THE EVOLUTION OF PHENOTYPIC DIVERSITY IN A LAKE TANGANYIKA CICHLID FISH

DISSERTATION

Zur Erlangung des akademischen Grades
Dr. rer. nat.

an der Naturwissenschaftlichen Fakultät der
Karl-Franzens-Universität Graz
Institut für Zoologie

vorgelegt von

Mag. Karin Mattersdorfer

Unter der Betreuung von
Univ.-Ass. Dr. Kristina Sefc

Graz, September 2011

Many thanks to

Kristina Sefc for the ambitious supervision and the opportunity to be part of a challenging and interesting scientific project.

Stephan Koblmüller for advice and support and for answering plenty of questions.

Caroline Hermann for being so much more than a colleague.

My parents and the rest of my family for their guidance and support in all aspects of life.

Johannes for believing in me, looking after me, cheering me up, and simply for being with me.

Walter Salzburger for providing me with primers for SNP analyses.
Britta Meyer and Bernd Egger for advice concerning the SNP and AFLP analyses, respectively.
Christian Sturmbauer and Steven Weiss for useful comments and advice.

Philipp for solving technical problems, for making me laugh, and for chocolate, apples, and coca cola.

All colleagues and friends from the institute, especially Selma, Wolfi, Lisbeth, Miriam, and Kathi, who made this a great place to work.

All my friends for sharing ups and downs with me not only in the last three years.

Contents

ABSTRACT

Lake Tanganyika's cichlid genus *Tropheus*, with over 100 allo- or parapatrically distributed color morphs, represents an outstanding example for phenotypic diversity and thus serves as a promising system for studies of color pattern evolution. For Lake Malawi and Victoria cichlids sexual selection has been shown to drive the evolution of novel color patterns. This has also been discussed for *Tropheus*, but without conclusive evidence. Here the role of sexual selection is hard to asses as *Tropheus* lack some of the traits typically associated with sexually selected species. Another force that has gained increasing attention in potentially creating new phenotypes is hybridization. In *Tropheus* the opportunity for it could have been provided when due to recurrent lake level fluctuations formally allopatric populations came into secondary contact. However, the mechanisms behind phenotypic variability in *Tropheus* remain largely unclear. In fact nothing is known about the underlying genetics and only assumptions can be made regarding the forces triggering it.

A genome scan, carried out on three *T. moorii* color morphs in the present study, as a first step towards the characterization of the genetic basis underlying coloration, revealed signs of selection acting on a few genes or linked loci that likely triggered the rapid evolution of different color patterns. An analysis of body color through photos and pigment extraction revealed general differences between the three morphs.

To assess the role of hybridization in promoting color diversification in *Tropheus* two contact zones were investigated for signs of admixture. Although the two scenarios differed in the circumstances and consequences of the hybridization event, in both cases genetic admixture was confirmed. This led to the conclusion that Lake Tanganyika's water level fluctuations likely promoted the phenotypic diversification through hybridization between formally separated morphs in the course of secondary contact.

ZUSAMMENFASSUNG

Die im Tanganjikasee vorkommende Cichlidengattung *Tropheus* ist, mit über 100 allo- oder parapatrischen Farbmorphen, ein besonderes Beispiel für phänotypische Diversität, und stellt daher ein viel versprechendes System für Studien über die Evolution von Farbmustern dar. Für Cichliden des Malawi- und Viktoriasees wurde gezeigt, dass sexuelle Selektion die Evolution neuer Farbmuster antreibt. Dies wurde auch für *Tropheus* diskutiert, jedoch bis jetzt ohne eindeutige Beweise. Die Bedeutung sexueller Selektion ist hier schwer zu erfassen, da *Tropheus* einige der für sexuell selektierte Arten typischen Merkmale nicht besitzt. Eine weitere Triebkraft, die immer mehr Aufmerksamkeit wegen ihres Potentials neue Phänotypen zu kreieren erlangt hat, ist Hybridisierung. In *Tropheus* könnte diese im Zuge von sekundärem Kontakt, herbeigeführt durch Seespiegelschwankungen, stattgefunden haben. Die genauen Mechanismen, die hinter der phänotypischen Variabilität in *Tropheus* stehen sind dennoch ungeklärt. Es ist nichts über die zugrundeliegende Genetik bekannt, und bezüglich der treibenden Kräfte kann nur spekuliert werden.

Mit einem Genome Scan, hier ausgeführt an drei *T. moorii* Farbmorphen, als erster Schritt zur Charakterisierung der genetischen Basis der Färbung, konnten Loci unter Selektion detektiert werden. Diese haben wahrscheinlich die schnelle Evolution verschiedener Farbmuster gesteuert. Eine Analyse der Farbe mittels Fotos und Pigmentextraktion zeigte generelle Unterschiede zwischen den drei Morphen auf.

Um die Rolle von Hybridisierung bei der Diversifizierung von Farbe zu erfassen, wurden zwei Kontaktzonen auf Anzeichen von Introgression untersucht. Die Szenarien unterschieden sich zwar in den Umstände und Konsequenzen der Hybridisierung, doch wurde die genetische Vermischung in beiden Fällen bestätigt. Daraus kann man folgern, dass Seespiegelschwankungen phänotypische Diversifizierung durch Hybridisierung zwischen einst getrennten Morphen sehr wahrscheinlich förderten.

CHAPTER 1

General Introduction

Tropheus moorii

Cichlids - model organisms for evolutionary research

With more than 3 000 species the Cichlid fish family (order: Perciformes) is the most species-rich vertebrate family existent. Cichlids occur in various regions ranging from Central and South America across Africa to Madagascar and southern India. Over their whole distribution range they have been demonstrated to have a high potential for rapid radiation (e.g. McKaye et al. 2002) and even sympatric speciation has been shown (e.g. Schliewen et al. 1994, 2006; Barluenga et al. 2006) but it is the species flocks of the East African Rift Valley Lakes that outstand most. This can be attributed to the fact that in Lakes Tanganyika, Malawi and Victoria, the three biggest of the Rift Valley Lakes, about 2 000 cichlid species have evolved within a short period of time. Alone in the past 5 Million years between 1 000 and 2 000 speciation events have occurred (Seehausen 2006). Thus the cichlid fish radiations in the Great Lakes of East Africa are the most diverse extant animal radiations and this makes them a very special model system to investigate theories of speciation and adaptive radiations (e.g. Kocher 2004; Seehausen 2006).

Cichlids in general, but especially East African cichlids are known for their exceptional phenotypic diversity, including various body shapes, color patterns, and morphological characters, triggered by the specialization to different trophic or ecological niches (e.g. Fryer & Iles 1972; Stiassny & Meyer 1999; Schluter 2000; Kocher 2004; Seehausen 2006). A lot of research focused on the preconditions that allowed the explosive adaptive radiation of the cichlid fishes and several factors, including intrinsic characters and extrinsic factors, are believed to have triggered the outstanding diversity: Cichlids are e.g. highly tolerant for diluted minerals, which are a component of the lake water (Poll 1986). However, the so called key innovation is a specialized anatomy, i.e. the existence of a second set of jaws: Apart from the oral jaws they possess pharyngeal jaws which derived from the 5th gill arch and serve as a tool to crush, macerate, slice or pierce the food before the ingestion (Salzburger & Meyer 2004). This anatomical character which is shared by all cichlids (Liem 1973, 1980) and some other closely related fish families is very versatile and only slight modifications in these jaws and their teeth provide the opportunity for the accession of new food sources and thus for the rapid occupation of new ecological niches (e.g. Stiassny & Jensen 1987; Stiassny 1991; Sturmbauer 2000).

The second key factor that is believed to have contributed to the evolutionary success of the East African cichlids is the existence of a highly specialized breeding behavior including various types of brood care ranging from substrate brooding to highly specialized

mouthbrooding (e.g. Fryer & Iles 1972; Goodwin et al. 1998; Stiassny & Meyer 1999; Klett & Meyer 2002). Moreover, for some cichlid species of Lake Malawi and Lake Victoria color pattern diversification has been discussed or even shown to be triggered by inter- and/or intrasexual selection (e.g. Seehausen & van Alphen 1999; Seehausen et al. 1999; Knight & Turner 2004, Pauers et al. 2004; Maan et al. 2004; Dijkstra et al. 2007). However, in Lake Tanganyika, the oldest of the lakes, this correlation has not been unambiguously demonstrated.

On the extrinsic level geologic and environmental events, like lake level fluctuations, need to be mentioned as they led to the creation of novel niches and thus certainly triggered the explosive speciation and adaptive radiation in East African cichlids (e.g. Fryer & Iles 1972; Rossiter 1995; Sturmbauer 1998). After all, it is the interplay between intrinsic and extrinsic factors that promotes a radiation as extrinsic factors provide the opportunity and intrinsic factors the potential for it (Martens et al. 1994; Sturmbauer 1998; Koblmüller et al. 2008).

Lake Tanganyika and its cichlids

Lake Tanganyika, one of the Great African Rift Valley Lakes, is listed among the deepest and oldest inland waters existent (e.g. Sturmbauer 2000). Despite of the fact that Lake Tanganyika is much older (the estimated age is nine to 12 million years; Cohen et al. 1993, 1997) than Lakes Malawi and Victoria, it comprises the smallest number of endemic cichlids, compared to the other two Great Lakes, with the expected number being ~ 250 species (Snoeks 2000; Turner et al. 2001). However, in terms of morphological, ecological and behavioral diversity, Lake Tanganyika's cichlids easily surpass their relatives within Lakes Malawi and Victoria (Rossiter 1995; Salzburger et al. 2002a). Sturmbauer & Meyer (1992) e.g. stated that a single Lake Tanganyika cichlid genus exhibits about twice as much genetic variation as the total cichlid flock of Lake Malawi.

In the last years several studies have been carried out on Lake Tanganyika's cichlid species assemblage in order to shed light on this extraordinary radiation and the taxonomic status of its fishes. Poll (1986) and Takahashi (2003) revealed 12 and 16 tribes, respectively, based on morphological data. With the use of molecular genetic techniques the calculation of fine-scale phylogenies became possible thus enabling big progress concerning the characterization of the adaptive evolution and speciation within these species flocks. Although some disagreement existed, molecular studies largely supported the morphological

classification (Koblmüller et al. 2008). Summarized, the use of molecular markers revealed that Lake Tanganyika's species flock has a polyphyletic origin as it was colonized by several lineages independently 5 - 6 million ya (e.g. Nishida 1991; Takahashi et al. 2001; Salzburger et al. 2002a, b; Koblmüller et al. 2008), at a time that is characterized with the establishment of a true lacustrine habitat with deep-water conditions (Cohen et al. 1993, 1997). Seven seeding lineages which are thought to be the source of the 'primary lacustrine radiation' were identified. These were the ancestors of *Boulengerochromis microlepis*, of the genus *Hemibates* (Koblmüller et al. 2005), and of the Trematocarini, Bathybatini, Eretmodini, and Lamprologini. The seventh seeding lineage represented the ancestors of the so called C-lineage (including Tropheini, Haplochromini, Orthochromini, Cyprichromini, Ectodini, Perissodini, Cyphotilapini, and Limnochromini; Clabaut et al. 2005). A 'secondary radiation' was dated to roughly 2.5 - 3 million ya (Koblmüller et al. 2004, 2005, 2008; Duftner et al. 2005; Seehausen 2006) and the diversification of the Perissodini, one of the mouthbrooding tribes, was dated slightly later (1.5 - 2 million ya; Koblmüller et al. 2007). In a recent review by Koblmüller et al. (2008) evidence from several studies was combined and the Lake Tanganyika species assemblage was illustrated in a schematic molecular phylogeny (Fig. 1).

Tropheus - One of Lake Tanganyika's most popular cichlids

This thesis deals with a very special member of the Lake Tanganyika cichlid species assemblage: The genus *Tropheus* which belongs to the tribe Tropheini that in turn is phylogenetically nested within the Haplochromini (Salzburger et al. 2005; Koblmüller et al. 2008; Koblmüller et al. 2010; Fig. 1). *Tropheus* occurs with high abundance in the rocky habitats of the upper littoral zone. Rocks provide members of this genus with food (they feed on epilithic algae) and shelter from predators and thus sandy bottom habitat is strictly avoided. Several genetic and ecological studies strongly suggested that *Tropheus* is not able to disperse over large distances, unsuitable habitat, and open water due to strong habitat specificity and site fidelity (e.g. Brichard 1978; Yanagisawa & Nishida 1991; Sturmbauer & Dallinger 1995; Baric et al. 2003; Sturmbauer et al. 2005; Egger et al. 2007; Sefc et al. 2007; Koblmüller et al. 2011) thus leading to a considerable degree of genetic differentiation even between adjacent populations.

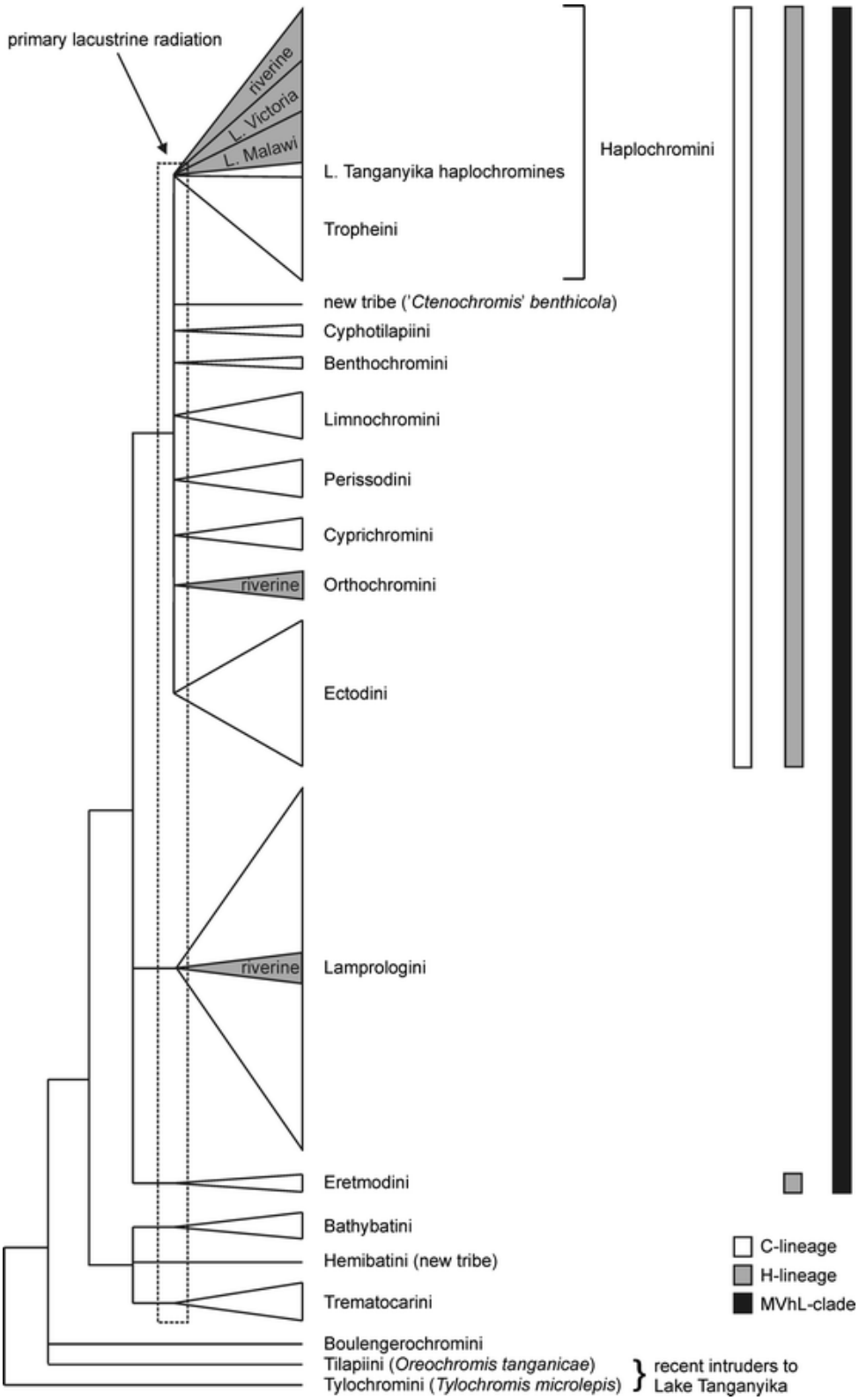


Figure 1 Schematic molecular phylogeny of Lake Tanganyika's cichlids. Triangles indicate lineages that underwent radiation. The size of the triangles corresponds to the number of species within the lineage (except for the species-rich non-Tanganyika haplochromines). Species that do not occur in Lake Tanganyika are represented through grey triangles. It is assumed that the 'primary lacustrine radiation' took place 5 - 6 million ya coinciding with the establishment of a true lacustrine habitat with deep-water conditions. Bars on the right indicate taxonomic groups that have been proposed based on combined mitochondrial and nuclear DNA sequence data (C-lineage; Clabaut et al. 2005), allozyme data (H-lineage; Nishida 1991) and SINEs (MVhL-lineage; Takahashi et al. 2001). Modified from Koblmüller et al. 2007.

Sexually monomorphic *Tropheus* are maternal mouthbrooders and display a complex social system with males and females defending territories of their own (e.g. Schürch & Taborsky 2005). For a period of stable pair-bonding before spawning, however, females are allowed to feed in a male's territory (Yanagisawa & Nishida 1991) and clutches are sired by a single male each (Egger et al. 2006).

The most conspicuous character of this genus is its outstanding wealth of phenotypic diversity in terms of body coloration with over 100 mostly allo- and parapatrically distributed color variants occurring all around Lake Tanganyika (Konings 1998; Schupke 2003; Fig. 2). Even on small geographic scales *Tropheus* display a considerable high level of color diversity (e.g. Koblmüller et al. 2011). The taxonomic status of the members of the *Tropheus* genus is still largely unclear and previous classification attempts do not reflect the existing variation within the genus. Currently, based on morphological data, six nominal species are described which are *Tropheus moorii* BOULENGER 1898, *T. annectens* BOULENGER 1900, *T. duboisi* MARLIER 1959, *T. brichardi* NELISSEN and THYS VAN DEN AUDENAERDE 1975, *T. kasabae* NELISSEN 1977, and *T. polli* AXELROD 1977. However, this classification was done using individuals from only few localities and without information on the outstanding number of color variants existent and several populations from different lake regions cannot be unambiguously assigned to these six species (Snoeks et al. 1994). Furthermore, *T. duboisi* is the only one of these species that is a distinct phenotypic and genetic entity (Snoeks et al. 1994; Sturmbauer et al. 2005). In an attempt to improve the taxonomic resolution of the genus Konings (1998) suggested five additional taxa (*T.* sp. 'black', *T.* sp. 'red', *T.* sp. 'Ikola', and *T.* sp. 'Mpimbwe'). Schupke (2003) on the other hand suggested 13 *Tropheus* color lineages comprising over 100 populations. However, these classifications have not yet been totally validated by scientific studies. To handle the taxonomic uncertainties it is meanwhile well-established that populations are identified by their locality instead of their species (e.g. Baric et al. 2003; Sturmbauer et al. 2005; Sefc et al. 2007; Egger et al. 2007; Koblmüller et al. 2011). A handful of studies exist on the phylogenetic and phylogeographic status of *Tropheus*. Most of them used mitochondrial sequence data and revealed that the distribution of lineages is very complex (Sturmbauer & Meyer 1992; Baric et al. 2003, Sturmbauer et al. 2005). The existence of 12 mitochondrial lineages was suggested by Sturmbauer et al. (2005) which are thought to have evolved in the course of three major radiation events triggered by severe lake level fluctuations (Baric et al. 2003; Sturmbauer et al. 2001; 2005). An additional more recent fluctuation, linked to a drop of the water level in the late Pleistocene (~20 000 Ya; e.g. Scholz et al. 2003; McGlue et al.

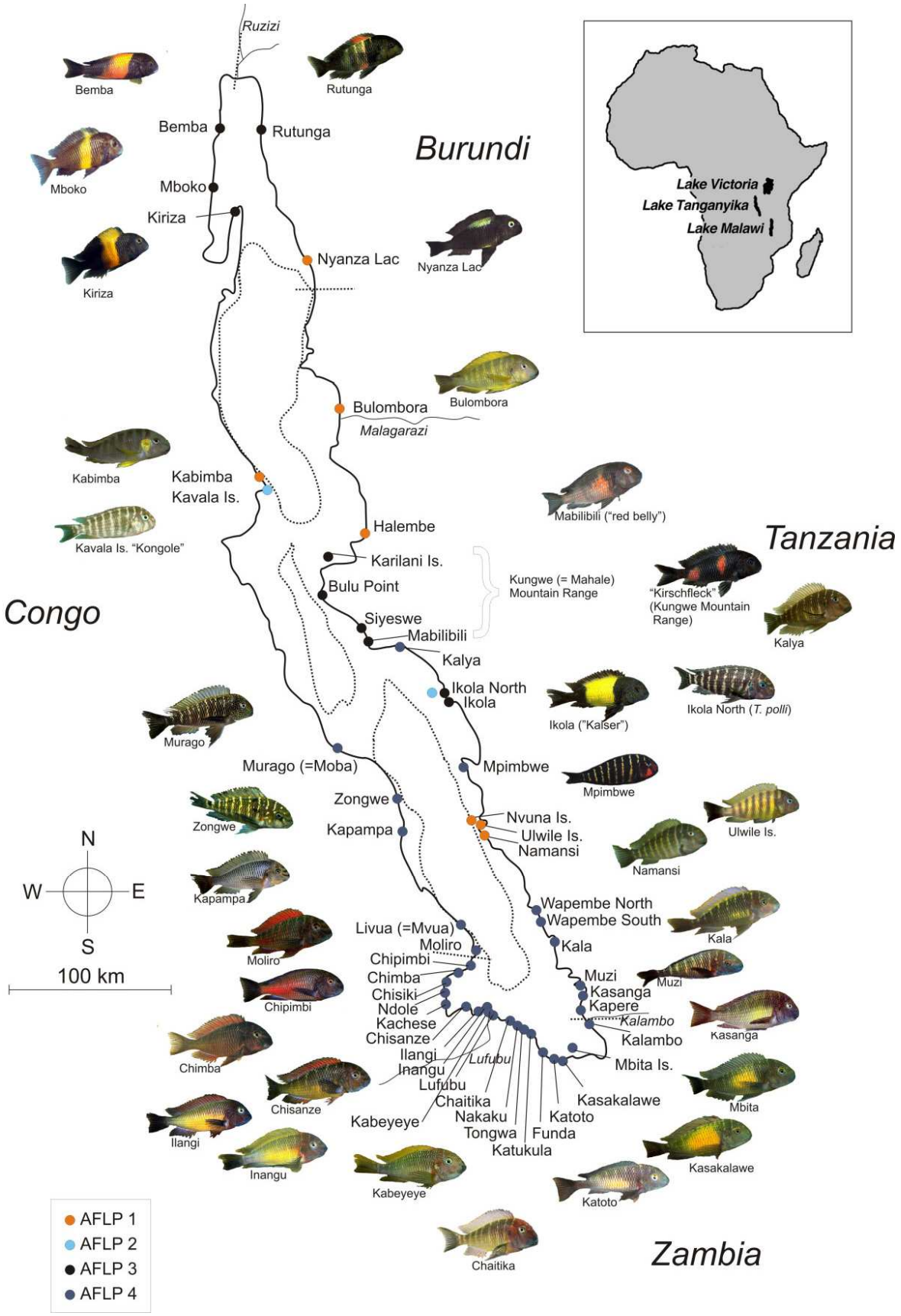


Figure 2 Map of Lake Tanganyika showing the 51 sample sites used in the study of Egger et al. (2007). The graph illustrates the phenotypic diversity in *Tropheus* around the Lake. Photographs of fishes are labeled by the locality where the respective morphs occur. Names in parenthesis correspond to the designations used in the aquarium trade. The dots marking sample sites were colored representing the distribution of the four major AFLP clades. The graph was taken from Egger et al. (2007).

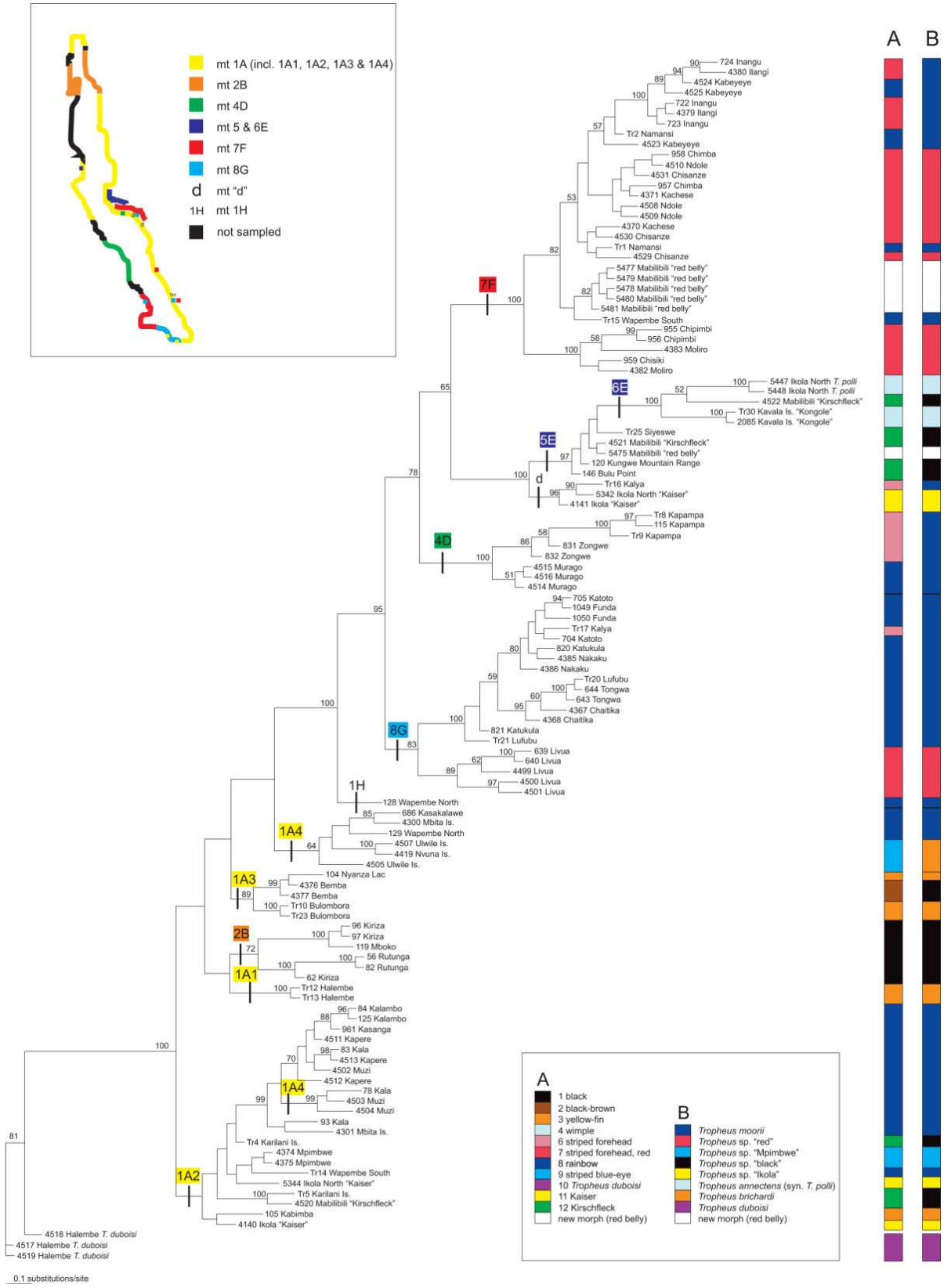


Figure 3 Bayesian inference tree based on mitochondrial control region sequences. MtDNA lineages defined by Sturmbauer et al. (2005) are indicated above branches and their approximate distribution is shown in the map (top left; adapted from Sturmbauer et al. 2005). Posterior probability values are shown at the respective nodes. The bars on the right show the assignment of the samples to (A) color lineages (Schupke 2003) and to (B) the species classification suggested by Konings (1998). Same-colored bar sections comprising paraphyletic clades are separated by black lines to highlight the inconsistency with tree topology. The phylogeny was calculated by and taken from Egger et al. (2007).

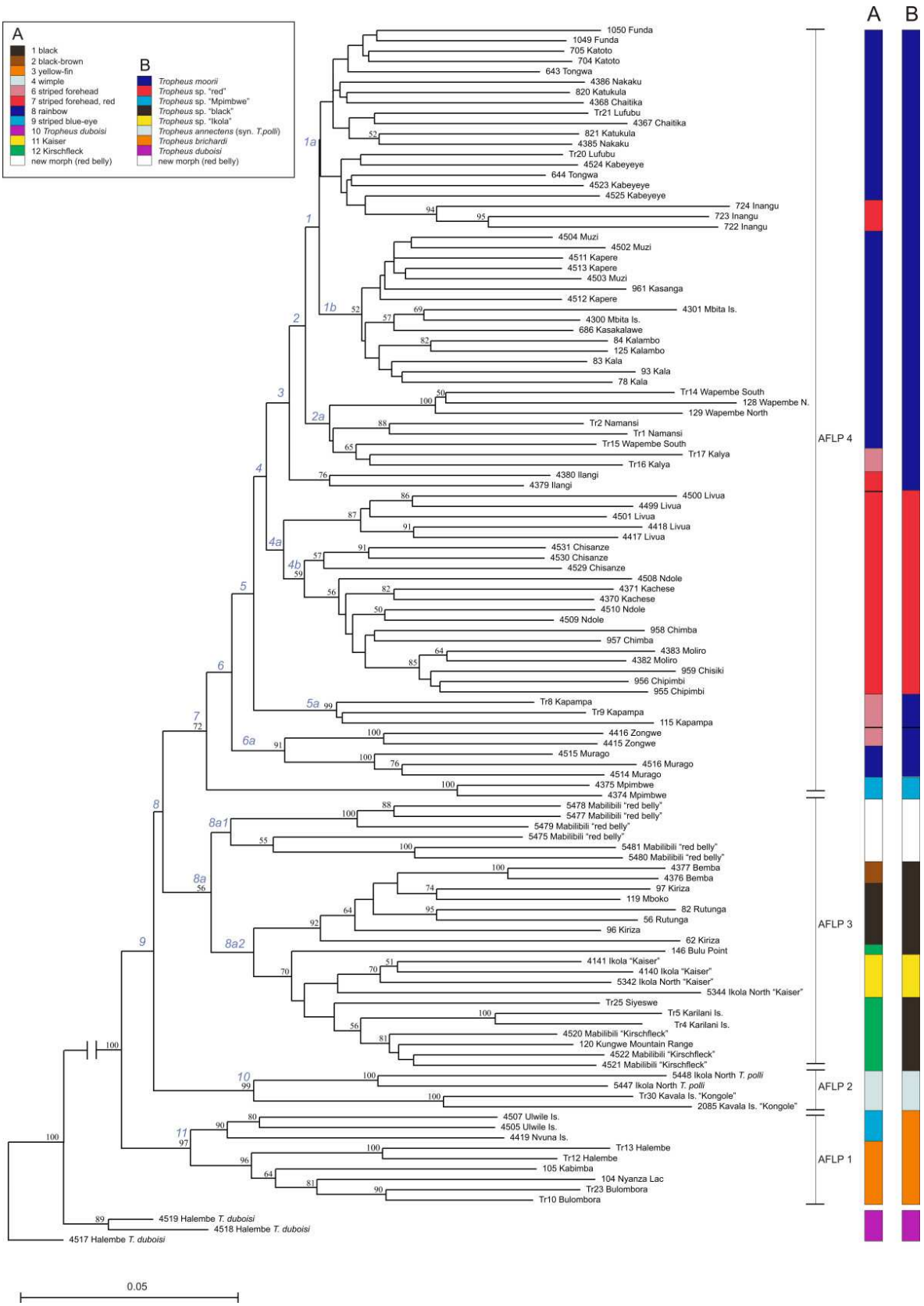


Figure 4 Neighbor joining tree based on AFLP genotype distances. Bootstrap support > 50% is shown at the respective nodes. On the right side the four well-supported AFLP clades are shown and the colored bars represent the assignment of the samples to (A) color lineages (Schupke 2003) and to (B) the species classification suggested by Konings (1998). Same-colored bar sections comprising paraphyletic clades are separated by black lines to highlight the inconsistency with tree topology. The geographic distribution of the AFLP clades is shown in Fig. 2. The phylogeny was calculated by and taken from Egger et al. (2007).

2008), is said to have shaped the present distribution of the different lineages (Sturmbauer et al. 2001, 2005). A recent study on *Tropheus* used mitochondrial as well as amplified fragment length polymorphism (AFLP) data to shed light on the phylogeographic status of individuals from 51 localities covering almost the complete shoreline of Lake Tanganyika (Egger et al. 2007; Fig. 2). This study revealed large discrepancies between the mtDNA and the AFLP phylogeny: The distribution of mitochondrial haplotypes was highly complex and largely disagreeing with suggested species classifications (e.g. Konings 1998; Fig. 3) and the proposed classification based on color pattern similarities (Schupke 2003; Fig. 3). AFLP data in contrast yielded a phenogram which was largely consistent with species classifications, color pattern similarities and often also with the current geographic distribution of populations (Fig. 2 and 4). Thus the authors stated that mtDNA phylogeographic patterns possibly reflected large-scale migration events triggered by major lake level fluctuations, whereas cohesion among local groups reflected by AFLP data was suggested to have resulted from gene flow between neighboring populations, triggered by frequent minor lake level fluctuations. Both types of molecular markers however, showed evidence for hybridization and introgression which led the authors to the conclusion that lake level fluctuations drove the exceptional diversification of *Tropheus* morphs through fragmentation of populations on the one hand but also through hybridization between differentiated and formally separated populations in periods of secondary contact.

Tropheus and its color diversity

To find the underlying mechanisms of diversification and subsequent speciation is a very important and challenging issue in evolutionary biological research. Especially variation on the within-species level provides an opportunity to study the processes driving both, genetic and phenotypic diversity (Koblmüller et al. 2011). A trait which has been the focus of plenty of studies on animals is body coloration (e.g. Endler 1983; Seehausen et al. 1999; Galeotti et al. 2003; Salzburger 2009; Croucher et al. 2011). These studies show that body coloration may be influenced by natural as well as sexual selection. Alternatively, color divergence could merely be a side-product of reproductive isolation between populations due to other cues than color (e.g. Plenderleith et al. 2005; Seehausen et al. 2008) or it might be caused by random genetic drift alone (e.g. Brakefield 1990; Arnegard et al. 1999; Hoffman et al. 2006).

The high level of diversity in terms of body coloration makes *Tropheus* a promising system for studies of color pattern evolution. Body coloration plays a crucial role in this genus

not only during courtship and mating but also for communication during other social interactions, and both genders exhibit the same repertoires of color patterns signaling motivation and social status (Sturmbauer & Dallinger 1995). However, the origin of the high amount of color polymorphism in *Tropheus* yet remains largely unknown. In fact nothing is known about the genetic basis underlying it and only assumptions can be made regarding the forces triggering it. Concerning the genetic basis the rapidity of the diversification suggests the presence of only few genetic'switches', such that color patterns can be altered by minor changes in the genome. In **Chapter 2** I present the results from a genome scan carried out on populations of three closely related color morphs of *Tropheus moorii* (the Blue, the Red and the Yellow-Blotch morph) using a large set of AFLP loci. The aim of a genome scan is to detect candidate loci that are potentially under selection or linked to genome regions under selection and these loci are considered responsible for divergence between populations. As all around Lake Tanganyika *Tropheus* populations are adapted to the same kind of habitat, they show only subtle morphological disparities, both in external (Snoeks et al. 1994; Maderbacher et al. 2008; Herler et al. 2010) and in internal morphology (Postl et al. 2008). Thus the only obvious difference between *T. moorii* populations is their body coloration making it very plausible that 'outlier' loci detected to be under selection through the genome scan are indeed associated with body coloration. This assumption is furthermore encouraged as we investigated color morphs that are very closely related to each other. **Chapter 2** presents the so far first genome scan study that focuses on a trait (i.e. body coloration) that is possibly not shaped by natural selection, as it doesn't obviously represent an adaptation to some ecological and/or environmental conditions. This assumption is based on the fact that variability in environmental conditions between habitats of different *T. moorii* morphs is similar to that between habitats of populations within a morph. Thus it is possible that coloration and outlier loci associated with it in this study are under sexual selection.

In addition to the genome scan **Chapter 2** deals with a general characterization of the differences in coloration between the morphs and sexes. Therefore we first took digital pictures of whole individuals (representative males and females of each morph) and compared them between morphs. Secondly, pictures of body regions were taken to make comparisons between morphs on the one hand but also within morphs and within individuals on the other hand. Finally, concerning the yellow/red coloration, underlying pigments were extracted from different body regions in order to test which pigments were present and to characterize the differences between the three morphs.

In **Chapters 3** and **4** the focus lies on the forces driving the evolution of the outstanding color pattern diversity in *Tropheus moorii*. Like for some Malawi and Victoria cichlids sexual selection as a force driving the evolution of color diversity has been discussed for *Tropheus* (Sturmbauer & Meyer 1992; Salzburger et al. 2006) but without conclusive evidence so far (Egger et al. 2008, 2010; Sefc 2008; Steinwender et al. 2011).

Chapters 3 and **4** of this thesis deal with hybridization, a factor that has also been discussed to be triggering color pattern diversification in cichlids (Salzburger et al. 2002b; Smith et al. 2003; Schelly et al. 2006; Stelkens et al. 2009). Traditionally, hybridization in animals was seen as a destructive force, diminishing species diversity (Mayr 1963). However, it is now known that hybridization and introgression could very well lead to an increase of genetic diversity in the hybrid population and furthermore to the formation of a stable population with a new, unique phenotype. This subsequently, could end up in the formation of a new species (Smith et al. 2003). Phylogeographic data indicated that in *Tropheus* opportunity for hybridization was provided several times in history due to recurrent cycles of lake level fluctuations (Sturmbauer & Meyer 1992; Baric et al. 2003, Sturmbauer et al. 2005; Egger et al. 2007; Sefc et al. 2007) which generally had severe impact on rocky habitats and species communities within them in East African Rift Valley Lakes (e.g. Sturmbauer & Meyer 1992, Johnson et al. 1996; Sturmbauer et al. 1997; Nagl et al. 2000). A major decline of the water level would seriously affect shallow water inhabitants like *Tropheus*, especially in flat shores: Populations would be displaced from their former location and secondary contact between previously allopatric populations and subsequent hybridization between them could happen. If and to what extent hybridization in the scenario of secondary contact would happen is dependent on the degree of assortative mating preferences which have shown to be variable for some *Tropheus moorii* populations. Differing strength of positive color assortative mating preferences has been shown in an artificially admixed population in the lake: Strong color assortative mating was demonstrated for highly distinct *Tropheus* morphs and isolation was weak between more similar morphs (Salzburger et al. 2006). Laboratory mate-choice and consecutive breeding pond experiments, conducted on populations of color morphs investigated in the two following studies, also revealed variable degrees of assortative mating preferences (Egger et al. 2008, 2010; Hermann et al. unpublished). Details on the results are described in the separate introduction for each chapter.

Hybrid status has already been discussed for certain *Tropheus* populations in southern Lake Tanganyika, including populations analyzed in **Chapters 3** and **4**. Here two contact zones in southern Lake Tanganyika are investigated for signs of hybridization using large

scale data sets including many samples and mitochondrial as well as nuclear molecular genetic markers. **Chapter 3** deals with previously detected evidence for introgression and hybridization between two color morphs, the Blue and the Yellow-Blotch lineage, west of a large sandy bay that actually separates the two morphs. The aim was to confirm signs of introgression detected in previous studies (Baric et al. 2003; Sturmbauer et al. 2005; Sefc et al. 2007) with a large data set and to reconstruct the circumstances of the hybridization. In **Chapter 4** I investigate a second contact zone where several orange to yellow populations north of a river estuary were suggested to result from hybridization between the adjacent Red and Blue morph (homoplasy excess test; Egger et al. 2007). Here again the aims were to confirm this signs of introgression using a large data set and shed lights on the circumstances that led to the observed introgression patterns. These two studies are the first that investigate hybridization in *Tropheus moorii* on the population level using a large scale data set.

REFERENCES

Arnegard ME, Markert JA, Danley PD, Stauffer JR, Ambali AJ, Kocher TD (1999) Population structure and colour variation of the cichlid fish *Labeotropheus fuelleborni* Ahl along a recently formed archipelago of rocky habitat patches in southern Lake Malawi. Proceedings of the Royal Society of London Series B-Biological Sciences 266:119-130

Baric S, Salzburger W, Sturmbauer C (2003) Phylogeography and evolution of the Tanganyikan cichlid genus *Tropheus* based upon mitochondrial DNA sequences. Journal of Molecular Evolution 56:54-68

Barluenga M, Stolting KN, Salzburger W, Muschick M, Meyer A (2006) Sympatric speciation in Nicaraguan crater lake cichlid fish. Nature 439:719-723

Brakefield PM (1990) Genetic drift and patterns of diversity among color-polymorphic populations of the homopteran *Philaenus spumaris* in an island archipelago. Biological Journal of the Linnean Society 39:219-237

Brichard P (1978) An example of the isolation of rocky substrates among sand bottoms in the north of Lake Tanganyika, East Africa. Revue de Zoologie Africaine 92:518-524

Clabaut C, Salzburger W, Meyer A (2005) Comparative phylogenetic analyses of the adaptive radiation of Lake Tanganyika cichlid fish: Nuclear sequences are less homoplasious but also less informative than mitochondrial DNA. Journal of Molecular Evolution 61:666-681

Cohen AS, Soreghan MJ, Scholz CA (1993) Estimating the age of ancient lakes: an example from Lake Tanganyika, East African rift system. Geology 21:511-514

Cohen AS, Lezzar K-E, Tiercelin J-J, Soreghan M (1997) New palaeogeographic and lake-level reconstructions of Lake Tanganyika: implications for tectonic, climatic and biological evolution in a rift lake. Basin Research 9:107-132

Croucher PJP, Oxford GS, Lam A, Gillespie RG (2011) Stabilizing selection maintains exuberant colour polymorphism in the spider *Theridion californicum* (Araneae, Theridiidae). Blackwell Publishing Ltd, p no-no

Dijkstra PD, Seehausen O, Pierotti MER, Groothuis TGG (2007) Male-male competition and speciation: aggression bias towards differently coloured rivals varies between stages of speciation in a Lake Victoria cichlid species complex. Journal of Evolutionary Biology 20:496-502

Duftner N, Koblmueller S, Sturmbauer C (2005) Evolutionary relationships of the Limnochromini, a tribe of benthic deepwater Cichlid fish endemic to Lake Tanganyika, East Africa. Journal of Molecular Evolution 60:277-289

Egger B, Obermueller B, Phiri H, Sturmbauer C, Sefc KM (2006) Monogamy in the maternally mouthbrooding Lake Tanganyika cichlid fish *Tropheus moorii*. Proceedings of the Royal Society B-Biological Sciences 273:1797-1802

Egger B, Koblmueller S, Sturmbauer C, Sefc KM (2007) Nuclear and mitochondrial data reveal different evolutionary processes in the Lake Tanganyika cichlid genus *Tropheus*. Bmc Evolutionary Biology 7:-

Egger B, Obermueller B, Eigner E, Sturmbauer C, Sefc KM (2008) Assortative mating preferences between colour morphs of the endemic Lake Tanganyika cichlid genus *Tropheus*. Hydrobiologia 615:37-48

Egger B, Mattersdorfer K, Sefc KM (2010) Variable discrimination and asymmetric preferences in laboratory tests of reproductive isolation between cichlid colour morphs. J Evol Biol 23:433-9

Endler JA (1983) Natural and sexual selection on color patterns in poeciliid fishes. Environmental Biology of Fishes 9:173-190

Fryer G & Iles TD (1972) The cichlid fishes oft he great lakes of Africa. Their biology and evolution. London, UK: Oliver and Boyd.

Galeotti P, Rubolini D, Dunn PO, Fasola M (2003) Colour polymorphism in birds: causes and functions. Journal of Evolutionary Biology 16:635-646

Goodwin NB, Balshine-Earn S, Reynolds JD (1998) Evolutionary transitions in parental care in cichlid fish. Proceedings of the Royal Society of London Series B 265:2265-2272

Herler J, Kerschbaumer M, Mitteroecker P, Postl L, Sturmbauer C (2010) Sexual dimorphism and population divergence in the Lake Tanganyika cichlid fish genus *Tropheus*. Frontiers in Zoology 7:-

Hoffman EA, Schueler FW, Jones AG, Blouin MS (2006) An analysis of selection on a colour polymorphism in the northern leopard frog. Molecular Ecology 15:2627-2641

Johnson TC, Scholz CA, Talbot MR, Kelts K, Ricketts RD, Ngobi G, Beuning K, Ssemmanda I, McGill JW (1996) Late pleistocene desiccation of Lake Victoria and rapid evolution of cichlid fishes. Science 273:1091-1093

Klett V, Meyer A (2002) What, if anything, is a Tilapia? - Mitochondrial ND2 phylogeny of Tilapiines and the evolution of parental care systems in the African cichlid fishes. Molecular Biology and Evolution 19:865-883

Knight ME, Turner GF (2004) Laboratory mating trials indicate incipient speciation by sexual selection among populations of the cichlid fish *Pseudotropheus zebra* from Lake Malawi. Proceedings of the Royal Society of London Series B-Biological Sciences 271:675-680

Koblmüller S, Salzburger W, Sturmbauer C (2004) Evolutionary relationships in the sand-dwelling cichlid lineage of lake Tanganyika suggest multiple colonization of rocky habitats and convergent origin of biparental mouthbrooding. Journal of Molecular Evolution 58:79-96

Koblmüller S, Duftner N, Katongo C, Phiri H, 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, Egger B, Sturmbauer C, M SK (2007) Evolutionary history of Lake Tanganyika's scale-eating cichlid fishes. Molecular Phylogenetics and Evolution 44:1295-1305

Koblmüller S, Sefc KM, Sturmbauer C (2008) The Lake Tanganyika cichlid species assemblage: recent advances in molecular phylogenetics. Hydrobiologia 615:5-20

Koblmüller S, Egger B, Sturmbauer C, Sefc KM (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

Koblmüller S, Salzburger W, Obermuller B, Eigner E, Sturmbauer C, Sefc KM (2011) Separated by sand, fused by dropping water: habitat barriers and fluctuating water levels steer the evolution of rock-dwelling cichlid populations in Lake Tanganyika. Molecular Ecology 20:2272-2290

Kocher TD (2004) Adaptive evolution and explosive speciation: the cichlid fish model. Nat Rev Genet 5:288-298

Konings A (1998) Tanganyika cichlids in their natural habitats. El Paso: Cichlid Press, 272 pp.

Liem KF (1973) Evolutionary strategies and morphological innovations - Cichlid pharyngeal jaws. Systematic Zoology 22:425-441

Liem KF (1980) Adaptive significance of intraspecific and interspecific differences in the feeding repertoires of cichlid fishes. American Zoologist 20:295-314

Maan ME, Seehausen O, Soderberg L, Johnson L, Ripmeester EAP, Mrosso HDJ, Taylor MI, van Dooren TJM, van Alphen JJM (2004) Intraspecific sexual selection on a speciation trait, male coloration, in the Lake Victoria cichlid *Pundamilia nyererei*. Proceedings of the Royal Society of London Series B-Biological Sciences 271:2445-2452

Maderbacher M, Bauer C, Herler J, Postl L, Makasa L, Sturmbauer C (2008) Assessment of traditional versus geometric morphometrics for discriminating populations of the *Tropheus moori*i species complex (Teleostei : Cichlidae), a Lake Tanganyika model for allopatric speciation. Journal of Zoological Systematics and Evolutionary Research 46:153-161

Martens K, Coulter G, Goddeeris B (1994) Speciation in ancient lakes - 40 years after Brooks. Ergebnisse der Limnologie 0:75-96

Mayr E (1963) Animal Species and Evolution, Belknap Press

McGlue MM, Lezzar KE, Cohen AS, Russell JM, Tiercelin J-J, Felton AA, Mbede E, Nkotagu HH (2008) Seismic records of late Pleistocene aridity in Lake Tanganyika, tropical East Africa. Journal of Paleolimnology 40:635-653

McKaye, K. R. et al. (2002) Behavioral, morphological and genetic evidence of divergence of the Midas cichlid species complex in two Nicaraguan crater lakes. Cuademos de la Investigación de la UCA 12: 19-47

Nagl S, Tichy H, Mayer WE, Takezaki N, Takahata N, Klein J (2000) The origin and age of haplochromine fishes in Lake Victoria, East Africa. Proceedings of the Royal Society of London Series B-Biological Sciences 267:1049-1061

Nishida M (1991) Lake Tanganyika as an evolutionary reservoir of old lineages of East African cichlid fishes: Inferences from allozyme data. Experientia 47:974-979

Pauers MJ, McKinnon JS, Ehlinger TJ (2004) Directional sexual selection on chroma and within-pattern colour contrast in *Labeotropheus fuelleborni*. Proceedings of the Royal Society of London Series B-Biological Sciences 271:S444-S447

Plenderleith M, van Oosterhout C, Robinson RL, Turner GF (2005) Female preference for conspecific males based on olfactory cues in a Lake Malawi cichlid fish. Biology Letters 1:411-414

Poll M (1986) Classification des Cichlidae du lac Tanganyika: tribus, genre et especes. Academie Royale de Belgique Memoires de la Classe des Sciences 45:1-163

Postl L, Herler J, Bauer C, Maderbacher M, Makasa L, Sturmbauer C (2008) Geometric morphometrics applied to viscerocranial bones in three populations of the Lake Tanganyika cichlid fish *Tropheus moorii*. Journal of Zoological Systematics and Evolutionary Research 46:240-248

Rossiter A (1995) The cichlid fish assemblages of Lake Tanganyika: ecology, behaviour and evolution of its species flock. Advances in Ecological Research 26:187-252

Salzburger W, Meyer A, Baric S, Verheyen E, 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, Sturmbauer C (2002b) Speciation via introgressive hybridization in East African cichlids? Molecular Ecology 11:619-625

Salzburger W, 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, Meyer A (2005) Out of Tanganyika: Genesis, explosive speciation, key-innovations and phylogeography of the haplochromine cichlid fishes. Bmc Evolutionary Biology 5:15

Salzburger W, Niederstatter H, Brandstatter A, Berger B, Parson W, Snoeks J, Sturmbauer C (2006) Colour-assortative mating among populations of *Tropheus moorii*, a cichlid fish from Lake Tanganyika, East Africa. Proceedings of the Royal Society B-Biological Sciences 273:257-266

Salzburger W (2009) The interaction of sexually and naturally selected traits in the adaptive radiations of cichlid fishes. Molecular Ecology 18:169-185

Schelly R, Salzburger W, Koblmueller S, Duftner N, 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 UK, Tautz D, Pääbo S (1994) Sympatric speciation suggested by monophyly of crater lake cichlids. Nature 368:629-632

Schliewen UK, Kocher TD, McKaye KR, Seehausen O, Tautz D (2006) Evidence for sympatric speciation? Nature 444:E12-E13

Schluter D (2000) The ecology of adaprive radiation. Oxford, UK: Oxford University Press.

