This electronic thesis or dissertation has been downloaded from Explore Bristol Research, http://research-information.bristol.ac.uk

Author:
Hsu, Ling-Lan

Title:
Phylogenetic and population genetic structure of riverine Astatotilapia cichlid fishes of East Africa.

General rights
Access to the thesis is subject to the Creative Commons Attribution - NonCommercial-No Derivatives 4.0 International Public License. A copy of this may be found at https://creativecommons.org/licenses/by-nc-nd/4.0/legalcode This license sets out your rights and the restrictions that apply to your access to the thesis so it is important you read this before proceeding.

Take down policy
Some pages of this thesis may have been removed for copyright restrictions prior to having it been deposited in Explore Bristol Research. However, if you have discovered material within the thesis that you consider to be unlawful e.g. breaches of copyright (either yours or that of a third party) or any other law, including but not limited to those relating to patent, trademark, confidentiality, data protection, obscenity, defamation, libel, then please contact collections-metadata@bristol.ac.uk and include the following information in your message:

- Your contact details
- Bibliographic details for the item, including a URL
- An outline nature of the complaint

Your claim will be investigated and, where appropriate, the item in question will be removed from public view as soon as possible.

This electronic thesis or dissertation has been downloaded from Explore Bristol Research, http://research-information.bristol.ac.uk

Author:
Hsu, Ling-Lan

Title:
Phylogenetic and population genetic structure of riverine Astatotilapia cichlid fishes of East Africa.

General rights
Access to the thesis is subject to the Creative Commons Attribution - NonCommercial-No Derivatives 4.0 International Public License. A copy of this may be found at https://creativecommons.org/licenses/by-nc-nd/4.0/legalcode This license sets out your rights and the restrictions that apply to your access to the thesis so it is important you read this before proceeding.

Take down policy
Some pages of this thesis may have been removed for copyright restrictions prior to having it been deposited in Explore Bristol Research. However, if you have discovered material within the thesis that you consider to be unlawful e.g. breaches of copyright (either yours or that of a third party) or any other law, including but not limited to those relating to patent, trademark, confidentiality, data protection, obscenity, defamation, libel, then please contact collections-metadata@bristol.ac.uk and include the following information in your message:

- Your contact details
- Bibliographic details for the item, including a URL
- An outline nature of the complaint

Your claim will be investigated and, where appropriate, the item in question will be removed from public view as soon as possible.

Phylogenetic and population genetic structure of riverine *Astatotilapia* cichlid fishes of East Africa.

Ling-Lan Hsu

A dissertation submitted to the University of Bristol in accordance with the requirements for award of the degree of Masters by Research in the Faculty of Life Sciences.

January 2019

Student Declaration

I declare that the work in this dissertation was carried out in accordance with the requirements of the *University's Regulations and Code of Practice for Research Degree Programmes* and that it has not been submitted for any other academic award. Except where indicated by specific reference in the text, the work is the candidate's own work. Work done in collaboration with, or with the assistance of, others, is indicated as such. Any views expressed in the dissertation are those of the author.

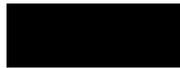
Signed:

Ling-Lan Hsu

Date:

13 / 01 / 2019

Student Number:



Supervisor:

Professor Martin Genner, University of Bristol

Professor Mark Beaumont, University of Bristol

Thesis Abstract

East African freshwater cichlid fish radiations undergo rapid species diversification and thus contain exceptional species richness and phenotypic diversity. However, the origins of the genetic variants that have undergone divergent selection during these radiations are unclear. Although much of the variation likely stems from standing genetic variation present in shared ancestors, it has been proposed that riverine species could promote the sharing of genetic variation by acting as "gene transporters". However, supporting this hypothesis would require both dispersal across riverine boundaries, and hybridizations between species at locations where they overlap.

In Chapter 1, I introduce some background knowledge of East African cichlid radiations, and review important research about the evolution and species diversification of East African cichlid, including some hypotheses based on biogeographic evidence. In Chapter 2, I present a study using ddRAD sequencing data to reconstruct the phylogenetic relationships of riverine haplochromines within East Africa, and their close relatives within the Lake Malawi, Tanganyika and Victoria radiations. The results demonstrate clear evidence of the presence of multiple riverine species being present in East Africa, several that apparently lack formal taxonomic identities. The results also reveal very strong population genetic structure both within and among populations of riverine species, but we could find no clear evidence of elevated levels of recent gene flow between sympatric riverine species. However, patterns of coancestry are potentially indicative of historical gene flow among riverine and lacustrine species in the region. Based on the results we conclude that East African rivers harbour an unexpected diversity of cichlid species, and that multiple extant species may have contributed to genetic diversity present within the lake radiations. However, we suggest that limited dispersal among populations coupled within strong assortative mating of these species in sympatry may strongly limit their role as active gene transporters of contemporary genetic diversity.

Acknowledgements

I would like to express my sincere thanks to many people for their supporting during my Masters by Research at University of Bristol.

First to my supervisor, Professor Martin Genner, for his generous help and tremendous academic support throughout my work. The door to his office was always open whenever I had any questions in my experiments or writings.

I would also like to thank all my lab colleagues who involved and supported me in this project. Especially Dr. Rupert Collins, who really helped me a lot in the lab works and computer programming for data analysis.

Thanks also to Dr. Gary Barker, for his involvement in my six-month assessment viva and constructive reviews in the midpoint of my research project. Thanks again to Dr. Gary Barker, together with Dr. Antonia Ford and Dr. Jon Bridle, who were the examiners and independent chair of my final viva.

Finally, I would like to thank my parents for supporting me studying abroad, and all my friends I met and knew in Bristol for accompanying me in this special life period.

Contents

Student Declaration iii

Thesis Abstract ii

Acknowledgements iv

List of Tables vi

List of Figures vii

List of Appendices viii

Acronyms ix

Chapter One

1.1 Adaptive radiation 1

1.2 Hybridization fuels adaptive radiations 3

1.3 East African Cichlidae 5

1.4 Phylogenetics and population genetics of East African cichlids 9

1.5 Riverine conduits as gene flow pathway 10

1.6 Clarifying the evolutionary relationships of East African river cichlids 12

1.7 Key research aims 14

Chapter Two

Abstract 15

2.1 Introduction 17

2.2 Materials and Methods 19

2.3 Results 266

2.4. Discussion 32

References 377

Appendices 50

List of Tables

Table 2.1. Sampling information of *Astatotilapia* sequenced in the study. 22

Table 2.2. Pairwise *post-hoc* permutation test matrixes 31

List of Figures

Fig. 2.1. Sampling locations of the sequenced specimens. 21

Fig. 2.2. Maximum-likelihood phylogenetic tree 27

Fig. 2.3. Co-ancestry matrix with population structure 30

Fig. 2.4. Boxplots of the genetic distance between population groups........................ 31

List of Appendices

Table S1. GPS coordinates of collection sites. 50

Table S2. Scripts for analysis programming. 51

Table S3. Adapter barcodes and PCR indices. 58

Table S4. Deleted samples filtered by low quality. 59

Table S5. Percentage of reads mapped against the reference genome 60

Table S6. Pairwise genetic distance for all sympatric and allopatric populations. 62

Fig. S1. Photos of selected haplochromines in this study. 66

Fig. S2. Testing optimal threshold of min numbers of samples per locus. 69

Acronyms

RADseq: restriction-site associated DNA sequencing

ddRADseq: double digest restriction-site associated DNA sequencing

SCG: sympatric congeneric populations

ACG: allopatric congeneric populations

ACS: allopatric conspecific populations

ADG: allopatric different genera

Chapter One

General Introduction

1.1 Adaptive radiation

Since Darwin and Wallace published their papers "On the tendency of species to form varieties" and "On the perpetuation of varieties and species by natural means of selection" in 1858, biologists have increasingly recognised the role of ecological variation in evolution (Darwin 2004). Adaptive radiation is a process where organisms simultaneously and rapidly diversify into new forms from the same common ancestry (Schluter 1996). The process is particularly strongly associated with the available of new ecological opportunity, triggered by environmental changes that create new ecological niches (Schluter 1996, 2001). When colonizing species find themselves in territories with few competitors or predators, and many available resources, adaptation to different portions of the available resource spectrum takes place (Simpson 1955; Givnish and Sytsma 2000; Schluter 2001).

According to Schluter in his book "The ecology of adaptive radiation", four features can help identify an adaptive radiation (Schluter 2000). First, the species within the radiation should have common ancestry. Second, there should be evidence of correlation between the environment and the phenotype of these species. Third, the traits that differ among species should have utility within the environment. Finally, the species should have diversified rapidly. Evidence of rapid diversification includes phylogenetic evidence of the evolution of bursts of new eco-morphologically divergent species, ideally around the time of the formation of the new ecological opportunity.

There are now many recognised cases of adaptive radiation across animals and plants, including Galápagos finches (Freeland and Boag 1999), East African cichlids (Salzburger et al. 2002a; Smith et al. 2003), Caribbean *Anolis* lizards (Losos 2007), among others (DeSalle and Giddings 1986; Shaw 2002).

Examples of adaptive radiation

Perhaps the most famous and well-studied example of adaptive radiation are the Darwin's finches. During the voyage to the Galápagos Islands, Darwin discovered a group of passerine birds with diverse beak forms and body sizes. This system has now been extensively studied from ecological and evolutionary perspectives (Bowman 1961; Grant 1999; Petren et al. 1999). Strikingly, the different beak sizes and depths show strong correlations with the food sources, including seeds with different hardness, insects, or vegetation (Bowman 1961). Phylogenetic studies show that the species collected on the islands form a clearly monophyletic clade with respect to mainland passerine birds (Petren et al. 1999). The evidence suggests that common ancestor likely occupy the islands within the last three million years ago, and radiated into diverse forms and functions in a relatively short period of evolutionary time.

Another famous example of adaptive radiation is the Caribbean *Anolis* lizards. *Anolis* is the largest amniote genus, and has experienced extensive adaptive radiation and convergent evolution (Nicholson et al. 2012). Species in the islands of West Indies have adapted to diverse microhabitats, with some species favouring bushes, others the trunk and some the tree crown. This habitat divergence has been accompanied by morphological changes (Losos 2007), including changes in limb lengths which strongly associate with the environmental substrate they most frequently encounter (Losos et al.

1998; Losos 2007). This thesis focusses on an another spectacular adaptive radiation, African cichlids, which dominate the large lakes of the Great East African Rift (Kornfield and Smith 2000; Salzburger et al. 2005).

1.2 Hybridization fuels adaptive radiations

Hybridization continuum

The Biological Species Concept, proposed by Ernst Mayr in 1942, defines species as groups of potentially interbreeding populations that are completely reproductive-isolated from other groups (Mayr 1947; Mayr 1970, 1982; De Queiroz 2005). Although this species concept has built the framework of many evolutionary studies, Mayr's focus on discontinuity and complete reproductive isolation has long been questioned. Many natural phenomena (e.g. speciation) may be better explained by continua instead of discontinuity (Grant and Grant 1992; Schluter and Nagel 1995; Seehausen et al. 1997; Ford et al. 1998; Lu and Bernatchez 1999; Gibbs et al. 2000; Berlocher and Feder 2002; Drès and Mallet 2002; Arnold 2004; Isaac et al. 2004; Mallet et al. 2007). Hybridization between closely related species is now known to be extremely common in natural systems, suggesting that complete reproductively isolation may not be a useful criterion for speciation. It is also known that the extent of hybridization among taxa can vary considerably, generating a wide range of levels of genetic divergence between the parental forms (Harrison and Larson 2014). The point of zero genetic compatibility between parental types, where post-zygotic reproductive isolation is complete, may take millions of years to happen (Price and Bouvier 2002; Harrison and Larson 2014; Grant and Grant 2016). Among species that are closely related, often hybrids are viable

and fertile. This observation has led to suggestions that gene flow may be an important source of adaptive genetic variation. Indeed, there are several groups of organisms that are now thought to have undergone hybridization at an unusually high rate, potentially accelerating the process of adaptive radiation (Lewontin and Birch 1966; Abbott et al. 2013).

Introgressive hybridization

Natural hybridization is often most evident in cases where previously geographically isolated populations ("allopatric populations") re-join in secondary contact. Historically, this was considered as a rare exception instead of a common event (Mayr 1982, 1999). Improvements of molecular techniques have provided a weight of evidence countering this view, and many species are now known to hybridize quite regularly (Endler 1977; Barton and Hewitt 1989; Mallet 2005). Where hybridization in the wild generates fertile and viable offspring that backcross into parental populations, the hybridization is considered to be "introgressive" (Song et al. 2011; Dasmahapatra et al. 2012). Hybridization and subsequent introgression can, in some cases, provide opportunities to form new lineages that may potentially become new species, and can in principle create a variety of recombinants that form the starting point of an adaptive radiation (Seehausen 2004; Harrison and Larson 2014).

Evidence of hybridization has now been found in many population genetic studies of adaptive radiations. For example, introgressive hybridization appears to be common and widespread in the Darwin's finches on the Galápagos Islands (Grant and Grant 2016). Genomic analyses show that hybridization events have happened throughout most of their history, leading to frequent genetic exchanges (Lamichhaney et al. 2015).

Similarly, a survey of two redfish species, *Sebastes fasciatus* and *Sebastes mentella,* showed frequent introgression between them, and these events are thought to have contributed to their genetic distinctness and population structure (Roques et al. 2001).

1.3 East African Cichlidae

Cichlids belongs to the family Cichlidae in the order Perciformes. They are one of the largest vertebrate families, with perhaps as many as 3000 species. However, the number of species in this family remains unknown, since many species are still undescribed and new species have been discovered every year. Cichlids are widely distributed globally, being naturally found in South/Central America, mainland Africa, Madagascar, the Middle East and the Indian subcontinent (Nelson et al. 2016). Across the continents, they have repeatedly undergone adaptive radiation and sympatric speciation (Kornfield and Smith 2000). Most cichlids are found living in relatively shallow depths (< 50m), and are almost exclusively found in freshwater habitats. There are no fully marine cichlid species, although several species are able to tolerate euryhaline brackish conditions. Parental care systems vary extensively among species (McKaye and McKaye 1977; Ribbink et al. 1981), with the most common mechanisms being i) pair-forming substrate spawning brood guarders, and ii) non-pair forming, maternal mouthbrooders, where males contribute only genes to their offspring (Oppenheimer 1970).

Adaptive radiation and speciation of East African cichlids

The cichlid radiations in the East Africa rift valley lakes, Victoria, Malawi and

Tanganyika are considered to be most diverse of all the cichlid faunas (Kornfield 1978; Kornfield and Smith 2000). Thus, the lake cichlids, and those in the connected river systems, have been the most attractive to evolutionary biologists. The most important reasons are the unusually large scale and rapid rates of adaptive radiation that have taken place in these lakes, and the evidence of remarkable patterns of replicated convergent evolution (Meyer 1993; Kornfield and Smith 2000; Seehausen 2006; Brawand et al. 2014).

