (Schliewen et al. in prep). All deviations from the present taxonomy are supported by well resolved phylogenetic trees (present study) as well as meristics (unpublished results U. K. Schliewen and J. Wedekind). In *Steatocranus*, the type material of *S. tinanti* Poll, 1939 is described from below Kinshasa. Based on this, only specimens that were sampled in the upstream lower Congo region at sampling sites 1 to 4 (Fig. 1) will be designated as *S. tinanti*. Specimens from the central part of the lower Congo are called *S.* sp. aff. *tinanti* "ultraslender", specimens from Inga (Fig. 1, sampling site 8) *S*. sp. aff. *tinanti* "Inga" and those distributed in the rapids of Yalala are called *S.* sp. aff. *tinanti* "intermediate" is (Fig. 1). These informal names correspond with distinct genetic clusters supported by morphological data. In *S. casuarius* Poll, 1939, the type material was collected from below Kinshasa. Specimens from central lower Congo sampling locations are in the following called *S*. sp. aff. *casuarius* "brown pearl", as they clearly differ in coloration from *S. casuarius* exhibiting a smaller dark center in all scales. For *Steatocranus gibbiceps* Boulenger, 1899 the type locality is Matadi (Boulenger 1899). The occurrence of *S. gibbiceps* below the Yalala rapids (Fig. 1) could, however, not be verified despite substantial efforts. It is therefore likely that the type location information refers to the port of shipment ("Matadi") rather than the

collection site. All *S. gibbiceps* specimens are called *S*. cf. *gibbiceps* in the following sections with remarks on the collection localities. In *S. mpozoensis* Roberts & Stewart, 1976 the type material was described from the Mpozo River, a tributary to the lower Congo. Specimens from downstream sampling locations in Boma (Fig. 1) appear reciprocally monophyletic to those from the Mpozo and are therefore called here *S*. cf. *mpozoensis*.

Systematics and taxonomy of haplochromine cichlids remain complicated, and species of various riverine lineages have been placed in poorly diagnosable genera. Here, a pragmatic approach is adopted, naming all riverine haplochromine species in the Congo basin *Haplochromis* except for members of genera endemic to River Fwa (*Schwetzochromis neodon* Poll, 1948 and *Cyclopharnyx schwetzi* (Poll, 1948)), all rheophilic haplochromines currently assigned to the genus *Orthochromis* Greenwood, 1954, and the serranochromine genera *Pharyngochromis* Greenwood, 1979, *Sargochromis* Regan, 1920, *Chetia* Trewavas, 1961 and *Serranochromis* Regan, 1920. Obviously, a phylogenetic classification of more than a thousand haplochromine species is not possible at the moment, but I follow the argumentation of van Oijen (1996), who argued, that most haplochromine genera a defined on the basis of clinal rather than discrete characters.

Summary

Fishes account for the highest diversity of all vertebrate groups on our planet. Among them one group clearly outnumbers all others - cichlids (Perciformes, Cichlidae). Within the East African Lakes a stunning number of cichlid species evolved most likely within a very short time frame. Many studies dealing with these lacustrine radiations were conducted within the last decade rendering the cichlids of the so-called East African radiations to text book examples in evolutionary biology. Riverine cichlids, distributed in all major African river systems, are considerably less well studied, most likely because they are generally species-poor and exhibit limited morphological diversity. Recent studies highlighted the impact of riverine cichlids on species diversity of the lacustrine radiations and as potential seeding lineages of the East African Lakes. The present study focuses on cichlids of the lower Congo River, one of the most spectacular habitats for animal life on earth. One primary aim was to establish the lower Congo cichlids as the first model system in speciation research based on riverine species and to emphasize their contribution to general cichlid species diversity. The reconstruction of robust phylogenetic trees and reliable placement of the lower Congo cichlids within a framework of African cichlids were the first goals of the present study. The lower Congo cichlid genera are assigned to different cichlid tribes: "*Haplochromis*", *Lamprologus* and *Steatocranus* belong to the haplotilapiines, a tribe comprising the megadiverse lacustrine radiations and tilapiine genera, and *Nanochromis* and *Teleogramma* belong to chromidotilapiines, an ancient West/Central African mainly riverine cichlid lineage. Multi-locus phylogenetic trees of both haplotilapiines and chromidotilapiines and more detailed of haplochromines, a subclade of the haplotilapiines, were reconstructed, allowing the identification of biogeographic coherences and general relationships of the lower Congo genera with the remaining African cichlids (Chapter 2, Chapter 3 and Chapter 4). Age estimates allowed setting phylogenetic splitting events in context with the Palaeohistory of the African continent. Species differentiation along the lower Congo River and potential timing of initial colonization of the lower Congo rapids were inferred by including (nearly) all to date known species of the genera *Steatocranus* and *Nanochromis*. Intrageneric relationships were reconstructed based on extensive multi-locus AFLP datasets in combination with mitochondrial and nuclear sequence data. I showed, that *Steatocranus* and *Nanochromis* species predominantly differentiated allopatrically within the lower Congo

River about 5mya (Chapter 5). The rapids were colonized in each genus at least twice from of surrounding lakes and rivers. The existence of various levels of gene flow between adjacent and more distantly related species especially in *Steatocranus* species (Chapter 6) and within the haplochromines (Chapter 4) underlines the non-destructive and potentially even beneficial role of hybridization in cichlids. A rather web-like evolution in the genus *Steatocranus* challenges the general applicability of bifurcating trees especially in species known to hybridize.

The present study provides a phylogenetic framework for a complex system that may serve as a link between African riverine cichlid diversity and the megadiverse cichlid radiations of the East African lakes. Based on phylogenetic reconstructions, the genetic structure and the time and origin of colonization of two Congolean cichlid genera were inferred. Furthermore, the substantial impact of hybridization on riverine and lacustrine cichlids was shown.

This work was supported by grants of the Deutsche Forschungsgemeinschaft to B.M. (DFG MI649/8-1) and UKS (SCHL567/4-1) and a graduate student grant of the University of Bonn to J.S..

Contents

CHAPTER 1

General introduction

Speciation theory

Species diversity is an observable fact, the origin and underlying causes for biodiversity, however, remain unclear in most cases (Coyne & Orr 2004). Many studies address theoretical aspects of speciation (Kondrashov & Mina 1986; Kondrashov & Shpak 1998; Dieckmann & Doebeli 1999; Gavrilets 2000; Turelli et al. 2001; Doebeli & Dieckmann 2003), but empirical evidence is still rare (Schliewen 1994; Coyne & Orr 2004; Savolainen et al. 2006), as is a uniformly adopted species concept (Coyne & Orr 2004). Geographically, one can allocate speciation scenarios into three categories: (1) allopatric (distinct distributions, Mayr 1942; Mayr 1963), (2) parapatric (partly overlapping distributions, Endler 1977) and (3) sympatric (fully overlapping distributions, Turelli et al. 2001) speciation. Whereas allopatric speciation can in theory operate as a function of time driven by drift and mutations, parapatric and sympatric speciation require selective forces driving divergence and hindering admixture (like assortative mating, Coyne & Orr 2004). It might therefore be on first sight counterintuitive to assume that gene flow and hybridization can also promote speciation (Seehausen 2004; Mallet 2007).

Hybridization and hybrid speciation

Hybridization has long been viewed as a process hampering speciation rather than facilitating it. Given that at least 10 % of all animal species (and about 25 % of all plant species) are known to hybridize (Mallet 2005), the potential impact of hybridization on biodiversity can not be ignored. In fact, the impact of hybridization on speciation might be substantial. There is growing evidence that homoploid hybrid speciation (without a change in chromosome numbers) is contributing to the creative potential of natural hybridization, despite the high complexity of the underlying process (Seehausen 2004; Schliewen & Klee 2004; Nolte et al. 2005; Mavarez et al. 2006; Mavarez & Linares 2008). Recent hypotheses emphasize the potential of hybridization in shaping the diversity of adaptive radiations (Seehausen 2004). For example, the *hybrid swarm hypothesis* states that a population that varies in functional traits due to hybridization may be capable to reach fitness peaks (or "niches") significantly faster by ecological selection than non-hybrid species. The *syngameon*

hypothesis, emphasizes the potential of occasional gene flow between incipient species, which can also lead to new adaptive trait combinations (Seehausen 2004). The success of hybrid speciation depends critically on the production of recombinant genotypes that can outperform their parents in certain habitats (Burke & Arnold 2001). If hybrids are able to survive in competition with their parents, they must have new and advantageous traits. Theoretical considerations suggest that adaptive peaks must be common, scattered and unoccupied in the fitness landscape in order to enable hybrid speciation. To overcome adaptive valleys, the process must also be saltatorial (Mallet 2007). Common sense and prevailing opinion suggest that evolution normally occurs through small adaptive changes (adjustments) rather than saltation. In this context, hybridization can be considered as a multi-locus "macro-mutation" that is capable to generate major phenotypic shifts and thus allows hybrids to occupy available ecological niches or adaptive peaks (Mallet 2007). *Transgressive segregation* (the exhibition of extreme phenotypes) is one mechanism that rapidly generates potentially favourable phenotypes (Rieseberg et al. 1999). To establish a new adaptive lineage from transgressive phenotypes requires the prevention of recombination that could break down the acquired novel gene complexes. This can be achieved if the emerging hybrids can separate spatially or ecologically from their parental species or if chromosomal rearrangements prevent recombination (Nolte & Tautz 2010). Selection favouring extreme phenotypes (Burke & Arnold 2001) plus the availability of vacant resources might thus lead to a stable reproductively isolated diploid hybrid lineage.

Model systems and available data

In animals, homoploid hybrid speciation seems to be quite common in groups known for their high propensity to speciate, such as freshwater fish and butterflies (Mallet 2005; Nolte et al. 2005; Mavarez et al. 2006; Huson & Bryant 2006; Mavarez & Linares 2008; Lucek et al. 2010). In butterflies, the potential hybrid species *Heliconius heurippa* exhibits an intermediate wing pattern compared to parental species, which was repeatable through experimental crossings (Mavarez et al. 2006). *Cottus* sp., an invasive sculpin in the Rhine and Sieg River opened up a new riverine habitat with parental species (*C. rhenanus* and *C. perifretum*) distributed in smaller adjacent tributaries (Nolte et al. 2005; Nolte et al. 2006). Several other examples are known from African cichlid fish (Cichlidae, Perciformes) (Salzburger et al. 2002; Seehausen 2004; Schliewen & Klee 2004). Cichlids are distributed in Africa, Madagascar, South and Middle America, parts of the Middle East and along the coast

of India, but especially the species-rich adaptive radiations of the East African Lakes Victoria, Malawi and Tanganyika are popular systems in speciation research (Kocher 2004; Seehausen 2006). With more than 2000 estimated species, Cichlids rank among the most diverse vertebrate groups on earth (Turner et al. 2001). Especially their ability to adapt and radiate in short time resulted in numerous studies on different aspects of speciation and the evolution of adaptive traits, e.g. (Terai et al. 2003a; Sugie et al. 2004; Streelman et al. 2007; Salzburger 2009). Their morphological, behavioural and ecological diversity restricted to single water bodies is considered ideal to study patterns and processes of speciation. It has been shown in recent studies that hybridization is a common phenomenon in cichlids, which can shape even long separated species (Stelkens et al. 2010). The influence of riverine cichlids on the East African radiations has repeatedly been discussed (Seehausen et al. 2003; Salzburger et al. 2005; Koblmüller et al. 2008a; Koblmüller et al. 2008b) and was recently shown by Joyce et al. (Joyce et al. 2011). Nevertheless, riverine African cichlid species remain poorly studied (Katongo et al. 2005; but see Joyce et al. 2005; Katongo et al. 2007). One reason might be that many riverine cichlid genera are species-poor and exhibit limited morphological diversity, making them less attractive study subjects. Similar to small Crater Lake radiations (Schliewen 1994; Schliewen & Klee 2004), small species flocks in well-circumscribed river stretches are, however, attractive for speciation studies, as they offer the possibility to be analysed *in toto*.

Cichlids of the lower Congo

The Congo River is a hotspot of aquatic diversity showing the highest fish species richness of any river system on the African continent (ca. 700 described species, see also (Teugels & Guegan 1994)) with dozens of endemic cichlid species (Roberts & Stewart 1976; Thieme et al. 2005; Schwarzer et al. 2009). Before reaching the Atlantic Ocean, all water collected in a drainage basin encompassing one eighth of the African continent (~ 3.8 mil km^2) is squeezed through an intermittently narrow and deep (up to 200 m) rocky channel, creating the world's most extensive rapids (Runge 2008). This ~ 350 km long river section extending from the first rapids near Kinshasa down to the last one upstream of Matadi is geologically young. Its origin is most likely related to a river capture event, i.e. a small coastal river hypothetically tapped the interior Congo basin (sometimes referred to as "Palaeolake Congo"), and subsequently created a novel outlet for the whole Congo drainage (Runge 2008). Sediment analyses, however, indicated that sediment deposition occurred more likely under non-

lacustrine, semi-arid environmental conditions (Runge 2008). The basin was originally a nearly flat area covered by rain forest. The forest was replaced by savanna vegetation after an aridification during the Last Glacial Maximum 26,500-19,000/20,000 ya (Runge 2001). Age estimates of the origin of the lower Congo rapids are imprecise, varying from early estimates of 0.4 mya (Colyn 1991) to 34 mya (Leturmy et al. 2003; Lucazeau et al. 2003). More recent analyses of Congo offshore deposits suggest an origin of the modern lower Congo at the Miocene-Pliocene transition at approximately 5 mya (Ferry et al. 2004), after the southern African continent had been affected by a significant uplift inducing a progressive rearrangement of the watersheds (Lavier et al. 2001). Dozens of fish species are endemic to the lower Congo (Roberts & Stewart 1976), including representatives of the cichlid genera *Steatocranus*, *Nanochromis, Lamprologus*, *Teleogramma* and "*Haplochromis*" (Fig. 1). The distribution of all five genera except for the nearly pan-African group *"Haplochromis"* is restricted to the Congo basin (van Oijen et al. 1991). Most species of the lower Congo River are endemic to the whole river stretch or even locally endemic to small subparts of the river (Roberts & Stewart 1976). The most speciose lineage in the Congo River is the haplotilapiine genus *Steatocranus* with 13-14 currently recognized species (Schliewen 2006a; Schliewen 2006b). While some occur parapatrically in the lower Congo rapids (e.g. *S. mpozoensis*, *S. glaber*), other species occur over a larger river stretch from Pool Malebo down to the Yalala rapids (Fig. 1, e.g. the *S. tinanti*, *S. casuarius*, *S. gibbiceps* species groups). Five species, (e.g. *S. rouxi* and *S. ubanguiensis*) are found in rapids of major Congo tributaries and in the area of rapids of the Upper Congo at Kisangani, all locations being at the outer margins of the Cuvette Centrale (Fig. 1). As far as it is known, all *Steatocranus* species are confined to rocky areas of larger rivers, and the genus is notably absent in smaller rivers and streams. While *Steatocranus* species of the lower Congo rapids exhibit a remarkable diversity in dentition and body shape, other taxa within the Congo are relatively uniform (Roberts & Stewart 1976). The genus *Nanochromis* comprises only taxa of the central part of the Cuvette Centrale, of Pool Malebo and the lower Congo rapids, while its sister-genus *Congochromis* is widely distributed in the Congo basin, but species within this genus have not colonized the lower Congo rapids (Schliewen & Stiassny 2006; Stiassny & Schliewen 2007). Four species, (*N. consortus*, *N. splendens*, *N. parilus* and *N. minor*) are endemic to the lower Congo, two to Lac MaiNdombe in the Cuvette Centrale (*N. wickleri*, *N. transvestitus*) and three to the rest of the Cuvette Centrale (Fig. 1, *N*. sp. aff. *consortus*, *N*. sp. "Kasai", *N. nudiceps*). While some

of the four lower Congo species occur in sympatry in the central parts of the lower Congo rapids at Inga (Roberts & Stewart 1976), and exhibit overlapping distributions with the more widespread *N. parilus*, the species of the Cuvette Centrale occupy allopatric areas within these vast lowland areas. As far as it is known, all species are restricted to the rock/sand interface of larger rivers and lakes and are notably absent from smaller rivers and streams. The highly rheophilic *Teleogramma* species occur in the lower Congo rapids (*T. depressa*, *T. gracile*, *T. brichardi*) and with a single species, *T. monogramma*, in the rapids of the Kasai River, a major southern affluent of the Congo (Fig. 1). While *T. gracile* is confined to the central stretch of the lower Congo, *T. brichardi* is known only from the upper portion close to Pool Malebo. In contrast, the distribution of *T. depressa* overlaps with both previous mentioned species. The large-scaled and apparently more generalized *T. monogramma* is spatially isolated from the lower Congo species. All non-Tanganyika Congolian lamprologines are placed in the genus *Lamprologus*, which not only comprises all riverine taxa, but also 11 taxa from Lake Tanganyika. Four species occur with partial overlap in different parts of the lower Congo (*L. teugelsi, L. tigripictilis, L. lethops* and *L. werneri*); other species are widespread in the northeastern part of the Congo basin (*L. mocquardii*), in the central and upper parts including Pool Malebo (*L. congoensis*), in the Cuvette Centrale (*L. tumbanus*), and in the upper Congo (Lualaba) drainage (*L. symoensi*). All other lamprologines (approx. 80 species) are confined to Lake Tanganyika and one affluent, the Malagarazi River. In contrast to the several hundred haplochromine species existing in Lakes Victoria, Malawi and Tanganyika (Turner et al. 2001), as well as in northern, eastern and southern Africa (Koblmüller et al. 2008a), species-richness in the lower Congo area is moderate with six presently known species ("*H*aplochromis" *polli*, "*H.*" *demeusii*, "*H*." *fasciatus*, "*H*." *bakongo*, "*H*." *snoeksi* and "*H*." sp. "Sanzikwa"). All "*Haplochromis*" species of the lower Congo appear to be distributed allopatrically. "*Haplochromis*" *bakongo*, "*H*." *snoeksi* and "*H*." sp. "Sanzikwa" are distributed in smaller lower Congo tributaries, whereas "H." *polli*, "*H.*" *demeusii*, "*H*." *fasciatus* occur in sandy still-water areas in the mainstream.

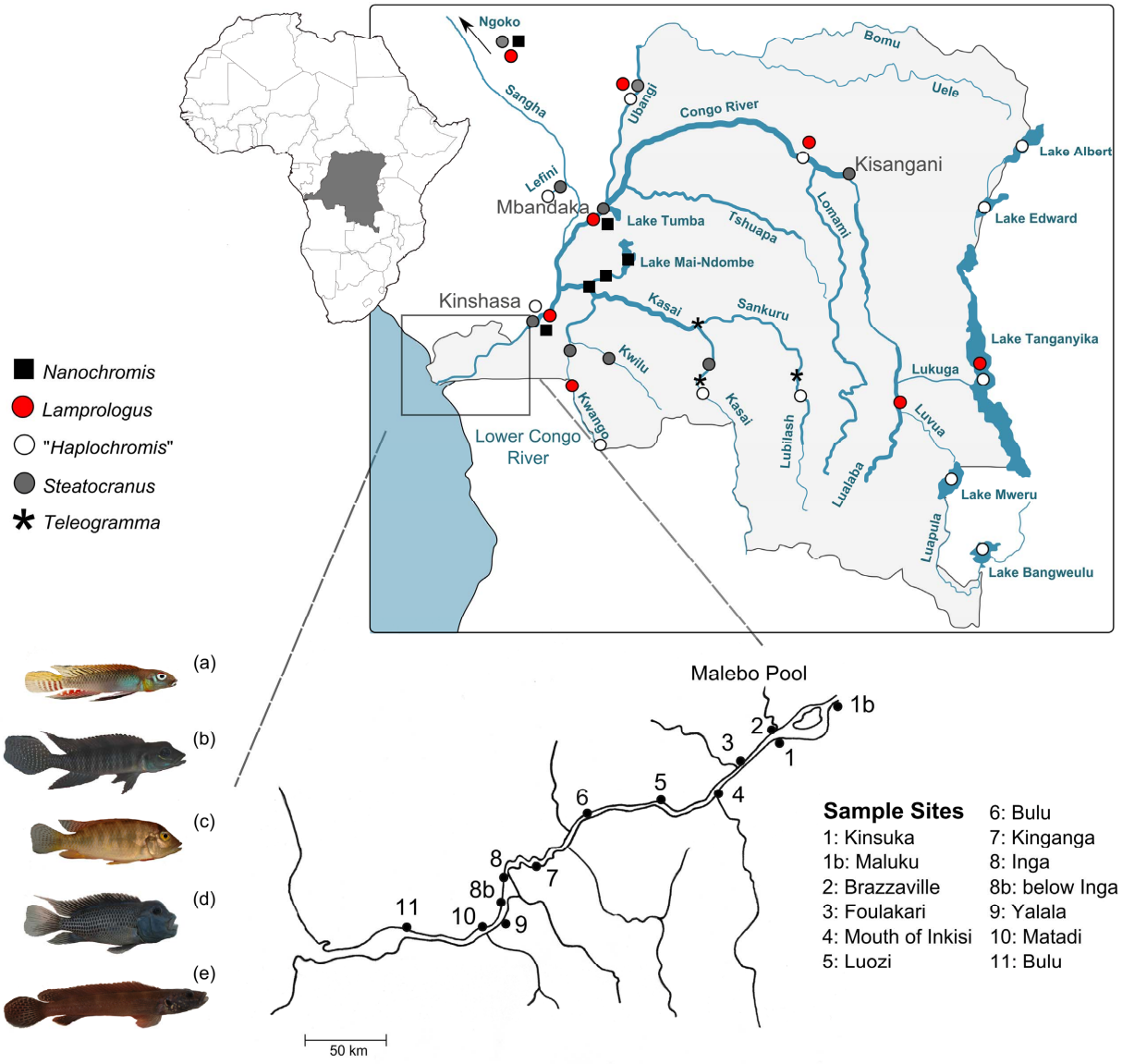


Fig. 1 *Cichlid genera of the lower Congo rapids and their distributions in the Congo basin*

The cichlid genera *Nanochromis*, *Lamprologus*, "*Haplochromis*", *Steatocranus* and *Teleogramma* are distributed along nearly the whole lower Congo River stretch. Species distributions, however, differ along the lower Congo and changes in species composition are present between upstream sampling sites (1-4), central sampling sites (5-8) and downstream sampling sites (8a – 11). The larger map roughly indicates distributions of species belonging to the above mentioned genera in the Congo Basin. Whereas *Steatocranus*, *Nanochromis*, *Teleogramma* and *Lamprologus* species are restricted to the Congo Basin, "*Haplochromis*" species cover a wider distribution range (specified in Chapter 4). "*Haplochromis*" species occur in all East African Lakes and *Lamprologus* species in L. Tanganyika.

Aim and structure of the present thesis

The cichlid species endemic to the 350 km stretch of the lower Congo River provide a unique possibility to study species differentiation and riverine speciation processes. A prerequisite, however, are accurate phylogenies, knowledge of gene-flow and species distribution patterns along the lower Congo river and relative divergence times between species pairs. Phylogenetically, African cichlids can be divided in five major monophyletic sublineages: “tylochromines”, “hemichromines”, “chromidotilapiines”, “pelmatochromines” and “haplotilapiines”. A sixth lineage, the monotypic genus *Heterochromis*, is either regarded as a distant outgroup or as the sistertaxon to all remaining African cichlids (Oliver 1984; Stiassny 1991; Lippitsch 1995; Kocher et al. 1995; Sultmann et al. 1995; Streelman & Karl 1997; Streelman et al. 1998; Mayer et al. 1998; Farias et al. 2000; Klett & Meyer 2002; Salzburger et al. 2002a; Schliewen & Stiassny 2003; Sparks & Smith 2004). The haplotilapiines (Schliewen & Stiassny 2003) comprise most of the species diversity of African cichlids, as the speciose Haplochromines are assigned to that tribe. The Congolian genera *Steatocranus*, *Lamprologus* and “*Haplochromis*” belong to that lineage (sensu Schliewen & Stiassny 2003), whereas *Nanochromis* and *Teleogramma* are assigned to the less species-rich chromidotilapiine cichlids (Takahashi & Nakaya 2002; Stiassny & Schliewen 2007). For both lineages, no resolved phylogenetic trees existed prior to this study (Klett & Meyer 2002; Schliewen & Stiassny 2003; Terai et al. 2003b; Stiassny & Schliewen 2007). The first goal of the present study was to place the lower Congo cichlids within a phylogenetic framework by reconstructing robust phylogenies of African cichlids (Chapters 2-4). The second goal was to solve intrageneric relationships within the genera to deduct hypotheses for the colonization and age of the lower Congo River and to detect potential differentiation and gene flow patterns along and within presumed species along the lower Congo River (Chapter 5). This was done, based on the ecologically differing cichlid genera *Steatocranus* and *Nanochromis* (see section above). The third goal was to infer the role of hybridization and gene-flow within cichlids, focussing on lower Congo and other riverine cichlid species (Chapter 3 and Chapter 6). This was done based on a phylogenetic approach combined with an experimental approach to detect homoplasy excess (specified in Chapter 6). The last chapter (chapter 7) gives a general discussion and a short outlook.

CHAPTER 2

The root of the East African cichlid radiations

Julia Schwarzer [1,2], Bernhard Misof [1], Diethard Tautz [3]& Ulrich K. Schliewen [2]

(1) Zoologisches Forschungsmuseum Alexander Koenig, Adenauerallee 160, 53113 Bonn, Germany

(2) Bavarian State Collection of Zoology, Münchhausenstr. 21, 81247 München, Germany

(3) Max-Planck-Institut für Evolutionsbiologie, August-Thienemann-Str. 2 24306 Plön, Germany

This is the author's version of a work originally published by BioMed Central Ltd. in: BMC Evolutionary Biology 2009, 9: 186

Abstract

For decades cichlid fishes (Perciformes: Cichlidae) of the East African cichlid radiations (Teleostei: Cichlidae) have served as natural experimental panels for the study of speciation processes and the search for potential speciation "key traits". Despite numerous phylogenetic studies dealing with their intragroup relationships, surprisingly little is known about the phylogenetic placement and time of origin of this enigmatic group. We used multilocus DNA-sequence data of five nuclear and four mitochondrial genes and refined divergence time estimates to fill this gap of knowledge. In concordance with previous studies, the root of the East African cichlid radiations is nested within the so called Tilapias, which is a paraphyletic assemblage. For the first time, we clarified tilapiine intragroup relationships and established three new monophyletic groups: "oreochromines", "boreotilapiines" and a group with a distribution center in East/Central Africa "austrotilapiines". The latter is the founder lineage of the East African radiations and emerged at the Miocene/Oligocene border at about 14 to 26 mya. Our results provide the first resolved hypothesis for the phylogenetic placement and age of origin of the megadivers East African cichlid radiations as well as for the phylogenetic placement of world's second most important aquaculture species, the Nile Tilapia, *Oreochromis niloticus*. Our analysis constitutes not only a first robust basis for African cichlid phylogenetics and systematics, but provides a valid and necessary framework for upcoming comparative phylogenomic studies in evolutionary biology and aquaculture.

Introduction

African cichlid fishes (Perciformes: Cichlidae) constitute the most species rich vertebrate model system in evolutionary biology and ecology, reviewed in (Seehausen 2006). The spectacular radiations of the East African rift valley Lakes Malawi and Tanganyika, L. Victoria and surrounding smaller lakes and rivers, are best known for their exceptional diversity and efficient habitat and resource exploitation (Kocher 2004). Numerous studies on different aspects of speciation and the evolution of adaptive traits are based on East African cichlids, e.g. (Terai et al. 2003a; Sugie et al. 2004; Streelman et al. 2007; Salzburger 2009). Identification of key factors (Coyne & Orr 2004) associated with the enormous evolutionary success of these radiations might improve our general understanding of speciation processes. For this a resolved phylogenetic framework is crucial (Coyne & Orr 2004). Nevertheless, the closest relatives of the East African cichlid Radiations (EAR) are still unknown, confusing interpretations of evolutionary trends in this group. This lack of knowledge can especially hinder comparative genomic studies and meta analyses, (e.g. Terai et al. 2003a; Kocher 2004; Seehausen 2006; Streelman et al. 2007; Salzburger 2009) which must rely on poorly resolved or poorly supported tree topologies. The monophyletic origin of African cichlids is supported by molecular and morphological analyses, as are five major monophyletic sublineages (Tylochromini, hemichromines, chromidotilapiines, pelmatochromines and haplotilapiines). A sixth lineage, the monotypic genus *Heterochromis*, is either regarded as a distant outgroup or as the sistertaxon to all remaining African cichlids (Oliver 1984; Stiassny 1991; Lippitsch 1995; Kocher et al. 1995; Sultmann et al. 1995; Streelman & Karl 1997; Streelman et al. 1998; Mayer et al. 1998; Farias et al. 2000; Klett & Meyer 2002; Salzburger et al. 2002a; Schliewen & Stiassny 2003; Sparks & Smith 2004). It is further established, that (1) the EAR, including the Nile Tilapia (*Oreochromis niloticus*), world's second most important aquaculture species (Contreras-Sanchez & Fitzsimmons 2006), are placed within the so called haplotilapiines (Schliewen & Stiassny 2003), of which internal relationships remain largely unresolved; and, that (2) the root of the EAR is placed somewhere within a large subgroup of cichlid fishes, the so called “Tilapias” or “Tilapiines” (Schliewen & Stiassny 2003; Terai et al. 2003b). Tilapiines are a widespread paraphyletic species assemblage including a few speciose and phenetically similar genera, i.e. *Tilapia*, *Oreochromis* and *Sarotherodon*, as well as several less speciose and in some cases monotypic genera as *Alcolapia*, *Tristramella*, *Danakilia, Iranocichla*, *Steatocranus*, *Gobiocichla* and

Chilochromis (Trewavas 1983; Klett & Meyer 2002; Schliewen & Stiassny 2003). Divergence time estimates for splits within the African cichlids are scarce and sometimes contradictory depending on the source of data. For example, fossil calibrated dating has resulted in much younger age estimates than Gondwana separation based dating, e.g. (Vences et al. 2001; Genner et al. 2007; Azuma et al. 2008). Reliable age estimates are not only required to link phylogenetic divergence with the palaeo-geographical background but also to appraise the speed of evolutionary change associated with rapid speciation events. Until now age estimates for the origin of the East African radiations have been mainly based on geological information, e.g. on lake ages, assuming that divergence of endemic clades took place after the formation of the lacustrine habitats (Salzburger et al. 2005; Koblmüller et al. 2008b). Other estimates based on Gondwana fragmentation yield rather imprecise ages for terminal nodes (Genner et al. 2007; Azuma et al. 2008) varying between 22 and 62 mya for the root of the EAR. The present study is designed to fill the gap between the rapidly increasing knowledge of various aspect shaping African cichlid evolution and the lack of a reliable phylogenetic background and divergence time estimates. In particular we intend to (i) establish a robust phylogeny for the paraphyletic group of Tilapias, (ii) identify the root of East African cichlid radiations, and finally (iii) estimate the root age of the primary East African radiation.

Methods

Samples and Sequences

A total of 63 specimens of 54 species were included (Table S1), representing all major groups of African cichlids, with focus on haplotilapiines sensu Schliewen & Stiassny (Schliewen & Stiassny 2003). To serve as nested outgroups, members of all basal African lineages (Tylochromini, chromidotilapiines, pelmatochromines, hemichromines) were added. As several recent molecular and morphological studies support the basal position of *Heterochromis multidens* with respect to the rest of the African cichlid radiation (Stiassny 1990; Lippitsch 1995; Salzburger et al. 2002a) this taxon served as outgroup. Total genomic DNA was isolated from fin clips or muscle tissue using the Qiagen Tissue Extraction Kit (DNeasy Tissue Extraction Kit) following the manufacturer's protocol. The following mitochondrial markers were amplified and sequenced: partial mitochondrial 12S and 16S genes, the connecting part between the above mentioned fragments and ND2. Additionally, four nuclear protein coding genes (ENCI, Ptr, Sh3px3 and Tmo4c4) and the first intron of the

ribosomal protein coding gene S7 were amplified and sequenced. Alignment of the sequences was conducted using BioEdit (ClustalW) and MUSCLE v. 3.6. Coding genes were translated into amino acid sequences to check for stop-codons or frame shifts and datasets were checked separately for saturation at each codon position. Base frequencies were equal for all markers (Chi-square tests, df = 183, all $p > 0.9$). A control for ambiguous alignment positions was conducted using ALISCORE v. 0.2 under default settings (Misof & Misof 2009). ALISCORE checks for random similarity of sequences using MCMC and a sliding windows approach. Under this regime similarity profiles based on pairwise comparisons of sequences were calculated. Ambiguous positions were summarized in a consensus profile along the alignment (Misof & Misof 2009) and subsequently removed from all analyses. The combined dataset of all sequenced markers resulted in a data matrix of in total 6176bp comprised of 12SrRNA: 349bp, 16SrRNA: 543bp, 12S/16S: 1245bp, ND2: 1014bp, ENCI: 725bp, Ptr: 691bp, Sh3px3: 681bp, Tmo4c4: 425bp and S7 (first intron): 503bp. In addition, a second dataset of 263 ND2 sequences (900bp) retrieved from Genbank and 38 newly sequenced ND2 sequences was generated (Table S2), resulting in high coverage over all major African cichlid tribes, some of which are not present in data set A. ND2 was chosen, because this marker was available on GenBank for a representative sampling of African cichlids. The third Codon position was saturated between in- and outgroups in dataset B, but because the focus of this analysis was the identification of terminal clades (younger splits), they were not excluded. Data were partitioned according to 1^{st}, 2^{nd} and 3^{rd} codon position and all parameters were estimated separately. A ML phylogeny was constructed with RAXML v. 7.0.3 using the fast rapid hill climbing bootstrap algorithm with 1000 replicates and following ML search. Branches not supported by 50 % bootstrap value were collapsed.

Phylogenetic reconstruction

Bayesian Inference (BI) and Maximum Likelihood (ML) approaches were used for phylogenetic inferences. The dataset was partitioned according to coding vs. non-coding and mitochondrial vs. nuclear genes yielding four partitions, i.e. two partitions for mitochondrial genes (rRNA and 1^{st} and 2^{nd} codon position of ND2) and two for nuclear genes (Exons and Intron). The third codon-position of ND2 was excluded from phylogenetic analyses (dataset A), as previous test showed saturation between Haplotilapiine and basal taxa (data not shown). For each partition model parameters were estimated separately. For BI, best fitting models of sequence evolution were estimated using the Bayes Factor Test (Nylander et al.

2004). Bayesian analyses were performed using MrBayes v. 3.1.2 (Huelsenbeck & Ronquist 2001) with eight parallel runs each with 10^6 generations starting with random trees and sampling of trees every 1000 generations. To ensure convergence the first 10^5 generations of each run were treated as burn-in and excluded. The remaining trees from all Bayesian analyses were used to build a 50 % majority rule consensus tree. The program RAxML v. 7.0.3 (Stamatakis 2006) was used for Maximum Likelihood analyses. Branch support was evaluated for the best scoring ML tree using non-parametric bootstrapping (BS) consisting of 1000 pseudoreplicates (using RAxML) and Bayesian posterior probabilities (BPP).

Testing alternative phylogenetic hypotheses

To test for unreliably placed taxa the leaf stability index (Thorley & Wilkinson 1999) was calculated for all taxa based on 1000 bootstrap trees using the program phyutility v. 2.2 (Smith & Dunn 2008). This index is a good measure of the consistency of a taxon's position relative to other taxa across bootstrap replicates. Using the same program branch attachment frequencies were calculated for lower supported clades (BS < 90) using 1000 bootstrap trees and the ML topology as well as the first 2000 BI topologies (after burn-in) and the BI topology. Following Seehausen (Seehausen 2004), we applied a tree based method to test for homoplasy excess in our dataset, possibly introduced by taxa of ancient hybrid origin. The inclusion of a hybrid taxon would be expected to increase internal conflict in the tree and diminish support values for affected nodes owing the reticulate nature of the process (Seehausen 2004). To test for this possibility, each taxon was successively removed from the dataset (N = 63 experiments) and subsequently a likelihood run (using RAXML), under the GTR + Γ model with 1000 rapid bootstrap replicates was conducted for each resulting dataset. The resulting trees and bootstrap support values for the focus clades were checked manually.

Divergence time estimates

Date estimates were calibrated using two age constraints. One calibration point (O) was based on the fossil record of *Oreochromis lorenzoi*† (Carnevale et al. 2003) from the Early Miocene of the Baid Formation (5.98 mya, Krijgsman et al. 1999). The second calibration point, assigned to the split between *Tylochromis* and the remaining African cichlids (except *Heterochromis multidens*), corresponds with the 95 % credible interval estimates for African Cichlidae from Azuma et al. (2008, exact dates were provided by Y. Kumazawa. and Y. Azuma, pers. comm.). Either estimates based on non-cichlid teleostean fossils (A_1 53-84 mya)

or Gondwana fragmentation (A2 71-89 mya) were taken. An exponential prior using a zero offset of 5.58 mya (marking the minimum age) with a mean of 1 was used for the fossil calibration point and a uniform prior with upper and lower bounds either from 53 to 84 mya (A_1), 71 to 89 mya (A_2) or a combination of both with 53 to 89 mya (A_3) (Azuma et al. 2008) were fixed prior to analyses. As the distinction between *Oreochromis* and *Sarotherodon* is based on characters that are often not preserved in fossils (Trewavas 1983), at least two possible placements for the *Oreochromis lorenzoi*† fossil in the phylogenetic framework exist. Most conservative is a placement at the base of all mothbrooding tilapiines (O_1) or, less so is a placement at the base of the genus *Oreochromis* (point O_2). *Oreochromis lorenzoi*† (Carnevale et al. 2003) is in our point of view one of the few reliable cichlid fossils suitable for calibration, as the type specimens are in a well preserved state and all key traits necessary for identification are recognizable. Its phylogenetic placement within the African cichlid phylogeny is less ambiguous that for other fossils, as the Oreochromines are a clearly monophyletic group (Fig. 2). Unfortunately this is not the case for most other African cichlid fossils, which often lack diagnostic characters necessary for a precise assignment to cichlid tribes (for more detailed discussion additional file 7). Divergence time analyses were conducted using a log-normal distributed relaxed molecular clock MCMC approach (Drummond et al. 2006) as implemented in BEAST v. 1.4.8 (Drummond & Rambaut 2007). For all calculations data were partitioned as described earlier and the BI topology was used as starting tree. Separate substitution models were used for each partition based on the results of the Bayes Factor test. A pure birth model (Yule) was assigned as prior for the branching process and two independent and identical runs were conducted for each BEAST setup for 30^6 generations. Convergence of parameters was checked using Tracer v. 1.4. The first 10 % of generations were discarded as burn-in and the effective sample size (ESS) was checked for good mixing of the MCMC. All exceeded 200 for all model parameters. Divergence dates were also estimated using penalized likelihood (Sanderson 2002) as implemented in the program R8s v. 7.1 (Sanderson 2003). The optimal smoothing parameter was 63 for each run determined by a cross-validation approach (Sanderson 2002). All runs were conducted several times with different sets of constraints to evaluate the influence of different calibration points. As expected inclusion of the fossil calibration point lead to slightly younger but also narrower confidence intervals for all ages (Fig. S3, SI). Two alternative placements of *Oreochromis lorenzoi*† within the topology resulted in slightly

different age estimates, with younger ages when the calibration point was set at the root of all Oreochromines. Using the penalized likelihood approach no difference in age estimates was observed for different placements of *Oreochromis lorenzoi*†. Overall, age estimates largely overlap independent of used priors (Fig. S3).

Results

The concatenated dataset included 56 taxa each with 6176 bp DNA sequence data derived from four mitochondrial and five nuclear loci (dataset A, Table S1). Of these, 394 bp were excluded from all analyses due to alignment ambiguities in non-coding genes and saturation in the 3rd codon position of the mitochondrial ND2 locus, resulting in a final alignment of 5782 bp. A second dataset (B) was composed of 301 taxa and 993bp of ND2 (Table S2). The 3rd codon position was not excluded in this dataset, as taxon assignment to terminal groups rather than basal resolution was the focus. Parameters were estimated separately for each codon position. Dataset A had 1783 variable sites and empirical base frequencies of A = 0.269, C = 0.252, G = 0.228, T = 0.251. Dataset B had 707 variable sites and empirical base frequencies of A = 0.262, C = 0.357, G = 0.118, T = 0.262. The Bayes factor test (Nylander et al. 2004) identified the HKY model as best fitting model for all partitions except for nuclear exons (ENC1, Ptr, Sh3px3, Tmo4c4), which were assigned to GTR + Γ. As expected, nuclear genes gave a better resolution in the more basal splits whereas mitochondrial genes provided increased resolution in terminal groups. The leaf stability index revealed an unstable placement of *Tilapia mariae* (0.67 vs. 0.87 as next higher value) whereas all other taxa were comparatively highly supported. Exclusion of this taxon from further analyses increased the leaf stability index significantly (Wilcoxon matched pairs signed rank test, N = 62, z = -6.164, $p < 0.001$, leaf stability for all taxa > 0.90). Furthermore, exclusion of the ambiguous *T. mariae* yielded a clear increase of BPP and BS support values in affected clades. This effect was not evident by consecutive exclusion of all other taxa (Fig. 1), thus *T. mariae* was excluded from all main analyses. Nevertheless, the topology of the remaining consensus trees in both ML and BI analyses remained unaffected.

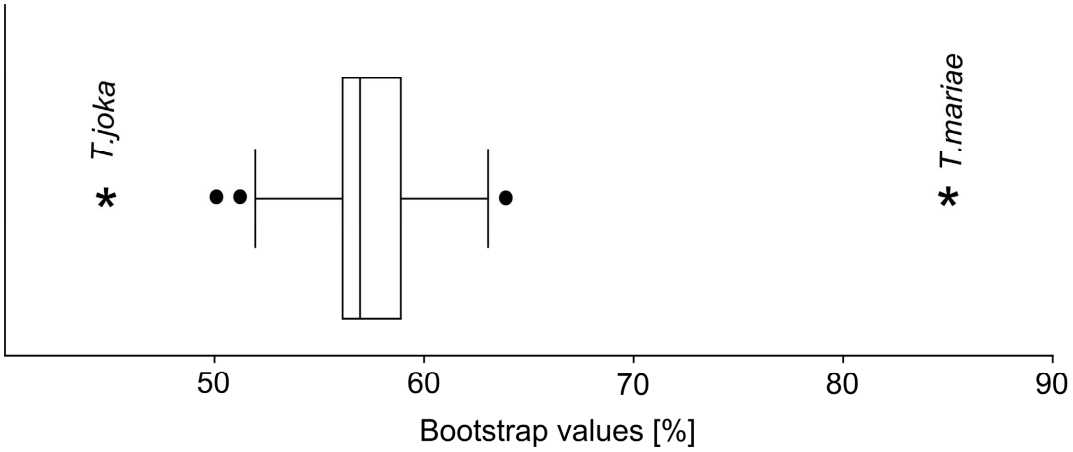


Figure 1 *Boxplot showing the results of the Homoplasy excess test*

The boxplot shows the distribution of bootstrap support values [%] for the austrotilapiines. Each specimen was removed iteratively from the dataset (resulting in N= 63 experiments) and 1000 bootstrap replicates were calculated using ML. Outliers are shown as asterisks. Bootstrap support values clearly increased (from initially 56 to 86) after exclusion of *T. mariae*. Removal of all other taxa did not cause this effect.

Phylogenetic relationships

Trees obtained from ML and BI analyses were highly congruent and nodes were supported for all major clades. Both approaches corroborated the monophyly of the haplotilapiines (100/1.00) whereas sistergroup relationship of this group within the African cichlids gained low BS and BPP values (45/0.74, Fig. 2). Within the haplotilapiines the following topology was highly supported (BS and BPP = 99): (a) *Etia nguti* was sistergroup to all other haplotilapiines. (b) The mouthbrooding genera *Oreochromis*, *Sarotherodon*, *Iranocichla* and *Tristramella* formed a monophyletic group, hereafter named „oreochromines'', after the most species rich genus within this group *Oreochromis*. The oreochromines were sistergroup to the substrate-brooders (clades BI, BII, AII and AIII) as well as to clade AI, comprising substrate and moothbrooding representatives of the East African radiations (Fig. 2). (c) A clade comprising clade AI (100/1.00), *Tilapia* sensu stricto (AII, 98/1.00), and *Steatocranus* from the Congo Basin (AIII, 100/1.00) formed the sistergroup to remaining haplotilapiines distributed mainly in the East/Central/Southern part of Africa. In recognition, this group is called “austrotilapiines”, in contrast to the “boreotilapiines” with a predominantly West/Central African distribution (Fig. 2). Within the boreotilapiines, a clade consisting of *Gobiocichla wonderi*, *Tilapia brevimanus*, *Tilapia busumana* and “*Steatocranus” irvinei* (BI, 100/1.00) and a clade comprising the *Tilapia* (Coptodon) subgenus (sensu Thys van den

Audenaerde 1971) as well as *T. joka* and *T. buttikoferi* (BII, 100/1.00, Fig. 1) appeared monophyletic and emerged as well supported sistergroups (96/1.00). Within the austrotilapiines sistergroup relationships were consistent and moderately well supported (86/1.00 and 87/1.00 respectively). All major clades were confirmed as monophyletic in a larger phylogenetic framework based on ND2 (Fig. S1).

The phylogenetic placement of the East African Radiations

Clade AI, comprising the EAR, appeared as sistergroup to the remaining austrotilapiines (Fig. 2). The mitochondrial dataset supported a sistergroup-relationship between AI and the Congolian genus *Steatocranus* (AIII), though with low support values (63/0.80), whereas the nuclear dataset in accordance with the concatenated dataset, favored the above mentioned relationship (69/0.95 and 87/1.00 respectively). Discordant phylogenetic signal was evident in 6 % and 7 % of the bootstrap replicates, favoring either a placement as sistergroup to monophyletic boreotilapiines and austrotilapiines (6 %) or a sistergroup relationship to boreotilapiines alone (7 %). All remaining hypotheses were supported with less than 1 % (Table S3). The 6 % signal was only detectable in the nuclear non-coding intron S7: without this marker the signal was hardly detectable (Table S3). No conflicting signal was detectable in 2000 randomly chosen BI topologies.

Divergence time estimates

Divergence time estimates yielded broadly consistent results (Table 1). Preliminary analyses indicated a younger age for node A (Fig. 3) than represented by prior A2 (71-89 mya, Gondwana calibration from Azuma et al. 2008) and the age estimates for most recent ancestor of *Oreochromis* (Node O2, Fig. 3) were younger (minimum age 4.18 mya, Table 1) than the age of the *Oreochromis lorenzoi†* fossil (Carnevale et al. 2003). Thus, final analyses were performed using priors O_1 (lower bound 5.98 mya at the base of all oreochromines, Fig. 3) and A_1 (53-84 mya, teleost fossil calibration from Azuma et al. 2008). The mean standard deviation width of the 95 % highest posterior density (HPD) was 12.07-5.32 mya and the precision of the estimate was highly correlated with node age (Pearson correlation, $p < 0.001$, $r = 0.703$, $N = 21$), pointing to more precise younger ages. The age of the most recent common ancestor of the haplotilapiines was estimated at about 37 (28-46) mya (Fig. 3, node C). Mean ages for the three major clades within the haplotilapiines were estimated at about 25 (19-32) mya for both, austrotilapiines and boreotilapiines (nodes F and G) and 13 (9-17) mya for the constrained oreochromines (node O_1, Fig. 3). The age for the East African

radiations, including the ancient lineages Bathybatini and *Boulengerochromis* was estimated at 20 (14-26) mya (node K) and the subclade comprising the H-lineage and „lamprologines'' was estimated to have emerged at 15 (11-20) mya (node P). In a second analysis Gondwana estimates, following Azuma et al. (2008), were included for calibration point A (A_3: 53-89 mya, Table 1). Results were highly congruent with the first run using fossil calibrations even though confidence intervals increased. The alternative algorithm based on penalized likelihood revealed highly congruent results with those obtained by the Bayesian approach (Table 1).

Discussion

With this well supported phylogeny and consistent divergence time estimates for the ancestors of the most diverse group of African cichlids a stable foundation is laid for further studies on this prime model system in evolutionary biology. Our results clearly show that the genus *Tilapia* is paraphyletic, and that previously proposed tilapiine subgenera, summarized in (Thys van den Audenaerde 1968b), need revision. As this is beyond the present study, we propose in accordance with good practice in cichlid taxonomy to use the genus name *Tilapia* Smith, 1840 only for *Tilapia* sensu stricto, i. e. the small ingroup of southeastern species containing the type species *Tilapia sparrmanii*, *T. ruweti*, *T. baloni*, and *T. guinasana*. Pending a thorough revision all other members should be referred to as „Tilapia'' (in quotation marks). The informal designation of identified clades etiines, oreochromines, austrotilapiines and boreotilapiines will facilitate discussion of haplotilapiine monophlyetic groups in the absence of a full taxonomic revision and renders the previously used term “Tilapiini” meaningless in the phylogenetic context. A list of all currently valid tilapiine species level taxa and their placement with respect to the newly named clades is provided (additional file 6) and will be available in a regularly updated version under www.zsm.mwn.de/ich/resources.htm.

Phylogenetic relationships of African cichlids

Resolving relationships of African cichlids has always been challenging. While phylogenetic relationships between and within the great African lakes radiations (Kocher 2004; Seehausen 2006) and riverine haplochromines (Salzburger et al. 2005; Joyce et al. 2005; Koblmüller et al. 2008a) are comparatively well understood, little was known about the broader

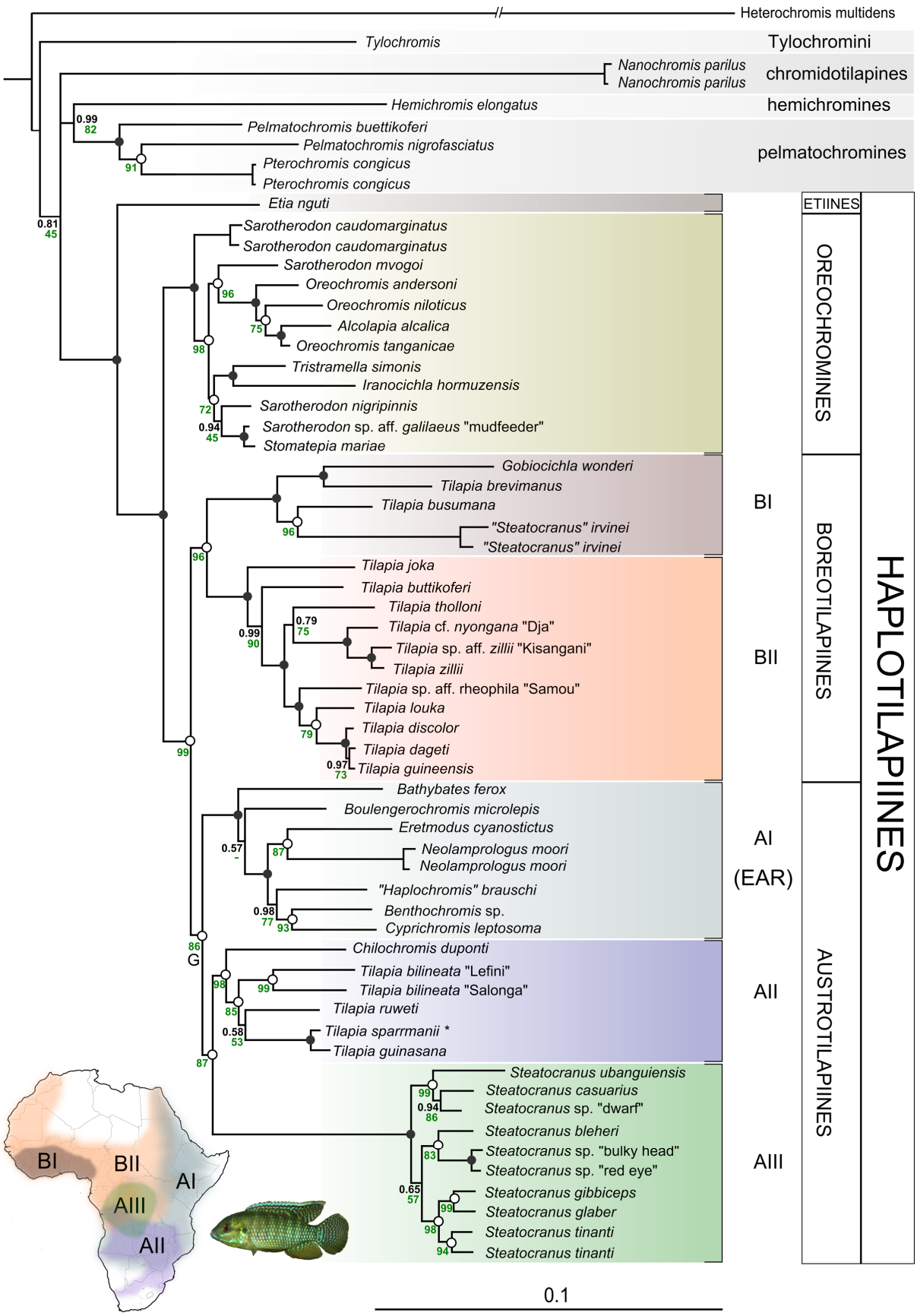


Figure 2 (caption overleaf)

Figure 2 *Consensus BI Tree of the African cichlid phylogeny*

Consensus tree (50 % majority rule) of the African cichlid phylogeny based on the concatenated dataset. The dataset comprises mitochondrial and nuclear sequences of nine independent markers. Green numbers at nodes refer to bootstrap-values (BS, 1000 replicates) of the ML run and black numbers to Bayesian posterior probabilities (BPP). Filled circles represent a 100 % BS support and 1.00 BPP and empty circles 1.00 BPP and lower BS values. Major groups within the phylogeny were named either based on the center of their geographic distribution (austrotilapiines and boreotilapiines) or based on taxonomic aspects (oreochromines). The asterisk (*) in the tree marks the type species of the genus *Tilapia*. The leaf stability index exceeded 0.95 for all specimens, except for clade AI (all taxa 0.90). Note that for clade AI only representatives of the EAR are included. The results presented here were verified using a more detailed taxon sampling based on ND2 (see Fig. S1). The map in the lower left corner shows major distribution ranges for austro- and boreotilapiines. The fish on the picture is *T. ruweti.*

phylogenetic framework for the most speciose group of cichlids (Terai et al. 2003b; Seehausen 2006).

Most often the so called Tilapiines were discussed as precursors of the East African cichlid radiations (Terai et al. 2003b; Seehausen 2006). Several morphological studies classified different tilapiine genera into various numbers of subgenera largely based on overall similarity of character states than on unambiguous apomorphies (Thys van den Audenaerde 1968b; Trewavas 1983). However, the diversity of this heterogeneous group was comparatively poorly represented in molecular phylogenetic studies (Mayer et al. 1998), but see (Nagl et al. 2001; Klett & Meyer 2002; Sparks & Smith 2004). A recent work based on the mitochondrial ND2 marker (Klett & Meyer 2002) accentuated the paraphyletic origin of the genera *Tilapia* and *Sarotherodon*, but did not recover well-supported deeper phylogenetic relationships. We present the first largely resolved phylogeny of African cichlids with emphasis on tilapiine cichlids including 47 ingroup and 7 outgroup species (Fig. 2). Phylogenetic analyses revealed congruent and largely well resolved topologies supporting a monophyletic origin of the haplotilapiines, comprised of all tilapiine cichlids as well as the East African radiations. Relationships of haplotilapiines to basal African cichlid tribes were only weakly supported, possibly due to high genetic distances compounding homoplastic signal amongst the most ancient nodes. In accordance to previous results the sistergroup to all remaining haplotilapiines was the monotypic taxon *Etia nguti* from the Cross River in Cameroon (Schliewen & Stiassny 2003). Earliest divergence within the haplotilapiines

separates the mouthbrouding, almost panafrican Oreochromines from predominantly substrate brooding tilapiines and the EAR. The latter form two monophyletic clades with largely non-overlapping distribution, one with a center in West/Central Africa (boreotilapiines), and one in East/CentralAfrica (austrotilapiines, Fig. 2).

Table 1 *Date estimates resulting from different molecular clock approaches*

Single dating points (mean height) and confidence intervals (95 %HPC) are shown for runs with (1) and without (2) the cichlid fossil calibration point. Prior A was constrained either with 53 to 84 mya (run 1) or with 53-89 mya (run2). § Letters correspond to node labels in Fig. 3.

Node[§]	Date estimates in Myr						Penalized Likelihood	
	Bayesian Inference							
	This study [1]		This study [2]		Genner et al. [Gondwana]		This study[1]	This study[2]
A_1	**56.7**	(53.0, 64.2)	**66.5**	(53.0, 85.2)	63.7 (N)	(46.6, 79.6)	53.0	53.4
B	**36.8**	(28.0, 45.9)	**46.9**	(32.9, 63.2)			37.1	37.4
C	**30.6**	(23.1, 37.9)	**39.6**	(27.9, 54.0)	46.4 (M)	(31.9, 61.7)	28.4	28.6
D	**27.6**	(21.0, 34.5)	**35.8**	(24.9, 48.9)			24.6	24.8
E	**25.5**	(19.0, 31.7)	**33.0**	(22.6, 45.0)			23.1	23.2
F	**25.3**	(18.9, 31.8)	**32.8**	(22.4, 44.9)			22.6	22.8
G	**20.2**	(14.4, 26.0)	**26.1**	(17.5, 36.7)	35.6 (L)	(22.3, 50.6)	18.1	18.2
H	**15.4**	(10.6, 20.4)	**20.0**	(12.8, 28.4)	29.5 (K)	(17.7, 43.2)	14.3	14.3
O_1	**12.8**	(8.9, 16.8)	**21.4**	(12.9, 31.1)			19.4	19.5

Remarkably a comparable distribution pattern is evident in cyprinodont killifish (Murphy & Collier 1997; Collier et al. 2009), explained by a marine incursion in the late Palaeocene at about 92-52 mya (Reyment & Dingle 1987; Giresse 2005) separating West Africa from the East and Central part. However, estimates for the haplotilapiine clades are substantially younger with 28 (21-35) mya (Fig. 3, node E) for the separation of boreo- and austrotilapiines and a subsequent diversification at 25 (19-32) mya (Nodes F and G). This even holds true without a fossil prior (Table 1). The estimated ages are concordant with the East African aridification at 33-20 mya (Davis et al. 2002; Loader et al. 2007), which influenced distribution patterns of the African fauna and flora, i.e. rainforest trees (Couvreur et al. 2008) and caecilian amphibians (Loader et al. 2007). However, the influence of the drought on freshwater systems possibly inhabited by the ancestors of the austro- and

boreotilapiines is not known at this time, leaving room for speculations about the evolution of this distribution pattern. The position of *Tilapia mariae* remained ambiguous in our analyses, which is reflected by a low leaf stability index. The predominant phylogenetic signal resulted in its placement as sistertaxon to boreotilapiines, but depending on the algorithm used, it was also sometimes resolved as sister to austrotilapiines (Fig. S2). A possible explanation for this could be an ancient hybrid origin of *T. mariae*, causing discordant phylogenetic signals in our dataset. Indeed, the distribution of the clade represented by *T. mariae* and its sistertaxon *T. cabrae* is intermediate between austrotilapiines and boreotilapiines (Fig. 2). A more detailed analysis is necessary to elucidate this pattern.

The origin of the East African radiation

The root of the East African radiations (EAR) within the substrate brooding tilapiine cichlids (Fig. 2) is corroborated with high support values. These results are consistent with earlier analyses based on limited taxon sampling or fewer loci, e.g. (Mayer et al. 1998; Terai et al. 2003b), proposing a closer sistergroup relationship of *Tilapia*/*Steatocranus* to the EAR than the mouthbrooding Oreochromines. Whereas the *Steatocranus* radiation of the Congo Basin forms a monophyletic clade, the genus *Tilapia* is clearly paraphyletic (Klett & Meyer 2002). *Tilapia* taxa included in the study of Terai et al. (Terai et al. 2003b) have affinities with the more distantly related boreotilapiines in our analyses. A closer phylogenetic relationship of the EAR to the austrotilapiines, comprising the Congolean *Steatocranus* and a clade composed of *Tilapia* s.str. *T. bilineata* and *Chilochromis*, is corroborated (Fig. 2). Biogeographically this is plausible, because austrotilapiines and the EAR largely overlap in their distribution around the East African Lakes. In this area, the only representative of the boreotilapiines present is *T*. (Coptodon) *rendalli*. The divergence of the EAR clade was estimated at 20 (14-16) mya (Fig. 3, Node K) including the ancient lineages and at 15 (11- 20) mya (Fig. 3, node P) for the more derived lacustrine and riverine radiations. Though only slightly overlapping, the latter age estimate would be congruent with an origin of the derived lineages in an emerging Lake Tanganyika, estimated at 9-12 mya (Cohen et al. 1993). Alternative age estimates using Gondwana fragmentation calibrations or an alternative dating algorithm (penalized likelihood) point to an older age for this node at 20 (13-28) mya and 14 mya, respectively (Table 1).