Scholz CA, King JW, Ellis GS, Swart PK, Stager JC, Colman SM (2003) Paleolimnology of Lake Tanganyika, East Africa, over the past 100 k yr. Journal of Paleolimnology 30:139-150

Schupke P (2003) African Cichlids II: Tanganyika I: *Tropheus*. Aqualog, ACS GmbH, Rodgau.

Schürch R, Taborsky B (2005) The functional significance of buccal feeding in the mouthbrooding cichlid *Tropheus moorii*. Behaviour 142:265-281

Seehausen O, Mayhew PJ, van Alphen JJM (1999) Evolution of colour patterns in East African cichlid fish. Journal of Evolutionary Biology 12:514-534

Seehausen O, van Alphen JJM (1999) Can sympatric speciation by disruptive sexual selection explain rapid evolution of cichlid diversity in Lake Victoria? Ecology Letters 2:262-271

Seehausen O (2006) African cichlid fish: a model system in adaptive radiation research. Proceedings of the Royal Society of London Series B 273:1987-1998

Seehausen O, Terai Y, Magalhaes IS, Carleton KL, Mrosso HDJ, Miyagi R, van der Sluijs I, Schneider MV, Maan ME, Tachida H, Imai H, Okada N (2008) Speciation through sensory drive in cichlid fish. Nature 455:620-U23

Sefc KM, Baric S, Salzburger W, 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

Sefc KM (2008) Variance in reproductive success and the opportunity for selection in a serially monogamous species: simulations of the mating system of *Tropheus* (Teleostei: Cichlidae). Hydrobiologia 615:21-35

Smith PF, Konings A, Kornfield I (2003) Hybrid origin of a cichlid population in Lake Malawi: implications for genetic variation and species diversity. Molecular Ecology 12:2497-2504

Snoeks J, Rueber L, Verheyen E (1994) The Tanganyika Problem: comments on the taxonomy and distribution patterns of its cichlid fauna. Adv. Limn. 44:357-374

Snoeks J (2000) How well known is the ichthyodiversity of the Large East African Lakes? Advances in Ecological Research 31:17-38

Steinwender B, Koblmüller S, Sefc K (2011) Concordant female mate preferences in the cichlid fish *Tropheus moorii*. Hydrobiologia:1-10

Stelkens RB, Schmid C, Selz O, Seehausen O (2009) Phenotypic novelty in experimental hybrids is predicted by the genetic distance between species of cichlid fish. Bmc Evolutionary Biology 9:Article No.: 283

Stiassny MLJ, Jensen JS (1987) Labroid intrarelationships revisited morphological complexity key innovations and the study of comparative diversity. Bulletin of the Museum of Comparative Zoology 151:269-319

Stiassny MLJ (1991) Phylogenetic interrelationships of the family Cichlidae: an overview. In Keenleyside MHA (ed.), Cichlid Fishes, Behavior, Ecology and Evolution. London, UK: Chapman & Hall, 1-35.

Stiassny MLJ, Meyer A (1999) Cichlids of the Rift lakes. Scientific American 280:64-69

Sturmbauer C, Meyer A (1992) Genetic divergence, speciation and morphological stasis in a lineage of African cichlid fishes. Nature 358:578-581

Sturmbauer C, Dallinger R (1995) Diurnal variation of spacing and foraging behaviour in *Tropheus moorii* (Cichlidae) in Lake Tanganyika, Eastern Africa. Netherlands Journal of Zoology 45:386-401

Sturmbauer C, Verheyen E, Rueber L, Meyer A (1997) Phylogeographic patterns in populations of cichlid fishes from rocky habitats in Lake Tanganyika. Academic Press, Inc.; Academic Press Ltd.

Sturmbauer C (1998) Explosive speciation in cichlid fishes of the African Great Lakes: a dynamic model of adaptive radiation. Journal of Fish Biology 53:18-36

Sturmbauer C (2000) Modellsysteme der molekularen Evolutionsbiologie: Die Seen Ostafrikas und ihre Buntbarsche; Biologie in unserer Zeit 6: 2-11.

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

Sturmbauer C, Koblmueller S, Sefc KM, Duftner N (2005) Phylogeographic history of the genus *Tropheus*, a lineage of rock-dwelling cichlid fishes endemic to Lake Tanganyika. Hydrobiologia 542:335-366

Takahashi K, Terai Y, Nishida M, Okada N (2001) Phylogenetic relationships and ancient incomplete lineage sorting among cichlid fishes in Lake Tanganyika as revealed by analysis of the insertion of retrotransposons. Molecular Biology and Evolution 18:2057-2066

Takahashi T (2003) Systematics of Tanganyikan cichlid fishes (Teleostei: Perciformes). Ichthyological Research 50:367-382

Turner GF, Seehausen O, Knight ME, Allender C, Robinson RL (2001) How many species of cichlid fishes are there in African lakes? Molecular Ecology 10:793-806

Yanagisawa Y, Nishida M (1991) The Social and Mating System of the Maternal Mouthbrooder *Tropheus moorii* (Cichlidae) in Lake Tanganyika. Japanese Journal of Ichthyology 38:271-282

CHAPTER 2

The basis of color pattern divergence in Lake Tanganyika's *Tropheus moorii*

ABSTRACT

The search for the basis of phenotypic diversity in natural populations has become a challenging task for evolutionary biologists within the last decade. Multiple studies use genome wide scans to detect so called outlier loci which show levels of differentiation that deviate from the rest of the genome. Thus they reflect genes being under selection or loci linked to them. In this study we investigated 11 closely related populations of the Lake Tanganyika cichlid *Tropheus moorii*, which belong to three color morphs. The aim was to find color related outlier loci and thus show that coloration in *T. moorii* has not evolved due to neutral forces alone but instead underlies selection. We examined a large set of genome wide distributed AFLP markers for outlier detection between morphs and also within morphs to distinguish color related outliers from population-specific ones. Additionally, as a first step to characterize the phenotypic differences between the three morphs, we carried out body color analyses, including whole individual and body part photography and, concerning the red/yellow coloration, skin-pigment extraction. We found a total of 11 between-morph outlier loci that fulfilled our criteria for reliable outliers. Of these, six were predominantly found in Blue vs. Yellow-Blotch comparisons and five only in Red vs. Blue, and Red vs. Yellow-Blotch comparisons, respectively. Color analyses highlighted body color differences between morphs, and skin pigment analyses showed that yellow to red coloration in this species is merely due to different types of carotenoids, as no pteridines were found in any sample. This is the first study that uses the genome scan approach to detect outlier loci associated with a trait that doesn't obviously represent adaptation to some environmental or ecological condition. Together with the body color analyses it serves as a valuable method for a basic characterization of coloration and the forces underlying the evolution of it in *T. moorii*. The results lay the foundation for further investigations of the genetic basis of coloration in this species.

INTRODUCTION

A big challenge for evolutionary biologists is to answer the question what underlies phenotypic diversity in natural populations. One of the promising approaches concentrates on unraveling the genetic basis of certain traits, with the focus lying on two main issues: (i) The identification of the genes or genomic regions underlying phenotypic traits and (ii) the identification of the forces contributing to phenotypic diversity, i.e. the relative influence of neutral versus selective processes. Discerning between these influences is a crucial step towards understanding the evolution of species (Gagnaire et al. 2009). Particularly in wild populations information about the genes underlying phenotypic diversity is still largely lacking. However, in the last years several methods for examining the molecular basis of certain traits have been developed to address one or both of the issues mentioned above, including candidate gene analysis, quantitative trait loci (QTL) mapping, and genome scans. There are several promising studies that focus on the analysis of candidate genes (e.g. Rosenblum et al. 2004; Hoekstra et al. 2006; reviewed in Noor & Feder 2006) but this approach requires information on the presence and DNA sequence of the genes potentially involved with the phenotypic trait in question. For an association of phenotypes with genotypes QTL mapping has been shown to be a valuable approach (e.g. Streelman et al. 2003; Stemshorn et al. 2005; McClelland & Naish 2010; Wringe et al. 2010; Sahana et al. 2011). But once again the requirements for this method are not easily complied with: Relying on crosses between phenotypes and the production of a large number of F2 offspring, it requires well-studied species that are easy to cross and manipulate (Stinchcombe & Hoekstra 2007). This is a problem in species with long generation times and small offspring numbers. However, in the recent past genome scans as population genomic approaches to detect genes influenced by selection or loci linked to them have become accessible even for non-model species. In several studies genome scans have proved as a valuable method to detect genetic loci being under selection and thus putatively underlying a certain phenotypic adaptive trait (e.g. Bonin et al. 2006; Jump et al. 2006, Egan et al. 2008; Nosil et al. 2008; Minder & Widmer 2008; Parisod & Joost 2009; Paris et al. 2010; Midamegbe et al. 2011). Those studies have shown that it is possible to detect signs of selection without a phenotype, simply by looking at genomic data. Moreover they unraveled that due to the effects of e.g., genetic hitchhiking and/or linkage selection affects non-coding regions throughout the genome as well as coding regions and thus contributes to variation within and between species. This stays in contrast to the neutral theory which states that most of this variation is selectively

neutral, i.e., it does not have effects on the organism's fitness (Kimura 1983). Thus to detect signals of selection in putatively neutral genomic regions it is necessary to investigate a sufficient number of molecular markers which are randomly distributed across the whole genome. This is provided by the amplified fragment length polymorphism (AFLP) technique (Vos et al. 1995). This approach has been used very often to detect selection because it represents a favorable way to easily scan hundreds of markers scattered throughout the genome of non-model species (e.g. Wilding et al. 2001; Campbell & Bernatchez 2004; Bonin et al. 2006; Chen et al. 2009; Paris et al. 2010; Midamegbe et al. 2011). The idea of genome scans is that demography and neutral evolutionary history of populations have the same effect on neutral loci across the genome. In contrast, loci under selection and loci closely linked to them will possess an outlier pattern of differentiation (Luikart et al. 2003). One of the approaches for outlier detection uses F_{ST} as a measure for variation between populations or sub-populations. Thus genomic regions that are under directional or diversifying selection, which drives differentiation between populations, are predicted to exhibit larger differentiation, and hence a larger F_{ST} value than loci that are neutral. Vice versa regions that are influenced by balancing selection, which tends to homogenize allele frequencies between populations and thus counteracts differentiation, should exhibit less differentiation, i.e. a smaller F_{ST} than neutral loci. Therefore the aim of genome scans is to detect those 'outlier' loci and discern them from background neutral-variability caused by drift and mutations. This is a very challenging task because it is often not easy to pick out 'true outliers' while avoiding false positives (Paris et al. 2010).

A good example for studies on population differentiation is the Lake Tanganyika cichlid genus *Tropheus*. It belongs to the tribe Tropheini, which is phylogenetically nested within the Haplochromini (Salzburger et al. 2005, Koblmüller et al. 2008; Koblmüller et al. 2010). All around Lake Tanganyika *Tropheus* populations occur in the same kind of habitat: as specialized rock-dwellers and algae grazers they are highly stenotopic to the shallow areas of the shoreline with light conditions and habitat structure being similar all over their distribution range. Being strictly dependant on rocky ground, major habitat discontinuities like long stretches of sand act as strong barriers to dispersal between populations and thus to a severe reduction of gene flow (Sefc et al. 2007). But also genetic differentiation between populations along mostly continuous shoreline was found to be significant, due to strong behavioral philopatry and/or minor habitat discontinuities in this species (Sefc et al. 2007; Koblmüller et al. 2011). This is also shown in two more recent studies on particular *T. moorii* populations in southern Lake Tanganyika using mitochondrial DNA, AFLPs, and

microsatellite data, where strong population structure between populations even across short geographic distances was revealed (Mattersdorfer et al. unpublished; see Chapters 3 and 4). However, genetically differentiated *Tropheus moorii* populations show only subtle morphological disparities, both in external (Snoeks et a. 1994; Maderbacher et al. 2008; Herler et al. 2010) and in internal morphology (Postl et al. 2008). This is probably due to the high degree of specialization to the same habitat and the same trophic niche (Postl et al. 2008). The only obvious difference between *T. moorii* populations is their body coloration which varies only slightly between geographically close populations along stretches of continuous shoreline, but prominently between populations that are geographically further apart or separated by a habitat barrier, like e. g., a long stretch of sand. Thus body coloration differs from location to location, leading to a high degree of phenotypic diversity in this species.

Body coloration in vertebrates is caused by different pigments that reside in specialized organelles of pigment cells, the chromatophores, which developed from the neural crest. In teleost fish five types of chromatophores have been identified so far: (i) Dark melanophores, containing the dark pigment eumelanin. (ii) The yellow to red xanthophores and erythrophores, containing pteridine and/or carotenoid pigments. Generally it can be said that xanthophores contain yellow, whereas erythrophores contain red pigment cells. However, yellow as well as red coloration can be produced by plenty pteridine and carotenoid pigments and these can be found in both chromatophore types (reviewed in Mills & Patterson 2009). Thus, as we cannot distinguish between the two types of chromatophores, we speak of xanthophores/erythrophores in this study. (iii) The silvery-blue reflecting iridophores which contain purines as reflecting platelets. (iv) The whitish leucophores which also contain purines but as whitish leucosomes, and (v) the cyanophores that contain a blue pigment (reviewed in Bagnara & Matsumoto 2006). Cyanophores have been found so far only in callionymid fishes (Goda & Fuji 1995; Bagnara et al. 2007). On the basis of these chromatophore types the actual color pattern of a fish is then determined by their spatial arrangement, their combinatory effects and also by the distribution of pigments within them (Hirata et al. 2003; Grether et al. 2004; Hirata et al. 2005) thus leading to a high degree of complexity. Although first studies on inheritance of pigmentation in teleost fish can be dated far back in time (e.g. Winge 1927), only recently genes involved in pigmentation have been identified owing to extensive research on the zebrafish *Danio rerio* (e.g. Kelsh et al. 1996; Odenthal et al. 1996; Parichy et al. 1999, 2000, 2007; Dutton et al. 2001, Elworthy et al. 2003; Mellgren et al. 2005) and on the medaka *Oryzias latipes* (e.g. Koga & Hori 1997; Kelsh

et al. 2004a; reviewed in Takeda & Shimada 2010). The existence of a large collection of natural and induced mutants of these easily bread freshwater fish has made them valuable model species for numerous studies on development and genetics (Furutani-Seiki & Wittbrodt 2004). Based on these fish model studies a series of candidate genes for color determination are thus existent and these have already been applied to cichlids in several studies. Sugie et al. (2004) e.g., cloned cichlid homologs of *tyrosinase, endothelin receptor b1*, *mitf* and *Aim1* to elucidate mechanisms underlying pigment pattern diversity in African cichlids. They showed a fast evolution of the *mitf* gene and suggested that this acceleration might have occurred together with the Great Lake cichlids pigment pattern diversification. Two other studies identified a single region of the cichlid genome underlying the orange blotch (OB) phenotype in Lake Malawi cichlids via linkage mapping: They found that the OB pattern is based on a mutation in the cis-regulatory region of the *Pax7* gene resulting in significant expression increase (Streelman et al. 2003; Roberts et al. 2009). Terai et al. (2002, 2003) suggested the cichlid homolog of the zebrafish *hagoromo* (*hag*) gene to be involved in determining cichlid coloration: In addition to accelerated protein evolution in a domain of *hag* in East African cichlids they also found alternatively spliced mRNA variants in this gene in the haplochromines. Both findings are concordant with the expectation that genes involved in color pattern diversity must have either evolved rapidly or undergone dynamic changes that increased their functional potential. Salzburger et al. (2007) found out that the transcription factor *colony-stimulating factor I receptor a* (*csfIra*) is expressed in the xanthophores present in egg spots of haplochromines and ectodines. Moreover they revealed that *csfIra* bears the signature of adaptive sequence evolution in the egg-spot exhibiting haplochromine lineage. Another very recent study of Gunter et al (2011) compared transcription in the blue skin of males to that in the yellow skin of females in a Lake Malawi cichlid to get insights into the molecular basis of cichlid coloration. They found five genes to be definitely differentially expressed. This list included the *Coatomer protein complex, subunit zeta-1* (*Copz-1*) which is known to be an important determinant of pigmentation in humans and zebrafish. Also studies concerning the possible number of genes controlling coloration in cichlid fish have already been conducted: Barson et al. (2007) estimated four to seven loci to be responsible for body color differences between males of two Lake Malawi cichlid species. A similar study by Magalhaes & Seehausen (2010) on males of two Lake Victoria sibling cichlid species showed that red coloration is likely controlled by two to four genes whereas yellow coloration seems to be determined by only one gene with complete dominance. Finally, Gunter et al. (2011)

suggested five genes to underlie color differences between males and females of a Lake Malawi cichlid (as mentioned above).

Thus several approaches have already been pursued towards identifying genes involved in cichlid coloration and estimating their possible number. It is of special importance to identify the molecular mechanisms that drive the evolution of cichlid color patterns because this will be helpful in understanding their explosive radiation. An explanation for their species richness on the genomic level could be the fish-specific genome duplication (e.g. Taylor et al. 2001a, b; Braasch et al. 2006, 2007, 2008, 2009a, b, c): Due to this phenomenon, which occurred 320-350 million years ago, teleost fishes exhibit additional copies of many pigmentation genes. Divergent evolution in such duplicated genes, including lineage-specific gene loss, sub-functionalization, and regulatory evolution (Gunter et al. 2011) could have played a major role in pigmentation diversity in teleost fishes and thus also in cichlids. However, in our study species *Tropheus moorii*, nothing is known yet about the identity or number of genes underlying differences in body coloration.

The genus *Tropheus* currently comprises about 120 mostly allo-or parapatric color morphs all around the lake (Konings 1998; Schupke 2003) which evolved rapidly within the last several hundred thousand years in three major radiation events triggered by severe lake level fluctuations (Sturmbauer et al. 2005; Baric et al. 2003). Another more recent fluctuation 18 000 to 12 000 ya is said to have shaped the present distribution of the lineages (Sturmbauer et al. 2001). Body coloration plays a crucial role in this genus not only during courtship and mating but also for communication during other social interactions, and both genders exhibit congruent repertoires of color patterns signaling motivation and social status (Sturmbauer & Dallinger 1995). However, the origin of the high amount of color polymorphism in *Tropheus* yet remains largely unknown. In some cichlid species of Lakes Victoria and Malawi color divergence has been suggested to be triggered by sexual selection on male nuptial coloration (e.g. Seehausen & van Alphen 1999; Seehausen et al. 1999; Knight & Turner 2004). This has also been discussed for *Tropheus* (Sturmbauer & Meyer 1992; Salzburger et al. 2006) but without conclusive evidence so far. The role of sexual selection in *Tropheus* is hard to assess as *Tropheus* lack some traits typically associated with sexually selected species, like e.g. sexual dimorphism. Further investigations especially on mate choice are needed to get insights into the potential of sexual selection and its possible association with color diversification in this genus (Sefc 2008). Also hybridization has been discussed as a source of new phenotypic variants in cichlids (Salzburger et al. 2002; Smith et al. 2003; Stelkens et al. 2009). Phylogeographic data indicate that in *Tropheus* opportunity for hybridization was

provided due to recurrent cycles of lake level fluctuations (Sturmbauer & Meyer 1992; Baric et al. 2003; Sturmbauer et al. 2005; Egger et al. 2007; Sefc et al. 2007). Several studies showed that water fluctuations generally had severe impact on rocky habitats and the species communities within them in East African rift lakes (Sturmbauer & Meyer 1992; Johnson et al. 1996; Sturmbauer et al. 1997; Sturmbauer 1998; Rüber et al. 1998; Nagl et al. 2000). A severe decline of the water level can lead to secondary contact between formally allopatric morphs and thus to hybridization between them. In *Tropheus moorii* hybrid status has been discussed and demonstrated for some populations in southern Lake Tanganyika (Egger et al. 2007; Mattersdorfer et al. unpublished; see Chapters 3 and 4) but the role of hybridization for the evolution of new color variants is not clear yet. Alternatively, color divergence may merely be a side-product of reproductive isolation between populations due to other cues than color (e.g. Plenderleith et al. 2005; Seehausen et al. 2008) or of random genetic drift (Arnegard et al. 1999). Thus the origin of color diversity in *T. moorii* remains largely unknown. Especially when it comes to the genetic basis underlying it we still lack knowledge.

In this study we hypothesize that differences in body coloration between populations of *T. moorii* have not diverged by neutral forces as genetic drift and mutations alone, but in contrast may be traced back to some genes or genomic regions that are subject to selection. Thus, to detect signals of putatively color-related selection, we carried out a whole-genome scan using amplified fragment length polymorphisms (AFLPs), on 11 *Tropheus moorii* populations belonging to the same AFLP clade (Egger et al. 2007) but to three distinct color morphs from southern Lake Tanganyika. Because of the rapid evolution of different color morphs in this species we would expect to find only a few outlier loci pointing to a small set of genes which could act as switches between different color patterns. In addition to between-color-morph comparisons we also did within-morph comparisons to distinguish loci being outliers due to different body color of the fish from loci being outliers due to some location/population-specific effect. Among the several methods available for outlier detection we primarily chose the program BAYESCAN (Foll & Gaggiotti 2008). For some analyses, to further ensure reliability, we additionally used the selection detection workbench MCHEZA (Antao & Beaumont 2011) which implies the popular DFDIST approach for outlier detection (Beaumont & Nichols 1996; Beaumont & Balding 2004). We considered the results from both analyses in our final set of outlier loci. As to our knowledge this is the first genome scan study that focuses on a trait (i.e. body coloration), which is possibly not shaped by natural selection, as it doesn't obviously represent an adaptation to some ecological and/or environmental condition. This assumption is based on the fact that variability in environmental conditions

between habitats of different *T. moorii* morphs is similar to that between habitats of populations within a morph. Thus it is possible that coloration and outlier loci associated with it in this study are under sexual selection.

In addition to the genome scan, a general characterization of the differences in coloration between the morphs and sexes was done using representative males and females of each morph by taking digital pictures of whole individuals on the one hand and of body regions on the other hand. Furthermore, concerning the yellow/red coloration, underlying pigments were extracted from different body regions in order to test which pigments are present and with which approximate amount.

MATERIALS AND METHODS

Taxon sampling and DNA extraction

Genetic data was obtained from 280 *Tropheus moorii* individuals from 11 different locations in southern Lake Tanganyika (Fig. 1; Tab.1). These populations, belonging to three different color morphs, are geographically and genetically separated but still closely related as they are resolved within the same AFLP clade in a phylogeny of *Tropheus* (Egger et al. 2007, Fig.1). The fish were either collected in the lake during field trips between the years 2003 and 2009, or purchased from ornamental fish traders. Finclips for DNA extraction were taken from each individual and were preserved in 99% ethanol. DNA was isolated using proteinase K digestion followed by a protein precipitation step using ammonium acetate. DNA concentration was measured with a NanoPhotometer^{TM} (IMPLEN) and if required, extracts were diluted with deionized water yielding concentrations of approximately 6 ng/µl for each sample.

Table 1 Color classification and number of individuals in studied populations

Population Code	Location	Body Color Classification	Number of individuals genotyped
Moli	Moliro	Red	26
Kate	Katete	Red	30
Chim	Chimba	Red	31
Ndol	Ndole	Red	5
Kach	Kachese	Red	17
Chai	Chaitika	Blue	24
Naka	Nakaku	Blue	30
Fund	Funda	Blue	28
TanL	Tanganyika Lodge	Yellow-Blotched	30
KasL	Kasakalawe Lodge	Yellow-Blotched	30
Mbit	Mbita Island	Yellow-Blotched	29

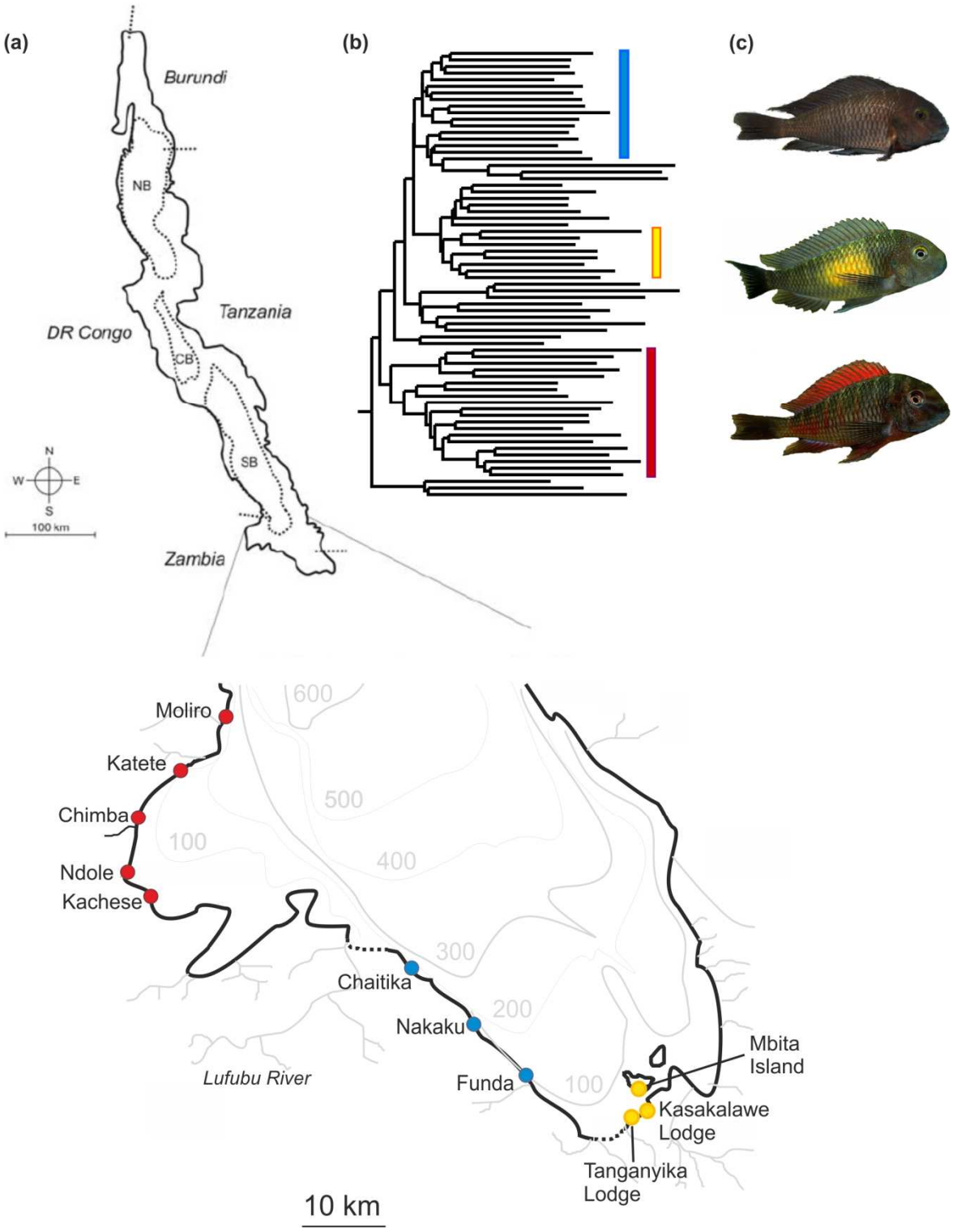


Figure 1 (a) Map of southern Lake Tanganyika showing the locations of the investigated populations. (b) Phylogenetic relationships between tested morphs based on amplified fragment length polymorphism (AFLP). Data from Egger et al. (2007). (c) Phenotypes of the three study morphs.

AFLP analysis

Amplified fragment length polymorphism analyses followed the original protocol of Vos et al. (1995), modified as in Egger et al. (2007) using *Eco*RI and *Mse*I as restriction enzymes. Restriction digestion of 60 ng of DNA (10 µl) was performed using 0.5 µl *Mse*I (10 units/µl, New England Biolabs), 0.25 µl *Eco*RI (20 units/µl, New England Biolabs), 5 µl enzyme buffer (10x), 0.5 µl BSA (100x) and high performance liquid chromatography (HPLC) water yielding a total volume of 50 µl per sample. Incubation took place for three hours at 37°C. For the ligation reaction 1 µl of *Eco*RI adaptor (5 pmol/µl), 1 µl of *Mse*I adaptor (50 pmol/µl), 1 µl of T4 ligase buffer, 0.2 µl of T4 DNA ligase (400 cohesive-end ligation units/µl, New England Biolabs) and 6.8 µl of HPLC water were added to the restriction digestion product, and incubation was performed at 22°C over night. Adapter sequences are shown in Table 2. After dilution of the ligation product with 120 µl HPLC water, 3 µl of it were used in the preselective amplification with 0.4 µl of each pre-selective primer (10 µM), 2 µl 10 x dNTP mix (10 µM), 2 µl 10 x $MgCl_2$ buffer , 12.1 µl HPLC water and 0.1 µl *Taq* DNA polymerase (5 units/µl, BioTherm^{TM}) yielding a total volume of 20 µl. Preselective primers were composed of the adaptor primer sequence with one single selective nucleotide at the 3' end (for *Eco*RI: pre A, for *Mse*I: pre C). Sequences are shown in Table 2. The temperature profile for the preselective PCR was as follows: 2 min at 72°C followed by 20 cycles of 20 sec at 94°C, 30 sec at 56°C and 2 min at 72°C. Then a holding step at 60°C followed for 30 min. For subsequent selective amplification 1 µl of 1:10 diluted preselective PCR product was mixed with 6.6 µl HPLC water, 0.1 µl *Taq* DNA polymerase (5 units/µl, BioTherm^{TM}), 0.8 µl 10 x dNTP mix (10 µM), 1 µl 10 x $MgCl_2$ buffer and 0.25 µl of each selective primer (*Mse*I and *Eco*RI, both 10 µM). The total volume was 10 µl.

The 18 primer combinations used for selective amplification were: EcoRI-ACA/MseI-CAA, EcoRI-ACA/MseI-CAG, EcoRI-ACA/MseI-CAC, EcoRI-ACA/MseI-CAT, EcoRI-ACT/MseI-CAT, EcoRI-ACT/MseI-CAA, EcoRI-ACT/MseI-CAG, EcoRI-ACT/MseI-CAC, EcoRI-ACC/MseI-CAA, EcoRI-ACC/MseI-CAG, EcoRI-ACC/MseI-CAC, EcoRI-ACC/MseI-CAT, EcoRI-ACC/MseI-CTG, EcoRI-ACA/MseI-CTG, EcoRI-ACA/MseI-CTA, EcoRI-ACA/MseI-CTC, EcoRI-ACT/MseI-CTA, and EcoRI-ACT/MseI-CTC.

Selective amplification took place with the following temperature profile: 2 min of 94°C followed by 10 cycles with 20 sec at 94°C, 30 sec at annealing temperature (touchdown settings: decrease of 1°C per cycle starting at 65°C) and 2 min at 72°C. Then there followed

25 cycles with 20 sec at 94°C, 30 sec at 56°C and 2 min at 72°C and finally followed by a holding step at 60°C for 30 min.

Table 2 Sequences of adaptors and primers used for ligation and PCR

	Name	Sequence
Adaptors	EcoRI ad A	5'-ctc gta gac tgc gta cc-3'
	EcoRI ad B	5'-aat tgg tac gca gtc tac-3'
	MseI ad A	5'-gac gat gag tcc tga g-3'
	MseI ad B	5'-tac tca gga ctc at-3'
Pre-selective Primers	EcoRI – pre A	5'-gac tgc gta cca att ca-3'
	MseI – pre C	5'-gat gag tcc tga gta ac-3'
Selective Primers	EcoRI-ACA	5'-gac tgc gta cca att cac a-3'
	EcoRI-ACT	5'-gac tgc gta cca att cac t-3'
	EcoRI-ACC	5'-gac tgc gta cca att cac c-3'
	MseI-CAA	5'-gat gag tcc tga gta aca a-3'
	MseI-CAG	5'-gat gag tcc tga gta aca g-3'
	MseI-CAC	5'-gat gag tcc tga gta aca c-3'
	MseI-CAT	5'-gat gag tcc tga gta aca t-3'
	MseI-CTG	5'-gat gag tcc tga gta act g-3'
	MseI-CTA	5'-gat gag tcc tga gta act a-3'
	MseI-CTC	5'-gat gag tcc tga gta act c-3'

As there were too many samples to be processed in a single restriction - ligation procedure and PCR, respectively, the samples were processed in several batches. To detect restriction - ligation specific or PCR specific peaks (PCR noise) every batch contained at least eight replicate samples that had already been processed in the first set of reactions. Thus artifact, batch-specific peaks could be later on identified and excluded. Furthermore to check the potential of contamination of chemicals and the possibility of instrument noise from the sequencer entering the data set one negative control (consisting of reaction chemicals but water instead of DNA) per restriction - ligation and per PCR reaction was included in the analysis.

Fluorescently labeled products of selective amplification were separated and visualized using an ABI 3130xl automated sequencer (Applied Biosystems) along with an internal size standard (GeneScan-500 ROX, Applied Biosystems).

AFLP fragment scoring

Fragment size and peak height detection was performed using Genemapper v.3.7 (Applied Biosystems) with only fragments between 90 and 500 base pairs being considered. Every bin in each fingerprint was checked by eye and preprocessed as follows: (i) Bin positions were set manually in order to adjust misaligned bins. It was of special importance that fragments within one bin had to be definitely distinguishable from fragments in adjacent bins in order to avoid size homoplasy (Arrigo et al. 2009). (ii) Bins containing obviously PCR-specific peaks, which could be detected through the replicate samples, were removed. (iii) Bins containing exclusively ambiguous peaks with very low intensities were considered as non-informative loci and thus deleted. (iv) As peak quality tends to decrease in some samples towards the higher fragment-length areas of the fingerprint, and thus presence or absence of peaks is ambiguous, such bins were removed. (v) Partially failed or failed profiles, that is, for example, profiles with repeadetly very low peak heights or read-length relative to the average, were excluded. Such preprocessed but still unnormalized peak-height data was then exported from Genemapper as a tab-delimited text file for automated normalization and final scoring in the R-script (R Development Core Team 2009) program AFLPSCORE 1.4a (Whitlock et al. 2008). This recently developed method provides an objective way for scoring AFLP profiles by identifying the optimum parameter values, that is, the optimum thresholds, via the calculation of two error rates: First the 'mismatch error rate' is calculated which represents the percentage of differences in the AFLP profile between the original and the replicate of a duplicated sample. Therefore 56 samples were fingerprinted twice starting from independent restriction digestion reactions. This resulted in a minimum (depending on how many replicate samples yielded good quality profiles) of 32 replicate pairs per selective primer combination. Secondly the 'Bayesian error rate' is computed, which describes the error process at the allelic level. Briefly, it is the probability of miss-scoring a 1 allele as a 0 allele. Both error rates are used to optimize the two parameters chosen for scoring which are (i) the locus selection threshold, that determines the minimum peak height at loci, averaged over all samples, that are retained in the analysis; and (ii) the phenotype calling threshold, which determines whether an allele at a locus is scored present or absent in a given sample. Optimum thresholds are those that minimize the error rates while retaining a reasonable number of loci. The phenotype calling threshold can be given as an absolute or a relative value. Depending on the number of retained loci and the achieved error rates the decision between absolute or relative phenotype calling was done for each primer combination independently. The analysis with

AFLPSCORE was performed for each primer combination separately resulting in one binary (1/0) matrix per primer pair. The combination of manual preprocessing of the data and automated scoring implemented in AFLPSCORE provides a good method for the improvement of AFLP data quality although of course the number of marker bands per primer combination is reduced in relation to totally automated scoring. For further analyses the matrices were assembled resulting in a final binary data matrix with 1160 characters.

AFLP diversity

We used AFLP-SURV version 1.0 (Vekemans 2002; Vekemans et al. 2002) to estimate allele frequencies, calculate the total number of segregating loci (i.e. fragments that are not always present nor always absent in all individuals), the proportion of polymorphic loci in terms of at least 5% presence or absence of the band in each population, estimated heterozygosity values for each population (i.e. Nei's gene diversity), the observed value of F_{ST} over all populations, and a distance matrix of F_{ST} values between every pair of populations. Under the assumption of Hardy-Weinberg equilibrium (based on the analysis of microsatellite loci, e.g., see Chapter 3) the program estimates allele frequencies from the binary presence/absence matrix, assuming that markers are dominant with two alleles per locus (presence of the band: dominant; absence of the band: recessive). We used the default option, that is, Bayesian method with non-uniform prior distribution of allele frequencies (Zhivotovsky 1999) as recommended by the authors. Such obtained allele frequencies were used for subsequent calculations. AFLP-SURV computes all statistics of genetic diversity and population genetic structure following the method of Lynch and Milligan (1994). The significance of the observed F_{ST} values was tested with 5 000 random permutations where individuals are randomly permuted among the populations and F_{ST} is recalculated for each permutation. The observed value of F_{ST} is then compared to the distribution of randomized F_{ST} values. As AFLP-SURV only gives a *P*-value for the overall observed F_{ST} but not for each F_{ST} between single pairs of populations in the overall computation, calculations of F_{ST} were additionally conducted population-pair-wise to get these *P*-values. For all AFLP-SURV analyses the geographically very close populations from Ndole and Kachese were pooled because of low sample size in both. This resulted in a total of ten populations.

Outlier detection

To detect outlier loci with significantly higher differentiation than under neutral expectations and thus representing candidate loci for selection, we primarily used the program BAYESCAN (freely available at http://www-leca.ujf-grenoble.fr/log-iciels.htm). This hierarchical Bayesian model-based method was first described in Beaumont & Balding (2004) and has recently been improved by Foll & Gaggiotti (2008) for use with dominant markers. It directly estimates the posterior probability that a locus is subject to selection. BAYESCAN assumes that the investigated populations are subpopulations that split from a hypothetical ancestral population and evolved in isolation since then. The subpopulations may have been influenced by various degrees of genetic drift and hence they will show differences concerning the amount of differentiation from the ancestral allele frequency. Thus the primary advantage of this method over other methods commonly used for outlier detection is that it implies the estimation of population specific F_{ST} coefficients (represented as β) and therefore allows for different demographic scenarios and variable amounts of genetic drift in each subpopulation. The method is based on a logistic regression model in which each logit value of genetic differentiation F_{ST} (*i*,*j*) for locus *i* in population *j* is decomposed as a linear combination of the coefficients of the logistic regression, α_i and β_j, corresponding, respectively, to a locus effect (i.e. selection) and a population effect (i.e. genetic drift). The posterior probability of a locus *i* being under selection is estimated by using two alternative models, one that includes α_i (i.e. $\alpha_i \neq$ zero, indicating selection) and one that excludes it (i.e. α_i = zero, indicating neutrality). For the estimation of the posterior probabilities of the two models a combination of a MCMC and a reversible jump MCMC algorithm is used (Foll & Gaggiotti 2008). Counting the number of times that α_i is included in the model, the posterior probability that a locus is under selection, corresponding to P ($\alpha_i \neq 0$), is estimated. Since this probability cannot directly be compared to a *P*-value as in the FDIST software by Beaumont & Nichols (1996) and Beaumont & Balding (2004), the decision which model is chosen is performed by a 'Bayes Factor' which provides a scale of evidence in favor of one model versus the other (Foll & Gaggiotti 2008). We made the final decision that a locus is under selection by looking at the $\log_{10}$ of the Bayes Factor and considered a locus with a $\log_{10}(BF) > 0.5$ (which would correspond to P ($\alpha_i \neq 0$) > 0.76) as substantially under selection following Jeffreys' scale of evidence (Jeffreys 1961). With an increasing $\log_{10}(BF)$ the reliability that a locus is an outlier also increases. A $\log_{10}(BF) > 2$, for example (corresponding to P ($\alpha_i \neq 0$) > 0.99), would indicate a decisive signal of selection (Foll & Gaggiotti 2008).

A recent study of Pérez-Figueroa et al. (2010), comparing three different methods for the detection of selective loci using dominant markers, showed that BAYESCAN appeared to be the most efficient one under different scenarios in both, detecting a high percentage of true outlier loci, and discovering a very low amount of false positives (less than 1%). The authors stated another big advantage of BAYESCAN: through the prior distribution it takes all loci into account in the analysis, and thus resolves the problem of multiple testing of a large number of genomic locations, which is a problem with other methods used for outlier detection. Also another study was published recently that used simulations to test a new version of BAYESCAN which includes the distribution of band intensities at a given locus for estimation of allele frequencies (Fischer et al. 2011; method described in Foll et al. 2010). The main focus of the authors was to show that the power of outlier detection increases when implementing the new approach using band intensity data instead of mere presence/absence information. But also they showed that the specificity of BAYESCAN to detect outliers (i.e. 1 – false positive rate) lies at 99,9% even when using AFLP as binary data. This indicates that the false positive rate should be indeed very low in our data set as well. Concerning other markers BAYESCAN seems to perform well too as shown in a comparative simulation study by Narum et al. (2011) using SNP loci: BAYESCAN yielded the overall lowest false positive rate among the methods compared and also a high power to detect loci under selection (i.e. a low false negative rate) depending on the scenario.

In our BAYESCAN analysis the estimation of model parameters was automatically tuned on the basis of pilot runs which also play the role of a burn-in period. We conducted 10 pilot runs with a length of 2 000 iterations each and an additional burn-in of 10 000 iterations, as preceding tests indicated that this was enough to achieve convergence in the MCMC (Foll & Gaggiotti 2008). As suggested by the authors we used a sample size of 10 000 and a thinning interval of 50 resulting in a total length of the chain of 500 000. We carried out pair-wise BAYESCAN analysis between (i) the pooled sets of color morphs, i.e. all Blue populations pooled together, all Red populations pooled together and all Yellow-Blotch populations pooled together resulting in three big populations (Blue, Red, Yellow-Blotch), and (ii) the single populations of different color morphs. These two approaches were done in order to detect loci that were outliers presumably due to different body coloration. For comparative reasons we additionally carried out analysis between (iii) the single populations within a color morph in order to detect loci that are outliers due to some location/population-specific effect. That resulted in a total of 42 possible pair-wise comparisons (three pooled comparisons and 39 single population comparisons). For the within color morph comparisons

the geographically close populations from Ndole and Kachese were pooled together because of low sample size in each of them. From between-morph population comparisons they were even excluded. To avoid a downward bias of the overall F_{ST}, only loci that were polymorphic (in terms of at least 5% presence or absence of the band) between the pooled sets of color morphs for between-morph outlier detection, and within a color morph for within-morph outlier detection, respectively, were retained in the analysis. This is important because a large amount of monomorphic loci could lead to some false positives under directional selection because loci with a high F_{ST} appear more extreme as the F_{ST} is biased downward (pers. comm. with M. Foll).

To further ensure the reliability of between-morph outliers being detected by BAYESCAN we used another outlier detection approach which we applied only on the pooled pair-wise between-color-morph comparisons. The method uses the DFDIST approach for outlier detection (Beaumont & Nichols 1996; Beaumont & Balding 2004) and implies it in the selection detection workbench MCHEZA (Antao & Beaumont 2011; freely usable at http://popgen.eu/soft/mcheza/). MCHEZA is based on the workbench LOSITAN that is used for outlier detection with co-dominant markers (Antao et al. 2008). In brief, the DFDIST method works in several steps: First an empirical distribution of F_{ST} values among loci for pairs of populations is generated using a Bayesian method developed by Zhivotovsky (1999) for estimating allele frequencies from the proportion of recessive phenotypes in the sample. Furthermore the Weir & Cockerham (1984) F_{ST} between the subgroups defined in the sample is estimated. Then, to generate a null distribution of F_{ST} values, it performs coalescent simulations. The suchlike simulated data provide a baseline F_{ST} distribution against which potential outlier loci can be statistically evaluated. The decision if a locus is an outlier or not is made by the probability that the simulated F_{ST} is smaller than the sample F_{ST} (P (simulated F_{ST} < sample F_{ST})). An important feature in MCHEZA is that it implies a support for multitest correction based on false discovery rates which is crucial to avoid false positives. We ran MCHEZA with the following settings: 10 000 simulations were conducted with a confidence interval of 95%. We left the false discovery rate at the default of 0.1. We also left the Theta value (defined by two times the mutation rate (per site per generation) times the number of heritable units in the population), the Beta-a and the Beta-b value (the priors of the beta distribution; see Zhivotovsky 1999) at the defaults of 0.1, 0.25 and 0.25, respectively. We chose the function 'Neutral mean F_{ST}' to remove potential selected loci in a first simulation run for computing the initial mean F_{ST} uninfluenced by outliers. Also we used the function 'Force mean F_{ST}' which tries to simulate a precise mean F_{ST} by running a bisection algorithm

over repeated simulations. Both options are recommended by the authors. We conducted three runs comparing only the pooled data sets, again with only polymorphic markers left in the data set.

In an attempt to maximize the reliability of between-morph outliers detected in our study we only considered those loci as potentially under selection that were detected as outliers (i) in the pooled comparisons with both methods as well as (ii) in at least five of the nine cross-morph population pair-wise comparisons with the BAYESCAN method. The thresholds we chose for loci being considered as outliers were a $\log_{10}$ of the Bayes Factor bigger than 0.5 (BAYESCAN) and a probability of the simulated F_{ST} being smaller than the sample F_{ST} of at least 0.95 (MCHEZA), respectively.

Test for linkage disequilibrium

As a first approach to obtain information about the genomic distribution of the detected outlier loci, we used the program DIS (developed by Jim Mallet) to estimate pair-wise linkage disequilibria between them. The program assumes Hardy-Weinberg equilibrium at each dominant locus and, following the equation of Hill (1974), calculates maximum-likelihood estimates of the pair-wise gametic disequilibrium coefficient *D* between two loci within a population. Also it computes estimates of the pair-wise gametic correlation coefficient *R*, which corrects for variable allele frequencies, between pairs of loci following the method described in Dasmahapatra et al. (2002). Thus inferences of the likelihood of two loci being in physical linkage can be drawn by looking at *R*. Analyses of linkage disequilibrium were carried out first, on our final sets of six (Blue vs. Yellow-Blotch), five (Red vs. Blue), and four (Red vs. Yellow-Blotch) outlier loci, respectively, and additionally on 40 randomly selected neutral loci (as the program is limited to a maximum of 40 loci). We expected that, if physical linkage between outlier loci existed, then their level of linkage disequilibrium should be increased compared to neutral loci. This could, for example, be caused by selective sweeps (Kim & Nielsen 2004). To test for significant differences in linkage disequilibrium between the outlier sets and putatively neutral markers we compared the mean gametic correlation coefficient *R* measured for each population across outliers to the distribution of 1 000 averaged values, which were calculated from random draws of an equal number of neutral markers. We did this using an ad hoc R-script (R Development Core Team 2009).

As the gametic correlation coefficient *R* cannot be computed when at least one of the two loci is fixed for either allele, we manipulated our input files, such that we changed the

genotype of one randomly chosen individual in a population at a fixed locus. This way, unobserved alleles were represented as rare alleles and we don't expect those slight alterations to change the results and conclusions of this test.

Color analyses

Characterization of differences in coloration between the morphs was done using three approaches: (i) Digital pictures of whole individuals were either taken from former publications (Photos of the Yellow-Blotch and the Red morph were used in Egger et al. 2007, provided by Peter Berger) or taken with an Olympus digital camera (Blue Morph). Overall amounts of yellow/red, dark, and blue coloration were inferred visually. (ii) Digital pictures of body regions were taken from one representative individual of each color morph with an Olympus digital camera (Olympus E-1 – Zuiko Digital 14-54mm f 2.8-3.5) mounted on an Olympus SZX-ILLB2-200 binocular eyepiece using the software Olympus Viewer and Studio (Copyright 2003 Olympus Corporation). (iii) Concerning the yellow/red coloration, underlying pigments were extracted from different body regions of representatives of each color morph in order to test which pigments were present and with which amount. Yellow to red coloration in fish is known to be based on one or both of the following color pigments: Carotenoids and pteridines. These reside in pigmentary organelles within special pigment cells: The xanthophores and the erythrophores. Most animals have lost the ability to synthetisize carotenoids *de novo* (Goodwin 1984; Schiedt 1989) and thus they have to be obtained through food. In contrast to that, pteridines are derived from purines, which are synthetisized *de novo* within the pterinosomes of the pigment cells (Hurst 1980; Braasch et al. 2008). To find out if one or both of these pigments were present in our *Tropheus moorii* morphs, and if so, to what approximate extend in which region of the body, we carried out skin-pigment analysis.

