On the genetic architecture of an adaptive radiation

Inauguraldissertation
der Philosophisch-naturwissenschaftlichen Fakultät
der Universität Bern

vorgelegt von

Anna Fiona Feller

von Köniz, Bern

LeiterInnen der Arbeit:
Prof. Dr. Ole Seehausen & Prof. Dr. Catherine Peichel
Institut für Ökologie und Evolution

On the genetic architecture of an adaptive radiation

Inauguraldissertation
der Philosophisch-naturwissenschaftlichen Fakultät
der Universität Bern

vorgelegt von

Anna Fiona Feller

von Köniz, Bern

LeiterInnen der Arbeit:
Prof. Dr. Ole Seehausen & Prof. Dr. Catherine Peichel
Institut für Ökologie und Evolution

Von der Philosophisch-naturwissenschaftlichen Fakultät angenommen.

Bern, 29.06.2021

Der Dekan
Prof. Dr. Zoltán Balogh

Dedicated to my fun and loving family

Table of contents

Abstract

How biodiversity evolves is of central interest to evolutionary biology. Adaptive radiations provide prime systems to investigate processes of biological diversification and speciation. The Lake Victoria haplochromine cichlid adaptive radiation stands out due to the exceptional speed with which a spectacular diversity of species have evolved within this lake. Several coinciding factors were likely key to the emergence of this radiation including ecological opportunity, sexual selection, and intrinsic genomic features. Furthermore, the radiation was seeded by a hybrid lineage derived from ancient hybridisation between two divergent species, which fuelled the radiation by providing large amounts of genetic variation.

In my thesis, I investigate the 'genetic architecture' (the distribution and effects of genetic variants in the genome) that underpin traits or genomic features that may have played key roles in the rapid diversification and in the evolution and maintenance of reproductive barriers in the Lake Victoria cichlid adaptive radiation.

In Chapter 1, I compare the genetic architecture of male nuptial colour, a trait under sexual selection, in a sympatric pair that exhibits some gene flow vs a non-sympatric pair. In the sympatric pair, I find the presence of moderate to large effect loci and some evidence for linkage/pleiotropy, which might provide a genetic architecture that is robust to gene flow. The absence of such an architecture in the non-sympatric pair could explain why these species seem to not be able to persist in sympatry.

In Chapter 2, I investigate the genetic architecture of a whole suite of traits that distinguish representative species of two different trophic levels. Transition between trophic levels have occurred multiple times and rapidly in this radiation. I find a distributed architecture with little evidence for linkage/pleiotropy (but some moderate to large effect loci). This suggests that at the stage where these transitions occurred, genome-wide LD had to build between multiple unlinked genomic regions underpinning traits that had to change in a concerted fashion, which would have required both divergent selection and reproductive isolation.

In Chapter 3, I report a novel sex determining chromosome in African cichlids, that acts as XY in one and ZW in another lineage. This highlights the high evolvability of sex determination in these radiations, which might also play a role in speciation.

Finally, Chapter 4 is the first empirical test of the hypothesis that postzygotic intrinsic incompatibilities could be important in speciation from a hybrid swarm, in a process where they become sorted between subpopulations of a hybrid swarm and contribute to reproductive isolation between the emerging species. I screen whole genomes of 94 species of the Lake Victoria radiation and several hybrid crosses for genotype ratio distortions. I find signatures that could be consistent with postzygotic incompatibilities, and significant overlap of putative incompatibility regions with regions of high differentiation between sympatric sister species.

I end with a general discussion of my findings and suggestions for future work.

General Introduction

General Introduction

The diversity observed across the biological world is spectacular. To understand the processes and mechanisms that generate and maintain this biological diversity is a major goal of evolutionary biology. In the epoch of the Anthropocene (Waters *et al.*, 2016), understanding how biodiversity evolves -how it has arisen, how it is lost, and how it could be preserved- has become more important than ever. The impact of our species on this planet is profound (Waters *et al.*, 2016; Otto, 2018) and biodiversity is being lost at an unprecedented rate (Ceballos *et al.*, 2015; Tilman *et al.*, 2017). However, we are dependent on the functionality of our ecosystems (Millennium Ecosystem Assessment, 2005; Costanza *et al.*, 2014), and the effective functioning of ecosystems is tightly linked to, and likely often dependent on, biodiversity (Cardinale *et al.*, 2012; van der Plas, 2019).

Neither the distribution of biodiversity nor the tempo of biological diversification is uniform across space (Wagner *et al.*, 2014; Schluter & Pennell, 2017; Rabosky *et al.*, 2018), time (Uyeda *et al.*, 2011), or different lineages. For instance, while there are only four species of horseshoe crabs (Sekiguchi, 1988), there are many hundreds species of cichlid fishes (Froese & Pauly, 2021). A general estimate for background speciation rates (i.e. the average time to when a new species arises from an existing lineage) across the tree of life is ~2 million years (Hedges *et al.*, 2015). But in some instances, this has been much faster. Some of the most exceptional bursts of diversification have happened as part of adaptive radiations (Simpson, 1953; Schluter, 2000; Gillespie *et al.*, 2020). Adaptive radiation is the diversification of a lineage into an array of many ecologically diverse forms, usually associated with rapid speciation (Simpson, 1953; Schluter, 2000). Speciation is the evolution of populations that are reproductively isolated from one another, allowing them to persist as separate populations in the absence of geographic barriers (Coyne & Orr, 2004). Adaptive radiations thus provide ideal systems to study the processes and mechanisms that generate biodiversity and that drive the evolution and maintenance of reproductive barriers.

In my thesis I investigated the distribution and phenotypic effects of genetic variants in the genome (the 'genetic architecture') that underpin traits (or genomic features) that may have played key roles in the rapid diversification and to the evolution of reproductive barriers of cichlid fish in the Lake Victoria adaptive radiation.

Adaptive radiation

Classical definitions of adaptive radiation range from rather broad definitions such as "evolutionary divergence of members of a single monophyletic lineage into a variety of adaptive forms" (Futuyma, 1998) to more narrow definitions such as "the evolution of ecological diversity within a rapidly multiplying lineage" with the four specific criteria of common ancestry, phenotype-environment correlation, trait utility, and rapid speciation around the time that ecological diversification occurs (Schluter, 2000). Iconic examples of adaptive radiations include Darwin's finches (Grant & Grant, 2008), Hawaiian honeycreepers (Amadon, 1950), *Anolis* lizards (Losos, 2009), Hawaiian silverswords (Carlquist *et al.*, 2003), and African Great Lakes cichlids (Fryer & Iles, 1972), to name just a few. Recent efforts have been made to identify commonalities and differences among the very diverse forms and trajectories of adaptive radiations (Stroud & Losos, 2016; Gillespie *et al.*, 2020). One common prerequisite to adaptive radiation is ecological opportunity and usually, adaptive radiations are at some stage associated with divergent natural selection that shapes adaptation to different ecological niches (Gillespie *et al.*, 2020). Ecological

opportunity can arise by the colonization of a new area with few other species or with underutilized niches, or by the evolution of a key innovation that allows a novel way of accessing resources in a given environment, or by the extinction of another lineage, freeing a previously occupied environment or niche (Simpson, 1953).

Adaptive radiation have proven particularly useful to study the evolutionary processes of diversification and speciation due to their high species diversity, which in many cases also allows to study highly replicated cases of the evolution of similar phenotypes (e.g. Hudson *et al.*, 2011; Brawand *et al.*, 2014). Furthermore, they often feature high levels of endemism, as is particularly apparent in the case in the East African cichlid radiations, in which most of the hundreds of species that have evolved within them are unique to one lake (Fryer, 1996; Seehausen, 2015). Finally and most importantly, the cichlid radiations (and other 'classic' adaptive radiations (see Martin & Richards, 2019)) feature high levels of sympatric species richness (Seehausen, 2015).

The adaptive radiations of the African Great Lakes cichlids

Among the most spectacular adaptive radiations are the East African cichlid radiations. Cichlids belong to the class of ray-finned fishes (Actinopterygians), which comprise over 30,000 described species (Froese & Pauly, 2021), making up approximately half of all described vertebrate species. With currently ~1,700 described species (Froese & Pauly, 2021) (and many more to be described), cichlids constitute one of the most speciose fish families (Cichlidae). The largest congregation of all cichlid species is found in the East African Great Lakes region. Within each of the three largest lakes in this region, 200 to several hundreds species of cichlids have evolved in the most species-rich adaptive radiations known in vertebrates (Kocher, 2004; Seehausen, 2006; Henning & Meyer, 2014; Salzburger, 2018). In Lake Tanganyika, this has happened in less than 10 million years (Cohen *et al.*, 1993; Ronco *et al.*, 2020), in Lake Malawi in as few as 1-5 million years (Ivory *et al.*, 2016; Malinsky *et al.*, 2018), in Lake Victoria in only 15,000 years (Stager & Johnson, 2008; McGee *et al.*, 2020; Marques *et al.*, *in prep*).

The observation of the exceptionally rapid evolution of high numbers of species have led many to believe that all an African cichlid needs to radiate is a body of water. However, most African cichlid lineages have not undergone any adaptive radiation (Wagner *et al.*, 2012). A large comparative study of 46 African lakes and more than 180 cichlid colonisations by Wagner *et al.*, 2012 suggests that both ecological opportunity and lineage-intrinsic factors determine whether a lineage radiates upon colonization of a lake or not. One such lineage-intrinsic factor is the presence of sexual dichromatism, a proxy for the presence of strong sexual selection pressures (Wagner *et al.*, 2012). Recent work suggests that certain genomic features might also constitute important intrinsic factors that can facilitate rapid diversification in cichlids, such as the presence of a large number of structural variants (e.g. indels), many of which of more ancient origin than the radiation itself (McGee *et al.*, 2020). All three major cichlid radiations have now been shown to have evolved from hybrid populations seeded by distantly related species (Meier *et al.*, 2017; Irisarri *et al.*, 2018; Svardal *et al.*, 2020), providing large amounts of standing genetic variation upon which selection could then act in many different ways. Furthermore, lineages of hybrid origin have repeatedly evolved in the course of these radiations (Genner & Turner, 2012; Keller *et al.*, 2013; Meier *et al.*, 2017, 2018).

Classic adaptive radiations (see Martin & Richards, 2019) generally feature high amounts of geographic overlap (sympatry) among the closely related species that evolved in

them (Gavrilets & Losos, 2009). What sets the East African cichlid radiations apart from other radiations is how rapidly they evolved reproductive isolation that is robust to full sympatry (Bezault *et al.*, 2011; Wagner *et al.*, 2013; Seehausen & Wagner, 2014; Seehausen, 2015) (in contrast to adaptive diversification via intraspecific polymorphism without speciation, (see Smith & Skúlason, 1996), or to nearly completely allopatric radiations such as stickleback (see Hendry *et al.*, 2009; Bolnick, 2011)), making them ideal systems to study the evolution of reproductive isolation.

Speciation and the evolution of barriers to gene flow

Speciation is the process by which new species arise. It is characterized by the evolution of reproductive barriers that reduce or prevent populations from exchanging genes (gene flow), allowing diverging lineages to accumulate phenotypic and/or genetic distinctiveness that can persist in the absence of physical / geographic barriers (Mayr, 1963; Coyne & Orr, 2004). Such reproductively isolating barriers can manifest at different stages of the (sexual) reproductive process. They can act before mating or before fertilization happens (prezygotic barriers), or after fertilization has occurred (postzygotic barriers) (Coyne & Orr, 2004). Prezygotic barriers in animals can result from differences in mating behaviour or in ecology (e.g. when and where individuals mate), or by incompatibilities of the reproductive organs and gametes. Postzyotic barriers manifest in reduced hybrid viability or fertility. These barriers can be due to interactions with the environment (extrinsic) or independent of the environment (intrinsic) (Coyne & Orr, 2004).

While much of the early work in speciation has focused on intrinsic postzygotic incompatibilities (Haldane, 1922; Dobzhansky, 1934; Stebbins, 1958), the past couple of decades have seen major advances in research on prezygotic and extrinsic postzyogtic barriers (e.g. Schluter, 2001; Seehausen *et al.*, 2008; Nosil, 2012). The observation that prezgyotic barriers can evolve more rapidly than postzygotic incompatibilities (Coyne & Orr, 1989) and that they seem to be prevalent in many systems (e.g. Grant & Grant, 1997; Irwin, 2000; Lowry *et al.*, 2008; Maan & Seehausen, 2010), has led to the prevalent view that they constitute the more common barriers. However, most species are likely separated by multiple barriers and how and when they act over time may be highly dynamic (Coyne & Orr, 2004; Seehausen *et al.*, 2014; Sobel & Chen, 2014; Butlin & Smadja, 2018; Coughlan & Matute, 2020; Rometsch *et al.*, 2020).

A long-standing debate concerns the geographic mode of speciation (Gavrilets, 2003; Coyne & Orr, 2004; Jiggins, 2006; Bolnick & Fitzpatrick, 2007; Foote, 2018). During sympatric speciation, reproductive isolation evolves without geographic barriers whereas in allopatric speciation, geographic barriers initiate reproductive isolation followed by population divergence. While theoretically possible (Dieckmann & Doebeli, 1999; Higashi *et al.*, 1999; Kondrashov & Kondrashov, 1999; Gavrilets, 2004), there are few widely accepted empirical cases of primary sympatric speciation (Coyne & Orr, 2004; Bolnick & Fitzpatrick, 2007), and distinguishing scenarios where divergence with gene flow has primarily occurred in sympatry versus upon secondary contact may often not be possible even with modern genomic data (Foote, 2018).