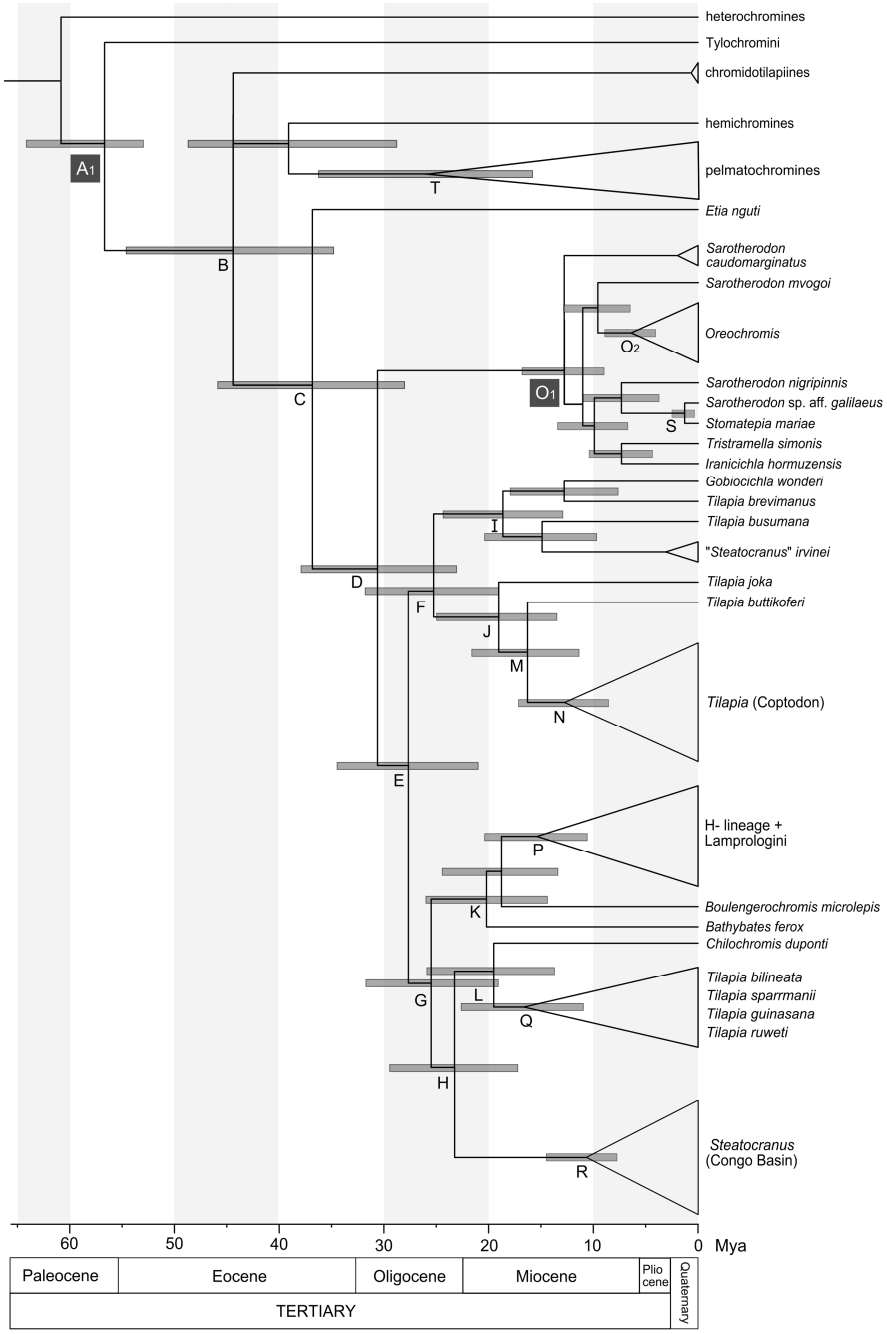


Fig. 3 *Chronogram showing divergence time estimates*

The chronogram was calculated based on the BI consensus tree. Divergence times were estimated using a partitioned Bayesian analysis implemented in BEAST. The following time constraints were used: A_1 53-84 mya (uniform prior), published age estimate based on non-cichlid fossils (Azuma et al. 2008) and O_1 5.98 mya (lower bound), the age estimate for *Oreochromis lorenzoi*† (Carnevale et al. 2003). The chronogram shows 95 % credibility intervals (HPC, grey bars). For nodes marked with letters, age estimates (95 % HPC and mean heights) are given in Table 1. Calibration points (O_1 and A_1) are marked with black squares. For simplification clear monophyletic groups were combined (shown as triangles).

These estimates favor the alternative hypothesis of an origin of derived lineages prior to the formation of Lake Tanganyika, in surrounding rivers or peripheral palaeolakes and subsequent independent colonization (Genner et al. 2007). Possibly, an increased taxon sampling with a multi-locus dataset would render more precise age estimates and remove this remaining uncertainty.

Conclusions

Here, we provide the first reliable phylogenetic placement of one of the most important model organism in evolutionary biology, the East African cichlids. We show that they are sister group to geographically proximate tilapiine cichlids with a main distribution center in East/Central Africa and that the whole group emerged in late Oligocene/early Miocene. The dataset provided here constitutes not only a stable basis for critical testing of divergence dates for basal EAR lineages from their tilapiine precursors (Koblmüller et al. 2005) but also a critical template for future phylogenomic and comparative studies based on African cichlids.

Authors contributions

JS and UKS designed the study. JS carried out the molecular work. JS, BM and UKS designed and conducted the analyses. All authors contributed to the preparation of the manuscript. They read and approved the final version.

Acknowledgements

We gratefully thank K. Borkenhagen, J. Friel, D. Golani, L. Rüber and R. Schelly for providing tissue samples and sequences. Y. Kumazawa and Y. Azuma kindly provided their age estimates for African cichlids. J. Frommen and F. Herder gave valuable comments on earlier versions of the manuscript and helped with statistical questions. K. Meusemann helped with LaTex. The picture in Fig. 2 was taken by J. Geck. This work was supported by grants of the Deutsche Forschungsgemeinschaft to B.M. (DFG MI649/8-1) and UKS (SCHL567/4-1) and a graduate student grant of the University of Bonn to J.S..

CHAPTER 3

Phylogenetic relationships of chromidotilapiines (Teleostei: Cichlidae) with emphasis on the genus *Teleogramma*

Julia Schwarzer [1,2], Anton Lamboj[3], Kathrin Langen[4], Bernhard Misof [1], & Ulrich K. Schliewen [2]

(1) Zoologisches Forschungsmuseum Alexander Koenig, Adenauerallee 160, 53113 Bonn, Germany

(2) Bavarian State Collection of Zoology, Münchhausenstr. 21, 81247 München, Germany

(3) Institut für Zoologie der Universität Wien, Abteilung für Evolutionsbiologie, Althahnstr. 14, A-1090 Wien, Austria

(4) Institute for Evolutionary Biology and Ecology, University of Bonn, An der Immenburg 1, D-53121 Bonn, Germany,

This is the author's version of a work prepared for publication in Molecular Phylogenetics and Evolution

Abstract

Cichlid fishes (Perciformes: Cichlidae) of the East African cichlid radiations (Teleostei: Cichlidae) are well studied and serve as prime model systems in the study of speciation and adaptive radiations. Ancestral lineages within the African cichlids remain, however, poorly studied. Here we present the first molecular phylogeny of the most species rich west-central African group of cichlids, the chromidotilapiines. Apart from resolving internal relationships within the chromidotilapiines, the enigmatic genus *Teleogramma* could be phylogenetically placed for the first time and the sistergroup of the congolian species complexes of *Nanochromis* and *Congochromis* could be identified. Cytonuclear discordances, however, partly mask phylogenetic patterns and complicated final conclusions.

Introduction

The center of cichlid species richness is located in the East African Great Lakes Tanganyika, Malawi and Victoria (Kocher 2004; Seehausen 2006). The search for the origin of these megadiverse species flocks has fuelled a large number of molecular phylogenetic studies. These, however, covered only a small fraction of the basal phylogenetic lineages of cichlids (Schliewen & Stiassny 2003; Schwarzer et al. 2009). Although there is increasing evidence that the African Cichlidae represent a monophylum, the Pseudocrenilabrinae (excluding *Heterochromis multidens*, Kullander 1998), neither relationships among the five major lineages (comprising the chromidotilapiines, pelmatochromines, Tylochromini, haplotilapiines and hemichromines), nor the intrarelationships of the most speciose and widespread non-haplotilapiine clade, the chromidotilapiines (Schliewen & Stiassny 2006), is resolved with adequate taxon sampling and statistical confidence. Thys van den Audenaerde (Thys van den Audenaerde 1968a) presented the first systematic revision of all species that were once classified in Regan´s genus *Pelmatochromis*. Based on a phenetic approach he recognized several informal "strains" and "species-groups" within the former *Pelmatochromis,* however, without giving a formal generic rank to any of those, or explicitly reassigning species to other genera. Instead, he created a novel subgenus, *Pelvicachromis,* and suggested to use for the remaining *Pelmatochromis* existing genus names as i.e. *Pelmatochromis*, *Chromidotilapia*, and *Nanochromis.* However, he considered his revision as preliminary and recognized that *Pelmatochromis* is a polyphyletic assemblage including problematic species that might rather belong to *Tilapia* or *Hemichromis* than to *Pelmatochromis.* Several of these problematic taxa were reassigned later to other genera, i.e. *P. ruweti* to *Tilapia* (Thys van den Audenaerde 1968b) and *P. exsul* to *Hemichromis* (Trewavas 1973)*;* the subgenus *Pelmatochromis* (*Pelmatochromis*) was transferred to *Tilapia* as a subgenus (Thys van den Audenaerde 1968b), but resurrected as a full genus by Trewavas (1973). With resurrecting *Chromidotilapia* to generic rank by Trewavas (1973), the remainder of Thys´ subgenera, *Pelvicachromis* and *Nanochromis* implicitly gained full generic rank, too. Greenwood (1987) presented the first cladistic phylogenetic review based mainly on osteological investigations of all genera of *Pelmatochromis* - related cichlids that were recognized at that time: *Pelmatochromis*, *Pterochromis*, *Thysochromis*, *Chromidotilapia*, *Nanochromis* and *Pelvicachromis*.

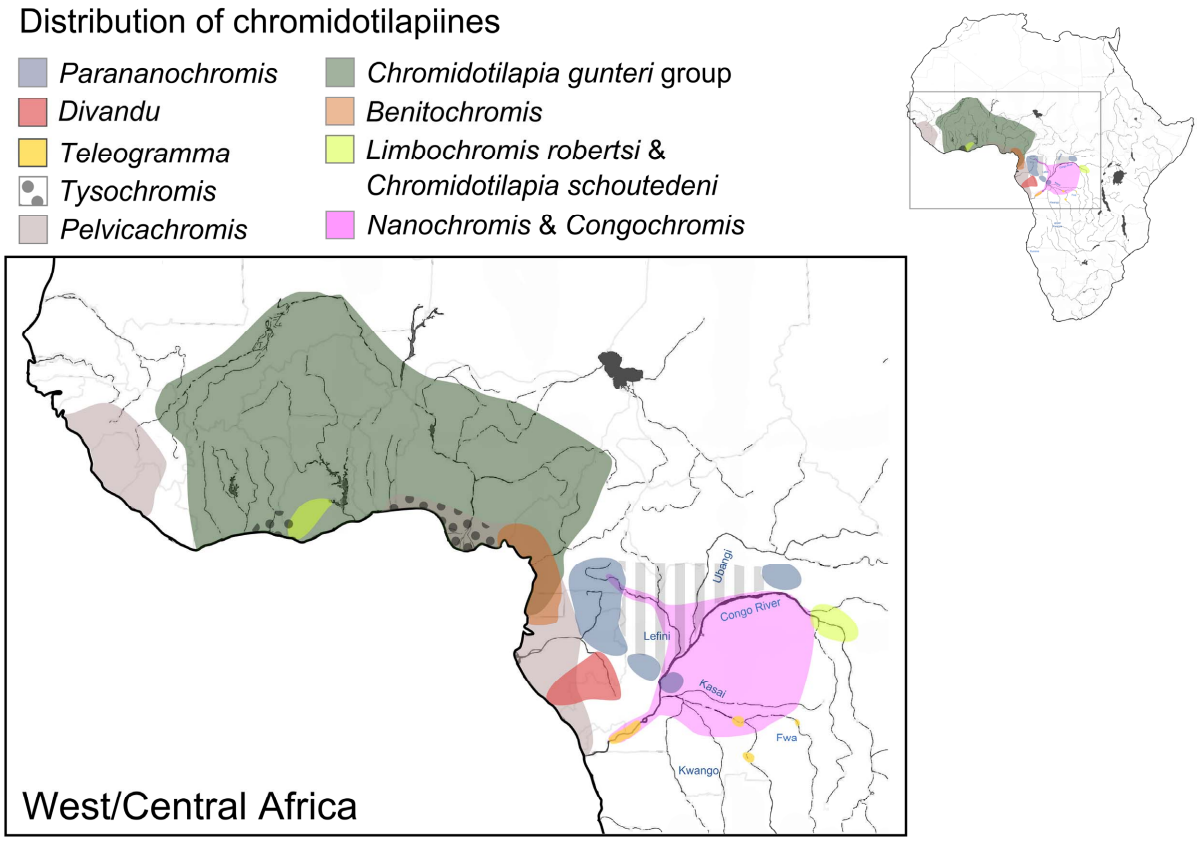


Figure 1 *Distribution ranges of chromidotilapiine lineages*

The distribution center of chromidotilapiines in West/Central Africa is shown enlarged. The hatched area indicates the proposed distribution of *Parananochromis* based on recently collected specimens in the Congo area (Snoeks & Stiassny, pers. comm). Distribution ranges of *Nanochromis* and *Congochromis* species largely overlap, but no *Congochromis* species are known from the lower Congo.

The most important result of his study was the recognition of a monophyletic lineage comprising *Chromidotilapia*, *Nanochromis*, *Pelvicachromis*, *Thysochromis* and two newly described genera *Limbochromis* and *Parananochromis*; for this lineage he coined the term chromidotilapiines. Due to the lack of informative morphological characters, intrarelationships of chromidotilapiines remained unresolved, although he suggested that *Thysochromis* is the plesiomorphic sister taxon to the other chromidotilapiines, and that *Parananochromis* and *Nanochromis* could be sister taxa. In addition, he indicated the possibility of a further subdivision of *Nanochromis* and *Pelvicachromis* into two genera each upon investigation of more material. After Greenwoods study numerous descriptions of novel species and genera (Lamboj 1999; Lamboj & Snoeks 2000; Lamboj 2001; Lamboj 2002;

Lamboj & Stiassny 2003; Lamboj 2003; Lamboj 2004; Lamboj 2005; Lamboj & Schelly 2006) were published, but no phylogenetic studies were presented that would account for the drastically increased richness of the most speciose group of west-central African cichlids. Three cichlid genera with clear distribution centers in the Congo Basin (*Nanochromis*, *Congochromis*) or the lower Congo rapids (*Teleogramma*, Fig. 1) most likely belong to the chromidotilapiines. Their phylogenetic placement within that group however remains ambiguous. The genus *Teleogramma*, endemic to rapids of the lower Congo (three species) and the Kasai (one species), a large southern affluent of the Congo, was never placed within African Cichlidae with confidence (Stiassny 1991; Takahashi & Nakaya 2002). In fact, *Telegramma gracile*, the type species of the genus, was originally placed into the family Labridae (Boulenger 1899). Myers (1939) tentatively placed it into the Cichlidae by referring to the presence of only a single pair of nostrils in *Teleogramma*, a character state that is shared within Labroidei with the Pomacentridae. This familial allocation was accepted as a consensus among cichlid systematists. This hypothesis was tested by Takahashi & Nakaya (2002), who confirmed the placement of *Teleogramma* within the Cichlidae on the basis of numerous osteological, mycological and soft anatomical characters. Lippitsch (1995) suggested that *Teleogramma* forms a sistergroup to the hemichromines based on scale characteristics and Stiassny (1997) assumed that the genus either belongs to lamprologines [Lamprologini following Poll (1986)] or forms a sistergroup to them, based on the sharing of (1) a variable number of large canines on the anterior part of both premaxilla and dentary, (2) four or more anal fin spines and (3) an abrupt scale size change above the upper lateral line and a naked cheek. Takahashi & Nakaya (2002) revised this and provided evidence that the genus belongs to the group of “remaining African cichlids” or “RACs”, the clade of Pseudocrenilabrinae that according to Stiassny (1991) contains all genera except *Heterochromis. Teleogramma* shares with RACs two apomorphic characters, first the loss of the posterior supraneural, and second the presence of the opercular spot. Both Stiassny (1997) and Takahashi and Nakaya (2002) concluded that refined morphological and/or molecular based phylogenies are required for resolving the position of *Teleogramma* within the African cichlids.

The present study, based on multiple nuclear and mitochondrial genes and an extensive taxon sampling, was conceived to: (i) resolve internal phylogenetic relationships within chromidotilapiines, (ii) identify their phylogenetic position within African cichlids and provide

age estimates for their emergence, (iii) resolve the phylogenetic placement of the enigmatic genus *Teleogramma* within the Pseudocrenilabrinae.

Methods

Samples and Sequences

Tissue samples (fin clips) of 83 taxa of all currently described chromidotilapiine and pelmatochromine genera, *Teleogramma* and 11 African cichlids representing all remaining major pseudocrenilabrine lineages *sensu* Sparks & Smith (2004) (*Heterochromis*, tylochromines, hemichromines and haplotilapiines) were obtained from field collections, donations and from aquarium trade specimens (Table S1). Members of all basal African lineages (Tylochromini, hemichromines, pelmatochromines) were added. *Heterochromis multidens* served as outgroup with respect to the rest of the African cichlid radiations (Stiassny 1990; Lippitsch 1995; Salzburger et al. 2002a). Total genomic DNA was isolated from fin clips or muscle tissue using the Qiagen Tissue Extraction Kit (DNeasy Tissue Extraction Kit) following the manufacturer's protocol. The mitochondrial gene NADH dehydrogenase 2 (ND2) and 16S rRNA, three nuclear protein coding genes (ENCI, Ptr, Sh3px3) and the first intron of the ribosomal protein coding gene S7 were amplified and sequenced. Alignment of the sequences was conducted using BioEdit (ClustalW) and MUSCLE v. 3.6. Coding genes were translated into amino acid sequences to check for stop-codons or frame shifts and datasets were checked separately for saturation at each codon position. Base frequencies were equal for all markers (Chi-square tests, df = 246, all $p > 0.9$). A control for ambiguous alignment positions was conducted using ALISCORE v. 0.2 under default settings (Misof & Misof 2009). Under this regime similarity profiles based on pairwise comparisons of sequences were calculated. Ambiguous positions were summarized in a consensus profile along the alignment (Misof & Misof 2009) and subsequently removed from all analyses. The combined dataset of all sequenced markers resulted in a data matrix of in total 3779 bp comprised of 16SrRNA: 515bp, ND2: 662bp, ENCI: 707bp, Ptr: 688bp, Sh3px3: 679bp and S7 (first intron): 528bp.

Phylogenetic reconstruction

Bayesian Inference (BI) and Maximum Likelihood (ML) approaches were used for phylogenetic inferences. The dataset was partitioned according to coding vs. non-coding and mitochondrial vs. nuclear genes yielding four partitions, i.e. two partitions for mitochondrial

genes (rRNA and 1st and 2nd codon position of ND2) and two for nuclear genes (exons and introns). The third codon-position of ND2 was excluded from phylogenetic analyses, as previous tests showed saturation between chromidotilapiines and basal taxa (data not shown). For each partition model parameters were estimated separately. For BI, best fitting models of sequence evolution were estimated using the Bayes factor test (Nylander et al. 2004). Bayesian analyses were performed using MrBayes v. 3.1.2 (Huelsenbeck & Ronquist 2001) with four parallel runs each with 10^6 generations starting with random trees and sampling of trees every 1000 generations. To ensure convergence the first 10^5 generations of each run were treated as burn-in and excluded. The remaining trees from all Bayesian analyses were used to build a 50 % majority rule consensus tree. The program RAxML v. 7.0.3 (Stamatakis 2006) was used for Maximum Likelihood analyses. Branch support was evaluated for the best scoring ML tree using non-parametric bootstrapping (BS) consisting of 1000 pseudoreplicates (using RAxML) and Bayesian posterior probabilities (BPP). Mitochondrial and nuclear genes were analyzed separately using the above described ML approach to be able to detect cyto-nuclear discordances. The leaf stability index (Thorley & Wilkinson 1999) was calculated using phyutility v. 2.2 (Smith & Dunn 2008) based on 1000 ML bootstrap trees.

Dating and diversification rates

Divergence times for the chromidotilapiines were estimated using a Bayesian approach with a relaxed-clock implemented in BEAST v. 1.5.3 (Drummond & Rambaut 2007). The ML tree was used as starting tree. The Yule model was selected as tree prior and an uncorrelated lognormal model was used to estimate rate variation along branches. The same priors used in Schwarzer et al. (2009) were applied: an exponential prior (zero offset 5.98 mya) at the root of all oreochromines (following Carnevale et al. 2003) and a uniform prior (53-84 mya) at the root of all African cichlids except *Heterochromis* (Azuma et al. 2008). For a detailed discussion on the choice of priors see Schwarzer et al. (Schwarzer et al. 2009). The analysis was run 30^6 generations and the effective sample size was checked using Tracer v. 1.4 (Rambaut & Drummond 2007).

Results

Due to alignment ambiguities within the S7 intron (16 positions) and saturation in the 3rd codon position of the mitochondrial ND2 locus 347 positions were excluded from further analyses. The final alignment had 3779 bp. The Bayes factor test identified the HKY model as best fitting model for all partitions except for nuclear exons (ENC1, Ptr and Sh3px3) and 16S which were assigned to GTR + Γ. The final dataset had 739 variable sites and empirical base frequencies of A = 26, C = 26, G = 23, T = 25. The nuclear and mitochondrial datasets alone had 359 and 381 variable sites and base frequencies of A = 25, C = 24, G = 25, T = 26 and A = 26, C = 31, G = 18, T = 25, respectively. Nuclear genes gave a better resolution in the more basal splits whereas mitochondrial genes provided increased resolution in terminal groups. The leaf stability index, representing the stability of single taxa within the concatenated dataset revealed a comparatively high support for all included chromidotilapiines (between 0.93 and 0.95). Highest values were present for all *Parananochromis* and *Divandu albimarginatus* and lowest values for the phylogenetic positioning of *Teleogramma*, *Chromidotilapia schoutedeni* and *C. guntheri*, *Thysochromis ansorgii*, *Nanochromis minor* and both included *Pelvicachromis* species.

Phylogenetic patterns

Trees reconstructed based on the concatenated dataset yielded overall a good resolution for larger clades and recent relationships (Fig. 2). ML and BI analyze revealed congruent topologies for all major clades. Both approaches corroborated the monophyly of the chromidotilapiines whereas sistergroup relationship of this group within the African cichlids gained low BS and BPP values (46/0.88, Fig. 2). Within the chromidotilapiines two reciprocally monophyletic internal clades are present. One is composed of *Divandu albimarginatus* and *Parananochromis* (100/1.0) and the second (91/1.0) of *Teleogramma*, *Thysochromis*, *Pelvicachromis*, *Chromidotilapia guntheri*, *Benitochromis*, *Limbochromis robertsi*, *C. schoutedeni* and the two mainly Congolian genera *Nanochromis* and *Congochromis* (Fig. 2). Within the first clade the following topology was highly supported: *Divandu albimarginatus* appears as sistergroup to a monophyletic clade composed of *Parananochromis* (100/1.0); *P. brevirostris* and *P. longirostris* cluster together (70/0.96) as sistergroup (100/1.0) to *P.* sp. “Amba”, *P*. sp. “Lefini” and *P. ornatus*, which from a monophyletic group (100/1.0).

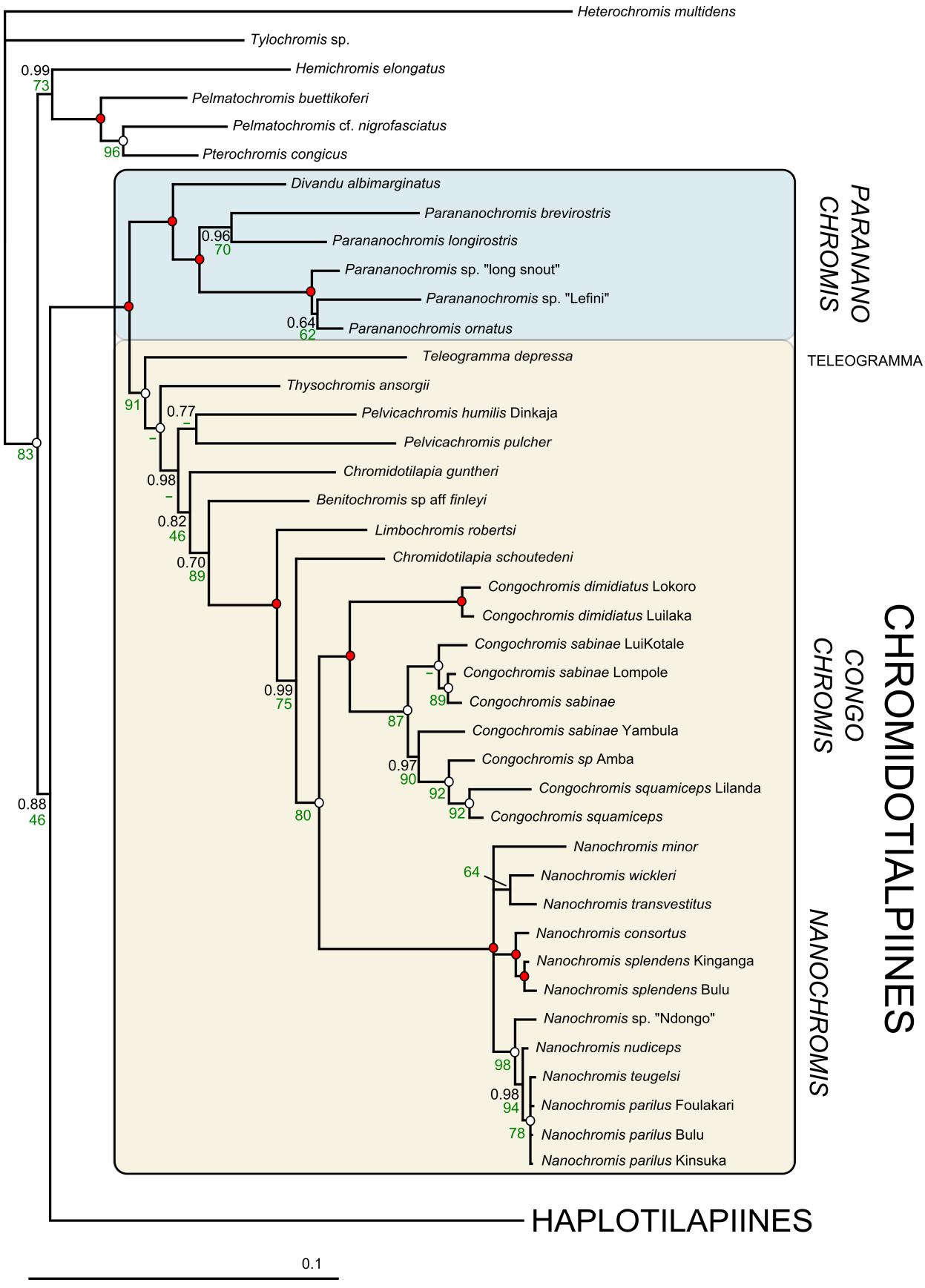


Figure 2 *(caption overleaf)*

Figure 2 *Consensus BI tree of the concatenated dataset based on six genes*

Consensus tree (50 % majority rule) of chromidotilapiines including representatives of the remaining major African cichlid tribes (pelmatochromines, Tylochromini, haplotilapiines and hemichromines) as well as *Heterochromis multidens* as outgroup. The dataset comprises mitochondrial and nuclear sequences of six independent markers. Green numbers at nodes refer to bootstrap-values (BS, 1000 replicates) of the ML run and black numbers to Bayesian posterior probabilities (BPP). Red circles represent a 100 % BS support and 1.00 BPP and white circles 1.00 BPP and lower BS values. The leaf stability index exceeded 0.93 for all specimens.

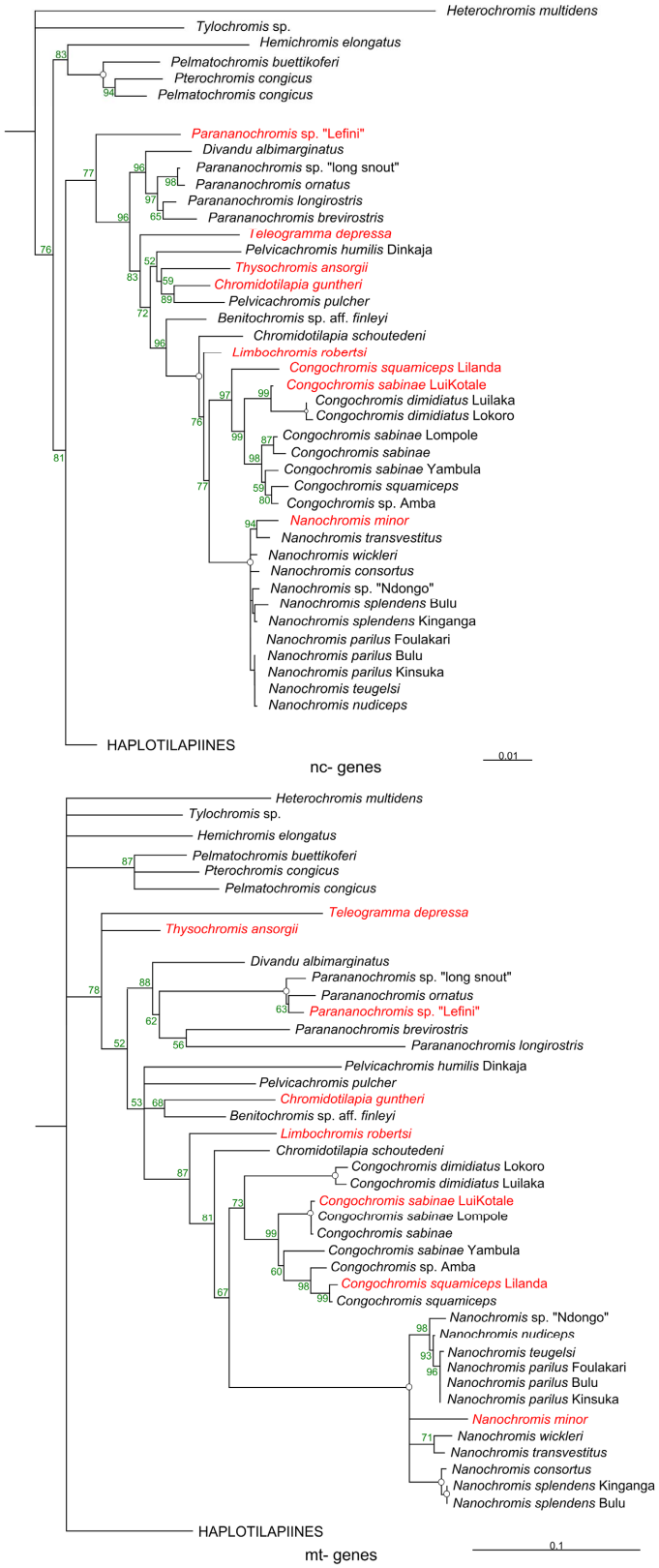


Figure 3 *ML trees generated based on mt (left) and nc genes (right)*

Maximum likelihood phylogeny calculated separately based on mitochondrial (1177bp) or nuclear genes (2602bp). Green numbers at nodes refer to bootstrap-values (BS, 1000 replicates). Species names marked in red indicate differences between the trees.

In the second clade *Teleogramma* was a sistergroup to the remaining taxa (91/1.0). Basal sistergroup relationships in this clade are mainly very weakly supported in the ML tree (all < 50), but comparatively well in the BI tree (0.98 - 1.0, Fig. 2). Well supported was the monophyly of a clade composed of *Limbochromis robertsi*, *Chromidotilapia schoutedeni*, *Congochromis* and *Nanochromis*. The two included *Chromidotilapia* species appeared paraphyletic based on our data (100/1.0, Fig. 2). *Nanochromis* and *Congochromis* species formed monophyletic sistergroups (80/1.00, monophyly of the clades: 100/1.0 both) with *C. schoutedeni* as sister species (75/0.99).

Cytonuclear discordances

The separate analyses of mtDNA and nuclear genes revealed differences concerning the position of *Teleogramma*, *Thysochromis*, *Parananochromis* sp. "Lefini", *Chromidotilapia guntheri*, *Limbochromis robertsi*, *Congochromis squamiceps* "Lilanda", *C. sabinae* "Lui Kotale" and *Nanochromis minor* (Fig. 3). In the mtDNA dataset *Teleogramma* appeared together with *Thysochromis ansorgii* unresolved as sistergroup to all remaining chromidotilapiines, whereas in the nuclear dataset *Teleogramma* forms the sistergroup to the chromidotilapiine clade 2, including *T. ansorgii*, which clusters with *Pelvicachromis* and *C. guntheri* (Fig. 4). *Parananochromis* sp. "Lefini" forms a well supported monophyletic clade with *P. ornatus* and *P*. sp. "Amba" (BS = 100) whereas the taxon appears as sistergroup to all chromidotilapiines in the nuclear dataset (BS = 77, Fig. 3). Whereas the monophyly of a clade comprising *Chromidotilapia schoutedeni*, *Limbochromis robertsi*, *Nanochromis* and *Congochromis* is supported in both datasets with high statistical support (nuclear DNA: BS = 100 and mtDNA: BS = 87), discordant signal is present concerning the internal sistergroup relationships. Whereas based on the mtDNA dataset *C. schoutedeni* appears as sistergroup to *Congochromis* and *Nanochromis* species (BS = 81), *Limbochromis robertsi* is their closest relative based on the nuclear dataset (BS = 76, Fig. 4). Intrageneric topological differences were also present within *Congochromis* and *Nanochromis* (Fig. 3).

Dating and diversification rates

The median age for the most recent common ancestor (MRCA) of chromidotilapiines was estimated at 38.1 (27.4 – 48.7) mya. The two internal clades (CHCL1 and CHCL2) were estimated to have emerged 28.4 (17.7 – 39) mya and 35.5 (25.6 – 46.2) mya. The second chromidotilapiine clade without *Teleogramma* (node CH3, Fig. 4) and *Tysochromis ansorgii* (node CH4, Table 1, Fig. 3) were estimated at 32.9 (23.3 – 42.7) mya and 30.7 (21.9 – 40.2)

mya respectively. The age of the included *Pelvicachromis* species was estimated at 25.9 (14.2 – 26.7) mya. The splitting of *Limbochromis robertsi* from the Congo Basin clade (node CON, Fig. 4) was estimated at 26.4 (18.8 – 35.3) mya and the MRCA of *Chromidotilapia schoutedeni*, *Congochromis* and *Nanochromis* (node CON 2, Table 1, Fig. 4) at 20 (13.9 – 27.2) mya. The MRCA of *Nanochromis* and *Congochromis* (node CON3) was estimated at 18 (12.2 – 24.4) mya and the age for *Nanochromis* and *Congochromis* at 8.12 (5.1 – 12.1) mya (node Node NA, Table 1, Fig.4) and 14.9 (9.6 – 20.4) mya, respectively.

Discussion

Phylogenetic relationships between and within the great African lakes radiations (Kocher 2004; Seehausen 2006) are comparatively well understood. Recently, the phylogenetic framework of these radiations within the haplotilapiines was partly resolved (Schwarzer et al. 2009). But still a lack of knowledge concerning broader phylogenetic relationships existed for the most speciose west-central-African lineage, the so called chromidotilapiines. This tribe comprises riverine and lacustrine species with a distribution center in West Africa, Lower Guinea and the Congo basin (Fig. 1). There is increasing evidence that Congolian riverine taxa might play an important role for the evolution of African cichlids *in toto* (Seehausen et al. 2003; Schwarzer et al. 2009). Beside others the endemic lower Congo genera *Nanochromis* and *Teleogramma* belong to this tribe. With the present study the monophyletic origin of the chromidotilapiines and the generic assignment of all suggested genera (Greenwood 1987, including *Teleogramma*) to the tribe could be verified.

Age and phylogenetic placement of chromidotilapiines

Phylogenetic analyses revealed congruent well resolved topologies supporting a monophyletic origin of chromidotilapiines with largely resolved internal relationships (Fig. 2). The genera *Pelvicachromis*, *Parananochromis*, *Congochromis* and *Nanochromis* appear monophyletic, whereas *Chromidotilapia* is clearly paraphyletic (Fig. 2). Relationships to basal African cichlid tribes (as hemichromines, pelmatochromines and haplochromines) were, however, only weakly supported. One potential reason might be that homoplastic signal amongst the most ancient nodes accumulated due to high genetic distances. The estimated origin for the chromidotilapiines dates back to late Eocene, early Oligocene (27 – 48.7) mya. Clade 1, comprising the monotypic genus *Divandu* and *Parananochromis* emerged with

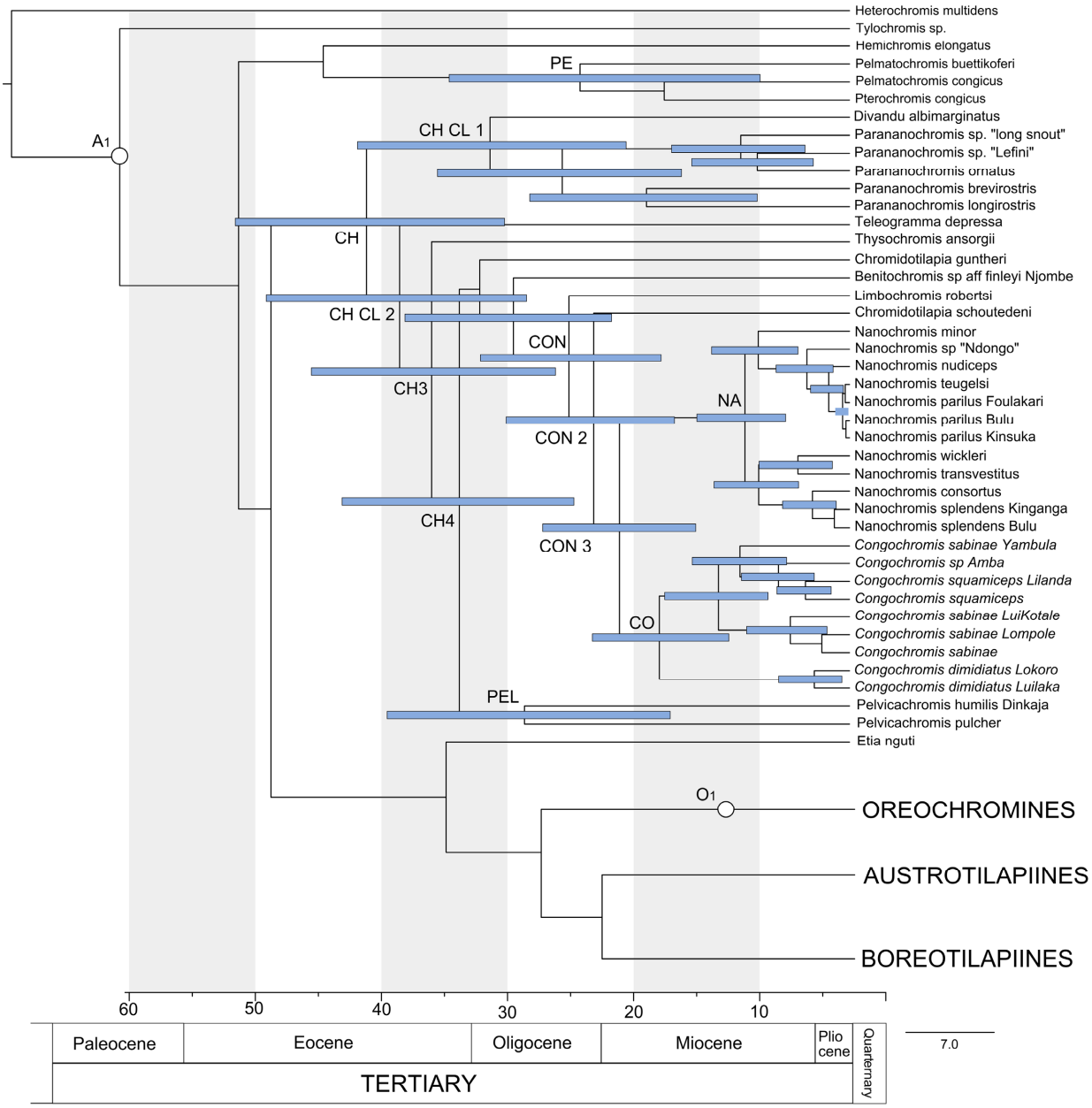


Figure 4 *Chronogram showing age estimates for chromidotilapiines*

The chronogram was calculated based on the BI consensus tree. Divergence times were estimated using a partitioned Bayesian analysis implemented in BEAST. Time constraints were used following (Schwarzer et al. 2009): A_1 53-84 mya (uniform prior), published age estimate based on non-cichlid fossils (Azuma et al. 2008) and O_1 5.98 mya (lower bound), the age estimate for *Oreochromis lorenzoi* † (Carnevale et al. 2003). The chronogram shows 95 % credibility intervals (HPC, blue bars). For nodes marked with letters, age estimates (95 % HPC and mean heights) are given in Table 1. The position of calibration points (O_1 and A_1) are indicated with circles.

28.4 (17.7 – 39) mya slightly more recent than clade 2 comprising the Congolian *Nanochromis*, *Congochromis* and *Teleogramma* as well as *Chromidotilapia*, *Limbochromis*, *Benitochromis*, *Pelvicachromis* and *Thysochromis* 35.5 (25.6 – 46) mya (Fig. 4).

Parananochromis species previously were assumed to be restricted to the Ogooué and Dja River system in Gabon and Cameroon (Lamboj & Stiassny 2003), but recent collections extend their distribution range to rivers Itimbiri and Lefini in the central Congo system (Snoeks & Stiassny, pers. comm., Fig. 1). Palaeogeographic hypotheses based on coastal sediment cores indicate, that the ancient (Oligocene) proto-Congo drained west through the Ogooué valley into the Atlantic, rather than through the present lower Congo or Niari-Kouilou (Leturmy et al. 2003). Age estimates for the MRCA of *Parananochromis* corroborate this (Table 1), supporting an ancient connection of northern and northwestern Congo tributaries (rivers Itimbiri, Lefini, Sangha and Dja).

Table 1 *Age estimates*

Priors A_1 and O_1 were taken from Schwarzer et al. (Schwarzer et al. 2009b) and resulting age estimates were compared with publishes studies when possible (Genner et al. 2007). Age estimates given for node A_1 from Genner et al. (Genner et al. 2007) correspond to their dataset calculated with Gondwana priors. Letters correspond to nodes are marked in Figure 2.

Node	Date estimates in Myr					
	Baysian Inference (95% credibility intervals)					
	This study		Schwarzer et al. (2011)		Genner et al. (2007)	
A_1	**56.4**	(53.0, 67.6)	**55.5**	(53, 63.9)	**63.7 (N)**	(46.6, 79.6)
O_1	**9.1**	(5.9, 12.7)	**12.3**	(10.2, 15.9)		
PE	**14.1**	(5.6, 25.9)	**23.3**	(18.1, 29.5)		
CH	**38.1**	(27.4, 48.7)				
CHCL1	**28.4**	(17.7, 39.0)				
CHCL2	**35.5**	(25.6, 46.2)				
CH3	**32.9**	(23.3, 42.7)				
CH4	**30.7**	(21.9,40.2)				
PEL	**25.9**	(14.2, 36.7)				
CON	**26.4**	(18.9, 35.3)				
CON2	**20**	(13.9, 27.2)				
CON3	**18**	(12.2, 24.4)	**27.9**	(23.5, 33.5)		
CO	**14.9**	(9.6, 20.4)	**22**	(17.9, 26.7)		
NA	**8.12**	(5.1, 12.1)	**8.4**	(6.5, 10.4)		

The phylogenetic position and estimated age of the splitting of *Teleogramma* from the remaining chromidotilapiine species in clade 2 is surprising, as present day representatives of *Teleogramma*, are endemic to the lower Congo rapids (three species) and the Kasai (one species). *Teleogramma* appeared as sistergroup to a chromidotilapiine subgroup, without *Parananochromis* and *Divandu* in the nuclear and the concatenated dataset (Fig. 2, Fig. 3), but appears (together with *Thysochromis*) as sistergroup to all chromidotilapiines based on mtDNA (Fig. 3). Resolution of the mtDNA tree, however, was low and it seems likely that the pattern was caused by insufficient resolution rather than by real discordant signal (caused e.g. by hybridization, Seehausen 2004). All *Teleogramma* species are strongly rheophilic and are characterized by a flattened and elongated body and a reduced swim bladder (Roberts & Stewart 1976). Three of four species inhabit the rapids of the lower Congo River, which are estimated to have formed about 5 mya (Ferry et al. 2004) and it is likely that the lower Congo *Teleogramma* radiation is not older than that. Beyond this background, this splitting of *Teleogramma* from the remaining chromidotilapiines (excl. *Parananochromis* & *Divandu*) was estimated to have appeared at the Eocene/Oligocene border, *Teleogramma* precursors now distributed in the Kasai most likely seeded the small radiation in the lower Congo River.

Radiations in the Congo Basin

The two chromidotilapiine genera *Nanochromis* and *Congochromis* form small species assemblages with a distribution center in the Congo basin, but exhibit clearly allopatric patterns. *Nanochromis* is restricted to the central part of the Cuvette Centrale (CC), Pool Malebo and the lower Congo rapids whereas *Congochromis* is distributed in the Congo-basin and the Ogooué system (*C. sabinae*), but is absent in the lower Congo rapids (Schliewen & Stiassny 2006; Stiassny & Schliewen 2007).

Intrageneric relationships within *Congochromis* were resolved based on the multi-locus approach with the widely distributed (northern Congo) *C. dimidiatus* as sistergroup to a mixed clade composed of *C. sabinae*, *C. squamiceps* and *C.* sp. “Amba” (Fig. 2). Relationships within the younger *Nanochromis* clade remain unresolved (Schwarzer et al. 2011). An explosive radiation or hybridization prior to species formation might be responsible for this pattern (Seehausen 2004). Earlier results focussing on intragroup relationships in *Nanochromis* based on AFLP markers yielded a better resolution but gave also hints towards hybridization in this genus (Schwarzer et al. 2011). Based on the concatenated dataset and the nuclear dataset the sistergroup to *Nanochromis* and *Congochromis* is *Chromidotilapia*

schoutedeni from tributaries of the middle Congo River/Lualaba River (upper central Congo) (Poll & Thys van den Audenaerde 1967). Based on the mtDNA, however, *Limbochromis robertsi*, endemic to the upper tributaries of rivers Pra, Ankasa and Ankobra basins in Ghana (~ 2000 km distance from the Congo basin, Fig. 1) appears as sistergroup to *Nanochromis* and *Congochromis*. Both conflicting phylogenetic alternatives are well supported (Fig. 4), indicating ancient reticulate signal or ancient incomplete lineage sorting.

Conclusions

The chromidotilapiines are the most diverse group of West/Central African cichlids. With the present study we largely resolve internal relationships and place for the first time the Congolian genera *Nanochromis* and the rheophilic *Teleogramma* in a phylogenetic framework. Age estimates of the widely distributed *Parananochromis* species support a proposed ancient connection of northern and north-western Congo tributaries. With the present dataset we elucidate phylogenetic relationships in one of the oldest African cichlid tribes (late Eocene/early Oligocene), which significantly contributes to the understanding of the evolution of the whole African cichlids and the emergence of their spectacular diversity.

Authors contributions

JS, UKS and AL designed the study. JS and KL carried out the molecular work. JS, BM and UKS designed and JS conducted the analyses. All authors contributed to the preparation of the manuscript. They read and approved the final version.

CHAPTER 4

Transcontinental hybridization among haplochromine cichlids (Cichlidae)

Julia Schwarzer [1,2] ,Swartz E[3], Vreven E[4], Snoeks J[4,5], Stiassny MLJ[6], Misof B[1] & Schliewen UK[2]

(1) Zoologisches Forschungsmuseum Alexander Koenig, Adenauerallee 160, 53113 Bonn, Germany

(2) Bavarian State Collection of Zoology, Münchhausenstr. 21, 81247 München, Germany

(3) South African Institute for Aquatic Biodiversity, Private Bag 1015, Grahamstown 6140, South Africa

(4) Laboratory of Animal Diversity and Systematics - Ch. Deberiotstraat 32 - B-3000 Leuven - Belgium

(5) Royal Museum for Central Africa, Leuvensesteenweg 13, 3080 Tervuren - Belgien

(6) American Museum of Natural History, Department of Ichthyology, Central Park West at 79th Street, New York, NY 10024-5192

This is the author's version of a work prepared for publication in: Proceedings of the National Academy of Sciences of the United States of America

Abstract

Hybridization has long been viewed as a process hindering diversity and speciation rather than supporting it. It has recently been shown, that the megadiverse cichlid radiations of the East African lakes are characterised by introgression of riverine representatives. Here we provide a first comprehensive mtDNA and nuclear DNA (AFLP) set focussing on an extensive sampling of riverine species including those from the lower Congo and Angolan Rivers. Reconstruction of phylogenetic hypotheses generated a paradox of clearly discordant nuclear and mitochondrial phylogenetic signal as closely related mtDNA haplotypes distributed thousands of kilometers apart and crossing major African watersheds, whereas neighboring species may carry drastically distinct haplotypes. Supported by the assessment of reticulate phylogenetic signal in the nuclear data, a novel phylogenetic network hypothesis is provided for one of the most species rich group of vertebrates, the haplochromine cichlids. Taking into account the complex palaeohistory of African water bodies and rivers a strong hybrid contribution of different lineages from different watersheds undoubtedly shaped each of the major haplochromine radiations in Lake Tanganyika, Victoria, Malawi and palaeo-Lake Magkadigadi, as well as the origin of a miniature species flock in the Congo basin (“Lac Fwa”).

Introduction

Cichlid fishes of the haplochromine lineage gave rise to one of the most spectacular vertebrate radiations on our planet, the cichlid radiations endemic to the east African Great Lakes and southern Africa (Turner et al. 2001; Joyce et al. 2005). Establishing phylogenetic hypotheses for haplochromine cichlids has proven to be difficult due to limitations in taxon sampling and phylogenetically informative characters. Until recently, phylogenetic analyses of East African cichlid radiations have implicitly supported the monophyly of each of the major haplochromine species flocks ("modern haplochromines" including the Lake Victoria superflock (Verheyen et al. 2003), southern African serrannochromines, Lake Malawi haplochromines and Tropheini of Lake Tanganyika, (Albertson et al. 1999; Verheyen et al. 2003; Joyce et al. 2005; Koblmüller et al. 2010)). However, Joyce et al. (2011) and Seehausen et al. (2003) falsified the assumed monophyly for Lakes Malawi and Victoria species flocks, based on multilocus nuclear data and after inclusion of riverine haplochromines. The only available comprehensive analysis of haplochromines including several important riverine lineages has built exclusively on mitochondrial DNA (Salzburger et al. 2005; Koblmüller et al. 2008a), which is unsuitable to detect reticulate signal due to their maternal inheritance (Chan & Levin 2005). Multilocus nuclear data sets that might overcome this limitation are only available for a subset of haplochromine taxa focussing on particular subgroups from the Lake Victoria region (Seehausen et al. 2003), Lake Malawi (Albertson et al. 1999) or Lake Tanganyika (Koblmüller et al. 2010). In these analyses riverine haplochromines inhabiting different regions the Congo basin and Angola are represented only by very few taxa (Congo basin) or are not represented at all (Angola). This is surprising, as the origin of diversification of haplochromines is assumed to be in Lake Tanganyika (Salzburger et al. 2005), which is part of the Congo basin. Furthermore, several studies have identified selected congolian haplochromines as sistertaxa to modern haplochromine sublineages (Seehausen et al. 2003), or as sistergroup to members of the serranochromine palaeo-Lake Makgadikgadi flock (Joyce et al. 2005; Salzburger et al. 2005; Koblmüller et al. 2008a).

Hybridization as a potential force promoting phenotypic diversification has been repeatedly suggested in theoretical considerations (Rieseberg & Linder 1999; Seehausen 2004; Bell & Travis 2005) and shown in empirical studies, including cichlids (Albertson & Kocher 2005; Parnell et al. 2008; van der Sluijs et al. 2008). Other studies have provided phylogenetic evidence for massive introgression and hybridization among ancient lineages in evolving

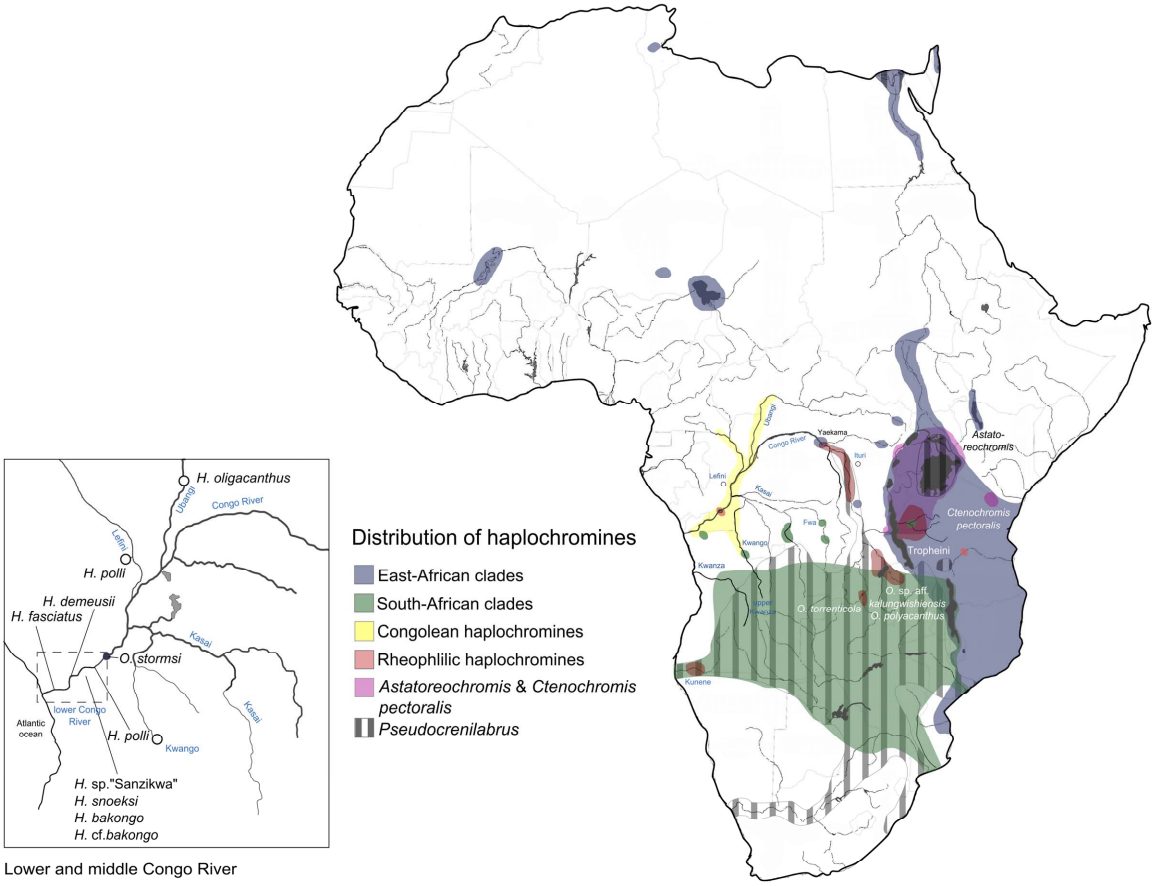


Figure 1 *Distributions of the major haplochromine clades*

Colour codes correspond to distribution ranges of haplochromine clades. Sampling locations along the lower and middle Congo River are presented in larger scale.

species flocks (Salzburger et al. 2002b; Schliewen & Klee 2004; Schelly et al. 2006; Herder et al. 2006; Mallet et al. 2007), but phylogenetic tests for a potential contribution of Hybridization to the evolution of haplochromine lineages which have given rise to the most spectacular vertebrate radiations remain scarce (but see Seehausen et al. 2003; Joyce et al. 2011). Here we use an extensive taxon sampling focussing on riverine haplochromines, including for the first time multiple Angolan and all to date available haplochromine species from the Congo area. Phylogenetic hypotheses were reconstructed based on more than 2000 AFLP loci as well as two mitochondrial genes. This, in combination with an experimental approach based on Seehausen (Seehausen 2004), enabled us to assess and quantify reticulate signal in the dataset and hereby appraise the impact of riverine lineages and ancient reticulate evolution on the genesis of the megadiverse cichlid lineages of the East African Lakes.

Methods

Sampling focussed on an extensive and representative coverage of all major haplochromine lineages (Table S1) and biogeographic regions (Fig. 1). *Tilapia bilineata* and *T. sp. aff. bilineata* as well as *Lamprologus* were chosen as outgroups based on results in Schwarzer et al. (2009). All voucher specimens were initially preserved in formalin and later transferred into 70% ethanol. Of these, tissue samples were directly preserved in 92-96% ethanol. Genomic DNA was isolated from fin clips or muscle tissue using the Qiagen Tissue Extraction Kit (DNeasy Tissue Extraction Kit) following the manufacturer's protocol. The concentration of all extracts was measured and adjusted to approx. 30 ng/µl. The mitochondrial ND2 and part of the cytochrome b gene (Cytb) were amplified and sequenced for 69 taxa (48 species) as described in Schwarzer et al. (2011). AFLP genotypes were obtained from the same individuals as sequenced for ND2 and cytb (N = 69). A modified protocol of the original AFLP method (Vos et al. 1995) as suggested in Herder *et al*. (2006) was used. The following twelve *EcoRI/MseI* primer pairs with three selective bases were used for selective AFLP amplification: ACA*-CAA, ACA*-CTT, ACC*-CTA, ACC*-CAG, AGC*-CAG, AGC*-CTT, ACC*-CTG, AGG*-CAC, AGG*-CTA, ACT*-CAA, ACT*-CAT, ACT*-CTC. Bands were visualized on an AB 3130 sequencer (Applied Biosystems) and Genemapper® v. 4.0. software using the size standard ROX 500 XL. Peaks between 100 and 499 bases could be scored unambiguously for presence/absence. The analysis was conducted automatically using Genemapper v. 4.0. Considering the standard error of automated sequencers, pairs of neighbouring bins whose minimum distance between each other was less than 0.25 bp and also bins containing fragments differing more than 0.65 bp in size were removed from the dataset. Eight individuals were genotyped twice to test for reproducibility. The error rate per individual was calculated as the ratio between observed number of differences and the total number of scored fragments (Pompanon et al. 2005), resulting in a mean error rate of 0.03.

Phylogenetic inference

Sequence alignment was conducted using ClustalW. Coding genes were translated into amino acid sequences to check for stop-codons or frame shifts and datasets were checked separately for saturation at each codon position. Base frequencies were equal for all markers (Chi-square tests, df = 183, all $p > 0.9$). Masking of ambiguous alignment positions was done using ALISCORE v. 0.2 under default settings (Misof & Misof 2009). We used for both datasets a partition separating first and second codon positions from the third. The GTR +Γ

best fitted for the first two whereas the HKY +Γ model was assigned to third codon positions based on results from the Bayes factor test (Nylander et al. 2004). Bayesian analyses were performed using MrBayes v. 3.1.2 (Huelsenbeck & Ronquist 2001) with two parallel runs each with 10^6 generations starting with random trees and sampling of trees every 1000 generations. To ensure convergence the first 10^5 generations of each run were treated as burn-in and excluded. The remaining trees from all Bayesian analyses were used to build a 50 % majority rule consensus tree. A Maximum Likelihood (ML) analysis was conducted using RAxML v. 7.0.3 (Stamatakis 2006) using the GTR+Γ model and the rapid bootstrap algorithm with following search for the best-scoring ML tree. Branch support was evaluated based on non-parametric bootstrapping (BS) consisting of 1000 pseudoreplicates. For the AFLP data a neighbour joining tree based on Link et al. (1995) distances was calculated using TREECON v. 1.3. The Link et al. algorithm takes only shared and unique bands into account while absent band are ignored. Bootstrap values were calculated based on 1000 pseudoreplicates. Phylogeny sequence data was incorporated into a bigger phylogenetic framework (N=145 taxa) using available ND2 and cytb sequences from Genebank (Table S2).

Inferring hybrid signal

Following Seehausen (2004), we applied a tree based method to test for homoplasy excess (HET) in our dataset. The expectation is that the inclusion of hybrid taxa increases the conflict in the dataset and reduces support values for affected nodes in a phylogenetic tree more than the inclusion of non-hybrid taxa (Seehausen 2004). The exclusion of a hybrid taxon from the dataset should therefore increase support values for affected nodes. This detection of potential hybrid signal focuses on the AFLP dataset, as hybridization cannot be detected in maternally inherited mitochondrial markers (Chan & Levin 2005). All clades showing discordant signal between the nuclear (nc) and mitochondrial (mt) trees as well as all monophyletic clades (in the nc- tree) were successively removed from the dataset resulting in 86 removal experiments (Fig. S1). Subsequently, distance trees based on the Link et al. algorithm (1995) were built for each reduced dataset with 500 bootstrap replicates using TREECON v. 1.3. The resulting trees and bootstrap support values were checked manually for all remaining clades. Results of the HET were visualized in boxplots for major phylogenetic nodes with initially low BS support values. To test for random effects on BS support, additional removal experiments were conducted with a certain number of randomly chosen taxa. The number of excluded taxa depended on the number of individuals

that caused an effect on node support and ranged from N = 1 to N = 6. For each N the random removals were repeated 100 times. Tree generation and BS evaluation was conducted as described above. A heat map based on bootstrap outliers was generated representing the change of bootstrap support values for all removal experiments over the whole dataset. Outliers were defined following (Tukey 1977), as data points located outside of the 1.5* inter-quartile distance displayed (in boxplots) as whiskers.