Each of the three Great East African lakes, Victoria, Malawi and Tanganyika contains several hundred cichlid fish species. In Lake Victoria, it is thought that over 500 endemic species evolved within the past 15,000-100,000 years (Johnson et al. 1996). In Lake Malawi, the radiation of over 700 species took less than 5 million years, while the radiation in Lake Tanganyika of over 300 species took place within the last 10-12 million years (Nishida 1991; Meyer 1993). In each of these lakes there has been diversification in behaviour, body coloration, and body morphology, as well as considerable ecological specialization (Kocher 2004). It is thought that presence of similar habitats and ecological opportunities in different lake systems shaped convergent evolution, producing frequent occurrence of similar ecotypes in different lakes (Kocher 2004; Genner and Turner 2005). It has been suggested that the evolution of these lake radiations may have gone through stages. The first stage represents the divergence of habitat specialists, for example diverging into rock- or sand-dwelling clades. The second stage was divergence of different feeding apparatus, resulting in diverse jaw and tooth shapes. Finally, newly evolved species might primarily differ by their colour patterns, which indicates the preferences of sexual selection within populations. However, recent work suggests that divergence in habitat, morphology and

breeding colours can take place simultaneously (Malinsky et al. 2015).

These cichlid radiations may be characterized by rapid sympatric speciation, since populations within each lake are not completely isolated by geographical barriers (Brawand et al. 2014). Two associated selective forces need to be highlighted in this situation, namely ecological selection and sexual selection (Kocher 2004). In principle, when a threshold is reached in genetic correlation between ecological traits that convey enhanced ecological fitness, and the traits that are favoured by sexually selection, then sympatric ecological speciation may occur rapidly.

The role of hybridization

Hybridization has potential to shape the genetic diversity of the three lacustrine radiations, through admixture among species within the radiation, but also there is the possibility of admixture following occasional colonization events from external sources (Joyce et al. 2011; Loh et al. 2013). Within radiations, interspecific introgression has long been suspected within both Malawi and Victoria (Salzburger et al. 2002a), and new genomic evidence has confirmed this has been widespread (Malinsky et al. 2018a). Such hybridization can facilitate adaptive radiation when coincident with ecological opportunities as it allows the production of novel transgressive phenotypes (Stelkens et al. 2010; Meier et al. 2017). Such phenotypic novelty, distinct from the parental species in body or jaw morphological forms (Kornfield and Smith 2000; Kocher 2004; Seehausen 2006), may allow species to either alter their ecological niche, or for admixed populations to occupy ecological niches distinct from both parental populations.

The possibility of rare colonisation from riverine sources, followed by hybridization

events, may be an important factor of the rapid formation of cichlid radiations, as it allows the transfer of potentially new adaptive alleles that selection can act upon (Grant and Grant 2002; Nichols et al. 2015). Particularly when a lake is newly formed, and colonisation from multiple sources takes place, selective forces against the hybrids may be quite weak, allowing the formation of a hybrid swarm. Perhaps the strongest evidence of this comes from the huge group of haplochromines in Lake Victoria and nearby rift lakes. Here, it appears that ancient hybridization between multiple ancestral haplochromine lineages has formed the genomic diversity of the cichlids in the Lake Victoria radiation, which has subsequently segregated among species during adaptive radiation (Seehausen et al. 2003; Meier et al. 2017).

The process has also been proposed for seeding the genomic diversity of the Lake Malawi system. It has been suggested that the 'mbuna' haplochromine group in Lake Malawi has arisen from introgressive hybridization between two geographically distinct lineages of *Astatotilapia*-like phenotypes (Joyce et al. 2011). However, Malinsky (2018a), using evidence from whole genome sequences, found no evidence to suggest that hybridization was involved in formation of the mbuna. This does not rule out the possibility of multiple ancestors of the Lake Malawi radiation, however. Genner et al. (2015), using mitochondrial DNA identified a potential ancestor of the Lake Malawi radiation in the geographically adjacent Ruaha river system (*Astatotilapia* sp. "Ruaha blue"). Malinsky et al. (2018a) showed that this species was not resolved as a sister species to the radiation using whole genome sequences. Together, this may be suggestive of an ancient partial genomic contribution of the distinct Ruaha species to the Lake Malawi fauna, in addition to the diversity in the flock seeded by an *A. calliptera*-like ancestor.

1.4 Phylogenetics and population genetics of East African cichlids

Phylogenetic evidence demonstrates that the cichlid species radiations in the three Great Lake regions have evolved independently (Seehausen et al. 2003). The radiations within Lake Malawi and Victoria are both from the haplochromine tribe. By contrast, Lake Tanganyika contains radiations belonging to several tribes. It is possible that the Lake Tanganyika radiation has been colonized by the tribes independently from river systems (Nishida 1991; Salzburger et al. 2002b), although it is also possible that tribes originated in the lake itself (Irisarri et al. 2018).

Although there is evidence that recent riverine lineages have involved in the evolution of haplochromine lacustrine radiations (Schwarzer et al. 2012), gene flow between lake and river systems should not be considered a one-way process, as species with lacustrine ancestry might recolonize the rivers (Sturmbauer et al. 2010). Indeed, it has long been proposed that the haplochromine tribe originated within Lake Tanganyika and thereafter colonised river systems and the other Great Lakes in the region (Salzburger et al. 2005). The current distribution may be a consequence of alternative periods of wetness and droughts over the Late Cenozoic in the Eastern Africa (Cohen et al. 1993; DeMenocal 2004; Trauth et al. 2007; Trauth et al. 2010).

The importance of climatic conditions in affecting cichlid species distributions has been best emphasised from studies of lake level fluctuations on Lake Tanganyika. Here, the climate driven changes in water level have led to the fragmentation and reconnection of habitat, and the cichlid populations on those habitats (Cohen et al. 1997). Lake level fluctuations in the Holocene was even more severe in Lake Malawi and Victoria than in Lake Tanganyika (Owen et al. 1990), for example, there was a large drought around

Lake Malawi has fluctuated in water level extensively over the last 1 million years (Ivory et al. 2016; Malinsky and Salzburger 2016). It is thus plausible that earlier radiations were followed by lineage extinction, until the resilient radiations created modern species diversity that we are able to observe today (Seehausen 2006).

Despite historical factors, contemporary evolution associated with environmental perturbation might also affect the population structure of African cichlid fishes. One prominent example comes from Lake Victoria, where increased water turbidity has taken place associated with eutrophication of the catchment. As a consequence, visually mediated assortative mating between species of rock cichlids has broken down, promoting hybridization (Seehausen et al. 1997).

Riverine species often show consideration genetic structure within and among river basins (Hurwood et al. 2008). Such genetic structure can be the result of barriers to dispersal, such as stretches of unsuitable habitat, or adaptation to local environmental conditions. However, such structure may be due to historical events, such as drainage rearrangements, or large-scale droughts that fragment populations. As a consequence of the presence of this intrinsic genetic structure within rivers, riverine ancestors of lacustrine cichlid radiations are likely to have strong genetic differences.

1.5 Riverine conduits as gene flow pathway

One of the more notable features of African cichlids is the extent of convergence of phenotypic traits between radiations, including habitat preferences, trophic morphology (jaws and teeth), and colour patterns (Salzburger et al. 2002b; Kocher 2004; Genner

and Turner 2005). These shared morphological variations have evolved *de-novo* in each lake, and played an important role in the divergence of closely related species. However, it is unlikely that the diversity of traits within each lake have evolved from *de-novo* mutation. Instead, recent studies in humans (Green et al. 2010) and fishes (Colosimo et al. 2005; Loh et al. 2008; Seehausen et al. 2008; Roberts et al. 2009) draw special attention to the importance of shared ancestral genetic variation for promoting adaptation. Such ancestral genetic variants could explain the repeated nature of cichlid radiations, as well as the rapid evolutionary rate of cichlid radiations. The most likely explanation for the presence of standing genetic variation is that ancestral polymorphisms were present in the colonising lineages of all lakes (Loh et al. 2013; Meier et al. 2017). Another non-mutually exclusive hypothesis suggests that standing genetic variation can be moved between populations in different waterbodies through riverine conduits. This “transporter” hypothesis is first proposed for marine and stream-resident three-spine stickleback populations (Schluter and Conte 2009). Phylogenies have shown that marine stickleback populations are the ancestral form to all freshwater populations, and studies provide evidence that hybridization events and gene flow between marine and freshwater-adapted populations accelerate the process of multitrait parallel evolution and speciation.

Similar features are also found in the East African cichlid populations. Evidence has been found that the genetic variation of East African cichlids may have been moved between radiations by river species (Loh et al. 2013). Specifically, this evidence comes from the presence of genetic variation shared between riverine species and lake radiations (Loh et al. 2013). Thus, it is possible that rivers act as gene flow pathways between lakes. However, such a pathway would require both hybridization between

river cichlid species with overlapping geographic ranges, and between lake and river cichlids. Both scenarios are plausible. Multiple riverine cichlid species are often found in direct sympatry, and lacustrine cichlids sometimes partially overlap their habitats with riverine ones - lake cichlid radiations often stretches into the rivers. It would also require the movement of riverine cichlid species across watershed boundaries. Although cichlids tend to have low level of migration, they are often found at the boundary regions of catchments, where flooding can periodically connect catchments. Thus, it may be reasonable to suppose that gene flow among lakes faunas through dispersal and interspecific hybridization could provide standing genetic variation permitting rapid diversification and parallel evolution.

1.6 Clarifying the evolutionary relationships of East African river cichlids

A full evaluation of the role of riverine cichlid species as sources of genetic diversity of East African lacustrine radiations requires an understanding of their diversity. However, many taxa have been poorly sampled for phylogenetic studies, and it is clear that new surveys are likely to reveal new cichlid diversity outside of the Great Lakes (Genner et al. 2015; Malinsky and Salzburger 2016). The poor understanding of riverine haplochromine species diversity of the region is exemplified by knowledge of the diversity in Tanzania. Here, most *Astatotilapia* populations in the central drainage basins have been referred to as *Astatotilapia bloyeti* and *Astatotilapia stappersi* (e.g. (Eccles 1992)). However, the region likely contains several more species (e.g. Genner et al. 2015; Malinsky et al. 2018a)). Thus, there is a need for more studies to explore the genetic diversity of East African riverine cichlids, to further understanding of both

the relationships of river and lake populations, and our formal understanding of the systematics of the system.

Our understanding of the processes driving speciation and adaptive radiation have been dramatically improved over recent years by next generation sequencing of cichlid genomes. For example, genomic data enables the reconstruction of historical demography of studied populations, while enabling the identification of genetic markers undergoing divergence through selection or drift. However, sequencing full genomes remains expensive for most phylogenetic or population genetic applications. Therefore, alternative methods have emerged to enable phylogenetic and population genetic inference from genome-wide data, but from only subsets of the genome. One prominent method is restriction-site associated DNA sequencing (RADseq), an approach that enables the sequencing of anonymous fractions of target genomes after they are digested by restriction enzymes, amplified in a PCR step, and subject to size selection (Davey and Blaxter 2010). This method combines a reduced representation library construction for polymorphism discovery, yet retains the ability to identify and score a large number of genetic markers. To improve RADseq by increasing the number of markers sequenced, Peterson, et al. (2012) established double digest restriction-site associated DNA sequencing (ddRADseq). This method uses two different restriction enzymes, and enables generation of more than ten thousand sequences homologous across the samples. Given each ddRADseq run can simultaneously tagged sequences from hundreds of individuals, the method has enabled researchers to readily identify unlinked SNPs for downstream genetic analyses.

1.7 Key research aims

This study focusses on the haplochromine populations of the river systems of East Africa, and quantified their evolutionary relationships with representatives of the lacustrine faunas. The aim of the work is to a) establish the phylogenetic relationships of the focal populations, and b) test for evidence of hybridization among riverine *Astatotilapia* species where they occur in sympatry. The results are interpreted with respect to existing knowledge of the phylogeny of cichlids of the region, and the plausibility of riverine habitats acting as conduits of genetic diversity between the lake faunas as proposed by the "transporter" hypothesis.

Chapter Two

Genetic structure of *Astatotilapia* cichlid fishes of Tanzania

Abstract

The East African lacustrine cichlid fish radiations contain exceptional species richness and phenotypic diversity. However, the origins of the genetic variants that have undergone divergent selection during these radiations are unclear. Although much of the variation likely stems from standing genetic variation present in shared ancestors, it has been proposed that riverine species could promote the sharing of genetic variation by acting as "gene transporters". However, this would require both dispersal across riverine boundaries, and hybridization between species at locations where they overlap. To test this hypothesis, we used ddRAD sequencing data to reconstruct the phylogenetic relationships of riverine haplochromines within East Africa, and their close relatives within the Lake Malawi, Tanganyika and Victoria radiations. The results reveal clear evidence of the presence of multiple riverine species being present in East Africa, several which apparently lack formal taxonomic identities. The results also reveal very strong population genetic structure both within and among populations of riverine species, but we could find no clear evidence of elevated levels of recent gene flow between sympatric riverine species. However, patterns of coancestry are potentially indicative of historical gene flow among riverine and lacustrine species in the region. On the basis of these results we conclude that East African rivers harbour an unexpected diversity of cichlid species, and that multiple extant species may have contributed to

genetic diversity present within the lake radiations. However, we suggest that limited dispersal among populations coupled within strong assortative mating of these species in sympatry may strongly limit their role as active gene transporters of contemporary genetic diversity.

2.1 Introduction

Groups of organisms that have undergone rapid speciation provide useful opportunities to resolve the mechanistic drivers of the process. Some of the most diverse radiations of vertebrates are of cichlid fishes in the East African rift valley lakes, where an estimated 2000 cichlid species have evolved in the past 10 million years (Nishida 1991; Meyer 1993; Johnson et al. 1996). The highest levels of species richness are within Lakes Malawi, Tanganyika and Victoria, each of which contains a phenotypically diverse radiation of several hundred endemic species (Kocher 2004). Research into these systems has demonstrated the importance of genetic variation for underpinning the functional morphological and ecological diversity that is present (Sturmbauer and Meyer 1992; Kornfield and Smith 2000; Kocher 2004; Seehausen 2006; Brawand et al. 2014). Moreover, research has also demonstrated that the same genetic variants have repeatedly been favoured in cases of parallel evolution between lakes (Salzburger et al. 2002b; Kocher 2004; Genner and Turner 2005) . However, the origins of the genetic variation present within the lake radiations has been unclear.

A key issue is the extent the underlying genetic diversity present within the modern radiations represents ancient standing genetic variation from ancestral colonisers (Loh et al. 2008; Seehausen et al. 2008; Roberts et al. 2009; Loh et al. 2013) or novel mutation that has evolved *in-situ*. Undoubtedly, riverine cichlids have been critical in the provision of genetic material for lacustrine radiation. There is now increasing evidence that there have been multiple independent colonisations of each great lake by riverine ancestors (Verheyen et al. 1994; Joyce et al. 2011), followed by within-lake hybridization, which provided novel opportunities for diversification (Meier et al. 2017). In principle, repeated colonisers could continue to provide novel genetic

diversity during the evolution of systems, and are not necessarily restricted to events early in the evolution of the radiation.

The evidence that genetic variation can be repeatedly be transferred between riverine and lacustrine during adaptive radiations opens up the possibility that genetic variation could be moved between radiations by riverine species (Loh et al. 2013). This "transporter" hypothesis has been proposed as a mechanism promoting extremely rapid adaptation to freshwater environments from marine habitats in sticklebacks (Schluter and Conte 2009; Lescak et al. 2015). However, it is unclear whether this would readily operate within cichlids. This is in part because there are no widely distributed riverine haplochromine species overlap between the great lakes. Thus, support for the hypothesis would require evidence of hybridization among riverine species at locations where the species overlap. Additionally, it would require gene flow among populations within species over relevant timescales, including across catchment boundaries.