Under the classic biogeographic definition of sympatric speciation, where reproductive isolation arises within populations whose individuals are within each other's 'cruising range' (Mayr, 1963), much of speciation in the East African cichlid radiations would qualify as sympatric. Under an alternative definition by Gavrilets, 2004 however, where the crucial condition is that mating is independent of the birthplace of the mating partners, speciation in the

African cichlid radiations would often rather qualify as ecologically parapatric as they tend to mate non-randomly with respect to their birthplace (with much of the distribution occurring along the depth gradient; Seehausen *et al.*, 2008; Seehausen & Magalhaes, 2010). The geographic mode of speciation in these systems thus presents itself as a continuum that ranges from strictly sympatric to geographically sympatric but ecologically parapatric, to geographically parapatric, and allopatric (see e.g. Seehausen & Magalhaes, 2010).

Reproductive isolation barriers in African cichlids

Both ecological and sexual selection have been implicated in the evolution (and for the persistence) of reproductive isolation barriers in East African cichlids. Premating isolation mechanisms play a central role: behavioural mate choice is often strongly species-assortative (Knight *et al.*, 1998; Seehausen *et al.*, 1998b; Van Oppen *et al.*, 1998), and based on male nuptial colour (Seehausen *et al.*, 1997; Seehausen & van Alphen, 1998; Maan *et al.*, 2004; Selz *et al.*, 2014), which is often the most striking difference between young and ecologically similar species (Seehausen, 1996, 1997; Albertson *et al.*, 1999; Seehausen *et al.*, 1999a).

Reproductive isolation can evolve under ecologically-based divergent natural selection (i.e. via ecological speciation, Schluter, 2001; Nosil, 2012), but few well-documented examples exist (see Ford *et al.*, 2016 for one). Seehausen *et al.*, 2008 demonstrated sensory drive speciation in *Pundamila*, where reproductive isolation evolved as by-product of divergent adaptation in the visual system to differing light environments. Disruptive ecological selection on morphological traits is (still) prevalent in Lake Victoria cichlids, but the mechanism that dominates the speciation process may differ among genera (van Rijssel *et al.*, 2018). In sympatric *Pundamilia* sister species pairs, disruptive selection on feeding morphology only becomes important after reproductive isolation has evolved and a certain degree of genomic differentiation is present, while in sympatric *Neochromis* sister species pairs, disruptive selection on feeding morphology is more prevalent even when reproductive isolation is still weak (van Rijssel *et al.*, 2018). (Most speciation in *Neochromis* is likely allopatric though; Seehausen & van Alphen, 1999.)

Intrinsic postzygotic isolating mechanisms have received little consideration to date and what role they play is currently unclear (Rometsch *et al.*, 2020). However, observations of skewed sex ratios in hybrids (Crapon De Caprona & Fritzsch, 1983; and reviewed in Rometsch *et al.*, 2020) and reduced hybrid survival (although mostly confined to the more divergent crosses; Stelkens *et al.*, 2010, 2015) suggest that mechanism such as genetic conflicts (which might e.g. lead to rapid transitions between sex determining systems; Seehausen *et al.*, 1999b; Lande *et al.*, 2001; Roberts *et al.*, 2009) and intrinsic postzygotic incompatibilities are also at least sometimes acting in these radiations.

While cichlids are isolated well enough by premating barriers to feature considerable genetic differentiation in sympatry (Bezault *et al.*, 2011; Wagner *et al.*, 2013), in many cases there is likely some gene flow between them (Smith *et al.*, 2003; Gante *et al.*, 2016; Meier *et al.*, 2018). This leads to some theoretical predictions about the genetic architecture of traits involved in reproductive isolation that has to be robust to the potentially homogenising effects of gene flow.

Genetic architectures and tools to study them

The definition of genetic architecture I use here is slightly modified from the one by (Gratten *et al.*, 2014) and refers to the distribution, i.e. number and location, of loci in the genome

that underpin a certain phenotype, their interactions, and how large the effect of each of these loci is on the phenotype.

Predictions can be drawn from theoretical work about the genetic architectures underpinning reproductive isolation traits in sympatric species that experience some gene flow. When there is gene flow, recombination can break up favourable combinations of alleles, quickly leading to the loss of distinct adaptive phenotypes (Felsenstein, 1981). If genomic regions underlying a phenotype were somehow less prone to recombination however, this could allow the retention of genotypic and phenotypic distinctiveness among the sympatric species despite some gene exchange. Theoretical studies suggest tight physical linkage among locally adapted alleles, pleiotropic genes, or few loci of large effect (as opposed to many distributed loci of small effect) might confer genetic architectures that are more stable in the face of gene flow (Gavrilets, 2004; Bürger & Akerman, 2011; Yeaman & Whitlock, 2011). Inversions could have a similar effect by completely supressing recombination between the loci captured in them (Kirkpatrick & Barton, 2006). Clusters of tightly linked loci that each code for a different trait can make up supergenes that can allow adaptive variation to co-segregate within species, and once established, can even spread between species by introgressive hybridisation (reviewed in Schwander *et al.*, 2014). To understand how such co-adapted haplotypes evolve is still challenging, however, and they may be difficult to evolve de-nuovo in the face of gene flow (Charlesworth & Charlesworth, 1976; Felsenstein, 1981; Yeaman, 2013; Schwander *et al.*, 2014).

Tools to study genetic architectures include Genome Wide Association Studies (GWAS) and Quantitative Trait Locus (QTL) mapping. Both are statistical methods that can be used to link phenotypic to genotypic data. Examples of studies that have used either one of these methods to identify the genetic basis of key traits in adaptive radiations include studies of beak size in Darwin's finches (Chaves *et al.*, 2016), plate number (and many other traits) in stickleback (reviewed in Peichel & Marques, 2017), oral jaw morphology in Lake Malawi cichlids (Albertson *et al.*, 2003), wing colour patterns in *Heliconius* (reviewed in Jiggins, 2017; Bainbridge *et al.*, 2020), or colour and trophic traits in Midas cichlids (Kautt *et al.*, 2020). GWAS are typically performed on natural populations, across high density SNP panels, requiring large sample sizes to obtain enough statistical power (Hong & Park, 2012; Korte & Farlow, 2013). QTL mapping takes advantage of (meiotic) recombination events in offspring between two species or strains that differ in the trait of interest to locate genomic regions that are associated with variation in a given phenotype (see Figure 1 for a conceptual overview of QTL mapping in an experimental hybrid cross). The offspring are phenotyped and genotyped at markers (i.e. SNPs) that distinguish the two parental lines, and in the subsequent analyses, one then tests for statistical association of a phenotype with a marker allele (Figure 1). While power to detect small effect loci is also limited by sample sizes, QTL mapping provides a valuable method to obtain a general picture of the genetic architecture (i.e. the presence and approximate location of at least the moderate to large effect loci) underlying the traits of interest.

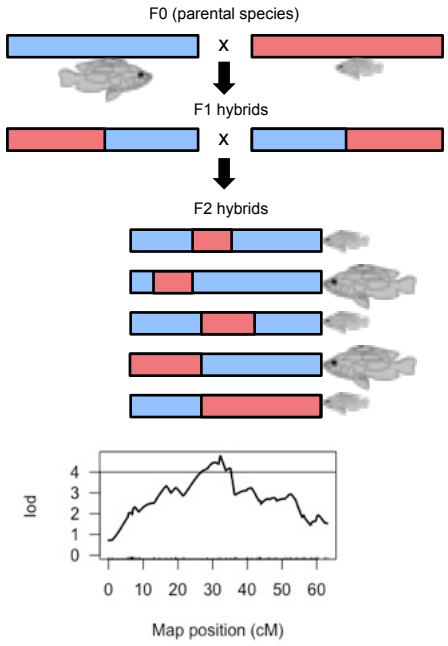


Figure 1 | QTL mapping in an interspecific second-generation (F2) hybrid cross. Individuals of two (F0) parental species are crossed and the resulting first-generation (F1) hybrids are inter-crossed again, resulting in F2 hybrid individuals that contain different fractions of the parental species' genomes (and thus also feature different phenotypic values). In standard interval mapping for instance, the probability that a marker (or a position between two adjacent markers) is associated with a locus that affects the trait is determined using maximum likelihood estimation under a mixture model. Conceptually, standard interval mapping is similar to performing an anova at each marker, but it is much more powerful as it models missing genotypes, also at locations between markers, giving a better estimation of the location of a putative QTL (see e.g. Broman & Sen, 2009). 'lod' is the likelihood ratio test statistic which is plotted as a function of genomic position in 'cM' (recombination units). Only one chromosome is shown. The horizontal line indicates a putative significance threshold, which can be determined by permutation (see e.g. Broman *et al.*, 2003). Figure inspired by and adapted from Mackay, 2001.

This thesis – on the genetic architecture of the Lake Victoria cichlid adaptive radiation

The broad aim of this thesis was to contribute to our understanding of factors that allow rapid speciation and adaptive radiation. In each of the following four chapters I investigated the genetic architecture of one component that has likely importantly contributed to the emergence of many species and their persistence in close geographic proximity and to the rapid evolution of trophic diversity in the Lake Victoria adaptive radiation.

Among the East African cichlid radiations the Lake Victoria radiation stands out as exceptionally rapid, faster than any other radiation by orders of magnitude (McGee *et al.*, 2020). After complete dessication during the Pleistocene, Lake Victoria only refilled with water to become the world's largest tropical lake approximately 15,000 years ago (Johnson *et al.*, 1996, 2000; Stager & Johnson, 2008). This new young lake was colonized by a lineage of hybrid origin (Meier *et al.*, 2017), which subsequently radiated into over 500 distinct species (Seehausen, 1996). In the following paragraphs I will describe four patterns/observations in this radiation that provided the foundation to the four chapters in my thesis.

Replicated male nuptial coloration patterns

Lake Victoria haplochromine cichlids are strongly sexually dimorphic. Females are smaller and cryptic in colour while males are larger and very colourful. Females are the sole investors into parental care (they are mouthbrooders) and they choose their mates based on male nuptial colour (Seehausen & van Alphen, 1998; Seehausen *et al.*, 1999b; Maan *et al.*, 2004; Selz *et al.*, 2014). This process is likely affected by the visual environment (Seehausen *et al.*, 1997; Maan *et al.*, 2010). In many species, males fiercely defend their territories to which they try to attract females against males of their own and other species (Seehausen & Schluter, 2004; Dijkstra *et al.*, 2007). Hence, inter- and intra-specific sexual selection is prevalent in this system, and premating behavioural isolation an important barrier to gene flow. These barriers are not

fully isolating however, and there is some gene flow among sympatric species (Seehausen, 1997; Seehausen *et al.*, 2008; Magalhaes *et al.*, 2010; Konijnendijk *et al.*, 2011; Meier *et al.*, 2018).

One striking observation is the occurrence of many sister species pairs that are ecologically highly similar but very different in male nuptial coloration. Often, one species is blue and the other red. In sympatry, the red species of such a pair usually has a red dorsum, while red species with a red chest are not usually seen in sympatry with their blue sister species (Seehausen, 1996, 1997, 2009; Seehausen *et al.*, 1998a). Neither geographic isolation or dispersal limitation could explain why the red-chest species is (almost) never seen in sympatry with its blue sister species (Seehausen, 1996; Seehausen *et al.*, 1998a). We hypothesised that if the genetic architecture of the red-chest species is not stable against gene flow, this might explain why it cannot persist in sympatry with its blue sister species. Or vice versa: a more stable 'simple' genetic architecture might allow the red-dorsum species to persist in sympatry despite some gene flow. In ***Chapter 1*** I tested this hypothesis by performing QTL mapping analyses of male nuptial coloration in a sympatric blue vs red-dorsum pair and in a non-sympatric blue vs red-chest pair.

Rapid and repeated transitions between trophic levels

Lake Victoria species can be grouped into at least 14 different trophic groups. There are detritivores, phytoplanktivores, algae grazers (epilithic or epiphytic), molluscivores (oral shellers/crushers or pharyngeal crushers), zooplanktivores, insectivores, prawn eaters, crab eaters, piscivores, paedophages, scale eaters, and parasite eaters (Greenwood, 1974; Witte & Van Oijen, 1990; Seehausen, 1996). Each trophic group except one (the scale eaters) contains several to many species (Greenwood, 1974; Witte & Van Oijen, 1990; Seehausen, 1996). They thus make up whole trophic networks spanning three trophic levels. Such rapid radiation across three trophic levels and especially the repeated evolution of herbivory from insectivorous ancestors is unparalleled compared to other freshwater fish radiations (e.g. stickleback; Bell & Foster, 1994), charr (Jonsson & Jonsson, 2001), whitefish (Lu & Bernatchez, 1999; Hudson *et al.*, 2007), sculpins (Sideleva, 2003), neither of which evolved any herbivorous species), and probably constitutes one key factor to explain the evolution of the spectacular species richness in this system. Furthermore, recent work suggests that transitions between trophic levels may even have happened in both directions (McGee *et al.*, 2020). This is remarkable because transitions between trophic types requires the simultaneous change of a whole suite of morphological, physiological and behavioural adaptations. We hypothesized that these rapid transitions may have been facilitated by a highly linked or pleiotropic (i.e. supergene-like) genetic architecture. In ***Chapter 2*** I tested this hypothesis by investigating the genetic architecture of trophic morphology that differentiates members of an invertivorous lineage from members of a herbivorous lineage.