Results

The final AFLP (nc-) dataset was composed of 69 taxa (66 haplochromines, 3 outgroup taxa) with 2106 AFLP loci. Of these 1984 (1889 without outgroups) fragments were polymorphic. The ND2 dataset consisted of 1022bp and cytb dataset of 405bp (total = 1427bp) with 399 (ND2) and 130 (Cytb) variable sites and empirical base frequencies of A = 0.26, C = 0.35, G = 0.12, T = 0.27 and A = 0.24, C = 0.30, G = 0.17, T = 0.29, respectively. The large mitochondrial dataset (dataset LD) consisted of 1020 positions of ND2 and 380 of cytb with 476 (ND2) and 147 (cytb) variable sites with base frequencies of A=0.26, C= 0.35, G= 0.12, T=0.27 and A=0.23, C= 0.31, G= 0.16, T=0.30. One individual, *H. snoeksi,* failed to amplify for one AFLP primer combination (ACT-CAT*). The AFLP analysis was subsequentely also conducted with a reduced dataset of eleven primer-combinations, showing no topological differences (data not shown).

mtDNA and AFLP based phylogenetic hypotheses

Analyses of AFLPs (nuclear DNA) and mitochondrial (mtDNA) genes yielded well resolved phylogenetic hypotheses. In the AFLP dataset eleven well supported phylogenetic clades are present (Fig. 2). A mostly congruent biogeographic division into an eastern, Congolean and southern group can be derived based on the AFLP topology. *Haplochromis* cf. *bakongo* and *H*. *snoeksi* from lower Congo tributaries, however, appear closer to the southern clade (Fig. 2), rendering a Congolean clade paraphyletic (Fig. 2). Based on the AFLP dataset the single included *Pseudocrenilabrus* captures a position as sistergroup to all remaining haplochromines (Fig. 2, BS = 99). The rheophilic haplochromines composed of members currently assigned to *Orthochromis*, i.e. *O. stormsi*, O. cf. *stormsi* “Kisangani”, O. *polyacanthus* and *O.* sp. aff. *kalungwishiensis* then form a sistergroup to the Tropheini from Lake Tanganyika, the East African clades, the Congo clade, *H. snoeksi* and *H*. cf. *bakongo* and the southern clades (BS = 78). Members of the East African clades (BS = 92) appear as

sistergroup to the Tropheini (BS = 99) from Lake Tanganyika (Fig. 2, BS = 62). Based on the nc-dataset *Astatoreochromis alluaudi* and *"H." burtoni* capture a sistergroup position to the remaining East African clades (BS = 92, BS = 100). Based on the mt-tree, however, *A. alluaudi* is sistergroup to the East African clades including the Tropheini (BS <50) and "*H." burtoni* and the L. Victoria "superflock" (BS = 51, BPP= 0.64). Lake Malawi haplochromines form a monophyletic group in both trees (BS = 100, BS = 81/0.93). Interestingly, *"H."* sp. "Yaekama" falls into the Lake Victoria clade based on both trees (Fig. 2, BS = 100 and 100/1.0) even though it is distributed in the upper Congo drainage (near Kisangani). The whole Eastern clade (BS=62) appears as sistergroup to the Congo clade (BS = 100), "*H." snoeksi* and "*H*." cf. *bakongo* (BS = 76) and the southern clades (BS = 60) based on the nc-dataset. The integrity of the southern clades is supported by a low bootstrap support only, as is the sistergroup relationship of *"H."* cf. *bakongo* and "*H*." *snoeksi* to the South African clades (Fig. 2, BS=60). Within the Congo clade, the Pool Malebo and central Congo basin *"H." polli* and "*H." oligacanthus* and the three Lower Congo rapids species "*H." fasciatus*, "*H*." *demeusii* and "*H*." sp. "Sanzikwa" appear monophyletic (BS = 100). Within the southern clades, all included species from rivers Fwa, Kasai and Kwango form a monophyletic group (BS = 79) which is sistergroup (BS = 60) to species from the Angolan Kwanza system and *"O." torrenticola* (Fig. 2, BS = 83). Except for the *O. torrenticola* and the *Serranochromis* sp. "red scales" clades none of the major clades supported by the nc-tree is recovered in the mt-phylogeny (Fig. 2).

Two well supported, but geographically heterogenous clades are recovered in the mt-phylogeny. One is composed of members of the East African clade and "*H." fasciatus*/"*H." demeusii* and "*H*." sp. "Sanzikwa" from the lower Congo River (BS = 79/0.96) with the Tropheini as sistergroup (79/0.96) and the other one of "*H." snoeksi*/"*H*." cf. *bakongo*/"*H*." sp "Lefini" (92/1.0) and species from River Fwa (92/1.0) as sistergroup to *Serranochromis* sp. "red scales", *O. stormsi*/ *O*. cf. *stormsi* "Kisangani" and *O. polyacanthus,* members of the southern clades and the congolean clade composed of "*H." polli* and "*H." oligacanthus* (81/0.95). *"H." polli* and "*H." oligacanthus* appear as sistergroup to species from South African Rivers Kwango, Kasai, upper Kwanza and the proposed ancient Lake Magkadigadi radiation (68/0.99, Fig. 2). Phylogenetic patterns based on the mt-topology were recovered

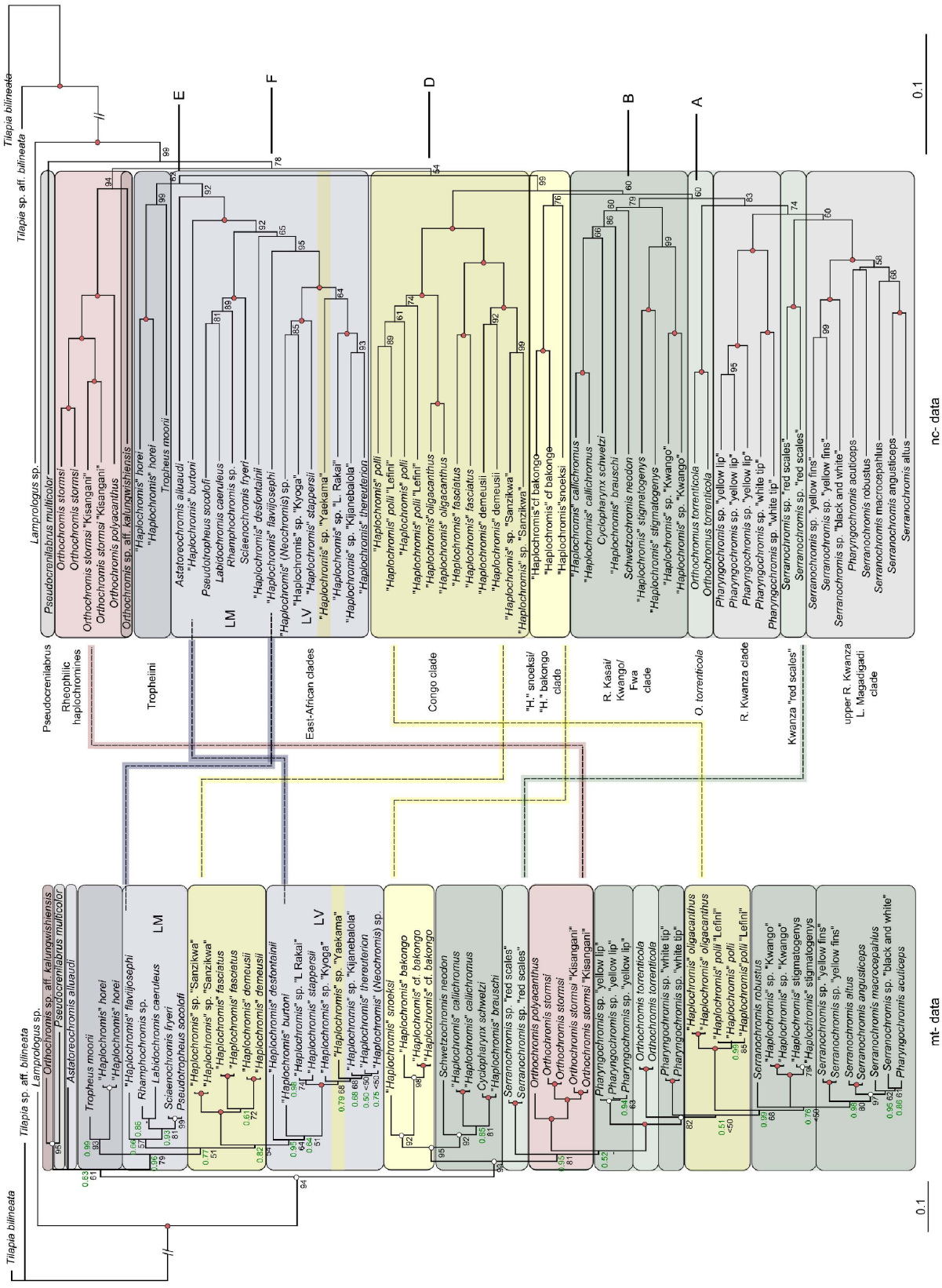


Figure 2 *(caption overleaf)*

Figure 2 *Phylogenies based on mt- genes and AFLP data*

The datasets comprise mitochondrial sequences of ND2 and Cytochrome b and over 2000 AFLP markers. Black numbers on nodes refer to bootstrap-values (BS, 1000 pseudo-replicates) of the ML run (left side) or the neighbour joining tree (right side). Green numbers refer to posterior probabilities from the Bayesian inference (BPP). Filled circles represent 100% BS (right side) or 100/1.0 BS- and BPP support (left side). Empty circles on the mt-tree indicate 1.0 BPP and a lower bootstrap value. Biogeographic affiliations within the phylogeny are marked with coloured frames (corresponding to Fig. 1). Green = South African clades, yellow= Congolian clades, blue = East African clades and the red frames refer to the *Orthochromis* species. Detailed distribution ranges are shown in Fig. 1. Species or clades that are placed differently in AFLP and mt-trees are connected by dotted lines.

in a more extensive phylogenetic analysis based on ND2 and Cytb, supporting the reliability of the main mt- analyses despite limited taxon sampling (Fig. S2).

Cytonuclear discordance and homoplasy excess test

Cytonuclear discordances indicating hybridization events are present throughout the whole haplochromine phylogeny. Major discrepancies between the mt-DNA and AFLP phylogenetic hypotheses are: (1) The Congolean species (without "*H." snoeksi* and "*H*." cf. *bakongo*) appear monophyletic in the nc-tree (Fig. 2), but based on the mtDNA dataset the two Congolean subclades ("*H." polli*/"*H." oligacanthus* and "*H*." *fasciatus*/"*H." demeusii*/"*H*." sp. "Sanzikwa") are nested within southern clade species or east African clades, as sistergroup to members of the L. Victoria superflock. *"Haplochromis" snoeksi* and "*H*." cf. *bakongo* form a sistergroup to the southern clades based on nc-data, but in the mt-tree they are sistergroup to species from River Fwa (*Schwetzochromis neodon*, "*H." callichromus*, "*H." brauschi* and *Cyclopharynx schwetzi*, Fig.2). Within the River Fwa species "*H." brauschi* appears as sister to *C. schwetzi* and "*H." callichromus* based on the nc- (BS = 86), but to *C. schwetzi* based on the mtDNA dataset (100/1.0). (2) The rheophilic haplochromines *O. stormsi*/*O. stormsi* "Kisangani", *O. polyacanthus* and *O. kalungwishiensis* are based on the nc-dataset monophyletic (BS = 94) but *O. stormsi*/*O. stormsi* "Kisangani" and *O. polyacanthus* are nested within a clade of southern species and the "*H." polli*/"*H." oligacanthus* clade (81/0.95) and *O. kalungwishiensis* appears as sistergroup *Pseudocrenilabrus multicolor* based on the mtDNA dataset (95/1.0). (3) *Astatoreochromis alluaudi* forms the sistergroup to remaining species from East- African clades based on

nuclear DNA data (BS = 92), but remains unresolved at a basal position based on the mtDNA dataset.

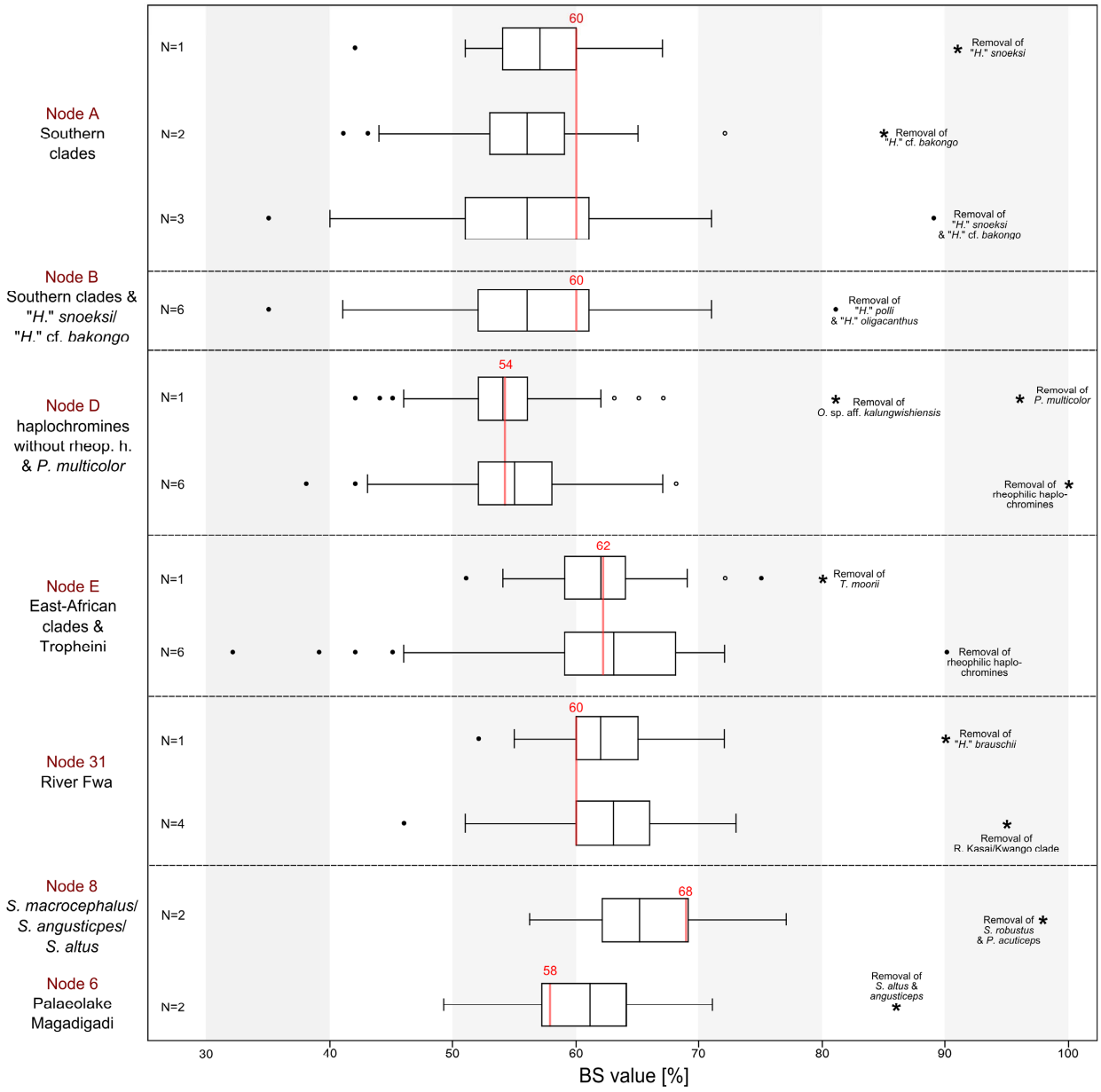


Figure 3 *Homoplasy excess tests of nodes with low BS support values in the AFLP tree*

The boxplots show the distribution of bootstrap support values [%] for selected basal nodes with initially low BS support values. The zero-distributions (derived from 100 randomly conducted removals) are shown as empty boxes. The zero-distributions conducted with a certain number of randomly chosen taxa. The number of excluded taxa depended on the number of individuals that caused an effect on node support and ranged from N = 1 to N = 6. For each N the random removals were repeated 100 times.

Further cytonuclear discordances within the East-African clades are present concerning the position of "*H." flaviijosephi* and "*H." burtoni. "H." flaviijosephi* is sister to members of the L. Victoria superflock based on nc- (BS = 95) but to the L. Malawi clade based on mt- data (66/0.57). *"H." burtoni* appears as sister to the remaining east African clades (without *A. alluaudi*) based on the nc- (BS = 92), but is nested in a clade composed of to the L. Victoria "superflock" and *H. desfontanii* based on the mtDNA dataset (64/0.95). Within the southern clade, several discordances are obvious concerning within-group relationships (Fig. 2). *Serranochromis robustus* clusters within the upper Kwanza/L. Magkadigadi clade based on nuclear DNA data (BS = 100) but not based on mtDNA data (Fig. 2), where it appears in a clade with Kwango/Kasai and Kwanza/L. Magkadigadi species (< 50/0.76). Discordances concerning whole South African clades are, except for species from River Fwa, comparatively weakly supported (Fig. 2).

In 86 removal experiments (Fig. S1) effects on different BS support values are evident across the whole haplochromine phylogeny (Fig. 3). Strong effects (more than 50% in- or decrease based on the mean bootstrap value) exist concerning the support values of five nodes (6, 31, 70, A, D, Fig. 4, Fig. S1) and medium effects (25% in-or decrease of BS support) are present on six additional nodes (7, 8, 20, 29, B and E, Fig.4, Fig. S1). Removals causing an increase of BS support (indicating a decrease of homoplastic signal in the dataset) are caused mainly by members of the following subclades: Upper Kwanza/L. Magkadigadi, "*H." snoeksi*/ "*H*." cf. *bakongo* and Congo clades, Tropheini, Lake Victoria superflock, riverine haplochromines and *P. multicolor* (Fig. 3). Box plots were generated for nodes yielding an initially low BS support in the AFLP tree (Fig. 4, indicated in Fig. 2). An exclusion of *"H."* cf. *bakongo* and "*H*." *snoeksi* entails an increase of BS support for Node A comprising the southern clades (Fig. 4, BS = 60 to 89). Exclusion of "*H." polli* and "*H." oligacanthus* increase the BS support of Node B, comprising the southern clades and "*H."* cf. *bakongo*/"*H*." *snoeksi* (Fig. 4, from BS = 60 to 81) and the removal of *T. moorii* (but not *"H." horei*) and of the rheophilic haplochromines leads to an increase of BS-value for the East African clades and Tropheini (Node E, Fig. 4, BS = 62 to 80 or 90). The BS support for the node comprising all haplochromines excl. the rheophilic haplochromines and *Pseudocrenilabrus* increases, when all rheophilic haplochromines, *O. kalungwishiensis* or *P. multicolor* are removed (Node D, Fig. 4, from BS = 54 to 82, 96 or 99 respectively). Effects on node support values within species from R. Fwa are only present when "*H." brauschi* is removed (Node 29 and 31, Fig. 3, Fig. 4).

Discussion

Haplochromine cichlids are well known for their ability to rapidly adapt and form megadiverse lacustrine adaptive radiations (Seehausen 2000; Seehausen 2006; Koblmüller et al. 2008a). Recent molecular studies indicate a significant impact (through hybridization) of riverine haplochromines on L. Malawi (Joyce et al. 2011). The inclusion of additional riverine haplochromines covering almost the whole range of their distribution in the present study, however, highlights for the first time to which large extent hybridization has shaped the evolution of haplochromines. Nuclear data, based on more than 2000 AFLP markers, reflect close relationships of geographically adjacent haplochromine species and clades (Fig.1 and Fig. 2). The mtDNA-data, however, yield clearly conflicting phylogenetic hypotheses. Theoretically, this might either be explained by ancient shared polymorphisms as a result of incomplete lineage sorting or introgression after secondary contact (Seehausen 2004). A strong argument against incomplete lineage sorting is the unequal spatial distribution of well separated mtDNA haplotypes (mean sequence divergence 0.086, s.d. = 0.029, Fig. 2). Further evidence comes from removal experiments showing that homoplasy excess is induced by single species or clades and not randomly (Fig. 3, Fig. 4). Stelkens et al. (2010) demonstrate experimentally that viable hybrids may be produced even among distantly related haplochromine species. They state that after lineage separation evolution of pre-mating isolation rapidly increases in haplochromine cichlids, but later nearly stagnates with increasing genetic distance. Surprisingly, viable hybrid offspring can be produced up to an upper temporal limit of estimated species divergence of 4.4/8.5/18.4 mya (Stelkens et al. 2010, depending on the underlying molecular clock algorithm and priors).

In general, the likelihood for interspecific gene flow should be higher in species with spatial proximity. Phylogenetically closely related mtDNA haplotypes are shared between parts of the Congolean, southern African and the rheophilic haplochromines *O. stormsi* and *O. polyacanthus* (Fig. 1, Fig. 2), and genetic exchange is indicated based on nuclear data as well (Fig. 3, Node B, Fig. 4). This can be explained by a potential connection of the Congolean species to southern (Zambesian) haplochromines, which could have existed through upper reaches of southern Congo tributaries (e.g. rivers Kwango and Kasai). A second mtDNA haplotype clade is shared by species from R. Fwa and Rivers Inkisi and Kwilu (*"H." snoeksi* and *H*. cf. *bakongo*), affluents of the lower Congo River, approx. 900 km away. Neighbouring species to the R. Fwa (from rivers Kwango and Kasai) carry clearly distinct mtDNA haplotypes

(Fig. 2). The Fwa haplochromines appear monophyletic and are locally endemic indicating some kind of, potentially ecological, isolation (see also Roberts & Kullander 1994). Seehausen et al. (2003) suggested *Haplochromis brauschi* from R. Fwa and *H. flaviijosephi* as sistergroup to the LV superflock. The inclusion of additional species from R. Fwa and more nuclear markers in the present study contradicts this inference and renders an East African/nilotic origin of the LV superflock more likely, possibly with a hybrid contribution of different lineages (Fig. 2).

Transcontinental dispersal of East African haplotypes into the lower Congo

A close sistergroup relationship of mtDNA haplotypes is evident between east African haplochromines and a restricted group of narrow endemics of the lower Congo rapids (Fig. 2). Interestingly, "*Haplochromis"* species of the lower Congo rapids are parapatrically distributed, corresponding to a trisection of the river stretch (Schwarzer et al. 2011) in upstream ("*H*." *polli*), central ("*H*." *demeusii*) and downstream ("*H*." *fasciatus*) lower Congo. The distribution of "*H*." *polli* expands further upstream into River Lefini (Fig. 1). The restriction of "*H*." *fasciatus*/"*H*." *demeusii* to the central and downstream stretch of the lower Congo makes it even more surprising that they, contrary to their closest relatives (according to ncDNA data) "*H*." *polli*/"*H*." *polli* "Lefini" and "*H*." *oligacanthus*, share haplotypes closely related to eastern African haplochromines radiations (Fig. 2). Surprisingly, there is no trace of introgression (indicated by homoplasy excess) detectable in the ncDNA dataset (Fig. 4). A complete replacement of mtDNA haplotypes can occur without any evidence of nuclear introgression. This was for example shown for arctic char (Wilson & Bernatchez 1998), mountain hare (Alves et al. 2006, Melo-Ferreira et al. 2005) and green pond frogs (Liu et al. 2010). The mtDNA haplotype can become fixed by chance (drift) or by positive selection (Ballard & Whitlock 2004). As there are no obvious indications for a selective advantage, potentially drift, through spatial isolation of the downstream lower Congo species (indicated by Schwarzer et al. 2011), triggered the retention of the ancient, eastern mtDNA haplotypes. An early connection of the Congo River with eastern Africa including the L. Victoria region and the East African rift valley is indicated by the recently detected multiple occurrence of "modern" haplochromines in the Upper Congo River: Both, *"H."* sp. "Yaekama" from the upper Congo near of Kisangani, as well as a recently discovered "*Haplochromis"* from the River Ituri carry ocellated egg-spots typical of east African haplochromines of the modern Haplochromines (Salzburger et al. 2005). Both species cluster

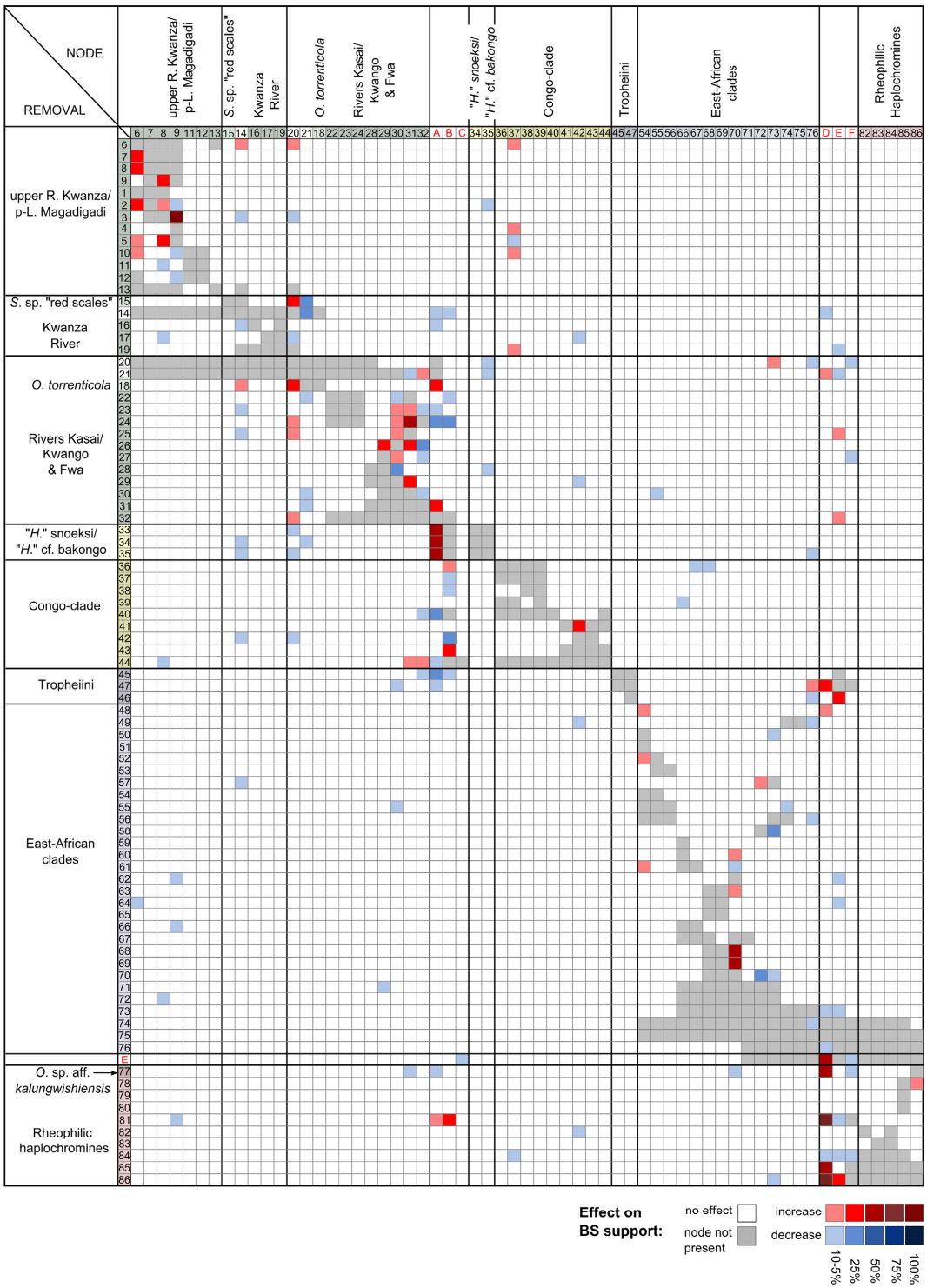


Figure 4 *Heatmap of all removal experiments representing changes in BS support values*

A heat map based on bootstrap outliers was generated representing the change of bootstrap support values for all removal experiments over the whole dataset. Outliers were defined as data points located outside of the 1.5* inter-quartile distance displayed (in boxplots) as whiskers. Each clade and single species were removed iteratively from the dataset (resulting in N = 86 experiments, Fig. S1). Subsequently NJ-trees based on 500 bootstrap replicates were recalculated using Treecon v. 1.3. Numbers correspond to node names and removal experiments specified in Fig. S1.

with members of the L. Victoria superflock in mtDNA and ncDNA in our dataset (Fig. 2), hereby defining the present day most western distribution of "modern" east African haplochromines.

Lacustrine origin of species diversity?

The *hybrid swarm hypothesis* predicts that hybridization between distantly related lineages can favour the onset of rapid adaptive radiations (Seehausen 2004). Empirical evidence comes from different studies in animals and plants (Rieseberg et al. 1999; Shaw 2002; Hudson et al. 2011), and initial hybridization most likely shaped cichlid radiations of palaeo-Lake Magkadigadi (Joyce et al. 2005) and L. Malawi (Joyce et al. 2011). A sistergroup relationship of the upper River Kwanza species and the palaeo-Lake Magkadigadi haplochromines is supported based on our data (Fig. 2). The occurrence of closely related mtDNA haplotypes from Congolian and lower Kwanza haplochromines (Fig. 2) and homoplasy excess present in palaeo-Lake Magadigadi species (Fig. 4) supports the hypothesized hybrid origin of the ancient lake radiation. A biogeographic connection could have existed through the Kwanza and Congo system and southern African Rivers. This strongly indicates an originally riverine origin of the southern clades, including the proposed palaeolake Magkadigadi radiation (Joyce et al. 2005).

The included four (out of five existing) species from River Fwa appear monophyletic (with a low BS support) and are closely related to neighbouring species from Rivers Kwango and Kasai based on the nuclear DNA dataset (Fig. 2). The removal of R. Kasai and R. Kwango species as well as of the Fwa species "*H." brauschi* entails an increase of BS-support for R. Fwa monophyly (Fig. 3, Node 31 in Fig. 4). This indicates that they carry a high degree of reticulate ("homoplastic") genetic signal suggesting a hybrid origin of the radiation influenced by distantly related southern and Congolean lineages.

Finally, removal of the basal haplochromines lineages (based on nuclear DNA data) *O.* sp. aff. *kalungwishiensis*, *O. stormsi/O. polyacanthus* and *Pseudocrenilabrus* leads to an increase of BS support for the whole remaining haplochromine phylogeny (Node D, Fig 3). Exclusion of *O.* sp. aff. *kalungwishiensis*, and *O. stormsi/O. polyacanthus* entails the increase of BS support for the East African clades including Tropheini (Node E, Fig 3, Fig. 4). *Orthochromis* sp. aff. *kalungwishiensis*, *P. multicolor* and *Asatoreochromis alluaudi* carry mtDNA haplotypes ancestral to those of the East African clades including Tropheini, whereas *O. stormsi* and *O. polyacanthus* species possess southern mtDNA haplotypes. Salzburger et al.

(2002a; 2005) proposed that the origin of haplochromine diversity originated in L. Tanganyika (LT). According to their "Out of Tanganyika hypothesis" members of an ancestral Tanganyika lineage left Lake Tanganyika, spread through eastern and southern Africa, colonized the Lake Victoria region, Lake Malawi and south-eastern Africa (including palaeo-Lake Magkadigadi) basins and explosively speciated faster than any other vertebrate lineage (Salzburger et al. 2005; Koblmüller et al. 2008a). Results of the present study highlights the involvement of ancient (*Orthochromis* and *Pseudocrenilabrus*) and recent riverine lineages on the evolution of lacustrine haplochromine radiations through hybridization. This does, however, not generally doubt a lacustrine origin of the East African clades, but clearly draws attention to a much more complex history of the megadiverse haplochromine radiations than previously assumed.

Repeated wet and dry periods in the Holocene (Gasse 2000; Russell & Johnson 2005) also affected the entire East African region as shown from sediment core analyses of L. Victoria and L. Tanganyika (Cohen et al. 1993; Johnson et al. 2000). In addition, at the Miocene-Pliocene transition the southern African continent was significant uplifted inducing a progressive rearrangement of watersheds (Lavier et al. 2001) involving the drainage of the present day upper Congo affluents, as well as initializing the origin of the modern lower Congo River rapids. This ongoing hydrological rearrangement undoubtedly has shaped the spatial distribution of haplochromines by allowing for novel connections of previously isolated watersheds, hereby enabling extensive hybridization between of previously separate lineages

Conclusions

Hybridization among even nowadays distant haplochromine lineages massively shaped the evolution of haplochromines. Our results demonstrate a transcontinental spread of mtDNA haplotyopes present within haplochromines, hereby falsifying simplistic assumptions about the proposed monophyletic origin of major haplochromine lineages. These analyses ask for more rigorous phylogenetic analyses including tests of multiple reticulate hypotheses on the basis of a fully representative taxon sampling of both riverine and lacustrine lineages. Only by doing so, it will be possible to assess the role of phylogenetic constraints as decisive factors shaping the origin and speciation of megadiverse haplochromine species flocks.

Authors contributions

JS and UKS designed the study. JS carried out the molecular work. JS, BM and UKS designed and JS conducted the analyses. ES, ES, JS and MLJS provided crucial samples. All authors contributed to the preparation of the manuscript. They read and approved the final version.

Acknowledgments

This work was supported by grants of the Deutsche Forschungsgemeinschaft to B.M. (DFG MI649/8-1) and U.K.S. (SCHL567/4-1) and a graduate student grant of the University of Bonn as well as a travel grant from the supporters of the Zoologisches Forschungsmuseum A. Koenig to J.S. We gratefully thank O. Seehausen and L Rüber for providing tissues. For indispensable assistance in the field we thank R. Schelly, D. Neumann, J. Frommen, P. Alibert, A. Dunz, M. Levy, V. Mamonekene, A. Ibal-Zamba, J. Punga, U. Ali-Patho, P. Mongindo, C. Danadu, F. Bapeamoni and T. Kadangé Ngongo. Special thanks go to K. Langen, F. Eppler and B. Müller who helped with lab routines.

CHAPTER 5

Time and origin of cichlid colonization of the lower Congo rapids

Julia Schwarzer[1,2*], Bernhard Misof[2], Seraphin N. Ifuta[3]& Ulrich K. Schliewen[1]

(1) Zoologisches Forschungsmuseum Alexander Koenig, Adenauerallee 160, 53113 Bonn, Germany

(2) Bavarian State Collection of Zoology, Münchhausenstr. 21, 81247 München, Germany

(3) Department of Biology, ISP-GOMBE, Kinshasa-Gombe, Democratic Republic of Congo

This is the author's version of a work originally published by the public library of Science in:

PLoS ONE 20116(7): e22380.

Abstract

Most freshwater diversity is arguably located in networks of rivers and streams, but in contrast to lacustrine systems riverine radiations are largely understudied. The extensive rapids of the lower Congo River is one of the few river stretches inhabited by a locally endemic cichlid species flock as well as several species pairs, for which we provide evidence that they have radiated *in situ*. We use more that 2,000 AFLP markers as well as multilocus sequence datasets to reconstruct their origin, phylogenetic history, as well as the timing of colonization and speciation of two Lower Congo cichlid genera, *Steatocranus* and *Nanochromis*. Based on a representative taxon sampling and well resolved phylogenetic hypotheses we demonstrate that a high level of riverine diversity originated in the lower Congo within about 5 mya, which is concordant with age estimates for the hydrological origin of the modern lower Congo River. A spatial genetic structure is present in all widely distributed lineages corresponding to a trisection of the lower Congo River into major biogeographic areas, each with locally endemic species assemblages. With the present study, we provide a phylogenetic framework for a complex system that may serve as a link between African riverine cichlid diversity and the megadiverse cichlid radiations of the East African lakes. Beyond this we give for the first time a biologically estimated age for the origin of the lower Congo River rapids, one of the most extreme freshwater habitats on earth.

Introduction

The rapids of the Lower Congo River rank among the most spectacular habitats for animal life on earth. The Congo basin harbours the highest fish species richness of any river system on the African continent (ca. 700 described species, see also Teugels & Guegan 1994), and especially the lower Congo exhibits a remarkable hydrological, spatial and ichthyological complexity along a short and narrow river stretch (Robert 1946; Roberts & Stewart 1976; Jackson et al. 2009). Before reaching the Atlantic Ocean, all water collected in a drainage basin encompassing one eighth of the African continent (ca. 3.8 mil km^2) is flushed through an intermittently narrow and deep (up to 200 m) rocky channel, creating the world's most extensive rapids (Runge 2008). This approx. 350 km long river section extending from the first rapids near Kinshasa down to the last one upstream from Matadi is geologically young. Its origin is most likely related to a river capture event, i.e. a small coastal river hypothetically tapped the interior Congo basin (sometimes referred to as "Palaeo-lake Congo"), and subsequently created a novel outlet for the whole Congo drainage (Runge 2008). Probably before this event, in the Pliocene, the Atlantic Rise had dammed the course of the Congo River hereby creating a large endorheic basin ("lake") in the western Congo basin, which nowadays is thought to survive in part as Malebo Pool (Runge 2008). Previous assumptions that this lake covered the whole Cuvette Centrale are unlikely (Runge 2008). Age estimates of the river capture are imprecise, varying from early estimates as young as 0.4 mya (Colyn 1991) to 34 mya (Leturmy et al. 2003; Lucazeau et al. 2003). Dozens of fish species are endemic to the lower Congo (Roberts & Stewart 1976) including representatives of the cichlid genera *Steatocranus*, *Nanochromis, Lamprologus*, *Teleogramma* and "*Haplochromis*". The distribution of all five genera except for the nearly pan-African catch-all genus *"Haplochromis"* is restricted to the Congo basin (van Oijen et al. 1991).

In general, cichlids (Perciformes, Cichlidae) are among the most species rich vertebrate groups. Most of their diversity evolved in the great lakes of East Africa, e.g. Lake Malawi, L. Victoria and L. Tanganyika (ca. 1500 species, Turner et al. 2001; Kocher 2004). Their morphological, behavioural and ecological diversity confined to these single water bodies is considered ideal to study patterns and processes of speciation, which is the reason that African lacustrine cichlid species flocks have been established as evolutionary model systems (Kocher 2004; Seehausen 2006). In contrast, riverine African cichlid species have been poorly studied. One reason might be that many riverine genera are species poor and exhibit limited

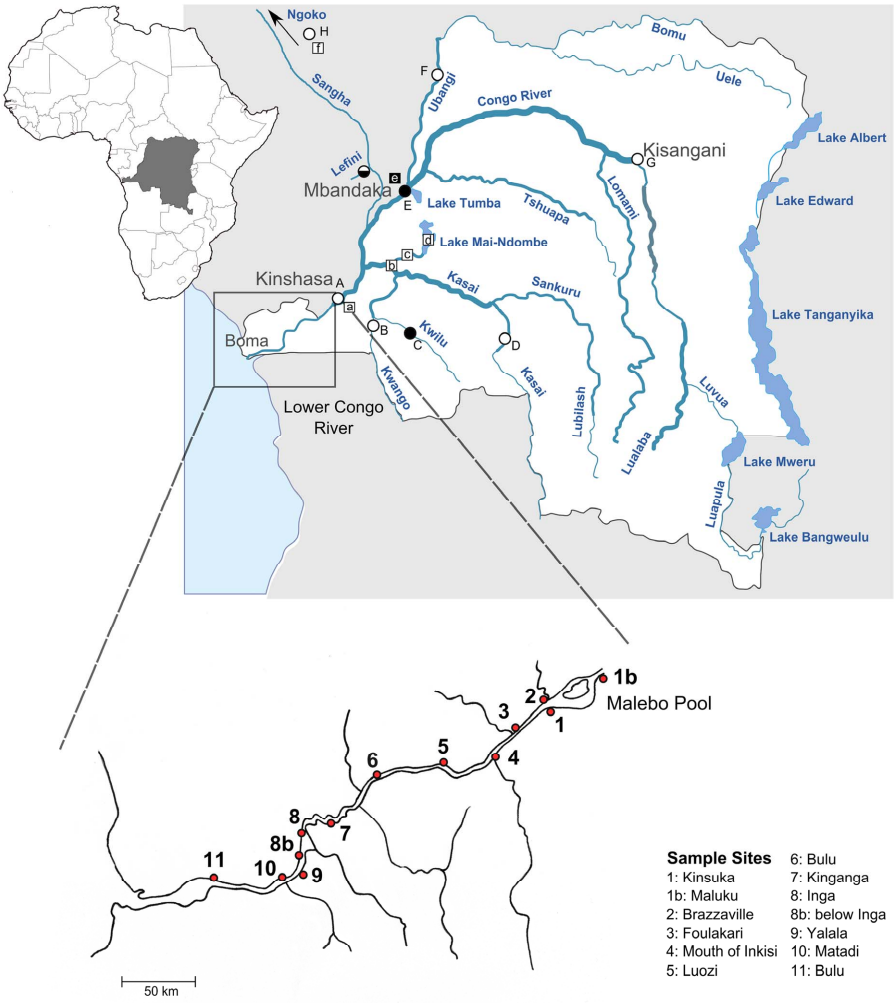


Figure 1 *Location map of sampling sites*

Circles correspond to *Steatocranus* and squares to *Nanochromis* sampling sites. The filled circle and square mark the location of unsamples species. The half filled circle indicate that the species was included in the sequence but not in the AFLP analysis. Letters correspond to non-lower-Congo species distributions: (a) *Nanochromis parilus*, (b) *N. teugelsi*, (c) *N. nudiceps*, (d) *N. transvestitus*, *N. wickleri*, (e) *N*. sp. "Mbandaka", (f) *N*. sp. "Ndongo", (A) *S*. sp. "dwarf", *S*. sp. "bulky head", *S. bleheri*, *S*. sp. "Maluku" (B) *S*. sp. "red eye", (C) S. sp. "Kwilu" (D) *S. rouxi*, (E) *S*. sp. "Mbandaka", *S*. sp "Maluku" (F) *S. ubanguiensis,* (G) *S*. sp. "Kisangani" and (H) *S*. sp. "Nki. Sampling locations along the lower Congo are presented in larger scale and marked by red circles. Numbers along the lower Congo correspond to the sample site legend in the lower right side. White background highlights the territory of the Democratic Republic of Congo.

morphological diversity, which makes them less attractive study subjects. Up to now the single notable exception is a species complex of southern African rivers, the serranochromines, which may have originally radiated under lacustrine conditions in the

now extinct Lake Palaeo-Makgadikgadi (Joyce et al. 2005). A convincing explanation for the low cichlid species numbers observed in riverine systems is still lacking. One proposed hypothesis is that fluvial systems already inhabited by ecologically diverse fish assemblages generally lack the multiplicity of ecological opportunities necessary for the formation of adaptive radiations (Joyce et al. 2005). If this is correct, riverine diversification should be best explained by vicariance and geographic isolation and less so by ecological differentiation (Katongo et al. 2005; Joyce et al. 2005; Katongo et al. 2007). However, the general applicability of this hypothesis remains untested.

Species of the two lower Congo cichlid genera *Steatocranus* and *Nanochromis* show a scattered distribution of few predominantly allopatric species within the Congo basin (Fig. 1) with a remarkable peak of recently discovered species richness (described and undescribed) endemic to the lower Congo ($N_{Steatocranus}$ =10, $N_{Nanochromis}$ =3). Apart from the lower Congo, *Nanochromis* species are distributed mainly south to the central Congo basin in Lakes Tumba and Mai Ndombe and adjacent rivers (*N. wickleri* and *N. transvestitus*, *N. nudiceps*) or in the Kasai River drainage (*N. teugelsi*). One undescribed species (*N.* sp. “Ndongo”) has recently been discovered in rivers Ngoko and Sangha (pers. obs.), both forming a northwestern Congo tributary, and another yet undescribed species (*N*. sp. “Mbandaka”) is known from the Congo mainstream around Mbandaka (Numrich 2003). *Nanochromis parilus* is distributed in the lower Congo but also above Malebo Pool e.g. at Maluku. *Steatocranus* species occurring outside the lower Congo are distributed either in northern tributaries (*S*. sp. “Nki”, *S. ubanguiensis* and *S*. sp. “Lefini” from Ngoko, Ubangi and Lefini rivers) or south to the Congo mainstream (*S. rouxi*, *S*. sp. “red eye”, S. sp. “Kwilu” from Kasai, Kwango and Kwilu rivers), or in the Congo proper (*S*. sp. “dwarf”, *S*. sp. “bulky head”, *S. bleheri*, *S*. sp. “Maluku”, *S*. sp. “Mbandaka” and *S*. sp. “Kisangani” from around Malebo Pool, Maluku, Mbandaka and Kisangani). The haplotilapiine genus *Steatocranus* consists of rheophilic species whereas within the chromidotilapiine genus *Nanochromis* adaptations to high current are less obvious (Roberts & Stewart 1976; Schliewen & Stiassny 2006; Schwarzer et al. 2009). Habitat preferences differ between the mainly rock-dwelling *Steatocranus* and the more sand-dwelling *Nanochromis* (pers. obs.). Lower Congo *Steatocranus* are characterized by divergence in trophic traits indicating ecologically differentiated trait utility, i.e. dentition used for algae scraping (“aufwuchs feeding”), molluscivory and drift feeders (Roberts & Stewart 1976, pers. obs.). Recent surveys by different teams along multiple locations along

the Lower Congo discovered that the species distribution of cichlids along the Lower Congo is not homogeneous. Both *Nanochromis* and *Steatocranus* species can be confined to short rapids stretches, and partially occur syntopically with close congenerics, whereas other species are less restricted in their distribution and/or represent the single genus representative in a selected rapids stretch. Greenwood (1984) has defined "species flocks" as systems in which multiple species have diversified from a single common ancestor in a geologically restricted area. This definition is assessed by three diagnostic criteria characterising a fish species flock: (1) a geographical circumscription, (2) a high level of endemism and (3) a close phylogenetic relationship (e.g. Salzburger & Meyer 2004). Following this definition, the lower Congo River species assemblages of the genera *Steatocranus* and *Nanochromis* should each qualify as riverine cichlid "species flocks".

Based on extensive AFLP and DNA sequence data and an almost complete taxon sampling, we use a phylogenetic approach to decipher age, origin and pattern of local diversification of these two distantly related lower Congo cichlid genera. We provide for the first time age estimates for the colonization of the lower Congo rapids, which also serve as the minimum age of their geological formation.

Methods

Sampling

Samples from the lower Congo River were collected during low water season (June – August) between 2005 and 2008 in the Democratic Republic of Congo and in 2004 in the Republic of Congo. Our sampling focussed on sequentially arranged regions along the lower Congo River (Fig. 1). In addition, almost all known *Steatocranus* and *Nanochromis* species from surrounding rivers were included in the analyses (Table S1). Based on the phylogenetic analysis of Schwarzer et al. (2009), the haplotilapiine cichlids *Tilapia busumana* (West Africa), *Tilapia cf. bilineata* (Central Congo basin), *Eretmodus cyanostictus* (L. Tanganyika) and *Lamprologus mocquardi* (Central Congo Basin) were chosen as outgroups for the *Steatocranus* dataset, and the Congolian chromidotilapiine genera *Teleogramma* and *Congochromis* (Schliewen & Stiassny 2006) served as outgroups for the *Nanochromis* dataset.

Molecular methods

The mitochondrial gene ND2 was amplified using primers ND2Met and ND2Asn and sequenced using primers ND2Met and ND2Trp (Kocher et al. 1995). ND2 datasets consisted of 1029bp for *Steatocranus* (N = 133) and 980 bp for *Nanochromis* (N = 78). Additionally the mitochondrial 16S, the first nuclear intron of S7 (Chow & Hazama 1998) and the ribosomal genes ENC1, Ptr and Sh3px3 (Li et al. 2008) were sequenced for selected *Nanochromis* and *Steatocranus* species (Table S2) in order to incorporate the data in a larger cichlid phylogenetic framework from Schwarzer et al. (2009a). Amplifications were performed in 10 µl volumes containing 5 µl Multiplex Mix (Qiagen), genomic DNA 1 µl, 0.8 µl of each Primer (2,5 nmol), Q-Solution (Qiagen) and water. Amplifications of all fragments were carried out in 40 cycles according to the temperature profile: 15 min at 95 °C (initial denaturation), 30 s at 95 °C, 30 s at 60 °C, 90 s at 72 °C, and finally 10 min at 72 °C. PCR products were purified with ExoSAP-IT (USB) and diluted with 10 µl - 20 µl HPLC water, depending on product concentration. Sequencing was performed according to standard methods, using Big Dye 3.1. (Applied Biosystems). DNA sequences were read using an ABI 3130xl DNA sequencer (Applied Biosystems). Chromatograms were assembled using SeqMan v. 4.03 included in the Lasergene software package (DNASTAR) and manually proof read. Alignments were conducted using the Clustal W algorithm implemented in BioEdit v. 7.0.4.1. The multilocus dataset (N = 65) of all sequenced markers resulted in a data matrix of in total 4108 bp comprised of S7 (first intron): 528 bp, 16SrRNA: 513 bp, ND2: 993 bp, ENCI: 707 bp, Ptr: 688 bp and Sh3px3: 679 bp. A modified protocol of the original AFLP method (Vos et al. 1995) as suggested in Herder et al. (2006) was used. The following twenty *EcoRI/MseI* primer pairs with three selective bases were used for selective AFLP amplification: ACA*-CAA; AGG*-CTG; ACC*-CTA; ACT*-CAT; ACA*-CTT; AGG*-CAC; AGC*-CAG; ACT*-CTC; ACC*-CAC; AGG*-CTA; AGC*-CTT; ACT*-CAA; ACA*-CAC; AGG*-CAG; ACC*-CTG; ACT*-CTG; AGC*-CAT; AGG*-CTT; ACA*-CAG; ACT*-CAC. Bands were visualized on an AB 3130 sequencer (Applied Biosystems) and Genemapper® v 4.0. software using the size standard ROX 500 XL. Peaks between 100 and 499 bases could be scored unambiguously for presence/absence. The analysis was conducted automatically using Genemapper® v 4.0. Eight to eleven individuals were genotyped two (N = 8) to three times (N = 3) to test for reproducibility. Considering the standard error of automated sequencers, pairs of neighbouring bins whose minimum distance between each other was less than 0.25 bp and also bins containing fragments

differing more than 0.65 bp in size were removed from the dataset. In the *Steatocranus* dataset samples were run in two batches. Therefore bins with fragments that differed by more than 20% relative frequency between the two runs were removed. This step in primary data acquisition decreases rather than increases the likelihood of detection of population structure and was chosen to prune the data set from plate specific effects. The error rate per individual was calculated as the ratio between observed number of differences and the total number of scored fragments (Pompanon et al. 2005). The mean error rate for the *Steatocranus* and *Nanochromis* AFLP datasets was 3% and 2% respectively.

Phylogenetic inference

Alignment of the sequences was conducted using BioEdit (ClustalW), followed by a masking of ambiguous alignment positions using ALISCORE v.0.2 under default settings (Misof & Misof 2009). A ML analysis was conducted for all datasets with RAxML v. 7.0.3 (Stamatakis 2006) using the GTR+Γ model and the rapid bootstrap algorithm with a subsequent search for the best-scoring ML tree. Branch support was evaluated with 1000 non-parametric bootstrap (BS) pseudo-replicates. Model parameters were estimated separately for each possible partition (genes and codon positions separately). For BI, best-fitting models of sequence evolution were estimated using the Bayes Factor Test (Nylander et al. 2004). Bayesian analyses were performed using MrBayes v. 3.1.2 (Huelsenbeck & Ronquist 2001) with 20^6 generations starting with random trees and sampling of trees every 500 generations. To ensure convergence the first 20^5 generations were treated as burn-in and excluded. The remaining trees from all Bayesian analyses were used to build a 50% majority rule consensus tree. For the AFLP data a neighbour-joining tree based on Link et al. (1995a) distances was calculated using TREECON v. 1.3b (Van de Peer & Dewachter 1993). Bootstrap values were calculated based on 1000 pseudoreplicates.

Dating and diversification rates

Divergence times of *Steatocranus* and *Nanochromis* were estimated using a relaxed-clock Bayesian approach implemented in BEAST v. 1.5.3 (Drummond & Rambaut 2007). To set calibration points, *Nanochromis* and *Steatocranus* sequences were integrated in an already published dataset (Schwarzer et al. 2009) based on multiple genes. The ML tree was used as starting tree. The Yule model was selected as tree prior and an uncorrelated lognormal model was used to estimate rate variation along branches. The same priors used in Schwarzer et al. (2009) were applied: an exponential prior (zero offset 5.98 mya) at the root

of all oreochromines (following Carnevale et al. 2003) and a uniform prior (53-84 mya) at the root of all African cichlids except *Heterochromis* (Azuma et al. 2008). For a detailed discussion on the choice of priors see Schwarzer et al. (2009). The analysis was run 30^6 generations and the effective sample size was checked using Tracer v. 1.4. (Rambaut & Drummond 2007).

Results

Molecular phylogenetics

AFLPs provided an almost fully resolved phylogenetic tree for both lower Congo cichlid genera. In the *Steatocranus* dataset one primer combination (ACT*-CTC) was excluded from further analysis as all samples showed off-scale peaks. The final AFLP datasets were composed of 145 (141 *Steatocranus,* 4 outgroup taxa) and 76 (70 *Nanochromis*, 6 outgroup taxa) specimens with 3031 (*Steatocranus*) and 3658 AFLP loci (*Nanochromis*) respectively. Of these 2101 (1706 without outgroups) fragments of the *Steatocranus* and 2182 (1566 without outgroups) of the *Nanochromis* AFLP dataset were polymorphic. The ND2 alignments had 293 (*Nanochromis*) and 285 (*Steatocranus*) variable sites and empirical base frequencies of A = 0.26, C = 0.35, G = 0.11, T = 0.28 and A = 0.26, C = 0.35, G = 0.11, T = 0.28, respectively. The concatenated multigene dataset consisted of two mitochondrial (ND2 and 16S) and four nuclear loci (ENC1, Ptr, Sh3px3 and S7) for 65 selected taxa (Table S2). 347 bp were excluded from further analyses due to alignment ambiguities within the S7 intron (16 positions) and due to saturation in the 3rd codon position of the mitochondrial ND2 locus. The final alignment had 3761 bp. Based on the Bayes factor test, we used a partition separating Exons, 16S and ND2 (first and second codon position) and the S7 intron. The HKY model resulted as best fitting model for all partitions except for nuclear exons (ENC1, Ptr and Sh3px3), which were assigned to GTR + Γ. The dataset had 605 variable sites and empirical base frequencies of A = 0.26, C = 0.26, G = 0.23, T = 0.25.

Intragroup phylogenetic patterns

Phylogenetic analyses of the AFLP and the mtDNA dataset based on ND2 yielded mostly congruent results concerning intrageneric relationships (Fig. 2). The lower Congo species were polyphyletic with respect to central and upper Congo taxa, since in both cases two distantly related lower Congo lineages were present.

In *Nanochromis* one phylogenetic group is composed of the central and downstream lower Congo endemics *N. consortus* and *N. splendens* (BS in nc- and mt-phylogenies 100 and 95 respectively) and a second of *N. parilus* (from the upper and central lower Congo and Maluku)*, N. nudiceps* (from a river close to Lake Mai Ndombe) and *N. teugelsi* (from the lower Kasai, BS = 97 and = 77, Fig. 2). *Nanochromis minor* from the central lower Congo sampling location Kinganga and *N. transvestitus* from Lake Mai Ndombe appears as sistergroup to all other *Nanochromis* in the AFLP dataset (BS = 100) but as sistergroup to the *N. splendens*/*N. consortus* group in the mtDNA dataset (including *N. wickleri*, BS = 69). The phylogenetic positions of *N.* sp "Ndongo" from the Sangha drainage and *N. teugelsi* from the Kasai also remain unresolved. *Nanochromis* sp "Ndongo" either clusters with *N. parilus/N. nudiceps/N. teugelsi* in the mtDNA dataset (BS = 77) or appears as sistergroup to all *Nanochromis* except for *N. minor* and the central Congo species *N. wickleri* and *N. transvestitus* in the AFLP tree (BS = 91). *Nanochromis teugelsi* from the Lower Kasai, a southern tributary to the Congo, is part of the *N. parilus* clade in the mtDNA dataset (BS = 77) but appears as weakly supported (BS = 52) sistergroup to either *N. parilus* or *N. nudiceps* (BS = 46) in the AFLP dataset.

Steatocranus splits into three major monophyletic groups. One is composed of the lower Congo endemics *S. casuarius* and *S.* sp. aff. *casuarius* "brown pearl", *Steatocranus* sp. "Maluku" from upstream of Pool Malebo*, S*. sp. "dwarf" and species from the northern tributaries Ubangi and Ngoko and from the upper Congo near Kisangani (*S*. sp. "Kisangani", *S. ubanguiensis*, *S*. sp. "Nki" and based on the mt- tree also *S*. sp. "Lefini", Fig. 2). *Steatocranus* sp. "Maluku" appears as sistergroup to the *S.* cf. *casuarius* clade (BS = 64) in the nc- but to *S*. sp. "dwarf" in the mt- dataset (BS = 96). The second major group is composed of the lower Congo endemics *S.* cf. *gibbiceps, S. mpozoensis*, *S. glaber* and four distinct *S.* cf. *tinanti* clades (BS = 100 in the nc and BS = 91 in the mt-dataset, Fig. 2). The relationship of *S. glaber* is ambiguous as it is sister to *S. mpozoensis* in the mt- dataset (BS = 95) but to *S. gibbiceps* in the AFLP dataset (BS = 90). *Steatocranus* species occurring outside of the Lower Congo roughly form two phylogenetic clusters according to their major geographic distribution either north ("North" clade, Congo mainstream/Ubanghi/Sangha/Lefini) or south ("South" clade, Kasai/Kwango and Pool Malebo) to the Congo mainstream.

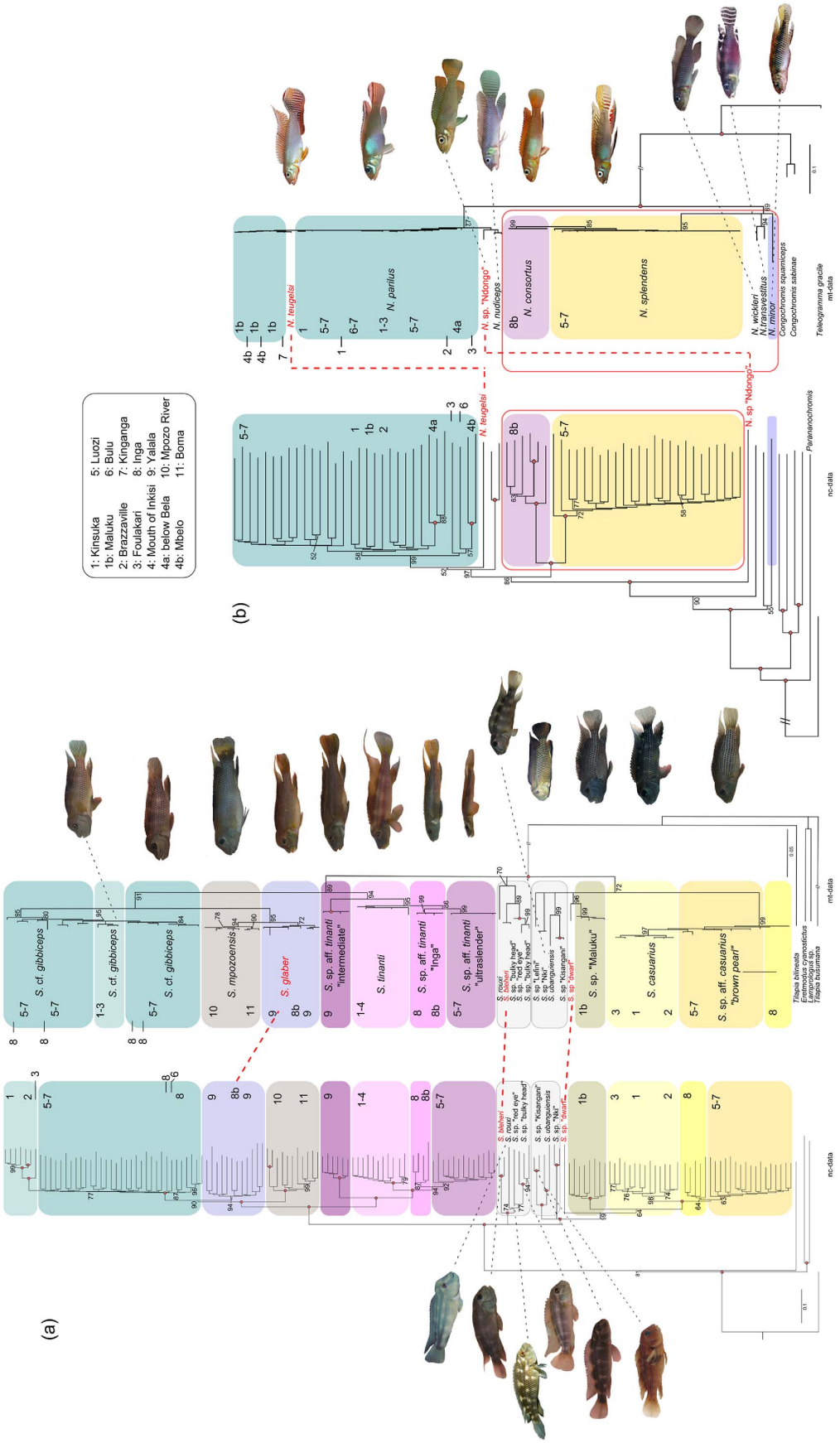


Figure 2 *(caption overleaf)*

Figure 2 *Phylogenetic trees based on nc - and mt - datasets*

The datasets comprises mitochondrial sequences of ND2 and over 2000 AFLP markers for (a) *Steatocranus* and (b) *Nanochromis*. Black numbers at nodes refer to bootstrap-values (BS, 1000 pseudo-replicates) of the ML run (right side) or the neighbour joining tree (left side). Filled circles represent a 100% BS support. Major groups within the phylogeny are marked with coloured frames. Species that are placed differently in AFLP and mt-trees are marked in red. The red frames in *Nanochromis* trees indicate differences in sistergroup-ralationships.

The third major group, composed of species from the southern Congo tributaries and from the Congo mainstream (*S. rouxi*, *S.* sp. "red eye", *S.* sp. "bulky head", and *S. bleheri*), appears polyphyletic in all datasets. In 1000 bootstrap replicates they appear in almost equal measure as sistergroup to all remaining *Steatocranus* (BS = 47, Fig. S1), or alternatively, to the lower Congo *S.* cf. *gibbiceps, S. glaber, S. mpozoensis* and *S.* cf. *tinanti* (BS = 44, Fig. S1). Within this monophyletic group, *S. rouxi* from the upper Kasai River appears as sistergroup to the remaining taxa from southern tributaries in the mt- but to *S*. sp. "red eye" and *S*. sp. "bulky head" in the nc- phylogeny (Fig. 2).

Dating and diversification rates

The age of the most recent common ancestor (MRCA) of *Nanochromis* is estimated at 8.36 (6.5-10.4) mya (Fig. 3, Node N). Median ages for clades containing lower Congo taxa (Node LC 1: *N. splendens* and *N. consortus,* Node LC 2*: *N. parilus*, *N. teugelsi* and *N. nudiceps*) are estimated at 2.67 (1.5-3.9) mya and 1.6 (0.7-2.5) mya. The two lacustrine species from the central Congo, *Nanochromis wickleri* and *N. transvestitus* (Node C1), diverged 4.4 (3-6.2) mya based on our data (Fig. 3, Table 1). The age of the MRCA of the genus *Steatocranus* is estimated at 7.7 (6.1-9.6) mya (Fig. 3, Node S). For the lower Congo *Steatocranus* clades age estimates were as follows: node LC 3: *S. casuarius* species complex 0.94 (0.3-1.7) mya, node LC 4: *S.* cf. *gibbiceps*, *S. glaber*, *S. mpozoensis* and *S.* cf. *tinanti* 4.48 (3.3-5.8) mya, node LC 4a: *S.* cf. *gibbiceps*, *S. glaber* and *S. mpozoensis* 3.43 (2.4-4.6) mya, node LC 4b: *S. gibbiceps* species complex 1.42 (0.7-2.3) mya, node LC 4c: *S. tinanti* species complex 3.03 (4.5-7.6) mya (Fig. 3). The age for MRCA of the clade containing *S*. sp "dwarf", *S*. sp. "Maluku", *S. casuarius*, *S*. sp. aff. *casuarius* "brown pearl" and species from the northern tributaries (Node C2) was estimated at 5.3 (4.1-6.7) mya. The MRCA of the species from the northern tributaries alone was estimated at 4.8 (3.2-5.5) mya (Fig. 3, Node C3). The estimation of dates for the origin of

species from surrounding lakes and rivers is problematic as ambiguous signal masked the phylogenetic relationships at the base of the lower Congo clades (see section above).