Skin-pigment analysis was conducted on a various number of representatives of each color morph, depending on availability of fish: One *T. moorii* Chipimbi female and three T. *moorii* Moliro males, representing the Red morph, one *T. moorii* Chaitika male and one *T. moorii* Nakaku female, representing the Blue morph, and two *T. moorii* Mbita individuals, one male and one female, representing the Yellow-Blotch morph. We did not use individuals from Chipimbi for our DNA analyses as there were wasn't a sufficient number of samples available. However, they represent members of the Red morph (Chipimbi is geographically located between Katete and Moliro) thus we utilized one sample of this population for the

color analyses. Individuals used for the extraction were fishes that had died due to natural causes some time in the past and had been stored at -20°C since then. The body was divided into 11 compartments excluding the head. Tissue from fins was also included in the analysis. Therefore the dorsal fin was divided into four fractions, and the anal fin into two fractions. The pectoral fin was taken on the whole. This resulted in 18 tissue samples for pigment extraction (Fig. 2). Additionally one pure muscle tissue sample was taken to make sure that no pigments were present in the muscle.

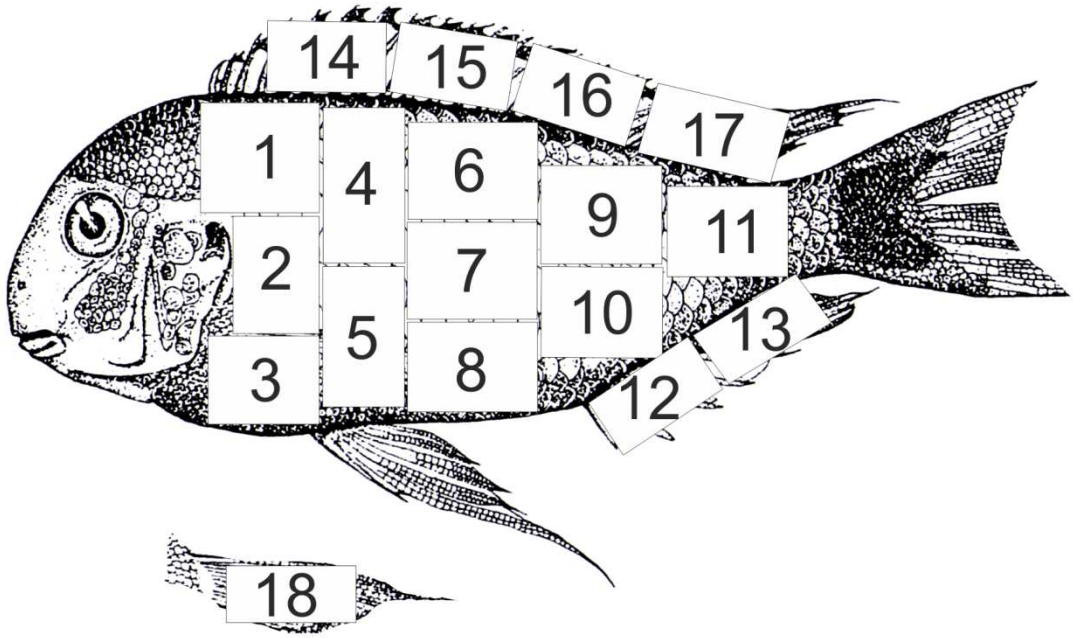


Figure 2 The 18 compartments from which pigments were extracted in each fish. The picture was modified from Poll (1986).

For the 11 ‘body-samples’ the skin, partially including scales, was stripped from the body with surgical instruments. Special care was taken in removing as much as possible of the muscle tissue beyond the skin. For ‘fin-samples’ fins were cut into compartments approximately the same size, including spines. Samples were dried for a few minutes, weighted and put in 1.5 ml Eppendorf tubes each. Carotenoid and pteridine extraction followed the protocol of Grether et al. (2001). Carotenoids were extracted in acetone with ≥ 99.8% purity. Therefore 100 µl of acetone were added to each tissue sample and extraction took place over night in a thermomixer at 650 rpm and room temperature. After that the acetone, now containing extracted carotenoids, was transferred into a new vial. To concentrate the carotenoids and to remove the acetone, samples were evaporated under a flow of nitrogen for approximately 20 minutes. Carotenoids were redissolved in 10 to 300 µl hexane (depending on color intensity) for absorption measurements.

Pteridines were extracted from the carotenoid-stripped skin using 30% acidified (with HCL to pH 2) ethanol. Therefore 20-50 µl (depending on tissue size) of acidified ethanol were added to each sample and again extraction took place over night in a thermomixer at 650 rpm and room temperature. As a positive control the same procedure was applied to five heads of *Drosophila sp*. flies as the pteridine drosopterin is known to be present in their eyes (e.g. Forrest & Mitchell 1955). Absorption measurements were directly conducted from the extract. Both extractions were also conducted with a null sample, including everything but tissue, as a negative control. Absorption measurements were carried out with a NanoPhotometer™ (IMPLEN) using a volume of 3 µl pigment solution for each measure. Absorption spectra were recorded between 300 and 600 nm. It was not our intention to measure absolute concentrations but instead to (i) merely test if both classes of pigments were present or only one of them, and (ii) to infer the approximate relative amount of pigments compared between different body regions and between different color morphs. Thus absorbance (A) of extracts was noted at the peak of absorption in each skin- and fin-sample and compared with other samples and other morphs taking into account the weight of the tissue and the volume of the solvent (i.e. hexane for carotenoids and 30% acidified ethanol for pteridines, respectively).

RESULTS

DNA analysis

Genetic diversity

The 18 primer combinations yielded a total of 1160 AFLP loci of which 1068 (92.1%) were segregating within the overall data set. The mean number of AFLP bands per individual was 279 and per primer combination it was 64 loci, ranging from 22 to 116. The mean proportion of polymorphic markers per population was 39.6%, ranging from 36.5% to 43.4%, and the average expected heterozygosity (i.e. Nei's gene diversity) was 0.124 ranging from 0.107 to 0.137 (Tab. 3). The observed F_{ST} value over all populations was 0.1944 ($p < 0.0001$). Pair-wise F_{ST} values were overall high between populations belonging to different color morphs and lower between populations of the same color morph (Tab. 4). Highest mean genetic distances were found between Red and Yellow-Blotch populations with 0.309 (range: 0.266 – 0.350) followed by Red – Blue with 0.238 (range: 0.202 - 0.277) and then Blue – Yellow-Blotch with 0.101 (range: 0.076 – 0.122). *P*-values between every single population-pair were highly significant (< 0.0001, Tab.4) even in cases where F_{ST} values were low.

Table 3 Genetic diversity measures of the ten sampled populations calculated with AFLP-SURV v1.0

Population	n individuals	PLP [%]*	Hj^{+}	Hj^{+} S.E.
Moli	26	40.8	0.13	0.0048
Kate	30	43.4	0.133	0.0047
Chim	31	41.6	0.132	0.0049
Nd-Ka	22	40.9	0.137	0.0049
Chai	24	37.2	0.125	0.0048
Naka	30	40.3	0.124	0.0048
Fund	28	38.5	0.123	0.0048
TanL	30	37.7	0.109	0.0045
KasL	30	38.9	0.118	0.0046
Mbit	29	36.5	0.107	0.0046

* Proportion Loci Polymorphic, that is, percentage of polymorphic loci at the 5% level.

$^{+}$ Expected heterozygosity under Hardy-Weinberg genotypic proportions, that is, Nei's gene diversity.

Table 4 Matrix of genetic distances between all ten populations calculated with AFLP-SURV version 1.0. F_{ST}-values are in bottom left corner, *P*-values in upper right corner. Calculation of *P*-values was based on 5000 permutations

	Moli	Kate	Chim	Nd-Ka	Chai	Naka	Fund	TanL	KasL	Mbit
Moli		<0.0001	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001
Kate	0.030		<0.0001	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001
Chim	0.098	0.056		<0.0001	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001
Nd-Ka	0.132	0.090	0.038		<0.0001	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001
Chai	0.262	0.241	0.216	0.202		<0.0001	<0.0001	<0.0001	<0.0001	<0.0001
Naka	0.274	0.247	0.225	0.214	0.008		<0.0001	<0.0001	<0.0001	<0.0001
Fund	0.277	0.255	0.228	0.216	0.032	0.032		<0.0001	<0.0001	<0.0001
TanL	0.337	0.324	0.298	0.278	0.101	0.109	0.076		<0.0001	<0.0001
KasL	0.323	0.311	0.289	0.266	0.105	0.108	0.078	0.012		<0.0001
Mbit	0.350	0.337	0.311	0.291	0.114	0.122	0.095	0.025	0.037	

Outlier detection

Between color morph outliers After removing monomorphic loci at the 5% level between pooled sets of color morphs 419 (Blue vs. Yellow-Blotch), 513 (Blue vs. Red) and 526 loci (Red vs. Yellow-Blotch), respectively, remained in the analysis. We detected a total of 11 outliers that fulfilled all criteria, that is, detection in the pooled comparison by BAYESCAN and MCHEZA and detection in at least five of the population-pair-wise analyses by BAYESCAN with a $\log_{10}(BF) > 0.5$ (Tab. 5). Those markers were distributed over seven primer combinations. Outlier loci were named by the last two selective bases of the primers with which they were obtained (*Eco*RI in capital letters, *Mse*I in lower-case letters) plus the length of the fragment. According to the color morphs between which they were found, the outlier loci could be divided in two sets: (1) Six of the 11 outliers were found in comparisons between Blue and Yellow-Blotch populations (CAag104, CAag295, CAac334, CCat103, CCat322, CAtg200). Two of these were found exclusively in Blue vs. Yellow-Blotch comparisons (CAag104 and CCat103) and four were also found between other morphs: CAag295 was also detected in the pooled MCHEZA analysis between Blue and Red as well as between Katete and Nakaku. CAac334 and CAtg200 were detected in both pooled analyses between Red and Yellow-Blotch and additionally in four and three population pair-wise comparisons, respectively. CCat322 was found also in the pooled MCHEZA comparison between Blue and Red. (2) Five outliers were detected between Blue and Red populations (CTaa236, CTag236, CCat240, CAtg317, CAtc464) and four of them also between Red and

Yellow-Blotch populations (CTaa236, CTag236, CCat240, CAtg317). None of these five loci were detected between the Blue – Yellow-Blotch analyses.

Table 5 $Log_{10}BF$ values (BAYESCAN) and probability values (MCHEZA, in italic letters) of the final set of 11 AFLP markers detected to be under selection in comparisons between the three color morphs. Results from the pooled comparisons are in bold letters

Blue – Yellow-Blotch	(total number of loci: 419)										
	CAag 104	**CAag 295**	**CAac 334**	**CCat 103**	**CCat 322**	**CAtg 200**	**CTaa 236**	**CTag 236**	**CCat 240**	**CAtg 317**	**CAtc 464**
Pooled BS	**1.48**	**2.23**	**1.48**	**0.80**	**1.66**	**1.50**					
Pooled Mcheza	***1.00***	***1.00***	***1.00***	***0.99***	***0.99***	***1.00***					
Chai-TanL	0.93	0.76	0.58		2.01	0.85					
Chai-KasL	0.52	0.60		1.43	1.71						
Chai-Mbit	1.07	0.57		1.36	1.57	2.02					
Naka-TanL	1.01	3.22	0.91		1.42						
Naka-KasL	0.68	3.22		0.57	1.35						
Naka-Mbit	1.14	2.52	0.70		1.11	1.27					
Fund-TanL	0.86		1.69			1.23					
Fund-KasL			0.93	0.69		0.66					
Fund-Mbit	1.09		1.59	0.66		2.55					
Red - Blue		(total number of loci: 513)									
	CAag 104	**CAag 295**	**CAac 334**	**CCat 103**	**CCat 322**	**CAtg 200**	**CTaa 236**	**CTag 236**	**CCat 240**	**CAtg 317**	**CAtc 464**
Pooled BS							**1.33**	**3.22**	**2.42**	**2.04**	**1.43**
Pooled Mcheza		*0.95*			*0.97*		***0.98***	***1.00***	***1.00***	***0.99***	***0.97***
Moli-Chai							0.55	1.39	1.32		
Moli-Naka								1.23		1.30	1.30
Moli-Fund							1.18	0.58	1.14		1.16
Kate-Chai							0.93	1.83	1.69	0.72	
Kate-Naka		0.71					0.67	1.82	0.61	1.79	1.02
Kate-Fund							1.70	0.88	1.60	1.59	0.85
Chim-Chai							1.19	2.15	2.09	0.98	
Chim-Naka							0.91	2.08	0.84	2.09	1.35
Chim-Fund							1.93	1.25	2.07	1.97	1.14
Red – Yellow-Blotch	(total number of loci: 526)										
	CAag 104	**CAag 295**	**CAac 334**	**CCat 103**	**CCat 322**	**CAtg 200**	**CTaa 236**	**CTag 236**	**CCat 240**	**CAtg 317**	**CAtc 464**
Pooled BS			1.00			0.75	**1.43**	**2.16**	**1.40**	**1.08**	
Pooled Mcheza			*0.96*			*0.96*	***0.97***	***0.99***	***0.98***	***0.96***	
Kate-TanL			1.03				1.03	0.96	0.92	0.98	
Kate-KasL							1.08	1.03		1.08	
Kate-Mbit			0.73			0.79	0.92	0.77	0.81		
Moli-TanL											
Moli-KasL							0.80	0.73		0.77	
Moli-Mbit			0.52			0.56	0.57	0.57	0.61		
Chim-TanL			0.54				1.13	1.07	1.11	1.16	
Chim-KasL							1.15	1.21		1.11	
Chim-Mbit						0.97	1.07	0.89	1.02		

An example for an output plot created by BAYESCAN shows the results of the pooled Blue vs. Yellow-Blotch comparison (Fig. 3). With the threshold set to a $\log_{10}$(BF) of 0.5, 11 outliers were detected in this analysis. Six of these were within our final set of outlier loci and were labeled in the graph.

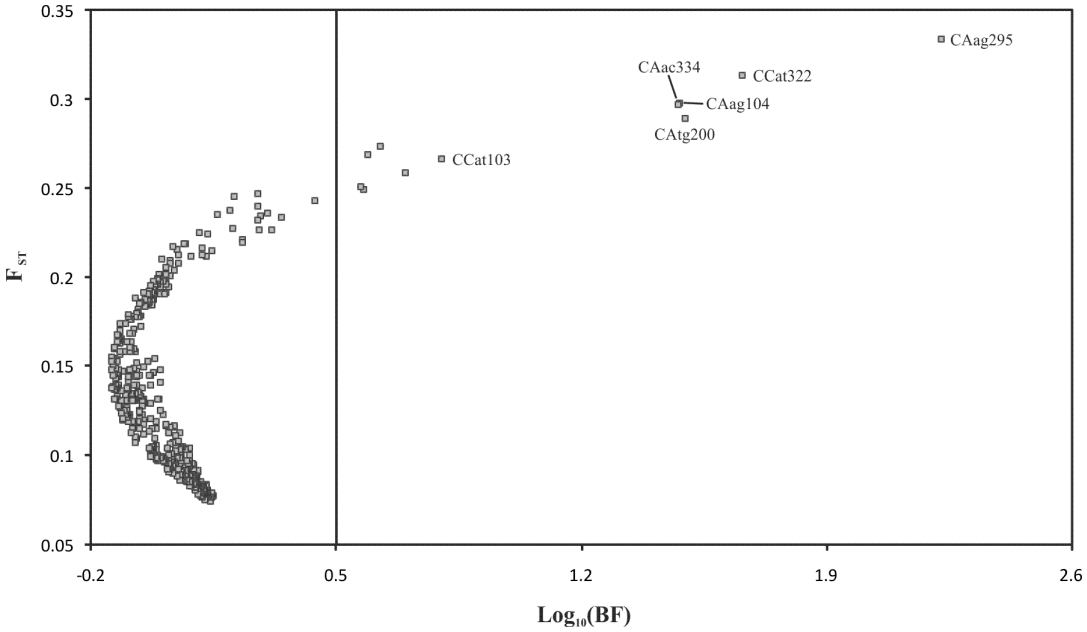


Figure 3 BAYESCAN plot of 419 polymorphic AFLP loci from the pooled Blue vs. Yellow-Blotch comparison. F_{ST} values are plotted against the log10 of the Bayes Factor. The vertical line at 0.5 indicates the threshold above which loci were identified as outliers. Labeled loci are those that were within our final set of outlier loci.

Population-specific marker allele frequency estimates calculated with AFLP-SURV v1.0 for the final set of 11 outlier loci showed that for all outliers allele frequency estimates displayed an association with body coloration in most of the populations (Tab. 6): Differences between populations within a color morph were mostly very low. In contrast to that, differences between populations belonging to distinct color morphs were either low, if the locus was not under selection between the two morphs, or very high if it was so. However, it has to be pointed out here that a locus was sometimes not detected as being under selection between two morphs even when allele frequency differences were considerably high. In three cases we found a departure from the pattern of consistency of frequencies within a color morph: (i) At locus CAag295 the Nakaku population showed a rather high allele frequency (0.736) compared to the other two blue populations. (ii) The Katete population displayed a considerably lower allele frequency (0.053) than the other three red populations at locus CCat103. (iii) Finally, at locus CCat322 the allele frequency in the Funda population (0.095) was much lower than in the other two blue populations. This was probably due to some population specific feature not associated with body color (see discussion).

Within-morph outliers After removing monomorphic loci at the 5%-level 486 loci remained in the Red-morph analyses, 394 in the Blue-morph analyses and 132 in the Yellow-Blotch-morph analyses. With 49 outliers (10.1% of the tested loci), the highest number of outliers with a $Log_{10}(BF) > 0.5$ was detected between populations within the Red morph, 26 of them in more than one comparison. Analyses between populations within the Blue morph revealed 20 outliers (5.1%), eight of which in more than one comparison. Only five outlier loci (3.8%) were detected in analyses within the Yellow-Blotch morph, three of these in more than one comparison (Tab. 7). With the exception of three loci (CAag295, CCat103, CCat322) none of these outliers occurred as such in the between-morph comparisons.

Table 6 Allele frequencies of the AFLP loci detected under selection, estimated with AFLP-SURV v1.0 and comparisons between which they were detected. Values exceeding 0.2 are underlined to represent allele frequency differences between color morphs more clearly

Color Morph	Population	*CAag104*	*CAag295*	*CAac334*	*CCat103*	*CCat322*	*CAtg200*	*CTaa236*	*CTag236*	*CCat240*	*CAtg317*	*CAtc464*
Red	Moli	0.002	0.002	0.916	0.231	0.002	0.916	0.002	0.916	0.916	0.916	0.916
	Kate	0.019	0.002	0.914	0.053	0.002	0.914	0.002	0.914	0.914	0.914	0.804
	Chim	0.002	0.018	0.809	0.595	0.018	0.920	0.002	0.920	0.920	0.920	0.809
	Nd-Kach	0.003	0.023	0.905	0.540	0.003	0.905	0.045	0.903	0.905	0.781	0.905
Blue	Chai	0.023	0.387	0.354	0.541	0.589	0.642	0.786	0.002	0.002	0.044	0.066
	Naka	0.019	0.736	0.422	0.367	0.515	0.551	0.736	0.002	0.053	0.002	0.002
	Fund	0.057	0.268	0.464	0.319	0.095	0.619	0.919	0.020	0.002	0.002	0.002
Yellow-Blotch	TanL	0.559	0.001	0.001	0.051	0.001	0.051	0.932	0.001	0.001	0.001	0.018
	KasL	0.516	0.002	0.018	0.002	0.002	0.088	0.924	0.002	0.035	0.002	0.035
	Mbit	0.677	0.001	0.001	0.001	0.001	0.001	0.933	0.001	0.001	0.019	0.091
Detected between		B – Y	B – Y (B – R)*	B – Y (R – Y)*	B –Y	B –Y (B – R)*	B – Y (R – Y)*	B – R R – Y	B – R R – Y	B – R R – Y	B – R R – Y	B – R

R …. Red morph
B …. Blue morph
Y …. Yellow-Blotch morph
* Parentheses indicate that particular loci were also detected in one or more comparisons between these two color morphs but not often enough to be considered reliable (see Table 5).

Table 7 AFLP markers detected under selection in pair-wise BAYESCAN comparisons within color morphs. Underlined markers were detected in more than one comparison within the same color morph. Markers written in italic letters were repeatedly detected in population comparisons within more than one color morph. Markers that are written in bold letters were also detected in between morph comparisons. For each comparison outliers are ordered according to their log_{10}(BF) values

Color morph	Populations	log_{10}(BF)	ID of outliers detected
Red	Moli-Kate	> 2	CCac93
		1.5 – 2	---
		1 – 1.5	CTaa216, CCat237, CAtc240
		0.5 – 1	CAaa111, CTac388, CCag94, CCag406, CCat226, CAtg208, CAta274
	Moli-Chim	> 2	CTac293, CCac93, CAtc186.1
		1.5 – 2	CCac223.1
		1 – 1.5	CAtc240
		0.5 – 1	CAac173, CAat128, CAat298, CTat138, CTac239, CCaa232, CCac143, CCac145, CAtg208, CAtg415, CTta196, CTtc361
	Moli-Nd/Kach	> 2	CTac239, CCac223.1
		1.5 – 2	CAat128, CCac93, CAtc352
		1 – 1.5	CAac156, CTat215, CAtc186.1, CAtc240
		0.5 – 1	CAaa226, CTaa345, CTaa345.1, *CTag194*, CCaa232, CCag116, CCtg286, CAtg155, CAtg401, CAta149, CTta196, CTtc361
	Chim-Kate	> 2	CTac293, CTac388, **CCat103**, CAtc186.1
		1.5 – 2	---
		1 – 1.5	CCac223.1, CAtg205
		0.5 – 1	CCag94, CCac145, CCac223, CCtg272, CAtg170, CAtg312
	Chim-Nd/Kach	> 2	CAac156, CTac293
		1.5 – 2	CTat215
		1 – 1.5	---
		0.5 – 1	*CAac317*, CCag116, CCtg271, CCtg286
	Nd/Kach-Kate	> 2	CAac156, CTat215, CCac223.1
		1.5 – 2	CAat128, CCtg286, CAtc352
		1 – 1.5	CAaa226, *CTag194*, CTac239, CTac388, CAtg170, CAtc186.1
		0.5 – 1	CAat438, CTaa345.1, CCac342, CCat96, **CCat103**, CCat356, CAtg155, CAtg401, CTtc361
Blue	Chai-Naka	> 2	---
		1.5 – 2	---
		1 – 1.5	CCaa218
		0.5 – 1	CAaa189, CAat195, CTag193, CCag122, CCag205, CAtc402
	Chai-Fund	> 2	*CTag194*
		1.5 – 2	CAac333, **CCat322**
		1 – 1.5	*CAac317*
		0.5 – 1	CAac238, CAac356, CTac414, CCat166, CAta216, CTta247
	Naka-Fund	> 2	CAac333
		1.5 – 2	---
		1 – 1.5	**CCat322**, CAtc402
		0.5 – 1	**CAag295**, CAag296, CAac238, *CTag194*, CTac414, CCat166, CTta247, CTta344
Yellow-Blotch	TanL-KasL	> 2	---
		1.5 – 2	---
		1 – 1.5	CAta442
		0.5 – 1	CAat158
	TanL-Mbit	> 2	---
		1.5 – 2	---
		1 – 1.5	CAac244
		0.5 – 1	CAat158, CAat215
	Mbit-KasL	> 2	---
		1.5 – 2	---
		1 – 1.5	CAag129.1, CAac244, CAat215
		0.5 – 1	---

Linkage disequilibrium

Generally, no signs for linkage disequilibrium between pairs of loci were found in any set of outlier loci with *R*–values being overall low. Only in two single cases in the first set of six outlier loci between Blue and Yellow-Blotch populations we found deviations from the null hypothesis of linkage equilibrium with rather high and significant values of *R*. This concerned loci CAag104/CAag295 (R = 0.398; G = 5.0145; df = 1; P = 0.025) in the Funda population and loci CAag295/CAac334 (R = 0.504; G = 4.6825; df = 1; P = 0.031) in the Nakaku population (Tab. 8). However, after correction for multiple testing (Benjamini & Hochberg 1995) the significance in both cases vanished. Apart from that we wouldn't have considered these deviations from linkage equilibrium indicative of physical linkage as in both cases it only occurred in one population. No signs for linkage disequilibrium were found in the other two outlier sets (data not shown). Moreover comparisons focusing on the level of linkage disequilibrium (represented through the mean *R*–values) between any of the three sets of outliers and that of putatively neutral markers yielded no significant differences within each population. *P*-values ranged from 0.149 to 0.9 in the Blue vs. Yellow-Blotch outlier set, from 0.323 to 0.953 in the Red vs. Blue outlier set, and from 0.339 to 0.843 in the Red vs. Yellow-Blotch outlier set (Tab. 9).

Table 8 *R*-values computed with the program DIS for every locus-pair-wise calculation within each population for the first set of six outlier loci detected between Blue and Yellow-Blotch color morphs. None of the *R*-values was significant after correction for multiple testing

	Population					
Locus-pair	**Chai**	**Naka**	**Fund**	**TanL**	**KasL**	**Mbit**
CAag104/CAag295	-0.116	0.077	0.398	0.108	0.126	-0.193
CAag104/CAac334	0.197	0.152	0.258	-0.157	0.126	0.091
CAag104/CCat103	0.134	0.171	-0.165	0.193	0.126	0.091
CAag104/CCat322	0.121	0.126	-0.077	0.108	0.126	0.091
CAag104/CAtg200	0.108	0.117	0.188	0.193	0.044	0.091
CAag295/CAac334	0.056	0.504	-0.06	-0.017	-0.017	-0.017
CAag295/CCat103	-0.311	-0.773	-0.174	-0.03	-0.017	-0.017
CAag295/CCat322	0.162	-0.57	-0.194	-0.017	-0.017	-0.017
CAag295/CAtg200	0.31	0.308	0.236	-0.03	-0.04	-0.017
CAac334/CCat103	-0.025	-0.449	0.251	-0.03	-0.017	-0.017
CAac334/CCat322	0.383	-0.084	0.107	-0.017	-0.017	-0.017
CAac334/CAtg200	-0.997	0.236	-0.835	-0.03	-0.04	-0.017
CCat103/CCat322	-0.761	-0.51	-0.147	-0.03	-0.017	-0.017
CCat103/CAtg200	0.18	-0.103	-0.303	-0.054	-0.04	-0.017
CCat322/CAtg200	-0.614	0.17	0.25	-0.03	0.423	-0.017

Table 9 Mean *R*-values across pair-wise outlier comparisons and across pair-wise neutral comparisons for each population for the three sets of outlier loci. *P*-values yielded from comparing the level of linkage disequilibrium (represented through the mean *R*-values) between outlier loci and putatively neutral loci for each population for the three sets of outlier loci. Differences were non-significant in any comparison

Blue vs. Yellow-Blotch							
Population	**Chai**	**Naka**	**Fund**	**TanL**	**KasL**	**Mbit**	
Mean *R* outlier loci	-0.078	-0.042	-0.018	0.011	0.050	0.000	
Mean *R* neutral loci	0.007	-0 013	-0.031	-0.021	-0.008	-0.016	
P-value	0.9	0.786	0.419	0.349	0.149	0.365	
Red vs. Blue							
Population	**Moli**	**Kate**	**Chim**	**Ndol/Kach**	**Chai**	**Naka**	**Fund**
Mean *R* outlier loci	-0.118	-0.109	-0.107	-0.116	0.018	0.023	0.015
Mean *R* neutral loci	0.006	-0.003	-0.003	-0.010	0.007	-0 013	-0.031
P - value	0.953	0.923	0.898	0.882	0.466	0.446	0.323
Red vs. Yellow-Blotch							
Population	**Moli**	**Kate**	**Chim**	**Ndol/Kach**	**TanL**	**KasL**	**Mbit**
Mean *R* outlier loci	-0.087	-0.081	-0.079	-0.078	0.022	0.024	0.023
Mean *R* neutral loci	0.006	-0.003	-0.003	-0.010	-0.021	-0.008	-0.016
P - value	0.843	0.814	0.771	0.722	0.339	0.38	0.379

Color analysis

As no differences in coloration between males and females of the same color morph could be detected, neither through photography nor pigment analysis, we only speak of color morphs in the following results, neglecting the gender of the investigated fish.

Photography of whole individuals and body parts

Whole individual photos showed obvious differences in coloration between morphs and also between body regions within an individual. Yellow-Blotch individuals displayed a dark basic body color with a yellowish background overlain by melanin in most body regions except the flanks where a conspicuous yellow blotch was present. Fins were rather dark (Fig. 4a). The second morph was the 'Blue' morph, which actually doesn't display a blue body but due to the presence of some bluish areas especially in the dorsal fin this appellation has established itself. In fact Blue morph fish displayed a similar although a bit darker basic body color than the Yellow-Blotch morph. However, here the yellowish background was overlain by dark melanin all over the body. There were also several spots of red especially at the bases of the scales and a few bluish ones. The ventral fin was overall dark with blue spots but also reddish elements (Fig. 4b). The Red morph was generally darker than the two other morphs and no

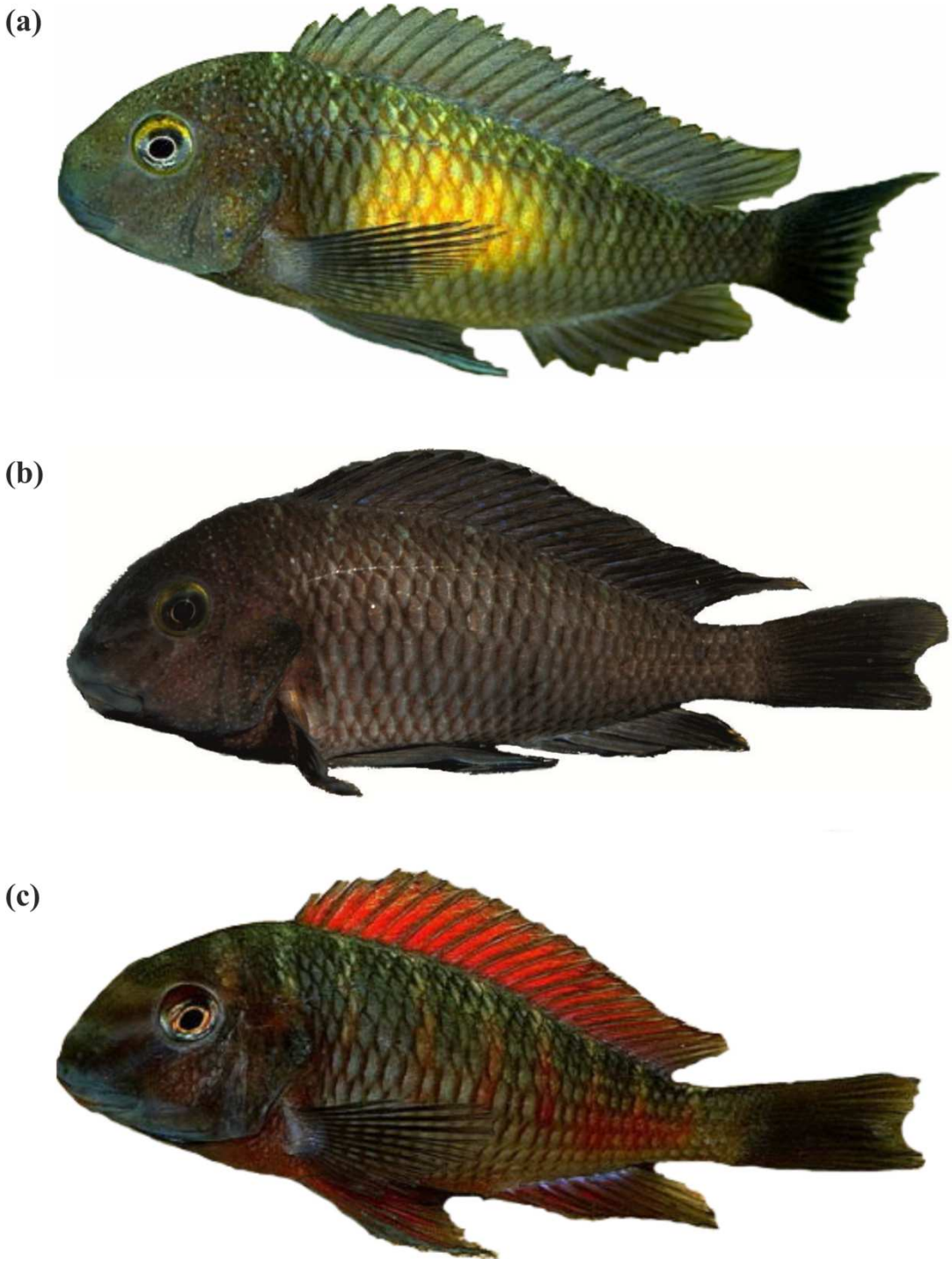


Figure 4 Whole-individual photos from representatives of the three morphs investigated. (a) a Mbita individual representing the Yellow-Blotch morph, (b) a fish from Chaitika representing the Blue morph and (c) a Moliro individual representing the Red morph.

bright yellow or blue color was present. Instead red coloration was predominant in variable amounts depending on the body region. The ventral fin was strikingly red (Fig. 4c).

Pictures from body regions of all three morphs showed that the distribution of chromatophores over the body was variable. Yellow to red xanthophores/erythrophores could be seen in almost all body regions except the ventral parts sometimes. They were overlain by varying numbers of melanophores. In all three color morphs we found variable amounts of dark melanophores all over the body, with highest concentrations at the bases of scales and in regions that were overall darker. In Yellow-Blotch fish a high frequency of melanophores was present together with xanthophores/erythrophores in darker regions of the body like in the area right below the dorsal fin behind the head (Fig. 5a). In contrast, only few melanophores were present within the yellow blotch on the flank bringing out the yellow xanthophores/erythrophores beneath (Fig. 5b). The Blue morph also displayed a variable distribution of melanophores all over the body with again a high concentration for example right below the dorsal fin behind the head (Fig. 5c). We found several spots with fewer melanophores but high frequencies of reddish xanthophores/erythrophores at the basis of scales at the flanks (Fig. 5d). Red morph fish had an overall high amount of melanophores and also red xanthophores/erythrophores. The frequency of both types of chromatophores was especially high at the bases of the scales thinning out towards the middle part but again concentrating at the edge of the scales (Fig. 5e). The amount of xanthophores/erythrophores reached an extreme extent for example in the region with the broad red bar right before the caudal fin (Fig. 5f) or even extremer in the dorsal fin.

Pigment extraction

No pteridines were found in any parts of the body in any of the three investigated color morphs. In contrast to that carotenoids were found in all three morphs in almost all body regions with variable amounts. The absorption spectra varied in their pattern: Either they showed three clearly distinct peaks with the first low intensity peak at ca. 415 nm, followed by two higher intensity peaks at ca. 440 and 470 nm (Fig. 6a), or they displayed two distinct peaks at 440 and 470 nm plus a shoulder at ca. 415 nm (Fig. 6b). Also some spectra showed two not so clearly distinct peaks at ca. 440 and 470 nm (Fig. 6c) or even only one broader peak, approximately between 440 and 490 nm (Fig. 6d). This can be attributed to the presence of different kinds of carotenoids within a certain sample. For example, spectra with only a

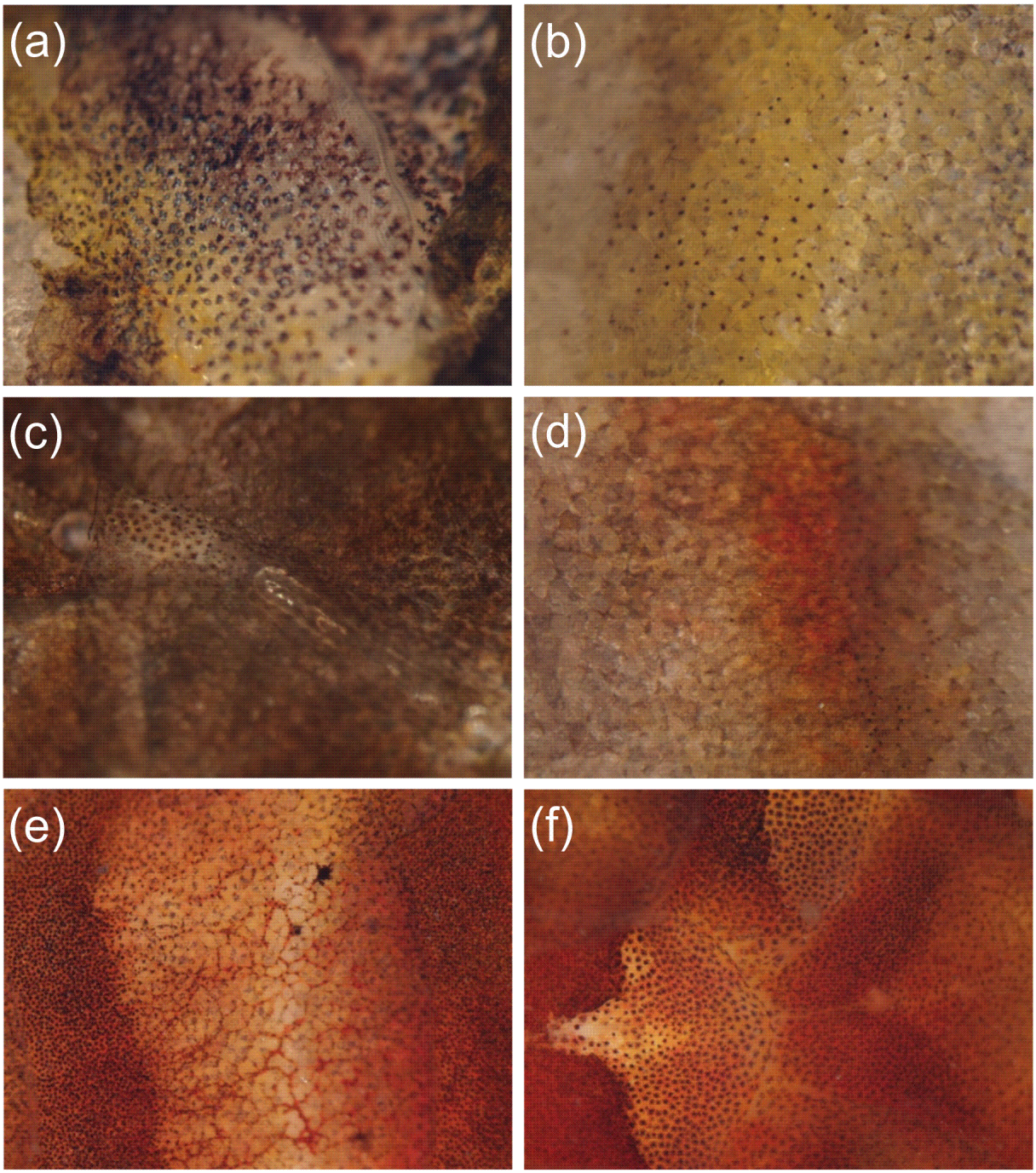


Figure 5 Detailed pictures of body regions from representatives of the three color morphs taken through a binocular. Dark spots represent melanophores and yellow to red spots or regions represent xanthophores/erythrophores. (a) and (b) show details from an Mbita individual representing the Yellow-Blotch morph. Picture (a) was taken in the darker dorsal body region (region 4) and picture (b) in the lateral region where the yellow blotch is located (region 7). (c) and (d) show details from a Chaitika fish representing the Blue morph. Picture (c) was taken in the dark dorsal region directly behind the head (region 1) and picture (d) at the flank (region 7). (e) and (f) show photographs from a Moliro individual representing the Red morph. Picture (e) was taken in the dorsal region (region 6) and picture (f) in the backmost region right before the caudal fin (region 11). Pictures (a – d) were taken with a 650-fold magnification, picture (e) with a 400-fold and picture (f) with a 500-fold magnification.

single peak are typical for keto-carotenoids (Veermann 1970; Foppen 1971; Rempeters 1981; cf. Fig. 6d), whereas spectra with three clearly distinct peaks are typical for carotenoids with an α- and a β- ionone ring (Hager & Meyer-Bertenrath 1966; Rempeters 1981; cf. Fig. 6a). Spectra with two peaks and a shoulder before them are typical for carotenoids with two β-ionone rings (Hager & Meyer-Bertenrath 1966; Rempeters 1981; cf. Fig. 6b) However, as it was not our aim to characterize the types of carotenoids we only focused on the overall patterns of the spectra and compared them between morphs and between the different body regions of one fish. Interestingly we detected that Yellow-Blotch and Blue morph fish showed pattern a or b in most of the 18 regions (Fig. 7; 97% of all spectra in the Yellow-Blotch morph and 74% in the Blue morph; averaged over two individuals). Spectra c and d were found in Blue morph samples only in fins (Fig. 7; 26 % of all spectra averaged over two individuals) which is concordant with the presence of orange to red spots in fins. In contrast to that Red morph fish mostly displayed pattern c or d in body and fin samples, which is typical for red keto-carotenoids (Fig. 7; 84% of all spectra, averaged over four individuals). Congruently, the color of the extract of all Yellow-Blotch samples was yellow, which was also the case for most of the Blue morph samples although some fin-sample extracts tended to be rather orange. The Red fish samples in contrast showed almost always an orange to red or deeply red coloration and only very few extracts were yellowish. Thus it seems that carotenoids producing a yellow coloration tend to show an absorption spectrum that reflects that of Fig. 6 a or b whereas carotenoids that produce red coloration, like keto-carotenoids, tend to display the pattern shown in Fig. 6 c or d.

The approximate amount of carotenoids, represented through the absorbance at the peak of absorption, differed highly between body regions of one fish, between morphs and also to some degree between individuals of the same morph. However, we are aware, that we cannot directly compare the amount or concentration of carotenoids based on absorbance values, between morphs with different carotenoid types present. That is because extinction coefficients differ between carotenoids. This counts especially for Red morph fish as obviously different carotenoids are present in their skin compared to Blue and Yellow-Blotch fish. For all comparisons we used the absorbance value at the 470 nm peak as this peak always had a clear shape. Maximum absorbance values per mg tissue within a morph ranged from 0.442 in the Red morph, to 0.092 in the Yellow-Blotch morph to 0.065 in the Blue morph. The highest absorbance value in Red fish was found in a dorsal fin sample. In contrast to that the highest absorbance values in Yellow-Blotch and Blue fish were found in a body sample. Absorbance values were averaged (i) over the four dorsal fin samples and the two

anal fin samples for fin-sample values and (ii) over 11 body regions for the body sample values. Additionally the values from the individuals representing the same morph were averaged resulting in only one value for the fins and one value for the body per morph. Mean absorbance values per mg tissue for the fin samples varied extremely from 0.003 in the Yellow-Blotch morph to 0.098 in the red morph. Concerning the body, mean values ranged from 0.013 in the Blue morph to 0.028 in the red morph, so here differences between morphs were not as pronounced as in the fin samples (Tab. 10).

Table 10 Absorbance measured from extracts diluted in 40 µl hexane and calculated per mg tissue. Values are averaged over 11 body samples and six fin samples and over the representatives of each morph in the three investigated color morphs

Region	Yellow-Blotch	Blue	Red
Body	0.019	0.013	0.028
Fins	0.003	0.020	0.098

However, also a large amount of variation in absorbance within the body of one fish and partly also between representatives of the same morph was present. Overall, both representatives of the Yellow-Blotch morph for example, showed highest absorbance values in the anterior part of the body, especially in the dorsal to lateral region directly behind the head, an area which actually looks very dark. But also absorbance directly below the dorsal fin and in the central part of the body, where the yellow blotch is located, was quite high, at least in the first Mbita sample. Absorbance was medium at ventral body parts and very low in the backmost body region (Tab. 11). In both representatives of the Blue morph absorbance was highest in the dorsal body region directly below the dorsal fin and medium to low in the rest of the body (Tab. 11). Also in the Red morph absorbance values were overall high in the dorsal region below the dorsal fin but also in the middle and backmost body parts. However, this pattern was rather variable between the four different individuals representing the Red morph (Tab. 11).

Table 11 Carotenoid absorbance values measured in the 11 body parts from all representatives of the three color morphs. The four highest values in each individual are written in bold letters. Mean values are averaged over the representatives of a color morph for each body part and are written in italic letters

	Yellow-Blotch			**Blue**			**Red**				
Body part	**Mbita 1**	**Mbita 2**	*Mean*	**Chai=tika**	**Naka=ku**	*Mean*	**Chi=pimbi**	**Moliro 1**	**Moliro 2**	**Moliro 3**	*Mean*
1	**0.092**	**0.013**	*0.053*	**0.046**	**0.016**	*0.031*	0.060	0.009	**0.011**	0.029	*0.027*
2	**0.047**	**0.023**	*0.035*	**0.019**	0.003	*0.011*	0.043	0.002	0.003	**0.045**	*0.023*
3	0.027	N/A	*0.027*	0.014	**0.005**	*0.010*	0.060	0	0.009	**0.039**	*0.027*
4	0.039	**0.012**	*0.026*	**0.018**	**0.016**	*0.017*	**0.113**	0.007	0.007	0.008	*0.034*
5	0.014	0.005	*0.010*	0.008	0.002	*0.005*	0.047	0.008	0.002	0.022	*0.020*
6	0.011	0.009	*0.010*	**0.065**	**0.013**	*0.039*	**0.099**	**0.071**	**0.009**	**0.041**	*0.055*
7	**0.046**	0.004	*0.025*	0.014	0.003	*0.009*	0.018	0.004	**0.014**	0.010	*0.012*
8	0.009	N/A	*0.009*	0.008	0.001	*0.005*	0.002	0.002	N/A	0.015	*0.006*
9	0.004	0.011	*0.008*	0.010	0.003	*0.007*	**0.078**	**0.026**	0.007	0.029	*0.035*
10	0.008	0.004	*0.006*	0.007	0.003	*0.005*	**0.069**	**0.020**	0.002	0.033	*0.031*
11	0.002	0	*0.001*	0.013	0.003	*0.008*	0.029	**0.048**	**0.018**	**0.051**	*0.037*

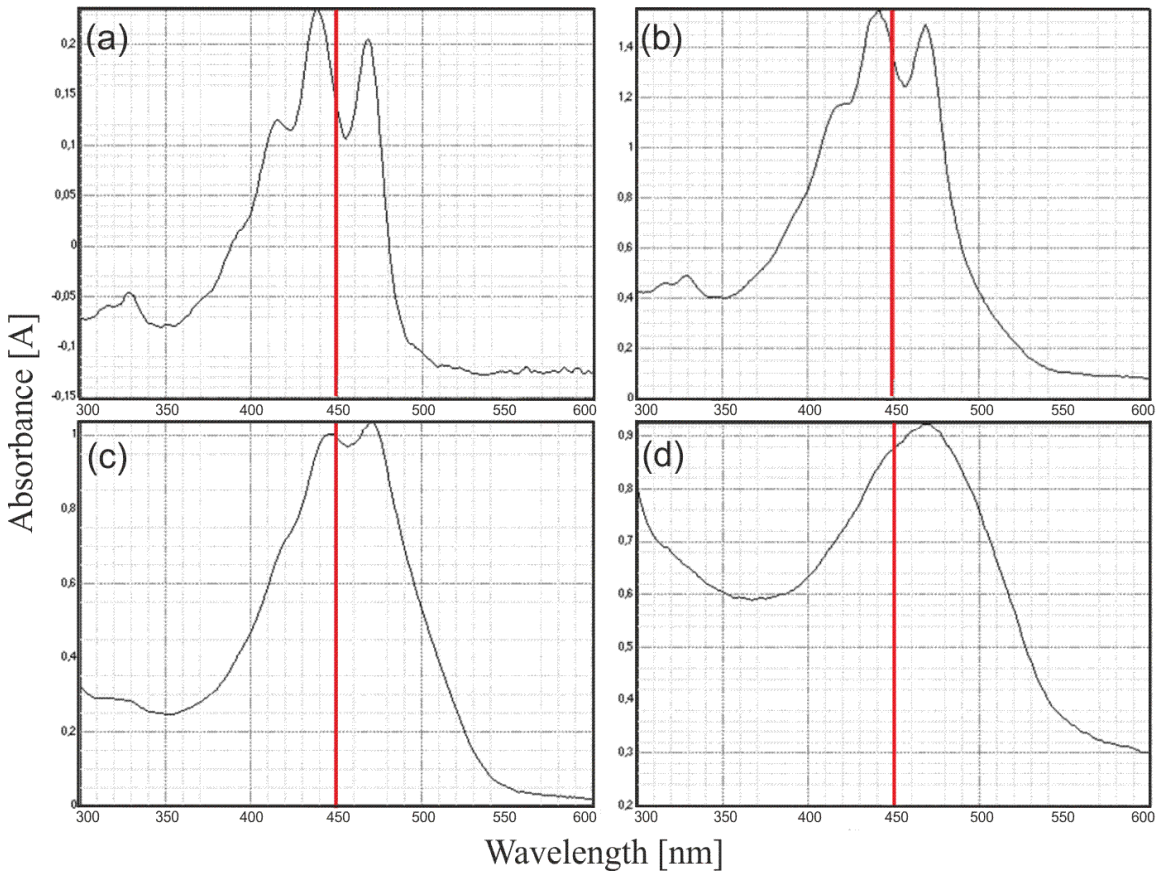


Figure 6 The four different carotenoid absorption spectra that were found in our analyzed individuals. Wavelength [nm] is plotted against Absorbance [A]. The red line indicates a wavelength of 450 nm. **(a)** A spectrum with three peaks **(b)** Absorption spectrum with a shoulder followed by two peaks **(c)** A spectrum with two not so well separated peaks **(d)** Absorption spectrum with only one broad peak. Note that y-axes are not equally scaled.