Multiple sex determining systems

Unlike mammals and birds, cichlids (as many other fishes; Devlin & Nagahama, 2002) do not have easily identifiable heteromorphic sex chromosomes. Advances in DNA sequencing technologies and analyses over the past few decades have made it possible to identify younger and less differentiated sex chromosomes (Abbott *et al.*, 2017). This has revealed a stunning variety in sex determining systems among African cichlids (Gammerdinger & Kocher, 2018; Böhne *et al.*, 2019). More than 10 different chromosomes have been found to be involved in sex determination, and in some cases, these are B-chromosomes (supernumerary chromosomes (Wilson, 1907)), there are XY as well ZW heterogametic systems and sometimes these even

segregate within the same lineage or species (Gammerdinger & Kocher, 2018). Looking at the few existing investigations of sex determination in Lake Victoria cichlids indicates that they are probably no exception to this: QTLs for sex have been found on three different linkage groups (Kudo *et al.*, 2015; Feulner *et al.*, 2018), B chromosomes likely play a role (Kudo *et al.*, 2015), and analyses of sex ratio distortion patterns in early crossing experiments with colour-polymorphic *Neochromis omnicaeruleus* suggests that there are multiple autosomal sex modifiers segregating in single populations of this species (Seehausen *et al.*, 1999b). Sex ratio distortions as observed in *N. omnicaeruleus* could occur due to intragenomic conflicts over sex ratio. This could cause the rapid evolution of different sex-determining systems (Werren & Beukeboom, 1998), which in turn could lead to the emergence of new species (see e.g. Seehausen *et al.*, 1999b; Lande *et al.*, 2001; Kocher, 2004). New sex chromosomes can also evolve due to sexually antagonistic selection (Van Doorn & Kirkpatrick, 2007) resulting from female mate choice or male-male competition (Arnqvist & Rowe, 2005), both of which are prevalent in Lake Victoria cichlids (see above). In ***Chapter 3*** we took advantage of the RAD sequencing data generated for the crosses used in Chapter 1 as well as for a Lake Malawi species cross (Feller *et al.*, 2020) to characterise sex determination among the crossed species using a combination of QTL mapping and allele frequency analyses.

Old incompatibilities are likely to be segregating in this system

In light of the fact that most Lake Victoria cichlid species can be crossed relatively easily in the lab and that prezygotic barriers have been shown to be important (see above), it may seem counter-intuitive to look for postzygotic incompatibilities in this system. However, the age of the founding lineages of the hybrid swarm that seeded the Lake Victoria radiation makes it highly likely that there were incompatibilities segregating between them (Stelkens *et al.*, 2010; Meier *et al.*, 2017), and some still may be segregating among the radiation species (Stelkens *et al.*, 2010, 2015; Meier *et al.*, 2017). It has been suggested that postzygotic incompatibilities could become sorted between subpopulations of a hybrid swarm and contribute to reproductive isolation between the emerging species (and will then no longer be purged) (Seehausen, 2013). Meier *et al.*, 2017 was the first study to test predictions derived from this verbal model, and found statistical patterns of ancestry that were consistent with it. ***Chapter 4*** constitutes the first experimental test of this hypothesis. I combined multiple approaches to look for signatures of postzygotic incompatibilities by screening for genotype ratio distortions among the hybrid crosses used in the previous three chapters and among many species of the Lake Victoria radiation using whole genome sequencing data.

References

Abbott, J.K., Nordén, A.K. & Hansson, B. 2017. Sex chromosome evolution: Historical insights and future perspectives. *Proc. R. Soc. B Biol. Sci.* **284**.

Albertson, R.C., Markert, J.A., Danley, P.D. & Kocher, T.D. 1999. Phytogeny of a rapidly evolving clade: The cichlid fishes of Lake Malawi, East Africa. *Proc. Natl. Acad. Sci. U. S. A.* **96**: 5107–5110.

Albertson, R.C., Streelman, J.T. & Kocher, T.D. 2003. Directional selection has shaped the oral jaws of Lake Malawi cichlid fishes. *Proc. Natl. Acad. Sci.* **100**: 5252–5257.

Amadon, D. 1950. The Hawaiian honeycreepers (Aves, Drepaniidae). *Bull. Am. Museum Nat. Hist.* **95**: 151–262. New York, NY.

Arnqvist, G. & Rowe, L. 2005. *Sexual Conflict*. Princeton University Press, Princeton, NJ.

Bainbridge, H.E., Brien, M.N., Morochz, C., Salazar, P.A., Rastas, P., Nadeau, N. 2020. Limited genetic parallels underlie convergent evolution of quantitative pattern variation in mimetic butterflies. *Journal of Evolutionary Biology* **33***:* 1516-1529.

Bell, M.A. & Foster, S.A. 1994. *The Evolutionary Biology of the Threespine Stickleback*. Oxford University Press, Oxford, UK.

Bezault, E., Mwaiko, S. & Seehausen, O. 2011. Population genomic tests of models of adaptive radiation in Lake Victoria region cichlid fish. *Evolution* **65**: 3381–3397.

Böhne, A., Weber, A.A., Rajkov, J., Rechsteiner, M., Riss, A., Egger, B., *et al.* 2019. Repeated Evolution Versus Common Ancestry : Sex Chromosome Evolution in the Haplochromine Cichlid Pseudocrenilabrus philander. *GBE* **11**: 439–458.

Bolnick, D.I. 2011. Sympatric speciation in threespine stickleback: Why not? *Int. J. Ecol.* **2011**.

Bolnick, D.I. & Fitzpatrick, B.M. 2007. Sympatric speciation: Models and empirical evidence. *Annu. Rev. Ecol. Evol. Syst.* **38**: 459–487.

Brawand, D., Wagner, C.E., Li, Y.I., Malinsky, M., Keller, I., Fan, S., *et al.* 2014. The genomic substrate for adaptive radiation in African cichlid fish. *Nature* **513**: 375–381.

Broman, K.W. & Sen, S. 2009. *A Guide to QTL Mapping with R/qtl*. Springer New York, New York, NY.

Broman, K.W., Wu, H., Sen, S. & Churchill, G.A. 2003. R/qtl: QTL mapping in experimental crosses. *Bioinformatics* **19**: 889–890.

Bürger, R. & Akerman, A. 2011. The effects of linkage and gene flow on local adaptation: A two-locus continent-island model. *Theor. Popul. Biol.* **80**: 272–288.

Butlin, R.K. & Smadja, C.M. 2018. Coupling, reinforcement, and speciation. *Am. Nat.* **191**: 155–172.

Cardinale, B.J., Duffy, J.E., Gonzalez, A., Hooper, D.U., Perrings, C., Venail, P., *et al.* 2012. Biodiversity loss and its impact on humanity. *Nature* **486**: 59–67.

Carlquist, S., Baldwin, B.G., Carr, G.D. & (eds). 2003. *Tarweeds and Silverswords: Evolution of the Madiinae*. Missouri Botanical Garden Press, St. Louis, MO.

Ceballos, G., Ehrlich, P.R., Barnosky, A.D., García, A., Pringle, R.M. & Palmer, T.M. 2015. Accelerated modern human-induced species losses: Entering the sixth mass extinction. *Sci. Adv.* **1**: 9–13.

Charlesworth, D. & Charlesworth, B. 1976. Theoretical genetics of Batesian mimicry II. Evolution of Supergenes. *J. Theor. Biol.* **55**: 305–324.

Chaves, J.A., Cooper, E.A., Hendry, A.P., Podos, J., De León, L.F., Raeymaekers, J.A.M., *et al.* 2016. Genomic variation at the tips of the adaptive radiation of Darwin's finches. *Mol. Ecol.* **25**: 5282–5295.

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

Costanza, R., de Groot, R., Sutton, P., van der Ploeg, S., Anderson, S.J., Kubiszewski, I., *et al.* 2014. Changes in the global value of ecosystem services. *Glob. Environ. Chang.* **26**: 152–158. Elsevier Ltd.

Coughlan, J.M. & Matute, D.R. 2020. The importance of intrinsic postzygotic barriers throughout the speciation process. *Philos. Trans. R. Soc. Lond. B. Biol. Sci.* **375**: 20190533.

Coyne, J.A. & Orr, H.A. 1989. Patterns of Speciation in Drosophila. *Evolution* **43**: 362–381.

Coyne, J.A. & Orr, H.A. 2004. *Speciation*. Sinauer Associates, Sunderland, MA.

Crapon De Caprona, M.-D. & Fritzsch, B. 1983. Interspecific Fertile Hybrids of Haplochromine Cichlidae (Teleostei) and Their Possible Importance for Speciation. *Netherlands J. Zool.* **34**: 503–538.

Devlin, R.H. & Nagahama, Y. 2002. Sex determination and sex differentiation in fish: an overview of genetic, physiological, and environmental influences. *Aquaculture* **208**: 191–364.

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

Dijkstra, P.D., Seehausen, O., Pierotti, M.E.R. & Groothuis, T.G.G. 2007. Male-male competition and speciation: aggression bias towards differently coloured rivals varies between stages of speciation in a Lake Victoria cichlid species complex. *J. Evol. Biol.* **20**: 496–502.

Dobzhansky, T. 1934. Studies on hybrid sterility. *Genetics* **21**: 113–135.

Feller, A.F., Selz, O.M., McGee, M.D., Meier, J.I., Mwaiko, S. & Seehausen, O. 2020. Rapid generation of ecologically relevant behavioral novelty in experimental cichlid hybrids. *Ecol. Evol.* **10**: 7445–7462.

Felsenstein, J. 1981. Skepticism Towards Santa Rosalia, or Why are There so Few Kinds of Animals? *Evolution* **35**: 124.

Feulner, P.G.D., Schwarzer, J., Haesler, M.P., Meier, J.I. & Seehausen, O. 2018. A Dense Linkage Map of Lake Victoria Cichlids Improved the Pundamilia Genome Assembly and Revealed a Major QTL for Sex-Determination. *G3; Genes|Genomes|Genetics* **8**: 2411–2420.

Foote, A.D. 2018. Sympatric Speciation in the Genomic Era. *Trends Ecol. Evol.* **33**: 85–95. Elsevier Ltd.

Ford, A.G.P., Rüber, L., Newton, J., Dasmahapatra, K.K., Balarin, J.D., Bruun, K., *et al.* 2016. Niche divergence facilitated by fine-scale ecological partitioning in a recent cichlid fish adaptive radiation. *Evolution* **70**: 2718–2735.

Froese, R. & Pauly, D. 2021. http://www.fishbase.org.

Fryer, G. 1996. Endemism, speciation and adaptive radiation in great lakes. *Environ. Biol. Fishes* **45**: 109–131.

Fryer, G. & Iles, T.D. 1972. *The cichlid fishes of the great lakes of Africa: their biology and evolution*. Oliver & Boyd, Edinburgh.

Futuyma, D.J. 1998. *Evolutionary biology*. Sinauer, Sunderland, MA.

Gammerdinger, W.J. & Kocher, T.D. 2018. Unusual Diversity of Sex Chromosomes in African Cichlid Fishes. *Genes* **9**: 480.

Gante, H.F., Matschiner, M., Malmstrøm, M., Jakobsen, K.S., Jentoft, S. & Salzburger, W. 2016. Genomics of speciation and introgression in Princess cichlid fishes from Lake Tanganyika. *Mol. Ecol.* **25**: 6143–6161.

Gavrilets, S. 2004. *Fitness landscapes and the origin of species*. Princeton University Press, Princeton, NJ.

Gavrilets, S. 2003. Perspective: Models of speciation - What have we learned in 40 years? *Evolution* **57**: 2197–2215.

Gavrilets, S. & Losos, J.B. 2009. Adaptive Radiation: Contrasting Theory with Data. *Science* **323**: 732–737.

Genner, M.J. & Turner, G.F. 2012. Ancient Hybridization and Phenotypic Novelty within Lake Malawi's Cichlid Fish Radiation. *Mol. Biol. Evol.* **29**: 195–206.

Gillespie, R.G., Bennett, G.M., De Meester, L., Feder, J.L., Fleischer, R.C., Harmon, L.J., *et al.* 2020. Comparing Adaptive Radiations Across Space, Time, and Taxa. *J. Hered.* **111**: 1–20.

Grant, P.R. & Grant, B.R. 1997. Genetics and the origin of species. *Proc. Natl. Acad. Sci. USA* **33**: 7768–7775.

Grant, P.R. & Grant, B.R. 2008. *How and Why Species Multiply: The Radiations of Darwin's Finches*. Princeton University Press, Princeton, NJ.

Gratten, J., Wray, N.R., Keller, M.C. & Visscher, P.M. 2014. Large-scale genomics unveils the genetic architecture of psychiatric disorders. *Nat. Neurosci.* **17**: 782–790. Nature Publishing Group.