Discussion

Potential colonization of the lower Congo rapids

Phylogenetic inferences based on a fully representative taxon sampling revealed that the ancestors of lower Congo species of both studied genera reached the rapids in at least two allochronic events and then further differentiated within the lower Congo. Though the age estimate for the MRCA of *Steatocranus,* at 7.7 (6.1-9.6) mya, is slightly younger than in the previous study by Schwarzer et al. (2009, 7.4-14.1 mya) the 95% confidence intervals still highly overlap and are on average smaller (Table 1). The inclusion of more terminal taxa combined with a different gene composition used for the analysis might have caused the differences. Age estimates, however, should in the present context not be seen as fixed numbers but as rough chronological placements of evolutionary processes.

Apparent cyto-nuclear tree incongruence hampers the reconstruction of some colonization patterns in both genera (Fig. 2). In *Steatocranus* a well supported relationship is identified between the "North" clade species and the lower Congo *S.* cf. *casuarius* clade (including *S*. sp. "Maluku" and *S*. sp. "dwarf"), indicating that the ancestors of present day *S.* cf. *casuarius* colonized the lower Congo (at about 3.23 mya) from Northern populations distributed presently in the upper Congo near Kisangani, downstream of Mbandaka and in northern Congo tributaries (Ubangi and Sangha Rivers). The phylogenetic signal concerning the position of the Southern populations is ambiguous. Two alternative topologies are most frequently found among the bootstrap replicates (Fig. S1), supporting two slightly different colonization scenarios with regard to the lower Congo: In the first scenario, a simultaneous dispersal of southern precursors into the lower Congo as well as into the northern tributaries took place, i.e. seeding both the present day *S*. cf. *tinanti*/*S. mpozoensis*/*S. glaber*/*S*. cf. *gibbiceps* clade (ca. 4.48 mya, Fig. 3) and the northern clade (BS = 47, Fig. S1). Subsequently a secondary colonization wave from northern tributary species and from the Congo mainstream then founded the younger lower Congo *Steatocranus* cf. *casuarius* clade (ca. 3.23 mya, Fig. 3). In the second scenario, an early vicariance event separated already existing southern and northern tributary populations which then founded the *S*. cf. *tinanti*/*S. mpozoensis*/*S. glaber*/*S*. cf. *gibbiceps* clade (BS = 44, Fig. S1) from the South and later the *Steatocranus* cf. *casuarius* clade from the North (Fig. 4).

In *Nanochromis* the nc-dataset indicates an initial colonization from the lineage comprising central lacustrine species *N. transvestitus, N. wickleri* and the lower Congo species *N. minor*. This scenario resembles that of *Steatocranus*, with older southern populations seeding the lower Congo (mt-signal, Fig. 2) and (potentially) the Northern tributaries (nc - signal, Fig. 2) followed by a second colonization wave from central Congo (ca. 1.6 mya) and potentially from the North (*N*. sp. “Ndongo”, mt - signal, Fig. 2).

Both younger phylogenetic clades (*S*. cf. *casuarius* and *N. parilus*) exhibit a distribution limited to the upstream and central lower Congo. This parallel pattern implying two colonization waves into the lower Congo in both genera (Fig. 4) might indicate that geological factors have shaped the evolution of these distantly related genera in parallel. One palaeogeological event that may have played a part was the gradual coastal uplift of the West African continental margin (“Atlantic Rise”) since the early Miocene (Lavier et al. 2001) which potentially went along with changes in drainage pattern of the greater lower Congo region (Lucazeau et al. 2003).

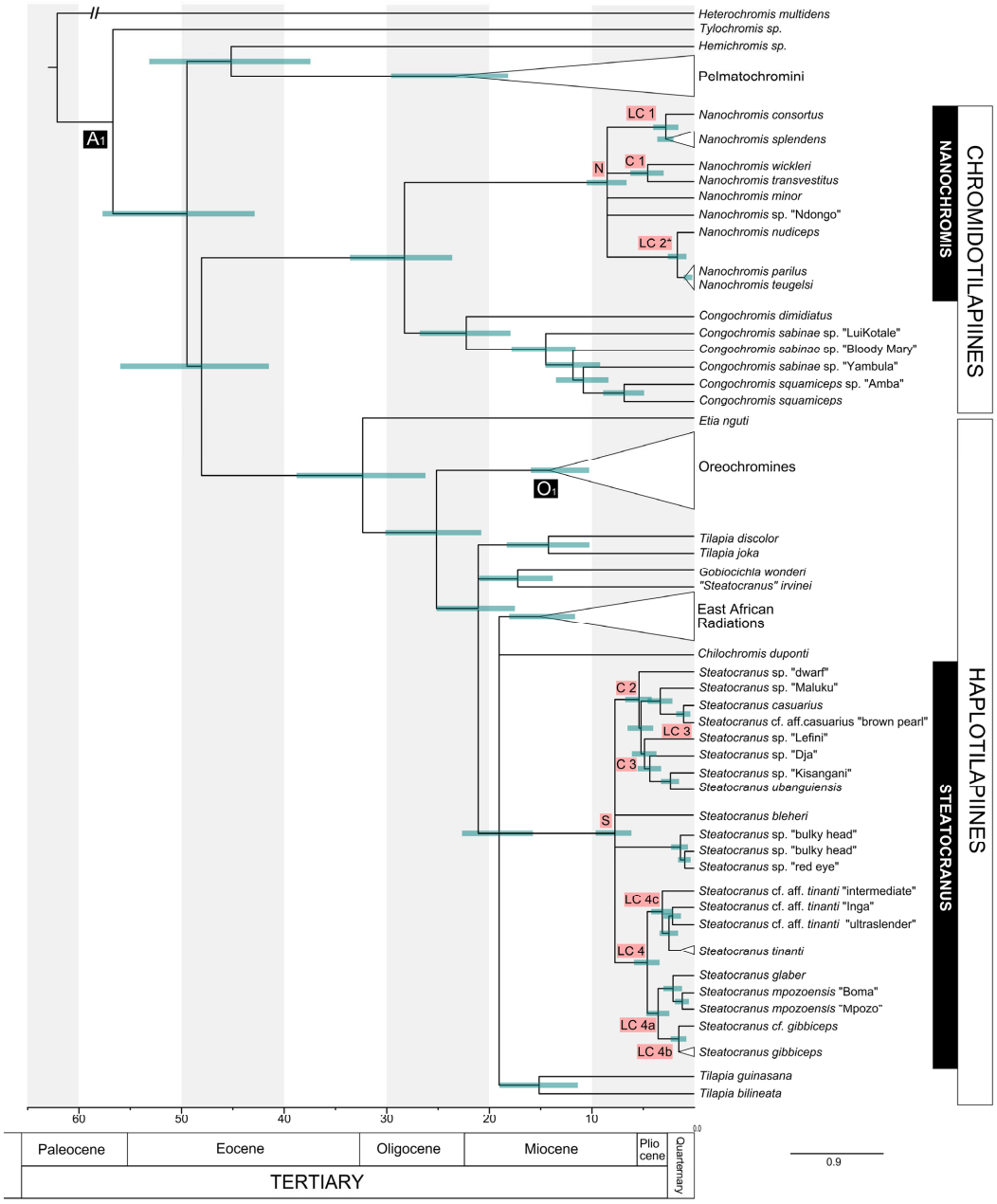


Figure 3 *Chronogram showing divergence time estimates for Steatocranus and Nanochromis*

The chronogram was calculated based on the ML tree. A partitioned Bayesian analysis implemented in BEAST was used to estimate divergence times. Time constraints were used following Schwarzer et al. (2009): A_1 53–84 mya (uniform prior), published age estimate based on non-cichlid fossils (Azuma et al. 2008) and O_1 5.98 mya (lower bound), the age estimate for *Oreochromis lorenzoi*† (Carnevale et al. 2003). The chronogram shows 95% credibility intervals (HPC, green bars). For nodes marked with letters, age estimates (95% HPC and mean heights) are given in Table 1. The asterisk marks the non-endemic lower Congo clade including *N. parilus* and *N. teugelsi*. *Nanochromis parilus* is distributed in the lower Congo but can also be found at Maluku upstream of Malebo pool. For simplification clear monophyletic groups were combined and shown as triangles.

The geological age of the lower Congo River drainage

Analyses of fluvial sediments in the Gulf of Guinea show that major offshore deposits were not present in the region of the present day Congo mouth until relatively recently, but that major parts of central Africa´s interior drainage discharged in the Cretaceous and Oligocene (65 to 36 mya) into the Atlantic ocean through the Ogooué valley in Gabon and the Cuanza system in Angola (Leturmy et al. 2003; Lucazeau et al. 2003). This indicates that the modern lower Congo rapids cannot be older than 35 mya, and additional analyses of Congo offshore deposits suggest an origin of the present day Congo discharge at the Miocene-Pliocene transition at approximately 5 mya (Ferry et al. 2004), after the southern African continent had been affected by a significant uplift inducing a progressive rearrangement of the watersheds (Lavier et al. 2001). Our age estimates of cichlid radiations of the lower Congo are fully concordant with this timing, as two phylogenetically independent cichlid lineages radiated within the last 5 mya. Median age estimates for the lower Congo cichlid endemics included here range from 1.6-2.67 mya for *Nanochromis* and from 0.94-4.48 mya for *Steatocranus* (Fig. 3). These results corroborate the view that the lower Congo cichlid radiation was closely linked to the establishment of new habitat availability and thus represents an autochthonous radiation within the lower Congo River.

Biogeographic differentiation along the Lower Congo rapids

Our phylogenetic data correspond to a subdivision of the lower Congo into three biogeographic areas (Fig. 1, see also Robert 1946): (A) upstream (from Malebo Pool to Mbelo, 133 km) (B) central (from Luozi to Kinganga, 129 km) and (C) downstream (from Inga to Boma, 88 km). Each of these river sections is characterized by unique geomorphological settings (Robert 1946) as well as different species assemblages with various degrees of local endemism not restricted to cichlids (Roberts & Stewart 1976; Vreven & Stiassny 2009 and pers. obs.). The upstream lower Congo (“northern rapids”, Robert 1946) starts with the steep Livingston falls, separating the lower Congo from the Cuvette Centrale, and stretches around 133 km downstream intersected by several smaller rapids. The transition from upstream to the central lower Congo is characterized by a change in sediments from Proterozoic to Precambrian quarzites and schists (Runge 2008) and the presence of rapids at Mbelo and Bela (Fig. 1). The central lower Congo (ca. 129 km long) is a large navigable tract characterized by a wider lake-like river channel that is occasionally narrow and very deep (up to 200 m around Bulu, Jackson et al. 2009).

Table 1 *Age estimates*

Priors A_1 and O_1 were taken from Schwarzer et al. (2009) and resulting age estimates were compared with publishes studies (Genner et al. 2007; Schwarzer et al. 2009) when possible. LC2* contains *Nanochromis parilus* and *N. teugelsi*, which are not endemic to the lower Congo. Two or three asterisks (** or ***) mark nodes whose node ages were estimated either without *S. mpozoensis* (**) or *S*. sp. "Maluku" (***) and are thus not one to one equivalent to nodes LC 4, LC 4a and C2 in this study. Age estimates given for node A_1 from Genner et al. (2007) correspond to their dataset calculated with Gondwana priors.

Node	Date estimates in Myr					
	Baysian Inference (95% credibility intervals)					
	This study		Schwarzer et al. (2009)		Genner et al. (2007)	
A_1	**55.5**	(42.8, 57.6)	**56.7**	(53.0, 64.2)	63.7 (N)	(46.6, 79.6)
O_1	**12.8**	(10.2, 15.9)	**12.8**	(8.9, 16.8)		
N	**8.36**	(6.5, 10.4)				
S	**7.7**	(6.1, 9.6)	**10.7**	(7.4, 14.1)		
LC1	**2.7**	(1.5, 3.9)				
LC2*	**1.6**	(0.7, 2.5)				
LC3	**0.9**	(0.3, 1.7)				
LC4	**4.5**	(3.3, 5.8)	**5.7****	(3.6, 8.4)		
LC4a	**3.4**	(2.4, 4.6)	**3.1****	(1.4, 5.1)		
LC4b	**1.4**	(0.7, 2.3)				
LC4c	**3.0**	(4.5, 7.6)				
C1	**4.4**	(3.0, 6.2)				
C2	**5.3**	(4.1, 6.7)	**6.9*****	(4.1, 10.2)		
C3	**4.8**	(3.2, 5.5)				

The downstream lower Congo (ca.88 km long, "southern rapids", Robert 1946) is the steepest river section and mainly characterized by the presence of the huge rapids at Inga and Yalala. A spatial genetic differentiation is apparent in most *Steatocranus* and *Nanochromis* clades whose distribution exceeds one of these river stretches (Fig. 2). The most obvious mechanism shaping the lower Congo species diversity is the complexity of the river itself, with alternating stretches of rapids and deep river habitats (Jackson et al. 2009). According to our data, the rapids at Inga and Yalala have provided the strongest barriers to dispersal, as supported for example by the elevated degrees of local endemism and ancient splits leading to differences in species composition in both genera. Below Nziya (close to Inga), no *Nanochromis* species occur, and apart from *Steatocranus glaber* only the locally

endemic *S*. cf. aff. *tinanti* “intermediate” is present. Downstream of the Yalala rapids only a single *Steatocranus* species (*S. mpozoensis*) is found, even though Matadi (downstream of Yalala) was given as the type locality for *S. gibbiceps* (Boulenger 1899). However, its occurrence could not be verified despite substantial efforts and the type location information likely refers to the port of shipment (“Matadi”) rather than the true collection site. Fine-scale differentiation in two other distantly related lower Congo cichlids (Markert et al. 2010) matches the above described spatial pattern for the central/downstream lower Congo area. *Lamprologus tigripictilis* forms two well separated populations above and below the Inga rapids and the highly rheophilic cichlid genus *Teleogramma* exhibits a pronounced population structure shaped by smaller rapids (Isangila and Fwamalo) upstream of Inga but is absent below the rapids. The *S*. cf. *tinanti*, *S*. cf. *gibbiceps*/*S. glaber*/*S. mpozoensis* and *N. splendens*/*N. consortus* species groups diverged roughly 3 mya (3.03, 3.43 and 2.67 mya respectively, Fig. 3) and both lineages evolved locally endemic species in and below the rapids of Inga (e.g. *S. glaber*, *S. mpozoensis, S*. cf. aff. *tinanti* “Inga”, and *N. consortus*). Within the *S*. cf. *tinanti* species complex a phylogenetic split (node age 3.03 mya) separates the Yalala rapids endemic in the downstream part of the lower Congo (*S*. cf. aff. *tinanti* “intermediate”) from the remaining three clades (Fig. 2). At first glance, this basal phylogenetic split between a single downstream and the remaining upstream species appears implausible. In a strongly flowing (continuous) river stretch like the lower Congo a gradual progression and differentiation of populations in flow direction appears more likely. A potential scenario (“large Inga waterfall hypothesis”) explaining this and the high degree of endemism below the Inga rapids is, that following the first colonization wave early colonists of *Steatocranus* and *Nanochromis* have remained strongly isolated in downstream regions since about 3 mya (e.g. by a waterfall at Inga). Riverbed erosion may then have worn down this waterfall resulting in the present day Inga rapids (rather than falls), which are penetrable to upstream movement by fishes.

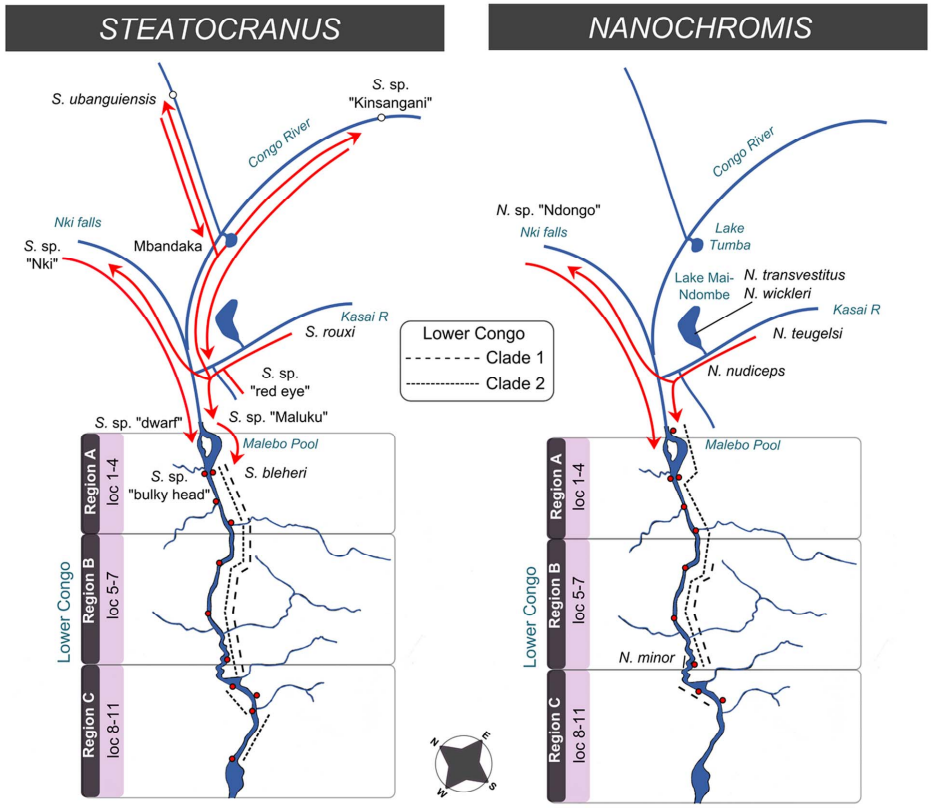


Figure 4 *Potential colonization scenarios for Steatocranus and Nanochromis*

Potential colonization scenarios are shown separately for both genera. The red arrows indicate potential dispersal routes of *Steatocranus* and *Nanochromis* precursors. Non-lower Congo taxa are written in black. Current distribution ranges of the major lower Congo phylogenetic clades (see Fig. 2) are represented by different kinds of dotted lines (see figure legend). Clade 1 (*Steatocranus*): *S.* cf. *casuarius* species pair, clade 2 (*Steatocranus*): *S.* cf. *tinanti/ S.* cf. *gibbiceps/ S. glaber* and *S. mpozoensis*, clade1 (*Nanochromis*): *N. splendens* and *N. consortus*, and clade 2 (*Nanochromis*): *N. parilus.* The trisection of the lower Congo is shown in combination with sampling sites (indicated by red dots) along the lower Congo. The upper part of the Congo River is presented highly simplified for a better understanding.

The formation of the present day lower Congo offered not only new habitat opportunities but also structurally new habitat types, e.g. by the combination of extreme currents and turbidity (Roberts & Stewart 1976). Differing spatial structures depending on the species-(group) or genus (Fig. 2) indicate that apart from a prominent role played by extrinsic habitat features of the lower Congo rapids, intrinsic factors shaped the species divergence. Often a synergetic composition of both extrinsic (e.g. habitat composition, physical barriers to gene

flow) and intrinsic factors (e.g. dispersal capabilities or ecological adaptations) is responsible for species differentiation (Coyne & Orr 2004). Analogous to the famous “Mbuna” from Lake Malawi (Ribbink et al. 1983; Konings 2007), the rock-dwelling and strictly rheophilic *Steatocranus* (Roberts & Stewart 1976) apparently exhibit higher site fidelity and limited dispersal capabilities compared to the sand-dwelling and less rheophilic *Nanochromis*. Further, intrageneric differences within *Steatocranus* point to alternative ecological adaptations (e.g. differences in dentition and body shape, Roberts & Stewart 1976). Most obvious differences are present between the low-bodied *S*. cf. *tinanti* and *S. mpozoensis* and the higher bodied *S*. cf. *gibbiceps* or *S*. cf. *casuarius* (Roberts & Stewart 1976), indicating differential adaptations to the life in strong current, e.g. on top (*S.* cf. *tinanti, S. mpozoensis*) or among stones (*S*. cf. *gibbiceps*, *S. casuarius*, pers. obs.). However, quantitative ecological data and a denser sampling are needed to differentially analyze causes and factors responsible for the riverine lower Congo species richness. Ecomorphological differentiations in dentition and body shape described for the rock dwelling Eretmodini (genera *Tanganicodus*, *Eretmodus* and *Spathodus*) from Lake Tanganyika correspond to those found in *Steatocranus* (Roberts & Stewart 1976; Ruber & Adams 2001). This might indicate either a synapomorphic developmental basis for character evolution at the base of austrotilapiines (Schwarzer et al. 2009) or a rapid genetically independent origin of trophic adaptations. This highlights the importance of including possible riverine founder species, when analysing the origins of the enormous adaptability of the megadiverse East African cichlid radiations.

The cichlid species flocks of the lower Congo rapids

The use of the term “species flock” is controversial and has been much debated (Greenwood 1984). Fish species flocks were typically discussed with respect to speciose groups in closed lacustrine environments, of which the cichlid radiations of East African rift valley lakes and Cameroonian crater lakes are the most famous examples (Kocher 2004; Salzburger & Meyer 2004; Schliewen 2005). In contrast, riverine species flocks were rarely studied and to date only a few examples exist. Sullivan et al. (2002; 2004), Feulner et al. (2007; 2008) and Kullander et al. (2010) gave examples for riverine species flocks of weakly electric fish (Mormyridae) in the Ogooué River system in Gabon and the lower Congo River and South American cichlids restricted to the upper Rio Uruguay system, respectively. The distributions of these species flocks were either very broad and encompassed several river systems (Sullivan et al. 2002; 2004) or intrageneric sampling was not complete with regard to known

taxa and sampled areas (Feulner et al. 2007; 2008). To our knowledge, the South American pike cichlids of the *Crenicichla missioneira* species group currently represent the best candidate for a riverine species flock, even though the phylogenetic study (Kullander et al. 2010) was based on only a single mitochondrial marker and the taxon sampling did not contain all important members. Here, we present the first evidence for a riverine species flock and several species pairs endemic to an exceptionally complex habitat.

The *S*. cf. *tinanti*/ *S*. cf. *gibbiceps*/ *S. glaber* and *S. mpozoensis* species group is endemic to the lower Congo rapids, a defined geographic area, and forms closely related assemblages, thereby meeting the three species flock criteria. *Nanochromis splendens* and *N. consortus* and *S. casuarius* and the yet undescribed species *S*. cf. aff. *casuarius* "brown pearl" each form species pairs endemic to the lower Congo rapids. Both genera colonized the lower Congo rapids in at least two allochronic events each forming independent closely related riverine cichlid species assemblages (Fig. 2) within a timeframe of 5 mya.

Conclusion

The rapids of the lower Congo River, inhabited by a remarkable diversity of cichlids and other fishes, provide an outstanding example for the underestimated diversity of riverine and especially rapids systems. Like lakes and islands, rapids provide novel ecological opportunities for riverine organisms and often form after catastrophic geomorphological events. Due to steep selectional gradients between average riverine conditions and rapids habitats, invading species facing multiple unutilized ecological opportunities may rapidly adapt to these extreme conditions and form locally endemic species assemblages. Our data suggest that multiple (minimum two) allochronic colonization events seeded the present day diversity of the lower Congo *Steatocranus* and *Nanochromis* species, which subsequently evolved into small species flocks. Cichlid species diversity in the lower Congo is arranged sequentially and differentiated ecologically. This study provides a phylogenetic primer for the study of this complex system, that may serve as a link between riverine diversity and the megadiverse cichlid radiations of the East African lakes, which - based on the inherent logic of lakes being interconnected only by rivers - surely were seeded by riverine cichlids (Kocher 2004). In particular the close phylogenetic relationship of the *Steatocranus* species to the East African cichlid radiations (Schwarzer et al. 2009) renders this system appealing.

Authors contributions

JS and UKS conceived and designed the experiments. JS performed the experiments. JS, BM and UKS analyzed the data. SNI provided crucial material. All authors contributed to the preparation of the manuscript. They read and approved the final version.

Acknowledgments

Our thank goes to J. Friel, L. Rüber, A. Lamboj, E. Vreven for providing tissue samples. We gratefully thank R. Schelly and M. Stiassny for company in the field, tissue samples and valuable information on selected specimens. For indispensable assistance in the field we thank D. Neumann, J. Frommen, P. Alibert, A. Dunz, M. Levy, V. Mamonekene, A. Ibal-Zamba, J. Punga, U. Ali-Patho, P. Mongindo, C. Danadu, F. Bapeamoni and T. Kadangé Ngongo. Special thanks go to K. Langen, F. Eppler and B. Müller who helped with lab routines. J. Runge helped with questions concerning Palaeogeography. L. Wescott and R. Schelly proofread the manuscript. L. Rüber and W. Salzburger gave valuable comments on the manuscript.

CHAPTER 6

Speciation within genomic networks: A case study based on Steatocranus cichlids from the Congo basin

Julia Schwarzer [1,2], Bernhard Misof [1] & Ulrich K. Schliewen [2]

(1) Zoologisches Forschungsmuseum Alexander Koenig, Adenauerallee 160, 53113 Bonn, Germany

(2) Bavarian State Collection of Zoology, Münchhausenstr. 21, 81247 München, Germany

This is the author's version of a work submitted to: Journal of Evolutionary Biology.

Abstract

Phylogenetic methods largely rely on the reconstruction of bifurcating trees. Conclusions drawn from these data, however, can be affected by reticulate effects potentially remaining undetected, due to the usage of single markers or the reliance on well resolved trees. As hybridization is a much more common phenomenon as previously thought it is advisable, not only to use multi-locus datasets, but to calculate networks and, best, conduct a follow-up explorative analyses to evaluate the full information content in a dataset. The cichlid genus *Steatocranus*, a close relative to members of the East African cichlid radiations, forms a riverine radiation in the lower Congo rapids. There are indications from previous phylogenetic analyses that hybridization occurred in this genus. In the present study we analyse an already published AFLP dataset with more than 2000 loci with removal experiments to detect gene flow and potential formation of genomic networks. A high degree of gene-flow connecting adjacent populations but also distantly related species is evident, indicating that the evolution of these species is best represented by a network rather than a tree. We give to our knowledge, here, the first example of a reticulate network in vertebrates.

Introduction

The “Tree of Life” metaphor seduced us to believe that the evolution of organismic diversity is exclusively a stepwise process that has generated novel species through a strictly bifurcating speciation process. Consequently, hybridization has long been viewed as a phylogenetic accident that blurs species differences through gene flow and hampers speciation rather than facilitating it. However, there is growing evidence for inter-specific gene flow indicating that introgression and hybridization contribute substantially to speciation in plants and animals (Rieseberg & Wendel 1993; Rieseberg et al. 2003; Nolte et al. 2005; Mallet 2007; Larsen et al. 2010). Especially if reduced contact and competition with parental lineages coincides with the occurrence of new ecological opportunities, for example after colonization of novel habitats, hybrid fitness can be high and gene flow can act even as a promoter for speciation (Danley et al. 2000; Willis et al. 2006; Jiggins et al. 2008; Joyce et al. 2011; Brelsford et al. 2011). In this sense, Seehausen (2004) based on Templeton (1981) proposed that the rapid origin of species of mega-diverse radiations, e.g. cichlids in the East African Great Lakes, is based on large scale Hybridization events among originally allopatric lineages forming secondarily hybrid swarms in newly colonized areas. After primary formation of lineages within an adaptive radiation, additional Hybridization events between primary lineages could then further trigger the evolution of functional novelty and speciation in adaptive radiations ("syngameon hypothesis", Seehausen 2004). Multiple examples for hybridization are known from African cichlid fish (Salzburger et al. 2002; Seehausen 2004; Schliewen & Klee 2004) and the impact of introgression and hybridization of riverine species on speciation of lacustrine radiations came recently into focus (Joyce et al. 2011). In this context we present a study on the riverine *Steatocranus* species, endemic to Congo basin, which are closely related to the mega-divers lacustrine African cichlid radiations (Schwarzer et al. 2009). Our analyses of molecular divergences corroborate the idea that multiple introgression and hybridization produce a genomic network which potentially promoted divergence and speciation.

The reconstruction of phylogenetic hypotheses are essential to understand the historical origin of diversity (Linder & Rieseberg 2004), but it became increasingly obvious that single genes are potentially misleading, as even good species may continue to exchange genetic material through hybridization (Chan & Levin 2005). One approach to detect reticulate signal in multigene data sets is the comparison of phylogenies based on nuclear genes and genes of

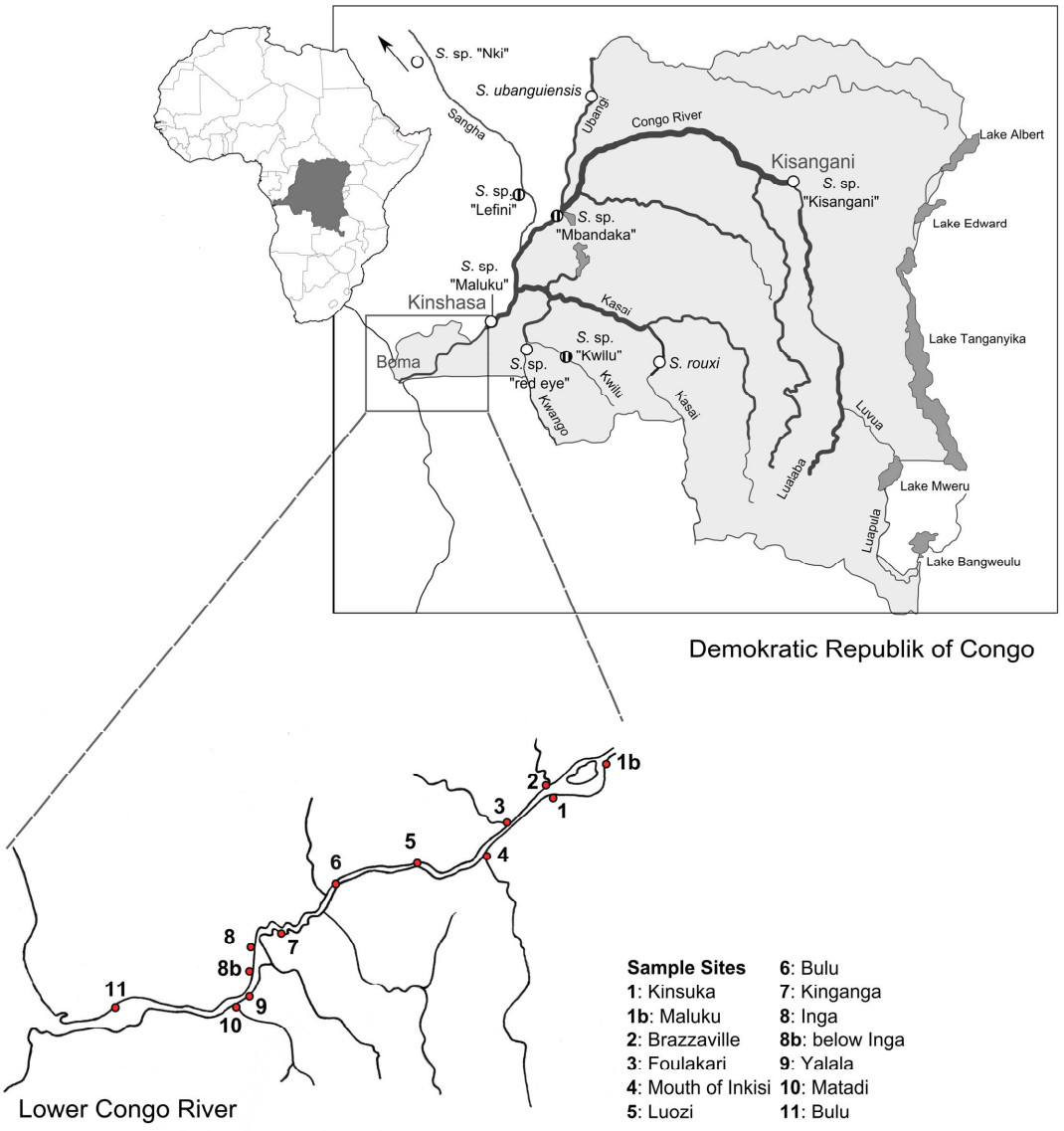


Figure 1 *Location map of sampling sites*

Overview of *Steatcranus* sampling sites in the Congo basin and, in more detail, the lower Congo River. White circles correspond to taxa from Northern and Southern tributaries that were included in the analyses and striped circles to those that were not included. The smaller circles on the more detailed scheme of the lower Congo River correspond to sampling sites from expeditions from 2001 to 2008.

maternally inherited organelles (Sullivan et al. 2004; Schliewen & Klee 2004; Herder et al. 2006). Conflict signal can be visualized by the application of phylogenetic networks (Huson & Bryant 2006). However, weak hybrid signal, e.g. through ancient or rare hybridization events, although having profound effects on hybrid speciation (Jiggins et al. 2008; Salazar et al. 2010), might be masked by the dominance of the main phylogenetic signal. Beside the

creation of multi-locus datasets, a careful evaluation of the underlying signal should, therefore, be conducted to capture as much information content as possible.

The riverine cichlid genus S*teatocranus* underwent a radiation in the lower Congo rapids resulting in a riverine species flock (Schwarzer et al. 2011). Schwarzer et al. (2011) further showed that this radiation is about 5my old, concordant with the proposed origin of the lower Congo rapids (Lavier et al. 2001). Phylogenetic analyses based on AFLP and sequence data suggested the presence of many partly parapatric species within the species flock delimited with reasonable bootstrap support. Cyto-nuclear discordance, however, indicated gene flow among species and suggested a role for Hybridization in shaping the extant *Steatocranus* species diversity.

Here, based on previously published AFLP and mt-datasets of *Steatocranus* (Schwarzer et al. 2011) four lines of evidence were applied to infer the amount and origin of reticulate signal in the data set: (1) cyto-nuclear discordance, (2) conflict signal indicated by NeighborNet analyses (3) homoplasy excess tests, and (4) principal component analyses.

Methods

Study system

The present study is based on *Steatocranus*, a rheophilic cichlid genus endemic to the Congo basin. A species flock with around ten species (described and undescribed) inhabits the lower Congo River, showing different distribution ranges and degrees of genetic differentiation (Schwarzer et al. 2011). For comprehensibility we summarize the species under following generic terms: *S*. cf. *tinanti* (*S. tinanti*, *S*. sp. aff. *tinanti "*ultraslender", S. sp. aff. *tinanti* "Inga" and S. sp. aff. *tinanti* "intermediate"), *S*. cf. *casuarius* (*S. casuarius* and *S*. sp. aff. *casuarius* "brown pearl"), *S*. cf. *gibbiceps* (*S*. cf. *gibbiceps* from upstream lower Congo and *S*. cf. *gibbiceps* from central lower Congo), *S. glaber* and *S. mpozoensis*. The lower Congo can roughly be divided in three parts: upstream, central and downstream based on differing species assemblages (Schwarzer et al. 2011). *Steatocranus* occurring upstream of the lower Congo are distributed either in northern tributaries (*S*. sp. "Nki", *S. ubanguiensis* and *S*. sp. "Lefini" from Ngoko, Ubangi and Lefini Rivers) or south to the Congo mainstream (*S. rouxi*, *S*. sp. "red eye", S. sp. "Kwilu" from Kasai, Kwango and Kwilu Rivers), or in the Congo proper (*S*. sp. "dwarf", *S*. sp. "bulky head", *S. bleheri*, *S*. sp. "Maluku", *S*. sp. "Mbandaka" and *S*. sp. "Kisangani" from around Malebo Pool, Maluku, Mbandaka and Kisangani, Fig. 1).

NeighborNet and Principal Components analyses

All analyses are based on an AFLP dataset composed of 3031 loci for 141 *Steatocranus* taxa (without outgroups) taken from Schwarzer et al. (2011). Out of these 1706 fragments were polymorphic. To extract and simplify the information content of the AFLP data a Principal Component Analysis (PCA) was conducted using PAST v. 2.04 (Hammer et al. 2001). The PCA is especially useful to depict structure in genetic datasets, as the variation can be summarized into few synthetic variables. Another advantage is that the analysis is exploratory and no underlying genetic model is assumed (Jombart et al. 2009). Results were visualized in a distance biplot with no scaling of the data (Legendre & Legendre 1998). Based on the information content, we applied an arbitrary threshold of 2% for explanatory power of PC axes, resulting in six informative axes.

A NeighborNet analysis (Bryant & Moulton 2004) based on a Link et al. distance matrix (Link et al. 1995) was inferred using SplitsTree4 v. 4.10 (Huson & Bryant 2006). The algorithm takes only shared and unique bands into account while absent band are ignored. Since the absence of a band cannot be unambiguously equated with a loss of a particular restriction site this approach is conservative compared to alternative distance measures (like e.g. Nei & Li 1979). In a NeighborNet, hybrid taxa are expected to be found at the intersection of the two parental splits, as they share genetic similarities with each of the two parents (Dixon et al. 2009). Conflicting signal is represented as “boxes” at the base of affected nodes. The LS fit value reflects how well the network represents the underlying data.

Inferring hybrid signal

Following Seehausen (2004), we applied a tree based method to test for homoplasy excess (HET) in our dataset. The expectation is that the inclusion of hybrid taxa increases the conflict in the dataset and reduces support values for affected nodes in a phylogenetic tree more than the inclusion of non-hybrid taxa (Seehausen 2004). This detection of potential hybrid signal focuses on the AFLP dataset, as hybridization can not be detected in maternal inherited mitochondrial markers (Chan & Levin 2005). Clades showing discordant signal between the nuclear (nc) and mitochondrial (mt) trees as well as all well supported clades (BS > 90) were successively removed from the dataset resulting in 45 removal experiments (Fig. S1). Subsequently a distance tree based on the Link et al. algorithm (Link et al. 1995) was built for each reduced dataset with 500 bootstrap replicates using TREECON v. 1.3. The resulting trees and bootstrap support values were checked manually for all remaining clades.

Alternative phylogenetic positions of ambiguous taxa and/or clades were recorded and non-random signal deduced by calculating branch attachment frequencies (BAF) based on 1000 bootstrap trees using the program Phyutility v. 2.2. (Smith & Dunn 2008). Based on the obtained values a heat map representing the change of bootstrap support values over the whole dataset was generated. Outliers were defined as data points lying out of the range of the whiskers (calculated as 1.5* inter-quartile distance following Tukey 1977). Boxplots were generated for nodes showing cyto-nuclear discordance between AFLP and mitochondrial trees (Schwarzer et al. 2011).

Results

NeighborNet analysis

The NeighborNet gives a good representation of the phylogenetic signal (fit value = 99.92 %). The analysis based on 3031 AFLP loci revealed three main clusters, and several smaller subclusters (Fig. 2). One main cluster contains *S*. cf. *gibbiceps*, *S. glaber*, *S. mpozoensis* and *S*. cf. *tinanti* samples (GGMT, Fig. 2). Within this cluster each species forms a distinct subcluster, whereas relationships between the subclusters are affected by reticulate signal (indicated through "boxes" in the network). The second main cluster comprises *S. bleheri*, *S. rouxi*, *S*. sp. "red eye" and *S*. sp. "bulky head" from the Southern tributaries (STR, Fig. 2.). A clear conflict is present concerning the position of *S. bleheri*. The third cluster contains *S*. cf. *casuarius*, *S*. sp. "Maluku", *S*. sp "dwarf" (CMD) and *S*. sp. "Nki", *S*. sp. "Kisangani" and *S. ubanguiensis* from the Northern tributaries (NT). The whole third cluster will be abbreviated as CMDN in the following (Fig. 2).

Principal component analysis

In the Principal Component Analysis (PCA), ca. 44% of the variation is explained by the first six components. Principal Component (PC) 1 accounts for 22% and PC 2 for 9% of the total variation. The remaining four PC axes explain less than 5% but more than 2% of the variation (Fig. S2). On PC1 gross phylogenetic patterns are obvious: GGMT taxa are separated from CMDN taxa and specimens from northern and southern tributaries hold an intermediate position (Fig. S2) with *S*. sp. "bulky head" being located between *S. bleheri* and *S*. sp "red eye". Principal component 2 separates *S*. cf. *tinanti* from *S*. cf. *gibbiceps*, *S. glaber*, *S. mpozoensis* and the Northern and Southern tributaries and *S*. cf. *casuarius*, S. sp "Maluku" and S. sp "dwarf" from the rest. Specimens of *S*. cf. *tinanti* from Yalala (Fig. 1) are separated

from other *S*. cf. *tinanti* specimens by PC 2 and 6. On PC4 *S*. cf. *gibbiceps* from upstream (loc. 1-3, Fig. 1) are separated from the remaining *S.* cf. *gibbiceps*.

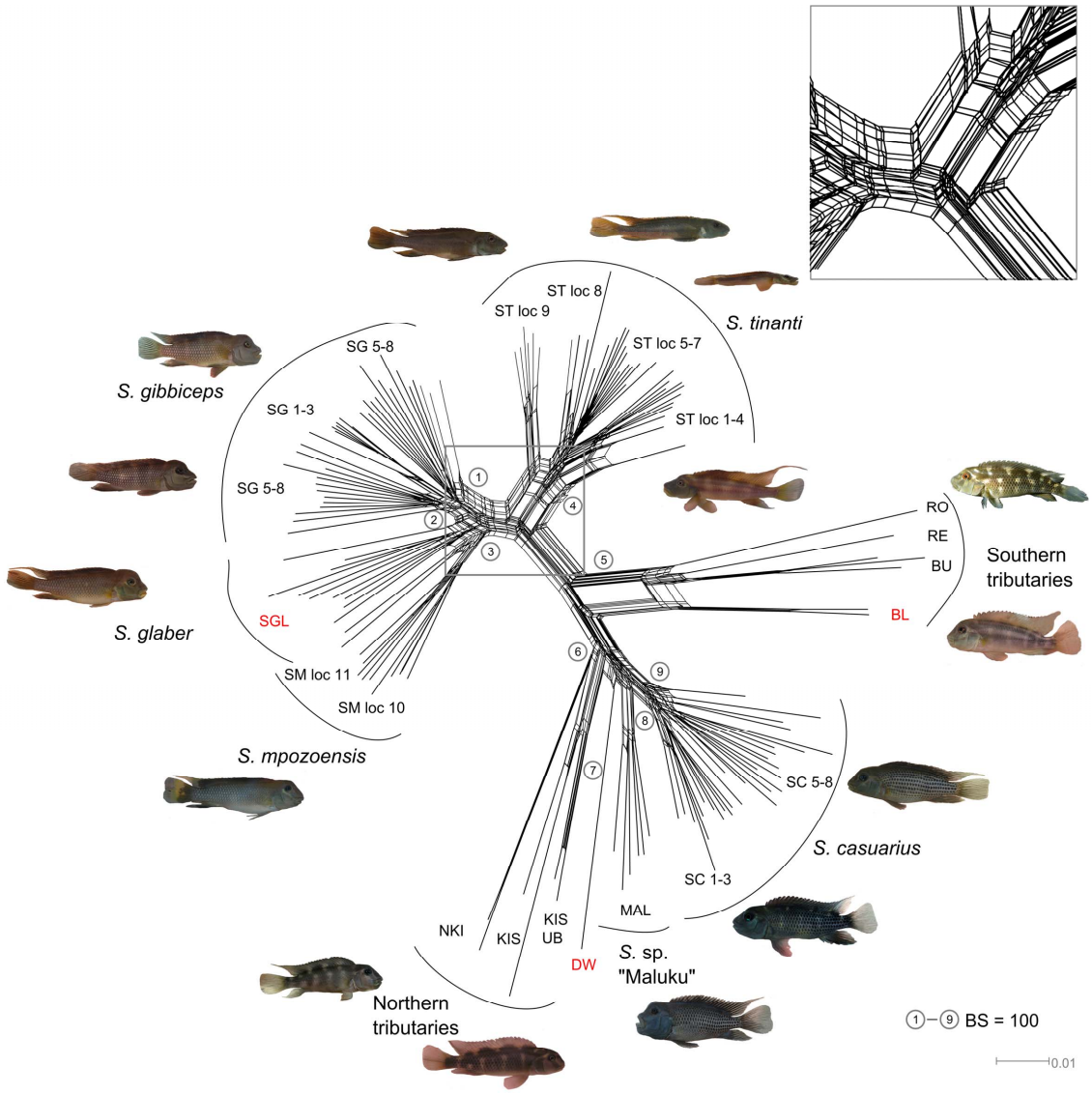


Figure 2 *Neighbornet graph based on AFLP data*

The NeighborNet was calculated based on Link et al., distances. The taxa indiacted in red are those that show cyto-nuclear discordances between nc- and mt- phylogenies in Schwarzer et *al.* (2011). The boxes at the base of the GGMT clade and indicate conflicting signal in the dataset.

Homoplasy excess test

Results of 45 removal experiments (Fig. S1) reveal scattered effects on node support across the whole phylogeny (Fig. 3). Strongest effects (more than 50% in- or decrease based on the mean bootstrap value) are detectable if clades including or comprising *S*. cf. *gibbiceps* are

removed (Fig. 3). Removal of the *S*. cf. *gibbiceps* subgroup from the central lower Congo leads to an increase of bootstrap (BS) support of several nodes within the CMDN clade (Fig 3). By the exclusion of *S*. cf. *gibbiceps* (central) the sistergroup relationship between *S. glaber* and *S. mpozoensis* increases by more than 100% of the initial value from BS = 6 to 81 (Fig. 3, Fig. 4a). At the same time support for the alternative relationship of *S.* cf. *gibbiceps* with *S. glaber* decreases from BS 90 to 12 (Fig. 4a). Further effects are the increase of BS-support of the sistergroup relationship of *S.* cf. *casuarius* and *S*. sp "Maluku" from BS = 65 to 94, of the *S*. cf. *casuarius* clade from sampling location 8 (Inga) from BS = 62 to 79 and of the BS support for a monophyletic *S*. cf *casuarius* clade from sampling locations 5 to 8 (central lower Congo) from BS = 30 to 93 (Fig. 3, Table S1). Simultaneously, support for alternative topologies uniting *S*. cf. *casuarius* populations from lower Congo sampling sites (loc. 1-7, Fig. 1) decreases from BS = 29 to 5. The BS support value for a monophyletic *S. casuarius* clade from sampling site 2 (Fig. 1) increases from 73 to 92 (Table S1). Within the Southern tributary clade (STR) the exclusion of *S.* cf. *gibbiceps* (central) entails an increase of the BS support for the relationship of *S. rouxi*/S. sp. "red eye"/S. sp. "bulky head" from BS = 75 to 99 (Fig. 4b). Removal of *S.* cf. *casuarius* from Inga effect the support for the sistergroup relationship of *S. casuarius* and *S*. sp "Maluku" as the BS value increases from to 65 to 80. The BS support value for *S.* cf. *casuarius* from lower Congo sampling locations 5-7 (Fig. 1) increases from 59 to 87.

Exclusion of *S*. sp. "red eye" leads to a decrease of BS support for the node comprising the two southern tributary species *S. rouxi* and *S*. sp. "bulky head" to BS= 27, whereas the support for the relationship of *S. bleheri* and *S.* sp. "bulky head" increases (BS = 15 to 78). Removal of *S*. sp. "bulky head" from the dataset leads to an increased BS support for a sistergroup relationship of *S. rouxi* and *S*. sp. "red eye" (from BS = 75 to 97) and the exclusion of *S. bleheri* leads to an increase of bootstrap support (from BS = 77 to 84, Node B, Fig. 4b) for the sistergroup relationship of S. sp. "red eye" and S. sp. "bulky head". Removal of *S.* sp. "dwarf" increases the BS support for the monophyly of the NT species from BS = 51 to 62 and the removal of the northern tributary species *S*. sp. "Kisangani" to BS = 95 (Fig 4c).

Branch attachment frequencies

In cases with low BS support or cyto-nuclear discordant signal alternative topologies were evaluated in 1000 bootstrap replicates. Phylogenetic relationships of a monophyletic clade composed of species from Southern tributaries and Malebo Pool (clade STR, *S. bleheri*, *S.*

rouxi, *S*. sp "red eye" and *S*. sp "bulky head") remain ambiguous. The analysis of 1000 bootstrap replicates (BS) reveals that 50% of the trees support a sistergroup relationship with all remaining *Steatocranus* species (Node A) whereas 41% support a sistergroup relationship with clade GGMT (Node A2, *S.* cf. *gibbiceps*/*S. glaber*/*S. mpozoensis* and the *S.* cf. *tinanti* species). In the remaining 9% of all bootstrap trees the Southern tributaries appear as sistergroup to clade CMDN (composed of *S. casuarius* and *S*. cf. *casuarius* , *S.* sp. "dwarf", *S.* sp. "Maluku" and species from Northern tributaries, Node A3, Fig. S3). *Steatocranus bleheri* appears as sistergroup to remaining taxa from the Southern tributaries (congruent with the nc signal) in 74% of the bootstrap trees. Eleven percent support a relationship to *S*. sp "red eye" and *S*. "bulky head" (congruent with the mtDNA signal) and 15% a closer relationship to *S*. sp "bulky head". *Steatocranus* species from Northern tributaries (NT) to the Congo River appear monophyletic, but with low BS support (Fig. 1, BS = 51%). The relationship of *S*. sp "Nki" and *S*. sp "Lefini" to the remaining NT taxa is not resolved. Alternative topologies in 1000 bootstrap replicates indicate a sistergroup relationship of *S*. sp "Nki" to the complete CMDN clade (node **B2** 31%) or to the CMD clade (node **B3** 18% Fig. S4).

Discussion

Inter- and intraspecific gene flow is increasingly recognized as an important factor shaping speciation. Several examples of hybrid species are known from natural populations (Nolte et al. 2006; Mavarez & Linares 2008; Lucek et al. 2010; Brelsford et al. 2011) and even the origin of whole adaptive radiations of ancestral hybrid origin was repeatedly postulated (Joyce et al. 2011; Hudson et al. 2011). Conflicting genetic information caused by hybridization can be reflected in phylogenetic incongruences between different genes (Seehausen 2004) and visualized in networks (Huson & Bryant 2006). In frequently hybridizing groups, like plants or corals (Willis et al. 2006; Soltis & Soltis 2009), network rather than tree generation are common practice in phylogenetic analyses (Linder & Rieseberg 2004). To robustly deduce hybrid signal from networks, however, additional experimental approaches are necessary. In the present study we applied a stepwise jackknifing approach based on single taxa or clades to assess approximate ancient or possibly ongoing gene flow patterns.

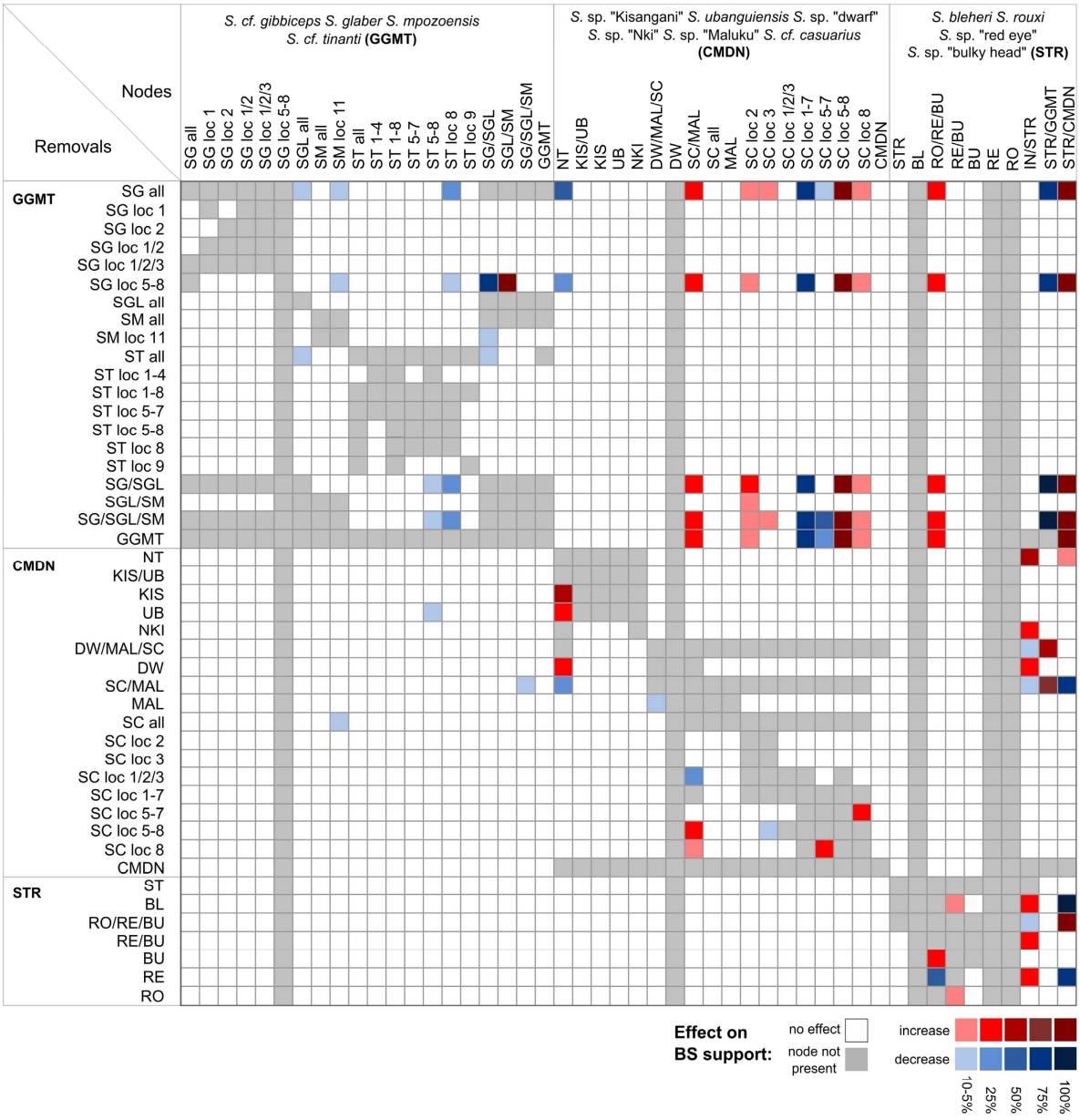


Figure 3 *Heatmap – overview of all removal experiments*

Increase of bootstrap support is indicated in red and decrease is blue. The intensity of the colour corresponds to the strength of the effect referring to the median of the affected Node. Only "outlier" values are shown, defined as bootstrap values lying outside the range of the 25% and 75% quartiles and whiskers. White boxes indicate no change in BS- values and grey boxes the absence of a node. SG = *S*. cf. *gibbiceps*, SGL = *S. glaber*, SM = *S. mpozoensis*, ST = *S*. cf. *tinanti*, SC = *S*. cf. *casuarius*, GGMT = *S*. cf. *gibbiceps*/*S. glaber*/ *S. mpozoensis*/ *S*. cf. *tinanti* clade, NT = Northern tributaries, KIS= *S*. sp. "Kisangani", UB = *S. ubanguisensis*, NKI = *S.* sp. "Nki", DW = *S*. sp. "dwarf", MAL = *S*. sp "Maluku", CMDN = *S*. cf. *casuarius*/ *S*. sp "Maluku"/ *S*. sp. "dwarf"/ Nothern tributaries clade, STR = Southern tributaries, BL = *S. bleheri*, RO = *S. rouxi*, RE = *S*. sp. "red eye", BU = *S*. sp "bulky head".

Potential sources and interpretation of conflict signal

Adopting a neutral theory of molecular evolution (Kimura 1983), conflict signal in phylogenetic datasets can accumulate solely by chance (due to the stochastical nature of the substitution process). Potential source of conflict signal might also arise through background "noise" caused by methological uncertainties. It is, however, especially in our context, necessary to separate "real" conflict signal due to gene flow from randomly occurring homoplasies. By using a combined methodical approach and focusing only on strong effects (e.g. outlier in the HET) we can solve this issue. The retention of ancient shared polymorphisms or incomplete lineage sorting can cause similar patterns compared with hybridization or gene flow (Seehausen 2004). The first (as a function of time) is expected to be distributed randomly among closely related species and is expected to mainly affect recently diverged species or populations rather than deeper phylogenetic splits (Maddison & Knowles 2006). In contrast, gene flow can be assumed to be unequally distributed (e.g. higher between neighboring populations than among distantly located ones). The *Steatocranus* radiation is estimated to have emerged about 7 mya (Schwarzer et al. 2011), which is old compared to the species rich cichlid radiations of e.g. Lakes Victoria and Malawi (Genner et al. 2007). Furthermore NeighborNet and previous DNA sequence based analyses of a reduced data set (Schwarzer et al. 2011) resulted in mainly well resolved clades indicating that a large fraction of the observed gene flow is not of recent origin (Fig. 2). This, in combination with the unequally distributed patterns of gene flow correlates makes the assumption that ancient shared polymorphism or incomplete lineage sorting explains the observed patterns rather unlikely.

Are *Steatocranus* species connected by ancient and ongoing geneflow?

Our results strongly support that *Steatocranus* species form a genomic network with the widely distributed *S.* cf. *gibbiceps* serving as vector between the distantly related and localized species assemblages along the lower Congo (Fig. 5). Additional reticulate signal is present among distantly related and more recently evolved *Steatocranus* species. Main reticulate patterns can be derived from results of the homoplasy excess test (HET, Fig. 3), further supported by those from branch attachment frequencies (BAF) principal components (PC) and the NeigbourNet analysis. In the following, we will discuss these patterns and their potential impact on our study system (1) within potential founder species, (2) between founder species and lower Congo taxa and (3) within the lower Congo *Steatocranus* species.

Gene flow between founder species has the potential to increase initial evolutionary diversity of colonizing species through acquisition of new beneficial traits e.g. through recombination and transgressive segregation (Lewontin & Birch 1966; Nolte & Sheets 2005; Stelkens et al. 2009b; Stelkens & Seehausen 2009b; Larsen et al. 2010; Aghnoum & Niks 2011). The lower Congo rapids were colonized in at least two waves (Fig. 1) at about 0.94 (0.3-1.7, *S.* cf. *casuarius*) and 4.48 (3.3-5.8, GGMT clade) mya apparently right after the assumed initial formation of the modern lower Congo River (Ferry et al. 2004; Schwarzer et al. 2011).

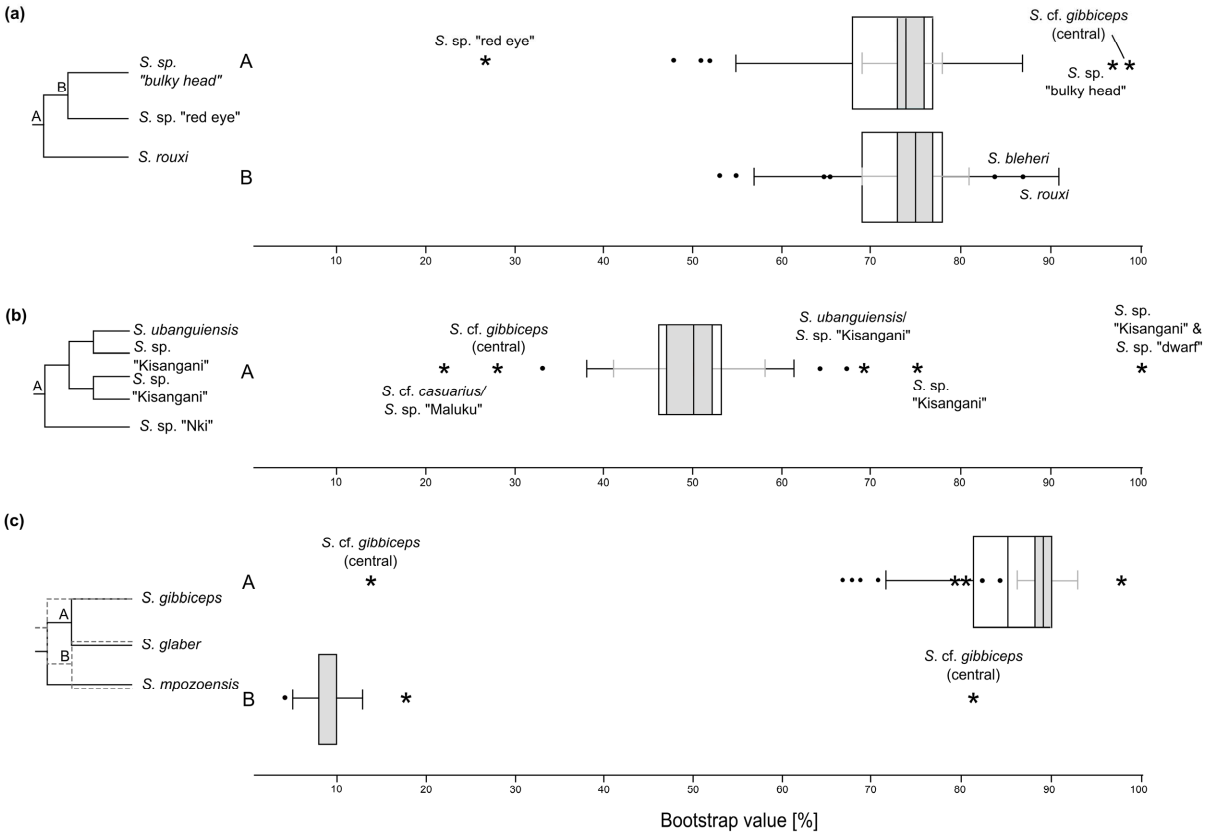


Figure 4 *Boxplots of homoplasy excess test*

The boxplots show the distribution of bootstrap support values (%) for (a) alternative topologies concerning the phylogenetic position of *S. glaber* (b) nodes within the Southern and (c) Northern tributaries. The shown boxplots refer to discordances between mt- and AFLP datasets (Schwarzer et al. 2011). Each clade was removed iteratively from the dataset (resulting in N = 45 experiments, Fig. S1) and NJ-trees based on 500 bootstrap replicates were recalculated using Treecon. Outliers are shown as asterisks. The zero-distributions (derived from 100 randomly conducted removals) are shown as empty boxes whereas the actual distributions of BS- support values (based on the 45 removal experiments, Fig. S1) are marked in grey.