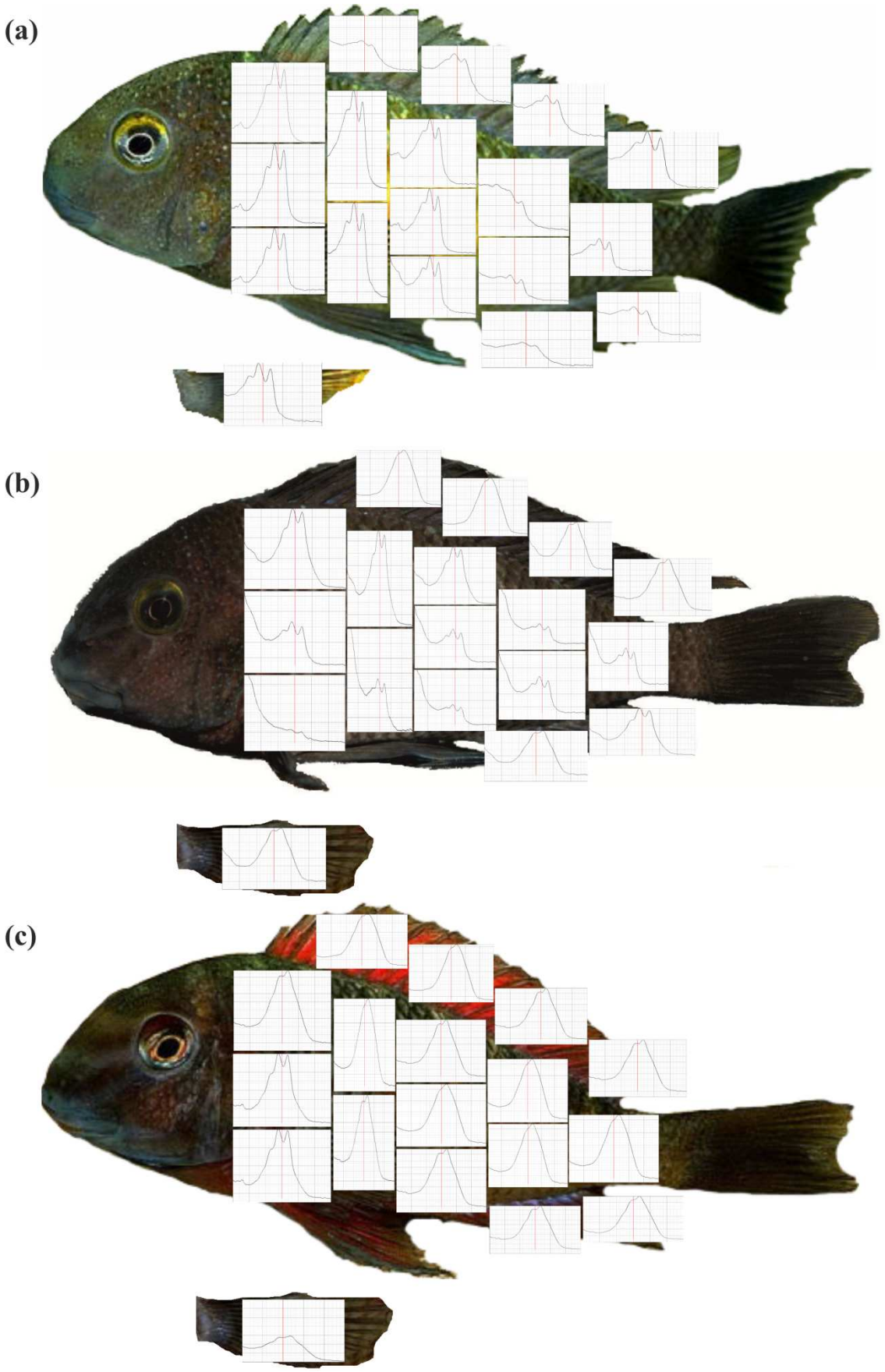


Figure 7 Absorbance spectra yielded from the 18 samples used for pigment extraction. Results are shown for one representative of each color morph. (a) The Mbita 1 sample. (b) The Nakaku sample. (c) The Moliro 3 sample.

DISCUSSION

Outlier loci

We found a total of 11 outlier loci between color morphs. Thus we confirm our hypothesis that body coloration is influenced by selection in contrast to the theory that different color patterns have evolved by genetic drift alone. The 11 outliers could be split up into two sets: (i) Six outlier loci that were detected mainly between the Blue and the Yellow-Blotch morph and (ii) five outliers that were found between the Red and the Blue morph, four of those being also detected between the Red and the Yellow-Blotch morph (thus we refer to three sets in some of the analyses and also in the discussion). The presence of only a small number of loci under selection between color morphs is congruent with the rapid evolution of different color patterns in *Tropheus moorii*: the few genes associated with our outlier loci could act as switches between color patterns and thus facilitate the rapid evolution of different color morphs. Moreover the detection of a low number of genes associated with body color in our study morphs is consistent with other studies estimating similarly low numbers of genes controlling for coloration differences in cichlids (Barson et al. 2007; Magalhaes & Seehausen 2010; Gunter et al. 2011; see introduction). No signs of linkage disequilibrium were found between pairs of loci within any set of outliers, indicated through the overall low and non-significant values of *R* and also through the lack of significant differences in the level of linkage disequilibrium between the outlier sets and the putatively neutral set. Thus we conclude that the six outliers detected in Blue vs. Yellow-Blotch comparisons as well as the five and four outliers detected in Red vs. Blue and Red vs. Yellow-Blotch comparisons, respectively, were independently distributed across the genome. An extreme case could have been that all loci represent selection acting on only one genomic region because they are all within one linkage group. This could be e.g., due to an inversion (Noor et al. 2001; Feder et al. 2003). As nothing of that kind could be detected in our analyses we draw the conclusion that our final set of 11 outliers is indeed indicating selection acting on 11 different, independent regions in the genome.

Limitations of the outlier detection method

In their study comparing three different methods for outlier detection Pérez-Figueroa et al. (2010) showed that the power to detect loci under directional selection decreases with higher differentiation resulting from neutral processes like drift, because in that case selected loci

would not exceed the background. They stated that the most favorable situation for outlier detection is that of a low F_{ST} and showed that the efficiency to detect outliers under directional selection is already much lower with a neutral F_{ST} of 0.1 compared to a neutral F_{ST} of 0.025. However, BAYESCAN still worked well in the F_{ST} = 0.1 scenario showing a positive relationship between the percentage of outliers detected and the true number of selected loci. Moreover within all outliers detected the percentage of selected loci was high in the BAYESCAN analysis, compared to the other two methods. Only the total number of selected loci detected was quite low compared to the scenario with the lower F_{ST}. For an F_{ST} of 0.3 the investigated genome scan methods, including BAYESCAN, did not work at all.

In our study we have four different levels of differentiation: (i) High differentiation between the Red and the Yellow-Blotch morph (mean F_{ST} = 0.309), (ii) medium differentiation between the Blue and the Red morph (mean F_{ST} = 0.238), (iii) medium to low differentiation between the Blue and the Yellow-Blotch morph (mean F_{ST} = 0.101), and (iv) very low differentiation within morphs (mean F_{ST} = 0.041). We would thus expect to find the highest number of outliers in comparisons with low differentiation, which was in fact the case. The highest number of outliers was found within morphs, although these outliers were not reassured through replicates. Disregarding within-morph outliers we found the highest number of reliable outliers (n = 6) in Blue vs. Yellow-Blotch comparisons, fewer in Red vs. Blue comparisons (n = 5) and again fewer in Red vs. Yellow-Blotch comparisons (n = 4). Thus in our study the number of outliers detected decreased marginally with an increasing level of differentiation, as expected. But in contrast to the study of Pérez-Figueroa et al. (2010), in our case outlier detection did not fail even in a high differentiation scenario. However, we are aware that the true number of outliers between morphs is probably higher, as with the lowest F_{ST} being 0.101 we are not in the most favorable situation for detecting loci under directional selection. Moreover some potentially true outliers may have been lost due to our stringent criteria for an outlier being considered as a true outlier. But ultimately all of this gives us the assurance that the outliers that we considered as true outliers in our analyses are in fact true outliers and not false positives which is crucial for the whole study.

Within morph vs. between morph outliers

With the exception of three loci (CAag295, CCat103, CCat322) none of the within morph outliers occurred as such in the between-morph comparisons suggesting that the latter are not location/population specific but indeed likely color specific. The fact that the three loci in question appeared also as outliers within a color morph is congruent with the deviation from

the normal pattern of allele frequencies of certain populations within a morph at those loci (Tab. 5). CAag295 was detected as an outlier between Nakaku and Funda which is not surprising given the high allele frequency difference between the two populations. Similarly CAat103 was detected between Chimba and Katete and also between Ndole/Kachese and Katete, associated with the extremely low allele frequency at that locus in the Katete population. Finally, CCat322 was detected between Funda and Chaitika and also Funda and Nakaku. Again this is coherent with the outstanding low allele frequency of the Funda population at this locus (Tab. 5). The fact that the highest number of within-morph outlier loci was found in comparisons between Red morph populations is not surprising as these are geographically most distant from each other. Thus the probability that habitats differ in some way between the investigated Red populations is rather high and that is reflected through the numerous population/location specific outliers.

Two sets of outlier loci

The presence of two mainly non-overlapping sets of outlier loci suggests that the changes in color pattern between the Yellow-Blotch and the Blue morph are caused by different genetic mechanisms than the changes between the Red morph and the Blue morph, and the Red morph and the Yellow-Blotch morph, respectively. Especially the second set of loci was strictly limited to comparisons between the Red and the Blue morph and between the Red and the Yellow-Blotch morph, respectively, thus indicating that these loci are strongly associated with the red coloration. Our pigment analyses revealed that all three morphs possess carotenoids in their skin. Yellow-Blotch and Blue morph fish carotenoids show similar absorption spectra suggesting that they feature the same or similar kinds of carotenoids. Red morph fish carotenoids instead show different absorption spectra indicating the presence of different carotenoid types than the other two morphs (see further discussion below). Thus, assuming that Blue and Yellow-Blotch fish have predominantly similar carotenoids, the six outliers most reliably detected between Blue and Yellow-Blotch maybe point to genomic regions that somehow control the distribution of Xanthophores/Erythrophores (e.g., in the Yellow-Blotch morph high concentrations of carotenoids are present in the flanks where the yellow blotch is located), or the distribution of melanophores (these are distributed more area-wide with high concentrations all over the body in Blue morph fish). Kelsh (2004b) reviewed three models that describe how the distribution of chromatophores and thus pigment patterning might be influenced: The first model proposes that pigment cell distribution might

be determined by long-range signaling influences, e.g., reaction-diffusion mechanisms. In this model the generation of diffusible activator and inhibitor molecules from a localized source could lead to the formation of a certain pattern of activator concentration. The final pigmentation pattern could then be determined by the size of the developmental field and the diffusion coefficients of the activator and inhibitor molecules, as pigment cells are suggested to react to the local levels of activation. The second model suggests that pigment patterning is determined by interactions with the environment, that is, local tissue. The skin, e.g., could be lined with sites that are adhesive or repulsive to certain chromatophore types beforehand resulting in a certain pattern of pigment cell distribution. The third model focuses on the interactions between different chromatophore types which have been shown to play an important role in zebrafish pigment patterning: in this fish model adult stripe formation is suggested to be critically-dependent upon melanophore-xanthophore interactions. Thus there exist several potential explanations for the observed, obviously carotenoid-type-independent differences between the Yellow-Blotch and the Blue morph fish in our study.

The second set of five outliers is apparently associated with the Red morph. The facts that outliers between Red and Blue and outliers between Red and Yellow-Blotch greatly overlap, and that the Blue morph and the Yellow-Blotch morph are not differentiated at these loci (see allele frequencies in Tab. 6) suggest similar genetic mechanisms in color change from both morphs to red. This might be attributed to the presence of different carotenoids in the Red morph than in the two other morphs, as mentioned above and switches from both morphs to red are obviously controlled by the same genomic regions. One consideration could be that either Blue/Yellow-Blotch fish or Red fish metabolize dietary carotenoids to other carotenoids before transporting them into the skin as shown in other fishes (e.g. Chatzifotis et al. 2005; reviewed for marine animals in Maoka 2011). Also it could be the case that all three morphs do that but with a different product in the Red morph than in the other two morphs.

Color analysis

Lack of pteridines

Our first finding was that no pteridines were present in any investigated individual. This is not surprising as pteridines generally play a smaller role in coloration compared to carotenoids (Anderson 2000). For example, the red coloration of the male three-spined stickleback *Gasterosteus aculeatus* was shown to be produced by at least two different types of carotenoids but no pteridines were found in the skin (Wedekind et al. 1998). Furthermore, a study of Grether et al. (2001), showed, that in the guppy *Poecilia reticulata* pteridines exist together with carotenoids in males but females again lack pteridines and possess only carotenoids. Such gender specificity could not be detected in our study as both sexes lacked pteridines. As pteridines can be synthesized de novo and carotenoids have to be gained through food, Grether et al. (2001) hypothesized, that in low-carotenoid-availability environments the presence of pteridines should increase. However, the study results did not confirm this hypothesis as males in low-carotenoid-availability environments possessed, together with less carotenoids, also less pteridines. This was possibly due to the fact that these guppy males try to maintain a particular hue in their spot which is only given by a certain ratio of the two pigments (Grether et al. 2001; Grether et al. 2005). Also, the authors brought up another issue for discussion: they stated that the metabolic costs for pteridine production could be sufficiently high that it also decreases with decreasing algae availability (although they had no relevant data on the costs for pteridine production). Even though this hypothesis could not be confirmed (Grether at al. 2005) it is still interesting concerning our study: Assuming first, that the costs for pteridine production are in fact high and secondly, that carotenoid rich food is not scarce throughout the distribution range of *Tropheus*, (which is indeed the case as algae grow everywhere in the littoral zone) then our *Tropheus moorii* individuals should rather use carotenoids from food than synthesize pteridines de novo. Moreover, carotenoids are obviously sufficient to yield the particular hue in our study fish.

Variation in carotenoid types, distribution and amount

One interesting finding of our color analysis is that variability in red pigmentation among the three color morphs is purely carotenoid-based as no pteridines were found in any part of the body in any of the investigated fishes. Thus differences in red coloration obviously rely on the

presence of different types and variable amounts of carotenoids and also on differences in their distribution over the body. We found variation in the overall carotenoid absorption spectra between the morphs: Yellow-Blotch and Blue morph fish displayed overall similar absorption spectra whereas absorption spectra of the Red morph were generally different. That is in agreement with the closer genetic relationship between the Yellow-Blotch and the Blue morph than between each of them with the Red morph. This was shown in this study through the F_{ST} values but also in phylogenies including these morphs, based on AFLP and also mitochondrial DNA data (Egger et al. 2007; Mattersdorfer et al. unpublished; see Chapters 3 & 4). Thus, obviously the Yellow-Blotch and the Blue morph are not only genetically more closely related but are also more similar concerning their carotenoid types. This is congruent with the observation that Yellow-Blotch fish look overall similar to Blue morph fish as they display the same basic color. The obvious exception is that the latter do not display the yellow blotch which is due to higher amounts of melanin in this region and probably also to a bit lower amounts of carotenoids, as the yellow coloration is never as bright as that of the Yellow-Blotch fish even when melanophores are contracted.

Carotenoid metabolism

Talking about carotenoids, it is important to know how these pigments get into the skin of a fish. Carotenoids are synthesized in nature by plants, some bacteria, algae and fungi. Fish, as most animals are not able to synthesize carotenoids *de novo,* thus they have to be gained through food and are then directly accumulated or partly modified through metabolic reactions like e.g., oxidation or reduction (reviewed for marine animals in Maoka 2011). It is important to note that the metabolism of carotenoids is not following the same pathway in different fish species, thus the modification of carotenoids is not universal in fish tissues (Chatzifotis et al. 2005). For example, the yellow xanthophyl zeaxanthin can be converted to red astaxanthin in goldfish, whereas rainbow trout *Oncorhynchus mykiss* transforms astaxanthin to zeaxanthin, which is just the other way round. Other fish species do not convert xanthophylls at all (reviewed in Chatzifotis et al. 2005). Nothing is known about the pathway of carotenoids from food to skin pigments in our study species *Tropheus moorii*. Yet, a general pathway of carotenoid transportation in fish is summarized in Baron et al. (2008): After absorbance of fat-soluble dietary carotenoids through the intestinal mucosa they are transported together with serum lipoproteins within the bloodstream, metabolized to other carotenoid types and finally deposited in skin chromatophores. Studies on rainbow trout

Oncorhynchus mykiss (Hardy et al. 1990) and on atlantic salmon *Salmo salar* (Torrissen et al. 1992) showed that the main organ where metabolism and excretion of carotenoids happens is the liver. Unfortunately we know nothing about the types of carotenoids present in our study fish but given the fact that we have a morph with deep red skin coloration and two morphs with yellow or at least partly yellowish coloration together with the fact that diet is the same for all morphs particularly in captivity and maybe also in the lake, it is obvious that different metabolic transformations of same dietary carotenoids take place in the different morphs, especially in Red fish vs. Yellow-Blotch and Blue fish.

CONCLUSION

In summary this study shows first, that genome scans provide a valuable approach for the detection of selected loci even when they are associated with a trait that is not adaptive and secondly, that coloration in *Tropheus moorii* obviously did not evolve by neutral forces alone but also through selection. We found signs of selection acting on a few genes or loci linked to genes that likely triggered the rapid evolution of different color patterns in this species. In addition to the genome scan the analyses on body coloration allowed us a first characterization of the nature of the body color differences between morphs, with a special focus on the yellow/red coloration. However, these are only first steps as, for example, identification of carotenoid types or accurate carotenoid concentration measures were not done. Yet the results give us a valuable overview. Also, the genome scan for outlier loci between color morphs is merely a first, but important step towards the characterization and identification of genes being subject to selective forces and contributing to color pattern differences in *T. moorii*. Future work could focus on the sequencing of outlier loci and the mapping of them within the genome, given that the genomic resources for cichlids should be available soon.

REFERENCES

Anderson S (2000) Salmon Color and the Consumer. IIFET 2000 Proceedings

Antao T, Lopes A, Lopes RJ, Beja-Pereira A, Luikart G (2008) LOSITAN: A workbench to detect molecular adaptation based on a F-st-outlier method. BMC Bioinformatics 9

Antao T, Beaumont MA (2011) Mcheza: a workbench to detect selection using dominant markers. Bioinformatics 27:1717-1718

Arnegard ME, Markert JA, Danley PD, Stauffer JR, Ambali AJ, Kocher TD (1999) Population structure and colour variation of the cichlid fish *Labeotropheus fuelleborni* Ahl along a recently formed archipelago of rocky habitat patches in southern Lake Malawi. Proceedings of the Royal Society of London Series B-Biological Sciences 266:119-130

Arrigo N, Tuszynski JW, Ehrich D, Gerdes T, Alvarez N (2009) Evaluating the impact of scoring parameters on the structure of intra-specific genetic variation using RawGeno, an R package for automating AFLP scoring. BMC Bioinformatics 10:Article No.: 33

Bagnara JT, Matsumoto J (2006) Comparative anatomy and physiology of pigment cells in non-mammalian tissues. In The Pigmentary System: Physiology and Pathophysiology. Oxford University Press, New York, NY

Bagnara JT, Fernandez PJ, Fujii R (2007) On the blue coloration of vertebrates. Pigment Cell Research 20:14-26

Baric S, Salzburger W, Sturmbauer C (2003) Phylogeography and evolution of the Tanganyikan cichlid genus *Tropheus* based upon mitochondrial DNA sequences. Journal of Molecular Evolution 56:54-68

Baron M, Davies S, Alexander L, Snellgrove D, Sloman KA (2008) The effect of dietary pigments on the coloration and behaviour of flame-red dwarf gourami, *Colisa lalia*. Animal Behaviour 75:1041-1051

Barson NJ, Knight ME, Turner GF (2007) The genetic architecture of male colour differences between a sympatric Lake Malawi cichlid species pair. Journal of Evolutionary Biology 20:45-53

Beaumont MA, Nichols RA (1996) Evaluating loci for use in the genetic analysis of population structure. Proceedings of the Royal Society of London Series B-Biological Sciences 263:1619-1626

Beaumont MA, Balding DJ (2004) Identifying adaptive genetic divergence among populations from genome scans. Molecular Ecology 13:969-980

Benjamini Y, Hochberg Y (1995) Controlling the false discovery rate - a practical and powerful approach to multiple testing. Journal of the Royal Statistical Society Series B-Methodological 57:289-300

Bonin A, Taberlet P, Miaud C, Pompanon F (2006) Explorative genome scan to detect candidate loci for adaptation along a gradient of altitude in the common frog (*Rana temporaria*). Molecular Biology and Evolution 23:773-783

Braasch I, Salzburger W, Meyer A (2006) Asymmetric evolution in two fish-specifically duplicated receptor tyrosine kinase paralogons involved in teleost coloration. Molecular Biology and Evolution 23:1192-1202

Braasch I, Schartl M, Volff JN (2007) Evolution of pigment synthesis pathways by gene and genome duplication in fish. Bmc Evolutionary Biology 7

Braasch I, Volff JN, Schartl M (2008) The evolution of teleost pigmentation and the fish-specific genome duplication. Journal of Fish Biology 73:1891-1918

Braasch I, Brunet F, Volff JN, Schartl M (2009a) Pigmentation Pathway Evolution after Whole-Genome Duplication in Fish. Genome Biology and Evolution 1:479-493

Braasch I, Liedtke D, Volff JN, Schartl M (2009b) Pigmentary function and evolution of tyrp1 gene duplicates in fish. Pigment Cell & Melanoma Research 22:839-850

Braasch I, Volff JN, Schartl M (2009c) The Endothelin System: Evolution of Vertebrate-Specific Ligand-Receptor Interactions by Three Rounds of Genome Duplication. Molecular Biology and Evolution 26:783-799

Campbell D, Bernatchez L (2004) Generic scan using AFLP markers as a means to assess the role of directional selection in the divergence of sympatric whitefish ecotypes. Molecular Biology and Evolution 21:945-956

Chatzifotis S, Pavlidis M, Jimeno CD, Vardanis G, Sterioti A, Divanach P (2005) The effect of different carotenoid sources on skin coloration of cultured red porgy (*Pagrus pagrus*). Aquaculture Research 36:1517-1525

Chen L, Yang G (2009) A genomic scanning using AFLP to detect candidate loci under selection in the finless porpoise (*Neophocaena phocaenoides*). Genes & Genetic Systems 84:307-313

Dasmahapatra KK, Blum MJ, Aiello A, Hackwell S, Davies N, Bermingham EP, Mallett T (2002) Inferences from a rapidly moving hybrid zone. Evolution 56:741-753

Dutton KA, Pauliny A, Lopes SS, Elworthy S, Carney TJ, Rauch J, Geisler R, Haffter P, Kelsh RN (2001) Zebrafish colourless encodes sox10 and specifies non-ectomesenchymal neural crest fates. Development 128:4113-4125

Egan SP, Nosil P, Funk DJ (2008) Selection and genomic differentiation during ecological speciation: Isolating the contributions of host association via a comparative genome scan of *Neochlamisus bebbianae* leaf beetles. Evolution 62:1162-1181

Egger B, Koblmüller S, Sturmbauer C, Sefc KM (2007) Nuclear and mitochondrial data reveal different evolutionary processes in the Lake Tanganyika cichlid genus *Tropheus*. BMC Evolutionary Biology 7:-

Elworthy S, Lister JA, Carney TJ, Raible DW, Kelsh RN (2003) Transcriptional regulation of mitfa accounts for the sox10 requirement in zebrafish melanophore development. Development 130:2809-2818

Feder JL, Roethele FB, Filchak K, Niedbalski J, Romero-Severson J (2003) Evidence for inversion polymorphism related to sympatric host race formation in the apple maggot fly, *Rhagoletis pomonella*. Genetics 163:939-953

Fischer MC, Foll M, Excoffier L, Heckel G (2011) Enhanced AFLP genome scans detect local adaptation in high-altitude populations of a small rodent (*Microtus arvalis*). Molecular Ecology 20

Foll M, Gaggiotti O (2008) A Genome-Scan Method to Identify Selected Loci Appropriate for Both Dominant and Codominant Markers: A Bayesian Perspective. Genetics 180:977-993

Foll M, Fischer MC, Heckel G, Excoffier L (2010) Estimating population structure from AFLP amplification intensity. Molecular Ecology 19:4638-4647

Foppen FH (1971) Tables for the identification of carotenoid pigments. Chromatographic Reviews 14:133-298

Forrest HS, Mitchell HK (1955) Pteridines from Drosophila 3 - Isolation and identification of 3 more pteridines. Journal of the American Chemical Society 77:4865-4869

Furutani-Seiki M, Wittbrodt J (2004) Medaka and zebrafish, an evolutionary twin study. Mechanisms of Development 121:629-637

Gagnaire PA, Albert V, Jonsson B, Bernatchez L (2009) Natural selection influences AFLP intraspecific genetic variability and introgression patterns in Atlantic eels. Molecular Ecology 18:1678-1691

Goda M, Fujii R (1995) Blue chromatophores in two species of callionymid fish. Zoological Science 12:811-813

Goodwin TW (1984) The biochemistry of the carotenoids. Chapman & Hall, London, England

Grether GF, Hudon J, Endler JA (2001) Carotenoid scarcity, synthetic pteridine pigments and the evolution of sexual coloration in guppies (*Poecilia reticulata*). Proceedings of the Royal Society B-Biological Sciences 268:1245-1253

Grether GF, Kolluru GR, Nersissian K (2004) Individual colour patches as multicomponent signals. Biological Reviews 79:583-610

Grether GF, Cummings ME, Hudon J (2005) Countergradient variation in the sexual coloration of guppies (*Poecilia Reticulata*): Drosopterin synthesis balances carotenoid availability. Evolution 59:175-188

Gunter HM, Clabaut C, Salzburger W, Meyer A (2011) Identification and Characterization of Gene Expression Involved in the Coloration of Cichlid Fish Using Microarray and qRT-PCR Approaches. Journal of Molecular Evolution 72:127-137

Hager A, Meyer-Bertenrath.T (1966) Die Isolierung und quantitative Bestimmung der Carotinoide und Chlorophylle von Blättern, Algen und isolierten Chloroplasten mit Hilfe dünnschichtchromatographischer Methoden. Planta 69:198-217

Hardy RW, Torrissen OJ, Scott TM (1990) Absorption and distribution of C-14-labeled canthaxanthin in rainbow trout (*Oncorhynchus mykiss*). Aquaculture 87:331-340

Herler J, Kerschbaumer M, Mitteroecker P, Postl L, Sturmbauer C (2010) Sexual dimorphism and population divergence in the Lake Tanganyika cichlid fish genus *Tropheus*. Frontiers in Zoology 7:-

Hill WG (1974) Estimation of linkage disequilibrium in randomly mating populations. Heredity 33:229-239

Hirata M, Nakamura K, Kanemaru T, Shibata Y, Kondo S (2003) Pigment cell organization in the hypodermis of zebrafish. Developmental Dynamics 227:497-503

Hirata M, Nakamura K, Kondo S (2005) Pigment cell distributions in different tissues of the zebrafish, with special reference to the striped pigment pattern. Developmental Dynamics 234:293-300

Hoekstra HE, Hirschmann RJ, Bundey RA, Insel PA, Crossland JP (2006) A single amino acid mutation contributes to adaptive beach mouse color pattern. Science 313:101-104

Hurst DT (1980) An introduction to the chemistry and biochemistry of pyrimidines, purines, and pteridines. John Wiley, New York, NY

Jeffreys H (1961) Theory of Probability. 3rd edition. Clarendon Press, Oxford, UK

Johnson TC, Scholz CA, Talbot MR, Kelts K, Ricketts RD, Ngobi G, Beuning K, Ssemmanda I, McGill JW (1996) Late pleistocene desiccation of Lake Victoria and rapid evolution of cichlid fishes. Science 273:1091-1093

Jump AS, Hunt JM, Martinez-Izquierdo JA, Penuelas J (2006) Natural selection and climate change: temperature-linked spatial and temporal trends in gene frequency in Fagus sylvatica. Molecular Ecology 15:3469-3480

Kelsh RN, Brand M, Jiang YJ, Heisenberg CP, Lin S, Haffter P, Odenthal J, Mullins MC, vanEeden FJM, FurutaniSeiki M, Granato M, Hammerschmidt M, Kane DA, Warga RM, Beuchle D, Vogelsang L, NussleinVolhard C (1996) Zebrafish pigmentation mutations and the processes of neural crest development. Development 123:369-389

Kelsh RN, Inoue C, Momoi A, Kondoh H, Furutani-Seiki M, Ozato K, Wakamatsu Y (2004a) The Tomita collection of medaka pigmentation mutants as a resource for understanding neural crest cell development. Mechanisms of Development 121:841-859

Kelsh RN (2004b) Genetics and evolution of pigment patterns in fish. Pigment Cell Research 17:326-336

Kim Y, Nielsen R (2004) Linkage disequilibrium as a signature of selective sweeps. Genetics 167:1513-1524

Kimura M (1983) The neutral theory of molecular evolution. Cambridge University Press, New York, NY

Knight ME, Turner GF (2004) Laboratory mating trials indicate incipient speciation by sexual selection among populations of the cichlid fish *Pseudotropheus zebra* from Lake Malawi. Proceedings of the Royal Society of London Series B-Biological Sciences 271:675-680

Koblmüller S, Schliewen UK, Duftner N, Sefc KM, Katongo C, Sturmbauer C (2008) Age and spread of the haplochromine cichlid fishes in Africa. Molecular Phylogenetics and Evolution 49:153-169

Koblmüller S, Egger B, Sturmbauer C, Sefc KM (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

Koblmüller S, Salzburger W, Obermüller B, Eigner E, Sturmbauer C, Sefc KM (2011) Separated by sand, fused by dropping water: habitat barriers and fluctuating water levels steer the evolution of rock-dwelling cichlid populations in Lake Tanganyika. Molecular Ecology 20:2272-2290

Koga A, Hori H (1997) Albinism due to transposable element insertion in fish. Pigment Cell Research 10:377-381

Konings A (1998) Tanganyika cichlids in their natural habitat. Cichlid Press, El Paso, TX

Luikart G, England PR, Tallmon D, Jordan S, Taberlet P (2003) The power and promise of population genomics: From genotyping to genome typing. Nature Reviews Genetics 4:981-994

Lynch M, Milligan BG (1994) Analysis of population genetic structure with RAPD markers. Molecular Ecology 3:91-99

Maderbacher M, Bauer C, Herler J, Postl L, Makasa L, Sturmbauer C (2008) Assessment of traditional versus geometric morphometrics for discriminating populations of the *Tropheus moorii* species complex (Teleostei : Cichlidae), a Lake Tanganyika model for allopatric speciation. Journal of Zoological Systematics and Evolutionary Research 46:153-161

Magalhaes IS, Seehausen O (2010) Genetics of male nuptial colour divergence between sympatric sister species of a Lake Victoria cichlid fish. Journal of Evolutionary Biology 23:914-924

Maoka T (2011) Carotenoids in Marine Animals. Marine Drugs 9:278-293

McClelland EK, Naish KA (2010) Quantitative trait locus analysis of hatch timing, weight, length and growth rate in coho salmon, *Oncorhynchus kisutch*. Heredity 105:562-573

Mellgren EM, Johnson SL (2005) kitb, a second zebrafish ortholog of mouse Kit. Development Genes and Evolution 215:470-477

Midamegbe A, Vitalis R, Malausa T, Delava E, Cros-Arteil S, Streiff R (2011) Scanning the European corn borer (*Ostrinia* spp.) genome for adaptive divergence between host-affiliated sibling species. Molecular Ecology 20:1414-1430

Mills MG, Patterson LB (2009) Not just black and white: Pigment pattern development and evolution in vertebrates. Seminars in Cell & Developmental Biology 20:72-81

Minder AM, Widmer A (2008) A population genomic analysis of species boundaries: neutral processes, adaptive divergence and introgression between two hybridizing plant species. Molecular Ecology 17:1552-1563

Nagl S, Tichy H, Mayer WE, Takezaki N, Takahata N, Klein J (2000) The origin and age of haplochromine fishes in Lake Victoria, East Africa. Proceedings of the Royal Society of London Series B-Biological Sciences 267:1049-1061

Narum SR, Hess JE (2011) Comparison of F-ST outlier tests for SNP loci under selection. Molecular Ecology Resources 11

Noor MAF, Grams KL, Bertucci LA, Reiland J (2001) Chromosomal inversions and the reproductive isolation of species. Proceedings of the National Academy of Sciences of the United States of America 98:12084-12088

Noor MAF, Feder JL (2006) Speciation genetics: evolving approaches. Nature Reviews Genetics 7:851-861

Nosil P, Egan SP, Funk DJ (2008) Heterogeneous genomic differentiation between walking-stick ecotypes: "Isolation by adaptation" and multiple roles for divergent selection. Evolution 62:316-336

Odenthal J, Rossnagel K, Haffter P, Kelsh RN, Vogelsang E, Brand M, vanEeden FJM, FurutaniSeiki M, Granato M, Hammerschmidt M, Heisenberg CP, Jiang YJ, Kane DA, Mullins MC, NussleinVolhard C (1996) Mutations affecting xanthophore pigmentation in the zebrafish, *Danio rerio*. Development 123:391-398

Parichy DM, Rawls JF, Pratt SJ, Whitfield TT, Johnson SL (1999) Zebrafish sparse corresponds to an orthologue of c-kit and is required for the morphogenesis of a subpopulation of melanocytes, but is not essential for hematopoiesis or primordial germ cell development. Development 126:3425-3436

Parichy DM, Mellgren EM, Rawls JF, Lopes SS, Kelsh RN, Johnson SL (2000) Mutational analysis of endothelin receptor b1 (rose) during neural crest and pigment pattern development in the zebrafish *Danio rerio*. Developmental Biology 227:294-306

Parichy DM (2007) Homology and the evolution of novelty during *Danio* adult pigment pattern development. Journal of Experimental Zoology Part B-Molecular and Developmental Evolution 308B:578-590

Paris M, Boyer S, Bonin A, Collado A, David JP, Despres L (2010) Genome scan in the mosquito *Aedes rusticus*: population structure and detection of positive selection after insecticide treatment. Molecular Ecology 19:325-337

Parisod C, Joost S (2010) Divergent selection in trailing- versus leading-edge populations of *Biscutella laevigata*. Annals of Botany 105:655-660

Pérez-Figueroa A, Garcia-Pereira MJ, Saura M, Rolan-Alvarez E, Caballero A (2010) Comparing three different methods to detect selective loci using dominant markers. Journal of Evolutionary Biology 23:2267-2276

Plenderleith M, van Oosterhout C, Robinson RL, Turner GF (2005) Female preference for conspecific males based on olfactory cues in a Lake Malawi cichlid fish. Biology Letters 1:411-414

Poll M (1986) Classification des Cichlidae du lac Tanganyika: tribus, genre et especes. Academie Royale de Belgique Memoires de la Classe des Sciences 45:1-163

Postl L, Herler J, Bauer C, Maderbacher M, Makasa L, Sturmbauer C (2008) Geometric morphometrics applied to viscerocranial bones in three populations of the Lake

Tanganyika cichlid fish *Tropheus moorii*. Journal of Zoological Systematics and Evolutionary Research 46:240-248

R Development Core Team (2009) R: A language and environment for statistical computing. Vienna, Austria

Rempeters G, Henze M, Anders F (1981) Carotenoids and pteridines in the skin of interspecific hybrids of *Xiphophorus*. Comparative Biochemistry and Physiology 69:91-98

Roberts RB, Ser JR, Kocher TD (2009) Sexual Conflict Resolved by Invasion of a Novel Sex Determiner in Lake Malawi Cichlid Fishes. Science 326:998-1001

Rosenblum EB, Hoekstra HE, Nachman MW (2004) Adaptive reptile color variation and the evolution of the Mc1r gene. Evolution 58:1794-1808

Rueber L, Verheyen E, Sturmbauer C, Meyer A (1998) Lake level fluctuations and speciation in rock-dwelling cichlid fish in Lake Tanganyika, East Africa. In Grant P (ed) Evolution of islands. Oxford University Press, Oxford, UK

Sahana G, Guldbrandtsen B, Lund MS (2011) Genome-wide association study for calving traits in Danish and Swedish Holstein cattle. Journal of Dairy Science 94:479-486

Salzburger W, Baric S, Sturmbauer C (2002) Speciation via introgressive hybridization in East African cichlids? Molecular Ecology 11:619-625

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

Salzburger W, Niederstatter H, Brandstatter A, Berger B, Parson W, Snoeks J, Sturmbauer C (2006) Colour-assortative mating among populations of *Tropheus moorii*, a cichlid fish from Lake Tanganyika, East Africa. Proceedings of the Royal Society B-Biological Sciences 273:257-266

Salzburger W, Braasch I, Meyer A (2007) Adaptive sequence evolution in a color gene involved in the formation of the characteristic egg-dummies of male haplochromine cichlid fishes. BMC Biology 5

Schiedt K (1989) New aspects of carotenoid metabolism in animals. In Carotenoids: Chemistry and biology. Plenum Press, New York, NY

Schupke P (2003) African Cichlids II: Tanganyika I: Tropheus. Aqualog, Rodgau, Germany

Seehausen O, Mayhew PJ, Van Alphen JJM (1999) Evolution of colour patterns in East African cichlid fish. Journal of Evolutionary Biology 12:514-534

Seehausen O, van Alphen JM (1999) Can sympatric speciation by disruptive sexual selection explain rapid evolution of cichlid diversity in Lake Victoria? Ecology Letters 2:262-271

Seehausen O, Terai Y, Magalhaes IS, Carleton KL, Mrosso HDJ, Miyagi R, van der Sluijs I, Schneider MV, Maan ME, Tachida H, Imai H, Okada N (2008) Speciation through sensory drive in cichlid fish. Nature 455:620-U23

Sefc KM, Baric S, Salzburger W, 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

Sefc KM (2008) Variance in reproductive success and the opportunity for selection in a serially monogamous species: simulations of the mating system of *Tropheus* (Teleostei: Cichlidae). Hydrobiologia 615:21-35

Smith PF, Konings A, Kornfield I (2003) Hybrid origin of a cichlid population in Lake Malawi: implications for genetic variation and species diversity. Molecular Ecology 12:2497-2504

Snoeks J, Rueber L, Verheyen E (1994) The Tanganyika Problem: comments on the taxonomy and distribution patterns of its cichlid fauna. Adv. Limn. 44:357-374

Stelkens RB, Schmid C, Selz O, Seehausen O (2009) Phenotypic novelty in experimental hybrids is predicted by the genetic distance between species of cichlid fish. Bmc Evolutionary Biology 9:Article No.: 283

Stemshorn KC, Nolte AW, 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

Stinchcombe JR, Hoekstra HE (2007) Combining population genomics and quantitative genetics: finding the genes underlying ecologically important traits. Heredity 100:158-170

Streelman JT, Albertson RC, Kocher TD (2003) Genome mapping of the orange blotch colour pattern in cichlid fishes. Molecular Ecology 12:2465-2471

Sturmbauer C, Meyer A (1992) Genetic divergence, speciation and morphological stasis in a lineage of African cichlid fishes. Nature 358:578-581

Sturmbauer C, Dallinger R (1995) Diurnal variation of spacing and foraging behaviour in *Tropheus moorii* (Cichlidae) in Lake Tanganyika, Eastern Africa. Netherlands Journal of Zoology 45:386-401

Sturmbauer C, Verheyen E, Rueber L, Meyer A (1997) Phylogeographic patterns in populations of cichlid fishes from rocky habitats in Lake Tanganyika. In Kocher TD, Stepien C (eds) Molecular phylogeny of fishes. Academic Press, New York, NY.

Sturmbauer C (1998) Explosive speciation in cichlid fishes of the African Great Lakes: a dynamic model of adaptive radiation. Journal of Fish Biology 53:18-36

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

Sturmbauer C, Koblmüller S, Sefc KM, Duftner N (2005) Phylogeographic history of the genus *Tropheus*, a lineage of rock-dwelling cichlid fishes endemic to Lake Tanganyika. Hydrobiologia 542:335-366

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

Takeda H, Shimada A (2010) The Art of Medaka Genetics and Genomics: What Makes Them So Unique? Annual Review of Genetics 44:217-241

Taylor JS, Van de Peer Y, Braasch I, Meyer A (2001a) Comparative genomics provides evidence for an ancient genome duplication event in fish. Philosophical Transactions of the Royal Society of London Series B-Biological Sciences 356:1661-1679

Taylor JS, Van de Peer Y, Meyer A (2001b) Revisiting recent challenges to the ancient fish-specific genome duplication hypothesis. Current Biology 11:R1005-R1007

Terai Y, Morikawa N, Kawakami K, Okada N (2002) Accelerated evolution of the surface amino acids in the WD-repeat domain encoded by the hagoromo gene in an explosively speciated lineage of east African cichlid fishes. Molecular Biology and Evolution 19:574-578

Terai Y, Morikawa N, Kawakami K, Okada N (2003) 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

Torrissen OJ, Ingebrigtsen K (1992) Tissue distribution of C-14 astaxanthin in the atlantic salmon (*Salmo salar*). Aquaculture 108:381-386

Veerman A (1970) The pigments of *Tetranychus cinnabarinus* Boisd. (*Acari, Tetranychidae*). Comparative Biochemistry and Physiology 36:749-763

Vekemans X (2002) AFLP-SURV. Version 1.0. Distributed by the author, Laboratoire de Génétique et Ecologie Végétal, Université Libre de Bruxelles, Belgium

Vekemans X, Beauwens T, Lemaire M, Roldan-Ruiz I (2002) Data from amplified fragment length polymorphism (AFLP) markers show indication of size homoplasy and of a relationship between degree of homoplasy and fragment size. Molecular Ecology 11:139-151

Vos P, Hogers R, Bleeker M, Reijans M, Vandelee T, Hornes M, Frijters A, Pot J, Peleman J, Kuiper M, Zabeau M (1995) AFLP: A new technique for DNA fingerprinting. Nucleic Acids Research 23:4407-4414

Wedekind C, Meyer P, Frischknecht M, Niggli UA, Pfander H (1998) Different carotenoids and potential information content of red coloration of male three-spined stickleback. Journal of Chemical Ecology 24:787-801

Weir BS, Cockerham CC (1984) Estimating F-statistics for the analysis of population structure. Evolution 38:1358-1370

Whitlock R, Hipperson H, Mannarelli M, Butlin RK, Burke T (2008) An objective, rapid and reproducible method for scoring AFLP peak-height data that minimizes genotyping error. Molecular Ecology Resources 8:725-735

Wilding CS, Butlin RK, Grahame J (2001) Differential gene exchange between parapatric morphs of *Littorina saxatilis* detected using AFLP markers. Journal of Evolutionary Biology 14:611-619

Winge O (1927) The location of eighteen genes in *Lebistes reticulatus*. Journal of Genetics 18:1-43

Wringe BF, Devlin RH, Ferguson MM, Moghadam HK, Sakhrani D, Danzmann RG (2010) Growth-related quantitative trait loci in domestic and wild rainbow trout (*Oncorhynchus mykiss*). BMC Genetics 11:Article No.: 63

Zhivotovsky LA (1999) Estimating population structure in diploids with multilocus dominant DNA markers. Molecular Ecology 8:907-913

CHAPTER 3

Asymmetric introgression between two *Tropheus moorii* color morphs in southern Lake Tanganyika

ABSTRACT

To get a better understanding of the mechanisms that underlie phenotypic diversification, it is of special interest to identify the forces contributing to phenotypic variation. A stunning example for the rapid evolution of phenotypic diversity are the cichlid species flocks of the east African Rift Valley Lakes. In Lake Tanganyika, *Tropheus moorii* represents a striking example of intraspecific geographic color pattern variation with numerous allo- and parapatric color morphs. *Tropheus* use coloration for communication during courtship, mating, and other social interactions. Sexual selection has been considered as a driving force in the rapid evolution of color variants in *Tropheus*, but so far without conclusive evidence. Here, we investigate hybridization as a further potential source of new variants: Admixture between existing morphs could have given rise to novel patterns when, in the wake of Lake Tanganyika's water level fluctuations, allopatric color morphs came into secondary contact. We study nine *T. moorii* populations around a large sandy bay in southern Lake Tanganyika, the Mbete Bay, which separates two ancient mitochondrial lineages, the Yellow-Blotch morph and the Blue morph. Using a large scale mtDNA data set we test previously found sings of mtDNA introgression between the two lineages and try to confirm the reported asymmetry of admixture as it was only found west of the bay, where fish even display an intermediate color phenotype. Then we test the hypothesis that the previously detected asymmetry in female mate choice behavior between the two morphs could have shaped the observed asymmetric mtDNA introgression pattern. Thus we use a large nuclear data set including amplified fragment length polymorphism (AFLP) and microsatellite (SSR) markers to detect possible signs of hybridization on both sides of the bay which would support the mate-choice-shapes-mtDNA-introgression-pattern-hypothesis. Furthermore we carry out coalescence based estimates of splitting times to infer the chronology of the colonization of the contact zone. We confirmed asymmetric mtDNA hybridization and showed that nuclear admixture was also restricted to populations west of the bay. This rules out asymmetric mate choice as the reason for the observed uni-directional hybridization and provides space for alternative scenarios. Coalescence based analyses suggested that the Yellow-Blotch morph colonized the area west of Mbete Bay during a high lake level scenario and only after that the Blue morph expanding from the north introgressed into the Yellow-Blotch population. Our study highlights that interpreting hybridization scenarios is a challenging task. However, it also shows that hybridization within *T. moorii* may very likely serve as a considerable contribution to the outstanding phenotypic diversity present within this taxon.

INTRODUCTION

In this study we investigated nine *Tropheus moorii* populations from southern Lake Tanganyika (Fig. 1) which belong to two phenotypically distinct and ancient mitochondrial lineages that are separated by a 7 km stretch of sand, the Mbete Bay, and whose split dates back to ~ 370 000 - 650 000 ya (Koblmüller et al. 2011; calculated from net divergence between lineages). Populations with a yellowish basic coloration overlain by dark melanin and with bluish elements (thus referred to as the 'Blue' morph) west of the bay represent the '8-G' lineage (Baric et al. 2003; Sturmbauer et al. 2005; Egger et al. 2007). Populations with a conspicuous yellow blotch on the flanks (thus referred to as the 'Yellow-Blotch' morph) east of the bay are resolved within the '1-A' lineage, which is further divided into four sub-lineages (1-A1 – 1-A4; Baric et al. 2003; Sturmbauer et al. 2005; Egger et al. 2007) with the study populations belonging to sub-clades 1-A2 and 1-A4. In a phylogenetic reconstruction based on AFLP markers the Blue and the Yellow-Blotch morph were resolved in sister clades within one major AFLP clade (AFLP 4 sub-clade 1a and 1b, respectively; Egger et al. 2007).