Greenwood, P.H. 1974. *Cichlid fishes of Lake Victoria, East Africa: the biology and evolution of a species flock*. John Wright and Sons Ltd., Stonebridge Press, Bristol.

Haldane, J.B.S. 1922. Sex ratio and unisexual sterility in hybrid animals. *J. Genet* **XII**: 101–109.

Hedges, S.B., Marin, J., Suleski, M., Paymer, M. & Kumar, S. 2015. Tree of life reveals clock-like speciation and diversification. *Mol. Biol. Evol.* **32**: 835–845.

Hendry, A.P., Bolnick, D.I., Berner, D. & Peichel, C.L. 2009. Along the speciation continuum in sticklebacks. *J. Fish Biol.* **75**: 2000–2036.

Henning, F. & Meyer, A. 2014. The evolutionary genomics of cichlid fishes: Explosive speciation and adaptation in the postgenomic Era. *Annu. Rev. Genomics Hum. Genet.* **15**: 417–441.

Higashi, M., Takimoto, G. & Yamamura, N. 1999. Sympatric speciation by sexual selection. *Nature* **402**: 523–526.

Hong, E.P. & Park, J.W. 2012. Sample Size and Statistical Power Calculation in Genetic Association Studies. *Genomics Inform.* **10**: 117.

Hudson, A.G., Vonlanthen, P., Müller, R. & Seehausen, O. 2007. Review: The geography of speciation and adaptive radiation in coregonines. *Adv. Limnol.* **60**: 111–146.

Hudson, A.G., Vonlanthen, P. & Seehausen, O. 2011. Rapid parallel adaptive radiations from a single hybridogenic ancestral population. *Proc. R. Soc. B Biol. Sci.* **278**: 58–66.

Irisarri, I., Singh, P., Koblmüller, S., Torres-Dowdall, J., Henning, F., Franchini, P., *et al.* 2018. Phylogenomics uncovers early hybridization and adaptive loci shaping the radiation of Lake Tanganyika cichlid fishes. *Nat. Commun.* **9**: 3159.

Irwin, D.E. 2000. Song variation in an avian ring species. *Evolution* **54**: 998–1010.

Ivory, S.J., Blome, M.W., King, J.W., McGlue, M.M., Cole, J.E. & Cohen, A.S. 2016. Environmental change explains cichlid adaptive radiation at Lake Malawi over the past 1.2 million years. *Proc. Natl. Acad. Sci. U. S. A.* **113**: 11895–11900.

Jiggins, C.D. 2006. Sympatric Speciation: Why the Controversy? *Curr. Biol.* **16**: 333–334.

Jiggins, C.D. 2017. What Can We Learn About Adaptation from the Wing Pattern Genetics of Heliconius Butterflies? In: *Diversity and Evolution of Butterfly Wing Patterns: An Integrative Approach* (T. Sekimura & H. F. Nijhout, eds), pp. 173–188. Springer Singapore, Singapore.

Johnson, T.C., Kelts, K. & Odada, E. 2000. The Holocene History of Lake Victoria. *AMBIO A J. Hum. Environ.* **29**: 2–11.

Johnson, T.C., Scholz, C.A., Talbot, M.R., Kelts, K., Ricketts, R.D., Ngobi, G., *et al.* 1996. Late pleistocene desiccation of Lake Victoria and rapid evolution of cichlid fishes. *Science* **273**: 1091–1093.

Jonsson, B. & Jonsson, N. 2001. Polymorphism and speciation in Arctic charr. *J. Fish Biol.* **58**: 605–638.

Kautt, A.F., Kratochwil, C.F., Nater, A., Machado-Schiaffino, G., Olave, M., Henning, F., *et al.* 2020. Contrasting signatures of genomic divergence during sympatric speciation. *Nature* **588**: 106–111

Keller, I., Wagner, C.E., Greuter, L., Mwaiko, S., Selz, O.M., Sivasundar, A., *et al.* 2013. Population genomic signatures of divergent adaptation, gene flow and hybrid speciation in the rapid radiation of Lake Victoria cichlid fishes. *Mol. Ecol.* **22**: 2848–2863.

Kirkpatrick, M. & Barton, N. 2006. Chromosome inversions, local adaptation and speciation. *Genetics* **173**: 419–434.

Knight, M.E., Turner, G.F., Rico, C., van Oppen, M.J.H. & Hewitt, G.M. 1998. Microsatellite paternity analysis on captive Lake Malawi cichlids supports reproductive isolation by direct mate choice. *Mol. Ecol.* **7**: 1605–1610.

Kocher, T.D. 2004. Adaptive evolution and explosive speciation: the cichlid fish model. *Nat. Rev. Genet.* **5**: 288–298.

Kondrashov, A.S. & Kondrashov, F.A. 1999. Interactions among quantitative traits in the course of sympatric speciation. *Nature* **400**: 351–354.

Konijnendijk, N., Joyce, D.A., Mrosso, H.D.J., Egas, M. & Seehausen, O. 2011. Community Genetics Reveal Elevated Levels of Sympatric Gene Flow among Morphologically Similar but Not among Morphologically Dissimilar Species of Lake Victoria Cichlid Fish. *Int. J. Evol. Biol.* **2011**: 1–12.

Korte, A. & Farlow, A. 2013. The advantages and limitations of trait analysis with GWAS: A review. *Plant Methods* **9**: 1. Plant Methods.

Kudo, Y., Nikaido, M., Kondo, A., Suzuki, H., Yoshida, K., Kikuchi, K., *et al.* 2015. A microsatellite-based genetic linkage map and putative sex-determining genomic regions in Lake Victoria cichlids. *Gene* **560**: 156–164.

Lande, R., Seehausen, O. & Alphen, J.J.M.V. 2001. Mechanisms of rapid sympatric speciation by sex reversal and sexual selection in cichlid fish. *Genetica* **112–113**: 435–443.

Losos, J.B. 2009. *Lizards in an Evolutionary tree: Ecology and Adaptive Radiation of Anoles*. University of California Press, Berkley, CA.

Lowry, D.B., Modliszewski, J.L., Wright, K.M., Wu, C.A. & Willis, J.H. 2008. Review. The strength and genetic basis of reproductive isolating barriers in flowering plants. *Philos. Trans. R. Soc. B Biol. Sci.* **363**: 3009–3021.

Lu, G. & Bernatchez, L. 1999. Correlated Trophic Specialization and Genetic Divergence in Sympatric Lake Whitefish Ecotypes (Coregonus clupeaformis): Support for the Ecological Speciation Hypothesis. *Society* **53**: 1491–1505.

Maan, M.., Seehausen, O. & van Alphen, J.J.M. 2010. Female mating preferences and male coloration change with water transparency in a Lake Victoria cichlid fish Female mating preferences and male coloration change with water transparency in a Lake Victoria cichlid fish. *Biol. J. Linn. Soc.* **99**: 398–406.

Maan, M. & Seehausen, O. 2010. Mechanisms of species divergence through visual adaptation and sexual selection: Perspectives from a cichlid model system. *Curr. Zool.* **56**: 285–299.

Maan, M.E., Seehausen, O., Söderberg, L., Johnson, L., Ripmeester, E.A.P., Mrosso, H.D.J., *et al.* 2004. Intraspecific sexual selection on a speciation trait, male coloration, in the Lake Victoria cichlid Pundamilia nyererei. *Proc. R. Soc. B Biol. Sci.* **271**: 2445–2452.

Mackay, T.F.C. 2001. Quantitative trait loci in Drosophila. *Nat. Rev. Genet.* **2**: 11–20.

Magalhaes, I.S., Mwaiko, S. & Seehausen, O. 2010. Sympatric colour polymorphisms associated with nonrandom gene flow in cichlid fish of Lake Victoria. *Mol. Ecol.* **19**: 3285–3300.

Malinsky, M., Svardal, H., Tyers, A.M., Miska, E.A., Genner, M.J., Turner, G.F., *et al.* 2018. Whole genome sequences of Malawi cichlids reveal multiple radiations interconnected by gene flow. *Nat. Ecol. Evol.* **2**: 1940–1955.

Marques, D.A., Haesler, M.P., Meier, J.I., Kishe,. M.A., McGee, M.D., Mwaiko, S., Wagner, C.E. & Seehausen, O. *in prep*. Whole genomes resolve age of the fastest adaptive radiation.

Martin, C.H. & Richards, E.J. 2019. The Paradox Behind the Pattern of Rapid Adaptive Radiation: How Can the Speciation Process Sustain Itself Through an Early Burst? *Annu. Rev. Ecol. Evol. Syst.* **50**: 569–593.

Mayr, E. 1963. *Animal species and evolution*. Harvard University Press, Cambridge, MA.

McGee, M.D., Borstein, S.R., Meier, J.I., Marques, D.A., Mwaiko, S., Taabu, A., *et al.* 2020. The ecological and genomic basis of explosive adaptive radiation. *Nature* **586**: 75–79.

Meier, J.I., Marques, D.A., Mwaiko, S., Wagner, C.E., Excoffier, L. & Seehausen, O. 2017. Ancient hybridization fuels rapid cichlid fish adaptive radiations. *Nat. Commun.* **8**.

Meier, J.I., Marques, D.A., Wagner, C.E., Excoffier, L. & Seehausen, O. 2018. Genomics of parallel ecological speciation in Lake Victoria cichlids. *Mol. Biol. Evol.* **35**: 1489–1506.

Millennium Ecosystem Assessment, M.E.A. 2005. *Ecosystems and Human Well- Being: Synthesis*. Island Press, Washington, D.C.

Nosil, P. 2012. *Ecological speciation*. Oxford University Press, Oxford, UK.

Otto, S.P. 2018. Adaptation, speciation and extinction in the Anthropocene. *Proc. R. Soc. B Biol. Sci.* **285**.

Peichel, C.L. & Marques, D.A. 2017. The genetic and molecular architecture of phenotypic diversity in sticklebacks. *Philos. Trans. R. Soc. B Biol. Sci.* **372**.

Rabosky, D.L., Chang, J., Title, P.O., Cowman, P.F., Sallan, L., Friedman, M., *et al.* 2018. An inverse latitudinal gradient in speciation rate for marine fishes. *Nature* **559**: 392-395.

Roberts, R.B., Ser, J.R. & Kocher, T.D. 2009. Sexual conflict resolved by invasion of a novel sex determiner in lake malawi cichlid fishes. *Science* **326**: 998–1001.

Rometsch, S.J., Torres-Dowdall, J. & Meyer, A. 2020. Evolutionary dynamics of pre- and postzygotic reproductive isolation in cichlid fishes. *Philos. Trans. R. Soc. Lond. B. Biol. Sci.* **375**: 20190535.

Ronco, F., Matschiner, M., Böhne, A., Boila, A., Büscher, H.H., El Taher, A., *et al.* 2020. Drivers and dynamics of a massive adaptive radiation in cichlid fishes. *Nature* **589:** 76–81.

Salzburger, W. 2018. Understanding explosive diversification through cichlid fish genomics. *Nat. Rev. Genet.* **19**: 705–717.

Schluter, D. 2001. Ecology and the origin of species. *Trends Ecol. Evol.* **16**: 372–380.

Schluter, D. 2000. *The Ecology of Adaptive Radiation*. Oxford University Press, New York, NY.

Schluter, D. & Pennell, M.W. 2017. Speciation gradients and the distribution of biodiversity. *Nature* **546**: 48–55.

Schwander, T., Libbrecht, R. & Keller, L. 2014. Supergenes and complex phenotypes. *Curr. Biol.* **24**: R288–R294.

Seehausen, O. 2006. African cichlid fish: a model system in adaptive radiation research. *Proc. R. Soc. B Biol. Sci.* **273**: 1987–1998.

Seehausen, O. 2013. Conditions when hybridization might predispose populations for adaptive radiation. *J. Evol. Biol.* **26**: 279–281.

Seehausen, O. 1997. Distribution of and reproductive isolation among color morphs of a rock-dwelling Lake Victoria cichlid (Haplochromis nyererei). *Ecol. Freshw. Fish* **6**: 59–66.

Seehausen, O. 1996. *Lake Victoria Rock Cichlids - taxonomy, ecology, and distribution*. Verduijn Cichlids, Zevenjuizen, NL.

Seehausen, O. 2015. Process and pattern in cichlid radiations - inferences for understanding unusually high rates of evolutionary diversification. *New Phytol.* **207**: 304–312.

Seehausen, O. 2009. Progressive levels of trait divergence along a ' speciation transect ' in the Lake Victoria cichlid fish Pundamilia. In: *Speciation and patterns of diversity* (R. K. Butlin, J. Bridle, & D. Schluter, eds), pp. 155–176. Cambridge University Press, Cambridge, UK.

Seehausen, O., Butlin, R.K., Keller, I., Wagner, C.E., Boughman, J.W., Hohenlohe, P.A., *et al.* 2014. Genomics and the origin of species. *Nat. Rev. Genet.* **15**: 176–192.

Seehausen, O., Lippitsch, E., Bouton, N. & Zwennes, H. 1998a. Mbibi, the rock-dwelling cichlids of Lake Victoria: description of three new genera and fifteen new species (Teleostei). *Ichthyol. Explor. Freshwaters* **9**: 129–228.