In this study, we explore the possibility that riverine ancestors are capable of transporting genetic diversity across continental Africa, testing for evidence of relatively recent interspecific hybridization at contact zones where riverine *Astatotilapia* species overlap. We focus the analysis on *Astatotilapia* from multiple sites across Tanzania, which is geographically intermediate between the Great Lakes. We first construct a phylogeny of the population present using ddRADseq data. We then explore patterns of genetic co-ancestry among populations, and test for differences in the genetic distances observed between sympatric and allopatric population pairs.

2.2 Materials and Methods

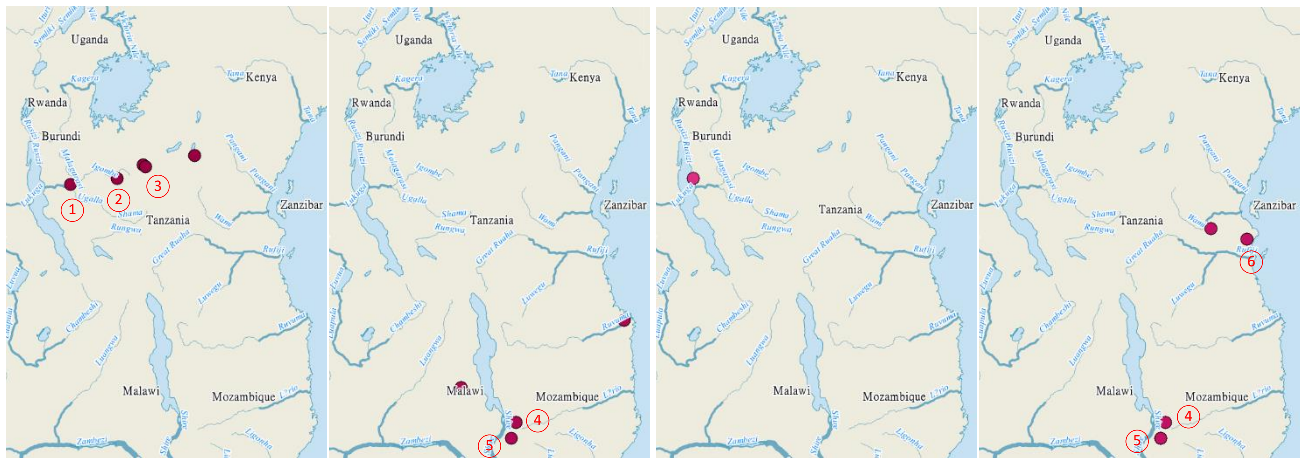
Sample selection

Specimens of haplochromine cichlids were sampled from East Africa between 2009 and 2017. These specimens were primarily *Astatotilapia* from rivers and small lakes, but also included taxa from Lake Victoria, Lake Malawi and the Lake Tanganyika Tropheini (Figure 2.1; Table 2.1; Figure S1). They were identified to species level using morphological traits by the collectors and then preserved in 95% ethanol. Ninety-six samples were selected among the specimens, that included cases where *Astatotilapia* species pairs are found in both sympatry and allopatry, and to cover the phylogenetic diversity of *Astatotilapia* in this region. *Ctenochromis pectoralis* was selected as outgroup. After reconstructing our molecular phylogenetic tree, samples that were initially misidentified were re-labelled to their real species name.

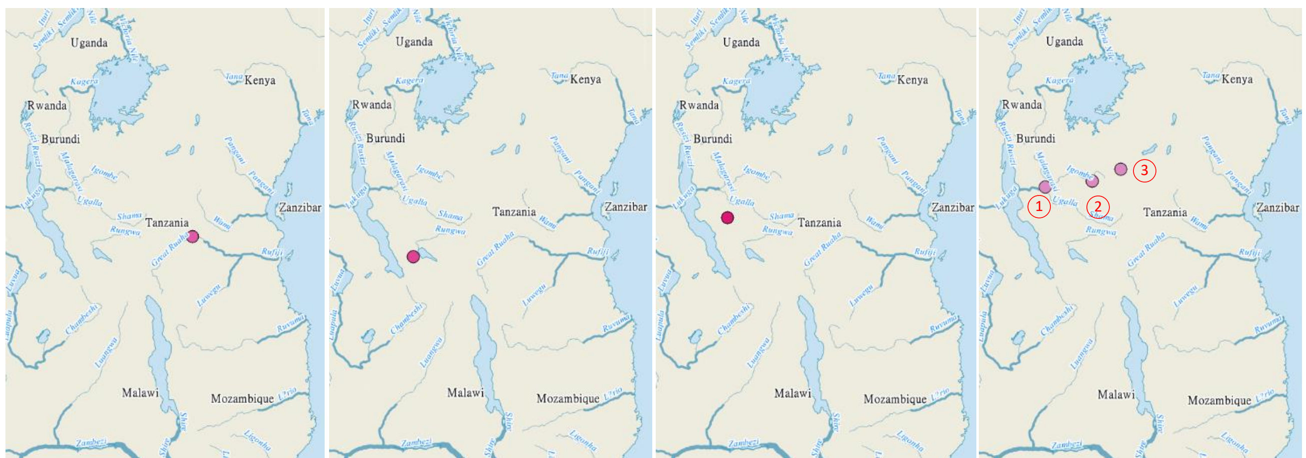
Library preparation and sequencing

Genomic DNA was extracted from fin clips using a modified CTAB method. Fin tissues were cut to approximately 4 mm^2 and placed into 1.5ml centrifuge tubes. 200µl of CTAB (hexadecyltrimethylammonium bromide) buffer and 2.5µl of proteinase K were added into the tube per sample. After 30 minutes of 60ºC incubation, 200µl chloroform were added inside the tubes. Products were then vortexed followed by a 5-min centrifuge in 14,600rpm. Took out the supernatant and placed into a new tube with 400µl 100% ethanol inside. They were vortexed to mix and then centrifuged 5 min in 14,600rpm again layering the liquid. Poured the supernatant and kept only the pellets adhered to the tube surface. Left the tubes in the hood to dry overnight. After adding 50µl of H_2O inside the tube, genomic DNA were extracted from the tissue samples.

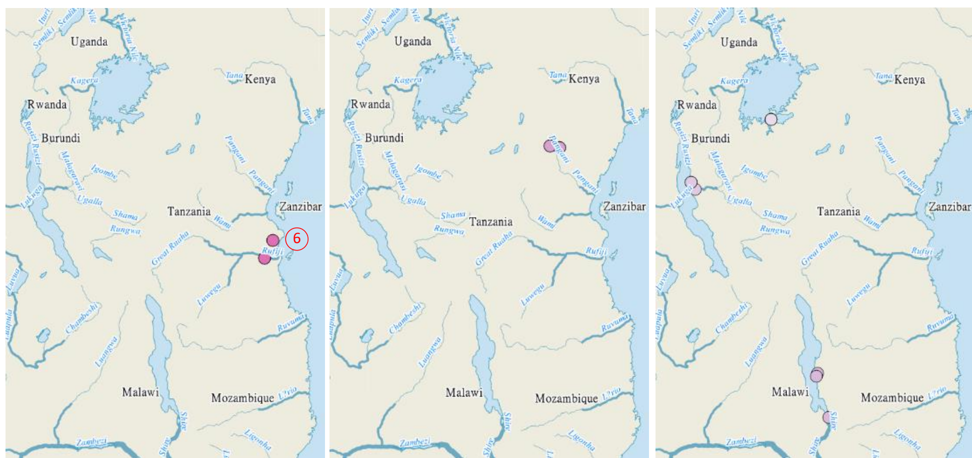
Purification of these DNA products were done by QIAquick PCR Purification Kit (QIAGEN.), and the quality of extracted DNA was measured by Nanodrop through the absorbance ratio at both 260/280 and 230/260 nm (Desjardins and Conklin 2010). DNA concentration was then quantified using a Qubit™ dsDNA HS Assay Kit (Thermo Fisher Scientific Inc.) in a Qubit 4 fluorometer (Thermo Fisher Scientific Inc.) and diluted to standardize the concentration of DNA to 200ng per sample. Sequences of fractioned genomes were generated using double digest restriction enzyme associated sequencing (ddRADseq) following the original protocol (Peterson, et al. 2012). Barcoded adapters were prepared from adapter stocks annealing together before using and were diluted ten times with ddH_2O to 0.4μM (concentration adjusted by the ligation molarity calculator provided in the original protocol). DNA was ligated to barcoded adapters (DNA 15μl, 1xCutSmart® Buffer 5μl, EcoRI 0.1μl, MspI 0.1μl, T4 ligase 0.5μl, ATP 0.5μl, adapter1 2μl, adapter2 2μl, H_2O 24.8μl, adapter1 and adapter2 listed in Table S2, program temperature: 3hr 37ºC, 15min 68ºC). Four replicates of 10-cycle PCR in total 20μl volume (ligated DNA 4μl, 2xPhusion Flash PCR Master Mix 10μl, primer1 1μl, primer2 1μl, BSA 0.5μl, H_2O 3.5μl, indexed primer1 and primer2 listed in Table S2, program temperature: 60s 98ºC, x10(10s 98ºC, 35s 55ºC, 90s 72ºC), 7 min 72ºC) were performed to amplify ddRAD fragments. DNA concentration was then quantified again with a Qubit™ dsDNA HS Assay Kit to help pooling 200 ng PCR products per sample into one tube. After doing AMPure XP bead purification (1.8x) was used to clean up the library. We then performed size selections with E-Gel SizeSelect II 2% Agarose Gels (Thermo Fisher Scientific Inc.) three times, targeting DNA ranging from 650bp – 765bp. The size-selected product was then sequenced using the Miseq Reagent Kit v3 (Illumina, Inc.).



A. bloyeti *A. calliptera* *A.* sp. "Chipwa" *A. gigliolii*



A. sp. "Ruaha Blue" *A.* sp. "pseudopaludinosus" *A. katavi* *A. stappersi*



Other species collected

from Lake Malawi,

Tanganyika and Victoria

A. sp. "Rufiji Blue" *Ctenochromis pectoralis*

Fig. 2.1. Sampling locations of the sequenced specimens. Source: https://qgis.org/en/site/ & https://anitagraser.com/2012/03/10/natural-earth-quick-start-kit/ (sympatric site numbers see Table 2.1)

Table 2.1. Sampling information of *Astatotilapia* sequenced in the study.

Species group	*Species*	*Sample site (Map code)*	**Total:** *96*
Riverine haplochromines	*Astatotilapia bloyeti*	Malagarasi River (①)	2
	Astatotilapia bloyeti	Lake Igombe (②)	3
	Astatotilapia bloyeti	Igogo dam	4
	Astatotilapia bloyeti	Mwamapuli dam (③)	4
	Astatotilapia bloyeti	Burungi	3
	Astatotilapia calliptera	Bua River	4
	Astatotilapia calliptera	Lake Chidya	4
	Astatotilapia calliptera	Lake Chiuta (④)	4
	Astatotilapia calliptera	Lake Chilwa (⑤)	2
	Astatotilapia gigliolii	Mindu Dam	4
	Astatotilapia gigliolii	Lake Mansi (⑥)	4
	Astatotilapia gigliolii	Lake Chiuta (④)	4
	Astatotilapia gigliolii	Lake Chilwa (⑤)	6
	Astatotilapia katavi	Milaca dam	4
	Astatotilapia sp. "Chipwa"	Lake Chipwa	4
	Astatotilapia sp. "pseudopaludinosus”	Lwiche River	4
	Astatotilapia sp. "Ruaha Blue"	Mtera Dam	4
	Astatotilapia sp. "Rufiji blue"	Lake Mansi (⑥)	2
	Astatotilapia sp. "Rufiji blue"	Rufiji River	4
	Astatotilapia stappersi	Malagarasi River (①)	2
	Astatotilapia stappersi	Lake Igombe (②)	5
	Astatotilapia stappersi	Mwamapuli dam (③)	4
Lake Tanganyika Tropheini	*Ctenochromis horei*	Malagarasi River	1
	Gnathochromis pfefferi	Kagera River	1
	Lobochilotes labiatus	Kagera River	1
Lake Malawi haplochromines	*Otopharynx speciosus*	Shire River	1
	Metriaclima zebra	Minos reef	1
	Rhamphochromis longiceps	Metangula market	1
Lake Victoria haplochromines	*Neochromis omnicaeruleus*	Makobe island	1
	Paralabidochromis chilotes	Makobe island	1
	Paralabidochromis sauvagei	Makobe island	1
Outgroup haplochromines	*Ctenochromis pectoralis*	Ruva River	3
	Ctenochromis pectoralis	Chemka Hot Spring	3

Data analyses

Raw sequence reads were first demultiplexed by indices embedded in PCR primers using CUTADAPT v1.16 (Martin 2011) (PCR indices see Table S3). PCR indices and associated adaptors in the data were then trimmed in the next step using the same package. The trimmed output was then demultiplexed again by the ligated adapter barcodes (adapter barcode see Table S3). The barcodes were trimmed off in the last step in order not to mix up the following phylogenetic results (See Table S2 for the programming scripts). Sequences were all aligned to the *Metriaclima zebra* reference genome UMD2a (Conte and Kocher 2015; available at https://www.ncbi.nlm.nih.gov/genome/2640) using IPYRAD v0.7.25 (Eaton and Overcast 2016) (IPYRAD step 1, see Table S2). In the next step, these sequences were filtered based on quality scores of base calls (IPYRAD step 2) and then clustered within each sample by setting clustering threshold to 0.85 (default setting) (IPYRAD step 3). After estimating sequencing error rate and heterozygosity (IPYRAD step 4), the obtained data were used to call the consensus of sequences within each cluster (IPYRAD step 5). Therefore, all loci have been reduced to get one consensus sequence per sample, and those with heterozygous sites were represented by IUPAC ambiguity codes. Additionally, these output consensus sequences were clustered across samples (IPYRAD step 6). Again, we use the default clustering threshold 0.85 to identify similar sequences between samples. Single nucleotide polymorphisms (SNPs) were discovered and were built in multiple formats after filtered by a quality check (IPYRAD step 7). We tested the number of samples per locus for output and chose 42 as threshold based on the results to retain loci in the dataset only when more than 42 individuals had that same read (See Table S2 for the programming scripts and the parameter setting for each step and Figure S2 for the

results). We reconstructed maximum-likelihood (ML) phylogenies using the SNP matrix and PHYML 3.0 (Guindon et al. 2010), with branch support resolved from 100 bootstrap replicates. All parameters were set default. Phylogenetic trees were viewed in FIGTREE v1.4.3 (available at http://tree.bio.ed.ac.uk/software/figtree/).

Population level co-ancestry coefficients were calculated and visualized using FINERADSTRUCTURE v0.3.2 (Malinsky et al. 2018b). This program infers clustered population structure based on shared ancestry among the individuals in the populations. It is specifically programmed for RADseq data, modified from the original package FINESTRUCTURE. Output loci format from IPYRAD v0.7.25 was first converted to finerad format by FINERAD_INPUT.PY (provided on https://github.com/edgardomortiz /fineRADstructure-tools). Generated data was applied to RADPAINTER, which is implemented in the FINERADSTRUCTURE package, to calculate the co-ancestry matrix. We then used the script FINESTRUCTURE to assign individuals to populations with 100,000 burn-in periods and 100,000 Markov chain Monte Carlo iterations. The same script was used to build a simple Bayesian tree with posterior population assignment probabilities. To plot the co-ancestry matrix, we applied the provided R scripts FINERADSTRUCTUREPLOT.R and FINESTRUCTURELIBRARY.R in R v3.5.2 (Scripts available at http://cichlid.gurdon.cam.ac.uk/fineRAD structure.html).