Congenerics of the lower Congo *Steatocranus* are distributed in tributaries north and south to the River, the Congo mainstream, as well as upstream of the lower Congo rapids, in Malebo Pool (Fig. 1). Within both groups reticulate signal is present based on our data (Fig. 2, Fig. 3), pointing to ancient hybridization events. For example cyto-nuclear discordances concerning the phylogenetic position of *S. bleheri* and *S.* sp. “dwarf” (both from Malebo pool) are associated either with Southern or Northern tributary species (Fig. 2). Gene flow has been present among the STR species *S. bleheri*, *S*. sp. “bulky head” and *S*. sp “red eye”, supported by HET (BL, Fig. 3, Fig 4b). The removal of *S*. sp. “dwarf” and *S*. sp. “Kisangani” raises the initial BS support from 51 to 100 for the monophyly of the remaining NT species, indicating the accumulation of homoplasious signal through those taxa.

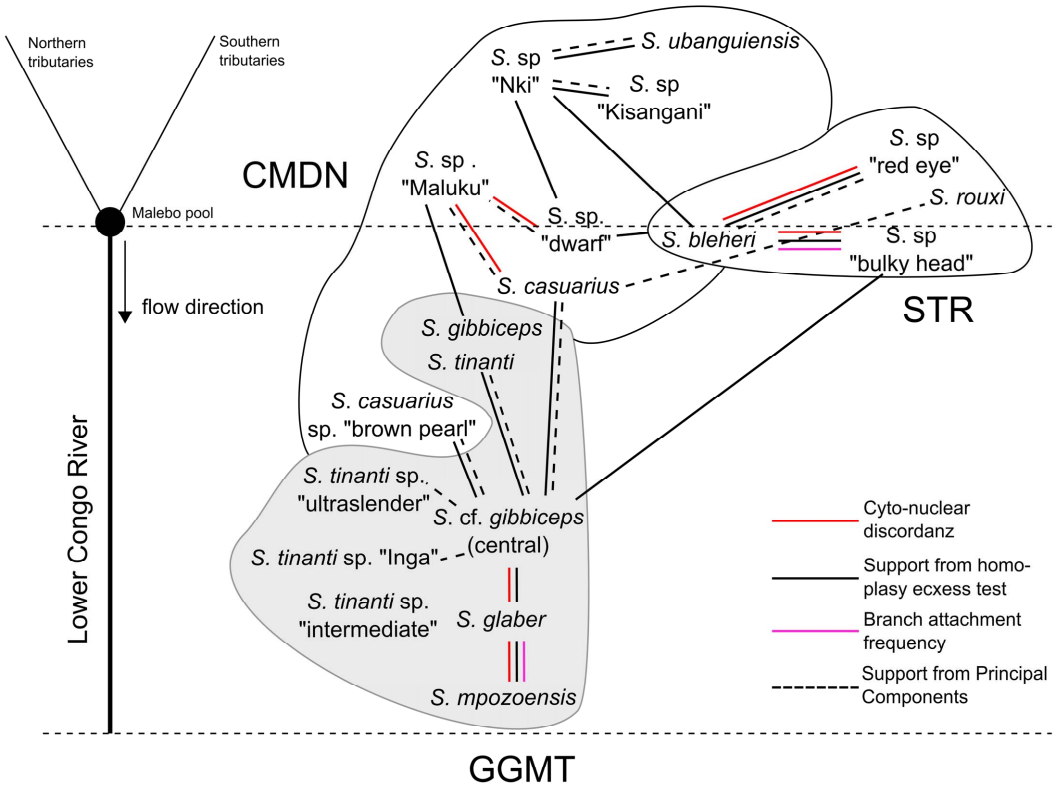


Figure 5 *Reticulate network (Scheme)*

A schematic image of connections (via gene flow) between and within the three major phylogenetic *Steatocranus* groups is shown. The lower Congo is sketched on the left side to display the approximate spatial placement of the taxa. Results supporting the linkage are displayed in different colors (cyto-nuclear discordance and branch attachment frequency) and/or different line types (homoplasy excess test, PCA).

Alternative topologies in 1000 bootstrap replicates indicate a relationship of *S*. sp "Nki" with members of the CMD (*S*. cf. *casuarius*, *S*. sp "Maluku" and *S*. sp "dwarf") clade (Fig. S2). Past gene flow can be assumed indicated by their present day allopatric distributions (except for *S. bleheri* and *S.* sp. "bulky head") and clear phylogenetic separation. Phylogenetic relationships of a monophyletic clade composed of species from Southern tributaries and Malebo pool (clade STR) remain ambiguous. The removal of *S*. cf. *gibbiceps* (central), *S. bleheri* and *S*. sp "Nki" from the dataset increases the BS support for a sistergroup relationship of the STR species to all remaining *Steatocranus* to BS = 96% (Fig. 4). There is no such effect if only *S*. cf. *gibbiceps* from upstream lower Congo localities 1 to 3, *S. glaber*, *S. mpozoensis* or any of the *S.* cf. *tinanti* clades are removed (Fig. 3). This highlights a non-randomly distributed impact of southern tributary taxa on distantly related lower Congo clades (Fig. 4), indicating selective hybridization events in the evolutionary history of *Steatocranus*. Initial colonization of the lower Congo by a hybrid swarm appears rather unlikely, as theory predicts that the inclusion of any species of the radiation will have similar homoplasy effects, as they share the same progenitors (Seehausen 2004). This is clearly not the case in *Steatocranus*.

After a radiation has accomplished a stage where two or more incipient species have formed, occasional or localized gene-flow between these species could facilitate and thereby accelerate further ecological diversification ("syngameon hypothesis", Seehausen 2004). There are strong indications that the local-endemic species *S. glaber* (restricted from below Inga to Yalala in the lower Congo, Fig.1) is a species of hybrid origin. This assumption is supported by results of the HET (Fig. 3), the evaluation of branch attachment frequencies (BAF) in 1000 BS replicates and PC (Fig. S2). Parental species are most likely *S*. cf. *gibbiceps* (central) and *S. mpozoensis*. The spatial proximity, with *S. glaber* occurring at Inga, an intermediate and partly overlapping areal of distributions of *S*. cf. *gibbiceps* and *S. mpozoensis* (Fig. 1) and yet unpublished morphometric and population genetic results further render a hybrid origin of *S. glaber* very likely.

Conclusions

With this case study on the lower Congo cichlid genus *Steatocranus,* we demonstrate that a high degree of reticulate signal can be present despite dominant phylogenetic signal favoring a well resolved and dichotomous phylogenetic hypothesis. We conclude that the course of speciation in *Steatocranus* has repeatedly been affected by hybridization events. An initial

colonization of the lower Congo River by a hybrid population appears, however, unlikely. To date few zoological studies exists, that critically tested for reticulate phylogenetic signal in their datasets (but see Seehausen 2004; Willis et al. 2006; Jiggins et al. 2008; Brelsford et al. 2011). The lower Congo *Steatocranus* are most likely just one of various yet undetected examples of reticulate networks in vertebrates. Their close relationship to the East-African radiations (Schwarzer et al. 2009) and the commonness of hybridization in this group (Danley et al. 2000; Salzburger et al. 2002; Schliewen & Klee 2004; Schelly et al. 2006; Stelkens & Seehausen 2009a; Stelkens et al. 2009a; Joyce et al. 2011) make a closer look at the impact of hybridization on these mega-diverse radiations and cichlids in general especially promising. This will certainly change and potentially further strengthened the role of hybridization in speciation research.

Authors contributions

JS and UKS designed the study. JS carried out the molecular work. JS, BM and UKS designed and conducted the analyses. All authors contributed to the preparation of the manuscript. They read and approved the final version.

Acknowledgments

This work was supported by grants of the Deutsche Forschungsgemeinschaft to B.M. (DFG MI649/8-1) and U.K.S. (SCHL567/4-1) and a graduate student grant of the University of Bonn as well as a travel grant from the supporters of the Zoologisches Forschungsmuseum A. Koenig to J.S. Our thank goes to R. Schelly and M. Stiassny for company in the field, tissue samples and valuable information on selected specimens. For indispensable assistance in the field we thank D. Neumann, J. Frommen, P. Alibert, A. Dunz, M. Levy, V. Mamonekene, A. Ibal-Zamba, J. Punga, U. Ali-Patho, P. Mongindo, C. Danadu, F. Bapeamoni and T. Kadangé Ngongo. Special thanks go to K. Langen, F. Eppler and B. Müller who helped with lab routines.

CHAPTER 7

General discussion

The present study introduces the riverine cichlid genera of the lower Congo rapids as potential model in speciation research. These cichlids provide an outstanding example for the underestimated diversity and evolutionary importance of riverine cichlids. Emphasis was laid on: (1) The phylogenetic placement and chronological classification of the Congolean genera within the African cichlids, (2) the reconstruction of intrageneric relationships and population differentiation within two selected lower Congo cichlid genera, and (3) the role of hybridization on riverine and lacustrine species flocks. Results are discussed in the light of the proposed Palaeohistory of East and Central Africa.

Riverine cichlid species in the light of African cichlid evolution

The largest fraction of species diversity in cichlids trace back to one group, the haplochromines, which exhibit a hotspot of diversity in the East African Lakes Malawi and Victoria (Kocher 2004; Seehausen 2006). Their high evolutionary success is associated with morphological key innovations like egg spots (Salzburger et al. 2005) and breeding strategies and can partly be correlated with lake-size and lake age (Seehausen 2006). Fundamental reasons for the diversity of this lineage in comparison to other, less diverse cichlid lineages remain unclear. Despite many studies on cichlids are based on the mega-diverse haplochromine radiations of the East African lakes (Sturmbauer & Meyer 1993; Seehausen & van Alphen 1999; Sturmbauer et al. 2001; Salzburger et al. 2002a; Seehausen & Schluter 2004; Salzburger & Meyer 2004; Spady et al. 2005; Salzburger et al. 2005; Sefc et al. 2007; Koblmüller et al. 2008a; Koblmüller et al. 2008b; Joyce et al. 2011) their seeding lineages were never reliably identified. In general, the relevance of riverine species was underrepresented in this context, as often not more than an incomplete subsample of riverine species was included in phylogenetic studies (Verheyen et al. 2003; Salzburger & Meyer 2004; Salzburger et al. 2005). Moreover, most studies rely on mitochondrial markers only (Salzburger et al. 2002a; Verheyen et al. 2003; Salzburger et al. 2005; Koblmüller et al. 2008a), leaving the potential impact of hybridization on phylogenetic reconstructions aside (Chan & Levin 2005). In recent studies, the inclusion of more (or relevant) riverine taxa reveal their substantial impact on the East African lacustrine radiations in lakes Victoria and

Malawi (Seehausen et al. 2003; Joyce et al. 2011). Adding additional riverine Congolese and Angolan taxa in the present study unveils extensive cyto-nuclear discordances among the whole haplochromine phylogeny (Chapter 4). Potential riverine founder species, including those from the lower Congo River, here capture a key role. Whereas lower Congo *Haplochromis* species cluster within a Congolese clade based on an extensive AFLP dataset, they spread within the East-African radiations and Southern haplochromines based on mitochondrial markers, indicating an impact on the origin of the Lake Victoria "superflock". The generally adopted hypothesis of a lacustrine origin of the haplochromines (Salzburger et al. 2002a; Salzburger et al. 2005) could not be ruled out based on the data presented in this study. Indications of unequal contributions to gene flow by two included species of the proposed ancestors to lacustrine diversity, the Tropheini, raise doubts on the simplicity of the hypothesis (Chapter 4). To tackle this question in more depth, a detailed taxon sampling concerning the lacustrine species would be necessary.

Adaptive radiations are thought to be confined mainly to lacustrine habitats (Seehausen 2006), as fluvial systems lack the multiplicity of ecological opportunities necessary for the formation of adaptive radiations (Joyce et al. 2005). This would explain the generally lower species diversity in rivers. One exception is known from South African serranochromine cichlids, which are, however, also proposed to be of lacustrine origin (in Palaeolake Magadigadi) (Joyce et al. 2005). Sullivan et al. (2002; 2004), Feulner et al. (2007; 2008) and Kullander et al. (2010) gave examples for riverine species flocks of weakly electric fish (Mormyridae) in the Ogooué River system in Gabon and the lower Congo River and South American cichlids restricted to the upper Rio Uruguay system, respectively. The distributions of these species flocks were either very broad and encompassed several river systems (2002; 2004) or intrageneric sampling was not complete with regard to known taxa and sampled areas (Feulner et al. 2007; Feulner et al. 2008). The lower Congo is estimated to have emerged about 5 mya (Ferry et al. 2004) coinciding with molecular age estimates for the origin of lower Congo *Steatocranus* and *Nanochromis* species ranging from 0.94 to 4.48 mya (median ages). This strongly indicates the emergence of their present day diversity in a riverine habitat. *Steatocranus* and *Nanochromis* species of the lower Congo River constitute the first known examples of species flocks originating in small-scale riverine environment (Chapter 5). Potential reasons for this might be caused by the complexity of the underlying habitat. The strong directionality of the River (through high current), in combination with

traversing rapids alternated by still water habitats offer a variety of ecological opportunities for colonizing populations. Furthermore, it promotes an initial allopatric separation of populations. A spatial genetic structure is present in all widely distributed cichlid species (see also Markert et al. 2010) corresponding to a trisection of the lower Congo River into biogeographic areas with locally endemic species assemblages. A high degree of endemism is present below the Inga rapids. A potential scenario explaining this, is that following the first colonization wave early colonists of *Steatocranus* and *Nanochromis* have remained strongly isolated in downstream regions of the River, e.g. by a waterfall at Inga. Riverbed erosion may then have worn down this waterfall resulting in its present day permeability. This assumption remains, however, speculative, as to date no supporting geological data is available.

Biogeographical implications

The geological history of East and Central Africa was affected by several major palaeoclimatic and tectonic events since the late Eocene/Oligocene concomitant with a structural change of waterbodies and river systems (Leturmy et al. 2003; Lucazeau et al. 2003). Droughts have been associated with lake level fluctuations in Lakes Malawi and Tanganyika at the last glacial maximum (Scholz & Rosendahl 1988) or in Lake Victoria at the late Pleistocene desiccation (ca. 15,600 – 14,700 years before present (yr bp), Johnson et al. 1996). The occurrence of a large Palaeolake in the late Holocene (ca. 2000 yr bp) most likely shaped the present day riverine cichlid species diversity in South African Rivers (Joyce et al. 2005). In contrast to lacustrine systems, the influence of these events on riverine systems was less well studied. Biogeographic inferences from riverine fauna and flora in combination with age estimates offer the potential to further underpin palaeo-geological hypotheses. Phylogenetic splitting of austro- and boreotilapiines for example coincides with a drought in East Africa in the Oligocene/early Miocene (Davis et al. 2002; Loader et al. 2007) which is assumed to have spaciously influenced distribution patterns of the African fauna and flora, e.g. in rainforest trees (Couvreur et al. 2008) and caecilian amphibians (Loader et al. 2007). Phylogenetic analyses of chromidotilapiines support a proposed connection between the Ogooué system and the Congo basin in the early Oligocene (Leturmy et al. 2003; Lucazeau et al. 2003). This ancient connection was assumed as major offshore deposits were absent in the region of the present day Congo mouth until relatively recent. Large parts of central Africas interior drainage discharged in the Cretaceous and Oligocene (65 to 36 mya) into the

Atlantic ocean through the Ogooué valley in Gabon and the Cuanza system in Angola (Leturmy et al. 2003; Lucazeau et al. 2003), indicating a minimum age of the modern lower Congo rapids of 35 mya. Recent analyses of Congo offshore deposits in line with molecular age estimated (Chapter 5) indicate an origin of the present day Congo discharge at the Miocene-Pliocene transition at approximately 5 mya (Ferry et al. 2004), after the southern African continent had been affected by a significant uplift inducing a progressive rearrangement of the watersheds (Lavier et al. 2001). This corroborates the view that the lower Congo cichlid radiation was closely linked to the establishment of new habitat availability and thus represents an autochthonous radiation within the lower Congo River. All age estimates strongly depend on used molecular markers, kind of molecular clock estimates and the choice of priors (Ho et al. 2011). Priors for all age estimates in the present study were based on a mixture of different genes, a relaxed molecular clock approach (Drummond et al. 2006; Drummond & Rambaut 2007), one reliable fossil prior (*Oreochromis lorenzoi* †) (Carnevale et al. 2003) and the 95 % credible interval estimates for African Cichlidae from Azuma et al. (2008) based on non-cichlid teleostean fossils or Gondwana fragmentation (for a detailed discussion on prior choice see Chapter 2, additional file 7).

The role of hybridization in cichlid speciation

Hybridization can roughly entail two consequences: (1) it can contribute to species diversity by enhancing evolutionary potential (Arnold 1997) or (2) it may lead to a collapse of existing species borders (Kirkpatrick & Ravigne 2002). Environmental factors can shape the outcome of hybridization (Seehausen et al. 2008). While hybridization accompanied by a loss of environmental structure might rather induce a collapse of species borders (Seehausen et al. 1997), hybridization in combination with accruing new ecological opportunities e.g. by colonization of novel or extreme habitats (Rieseberg et al. 2003; Schliewen & Klee 2004; Nolte et al. 2005) might favor the formation of new species. Hybridization seems to be a common phenomenon in cichlids (Salzburger et al. 2002; Seehausen 2004; Schliewen & Klee 2004), and is assumed to have significantly contributed to the species richness of the East African adaptive radiations (Seehausen 2004). Viable hybrids can occur even among distantly related cichlid species. Stelkens et al. (Stelkens et al. 2010) crossed 26 species pairs covering the entire East African haplochromine radiation and demonstrated that pre-mating isolation initially accumulates fast in cichlids but then nearly stagnates with increasing genetic distance. Stelkens et al. (2010) successfully produced viable offspring via species crossing up

to an upper border of proposed species divergence of 4.4/8.5/18.4 mya (depending on molecular clock algorithm and prior choice, Stelkens et al. 2010). The origin of the lower Congo River is assumed at 5mya providing new and extreme habitat opportunities to potential colonizers. *Steatocranus* and *Nanochromis* radiations of the lower Congo were apparently shaped by hybridization with adjacent congenerics. Concordant with Stelkens et al. (2010) hybridization occurred also among distantly related lineages prior to a colonization of the lower Congo (Chapter 5, 6). Concerning single taxa indications for hybridization were also present within ancient Haplotilapiine (*T. mariae*) and Chromidotilapiine (*C. schoutedeni*) lineages (Chapters 2, 3). In haplochromines, reticulate evolution is present across the whole group and among presently allopatric species (Chapter 4). Riverine *Orthochromis* as well as "*Haplochromis*" species from the Congo basin capture a key role in this context and raise doubts about the assumed colonization of the great African Lakes by Lake Tanganyika precursors (Salzburger et al. 2005).

A strong (dominant) phylogenetic signal might mask ancient or rare hybridization events and impede their detection. A careful evaluation of the underlying (weaker) signal should therefore be conducted to capture as much information as possible (Chapter 6). Reliance on networks rather than bifurcating trees is common practice in groups known to frequently hybridize, like plants (Linder & Rieseberg 2004; Soltis & Soltis 2009) or corals (Willis et al. 2006), as conflicting phylogenetic signal can be visualized (Bryant & Moulton 2004). A trade-off exists in distinguishing conflicting phylogenetic signal from background "noise", potentially caused by methological uncertainties. In the present study a combined methodical approach based on phylogenetic and experimental analyses was used, focusing on strong but not solely bifurcating tree-induced effects (e.g. bootstrap outlier in the homoplasy excess test, Chapter 4 and 6). The retention of ancient shared polymorphisms or incomplete lineage sorting can result in similar patterns compared with hybridization or gene flow (Seehausen 2004). Incomplete lineage sorting compared to gene flow, however, is assumed to cause random effects among closely related species, whereas gene flow patterns are assumed to be unequally distributed (e.g. higher between neighboring populations than among distantly located ones) and also detectable between older lineages. Age estimates for the origin of the analyzed cichlid genera are with ~7.7 mya for *Steatocranus,* ~ 8 mya for *Nanochromis* and ~5.3 mya for the earliest split in haplochromines (Koblmüller et al. 2008a) relatively old. The age in combination with the unequally distributed patterns of gene flow

(Chapter 4-6) makes the assumption that ancient shared polymorphism or incomplete lineage sorting as cause for the observed patterns rather unlikely. Based on results of the present study a population genetic approach on lower Congo *Steatocranus* with 17 microsatellites was conducted (Spieth 2011). Polymorphic microsatellites were established through cross-species amplification with primers designed for *Oreochromis niloticus* (Lee et al. 2005). The results of the study corroborate the fine-scale differentiation along the lower Congo rapids (Chapter 5) and a cut in species composition at Inga falls (Spieth 2011). The peculiarity of this small river-stretch around Inga is further highlighted by the restriction of the potential hybrid species, *Steatocranus glaber* to this area (Chapter 6). *Steatocranus glaber* is locally endemic from below the Inga rapids to Yalala in the lower Congo (Fig. 1). Results of the present study, partly overlapping spatial distributions with the proposed parental species *S.* cf. *gibbiceps* and *S. mpozoensis* (Fig. 1) and yet unpublished morphometric and population genetic results (Spieth 2011) render a hybrid origin of *S. glaber* very likely. Evidence for hybrid speciation in animals stems to date primarily from low resolution molecular data (Mallet 2007). Genomic mapping of ecological or speciation related hybrid traits has rarely been conducted in animals so far (but see Stemshorn et al. 2005; 2011), and the genetic background and perquisites leading to a "successful" homoploid hybrid speciation remain unknown (Nolte & Tautz 2010). It is therefore reasonable to draw significantly more attention to the evolutionary dynamics of hybrid genomes in future studies.

Prospects

The emergence of new habitat opportunities in the lower Congo River about 5 mya, most likely concomitant with further rearrangements of watersheds enabled and potentially maintained hybridization between Congolean cichlids. Strongest indications come from *Steatocranus* species. Based on the present results a study is ongoing based on 16 microsatellite markers and 379 individuals to resolve the population structure and the gene flow patterns within *Steatocranus* species with emphasis on the potential contact zone of *S. glaber* and hypothesized parental species (Spieth 2011). Allele distributions of microsatellites indicate that the population structure within *Steatocranus* is even more complex than assumed by the present study. The region around the huge rapids of Inga does not only mark a break in species composition but yields a high degree of local fine-scale

differentiation (Spieth 2011). Based on the results of the present study and the microsatellite survey, the goal of our future research will be, to study the genomic architecture and associated phenotypic and ecological divergence of multiple cichlid species assemblages around the rapids of Inga. Based on next generation sequencing (NGS) methods and extensive fine-scale sampling we expect to get a more detailed understanding of speciation processes in the light of present gene-flow.

Even though, it can be assumed that the diversity of lower Congo cichlids was covered by extensive sampling in field trips from 2002 to 2009, the lower Congo is in parts hardly accessible and sampling is time consuming and costly. Especially for *Nanochromis* it cannot be excluded that yet undiscovered species exist in the Congo basin. All until now newly discovered species, however, could be phylogenetically assigned in biogeographic clusters (e.g. in clusters comprising species from Northern or Southern tributaries to the lower Congo River). This makes this uncertainty for the study of lower Congo species controllable. A further drawback of the system lies in the heterogeneity and strong current of the River that prevents from a standardized recording of habitat parameters. To derive genotype-environment associations (GEAs Bierne et al. 2011) one has to rely on indirect estimators, like stable isotopes (Layman et al. 2007) and ecologically induced shape parameters (obtained e.g. by geometric morphometrics). Some clear advantages are, however, the conciseness of the system (less than 20 species in each genus), allowing to study genera *in toto* combined with the unusual species diversity which originated in a riverine habitat. This and the key role that the lower Congo cichlids capture in cichlid evolution (e.g. haplochromines and *Steatocranus*) makes them, nevertheless, an important, though untypical, model system for speciation research.

References

Aghnoum, R. and Niks, R. E. 2011 Transgressive segregation for very low and high levels of basal resistance to powdery mildew in barley. *Journal of Plant Physiology* 168, 45-50.

Albertson, R. C. and Kocher, T. D. 2005 Genetic architecture sets limits on transgressive segregation in hybrid cichlid fishes. *Evolution* 59, 686-690.

Albertson, R. C., Markert, J. A., Danley, P. D., and Kocher, T. D. 1999 Phylogeny of a rapidly evolving clade: the cichlid fishes of Lake Malawi, East Africa. *Proceedings of the National Academy of Sciences of the United States of America* 96, 5107-5110.

Alves, P. C., Harris, D., Melo-Ferreira, J., Branco, M., Ferrand, N., Suchentrunk, F., Melo-Ferreira, J., and Boursot, P. 2006 Hares on thin ice: Introgression of mitochondrial DNA in hares and its implications for recent phylogenetic analyses. *Molecular Phylogenetics and Evolution* 40, 640-641.

Arnold, M. L. 1997 *Natural Hybridization and Evolution*. New York: Oxford University Press.

Azuma, Y., Kumazawa, Y., Miya, M., Mabuchi, K., and Nishida, M. 2008 Mitogenomic evaluation of the historical biogeography of cichlids toward reliable dating of teleostean divergences. *BMC Evolutionary Biology* 8.

Ballard, J. W. O. and Whitlock, M. C. 2004 The incomplete natural history of mitochondria. *Molecular Ecology* 13, 729-744.

Bell, M. A. and Travis, M. P. 2005 Hybridization, transgressive segregation, genetic covariation, and adaptive radiation. *Trends in Ecology and Evolution* 20, 358-361.

Bierne, N., Welch, J., Loire, E., Bonhomme, F., and David, P. 2011 The coupling hypothesis: why genome scans may fail to map local adaptation genes. *Molecular Ecology* 20, 2044-2072.

Boulenger, G. A. 1899 Matériaux pour la faune du Congo. Zoologie. Série I. Poissons nouveaux du Congo. Quatrième Partie. Polyptères, Clupes, Mormyres, Characins. *Annales du Musée Royal du Congo Belge Sciences Zoologiques* 1, 59-96.

Brelsford, A., Mila, B., and Irwin, D. E. 2011 Hybrid origin of Audubon's warbler. *Molecular Ecology* 20, 2380-2389.

Bryant, D. and Moulton, V. 2004 Neighbor-Net: An agglomerative method for the construction of phylogenetic networks. *Molecular Biology and Evolution* 21, 255-265.

Burke, J. M. and Arnold, M. L. 2001 Genetics and the fitness of hybrids. *Anual Review of Genetics* 35, 31-52.

Carnevale, G., Sorbini, C., and Landini, W. T. 2003 Oreochromis lorenzoi, a new species of tilapiine cichlid from the late Miocene of central Italy. *Journal of Vertebrate Paleontology* 23, 508-516.

Chan, K. M. A. and Levin, S. A. 2005 Leaky prezygotic isolation and porous genomes: Rapid introgression of maternally inherited DNA. *Evolution* 59, 720-729.

Chow, S. and Hazama, K. 1998 Universal PCR primers for S7 ribosomal protein gene introns in fish. *Molecular Ecology* 7, 1255-1256.

Cohen, A. S., Soreghan, M. J., and Scholz, C. A. 1993 Estimating the age of formation of lakes: an example from Lake Tanganyika, East African Rift system. *Geology* 21, 511-514.

Collier, G. E., Murphy, W. M. J., and Espinoza, M. 2009 Phylogeography of the genus Epiplatys (Aplocheiloidea: Cyprinodontiformes). *Molecular Phylogenetics and Evolution* 50, 190-196.

Colyn, M. 1991 *L'importance zoogeographique de bassin du fleuve Zaire pour la speciation: le cas des primates simiens*. Tervuren, Belgique: Musée royal de l'Afrique centrale.

Contreras-Sanchez, W. M. and Fitzsimmons, K. 2006 Proceedings from the 7th international symposium on Tilapia aquaculture in Boca Del Rio, Vera Cruz, Mexico. (ed. World Aquaculture Society). Mexico.

Couvreur, T. L. P., Chatrou, L. W., Sosef, M. S. M., and Richardson, J. E. 2008 Molecular phylogenetics reveal multiple tertiary vicariance origins of the African rain forest trees. *BMC Biology* 6.

Coyne, J. A. and Orr, H. A. 2004 *Speciation*. Sunderland: Sinauer Associates.

Danley, P. D., Markert, J. A., Arnegard, M. E., and Kocher, T. D. 2000 Divergence with gene flow in the rock-dwelling cichlids of Lake Malawi. *Evolution* 54, 1725-1737.

Davis, C. C., Bell, C. D., Fritsch, P. W., and Mathews, S. 2002 Phylogeny of Acridocarpus-brachylophon (Malpighiaceae): Implications for tertiary tropical floras and afroasian biogeography. *Evolution* 56, 2395-2405.

Dieckmann, U. and Doebeli, M. 1999 On the origin of species by sympatric speciation. *Nature* 400, 354-357.

Dixon, C. J., Schonswetter, P., Suda, J., Wiedermann, M. M., and Schneeweiss, G. M. 2009 Reciprocal Pleistocene origin and postglacial range formation of an allopolyploid and its sympatric ancestors (Androsace adfinis group, Primulaceae). *Molecular Phylogenetics and Evolution* 50, 74-83.

Doebeli, M. and Dieckmann, U. 2003 Speciation along environmental gradients. *Nature* 421, 259-264.

Drummond, A. J., Ho, S. Y. W., Phillips, M. J., and Rambaut, A. 2006 Relaxed phylogenetics and dating with confidence. *Plos Biology* 4, 699-710.

Drummond, A. J. and Rambaut, A. 2007 BEAST: Bayesian evolutionary analysis by sampling trees. *BMC Evolutionary Biology* 7.

Endler, J. A. 1977 *Geographic variation, speciation, and clines*. Princeton: Princeton Press.

Farias, I. P., Orti, G., and Meyer, A. 2000 Total evidence: Molecules, morphology, and the phylogenetics of cichlid fishes. *Journal of Experimental Zoology* 288, 76-92.

Ferry, J. N., Babonneau, N., Mulder, T., Parize, O., and Raillard, S. 2004 Morphogenesis of Congo submarine canyon and valley: implications about the theories of the canyons formation. *Geodinamica Acta* 17, 241-251.

Feulner, P. G. D., Kirschbaum, F., Mamonekene, V., Ketmaier, V., and Tiedemann, R. 2007 Adaptive radiation in African weakly electric fish (Teleostei: Mormyridae: *Campylomormyrus*): a combined molecular and morphological approach. *Journal of Evolutionary Biology* 20, 403-414.

Feulner, P. G. D., Kirschbaum, F., and Tiedemann, R. 2008 Adaptive radiation in the Congo River: An ecological speciation scenario for African weakly electric fish (Teleostei; Mormyridae; *Campylomormyrus*). *Journal of Physiology-Paris* 102, 340-346.

Gasse, F. 2000 Hydrological changes in the African tropics since the Last Glacial Maximum. *Quaternary Science Reviews* 19, 189-211.

Gavrilets, S. 2000 Waiting time to parapatric speciation. *Proceedings of the Royal Society of London Series B-Biological Sciences* 267, 2483-2492.

Genner, M. J., Seehausen, O., Lunt, D. H., Joyce, D. A., Shaw, P. W., Carvalho, G. R., and Turner, G. F. 2007 Age of cichlids: New dates for ancient lake fish radiations. *Molecular Biology and Evolution* 24, 1269-1282.

Giresse, P. 2005 Mesozoic–Cenozoic history of the Congo Basin. *Journal of African Earth Sciences* 43, 301-315.

Greenwood, P. H. 1984 What is a species flock? In *Evolution of Fish Species Flocks* (eds. A. A. Echelle and I. Kornfield), pp. 13-19. Orono, Maine: University of Maine at Orono Press.

Greenwood, P. H. 1987 The genera of pelmatochromine fishes (Teleostei, Cichlidae). A phylogenetic review. *Bulletin of the British Museum (Natural History) Zoology* 53, 139-203.

Hammer, O., Harper, D. A. T., and Ryan, P. D. 2001 PAST: Paleontological Statistics software package for education and data analysis. *Palaeontologia Electronica* 4, 9-pp.

Herder, F., Nolte, A. W., Pfaender, J., Schwarzer, J., Hadiaty, R. K., and Schliewen, U. K. 2006 Adaptive radiation and hybridization in Wallace's Dreamponds: evidence from sailfin silversides in the Malili Lakes of Sulawesi. *Proceedings of the Royal Society B-Biological Sciences* 273, 2209-2217.

Ho, S. Y., Lanfear, R., Bromham, L., Phillips, M. J., Soubrier, J., Rodrigo, A. G., and Cooper, A. 2011 Time-dependent rates of molecular evolution. *Molecular Ecology* 20, 3087-3101.

Hudson, A. G., Vonlanthen, P., and Seehausen, O. 2011 Rapid parallel adaptive radiations from a single hybridogenic ancestral population. *Proceedings of the Royal Society B-Biological Sciences* 278, 58-66.

Huelsenbeck, J. P. and Ronquist, F. 2001 MrBAYES: Bayesian inference of phylogenetic trees. *Bioinformatics* 17, 754-755.

Huson, D. H. and Bryant, D. 2006 Application of phylogenetic networks in evolutionary studies. *Molecular Biology and Evolution* 23, 254-267.

Jackson, P. R., Oberg, K. A., Gardiner, N., and Shelton, J. M. 2009 Velocity Mapping in the Lower Congo River: A First Look at the Unique Bathymetry and Hydrodynamics of Bulu Reach. *Proceedings of the IAHR Symposium on River Coastal and Estuarine Morphodynamics 2009, 6th, Santa Fe, Argentina*, 1007-1014.

Jiggins, C. D., Salazar, C., Linares, M., and Mavarez, J. 2008 Hybrid trait speciation and Heliconius butterflies. *Philosophical Transactions of the Royal Society B-Biological Sciences* 363, 3047-3054.

Johnson, T. C., Kelts, K., and Odada, E. 2000 The Holocene history of Lake Victoria. *Ambio* 29, 2-11.

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

Jombart, T., Pontier, D., and Dufour, A. B. 2009 Genetic markers in the playground of multivariate analysis. *Heredity* 102, 330-341.

Joyce, D. A., Lunt, D. H., Bills, R., Turner, G. F., Katongo, C., Duftner, N., Sturmbauer, C., and Seehausen, O. 2005 An extant cichlid fish radiation emerged in an extinct Pleistocene lake. *Nature* 435, 90-95.

Joyce, D. A., Lunt, D. H., Genner, M. J., Turner, G. F., Bills, R., and Seehausen, O. 2011 Repeated colonization and hybridization in Lake Malawi cichlids. *Current Biology* 21, R108-R109.

Katongo, C., Koblmuller, S., Duftner, N., Makasa, L., and Sturmbauer, C. 2005 Phylogeography and speciation in the *Pseudocrenilabrus philander* species complex in Zambian Rivers. *Hydrobiologia* 542, 221-233.

Katongo, C., Koblmuller, S., Duftner, N., Mumba, L., and Sturmbauer, C. 2007 Evolutionary history and biogeographic affinities of the serranochromine cichlids in Zambian rivers. *Molecular Phylogenetics and Evolution* 45, 326-338.

Kimura, M. 1983 *The neutral theory of molecular evolution.* Cambridge.

Kirkpatrick, M. and Ravigne, V. 2002 Speciation by natural and sexual selection: models and experiments. *American Naturalist* 159, S22-S35.

Klett, V. and 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 fishes. *Molecular Biology and Evolution* 19, 865-883.

Koblmüller, S., Duftner, N., Katongo, C., Phiri, H., and Sturmbauer, C. 2005 Ancient divergence in bathypelagic Lake Tanganyika deepwater cichlids: Mitochondrial phylogeny of the tribe Bathybatini. *Journal of Molecular Evolution* 60, 297-314.

Koblmüller, S., Schliewen, U. K., Duftner, N., Sefc, K. M., Katongo, C., and Sturmbauer, C. 2008a Age and spread of the haplochromine cichlid fishes in Africa. *Molecular Phylogenetics and Evolution* 49, 153-169.

Koblmüller, S., Sefc, K. M., and Sturmbauer, C. 2008b The Lake Tanganyika cichlid species assemblage: recent advances in molecular phylogenetics. *Hydrobiologia* 615, 5-20.

Koblmüller, S., Egger, B., Sturmbauer, C., and Sefc, K. M. 2010 Rapid radiation, ancient incomplete lineage sorting and ancient hybridization in the endemic Lake Tanganyika cichlid tribe Tropheini. *Molecular Phylogenetics and Evolution* 55, 318-334.

Kocher, T. D., Conroy, J. A., Mckaye, K. R., Stauffer, J. R., and Lockwood, S. F. 1995 Evolution of NADH dehydrogenase subunit 2 in east African cichlid fish. *Molecular Phylogenetics and Evolution* 4, 420-432.

Kocher, T. D. 2004 Adaptive evolution and explosive speciation: The cichlid fish model. *Nature Reviews Genetics* 5, 288-298.

Kondrashov, A. S. and Mina, M. V. 1986 Sympatric speciation: when is it possible? *Biological Journal of the Linnean Society* 27, 201-223.

Kondrashov, A. S. and Shpak, M. 1998 On the origin of species by means of assortative mating. *Proceedings of the Royal Society of London Series B Biological Sciences* 265, 2273-2278.

Konings, A. 2007 *Malawi Cichlids in their natural habitat*, 4 edn. El Paso: Cichlid Press.

Krijgsman, W., Hilgen, F. J., and Marabini, S. V. G. 1999 New palaomagnetic and cyclostratigraphic age constrains on the Messinian of the Northern Apennines (Vena del Gesso Basin, Italy). *Memorie della Società Geologica Italiana* 54, 25-33.

Kullander, S. O. 1998 A phylogeny and classification of the South American Cichlidae (Teleostei: Perciformes). In *Phylogeny and Classification of Neotropical Fishes* (eds. L. R. Malabarba, R. E. Reis, R. P. Vari, Z. M. S. de Lucena, and C. A. S. de Lucena), pp. 461-498. Porto Alegre: Edipucrs.

Kullander, S. O., Noren, M., Fridriksson, G. B., and de Lucena, C. A. S. 2010 Phylogenetic relationships of species of *Crenicichla* (Teleostei: Cichlidae) from southern South America based on the mitochondrial cytochrome *b* gene. *Journal of Zoological Systematics & Evolutionary Research* 48, 248-258.

Lamboj, A. 2005 *Nanochromis sabinae*, a new cichlid species (Teleostei, Cichlidae) from the upper Congo River area and northeast Gabon. *Zootaxa* 827, 1-11.

Lamboj, A. 2004 *Pelvicachromis signatus* and *Pelvicachromis rubrolabiatus*, two new cichlid species (Teleostei, Perciformes) from Guinea, West Africa. *Zootaxa* 454, 1-12.

Lamboj, A. 2001 Revision des *Chromidotilapia batesii/finleyi*-Komplexes (Teleostei, Perciformes), mit der Beschreibung einer neuen Gattung und dreier neuer Arten. *Verhandlungen der Gesellschaft für Ichthyologie* 2, 11-47.

Lamboj, A. 2002 *Chromidotilapia mrac*, a new species of Cichlidae (Teleostei: Perciformes) from Gabon. *Ichthyological Exploration of Freshwaters* 13, 251-256.

Lamboj, A. 1999 Cichlids of the biospere reservation of Dimonika (Mayombe, Congo) with description of two new *Chromidotilapia* species (Teleostei, Perciformes). *Verhandlungen der Gesellschaft für Ichthyologie* 1, 139-156.

Lamboj, A. 2003 *Chromidotilapia melaniae* and *C. nana*, two new cichlid species (Perciformes, Cichlidae) from Gabon, Central Africa. *Zootaxa* 143, 1-15.

Lamboj, A. and Schelly, R. 2006 *Nanochromis teugelsi*, a new cichlid species (Teleostei: Cichlidae) from the Kasai Region and central Congo basin. *Ichthyological Exploration of Freshwaters* 17, 247-254.

Lamboj, A. and Snoeks, J. 2000 *Divandu albimarginatus*, a new genus and species of cichlid (Teleostei: Cichlidae) from Congo and Gabon, Central Africa. *Ichthyological Exploration of Freshwaters* 11, 355-360.

Lamboj, A. and Stiassny, M. L. J. 2003 Three new *Parananochromis* species (Teleostei, Cichlidae) from Gabon and Cameroon, Central Africa. *Zootaxa* 209, 1-19.

Larsen, P. A., Marchan-Rivadeneira, M. R., and Baker, R. J. 2010 Natural hybridization generates mammalian lineage with species characteristics. *Proceedings of the National Academy of Sciences of the United States of America* 107, 11447-11452.

Lavier, L. L., Steckler, M. S., and Brigaud, F. 2001 Climatic and tectonic control on the Cenozoic evolution of the West African margin. *Marine Geology* 178, 63-80.

Layman, C. A., Arrington, D. A., Montana, C. G., and Post, D. M. 2007 Can stable isotope ratios provide for community-wide measures of trophic structure? *Ecology* 88, 42-48.

Lee, B. Y., Lee, W. J., Streelman, J. T., Carleton, K. L., Howe, A. E., Hulata, G., Slettan, A., Stern, J. E., Terai, Y., and Kocher, T. D. 2005 A second-generation genetic linkage map of tilapia (*Oreochromis* spp.). *Genetics* 170, 237-244.

Legendre, P. and Legendre, L. 1998 *Numerical Ecology*, 2nd edition edn: Elsevier Science BV, Amsterdam.

Leturmy, P., Lucazeau, F., and Brigaud, F. 2003 Dynamic interactions between the gulf of Guinea passive margin and the Congo River drainage basin: 1. Morphology and mass balance. *Journal of Geophysical Research-Solid Earth* 108.

Lewontin, R. C. and Birch, L. C. 1966 Hybridization as a source of variation for adaptation to new environments. *Evolution* 20, 315-336.

Li, C. H., Lu, G. Q., and Orti, G. 2008 Optimal data partitioning and a test case for ray-finned fishes (Actinopterygii) based on ten nuclear loci. *Systematic Biology* 57, 519-539.

Linder, C. R. and Rieseberg, L. H. 2004 Reconstructing patterns of reticulate evolution in plants. *American Journal of Botany* 91, 1700-1708.

Link, W., Dixkens, C., Singh, M., Schwall, M., and Melchinger, A. E. 1995 Genetic diversity in European and Mediterranean faba bean germ plasm revealed by RAPD markers. *Theoretical and Applied Genetics* 90, 27-32.

Lippitsch, E. 1995 Scale and squamation character polarity and phyletic assessment in the family Cichlidae. *Journal of Fish Biology* 47, 91-106.

Liu, K., Wang, F., Chen, W., Tu, L., Min, M. S., Bi, K., and Fu, J. 2010 Rampant historical mitochondrial genome introgression between two species of green pond frogs, *Pelophylax nigromaculatus* and *P. plancyi*. *BMC Evolutionary Biology* 10.

Loader, S. P., Pisani, D., Cotton, J. A., Gower, D. J., Day, J. J., and Wilkinson, M. 2007 Relative time scales reveal multiple origins of parallel disjunct distributions of African caecilian amphibians. *Biology Letters* 3, 505-508.

Lucazeau, F., Brigaud, F., and Leturmy, P. 2003 Dynamic interactions between the Gulf of Guinea passive margin and the Congo River drainage basin: 2. Isostasy and uplift. *Journal of Geophysical Research* 108, 2384.

Lucek, K., Roy, D., Bezault, E., Sivasundar, A., and Seehausen, O. 2010 Hybridization between distant lineages increases adaptive variation during a biological invasion: stickleback in Switzerland. *Molecular Ecology* 19, 3995-4011.

Maddison, W. P. and Knowles, L. L. 2006 Inferring phylogeny despite incomplete lineage sorting. *Systematic Biology* 55, 21-30.

Mallet, J. 2005 Hybridization as an invasion of the genome. *Trends in Ecology & Evolution* 20, 229-237.

Mallet, J. 2007 Hybrid speciation. *Nature* 446, 279-283.

Mallet, J., Beltrán, M., Neukirchen, W., and Linares, M. 2007 Natural hybridization in heliconiine butterflies: the species boundary as a continuum. *BMC Evolutionary Biology* 7, 1-15.

Markert, J. A., Schelly, R. C., and Stiassny, M. L. J. 2010 Genetic isolation and morphological divergence mediated by high-energy rapids in two cichlid genera from the lower Congo rapids. *BMC Evolutionary Biology* 10.

Mavarez, J. and Linares, M. 2008 Homoploid hybrid speciation in animals. *Molecular Ecology* 17, 4181-4185.

Mavarez, J., Salazar, C. A., Bermingham, E., Salcedo, C., Jiggins, C. D., and Linares, M. 2006 Speciation by hybridization in *Heliconius* butterflies. *Nature* 441, 868-871.

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

Mayr, E. 1942 *Systematics and the origin of species*. New York: Columbia University Press.

Mayr, E. 1963 *Animal Species and Evolution*. Massachusetts: Belknap Press, Cambridge.

Melo-Ferreira, J., Boursot, P., Suchentrunk, F., Ferrand, N., and Alves, P. C. 2005 Invasion from the cold past: extensive introgression of mountain hare (*Lepus timidus*) mitochondrial DNA into three other hare species in northern Iberia. *Molecular Ecology* 14, 2459-2464.

Misof, B. and Misof, K. 2009 A Monte Carlo approach successfully identifies randomness in multiple sequence alignments: a more objective means of data exclusion. *Systematic Biology* 58, 21-34.

Murphy, W. J. and Collier, G. E. 1997 A molecular phylogeny for aplocheiloid fishes (Atherinomorpha, Cyprinodontiformes): The role of vicariance and the origins of annualism. *Molecular Biology and Evolution* 14, 790-799.

Myers, G. S. 1939 The possible identity of the Congo fish *Teleogramma* with the cichlid genus *Leptolamprolodus*. *Stanford Ichthyological Bulletin* 1, 160.

Nagl, S., Tichy, H., Mayer, W. E., Samonte, I. E., McAndrew, B. J., and Klein, J. 2001 Classification and phylogenetic relationships of African tilapiine fishes inferred from mitochondrial DNA sequences. *Molecular Phylogenetics and Evolution* 20, 361-374.

Nei, M. and Li, W. H. 1979 Mathematical-Model for Studying Genetic-Variation in Terms of Restriction Endonucleases. *Proceedings of the National Academy of Sciences of the United States of America* 76, 5269-5273.

Nolte, A. W., Freyhof, J., Stemshorn, K. C., and Tautz, D. 2005 An invasive lineage of sculpins, Cottus sp (Pisces, Teleostei) in the Rhine with new habitat adaptations has originated from hybridization between old phylogeographic groups. *Proceedings of the Royal Society B-Biological Sciences* 272, 2379-2387.

Nolte, A. W. and Sheets, H. D. 2005 Shape based assignment tests suggest transgressive phenotypes in natural sculpin hybrids (Teleostei, Scorpaeniformes, Cottidae). *Frontiers in Zoology* 2, 1-12.

Nolte, A. W., Freyhof, J., and Tautz, D. 2006 When invaders meet locally adapted types: rapid moulding of hybrid zones between sculpins (*Cottus*, Pisces) in the Rhine system. *Molecular Ecology* 15, 1983-1993.

Nolte, A. W. and Tautz, D. 2010 Understanding the onset of hybrid speciation. *Trends in Genetics* 26, 54-58.

Numrich, R. 2003 Zur Kenntnis einiger Nanochromis-Arten des zentralen Kongogebietes. *Aquaristik Fachmagazin* 171, 44-46.

Nylander, J. A. A., Ronquist, F., Huelsenbeck, J. P., and Nieves-Aldrey, J. L. 2004 Bayesian phylogenetic analysis of combined data. *Systematic Biology* 53, 47-67.

Oliver, M. K. 1984 Recognition of *Heterochromis multidens* (Pellegrin) from the central Zaire River system as the most primitive African cichlid. In: *Systematics of African cichlid fishes: determination of the most primitive taxon, and studies on the haplochromines of Lake Malawi (Teleostei: Chichlidae)* pp. 5-100: Yale University.

Parnell, N. F., Hulsey, C. D., and Streelman, J. T. 2008 Hybridization produces novelty when the mapping of form to function is many to one. *BMC Evolutionary Biology* 8, 1-11.

Poll, M. 1986 Classification des cichlidae du lac Tanganika: tribus, genres et aspeces. *Acad R Belg Mem Classe Sci Coll -8°-2° ser , tome XLV, fasc* 2, 1-163.

Poll, M. and Thys van den Audenaerde, D. F. E. 1967 Description de *Pelmatochromis schoutedeni* sp. n. du Congo oriental (Pisces, Cichlidae). *Revue de Zoologie et de Botanique Africaines* 75, 383-391.

Pompanon, F., Bonin, A., Bellemain, E., and Taberlet, P. 2005 Genotyping errors: Causes, consequences and solutions. *Nature Reviews Genetics* 6, 847-859.

Rambaut, A. and Drummond, A. J. 2007 Tracer. Available from: http://beast.bio.ed.ac.uk/Tracer.

Reyment, R. A. and Dingle, R. V. 1987 Palaeogeography of Africa during the cretaceous period. *Palaeogeography, Palaeoclimatology, Palaeoecology* 59, 93-116.

Ribbink, A. J., Marsh, B. A., Marsh, A. C., Ribbink, A. C., and Sharp, B. J. 1983 A preliminary survey of the cichlid fishes of rocky habitats in Lake Malawi. *South African Journal of Zoology* 18, 149-310.

Rieseberg, L. H., Archer, M. A., and Wayne, R. K. 1999 Transgressive segregation, adaptation, and speciation. *Heredity* 83, 363-372.

Rieseberg, L. H., Raymond, O., Rosenthal, D. M., Lai, Z., Livingstone, K., Nakazato, T., Durphy, J. L., Schwarzbach, A. E., Donovan, L. A., and Lexer, C. 2003 Major ecological transitions in wild sunflowers facilitated by hybridization. *Science* 301, 1211-1216.

Rieseberg, L. H. and Wendel, J. F. 1993 Introgression and its consequences in plants. In *Hybrid zones and the evolutionary process* (ed. R. G. Harrison). Oxford: Oxford University Press.

Rieseberg, L. H. and Linder, C. R. 1999 Hybrid classification: Insights from genetic map-based studies of experimental hybrids. *Ecology* 80, 361-370.

Robert, M. 1946 *Le Congo Physique*, 3 edn. Liege: H. Vaillant-Carmanna, S.a., Impr. de L´Academie.

Roberts, T. R. and Stewart, D. J. 1976 An ecological and systematic survey of fishes in the rapids of the lower Zaire or Congo River. *Bulletin of the Museum of Comparative Zoology* 147, 239-317.

Roberts, T. R. and Kullander, S. O. 1994 Endemic cichlid fishes of the Fwa River, Zaïre: systematics and ecology. *Ichthyological Exploration of Freshwaters* 5, 97-154.

Ruber, L. and Adams, D. C. 2001 Evolutionary convergence of body shape and trophic morphology in cichlids from Lake Tanganyika. *Journal of Evolutionary Biology* 14, 325-332.

Runge, J. 2001 On the age of stone-lines and hillwash sediments in the eastern Congo basin – palaeoenvironmental implications. *Palaeoecology of Africa* 27, 19-36.

Runge, J. 2008 The Congo River, Central Africa. In *Large Rivers: Geomorphology and Management*, pp. 293-308: J. Wiley & Sons Ltd.

Russell, J. M. and Johnson, T. C. 2005 A high-resolution geochemical record from Lake Edward, Uganda Congo and the timing and causes of tropical African drought during the late Holocene. *Quaternary Science Reviews* 24, 1375-1389.

Salazar, C., Baxter, S. W., Pardo-Diaz, C., Wu, G., Surridge, A., Linares, M., Bermingham, E., and Jiggins, C. D. 2010 Genetic evidence for hybrid trait speciation in *Heliconius* butterflies. *Plos Genetics* 6.

Salzburger, W., Meyer, A., Baric, S., Verheyen, E., and Sturmbauer, C. 2002a Phylogeny of the Lake Tanganyika cichlid species flock and its relationship to the Central and East African haplochromine cichlid fish faunas. *Systematic Biology* 51, 113-135.

Salzburger, W., Baric, S., and Sturmbauer, C. 2002b Speciation via introgressive hybridization in East African cichlids? *Molecular Ecology* 11, 619-625.

Salzburger, W. and Meyer, A. 2004 The species flocks of East African cichlid fishes: recent advances in molecular phylogenetics and population genetics. *Naturwissenschaften* 91, 277-290.

Salzburger, W., Mack, T., Verheyen, E., and Meyer, A. 2005 Out of Tanganyika: Genesis, explosive speciation, key-innovations and phylogeography of the haplochromine cichlid fishes. *BMC Evolutionary Biology* 5, 1-15.

Salzburger, W. 2009 The interaction of sexually and naturally selected traits in the adaptive radiations of cichlid fishes. *Integrative and Comparative Biology* 49, E149.

Sanderson, M. J. 2002 Estimating absolute rates of molecular evolution and divergence times: A penalized likelihood approach. *Molecular Biology and Evolution* 19, 101-109.

Sanderson, M. J. 2003 r8s: inferring absolute rates of molecular evolution and divergence times in the absence of a molecular clock. *Bioinformatics* 19, 301-302.

Savolainen, V., Anstett, M. C., Lexer, C., Hutton, I., Clarkson, J. J., Norup, M. V., Powell, M. P., Springate, D., Salamin, N., and Baker, W. J. 2006 Sympatric speciation in palms on an oceanic island. *Nature* 441, 210-213.

Schelly, R. C., Salzburger, W., Koblmüller, S., Duftner, N., and Sturmbauer, C. 2006 Phylogenetic relationships of the lamprologine cichlid genus *Lepidiolamprologus*

(Teleostei: Perciformes) based on mitochondrial and nuclear sequences, suggesting introgressive hybridization. *Molecular Phylogenetics and Evolution* 38, 426-438.

Schliewen, U. K. 1994 Sympatric speciation suggested by monophyly of crater lake cichlids. *Nature* 368, 629-633.

Schliewen, U. K. and Stiassny, M. L. J. 2003 *Etia nguti*, a new genus and species of cichlid fish from the River Mamfue, Upper Cross River basin in Cameroon, West-Central Africa. *Ichthyological Exploration of Freshwaters* 14, 61-71.

Schliewen, U. K. and Klee, B. 2004 Reticulate sympatric speciation in Cameroonian crater lake cichlids. *Frontiers in Zoology* 1, 1-12.

Schliewen, U. K. 2005 Cichlid species flocks in small Cameroonian lakes. In *Freshwater ecoregions of Africa: A conservation assessment* (eds. M. L. Thieme, R. Abell, M. L. J. Stiassny, P. Skelton, B. Lehner, E. Dinerstein, A. Kamdem Toham, N. Burgess, and D. Olson), pp. 58-60. Washington, DC: Island Press.

Schliewen, U. K. and Stiassny, M. L. J. 2006 A new species of *Nanochromis* (Teleostei: Cichlidae) from Lake Mai Ndombe, central Congo Basin, Democratic Republic of Congo. *Zootaxa* 1169, 33-46.

Schliewen, U. K. 2006a Buckelkopfcichliden aus dem Kongo Teil 2. *DATZ* 6.

Schliewen, U. K. 2006b Buckelkopfcichliden aus dem Kongo Teil1. *DATZ* 6.

Scholz, C. A. and Rosendahl, B. R. 1988 Low lake stands in Lakes Malawi and Tanganyika, East Africa, delineated with multifold seismic data. *Science* 240, 1645-1648.

Schwarzer, J., Misof, B., Tautz, D., and Schliewen, U. K. 2009b The root of the East African cichlid radiations. *BMC Evolutionary Biology* 9, 11.

Schwarzer, J., Misof, B., Ifuta, S. N., and Schliewen, U. K. 2011 Time and origin of cichlid colonization of the lower Congo rapids. *Plos One* 6, e22380.

Seehausen, O., vanAlphen, J. J. M., and Witte, F. 1997 Cichlid fish diversity threatened by eutrophication that curbs sexual selection. *Science* 277, 1808-1811.

Seehausen, O. and van Alphen, J. J. M. 1999 Color polymorphism and sex ratio distortion in a cichlid fish as an incipient stage in sympatric speciation by sexual selection. *Ecology Letters* 2, 367-378.

Seehausen, O. 2000 Explosive speciation rates and unusual species richness in haplochromine cichlid fishes: Effects of sexual selection. *Advances in Ecological Research* 31, 237-274.

Seehausen, O., Koetsier, E., Schneider, M. V., Chapman, L. J., Chapman, C. A., Knight, M. E., Turner, G. F., van Alphen, J. M., and Bills, R. 2003 Nuclear markers reveal unexpected genetic variation and a Congolese-Nilotic origin of the Lake Victoria cichlid species flock. *Proceedings of the Royal Society of London Series B Biological Sciences* 270, 129-137.

Seehausen, O. and Schluter, D. 2004 Male-male competition and nuptial-colour displacement as a diversifying force in Lake Victoria cichlid fishes. *Proceedings of the Royal Society of London Series B Biological Sciences* 271, 1345-1353.

Seehausen, O. 2004 Hybridization and adaptive radiation. *Trends in Ecology and Evolution* 19, 198-207.

Seehausen, O. 2006 African cichlid fishes: a model system in adaptive radiation research. *Proceedings of the Royal Society of London Series B Biological Sciences* 273, 1987-1998.

Seehausen, O., Takimoto, G., Roy, D., and Jokela, J. 2008 Speciation reversal and biodiversity dynamics with hybridization in changing environments. *Molecular Ecology* 17, 30-44.

Sefc, K. M., Baric, S., Salzburger, W., and Sturmbauer, C. 2007 Species-specific population structure in rock-specialized sympatric cichlid species in Lake Tanganyika, East Africa. *Journal of Molecular Evolution* 64, 33-49.

Shaw, K. L. 2002 Conflict between nuclear and mitochondrial DNA phylogenies of a recent species radiation: what mtDNA reveals and conceals about modes of speciation in Hawaiian crickets. *Proceedings of the National Academy of Sciences of the United States of America* 99, 16122-16127.

Smith, S. A. and Dunn, C. W. 2008 Phyutility: a phyloinformatics tool for trees, alignments and molecular data. *Bioinformatics* 24, 715-716.

Soltis, P. S. and Soltis, D. E. 2009 The Role of Hybridization in Plant Speciation. *Annual Review of Plant Biology* 60, 561-588.

Spady, T. C., Seehausen, O., Loew, E. R., Jordan, R. C., Kocher, T. D., and Carleton, K. L. 2005 Adaptive molecular evolution in the opsin genes of rapidly speciating cichlid species. *Molecular Biology and Evolution* 22, 1412-1422.

Sparks, J. S. and Smith, W. L. 2004 Phylogeny and biogeography of cichlid fishes (Teleostei: Perciformes: Cichlidae). *Cladistics* 20, 501-517.

Spieth, V. 2011 Gene flow patterns in the *Steatocranus* species of the lower Congo River. Diploma thesis, Rheinische Friedrich-Wilhelms Universität Bonn

Stamatakis, A. 2006 RAxML-VI-HPC: Maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. *Bioinformatics* 22, 2688-2690.

Stelkens, R. B., Schmid, C., Selz, O., and Seehausen, O. 2009 Phenotypic novelty in experimental hybrids is predicted by the genetic distance between species of cichlid fish. *BMC Evolutionary Biology* 9.

Stelkens, R. and Seehausen, O. 2009a Genetic distance between species predicts novel trait expression in their hybrids. *Evolution* 63, 884-897.

Stelkens, R. B. and Seehausen, O. 2009b Phenotypic divergence but not genetic distance predicts assortative mating among species of a cichlid fish radiation. *Journal of Evolutionary Biology* 22, 1679-1694.

Stelkens, R. B., Young, K. A., and Seehausen, O. 2010 The accumulation of reproductive incompatibilities in African cichlid fish. *Evolution* 64, 617-632.

Stemshorn, K. C., Nolte, A. W., and Tautz, D. 2005 A genetic map of *Cottus gobio* (Pisces, Teleostei) based on microsatellites can be linked to the physical map of *Tetraodon nigroviridis*. *Journal of Evolutionary Biology* 18, 1619-1624.

Stemshorn, K. C., Reed, F. A., Nolte, A. W., and Tautz, D. 2011 Rapid formation of distinct hybrid lineages after secondary contact of two fish species (*Cottus* sp.). *Molecular Ecology* 20, 1475-91.

Stiassny, M. L. J. 1990 *Tylochromis*, relationships and the phylogenetic status of the African Cichlidae. *American Museum Novitates* 1993, 1-14.

Stiassny, M. L. J. 1991 Phylogenetic intrarelationships of the family Cichliade: an overview. In *Cichlid fishes. Behaviour, ecology and evolution* (ed. M. H. A. Keenleyside), pp. 1-35. London: Chapman & Hall.

Stiassny, M. L. J. 1997 A phylogenetic overview of the lamprologine cichlids of Africa (Teleostei, Cichlidae): a morphological perspective. *South African Journal of Science* 93, 513-523.

Stiassny, M. L. J. and Schliewen, U. K. 2007 *Congochromis*, a new cichlid genus (Teleostei: Cichlidae) from Central Africa, with the description of a new species from the upper Congo River, Democratic Republic of Congo. *American Museum Novitates* 2007, 1-16.

Streelman, J. T. and Karl, S. A. 1997 Reconstructing labroid evolution with single-copy nuclear DNA. *Proceedings of the Royal Society B Biological Sciences* 264, 1011-1020.

Streelman, J. T., Peichel, C. L., and Parichy, D. M. 2007 Developmental genetics of adaptation in fishes: The case for novelty. *Annual Review of Ecology and Systematics* 38, 655-681.

Streelman, J. T., Zardoya, R., Meyer, A., and Karl, S. A. 1998 Multilocus phylogeny of cichlid fishes (Pisces: Perciformes): Evolutionary comparison of microsatellite and single-copy nuclear loci. *Molecular Biology and Evolution* 15, 798-808.

Sturmbauer, C., Baric, S., Salzburger, W., Ruber, L., and Verheyen, E. 2001 Lake level fluctuations synchronize genetic divergences of cichlid fishes in African lakes. *Molecular Biology and Evolution* 18, 144-154.

Sturmbauer, C. and Meyer, A. 1993 Mitochondrial phylogeny of the endemic mouthbrooding lineages of cichlid fishes from Lake Tanganyika in Eastern Africa. *Molecular Biology and Evolution* 10, 751-768.

Sugie, A., Terai, Y., Ota, R., and Okada, N. 2004 The evolution of genes for pigmentation in African cichlid fishes. *Gene* 343, 337-346.

Sullivan, J. P., Lavoue, S., and Hopkins, C. D. 2002a Discovery and phylogenetic analysis of a riverine species flock of African electric fishes (Mormyridae : Teleostei). *Evolution* 56, 597-616.

Sullivan, J. P., Lavoue, S., Arnegard, M. E., and Hopkins, C. D. 2004 AFLPs resolve phylogeny and reveal mitochondrial introgression within a species flock of African electric fish (Mormyroidea : Teleostei). *Evolution* 58, 825-841.