Seehausen, O. & Magalhaes, I.S. 2010. Geographical Mode and Evolutionary Mechanism of Ecological Speciation in Cichlid Fish. In: *In Search of the Causes of Evolution: From Field Observations to Mechanisms* (P. R. Grant & B. R. Grant, eds). Princeton University Press, Princeton, NJ.

Seehausen, O. & Schluter, D. 2004. Male-male competition and nuptial-colour displacement as a diversifying force in Lake Victoria cichlid fishes. *Proc. R. Soc. B Biol. Sci.* **271**: 1345–1353.

Seehausen, O., Terai, Y., Magalhaes, I.S., Carleton, K.L., Mrosso, H.D.J., Miyagi, R., *et al.* 2008. Speciation through sensory drive in cichlid fish. *Nature* **455**: 620–626.

Seehausen, O. & van Alphen, J.J.M. 1999. Can sympatric speciation by disruptive sexual selection explain rapid evolution of cichlid diversity in Lake Victoria? *Ecol. Lett.* **2**: 262–271.

Seehausen, O. & van Alphen, J.J.M. 1998. The effect of male coloration on female mate choice in closely related Lake Victoria cichlids (Haplochromis nyererei complex). *Behav. Ecol. Sociobiol.* **42**: 1–8.

Seehausen, O., van Alphen, J.J.M. & Witte, F. 1997. Cichlid fish diveristy threatened by eutrophication that curbs sexual selection. *Science* **277**: 1808–1811.

Seehausen, O., Van Alphen, J.J.M. & Witte, F. 1999a. Can ancient colour polymorphisms explain why some cichlid lineages speciate rapidly under disruptive sexual selection? *Belgian J. Zool.* **129**: 43–60.

Seehausen, O. & Wagner, C.E. 2014. Speciation in Freshwater Fishes. *Annu. Rev. Ecol. Evol. Syst.* **45**: 621–651.

Seehausen, O., Witte, F., Alphen, J.J.M. & Bouton, N. 1998b. Direct mate choice maintains diversity among sympatric cichlids in Lake Victoria. *J. Fish Biol.* **53**: 37–55.

Seehausen, van Alphen & Lande. 1999b. Color polymorphism and sex ratio distortion in a cichlid fish as an incipient stage in sympatric speciation by sexual selection. *Ecol. Lett.* **2**: 367–378.

Sekiguchi, K. 1988. *Biology of Horseshoe Crabs*. Science House, Tokyo.

Selz, O.M., Pierotti, M.E.R.R., Maan, M.E., Schmid, C. & Seehausen, O. 2014. Female preference for male color is necessary and sufficient for assortative mating in 2 cichlid sister species. *Behav. Ecol.* **25**: 612–626.

Sideleva, V.G. 2003. *The endemic fishes of Lake Baikal*. Backhuys Publishers, Leiden, NL.

Simpson, G.G. 1953. *The Major Features of Evolution*. Columbia Univeristy Press, New York, NY.

Smith, P.F., Konings, A. & Kornfield, I. 2003. Hybrid origin of a cichlid population in Lake Malawi: Implications for genetic variation and species diversity. *Mol. Ecol.* **12**: 2497–2504.

Smith, T.B. & Skúlason, S. 1996. Evolutionary significance of resource polymorphisms in fishes, amphibians, and birds. *Annu. Rev. Ecol. Syst.* **27**: 111–133.

Sobel, J.M. & Chen, G.F. 2014. Unification of methods for estimating the strength of reproductive isolation. *Evolution.* **68**: 1511–1522.

Stager, J.C. & Johnson, T.C. 2008. The late Pleistocene desiccation of Lake Victoria and the origin of its endemic biota. *Hydrobiologia* **596**: 5–16.

Stebbins, G.L. 1958. The Inviability, Weakness, and Sterility of Interspecific Hybrids. In: Advances in Genetics, Vol. 9 (Demerec M. ed), pp. 147–215, University of California, Davies, CA.

Stelkens, R.B., Schmid, C. & Seehausen, O. 2015. Hybrid Breakdown in Cichlid Fish. *PLoS One* **10**: e0127207.

Stelkens, R.B., Young, K.A. & Seehausen, O. 2010. The Accumulation of Reproductive Incompatibilities in African Cichlid Fish. *Evolution* **64**: 617–633.

Stroud, J.T. & Losos, J.B. 2016. Ecological Opportunity and Adaptive Radiation. *Annu. Rev. Ecol. Evol. Syst.* **47**: 507–532.

Svardal, H., Quah, F.X., Malinsky, M., Ngatunga, B.P., Miska, E.A., Salzburger, W., *et al.* 2020. Ancestral Hybridization Facilitated Species Diversification in the Lake Malawi Cichlid Fish Adaptive Radiation. *Mol. Biol. Evol.* **37**: 1100–1113.

Tilman, D., Clark, M., Williams, D.R., Kimmel, K., Polasky, S. & Packer, C. 2017. Future threats to biodiversity and pathways to their prevention. *Nature* **546**: 73–81.

Uyeda, J.C., Hansen, T.F., Arnold, S.J. & Pienaar, J. 2011. The million-year wait for macroevolutionary bursts. *Proc. Natl. Acad. Sci. U. S. A.* **108**: 15908–15913.

van der Plas, F. 2019. Biodiversity and ecosystem functioning in naturally assembled communities. *Biol. Rev.* **94**: 1220–1245.

Van Doorn, G.S. & Kirkpatrick, M. 2007. Turnover of sex chromosomes induced by sexual conflict. *Nature* **449**: 909–912.

Van Oppen, M.J.H., Turner, G.F., Rico, G., Robinson, R.L., Deutsch, J.C., Genner, M.J., *et al.* 1998. Assortative mating among rock-dwelling cichlid fishes supports high estimates of species richness from Lake Malawi. *Mol. Ecol.* **7**: 991–1001.

van Rijssel, J.C., Moser, F.N., Frei, D. & Seehausen, O. 2018. Prevalence of disruptive selection predicts extent of species differentiation in Lake Victoria cichlids. *Proc. R. Soc. B Biol. Sci.* **285**: 20172630.

Wagner, C.E., Harmon, L.J. & Seehausen, O. 2014. Cichlid species-area relationships are shaped by adaptive radiations that scale with area. *Ecol. Lett.* **17**: 583–592.

Wagner, C.E., Harmon, L.J. & Seehausen, O. 2012. Ecological opportunity and sexual selection together predict adaptive radiation. *Nature* **487**: 366–369.

Wagner, C.E., Keller, I., Wittwer, S., Selz, O.M., Mwaiko, S., Greuter, L., *et al.* 2013. Genome-wide RAD sequence data provide unprecedented resolution of species boundaries and relationships in the Lake Victoria cichlid adaptive radiation. *Mol. Ecol.* **22**: 787–798.

Waters, C.N., Zalasiewicz, J., Summerhayes, C., Barnosky, A.D., Poirier, C., Gałuszka, A., *et al.* 2016. The Anthropocene is functionally and stratigraphically distinct from the Holocene. *Science* **351**: aad2622-1-aad2622-10.

Werren, J.H. & Beukeboom, L.W. 1998. Sex determination, sex ratios, and genetic conflict. *Annu. Rev. Ecol. Syst.* **29**: 233–261.

Wilson, E.B. 1907. The supernumerary chromosomes of Hemiptera. *Science.* **26**: 870–871.

Witte, F. & Van Oijen, M.J.P. 1990. *Taxonomy, ecology and fishery of Lake Victoria haplochromine trophic groups*.

Yeaman, S. 2013. Genomic rearrangements and the evolution of clusters of locally adaptive loci. *Proc. Natl. Acad. Sci.* **110**: E1743–E1751.

Yeaman, S. & Whitlock, M.C. 2011. The genetic architecture of adaptation under migration-selection balance. *Evolution* **65**: 1897–1911.

Chapter 1 –

Genetic architecture of a key reproductive isolation trait differs between sympatric and non-sympatric sister species of Lake Victoria cichlids

PROCEEDINGS B

royalsocietypublishing.org/journal/rspb

Research

Cite this article: Feller AF, Haesler MP, Peichel CL, Seehausen O. 2020 Genetic architecture of a key reproductive isolation trait differs between sympatric and non-sympatric sister species of Lake Victoria cichlids. *Proc. R. Soc. B* **287**: 20200270.
http://dx.doi.org/10.1098/rspb.2020.0270

Received: 10 February 2020
Accepted: 17 March 2020

Subject Category:
Evolution

Subject Areas:
evolution, genetics

Keywords:
Lake Victoria, cichlids, sympatric speciation, genetic architecture, male coloration

Author for correspondence:
Anna F. Feller
e-mail: anna.feller@eawag.ch

Electronic supplementary material is available online at https://dx.doi.org/10.6084/m9.figshare.c.4902459..

THE ROYAL SOCIETY PUBLISHING

Genetic architecture of a key reproductive isolation trait differs between sympatric and non-sympatric sister species of Lake Victoria cichlids

Anna F. Feller[1,3], Marcel P. Haesler[1,3], Catherine L. Peichel[2] and Ole Seehausen[1,3]

[1]Division of Aquatic Ecology and Evolution, Institute of Ecology and Evolution, and [2]Division of Evolutionary Ecology, Institute of Ecology and Evolution, University of Bern, 3012 Bern, Switzerland
[3]Department of Fish Ecology and Evolution, Centre of Ecology, Evolution and Biogeochemistry, EAWAG Swiss Federal Institute of Aquatic Science and Technology, 6047 Kastanienbaum, Switzerland

AFF, 0000-0001-5786-7658; CLP, 0000-0002-7731-8944; OS, 0000-0001-6598-1434

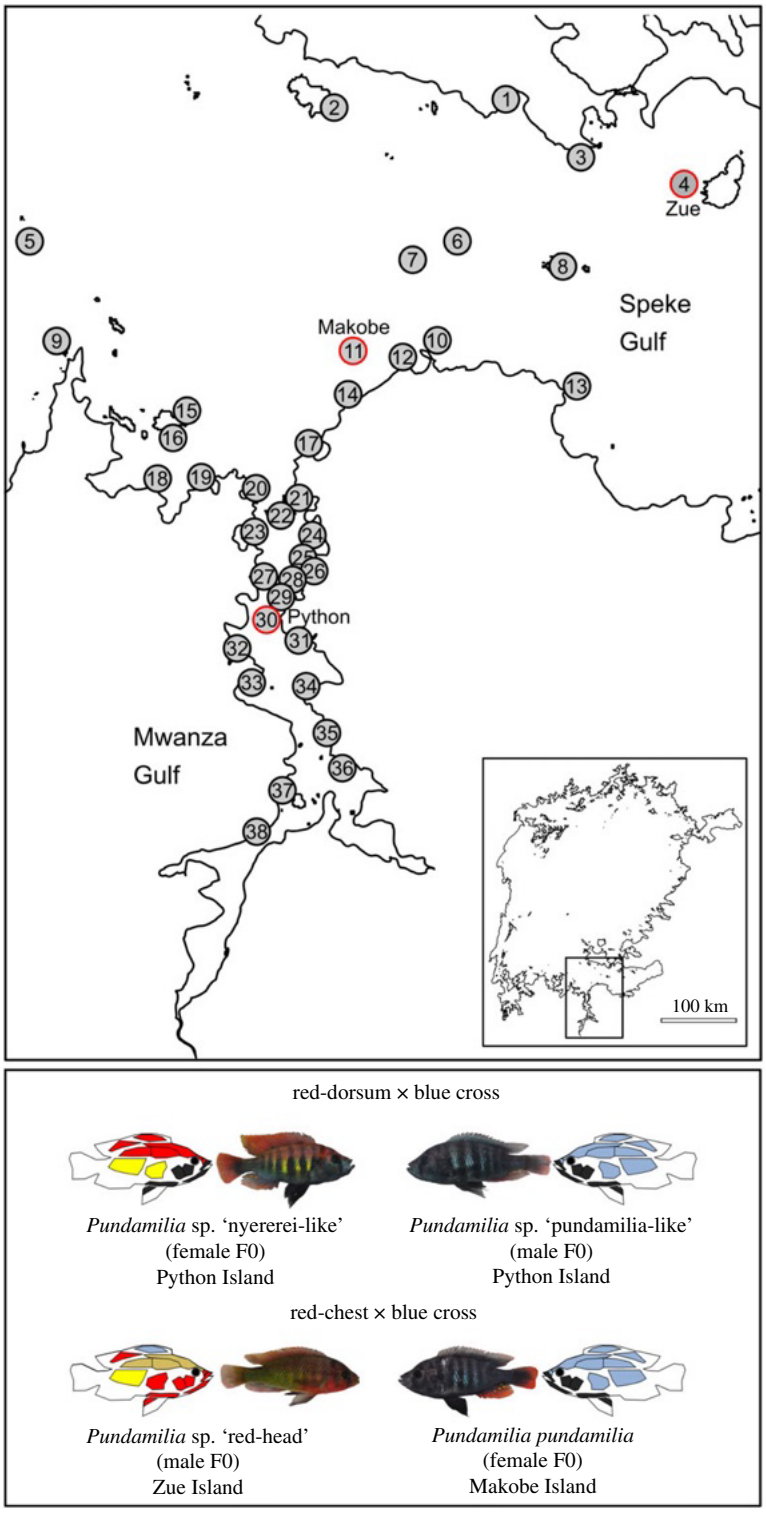
One hallmark of the East African cichlid radiations is the rapid evolution of reproductive isolation that is robust to full sympatry of many closely related species. Theory predicts that species persistence and speciation in sympatry with gene flow are facilitated if loci of large effect or physical linkage (or pleiotropy) underlie traits involved in reproductive isolation. Here, we investigate the genetic architecture of a key trait involved in behavioural isolation, male nuptial coloration, by crossing two sister species pairs of Lake Victoria cichlids of the genus *Pundamilia* and mapping nuptial coloration in the F2 hybrids. One is a young sympatric species pair, representative of an axis of colour motif differentiation, red-dorsum versus blue, that is highly recurrent in closely related sympatric species. The other is a species pair representative of colour motifs, red-chest versus blue, that are common in allopatric but uncommon in sympatric closely related species. We find significant quantitative trait loci (QTLs) with moderate to large effects (some overlapping) for red and yellow in the sympatric red-dorsum × blue cross, whereas we find no significant QTLs in the non-sympatric red-chest × blue cross. These findings are consistent with theory predicting that large effect loci or linkage/pleiotropy underlying mating trait differentiation could facilitate speciation and species persistence with gene flow in sympatry.