Nei's standard genetic distance (Nei 1978) and Slatkin's linearized pairwise F_{ST} values (Slatkin 1995) between riverine populations were estimated. We defined the sampled population pairs into four groups, "sympatric congeneric (SCG)", "allopatric congeneric (ACG)", "allopatric conspecific (ACS)" and "allopatric different genera (ADG). Mean genetic distances among these four groups were calculated and compared using ANOVA (See Table S2 for the programming scripts), followed by *post-hoc*

pairwise permutation one-way ANOVA test to discover gene flow between sympatric populations (See Table S2 for the programming scripts). Group ACS and group ADG were control groups. The former should have the shortest mean genetic distance and the latter should have the longest result.

2.3 Results

Sequencing and SNP discovery

Approximately 29.17 million reads were generated for all the 96 target individuals were produced by Illumina 250-bp paired-end sequencing with quality score (Q30) ranging from 57.62% – 65.45%. Thirteen samples with filtered sequence reads less than 30,000 were removed from the dataset (listed in Table S4). The percentage of reads mapped against the reference genome was about 63.66 % – 97.50 % per individual (Table S5), and in total 11,924 SNPs were called.

Phylogenetic analysis

The maximum likelihood phylogenetic tree constructed using all 11924 SNPs was rooted on the outgroup *Ctenochromis pectoralis*, and overall consistent with presence of nine distinct *Astatotilapia* lineages (Figure 2.2). The Tropheini were resolved as a sister lineage to a larger clade that contained all riverine *Astatotilapia*, and the haplochromines from the radiating Lake Malawi and Lake Victoria flocks. The Lake Malawi flock, including the habitat generalist *Astatotilapia calliptera*, was resolved as monophyletic. The Lake Victoria flock was resolved as being part of a broader Lake Victoria superflock, containing riverine haplochromines from Lake Rukwa (*Astatotilapia katavi*, *Astatotilapia* sp. "pseudopaludinosus"), the Rufiji river (*Astatotilapia* "Rufiji Blue") and the Malagarasi/Lake Tanganyika catchment (*Astatotilapia* sp. "Chipwa", Astatotilapia stappersi). There was a widely distributed *Astatotilapia bloyeti* was resolved as monophyletic, with a deep division evident between the Lake Burungi population and those in the Malagarasi system (Igogo Dam, Igombe Dam, Malagarasi River, Mwamapuli Dam).

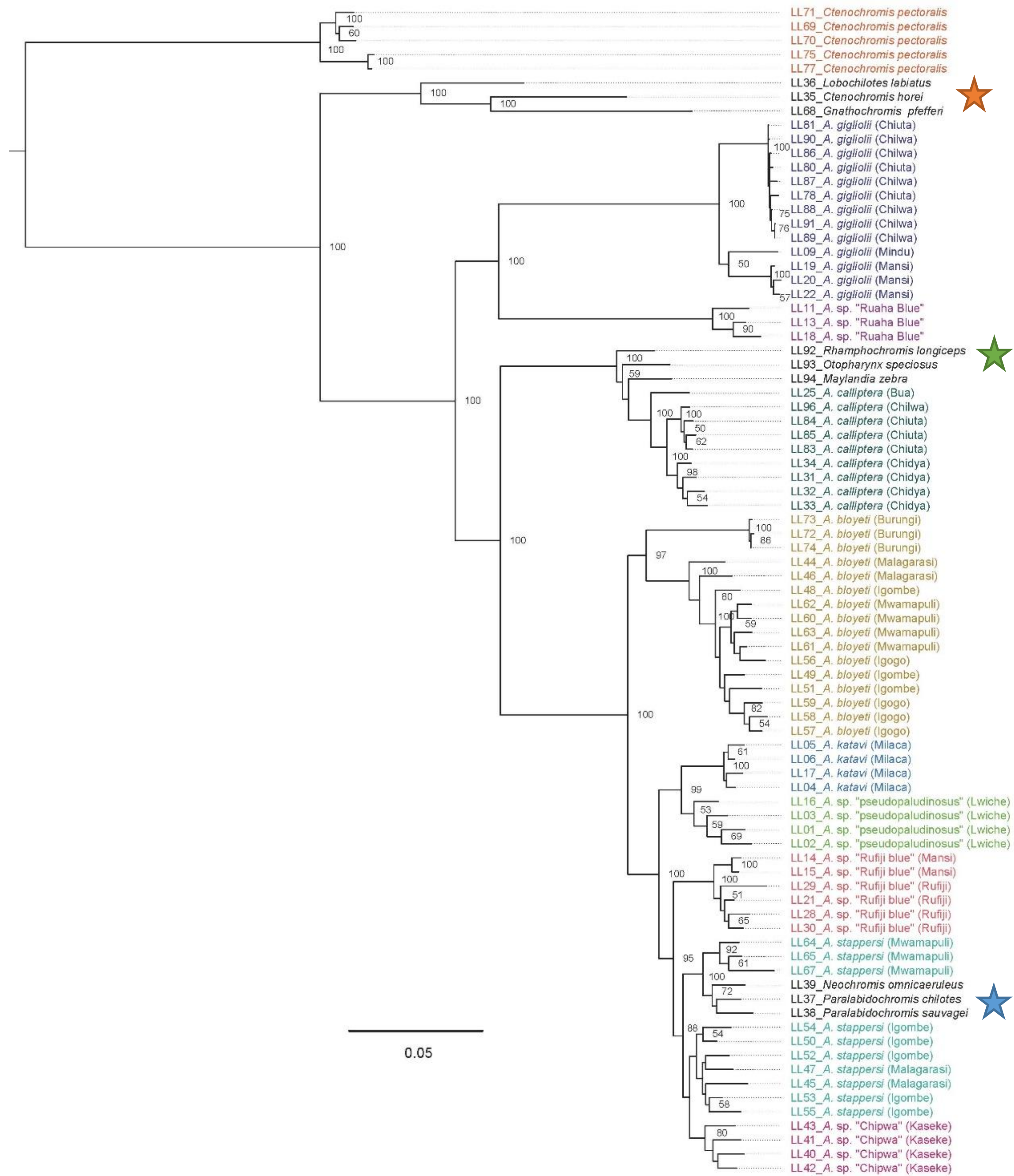


Fig. 2.2. Maximum-likelihood phylogenetic tree estimating relationships among 83 caught individuals**.** Percentage of bootstrap values are given along the branches and only values greater than 50% are shown. Star signs in different colours, orange, green and blue, represents endemic cichlid species of Lake Tanganyika, Malawi and Victoria, respectively.

The broadly distributed *A. gigliolii* also showed strong spatial structure across its range, with the geographically proximate southern Lake Chilwa and Chiuta populations, being strongly differentiated from the northern Lake Mansi and Mindu Dam populations. The phylogeny also resolved a monophyletic *A*. "Ruaha Blue" as a sister clade to the *A. gigliolii*, while the Lake Malawi radiation was resolved as sister to a clade containing *A. bloyeti* and the broader Lake Victoria superflock.

Population structure and hybridization events

The co-ancestry matrix generated by FINERADSTRUCTURE was broadly consistent with the phylogeny, with the highest levels of co-ancestry among individuals from the same species and population groupings (Figure 2.3). Cells with deeper colours indicate closer relationships of two individuals having more numbers of similar genome than others. Even within the same species, individuals collected from different sites may also affect their genomic similarities, causing the ones collected from closer areas having deeper colours. For example, the distinct colour blocks among *A. gigliolii* individuals (Figure 2.3). Importantly, there were no signals of high-levels of coancestry (i.e. values > 100) between populations of different species. However, patterns of coancestry did not uniformly follow the pattern observed in the phylogeny, with relatively high level of coancestry visible among some relatively distantly related species. Specifically, *A. gigliolii* was not uniformly related to individuals from the larger clade containing the Lake Malawi and Lake Victoria flocks. Instead, it had closer ancestry to Lake Malawi endemics, *A .calliptera* and *A*. "Rufiji Blue", than other species in the group. Similarly, *A*. "Ruaha Blue" showed stronger coancestry with the Lake Malawi endemics and *A. calliptera*, than the Lake Victoria flocks. Intriguing, the Lake Tanganyika Tropheini

also showed relatively strong coancestry with *A*. "Ruaha Blue" in particular, and also the Lake Malawi endemics and *A. calliptera*, relative to the broader Lake Victoria superflock.

Mean genetic distances between populations in each of defined groups ("sympatric congeneric (SCG)", "allopatric congeneric (ACG)", "allopatric conspecific (ACS)" and different genera (DG)") showed significant differences (ANOVA; F_{ST}, $F_{3, 249}$= 25.88, $p < 0.001$; Nei's distance, $F_{3, 249}$= 149.4; $p < 0.001$) (Original data per group in Table S5). *Post-hoc* Tukey's HSD tests showed that genetic distance was greatest between different genera, and most similar between populations of conspecifics, but we found no clear evidence of differences in the genetic distances present between sympatric or allopatric congeneric species. Permutation one-way ANOVA show the same pattern as the normal one, showing little differences in the genetic distances between the two groups, ACG and SCG (Figure 2.4, Table 2.2).

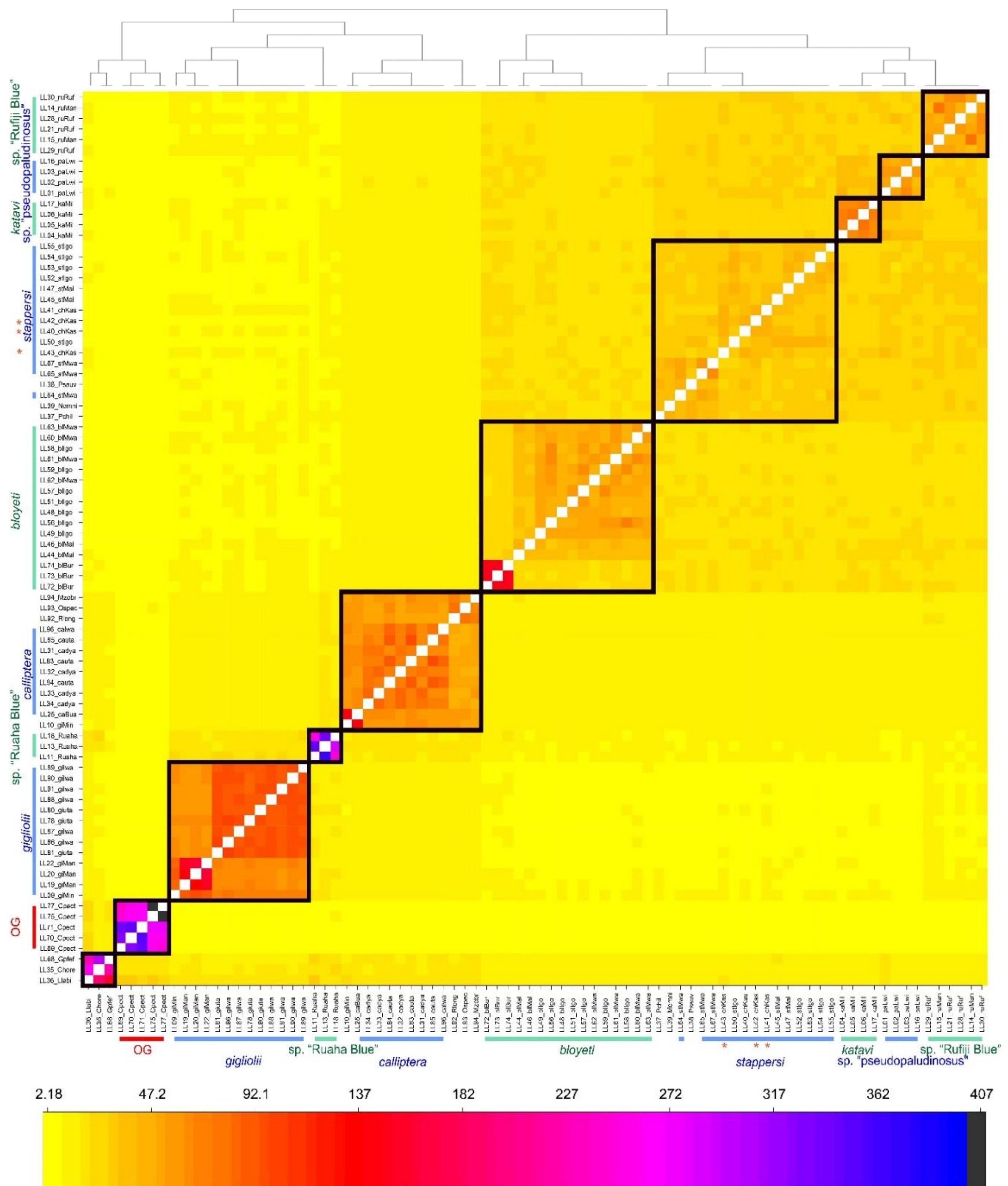


Fig. 2.3. Co-ancestry matrix with population structure from fineRADstructure. The colour of each cell indicates the expected number of similar genome from one individual to another. Species name is provided next to each strain in abbreviation. Red "*" sign shown out of the matrix indicates *A.* sp. "Chipwa".

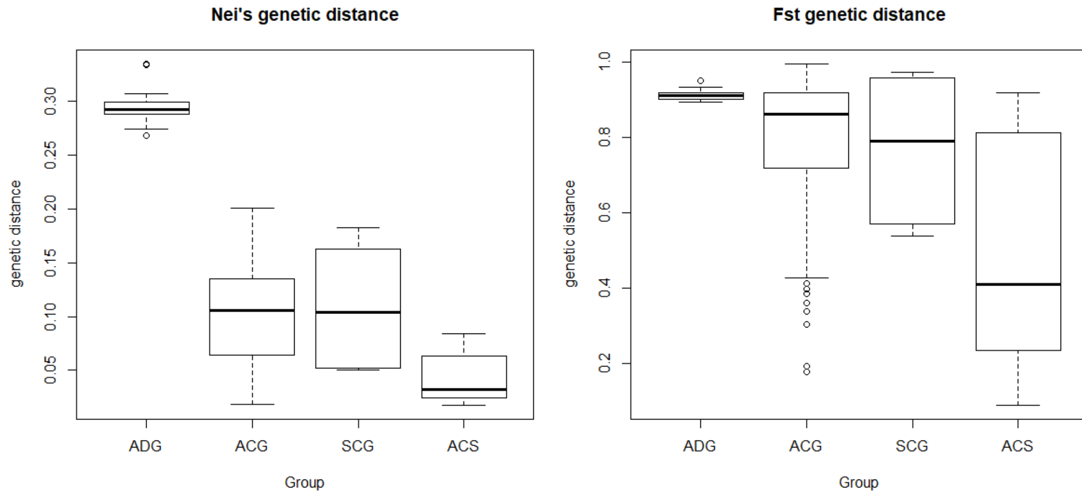


Fig. 2.4. Boxplots of the genetic distance (Nei's and F_{ST}) among populations within each of four groups. ADG = allopatric different genera, SCG = sympatric congeneric, ACG = allopatric congeneric, ACS = allopatric conspecific.

Table 2.2. Results of pairwise post-hoc permutation test matrixes comparing the extent of interspecific genetic divergence present among groups of East African river cichlids. Adjusted p-values calculated from Nei's standard genetic distance were in the lower diagonal, and p-values from Slatkin's linearized pairwise F_{ST} were in the upper one. Non-significant results were marked in red colours.