Sultmann, H., Mayer, W. E., Figueroa, F., Tichy, H., and Klein, J. 1995 Phylogenetic analysis of cichlid fishes using nuclear-DNA markers. *Molecular Biology and Evolution* 12, 1033-1047.

Takahashi, T. and Nakaya, K. 2002 Description and familial allocation of the African fluvial genus *Teleogramma* to the Cichlidae. *Ichthyological Research* 49, 171-180.

Templeton, A. R. 1981 Mechanism of speciation - a population genetic approach. *Annual Review of Ecology and Systematics* 12, 23-48.

Terai, Y., Morikawa, N., Kawakami, K., and Okada, N. 2003a The complexity of alternative splicing of hagoromo mRNAs is increased in an explosively speciated lineage in East African cichlids. *Proceedings of the National Academy of Sciences of the United States of America* 100, 12798-12803.

Terai, Y., Takahashi, K., Nishida, M., Sato, T., and Okada, N. 2003b Using SINEs to probe ancient explosive speciation: "Hidden" radiation of African cichlids? *Molecular Biology and Evolution* 20, 924-930.

Teugels, G. G. and Guegan, J.-F. 1994 Diversite biologique des poissons d'eaux douces de la Basse-Guinee et de l'Afrique Centrale. In *Biological diversity in African fresh- and brackish water fishes: geographical overviews* (eds. G. G. Teugels, J.-F. Guégan, and J.-J. Albaret), pp. 67-85. Tervuren, Belgie, Paris: Koninklijk Museum Voor Midden-Afrika, ORSTOM.

Thieme, M. L., Abell, R., Stiassny, M. L. J., Skelton, P., Lehner, B., Dinerstein, E., Kamdem Toham, A., Burgess, N., and Olson, D. 2005 *Freshwater ecoregions of Africa and Madagascar. A conservation assessment*. Washington, Covelo, London: Island Press.

Thorley, J. L. and Wilkinson, M. 1999 Testing the phylogenetic stability of early tetrapods. *Journal of Theoretical Biology* 200, 343-344.

Thys van den Audenaerde, D. F. E. 1968a A preliminary contribution to a systematic revision of the genus *Pelmatochromis* Hubrecht sensu lato (Pisces, Cichlidae). *Revue de Zoologie et de Botanique Africaines* 77, 349-391.

Thys van den Audenaerde, D. F. E. 1968b *An annotated bibliography of Tilapia (Pisces, Cichlidae)*. Tervuren, Belgique: Musée royal de l'Afrique centrale.

Thys van den Audenaerde, D. F. E. 1971 Some new data concerning the *Tilapia*-species of the subgenus *Coptodon* (Pisces, Cichlidae). *Revue de Zoologie et de Botanique Africaines* 84, 160-168.

Trewavas, E. 1973 I. On the cichlid fishes of the genus *Pelmatochromis* with proposal of a new genus for *P. congicus*: on the relationship between *Pelmatochromis* and *Tilapia* and the recognition of *Sarotherodon* as a distinct genus; and a new species of cichlid fishes of rivers Quanza and Bengo, Angola, with a list of the known Cichlidae of these rivers and a note on *Pseudocrenilabrus natalensis* Fowler. *Bulletin of the British Museum (Natural History) Zoology* 25, 1-26.

Trewavas, E. 1983 *Tilapiine fishes of the genera Sarotherodon, Oreochromis and Danakilia*. London: British Museum (Natural History).

Tukey, J. W. 1977 *Exploratory Data Analysis*.

Turelli, M., Barton, N. H., and Coyne, J. A. 2001 Theory and speciation. *Trends in Ecology & Evolution* 16, 330-343.

Turner, G. F., Seehausen, O., Knight, M. E., Allender, C. J., and Robinson, R. L. 2001 How many species of cichlid fishes are there in African lakes? *Molecular Ecology* 10, 793-806.

Van de Peer, Y. and Dewachter, R. 1993 Treecon - A Software package for the construction and drawing of evolutionary trees. *Computer Applications in the Biosciences* 9, 177-182.

van der Sluijs, I., van Dooren, T. J. M., Seehausen, O., and van Alphen, J. J. M. 2008 A test of fitness consequences of hybridization in sibling species of Lake Victoria cichlid fish. *Journal of Evolutionary Biology* 21, 480-491.

van Oijen, M. J. P. 1996 The generic classification of the haplochromine cichlids of Lake Victoria, East Africa. *Zoologische Verhandelingen (Leiden)* 302, 57-110.

van Oijen, M. J. P., Snoeks, J., Skelton, P. H., Maréchal, C., and Teugels, G. G. 1991 Haplochromis. In *Check-list of the Freshwater Fishes of Africa (CLOFFA) v. 4* (eds. J. Daget, J.-P. Gosse, G. G. Teugels, and D. F. E. Thys van den Audenaerde), pp. 100-184. Paris: ISNB Bruxelles, MRAC Tervuren, ORSTOM.

Vences, M., Freyhof, J., Sonnenberg, R., Kosuch, J., and Veith, M. 2001 Reconciling fossils and molecules: Cenozoic divergence of cichlid fishes and the biogeography of Madagascar. *Journal of Biogeography* 28, 1091-1099.

Verheyen, E., Salzburger, W., Snoeks, J., and Meyer, A. 2003 Origin of the superflock of cichlid fishes from Lake Victoria, East Africa. *Science* 300, 325-329.

Vos, P., Hogers, R., Bleeker, M., Reijans, M., Vandelee, T., Hornes, M., Frijters, A., Pot, J., Peleman, J., Kuiper, M., and Zabeau, M. 1995 Aflp - a New Technique for DNA-Fingerprinting. *Nucleic Acids Research* 23, 4407-4414.

Vreven, E. J. and Stiassny, M. L. J. 2009 *Mastacembelus simbi*, a new dwarf spiny eel (Synbranchiformes: Mastacembelidae) from the lower Congo River. *Ichthyological Exploration of Freshwaters* 20, 213-222.

Willis, B. L., Van Oppen, M. J. H., Miller, D. J., Vollmer, S. V., and Ayre, D. J. 2006 The role of hybridization in the evolution of reef corals. *Annual Review of Ecology Evolution and Systematics* 37, 489-517.

Wilson, C. C. and Bernatchez, L. 1998 The ghost of hybrids past: fixation of arctic charr (*Salvelinus alpinus*) mitochondrial DNA in an introgressed population of lake trout (*S. namaycush*). *Molecular Ecology* 7, 127-132.

Appendix

Chapter 2: The root of the East African cichlid radiations

Table S1 *List of all taxa and genes (with GB accession numbers) included in dataset A*

Species	16SrRNA	12SrRNA	12S/16S	ND2	Tmo4c4	SH3PX2	ENCI	Ptr	S7 1st Intron
Heterochromis multidens	GQ167968	GQ167842	GQ167905	GQ167779	GQ168156	GQ168219	GQ168282	GQ168031	GQ168093
Tylochromis sp.	GQ167998	GQ167872	GQ167935	GQ167809	GQ168186	GQ168249	GQ168312	GQ168060	GQ168123
Nanochromis parilus	GQ168003	GQ167877	GQ167940	GQ167814	GQ168191	GQ168254	GQ168317	GQ168065	GQ168128
Nanochromis parilus	GQ168004	GQ167878	GQ167941	GQ167815	GQ168192	GQ168255	GQ168318	GQ168066	GQ168129
Hemichromis elongatus	GQ168001	GQ167875	GQ167938	GQ167812	GQ168189	GQ168252	GQ168315	GQ168063	GQ168126
Pelmatochromis buettikoferi	GQ167972	GQ167846	GQ167909	GQ167783	GQ168160	GQ168223	GQ168286	GQ168035	GQ168097
Pelmatochromis nigrofasciatus	GQ167973	GQ167847	GQ167910	GQ167784	GQ168161	GQ168224	GQ168287	GQ168036	GQ168098
Pterochromis congicus	GQ167974	GQ167848	GQ167911	GQ167785	GQ168162	GQ168225	GQ168288	GQ168037	GQ168099
Pterochromis congicus	GQ167996	GQ167870	GQ167933	GQ167807	GQ168184	GQ168247	GQ168310	GQ168058	GQ168121
Etia nguti	GQ167966	GQ167840	GQ167903	GQ167777	GQ168154	GQ168217	GQ168280	GQ168029	GQ168091
Sarotherodon caudomarginatus	GQ167975	GQ167849	GQ167912	GQ167786	GQ168163	GQ168226	GQ168289	GQ168038	GQ168100
Sarotherodon caudomarginatus	GQ168008	GQ167882	GQ167945	GQ167819	GQ168196	GQ168259	GQ168322	GQ168070	GQ168133
Sarotherodon mvogoi	GQ168000	GQ167874	GQ167937	GQ167811	GQ168188	GQ168251	GQ168314	GQ168062	GQ168125
Oreochromis andersoni	GQ167994	GQ167868	GQ167931	GQ167805	GQ168182	GQ168245	GQ168308	GQ168056	GQ168119
Oreochromis niloticus	GQ167969	GQ167843	GQ167906	GQ167780	GQ168157	GQ168220	GQ168283	GQ168032	GQ168094
Alcolapia alcalica	GQ167970	GQ167844	GQ167907	GQ167781	GQ168158	GQ168221	GQ168284	GQ168033	GQ168095
Oreochromis tanganicae	GQ167971	GQ167845	GQ167908	GQ167782	GQ168159	GQ168222	GQ168285	GQ168034	GQ168096
Tristramella simonis	GQ168002	GQ167876	GQ167939	GQ167813	GQ168190	GQ168253	GQ168316	GQ168064	GQ168127
Iranocichla hormuzensis	GQ168019	GQ167893	GQ167956	GQ167830	GQ168207	GQ168270	GQ168333	GQ168081	GQ168144
Sarotherodon nigripinnis	GQ167976	GQ167850	GQ167913	GQ167787	GQ168164	GQ168227	GQ168290	GQ168039	GQ168101
Sarotherodon sp. aff. galilaeus	GQ167977	GQ167851	GQ167914	GQ167788	GQ168165	GQ168228	GQ168291	GQ168040	GQ168102
Stomatepia mariae	GQ167985	GQ167859	GQ167922	GQ167796	GQ168173	GQ168236	GQ168299	GQ168048	GQ168110
Gobiocichla wonderi	GQ167967	GQ167841	GQ167904	GQ167778	GQ168155	GQ168218	GQ168281	GQ168030	GQ168092
Tilapia brevimanus	GQ168017	GQ167891	GQ167954	GQ167828	GQ168205	GQ168268	GQ168331	GQ168079	GQ168142
Tilapia busumana	GQ167987	GQ167861	GQ167924	GQ167798	GQ168175	GQ168238	GQ168301	GQ168049	GQ168112
"Steatocranus" irvinei	GQ167981	GQ167855	GQ167918	GQ167792	GQ168169	GQ168232	GQ168295	GQ168044	GQ168106
"Steatocranus" irvinei	GQ167995	GQ167869	GQ167932	GQ167806	GQ168183	GQ168246	GQ168309	GQ168057	GQ168120
Tilapia joka	GQ167992	GQ167866	GQ167929	GQ167803	GQ168180	GQ168243	GQ168306	GQ168054	GQ168117
Tilapia buttikoferi	GQ167986	GQ167860	GQ167923	GQ167797	GQ168174	GQ168237	GQ168300	-	GQ168111
Tilapia tholloni	GQ167993	GQ167867	GQ167930	GQ167804	GQ168181	GQ168244	GQ168307	GQ168055	GQ168118
Tilapia cf. nyongana "Dja"	GQ168016	GQ167890	GQ167953	GQ167827	GQ168204	GQ168267	GQ168330	GQ168078	GQ168141
Tilapia sp. aff. zillii "Kisangani"	GQ168018	GQ167892	GQ167955	GQ167829	GQ168206	GQ168269	GQ168332	GQ168080	GQ168143
Tilapia zillii	GQ168025	GQ167899	GQ167962	GQ167836	GQ168213	GQ168276	GQ168339	GQ168087	GQ168150
Tilapia ap. aff. rheophila "Samou"	GQ168014	GQ167888	GQ167951	GQ167825	GQ168202	GQ168265	GQ168328	GQ168076	GQ168139
Tilapia louka	GQ168011	GQ167885	GQ167948	GQ167822	GQ168199	GQ168262	GQ168325	GQ168073	GQ168136
Tilapia discolor	GQ167990	GQ167864	GQ167927	GQ167801	GQ168178	GQ168241	GQ168304	GQ168052	GQ168115
Tilapia dageti	GQ168010	GQ167884	GQ167947	GQ167821	GQ168198	GQ168261	GQ168324	GQ168072	GQ168135
Tilapia guineensis	GQ168026	GQ167900	GQ167963	GQ167837	GQ168214	GQ168277	GQ168340	GQ168088	GQ168151
Bathybates ferox	GQ168021	GQ167895	GQ167958	GQ167832	GQ168209	GQ168272	GQ168335	GQ168083	GQ168146
Boulengerochromis microlepis	GQ168009	GQ167883	GQ167946	GQ167820	GQ168197	GQ168260	GQ168323	GQ168071	GQ168134
Eretmodus cyanostictus	GQ168020	GQ167894	GQ167957	GQ167831	GQ168208	GQ168271	GQ168334	GQ168082	GQ168145
Neolamprologus moorii	GQ167999	GQ167873	GQ167936	GQ167810	GQ168187	GQ168250	GQ168313	GQ168061	GQ168124
Neolamprologus moorii	GQ168022	GQ167896	GQ167959	GQ167833	GQ168210	GQ168273	GQ168336	GQ168084	GQ168147

Species	16SrRNA	12SrRNA	12S/16S	ND2	Tmo4c4	SH3PX2	ENCI	Ptr	S7 1st Intron
"Haplochromis" brauschi	GQ168007	GQ167881	GQ167944	GQ167818	GQ168195	GQ168258	GQ168321	GQ168069	GQ168132
Benthochromis sp.	GQ168023	GQ167897	GQ167960	GQ167834	GQ168211	GQ168274	GQ168337	GQ168085	GQ168148
Cyprichromis leptosoma	GQ168024	GQ167898	GQ167961	GQ167835	GQ168212	GQ168275	GQ168338	GQ168086	GQ168149
Chilochromis duponti	GQ167965	GQ167839	GQ167902	GQ167776	GQ168153	GQ168216	GQ168279	GQ168028	GQ168090
Tilapia bilineata "Lefini"	GQ167964	GQ167838	GQ167901	GQ167775	GQ168152	GQ168215	GQ168278	GQ168027	GQ168089
Tilapia bilineata "Salonga"	GQ168013	GQ167887	GQ167950	GQ167824	GQ168201	GQ168264	GQ168327	GQ168075	GQ168138
Tilapia ruweti	GQ167988	GQ167862	GQ167925	GQ167799	GQ168176	GQ168239	GQ168302	GQ168050	GQ168113
Tilapia sparrmanii	GQ167989	GQ167863	GQ167926	GQ167800	GQ168177	GQ168240	GQ168303	GQ168051	GQ168114
Tilapia guinasana	GQ167991	GQ167865	GQ167928	GQ167802	GQ168179	GQ168242	GQ168305	GQ168053	GQ168116
Steatocranus ubangiensis	GQ168015	GQ167889	GQ167952	GQ167826	GQ168203	GQ168266	GQ168329	GQ168077	GQ168140
Steatocranus casuarius	GQ167979	GQ167853	GQ167916	GQ167790	GQ168167	GQ168230	GQ168293	GQ168042	GQ168104
Steatocranus sp. "dwarf"	GQ167983	GQ167857	GQ167920	GQ167794	GQ168171	GQ168234	GQ168297	GQ168046	GQ168108
Steatocranus bleheri	GQ167978	GQ167852	GQ167915	GQ167789	GQ168166	GQ168229	GQ168292	GQ168041	GQ168103
Steatocranus sp. "bulky head"	GQ167982	GQ167856	GQ167919	GQ167793	GQ168170	GQ168233	GQ168296	GQ168045	GQ168107
Steatocranus sp. "redeye"	GQ167997	GQ167871	GQ167934	GQ167808	GQ168185	GQ168248	GQ168311	GQ168059	GQ168122
Steatocranus gibbiceps	GQ167980	GQ167854	GQ167917	GQ167791	GQ168168	GQ168231	GQ168294	GQ168043	GQ168105
Steatocranus glaber	GQ168005	GQ167879	GQ167942	GQ167816	GQ168193	GQ168256	GQ168319	GQ168067	GQ168130
Steatocranus tinanti	GQ167984	GQ167858	GQ167921	GQ167795	GQ168172	GQ168235	GQ168298	GQ168047	GQ168109
Steatocranus tinanti	GQ168006	GQ167880	GQ167943	GQ167817	GQ168194	GQ168257	GQ168320	GQ168068	GQ168131
Tilapia mariae	GQ168012	GQ167886	GQ167949	GQ167823	GQ168200	GQ168263	GQ168326	GQ168074	GQ168137

Table S2 *List of all taxa and accession numbers for ND2 included in dataset B*

No. §	Species	GenBank accession number
1	*Satanoperca sp.*	AB018971.2
2	*Oxylapia polli*	AF317275.1
3	*Amphilophus sp.*	AB018970.2
5	*Altolamprologus* sp. "shell"	EF191107.1
6	*Asprotilapia leptura*	AY337772.1
8	*Astatoreochromis alluaudi*	AY930075.1
11	*Aulonocara stuartgranti*	EU661720.1
12	*Aulonocranus dewindti*	AY337782.1
13	*Baileychromis centropomoides*	AY682510.1
14	*Bathybates fasciatus*	AY663733.1
15	*Bathybates ferox*	AY663737.1
16	*Bathybates graueri*	AY663726.1
17	*Bathybates hornii*	AY663735.1
18	*Bathybates leo*	AY663731.1
19	*Bathybates minor*	AY663720.1
20	*Bathybates vittatus*	AY663728.1
24	*Benthochromis melanoides*	AY682512.1
25	*Benthochromis melanoides*	AY682513.1
27	*Benthochromis cf. tricoti*	AY682514.1
31	*Boulengerochromis microlepis*	AF317229.1
32	*Boulengerochromis microlepis*	GQ167820
33	*Buccochromis heterotaenia*	EU661719.1
35	*Callochromis pleurospilus*	AY337771.1
37	*Cardiopharynx schoutedeni*	AY337791.1
38	*Cheilochromis euchilus*	AY930092.1
43	*Chilochromis duponti*	GQ167776
45	*Chromidotilapia guntheri*	AF317270.1
57	*Nyassachromis prostoma*	EU661715.1
59	*Cunningtonia longiventralis*	AY682516.1
60	*Cyathopharynx furcifer*	AY337781.1
63	*Cyprichromis cf. leptosoma* "blue"	AY740338.1
64	*Cyprichromis cf. leptosoma* "gold"	AY740344.1
65	*Cyprichromis cf. leptosoma* "jumbo"	AY740340.1
71	*Cyprichromis cf. leptosoma* "yellow"	AY740342.1
73	*Cyprichromis microlepidotus*	AY740354.1
74	*Cyprichromis pavo*	AY740373.1
75	*Cyprichromis zonatus*	AY740377.1
76	*Cyrtocara moorii*	AY930089.1
82	*Eclectochromis ornatus*	EU661717.1
83	*Ectodus descampsii*	AY337790.1
85	*Etia nguti*	GQ167777
86	*Gnathochromis permaxillaris*	AY682522.1
89	*Gobiocichla wonderi*	GQ167778
90	*Grammatotria lemairii*	AY337787.1
91	*Greenwoodochromis bellcrossi*	AY682524.1
92	*Greenwoodochromis christyi*	AY682528.1

No. §	Species	GenBank accession number
96	*Haplochromis brauschi*	AY930095.1
97	*Haplochromis burtoni*	AF317266.1
99	*Haplochromis degeni*	AY930064.1
100	*Haplochromis gracilior*	AY930078.1
101	*Haplochromis horei*	AY930100.1
103	*Haplochromis insidiae*	AY930077.1
104	*Haplochromis"oligacanthus*	AF416779.1
107	*Haplochromis phytophagus*	AY930076.1
112	*Haplochromis sauvagei*	AY930063.1
116	*Haplochromis* "sp."Kisangani"	AY930062.1
120	*Haplochromis stappersii*	AY930046.1
121	*Haplochromis stormsi*	AY930057.1
122	*Haplotaxodon microlepis*	EF437498.1
123	*Haplotaxodon trifasciatus*	EF437487.1
124	*Hemibates stenosoma*	AY663719.1
126	*Hemichromis elongatus*	AY663714.1
127	*Hemichromis sp.*	GQ167812
128	*Heterochromis multidens*	GQ167779
129	*Iranocichla hormuzensis*	AF317278.1
130	*Julidochromis ornatus*	DQ093111.1
131	*Konia dikume*	AJ845105.1
132	*Konia eisentrauti*	AJ845103.1
134	*Lamprologus callipterus*	EF191085.1
136	*Lamprologus congoensis*	AF317272.1
137	*Lamprologus kungweensis*	EF191084.1
138	*Lamprologus laparogramma*	EF191088.1
139	*Lamprologus lemairii*	AY740386.1
140	*Lamprologus meleagris*	EF191098.1
141	*Lamprologus ocellatus*	EF191115.1
142	*Lamprologus ornatipinnis*	EF191109.1
143	*Lamprologus speciosus*	EF191102.1
144	*Lepidiolamprologus attenuatus*	AY682532.1
145	*Lestradea stappersii*	AY337792.1
147	*Limnochromis abeelei*	AY682535.1
149	*Limnochromis auritus*	AY337766.1
150	*Limnochromis staneri*	AY682540.1
152	*Melanochromis auratus*	AY930069.1
153	*Metriaclima zebra*	DQ093114.1
154	*Microdontochromis rotundiventralis*	AY337793.1
155	*Microdontochromis tenuidentatus*	AY337784.1
156	*Myaka myaka*	AJ845107.1
161	*Neolamprologus brevis*	EF191094.1
163	*Neolamprologus calliurus*	DQ093112.1
164	*Neolamprologus caudopunctatus*	EF191122.1
165	*Neolamprologus christyi*	AY740389.1
166	*Neolamprologus devosi*	EF437476.1
167	*Neolamprologus fasciatus*	EF191120.1

No. §	Species	GenBank accession number
168	*Neolamprologus leloupi*	EF191103.1
169	*Neolamprologus marunguensis*	AY740390.1
170	*Neolamprologus multifasciatus*	EF191091.1
171	*Neolamprologus niger*	AY740391.1
172	*Neolamprologus nigriventris*	AY740392.1
173	*Neolamprologus olivaceous*	AY740393.1
174	*Neolamprologus palmeri*	AY740394.1
175	*Neolamprologus pulcher*	AY740395.1
176	*Neolamprologus similis*	EF191099.1
178	*Neolamprologus wauthioni*	EF191118.1
181	*Oreochromis variabilis*	AF317241.1
182	*Ophthalmotilapia boops*	AY337773.1
183	*Ophthalmotilapia nasuta*	AY337783.1
184	*Ophthalmotilapia ventralis*	AY337774.1
187	*Alcolapia alcalica*	GQ167781
188	*Oreochromis amphimelas*	AF317230.1
189	*Oreochromis andersonii*	AF317231.1
190	*Oreochromis andersonii*	GQ167805
191	*Oreochromis esculentus*	AF317232.1
192	*Oreochromis karongae*	DQ465030.1
193	*Oreochromis leucostictus*	AF317233.1
194	*Oreochromis macrochir*	AF317235.1
195	*Oreochromis mossambicus*	AF317234.1
196	*Oreochromis "niloticus"*	GQ167780
197	*Oreochromis niloticus*	AF317237.1
198	*Oreochromis niloticus*	AB018974.2
199	*Oreochromis niloticus vulcani*	AF317242.1
200	*Oreochromis schwebischi*	AF317238.1
201	*Oreochromis tanganicae*	AF317240.1
202	*Oreochromis tanganicae*	GQ167782
203	*Oreochromis urolepis*	AF317239.1
204	*Orthochromis kasuluensis*	AY930049.1
205	*Orthochromis luichensis*	AY930052.1
206	*Orthochromis malagaraziensis*	AY930054.1
207	*Orthochromis malagaraziensis*	AY930056.1
208	*Orthochromis mazimeroensis*	AY930053.1
209	*Orthochromis mosoensis*	AY930055.1
210	*Orthochromis rubrolabialis*	AY930051.1
211	*Orthochromis rugufuensis*	AY930050.1
212	*Orthochromis uvinzae*	AY930048.1
215	*Otopharynx walteri*	EU661716.1
217	*Paracyprichromis brieni*	AY740352.1
218	*Paracyprichromis brieni*	AY740378.1
219	*Paracyprichromis nigripinnis*	AY740339.1
221	*Pelmatochromis buettikoferi*	GQ167783
225	*Pelvicachromis pulcher*	AF317271.1
232	*Perissodus eccentricus*	EF437511.1

No. §	Species	GenBank accession number
233	*Perissodus microlepis*	EF437483.1
236	*Petrochromis macrognathus*	AY930068.1
237	*Petrotilapia nigra*	EU661721.1
239	*Pharyngochromis acuticeps*	EF393695.1
241	*Plecodus elaviae*	EF437503.1
242	*Plecodus multidentatus*	EF437505.1
244	*Plecodus straeleni*	EF437482.1
245	*Protomelas annectens*	EU661718.1
246	*Protomelas similis*	EU661714.1
248	*Pseudocrenilabrus multicolor*	AY930070.1
251	*Pseudocrenilabrus philander*	AY602993.1
254	*Pseudotropheus livingstonii*	AY930061.1
256	*Pseudotropheus tropheops*	AY740384.1
258	*Pterochromis congicus*	GQ167785
259	*Pterochromis congicus*	GQ167807
260	*Pungu maclareni*	AJ845101.1
261	*Reganochromis calliurus*	AY682544.1
266	*Sargochromis carlottae*	EF393683.1
268	*Sargochromis codringtonii*	EF393717.1
270	*Sargochromis giardi*	AY930098.1
272	*Sargochromis mellandi*	EF393711.1
277	*Sarotherodon caroli*	AJ845112.1
278	*Sarotherodon caudomarginatus*	AF317243.1
279	*Sarotherodon caudomarginatus*	GQ167786
280	*Sarotherodon galilaeus*	AJ845088.1
281	*Sarotherodon galilaeus multifasciatus*	AJ845087.1
282	*Sarotherodon galilaeus sanagaensis*	AJ845085.1
283	*Sarotherodon linnellii*	AJ845115.1
284	*Sarotherodon lohbergeri*	|AJ845109.1
285	*Sarotherodon melanotheron*	AF317245.1
286	*Sarotherodon mvogoi*	GQ167811
287	*Sarotherodon nigripinnis*	AJ845084.1
288	*Sarotherodon nigripinnis*	GQ167787
289	*Sarotherodon occidentalis*	AF317246.1
290	*Sarotherodon "sp."bighead"*	AJ845091.1
291	*Sarotherodon "sp."mudfeeder"*	AJ845092.1
292	*Sarotherodon steinbachi*	AJ845111.1
294	*Serranochromis altus*	EF393697.1
296	*Serranochromis angusticeps*	EF393687.1
300	*Serranochromis macrocephalus*	EF393705.1
302	*Serranochromis robustus*	EF393712.1
308	*Serranochromis stappersi*	EF393699.1
310	*Serranochromis thumbergi*	EF393704.1
311	*Simochromis diagramma*	AY930087.1
312	*Simochromis marginatus*	AY930088.1
314	*Spathodus erythrodon*	AF317267.1
315	*Steatocranus bleheri*	GQ167789

No. §	Species	GenBank accession number
316	*Steatocranus casuarius*	AF317247.1
318	*Steatocranus irvinei*	GQ167792
320	*Steatocranus* sp."redeye"	GQ167808
321	*Steatocranus tinanti*	AF317248.1
322	*Steatocranus tinanti*	GQ167795
323	*Steatocranus* sp. "bulky head"	GQ167793
324	*Stomatepia mariae*	AJ845097.1
325	*Stomatepia mariae*	GQ167796
326	*Stomatepia mongo*	AJ845095.1
327	*Stomatepia pindu*	AJ845099.1
331	*Telmatochromis vittatus*	AY740396.1
334	*Thysochromis ansorgii*	AY663713.1
335	*Thysochromis ansorgii*	AF317263.1
336	*Tilapia bilineata* "Lefini"	GQ167775
337	*Tilapia brevimanus*	AF317249.1
338	*Tilapia busumana*	AF317250.1
339	*Tilapia busumana*	GQ167798
340	*Tilapia buttikoferi*	AF317251.1
341	*Tilapia buttikoferi*	GQ167797
342	*Tilapia cabrae*	AF317252.1
343	*Tilapia cessiana*	AF317253.1
344	*Tilapia cf. rheophila*	GQ167825
345	*Tilapia coffea*	AF317254.1
346	*Tilapia dageti*	GQ167821
347	*Tilapia discolor*	AF317255.1
348	*Tilapia discolor*	GQ167801
349	*Tilapia guinasana*	GQ167802
350	*Tilapia guineensis*	AF317256.1
351	*Tilapia guineensis*	GQ167837
352	*Tilapia joka*	GQ167803
353	*Tilapia louka*	AF317257.1
354	*Tilapia louka*	GQ167822
355	*Tilapia mariae*	AF317258.1
356	*Tilapia mariae*	GQ167823
358	*Tilapia rendalli*	AF317259.1
359	*Tilapia ruweti*	GQ167799
360	*Tilapia sparrmanii*	AF317260.1
361	*Tilapia sparrmanii*	GQ167800
362	*Tilapia tholloni*	GQ167804
363	*Tilapia walteri*	AF317261.1
364	*Tilapia zillii*	AF317262.1
365	*Trematocara macrostoma*	AY663715.1
366	*Trematocara unimaculatum*	AF317268.1
368	*Tristramella simonis*	AF317276.1
369	*Tristramella simonis*	GQ167813
370	*Tropheus brichardi*	AY930086.1
371	*Tropheus duboisi*	AY930085.1

No. §	Species	GenBank accession number
372	*Tropheus moorii*	AY930066.1
373	*Tropheus moorii*	AY930067.1
375	*Tropheus polli*	AY930084.1
376	*Tylochromis leonensis*	AF317274.1
377	*Tylochromis polylepis*	AB018973.2
380	*Xenochromis hecqui*	EF437514.1
381	*Xenotilapia bathyphila*	AY337789.1
382	*Xenotilapia caudafasciata*	AY337777.1
384	*Xenotilapia flavipinnis*	AY337794.1
385	*Xenotilapia longispinis*	AY337778.1
386	*Xenotilapia melanogenys*	AY682517.1
387	*Xenotilapia ochrogenys*	AY337767.1
388	*Xenotilapia sima*	AY337785.1
389	*Xenotilapia papilio*	AY337776.1
390	*Xenotilapia spiloptera*	AY337788.1
392	*Astatoreochromis alluaudi*	EU753923.1
393	*Chetia brevicauda*	EU753924.1
394	*Chetia brevis*	EU753925.1
395	*Chetia flaviventris*	EU753926.1
398	*Haplochromis albolabris*	EU753929.1
399	*Haplochromis bloyeti*	EU753930.1
400	*Haplochromis brauschi*	EU753931.1
401	*Haplochromis burtoni*	EU753932.1
402	*Haplochromis buysi*	EU753933.1
403	*Haplochromis calliptera*	EU753934.1
404	*Haplochromis horei*	EU753935.1
405	*Orthochromis machadoi*	EU753936.1
406	*Haplochromis oligacanthus*	EU753937.1
407	*Ctenochromis pectoralis*	EU753938.1
409	*Haplochromis phytophagus*	EU753940.1
410	*Haplochromis polli*	EU753941.1
411	*Haplochromis rudolfianus*	EU753942.1
412	*Haplochromis squamipinnis*	EU753943.1
413	*Haplochromis sp*."Kanyaboli"	EU753944.1
414	*Haplochromis* sp."Fayoum"	EU753945.1
415	*Haplochromis sp. Mburo Black*	EU753946.1
416	*Nimbochromis venustus*	EU753947.1
417	*Nimbochromis livingstonii*	EU753948.1
419	*Pseudocrenilabrus* sp. "Lufubu"	EU753950.1
420	*Pseudocrenilabrus* sp."blue"	EU753951.1
421	*Pseudocrenilabrus* sp."orange"	EU753952.1
422	*Pseudocrenilabrus sp.*	EU753953.1
423	*Sargochromis coulteri*	EU753954.1
426	*Schwetzochromis neodon*	EU753957.1
427	*Serranochromis angusticeps*	EU753958.1
429	*Serranochromis stappersi*	EU753960.1
430	*Serranochromis thumbergi*	EU753961.1

No. §	Species	GenBank accession number
432	*Rhamphochromis cf. longiceps*	AF305246.1
433	*Rhamphochromis* sp."maldeco"	AF305254.1
434	*Rhamphochromis leptosoma*	AF305253.1
435	*Rhamphochromis esox*	AF305252.1
436	*Rhamphochromis* sp."big"	AF305251.1
437	*Rhamphochromis macrophthalmus*	AF305250.1
438	*Rhamphochromis* sp."brown"	AF305247.1
440	*Rhamphochromis macrophthalmus*	AF305249.1
441	*Diplotaxodon 'similis'*	AF305275.1
442	*Diplotaxodon 'similis'*	AF305274.1
447	*Diplotaxodon greenwoodi*	AF305269.1
448	*Diplotaxodon macrops*	AF305268.1
450	*Diplotaxodon 'brevimaxillaris'*	AF305265.1
452	*Diplotaxodon macrops*	AF305267.1
453	*Diplotaxodon 'holochromis'*	AF305262.1
454	*Diplotaxodon* sp."deep"	AF305263.1
457	*Cyclopharynx fwae*	AY930099.1
458	*Triglachromis otostigma*	AY682546.1
459	*Triglachromis otostigma*	AY337769.1
460	*Iranicichla hormuzensis*	GQ167830

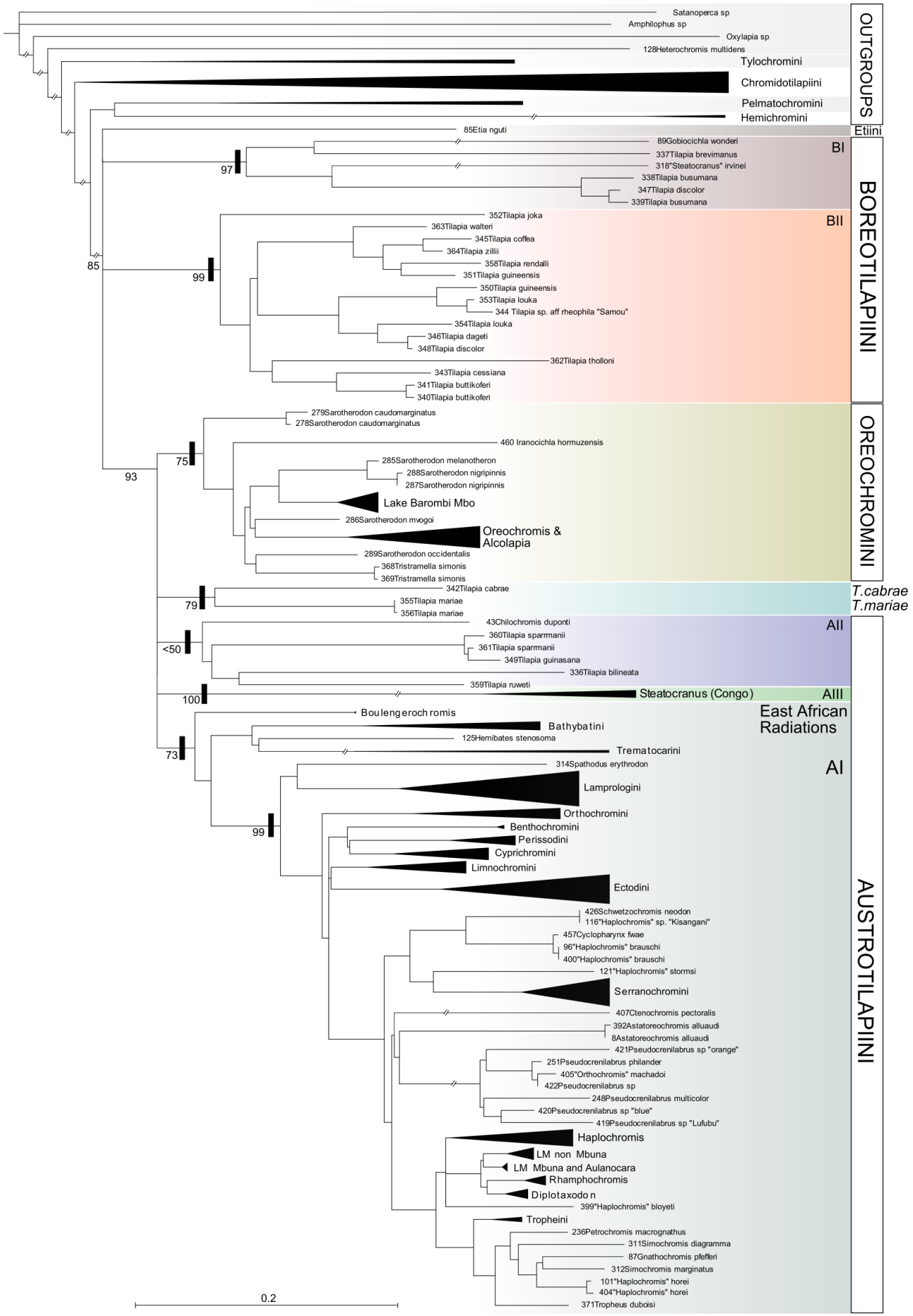
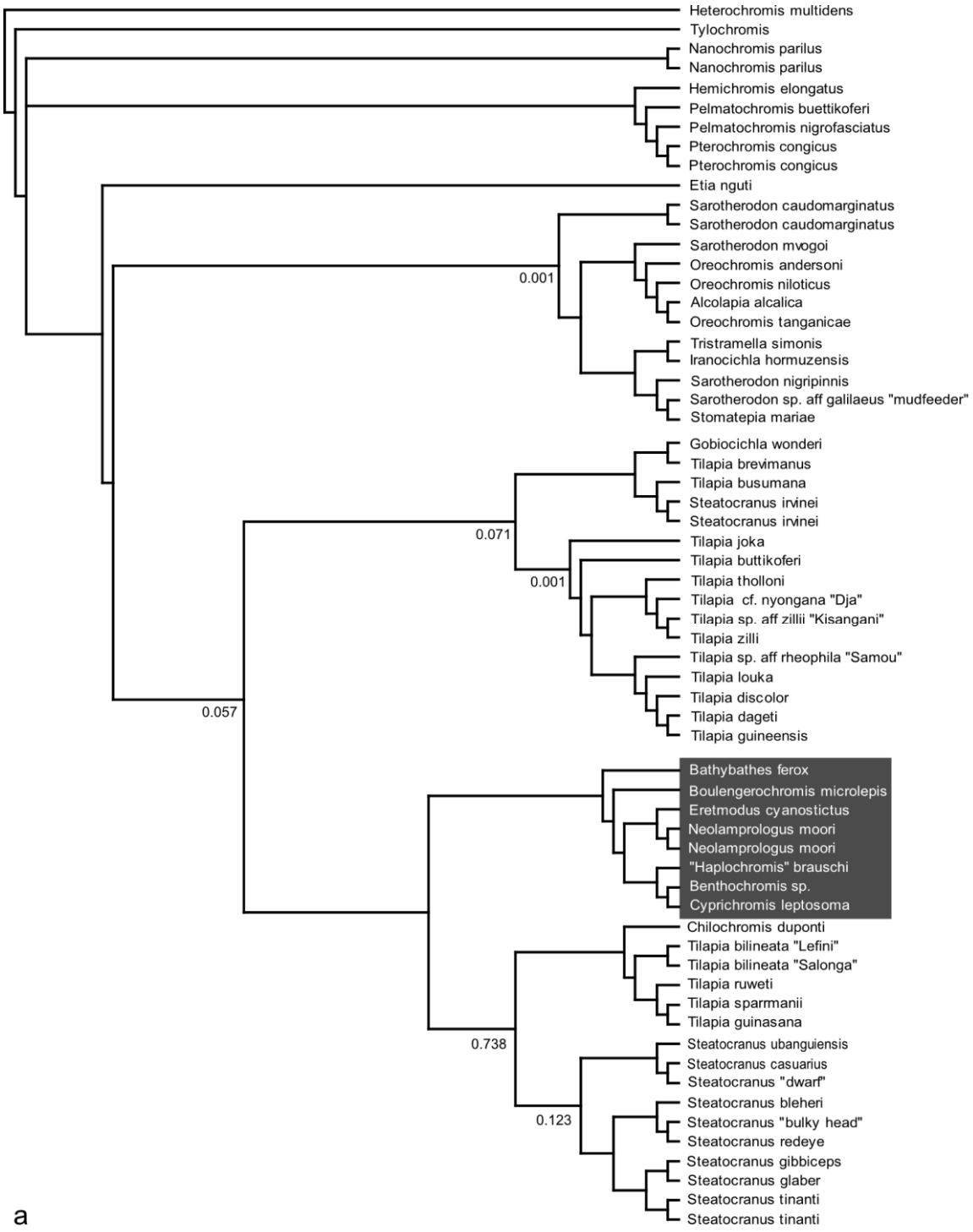


Figure S1 (caption overleaf)

Figure S1 *Maximum likelihood Phylogeny based on dataset B*

Maximum likelihood phylogeny for dataset B based on 992bp of ND2. Sequences were taken from GenBank (N=263) and additional taxa from dataset A (N=38) were supplemented. Focus clades are marked with black bars and BS support values are given only for those clades. All focus clades (well supported clades from dataset A) were recovered as monophyletic in this tree, despite lower data density and higher taxon sampling. One sequence of *Tilapia discolor* taken from GenBank is nested within *T. busumana* in clade BI instead of being sister to our conspecific and positively identified T. discolor. As no specimen vouchers of this specimen are available, we assume that either misidentification or mitochondrial introgression of sympatric *T. busumana* is the reason for this discrepancy.


Heterochromis multidens
Tylochromis
Nanochromis parilus
Nanochromis parilus
Hemichromis elongatus
Pelmatochromis buettikoferi
Pelmatochromis nigrofasciatus
Pterochromis congicus
Pterochromis congicus
Etia nguti
Sarotherodon caudomarginatus
Sarotherodon caudomarginatus
Sarotherodon mvogoi
Oreochromis andersoni
Oreochromis niloticus
Alcolapia alcalica
Oreochromis tanganicae
Tristramella simonis
Iranocichla hormuzensis
Sarotherodon nigripinnis
Sarotherodon sp. aff galilaeus "mudfeeder"
Stomatepia mariae
Gobiocichla wonderi
Tilapia brevimanus
Tilapia busumana
Steatocranus irvinei
Steatocranus irvinei
Tilapia joka
Tilapia buttikoferi
Tilapia tholloni
Tilapia cf. nyongana "Dja"
Tilapia sp. aff zillii "Kisangani"
Tilapia zilli
Tilapia sp. aff rheophila "Samou"
Tilapia louka
Tilapia discolor
Tilapia dageti
Tilapia guineensis
Bathybathes ferox
Boulengerochromis microlepis
Eretmodus cyanostictus
Neolamprologus moori
Neolamprologus moori
"Haplochromis" brauschi
Benthochromis sp.
Cyprichromis leptosoma
Chilochromis duponti
Tilapia bilineata "Lefini"
Tilapia bilineata "Salonga"
Tilapia ruweti
Tilapia sparrmanii
Tilapia guinasana
Steatocranus ubanguiensis
Steatocranus casuarius
Steatocranus "dwarf"
Steatocranus bleheri
Steatocranus "bulky head"
Steatocranus redeye
Steatocranus gibbiceps
Steatocranus glaber
Steatocranus tinanti
Steatocranus tinanti
0.001
0.071
0.001
0.057
0.738
0.123
a

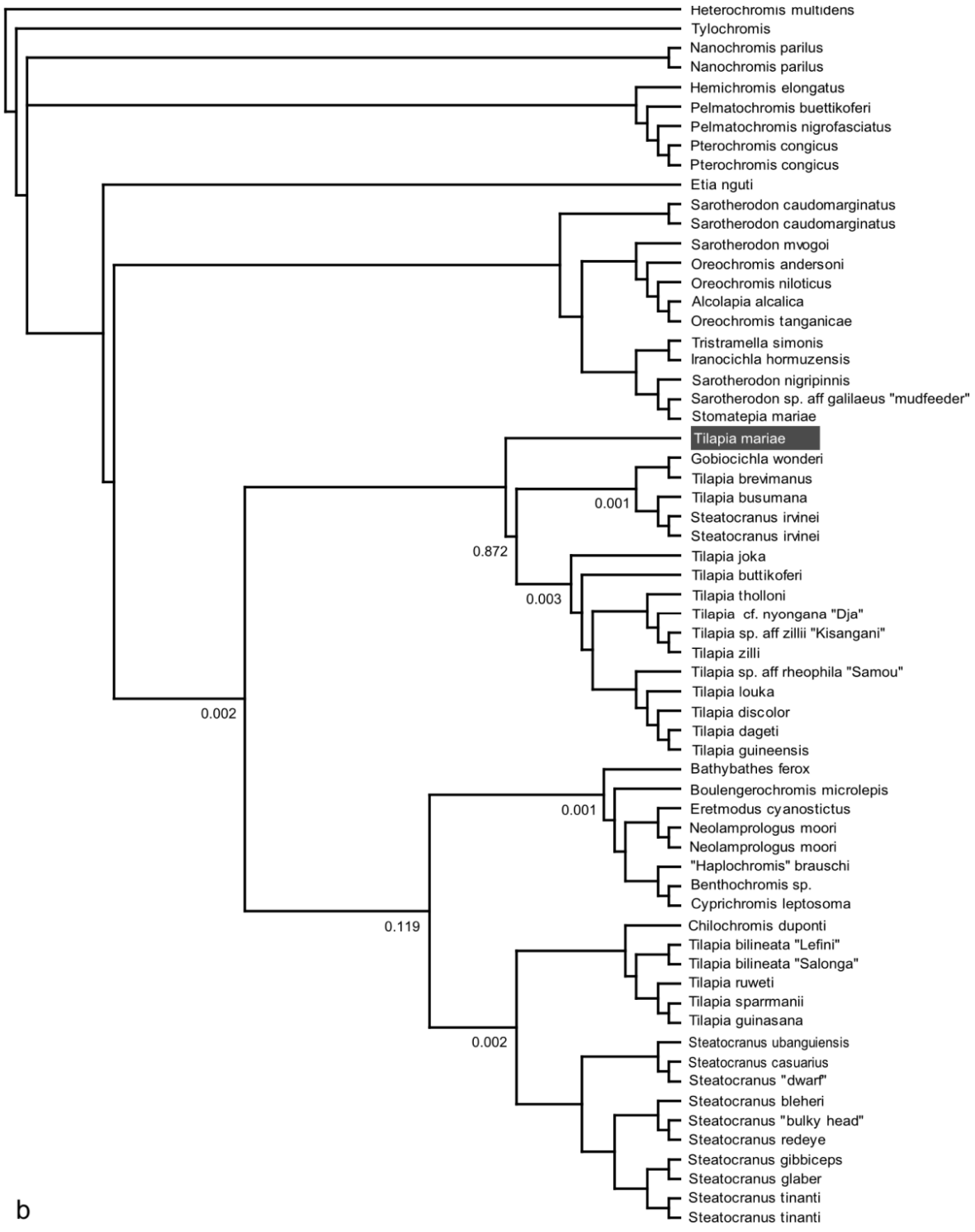


Figure S2 *Branch attachment frequencies in bootstrap replicates*

Alternative positions of the East African radiations (a) and the single unstable taxon *T. mariae* (b) in 1000 bootstrap topologies. The numbers, plotted on the ML tree, indicate fractions of bootstrap trees in which alternative branching patterns occur.

Table S3 *Results of the approximately unbiased test*

Results of the approximately unbiased (AU) test for alternative phylogenetic placements of the EAR with and without the nuclear intron S7. The topologies tested were taken from the branch attachment frequency test (1-5), were consensus topologies based on solely mitochondrial or nuclear markers (6-7) and a topology with a polytomy at the base of the austrotilapiini was tested (8).

Topology	AU test		
	complete dataset	without S7 intron	
(OUT,(ETIA,(OREO,((**AIII, AII), AI)**,(BI,BII)))))	**0.867**	**0.767**	
(OUT,(ETIA,(OREO,((**AIII, AI), AII)**,(BI,BII)))))	**0.300**	**0.548**	
(OUT,(ETIA,(OREO,**(AI,((BII,BI),(AII,AIII))**)))));	**0.236**	0.036	
(OUT,(ETIA,(OREO,**((AI,(BII,BI)),(AII,AIII))**)))	**0.148**	**0.163**	
(OUT,(ETIA,(**(AI, OREO)**,((BI,BII),(AII,AIII)))))	0.002	0.001	
(OUT,(Etia,(OREO,((((AII,AIII),AI),BI),BII))))	> 0.001	> 0.001	**mitochondrial markers**
(HET,(((ETIA,OUT),BI),(OREO,(AII,(AIII,AI)))),BII)	0.001	0.026	**nuclear markers**
(OUT,(ETIA,(OREO,(**(AIII, AI, AII)**,(BI,BII)))))	0.002	0.006	**polytomy**

Additional file 6 *Informal classification of African cichlid genera with special reference to species level taxa previously referred to as "Tilapiini"*

Subfamily Heterochrominidinae *Heterochromis*

Subfamily Pseudocrenilabrinae

Tylochromini: *Tylochromis*

Hemichromini: *Anomalochromis, Hemichromis*

Pelmatochromini: *Pelmatochromis, Pterochromis*

Chromidotilapiini: *Benitochromis*, *Chromidotilapia, Congochromis*, *Divandu*, *Limbochromis*, *Nanochromis*, *Parananochromis*, *Pelvicachromis*, *Teleogramma*, *Thysochromis*

Haplotilapiini

Etiini: *Etia nguti*

Oreochromini: *Alcolapia alcalica*, *Alcolapia grahami*, *Alcolapia latilabris, Alcolapia ndalani, Danakilia franchettii, Iranocichla hormuzensis, Konia eisentrauti, Konia dikume, Myaka myaka, Oreochromis amphimelas, Oreochromis andersonii, Oreochromis angolensis, Oreochromis aureus, Oreochromis chungruruensis, Oreochromis esculentus, Oreochromis hunteri, Oreochromis ismailiaensis, Oreochromis jipe, Oreochromis karomo, Oreochromis karongae, Oreochromis korogwe, Oreochromis lepidurus, Oreochromis leucostictus, Oreochromis lidole, Oreochromis macrochir, Oreochromis malagarsi, Oreochromis mortimeri, Oreochromis mossambicus, Oreochromis mweruensis, Oreochromis niloticus, Oreochromis pangani, Oreochromis placidus, Oreochromis rukwaensis, Oreochromis saka, Oreochromis salinicola, Oreochromis schwebischi, Oreochromis shiranus, Oreochromis spilurus, Oreochromis squamipinnis, Oreochromis tanganicae,*

Oreochromis upembae, Oreochromis urolepis, Oreochromis variabilis,

Pungu maclareni, Sarotherodon steinbachi, Sarotherodon linnellii, Sarotherodon caroli, Sarotherodon lohbergeri, Sarotherodon caudomarginatus, Sarotherodon mvogoi, Sarotherodon occidentalis, Sarotherodon melanotheron, Sarotherodon nigripinnis, Sarotherodon galilaeus, Sarotherodon tournieri, Stomatepia pindu, Stomatepia mariae, Stomatepia mongo, Tristramella intermedia, Tristramella sacra, Tristramella simonis,

Austrotilapiini:

Clade AI (East African Radiation):

Boulengerochromini: *Boulengerochromis microlepis*

Hemibatini: *Hemibates*

Bathybatini: *Bathybates*

Trematocarini: *Trematocara*

Eretmodini: *Eretmodus, Tanganicodus, Spathodus*

Lamprologini: *Altolamprologus*, *Chalinochromis*, *Julidochromis*, *Lamprologus*, *Lepidiolamprologus*, *Neolamprologus (incl. Variabilichromis)*, *Telmatochromis*

Orthochromini: *Orthochromis*

Ectodini: *Asprotilapia, Aulonocranus, Callochromis, Cardiopharynx, Cunningtonia, Cyathopharynx, Ectodus, Grammatotria, Lestradea, Microdontochromis, Opthalmotilapia*

Cyprichromini: *Cyprichromis*, *Paracyprichromis*

Perissodini: *Haplotaxodon*, *Perissodus (incl. Plecodus,Xenochromis)*

Limnochromini: *Baileychromis, Gnathochromis, Greenwoodochromis, Limnochromis, Reganochromis, Trematochromis benthicola*

Benthochromini: *Benthochromis*

Cyphotilapiini: *Cyphotilapia*

Haplochromini:

Tropheini: *"Ctenochromis" horei*, *«Gnathochromis» pfefferi, Limnotilapia, Lobochilotes, Petrochromis, Pseudosimochromis, Simochromis, Tropheus*

Serranochromini: *Chetia, Pharyngochromis, Sargochromis, Serranochromis, «Thoracochromis albolabris», «Thoracochromis» buysi*

Lake Malawi clade: all endemic Malawi genera

***Pseudocrenilabrus*-clade:** *Pseudocrenilabrus* (incl. *"Orthochromis" machadoi*)

***Ctenochromis* clade:** *Ctenochromis pectoralis*

Astatoreochromis clade: *Astatoreochromis*

Lake Victoria Superflock: all «*Haplochromis*»

"Paraphyletic Rest": *Cyclopharynx*, "*Haplochromis" (incl. Astatotilapia, "Ctenochromis" polli*, *"Ctenochromis"oligacanthus*, *Rheohaplochormis*, *Thoracochromis*, *"Schwetzochromis" polyacanthus*, *"S." stormsi*), *Schwetzochromis neodon*

Clade AII: *Chilochromis duponti, Tilapia baloni, Tilapia bilineata, T. guinasana, Tilapia sparrmanii, Tilapia ruweti*

Clade AIII: *Steatocranus* (except *"Steatocranus" irvinei*)

Boreotilapiini

Clade BI: *Gobiocichla ethewynnae, Gobiocichla wonderi*, *"Steatocranus" irvinei, Tilapia busumana*, *T. brevimanus*

Clade BII: *Tilapia. bakossiorum, Tilapia. bemini*, *Tilapia. buttikoferi*, *Tilapia. bythobates*, *Tilapia. cameronensis*, *Tilapia camerunensis*, *Tilapia cessiana*, *Tilapia coffea*, *Tilapia congica, Tilapia dageti, Tilapia deckerti, Tilapia discolor, Tilapia flava, Tilapia guineensis, Tilapia gutturosa, Tilapia imbriferna, Tilapia joka, Tilapia. kottae, Tilapia louka, Tilapia margaritacea, Tilapia nyongana, Tilapia rendalli, Tilapia spongotroktis, Tilapia tholloni, Tilapia thysi, Tilapia walteri, Tilapia zillii*

Clade C: *Tilapia cabrae*, *Tilapia mariae*

Incertae sedis: *Tilapia rheophila*

Disclaimer: This list is not to be considered as published in the sense of the International Code of Zoological Nomenclature, and statements made herein are not made available for nomenclatural purposes from this document

Additional file 7 *Supporting methodical information*

The following primers were used for amplification and sequencing: partial mitochondrial 12S and 16S genes using primers L1091 and H1478 (Kocher et al. 1989) and 16Sar-L and 16Sbr-H (Palumbi et al. 1991), the connecting part between the above mentioned fragments using primers fish12F1 and fish16SR1 as well as the internal primer fish 12SF2 for sequencing (Ruber et al. 2003) and ND2 using primers ND2Met and ND2Trp (Kocher et al. 1995). Additionally, four nuclear protein coding genes

(ENCI: primers ENCI_F85 and ENCI_R982, Ptr: primers Ptr_F458 and Ptr_R1248, Sh3px3: primers SH3PX3_F461 and SH3PX3_R1303 (Li et al. 2007) and Tmo4c4 (Streelmann et al. 1998) and the first intron of the ribosomal protein coding gene S7, using primers S7RPEX1F50 and S7RPEX2R50 (Chow & Hazama 1998), were amplified and sequenced. Amplifications were performed in 10 µl volumes containing 5 µl Multiplex Mix (Qiagen),

genomic DNA 1 µl, 0.8 µl of each Primer (2,5nmol), Q-Solution (Qiagen) and water. Amplifications of all fragments were carried out in 40 cycles according to the temperature profile: 15 min at 95 °C (initial denaturation), 30 s at 95 °C, 30 s at 55-60 °C, 60 - 90 s at 72 °C, and finally 10 min at 72 °C. PCR products were purified with ExoSAP-IT (USB) and diluted with 10 µl - 20 µl HPLC water, depending on roduct strength. Sequencing was performed according to standard methods, using Big Dye 3.1. (Applied Biosystems). DNA sequences were read using an ABI 3130xl DNA sequencer (Applied Biosystems).

Sequence data assemblage and reconstruction of alignments

Chromatograms were assembled using SeqMan v. 4.03 included in the Lasergene software package (DNASTAR) and proof read manually. Alignments were conducted using the Clustal W algorithm implemented in BioEdit v. 7.0.4.1 for coding genes and MUSLE v. 3.6 for non-coding genes and rRNA. Coding genes were translated into amino acid sequences to check for stop-codons or frame shifts and datasets were checked separately for saturation at each codon position. The alignments were checked for ambiguous positions using ALISCORE v. 0.2 under default settings (Misof & Misof 2009). ALISCORE checks for random similarity of sequences using MCMC and a sliding windows approach. Based on this similarity profiles based on pairwise comparisons of sequences were calculated. Ambiguous positions were summarized in a consensus profile along the alignment (Misof & Misof 2009) and subsequently removed from all analyses. Base frequencies were equal for all markers (Chi-square tests, df = 183, all $p > 0.9$). The list of genes and the information on missing data is given in Table S1.

The stabilities of taxa were assessed with leaf stabilities, as calculated by Phyutility v. 2.2 (Smith & Dunn 2008) (available at http://code.google.com/p/phyutility/).

Cross check of the topology in a bigger phylogenetic context

To verify the consistency of the major clades supported by our multilocus dataset, a large dataset (dataset B, Table S2) based on NADH dehydrogenase subunit 2 (ND2) consisting of 263 sequences from Genbank and 38 sequences from dataset A was conducted (see Table S2). This additional dataset possesses fully representative coverage for all major African cichlid groups, which are not present in data set A. ND2 was chosen, as this marker was available for a mayor part of African cichlids via Genbank. The data were aligned using the ClustalW algorithm implemented in Bioedit. The third Codon position was saturated between in- and outgroups, however as the focus of this analysis is the identification of

terminal clades (younger splits), they were not excluded. The data were partitioned (in 1st, 2nd and 3rd Codon position) and all parameters were estimated separately. A ML phylogeny was constructed with RAXML v. 7.0.3 using the fast rapid hill climbing bootstrap algorithm with 1000 replicates and following ML search. Branches not supported by > 50% bootstrap value were collapsed. All major clades of the analysis of dataset A were recovered using ND2 and a larger taxon sampling (Fig. S1).

Topology testing

Based on 1000 bootstrap and 2000 randomly chosen BI topologies branch attachment frequencies were calculated for the unstable taxon *Tilapia mariae* and the EAR using Phyutility v. 2.2 (Smith & Dunn 2008). Furthermore, statistical significance of likelihood differences between the best topology in which the EAR is nested within austrotilapiines (Figure 1) and alternative topologies was tested using the approximately unbiased test (AU test) (Shimodaira 2002) implemented in the program Consel (Shimodaira & Hasegawa 2001). Eight topologies were tested, including all alternative topologies obtained via the branch attachment frequency test as well as topologies constructed solely based on the mitochondrial or nuclear datasets. The results are given in Table S3. A bootstrap homplasy excess test was conducted. Bootstrap values for austrotilapiines increased by excluding *T. mariae.* The effect was clearly higher for focus taxon than for all other taxa iteratively excluded during the analysis (Figure S1).

Choice of priors for the age estimation

Using different approaches for dating cichlid divergence allow for a fairly exact placement within geological time periods, but can hardly provide precise values, due to a lack of adequate calibration points in the cichlid fossil record. Genner et al. (2007) highlighted a bias of divergence estimates towards younger ages using cichlid fossils compared to geological time constraints based on Gondwana fragmentation. Age estimates based on cichlid fossils were half as young as those based on Gondwana calibrations (Genner et al. 2007). However, constraining solely the root age might result in extremely high confidence intervals (Renner & Zhang 2004). An alternative approach is to use younger geological time constraints, e.g. the lake ages, assuming that divergence of endemic clades took place after the formation of lake basins (e.g. Salzburger et al. 2005, Koblmüller et al. 2008a). These approaches resulted in heterogeneous age estimates for the origin of the EAR, ranging from 5 to > 35 mya (Genner et al. 2007, Koblmüller et al. 2008b). Recently, however, molecular clock estimates

based on non-cichlid teleost fossils resulted in plausible and tighter time intervals for basal cichlid nodes and provide a novel source for calibration points in cichlids (Azuma et al. 2008). This study is based on these published time intervals, and only one cichlid fossil to calibrate a terminal node. *Oreochromis lorenzoi†* (Carnevale et al. 2003) is one of the few reliable cichlid fossils for calibration as Holo- and Paratypes are in a well preserved state and all key traits necessary for species identification are recognizable. Its phylogenetic placement within the African cichlid phylogeny is less ambiguous than for other fossils, as the oreochromines are a clearly monophyletic group (Fig. 1). Unfortunately this is not the case for most other African cichlid fossils, which often lack diagnostic characters necessary for a precise assignment to cichlid tribes. For example, the oldest cichlid fossil known to date is *Mahengechromis* from Tanzania dated at about 46 mya (Murray 2001, Murray 2000). Character states of key traits of this fossil are heterogeneous and a clear assignment to a cichlid tribe is not unambiguously possible (Murray 2001a, Murray 2001b). Another cichlid fossil is a specimen described as cf. *Tylochromis?* (sic) from the Jebel Qatrani Formation, Fayum, in Egypt, dated at late Eocene/early Oligocene (Murray 2004, Murray 2002). From this specimen only the pharyngeal jaw and teeth were preserved and the species determination was based on this. Here we follow a conservative approach using only one unambiguous fossil and test two alternative placements of this fossil at two slightly different nodes.

Influence of different priors

All BEAST runs were conducted several times with different sets of constraints to evaluate the influence of different calibration points. As expected inclusion of the fossil calibration point lead to, slightly younger but also narrower confidence intervals for all ages (Figure S4). Two alternative placements of *Oreochromis lorenzoi†* within the topology resulted in slightly different age estimates, with younger ages when the calibration point was set at the root of all oreochromines. Using the penalized likelihood approach no difference in age estimates was observed for different positioning of *Oreochromis lorenzoi†*. Overall, age estimates largely overlap independent from priors used (Fig. S3).