1. Background

The adaptive radiation of Lake Victoria haplochromine cichlids comprises approximately 500 endemic species that have evolved within the lake in perhaps as little as 15 000 years [1–4] and that are highly diverse in morphology, ecology, colour and behaviour [5–7]. Typically, however, closely related species are similar in morphology and ecology while they differ dramatically in male nuptial coloration [6,8,9]. Male nuptial coloration is considered a trait of key importance in the origin and maintenance of new species in these fish [9–11].

A highly recurrent pattern in male nuptial colour variation in pairs of closely related species of Lake Victoria cichlids is that males in one species are blue-grey on their body with any red colour confined to the fins, whereas males of the other species are yellow-red on the body [9–12]. The red colour can be confined to either dorsal parts of the body ('red-dorsum' type) or to the chest and lower head ('red-chest' type). When closely related species are sympatric, the red form generally has a red-dorsum, whereas many non-sympatric pairs of closely related species involve a red-chest and a blue form [6,8,13].

A representative case of sympatric red-dorsum and blue sister species is the young species pair of *Pundamilia* sp. 'nyererei-like' and *Pundamilia* sp.



location number	polymorphic	*P. nyererei*	*P. pundamilia*	*P.* sp. 'red-head'	location name
1			X		Nansio Island
2			X		Bwiru Island
3			X		Namatembi Island
4				X	Zue Island
5			X		Miandere Island
6				X	Mabibi Islands
7		X	X	*	Ruti Island
8		X	X		Vesi Islands
9			X		Bihiru Island
10			X		Ndurwa Point
11		X	X		Makobe Island
12		X	X		Igombe Island
13		X	X		Senga Point
14			X		Mponze Point
15		X	X		Juma Island
16			X		Juma Island
17			X		Bwiru Point
18			X		Wakahangara Point
19			X		Amranda Point
20		X	X		Karomo Bay
21		X	X		Gabalema Island
22		X	X		Hippo Island
23		X	X		Kilimo Island
24		X	X		Saa Nane Island
25		X	X		Anchor Island
26		X	X		Nyegezi Rocks
27		X	X		Ngoma Point
28		X	X		Kilomo Island
29		X	X		Nyamatala Island
30		X	X		Python Island
31	X				Shadi Rocks
32	X				Mashoro Bay
33	X				Matumbi Island
34	X				Luanso Island
35	X				small Marumbi Island
36	X				large Marumbi Island
37	X				Buyago Rocks 2
38	X				Buyago Rocks 1

Figure 1. Distribution and sampling locations of blue, red-dorsum and red-chest *Pundamilias*. The upper left panel shows the Mwanza gulf region in Lake Victoria (the inset figure indicates the location of this region on the full Lake Victoria map) with all sampling locations (circles with numbers). The table on the right indicates which species were recorded at each sampling location (data assembled from [6,8,13,14]). *P. pundamilia* records include records of *P.* sp. 'pundamilia-like', and *P. nyererei* records include records of *P.* sp. 'nyererei-like' [15,16]. At Ruti island (*), one single *P.* sp. 'red-head' male was caught by O.S. in 2001, but no other individuals of this species were seen in many hundreds of sampled individuals at this location in 10 years before or after this. Sampling locations 31–38 harbour colour-polymorphic *Pundamilia* populations, where species are not clearly identifiable. Highlighted in red are the catch locations of the four parental species of the two crosses, see lower left panel showing representative male individuals of these four species and their colour phenotypes. Lake Victoria contours were taken from the Lake Victoria Bathymetric map V6 by [17] and smoothed in Inkscape™ 0.92.

'pundamilia-like' at Python Island (figure 1). This pair may have evolved in sympatry from a hybrid population in only a few hundred generations [15,16,18]. *Pundamilia* sp. 'red-head' and *Pundamilia pundamilia* are representatives of a red-chest and blue species with overlapping geographical distributions that are never found together on the same rocky island, despite islands occupied by the different species lying within dispersal distance of each other (figure 1) [6,8,19].

Male nuptial colour in *Pundamilia* is under intra- and intersexual selection. Males fiercely compete for territories, and they use their bright colours to signal territory ownership to contestants. Negative frequency-dependent selection

generated by own-type aggression biases is likely involved in stabilizing the coexistence of different colour types [11,20–22]. Females, which are cryptically coloured and the sole investors into parental care, exhibit strong preferences for mates with particular nuptial coloration, which generates both directional sexual selection within and assortative mating between species [23–26]. Preference for male colour has been shown to be heritable and likely determined by few major genes or genomic regions [27,28] and to generate disruptive selection on male colour [29].

Despite strong preferences for bright male colour and assortative mating, there is some gene flow in fully sympatric red-dorsum versus blue pairs such as *Pundamilia* sp. 'nyererei-like' and *Pundamilia* sp. 'pundamilia-like' [15,16]. Nonetheless these two species persist in sympatry. With a genome-wide mean FST of 0.053 and hundreds of highly differentiated genomic regions, they show surprisingly strong genetic differentiation considering they have likely evolved in full sympatry in less than 200 generations [15,16]. Red-chest versus blue pairs like *Pundamilia* sp. 'red-head' and *Pundamilia pundamilia* experience little gene flow because they usually do not co-occur. Although they are neither strictly geographically isolated nor dispersal limited [6,8,19] (and figure 1), they do not seem to persist as two species in the same island.

Here we ask whether differences in the genetic architecture of the male nuptial colour motifs that differentiate these species could explain the difference in the distribution patterns of the species pairs. We tested the hypothesis that red-dorsum versus blue pairs persisting in sympatry despite some gene flow have an architecture that is robust against potentially homogenizing effects of gene flow, while the absence of such an architecture would make it difficult for red-chest versus blue pairs to persist in the presence of gene flow.

In a sympatric scenario with ongoing gene flow, recombination is expected to break up linkage disequilibrium between favourable combinations of alleles for local adaptation and reproductive isolation [30,31]. This is less likely to occur if divergently selected traits are coded by few genes with large effects, as this reduces the number of targets for recombination to break up, and increases the effectiveness of (correlational) selection because it is concentrated on fewer targets [32]. Indeed, theoretical work has shown that large effect alleles, or groups of tightly linked alleles with smaller effects that then act like a large effect locus, are less likely to be lost when adaptive multilocus phenotypes need to be maintained against gene flow [33–35]. Similarly, the physical linkage of multiple traits (or pleiotropy) could facilitate divergence with gene flow [34,36,37].

Recombinant males between cichlid (*Pundamilia*) species with either red or blue male nuptial coloration might have reduced fitness if they will be less likely to be chosen by females with strong preferences for either colour, as is the case in these species [23–28], or if intermediate coloration makes them targets of territorial aggression by males of both species. We thus predict large effect loci and/or physical linkage of several loci (or pleiotropy) for male nuptial colour to be present in sympatric red-dorsum versus blue cichlid pairs, as such an architecture could both facilitate the establishment of polymorphisms [38] and make it easier to retain phenotypic differentiation in sympatry with ongoing gene flow (see above). Furthermore, theoretical models that investigate the feasibility of sympatric speciation by sexual selection (alone) usually find such speciation feasible when assuming a simple genetic architecture (i.e. few additive loci with large effects) and that speciation becomes less likely when the number of loci underpinning reproductive isolation traits increases [39–41].

To compare the genetic architecture of red and yellow versus blue male nuptial colour motifs that do or do not persist with gene flow, we crossed *Pundamilia* sp. 'nyererei-like' and *Pundamilia* sp. 'pundamilia-like' (red-dorsum × blue) and *Pundamilia* sp. 'red-head' and *Pundamilia pundamilia* (red-chest × blue) in the laboratory and performed quantitative trait locus (QTL) mapping analyses on male nuptial coloration in the second generation (F2) hybrids.

We find several significant QTLs for the presence/absence of red and yellow colour in the sympatric red-dorsum × blue cross, with several traits mapping to the same region, whereas we find no significant QTLs in the non-sympatric red-chest × blue cross. We conclude that the presence of large effect loci, with physical linkage between some traits (or pleiotropy), likely makes up one key element for the rapid evolution of reproductive isolation and species persistence in sympatry despite some gene flow.

2. Material and methods

(a) Experimental crosses

The red-dorsum × blue cross was started with a *Pundamilia* sp. 'nyererei-like' female and a *Pundamilia* sp. 'pundamilia-like' male, both from laboratory bred strains established from fishes caught by OS at Python Island in Lake Victoria in 2003 (figure 1). The red-chest × blue cross was started with a *Pundamilia pundamilia* female caught by O.S. at Makobe Island in 2003 and a *Pundamilia* sp. 'red-head' male from a laboratory bred strain established from fishes caught by O.S. at Zue Island in 1993 (figure 1). Five to 6 days after spawning, the eggs were removed from the female's mouth and reared in an egg tumbler until hatching. After yolk sac resorption, the larvae were transferred to rearing aquaria that were part of a large recirculation system. After sexual maturity at the age of 1 to 2 years, two pairs of F1 individuals of each cross were then allowed to mate, and the eggs and larvae were reared the same way as the F1 generation. Because the average clutch size is just about 20–30 juveniles, we re-mated each F1 pair multiple times until we had obtained a total of approximately 300 F2 individuals in each cross. This procedure of repeatedly re-mating the same pairs took about 2 years. All F2s were reared to an age of at least 1 year before they were phenotyped. All four populations from which the grandparents were taken breed true in the laboratory in a common garden environment, hence the differences in coloration are heritable in both crosses.

(b) Colour photos and scoring of coloration

Sexually mature F2 males were removed from their holding tank and individually placed into one of six adjacent plexiglas photo cuvette compartments with transparent separations and a grey PVC background, inside a larger aquarium. This set-up induced territoriality in the males who could see each other and hence make them express full colour. Two Walimex pro Daylight 250S lamps were placed on either side of the cuvette and a first colour picture was taken of each fish with a Canon D60 camera equipped with a 50 mm lens (settings; P mode, ISO 200, auto focus). If a male failed to show territorial behaviour within an adjustment time of 1–2 days, it was moved back into its home tank and the process was repeated several weeks later. After good photos were obtained, each fish was sedated in MS222