The aim of the present study was to confirm and detect, respectively, signs of hybridization between the lineages. As already noted *Tropheus* normally do not cross large sandy areas and thus Mbete Bay acts as a barrier to gene flow between the two morphs. However, previous studies showed that the population immediately west of the bay in Katoto not only exhibited an intermediate color pattern (bluish but with a not as bright but still clearly visible yellow blotch; Kohda et al. 1996; Sefc et al. 2007) and displayed higher morphological similarity to the Yellow-Blotch Mbita morph than to the Blue Nakaku morph (Maderbacher et al. 2008), but, as already mentioned, also featured a considerable amount of individuals that carried haplotypes belonging to the eastern mitochondrial lineage (Yellow-Blotch; 1-A2 and 1-A4; Baric et al. 2003; Sturmbauer et al. 2005; Sefc et al. 2007; all analyses were conducted on the first part of the mitochondrial control region). Populations further north already possessed fewer introgressed haplotypes and displayed the typical bluish phenotype. This remarkable level of mtDNA introgression across Mbete Bay is particularly interesting because of two reasons: (i) The bay clearly separates mtDNA lineages in the cichlid species *Variabilichromis moorii* (Duftner et al. 2006), *Eretmodus cyanostictus* (Sefc et al. 2007) and *Neolamprologus caudopunctatus* (Koblmüller et al. 2007) with no signs of introgression and moreover represents a general phenotypic break in intraspecific color variation in other cichlids (Kohda et al. 1996). (ii) mtDNA introgression was highly asymmetric as no western haplotypes (Blue; 8-G lineage) were found east of the bay, and all

investigated populations there displayed the typical Yellow-Blotch body coloration. As already mentioned in the general introduction, *Tropheus* morphs were shown to display differing strengths of positive color assortative mating preferences, with generally strong assortative preferences between highly distinct morphs, and weak isolation between more similar morphs (Salzburger et al. 2006). However, both, laboratory and consecutive breeding pond experiments showed high asymmetry of mating preferences in the *T. moorii* color morphs that were investigated in the present study: Mate choice experiments in lab aquaria revealed highly assortative preferences of Blue-morph Nakaku females with significant deviations from random mating but lack of discrimination in Yellow-Blotch-morph Mbita females (Egger et al. 2010). To find out if the observed mate preferences in the lab could be translated into actual reproductive success, Hermann et al. (unpublished) set up breeding populations consisting of males and females of both morphs, thus mimicking natural conditions in a hypothetical scenario of secondary contact. After one year reproductive success and mating preferences were inferred from the assignment of offspring to parents. Interestingly similar results as in the lab experiments were observed: Blue-morph females mated highly color assortatively with again significant deviations from random mating, whereas Yellow-Blotch-morph females even seemed to prefer heteromorphic males over their own, displaying also significant deviations from random mating. Male-male competitions between males of the two morphs, however, yielded no significant advantage of one male over the other (Hermann et al. unpublished) thus diminishing the possibility that Blue males generally dominate Yellow-Blotch males. The present study thus addresses two main questions: First, we ask how asymmetric mtDNA introgression across Mbete Bay came about. Therefore we set up two hypotheses: The first hypothesis is that the observed asymmetry is merely due to unidirectional migration across the bay from east to west but not vice versa. In other words fish disperse or dispersed from east to west, where they introgress or introgressed into the resident population but not in the other direction. The second hypothesis is that migration happens or happened both ways but introgression patterns were shaped by the asymmetric mate preferences mentioned before: Blue females (with highly assortative mate preferences) migrating from west to east, would not mate with resident Yellow-Blotch males. In contrast Yellow-Blotch females (which showed no mate discrimination or even preference for the heteromorphic male) migrating from east to west would very well mate with resident Blue males and bring their mtDNA into the population. The best way to address these hypotheses and to unravel the background of the observed asymmetry is the use of nuclear markers. If migration was indeed unidirectional this asymmetry would also be reflected in

nuclear DNA and we would expect to find nuclear introgression west of the bay but not vice versa, congruent with the mtDNA introgression pattern. However, if migration happens or happened both ways but mtDNA introgression was shaped by asymmetric mate preferences of the females, we would still expect to detect both-way nuclear introgression. The reason for this is that immigrant non-choosy Yellow-Blotch females from the east would together with mtDNA pass their nuclear DNA into the resident Blue populations west of the bay. In the other direction Blue males from the west would obtain matings with resident Yellow-Blotch females in the east which would lead to nuclear introgression into the Yellow-Blotch populations east of the bay. Thus to address this first question we applied the following approaches on nine populations around Mbete Bay: (i) analyses on the whole mtDNA D-loop control region and on more populations adjacent to the bay were conducted to confirm previous results concerning asymmetric mtDNA introgression which were carried out only on the first part of the control region, and on fewer populations. (ii) Amplified Fragment Length Polymorphism (AFLP) and microsatellite markers were used to investigate patterns of nuclear introgression in the populations around Mbete Bay and thus unravel the background of the observed mtDNA introgression asymmetry. Secondly, we ask for the time of the introgression event and furthermore, if it was indeed only observed west of Mbete Bay (mitochondrial and nuclear), we want to know which lineage was the one who was resident and which was the one who introgressed. Because although we always speak of introgression of eastern Yellow-Blotch haplotypes into western Blue haplotypes it could under certain circumstances also be the case that it happened the other way around from Blue into Yellow-Blotch. Therefore coalescent based analyses on the mtDNA data set were conducted to estimate splitting times between populations and parts of populations, respectively.

MATERIALS AND METHODS

Taxon sampling and DNA extraction

Genetic data was obtained from 295 (mtDNA), 261 (AFLP) and 301 (SSR) samples, respectively, from nine different locations in southern Lake Tanganyika (Fig. 1). Fish were collected in the lake during field trips between the years 2003 and 2009. Finclips for DNA extraction were taken from each individual and were preserved in 99% ethanol. From most of the samples DNA was isolated using proteinase K digestion followed by a protein precipitation step using ammonium acetate. DNA from samples with low quality tissue was extracted using a Wizard® SV Genome DNA Purification System (Promega) following the manufacturer's instructions. For mtDNA and SSR analysis DNA extracts were diluted 1:10 with deionized water before PCR reaction. For AFLP analysis DNA concentration was measured with a NanoPhotometer™ (IMPLEN) and if required, extracts were diluted with deionized water yielding concentrations of approximately 6 ng/µl for each sample.

mtDNA analysis

Amplification and sequencing

798 bp of mitochondrial control region were amplified from 295 individuals. MtDNA analyses were carried out following the protocol of Egger et al. (2007). DNA amplification via polymerase chain reaction (PCR) was performed with a total volume of 10 µl per sample, each consisting of 1 µl of diluted DNA extract, 1 µl of 10 x $MgCl_2$ buffer, 0.25 µl of 10 x dNTP mix (10 µM), 0.25 µl of each primer (10 µM), 01 µl of *Taq* DNA polymerase (5 units/µl, BioTherm™) and 7.15 µl of deionized water. The temperature profile for the amplification was as follows: After an initial denaturation at 94°C for 3 minutes, 45 cycles of 30 sec at 94°C, 30 sec at 53°C and 1 min 20 sec at 72°C were conducted followed by a final elongation step at 72°C for 7 min and a final hold at 8°C. The primers used for amplification were L15926-T (forward; modified from Kocher et al. 1989) and TDK-D (reverse; Lee et al. 1995) for the first part of the control region and SC-DL (forward; Lee et al. 1995) and TDK-DH4-T (reverse; Nevado et al. 2009) for the second part, which included the poly T region. Sequences of primers are shown in Table 1. After PCR a minigel electrophoresis was carried out in order to check whether amplification was successful and to approximately quantify the

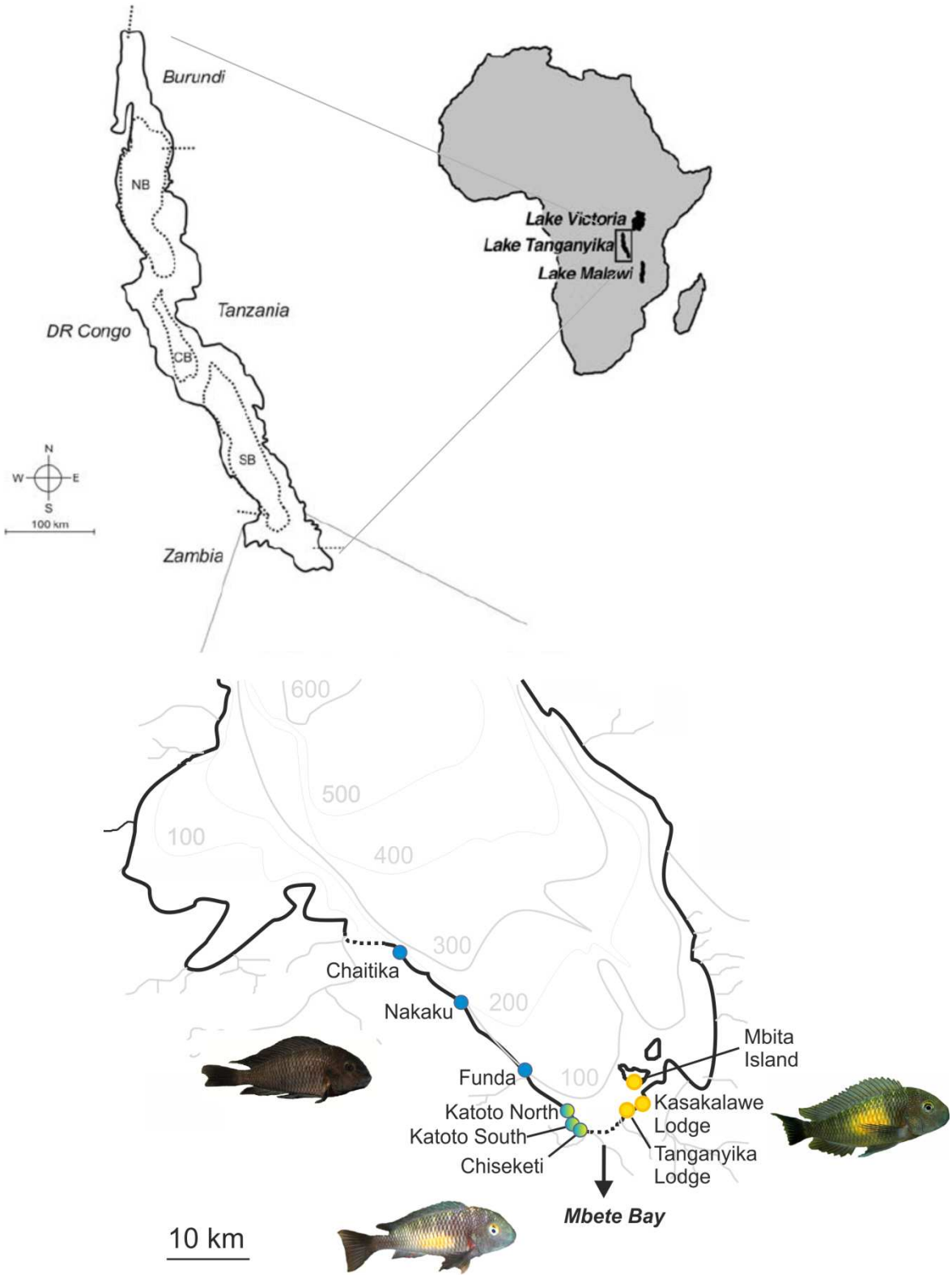


Figure 1 Lake Tanganyika with the nine sampling locations along the southwestern shoreline. Photographs show the three investigated *T. moorii* color morphs. Pictures were either taken from Egger et al. (2007) or by ourselves (Blue morph).

amplified DNA. Therefore 2 µl of PCR product were applied on a SYBR Safe stained gel of 2% D-1 low EEO agarose (Pronadisa) in 1 x TA buffer. PCR products were purified using ExoSAP-IT™ (USB corporation). Chain termination sequencing was carried out with a total volume of 8 µl per sample, each containing 1 µl of purified PCR product, 0.2 µl of Big Dye Termination Reaction Mix (Applied Biosystems), 1.6 µl of 5 x sequencing buffer (Applied Biosystems), 0.125 µl of the primer (10 µM) and 5.075 µl of deionized water with the following temperature profile: 3 min at 94°C as an initial denaturation followed by 35 cycles of 30 sec at 94°C, 20 sec at 50°C and 4 min at 60°C. The sequencing reaction was carried out with three of the four primers used in the amplification reaction: L15926-T for the first part of the control region (406 bp), SC-DL for the second part including the poly T region (244 bp of which 114 bp overlapped with the first part), and TDK-DH4-T for the second part after the poly T region (284 bp).Only if sequencing partly or totally failed with L15926-T, TDK-D was additionally used.

Table 1 Sequences of primers used for the amplification of the control region and chain termination sequencing

Primer name	Sequence	Reference
L15926-T	5'-cag cgc cag agc gcc ggt ctt g- 3'	modified from Kocher et al. 1989
TDK-D	5'-cct gaa gta gga acc aga tg- 3'	Lee et al. 1995
SC-DL	5'taa gag ccc acc atc agt tga- 3'	Lee et al. 1995
TDK-DH4-T	5'-tcc gtc tta aca tct tca gtg tta tgc- 3'	Nevado et al. 2009

After sequencing DNA fragments were purified with Sephadex™ G-50 (Amersham Biosciences) following the manufacturer's instructions and visualized on a 3130 x1 capillary sequencer (Applied Biosystems). Sequences were aligned by eye using MEGA software v.5 (Tamura et al. 2011). 22 bps were removed including the poly T region and adjacent nucleotides due to insufficient sequence quality in this region resulting in a total sequence length of 798 bp.

Data analysis

Indices for DNA sequence variation within each population were calculated using DnaSP v.5.0 (Librado & Rozas 2009). Pair-wise population differentiation based on haplotype frequencies (F_{ST}, calculated after the method of Weir & Cockerham 1984) and uncorrected genetic distances between haplotypes (Φ_{ST}, Excoffier et al. 2005) were calculated in ARLEQUIN v.3.1 (Excoffier et al. 2005). *P*-values were corrected for multiple testing using

the method of Benjamini and Hochberg (1995). MEGA software v.5 (Tamura et al. 2011) was used to construct a p-distance (proportion of nucleotide differences) based neighbor joining (NJ) tree to illustrate phylogenetic relationships among the sample individuals. Assessment of support for the tree topology was done by calculating bootstrap values from 1 000 replicates.

We used ARLEQUIN v.3.1 (Excoffier et al. 2005) to calculate mismatch distributions and estimate the expansion time parameter for a set of populations. Concerning mismatch distributions the assumption is that an observed distribution that does not deviate from a unimodal distribution of pair-wise differences among haplotypes and has a smooth distribution (Harpending 1994) indicates recent population expansion whereas populations at demographic equilibrium usually show a multimodal distribution (Slatkin & Hudson 1991; Rogers & Harpending 1992). The expansion time parameter $\tau = 2\mu t$, where μ stands for the mutation rate per haplotype per generation (Rogers & Harpending 1992), was computed for each population by a generalized non-linear least-square approach as proposed by Schneider & Excoffier (1999). The 95% Confidence Interval (CI) for mismatch distributions and parameters of demographic expansion was obtained by a parametric bootstrap approach (Schneider & Excoffier 1999) based on 10 000 replicates. Admixture of lineages and gene flow between neighboring populations, which are both present in several populations of our data set, can lead to inconsistencies in the test. These may represent the varying sensitivity of populations to divergent immigrant haplotypes thus mimicking long-term constant population size (Sefc et al. 2007). Therefore mismatch distributions were only calculated for the 'pure' Blue populations, with the term 'pure' referring to populations or individual members of populations not suspected to be introgressed, i.e. all individuals of Chaitika and Nakaku and only individuals with western haplotypes of Funda, Katoto North, Katoto South and Chiseketi. This approach should give us insights into the chronology of population expansion of the Blue morph and can be interesting as additional or comparable information to the coalescent analyses described below.

To estimate the time of population splitting ($\tau = t\mu$, where μ is the mutation rate per locus per generation) between pairs of populations, and thus to get information about the direction of introgression, we used the software IMa2 (freely available at http://genfaculty.rutgers.edu/hey/software#IMa2) by Jody Hey. The model assumes a split of an ancestral population into two daughter populations without any influence of selection, recombination, within-population structure, and gene-flow from unsampled populations. Ima2 utilizes a method originally described in Hey & Nielsen (2004, 2007) which is based on Markov Chain Monte Carlo (MCMC) simulations of gene genealogies. First, for each

population pair we conducted initial runs with a burn-in of 100 000 steps and 30 MCMC chains with geometric heating under the HKY model (Hasegawa et al. 1985), given a mutation rate of 0.0000255331 per year. This rate was calculated based on estimates for the molecular clock of the complete mitochondrial control region for East African cichlids of 1.78% (slow rate) to 3.11% (fast rate) per million years, described in Nevado et al. (2009). We used the fast rate for the calculation of the mutation rate and as the IMa2 analyses were only carried out on certain pairs of populations excluding populations from Chaitika and Nakaku, the length of the mtDNA sequence used for that was 821 bp. The initial runs were utilized for an effective exploration of the parameter space and to assure that parameters converged. Based on the results the boundaries of the parameters were adjusted according to the probability distribution's location and the heating parameters were changed if the exploration of the parameter space wasn't sufficient. After that final runs with the same number of burn-in steps, MCMC chains, and heating model were performed. Durations of runs were arranged such that ESS (Effective Sample Sizes) of every parameter were >50 (Hey & Nielsen 2004), and until the number of saved genealogies following the burn-in reached > 20 000, respectively, as recommended by the author. We scored values with the highest posterior probability as best estimates and assessed for each parameter the 90% highest posterior density (HPD) interval, for credibility intervals. This is represented by the shortest span that includes 90% of a parameter's probability density. Divergence parameters were finally translated into years using the mutation rate mentioned above. Ima2 analyses were conducted in order to get an idea which lineage (Blue or Yellow-Blotch) was resident in the area immediately west of the bay first and which lineage was the one that introgressed. Therefore we estimated splitting times between (i) the pooled Yellow-Blotch populations east of Mbete Bay and the individuals west of the bay possessing eastern haplotypes to get an idea of the time when the latter separated from the rest of their morph and (ii) between the 'pure' Blue members of the populations immediately north-west of the bay to estimate the time when the Blue morph expanded to the south and colonized the area adjacent to Mbete Bay. We explored models with different parameter settings: (i) we did not allow for migration in the model (-m0 option). (ii) We allowed for migration and tested both uniform and exponential priors (-j7 option) on migration rates. However, we only present the results from the analysis without migration in the model (i.e. pure isolation model) because runs allowing for migration with uniform priors did not yield reasonable results and runs using exponential priors did not even work properly. Ruling out migration is justifiable because we know from previous studies that gene flow is very low between *T. moorii* populations, even along short

distances, due to strong site philopatry and/or minor discontinuities in the habitat (Sefc et al. 2007; Koblmüller et al. 2011) and major habitat discontinuities like Mbete Bay totally restrict gene flow (Sefc et al. 2007). Moreover we were primarily interested in the relations of splitting times between the different sets of populations, i.e. the chronology of the splitting events, and only secondarily in the absolute values. To achieve that, the highest priority was to use the same model settings in all analyses and this was complied with in our analyses. Every pair-wise analysis was conducted three times with a different random number of seeds to assure for reproducibility of the results. If the replicate runs produced similar parameter estimates, which would be expected, estimates of the longest run were presented in the results.

AFLP analysis

Most of the AFLP genotypes were obtained from the same individuals as used for mitochondrial analysis. However, some samples had to be rejected from the AFLP analysis because of insufficient DNA quality. Depending on availability of samples those were replaced by other samples from the same location resulting in a total number of 261 individuals.

AFLP analysis and fragment scoring

The protocol for AFLP analysis is described in Chapter 2 and restriction/ligation and PCR reactions were conducted with the same adaptors and primers, respectively (Tab. 2) using a total of 18 selective primer combinations. Also the fragment scoring approach was the same as described in Chapter 2 resulting in a data matrix with 1160 loci.

Table 2 Sequences of adaptors and primers used for ligation and PCR

	Name	Sequence
Adaptors	EcoRI ad A	5'-ctc gta gac tgc gta cc- 3'
	EcoRI ad B	5'-aat tgg tac gca gtc tac- 3'
	MseI ad A	5'-gac gat gag tcc tga g- 3'
	MseI ad B	5'-tac tca gga ctc at- 3'
Pre-selective Primers	EcoRI – pre A	5'-gac tgc gta cca att ca- 3'
	MseI – pre C	5'-gat gag tcc tga gta ac- 3'
Selective Primers	EcoRI-ACA	5'-gac tgc gta cca att cac a- 3'
	EcoRI-ACT	5'-gac tgc gta cca att cac t- 3'
	EcoRI-ACC	5'-gac tgc gta cca att cac c- 3'
	MseI-CAA	5'-gat gag tcc tga gta aca a- 3'
	MseI-CAG	5'-gat gag tcc tga gta aca g- 3'
	MseI-CAC	5'-gat gag tcc tga gta aca c- 3'
	MseI-CAT	5'-gat gag tcc tga gta aca t- 3'
	MseI-CTG	5'-gat gag tcc tga gta act g- 3'
	MseI-CTA	5'-gat gag tcc tga gta act a- 3'
	MseI-CTC	5'-gat gag tcc tga gta act c- 3'

AFLP diversity and population structure analyses

AFLP-SURV v.1.0 (Vekemans 2002; Vekemans et al. 2002) was used to estimate allele frequencies, calculate the total number of segregating loci (i.e. fragments that are not always present nor always absent in all individuals), the proportion of polymorphic loci in terms of at least 5% presence or absence of the band in each population, estimated heterozygosity values for each population (i.e. Nei's gene diversity), and a distance matrix of F_{ST} values between every pair of populations. Details on methods and settings are described in Chapter 2. A neighbor joining tree was constructed in MEGA software v.5 (Tamura et al. 2011) based on restriction site distances (Nei & Li 1979) calculated in PAUP* v.4.0 (Swofford 2003). To test for hybridization and to determine the number of genetic clusters that best describes our AFLP data set we used a Bayesian assignment procedure implemented in the widely used program STRUCTURE v.2.3.3 (Pritchard et al. 2000). The original version has been modified for the use of dominant markers by Falush et al. (2007). STRUCTURE assigns individuals according to their genotypes at multiple loci into *K* clusters without using any a priori population information. Simplified, the model assumes that population structure introduces Hardy-Weinberg or linkage disequilibrium and tries to find groupings of populations that are not in disequilibrium, as far as this is possible (Pritchard et al. 2000). For each value of *K* it estimates a log Probability (Pr(X|*K*)), thus allowing the estimation of the more likely number of clusters. Also for each individual the likelihood to belong either to one or the other cluster is estimated (represented through Q, the estimate of membership). We ran STRUCTURE using the admixture-model with correlated allele frequencies (Falush et al. 2003) and let α (the Dirichlet parameter for the degree of admixture) be inferred from the data. Furthermore we set λ, the parameter of the distribution of allelic frequencies, to 1, according to the manuals advice. We conducted a burn-in of 50 000 and a MCMC of 250 000 iterations for the estimation of posterior probabilities. As different runs can produce different likelihood values we carried out 5 replicate runs for each estimation of *K*. The range of possible *K*s was tested from 1 to 10 populations. It should be noted that Pr(X|*K*) is, according to the authors, only an indication of the number of clusters and an ad hoc guide (Pritchard et al. 2000; Pritchard & Wen 2003). Thus we complementary applied another method to detect the number of clusters with the highest probability, which is described in Evanno et al. (2005). Implemented in the program STRUCTURE HARVESTER v.0.6.6 (Earl 2009), this approach calculates ΔK (the rate of change of the likelihood between successive *K* values) from the structure output

LnP(D) to detect the uppermost hierarchical level of genetic structure, and is freely usable at http://taylor0.biology.ucla.edu/struct_harvest/.

Microsatellite analysis

Analyses on 16 microsatellite loci and 301 samples were conducted by Caroline Hermann. Primer information and details on methods are described in her PhD thesis. Indices of microsatellite diversity were calculated with ARLEQUIN v.3.1 (Excoffier et al. 2005) and FSTAT v.2.9.3 (Goudet 1995), respectively. Tests for Hardy-Weinberg and linkage disequilibrium, and pair-wise population differentiation (F_{ST}) were also calculated in ARLEQUIN v.3.1. To determine the most likely number of genetic clusters and thus detect signs of hybridization STRUCTURE v.2.3.3 (Pritchard et al. 2000) was used and results were subsequently analyzed with STRUCTURE HARVESTER v.0.6.6 (Earl 2009; Evanno et al. 2005) to detect the uppermost hierarchical level of genetic structure.

RESULTS

mtDNA analysis

DNA sequence diversity

A total number of 160 mitochondrial haplotypes was found in 295 individuals. Genetic variation was high in all investigated *T. moorii* populations with an average number of 21.4 haplotypes per population (range: 17 – 24; Tab. 3; Fig. 2a), a mean haplotype diversity of 0.957 (range: 0.935 - 0.974; Tab. 3; Fig. 2b), a mean nucleotide diversity of 0.014 (range: 0.00761 – 0.02629; Tab. 3; Fig. 2c), and a mean average number of pair-wise nucleotide differences of 10.96 (Tab. 3). Haplotype diversity was very similar in all populations (coefficient of variation [CV] = 1.64%) and also numbers of haplotypes did not differ greatly (CV = 10.96%). In contrast to that populations differed highly in nucleotide diversity (CV = 50.07%). This was primarily due to the rather high values in the Katoto South (π = 0.02471) and the Chiseketi (π = 0.02629) population (Tab. 3; Fig. 2c). High degrees of genetic diversity in these two populations are concordant with the presence of introgressed haplotypes, i.e. the introduction of genetic variability from other populations.

Table 3 Population information and mtDNA sequence diversity in the nine investigated *T. moorii* populations. All indices were calculated with DnaSP v.5.0. Shown are sample size (*n*), number of haplotypes (*h*), haplotype diversity (H_E), nucleotide diversity (π), and average number of pair-wise nucleotide differences (*k*)

Population	Phenotype	*n*	*h*	H_E	π	*k*
West of Mbete Bay						
Chaitika	Blue	32	24	0.970	0.00761	6.012
Nakaku	Blue	32	20	0.962	0.00960	7.597
Funda	Blue	34	23	0.970	0.00767	6.066
Katoto N	Blue	36	23	0.935	0.01354	10.708
Katoto S	Intermediate	32	19	0.960	0.02471	19.544
Chiseketi	Intermediate	40	17	0.935	0.02629	20.795
East of Mbete Bay						
TanL	Yellow-Blotch	28	21	0.942	0.01048	8.291
KasL	Yellow-Blotch	30	23	0.968	0.01159	9.177
Mbita Island	Yellow-Blotch	31	23	0.974	0.01319	10.430
Total		295	160	0.987	0.0269	21.2565
Mean (s.d.)		32.8 (3.53)	21.4 (2.35)	0.957 (0.016)	0.014 (0.007)	10.96 (5.49)

Katoto N = Katoto North; Katoto S = Katoto South; TanL = Tanganyika Lodge; KasL = Kasakalawe Lodge

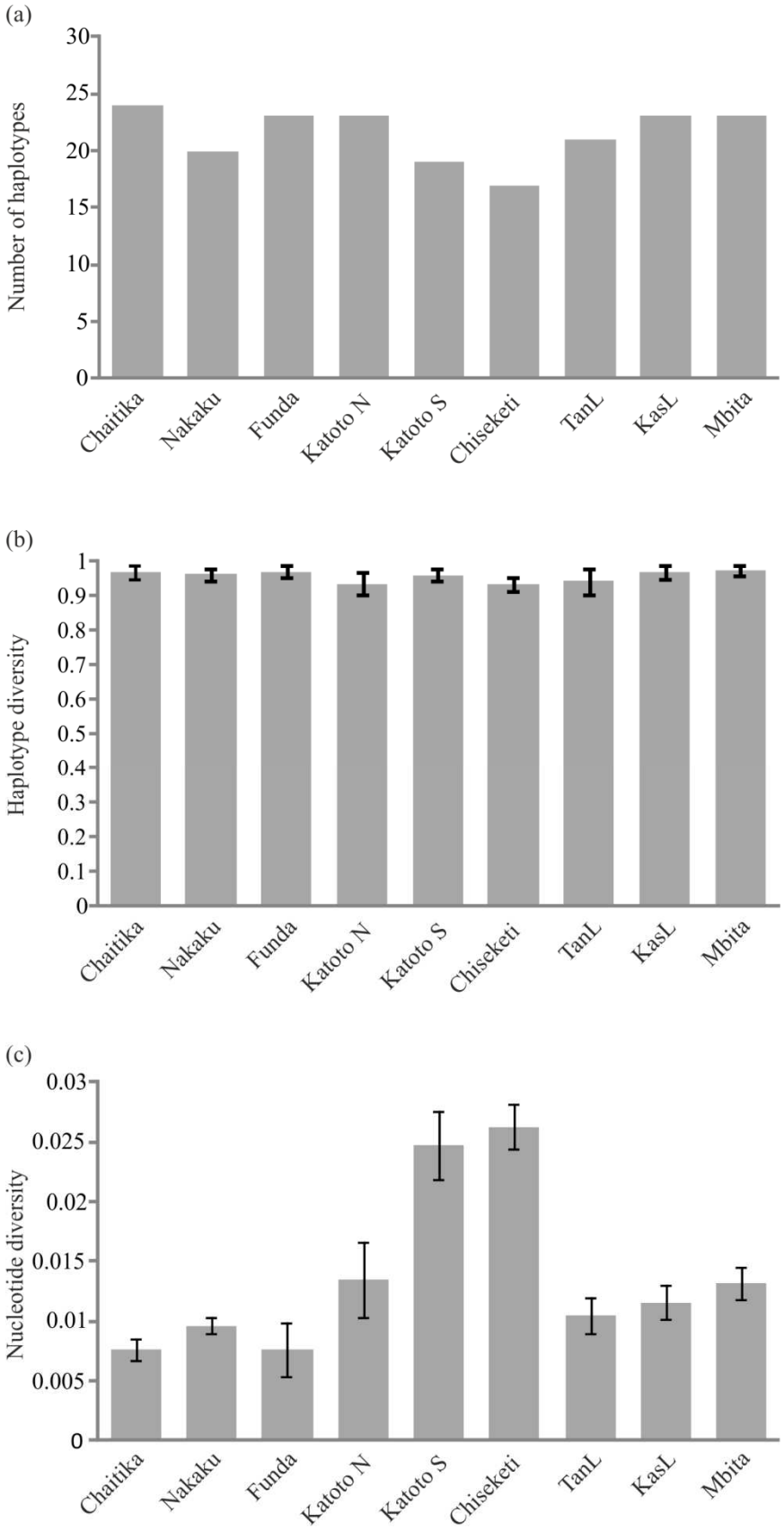


Figure 2 Mitochondrial sequence diversity represented by the number of haplotypes (a), haplotype diversity (b), and nucleotide diversity (c) in each of the investigated *T. moorii* populations. Error bars represent standard deviations.

Mitochondrial phylogenetic relationships

The p-distance based NJ tree constructed with MEGA software v.5 (Tamura et al. 2011) revealed clustering of samples into two major mtDNA lineages with high bootstrap support for the main nodes: Overall, populations from west of Mbete Bay formed one major cluster and populations east of the bay formed the other cluster (Fig. 3). However, there was a considerable number of samples that were collected west of the bay between Chiseketi and Funda but revealed clearly eastern haplotypes and thus clustered within the eastern lineage in our NJ tree (highlighted with arrows in Fig. 3). This was the case for one Funda individual (= 2.9% of all Funda samples), four Katoto North (= 11.1% of all Katoto North samples), 11 Katoto South (= 34.4% of all Katoto South samples), and 16 Chiseketi (= 40% of all Chiseketi samples) samples (Fig. 3, Tab. 4). Thus we found strong signs of mtDNA introgression in the population immediately west of Mbete Bay which declined with further distance from the bay. Vice versa no signs for hybridization were found east of the bay as all individuals sampled there were resolved within the eastern mtDNA lineage. Therefore our data strongly confirm the hypothesis of uni-directional mtDNA introgression into the populations immediately west of Mbete Bay, as already suggested in previous studies conducted only on the first part of the mtDNA control region and fewer populations (Baric et al. 2003; Sturmbauer et al. 2005; Sefc et al. 2007).

Population structure

Analyses of pair-wise population differentiation conducted with ARLEQUIN v.3.1 (Excoffier et al. 2005) revealed strong population structure between most of the investigated populations. F_{ST} values, which represent differentiation based on haplotype frequencies, ranged from -0.003 to 0.060 (mean F_{ST} = 0.033; Tab. 5). Almost all pair-wise analyses revealed highly significant population structure ($p < 0.001$) with the following exceptions: Differentiation between Chaitika and Nakaku was low but still significant ($p < 0.05$). The same could be observed between Katoto North and Katoto South, between Katoto South and Chiseketi, and between Kasakalawe Lodge and Mbita Island ($p < 0.05$ in all pair-wise comparisons). There was no significant differentiation between Katoto North and Chiseketi, Katoto South and Tanganyika Lodge, Katoto South and Kasakalawe Lodge, and between Tanganyika Lodge and Kasakalawe Lodge ($p > 0.05$ in all comparisons; Tab. 5). Φ_{ST} values representing differentiation based on uncorrected genetic distances between haplotypes ranged from -0.012

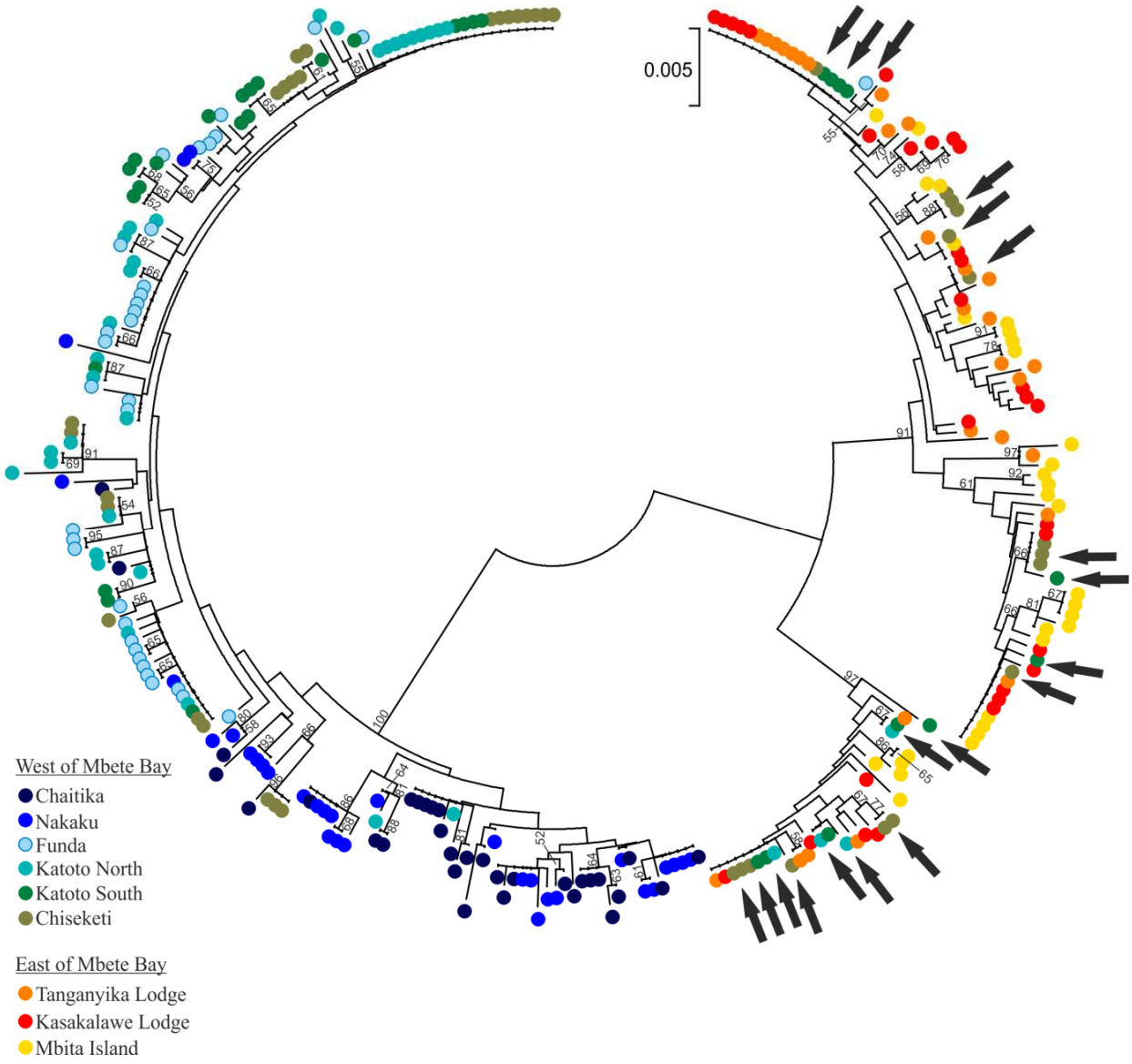


Figure 3 NJ tree of the 295 investigated *T. moorii* samples. The tree was calculated with p-distances based on 798 bp of mitochondrial control region with MEGA v.5 to illustrate phylogenetic relationships among the sample individuals. Cold colors represent individuals that were collected west of Mbete Bay and warm colors represent individuals collected east of the bay. Arrows point to individuals with introgressed haplotypes. Bootstrap values > 50 are shown at the respective branches.

to 0.786 (mean Φ_{ST} = 0.406; Fig. 5). In most cases population structure was highly significant ($p < 0.001$ and $p < 0.01$, respectively) except for the following: Differentiation was again low but still significant between Chaitika and Nakaku and between Katoto North and Katoto South ($p < 0.05$). There was no significant differentiation between Funda and Katoto North ($p > 0.05$), contrasting to the result from the F_{ST} analysis. But also differentiation between Katoto South and Chiseketi, between Tanganyika Lodge and Kasakalawe Lodge, and between Kasakalawe Lodge and Mbita Island were not significant ($p > 0.05$ in all comparisons), which is congruent with the low or non-significance in comparisons between these populations in the F_{ST} analysis. The low and non-significant differentiation between Katoto South and

Tanganyika Lodge and Katoto South and Kasakalawe Lodge, respectively, in the F_{ST} analysis was opposed by a high and significant Φ_{ST} value (Tab. 5). Thus, mitochondrial genetic differentiation between the investigated *T. moorii* populations was found to be generally very high, with a few exceptions which were either diminished by one or the other method of estimation, respectively, or concerned geographically directly adjacent populations. These results are concordant with previous studies on population structure in *T. moorii* (Sefc et al. 2007; Koblmüller et al. 2011).

Table 4 The investigated *T. moorii* populations with their respective sample size (n) and the number of individuals possessing western and eastern haplotypes, respectively. Pie charts illustrate the percentage of each of the two haplotype lineages in each population. Blue represents western haplotypes and yellow represents eastern haplotypes according to the body coloration of the two lineages. Populations are ordered corresponding to their geographical distribution from north-west to south-east and bold letters indicate populations with mtDNA introgression

Population	**n**	**n Western haplotypes**	**n Eastern haplotypes**	**%**
West of Mbete Bay				
Chaitika	32	32	0	
Nakaku	32	32	0	
Funda	**34**	**33**	**1**	
Katoto North	**36**	**32**	**4**	
Katoto South	**32**	**21**	**11**	
Chiseketi	**40**	**24**	**16**	
East of Mbete Bay				
Tanganyika Lodge	28	0	28	
Kasakalawe Lodge	30	0	30	
Mbita Island	31	0	31	

Table 5 Pair-wise mitochondrial population differentiation between the investigated *T. moorii* populations calculated with ARLEQUIN v.3.1. F_{ST} values are below diagonal, Φ_{ST} values above. Benjamini-Hochberg corrected significance levels: *** p < 0.001; ** p < 0.01; * p < 0.05

	Chaitika	Nakaku	Funda	KatN	KatS	Chiseketi	TanL	KasL	Mbita
Chai		0.047*	0.389***	0.288***	0.291***	0.282***	0.781***	0.776***	0.758***
Naka	0.018*		0.293***	0.212***	0.244***	0.242***	0.762***	0.758***	0.740***
Fund	0.026***	0.033***		0.031	0.160***	0.200***	0.786***	0.780***	0.761***
KatN	0.040***	0.050***	0.044***		0.081*	0.122**	0.705***	0.702***	0.687***
KatS	0.031***	0.038***	0.034***	0.016*		-0.012	0.478***	0.476***	0.466***
Chis	0.044***	0.051***	0.045***	0.009	0.019*		0.407***	0.406***	0.400***
TanL	0.040***	0.048***	0.044***	0.060***	0.016	0.052***		0.001	0.077**
KasL	0.027***	0.035***	0.031***	0.047***	0.014	0.036***	-0.003		0.022
Mbit	0.024***	0.032***	0.028***	0.045***	0.033***	0.042***	0.037***	0.015*	

Chai = Chaitika; Naka = Nakaku; Fund = Funda; KatN = Katoto North; KatS = Katoto South; Chis = Chiseketi; TanL = Tanganyika Lodge; KasL = Kasakalawe Lodge; Mbit = Mbita Island

Demographic history

Mismatch distributions calculated with ARLEQUIN v.3.1 (Excoffier et al. 2005) showed clear unimodality in the 'pure' Blue populations from Funda and Katoto South indicating population expansion (Fig. 4). Unimodality was overall also present in the Chaitika and Nakaku population, and within the individuals with western haplotypes in the Katoto North population, but with slight deviations, which may be due to gene flow between populations (Fig. 4). The mismatch distribution for the western haplotypes of the Chiseketi population was very flat and did not show unimodality, possibly again due to gene flow (Fig. 4). The time estimates of population growth revealed generally marginally older estimated expansion times for the northern populations (Chaitika: 95% CI of τ = 1.0 – 12.3; Nakaku: 95% CI of τ = 4.9 – 13.4; Funda: 95% CI of τ = 2.6 – 6.9; Fig. 4 and 5) than for the southern populations (Katoto North: 95% CI of τ = 1.9 – 10.6; Katoto South: 95% CI of τ = 1.6 – 6.6; Chiseketi: 95% CI of τ = 0.03 – 14.6; Fig. 4 and 5) although confidence intervals were mostly wide, especially for Chaitika and Chiseketi, and thus overlapped between populations. However, these results reflect the expansion of the Blue morph from north to south.

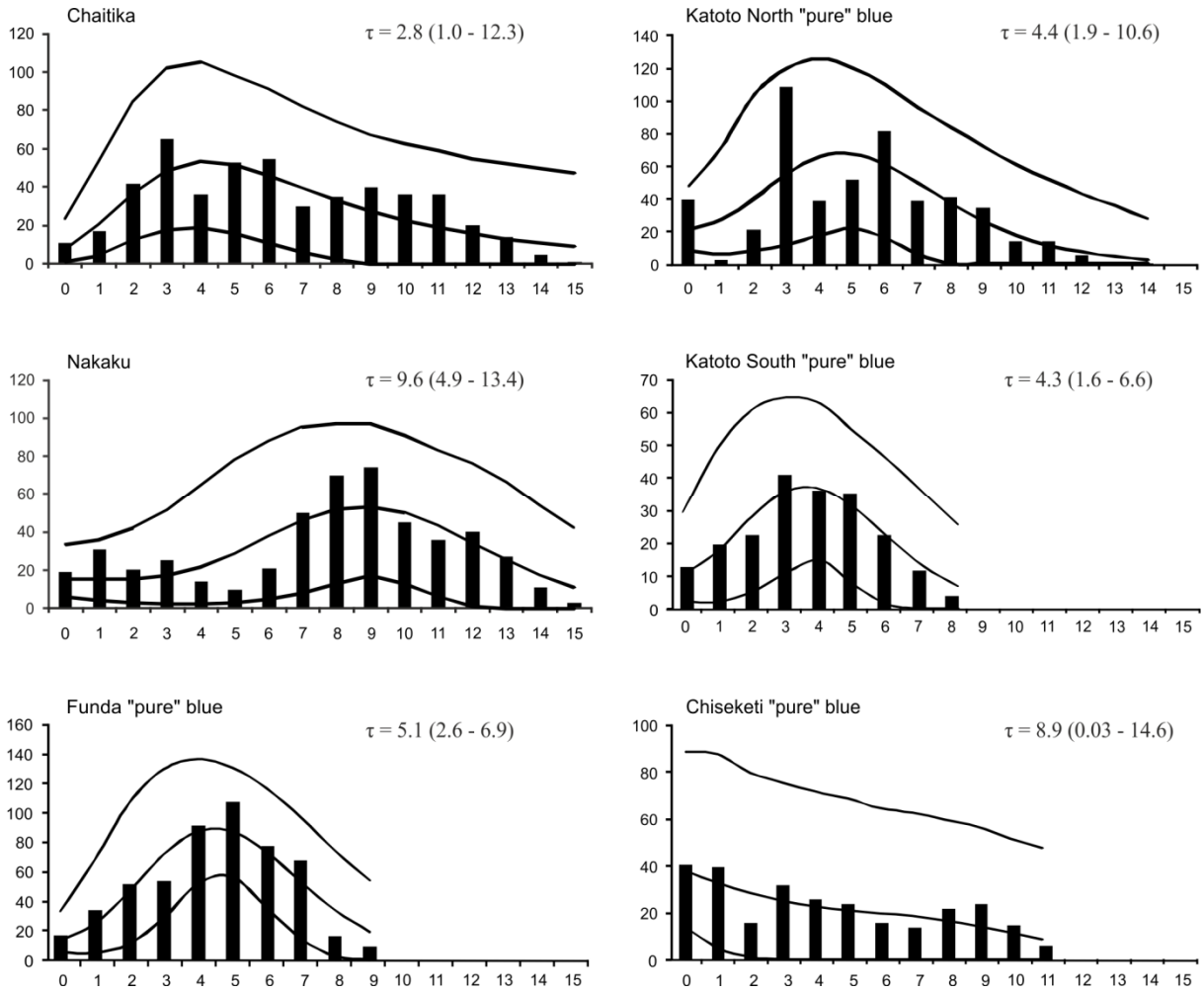


Figure 4 Results from tests for population expansion on 'pure' Blue *T. moorii* populations conducted with ARLEQUIN v.3.1. Observed mismatch distributions among haplotypes are represented by bar charts. Expected distributions, based on parameter estimates with their 95% confidence limits, are represented by lines. Parameter estimates of the time of population growth (τ) are given for each population with their respective 95% confidence intervals.

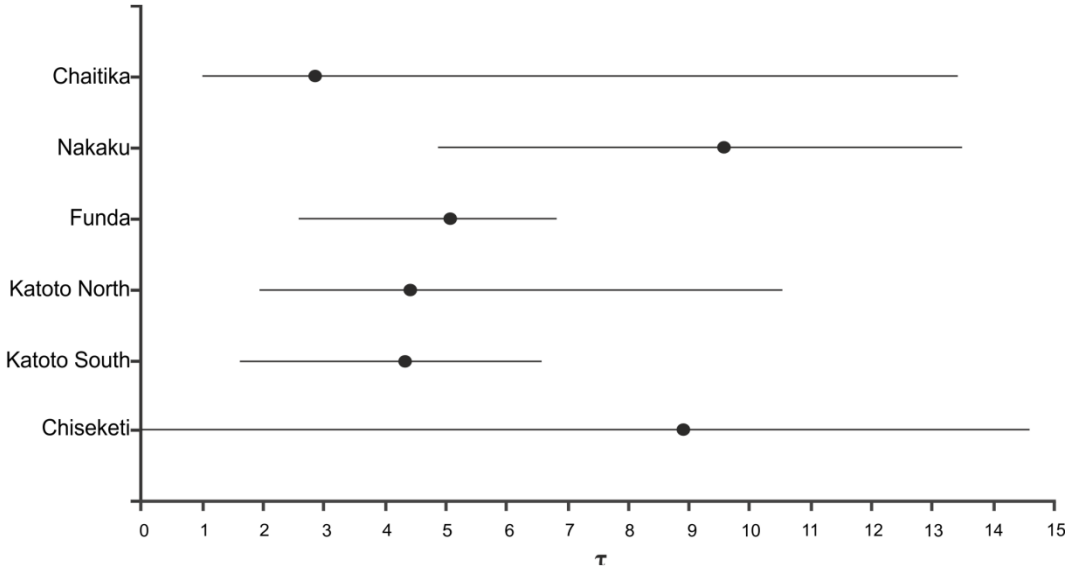


Figure 5 Parameter estimates of the time of population growth with their respective 95% confidence intervals given for all 'pure' Blue *T. moorii* populations. Calculations were conducted with ARLEQUIN v.3.1.

Coalescence-based analysis of splitting times

Ima2 analyses were used to get information about splitting times and thus to find out whether introgression happened from the Yellow-Blotch (eastern) lineage into the Blue (western) lineage or vice versa. For reasons already mentioned (see materials and methods) we present the results of all pair-wise analysis without allowing for migration. The split between the pooled Yellow-Blotch populations east of Mbete Bay and the individuals west of the bay possessing eastern (Yellow-Blotch) haplotypes was dated ~ 21 000 ya (Fig. 6a). Concerning the 'pure' Blue individuals from the populations immediately west of Mbete Bay the following splitting times were estimated: ~ 30 000 ya between Funda Blue and Katoto North Blue (Fig. 6b), ~ 30 000 ya between Katoto North Blue and Katoto South Blue (Fig. 6c), and ~ 17 000 ya between Katoto South Blue and Chiseketi Blue (Fig. 6d).