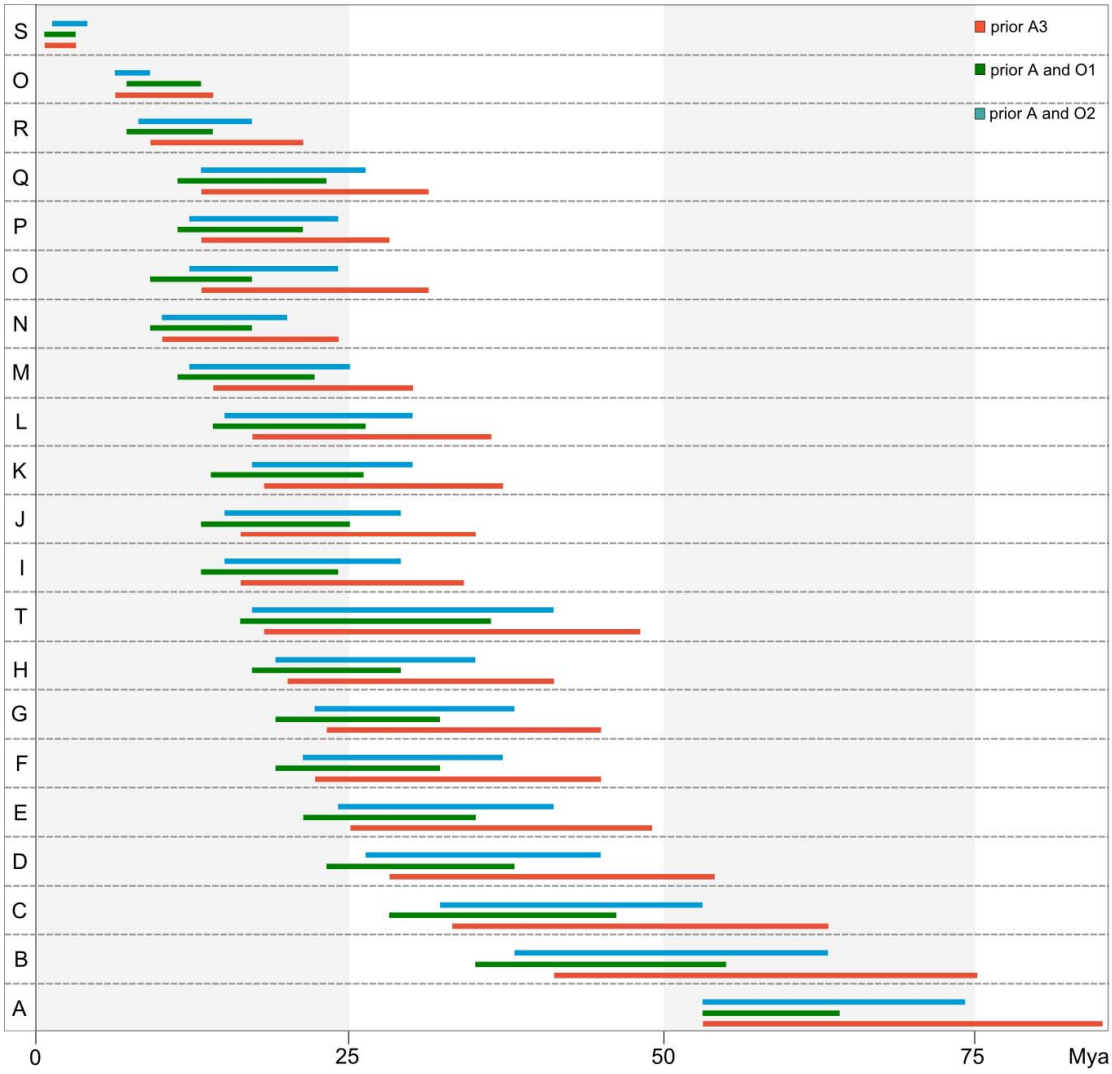


Figure S3 *Prior influence*

The effects of different age constraints on the estimation of divergence times using BEAST. Bars indicate age ranges (95 % credibility intervals) of different BEAST runs using either one single prior on the root (A3: 53-89 mya, based on published time intervals from Azuma et al. 2008) or two priors, including the *Oreochromis lorenzoi* fossil (lower bound 5.98 mya) at two possible positions (O1 and O2) in the phylogeny (Figure 2). Using solely the root prior increases credibility intervals and renders the whole age estimation older. Inclusion of the fossil prior shifts intervals to a younger age. Large overlaps in estimates unite all three results and increase the plausibility of the presented results. Alternative positions of the *Oreochromis lorenzoi* prior had no effect in age estimates using penalized likelihood (R8s).

References Appendix

Azuma, Y., Kumazawa, Y., Miya, M., Mabuchi, K., and Nishida, M. 2008 Mitogenomic evaluation of the historical biogeography of cichlids toward reliable dating of teleostean divergences. *BMC Evolutionary Biology* 8.

Carnevale, G., Sorbini, C., and Landini, W. T. 2003 Oreochromis lorenzoi, a new species of tilapiine cichlid from the late Miocene of central Italy. *Journal of Vertebrate Paleontology* 23, 508-516.

Chow S. and Hazama K. 1998 Universal PCR primers for S7 ribosomal protein gene introns in fish. *Molecular Ecology* 7, 1255-1256.

Genner, M. J., Seehausen, O., Lunt, D. H., Joyce, D. A., Shaw, P. W., Carvalho, G. R., and Turner, G. F. 2007 Age of cichlids: New dates for ancient lake fish radiations. *Molecular Biology and Evolution* 24, 1269-1282.

Koblmüller, S., Schliewen, U. K., Duftner, N., Sefc, K. M., Katongo, C., and Sturmbauer, C. 2008a Age and spread of the haplochromine cichlid fishes in Africa. *Molecular Phylogenetics and Evolution* 49, 153-169.

Koblmüller, S., Sefc, K. M., and Sturmbauer, C. 2008b The Lake Tanganyika cichlid species assemblage: recent advances in molecular phylogenetics. *Hydrobiologia* 615, 5-20.

Kocher T.D. et al. 1989 Dynamics of mitochondrial DNA evolution in animals - amplification and sequencing with conserved primers. *Proceedings of the National Academy of Sciences of the United States of America* 86, 6196-6200.

Kocher, T. D., Conroy, J. A., Mckaye, K. R., Stauffer, J. R., and Lockwood, S. F. 1995 Evolution of NADH dehydrogenase subunit 2 in east African cichlid fish. *Molecular Phylogenetics and Evolution* 4, 420-432.

Li, C. H., Lu, G. Q., and Orti, G. 2008 Optimal data partitioning and a test case for ray-finned fishes (Actinopterygii) based on ten nuclear loci. *Systematic Biology* 57, 519-539.

Misof, B. and Misof, K. 2009 A Monte Carlo approach successfully identifies randomness in multiple sequence alignments: a more objective means of data exclusion. *Systematic Biology* 58, 21-34.

Murray A.M. 2000 Eocene cichlid fishes from Tanzania, East Africa. *Journal of Vertebrate Paleontology* 20, 651-664.

Murray A.M. 2001a The oldest fossil cichlids (Teleostei : Perciformes): indication of a 45 million-year-old species flock. *Proceedings of the Royal Society of London Series B Biological Sciences* 268, 679-684.

Murray A.M. 2001b The fossil record and biogeography of the Cichlidae (Actinopterygii : Labroidei). *Biological Journal of the Linnean Society London* 74, 517-532.

Murray A.M. 2002 Lower pharyngeal jaw of a cichlid fish (Actinopterygii; Labroidei) from an early Oligocene site in the Fayum, Egypt. *Journal of Vertebrate Paleontology* 22, 453-455.

Murray A.M. 2004 Late Eocene and early Oligocene teleost and associated ichthyofauna of the Jebel Qatrani Formation, Fayum, Egypt. *Palaeontology* 47, 711-724.

Palumbi S. et al. (1991) *The simple fool´s guide to PCR*. University of Hawaii Press, Honolulu.

Renner S.S. and Zhang L.B. 2004 Biogeography of the *Pistia* clade (Araceae): Based on chloroplast and mitochondrial DNA sequences and Bayesian divergence time inference. *Systematic Biology* 53, 422-432.

Ruber L., Van Tassell J.L., Zardoya R. 2003 Rapid speciation and ecological divergence in the American seven-spined gobies (Gobiidae, Gobiosomatini) inferred from a molecular phylogeny. *Evolution* 57, 1584-1598.

Salzburger, W., Mack, T., Verheyen, E., and Meyer, A. 2005 Out of Tanganyika: Genesis, explosive speciation, key-innovations and phylogeography of the haplochromine cichlid fishes. *BMC Evolutionary Biology* 5, 1-15.

Shimodaira H, Hasegawa M 2001 CONSEL: for assessing the confidence of phylogenetic tree selection. *Bioinformatics* **17**:1246-1247.

Shimodaira H. 2002 An approximately unbiased test of phylogenetic tree selection. *Systematic Bioogy* 51, 492-508.

Smith, S. A. and Dunn, C. W. 2008 Phyutility: a phyloinformatics tool for trees, alignments and molecular data. *Bioinformatics* 24, 715-716.

Streelman J.T., Zardoya R., Meyer A., Karl S.A. 1998 Multilocus phylogeny of cichlid fishes (Pisces : Perciformes): Evolutionary comparison of microsatellite and single-copy nuclear loci. *Molecular Biology and Evololution* 15, 798-808.

Chapter 3: Phylogenetic relationships of chromidotilapines (Teleostei: Cichlidae) with emphasis on the genus Teleogramma

Table S1 *List of all taxa and genes (with GB accession numbers) included in the dataset*

Specimen information			nc-marker				mt- marker	
tribus	**genus**	**species**	**ENC1**	**Ptr**	**SH3Px3**	**S7**	**ND2**	**16S**
Heterochromini	*Heterochromis*	*multidens*	GQ168282	GQ168031	GQ168219	GQ168093	GQ167779	GQ167968
Tylochromini	*Tylochromis*	sp.	GQ168312	GQ168060	GQ168249	GQ168123	GQ167809	GQ167998
hemichromines	*Hemichromis*	*elongatus*	GQ168315	GQ168063	GQ168252	GQ168126	GQ167812	GQ168001
pelmatochromines	*Pelmatochromis*	*buettikoferi*	GQ168286	GQ168035	GQ168223	GQ168097	GQ167783	GQ167972
	Pelmatochromis	*nigrofasciatus*	GQ168287	GQ168036	GQ168224	GQ168098	GQ167784	GQ167973
	Pterochromis	*congicus*	GQ168288	GQ168037	GQ168225	GQ168099	GQ167785	GQ167974
CHROMIDOTILAPIINES	*Thysochromis*	*ansorgii*	subm. to NCBI	subm. to NCBI	subm. to NCBI	subm. to NCBI	subm. to NCBI	subm. to NCBI
	Benitochromis	sp. aff. finleyi	subm. to NCBI	subm. to NCBI	subm. to NCBI	subm. to NCBI	subm. to NCBI	subm. to NCBI
	Chromidotilapia	*guntheri*	subm. to NCBI	subm. to NCBI	subm. to NCBI	subm. to NCBI	subm. to NCBI	subm. to NCBI
	Chromidotilapia	*schoutedeni*	subm. to NCBI	subm. to NCBI	subm. to NCBI	subm. to NCBI	subm. to NCBI	subm. to NCBI
	Congochromis	*dimidiatus* "Lokoro"	subm. to NCBI	subm. to NCBI	subm. to NCBI	subm. to NCBI	subm. to NCBI	subm. to NCBI
	Congochromis	*dimidiatus* "Luilaka"	subm. to NCBI	subm. to NCBI	subm. to NCBI	subm. to NCBI	subm. to NCBI	subm. to NCBI
	Congochromis	*sabinae* "Bloody Mary"	subm. to NCBI	subm. to NCBI	subm. to NCBI	subm. to NCBI	subm. to NCBI	subm. to NCBI
	Congochromis	*sabinae* "Lompole"	subm. to NCBI	subm. to NCBI	subm. to NCBI	subm. to NCBI	subm. to NCBI	subm. to NCBI
	Congochromis	*sabinae* "LuiKotale"	subm. to NCBI	subm. to NCBI	subm. to NCBI	subm. to NCBI	subm. to NCBI	subm. to NCBI
	Congochromis	*sabinae* "Yambula"	subm. to NCBI	subm. to NCBI	subm. to NCBI	subm. to NCBI	subm. to NCBI	subm. to NCBI
	Congochromis	sp. "Amba shortsnout"	subm. to NCBI	subm. to NCBI	subm. to NCBI	subm. to NCBI	subm. to NCBI	subm. to NCBI
	Congochromis	*squamiceps*	subm. to NCBI	subm. to NCBI	subm. to NCBI	subm. to NCBI	subm. to NCBI	subm. to NCBI
	Congochromis	*squamiceps* "Lilanda"	subm. to NCBI	subm. to NCBI	subm. to NCBI	subm. to NCBI	subm. to NCBI	subm. to NCBI
	Divandu	*albimarginatus*	subm. to NCBI	subm. to NCBI	subm. to NCBI	subm. to NCBI	subm. to NCBI	subm. to NCBI
	Limbochromis	*robertsi*	subm. to NCBI	subm. to NCBI	subm. to NCBI	subm. to NCBI	subm. to NCBI	subm. to NCBI
	Nanochromis	*consortus*	subm. to NCBI	subm. to NCBI	subm. to NCBI	subm. to NCBI	subm. to NCBI	subm. to NCBI
	Nanochromis	*minor*	subm. to NCBI	subm. to NCBI	subm. to NCBI	subm. to NCBI	subm. to NCBI	subm. to NCBI

Specimen information			nc-marker				mt- marker	
tribus	genus	species	ENC1	Ptr	SH3Px3	S7	ND2	16S
	Nanochromis	*nudiceps*	subm. to NCBI	subm. to NCBI	subm. to NCBI	subm. to NCBI	subm. to NCBI	subm. to NCBI
	Nanochromis	*parilus* "Bulu"	subm. to NCBI	subm. to NCBI	subm. to NCBI	subm. to NCBI	subm. to NCBI	subm. to NCBI
	Nanochromis	*parilus* "Foulakari"	subm. to NCBI	subm. to NCBI	subm. to NCBI	subm. to NCBI	subm. to NCBI	subm. to NCBI
	Nanochromis	*parilus* "Kinsuka"	subm. to NCBI	subm. to NCBI	subm. to NCBI	subm. to NCBI	subm. to NCBI	subm. to NCBI
	Nanochromis	sp. "Ndongo"	subm. to NCBI	subm. to NCBI	subm. to NCBI	subm. to NCBI	subm. to NCBI	subm. to NCBI
	Nanochromis	*splendens* "Bulu"	subm. to NCBI	subm. to NCBI	subm. to NCBI	subm. to NCBI	subm. to NCBI	subm. to NCBI
	Nanochromis	*splendens* "Kinganga"	subm. to NCBI	subm. to NCBI	subm. to NCBI	subm. to NCBI	subm. to NCBI	subm. to NCBI
	Nanochromis	*splendens* "Luozi"	subm. to NCBI	subm. to NCBI	subm. to NCBI	subm. to NCBI	subm. to NCBI	subm. to NCBI
	Nanochromis	*teugelsi*	subm. to NCBI	subm. to NCBI	subm. to NCBI	subm. to NCBI	subm. to NCBI	subm. to NCBI
	Nanochromis	*transvestitus*	subm. to NCBI	subm. to NCBI	subm. to NCBI	subm. to NCBI	subm. to NCBI	subm. to NCBI
	Nanochromis	*wickleri*	subm. to NCBI	subm. to NCBI	subm. to NCBI	subm. to NCBI	subm. to NCBI	subm. to NCBI
	Parananochromis	*brevirostris*	subm. to NCBI	subm. to NCBI	subm. to NCBI	subm. to NCBI	subm. to NCBI	subm. to NCBI
	Parananochromis	*longirostris*	subm. to NCBI	subm. to NCBI	subm. to NCBI	subm. to NCBI	subm. to NCBI	subm. to NCBI
	Parananochromis	*ornatus*	subm. to NCBI	subm. to NCBI	subm. to NCBI	subm. to NCBI	subm. to NCBI	subm. to NCBI
	Parananochromis	sp. "Amba longsnout"	subm. to NCBI	subm. to NCBI	subm. to NCBI	subm. to NCBI	subm. to NCBI	subm. to NCBI
	Parananochromis	sp. "Lefini"	subm. to NCBI	subm. to NCBI	subm. to NCBI	subm. to NCBI	subm. to NCBI	subm. to NCBI
	Pelvicachromis	*humilis* "Dinkaya"	subm. to NCBI	subm. to NCBI	subm. to NCBI	subm. to NCBI	subm. to NCBI	subm. to NCBI
	Pelvicachromis	*pulcher* "red"	subm. to NCBI	subm. to NCBI	subm. to NCBI	subm. to NCBI	subm. to NCBI	subm. to NCBI
	Teleogramma	depressa?	subm. to NCBI	subm. to NCBI	subm. to NCBI	subm. to NCBI	subm. to NCBI	subm. to NCBI
HAPLOTILAPIINES	*Etia*	*nguti*	GQ168280	GQ168029	GQ168217	GQ168091	GQ167777	GQ167966
	Sarotherodon	*caudomarginatus*	GQ168289	GQ168038	GQ168226	GQ168100	GQ167786	GQ167975
	Sarotherodon	*mvogoi*	GQ168314	GQ168062	GQ168251	GQ168125	GQ167811	GQ168000
	Oreochromis	*andersoni*	GQ168308	GQ168056	GQ168245	GQ168119	GQ167805	GQ167994
	Oreochromis	*niloticus*	GQ168283	GQ168032	GQ168220	GQ168094	GQ167780	GQ167969
	Iranocichla	*hormuzensis*	GQ168333	GQ168081	GQ168270	GQ168144	GQ167830	GQ168019
	Stomatepia	*mariae*	GQ168299	GQ168048	GQ168236	GQ168110	GQ167796	GQ167985
	Tilapia	*joka*	GQ168306	GQ168054	GQ168243	GQ168117	GQ167803	GQ167992
	Tilapia	*discolor*	GQ168304	GQ168052	GQ168241	GQ168115	GQ167801	GQ167990
	Gobiocichla	*wonderi*	GQ168281	GQ168030	GQ168218	GQ168092	GQ167778	GQ167967
	"Steatocranus"	*irvinei*	GQ168295	GQ168044	GQ168232	GQ168106	GQ167792	GQ167981
	Bathybates	*ferox*	GQ168335	GQ168083	GQ168272	GQ168146	GQ167832	GQ168021
	Neolamprologus	*moorii*	GQ168313	GQ168061	GQ168250	GQ168124	GQ167810	GQ167999

Specimen information			nc-marker				mt- marker	
tribus	**genus**	**species**	**ENC1**	**Ptr**	**SH3Px3**	**S7**	**ND2**	**16S**
	Cyprichromis	*leptosoma*	GQ168338	GQ168086	GQ168275	GQ168149	GQ167835	GQ168024
	Chilochromis	*duponti*	GQ168279	GQ168028	GQ168216	GQ168090	GQ167776	GQ167965
	Tilapia	*bilineata "Salonga"*	GQ168327	GQ168075	GQ168264	GQ168138	GQ167824	GQ168013
	Tilapia	*guinasana*	GQ168305	GQ168053	GQ168242	GQ168116	GQ167802	GQ167991
	Steatocranus	*ubanguiensis*	GQ168329	GQ168077	GQ168266	GQ168140	GQ167826	GQ168015
	Steatocranus	*casuarius*	GQ168293	GQ168042	GQ168230	GQ168104	GQ167790	GQ167979
	Steatocranus	sp. "dwarf"	GQ168297	GQ168046	GQ168234	GQ168108	GQ167794	GQ167983
	Steatocranus	*bleheri*	GQ168292	GQ168041	GQ168229	GQ168103	GQ167789	GQ167978
	Steatocranus	sp. "bulky head"	GQ168296	GQ168045	GQ168233	GQ168107	GQ167793	GQ167982
	Steatocranus	sp. "red eye"	GQ168311	GQ168059	GQ168248	GQ168122	GQ167808	GQ167997
	Steatocranus	cf. gibbiceps	GQ168294	GQ168043	GQ168231	GQ168105	GQ167791	GQ167980
	Steatocranus	*glaber*	GQ168319	GQ168067	GQ168256	GQ168130	GQ167816	GQ168005
	Steatocranus	*tinanti*	GQ168320	GQ168068	GQ168257	GQ168131	GQ167817	GQ168006
	Steatocranus	*bleheri*	JF961622	JF961655	JF961688	JF961721	JF961464	JF961589
	Steatocranus	sp. "Maluku"	JF961628	JF961661	JF961694	JF961727	JF961502	JF961595
	Steatocranus	sp. aff. *casuarius* "brown pearl"	JF961627	JF961660	JF961693	JF961726	-	JF961594
	Steatocranus	sp. "Lefini"	JF961625	JF961658	JF961691	JF961724	JF961470	JF961592
	Steatocranus	sp. "Nki"	JF961626	JF961659	JF961692	JF961725	JF961471	JF961593
	Steatocranus	sp. "Kisangani"	JF961624	JF961657	JF961690	JF961723	JF961469	JF961591
	Steatocranus	sp. bulky head	JF961623	JF961656	JF961689	JF961722	-	JF961590
	Steatocranus	sp. aff. *tinanti "Inga"*	JF961631	JF961664	JF961697	JF961730	JF961527	JF961598
	Steatocranus	sp. aff. *tinanti "ultraslender"*	JF961630	JF961663	JF961696	JF961729	JF961516	JF961597
	Steatocranus	sp. aff. *tinanti "intermediate"*	JF961632	JF961665	JF961698	JF961731	JF961533	JF961599
	Steatocranus	*tinanti*	JF961629	JF961662	JF961695	JF961728	JF961507	JF961596
	Steatocranus	cf. *gibbiceps*	JF961635	JF961668	JF961701	JF961734	JF961545	JF961602
	Steatocranus	cf. *gibbiceps*	JF961636	JF961669	JF961702	JF961735	JF961576	JF961603
	Steatocranus	*mpozoensis*	JF961633	JF961666	JF961699	JF961732	JF961536	JF961600
	Steatocranus	*mpozoensis*	JF961634	JF961667	JF961700	JF961733	JF961542	JF961601

Chapter 4: Transcontinental hybridization among haplochromine cichlids (Cichlidae)

Table S1 *List of all taxa and genes (with GB accession numbers) included in the small mt- dataset*

No	genus	species	locality	ND2	CytB
1	*Pharyngochromis*	sp. "yellow lip"	R. Kwanza/Angola	subm. to NCBI	subm. to NCBI
2	*Pharyngochromis*	sp. "yellow lip"	R. Kwanza/Angola	subm. to NCBI	subm. to NCBI
3	*Pharyngochromis*	sp. "yellow lip"	R. Kwanza/Angola	subm. to NCBI	subm. to NCBI
4	*Serranochromis*	sp. "yellow fins"	R. Kwanza/Angola	subm. to NCBI	subm. to NCBI
5	*Serranochromis*	sp. "red scales"	R. Kwanza/Angola	subm. to NCBI	subm. to NCBI
6	*Serranochromis*	sp. "red scales"	R. Kwanza/Angola	subm. to NCBI	subm. to NCBI
7	*Pharyngochromis*	sp. "white tip"	R. Kwanza/Angola	subm. to NCBI	subm. to NCBI
8	*Pharyngochromis*	sp. "yellow fins"	R. Kwanza/Angola	subm. to NCBI	subm. to NCBI
9	*Serranochromis*	sp. "black and white"	R. Kwanza/Angola	subm. to NCBI	subm. to NCBI
10	*"Haplochromis"*	*oligacanthus*	middle Congo/R. Ubangi	subm. to NCBI	subm. to NCBI
11	*"Haplochromis"*	*oligacanthus*	middle Congo/R. Ubangi	subm. to NCBI	subm. to NCBI
12	*Rhamphochromis*	sp.	Lake Malawi	subm. to NCBI	subm. to NCBI
13	*"Haplochromis"*	sp. "L. Kijanebalola"	Uganda	subm. to NCBI	subm. to NCBI
14	*"Haplochromis"*	flaviijosephi	Jordan system		subm. to NCBI
15	*"Haplochromis"*	sp. "L. Rakai"	L. Rakai/Uganda	subm. to NCBI	subm. to NCBI
16	*"Haplochromis"*	*snoeksi*	Inkisi	subm. to NCBI	subm. to NCBI
17	*"Haplochromis"*	cf. b*akongo*	lower Congo	subm. to NCBI	subm. to NCBI
18	*"Haplochromis"*	cf. *bakongo*	lower Congo	subm. to NCBI	subm. to NCBI
19	*"Haplochromis"*	sp. "Kwango"	River Kwango	subm. to NCBI	subm. to NCBI
20	*"Haplochromis"*	sp. "Sanzikwa"	lower Congo/R. Sanzikwa	subm. to NCBI	subm. to NCBI
21	*"Haplochromis"*	*polli* "Lefini"	Lefini	subm. to NCBI	subm. to NCBI
22	*"Haplochromis"*	*polli* "Lefini"	Lefini	subm. to NCBI	subm. to NCBI
23	*"Haplochromis"*	*fasciatus*	lower Congo	subm. to NCBI	subm. to NCBI
24	*"Haplochromis"*	*fasciatus*	lower Congo	subm. to NCBI	subm. to NCBI
25	*Lamprologus*	sp.	lower Congo	subm. to NCBI	subm. to NCBI
26	*Orthochromis*	*stormsi*	Malebo pool	subm. to NCBI	subm. to NCBI
27	*Orthochromis*	*stormsi*	Malebo pool	subm. to NCBI	
28	*"Haplochromis"*	*demeusii*	lower Congo	subm. to NCBI	subm. to NCBI
29	*"Haplochromis"*	*demeusii*	lower Congo	subm. to NCBI	subm. to NCBI
30	*"Haplochromis"*	sp. "Sanzikwa"	lower Congo/R. Sanzikwa	subm. to NCBI	subm. to NCBI
31	*"Haplochromis"*	*polli*	lower Congo	subm. to NCBI	subm. to NCBI
32	*"Haplochromis"*	sp. "Yaekama"	upper Congo		subm. to NCBI
33	*Tilapia*	*bilineata*		subm. to NCBI	subm. to NCBI
34	*Pharyngochromis*	*acuticeps*	Asouth Africa	subm. to NCBI	subm. to NCBI
35	*Serranochromis*	*robustus*	Asouth Africa	subm. to NCBI	subm. to NCBI
36	*Serranochromis*	*altus*	Asouth Africa	subm. to NCBI	subm. to NCBI
37	*Serranochromis*	*angusticeps*	Asouth Africa	subm. to NCBI	subm. to NCBI
38	*Serranochromis*	*macrocephalus*	Asouth Africa	subm. to NCBI	subm. to NCBI
39	*"Haplochromis"*	*horei*	Lake Tanganyika	subm. to NCBI	subm. to NCBI
40	*"Haplochromis"*	*horei*	Lake Tanganyika	subm. to NCBI	subm. to NCBI
41	*"Haplochromis"*	*thereutherion*	Lake Victoria	subm. to NCBI	subm. to NCBI
42	*"Haplochromis"*	*stigmatogenys*	River Kasai	subm. to NCBI	subm. to NCBI
43	*Thoracochromis*	*callichromis*	River Fwa	subm. to NCBI	subm. to NCBI
44	*Cyclopharynx*	*schwetzi*	River Fwa	subm. to NCBI	subm. to NCBI
45	*"Haplochromis"*	*brauschi*	River Fwa	subm. to NCBI	subm. to NCBI
46	*Thoracochromis*	*callichromis*	River Fwa	subm. to NCBI	subm. to NCBI

No	genus	species	locality	ND2	CytB
47	*Schwetzochromis*	*neodon*	River Fwa	subm. to NCBI	subm. to NCBI
48	*"Haplochromis"*	*stigmatogenys*	River Kasai	subm. to NCBI	subm. to NCBI
49	*Orthochromis*	*stormsi* "Kisangani"	upper Congo	subm. to NCBI	
50	*Orthochromis*	*stormsi* "Kisangani"	upper Congo	subm. to NCBI	
51	*"Haplochromis"*	sp. "Kwango"	River Kwango	subm. to NCBI	subm. to NCBI
52	*"Haplochromis"*	sp. "Kyoga"	L. Kyoga Uganda	subm. to NCBI	subm. to NCBI
53	*Orthochromis*	*torrenticola*	Lufira	subm. to NCBI	subm. to NCBI
54	*Orthochromis*	*torrenticola*	Lufira	subm. to NCBI	subm. to NCBI
55	*Astatoreochromis*	*alluaudi*	around L. Victoria	subm. to NCBI	subm. to NCBI
56	*Orthochromis*	*polyacanthus*	upper and lower Congo River	subm. to NCBI	subm. to NCBI
57	*Neochromis*	*sp.*	Lave Victoria	subm. to NCBI	subm. to NCBI
58	*Orthochromis*	sp. aff. *kalungwishiensis*	Lake Mweru	subm. to NCBI	subm. to NCBI
59	*"Haplochromis"*	*burtoni*	L. Tanganyika and surrounding Rivers	subm. to NCBI	subm. to NCBI
60	*"Haplochromis"*	*stappersi*	L. Tanganyika and surrounding Rivers	subm. to NCBI	subm. to NCBI
61	*Labidochromis*	*caeruleus*	Lake Malawi	subm. to NCBI	subm. to NCBI
62	*Sciaenochromis*	*fryeri*	Lake Malawi	subm. to NCBI	subm. to NCBI
63	*Pseudotropheus*	*socolofi*	Lake Malawi	subm. to NCBI	subm. to NCBI
64	*Pseudocrenilabrus*	*multicolor*	Nile delta	subm. to NCBI	subm. to NCBI
65	*Tropheus*	*moorii*	Lake Tanganyika	subm. to NCBI	subm. to NCBI
66	*"Haplochromis"*	*desfontainii*	Lake Victoria	subm. to NCBI	subm. to NCBI
67	*Tilapia*	sp. aff. *bilineata*	River Lefini	subm. to NCBI	subm. to NCBI
68	*Haplochromis*	*polli*	lower Congo	subm. to NCBI	subm. to NCBI
69	*Pharyngochromis*	sp white tip	R. Kwanza/Angola	subm. to NCBI	subm. to NCBI

Table S2 *List of all taxa and genes (with GB accession numbers) included in the large mt- dataset*
Species order corresponds to phylogenetic clusters in Fig. S2. Taxon names marked in red indicate that these taxa appear also in the small mt- dataset.

Nb in tree	Genus	species	Cytochrome b	NADH 2
40	*Tilapia*	*bilineata*	subm. to NCBI	subm. to NCBI
32	*Lamprologus*	sp.	subm. to NCBI	subm. to NCBI
143	*Lamprologus*	*callipterus*	FJ706659.1	AF398226.1
142	*Julidochromis*	*marlieri*	EF679296.1	AF398230.1
144	*Telmatochromis*	*bifrenatus*	EF679271.1	AF398228.1
135	*Spathodus*	*erythrodon*	AF428156.1	AF398218.1
134	*Eretmodus*	*cyanostictus*	AF428155.1	AF398220.1
136	*Tanganicodus*	*irsacae*	Z21779.1	AF398219.1
49	*Cyclotilapia*	*frontosa*	EF679274.1	EF679242.1
86	*Limnochromis*	*auritus*	Z21775.1	AY337766.1
145	*Triglachromis*	*otostigma*	Z30004.1	AF398217.1
133	*Orthochromis*	*malagaraziensis*	AF428161.1	AF398232.1
30	*Benitochromis*	*melanoides*	EU753922.1	AY682513.1
31	*Benitochromis*	*tricoti*	AF428164.1	EU753962.1
104	*Plecodus*	*straeleni*	EF679290.1	EF679258.1
137	*Perissodus*	*microlepis*	AF428167.1	AF398222.1
50	*Cyprichromis*	*leptosoma*	AY740254.1	EF679243.1
141	*Paracyprichromis*	*brieni*	AY740249.1	AF398223.1
140	*Xenotilapia*	*sima*	AY337837.1	U07270.1|XSU07270
94	*Ophthalmotilapia*	*ventralis*	AY337826.1	AY337774.1
139	*Callochromis*	*macrops*	AY337851.1	U07242.1|CMU07242
56	*Ctenochromis*	*pectoralis*	EU753887.1	EU753938.1
57	*Ctenochromis*	*pectoralis*	EU753888.1	EU753939.1
15	*"Haplochromis"*	flaviijosephi	subm. to NCBI	
59	*"Haplochromis"*	*bloyeti*	EU753879.1	EU753930.1

Nb in tree	Genus	species	Cytochrome b	NADH 2
13	*Rhamphochromis*	sp.	subm. to NCBI	subm. to NCBI
26	*Astatotilapia*	*calliptera*	EU753883.1	EU753934.1
100	*Sciaenochromis*	*fryeri*	subm. to NCBI	subm. to NCBI
101	*Pseudotropheus*	*socolofi*	subm. to NCBI	subm. to NCBI
114	*Pseudotropheus*	*tropheops*	EU753905.1	EF585260.1
87	*Labidochromis*	*caeruleus*	EU753896.1	AY740383.1
99	*Labidochromis*	*caeruleus*	subm. to NCBI	subm. to NCBI
91	*Nimbochromis*	*livingstonii*	EU753897.1	EU753948.1
92	*Nimbochromis*	*venustus*	EU753898.1	EU753947.1
128	*"Haplochromis"*	*desfontainii*	subm. to NCBI	subm. to NCBI
39	*"Haplochromis"*	sp. "Yaekama"	subm. to NCBI	subm. to NCBI
82	*"Haplochromis"*	sp. "Kyoga"	subm. to NCBI	subm. to NCBI
14	*"Haplochromis"*	sp. "L. Kijanebalola"	subm. to NCBI	subm. to NCBI
71	*"Haplochromis"*	*thereutherion*	subm. to NCBI	subm. to NCBI
67	*"Haplochromis"*	sp. "L. Kanyaboli"	EU753893.1	EU753944.1
90	*Neochromis*	*sp.*	subm. to NCBI	subm. to NCBI
63	*"Haplochromis"*	*phytophagus*	EU753889.1	EU753940.1
65	*"Haplochromis"*	*rudolfianus*	EU753891.1	|EU753942.1
69	*"Haplochromis"*	sp. "Mburo black"	EU753895.1	EU753946.1
98	*"Haplochromis"*	*stappersi*	subm. to NCBI	subm. to NCBI
16	*"Haplochromis"*	sp. "L. Rakai"	subm. to NCBI	subm. to NCBI
70	*"Haplochromis"*	*squamipinnis*	EU753892.1	EU753943.1
66	*"Haplochromis"*	sp. "El Fayoum"	EU753894.1	EU753945.1
146	*"Haplochromis"*	sp. "Ituri"	subm. to NCBI	subm. to NCBI
147	*"Haplochromis"*	sp. "Ituri"	subm. to NCBI	subm. to NCBI
97	*"Haplochromis"*	*burtoni*	subm. to NCBI	subm. to NCBI
60	*"Haplochromis"*	*burtoni*	EU753881.1	EU753932.1
37	*"Haplochromis"*	sp. "Sanzikwa"	subm. to NCBI	subm. to NCBI
21	*"Haplochromis"*	sp. "Sanzikwa"	subm. to NCBI	subm. to NCBI
27	*"Haplochromis"*	*fasciatus*	subm. to NCBI	subm. to NCBI
28	*"Haplochromis"*	*fasciatus*	subm. to NCBI	subm. to NCBI
35	*"Haplochromis"*	*demeusii*	subm. to NCBI	subm. to NCBI
36	*"Haplochromis"*	*demeusii*	subm. to NCBI	subm. to NCBI
85	*Lobochilotis*	*labiatus*	AY301932.1	GQ995728.1
55	*"Haplochromis"*	*horei*	EU753884.1	EU753935.1
46	*"Haplochromis"*	*horei*	subm. to NCBI	subm. to NCBI
47	*"Haplochromis"*	*horei*	subm. to NCBI	subm. to NCBI
58	*Gnathochromis*	*pfefferi*	AY301929.1	GQ995721.1
138	*Gnathochromis*	*pfefferi*	AY301929.1	U07248.1|GPU07248
103	*Tropheus*	*moorii*	subm. to NCBI	subm. to NCBI
129	*Tropheus*	*moorii*	Z12035.1	GQ995810.1
25	*Astatoreochromis*	*alluaudi*	EU753872.1	EU753923.1
88	*Astatoreochromis*	*alluaudi*	subm. to NCBI	subm. to NCBI
68	*"Haplochromis"*	sp. "Lufubu"	EU753877.1	EU753928.1
93	*Orthochromis*	sp. aff. *kalungwishiensis*	subm. to NCBI	subm. to NCBI
112	*Pseudocrenilabrus*	sp. "Mweru orange"	EU753903.1	EU753952.1
95	*Orthochromis*	*machadoi*	EU753885.1	EU753936.1
113	*Pseudocrenilabrus*	sp. "Olushandja"	EU753904.1	EU753953.1
102	*Pseudocrenilabrus*	*multicolor*	subm. to NCBI	subm. to NCBI
108	*Pseudocrenilabrus*	*nicholsi*	AY600143.1	AY602994.1
107	*Pseudocrenilabrus*	*multicolor*	AY600141.1	AY930106.1
109	*Pseudocrenilabrus*	*philander*	AY600142.1	AY930047.1
110	*Pseudocrenilabrus*	sp. "Lufubu"	EU753901.1	EU753950.1
111	*Pseudocrenilabrus*	sp. "Lunzua"	EU753902.1	EU753951.1

Nb in tree	Genus	species	Cytochrome b	NADH 2
17	*"Haplochromis"*	*snoeksi*	subm. to NCBI	subm. to NCBI
29	*"Haplochromis"*	*bakongo*	subm. to NCBI	subm. to NCBI
18	*"Haplochromis"*	cf. *bakongo*	subm. to NCBI	subm. to NCBI
19	*"Haplochromis"*	cf. *bakongo*	subm. to NCBI	subm. to NCBI
48	*Cyclopharynx*	*fwae*	AF428158.1	AY930099.1
74	*Cyclopharynx*	*schwetzi*	subm. to NCBI	subm. to NCBI
75	*"Haplochromis"*	*brauschi*	subm. to NCBI	subm. to NCBI
131	*"Haplochromis"*	*brauschi*	EU753880.1	EU753931.1
73	*"Haplochromis"*	*callichromus*	subm. to NCBI	subm. to NCBI
76	*"Haplochromis"*	*callichromus*	subm. to NCBI	subm. to NCBI
77	*Schwetzochromis*	*neodon*	subm. to NCBI	subm. to NCBI
115	*Schwetzochromis*	*neodon*	EU753912.1	EU753957.1
5	*Serranochromis*	sp. "red scales"	subm. to NCBI	subm. to NCBI
6	*Serranochromis*	sp. "red scales"	subm. to NCBI	subm. to NCBI
89	*Orthochromis*	*polyacanthus*	subm. to NCBI	subm. to NCBI
33	*Orthochromis*	*stormsi*	subm. to NCBI	subm. to NCBI
34	*Orthochromis*	*stormsi*	no data	subm. to NCBI
96	*Orthochromis*	*stormsi*	AF428159.1	AF398231.1
79	*Orthochromis*	*stormsi* "Kisangani"	no data	subm. to NCBI
80	*Orthochromis*	*stormsi* "Kisangani"	no data	subm. to NCBI
4	*Serranochromis*	sp. "yellow fins"	subm. to NCBI	subm. to NCBI
9	*Pharyngochromis*	sp. "yellow fins"	subm. to NCBI	subm. to NCBI
10	*Serranochromis*	sp. "black and white"	subm. to NCBI	subm. to NCBI
51	*Chetia*	*brevicauda*	EU753873.1	EU753924.1
106	*Pharyngochromis*	*acuticeps*	EU753899.1	EF393692.1
125	*Sargochromis*	*coulteri*	EU753908.1	EU753955.1
41	*Pharyngochromis*	*acuticeps*	subm. to NCBI	subm. to NCBI
127	*Sargochromis*	*mellandi*	EU753910.1	EF393711.1
53	*Chetia*	*flaviventris*	EU753875.1	EU753926.1
54	*Chetia*	*flaviventris*	EU753876.1	EU753927.1
105	*Pharyngochromis*	*acuticeps*	EU753900.1	EU753949.1
123	*Sargochromis*	aff. *carlottae*	EU753911.1	EU753956.1
126	*Sargochromis*	*giardi*	EU753909.1	EF393714.1
45	*Serranochromis*	*macrocephalus*	subm. to NCBI	subm. to NCBI
119	*Serranochromis*	*macrocephalus*	EU753917.1	EF393705.1
117	*Serranochromis*	*angusticeps*	EU753914.1	EU753958.1
52	*Chetia*	*brevis*	EU753874.1	EU753925.1
43	*Serranochromis*	*altus*	subm. to NCBI	subm. to NCBI
44	*Serranochromis*	*angusticeps*	subm. to NCBI	subm. to NCBI
116	*Serranochromis*	*altus*	EU753913.1	EF393696.1
124	*Sargochromis*	*coulteri*	EU753907.1	EU753954.1
20	*"Haplochromis"*	sp. "Kwango"	subm. to NCBI	subm. to NCBI
81	*"Haplochromis"*	sp. "Kwango"	subm. to NCBI	subm. to NCBI
72	*"Haplochromis"*	*stigmatogenys*	subm. to NCBI	subm. to NCBI
78	*"Haplochromis"*	*stigmatogenys*	subm. to NCBI	subm. to NCBI
120	*Serranochromis*	*stappersi*	EU753919.1	EU753960.1
118	*Serranochromis*	*angusticeps*	EU753915.1	EU753959.1
42	*Serranochromis*	*robustus*	subm. to NCBI	subm. to NCBI
121	*Serranochromis*	*thumbergi*	EU753920.1	EU753961.1
122	*Serranochromis*	*thumbergi*	EU753921.1	EU753961.1
130	*Thoracochromis*	*albolabris*	EU753878.1	EU753929.1
132	*Thoracochromis*	*buysi*	EU753882.1	EU753933.1
62	*"Haplochromis"*	*horei*	EU753886.1	EU753937.1
11	*"Haplochromis"*	*oligacanthus*	subm. to NCBI	subm. to NCBI

Nb in tree	Genus	species	Cytochrome b	NADH 2
12	*"Haplochromis"*	*oligacanthus*	subm. to NCBI	subm. to NCBI
22	*"Haplochromis"*	*polli* "Lefini"	subm. to NCBI	subm. to NCBI
23	*"Haplochromis"*	*polli* "Lefini"	subm. to NCBI	subm. to NCBI
38	*"Haplochromis"*	*polli*	subm. to NCBI	subm. to NCBI
64	*"Haplochromis"*	*polli*	EU753890.1	EU753941.1
7	*Pharyngochromis*	sp. "white tip"	subm. to NCBI	subm. to NCBI
8	*Pharyngochromis*	sp. "white tip"	subm. to NCBI	subm. to NCBI
83	*Orthochromis*	*torrenticola*	subm. to NCBI	subm. to NCBI
84	*Orthochromis*	*torrenticola*	subm. to NCBI	subm. to NCBI
3	*Pharyngochromis*	sp. "yellow lip"	subm. to NCBI	subm. to NCBI
1	*Pharyngochromis*	sp. "yellow lip"	subm. to NCBI	subm. to NCBI
2	*Pharyngochromis*	sp. "yellow lip"	subm. to NCBI	subm. to NCBI

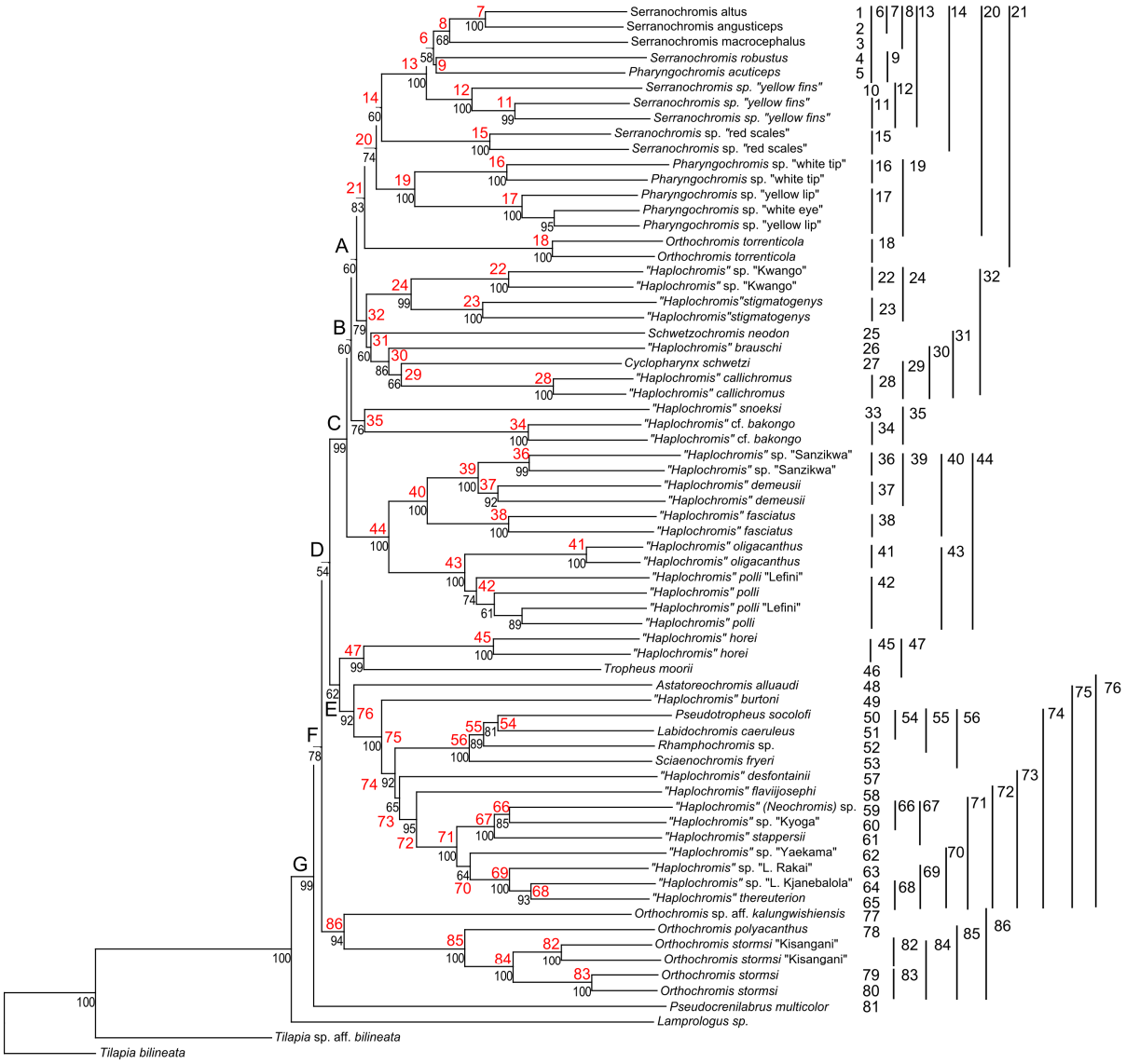


Figure S1 *Overview of BS- removals and nodes*

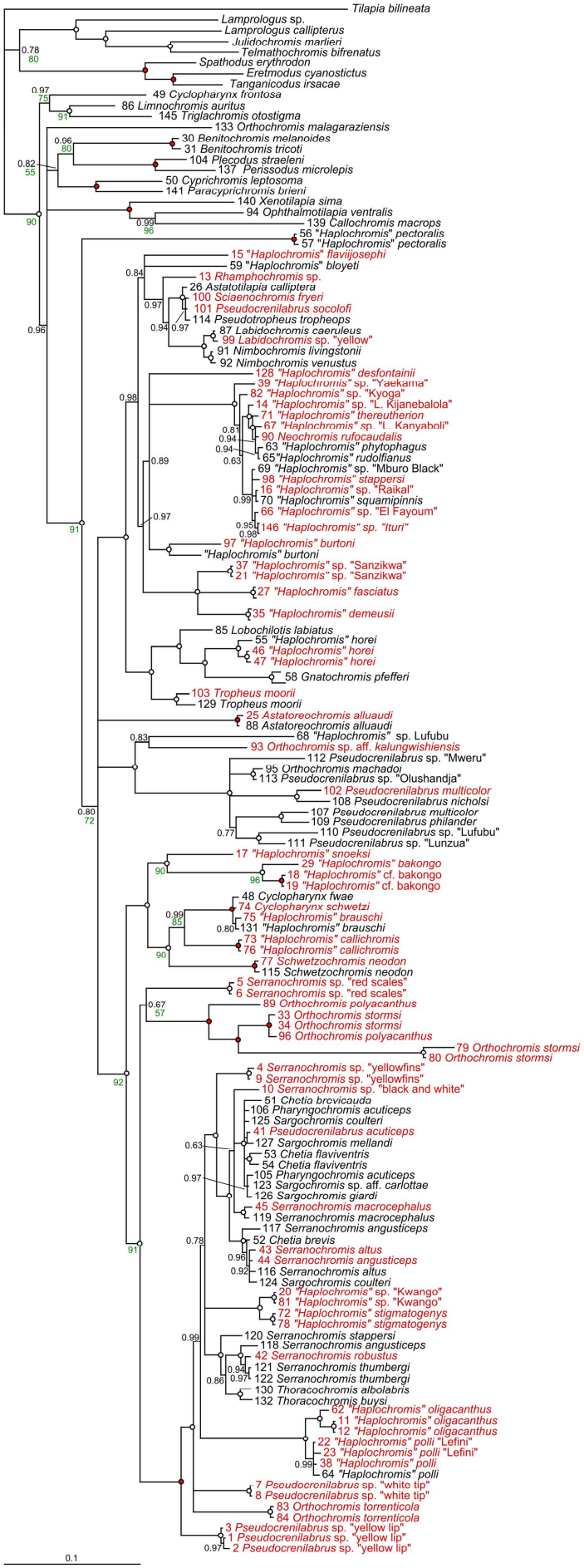

Tilapia bilineata
Lamprologus sp.
Lamprologus callipterus
Julidochromis marlieri
Telmathochromis bifrenatus
Spathodus erythrodon
Eretmodus cyanostictus
Tanganicodus irsacae
49 Cyclopharynx frontosa
86 Limnochromis auritus
145 Triglachromis otostigma
133 Orthochromis malagaraziensis
30 Benitochromis melanoides
31 Benitochromis tricoti
104 Plecodus straeleni
137 Perissodus microlepis
50 Cyprichromis leptosoma
141 Paracyprichromis brieni
140 Xenotilapia sima
94 Ophthalmotilapia ventralis
139 Callochromis macrops
56 "Haplochromis" pectoralis
57 "Haplochromis" pectoralis
15 "Haplochromis" flaviijosephi
59 "Haplochromis" bloyeti
13 Rhamphochromis sp.
26 Astatotilapia calliptera
100 Sciaenochromis fryeri
101 Pseudocrenilabrus socolofi
114 Pseudotropheus tropheops
87 Labidochromis caeruleus
99 Labidochromis sp. "yellow"
91 Nimbochromis livingstonii
92 Nimbochromis venustus
128 "Haplochromis" desfontainii
39 "Haplochromis" sp. "Yaekama"
82 "Haplochromis" sp. "Kyoga"
14 "Haplochromis" sp. "L. Kijanebalola"
71 "Haplochromis" thereutherion
67 "Haplochromis" sp. "L. Kanyaboli"
90 Neochromis rufocaudalis
63 "Haplochromis" phytophagus
65"Haplochromis" rudolfianus
69 "Haplochromis" sp. "Mburo Black"
98 "Haplochromis" stappersi
16 "Haplochromis" sp. "Raikal"
70 "Haplochromis" squamipinnis
66 "Haplochromis" sp. "El Fayoum"
146 "Haplochromis" sp. "Ituri"
97 "Haplochromis" burtoni
"Haplochromis" burtoni
37 "Haplochromis" sp. "Sanzikwa"
21 "Haplochromis" sp. "Sanzikwa"
27 "Haplochromis" fasciatus
35 "Haplochromis" demeusii
85 Lobochilotis labiatus
55 "Haplochromis" horei
46 "Haplochromis" horei
47 "Haplochromis" horei
58 Gnatochromis pfefferi
103 Tropheus moorii
129 Tropheus moorii
25 Astatoreochromis alluaudi
88 Astatoreochromis alluaudi
68 "Haplochromis" sp. Lufubu
93 Orthochromis sp. aff. kalungwishiensis
112 Pseudocrenilabrus sp. "Mweru"
95 Orthochromis machadoi
113 Pseudocrenilabrus sp. "Olushandja"
102 Pseudocrenilabrus multicolor
108 Pseudocrenilabrus nicholsi
107 Pseudocrenilabrus multicolor
109 Pseudocrenilabrus philander
110 Pseudocrenilabrus sp. "Lufubu"
111 Pseudocrenilabrus sp. "Lunzua"
17 "Haplochromis" snoeksi
29 "Haplochromis" bakongo
18 "Haplochromis" cf. bakongo
19 "Haplochromis" cf. bakongo
48 Cyclopharynx fwae
74 Cyclopharynx schwetzi
75 "Haplochromis" brauschi
131 "Haplochromis" brauschi
73 "Haplochromis" callichromis
76 "Haplochromis" callichromis
77 Schwetzochromis neodon
115 Schwetzochromis neodon
5 Serranochromis sp. "red scales"
6 Serranochromis sp. "red scales"
89 Orthochromis polyacanthus
33 Orthochromis stormsi
34 Orthochromis stormsi
96 Orthochromis polyacanthus
79 Orthochromis stormsi
80 Orthochromis stormsi
4 Serranochromis sp. "yellowfins"
9 Serranochromis sp. "yellowfins"
10 Serranochromis sp. "black and white"
51 Chetia brevicauda
106 Pharyngochromis acuticeps
125 Sargochromis coulteri
41 Pseudocrenilabrus acuticeps
127 Sargochromis mellandi
53 Chetia flaviventris
54 Chetia flaviventris
105 Pharyngochromis acuticeps
123 Sargochromis sp. aff. carlottae
126 Sargochromis giardi
45 Serranochromis macrocephalus
119 Serranochromis macrocephalus
117 Serranochromis angusticeps
52 Chetia brevis
43 Serranochromis altus
44 Serranochromis angusticeps
116 Serranochromis altus
124 Sargochromis coulteri
20 "Haplochromis" sp. "Kwango"
81 "Haplochromis" sp. "Kwango"
72 "Haplochromis" stigmatogenys
78 "Haplochromis" stigmatogenys
120 Serranochromis stappersi
118 Serranochromis angusticeps
42 Serranochromis robustus
121 Serranochromis thumbergi
122 Serranochromis thumbergi
130 Thoracochromis albolabris
132 Thoracochromis buysi
62 "Haplochromis" oligacanthus
11 "Haplochromis" oligacanthus
12 "Haplochromis" oligacanthus
22 "Haplochromis" polli "Lefini"
23 "Haplochromis" polli "Lefini"
38 "Haplochromis" polli
64 "Haplochromis" polli
7 Pseudocrenilabrus sp. "white tip"
8 Pseudocrenilabrus sp. "white tip"
83 Orthochromis torrenticola
84 Orthochromis torrenticola
3 Pseudocrenilabrus sp. "yellow lip"
1 Pseudocrenilabrus sp. "yellow lip"
2 Pseudocrenilabrus sp. "yellow lip"
0.1

Figure S2

Figure S2 *Phylogeny based on the large mitochondrial dataset*

Maximum likelihood phylogeny for the large mt-dataset based on 1020bp of ND2 and 380 of Cytb. Sequences were taken from GenBank (N=80) and additional taxa from the mt-dataset (N=65) were supplemented. Bootstrap support values are marked in green and BPP in black. Specimens taken from the small mt-dataset are marked in red.