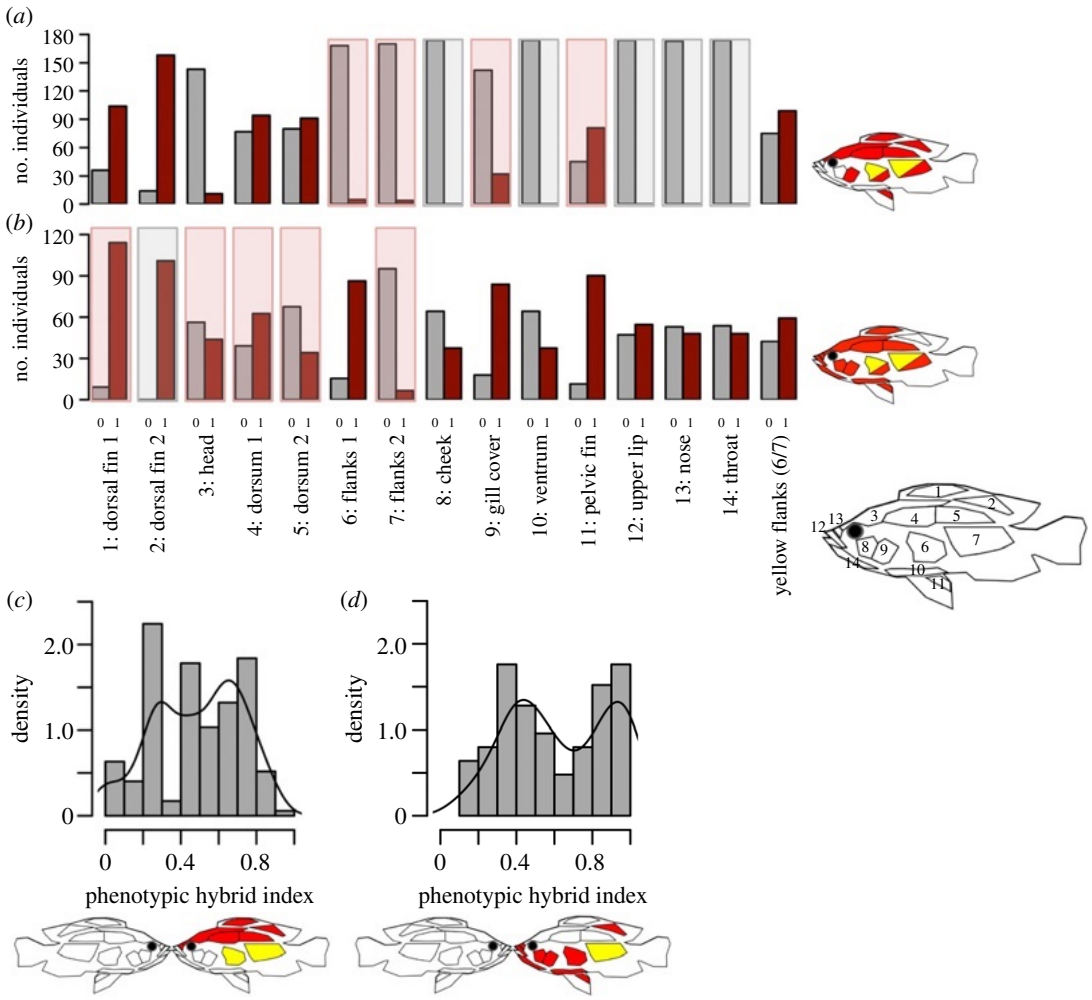


Figure 2. Phenotypic distributions in the two crosses. Distribution of absence (0; grey)/presence (1; red) scores for all traits in the (*a*) red-dorsum × blue cross and (*b*) red-chest × blue cross. A grey background indicates lack of variation in the F2 individuals in this trait (i.e. non-mappable traits), a red background indicates traits that are not differentiated between the two parental species but were variable in the F2 hybrids (transgressive traits). The sectors highlighted with colour on the inset cartoon fish are the traits that were mappable in the cross. Half red/half yellow indicates either colour could occur in this trait. (*c*,*d*) The distribution of the standardized phenotypic hybrid index, for the red-dorsum × blue cross (*c*) and for the red-chest × blue cross (*d*), which was calculated for every individual as the sum of presence scores in all traits differentiating the two parental species divided by the number of differentiating traits (see parts highlighted with colour in the inset fish figures). An index of 0 hence corresponds to the phenotype of the blue species, an index of 1 corresponds to the phenotype of the red species. Note that no red was scored on the nose on any F2 individual in the red-dorsum × blue cross. The comparison of the cartoon fish indicating mappable traits in (*a*,*b*) with the cartoon fish showing the parental types in (*c*,*d*) highlights which traits are transgressive in the F2 hybrids.

(50 mg l^{-1}), euthanized in MS222 (300 mg l^{-1}) and immediately put on melting ice for 10 min. This treatment had the effect of relaxation of the melanophores of system I that are under nervous control [42] and made colour better visible. The fish were then pinned down in a dissection bowl containing a wax layer and a ruler with a colour reference, and submerged in water (just covering the fish), and a second colour picture was taken from above with the same camera (same settings), now equipped with an 18–135 mm objective and with one Walimex pro Daylight 250S lamp for lighting. We applied a white balance filter to these second photos in Adobe Photoshop CC 2018 by setting a white point on the white part of the colour bar ruler.

We visually scored the presence/absence of red on 14 sectors of the body (henceforth referred to as traits) that are different between at least two of the parental species (see figures 1 and 2), and the presence/absence of yellow flanks (as 15th trait) from the second photo (when the fish was dead). The first pictures were only used as a cross-reference to confirm uncertain scores. In some individuals (mostly of the red-dorsum × blue cross), some traits could not be scored unambiguously and were scored as NA. In addition, we calculated a phenotypic hybrid index for each individual as the summed up presence scores (NAs treated as 0.5) in the subset of traits that differentiate the two respective parental species divided by the number of traits in the respective subset (figure 2*c*,*d*). The phenotypic hybrid index thus ranges from 0 (all traits like those of the blue species) to 1 (all traits like those of the red species).

(c) DNA extraction and RAD sequencing

Genomic DNA was extracted from finclips, which had been stored in 98% ethanol, using phenol–chloroform [43]. We prepared restriction-site associated DNA (RAD) sequencing libraries using *SbfI* as restriction enzyme, following [44] with some modifications (see electronic supplementary material, appendix S1). Single end-sequencing (100 bp for all others but 125 bp for the last two libraries, see electronic supplementary material, table S5) was done on an Illumina HiSeq 2500 platform either at the Genomic Technologies Facility of the University of Lausanne or at the Next Generation Sequencing Platform of the University of Bern. Each library was sequenced on a single lane. Bacteriophage PhiX genomic DNA was added to each library (4–12% of reads) to increase complexity in the first 10 sequenced base pairs and for base quality recalibration (see below).

(d) Sequence processing and genotyping

Demultiplexed, trimmed and filtered reads (see appendix S2 for details) were aligned to the anchored version of the *Pundamilia*

nyererei reference genome [45] with Bowtie2 v2.3.2 [46], allowing one mismatch. This was followed by base quality recalibration (see electronic supplementary material, appendix S2 for details) and subsetting to uniquely aligned reads. GATK Unified Genotyper [47] was used for genotyping (minimum base quality score set to 20). The resulting vcf files were filtered with Bcftools implemented in Samtools v.1.8 [48], Vcftools v.0.1.14 [49] and a custom Python script, to obtain bi-allelic SNPs (see electronic supplementary material, appendix S2 for details). In the red-dorsum × blue cross, 368 SNPs were homozygous fixed in the F0 and heterozygous in all four F1 parents, and were used in linkage map construction. In the red-chest × blue cross, 2358 SNPs were homozygous fixed in the F0 and heterozygous in two F1 parents (the other two F1s had low quality data that was removed during filtering), and were used in linkage map construction.

(e) Linkage map construction

We used JoinMap 4.0 [50] to build linkage maps for both crosses. We removed loci with extreme segregation distortion ($p < 0.01$), loci with greater than 20% missing genotypes and identical loci (i.e. SNPs within the same RAD locus) (greater than 0.950). Individuals were removed if they had greater than 30% missing data. The linkage maps were generated from 216 F2 individuals (173 males, 43 females) in the red-dorsum × blue cross and 171 F2 individuals (115 males, 56 females) in the red-chest × blue cross. We identified linkage groups based on an independent logarithm of odds (LOD) threshold of 5. Loci with suspicious linkage (recombination frequency greater than 0.6) were removed. The strongest cross-link (SCL) values in the maps are 4.7 (red-dorsum × blue cross) and 4.6 (red-chest × blue cross), and unlinked markers were excluded. To build the linkage maps, we used the Kosambi regression mapping algorithm with a LOD threshold of 1.0, a recombination threshold of 0.499, a goodness-of-fit threshold of 5.0 and no fixed order. We performed two rounds of mapping with a ripple after addition of each marker to the map (see [50]).

(f) QTL mapping

QTL mapping of male nuptial colour traits was performed in R/qtl [51,52]. We mapped the presence/absence of yellow colour on the flanks and of red colour in 14 body and fin locations (traits; see above and figure 2) in F2 males ($n = 174$ in the red-dorsum × blue cross and $n = 125$ in the red-chest × blue cross) as binary traits. Conditional genotype probabilities were calculated using the calc.genoprob function with a fixed stepsize of 1 (respectively 3 for scantwo; in cM), an assumed genotyping error rate of 0.05, and the Kosambi map function. The scanone and scantwo functions were used with the EM algorithm, and significance thresholds were determined by permutations ($n = 1000$). For the red and yellow traits, we used the binary model; for a multi-trait hybrid index (see below), we used the normal model. We consider $p < 0.05$ as significant, and $p < 0.1$ as marginally significant. For traits showing a significant effect of F1 family, mapping was additionally performed with family as an additive covariate (allowing the average phenotype in the two families to be different) and as both an additive and interactive covariate (additionally allowing the effect of the QTL between the two families to be different). For significant QTLs, approximate Bayesian credible intervals with a 0.95 probability coverage were calculated using the bayesint function, and the percentage of variation explained (PVE) of significant QTLs was calculated for each trait individually using the fitql function (with the HK algorithm since the EM algorithm was not available in this function for binary models).

The lower number of individuals in the red-chest × blue cross could result in reduced power of detecting QTLs of similar effect sizes as compared to the red-dorsum × blue cross, and the two linkage maps differ substantially in number and density of markers. To account for this, we repeated the single QTL analyses for the red-dorsum × blue cross after randomly downsampling to 125 (of 174) individuals using the sample function in R to randomly pick individuals and then subsetting the genotype–phenotype file to these individuals. For the red-chest × blue cross, we repeated the single QTL analyses after randomly downsampling markers on the linkage map to match the number of markers on each linkage group to those in our sparser red-dorsum × blue map, again using the sample function in R to randomly pick markers within each linkage group and then subsetting them to these markers. The procedure was repeated five times in each cross.

3. Results

(a) Linkage maps

The final map for the red-dorsum × blue cross contains 232 markers in 22 linkage groups (corresponding to the number of expected chromosomes in haplo-tilapiine cichlids [53,54]) with an average marker distance of 5.3 cM and a total map length of 1117.8 cM. The final map for the red-chest × blue cross contains 1198 markers in 22 linkage groups with an average marker distance of 1.2 cM and a total map length of 1360.8 cM.

(b) Phenotypic distributions

Seven traits differentiate the parental species in the red-dorsum × blue cross (figure 2*c*): both parts of the dorsal fin, both parts of the dorsum, head and nose are red versus blue, and both parts of the flank (considered as one trait) are yellow versus blue. However, nose was never scored as red in any F2. Red was also scored in some F2s for both parts of the flank, for gill cover, and for pelvic fin, although neither parental species has red in these traits (figures 1 and 2*c*). Altogether this resulted in 10 mappable traits for this cross (figure 2*a*). The F2 phenotypes range from no red or yellow, respectively, in any of the traits that differentiate the parental species (i.e. like *P.* sp. 'pundamilia-like'), to red or yellow, respectively, in all of these traits with the exception of nose (i.e. like *P.* sp. 'nyererei-like'), with most individuals being intermediate in expressing red in some but not all of these traits (figure 2*c*).

Ten traits differentiate the parental species in the red-chest × blue cross (figure 2*d*): the rear dorsal fin part, the frontal flank part, the pelvic fin and all parts on the head (except for the head part itself, which like the dorsum is greenish versus blue) are red or yellow (flanks) versus blue. The rear dorsal fin part was scored as red in all F2s and can hence not be mapped. Some F2s were scored as red on the front part of the dorsal fin as well as on both parts of the dorsum, on the head and on the rear flank part, even though neither parental species is red there (figures 1 and 2*d*). Altogether this resulted in 14 mappable traits for this cross (figure 2*b*). The F2 phenotypes range from no red or yellow, respectively, in any of these traits with the exception of the rear dorsal fin part (i.e. like *P. pundamilia*), to red or yellow, respectively, in all these traits (i.e. like *P.* sp. 'red-head'), with most individuals being intermediate in expressing red in some but not all of these traits. Yet, a large number of these F2 hybrids resemble *P.* sp. 'red-head' whereas fewer resemble *P. pundamilia*, suggesting more directional dominance effects in this cross than in the other cross.