Population pairwise permutation matrix (Nei's D \ F_{ST})				
	ADG	**ACG**	**SCG**	**ACS**
ADG		5.898e-03	7.026e-03	5.790e-06
ACG	0.000e+00		0.06296	3.700e-11
SCG	1.876e-06	0.09315		0.055900
ACS	3.549e-11	4.380e-10	3.415e-04	

2.4. Discussion

Unexpected diversity in East African rivers

This study, comprising populations of riverine haplochromines from Tanzania, represents one of only a handful of studies that have considered the phylogenetic relationships of these species. Importantly the structure of the phylogeny maps onto that of a phylogeny based on a smaller number of individuals, but using whole genome sequences (Malinsky et al. 2018a). Specifically, in both phylogenies *A. bloyeti* is resolved as sister to a broad group of species including *Astatotilapia* from Lake Rukwa, and those from the lower reaches of the Ruaha river (*A*. "Rufiji Blue). Moreover, in both phylogenies *A. gigliolii* (senior synonym of *A. tweddlei*, G. Turner pers. comm.), is sister to *Astatotilapia* from the upper reaches of the Ruaha (*A*. "Ruaha Blue"). Similarly, in both phylogenies *A. calliptera* falls within the Lake Malawi radiation. Finally, in both analyse the Lake Malawi group, as a whole, is sister to the larger "Lake Victoria superflock" group that includes *A. bloyeti*. Extremely similar relationships are notable from the RAD-based phylogeny of (Meier et al. 2017) that has several lineages that overlap with our study. Taken together, this universal congruence is indicative of our ddRAD data providing a strong resolution of the most plausible phylogenetic relationships of the focal haplochromine species.

A key finding from our work is that it provides strong evidence of for the distinctness of several riverine species, including some that are yet to undergo formal taxonomic description. The work supports the distinct identity of *A.* "Ruaha Blue", a recently identified endemic of the upper Ruaha region (Genner et al. 2015). The phylogeny also

highlights the distinctness *of A. katavi* from another species in the Rukwa catchment, *A*. “pseudopaludinosus”, consistent with the presence of multiple species in the Rukwa catchment as suggested by Seegers (1996). Similarly, the results support the distinction of *A*. “Rufiji blue”, a unique taxon currently known only from the lower Rufiji and Lake Mansi, which are approximately 60km apart. Intriguingly, the data suggests that there are two groups of *A. stappersi* populations. One *A. stappersi* group is found within the Malagarasi catchment (Lake Igombe and the Malagarasi river) and is closely related to the undescribed *A*. “Chipwa” from the same catchment (Meyer et al. 2015). The second *A. stappersi* group was resolved as sister to the Lake Victoria flock, and comprised only individuals sampled from the Mwamapuli Dam in the endorheic Lake Eyasi basin, which is geographically adjacent to the Lake Victoria basin. We consider it plausible that several of these genomically unique lineages recovered in this study represent distinct novel species. Confirmation will require further evaluation of their morphological characteristics, and in the cases where male colour has started to diverge (e.g. *A. bloyeti* populations) then ideally mate choice preferences would be established (Turner et al. 2001).

The role of riverine species in seeding lake radiations

Evidence of close evolutionary relationships between the lacustrine haplochromine species flocks and riverine taxa has fuelled the concept that riverine species have contributed to the lake radiations (Seehausen et al. 2003; Genner et al. 2015; Meier et al. 2017; Malinsky et al. 2018a). Our study highlights the close relationships of some riverine species to the lacustrine flocks. However, it identified uneven patterns of coancestry between the lacustrine flocks and putative ancestors. Three specific cases

were evident. 1) *Astatotilapia gigliolii* shared greater coancestry with lake Malawi species flock (including *A. calliptera*). Notably, these two lineages are found across Ruvuma system, including in full sympatry such as Lake Chiuta. Hence, it is possible that genetic exchange has taken place ancestrally, and the *A. gigliolii* has contributed to the Lake Malawi flock. 2) *Astatotilapia* "Ruaha Blue" shared greater coancestry with the Lake Malawi species flock (including *A. calliptera*), than the phylogenetically close *A. bloyeti* and the Lake Victoria superflock. Intriguingly, *A*. "Ruaha Blue" is found in a geographically adjacent catchment to the Lake Malawi flock, and there are suggestions of ancient riverine connections between the two catchments that is now separated by the Livingstone mountain range (Genner et al. 2013). Moreover, the Lake Malawi flock has unusually high similarity of mtDNA to *A*. "Ruaha Blue". Collectively, this is suggestive of ancient contributions by *A*. "Ruaha Blue" to the genetic structure of the Lake Malawi species flock. 3) *Astatotilapia* "Ruaha Blue" shared greater coancestry with the Lake Tanganyika Tropheini than *A. gigliolii*. Currently, *A*. "Ruaha Blue" and the Tropheini lineages are entirely allopatric, although historic connections between the Ruaha and Malagarasi make for a plausible route of historic connection between the two lineages. In summary, our data reveals uneven patterns of coancestry suggesting future research directions for the study of historical contributions to the genetic diversity of the lacustrine radiations.

Genetic structure within and among riverine cichlid species

Our results showed statistically significant genetic differences among allopatric populations of the same taxa (*p-value* < 0.05 in both Nei's D and F_{ST} dataset), with many allopatric populations of the same species being resolved as reciprocally

monophyletic in the phylogeny. Those population pairs that were not resolved as reciprocally monophyletic were typically geographically proximate. As an example*, A. gigliolii* in Lake Chilwa and Chiuta have been separated for a maximum of 10,000 years since the last high-stand that connected the lake systems (Thomas et al. 2009). Such a pattern was not ubiquitous, however. Specimens of *A. bloyeti* from Mwamapuli in Eyasi catchment were nested within a broader clade of *A. bloyeti* individuals sampled from the Malagarasi catchment (Lakes Igombe, Igogo and the Malagarasi river). Collectively, this evidence is suggestive of strong isolation between most conspecific populations, connected by occasional gene flow primarily within catchments, but with relatively ancient cross catchment episodes of genetic exchange. Such cross-catchment exchanges are plausible during flooding events that link headwater streams (Koblmüller et al. 2012), or river capture events that change the direction of flow following tectonic activity (Burridge et al. 2006).

Our results showed strong genetic distances between riverine species, both in sympatry ($p < 0.05$) and allopatry ($p < 0.05$). However, there was no clear evidence for high levels of recent interspecific gene flow from our analyses at locations where multiple species co-occur, or in allopatry. These results are consistent with strong assortative mating between the species that occupy river systems. Intriguingly, in the cases of where sympatric species were found, they differed strikingly in male breeding colours (Figure S1). Such a pattern is consistent with female choice for male colour being associated with assortative mating, as is well known for lacustrine haplochromine cichlids [e.g. (Seehausen et al. 1997; Knight et al. 1998)]. However, the maintenance of reproductive isolation within turbid impoundments is also consistent with other mechanisms promoting reproductive isolation. Potentially these species have diverged in preferred

breeding habitats, and have other mate choice cues including olfaction (Brock and Wagner 2018), auditory (Amorim et al. 2013) or even hydrodynamic courtship signals (Butler and Maruska 2015).

The relevance of rivers as transporters of genetic material.

Loh *et al.* (2013) suggested shared polymorphism in East African cichlids may be partially explained by transportation of genetic material through riverine corridors. If so, then we would expect to see hybridization between different lineages, and evidence of intraspecific gene flow within and across catchments. Our study has failed to find any clear support for extensive and *recent* sharing of alleles through hybridization, and instead found both strong interspecific and intraspecific genetic structuring within the riverine *Astatotilapia* fauna. Thus, we conclude that riverine East African cichlids are unlikely to act as continuous transporters of functional genetic material over contemporary timescales, as they may in three-spined sticklebacks (Schluter and Conte 2009). However, a considerable diversity of riverine cichlids, as evidenced here, does allow for the possibilities of multiple colonisations into lake systems over long-term evolutionary timescales, potentially leading to the sharing of adaptive alleles and the promotion of adaptive radiation.

References

Abbott, R., D. Albach, S. Ansell, J. W. Arntzen, S. J. Baird, N. Bierne, J. Boughman, A. Brelsford, C. A. Buerkle, and R. Buggs. 2013. Hybridization and speciation. Journal of Evolutionary Biology 26:229-246.

Amorim, M. C. P., A. N. da Ponte, M. Caiano, S. S. Pedroso, R. Pereira, and P. J. Fonseca. 2013. Mate preference in the painted goby: the influence of visual and acoustic courtship signals. Journal of Experimental Biology 216:3996-4004.

Arnold, M. L. 2004. Natural hybridization and the evolution of domesticated, pest and disease organisms. Molecular Ecology 13:997-1007.

Barton, N. and G. M. Hewitt. 1989. Adaptation, speciation and hybrid zones. Nature 341:497-503.

Berlocher, S. H. and J. L. Feder. 2002. Sympatric speciation in phytophagous insects: moving beyond controversy? Annual Review of Entomology 47:773-815.

Bowman, R. 1961. Morphological differentiation and adaptation in the Galápagos finches. Diferenciación morfológica y adaptación en los pinzones de las Galápagos. University of California Publications in Zoology 58:1-302.

Brawand, D., C. E. Wagner, Y. I. Li, M. Malinsky, I. Keller, S. Fan, O. Simakov, A. Y. Ng, Z. W. Lim, and E. Bezault. 2014. The genomic substrate for adaptive radiation in African cichlid fish. Nature 513:375-381.

Brock, C. D. and C. E. Wagner. 2018. The smelly path to sympatric speciation? Molecular ecology 27:4153-4156.

Burridge, C., D. Craw, and J. Waters. 2006. River capture, range expansion, and cladogenesis: the genetic signature of freshwater vicariance. Evolution 60:1038-1049.

Butler, J. M. and K. P. Maruska. 2015. The mechanosensory lateral line is used to assess

opponents and mediate aggressive behaviors during territorial interactions in an African cichlid fish. Journal of Experimental Biology 218:3284-3294.

Cohen, A. S., K. E. Lezzar, J. J. Tiercelin, and M. Soreghan. 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, A. S., M. J. Soreghan, and C. A. Scholz. 1993. Estimating the age of the formation of lakes - an example from Lake Tanganyika. Geology 21:511-514.

Colosimo, P. F., K. E. Hosemann, S. Balabhadra, G. Villarreal, M. Dickson, J. Grimwood, J. Schmutz, R. M. Myers, D. Schluter, and D. M. Kingsley. 2005. Widespread parallel evolution in sticklebacks by repeated fixation of ectodysplasin alleles. Science 307:1928-1933.

Conte, M. A. and T. D. Kocher. 2015. An improved genome reference for the African cichlid, Metriaclima zebra. BMC Genomics 16:724.

Darwin, C. 2004. On the origin of species, 1859. Routledge.

Dasmahapatra, K. K., J. R. Walters, A. D. Briscoe, J. W. Davey, A. Whibley, N. J. Nadeau, A. V. Zimin, D. S. T. Hughes, L. C. Ferguson, S. H. Martin, C. Salazar et al. 2012. Butterfly genome reveals promiscuous exchange of mimicry adaptations among species. Nature 487:94-98.

Davey, J. W. and M. L. Blaxter. 2010. RADSeq: next-generation population genetics. Briefings in Functional Genomics 9:416-423.

De Queiroz, K. 2005. Ernst Mayr and the modern concept of species. Proceedings of the National Academy of Sciences 102:6600-6607.

DeMenocal, P. B. 2004. African climate change and faunal evolution during the Pliocene-Pleistocene. Earth and Planetary Science Letters 220:3-24.

DeSalle, R. and L. V. Giddings. 1986. Discordance of nuclear and mitochondrial DNA phylogenies in Hawaiian Drosophila. Proceedings of the National Academy of Sciences 83:6902-6906.

Desjardins, P. and D. Conklin. 2010. NanoDrop microvolume quantitation of nucleic acids. Journal of visualized experiments: Journal of Visualised Experiments 2010: 2565

Drès, M. and J. Mallet. 2002. Host races in plant–feeding insects and their importance in sympatric speciation. Philosophical Transactions of the Royal Society of London B: Biological Sciences 357:471-492.

Eaton, D. and I. Overcast. 2016. ipyrad: interactive assembly and analysis of RADseq data sets. WWW document] URL http://ipyrad. readthedocs. io/[accessed 1 August 2016].

Eccles, D. H. 1992. Field guide to the freshwater fishes of Tanzania. FAO Species Identification Sheets for Fishery Purposes. FAO, Rome.

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

Ford, J. K., G. M. Ellis, L. G. Barrett-Lennard, A. B. Morton, R. S. Palm, and K. C. Balcomb III. 1998. Dietary specialization in two sympatric populations of killer whales (Orcinus orca) in coastal British Columbia and adjacent waters. Canadian Journal of Zoology 76:1456-1471.

Freeland, J. R. and P. T. Boag. 1999. The mitochondrial and nuclear genetic homogeneity of the phenotypically diverse Darwin's ground finches. Evolution 53:1553-1563.

Genner, M. J., E. Connell, A. Shechonge, A. Smith, J. Swanstrom, S. Mzighani, A. Mwijage, B. P. Ngatunga, and G. F. Turner. 2013. Nile tilapia invades the Lake

Malawi catchment. African Journal of Aquatic Science 38:85-90.

Genner, M. J., B. P. Ngatunga, S. Mzighani, A. Smith, and G. F. Turner. 2015. Geographical ancestry of Lake Malawi's cichlid fish diversity. Biology Letters 11.

Genner, M. J. and G. F. Turner. 2005. The mbuna cichlids of Lake Malawi: a model for rapid speciation and adaptive radiation. Fish and Fisheries 6:1-34.

Gibbs, H. L., M. D. Sorenson, K. Marchetti, M. d. L. Brooke, N. Davies, and H. Nakamura. 2000. Genetic evidence for female host-specific races of the common cuckoo. Nature 407:183-186.

Givnish, T. J. and K. J. Sytsma. 2000. Molecular evolution and adaptive radiation. Cambridge University Press.

Grant, B. R. and P. R. Grant. 2002. Simulating secondary contact in allopatric speciation: an empirical test of premating isolation. Biological Journal of the Linnean Society 76:545-556.

Grant, P. R. 1999. Ecology and evolution of Darwin's finches. Princeton University Press.

Grant, P. R. and B. R. Grant. 1992. Hybridization of bird species. Science 256:193-197.

Grant, P. R. and B. R. Grant. 2016. Introgressive hybridization and natural selection in Darwin's finches. Biological Journal of the Linnean Society 117:812-822.

Green, R. E., J. Krause, A. W. Briggs, T. Maricic, U. Stenzel, M. Kircher, N. Patterson, H. Li, W. Zhai, and M. H.-Y. Fritz. 2010. A draft sequence of the Neandertal genome. Science 328:710-722.

Guindon, S., J. F. Dufayard, V. Lefort, M. Anisimova, W. Hordijk, and O. Gascuel. 2010. New algorithms and methods to estimate Maximum-Likelihood phylogenies: Assessing the performance of PhyML 3.0. Systematic Biology 59:307-321.

Harrison, R. G. and E. L. Larson. 2014. Hybridization, introgression, and the nature of

species boundaries. Journal of Heredity 105:795-809.

Hurwood, D., E. Adamson, and P. Mather. 2008. Evidence for strong genetic structure in a regionally important, highly vagile cyprinid (Henicorhynchus lobatus) in the Mekong River Basin. Ecology of Freshwater Fish 17:273-283.