Table S1 *Specimen information*

Overview of samples used for the mt- and AFLP dataset

		Specimen information				Molecular datasets		
Taxa nb.	**Clade**	**Genus**	**Species**	**River**	**Locality**	**AFLP dataset**	**mt- dataset**	**Gene Bank accession**
S1	**Outgroup**	*Eretmodus*	*cyanotictus*		Lake Tanganyika	**X**	**X**	GQ167831.1
S2		*Lamprologus*	*werneri*	Congo River	Foulakari confluence	**X**	**X**	JF961463
S3		*Tilapia*	*bilineata*		Salonga	**X**	**X**	GQ167824
S4		*Tilapia*	*busumana*			**X**	**X**	GQ167798
S5	**Southern clade**	*Steatocranus*	*bleheri*	Congo River	"Tshikapa" in RC	**X**	**X**	JF961464
S6		*Steatocranus*	*bleheri*	Congo River	"Tshikapa" in RC	**X**	**X**	JF961465
S7		*Steatocranus*	*rouxi*	Kwango		**X**	**X**	JF961466
S8		*Steatocranus*	sp. "red eye"			**X**	**X**	GQ167808.1
S9		*Steatocranus*	sp. "bulky head"			**X**	**X**	GQ167793
S10		*Steatocranus*	sp. "bulky head"	Congo River	"Tshikapa" in RC	**X**	-	
S11		*Steatocranus*	sp. "bulky head"	Congo River	Kinsuka	**X**	**X**	JF961467
S12	**Northern clade**	*Steatocranus*	sp. "Kisangani"	Tshopo River	below falls in Kisangani	**X**	**X**	JF961468
S13		*Steatocranus*	sp. "Kisangani"	Tshopo River	below falls in Kisangani	**X**	**X**	JF961469
S14		*Steatocranus*	sp. "Kisangani"	Congo River	Lindi River, Chutes Soli	**X**	-	
S15		*Steatocranus*	*ubanguiensis*	Ubangui River	Ubangui River	**X**	**X**	GQ167826.1
S16		*Steatocranus*	sp. "Lefini"		Lefini River	-	**X**	JF961470
S17		*Steatocranus*	sp. "Nki"	Dja?	Nki falls	**X**	**X**	JF961471
S18		*Steatocranus*	sp. "Nki"	Dja?	Nki falls	**X**	**X**	JF961472
S19	***S.* sp. "dwarf"**	*Steatocranus*	sp. "dwarf"	Congo River	Congo River	**X**	**X**	GQ167794
S20	**S. casuarius**	*Steatocranus*	*casuarius*	Congo River	Kinsuka	**X**	**X**	JF961473
S21		*Steatocranus*	*casuarius*	Congo River	Kinsuka	**X**	**X**	JF961474
S22		*Steatocranus*	*casuarius*	Congo River	Kinsuka	**X**	**X**	JF961475
S23		*Steatocranus*	*casuarius*	Congo River	"Tshikapa" in RC	**X**	**X**	JF961476
S24		*Steatocranus*	*casuarius*	Congo River	"Tshikapa" in RC	**X**	**X**	JF961477

		Specimen information				Molecular datasets		
Taxa nb.	**Clade**	**Genus**	**Species**	**River**	**Locality**	**AFLP dataset**	**mt- dataset**	**Gene Bank accession**
S25		*Steatocranus*	*casuarius*	Congo River	Kinsuka	**X**	**X**	JF961478
S26		*Steatocranus*	*casuarius*	Congo River	Kinsuka	**X**	**X**	JF961479
S27		*Steatocranus*	*casuarius*	Congo River	Les Rapides, Brazzaville	**X**	**X**	JF961480
S28		*Steatocranus*	*casuarius*	Congo River	Les Rapides, Brazzaville	**X**	**X**	JF961481
S29		*Steatocranus*	*casuarius*	Congo River	Les Rapides, Brazzaville	**X**	**X**	JF961482
S30		*Steatocranus*	*casuarius*	Congo River	Foulakari confluence	**X**	**X**	JF961483
S31		*Steatocranus*	*casuarius*	Congo River	Foulakari confluence	**X**	**X**	JF961484
S32		*Steatocranus*	*casuarius*	Congo River	Foulakari confluence	**X**	**X**	JF961485
S33		*Steatocranus*	sp. aff. *casuarius* "brown pearl"	Congo River	Luozi village	**X**	**X**	JF961486
S34		*Steatocranus*	sp. aff. *casuarius* "brown pearl"	Congo River	Lenga Langa Camp	**-**	**X**	JF961487
S35		*Steatocranus*	sp. aff. *casuarius* "brown pearl"	Congo River	Lenga Langa Camp	**-**	**X**	JF961488
S36		*Steatocranus*	sp. aff. *casuarius* "brown pearl"	Congo River	Luozi - south side of channel	**X**	**X**	JF961489
S37		*Steatocranus*	sp. aff. *casuarius* "brown pearl"	Congo River	Bulu -north side of channel	**X**	**X**	JF961490
S38		*Steatocranus*	sp. aff. *casuarius* "brown pearl"	Congo River	Bulu beach-north side of channel	**X**	**X**	JF961491
S39		*Steatocranus*	sp. aff. *casuarius* "brown pearl"	Congo River	Bulu beach-north side of channel	**X**	**X**	JF961492
S40	**S. cf. casuarius**	*Steatocranus*	sp. aff. *casuarius* "brown pearl"	Congo River	Bulu -south side of channel	**X**	**-**	
S41		*Steatocranus*	sp. aff. *casuarius* "brown pearl"	Congo River	Bulu -south side of channel	**X**	**-**	
S42		*Steatocranus*	sp. aff. *casuarius* "brown pearl"	Congo River	Upstream Bulu South	**X**	**X**	JF961493
S43		*Steatocranus*	sp. aff. *casuarius* "brown pearl"	Congo River	Bulu - south side of channel	**X**	**X**	JF961494
S44		*Steatocranus*	sp. aff. *casuarius* "brown pearl"	Congo River	upstream Bulu rocky island	**X**	**X**	JF961495
S45		*Steatocranus*	sp. aff. *casuarius* "brown pearl"	Congo River	Kinganga	**X**	**X**	JF961496
S46		*Steatocranus*	sp. aff. *casuarius* "brown pearl"	Congo River	Kinganga	**X**	**-**	
S47		*Steatocranus*	sp. aff. *casuarius* "brown pearl"	Congo River	Bac Kinganga	**X**	**-**	
S48		*Steatocranus*	sp. aff. *casuarius* "brown pearl"	Congo River	Inga North	**X**	**X**	JF961497
S49		*Steatocranus*	sp. aff. *casuarius* "brown pearl"	Congo River	Inga middle	**X**	**X**	JF961498
S50		*Steatocranus*	sp. aff. *casuarius* "brown pearl"	Congo River	Inga middle	**X**	**-**	
S51		*Steatocranus*	sp. aff. *casuarius* "brown pearl"	Congo River	Inga, north of C05_10	**X**	**-**	
S52	**S. sp. "Maluku"**	*Steatocranus*	sp. "Maluku"	Congo River	Maluku	**X**	**X**	JF961499
S53		*Steatocranus*	sp. "Maluku"	Congo River	Maluku	**X**	**X**	JF961500
S54		*Steatocranus*	sp. "Maluku"	Congo River	Maluku	**X**	**X**	JF961501
S55		*Steatocranus*	sp. "Maluku"	Congo River	Maluku	**X**	**X**	JF961502
S56		*Steatocranus*	sp. "Maluku"	Congo River	Maluku	**X**	**X**	JF961503

		Specimen information				Molecular datasets		
Taxa nb.	**Clade**	**Genus**	**Species**	**River**	**Locality**	**AFLP dataset**	**mt- dataset**	**Gene Bank accession**
S57		*Steatocranus*	sp. "Maluku"	Congo River	Maluku	**X**	**X**	JF961504
S58		*Steatocranus*	*tinanti*	Congo River	Kinsuka	**X**	**X**	JF961505
S59		*Steatocranus*	*tinanti*	Congo River	Kinsuka	**X**	**X**	JF961506
S60		*Steatocranus*	*tinanti*	Congo River	Kinsuka	**X**	**X**	JF961507
S61		*Steatocranus*	*tinanti*	Congo River	Kinsuka	**X**	**X**	JF961508
S62	**S. tinanti**	*Steatocranus*	*tinanti*	Congo River	Kinsuka	**X**	**X**	JF961509
S63		*Steatocranus*	*tinanti*	Congo River	Les Rapides, Brazzaville	**X**	**X**	JF961510
S64		*Steatocranus*	*tinanti*	Congo River	Les Rapides, Brazzaville	**X**	**X**	JF961511
S65		*Steatocranus*	*tinanti*	Congo River	Les Rapides, Brazzaville	**X**	**X**	JF961512
S66		*Steatocranus*	*tinanti*	Congo River	Foulakari confluence	**X**	**X**	JF961513
S67		*Steatocranus*	*tinanti*	Congo River	Foulakari confluence	**X**	**X**	JF961514
S68		*Steatocranus*	*tinanti*	Congo River	mouth of Inkisi river	**X**	**X**	JF961515
S69		*Steatocranus*	sp. aff. *tinanti "ultraslender"*	Congo River	Luozi - south side of channel	**X**	**X**	JF961516
S70		*Steatocranus*	sp. aff. *tinanti "ultraslender"*	Congo River	Banda Nyenge	**-**	**X**	JF961517
S71		*Steatocranus*	sp. aff. *tinanti "ultraslender"*	Congo River	Rapid before Tadi middle	**X**	**X**	JF961518
S72		*Steatocranus*	sp. aff. *tinanti "ultraslender"*	Congo River	Rapid before Tadi middle	**X**	**X**	JF961519
S73		*Steatocranus*	sp. aff. *tinanti "ultraslender"*	Congo River	Bulu beach north	**X**	**X**	JF961520
S74		*Steatocranus*	sp. aff. *tinanti "ultraslender"*	Congo River	Bulu beach-north side of channel	**X**	**X**	JF961521
S75		*Steatocranus*	sp. aff. *tinanti "ultraslender"*	Congo River	Bulu beach north	**X**	**-**	
S76	**S. cf. tinanti I**	*Steatocranus*	sp. aff. *tinanti "ultraslender"*	Congo River	Bulu -south side of channel	**X**	**-**	
S77		*Steatocranus*	sp. aff. *tinanti "ultraslender"*	Congo River	Bulu -south side of channel	**X**	**-**	
S78		*Steatocranus*	sp. aff. *tinanti "ultraslender"*	Congo River	Bulu - south side of channel	**X**	**-**	
S79		*Steatocranus*	sp. aff. *tinanti "ultraslender"*	Congo River	Kinganga	**X**	**X**	JF961522
S80		*Steatocranus*	sp. aff. *tinanti "ultraslender"*	Congo River	Kinganga	**X**	**X**	JF961523
S81		*Steatocranus*	sp. aff. *tinanti "ultraslender"*	Congo River	Kinganga	**X**	**-**	
S82		*Steatocranus*	sp. aff. *tinanti "ultraslender"*	Congo River	Kinganga	**X**	**X**	JF961524
S83	**S. cf. tinanti II**	*Steatocranus*	sp. aff. *tinanti "Inga"*	Congo River	Inga middle	**X**	**X**	JF961525
S84		*Steatocranus*	sp. aff. *tinanti "Inga"*	Congo River	Inga middle	**X**	**X**	JF961526
S85		*Steatocranus*	sp. aff. *tinanti "Inga"*	Congo River	Inga around the corner, north - Point 50	**X**	**X**	JF961527
S86		*Steatocranus*	sp. aff. *tinanti "Inga"*	Congo River	Inga around the corner, north - Point 50	**X**	**X**	JF961528
S87	**S. cf. tinanti III**	*Steatocranus*	sp. aff. *tinanti "intermediate"*	Congo River	Yalala - North	**X**	**-**	
S88		*Steatocranus*	sp. aff. *tinanti "intermediate"*	Congo River	Yalala - North	**X**	**X**	JF961529

		Specimen information				Molecular datasets		
Taxa nb.	**Clade**	**Genus**	**Species**	**River**	**Locality**	**AFLP dataset**	**mt- dataset**	**Gene Bank accession**
S89		*Steatocranus*	sp. aff. *tinanti "intermediate"*	Congo River	Yalala rapid	**X**	**X**	JF961530
S90		*Steatocranus*	sp. aff. *tinanti "intermediate"*	Congo River	Yalala rapid II	**X**	**X**	JF961531
S91		*Steatocranus*	sp. aff. *tinanti "intermediate"*	Congo River	Yalala rapid II	**X**	**X**	JF961532
S92		*Steatocranus*	sp. aff. *tinanti "intermediate"*	Congo River	Yalala rapid II	**X**	**X**	JF961533
S93		*Steatocranus*	*mpozoensis*	Congo River	Mpozo - South	**X**	**X**	JF961534
S94		*Steatocranus*	*mpozoensis*	Congo River	Mpozo - South	**X**	**X**	JF961535
S95		*Steatocranus*	*mpozoensis*	Congo River	Matadi	**X**	**X**	JF961536
S96	**S. cf. tinanti IV**	*Steatocranus*	*mpozoensis*	Congo River	Matadi	**X**	**X**	JF961537
S97		*Steatocranus*	*mpozoensis*	Congo River	opposite Mpozo North	**X**	**X**	JF961538
S98		*Steatocranus*	*mpozoensis*	Congo River	oppsite Mpozo North	**X**	**X**	JF961539
S99		*Steatocranus*	*mpozoensis*	Congo River	Boma North	**X**	**X**	JF961540
S100		*Steatocranus*	*mpozoensis*	Congo River	Boma North	**X**	**X**	JF961541
S101		*Steatocranus*	*mpozoensis*	Congo River	Boma Pirogue	**X**	**X**	JF961542
S102		*Steatocranus*	*mpozoensis*	Congo River	Boma Pirogue	**X**	**X**	JF961543
S103		*Steatocranus*	cf. *gibbiceps*	Congo River	Kinsuka	**X**	**X**	JF961544
S104		*Steatocranus*	cf. *gibbiceps*	Congo River	Kinsuka	**X**	**X**	JF961545
S105		*Steatocranus*	cf. *gibbiceps*	Congo River	Kinsuka	**X**	**X**	JF961546
S106		*Steatocranus*	cf. *gibbiceps*	Congo River	Kinsuka	**X**	**X**	JF961547
S107	**S. cf gibbiceps I**	*Steatocranus*	cf. *gibbiceps*	Congo River	Les Rapides, Brazzaville	**X**	**X**	JF961548
S108		*Steatocranus*	cf. *gibbiceps*	Congo River	Les Rapides, Brazzaville	**X**	**X**	JF961549
S109		*Steatocranus*	cf. *gibbiceps*	Congo River	Mbelo, main chanel below Bela	**X**	**X**	JF961550
S110		*Steatocranus*	cf. *gibbiceps*	Congo River	Foulakari confluence	**-**	**X**	JF961551
S111		*Steatocranus*	cf. *gibbiceps*	Congo River	Luozi village	**X**	**X**	JF961552
S112		*Steatocranus*	cf. *gibbiceps*	Congo River	upstream Luozi - north side	**X**	**-**	
S113		*Steatocranus*	cf. *gibbiceps*	Congo River	upstream Luozi - north side	**X**	**X**	JF961553
S114		*Steatocranus*	cf. *gibbiceps*	Congo River	upstream Luozi - north side	**X**	**X**	JF961554
S115		*Steatocranus*	cf. *gibbiceps*	Congo River	Luozi - south side of channel	**X**	**-**	
S116	**S. cf gibbiceps II**	*Steatocranus*	cf. *gibbiceps*	Congo River	Luozi - south side of channel	**X**	**-**	
S117		*Steatocranus*	cf. *gibbiceps*	Congo River	Banda Nyenge	**-**	**X**	JF961555
S118		*Steatocranus*	cf. *gibbiceps*	Congo River	opposite Lenga Lenga (North)	**-**	**X**	JF961556
S119		*Steatocranus*	cf. *gibbiceps*	Congo River	opposite Lenga Lenga (North)	**-**	**X**	JF961557
S120		*Steatocranus*	cf. *gibbiceps*	Congo River	Luozi - north side of channel	**-**	**X**	JF961558

		Specimen information				Molecular datasets		
Taxa nb.	**Clade**	**Genus**	**Species**	**River**	**Locality**	**AFLP dataset**	**mt- dataset**	**Gene Bank accession**
S121		*Steatocranus*	cf. *gibbiceps*	Congo River	Rapid before Tadi middle	**X**	**X**	JF961559
S122		*Steatocranus*	cf. *gibbiceps*	Congo River	Bulu -north side of channel	**X**	**X**	JF961560
S123		*Steatocranus*	cf. *gibbiceps*	Congo River	Rapid before Tadi middle	**X**	**X**	JF961561
S124		*Steatocranus*	cf. *gibbiceps*	Congo River	Rapid before Tadi middle	**X**	**X**	JF961562
S125		*Steatocranus*	cf. *gibbiceps*	Congo River	Rapid before Tadi middle	**X**	**X**	JF961563
S126		*Steatocranus*	cf. *gibbiceps*	Congo River	downstream Bulu North	**X**	**X**	JF961564
S127		*Steatocranus*	cf. *gibbiceps*	Congo River	Upstream Bulu South	**X**	**X**	JF961565
S128		*Steatocranus*	cf. *gibbiceps*	Congo River	Upstream Bulu South	**X**	**X**	JF961566
S129		*Steatocranus*	cf. *gibbiceps*	Congo River	upstream Bulu rocky island	**X**	**X**	JF961567
S130		*Steatocranus*	cf. *gibbiceps*	Congo River	Upstream Bulu South	**X**	**X**	JF961568
S131		*Steatocranus*	cf. *gibbiceps*	Congo River	Bulu - south side of channel	**X**	**X**	JF961569
S132		*Steatocranus*	cf. *gibbiceps*	Congo River	Kinganga	**X**	**X**	JF961570
S146		*Steatocranus*	cf. *gibbiceps*	Congo River	Kinganga	**X**	**X**	JF961582
S133		*Steatocranus*	cf. *gibbiceps*	Congo River	Kinganga	**X**	**X**	JF961571
S134		*Steatocranus*	cf. *gibbiceps*	Congo River	Kinganga	**X**	**X**	JF961572
S135		*Steatocranus*	cf. *gibbiceps*	Congo River	Inga, north of C05_10	**X**	**X**	JF961573
S136		*Steatocranus*	cf. *gibbiceps*	Congo River	Inga middle	**X**	**X**	JF961574
S137		*Steatocranus*	cf. *gibbiceps*	Congo River	Inga around the corner, north - Point 50	**X**	-	
S138		*Steatocranus*	cf. *gibbiceps*	Congo River	Inga around the corner, north - Point 50	**X**	**X**	JF961575
S139		*Steatocranus*	cf. *gibbiceps*	Congo River	Inga around the corner, north - Point 50	**X**	**X**	JF961576
S140		*Steatocranus*	cf. *gibbiceps*	Congo River	Inga around the corner, north - Point 50	**X**	-	
S141		*Steatocranus*	cf. *gibbiceps*	Congo River	Inga North	**X**	**X**	JF961577
S142		*Steatocranus*	*glaber*	Congo River	Nziya_below Inga	**X**	**X**	JF961578
S143		*Steatocranus*	*glaber*	Congo River	Nziya_below Inga	**X**	**X**	JF961579
S144		*Steatocranus*	*glaber*	Congo River	Yalala - North	**X**	**X**	JF961580
S145		*Steatocranus*	*glaber*	Congo River	Yalala - North	**X**	**X**	JF961581
S147	**S. glaber**	*Steatocranus*	*glaber*	Congo River	Yalala rapid II	**X**	-	
S148		*Steatocranus*	*glaber*	Congo River	Yalala rapid II	**X**	**X**	JF961583
S149		*Steatocranus*	*glaber*	Congo River	Yalala rapid II	**X**	**X**	JF961584
S150		*Steatocranus*	*glaber*	Congo River	Yalala rapid	**X**	-	
S151		*Steatocranus*	*glaber*	Congo River	Yalala - North	**X**	**X**	JF961585
S152		*Steatocranus*	*glaber*	Congo River	Yalala - North	**X**	**X**	JF961586

		Specimen information				Molecular datasets		
Taxa nb.	**Clade**	**Genus**	**Species**	**River**	**Locality**	**AFLP dataset**	**mt- dataset**	**Gene Bank accession**
S153		*Steatocranus*	*glaber*	Congo River	Yalala - North	**X**	**X**	JF961587
S154		*Steatocranus*	*glaber*	Congo River	Yalala - North	**X**	**X**	JF961588
N1		*Teleogramma*	*gracile*	Congo River	Bulu - south side of channel	**X**	**X**	JF961380
N2	**Outgroup**	*Parananochromis*	sp.	Amba River	Amba River	**X**	**-**	
N3		*Congochromis*	sp.	Amba River	Amba River	**X**	**X**	JF961381
N4		*Congochromis*	sp.	Lilanda	Lilanda River	**X**	**X**	JF961382
N5		*Nanochromis*	*sabinae*			**X**	**-**	
N6		*Nanochromis*	cf. *sabinae*		Lui Kotale	**X**	**-**	
N7	***N. minor***	*Nanochromis*	*minor*	Congo River	Kinganga	**X**	**X**	JF961383
N8	**Lacustrine**	*Nanochromis*	*wickleri*	Lake Mai Ndombe	Xeno	**X**	**X**	JF961384
N9	***Nanochromis***	*Nanochromis*	*transvestitus*	Lake Mai Ndombe	Inongo	**X**	**X**	JF961385
N10	***N.* sp. "Ndongo"**	*Nanochromis*	sp. "Ndongo"	Ngoko River	near village Ndongo, Cameroon	**X**	**X**	JF961386
N11		*Nanochromis*	*splendens*	Congo River	Luozi - south side of channel	**X**	**X**	JF961387
N12		*Nanochromis*	*splendens*	Congo River	Luozi - south side of channel	**X**	**X**	JF961388
N13		*Nanochromis*	*splendens*	Congo River	Luozi - south side of channel	**X**	**X**	JF961389
N14		*Nanochromis*	*splendens*	Congo River	Bulu - south side of channel	**X**	**X**	JF961390
N15		*Nanochromis*	*splendens*	Congo River	Luozi - south side of channel	**X**	**X**	JF961391
N16		*Nanochromis*	*splendens*	Congo River	Bulu - south side of channel	**X**	**X**	JF961392
N17		*Nanochromis*	*splendens*	Congo River	upstream Bulu rocky island	**X**	**X**	JF961393
N18		*Nanochromis*	*splendens*	Congo River	upstream Bulu rocky island	**X**	**X**	JF961394
N19		*Nanochromis*	*splendens*	Congo River	upstream Bulu rocky island	**X**	**X**	JF961395
N20	***N. splendens***	*Nanochromis*	*splendens*	Congo River	upstream Bulu rocky island	**X**	**X**	JF961396
N21		*Nanochromis*	*splendens*	Congo River	Bulu -north side of channel	**X**	**X**	JF961397
N22		*Nanochromis*	*splendens*	Congo River	Bulu -north side of channel	**X**	**X**	JF961398
N23		*Nanochromis*	*splendens*	Congo River	Bulu -north side of channel	**X**	**X**	JF961399
N24		*Nanochromis*	*splendens*	Congo River	Bulu -south side of channel	**X**	**X**	JF961400
N25		*Nanochromis*	*splendens*	Congo River	Bulu -south side of channel	**X**	**X**	JF961401
N26		*Nanochromis*	*splendens*	Congo River	Rapid before Tadi middle	**X**	**X**	JF961402
N27		*Nanochromis*	*splendens*	Congo River	Luozi/Bulu South	**X**	**X**	JF961403
N28		*Nanochromis*	*splendens*	Congo River	Luozi/Bulu South	**X**	**X**	JF961404
N29		*Nanochromis*	*splendens*	Congo River	Bulu - south side of channel	**X**	**X**	JF961405
N30		*Nanochromis*	*splendens*	Congo River	Kinganga	**X**	**X**	JF961406

		Specimen information				Molecular datasets		
Taxa nb.	**Clade**	**Genus**	**Species**	**River**	**Locality**	**AFLP dataset**	**mt- dataset**	**Gene Bank accession**
N31		*Nanochromis*	*splendens*	Congo River	Kinganga	**X**	**X**	JF961407
N32		*Nanochromis*	*splendens*	Congo River	Kinganga	**X**	**X**	JF961408
N33		*Nanochromis*	*splendens*	Congo River	Kinganga	**X**	**X**	JF961409
N34		*Nanochromis*	*splendens*	Congo River	Kinganga	**X**	**X**	JF961410
N35		*Nanochromis*	*splendens*	Congo River	Kinganga	**X**	-	
N36		*Nanochromis*	*splendens*	Congo River	Kinganga	-	**X**	JF961411
N37		*Nanochromis*	*splendens*	Congo River	Kinganga	-	**X**	JF961412
N38		*Nanochromis*	*splendens*	Congo River	Kinganga	-	**X**	JF961413
N39		*Nanochromis*	*consortus*	Congo River	Nziya_below Inga	**X**	**X**	JF961414
N40	***N. consortus***	*Nanochromis*	*consortus*	Congo River	Nziya_below Inga	**X**	**X**	JF961415
N41		*Nanochromis*	*consortus*	Congo River	Nziya_below Inga	**X**	**X**	JF961416
N42		*Nanochromis*	*consortus*	Congo River	Nziya_below Inga	**X**	**X**	JF961417
N43		*Nanochromis*	*consortus*	Congo River	Nziya_below Inga	**X**	**X**	JF961418
N44		*Nanochromis*	*consortus*	Congo River	Nziya_below Inga	**X**	**X**	JF961419
N45	***N. nudiceps***	*Nanochromis*	*nudiceps*	Fimi River		**X**	**X**	JF961420
N46		*Nanochromis*	*nudiceps*	Fimi River		**X**	**X**	JF961421
N47	***N. teugelsi***	*Nanochromis*	*teugelsi*	Kasai River	Kasai River	**X**	**X**	JF961422
N48		*Nanochromis*	*parilus*	Congo River	Maluku	**X**	**X**	JF961423
N49		*Nanochromis*	*parilus*	Congo River	Maluku	**X**	**X**	JF961424
N50		*Nanochromis*	*parilus*	Congo River	Maluku	**X**	**X**	JF961425
N51		*Nanochromis*	*parilus*	Congo River	Maluku	**X**	**X**	JF961426
N52		*Nanochromis*	*parilus*	Congo River	Kinsuka	**X**	**X**	JF961427
N53		*Nanochromis*	*parilus*	Congo River	Kinsuka	**X**	**X**	JF961428
N54		*Nanochromis*	*parilus*	Congo River	Kinsuka	**X**	**X**	JF961429
N55	***N. parilus***	*Nanochromis*	*parilus*	Congo River	Kinsuka	**X**	**X**	JF961430
N56		*Nanochromis*	*parilus*	Congo River	Les Rapides, Brazzaville	**X**	**X**	JF961431
N57		*Nanochromis*	*parilus*	Congo River	Les Rapides, Brazzaville	**X**	**X**	JF961432
N58		*Nanochromis*	*parilus*	Congo River	Mbouno below Brazza	-	**X**	JF961433
N59		*Nanochromis*	*parilus*	Congo River	Foulakari confluence	-	**X**	JF961434
N60		*Nanochromis*	*parilus*	Congo River	Foulakari confluence	-	**X**	JF961435
N61		*Nanochromis*	*parilus*	Congo River	Foulakari confluence	**X**	**X**	JF961436
N62		*Nanochromis*	*parilus*	Congo River	Mbelo	**X**	**X**	JF961437
N63		*Nanochromis*	*parilus*	Congo River	Mbelo	**X**	**X**	JF961438

		Specimen information				Molecular datasets		
Taxa nb.	**Clade**	**Genus**	**Species**	**River**	**Locality**	**AFLP dataset**	**mt- dataset**	**Gene Bank accession**
N64		*Nanochromis*	*parilus*	Congo River	Below Bela	**X**	**X**	JF961439
N65		*Nanochromis*	*parilus*	Congo River	Below Bela	**X**	**X**	JF961440
N66		*Nanochromis*	*parilus*	Congo River	Below Bela	**X**	**X**	JF961441
N67		*Nanochromis*	*parilus*	Congo River	Luozi village	**X**	**X**	JF961442
N68		*Nanochromis*	*parilus*	Congo River	Luozi village	**X**	**X**	JF961443
N69		*Nanochromis*	*parilus*	Congo River	upstream Luozi - north side	**X**	**X**	JF961444
N70		*Nanochromis*	*parilus*	Congo River	Bulu -north side of channel	**X**	**X**	JF961445
N71		*Nanochromis*	*parilus*	Congo River	upstream Bulu rocky island	**X**	**X**	JF961446
N72		*Nanochromis*	*parilus*	Congo River	upstream Bulu rocky island	**X**	**X**	JF961447
N73		*Nanochromis*	*parilus*	Congo River	Bulu - south side of channel	**X**	**X**	JF961448
N74		*Nanochromis*	*parilus*	Congo River	Bulu -north side of channel	**X**	**X**	JF961449
N75		*Nanochromis*	*parilus*	Congo River	downstream Bulu North	**X**	**X**	JF961450
N76		*Nanochromis*	*parilus*	Congo River	Kinganga	**X**	**X**	JF961451
N77		*Nanochromis*	*parilus*	Congo River	Kinganga	**X**	**X**	JF961452
N78		*Nanochromis*	*parilus*	Congo River	Bac Kinganga	**X**	**X**	JF961453
N79		*Nanochromis*	*parilus*	Congo River	Kinganga	**X**	**X**	JF961454
N80		*Nanochromis*	*parilus*	Congo River	Kinganga	**X**	**X**	JF961455
N81		*Nanochromis*	*parilus*	Congo River	Bac Kinganga	**X**	**X**	JF961456
N82		*Nanochromis*	*parilus*	Congo River	Kinganga	**X**	**X**	JF961457

Table S2 *Overview of samples and genes used for molecular clock dataset*

	Specimen information			mt- marker		nc-marker			
Taxa nb.	**Genus**	**Species**	**Locality**	**ND2**	**16S**	**ENC1**	**Ptr**	**S7**	**Sh3PX3**
J02	*Chilochromis*	*duponti*		GQ167776.1	GQ167902.1	GQ168279.1	GQ168028.1	GQ168090.1	GQ168216
J03	*Etia*	*nguti*		GQ167777.1	GQ167966.1	GQ168280.1	GQ168029.1	GQ168091.1	GQ168217
J04	*Gobiocichla*	*wonderi*		GQ167778.1	GQ167967.1	GQ168281.1	GQ168030.1	GQ168092.1	GQ168218
J05	*Heterochromis*	*multidens*		GQ167779.1	GQ167968.1	GQ168282.1	GQ168031.1	GQ168093.1	GQ168219
J07	*Oreochromis*	*niloticus*		GQ167780.1	GQ167969.1	GQ168283.1	GQ168032.1	GQ168094.1	GQ168220.1
J11	*Pelmatochromis*	*buettikoferi*		GQ167783.1	GQ167972.1	GQ168286.1	GQ168035.1	GQ168097.1	GQ168223.1
J12	*Pelmatochromis*	*nigrofasciatus*		GQ167784.1	GQ167973.1	GQ168287.1	GQ168036.1	GQ168098.1	GQ168224.1

	Specimen information			mt- marker		nc-marker			
Taxa nb.	Genus	Species	Locality	ND2	16S	ENC1	Ptr	S7	Sh3PX3
J13	*Pterochromis*	*congicus*		GQ167785.1	GQ167974.1	GQ168288.1	GQ168037.1	GQ168099.1	GQ168225.1
J14	*Sarotherodon*	*caudomarginatus*		GQ167786.1	GQ167975.1	GQ168289.1	GQ168038.1	GQ168100.1	GQ168226.1
J18	*Steatocranus*	*bleheri*		GQ167789.1	GQ167978.1	GQ168292.1	GQ168041.1	GQ168103.1	GQ168229.1
J19	*Steatocranus*	*casuarius*		GQ167790.1	GQ167979.1	GQ168293.1	GQ168042.1	GQ168104.1	GQ168230.1
J20	*Steatocranus*	*gibbiceps*		GQ167791.1	GQ167980.1	GQ168294.1	GQ168043.1	GQ168105.1	GQ168231.1
J21	*"Steatocranus"*	*irvinei*		GQ167792.1	GQ167981.1	GQ168295.1	GQ168044.1	GQ168106.1	GQ168232.1
J22	*Steatocranus*	sp. "bulky head"		GQ167793.1	GQ167919.1	GQ168296.1	GQ168045.1	GQ168107.1	GQ168233.1
J23	*Steatocranus*	sp. "dwarf"		GQ167794.1	GQ167983.1	GQ168297.1	GQ168046.1	GQ168108.1	GQ168234.1
J24	*Steatocranus*	*tinanti*	Kinsuka	GQ167795.1	GQ167984.1	GQ168298.1	GQ168047.1	GQ168109.1	GQ168235.1
J25	*Stomatepia*	*mariae*		GQ167796.1	GQ167985.1	GQ168299.1	GQ168048.1	GQ168110.1	GQ168236.1
J31	*Tilapia*	*discolor*		GQ167801.1	GQ167990.1	GQ168304.1	GQ168052.1	GQ168115.1	GQ168241.1
J32	*Tilapia*	*guinasana*		GQ167802.1	GQ167991.1	GQ168305.1	GQ168053.1	GQ168116.1	GQ168242.1
J33	*Tilapia*	*joka*		GQ167803.1	GQ167992.1	GQ168306.1	GQ168054.1	GQ168117.1	GQ168243.1
J38	*Oreochromis*	*andersonii*		GQ167805.1	GQ167994.1	GQ168308.1	GQ168056.1	GQ168119.1	GQ168245.1
J43	*Steatocranus*	sp. "red eye"		GQ167808.1	GQ167997.1	GQ168311.1	GQ168059.1	GQ168122.1	GQ168248.1
J46	*Tylochromis*	sp.		GQ167809.1	GQ167998.1	GQ168312.1	GQ168060.1	GQ168123.1	GQ168249.1
J49	*Sarotherodon*	*mvogoi*		GQ167811.1	GQ168000.1	GQ168314.1	GQ168062.1	GQ168125.1	GQ168251.1
J50	*Hemichromis*	*elongatus*		GQ167812.1	GQ168001.1	GQ168315.1	GQ168063.1	GQ168126.1	GQ168252.1
J55	*Steatocranus*	*glaber*	Nziya_below Inga	GQ167816.1	GQ168005.1	GQ168319.1	GQ168067.1	GQ168130.1	GQ168256.1
J66	*Tilapia*	*bilineata* "Salonga"		GQ167824.1	GQ168012.1	GQ168327.1	GQ168075.1	GQ168138.1	GQ168264.1
J68	*Steatocranus*	*ubanguiensis*	Ubangui River	GQ167826.1	GQ168014.1	GQ168329.1	GQ168077.1	GQ168140.1	GQ168266.1
J74	*Iranocichla*	*hormuzensis*		GQ167830.1	GQ168018.1	GQ168333.1	GQ168081.1	GQ168144.1	GQ168270.1
J76	*Bathybates*	*ferox*		GQ167832.1	GQ168020.1	GQ168335.1	GQ168083.1	GQ168146.1	GQ168272.1
J77	*Variabilichromis*	*moorii*		GQ167833.1	GQ168021.1	GQ168336.1	GQ168084.1	GQ168147.1	GQ168273.1
J80	*Cyprichromis*	*leptosoma*		GQ167835.1	GQ168023.1	GQ168338.1	GQ168086.1	GQ168149.1	GQ168275.1
S5	*Steatocranus*	*bleheri*	"Tshikapa" in RC	JF961464	JF961589	JF961622	JF961655	JF961721	JF961688
S10	*Steatocranus*	sp. bulky head	"Tshikapa" in RC	-	JF961590	JF961623	JF961656	JF961722	JF961689
S13	*Steatocranus*	sp. "Kisangani"	below falls in Kisangani	JF961469	JF961591	JF961624	JF961657	JF961723	JF961690
S16	*Steatocranus*	sp. "Lefini"	Lefini River	JF961470	JF961592	JF961625	JF961658	JF961724	JF961691
S17	*Steatocranus*	sp. "Nki"		JF961471	JF961593	JF961626	JF961659	JF961725	JF961692
S50	*Steatocranus*	sp. aff. *casuarius* "brown pearl"	Inga middle	-	JF961594	JF961627	JF961660	JF961726	JF961693
S55	*Steatocranus*	sp. "Maluku"	Maluku	JF961502	JF961595	JF961628	JF961661	JF961727	JF961694
S60	*Steatocranus*	*tinanti*	Kinsuka	JF961507	JF961596	JF961629	JF961662	JF961728	JF961695

		Specimen information		mt- marker		nc-marker			
Taxa nb.	**Genus**	**Species**	**Locality**	**ND2**	**16S**	**ENC1**	**Ptr**	**S7**	**Sh3PX3**
S69	*Steatocranus*	sp. aff. *tinanti "ultraslender"*	Luozi - south side of channel	JF961516	JF961597	JF961630	JF961663	JF961729	JF961696
S85	*Steatocranus*	sp. aff. *tinanti "Inga"*	Inga around the corner, north - Point 50	JF961527	JF961598	JF961631	JF961664	JF961730	JF961697
S92	*Steatocranus*	sp. aff. *tinanti "intermediate"*	Yalala rapid II	JF961533	JF961599	JF961632	JF961665	JF961731	JF961698
S95	*Steatocranus*	*mpozoensis*	Matadi	JF961536	JF961600	JF961633	JF961666	JF961732	JF961699
S101	*Steatocranus*	*mpozoensis*	Boma Pirogue	JF961542	JF961601	JF961634	JF961667	JF961733	JF961700
S104	*Steatocranus*	cf. *gibbiceps*	Kinsuka	JF961545	JF961602	JF961635	JF961668	JF961734	JF961701
S139	*Steatocranus*	cf. *gibbiceps*	Inga around the corner, north - Point 50	JF961576	JF961603	JF961636	JF961669	JF961735	JF961702
N83	*Congochromis*	cf. *sabinae*	Yambula-Bakere	JF961458	JF961617	JF961650	JF961683	JF961749	JF961716
N84	*Congochromis*	spec.	Amba River	JF961459	JF961618	JF961651	JF961684	JF961750	JF961717
N85	*Congochromis*	spec.	Monkoto, Luilaka River	JF961460	JF961619	JF961652	JF961685	JF961751	JF961718
N4	*Congochromis*	spec.	Lilanda	JF961382	JF961604	JF961637	JF961670	JF961736	JF961703
N5	*Nanochromis*	*sabinae*		-	JF961605	JF961638	JF961671	JF961737	JF961704
N6	*Nanochromis*	*cf sabinae*	Lui Kotale	-	JF961606	JF961639	JF961672	JF961738	JF961705
N86	*Nanochromis*	*splendens*	Kinganga	JF961461	JF961620	JF961653	JF961686	JF961752	JF961719
N87	*Nanochromis*	*consortus*	Nziya	JF961462	JF961621	JF961654	JF961687	JF961753	JF961720
N7	*Nanochromis*	*minor*	Kinganga	JF961383	JF961607	JF961640	JF961673	JF961739	JF961706
N8	*Nanochromis*	*wickleri*	Xeno	JF961384	JF961608	JF961641	JF961674	JF961740	JF961707
N9	*Nanochromis*	*transvestitus*	Inongo	JF961385	JF961609	JF961642	JF961675	JF961741	JF961708
N10	*Nanochromis*	sp. Ndongo	Ngoko River Cameroon	JF961386	JF961610	JF961643	JF961676	JF961742	JF961709
N26	*Nanochromis*	*splendens*	Rapid before Tadi middle	JF961402	JF961611	JF961644	JF961677	JF961743	JF961710
N45	*Nanochromis*	nudiceps	Fimi River	JF961420	JF961612	JF961645	JF961678	JF961744	JF961711
N47	*Nanochromis*	*teugelsi*	Kasai River	JF961422	JF961613	JF961646	JF961679	JF961745	JF961712
N54	*Nanochromis*	*parilus*	Kinsuka	JF961429	JF961614	JF961647	JF961680	JF961746	JF961713
N60	*Nanochromis*	*parilus*	Foulakari confluence	JF961435	JF961615	JF961648	JF961681	JF961747	JF961714
N74	*Nanochromis*	*parilus*	Bulu -north side of channel	JF961449	JF961616	JF961649	JF961682	JF961748	JF961715

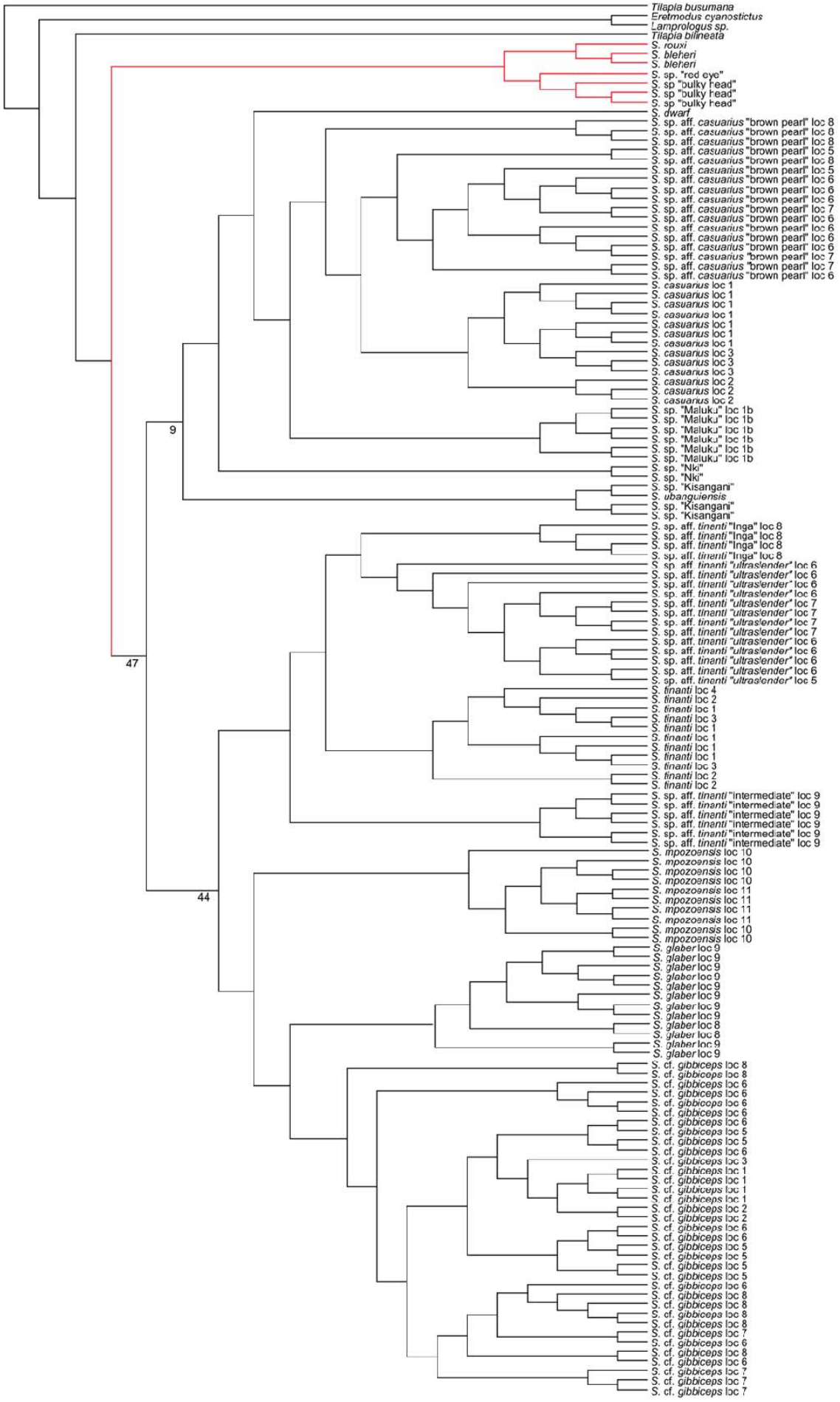


Figure S1 *(caption overleaf)*

Figure S1 *Branch attachment frequency*

Alternative positions of the unstable Southern clade in 1000 bootstrap topologies. The numbers, plotted on the neighbour joining tree (based on the AFLP dataset), indicate fractions of bootstrap trees in which alternative branching patterns occur.

Chapter 6: Speciation within genomic networks: A case study based on *Steatocranus* cichlids from the Congo basin

Table S1 *Overview of bootstrap support values in 45 removal experiments*

Support values for all major clades obtained in 45 removal experiments. Only nodes that were affected by removals are shown. The X stands for non existent nodes and red numbers indicate nodes that are not present in the consensus tree. Species abbrveations: SG = *S*. cf. *gibbiceps*, SGL = *S. glaber*, SM = *S. mpozoensis*, ST = *S*. cf. *tinanti*, SC = *S*. cf. *casuarius*, GGMT = *S*. cf. *gibbiceps*/*S. glaber*/ *S. mpozoensis*/ *S*. cf. *tinanti* clade, NT = Northern tributaries, KIS= *S*. sp. "Kisangani", UB = *S. ubanguisensis*, NKI = *S.* sp. "Nki", DW = *S*. sp. "dwarf", MAL = *S*. sp "Maluku", CMDN = *S*. cf. *casuarius*/ *S*. sp "Maluku"/ *S*. sp. "dwarf"/ Nothern tributaries clade, STR = Southern tributaries, BL = *S. bleheri*, RO = *S. rouxi*, RE = *S*. sp. "red eye", BU = *S*. sp "bulky head".

Clade		SGL all	SM all	ST loc 5-7	ST loc 5-8	ST loc 8	SG/SGL	SGL/SM	NT	SC/MAL	SC loc 2	SC loc 3	SC loc 1/2/3	SC loc 1-7	SC loc 5-7	SC loc 5-8	SC loc 8	RO/RE/BU	RE/BU	IN wt STR	STR/GGMT	STR/CMDN
	Nb	7	8	13	14	15	17	17b	21	28	31	32	33	34	35	35b	37	41	42	LCN	LCN b	LCN c
SG all	1	87	87	91	92	66	X	X	22	93	93	85	98	4	50	94	79	99	65	46	3	45
SG loc 1	2	92	98	91	94	87	88	6	52	64	73	78	98	27	58	33	63	75	75	44	44	12
SG loc 2	3	93	98	91	95	89	90	8	48	67	71	75	97	29	61	32	63	74	78	44	43	13
SG loc 1/2	4	93	99	90	93	88	90	5	51	64	71	73	98	30	60	32	61	75	76	42	45	13
SG loc 1/2/3	5	92	99	91	94	88	89	5	45	63	74	77	97	31	57	37	62	76	75	45	46	9
SG loc 5-8	6	84	87	90	92	69	12	81	28	94	92	81	98	5	53	93	79	99	72	50	6	41
SGL all	7	X	98	91	94	89	X	X	47	67	74	77	97	29	60	38	61	77	78	43	47	10
SM all	8	96	X	91	94	87	X	X	52	65	74	78	97	28	59	30	64	74	76	43	48	9
SM loc 11	9	93	X	88	94	86	82	11	50	63	77	78	97	30	61	28	65	69	75	48	43	9
ST all	10	80	100	X	X	X	80	3	56	61	72	74	98	22	63	26	62	74	69	54	39	7
ST loc 1-4	11	94	99	89	X	85	90	7	50	63	72	79	96	28	61	35	61	74	73	53	40	7
ST loc 1-8	12	92	98	X	X	X	90	6	52	59	76	76	98	26	58	42	63	76	66	49	41	10
ST loc 5-7	13	95	99	X	X	X	91	5	47	59	72	77	97	30	63	45	64	76	71	52	40	8
ST loc 5-8	14	90	98	X	X	X	84	7	47	60	74	77	97	23	58	36	60	78	72	49	40	11
ST loc 8	15	95	98	X	X	X	91	7	49	63	73	76	96	28	57	36	58	73	73	49	40	11
ST loc 9	16	95	98	91	95	88	88	7	51	69	75	79	97	28	61	29	66	70	75	50	41	9
SG/SGL	17	X	100	90	87	56	X	X	26	95	95	82	99	4	52	95	77	100	70	46	0	48

Clade		SGL all	SM all	ST loc 5-7	ST loc 5-8	ST loc 8	SG/SGL	SGL/SM	NT	SC/MAL	SC loc 2	SC loc 3	SC loc 1/2/3	SC loc 1-7	SC loc 5-7	SC loc 5-8	SC loc 8	RO/RE/BU	RE/BU	IN wt STR	STR/GGMT	STR/CMDN
	Nb	7	8	13	14	15	17	17b	21	28	31	32	33	34	35	35b	37	41	42	LCN	LCN b	LCN c
SGL/SM	18	X	X	91	91	85	X	X	43	72	81	77	98	22	59	40	66	81	79	47	39	14
SG/SGL/SM	19	X	X	87	87	51	X	X	58	95	94	83	98	3	21	95	78	100	70	47	0	48
GGMT	20	X	X	X	X	X	X	X	5	98	93	80	98	8	41	95	81	100	71	X	X	95
NT	21	92	99	89	94	86	87	9	X	67	75	75	98	26	60	32	63	75	81	71	11	18
KIS/UB	22	93	98	91	94	86	89	7	X	66	74	73	96	30	62	31	63	71	73	43	44	12
KIS	23	94	97	90	94	84	90	5	75	65	74	78	96	29	59	31	62	73	75	48	43	9
UB	24	94	98	91	87	87	89	8	69	66	71	76	97	27	59	38	59	72	77	49	40	11
NKI	25	92	98	90	95	87	88	8	X	61	73	76	97	30	64	31	62	73	75	61	26	13
DW/MAL/SC	26	91	98	91	91	87	86	6	X	X	X	X	X	X	X	X	X	75	75	27	66	6
DW	27	95	98	91	92	87	91	6	62	X	70	75	97	28	60	25	62	75	76	67	16	17
SC/MAL	28	93	98	93	95	89	89	8	26	X	X	X	X	X	X	X	X	73	80	19	77	3
MAL	29	94	99	90	94	87	88	6	45	X	72	72	97	28	62	30	64	73	77	42	50	7
SC all	30	93	91	91	93	89	89	5	41	X	X	X	X	X	X	X	X	74	78	39	52	9
SC loc 2	31	95	98	89	95	84	90	5	50	64	X	82	97	34	60	25	65	70	73	46	45	9
SC loc 3	32	94	97	90	96	86	89	7	46	71	96	X	97	28	65	47	66	70	75	45	43	12
SC loc 1/2/3	33	94	98	90	94	88	90	7	54	46	X	X	X	X	66	X	59	75	78	43	48	9
SC loc 1-7	34	94	98	90	96	88	91	6	58	X	X	X	X	X	X	X	X	75	79	52	39	9
SC loc 5-7	35	91	98	88	94	85	88	6	47	68	70	74	100	X	X	X	97	73	76	44	46	10
SC loc 5-8	36	92	99	90	95	88	89	6	44	90	68	78	X	X	X	X	X	75	77	45	45	10
SC loc 8	37	94	98	90	94	87	88	7	45	80	69	76	97	X	87	X	X	75	78	45	45	10
CMDN	38	92	99	92	92	88	89	6	X	X	X	X	X	X	X	X	X	73	75	X	X	X
STR	39	95	99	89	94	86	89	7	48	65	74	77	97	28	63	34	65	X	X	X	X	X
BL	40	94	99	90	93	87	88	6	52	64	72	79	95	30	60	38	62	X	84	68	32	0
RO/RE/BU	41	94	99	90	96	86	89	8	48	63	70	77	97	28	64	34	64	X	X	9	23	68
RE/BU	42	93	97	90	94	88	88	7	46	65	73	73	96	25	60	33	59	X	X	66	26	8
BU	43	94	99	89	94	86	87	7	49	63	75	78	97	31	65	26	64	97	X	42	48	10
RE	44	94	98	90	96	85	90	8	49	64	69	76	96	33	59	33	61	27	X	65	31	4
RO	45	93	99	91	93	88	88	6	49	66	72	75	96	33	59	34	61	X	87	31	57	12
BS		**94**	**98**	**89**	**95**	**86**	**90**	**6**	**51**	**65**	**73**	**76**	**96**	**29**	**59**	**30**	**62**	**75**	**77**	**47**	**44**	**0**
Mean		93	98	90	93	84	86	9	48	69	76	77	97	25	59	42	66	77	75	47	38	16
Median		93	98	90	94	87	89	7	49	65	74	77	97	28	60	34	63	75	75	46	42	10
25% Quartile		92	98	90	93	86	88	6	46	63	72	75	97	26	58	31	61	73	73	43	34	9
75% Quartile		94	99	91	95	88	90	8	52	69	75	78	98	30	62	39	65	76	78	50	46	13

Clade		SGL all	SM all	ST loc 5-7	ST loc 5-8	ST loc 8	SG/SGL	SGL/SM	NT	SC/MAL	SC loc 2	SC loc 3	SC loc 1/2/3	SC loc 1-7	SC loc 5-7	SC loc 5-8	SC loc 8	RO/RE/BU	RE/BU	IN wt STR	STR/GGMT	STR/CMDN
	Nb	7	8	13	14	15	17	17b	21	28	31	32	33	34	35	35b	37	41	42	LCN	LCN b	LCN c
Min		80	87	87	87	51	12	3	5	46	68	72	95	3	21	25	58	27	65	9	0	0
Max		96	100	93	96	89	91	81	75	98	96	85	100	34	87	95	97	100	87	71	77	95

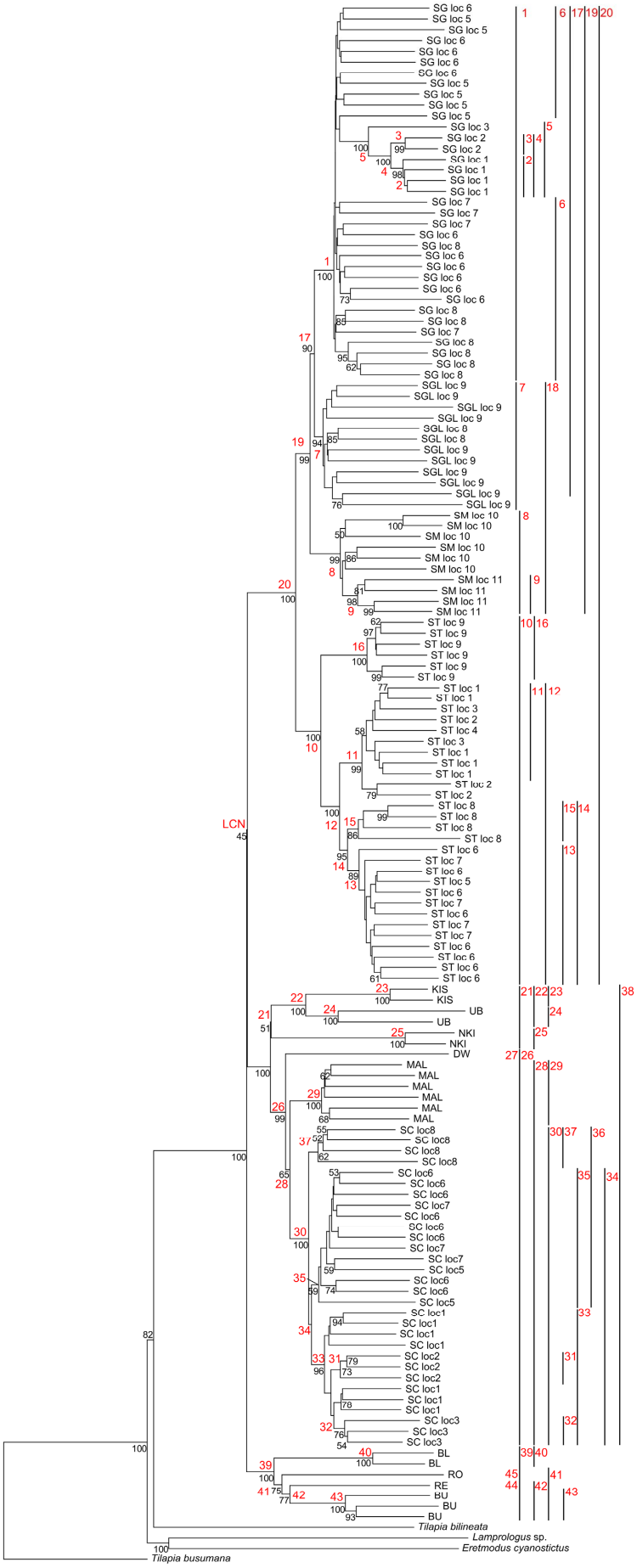


Figure S1 *(caption overleaf)*

Figure S1 *Overview of removal experiments of the homoplasy excess test*

Numbers on the tree correspond to those given in Table S1. LCN = lower Congo River *Steatocranus* & Northern tributaries, STEATO = *Steatocranus.*

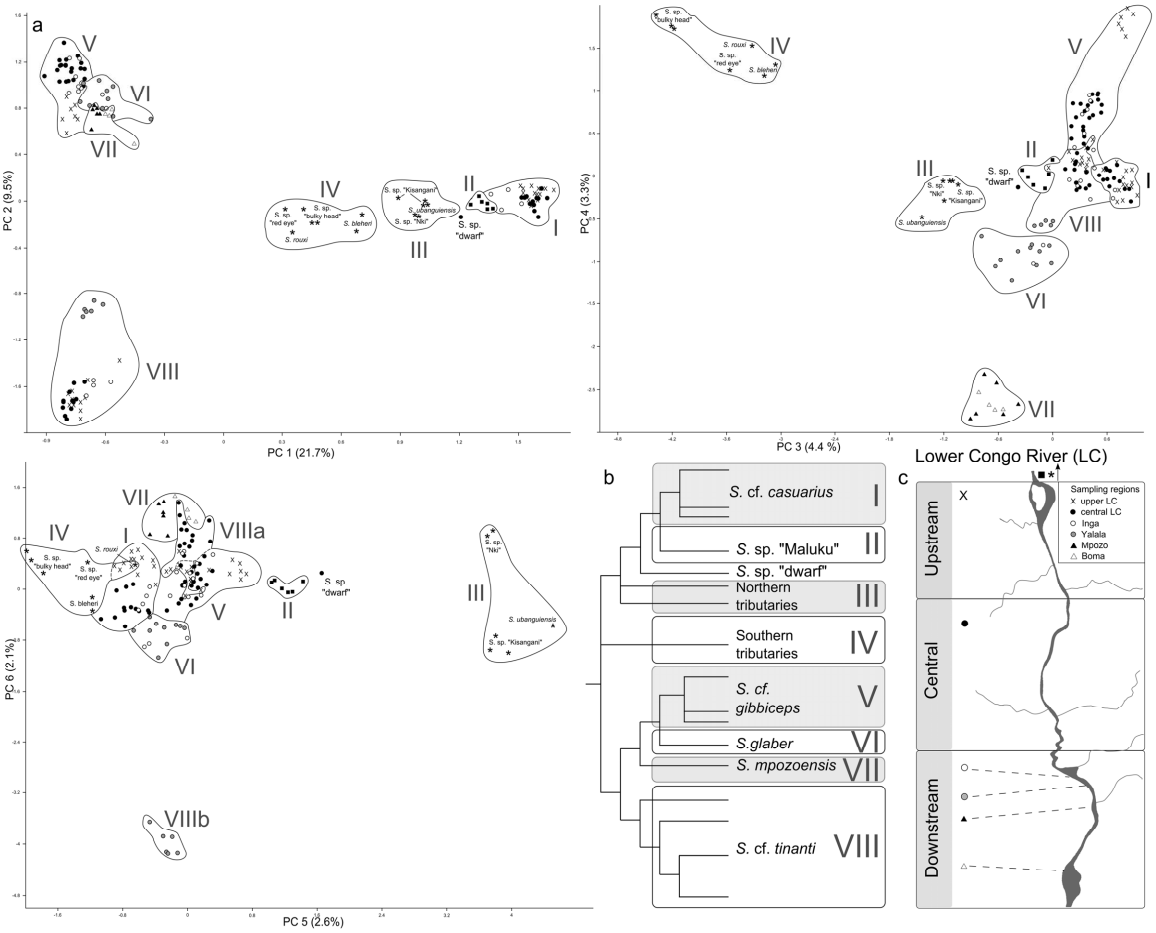


Figure S2 *Principal Components analysis based on AFLP data*

Results for the first six axes of the Principal components analysis (with more than 2% explanatory power). (a) Roman numerals correspond to phylogenetic clades indicated in (b) and symbols to the sampling locations where the individual was collected, given in (c).

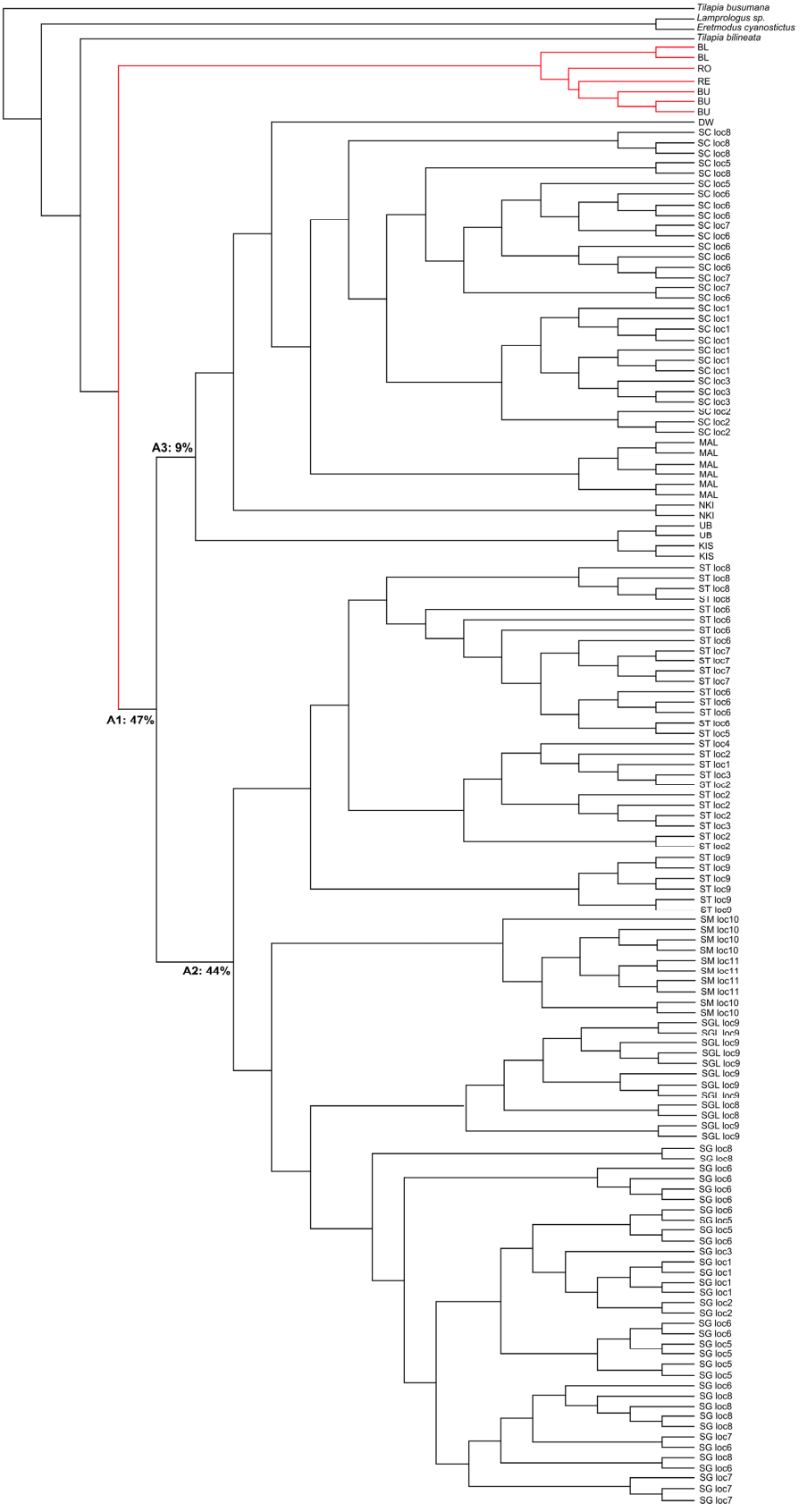


Figure S3 *(caption overleaf)*

Figure S3 *Branch attachment frequencies concerning the Southern tributary clade*

Alternative positions of the Southern tributary clade in 1000 bootstrap topologies. The numbers, plotted on the ML tree, indicate fractions of bootstrap trees (in percent) in which alternative branching patterns occur.

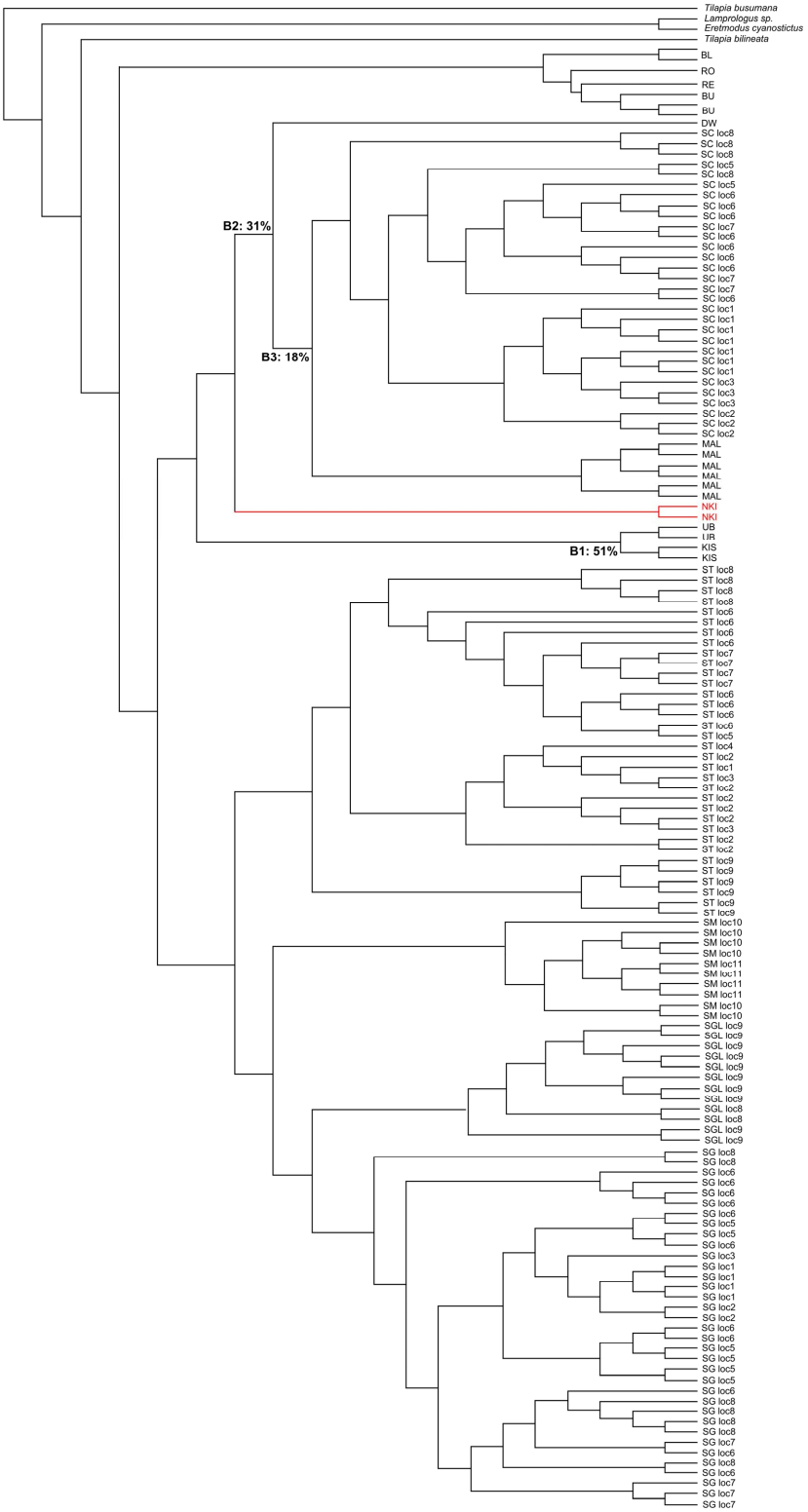


Figure S4 *(caption overleaf)*

Figure S4 *Branch attachment frequencies for the placement of Steatocranus sp. "Nki" from the Northern tributaries*

Alternative positions of the S. sp "Nki" in 1000 bootstrap topologies. The numbers, plotted on the ML tree, indicate fractions of bootstrap trees (in percent) in which alternative branching patterns occur.

Erklärung

Ich versichere, dass ich diese Arbeit selbständig verfasst, keine anderen Quellen und Hilfsmittel als die angegebenen benutzt und die Stellen der Arbeit, die anderen Werken dem Wortlaut oder Sinn nach entnommen sind, kenntlich gemacht habe. Diese Arbeit hat in dieser oder ähnlichen Form keiner anderen Prüfungsbehörde vorgelegen.

Bonn, den 18.08.2011