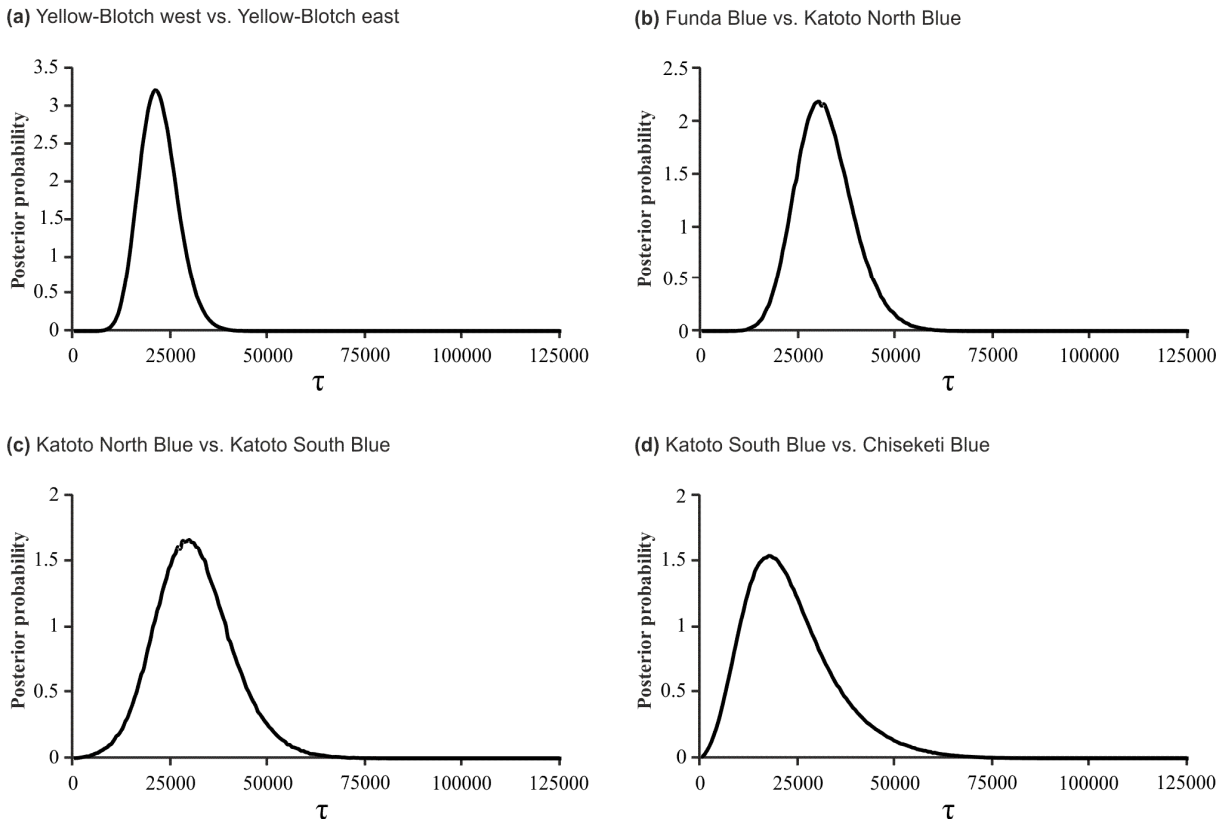


Figure 6 Posterior distributions of the estimated splitting times, i.e. the scaled times since divergence in years (τ) for four pairs of *T. moorii* populations calculated with IMa2. Distributions are shown for the estimated splitting time between the pooled Yellow-Blotch populations east of Mbete Bay and the individuals west of the bay possessing eastern (Yellow-Blotch) haplotypes (a), and between the 'pure' Blue individuals from Funda and Katoto North (b), Katoto North and Katoto South (c), and Katoto South and Chiseketi (d).

AFLP analysis

Genetic diversity

The 18 primer combinations yielded a total of 1160 AFLP loci of which 887 (76.5%) were segregating within the overall data set. The mean number of AFLP bands per individual was 273.3. The mean proportion of polymorphic markers per population was 38.03%, ranging from 36.5% to 40.3%. The average expected heterozygosity (i.e. Nei's gene diversity) was 0.118 ranging from 0.107 to 0.125 (Tab. 6).

Table 6 Genetic diversity measures of the nine investigated *T. moorii* populations calculated with AFLP-SURV v.1.0. Shown are sample size (*n*), percentage of polymorphic loci at the 5% level (*PLP*), and expected heterozygosity under Hardy-Weinberg genotypic proportions (i.e. Nei's gene diversity; H_j) with its standard error (H_j *S.E.*). Values are given for each population and for the whole data set

Population	n individuals	PLP [%]	H_j	H_j S.E.
West of Mbete Bay				
Chaitika	24	37.2	0.125	0.0049
Nakaku	30	40.3	0.124	0.0048
Funda	28	38.5	0.123	0.0048
Katoto North	29	38.8	0.121	0.0048
Katoto South	28	38.9	0.122	0.0048
Chiseketi	32	35.5	0.112	0.0046
East of Mbete Bay				
Tanganyika Lodge	31	37.7	0.109	0.0046
Kasakalawe Lodge	30	38.9	0.118	0.0047
Mbita Island	29	36.5	0.107	0.0046
Total	261	76.5	0.125	
Mean	29	38.0	0.118	0.00225

Phylogenetic relationships

Phylogenetic relationships, inferred from restriction site distances (Nei & Li 1979) and subsequent computation of a NJ tree, again revealed a clear separation between the 'pure' eastern populations (Yellow-Blotch morph; Tanganyika Lodge to Mbita Island) and the 'pure' western populations (Blue morph; Chaitika, Nakaku, and most of the Funda individuals; Fig. 7). Populations that had already shown signs of admixture in the mtDNA analysis occupied an intermediate position between the two 'pure' lineages, as expected. This was the case for some Funda individuals, all Katoto North, all Katoto South and all Chiseketi individuals

except three which even clustered within the Yellow-Blotch morph (Fig. 7). Thus, like mtDNA data also nuclear data revealed the separation of two color morphs and lineages, respectively, across Mbete Bay with the populations immediately west of the bay showing signs of admixture. There was no evidence for introgression from west to east, except perhaps in one Kasakalawe Lodge sample. Bootstrap values were very low (beyond 50%) for most of the nodes (Fig. 7). This is not surprising with hybrids being present in the data set, because they contain AFLP fragments of both parental clades. These appear as homoplasies and reduce bootstrap support for the parental clades (Seehausen 2004; Egger et al. 2007).

Population structure

The observed F_{ST} value over all populations was 0.0591 ($p < 0.0001$). Pair-wise F_{ST} values were moderate but, in contrast to the mtDNA data set, highly significant between all populations. Highest F_{ST} values were detected between the putatively 'pure' Blue populations west of Mbete Bay (i.e. Chaitika and Nakaku) and the putatively 'pure' Yellow-Blotch populations east of the bay (i.e. Tanganyika Lodge, Kasakalawe Lodge and Mbita Island) with an average value of 0.11 (range: 0.101 – 0.122; Tab. 7). F_{ST} values were lower in comparisons between adjacent populations and also in comparisons between the putative hybrid populations and populations of either 'pure' morph (Tab. 7). However, the highly significant population differentiation in all pair-wise comparisons revealed a considerable level of population structure between our investigated *T. moorii* populations on the nuclear DNA level using AFLP as genetic markers.

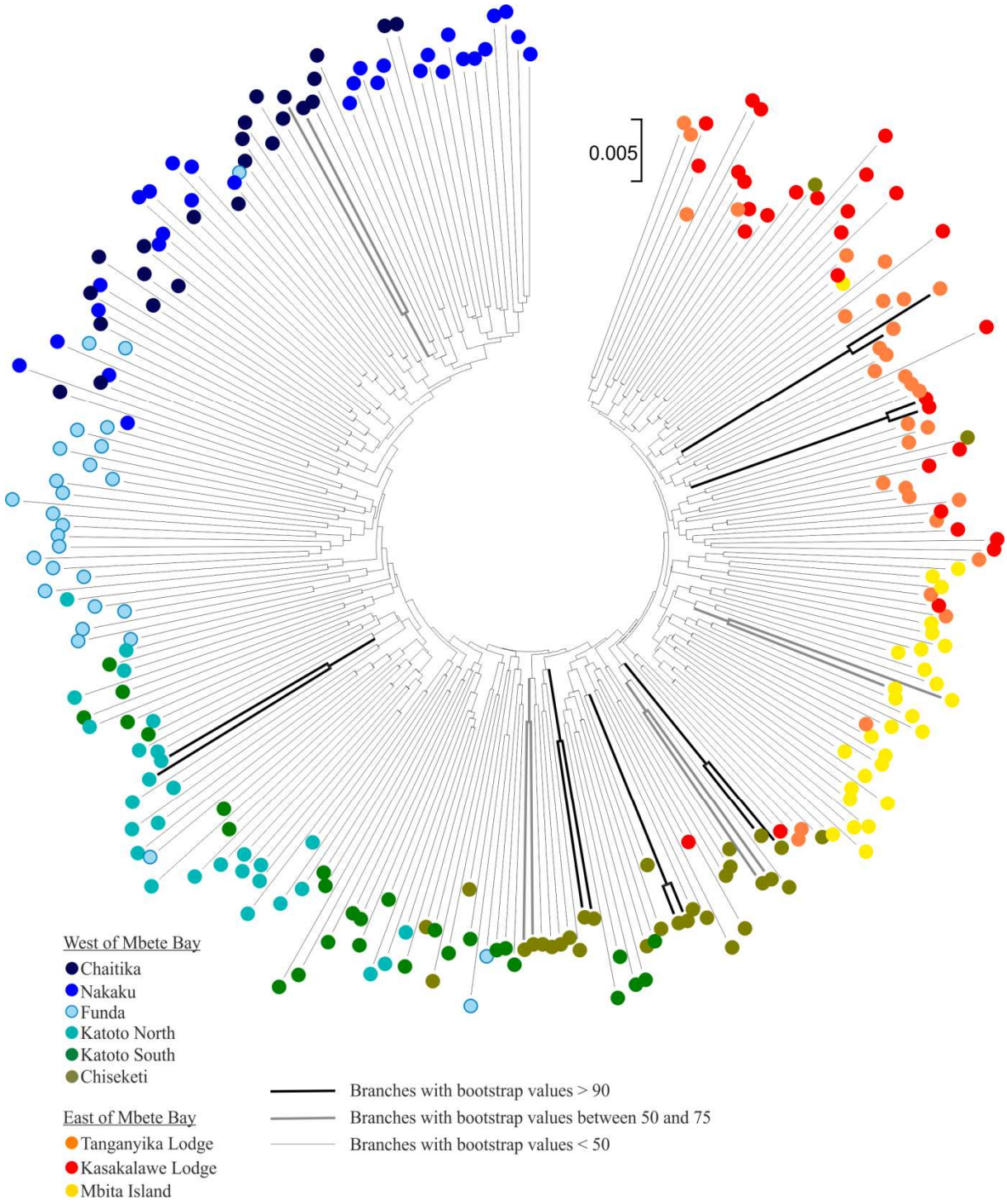


Figure 7 NJ tree of the 261 investigated *T. moorii* samples. The tree was calculated based on AFLP restriction site distances (Nei & Li 1978) with MEGA v.5 to illustrate phylogenetic relationships among the sample individuals. Cold colors represent individuals that were collected west of Mbete Bay and warm colors represent individuals collected east of the bay. Branches with bootstrap values > 50 are bold.

Table 7 Matrix of pair-wise genetic distances between the nine investigated populations calculated with AFLP-SURV v.1.0. F_{ST} values are in the bottom left corner, corresponding *P*-values in upper right corner. Calculation of *P*-values was based on 5000 permutations

	Chai	Naka	Fund	KatN	KatS	Chis	TanL	KasL	Mbit
Chai		<0.0001	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001
Naka	0.0079		<0.0001	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001
Fund	0.0321	0.0318		<0.0001	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001
KatN	0.0545	0.0577	0.0215		<0.0001	<0.0001	<0.0001	<0.0001	<0.0001
KatS	0.0536	0.0563	0.0209	0.0055		<0.0001	<0.0001	<0.0001	<0.0001
Chis	0.0730	0.0798	0.0409	0.0491	0.0284		<0.0001	<0.0001	<0.0001
TanL	0.1009	0.1082	0.0762	0.0753	0.0465	0.0224		<0.0001	<0.0001
KasL	0.1048	0.1078	0.0781	0.0616	0.0422	0.0321	0.0118		<0.0001
Mbit	0.1138	0.1215	0.0951	0.0998	0.0676	0.0461	0.0246	0.0371	

Chai = Chaitika; Naka = Nakaku; Fund = Funda; KatN = Katoto North; KatS = Katoto South; Chis = Chiseketi; TanL = Tanganyika Lodge; KasL = Kasakalawe Lodge; Mbit = Mbita Island

Test for hybridization

AFLP based Bayesian inference of population structure conducted with STRUCTURE v.2.3.3 (Pritchard et al. 2000) and subsequent analysis of the results with STRUCTURE HARVESTER v.0.6.6 (Earl 2009) revealed the highest peak in ΔK for $K = 2$ (Evanno et al. 2005; Fig. 8a). Thus the most likely number of genetic clusters for the whole data set was two, corresponding, overall, to a division between Blue-morph populations north-west of Mbete Bay and Yellow-Blotch-morph populations east of the bay (Fig. 8b). Interestingly we found the same pattern as already revealed by mtDNA analysis: There were no signs of hybridization between the two clusters in the populations east of the bay (Tanganyika Lodge to Mbita Island). However, populations immediately west of the bay showed strong admixture, with Chiseketi individuals possessing even more ‘eastern’ DNA than ‘western’ DNA. The proportion of eastern DNA in western populations decreased with further distance from Mbete Bay with Nakaku and Chaitika being almost ‘pure’ western (i.e. Blue; Fig. 8b and c).

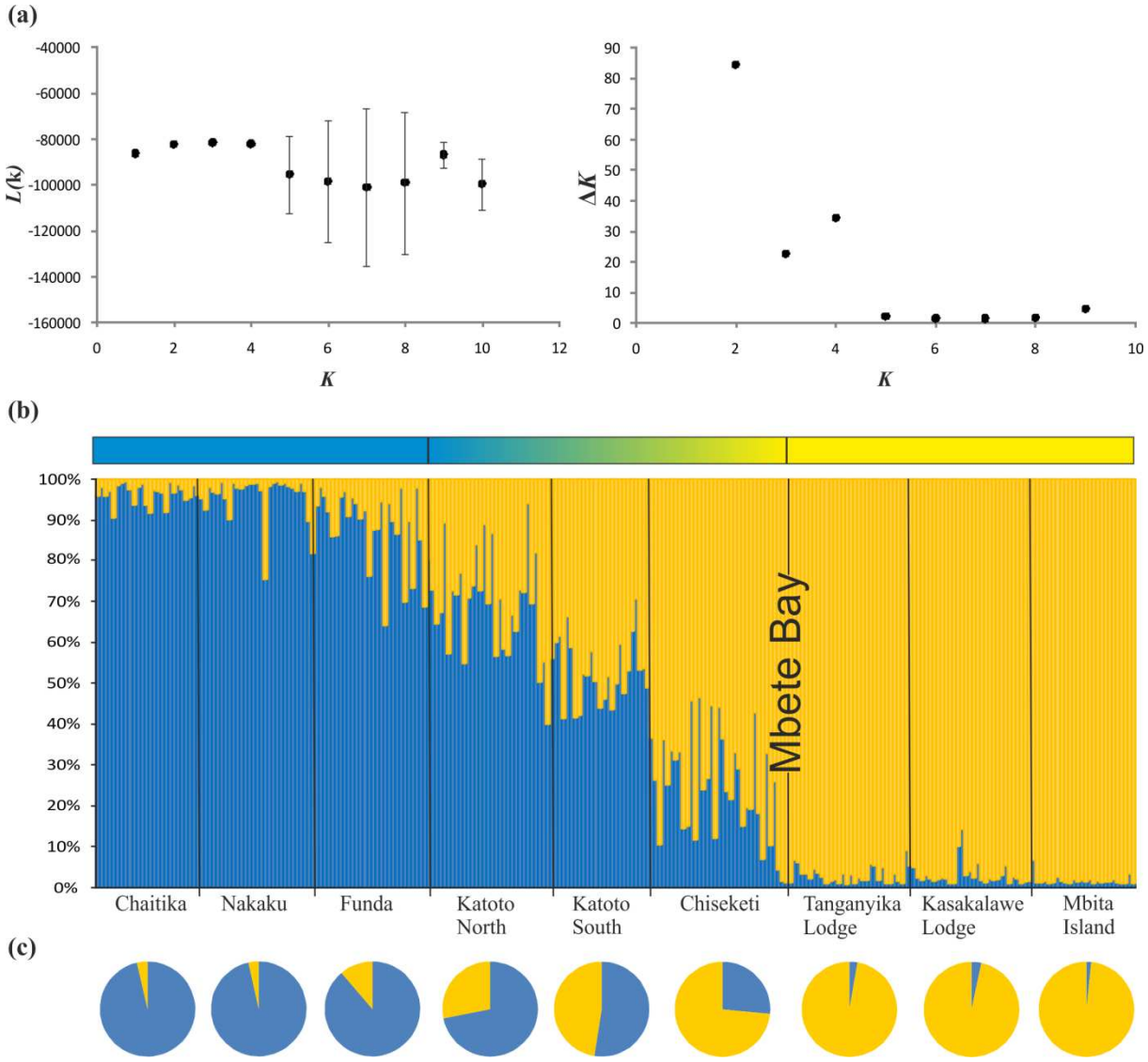


Figure 8 Results of the Bayesian clustering analysis of the nine investigated *T. moorii* populations using 1160 AFLP loci. Analyses were conducted with STRUCTURE v.2.3.3 and STRUCTURE HARVESTER v.0.6.6. (a) Left: the mean likelihood [L(K) ± S.D.] over five replicate runs assuming *K* clusters. Right: Δ*K*; the number of clusters with the highest Δ*K* peak is interpreted as the highest level of structuring. (b) Assignment of each individual to one of the two most probable clusters (as inferred from the Δ*K* statistic). Populations are ordered according to their geographical appearance from north-west to south-east. The color bar represents the body coloration classification of the individuals of each population (Blue – Intermediate – Yellow-Blotch). (c) Pie charts show the percentage of assignment to either of the two inferred clusters averaged over all individuals within each population.

Microsatellite analysis

Genetic diversity and population structure

Microsatellite loci were moderately to highly polymorphic with heterozygosities per locus and population ranging from 25% to 100%, with means of 84% expected and 84% observed heterozygosity across all sampled populations (Tab. 9). None of the 16 loci deviated significantly from Hardy-Weinberg equilibrium after correction for multiple testing (Benjamini & Hochberg 1995). Significant deviation from linkage equilibrium was observed three times, each concerning a different population and locus pair (Funda: UME003 & Pzeb2; Katoto North: Hchi36 & Ppun9; Tanganyika Lodge: Pzeb2 & Pmv3). However, since all three deviations only occurred in one single population we didn't consider them indicative of physical linkage and used all data for subsequent analyses. Almost all pair-wise analyses of population structure revealed highly significant genetic differentiation, with the only exception being differentiation between the geographically very close populations of Tanganyika Lodge and Kasakalawe Lodge. F_{ST} values ranged from 0.001 to 0.064 with an average value of 0.029 (Tab. 8).

Table 8 Pair-wise population differentiation between the nine investited *T. moorii* populations based on 16 microsatellite loci. Benjamini-Hochberg corrected significance levels: *** p < 0.001; ** p < 0.01; * p < 0.05

	Chaitika	Nakaku	Funda	KatN	KatS	Chiseketi	TanL	KasL	Mbita
Chaitika									
Nakaku	0.007**								
Funda	0.016***	0.023***							
Katoto N	0.024***	0.030***	0.013***						
Katoto S	0.022***	0.030***	0.008***	0.007***					
Chiseketi	0.025***	0.035***	0.018***	0.014***	0.012***				
TanL	0.048***	0.064***	0.047***	0.051***	0.037***	0.017***			
KasL	0.039***	0.049***	0.042***	0.043***	0.032***	0.014***	0.001		
Mbita	0.039***	0.055***	0.046***	0.048***	0.038***	0.025***	0.017***	0.008***	

Katoto N = Katoto North; Katoto S = Katoto South; TanL = Tanganyika Lodge; KasL = Kasakalawe Lodge; Mbita = Mbita Island

Test for hybridization

Microsatellite based Bayesian inference of population structure conducted with STRUCTURE v.2.3.3 (Pritchard et al. 2000) and subsequent analysis of the results with STRUCTURE HARVESTER v.0.6.6 (Earl 2009) revealed the highest peak in ΔK for K = 3 (Evanno et al.

2005; Fig. 9a). Thus the most likely number of genetic clusters for the whole data set was three, corresponding, overall, to a division between the populations east of Mbete Bay (the Yellow-Blotch morph) forming the first cluster, the four putative hybrid populations immediately west of the bay forming the second cluster, and the two 'pure' Blue morph populations north-west of the bay forming the third cluster (Fig. 9b). We also present the results for K = 2 for reasons of comparability with the AFLP analysis. This is justifiable because first, the mean likelihood value over the five replicates for K = 2 was only marginally smaller than for K = 3 (L(K2) = -24585.3 vs. L(K3) = -24310.8) , and secondly, K = 2 revealed the second highest peak in the ΔK analysis (Fig. 9a). Moreover, concerning our question, if introgression was observed on both sides of the bay or only west of the bay, both scenarios revealed the same answer: In the K = 3 scenario Yellow-Blotch populations east of Mbete Bay formed one genetic cluster with negligible western influence, whereas populations immediately west of the bay from Chiseketi to Funda represented a distinct cluster with a considerable degree of influence from the east, especially in Chiseketi. These populations also showed influence from the north-western Nakaku and Chaitika populations which represented the third cluster (Fig. 9b). In the K = 2 scenario again Yellow-Blotch populations east of Mbete Bay formed one cluster with no substantial influence from the west, whereas western populations, forming the second cluster, showed high degrees of admixture between the two clusters (Fig. 9c). After some additional STRUCTURE runs excluding certain loci we ascribe signs of western introgression in single individuals east of the bay (especially in the Mbita Island population) in both scenarios to locus-specific effects (homoplasy). As already observed in the AFLP data set the population in Chiseketi again showed an even higher assignment to the eastern, than to the western cluster (Fig. 9c and d). Overall, the degree of admixture decreased with further distance from the bay with Nakaku and Chaitika being already 'pure' Blue (Fig. 9c and d). Thus both scenarios revealed one way nuclear introgression in the populations immediately west of the Mbete Bay but no substantial signs of introgression east of the bay. These findings were congruent with the results from the mtDNA and AFLP analysis.

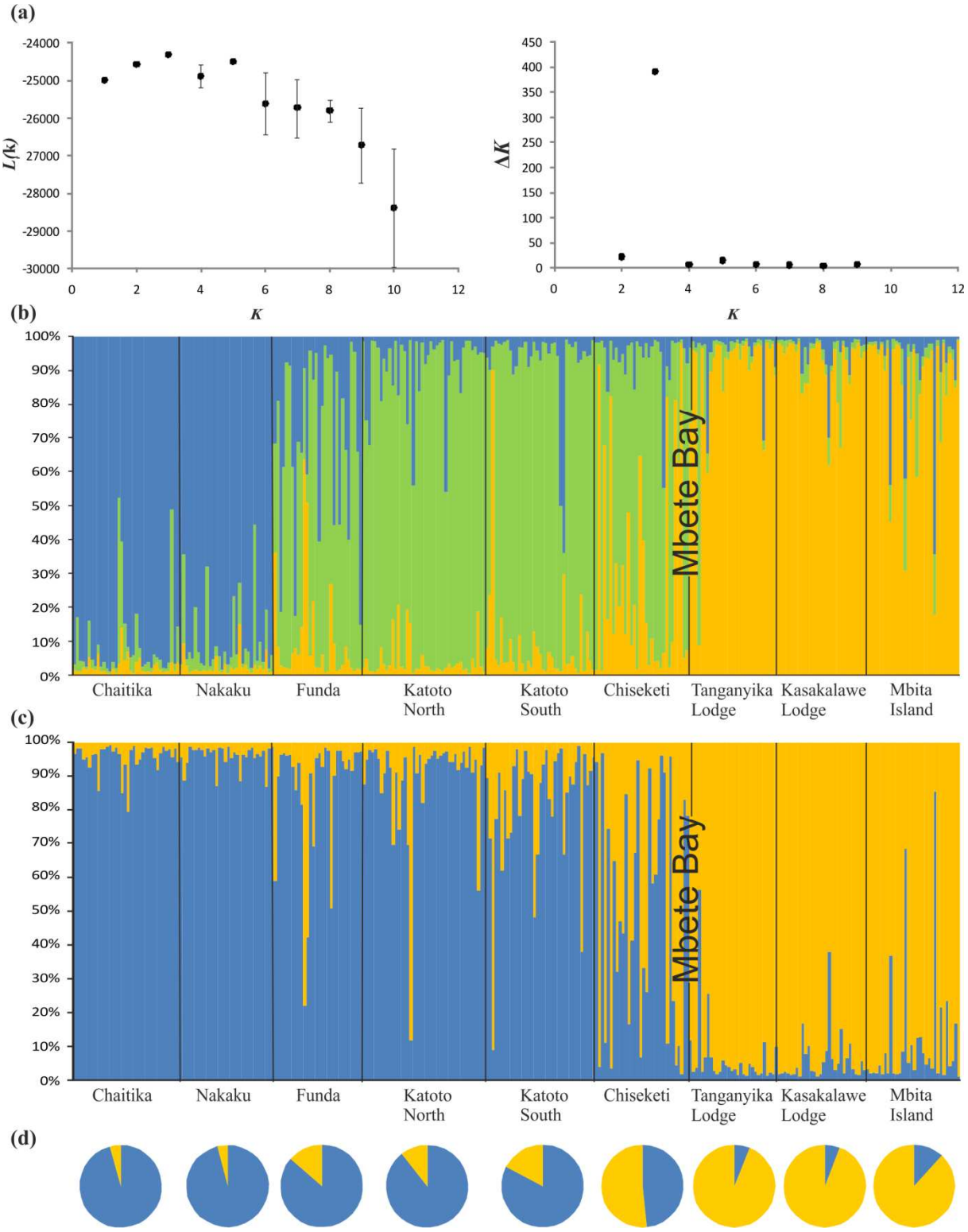


Figure 9 Results of the Bayesian clustering analysis of the nine investigated *T. moorii* populations using 16 microsatellite loci. Analyses were conducted with STRUCTURE v.2.3.3 and STRUCTURE HARVESTER v.0.6.6. (a) Left: the mean likelihood [L(K) ± S.D.] over five replicate runs assuming *K* clusters. Right: Δ*K*; the number of clusters with the highest Δ*K* peak is interpreted as the highest level of structuring. (b) Assignment of each individual to one of the three most probable clusters (as inferred from the Δ*K* statistic). (c) Assignment of individuals to two clusters, representing the second most probable scenario (as inferred from the Δ*K* statistic). (d) Pie charts show the percentage of assignment to either of the two clusters in the *K* = 2 scenario averaged over all individuals within each population.

Table 9 Microsatellite diversity in the nine investigated *T. moorii* populations calculated with ARLEQUIN v.3.1 and FSTAT v.2.9.3, respectively. For each population the sample size is given (n) and for each population and each locus the number of alleles (N_A), the observed heterozygosity (H_O), and the expected heterozygosity (H_E) are shown

			Locus																
Population	n		UME002	UME003	TmoM11	TmoM27	UNH130	UNH154	Pzeb3	UNH908	UNH2016	Pzeb2	Hchi36	Pmv3	Ppun9	Hchi6	Pmv17	Hchi1	Average
West of Mbete Bay																			
Chaitika	36	N_A	10	26	18	12	21	7	13	7	16	20	21	32	26	21	31	11	18.25
		H_O	0.67	0.94	0.89	0.85	1.00	0.58	0.75	0.67	0.86	0.97	0.89	0.97	0.94	0.89	1.00	0.84	0.86
		H_E	0.68	0.93	0.93	0.79	0.93	0.66	0.69	0.54	0.85	0.93	0.92	0.97	0.96	0.94	0.97	0.84	0.85
Nakaku	32	N_A	10	23	17	11	19	6	10	6	18	20	19	24	27	18	25	10	16.44
		H_O	0.59	0.91	0.94	0.80	0.94	0.69	0.72	0.69	0.91	0.94	0.97	0.94	0.97	0.97	0.84	0.78	0.85
		H_E	0.58	0.89	0.92	0.75	0.94	0.76	0.67	0.63	0.83	0.93	0.94	0.95	0.96	0.90	0.94	0.81	0.84
Funda	30	N_A	11	18	15	10	20	5	9	4	21	17	18	21	22	18	24	12	15.31
		H_O	0.73	1.00	0.90	0.77	0.90	0.43	0.87	0.53	0.83	0.97	0.88	0.83	0.97	0.93	0.93	0.88	0.84
		H_E	0.76	0.92	0.92	0.71	0.92	0.50	0.85	0.50	0.89	0.93	0.94	0.94	0.95	0.88	0.96	0.91	0.84
Katoto North	44	N_A	13	23	15	11	20	6	14	4	18	19	18	24	26	18	27	12	16.75
		H_O	0.80	0.98	0.93	0.93	0.82	0.64	0.86	0.43	0.76	0.98	0.98	0.91	0.95	0.98	0.93	0.68	0.85
		H_E	0.81	0.94	0.90	0.80	0.90	0.56	0.86	0.54	0.81	0.91	0.93	0.95	0.95	0.92	0.96	0.73	0.84
Katoto South	36	N_A	9	22	17	12	23	7	13	5	23	21	15	16	19	17	27	11	16.06
		H_O	0.78	0.97	0.92	0.78	0.97	0.50	0.94	0.25	0.83	1.00	0.94	0.92	0.92	0.97	1.00	0.69	0.84
		H_E	0.77	0.95	0.91	0.78	0.91	0.65	0.88	0.28	0.91	0.94	0.92	0.91	0.94	0.92	0.95	0.82	0.84
Chiseketi	32	N_A	12	20	17	10	22	6	12	4	21	19	16	18	20	17	23	9	15.38
		H_O	0.88	1.00	1.00	0.63	0.94	0.66	0.91	0.50	0.97	0.97	0.97	0.88	0.91	0.94	1.00	0.77	0.87
		H_E	0.83	0.94	0.91	0.69	0.89	0.67	0.88	0.57	0.92	0.93	0.91	0.92	0.95	0.92	0.96	0.73	0.85
East of Mbete Bay																			
Tanganyika Lodge	30	N_A	12	19	17	10	14	5	11	4	18	18	20	19	18	21	19	10	14.69
		H_O	0.87	0.93	0.80	0.67	0.87	0.47	0.83	0.27	0.87	0.93	0.88	0.93	0.93	0.93	1.00	0.81	0.81
		H_E	0.87	0.94	0.91	0.61	0.83	0.42	0.85	0.27	0.90	0.94	0.94	0.91	0.93	0.94	0.94	0.85	0.82
Kasakalawe Lodge	30	N_A	14	18	17	9	19	5	13	4	15	21	20	17	18	18	22	10	15.00
		H_O	0.93	1.00	0.93	0.83	0.83	0.47	0.90	0.47	0.87	0.97	0.97	0.90	0.93	0.97	0.97	0.87	0.86
		H_E	0.83	0.93	0.92	0.81	0.88	0.44	0.86	0.41	0.90	0.94	0.94	0.91	0.94	0.93	0.95	0.77	0.84
Mbita	31	N_A	9	19	17	12	18	7	9	3	14	21	17	18	19	20	22	11	14.75
		H_O	0.84	0.90	0.94	0.77	0.81	0.32	0.94	0.32	0.74	0.90	0.94	0.90	0.94	1.00	0.87	0.84	0.81
		H_E	0.83	0.91	0.92	0.84	0.88	0.32	0.79	0.30	0.83	0.94	0.93	0.93	0.94	0.94	0.95	0.84	0.82

DISCUSSION

Hybridization and introgression west of Mbete Bay

Our results indicate strong evidence for hybridization between two genetically and phenotypically divergent *Tropheus moorii* lineages, the Blue '8-G' lineage (Baric et al. 2003; Sturmbauer et al. 2005; Egger et al. 2007) and the Yellow-Blotch '1-A' lineage (Baric et al. 2003; Sturmbauer et al. 2005; Egger et al. 2007), which are separated by a large habitat barrier, the Mbete Bay. Our mtDNA data set confirmed previously reported uni-directionality of introgression in the populations immediately west of the bay (Baric et al. 2003; Sturmbauer et al. 2005; Sefc et al. 2007) as again no mtDNA hybridization could be observed within the Yellow-Blotch lineage east of the bay. Moreover the two populations with the highest amount of introgressed haplotypes, Chiseketi and Katoto South, showed elevated degrees of mitochondrial genetic diversity, in terms of nucleotide diversity, which is due to the introduction of genetic variability from other populations. Therefore the next step was to find out how this one-way mtDNA introgression came about and thus to confirm or reject either of the two hypotheses: (i) The observed asymmetry is merely due to unidirectional migration that happens or happened from east to west but not vice versa. (ii) Migration happens or happened both ways but female mate preferences shaped the observed mtDNA introgression pattern. The analyzed AFLPs and microsatellites, representing nuclear DNA markers, both confirmed uni-directionality of introgression, i.e. introgression in the populations west of Mbete Bay but only weak signs for hybridization east of the bay. These results rule out the hypothesis of asymmetric female mate preferences as the cause of unidirectional mtDNA introgression west of Mbete Bay and thus provide space for alternative explanations. There are two more possible reasons for the observed uni-directionality of mtDNA introgression. Both are again diminished by the fact that nuclear introgression was also found to be asymmetric in the same direction but for the sake of completeness I still want to mention them here: (i) It could be that mating between color morphs might have happened east and west of the bay at similar amounts but a selective sweep or strong stochastic drift led to the asymmetric distribution of introgressed mitochondrial lineages: Western haplotypes could have been eliminated in populations east of the bay, while eastern haplotypes were driven to high frequencies in populations west of the bay. However, this is also not very likely because of two reasons: First, as noted, in this case we would still find nuclear introgression on both sides of the bay, which we didn't. Secondly, a selective sweep as well as strong stochastic

drift following the introgression event would have led to a reduction of genetic diversity within the eastern haplotypes in the populations west of the bay compared to the source population (e.g. Kasakalawe Lodge). This was not the case as haplotype diversity was only marginally reduced among eastern haplotypes of the two populations comprising the highest number of introgressed haplotypes (Chiseketi: n = 16; h = 9; H_E = 0.917; Katoto South: n = 11; h = 7; H_E = 0.873) relative to the Kasakalawe Lodge population (n = 30; h = 23; H_E = 0.968) and nucleotide diversity was even slightly higher than within the Kasakalawe Lodge population (Chiseketi: π = 0.015; Katoto South: π = 0.014; Kasakalawe Lodge: π = 0.012). Thus no evidence exists for a selective sweep or strong stochastic drift. (ii) The observed asymmetric mitochondrial introgression pattern could have been caused by the fact that the fitness of reciprocal F1 hybrids is also asymmetric. In other words hybrids of Yellow-Blotch female/Blue male pairings could have a higher fitness than those of Blue female/Yellow-Blotch male pairings. Thus, although the latter also mate and reproduce, the offspring are just not viable for some reason. Again, two findings stand against this possibility: First, if Yellow-Blotch female/Blue male pairs had indeed higher fitness, nuclear introgression should be found on both sides of the bay, because immigrant Blue males from the west would produce offspring with high fitness with resident Yellow-Blotch females. Secondly, in two reciprocal *T. moorii* breeding populations of individuals of the Blue Nakaku and the Yellow-Blotch Mbita Island population (25 males of one morph and 75 females of the other morph in the first pond and vice versa in the second pond; Sturmbauer et al. unpublished) both pairings produced similar F1 numbers within a certain time. Also F2 have been produced from each of the reciprocal F1 but without statistical comparison of fitness so far. One possibility that could explain asymmetry in both, mitochondrial and nuclear introgression would be that single fish indeed only cross the bay from east to west and not in the other direction. However, we consider this very unlikely as there is no plausible explanation for that and data that would indicate some environmental influence that could lead to such unidirectional dispersal behavior, like e.g., water currents occurring exclusively from east to west, does not exist.

Demographic history

Initially we thought of gene flow happening from east to west, caused by migrating Yellow-Blotch individuals that brought their genes across Mbete Bay into the resident Blue populations. However, due to the fact that it is very unlikely that dispersal indeed happened only one-way, we have to think about other scenarios where this asymmetric pattern could

have arisen. One interesting option is that introgression did not happen from Yellow-Blotch into Blue as previously thought but from Blue into Yellow-Blotch. This could have happened at some time in the past in a scenario where the Yellow-Blotch morph was able to cross Mbete Bay in the course of a lake level that was higher than at present. Indeed there is evidence for such a high lake-level scenario in the late Pleistocene/early Holocene: There was a severe low-stand in Lake Tanganyika's water level during the last glacial maximum, which was dated between ~15 000 and ~35 000 ya (e.g. Cohen et al. 1997, 2007; Scholz et al. 2003, 2007; Felton et al. 2007; McGlue et al. 2008; Burnett et al. 2010). The LGM-induced aridity led to a drop of the water level by ~250 – 300 m (e.g. McGlue et al. 2008; Burnett et al. 2011), and after that arid period a long-term rise in lake level started (Burnett et al. 2011). Felton et al. (2007) dated an overflow of Lake Kivu through the Ruzizi River into Lake Tanganyika ~ 10 000 ya, or maybe prior to this. Previous geochemical data also suggested an overflow of Lake Tanganyika between 13 000 and 14 000 ya (pers. comm. with M. Talbot in Barker & Gasse 2003). In such a high lake-level period the surrounding escarpment at Mbete Bay would have been suitable *Tropheus* habitat with rocky underground. The Mbete Bay shoreline could have been colonized by expanding Yellow-Blotch populations from the east whereas populations of the Blue lineage might not yet have colonized the area at that time. Thus, a possible scenario could have been the following: IMa2 analyses yielded an estimated splitting time between the individuals with eastern haplotypes west of Mbete Bay and the pure Yellow-Blotch populations east of the bay of ~ 21 000 Ya. This time estimate coincides with recently calculated splitting times between other *T. moorii* populations of the Yellow-Blotch morph: Koblmüller et al. (2011) dated the splits of three geographically relatively close populations occurring east of Mbete Bay at ~ 15 000 – 30 000 years ago and associated an additional boost of population growth in these *Tropheus* populations with the time of rising lake level after the LGM. Thus our analysis suggests that the split of the Yellow-Blotch lineage across Mbete Bay occurred at the same time as the splits between the Yellow-Blotch populations east of the bay, which was presumably triggered by colonization of novel shoreline in the course of the water level rise after the LGM. With again dropping water level after the high-stand the sandy stretch of Mbete Bay became once again a dispersal barrier for *Tropheus* individuals but some individuals of the westernmost Yellow-Blotch population stayed in the rocky area immediately west of the bay. In the meantime the Blue morph could have started to expand further to the south. This was shown through an estimated splitting time between the pure blue individuals from Funda and Katoto North of ~ 30 000 ya and a similar dating for the split between Blue individuals from Katoto North and Katoto South.

The split between Chiseketi Blue and Katoto South Blue was dated at ~ 17 000 ya, i.e. after the estimated colonization by the yellow-blotch morph at 21 000 ya. Felton et al. (2007) dated the overflow of Lake Kivu into Lake Tanganyika at ~10 000 ya but as already mentioned the authors did not rule out the possibility that this event could have happened earlier. Adding the uncertainty around the molecular dates, it could be possible that the Yellow-Blotch population was already present west of the bay when the Blue morph expanded from the north and this could have led to the introgression from the Blue morph into the Yellow-Blotch morph in Katoto and Chiseketi at a time when Mbete Bay had already become an insuperable barrier once again.

Implications for color diversity and conclusion

Hybridization has been suggested as a force driving the generation of new taxa in cichlids (Salzburger et al. 2002; Seehausen 2004; Schelly et el. 2006; Stelkens et al. 2009) and is usually detected through the intermediate color pattern and/or body shape of the hybrids (Egger et al. 2007). In a previous molecular genetic study on *Tropheus* populations all over Lake Tanganyika several potential cases of admixture and introgression between morphs have been revealed and in some instances these events could have given rise to novel color phenotypes (Egger et al. 2007). Our study confirmed introgressive hybridization in one of these cases, between the Blue and the Yellow-Blotch *T. moorii* morph, which led to the generation of an intermediate color pattern and therefore demonstrates the potential of hybridization to create new phenotypes. Thus, considering the whole *Tropheus* genus, hybridization may very likely serve as a considerable contribution to the outstanding phenotypic diversity present within this taxon and future investigations may very likely reveal more cases of hybridization. Finally, this study shows that interpreting hybridization scenarios can be a very challenging task and highlights the importance of applying multiple approaches like the use of different genetic markers, the investigation of behavioral patterns, and the consideration of paleolimnological data. The combination of these data is particularly important when special introgression patterns occur as it is the case between the two morphs around Mbete Bay investigated in this study.

REFERENCES

Baric S, Salzburger W, Sturmbauer C (2003) Phylogeography and evolution of the Tanganyikan cichlid genus *Tropheus* based upon mitochondrial DNA sequences. Journal of Molecular Evolution 56:54-68

Barker P, Gasse F (2003) New evidence for a reduced water balance in East Africa during the Last Glacial Maximum: implication for model-data comparison. Quaternary Science Reviews 22:823-837

Benjamini Y, Hochberg Y (1995) Controlling the false discovery rate: a practical and powerful approach to multiple testing. Journal of the Royal Statistical Society B 57:289-300

Burnett AP, Soreghan MJ, Scholz CA, Brown ET (2011) Tropical East African climate change and its relation to global climate: A record from Lake Tanganyika, Tropical East Africa, over the past 90+ kyr. Palaeogeography Palaeoclimatology Palaeoecology 303:155-167

Cohen AS, Lezzar K-E, Tiercelin J-J, Soreghan M (1997) New palaeogeographic and lake-level reconstructions of Lake Tanganyika: implications for tectonic, climatic and biological evolution in a rift lake. Basin Research 9:107-132

Cohen AS, Stone JR, Beuning KRM, Park LE, Reinthal PN, Dettman D, Scholz CA, Johnson TC, King JW, Talbot MR, Brown ET, Ivory SJ (2007) Ecological consequences of early Late Pleisotcene megadroughts in tropical Africa. Proceedings of the National Academy of Sciences of the USA 104:16422-16427

Duftner N, Sefc KM, Koblmueller S, Nevado B, Verheyen E, Phiri H, Sturmbauer C (2006) Distinct population structure in a phenotypically homogeneous rock-dwelling cichlid fish from Lake Tanganyika. Molecular Ecology 15:2381-2395

Earl D (2009) Structure Harvester v.0.6.6, Available at http://taylor0.biology.ucla.edu/struct_harvest (accessed July 2011)

Egger B, Koblmueller S, Sturmbauer C, Sefc KM (2007) Nuclear and mitochondrial data reveal different evolutionary processes in the Lake Tanganyika cichlid genus *Tropheus*. Bmc Evolutionary Biology 7:-

Egger B, Mattersdorfer K, Sefc KM (2010) Variable discrimination and asymmetric preferences in laboratory tests of reproductive isolation between cichlid colour morphs. J Evol Biol 23:433-9

Evanno G, Regnaut S, Goudet J (2005) Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study. Molecular Ecology 14:2611-2620

Excoffier L, Laval G, Schneider S (2005) Arlequin (version 3.0): an integrated software package for population genetics data analysis. Evolutionary Bioinformatics Online 1:47-50

Falush D, Stephens M, Pritchard JK (2003) Inference of Population Structure Using Multilocus Genotype Data: Linked Loci and Correlated Allele Frequencies. Genetics 164:1567-1587

Falush D, Stephens M, Pritchard JK (2007) Inference of population structure using multilocus genotype data: dominant markers and null alleles. Molecular Ecology Notes 7:574-578

Felton AA, Russell JM, Cohen AS, Baker ME, Chesley JI, Lezzar KE, McGlue MM, Pigati JS, Quade J, Stager JC, Tiercelin JJ (2007) Paleolimnological evidence for the onset and termination of glacial aridity from Lake Tanganyika, Tropical East Africa. Palaeogeography Palaeoclimatology Palaeoecology 252:405-423

Goudet J (1995) FSTAT (Version 1.2): A computer program to calculate F-statistics. Journal of Heredity 86:485-486

Harpending HC (1994) Signature of ancient population growth in a low-resolution mitochondrial DNA mismatch distribution. Human Biology 66:591-600

Hasegawa M, Kishino H, Yano TA (1985) Dating of the human ape splitting by a molecular clock of mitochondrial DNA. Journal of Molecular Evolution 22:160-174

Hey J, Nielsen R (2004) Multilocus methods for estimating population sizes, migration rates and divergence time, with applications to the divergence of *Drosophila pseudoobscura* and *D. persimilis*. Genetics 167:747-760

Hey J, Nielsen R (2007) Integration within the Felsenstein equation for improved Markov chain Monte Carlo methods in population genetics. Proceedings of the National Academy of Science USA 104:2785-2790

Koblmüller S, Sefc KM, Duftner N, Warum M, Sturmbauer C (2007) Genetic population structure as indirect measure of dispersal ability in a Lake Tanganyika cichlid. Genetica 130:121-131

Koblmüller S, Salzburger W, Obermuller B, Eigner E, Sturmbauer C, Sefc KM (2011) Separated by sand, fused by dropping water: habitat barriers and fluctuating water levels steer the evolution of rock-dwelling cichlid populations in Lake Tanganyika. Molecular Ecology 20:2272-2290

Kocher TD, Thomas WK, Meyer A, Edwards SV, Pääbo S, Villablanca FX, Wilson AC (1989) Dynamics of mitochondrial DNA evolution in animals: amplification and sequencing with conserved primers. Proceedings of the National Academy of Sciences of the USA 86:6196-6200

Kohda M, Yanagisawa Y, Sato T, Nakaya K, Niimura Y, Matsumoto K, Ochi H (1996) Geographical colour variation in cichlid fishes at the southern end of Lake Tanganyika. Environmental Biology of Fishes 45:237-248

Lee W-J, Conroy J, Huntting Howell W, D KT (1995) Structure and evolution of teleost mitochondrial control regions. Journal of Molecular Evolution 41:54-66

Librado P, Rozas J (2009) DnaSP v5: a software for comprehensive analysis of DNA polymorphism data. Bioinformatics 25:1451-1452

Maderbacher M, Bauer C, Herler J, Postl L, Makasa L, Sturmbauer C (2008) Assessment of traditional versus geometric morphometrics for discriminating populations of the *Tropheus moorii* species complex (Teleostei : Cichlidae), a Lake Tanganyika model for allopatric speciation. Journal of Zoological Systematics and Evolutionary Research 46:153-161

McGlue MM, Lezzar KE, Cohen AS, Russell JM, Tiercelin J-J, Felton AA, Mbede E, Nkotagu HH (2008) Seismic records of late Pleistocene aridity in Lake Tanganyika, tropical East Africa. Journal of Paleolimnology 40:635-653

Nei M, Li WH (1979) Mathematical model for studying genetic variation in terms of restriction endonucleases. Proceedings of the National Academy of Sciences 76:5269-5273

Nevado B, Koblmüller S, Sturmbauer C, Snoeks J, Usano-Alemany J, Verheyen E (2009) Complete mitochondrial DNA replacement in a Lake Tanganyika cichlid fish. Molecular Ecology 18:4240-4255

Pritchard JK, Stephens M, Donnelly P (2000) Inference of population structure using multilocus genotype data. Genetics 155:945-959

Pritchard JK, Wen W (2003) Documentation for STRUCTURE software Version 2. Available from http://pritch.bsd.uchicago.edu.

Rogers AR, Harpending H (1992) Population growth makes waves in the distribution of pairwise genetic differences. Molecular Biology and Evolution 9:552-569

Salzburger W, Baric S, Sturmbauer C (2002) Speciation via introgressive hybridization in East African cichlids? Molecular Ecology 11:619-625

Salzburger W, Niederstatter H, Brandstatter A, Berger B, Parson W, Snoeks J, Sturmbauer C (2006) Colour-assortative mating among populations of *Tropheus moorii*, a cichlid fish from Lake Tanganyika, East Africa. Proceedings of the Royal Society B-Biological Sciences 273:257-266

Schelly R, Salzburger W, Koblmueller S, Duftner N, 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

Schneider S, Excoffier L (1999) Estimation of past demographic parameters from the distribution of pairwise differences when the mutation rates vary among sites: application to human mitochondrial DNA. Genetics 152:1079-1089

Scholz CA, King JW, Ellis GS, Swart PK, Stager JC, Colman SM (2003) Paleolimnology of Lake Tanganyika, East Africa, over the past 100 k yr. Journal of Paleolimnology 30:139-150

Scholz CA, Johnson TC, Cohen AC, al. e (2007) East African megadroughts between 135 and 75 thousand years ago and bearing on early-modern human origins. Proceedings of the National Academy of Sciences of the USA

Seehausen O (2004) Hybridization and adaptive radiation. Trends in Ecology and Evolution 19:198-207

Sefc KM, Baric S, Salzburger W, 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

Slatkin M, Hudson RR (1991) Pairwise comparisons of mitochondrial-DNA sequences in stable and exponentially growing populations. Genetics 129:555-562

Stelkens RB, Schmid C, Selz O, Seehausen O (2009) Phenotypic novelty in experimental hybrids is predicted by the genetic distance between species of cichlid fish. Bmc Evolutionary Biology 9:Article No.: 283

Sturmbauer C, Koblmüller S, Sefc KM, Duftner N (2005) Phylogeographic history of the genus *Tropheus*, a lineage of rock-dwelling cichlid fishes endemic to Lake Tanganyika. Hydrobiologia 542:335-366

Swofford DL (2003) PAUP* - Phylogenetic Analysis Using Parsimony and other methods, version 4.0 Sunderland: Sinauer.

Tamura K, Peterson D, Peterson N, Stecher G, Nei M, Kumar S (2011) MEGA5: Molecular Evolutionary Genetics Analysis using Maximum Likelihood, Evolutionary Distance, and Maximum Parsimony Methods. Molecular Biology and Evolution

Vekemans X (2002). AFLP-SURV. Version 1.0. Distributed by the author, Laboratoire Ge´ne´tique et Ecologie Ve´ge´tale, Universite´ Libre de Bruxelles, Belgium.