Irisarri, I., P. Singh, S. Koblmuller, J. Torres-Dowdall, F. Henning, P. Franchini, C. Fischer, A. R. Lemmon, E. M. Lemmon, G. G. Thallinger, C. Sturmbauer, and A. Meyer. 2018. Phylogenomics uncovers early hybridization and adaptive loci shaping the radiation of Lake Tanganyika cichlid fishes. Nature Communications 9: 3159.

Isaac, N. J., J. Mallet, and G. M. Mace. 2004. Taxonomic inflation: its influence on macroecology and conservation. Trends in Ecology & Evolution 19:464-469.

Ivory, S. J., M. W. Blome, J. W. King, M. M. McGlue, J. E. Cole, and A. S. Cohen. 2016. Environmental change explains cichlid adaptive radiation at Lake Malawi over the past 1.2 million years. Proceedings of the National Academy of Sciences 113:11895-11900.

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

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

Knight, M. E., G. F. Turner, C. Rico, M. J. H. van Oppen, and G. M. Hewitt. 1998. Microsatellite paternity analysis on captive Lake Malawi cichlids supports reproductive isolation by direct mate choice. Molecular Ecology 7:1605-1610.

Koblmüller, S., R. C. Albertson, M. J. Genner, K. M. Sefc, and T. Takahashi. 2012.

Cichlid Evolution: Lessons in Diversification 2012. International Journal of Evolutionary Biology 2012.

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

Kornfield, I. 1978. Evidence for rapid speciation in African cichlid fishes. Experientia 34:335-336.

Kornfield, I. and P. F. Smith. 2000. African cichlid fishes: model systems for evolutionary biology. Annual Review of Ecology and Systematics 31:163-196.

Lamichhaney, S., J. Berglund, M. S. Almén, K. Maqbool, M. Grabherr, A. Martinez-Barrio, M. Promerová, C.-J. Rubin, C. Wang, and N. Zamani. 2015. Evolution of Darwin's finches and their beaks revealed by genome sequencing. Nature 518:371-375.

Lescak, E. A., S. L. Bassham, J. Catchen, O. Gelmond, M. L. Sherbick, F. A. von Hippel, and W. A. Cresko. 2015. Evolution of stickleback in 50 years on earthquake-uplifted islands. Proceedings of the National Academy of Sciences of the United States of America 112:E7204-E7212.

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

Loh, Y.-H. E., L. S. Katz, M. C. Mims, T. D. Kocher, V. Y. Soojin, and J. T. Streelman. 2008. Comparative analysis reveals signatures of differentiation amid genomic polymorphism in Lake Malawi cichlids. Genome Biology 9:R113.

Loh, Y. H. E., E. Bezault, F. M. Muenzel, R. B. Roberts, R. Swofford, M. Barluenga, C. E. Kidd, A. E. Howe, F. Di Palma, K. Lindblad-Toh, J. Hey, O. Seehausen, W. Salzburger, T. D. Kocher, and J. T. Streelman. 2013. Origins of shared genetic variation in African cichlids. Molecular Biology and Evolution 30:906-917.

Losos, J. B. 2007. Detective work in the West Indies: integrating historical and experimental approaches to study island lizard evolution. AIBS Bulletin 57:585-597.

Losos, J. B., T. R. Jackman, A. Larson, K. de Queiroz, and L. Rodríguez-Schettino. 1998. Contingency and determinism in replicated adaptive radiations of island lizards. Science 279:2115-2118.

Lu, G. and L. Bernatchez. 1999. Correlated trophic specialization and genetic divergence in sympatric lake whitefish ecotypes (Coregonus clupeaformis): support for the ecological speciation hypothesis. Evolution 53:1491-1505.

Malinsky, M., R. J. Challis, A. M. Tyers, S. Schiffels, Y. Terai, B. P. Ngatunga, E. A. Miska, R. Durbin, M. J. Genner, and G. F. Turner. 2015. Genomic islands of speciation separate cichlid ecomorphs in an East African crater lake. Science 350:1493-1498.

Malinsky, M. and W. Salzburger. 2016. Environmental context for understanding the iconic adaptive radiation of cichlid fishes in Lake Malawi. Proceedings of the National Academy of Sciences 113:11654-11656.

Malinsky, M., H. Svardal, A. M. Tyers, E. A. Miska, M. J. Genner, G. F. Turner, and R. Durbin. 2018a. Whole genome sequences of Malawi cichlids reveal multiple radiations interconnected by gene flow. Nature Ecology & Evolution 2: 1940–1955

Malinsky, M., E. Trucchi, D. J. Lawson, and D. Falush. 2018b. RADpainter and fineRADstructure: population inference from RADseq data. Molecular Biology and Evolution 35:1284-1290.

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

Mallet, J., M. Beltran, W. Neukirchen, and M. Linares. 2007. Natural hybridization in heliconiine butterflies: the species boundary as a continuum. BMC Evolutionary Biology 7: 28.

Martin, M. 2011. Cutadapt removes adapter sequences from high-throughput sequencing reads. EMBnet.journal 17: 10-12.

Mayr, E. 1942. Systematics and the origin of species, from the viewpoint of a zoologist. Cambridge: Harvard University Press.

Mayr, E. 1970. Populations, species, and evolution: an abridgment of animal species and evolution. Harvard University Press.

Mayr, E. 1982. The growth of biological thought: Diversity, evolution, and inheritance. Harvard University Press.

Mayr, E. 1999. Systematics and the origin of species, from the viewpoint of a zoologist. Harvard University Press.

McKaye, K. R. and N. McKaye. 1977. Communal care and kidnapping of young by parental cichlids. Evolution 31:674-681.

Meier, J. I., D. A. Marques, S. Mwaiko, C. E. Wagner, L. Excoffier, and O. Seehausen. 2017. Ancient hybridization fuels rapid cichlid fish adaptive radiations. Nature Communications 8:14363.

Meyer, A. 1993. Phylogenetic relationships and evolutionary processes in East African cichlid fishes. Trends in Ecology & Evolution 8:279-284.

Nei, M. 1978. Estimation of average heterozygosity and genetic distance from a small number of individuals. Genetics 89:583-590.

Nelson, J. S., T. C. Grande, and M. V. Wilson. 2016. Fishes of the World. John Wiley & Sons.

Nichols, P., M. J. Genner, C. Van Oosterhout, A. Smith, P. Parsons, H. Sungani, J.

Swanstrom, and D. A. Joyce. 2015. Secondary contact seeds phenotypic novelty in cichlid fishes. Proceedings of the Royal Society of London B: Biological Sciences 282:20142272.

Nicholson, K. E., B. I. Crother, C. Guyer, and J. M. Savage. 2012. It is time for a new classification of anoles (Squamata: Dactyloidae). Zootaxa 3477:108.

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.

Oppenheimer, J. R. 1970. Mouthbreeding in fishes. Animal Behaviour 18:493-503.

Owen, R. B., R. Crossley, T. C. Johnson, D. Tweddle, I. Kornfield, S. Davison, D. H. Eccles, and D. E. Engstrom. 1990. Major low levels of Lake Malawi and their implications for speciation rates in cichlid fishes. Proceedings of the Royal Society Series B - Biological Sciences 240:519-553.

Peterson, B. K., J. N. Weber, E. H. Kay, H. S. Fisher, and H. E. Hoekstra. 2012. Double Digest RADseq: An Inexpensive Method for De Novo SNP Discovery and Genotyping in Model and Non-Model Species. PloS One 7: e37135.

Petren, K., B. R. Grant, and P. R. Grant. 1999. A phylogeny of Darwin's finches based on microsatellite DNA length variation. Proceedings of the Royal Society of London B: Biological Sciences 266:321-329.

Price, T. D. and M. M. Bouvier. 2002. The evolution of F1 postzygotic incompatibilities in birds. Evolution 56:2083-2089.

Ribbink, A. J., A. C. Marsh, and B. A. Marsh. 1981. Nest-building and communal care of young by Tilapia rendalli dumeril (Pisces, Cichlidae) in Lake Malawi. Environmental Biology of Fishes 6:219-222.

Roberts, R. B., J. R. Ser, and T. D. Kocher. 2009. Sexual conflict resolved by invasion of a novel sex determiner in Lake Malawi cichlid fishes. Science 326:998-1001.

Roques, S., J. M. Sevigny, and L. Bernatchez. 2001. Evidence for broadscale introgressive hybridization between two redfish (genus Sebastes) in the North-west Atlantic: a rare marine example. Molecular Ecology 10:149-165.

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

Salzburger, W., T. Mack, E. Verheyen, and A. Meyer. 2005. Out of Tanganyika: genesis, explosive speciation, key-innovations and phylogeography of the haplochromine cichlid fishes. BMC Evolutionary Biology 5:17.

Salzburger, W., A. Meyer, S. Baric, E. Verheyen, and C. Sturmbauer. 2002b. 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.

Schluter, D. 1996. Ecological causes of adaptive radiation. American Naturalist 148:S40-S64.

Schluter, D. 2000. The ecology of adaptive radiation. OUP Oxford.

Schluter, D. 2001. Ecology and the origin of species. Trends in Ecology & Evolution 16:372-380.

Schluter, D. and G. L. Conte. 2009. Genetics and ecological speciation. Proceedings of the National Academy of Sciences of the United States of America 106:9955-9962.

Schluter, D. and L. M. Nagel. 1995. Parallel speciation by natural selection. The American Naturalist 146:292-301.

Schwarzer, J., E. R. Swartz, E. Vreven, J. Snoeks, F. P. D. Cotterill, B. Misof, and U. K. Schliewen. 2012. Repeated trans-watershed hybridization among haplochromine cichlids (Cichlidae) was triggered by Neogene landscape evolution. Proceedings of the Royal Society B - Biological Sciences 279:4389-4398.

Seegers, L. 1996. The fishes of the Lake Rukwa drainage. Musee Royal de l’Afrique

Central, Tervuren, Belgique, Sciences Zoologique 258: 1- 396.

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

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

Seehausen, O., E. Koetsier, M. V. Schneider, L. J. Chapman, C. A. Chapman, M. E. Knight, G. F. Turner, J. J. M. van Alphen, and R. Bills. 2003. Nuclear markers reveal unexpected genetic variation and a Congolese-Nilotic origin of the Lake Victoria cichlid species flock. Proceedings of the Royal Society B - Biological Sciences 270:2637-2638.

Seehausen, O., Y. Terai, I. S. Magalhaes, K. L. Carleton, H. D. Mrosso, R. Miyagi, I. Van Der Sluijs, M. V. Schneider, M. E. Maan, and H. Tachida. 2008. Speciation through sensory drive in cichlid fish. Nature 455:620-626.

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

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

Simpson, G. G. 1955. Major features of evolution. Columbia University Press: New York.

Slatkin, M. 1995. A measure of population subdivision based on microsatellite allele frequencies. Genetics 139:457-462.

Smith, P. F., A. Konings, and I. Kornfield. 2003. Hybrid origin of a cichlid population in Lake Malawi: implications for genetic variation and species diversity.

Molecular Ecology 12:2497-2504.

Song, Y., S. Endepols, N. Klemann, D. Richter, F.-R. Matuschka, C.-H. Shih, M. W. Nachman, and M. H. Kohn. 2011. Adaptive introgression of anticoagulant rodent poison resistance by hybridization between old world mice. Current Biology 21:1296-1301.

Stelkens, R. B., K. A. Young, and O. Seehausen. 2010. The accumulation of reproductive incompatibilities in African cichlid fish. Evolution 64:617-633.

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

Sturmbauer, C., W. Salzburger, N. Duftner, R. Schelly, and S. Koblmuller. 2010. Evolutionary history of the Lake Tanganyika cichlid tribe Lamprologini (Teleostei: Perciformes) derived from mitochondrial and nuclear DNA data. Molecular Phylogenetics and Evolution 57:266-284.

Thomas, D. S. G., R. Bailey, P. A. Shaw, J. A. Durcan, and J. S. Singarayer. 2009. Late Quaternary highstands at Lake Chilwa, Malawi: Frequency, timing and possible forcing mechanisms in the last 44 ka. Quaternary Science Reviews 28:526-539.

Trauth, M. H., M. A. Maslin, A. L. Deino, A. Junginger, M. Lesoloyia, E. O. Odada, D. O. Olago, L. A. Olaka, M. R. Strecker, and R. Tiedemann. 2010. Human evolution in a variable environment: the amplifier lakes of Eastern Africa. Quaternary Science Reviews 29:2981-2988.

Trauth, M. H., M. A. Maslin, A. L. Deino, M. R. Strecker, A. G. N. Bergner, and M. Duhnforth. 2007. High- and low-latitude forcing of Plio-Pleistocene East African climate and human evolution. Journal of Human Evolution 53:475-486.

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

10:793-806.

Verheyen, E., R. Blust, and W. Decleir. 1994. Metabolic rate, hypoxia tolerance and aquatic surface respiration of some lacustrine and riverine African cichlid fishes (Pisces: Cichlidae). Comparative Biochemistry and Physiology Part A: Physiology 107:403-411.

Appendices

Table S1. GPS coordinates of collection sites.

Collected site	Latitude	Longitude
Bua River	-13.3084	33.5477
Burungi	-3.91563	35.8608
Chemka Hot Spring	-3.44433	37.1937
Igogo dam	-4.28855	33.7885
Kagera River	-4.90983	29.6847
Lake Chidya	-10.5972	40.1554
Lake Chilwa	-15.3719	35.5891
Lake Chilwa	-15.3716	35.5902
Lake Chipwa	-4.85285	29.7331
Lake Chiuta	-14.7250	35.7833
Lake Igombe	-4.85405	32.7451
Lake Mansi	-7.28230	39.0815
Lwiche River	-7.94215	31.5960
Makobe island	-2.36566	32.9226
Malagarasi River	-5.21151	29.8423
Malagarasi River	-5.09390	30.8485
Metangula market	-12.6896	34.8104
Milaca dam	-6.32324	31.0494
Mindu dam	-6.86467	37.6083
Minos reef	-12.8000	34.7800
Mtera Dam	-7.10469	35.8267
Mwamapuli dam	-4.35630	33.8767
Rufiji River	-7.99086	38.7493
Ruva River	-3.52920	37.5731
Shire River	-14.4774	35.2744

Table S2. Scripts for analysis programming.

Demultiplex PCR index and adapter barcode	**package: cutadapt**

```
>>>cutadapt --no-trim -a file:PCRIndex.txt -o trimmed-{name}.1 -p trimmed-{name}.2 ddRAD_S1_L001_R1_001.fastq ddRAD_S1_L001_R2_001.fastq #demultiplex by PCR indices
>>>cutadapt -a CGAGATCGGAAGAGC -o Index7_trim-{name}_R1_.fastq -p Index7_trim-{name}_R2_.fastq trimmed-PCR2_Idx_7_CAGATC.1.fastq trimmed-PCR2_Idx_7_CAGATC.2.fastq #trim PCR indices and associated adaptor junks >>>cutadapt -a file:AdaptorIndex.txt -o Index7_R1_-{name}.fastq -p Index7_R2_-{name}.fastq Index7_trim-1_R2_.fastq Index7_trim-1_R1_.fastq #demultiplex ligated adapter barcode and additionally trim off
```

Assembly ddRAD dataset	**package: ipyrad**

```
>>>ipyrad -n para42 #create a parameter file

------- ipyrad params file (v.0.7.23)-------------------------------
para42 ##[0][assembly_name]: Assembly name. Used to name output directories for assembly steps
./ ##[1][project_dir]: Project dir (made in curdir if not present)
##[2][raw_fastq_path]: Location of raw non-demultiplexed fastq files
##[3][barcodes_path]: Location of barcodes file
/Users/lh17359/Desktop/DEMULTIPLEX/TEST/*.fastq ##[4][sorted_fastq_path]: Location of demultiplexed/sorted fastq files
Reference ##[5][assembly_method]: Assembly method
/Users/lh17359/Desktop/DEMULTIPLEX/REF/M_zebra_UMD2a.fna
```

##[6][reference_sequence]: Location of reference sequence file

pairddrad ##[7][datatype]: Datatype (see docs): rad, gbs, ddrad, etc.

##[8][restriction_overhang]: Restriction overhang (cut1,) or (cut1, cut2)

5 ##[9][max_low_qual_bases]: Max low quality base calls (Q<20) in a read

33 ##[10][phred_Qscore_offset]: phred Q score offset (33 is default and very standard)

5 ##[11][mindepth_statistical]: Min depth for statistical base calling

3 ##[12][mindepth_majrule]: Min depth for majority-rule base calling

10000 ##[13][maxdepth]: Max cluster depth within samples

0.85 ##[14][clust_threshold]: Clustering threshold for de novo assembly

1 ##[15][max_barcode_mismatch]: Max number of allowable mismatches in barcodes

1 ##[16][filter_adapters]: Filter for adapters/primers (1 or 2=stricter)

35 ##[17][filter_min_trim_len]: Min length of reads after adapter trim

2 ##[18][max_alleles_consens]: Max alleles per site in consensus sequences

4, 4 ##[19][max_Ns_consens]: Max N's (uncalled bases) in consensus (R1, R2)

8, 8 ##[20][max_Hs_consens]: Max Hs (heterozygotes) in consensus (R1,

```
R2)
42 ##[21][min_samples_locus]: Min # samples per locus for output
20, 20 ##[22][max_SNPs_locus]: Max # SNPs per locus (R1, R2)
8, 8 ##[23][max_Indels_locus]: Max # of indels per locus (R1, R2)
0.5 ##[24][max_shared_Hs_locus]: Max # heterozygous sites per locus
(R1, R2)
0, 0, 5, 0 ##[25][trim_reads]: Trim raw read edges (R1>, <R1, R2>,
<R2)
0, 0, 0, 0 ##[26][trim_loci]: Trim locus edges (see docs) (R1>, <R1,
R2>, <R2)
p,s,n,v ##[27][output_formats]: Output formats (see docs)
##[28][pop_assign_file]: Path to population assignment file
------------------------------------------------------------------


>>>ipyrad -p params-para42.txt -s 1234567 #filter and cluster reads,
call the consensus of sequences within each cluster and cluster them
across samples, write output files in supported format
```

Illustrate population structure **package: fineRADstructure**

```
>>>python finerad_input.py --input para42.loci --minsample 2
#convert .loci matrix from ipyrad and exclude loci with less than 2
samples
>>>RADpainter paint para42.alleles.loci.min2.finerad #calculate co-
ancestry matrix
>>>finestructure -x 100000 -y 100000 -z 1000
para42.allele.loci.min2_chunks.out para42.allele.loci_chunks.mcmc
```

```
#assign individuals to populations
>>>finestructure –m T -x 10000 para42.allele.loci.min2_chunks.out para42.allele.loci_chunks.mcmc para42.allele.loci_chunks.mcmcTree
#build phylogeny
#Plotting output results, using R scripts released online. Source: http://cichlid.gurdon.cam.ac.uk/fineRADstructurePlot.R
```

Calculate genetic distance **program: R**

```
>>>library(devtools)
>>>library(adegenet)
>>>library(ape)
>>>obj1 <- read.structure("/Users/lh17359/Desktop/para42.str")
>>>obj2 <- import2genind("/Users/lh17359/Desktop/para42.str")
>>>all.equal(obj1,obj2)
#assign individuals to populations
>>>pop(obj1)<-
as.factor(rep(c('paLwi','kaMil','twMin','RuaBl','ManBl','paLwi','kaMil','RuaBl','twMan','Rufij','twMan','caBua','Rufij','cadya','Troph','chilo','sauva','omniM','chipw','blMal','stMal','blMal','stMal','blIgo','stIgo','blIgo','stIgo','blogo','blMwa','stMwa','Troph','Cteno','blBur','Cteno','twuta','cauta','twlwa','Rlong','Otosp','Mayze','calwa'),c(3,3,1,2,2,1,1,1,2,1,1,1,3,4,2,1,1,1,4,1,1,1,1,2,1,1,4,4,4,3,1,3,3,2,3,3,6,1,1,1,1)))
>>>library(pegas)
>>>library(hierfstat)
#calculate Nei's standard genetic distance
```

```
gpee <- genind2genpop(obj1)

Dgen <- dist.genpop(gpee,method=1)

>>>library(zvau)

>>>library(genepop)

#calculate Slatkin’s linearized pairwise FST values

>>>writeGenPop(obj1,file.name ="Fst.txt", comment = "")

>>>Fst("/Users/lh17359/Desktop/gene/Fst.txt", sizes = FALSE, pairs =
TRUE, dataType = "Diploid")

#identify each value to the four groups “ADG”, “ACG”, “ACS” and “SCG”
and create a new file with two columns “Group” and “genetic distance
value” before doing following steps

>>>distd<-read.csv(file.choose(),header=T)

>>>Nei<-distd$Nei

>>>Fst<-distd$Fst

>>>group<-distd$Group

#test if samples are from populations with equal variances in both
dataset

>>>bartlett.test(Nei,group)

>>>bartlett.test(Fst,group)

#perform ANOVA and Tukey’s HSD test on both dataset

>>>anovNei<-aov(Nei~group)

>>>summary(anovNei)

>>>TukeyHSD(anovNei)

>>>anovFst<-aov(Fst~group)

>>>summary(anovFst)
```

```
>>>TukeyHSD(anovFst)
```

One-way permutation test program: R

```
library(coin)
library(FSA)
library(rcompanion)
library(multcompView)
distd<-read.csv(file.choose(),header=T) #read genetic distance data
distd$Group = factor(distd$Group,
                  ordered=FALSE,
                  levels=unique(distd$Group)) #order groups
boxplot(Nei.s ~ Group, data = distd) #draw a boxplot of Nei's standard
genetic distance
boxplot(Fst ~ Group, data = distd) #draw a boxplot of Slatkin's
linearized pairwise FST values
independence_test(Nei.s ~ Group, data = distd) #permutation test
using Nei's standard genetic distance
independence_test(Fst ~ Group, data = distd) #permutation test using
Slatkin's linearized pairwise FST values
distd$Group = factor(distd$Group, levels = c("ADG", "ACG", "SCG",
"ACS")) #order groups by median (compare by boxplot)
NeiPT = pairwisePermutationTest(Nei.s ~ Group, data = distd,
method="fdr") #pairwise permutation test using Nei's standard genetic
distance
NeiPT #show result
cldList(p.adjust ~ Comparison, data = NeiPT, threshold = 0.05) #show
```

result displayed by letters

FstPT = pairwisePermutationTest(Fst ~ Group, data = distd, method="fdr") #pairwise permutation test using Slatkin’s linearized pairwise F_{ST} values

FstPT #show result

cldList(p.adjust ~ Comparison, data = FstPT, threshold = 0.05) #show result displayed by letters

NeiPM = pairwisePermutationMatrix(Nei.s ~ Group, data = distd, method="fdr") #pairwise test matrix using Nei’s standard genetic distance

NeiPM #show result

multcompLetters(NeiPM$Adjusted, compare="<", threshold=0.05, Letters=letters, reversed = FALSE) # Compact letter display output with pairwise permutation matrix

FstPM = pairwisePermutationMatrix(Fst ~ Group, data = distd, method="fdr") #pairwise test matrix using Slatkin’s linearized pairwise F_{ST} values

FstPM #show result

multcompLetters(FstPM$Adjusted, compare="<", threshold=0.05, Letters=letters, reversed = FALSE) # Compact letter display output with pairwise permutation matrix

Table S3. Adapter barcodes and PCR indices.

Adapter stock barcode	
GCATG_EcoRI_1.1	5'- ACACTCTTTCCCTACACGACGCTCTTCCGATCTGCATG-3'
AACCA_EcoRI_1.1	5'- ACACTCTTTCCCTACACGACGCTCTTCCGATCTAACCA-3'
CGATC_EcoRI_1.1	5'- ACACTCTTTCCCTACACGACGCTCTTCCGATCTCGATC -3'
TCGAT_EcoRI_1.1	5'- ACACTCTTTCCCTACACGACGCTCTTCCGATCTTCGAT -3'
TGCAT_EcoRI_1.1	5'- ACACTCTTTCCCTACACGACGCTCTTCCGATCTTGCAT -3'
CAACC_EcoRI_1.1	5'- ACACTCTTTCCCTACACGACGCTCTTCCGATCTCAACC -3'
GGTTG_EcoRI_1.1	5'- ACACTCTTTCCCTACACGACGCTCTTCCGATCTGGTTG -3'
AAGGA_EcoRI_1.1	5'- ACACTCTTTCCCTACACGACGCTCTTCCGATCTAAGGA -3'
GCATG_EcoRI_1.2	5'-[PHO]AATTCATGCAGATCGGAAGAGCGTCGTGTAGGGAAAGAGTGT -3'
AACCA_EcoRI_1.2	5'-[PHO]AATTTGGTTAGATCGGAAGAGCGTCGTGTAGGGAAAGAGTGT-3'
CGATC_EcoRI_1.2	5'-[PHO]AATTGATCGAGATCGGAAGAGCGTCGTGTAGGGAAAGAGTGT-3'
TCGAT_EcoRI_1.2	5'-[PHO]AATTATCGAAGATCGGAAGAGCGTCGTGTAGGGAAAGAGTGT-3'
TGCAT_EcoRI_1.2	5'-[PHO]AATTATGCAAGATCGGAAGAGCGTCGTGTAGGGAAAGAGTGT-3'
CAACC_EcoRI_1.2	5'-[PHO]AATTGGTTGAGATCGGAAGAGCGTCGTGTAGGGAAAGAGTGT-3'
GGTTG_EcoRI_1.2	5'-[PHO]AATTCAACCAGATCGGAAGAGCGTCGTGTAGGGAAAGAGTGT-3'
AAGGA_EcoRI_1.2	5'-[PHO]AATTTCCTTAGATCGGAAGAGCGTCGTGTAGGGAAAGAGTGT-3'
MspI_2.1	5'-GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCT-3'
MspI_2.2	5'-[PHO]CGAGATCGGAAGAGCGAGAACAA-3'
Primer Index	
PCR1	5'-AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACG-3'
PCR2_Idx_1	5'-CAAGCAGAAGACGGCATACGAGATCGTGATGTGACTGGAGTTCAGACGTGTGC-3'
PCR2_Idx_2	5'-CAAGCAGAAGACGGCATACGAGATACATCGGTGACTGGAGTTCAGACGTGTGC-3'
PCR2_Idx_3	5'-CAAGCAGAAGACGGCATACGAGATGCCTAAGTGACTGGAGTTCAGACGTGTGC-3'
PCR2_Idx_4	5'-CAAGCAGAAGACGGCATACGAGATTGGTCAGTGACTGGAGTTCAGACGTGTGC-3'
PCR2_Idx_5	5'-CAAGCAGAAGACGGCATACGAGATCACTGTGTGACTGGAGTTCAGACGTGTGC-3'
PCR2_Idx_6	5'-CAAGCAGAAGACGGCATACGAGATATTGGCGTGACTGGAGTTCAGACGTGTGC-3'
PCR2_Idx_7	5'-CAAGCAGAAGACGGCATACGAGATGATCTGGTGACTGGAGTTCAGACGTGTGC-3'
PCR2_Idx_8	5'-CAAGCAGAAGACGGCATACGAGATTCAAGTGTGACTGGAGTTCAGACGTGTGC-3'
PCR2_Idx_9	5'-CAAGCAGAAGACGGCATACGAGATCTGATCGTGACTGGAGTTCAGACGTGTGC-3'
PCR2_Idx_10	5'-CAAGCAGAAGACGGCATACGAGATAAGCTAGTGACTGGAGTTCAGACGTGTGC-3'
PCR2_Idx_11	5'-CAAGCAGAAGACGGCATACGAGATGTAGCCGTGACTGGAGTTCAGACGTGTGC-3'
PCR2_Idx_12	5'-CAAGCAGAAGACGGCATACGAGATTACAAGGTGACTGGAGTTCAGACGTGTGC-3'

Table S4. Deleted samples filtered by low quality.

Species	Collected sites	Total amounts
Astatotilapia gigliolii	Mindu Dam (3)	5
	Lake Mansi (1)	
	Lake Chiuta (1)	
Astatotilapia sp. "Ruaha Blue"	Mtera Dam (1)	1
Astatotilapia calliptera	Bua River (3)	5
	Lake Chiuta (1)	
	Lake Chilwa (1)	
Astatotilapia stappersi	Mwamapuli dam (1)	1
Ctenochromis pectoralis	Chemka Hot Spring (1)	1

Table S5. Percentage of reads mapped against the reference genome *Metriaclima zebra*. Abbreviations: M: reference sequence mapped reads. U: reference sequence unmapped reads. M+U: total reads used in mapping against reference genome.