Vekemans X, Beauwens T, Lemaire M, Roldan-Ruiz I (2002) Data from amplified fragment length polymorphism (AFLP) markers show indication of size homoplasy and of a relationship between degree of homoplasy and fragment size. Molecular Ecology 11:139-151

Weir BS, Cockerham CC (1984) Estimating F-statistics for the analysis of population structure. Evolution 38:1358-1370

CHAPTER 4

Hybridization as potential origin of novel color patterns in *Tropheus*

ABSTRACT

Phenotypic diversity is a very challenging topic for evolutionary biologists as the identification and characterization of the forces triggering it are not an easy task. The cichlid species of the East African Rift Valley Lakes represent a striking example for rapidly evolved phenotypic diversity. Among them Lake Tanganyika's *Tropheus moorii* especially stands out as this species complex comprises numerous allo- and parapatrically distributed color morphs. Coloration plays a crucial role in communication during mating and also other social interactions and the underlying mechanisms that triggered color diversification in this species have been discussed. Sexual selection, e.g., has been suggested to drive the rapid evolution of color variants in *Tropheus* but without conclusive evidence so far. Another factor generally often discussed to trigger phenotypic diversification is hybridization. In *Tropheus* opportunity for hybridization could have been provided through Lake Tanganyika's recurrent water level fluctuations which might have led to secondary contact between formally allopatric morphs. In the present study we test for signs of introgression from the two adjacent color morphs (Red morph and Blue morph) into a set of populations displaying an orange to yellow body coloration (Yellow morph). Therefore 13 *Tropheus moorii* populations in southern Lake Tanganyika belonging to the three color morphs were investigated using a large set of mitochondrial (control region) and nuclear (AFLP and SNP) data. MtDNA data revealed no signs of admixture as the Yellow morph populations totally clustered within the Red populations in a phylogenetic reconstruction. Nuclear DNA in contrast showed strong impact of the Blue morph into the putative hybrid's gene pool and phylogenetic reconstruction and analyses of population structure suggested an intermediate status of the Yellow morph populations between the Red and the Blue morph. Thus we found clear evidence for introgression from both adjacent morphs into the Yellow populations. The absence of Blue mtDNA in the hybrid populations can most plausibly be explained by genetic drift and subsequent total replacement of Blue mtDNA by Red mtDNA. This study confirms yet another incidence of hybridization in *T. moorii* and shows that genetic admixture in this species complex may very likely contribute to the outstanding phenotypic diversity. Moreover it once again highlights the importance of using multiple genetic markers to shed light on a complex topic such as hybridization.

INTRODUCTION

The second hybridization study concerns another contact zone on the western shore of Lake Tanganyika. Here, immediately north of the Lufubu river estuary between Linangu and Chisanze, some populations exist which display a yellow to orange body coloration (populations in closer proximity to the estuary are more yellow and body color gets more orange with further distance from it) and therefore are referred to as the 'Yellow' morph. Further north of them we find red-colored populations which we thus classify as the 'Red' morph. South of the river estuary populations display a yellowish basic color that is overlain by dark melanin with bluish elements, thus referred to as the 'Blue' morph. The Red and the Blue morph belong to two highly divergent mitochondrial lineages: The Red morph populations belong to the so called '7-F' lineage and the Blue morph populations to the '8-G' lineage (Baric et al. 2003; Sturmbauer et al. 2005, Egger et al. 2007). Furthermore they are resolved in two different clades by AFLP analysis, i.e. sub-clade 1a for the Blue morph and sub-clade 4b for the Red morph (both placed within major clade AFLP 4; Egger et al. 2007). Concerning the phylogenetic status of the Yellow morph populations within the whole *Tropheus* genus, previous studies revealed inconsistencies when using mtDNA or nuclear DNA data. Regarding mtDNA analyses, representatives of the Yellow morph clustered within the 7-F lineage together with Red morph samples from Moliro, Chimba, and Kachese (Baric et al. 2003; Sturmbauer et al. 2005; Egger et al. 2007), thus Lufubu River estuary seems to serve as a barrier to the distribution of mitochondrial lineages. In contrast to that Yellow-morph samples, were placed within the clade that contains the Blue morph (clade 3) when using AFLP data (Egger et al. 2007). The only exception were samples from Chisanze which were still grouped within the cluster containing the Red morph samples (clade 4b), which is not surprising as the Chisanze population is geographically and phenotypically closest to the Red morph. A homoplasy excess test (Seehausen 2004), carried out in the same AFLP study revealed some evidence that the Yellow morph populations are influenced by introgression from both, the Red and the Blue morph (Egger et al. 2007). With the aim to test if and to what extend hybridization between the Red and the Blue morph in the scenario of secondary contact could happen, mate choice and breeding pond experiments with representatives of these two morphs have already been carried out. Mate choice aquaria experiments conducted on individuals from Nakaku, representing the Blue morph and individuals from Moliro, representing the Red morph revealed highly assortative female preferences with significant deviations from random mating in the Red morph. Also in the Blue morph assortative

preferences were observed, although no significant deviation from random mating could be detected. However, this was attributed to the large proportion of negative trials, due to the use of females that were not receptive yet (Egger et al. 2008). These findings are concordant with another study on four artificially admixed *Tropheus* morphs, which has revealed strong assortative preferences between highly distinct morphs, and weak isolation between more similar morphs (Salzburger et al. 2006). Those results don't seem to conform to the suspected introgression. However, to find out if the observed mate preferences in the lab could be translated into actual reproductive success, Hermann et al. (unpublished) set up breeding populations consisting of males and females of both morphs (this time using Blue Chaitika and Red Chimba individuals), thus mimicking natural conditions in a hypothetical scenario of secondary contact. After one year reproductive success and mating preferences were inferred from the assignment of offspring to parents. The results deviated slightly from those observed in the lab experiments: Red females mated in fact highly color assortatively with significant deviation from random mating in all five ponds. Blue females in contrast mated color assortatively with significant deviation from random mating in only two ponds, whilst in the other three ponds both, assortative and dissasortative mating occurred. These results implicate at least the possibility of hybridization in a scenario of secondary contact and thus support the hypothesis of introgression between the two morphs.

In the present study we investigated 13 *Tropheus moorii* populations in southern Lake Tanganyika belonging to the Red, the Blue and the Yellow morph. The aim was to detect and confirm signs of hybridization of the Yellow morph using a large-scale data set with different genetic marker types. First, we used whole mitochondrial control region data to either support or complement results of previous studies which were carried out only on a small sample size. Secondly, we used nuclear markers in terms of amplified fragment length polymorphisms (AFLPs) and single nucleotide polymorphisms (SNPs) to confirm previous evidence of hybridization even if that evidence was absent in the mtDNA data set.

MATERIALS AND METHODS

Taxon sampling and DNA extraction

Genetic data was obtained from 304 (mtDNA), 269 (AFLP) and 277 (SNP) samples, respectively, from 13 (mtDNA), 12 (AFLP) and 11 (SNPs), respectively, different locations in southern Lake Tanganyika (Fig. 1). The three samples from the Lufubu population were only used for mtDNA analysis because of low DNA quality thus resulting in only 12 populations for the AFLP data set. SNP sequencing, in turn, also failed for samples from Ndole, thus resulting in only 11 populations for the SNP data set. Fish were either collected in the lake during field trips between the years 2003 and 2009, or purchased from ornamental fish traders. Finclips for DNA extraction were taken from each individual and were preserved in 99% ethanol. From most of the samples DNA was isolated using proteinase K digestion followed by a protein precipitation step using ammonium acetate. DNA from samples with low quality tissue was extracted using a Wizard® SV Genome DNA Purification System (Promega) following the manufacturer's instructions. For mtDNA and SNP analysis DNA extracts were diluted 1:10 with deionized water before PCR reaction. For AFLP analysis DNA concentration was measured with a NanoPhotometer™ (IMPLEN) and if required, extracts were diluted with deionized water yielding concentrations of approximately 6 ng/µl for each sample.

mtDNA analysis

Amplification and sequencing

802 bp of mitochondrial control region were amplified from 304 individuals. Sequences for the Chilanga, Linangu, and some Kachese individuals were provided by Peter Zoppoth. MtDNA analyses were carried out following the protocol of Egger et al. (2007) as described in Chapter 3. The primers used for amplification were again L15926-T (forward; modified from Kocher et al. 1989) and TDK-D (reverse; Lee et al. 1995) for the first part of the control region and SC-DL (forward; Lee et al. 1995) and TDK-DH4-T (reverse; Necado et al. 2009) for the second part, which included the poly T region. Sequences of primers are shown in Table 1. Purification of PCR product, subsequent sequencing, purification of the sequencing product, visualization, and alignment were conducted as described in Chapter 3.

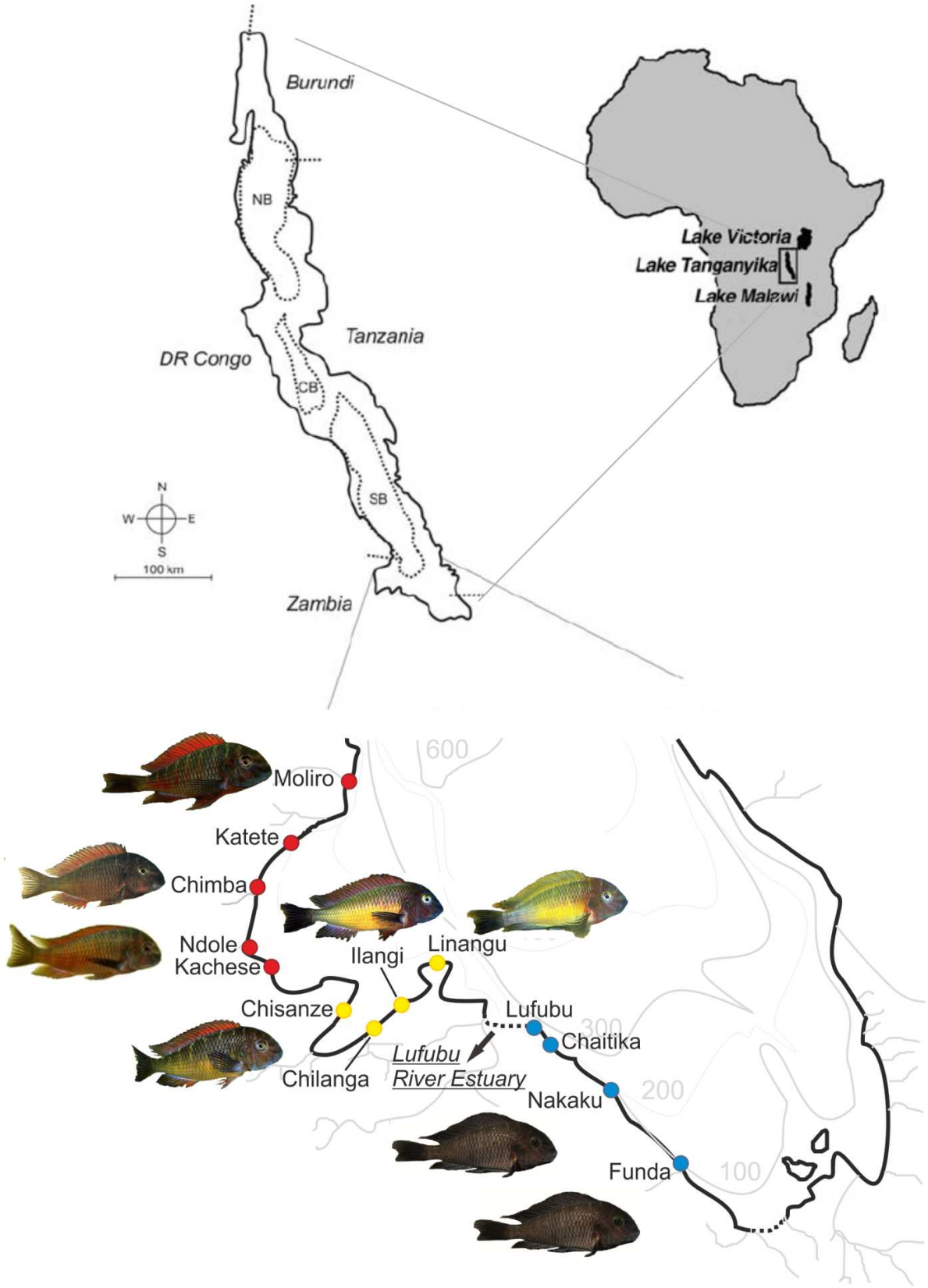


Figure 1 Lake Tanganyika with the 13 sampling locations along the southwestern shoreline. Representative photographs were either taken from Egger et al. (2007) or taken by ourselves (Blue morph) and show the investigated *T. moorii* color morphs.

The sequencing reaction was carried out with three of the four primers used in the amplification reaction: L15926-T for the first part of the control region (409 bp), SC-DL for the second part including the poly T region (244 bp of which 114 bp overlapped with the first part), and TDK-DH4-T for the second part after the poly T region (285 bp). Only if sequencing partly or totally failed with L15926-T, TDK-D was additionally used. 22 bps were removed including the poly T region and adjacent nucleotides due to insufficient sequence quality in this region resulting in a total sequence length of 802 bp.

Table 1 Sequences of primers used for the amplification of the control region and chain termination sequencing

Primer name	Sequence	Reference
L15926-T	5'-cag cgc cag agc gcc ggt ctt g- 3'	modified from Kocher et al. 1989
TDK-D	5'-cct gaa gta gga acc aga tg- 3'	Lee et al. 1995
SC-DL	5'taa gag ccc acc atc agt tga- 3'	Lee et al. 1995
TDK-DH4-T	5'-tcc gtc tta aca tct tca gtg tta tgc- 3'	Nevado et al. 2009

Data analysis

As in Chapter 3 indices for DNA sequence variation within each population were calculated using DnaSP v.5.0 (Librado & Rozas 2009). Pair-wise population differentiation based on haplotype frequencies (F_{ST}, calculated after the method of Weir & Cockerham 1984) and uncorrected genetic distances between haplotypes (Φ_{ST}, Excoffier et al. 2005) were calculated in ARLEQUIN v.3.1 (Excoffier et al. 2005). *P*-values were corrected for multiple testing using the method of Benjamini and Hochberg (1995). MEGA software v.5 (Tamura et al. 2011) was used to construct a p-distance (proportion of nucleotide differences) based neighbor joining tree (NJ) to illustrate phylogenetic relationships among the sample individuals. Assessment of support for the tree topology was done by calculating bootstrap values from 1 000 replicates.

AFLP analysis

Most of the AFLP genotypes were obtained from the same individuals as used for mitochondrial analysis. However, several samples had to be rejected from the AFLP analysis because of insufficient DNA quality. Depending on availability of samples some were replaced by other samples from the same location leading to a total of 269 individuals.

AFLP analysis and fragment scoring

The protocol for AFLP analysis is described in Chapter 2 and restriction/ligation and PCR reactions were conducted with the same adaptors and primers, respectively (Tab. 2) using a total of 18 selective primer combinations. Also the fragment scoring approach was the same as described in Chapter 2 resulting in a data matrix with 1160 loci.

AFLP diversity and population structure analyses

AFLP-SURV v.1.0 (Vekemans 2002; Vekemans et al. 2002) was used to estimate allele frequencies, calculate the total number of segregating loci (i.e. fragments that are not always present nor always absent in all individuals), the proportion of polymorphic loci in terms of at least 5% presence or absence of the band in each population, estimated heterozygosity values for each population (i.e. Nei's gene diversity), and a distance matrix of F_{ST} values between every pair of populations. Details are described in Chapter 2. As in Chapter 3 a neighbor joining tree (NJ) was constructed in MEGA software v.5 (Tamura et al. 2011) based on restriction site distances (Nei & Li 1979) calculated in PAUP* v.4.0 (Swofford 2003). To test for hybridization and to determine the number of genetic clusters that best describes our AFLP data set we used a Bayesian assignment procedure implemented in the widely used program STRUCTURE v.2.3.3 (Pritchard et al. 2000). The original version has been modified for the use of dominant markers by Falush et al. (2007). STRUCTURE assigns individuals according to their genotypes at multiple loci into *K* clusters without using any a priori population information. For details on the method see Chapter 3. We ran STRUCTURE using the same settings as in Chapter 3 (admixture model with correlated allele frequencies; infer α from the data; $\lambda = 1$; burn-in = 50 000; MCMC iterations = 250 000). We conducted again 5 replicate runs and tested the range of possible *K*s from 1 to 13 populations. For a final calculation of *K*, results were analyzed in STRUCTURE HARVESTER v.0.6.6 (Evanno et al. 2005; Earl 2009; freely usable at http://taylor0.biology.ucla.edu/struct_harvest/) as described in Chapter 3.

Table 2 Sequences of AFLP adaptors and primers used for ligation and PCR

	Name	Sequence
Adaptors	EcoRI ad A	5'-ctc gta gac tgc gta cc- 3'
	EcoRI ad B	5'-aat tgg tac gca gtc tac- 3'
	MseI ad A	5'-gac gat gag tcc tga g- 3'
	MseI ad B	5'-tac tca gga ctc at- 3'
Pre-selective Primers	EcoRI – pre A	5'-gac tgc gta cca att ca- 3'
	MseI – pre C	5'-gat gag tcc tga gta ac- 3'
Selective Primers	EcoRI-ACA	5'-gac tgc gta cca att cac a- 3'
	EcoRI-ACT	5'-gac tgc gta cca att cac t- 3'
	EcoRI-ACC	5'-gac tgc gta cca att cac c- 3'
	MseI-CAA	5'-gat gag tcc tga gta aca a- 3'
	MseI-CAG	5'-gat gag tcc tga gta aca g- 3'
	MseI-CAC	5'-gat gag tcc tga gta aca c- 3'
	MseI-CAT	5'-gat gag tcc tga gta aca t- 3'
	MseI-CTG	5'-gat gag tcc tga gta act g- 3'
	MseI-CTA	5'-gat gag tcc tga gta act a- 3'
	MseI-CTC	5'-gat gag tcc tga gta act c- 3'

Table 3 Sequences of primers used for amplification of SNP loci

Primer	Sequence	Locus	Reference
Cich2 (C2)	F: 5'-tta tgc tga ggt gtt tgg cct ac- 3'	TMO-4C4	Muenzel & Salzburger unpublished
	R: 5'-cca cag cac cct cct cat aaa t- 3'		
Cich5 (C5)	F: 5'-aca cta tca ctc ggg gct ttc- 3'	Rag1exon3	Muenzel & Salzburger unpublished
	R: 5'-tcc tgg aag atc ttg tag aat tca- 3'		
Cich6 (C6)	F: 5' -aag ggt tta tgt tca atc aa- 3'	Rag1intron2	Muenzel & Salzburger unpublished
	R: 5' -agg gct gga ata tct ggc gg- 3'		
Cich38a (38a)	F: 5' -agc agg gtt gac ctt ctc aa- 3'	Phpt1	Muenzel & Salzburger unpublished
	R: 5' -tgg cta aaa tcc ccg atg ta- 3'		
Ednrb1	F: 5' -aar gay tga tgr ctk ttc ag- 3'	Ednrb1	Lang et al. 2006
	R: 5' -gak gcc atg ttg ats cca at- 3'		

F = Forward; R = Reverse

SNP analysis

Five loci were analyzed for SNPs (Single Nucleotide Polymorphisms) in addition to mtDNA sequence and AFLP analysis (Tab. 3). We only present the results from the analyses of loci TMO-4C4, Rag1exon3, and Rag1intron2 because we found no polymorphism in the Phpt1 locus and sequences of the Ednrb1 locus were mostly of low quality and thus unambiguous identification of polymorphisms was not possible. We justify the usage of only three SNPs because the purpose of this approach was primarily to reassure the results from the AFLP analysis concerning hybrid status of the Yellow populations with another nuclear marker. Therefore three SNPs were sufficient. As far as possible the same samples were used as for mtDNA and AFLP analysis. However, sequencing did not work in equal measure for all loci

in all individuals and so some samples totally failed in the SNP analysis, including all samples from Ndole. This yielded two total data sets consisting of 277 (including missing data) and 265 individuals (excluding samples with missing data), respectively, from 11 populations. The use of two data sets was necessary as some of the analysis programs we used allowed for missing data (ARLEQUIN and INTROGRESS) and one did not (LEADMIX).

Amplification and sequencing

The protocol for sequencing of loci comprising a SNP was the same as described for the mtDNA control region (for details see Chapter 3). Sequences of the primers used for amplification and sequencing are shown in Table 3. Sequencing was conducted only in one direction, that is, forward for locus TMO-4C4 and reverse for loci Rag1exon3 and Rag1intron2.

Population genetic analyses

For each population and each locus allele frequencies were calculated manually. The average gene diversity over all loci for each population and expected and observed heterozygosities for each locus and each population were computed with ARLEQUIN v.3.1 (Excoffier et al. 2005). The same program was used to test for Hardy-Weinberg Equilibrium and deviations from Linkage Equilibrium. *P*-values were corrected for multiple testing using the method of Benjamini and Hochberg (1995). Tests for hybrid status of the Yellow populations north of Lufubu river estuary were carried out using the three SNP loci with two programs. First we used the R-script INTROGRESS v. 1.22 (R Development Core Team 2009; Gompert & Buerkle 2009, 2010). This individual based method estimates marker locus-specific ancestries, genome-wide admixture and, as a function of it, predicts the probability of a given genotype at a given locus using multinomial regression. Thus deviations at specific loci from neutral expectations (based on the genome-wide admixture) and with it signs of selection can be detected. As detection of selection was not the focus of our analysis and we were also not primarily interested in marker ancestries, we only present the results from the genome-wide admixture function of the program, which calculates hybrid indices for each individual of the putative hybrid populations. The hybrid index displays the proportion of alleles inherited from each of the parental populations (Buerkle 2005), taking into account uncertainty in inheritance, when marker differences are not fixed between parental populations (Gompert &

Buerkle 2009). It approximately corresponds to the Bayesian admixture proportion 'Q' in STRUCTURE (Pritchard et al. 2000) with the difference that for the calculation of hybrid indices the two parental populations must be defined a priori. Thus parental and putative hybrid populations should be identified before using INTROGRESS, with population genetic analyses like Bayesian assignment in STRUCTURE (Pritchard et al. 2000) and/or with phenotypic and distributional data (Gompert & Buerkle 2009). As parental populations we used a pool of the Red populations from Moliro to Kachese as parental population one, and a pool of the Blue populations from Chaitika to Funda as parental population two. The populations between Chisanze and Linangu were considered the hybrid population. These decisions were based on the result from the STRUCTURE analysis carried out with the AFLP markers. Overall admixture was assessed in the form of a maximum likelihood (ML) hybrid index. Secondly we used the program LEADMIX v.1.0, a Fortran program, written by Jinliang Wang. It implements the maximum-likelihood method described in Wang (2003) to estimate admixture proportions of the hybrid population represented through p_1, the proportion of all genes in the hybrid gene pool that derived from parental population P_1. Accordingly, the proportion derived from parental population P_2 can be calculated as $1-p_1$. The maximum-likelihood (ML) method was developed on the basis of the admixture model proposed by Bertorelle & Excoffier (1998). This model assumes that generations ago an ancestral population (P_0) split into the two parental populations (P_1 and P_2) , which evolved in isolation from each other for some time. Eventually the two populations got into secondary contact and created a hybrid population (P_h) exhibiting a combination of genes at random proportions from both of the parental populations. At the time where the sample is taken from the parental populations and the hybrid population, the three populations are assumed to already have evolved independently from each other for some generations. We used the same a priori defined parental and hybrid populations as in the INTROGRESS analysis to calculate the admixture proportion p_1 of the hybrid population with the following settings: We assumed that all contributing parental populations were sampled and conducted the full admixture model, with all ancestral populations and their differentiation considered. We set the minimum drift to 0.00001, the number of initial points in the maximum-likelihood search to 10, and the number of points in integration for parental and hybrid populations to 1000, all as suggested by the author. LEADMIX also calculates different drift parameters, but as we were only interested in the admixture proportions, those were not further considered.

RESULTS

mtDNA analysis

DNA sequence diversity

A total number of 191 mitochondrial haplotypes was found in 304 individuals. Genetic variation was generally high in all investigated *T. moorii* populations. However, the average number of haplotypes was rather low due to the small sample sizes in the populations from Ndole, Chilanga, Ilangi, and Lufubu, thus leading to a mean value of only 16.5 haplotypes per population (range: 3 – 24). Haplotype diversity was high in all populations with a mean value of 0.937 (range: 0.945 - 1; Tab. 4; Fig. 2), a mean nucleotide diversity of 0.011 (range: 0.00767 – 0.01869; Tab. 4; Fig. 2), and a mean average number of pair-wise nucleotide differences of 8.68 (Tab. 4).

Table 4 Population information and mtDNA sequence diversity in the 13 investigated *T. moorii* populations. All indices were calculated with DnaSP v.5.0. Shown are sample size (*n*), number of haplotypes (*h*), haplotype diversity (H_E), nucleotide diversity (π), and average number of pair-wise nucleotide differences (*k*)

Population	Phenotype classification	*n*	*h*	H_E	π	*k*
Moliro	Red	30	18	0.945	0.01469	11.607
Katete	Red	32	24	0.978	0.01869	14.762
Chimba	Red	33	20	0.958	0.00919	7.248
Ndole	Red	6	6	1	0.00828	6.533
Kachese	Red	24	21	0.989	0.00920	7.580
Chisanze	Orange/Yellow	34	20	0.955	0.00892	6.554
Chilanga	Orange/Yellow	12	10	0.970	0.01175	9.273
Ilangi	Yellow	9	8	0.972	0.01406	11.111
Linangu	Yellow	23	18	0.980	0.01250	9.850
Lufubu	Blue	3	3	1	0.01094	8.667
Chaitika	Blue	32	24	0.970	0.00761	6.012
Nakaku	Blue	32	20	0.962	0.00960	7.597
Funda	Blue	34	23	0.970	0.00767	6.066
Total		304	191	0.995	0.02814	22.0374
Mean (s.d.)		23.4 (11.7)	16.5 (7.21)	0.973 (0.017)	0.011 (0.003)	8.68 (2.59)

Haplotype diversity was very similar in all populations (coefficient of variation [CV] = 1.7%). Numbers of haplotypes differed substantially between populations (CV = 43.6%) but this was again due to the four populations with small sample size. Populations also differed in nucleotide diversity (CV = 29.6%). This can primarily be attributed to the high values in the

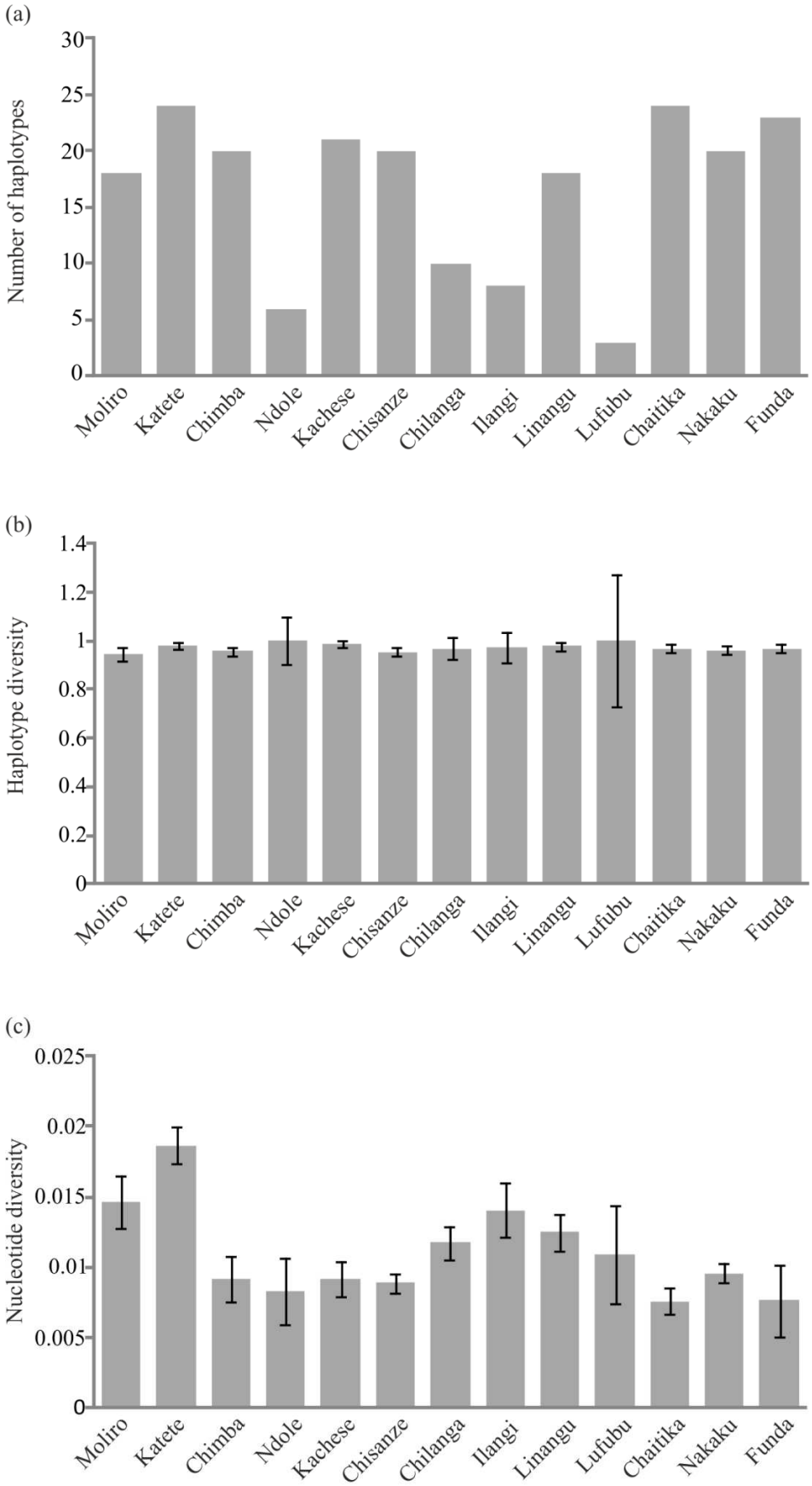


Figure 2 Mitochondrial sequence diversity represented through the number of haplotypes (a), haplotype diversity (b), and nucleotide diversity (c) in each of the investigated *T. moorii* populations. Error bars represent standard deviations

Moliro (π = 0.01469) and the Katete (π = 0.01869) population, but also to the relatively high nucleotide diversity in populations from Chilanga to Linangu (π = 0.01175 – 0.01406; Tab. 3; Fig.2). High degrees of genetic diversity in the Moliro and Katete populations can be attributed to the high stability of the populations: The habitat of these two populations is characterized by a very steep shore where lake level fluctuations had likely little impact on the communities. This is because populations were only displaced vertically and remained isolated in the same habitat as before without secondary contact to other populations. Thus they were able to evolve independently for a long time and accumulate mutations which results in a high level of nucleotide diversity.

Mitochondrial phylogenetic relationships

The p-distance based neighbor joining tree constructed with MEGA software v.5 (Tamura et al. 2011) revealed clustering of samples into two major mtDNA lineages with high bootstrap support for the main nodes (Fig. 3): Blue populations south of Lufubu River estuary formed one major cluster. One single individual from the Funda population was separated from the main cluster because this individual possessed an introgressed haplotype from the Yellow-Blotch lineage inhabiting the south-east of Lake Tanganyika. This color morph is actually separated from the Blue lineage by a large sandy bay but evidence for introgression has been found in the Blue populations immediately west of the bay (see Chapter 3). The second major cluster, which was further divided into several sub-clades, was formed by populations north of Lufubu River Estuary, comprising not only the Red populations from Moliro to Kachese but also the yellow to orange colored putative hybrid populations (the Yellow-morph) from Chisanze to Linangu (Fig. 3). Thus mtDNA revealed no evidence for hybrid status of the Yellow morph between the Red and the Blue morph. Furthermore, with the placement of samples into a Blue and a Red/Yellow main clade the mtDNA tree only roughly reflected the geographical distribution of the populations, as grouping of individuals within the main two clades was mostly not consistent with geographic distribution. In other words individuals belonging to the same population were not necessarily placed next to each other, neither were geographically adjacent populations.

Population structure

Analyses of pair-wise population differentiation conducted with ARLEQUIN v.3.1 (Excoffier et al. 2005) revealed significant population structure between most of the investigated populations. F_{ST} values, representing differentiation based on haplotype frequencies, ranged from -0.024 to 0.049 (mean F_{ST} = 0.023; Tab. 6). Most pair-wise analyses revealed significant population structure except comparisons including the Ndole and Lufubu population where none of the pair-wise analyses yielded significant differentiation ($p > 0.05$ in all comparisons). This was attributed to the very low sample size of only six and three individuals, respectively, in these two populations. The same could be observed for comparisons including the other two low sample size populations from Chilanga and Ilangi although in this case at least some comparisons revealed significant population structure ($p < 0.05$). Differentiation was also only slightly significant between Chimba and Kachese, Kachese and Linangu, Chisanze and Linangu, and between Chaitika and Nakaku ($p < 0.05$; Tab. 6). However, concerning the first three comparisons, this was opposed by highly significant Φ_{ST} values ($p < 0.01$ or $p < 0.001$). Between Nakaku and Chaitika Φ_{ST} was still only slightly significant ($p < 0.05$), which can be attributed to the fact that these two populations occur geographically adjacent along a continuous shoreline and thus differentiation between them is not as pronounced. Overall, Φ_{ST} values, representing differentiation based on uncorrected genetic distances between haplotypes ranged from -0.029 to 0.812 (mean Φ_{ST} = 0.466; Tab. 6). Almost all pair-wise analyses revealed significant differentiation again with the exception of some comparisons including the low sample size populations from Ndole and Lufubu, although several significant values were found even in these comparisons. Apart from that only population structure between Ilangi and Chilanga and between Linangu and Ilangi was found non-significant or only slightly significant ($p < 0.05$). This can again be attributed to the low sample size in the Chilanga and Ilangi population and/or to the geographical proximity of these populations. All other pair-wise analyses yielded highly significant population structure ($p < 0.01$ or $p < 0.001$; Tab. 6). Thus, overall, mitochondrial genetic structure was very high in the sampled populations concordant with previous studies investigating population structure in *T. moorii* (Sefc et al. 2007; Koblmüller et al. 2011).

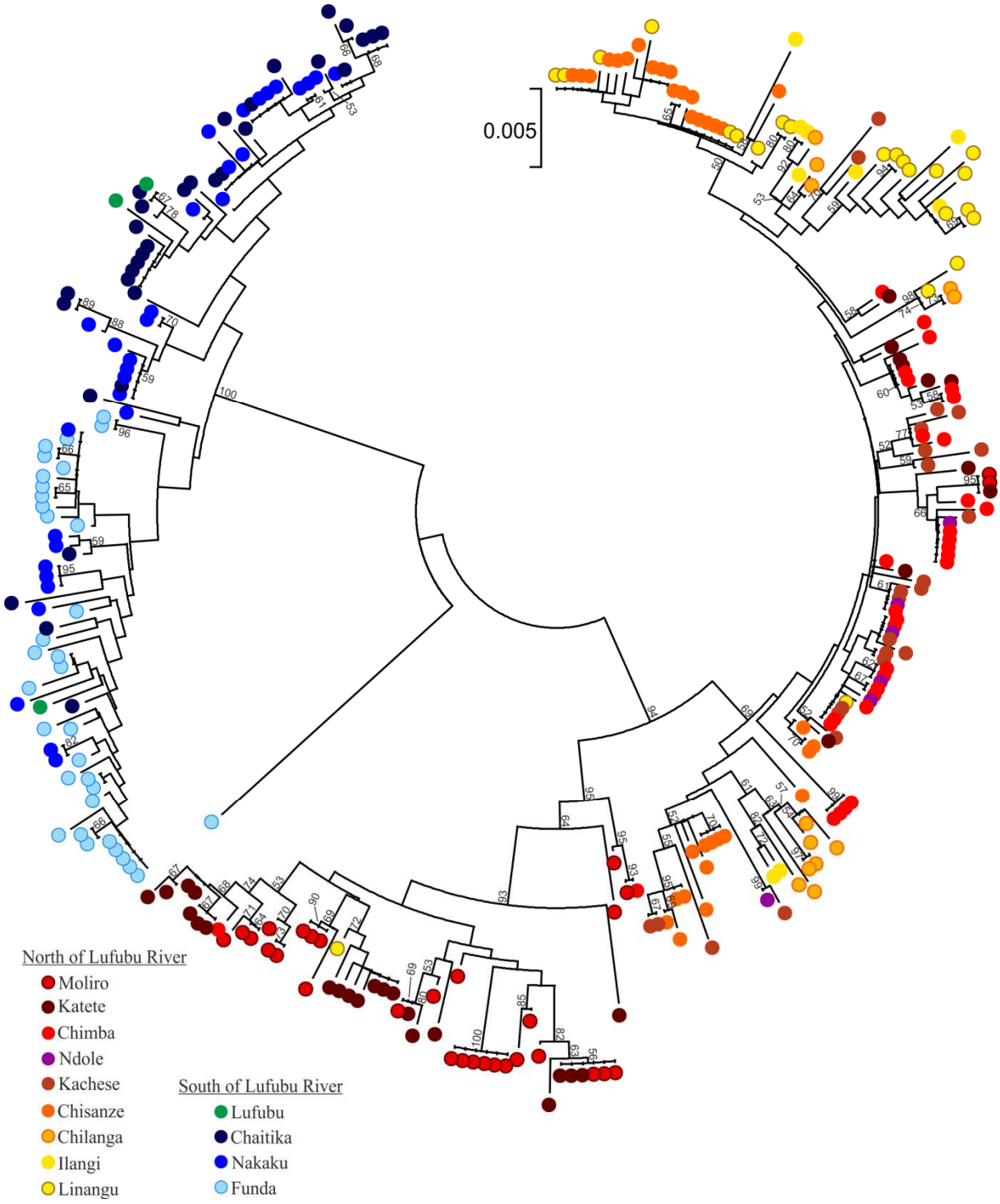


Figure 3 NJ tree of the 304 investigated *T. moorii* samples. The tree was calculated with p-distances based on 802 bp of mitochondrial control region with MEGA v.5 to illustrate phylogenetic relationships among the sample individuals. Cold colors represent individuals that were collected south of Lufubu River estuary and warm colors represent individuals collected north of the estuary. Individuals with putative hybrid status are represented through yellow to orange colors. Bootstrap values > 50 are shown at the respective branches.

AFLP analysis

Genetic diversity

For AFLP analysis only 12 populations were used due to insufficient DNA quality of the three Lufubu samples. The 18 primer combinations yielded a total of 1160 AFLP loci of which 1045 (90.1%) were segregating within the overall data set. The mean number of AFLP bands per individual was 283.1. The mean proportion of polymorphic markers per population was 40.19%, ranging from 36.8% to 43.4%. The average expected heterozygosity (i.e. Nei's gene diversity) was 0.134 ranging from 0.123 to 0.164 (Tab. 5).

Table 5 Genetic diversity measures of the twelve investigated *T. moorii* populations calculated with AFLP-SURV v.1.0. Shown are sample size (*n*), percentage of polymorphic loci at the 5% level (*PLP*), and expected heterozygosity under Hardy-Weinberg genotypic proportions (i.e. Nei's gene diversity; H_j) with its standard error (H_j *S.E.*). Values are given for each population and for the whole data set

Population	n	PLP [%]	H_j	H_j S.E.
Moliro	27	40.8	0.130	0.0048
Katete	30	43.4	0.134	0.0047
Chimba	31	41.6	0.132	0.0049
Ndole	5	40.6	0.164	0.0055
Kachese	19	42.5	0.134	0.0049
Chisanze	31	40.3	0.127	0.0050
Chilanga	12	36.8	0.138	0.0052
Ilangi	8	42.2	0.147	0.0053
Linangu	24	38.1	0.136	0.0051
Chaitika	24	37.2	0.125	0.0049
Nakaku	30	40.3	0.123	0.0048
Funda	28	38.5	0.123	0.0048
Total	269	90.1	0.156	
Mean (s.d.)	22.4 (9.3)	40.2 (2.1)	0.134 (0.01)	0.0050

Phylogenetic relationships

The AFLP neighbor joining tree, calculated in MEGA software v.5 (Tamura et al. 2011), based on restriction site distances (Nei & Li 1979), largely reflected the geographic distribution of the populations as individuals from the same population overall clustered together and geographically adjacent populations were placed next to each other in the tree (Fig. 4). It revealed two main clades: A Red clade uniting all samples from Moliro to Kachese, and a Blue/Yellow clade uniting all populations from Chisanze to Funda. The Red clade was further divided into two sub-clades of which one included all individuals from Moliro and Katete, and the other all Chimba, Ndole and Kachese samples. The Blue/Yellow

clade comprised a separate cluster for the Chisanze population, placed next to the Red morph, and a Blue/Yellow cluster with a bootstrap support of 78%. Within this cluster Blue populations were again revealed in a separate clade with 59% bootstrap support (Fig. 4). Thus, in contrast to the results obtained from mtDNA data, the Yellow populations north of Lufubu River estuary were not resolved within the Red lineage in the AFLP neighbor joining tree, but rather took an intermediate position between the Blue and the Red morph. The placement of the Yellow morph populations in the tree displayed a genetic gradient from Red to Blue that was highly congruent with the geographic distribution of the populations.

Population structure

The observed F_{ST} value over all populations was 0.1394 ($p < 0.001$). Pair-wise F_{ST} values were generally high and significant between all populations (Tab. 7). Highest differentiation levels were detected between the Blue populations south of Lufubu River estuary (i.e. Chaitika, Nakaku, Funda) and the Red populations north of the estuary (i.e. Moliro, Katete, Chimba, Ndole, Kachese) with an average value of 0.227 (range: 0.169 – 0.277; Tab. 7). Values for comparisons including Ndole were somewhat lower, most likely due to low sample size of the Ndole population. However, genetic differentiation was highly significant ($p < 0.001$) in all comparisons indicating a high level of population structure between our investigated *T. moorii* populations on the nuclear DNA level using AFLP as genetic markers.

Test for hybridization

AFLP based Bayesian inference of population structure conducted with STRUCTURE v.2.3.3 (Pritchard et al. 2000; Falush et al. 2007) and subsequent analysis of the results with STRUCTURE HARVESTER v.0.6.6 (Earl 2009) revealed the highest peak in ΔK for $K = 2$ (Evanno et al. 2005; Fig. 5 a). Thus the most likely number of genetic clusters for the whole data set was two, corresponding, overall, to a division between the Red and the Blue morph. Populations north of Lufubu River Estuary from Moliro to Kachese were purely assigned to the Red cluster and populations south of the Estuary from Chaitika to Funda were purely assigned to the Blue cluster. The orange to yellow populations from Chisanze to Linangu (the Yellow morph), which are geographically located in between, represented a genetic cline

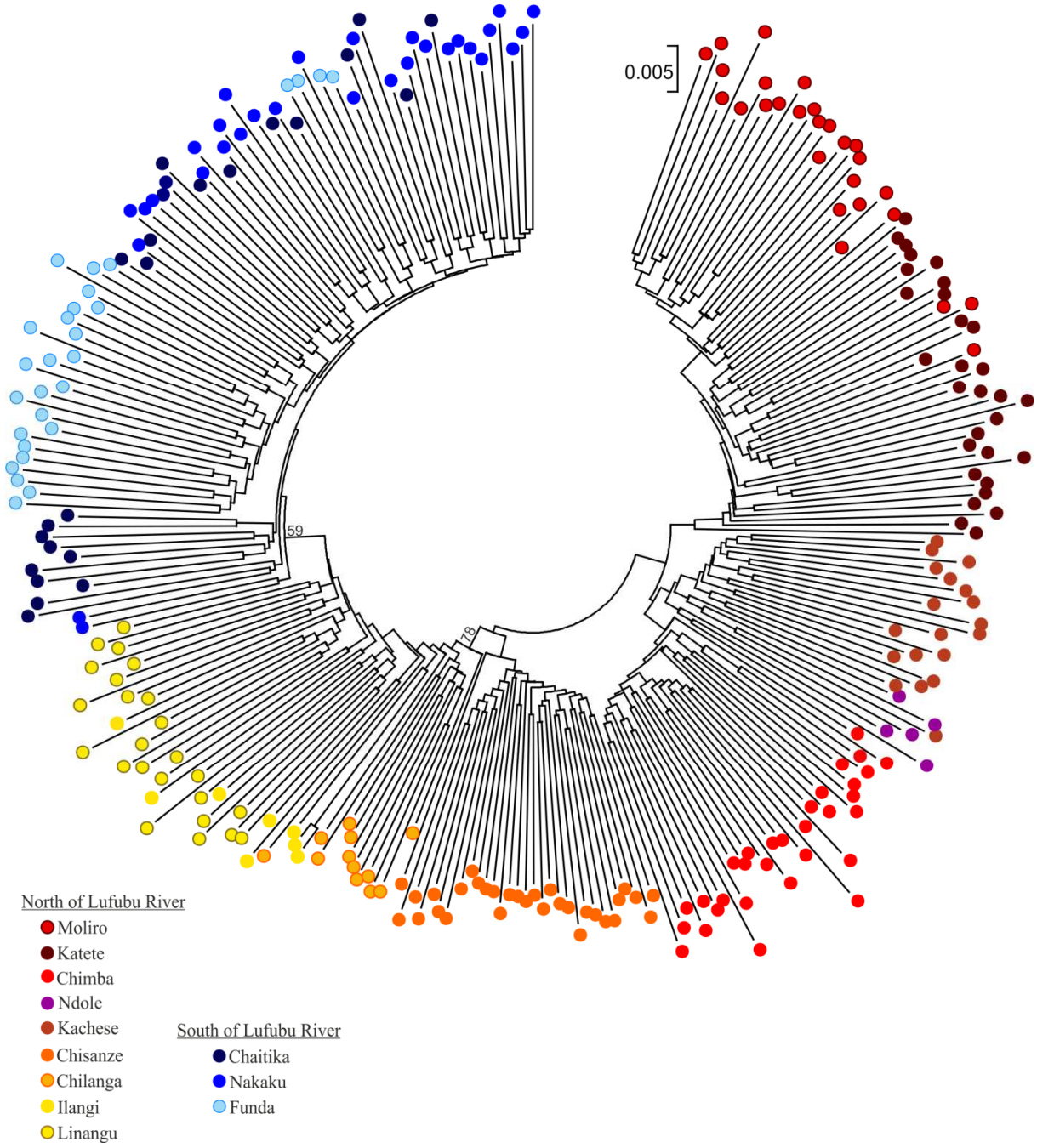


Figure 4 NJ tree of the 304 investigated *T. moorii* samples. The tree was calculated based on AFLP restriction site distances (Nei & Li 1978) with MEGA v.5 to illustrate phylogenetic relationships among the sample individuals. Cold colors represent individuals that were collected south of Lufubu River estuary and warm colors represent individuals collected north of the estuary. Bootstrap values > 50 are shown at the respective branches.

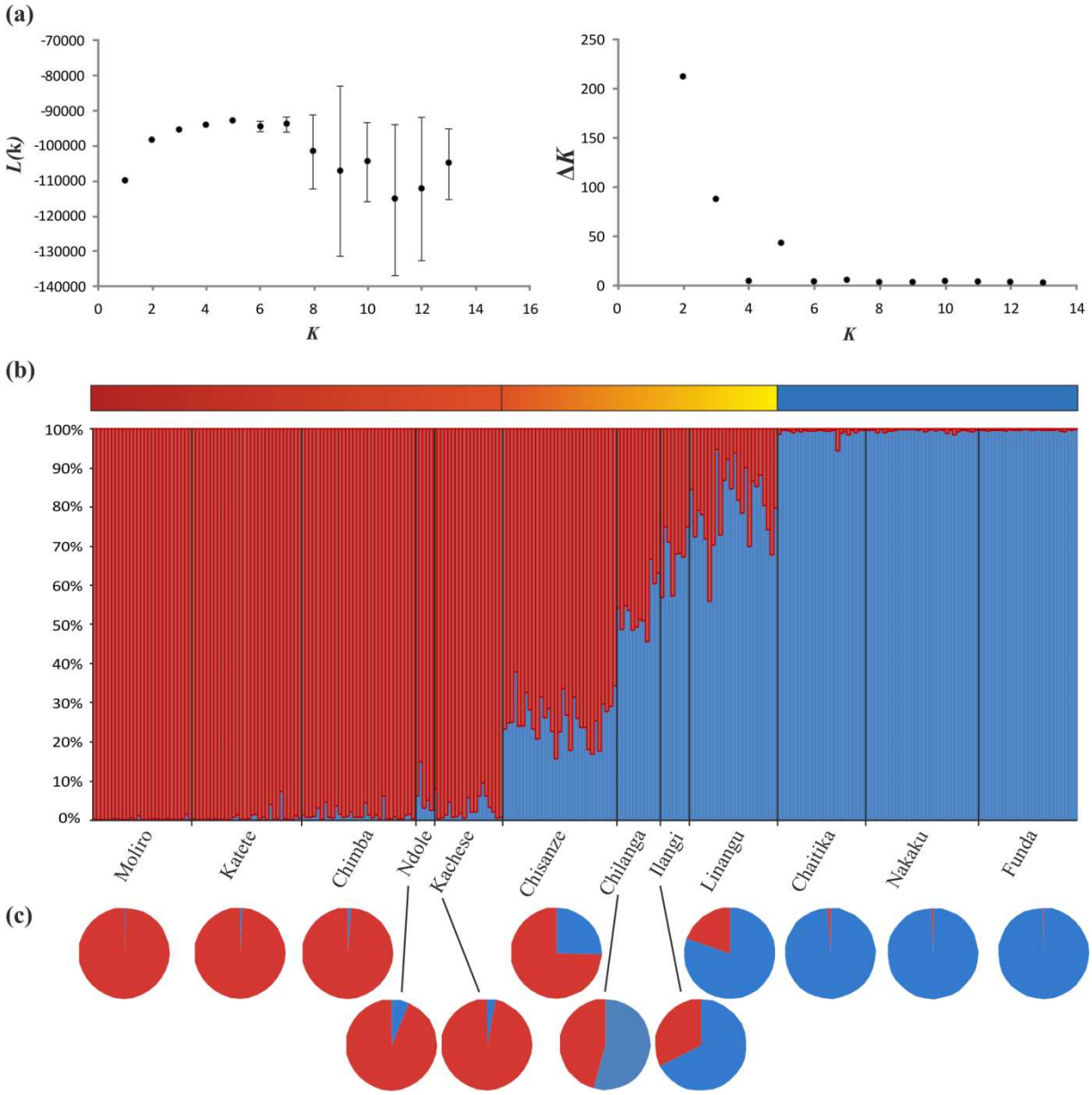


Figure 5 Results of the Bayesian clustering analysis of the 12 investigated *T. moorii* populations using 1160 AFLP loci. Analyses were conducted with STRUCTURE v.2.3.3 and STRUCTURE HARVESTER v.0.6.6. (a) Left: the mean likelihood [L(K) ± S.D.] over five replicate runs assuming *K* clusters. Right: Δ*K*; the number of clusters with the highest Δ*K* peak is interpreted as the highest level of structuring. (b) Assignment of each individual to one of the two most probable clusters (as inferred from the Δ*K* statistic). Populations are ordered according to their geographical appearance from north to south. The color bar represents the body coloration classification of the individuals of each population (Red – Orange – Yellow – Blue). (c) Pie charts show the percentage of assignment to either of the two inferred clusters averaged over all individuals within each population.