ID	reads passed filter	M	U	M+U	M/(M+U) (%)
LL01	53771	39049	2554	41603	93.86
LL02	64981	46414	3142	49556	93.66
LL03	84797	58242	3851	62093	93.80
LL04	202103	104415	4938	109353	95.48
LL05	191264	101414	5589	107003	94.78
LL06	105352	63938	3209	67147	95.22
LL09	67274	45604	2615	48219	94.58
LL10	34407	25921	740	26661	97.22
LL11	31759	17523	9324	26847	65.27
LL13	64115	45967	3051	49018	93.78
LL14	244469	123884	6385	130269	95.10
LL15	140401	82276	4816	87092	94.47
LL16	258902	143759	19431	163190	88.09
LL17	153572	88260	6837	95097	92.81
LL18	36513	28642	2594	31236	91.70
LL19	148589	81585	5750	87335	93.42
LL20	111028	64849	3733	68582	94.56
LL21	181945	96065	7191	103256	93.04
LL22	181118	94058	5463	99521	94.51
LL25	118476	74234	3815	78049	95.11
LL28	61596	42504	6540	49044	86.67
LL29	71819	46679	19456	66135	70.58
LL30	92328	59993	4896	64889	92.45
LL31	66418	46718	2561	49279	94.80
LL32	104021	65812	4374	70186	93.77
LL33	147364	84309	25776	110085	76.59
LL34	123612	74713	2342	77055	96.96
LL35	83917	38127	16018	54145	70.42
LL36	127687	72774	4754	77528	93.87
LL37	190008	102626	7191	109817	93.45
LL38	262121	125378	7025	132403	94.69

LL39	102336	65905	4284	70189	93.90
LL40	100708	61778	3001	64779	95.37
LL41	108357	65573	3319	68892	95.18
LL42	101362	61679	3192	64871	95.08
LL43	78822	50870	2674	53544	95.01
LL44	102229	65379	3385	68764	95.08
LL45	82402	52879	3086	55965	94.49
LL46	138100	78116	3930	82046	95.21
LL47	94209	41791	18041	59832	69.85
LL48	106003	65121	4036	69157	94.16
LL49	329250	151481	9883	161364	93.88
LL50	524930	223096	20427	243523	91.61
LL51	172433	92442	5117	97559	94.75
LL52	150064	83291	4653	87944	94.71
LL53	235124	115666	8829	124495	92.91
LL54	117423	68844	3344	72188	95.37
LL55	104391	62642	3005	65647	95.42
LL56	30556	23489	1510	24999	93.96
LL57	159154	82164	4074	86238	95.28
LL58	151877	80465	3994	84459	95.27
LL59	112019	44170	18530	62700	70.45
LL60	184590	102192	11785	113977	89.66
LL61	343270	159975	9196	169171	94.56
LL62	229390	126884	9480	136364	93.05
LL63	298872	147602	9367	156969	94.03
LL64	179983	101400	6375	107775	94.08
LL65	140404	79098	3911	83009	95.29
LL67	56204	40564	2597	43161	93.98
LL68	57072	40628	3369	43997	92.34
LL69	133958	73419	7271	80690	90.99
LL70	140371	75939	8010	83949	90.46
LL71	140449	49515	28260	77775	63.66
LL72	99494	58742	2835	61577	95.40
LL73	350684	151283	8152	159435	94.89
LL74	200042	99487	5206	104693	95.03
LL75	111487	63967	6306	70273	91.03

LL77	89512	53833	5249	59082	91.12
LL78	59154	43842	3022	46864	93.55
LL80	80150	53745	5477	59222	90.75
LL81	105694	68647	10175	78822	87.09
LL83	130204	51379	23381	74760	68.73
LL84	88594	58379	2347	60726	96.14
LL85	433688	185345	11633	196978	94.09
LL86	441481	181366	11254	192620	94.16
LL87	166421	86590	5335	91925	94.20
LL88	217387	103097	5813	108910	94.66
LL89	175539	91527	4826	96353	94.99
LL90	185273	92463	5375	97838	94.51
LL91	136596	73956	4154	78110	94.68
LL92	157413	83507	3188	86695	96.32
LL93	154165	87999	5958	93957	93.66
LL94	154148	83792	2152	85944	97.50
LL96	154853	80098	2482	82580	96.99

Table S6. Pairwise genetic distance for all sympatric and allopatric populations (Next page). Values were calculated by Nei's standard genetic distance in the lower diagonal, and by Slatkin's linearized pairwise F_{ST} in the upper one. Column was shown in different colours indicating the groups for ANOVA and Tukey's HSD analysis. Yellow represents "ACS", blue represents "SCG", grey represents "ADG", and no colour means "ACG". Abbreviations: paLwi: *A. sp.* "pseudopaludinosus" from Lwiche river. kaMil: *A. katavi* from Milaca dam. twMin: *A. gigliolii* from Mindu dam. RuaBl: *A. sp.* "Ruaha Blue". ManBl: *A. sp.* "Rufiji Blue" from Lake Mansi. twMan: *A. gigliolii* from Lake Mansi. Rufij: *A. sp.* "Rufiji Blue" from Rufiji river. caBua: *A. calliptera* from Bua river. cadya: *A. calliptera* from Lake Chidya. Troph: Tropheini. chilo: *Paralabidochromis chilotes.* sauva: *Paralabidochromis sauvagei.* omniM: *Neochromis omnicaeruleus*. chipw: *A. sp.* "Chipwa" from Lake Chipwa. blMal: *A. bloyeti* from Malagarasi river. stMal: *A. stappersi* from Malagarasi river. blIgo: *A. bloyeti* from Lake Igombe. stIgo: *A. stappersi* from Lake Igombe. blogo: *A. bloyeti* from Igogo dam. blMwa: *A. bloyeti* from Mwamapuli dam. stMwa: *A. stappersi* from Mwamapuli dam. Cteno: *Ctenochromis pectoralis*. blBur: *A. bloyeti* from Burungi. twuta: *A. gigliolii* from Lake Chiuta. cauta: *A. calliptera* from Lake Chiuta. twlwa: *A. gigliolii* from Lake Chilwa. Rlong: *Rhamphochromis longiceps*. Otosp: *Otopharynx speciosus.* Mayze: *Maylandia zebra.* calwa: *A. calliptera* from Lake Chilwa.

Pop.	paLwi	kaMil	twMin	RuaBl	ManBl	twMan	Rufij	caBua	cadya	chipw	blMal	stMal	blIgo
paLwi		0.4616	0.853	0.8581	0.5415	0.8988	0.4602	0.8442	0.8279	0.4139	0.5177	0.3856	0.5839
kaMil	0.0315		0.9202	0.8911	0.715	0.9283	0.6458	0.8912	0.8659	0.5979	0.6723	0.6324	0.7084
twMin	0.0923	0.0990		0.9194	0.9703	0.7795	0.9215	0.9956	0.9074	0.8687	0.8751	0.9196	0.875
RuaBl	0.1086	0.1156	0.0965		0.9356	0.9283	0.8669	0.9272	0.8736	0.8507	0.8601	0.8927	0.8688
ManBl	0.0403	0.0459	0.0933	0.1122		0.9522	0.3181	0.94	0.8751	0.5153	0.6756	0.5442	0.696
twMan	0.1811	0.1911	0.0843	0.1527	0.1820		0.9294	0.9712	0.9174	0.9011	0.9007	0.9186	0.9003
Rufij	0.0376	0.0427	0.0944	0.1120	0.0220	0.1841		0.8642	0.8497	0.4653	0.6184	0.4373	0.6522
caBua	0.0964	0.1045	0.0897	0.1010	0.0988	0.1595	0.0986		0.4943	0.8447	0.8314	0.8631	0.8095
cadya	0.1260	0.1343	0.1098	0.1178	0.1283	0.1791	0.1286	0.0406		0.8359	0.8177	0.845	0.8231
chipw	0.0353	0.0399	0.0925	0.1086	0.0372	0.1853	0.0348	0.0983	0.1285		0.567	0.1915	0.5955
blMal	0.0530	0.0587	0.0918	0.1092	0.0590	0.1806	0.0579	0.0939	0.1220	0.0537		0.5701	0.2134
stMal	0.0356	0.0417	0.0846	0.1014	0.0376	0.1715	0.0351	0.0892	0.1176	0.0261	0.0520		0.5952
blIgo	0.0563	0.0618	0.0959	0.1125	0.0624	0.1878	0.0604	0.0994	0.1293	0.0542	0.0277	0.0532	
stIgo	0.0311	0.0356	0.0897	0.1065	0.0312	0.1831	0.0290	0.0956	0.1267	0.0186	0.0500	0.0210	0.0505
blogo	0.0615	0.0663	0.1006	0.1168	0.0672	0.1926	0.0648	0.1042	0.1340	0.0598	0.0315	0.0574	0.0201
blMwa	0.0581	0.0640	0.0994	0.1151	0.0643	0.1906	0.0621	0.1022	0.1329	0.0570	0.0282	0.0554	0.0177
stMwa	0.0390	0.0421	0.0932	0.1109	0.0409	0.1861	0.0390	0.0981	0.1286	0.0301	0.0555	0.0317	0.0561
Cteno	0.2908	0.2992	0.2674	0.2865	0.2924	0.3330	0.2914	0.2735	0.2960	0.2907	0.2877	0.2793	0.2959
blBur	0.0651	0.0700	0.1100	0.1259	0.0735	0.1989	0.0698	0.1139	0.1426	0.0641	0.0631	0.0638	0.0645
twuta	0.1443	0.1531	0.0670	0.1268	0.1449	0.0611	0.1462	0.1284	0.1488	0.1456	0.1435	0.1350	0.1494
cauta	0.1251	0.1336	0.1078	0.1165	0.1268	0.1778	0.1276	0.0420	0.0186	0.1274	0.1208	0.1155	0.1283

Pop.	stIgo	blogo	blMwa	stMwa	Cteno	blBur	twuta	cauta	twlwa	calwa
paLwi	0.3049	0.6583	0.6339	0.4282	0.9017	0.7754	0.8988	0.8504	0.9333	0.8155
kaMil	0.4993	0.7587	0.7399	0.6025	0.9191	0.8557	0.9243	0.8898	0.9532	0.8753
twMin	0.8025	0.8953	0.8759	0.8897	0.9124	0.9965	0.8554	0.9464	0.8452	0.9896
RuaBl	0.8215	0.8818	0.8799	0.8696	0.9144	0.9634	0.9467	0.931	0.9492	0.9292
ManBl	0.3604	0.7567	0.7221	0.5937	0.9138	0.9299	0.9633	0.921	0.9698	0.9368
twMan	0.8633	0.9197	0.912	0.907	0.9306	0.9745	0.8218	0.9476	0.8623	0.9541
Rufij	0.3384	0.7067	0.6911	0.5238	0.9116	0.85	0.925	0.8841	0.9535	0.893
caBua	0.7773	0.8447	0.8479	0.8441	0.9024	0.9866	0.976	0.7069	0.9813	0.9201
cadya	0.793	0.8584	0.8449	0.831	0.9196	0.9169	0.925	0.3248	0.9443	0.2331
chipw	0.1788	0.6736	0.6452	0.3979	0.9101	0.7857	0.8984	0.8651	0.9351	0.826
blMal	0.4983	0.3559	0.3021	0.5563	0.8953	0.8122	0.9	0.8493	0.9384	0.8315
stMal	0.1027	0.6676	0.6506	0.3671	0.9009	0.8513	0.9238	0.8909	0.9516	0.8713
blIgo	0.5394	0.1905	0.1791	0.5833	0.9033	0.7937	0.9075	0.8469	0.9361	0.8096
stIgo		0.5974	0.5867	0.2397	0.894	0.6988	0.8522	0.803	0.9028	0.7616
blogo	0.0554		0.2349	0.6416	0.9167	0.8381	0.9204	0.8781	0.9454	0.8475
blMwa	0.0532	0.0186		0.6312	0.9155	0.8051	0.908	0.868	0.9394	0.8293
stMwa	0.0242	0.0600	0.0592		0.9024	0.8167	0.9106	0.8627	0.9386	0.8413
Cteno	0.2880	0.3015	0.2979	0.2914		0.9345	0.9258	0.9232	0.9506	0.8997
blBur	0.0607	0.0702	0.0662	0.0675	0.3069		0.9846	0.9573	0.9797	0.9767
twuta	0.1442	0.1541	0.1520	0.1475	0.3043	0.1621		0.9591	0.089	0.9828
cauta	0.1257	0.1327	0.1318	0.1278	0.2931	0.1413	0.1480		0.9656	0.452

Pop.	paLwi	kaMil	twMin	RuaBl	ManBl	twMan	Rufij	caBua	cadya	chipw	blMal	stMal	blIgo
twlwa	0.1828	0.1932	0.0836	0.1534	0.1841	0.0431	0.1859	0.1595	0.1788	0.1873	0.1827	0.1730	0.1900
calwa	0.1020	0.1098	0.0920	0.1046	0.1045	0.1626	0.1034	0.0426	0.0312	0.1037	0.0990	0.0938	0.1054

Pop.	stIgo	blogo	blMwa	stMwa	Cteno	blBur	twuta	cauta	twlwa	calwa
twlwa	0.1850	0.1949	0.1931	0.1877	0.3343	0.2008	0.0322	0.1779		0.9746
calwa	0.1016	0.1090	0.1079	0.1034	0.2762	0.1184	0.1318	0.0263	0.1622	

Astatotilapia bloyeti, collected from Mwamapuli dam. 02/08/2016

Astatotilapia bloyeti, collected from Burungi. 16/08/2015

Astatotilapia calliptera, collected from Bua river. 13/09/2012

Astatotilapia calliptera, collected from Lake Chilingali, Malawi catchment. July 2004.

Astatotilapia gigliolii, collected from Mindu dam. 29/01/2014

Astatotilapia gigliolii, collected from Ruvuma catchment. 2012

Fig. S1. (a) Photos of selected haplochromines in this study. *A. bloyeti, A. calliptera and A. gigliolii.* Catchment information was written below the photos.

Astatotilapia katavi, collected from Milaca dam. 28/07/2017

Astatotilapia sp. "pseudopaludinosus", collected from Lwiche river. 27/07/2017

Astatotilapia sp. "Ruaha Blue”, collected from Mtera dam. 16/01/2014

Astatotilapia sp. "Rufiji blue", collected from Rufiji river. 20/08/2013

Astatotilapia sp. "Chipwa", collected from Lake Chipwa. 28/07/2016

Astatotilapia stappersi, collected from Mwamapuli dam. 02/08/2016

Fig. S1. (b) Photos of selected Haplochromines in this study. *A. katavi, A. sp.* "pseudopaludinosus", *A. sp.* "Ruaha Blue", *A. sp.* "Rufiji Blue", *A. sp.* "Chipwa" *and A. stappersi.* Catchment information was written below the photos.

Ctenochromis horei, collected from Malagarasi river. 28/07/2016

Ctenochromis pectoralis, collected from Ruvu river (upper Pangani). 14/08/2015

Neochromis omnicaeruleus, collected from Makobe island, 05/08/2016

Otopharynx speciosus, collected from Lake Malawi, 2005.

Paralabidochromis chilotes, collected from Makobe island, 05/08/2016

Paralabidochromis sauvagei, collected from Makobe island, 05/08/2016

Fig. S1. (c) Photos of selected haplochromines in this study. *Ctenochromis horei, Ctenochromis pectoralis*, *Neochromis omnicaeruleus*, *Otopharynx speciosus*, *Paralabidochromis chilotes and Paralabidochromis sauvagei.* Catchment information was written below the photos.

Rhamphochromis longiceps, Salima, Malawi, August 2005.

Fig. S1. (d) Photos of selected haplochromines in this study. *Rhamphochromis longiceps.* Catchment information was written below the photos.

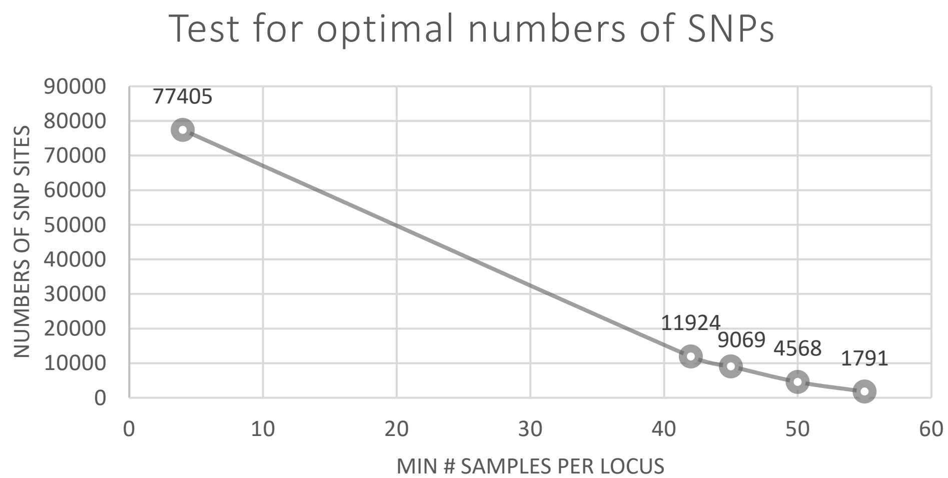


Fig. S2. Testing optimal threshold of min numbers of samples per locus. Five tests with different settings of minimal numbers of samples per locus, 4, 42, 45, 50, 55, were tested to find the appropriate one for following data analyses.