Table 6 Pair-wise mitochondrial population differentiation between the 13 investigated *T. moorii* populations calculated with ARLEQUIN v.3.1. F_{ST} values are below diagonal, Φ_{ST} values above. Benjamini-Hochberg corrected significance levels: *** p < 0.001; ** p < 0.01; * p < 0.05

	Moliro	Katete	Chimba	Ndole	Kachese	Chisanze	Chilanga	Ilangi	Linangu	Lufubu	Chaitika	Nakaku	Funda
Moliro		0.055**	0.417***	0.317***	0.409***	0.439***	0.335***	0.307***	0.343***	0.519**	0.651***	0.633***	0.663***
Katete	0.025**		0.373***	0.326**	0.386***	0.416***	0.334***	0.310***	0.312***	0.645***	0.721***	0.699***	0.731***
Chimba	0.041***	0.022***		-0.029	0.041**	0.235***	0.257***	0.211***	0.219***	0.775***	0.789***	0.764***	0.796***
Ndole	0.031	0.010	-0.024		-0.028	0.213**	0.176*	0.114*	0.144*	0.769*	0.803***	0.765***	0.810***
Kachese	0.032**	0.015**	0.018*	-0.001		0.188***	0.204***	0.134***	0.183***	0.768***	0.790***	0.762***	0.797***
Chisanze	0.049***	0.031***	0.037***	0.026	0.028**		0.227***	0.204***	0.185***	0.797***	0.805***	0.782***	0.812***
Chilanga	0.035*	0.017	0.023*	0.008	0.013	0.031*		0.111*	0.219***	0.707**	0.772***	0.740***	0.782***
Ilangi	0.042	0.023	0.029	0.015	0.019	0.037*	0.021		0.036	0.701**	0.782***	0.749***	0.792***
Linangu	0.033**	0.015**	0.021**	0.007	0.011*	0.021*	0.013	0.014		0.711***	0.764***	0.738***	0.778***
Lufubu	0.035	0.012	0.020	0	0.007	0.030	0.01	0.017	0.008		0.066	0.067	0.269
Chaitika	0.037***	0.020***	0.026***	0.013	0.017**	0.033***	0.019	0.025	0.017**	0.004		0.047*	0.389***
Nakaku	0.046***	0.028***	0.034***	0.022	0.025***	0.041***	0.027*	0.034	0.025**	0.025	0.018*		0.293***
Funda	0.042***	0.024***	0.030***	0.017	0.021**	0.037***	0.023*	0.029*	0.021**	0.02	0.026***	0.033***	

Table 7 Pair-wise population differentiation (F_{ST}) between the 12 investigated *T. moorii* populations based of AFLP data calculated with AFLP-SURV v.1.0. Calculation of *P*-values was based on 5000 permutations. Significance levels: *** p < 0.001

	Moliro	Katete	Chimba	Ndole	Kachese	Chisanze	Chilanga	Ilangi	Linangu	Chaitika	Nakaku	Funda
Moliro												
Katete	0.030***											
Chimba	0.098***	0.056***										
Ndole	0.118***	0.083***	0.044***									
Kachese	0.141***	0.098***	0.044***	0.032***								
Chisanze	0.181***	0.145***	0.102***	0.076***	0.063***							
Chilanga	0.176***	0.144***	0.105***	0.079***	0.095***	0.074***						
Ilangi	0.188***	0.161***	0.126***	0.083***	0.118***	0.103***	0.032***					
Linangu	0.212***	0.187***	0.164***	0.125***	0.160***	0.144***	0.063***	0.033***				
Chaitika	0.262***	0.241***	0.216***	0.169***	0.211***	0.186***	0.127***	0.101***	0.074***			
Nakaku	0.274***	0.247***	0.225***	0.180***	0.223***	0.198***	0.128***	0.109***	0.086***	0.008***		
Funda	0.277***	0.255***	0.228***	0.176***	0.227***	0.196***	0.129***	0.110***	0.089***	0.032***	0.032***	

between the Red and the Blue morph which was congruent with the populations' geographic location, i.e. samples that are located nearer to the Red morph showed a higher degree of genetic assignment to these populations and vice versa samples that are located nearer to the Blue morph exhibited more genetic similarities to these (Fig 5 b and c). These results are congruent with the resolution of the populations in the AFLP neighbor joining tree and reveal the hybrid status of the Yellow morph based on AFLPs as nuclear markers.

SNP analysis

Genetic diversity

The two total data sets consisted of 11 populations with 265 individuals without missing data and 277 individuals including missing data, respectively. In each of the three loci (TMO-4C4, Rag1exon3, Rag1intron2) one SNP was found. In the TMO-4C4 locus two alleles were existent exhibiting either G or T, with the Red populations from Moliro to Chimba being fixed for the G allele. The Kachese and Chisanze population already possessed a small frequency of the T allele (0.03; Tab. 8) and in populations from Chilanga to Funda the frequency of the T allele was variable ranging from 0.21 to 0.5. There was a general decrease of the G allele frequency from Red to Blue populations observable, although it was again relatively high in the Funda population (0.74; Tab. 8). However this SNP segregated the Red populations and the Chisanze population from the remaining data set. Moreover it clearly showed the influence of the Blue gene pool into the hybrid gene pool. The second SNP at Rag1exon3 was characterized by exhibiting an A/G polymorphism. Here, the Moliro and Katete population exhibited a very high frequency of the A allele (0.95 and 0.98, respectively; Tab. 8). It was still relatively high in the Chimba and Kachese population (0.81 and 0.68 respectively; Tab. 8) and variable in the hybrid and Blue populations (range: 0.32 to 0.88; Tab. 8). There was again a slight pattern of gradual decrease of the frequency of the A allele from Red to Blue populations observable although some populations presented outliers. In the Ilangi population e.g., the frequency of the A allele was again very high (0.88; Tab.8), which could be due to the small sample size in this population. Overall, however, the Rag1exon3 SNP separated the Red Moliro and Katete population from the remaining data set and again showed the influence of the Blue populations into the hybrid gene pool. Finally, the two alleles present at the Rag1intron2 locus exhibited C or T, respectively. This SNP locus was polymorphic in every population and was not segregating between sets of populations.

Frequencies of the C allele ranged from 0.44 to 0.72 (Tab. 8) but without any gradual pattern. Genetic diversity averaged over the three loci, calculated with ARLEQUIN v.3.1 (Excoffier et al. 2005) ranged from 0.16 to 0.49 (Tab. 9) and was overall lower in the Red populations than in the hybrid and Blue populations. A significant deviation from Hardy-Weinberg Equilibrium was only found in one case, i.e. for locus Rag1exon3 in the Linangu population (Tab. 9). However, significance vanished after correction for multiple testing, thus all loci in all populations were in Hardy-Weinberg Equilibrium. Tests for Linkage Disequilibrium (LD) revealed significant deviations from Linkage Equilibrium in ten out of eleven populations for loci Rag1exon3 and Rag1intron2, even after correction for multiple testing. Thus these two loci can be considered physically linked which is not a surprise as they are located adjacent to each other on the chromosome. Significant LD was also detected between loci TMO-4C4 and Rag1intron2 in the Funda population. However, as this deviation occurred only in one population we don't consider it as indicative of physical linkage. There were no signs of LD between loci TMO-4C4 and Rag1exon3. We used all three SNP loci in subsequent admixture analyses although we are aware that information yielded from the two linked markers is not independent from one another.

Test for hybridization

Assingment analyses using three SNP loci conducted with LEADMIX (Wang 2003) and INTROGRESS (Gompert & Buerkle 2009, 2010) to infer admixture proportions of the hybrid populations, revealed influence of both parental gene pools into the hybrid gene pool but with a higher contribution of the Blue parental population. Considering LEADMIX this was given by a p_1 value of 11,9%. P_1 corresponds to the proportion of all genes in the hybrid gene pool that derived from parental population one (the Red morph). Thus the proportion of genes derived from parental population two, the Blue morph, was 88.1% ($1 - p_1$). Similarly INTROGRESS showed a high influence of the Blue parental population into the hybrid gene pool: Over all three loci (overall admixture) the program yielded individual hybrid indices for population two ancestry (i.e. fraction of the genome inherited from the Blue parental population) that ranged from 0 to 1, with 0 meaning no assignment to parental population two (thus total assignment to population one, the Red morph), and 1 corresponding to total assignment to parental population two. Overall a slight trend towards a genetic cline that corresponds to the geographic distribution of the populations was observable: Hybrid indices were generally somewhat higher in the populations geographically nearer to the Blue morph than in the population nearest to the Red morph (see trend line in Fig. 6a) although each

population comprised individuals with total Blue ancestry as well as individuals with total Red ancestry. Moreover mean hybrid indices were only marginally different, ranging from 0.612 to 0.721 (Fig. 6b). However, hybrid indices were high in every population, suggesting a strong impact of the Blue morph into the Yellow populations. Thus SNP analysis largely confirmed the results obtained from AFLP data and thus also stay in contrast to the mtDNA pattern.

Table 8 Allele frequencies for the three SNP loci in the 11 investigated *T. moorii* populations. Pie charts illustrate the percentage of each of the two alleles in each population

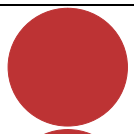
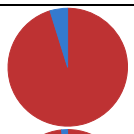
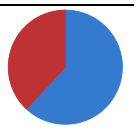
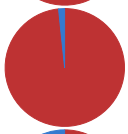
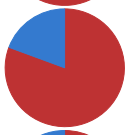
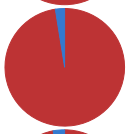
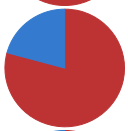
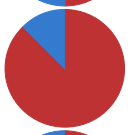
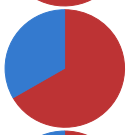
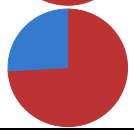
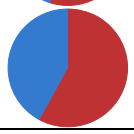
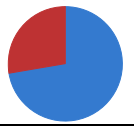
	Locus (Allele 1/ Allele 2)								
	TMO-4C4 (G/T)			**Rag1exon3 (A/G)**			**Rag1intron2 (C/T)**		
Population	G	T	%	A	G	%	C	T	%
Moliro	1.000	0.000		0.952	0.048		0.621	0.379	
Katete	1.000	0.000		0.983	0.017		0.633	0.367	
Chimba	1.000	0.000		0.806	0.194		0.452	0.548	
Kachese	0.975	0.025		0.676	0.324		0.441	0.559	
Chisanze	0.969	0.031		0.323	0.677		0.677	0.323	
Chilanga	0.792	0.208		0.500	0.500		0.667	0.333	
Ilangi	0.500	0.500		0.875	0.125		0.500	0.500	
Linangu	0.620	0.380		0.640	0.360		0.600	0.400	
Chaitika	0.583	0.417		0.521	0.479		0.625	0.375	
Nakaku	0.667	0.333		0.550	0.450		0.638	0.362	
Funda	0.742	0.258		0.581	0.419		0.724	0.276	

Table 9 Genetic diversity of three SNP loci in the 11 investigated *T. moorii* populations calculated with ARLEQUIN v.3.1. For each population the number of samples (*n*) and the average gene diversity over all loci (*Hj*) are given. For each locus and each population the number of alleles (N_A), the observed heterozygosity (H_O), the expected heterozygosity (H_E), and *P*-values for tests of Hardy-Weinberg Equilibrium (*HWE P-values*) are shown

			Locus (Allele 1/Allele 2)											
			TMO-4C4 (G/T)				Rag1exon3 (A/G)				Rag1intron2 (C/T)			
Population	*n*	H_j	N_A	H_O	H_E	*HWE P-values*	N_A	H_O	H_E	*HWE P-values*	N_A	H_O	H_E	*HWE P-values*
Moliro	31	0.17	1	N/A	N/A	N/A	2	0.097	0.094	1.000	2	0.483	0.479	1.000
Katete	31	0.16	1	N/A	N/A	N/A	2	0.033	0.033	1.000	2	0.467	0.472	1.000
Chimba	31	0.27	1	N/A	N/A	N/A	2	0.258	0.317	0.29	2	0.516	0.503	1.000
Kachese	20	0.25	2	0.050	0.050	1.000	2	0.529	0.452	0.45	2	0.529	0.508	1.000
Chisanze	32	0.30	2	0.063	0.062	1.000	2	0.516	0.444	0.44	2	0.387	0.444	0.44
Chilanga	12	0.44	2	0.417	0.344	1.000	2	0.667	0.522	0.563	2	0.333	0.464	0.518
Ilangi	9	0.38	2	0.556	0.529	1.000	2	0.250	0.233	1.000	2	0.500	0.533	1.000
Linangu	25	0.48	2	0.520	0.481	1.000	2	0.240	0.470	0.025	2	0.480	0.4900	1.000
Chaitika	24	0.49	2	0.500	0.497	1.000	2	0.375	0.510	0.238	2	0.417	0.479	0.667
Nakaku	30	0.46	2	0.400	0.452	0.682	2	0.567	0.503	0.711	2	0.448	0.470	1.000
Funda	32	0.39	2	0.387	0.389	1.000	2	0.387	0.495	1.000	2	0.414	0.407	1.000
Mean (s.d.)	25.2 (8.26)			0.361 (0.20)	0.350 (0.19)			0.356 (0.20)	0.370 (0.18)			0.452 (0.06)	0.477 (0.03)	

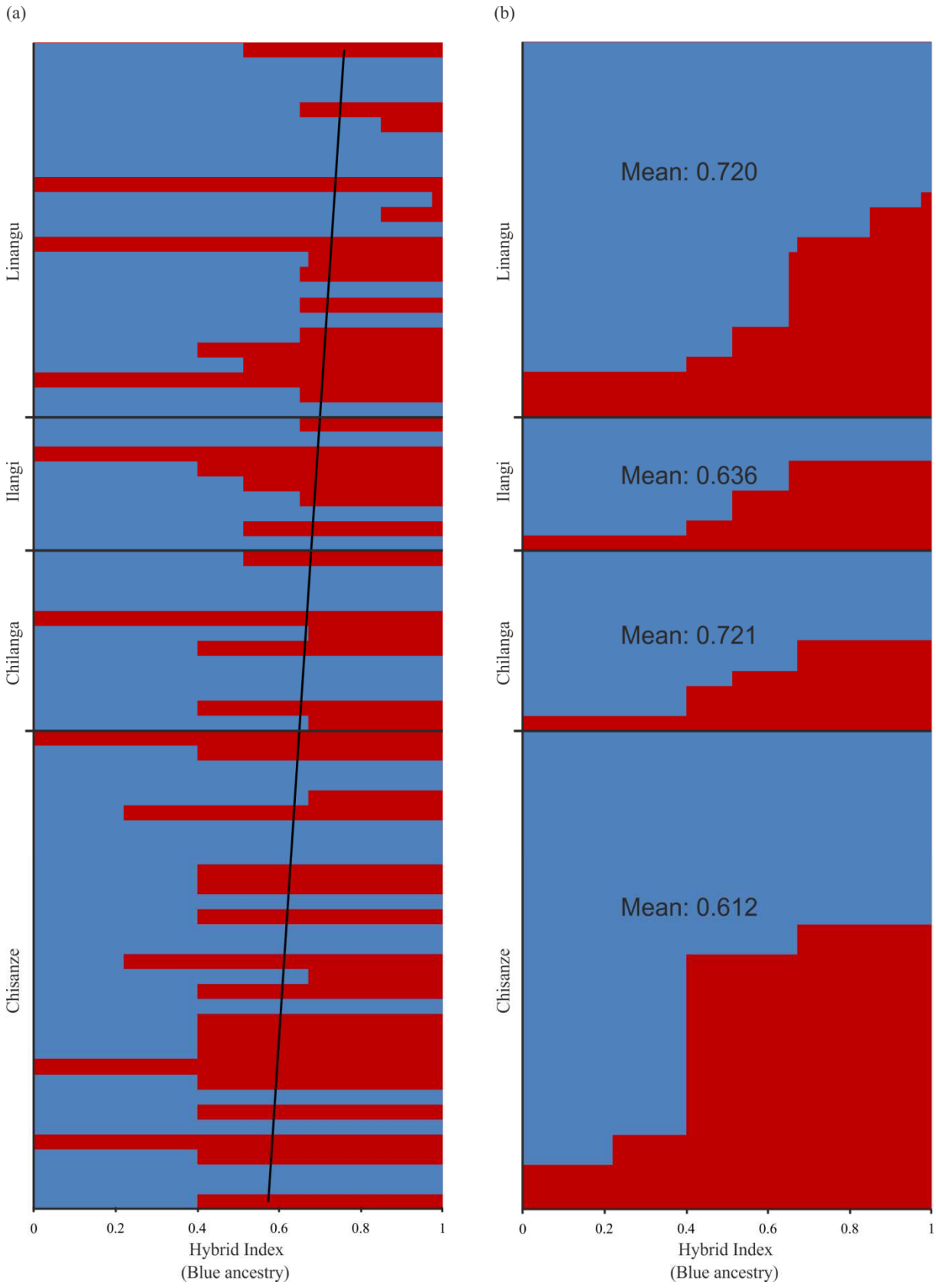


Figure 6 Admixture proportions in four hybrid populations of *T. moorii* over three SNP loci. Blue bars represent the hybrid index, in terms of. the fraction of the genome inherited by parental population 2 (Blue). (a) Admixture proportions are sorted by individuals and a trend line for Blue ancestry is given. (b) Individuals are sorted by hybrid index within each population. For each population the mean value of population 2 (Blue) ancestry over all individuals is given.

DISCUSSION

Evidence for hybridization in the Yellow morph populations

The present large-scale analysis on a set of *Tropheus moorii* populations from southern Lake Tanganyika, revealed evidence for hybridization in a contact zone of two phenotypically and genetically highly divergent lineages; on the one side there are the Blue populations which are resolved within the mitochondrial 8-G lineage (Baric et al. 2003; Sturmbauer et al. 2005; Egger et al. 2007) and within sub-clade 1a in a recent AFLP study (Egger et al. 2007), respectively. On the other side we find the Red populations, belonging to the mitochondrial 7-F lineage (Baric et al. 2003; Sturmbauer et al. 2005; Egger et al. 2007) and to AFLP sub-clade 4b (Egger et al. 2007), respectively. Phylogenetic status of the orange to yellow colored populations (the Yellow morph) in the region between the two color morphs was inconsistently resolved in previous studies when using mtDNA or AFLP data (Baric et al. 2003, Sturmbauer et al. 2005; Egger et al. 2007), and hybrid status of these populations has already been suggested based on a homoplasy excess test carried out on AFLP data (Egger et al. 2007). The aim of the present study was thus to shed light on the ambiguous Yellow morph status. With a large data set including many populations, high sample sizes, and multiple genetic markers we were able to infer strong evidence for introgression from both, the Blue and the Red morph, into the Yellow populations. This hybridization event must have happened in the course of secondary contact between the two morphs due to one of the recurrent cycles of lake level fluctuations that affected Lake Tanganyika in the past (e.g. Scholz et al. 2003; Cohen et al. 2007), and might have led to the generation of the Yellow morph populations north of Lufubu river estuary. However, we are aware that we cannot be sure if the Yellow populations are actually a ‘product’ of the Red and the Blue morph. Alternatively it might also be possible that the Yellow populations had already existed in the respective area and were introgressed by the Red and the Blue morph from both sides. Currently ongoing admixture pond experiments between the two morphs will help to shed light on that issue as we will be able to see the phenotype of the hybrids.

Inconsistencies between the mtDNA and the nuclear data set

In our study inconsistencies between the mtDNA data set and the nuclear data set were still present, concordant with previous studies. A possible explanation for the exclusive resolution

of the Yellow morph populations within the Red clade in previous mtDNA phylogenies could have been the low sample size used in these studies. Thus low levels of mitochondrial introgression between lineages could have been missed. However, also with the use of many populations with high sample sizes on the one hand, and of the whole mtDNA control region on the other hand we were not able to detect signs for introgression on the mtDNA level, as again all Yellow populations clustered exclusively within the Red morph clade. Thus we infer our evidence for genetic admixture merely from the nuclear data, where we found clear signals of hybridization: AFLP data revealed strong impact of the Blue gene pool into the Yellow gene pool with the Yellow populations representing a genetic cline between the Red and the Blue cluster, which is concordant with the geographic distribution of populations. These results are visualized through the position of the samples in the AFLP neighbor joining tree and through the STRUCTURE results. To find out if this hybridization signal was only an artifact of the AFLP data we additionally used SNP markers to infer hybridization. Again we found strong impact of the Blue gene pool into the Yellow morph populations, thus confirming our AFLP results.

Inconsistencies between the results from mitochondrial and nuclear data have been reported in several studies on cichlids (e.g. Schelly et al. 2006; Egger et al. 2007; Koblmüller et al. 2007; Koblmüller et al. 2010) and also in other organisms (e.g. Rush et al. 2009; Milá et al. 2010; Yannic et al. 2010; Turmelle et al. 2011). On the interspecific level such incongruities have mostly been attributed to recent speciation and resulting incomplete lineage sorting and/or hybridization (e.g. Machado & Hey 2003; Seehausen 2004; Egger et al. 2007; Nevado et al. 2009). On the intraspecific level explanations mostly involve asymmetric hybridization due to asymmetries in mate choice, sex-specific differences in dispersal or the variable influence of genetic drift. Asymmetric hybridization can explain the existence of mtDNA of only one of the parental species in the hybrid population (Wirtz 1999) and can be based on asymmetries in mate choice. This could have lead to inconsistencies between the mitochondrial and nuclear introgression pattern in the following way: If Blue females mated exclusively color assortatively they wouldn't bring any mitochondrial DNA into the hybrid population, whereas nuclear Blue DNA would be brought into the hybrid populations by Blue males. On the other side, Red females that would not differentiate between males of their own color morph or Blue males (or consequently Yellow hybrid males) would bring mitochondrial and nuclear DNA into the hybrid populations. This could explain the observed inconsistent pattern. However, results of mate choice (Egger et al. 2008) and subsequent breeding pond (Hermann et al. unpublished) experiments between representatives of the Blue and the Red

morph contradict this hypothesis: Overall both studies revealed highly color assortative preferences of the Red females whereas Blue females showed in fact a trend toward assortative mating but disassortative matings occurred as well. Thus the Blue females seem to be the ones that don't discriminate 100% between homo- and heteromorphic males, with regards to mating. Given this assumption Blue mtDNA should actually be found in the hybrid gene pool, instead of Red mtDNA. As the opposite was the case we need to search for alternative explanations for the observed introgression pattern that are not based on mate choice behavior. In a very recent study on the big brown bat *Epesicus fuscus* Turmelle et al. (2011) showed that strikingly contradicting results inferred by mtDNA data on the one hand and by nuclear DNA data on the other hand could be attributed to limited female dispersal combined with male-mediated gene flow. Sex-biased gene flow has been reported in numerous other studies including studies on cichlids (e.g. Knight et al. 1999; Taylor et al. 2003; Anseeuw et al. 2008; Carvajal-Vallejos et al. 2010) and male dispersal has generally been suggested to result in unidirectional hybridization (Wirtz 1999) and contrasting nuclear and mtDNA genetic patterns (Palumbi & Baker 1994). Concerning our study this would mean that from the Blue parental population only males disperse or dispersed northwards across Lufubu River estuary and hybridized there with the Red morph. On the north side of the estuary in contrast hybridization happened involving both sexes of the Red morph. Thus the hybrid pool contains both parental species' nuclear DNA but only Red mtDNA. However, our results on genetic structure contradict this hypothesis: If the scenario of male-biased dispersal was indeed true then we would overall find lower population structure in nuclear DNA than in mtDNA between populations. This was not the case as with both markers genetic structure was found to be congruently high, which indicates concordant dispersal behavior in both sexes. The third explanation could be that one of the two parental haplotype lineages disappeared due to genetic drift. A possible scenario could have been the following: The Yellow populations could have originally belonged to the Blue lineage, with high similarities to the Blue morph, in terms of both mitochondrial and nuclear DNA and probably also in terms of phenotype. This former close connection to the Blue morph is reflected through the results from the tests for hybridization using nuclear AFLP and SNP markers, where the Yellow morph populations largely showed a higher assignment to the Blue than to the Red parental species. Moreover in the AFLP phylogeny the Yellow populations even cluster together with the Blue morph. Then at some point in history the present Yellow populations became separated from their Blue relatives at a time when Lufubu River estuary became an un-crossable barrier, possibly due to a change in lake level. At this point in time the

populations still possessed Blue nuclear AND mtDNA. Then the Red morph populations from the north started to hybridize with the former Blue population, brought their nuclear and mtDNA into their gene pool, and after some time with increasing input from the Red gene pool the phenotype of the hybrid populations changed to yellow, as a result of the genetic admixture. Meanwhile, since there was no more input from the Blue populations that remained south of the estuary, the Blue mtDNA lineage vanished from the hybrid populations due to random genetic drift and was totally replaced by Red mtDNA in the course of a process called 'mitochondrial replacement' (e.g. Nevado et al. 2009). However, nuclear Blue DNA was largely retained. Random genetic drift has been discussed as the cause of the extinction of mitochondrial lineages in previous studies (e.g. Thulin & Tegelström 2002; Rush et al. 2009) and serves as a possible explanation for contrasting genetic patterns of nuclear and mitochondrial markers (Palumbi & Baker 1994). This is attributed to the fact that mtDNA is more sensible to genetic drift due to the reduced effective population size of mtDNA when compared with nuclear DNA. In our present study, random genetic drift leading to the observed incongruence between genetic patterns derived from nuclear and mitochondrial data and the resulting asymmetric introgression seems to be the most likely explanation. Although we are aware that this must not necessarily have been the true scenario as it is still merely an assumption. However, this study once more highlights the importance of using different types of genetic markers for phylogenetic reconstructions, as one marker type alone would have provided us with incomplete information only.

Implications for color diversity in Tropheus and conclusion

The results of the present study again show the potential of hybridization to create novel phenotypes. Here, in the second contact zone this is even more pronounced, as hybrids do not display an intermediate phenotype between the two parental morphs (as in the first contact zone; see Chapter 3) but instead a totally new color variant which is not expressed by either of the parental morphs. Hybridization as a promoter for the evolution of new taxa has been discussed in cichlids (Salzburger et al. 2002; Seehausen 2004; Schelly et al. 2006; Stelkens et al. 2009). In *Tropheus* several potential cases of admixture and hybridization between morphs have been reported and in some instances these events could have given rise to novel color phenotypes (Egger et al. 2007). This could have likely been the case in the contact zone examined in the present study, where the admixture of a Blue and a Red morph possibly led to the evolution of a Yellow morph. Thus, considering the whole *Tropheus* genus, hybridization may very likely serve as a considerable contribution to the outstanding phenotypic diversity

present within this taxon. Finally, as already discussed in Chapter 3, this study again demonstrates the difficulties of interpreting hybridization scenarios. Moreover it once more highlights the importance of combining multiple approaches like the use of different genetic markers and the investigation of behavioral patterns to shed light on the highly complex topic of hybridization.

REFERENCES

Anseeuw D, Maes GE, Busselen P, Knapen D, Snoeks J, Verheyen E (2008) Subtle population structure and male-biased dispersal in two *Copadichromis* species (Teleostei, Cichlidae) from Lake Malawi, East Africa. Hydrobiologia

Baric S, Salzburger W, Sturmbauer C (2003) Phylogeography and evolution of the Tanganyikan cichlid genus *Tropheus* based upon mitochondrial DNA sequences. Journal of Molecular Evolution 56:54-68

Benjamini Y, Hochberg Y (1995) Controlling the false discovery rate: a practical and powerful approach to multiple testing. Journal of the Royal Statistical Society B 57:289-300

Bertorelle G, Excoffier L (1998) Inferring admixture proportions from molecular data. Molecular Biology and Evolution 15:1298-1311

Buerkle CA (2005) Maximum-likelihood estimation of a hybrid index based on molecular markers. Molecular Ecology Notes 5:684-687

Carvajal-Vallejos FM, Duponchelle F, Torrico Ballivian JP, Hubert N, Nunez Rodriguez J, Berrebi P, Sirvas Cornejo S, Renno JF (2010) Population genetic structure of Cichla pleiozona (Perciformes: Cichlidae) in the Upper Madera basin (Bolivian Amazon): Sex-biased dispersal? Molecular Phylogenetics and Evolution 57:1334-1340

Cohen AS, Stone JR, Beuning KRM, Park LE, Reinthal PN, Dettman D, Scholz CA, Johnson TC, King JW, Talbot MR, Brown ET, Ivory SJ (2007) Ecological consequences of early Late Pleisotcene megadroughts in tropical Africa. Proceedings of th National Academy of Sciences of the USA 104:16422-16427

Earl D (2009) Structure Harvester v.0.6.6, Available at http://taylor0.biology.ucla.edu/struct_harvest, (accessed July 2011)

Egger B, Koblmueller S, Sturmbauer C, Sefc KM (2007) Nuclear and mitochondrial data reveal different evolutionary processes in the Lake Tanganyika cichlid genus *Tropheus*. Bmc Evolutionary Biology 7:-

Egger B, Obermueller B, Eigner E, Sturmbauer C, Sefc KM (2008) Assortative mating preferences between colour morphs of the endemic Lake Tanganyika cichlid genus *Tropheus*. Hydrobiologia 615:37-48

Evanno G, Regnaut S, Goudet J (2005) Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study. Molecular Ecology 14:2611-2620

Excoffier L, Laval G, Schneider S (2005) Arlequin (version 3.0): an integrated software package for population genetics data analysis. Evolutionary Bioinformatics Online 1:47-50

Falush D, Stephens M, Pritchard JK (2007) Inference of population structure using multilocus genotype data: dominant markers and null alleles. Molecular Ecology Notes 7:574-578

Gompert Z, Buerkle CA (2009) A powerful regression-based method for admixture mapping of isolation across the genome of hybrids. Molecular Ecology 18:1207-1224

Gompert Z, Buerkle A (2010) INTROGRESS: a software package for mapping components of isolation in hybrids. Molecular Ecology Resources 10:378-384

Knight ME, van Oppen MJH, Smith HL, Rico C, Hewitt GM, Turner GF (1999) Evidence for male-biased dispersal in Lake Malawi cichlids from microsatellites. Molecular Ecology 8:1521-1527

Koblmüller S, Duftner N, Sefc KM, Aibara M, Stipacek M, Blanc M, Egger B, Sturmbauer C (2007) Reticulate phylogeny of gastropod-shell-breeding cichlids from Lake Tanganyika - the result of repeated introgressive hybridization. Bmc Evolutionary Biology 7

Koblmüller S, Egger B, Sturmbauer C, Sefc KM (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

Koblmüller S, Salzburger W, Obermuller B, Eigner E, Sturmbauer C, Sefc KM (2011) Separated by sand, fused by dropping water: habitat barriers and fluctuating water levels steer the evolution of rock-dwelling cichlid populations in Lake Tanganyika. Molecular Ecology 20:2272-2290

Kocher TD, Thomas WK, Meyer A, Edwards SV, Pääbo S, Villablanca FX, Wilson AC (1989) Dynamics of mitochondrial DNA evolution in animals: amplification and sequencing with conserved primers. Proceedings of the National Academy of Sciences of the USA 86:6196-6200

Lang M, Miyake T, Braasch I, Tinnemore D, Siegel N, Salzburger W, Amemiya CT, Meyer A (2006) A BAC library of the East African haplochromine cichlid fish *Astatotilapia burtoni*. Journal of Experimental Zoology Part B: Molecular and Developmental Evolution 306B:35-44

Lee W-J, Conroy J, Huntting Howell W, D KT (1995) Structure and evolution of teleost mitochondrial control regions. Journal of Molecular Evolution 41:54-66

Librado P, Rozas J (2009) DnaSP v5: a software for comprehensive analysis of DNA polymorphism data. Bioinformatics 25:1451-1452

Machado CA, Hey J (2003) The causes of phylogenetic conflict in a classic Drosophila species group. Proceedings of the Royal Society of London Series B-Biological Sciences 270:1193-1202

Mila B, Carranza S, Guillaume O, Clobert J (2010) Marked genetic structuring and extreme dispersal limitation in the Pyrenean brook newt Calotriton asper (Amphibia: Salamandridae) revealed by genome-wide AFLP but not mtDNA. Molecular Ecology 19:108-120

Nei M, Li WH (1979) Mathematical model for studying genetic variation in terms of restriction endonucleases. Proceedings of the National Academy of Sciences 76:5269-5273

Nevado B, Koblmüller S, Sturmbauer C, Snoeks J, Usano-Alemany J, Verheyen E (2009) Complete mitochondrial DNA replacement in a Lake Tanganyika cichlid fish. Molecular Ecology 18:4240-4255

Palumbi SR, Baker CS (1994) Contrasting population-structure from nuclear intron sequences and mtDNA of humpback whales. Molecular Biology and Evolution 11:426-435

Pritchard JK, Stephens M, Donnelly P (2000) Inference of population structure using multilocus genotype data. Genetics 155:945-959

R Development Core Team (2009) R: A language and environment for statistical computing. Vienna, Austria

Rush AC, Cannings RJ, Irwin DE (2009) Analysis of multilocus DNA reveals hybridization in a contact zone between *Empidonax* flycatchers. Journal of Avian Biology 40:614-624

Salzburger W, Baric S, Sturmbauer C (2002) Speciation via introgressive hybridization in East African cichlids? Molecular Ecology 11:619-625

Salzburger W, Niederstatter H, Brandstatter A, Berger B, Parson W, Snoeks J, Sturmbauer C (2006) Colour-assortative mating among populations of *Tropheus moorii*, a cichlid fish from Lake Tanganyika, East Africa. Proceedings of the Royal Society B-Biological Sciences 273:257-266

Schelly R, Salzburger W, Koblmueller S, Duftner N, 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

Scholz CA, King JW, Ellis GS, Swart PK, Stager JC, Colman SM (2003) Paleolimnology of Lake Tanganyika, East Africa, over the past 100 k yr. Journal of Paleolimnology 30:139-150

Seehausen O (2004) Hybridization and adaptive radiation. Trends in Ecology & Evolution 19:198-207

Sefc KM, Baric S, Salzburger W, 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

Stelkens RB, Schmid C, Selz O, Seehausen O (2009) Phenotypic novelty in experimental hybrids is predicted by the genetic distance between species of cichlid fish. Bmc Evolutionary Biology 9:Article No.: 283

Sturmbauer C, Koblmüller S, Sefc KM, Duftner N (2005) Phylogeographic history of the genus *Tropheus*, a lineage of rock-dwelling cichlid fishes endemic to Lake Tanganyika. Hydrobiologia 542:335-366

Swofford DL (2003) PAUP* - Phylogenetic Analysis Using Parsimony and other methods, version 4.0 Sunderland: Sinauer.

Tamura K, Peterson D, Peterson N, Stecher G, Nei M, Kumar S (2011) MEGA5: Molecular Evolutionary Genetics Analysis using Maximum Likelihood, Evolutionary Distance, and Maximum Parsimony Methods. Molecular Biology and Evolution

Taylor MI, Morley JI, Rico C, Balshine S (2003) Evidence of genetic monogamy and female-biased dispersal in the parental mouthbrooding cichlid *Eretmodus cyanostictus* from Lake Tanganyika. Molecular Ecology 12:3173-3177

Thulin CG, Tegelstrom H (2002) Biased geographical distribution of mitochondrial DNA that passed the species barrier from mountain hares to brown hares (genus *Lepus*): an effect of genetic incompatibility and mating behaviour? Journal of Zoology 258:299-306

Turmelle AS, Kunz TH, Sorenson MD (2011) A tale of two genomes: contrasting patterns of phylogeographic structure in a widely distributed bat. Molecular Ecology 20:357-375

Vekemans, X. 2002. AFLP-SURV. Version 1.0. Distributed by the author, Laboratoire Ge´ne´tique et Ecologie Ve´ge´tale, Universite´ Libre de Bruxelles, Belgium.

Vekemans X, Beauwens T, Lemaire M, Roldan-Ruiz I (2002) Data from amplified fragment length polymorphism (AFLP) markers show indication of size homoplasy and of a

relationship between degree of homoplasy and fragment size. Molecular Ecology 11:139-151

Wang J (2003) Maximum-likelihood estimation of admixture proportions from genetic data. Genetics 164:747-765

Weir BS, Cockerham CC (1984) Estimating F-statistics for the analysis of population structure. Evolution 38:1358-1370

Wirtz P (1999) Mother species-father species: unidirectional hybridization in animals with female choice. Animal Behaviour 58:1-12

Yannic G, Dubey S, Hausser J, Basset P (2010) Additional data for nuclear DNA give new insights into the phylogenetic position of *Sorex granarius* within the *Sorex araneus* group. Molecular Phylogenetics and Evolution 57:1062-1071

SUMMARY

This thesis comprises three studies carried out on Lake Tanganyika's *Tropheus moorii* with the aim to shed light on the mechanisms that underlie the outstanding level of phenotypic diversity in terms of body color polymorphism in this species. The focus lied (i) on the genetic basis underlying this divergent trait and (ii) on the identification of the forces contributing to phenotypic divergence.

Chapter 2 dealt with the genetic basis of the color diversity in *T. moorii* as no work had been done in this study area so far. The rapidity of the diversification suggests the presence of only few genetic 'switches', such that color patterns can be altered by minor changes in the genome. We used AFLP data for a genome scan on populations of three closely related color morphs of *Tropheus moorii* and found a total of 11 between-morph outlier loci that fulfilled our criteria for reliable outliers. This part of the study showed first, that genome scans indeed provide a valuable approach for the detection of selected loci even when they are associated with a trait that is not adaptive and secondly, that coloration in *Tropheus moorii* obviously did not evolve by neutral forces alone but also through selection. As expected we found signs of selection acting on only a few genes or loci linked to genes that likely triggered the rapid evolution of different color patterns in this species. These results present an important first step towards the identification of genes being subject to selective forces and contributing to color pattern differences in *T. moorii*. Moreover the study presented in **Chapter 2** dealt with a general characterization of the nature of the body color differences between morphs, with a special focus on the yellow/red coloration. Although these were only first steps towards an accurate characterization of body coloration in *T. moorii* the results give us a valuable overview.

Chapters 3 and **4** focused on the forces contributing to phenotypic diversity in *T. moorii*. As for several cichlids of Lakes Malawi and Victoria (e.g. Seehausen & van Alphen 1999; Seehausen et al. 1999; Knight & Turner 2004; Maan et al. 2004; Pauers et al. 2004) sexual selection as a force driving color diversification has also been discussed for *Tropheus* but without conclusive evidence so far (Sturmbauer & Meyer 1992; Salzburger et al. 2006; Egger et al. 2008, 2010; Sefc 2008; Steinwender et al. 2011). **Chapters 3** and **4** of this thesis dealt with hybridization as yet another force that has been suggested to serve as a potential source of new variants in cichlids (Salzburger et al. 2002; Smith et al. 2003; Schelly et al. 2006; Stelkens et al. 2009). Hybridization between existing morphs could have given rise to novel patterns when, in the wake of Lake Tanganyika's water level fluctuations, allopatric

color morphs were brought into secondary contact. In the studies presented in the last two chapters of this thesis I investigated certain populations in southern Lake Tanganyika which were identified as potential hybrids by previous studies (Baric et al. 2003; Sturmbauer et al. 2005; Sefc et al. 2007; Egger et al. 2007). Both studies confirmed hybridization and thus allow for the possibility that admixture between formally separated populations could very well lead to the generation of novel color phenotypes in the whole *Tropheus* genus. However, the two studies are not totally congruent as the circumstances and consequences of the hybridization slightly differ: The study described in **Chapter 3** dealt with the hybridization between two ancient mitochondrial lineages which are actually separated by a large sandy bay. However, signs of mtDNA introgression were found in previous studies west of the bay but not east of it, thus indicating asymmetric introgression (Baric et al. 2003; Sturmbauer et al. 2005; Sefc et al. 2007). Moreover introgressed populations west of the bay display an intermediate phenotype between the two parental populations. Asymmetric introgression was confirmed by the present study using mtDNA as well as nuclear DNA and the conclusion was drawn that the observed asymmetry was most likely caused by changes in lake level combined with the non-simultaneous occupation of habitats by the two lineages. Thus, concerning color variation in *Tropheus*, this study revealed a very typical consequence of hybridization which is the generation of an intermediate color pattern (Egger et al. 2007). The second hybridization study presented in **Chapter 4**, revealed slightly different circumstances. First, the putative hybrids (previously identified by a homoplasy excess test; Egger et al. 2007) displayed none of the parental phenotypes but instead a new color variant. However, we don't know if the evolution of the observed new phenotype was indeed the result of the genetic admixture of the two adjacent morphs or if the respective populations had already existed as a separate entity (with a different color phenotype) when they were introgressed from both sides. Secondly, in this study mtDNA and nuclear DNA data revealed totally different results. No signs for hybridization were found using mtDNA data, but admixture was strongly confirmed by nuclear data. The observed incongruence between the two marker sets was assigned to the loss of one of the mtDNA lineages due to genetic drift.

Both hybridization studies presented in this thesis revealed first, that interpreting of admixture scenarios is a very challenging task and secondly, that it is crucial to combine multiple approaches, like the use different kinds of molecular genetic markers. Also the consideration of behavioral studies and paleolimnological data is of great importance when investigating contact zones.

Summarized the combined work in this thesis provides a valuable step towards the characterization of the genetics underlying, and the forces driving the stunning color diversity in Lake Tanganyika's *Tropheus moorii* and may serve as the foundation for future work on this extraordinary cichlid.

REFERENCES

Baric S, Salzburger W, Sturmbauer C (2003) Phylogeography and evolution of the Tanganyikan cichlid genus *Tropheus* based upon mitochondrial DNA sequences. Journal of Molecular Evolution 56:54-68

Egger B, Koblmueller S, Sturmbauer C, Sefc KM (2007) Nuclear and mitochondrial data reveal different evolutionary processes in the Lake Tanganyika cichlid genus *Tropheus*. Bmc Evolutionary Biology 7:-

Egger B, Obermueller B, Eigner E, Sturmbauer C, Sefc KM (2008) Assortative mating preferences between colour morphs of the endemic Lake Tanganyika cichlid genus *Tropheus*. Hydrobiologia 615:37-48

Egger B, Mattersdorfer K, Sefc KM (2010) Variable discrimination and asymmetric preferences in laboratory tests of reproductive isolation between cichlid colour morphs. J Evol Biol 23:433-9

Knight ME, Turner GF (2004) Laboratory mating trials indicate incipient speciation by sexual selection among populations of the cichlid fish *Pseudotropheus zebra* from Lake Malawi. Proceedings of the Royal Society of London Series B-Biological Sciences 271:675-680

Maan ME, Seehausen O, Soderberg L, Johnson L, Ripmeester EAP, Mrosso HDJ, Taylor MI, van Dooren TJM, van Alphen JJM (2004) Intraspecific sexual selection on a speciation trait, male coloration, in the Lake Victoria cichlid *Pundamilia nyererei*. Proceedings of the Royal Society of London Series B-Biological Sciences 271:2445-2452

Pauers MJ, McKinnon JS, Ehlinger TJ (2004) Directional sexual selection on chroma and within-pattern colour contrast in *Labeotropheus fuelleborni*. Proceedings of the Royal Society of London Series B-Biological Sciences 271:S444-S447

Salzburger W, Baric S, Sturmbauer C (2002) Speciation via introgressive hybridization in East African cichlids? Molecular Ecology 11:619-625

Salzburger W, Niederstatter H, Brandstatter A, Berger B, Parson W, Snoeks J, Sturmbauer C (2006) Colour-assortative mating among populations of *Tropheus moorii*, a cichlid fish from Lake Tanganyika, East Africa. Proceedings of the Royal Society B-Biological Sciences 273:257-266

Schelly R, Salzburger W, Koblmueller S, Duftner N, 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

Seehausen O, Mayhew PJ, Van Alphen JJM (1999) Evolution of colour patterns in East African cichlid fish. Journal of Evolutionary Biology 12:514-534

Seehausen O, van Alphen JM (1999) Can sympatric speciation by disruptive sexual selection explain rapid evolution of cichlid diversity in Lake Victoria? Ecology Letters 2:262-271

Sefc KM, Baric S, Salzburger W, 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

Sefc KM (2008) Variance in reproductive success and the opportunity for selection in a serially monogamous species: simulations of the mating system of *Tropheus* (Teleostei: Cichlidae). Hydrobiologia 615:21-35

Smith PF, Konings A, Kornfield I (2003) Hybrid origin of a cichlid population in Lake Malawi: implications for genetic variation and species diversity. Molecular Ecology 12:2497-2504

Steinwender B, Koblmüller S, Sefc K (2011) Concordant female mate preferences in the cichlid fish *Tropheus moorii*. Hydrobiologia:1-10

Stelkens RB, Schmid C, Selz O, Seehausen O (2009) Phenotypic novelty in experimental hybrids is predicted by the genetic distance between species of cichlid fish. Bmc Evolutionary Biology 9:Article No.: 283

Sturmbauer C, Meyer A (1992) Genetic divergence, speciation and morphological stasis in a lineage of African cichlid fishes. Nature 358:578-581

Sturmbauer C, Koblmüller S, Sefc KM, Duftner N (2005) Phylogeographic history of the genus *Tropheus*, a lineage of rock-dwelling cichlid fishes endemic to Lake Tanganyika. Hydrobiologia 542:335-366

Curriculum Vitae

Mag.[a] rer. nat. Karin Mattersdorfer
Klosterwiesgasse 73
8010 Graz
Österreich
Tel.: +43 650 4612309
e-mail: Karin.Mattersdorfer@inode.at

Personal information

6.3.1983	born in Friesach, Kärnten
	Nationality: Austrian
	Marital status: Unmarried

Education

2001 – 2005	Studies in Biology (Bacherlor's degree "Verhalten") at the Karl Franzens University of Graz, Department of Zoology
2005 – 2007	Studies in Zoology at the Karl Franzens University of Graz, Department of Zoology Graduation to "Magistra der Naturwissenschaften" (Masters degree) in June 2007 Title of the Master thesis: Molecular genetic analysis of the mating system of the socially monogamous Tanganyika cichlid *Variabilichromis moorii* with the use of microsatellites
2007 – 2008	Studies in Botany at the Karl Franzens University of Graz, Institute of Botany

since July 2008 PhD thesis: The evolution of phenotypic diversity in a Lake Tanganyika cichlid fish at the Department of Zoology, Karl Franzens University of Graz
Supervisor: Kristina M. Sefc

Field Trips

March 2008 Zambia: Research journey to Lake Tanganyika

September 2008 Zambia: Research journey to Lake Tanganyika

September 2009 Zambia: Research journey to Lake Tanganyika

Publications in peer reviewed journals

Sefc KM, Mattersdorfer K, Sturmbauer C, Koblmüller S (2008) High frequency of multiple paternity in broods of a socially monogamous cichlid fish with biparental nest defence. *Molecular Ecology* 17: 2531-2543

Egger B, Mattersdorfer K, Sefc KM (2009) Variable discrimination and asymmetric preferences in laboratory tests of reproductive isolation between cichlid colour morphs. *Journal of Evolutionary Biology* 23: 433-439

Conference presentations

Sefc KM, Hermann CM, Mattersdorfer K, Egger B, Sturmbauer C, Koblmüller S (2007) Genetic paternity analysis in broods of cichlid species with different social mating system. Poster, presented at the "XII European Congress of Ichthyology (ECI XII)" of the EIS (European Ichthyological Society) in Cavtat (Dubrovnik), Croatia.

Mattersdorfer K, Egger B, Sefc KM (2009) Reproductive isolation between cichlid color morphs: Behavioral and genetic studies in *Tropheus moorii*. Presented at the

"International Conference on Evolutionary Ecology of Fishes; Diversification, Adaptation and Speciation" of the "Leibniz Institut für Gewässerökologie und Binnenfischerei" in Erkner, Germany.

Mattersdorfer K, Egger B, Hermann CM, Steinwender B, Sefc KM (2009) Asymmetric cichlids: Behavioral and genetic studies of introgression and mate choice between two color morphs of *Tropheus moorii*. Poster, presented at the meeting "NOBIS 3 – Systematics Open" of the NOBIS (Network Of Biological Systematics) Austria in Vienna, Austria.

Mattersdorfer K, Koblmüller S, Sefc KM (2010) Color diversity in *Tropheus moorii*: Approaches to reveal the evolutionary and genetic basis. Presented at the "Cichlid Science 2010" meeting of the Institute of Zoology in Basel, Switzerland.