(c) QTL mapping results

We found significant QTLs for red and yellow colour for seven out of 10 mappable traits (figure 2*a*) in the sympatric red-

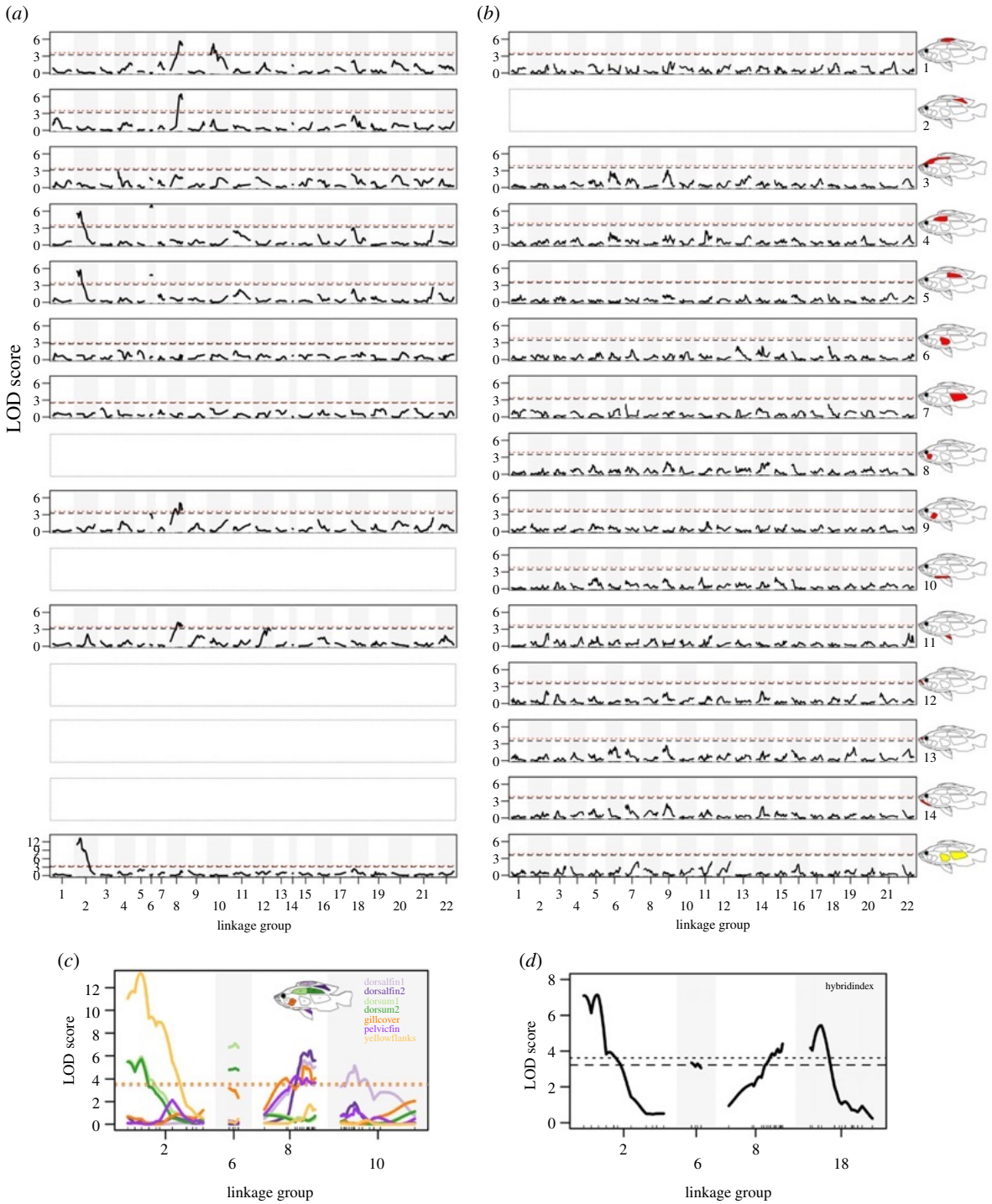


Figure 3. QTL mapping of the presence/absence of red (and yellow) male nuptial colour. LOD scores across the 22 *Pundamilia* chromosomes for all traits with presence scores for red and for yellow flanks (empty plots for non-mappable traits, figure 2) in (*a*) the red-dorsum × blue cross and (*b*) the red-chest × blue cross. The black dashed lines represent genome-wide significance thresholds of $p < 0.1$ for each trait, the red dotted lines for $p < 0.05$. (*c*) LOD scores across the chromosomes containing significant QTLs for red and yellow in the red-dorsum × blue cross (genome-wide significance thresholds of $p < 0.05$ shown for each trait); (*d*) shows the chromosomes containing significant QTLs for the hybrid index in the red-dorsum × blue cross. See also electronic supplementary material, table S1 and figure S1.

dorsum × blue cross (electronic supplementary material, table S1 and figure S1; figure 3*a*,*c*). For red in both sectors of the dorsal fin, the gill cover and the pelvic fin, we identified QTLs on Pun-LG8 with overlapping 95% confidence intervals, each with a PVE of 12.6–16.8%. A second QTL for red on dorsal fin sector 1 was found on Pun-LG10 with a PVE of 15.5%, and a second marginally significant QTL for red on the pelvic fin was found on Pun-LG12 with a PVE of 10.8%. Two QTLs for red on both dorsum parts were identified, one on Pun-LG2 and one on Pun-LG6, both with a PVE of 12.4–17.4%. Pun-LG2 also contains a QTL for yellow flanks, within the 95% confidence mapping interval for red-dorsum, with a PVE of 29.5%. In this cross, only red in the pelvic fin (a transgressive trait in this cross) showed a significant effect of family ($F = 26.901$, $p < 0.001$). Repeating the mapping with family as covariate for pelvic fin recovered the QTL on Pun-LG8 and revealed an additional marginally significant QTL on Pun-LG22. We also found four QTLs for the phenotypic hybrid index (figure 3*d*), one each on LG2 (17.2% PVE), LG6 (8.5% PVE), LG8 (10.97% PVE), all of them within the 95%

confidence intervals of the QTLs found for the individual traits and one on LG18 (13.3% PVE).

Two-dimensional two QTL scans revealed several additional putative QTLs in this cross (electronic supplementary material, table S2). Repeating the single QTL mapping with 125 individuals randomly sampled from 174 five times recovered 82% of the expected total of 50 significant QTL results (i.e. 10 significant QTLs in the original dataset times five); 36 were recovered as significant and five as marginally significant (electronic supplementary material, table S3).

We found no significant QTLs for any of the 14 mappable traits (figure 2*b*) nor for the hybrid index in the non-sympatric red-chest × blue cross (electronic supplementary material, table S1; figure 3*b*). Several traits showed a significant effect of family in this cross: cheek ($F = 16.478$, $p < 0.001$), ventrum ($F = 29.407$, $p < 0.001$), upper lip ($F = 29.079$, $p < 0.001$), nose ($F = 18.232$, $p = 0.001$), throat ($F = 46.281$, $p < 0.001$) and yellow flanks ($F = 5.6352$, $p = 0.025$). Repeating the mapping with family as covariate for these traits; however, only detected one marginally significant QTL for red on the throat on Pun-LG7 with a PVE of 8.3%. A QTL for red on the head (a transgressive trait in this cross) reached the 0.1 significance threshold when repeating the mapping with a subsampled linkage map in two out of five such mapping rounds (electronic supplementary material, tables S1 and S4). Two-dimensional two QTL scans for this cross revealed two potentially interacting QTLs each for red on the cheek and throat (Pun-LG1 and Pun-LG4 for both traits) and for yellow on the flanks (Pun-LG3 and Pun-LG11) (electronic supplementary material, table S2).

4. Discussion

We investigated the genetic architecture of a trait complex of key importance to speciation in Lake Victoria cichlid fish, male nuptial colour motifs that feature importantly in behavioural reproductive isolation. In a cross between two sympatric species, representative in their mating trait motifs of many closely related sympatric species pairs, we found significant QTLs with moderate to large effects for red and yellow colour traits, with several traits mapped to the same genomic regions. These results are consistent with genetic architectures predicted to facilitate differentiation and persistence of differentiation in traits contributing to reproductive isolation in sympatry with ongoing gene flow [32–34,36–38]. By contrast, we did not find any significant QTLs in a cross between two species representative in their mating trait motifs of closely related species that are usually seen to occupy different islands but do not occur in sympatry. This is consistent with our hypothesis of a genetic architecture that makes phenotypic differentiation not robust to gene flow. We argue that these differences in genetic architecture of superficially similar trait differences could help explain why species with the red-dorsum nuptial colour motif are often sympatric with blue sister species, whereas those with the red-chest motif seem unable to retain differentiation from their blue relatives in sympatry.

The difference between the two crosses in the presence/absence of QTLs with moderate to large effects cannot simply be explained by a difference in power to detect QTLs: repeatedly and randomly downsampling the red-dorsum × blue cross F2 individuals to match the lower sample size in our red-chest × blue cross (a lower sample size decreases power to detect QTLs) did not significantly change the results, nor did downsampling the markers on the linkage map in the red-chest × blue cross to match the sparser red-dorsum × blue cross-linkage map (a sparser marker density lowers the significance threshold). Another statistical bias [55], where low sample sizes lead to overestimation of effect sizes or PVE, mainly due to the difficulties of statistically detecting loci with small effects, cannot be ruled out. However, this would be expected to affect both of our crosses and should thus not confound the comparison between the two. The different direction of our two crosses (red female × blue male in the red-dorsum × blue cross, blue female × red male in the red-chest × blue cross) should also not affect our results: the lab-strain populations from which the grandparents were taken have stable male colour. Also, none of the QTLs map to known sex determining chromosomes (Pun-LG10 in the red-chest × blue cross [45]), which is also consistent with earlier studies of experimental crosses of the same red-dorsum versus blue species pair [56–58]. Furthermore, in both crosses, both parental phenotypes (considering traits differentiating the two species, figure 2*c*,*d*) are recovered (with the exceptions that one trait (nose) was never scored as red in any F2s in the red-dorsum × blue cross, and one trait (dorsal fin 2) was scored as red in all F2s of the red-chest × blue cross). Additionally, most individuals are intermediate in the number of traits in which red is expressed in both crosses, albeit with signs of more red dominance in the red-chest × blue cross.

A previous crossing experiment [56] estimated that the difference in the amount of red on the body (mostly flank and dorsum) of males between *Pundamilia* sp. 'nyererei-like' and *Pundamilia* sp. 'pundamilia-like' (our red-dorsum × blue cross) is likely controlled by at least 2–4 loci (with effects of dominance and epistasis). Furthermore, they estimated one gene with complete dominance for yellow flank and epistatic interaction with red on flank and dorsum. Our results conform to these estimates quite well: we found two significant QTLs each for red on the dorsum (LG2 and LG6) and for red on the dorsal fin (LG8 and LG10), each with a PVE of 12–17%. The interval of the QTL for red on the dorsum on LG2 also contains a major QTL for yellow on the flanks with a PVE of nearly 30%, and the significant QTLs for pelvic fin and gill cover overlap with the interval of the QTL on LG8 for red on the dorsal fin, suggesting either linkage of several loci or pleiotropic effects of a single locus. Two-dimensional 2-QTL scans indicate the presence of additional loci contributing to red/yellow colour. Although our modest sample sizes do not allow us to detect small effect QTLs, they are likely present, as our QTLs do not explain all of the variance in our mapped traits. However, the main contribution to variance in red and yellow male colour in this cross comes from these four QTL regions. A first screen of the genes closest to the QTL peaks has not yet revealed any obvious candidate genes. To follow-up on screening for candidate genes across the mapping intervals, each of which contains many dozens to hundreds of genes, will be a topic of future work.

The presence of moderate to large effect QTLs should make phenotypic differentiation and maintenance of differentiation in sympatry more likely than a more dispersed architecture of many loci with small effects under the opposing effects of disruptive selection and gene flow e.g. [32–34,38]. Furthermore, the theoretical models of sympatric

speciation by sexual selection suggest such a process is more likely when reproductive isolation is based on traits underlain by fewer loci [39–41]. Additionally, we find that several traits (figure 3*c*) map to the same chromosomal region. Although our current dataset does not allow us to determine whether this is due to a single pleiotropic locus or due to several tightly linked loci, both pleiotropy and physical linkage favour divergence and persistence of phenotypic differentiation despite gene flow [34,36,37]. Linkage (or pleiotropy) of divergent adaptive traits has also been observed in Midas cichlids, which are undergoing rapid sympatric divergence [59]. Our findings are also similar to other systems such as *Heliconius* [60], *Mimulus* [61] or *Drosophila* [62], in which traits involved in reproductive isolation in the presence of gene flow are underpinned by large effect or pleiotropic loci. However, none of these other studies made a direct comparison of the genetic architecture of corresponding key traits for reproductive isolation in species pairs that persist (and probably evolved) in sympatry and others that do not persist in sympatry, as we have done here.

In our cross between the species that do not persist in sympatry, i.e. *Pundamilia* sp. 'red-head' and *Pundamilia pundamilia* (red-chest × blue), we cannot directly infer the type of genetic architecture underpinning red, or yellow due to the absence of any significant QTLs. Most likely, the effects of potential QTLs are too small to be detected with our sample size, suggesting the presence of a larger number of small effect loci. In red-chest versus blue species without direct geographical contact and with little gene flow that is restricted to rare (but documented, figure 1) long distance dispersal events, trait differentiation due to more and smaller effect mutations is more likely to evolve and persist because recombination will not erode associations between them.

What we cannot yet resolve is whether the genetic architecture for male colour in our red-dorsum versus blue pair has evolved in the face of gene flow, i.e. through selection for the clustering of small effect loci, for instance through genomic rearrangements [63], or was already present, allowing the two species to speciate (and now persist) in sympatry.

5. Conclusion

The presence of large effect loci, and of physical linkage or pleiotropy, underlying traits involved in behavioural reproductive isolation, such as male nuptial coloration, may enable sister species pairs to differentiate and persist in sympatry despite some gene flow. One hallmark of the East African cichlid radiations is the rapid evolution of strong (behavioural) reproductive isolation that is robust to full sympatry in many closely related species [64]. If the genetic architecture of male nuptial coloration in sympatric *Pundamilia* species we report here is representative for other Lake Victoria cichlid species that live sympatrically, this may help explain how speciation in this system could have led to the rapid emergence of communities with many closely related species that persist in sympatry, and why some phenotypic motifs regularly distinguish sympatric species while others are confined to allopatric species.

Ethics. Fish experimentation and euthanasia were authorized by the veterinary offices of the cantons of Lucerne (licence number: LU04/07) and Bern (BE 18/15 & BE 65/18).

Data accessibility. Raw read (fastq) files for all genotyped individuals of the red-dorsum × blue cross are deposited in the NCBI short read archive (SRA accession PRJNA612290). The fastq files for the red-chest × blue cross have previously been deposited by [45] (SRA accession PRJNA439430). The filtered vcf files, linkage maps, phenotype–genotype tables, and additional information on sequencing data for both crosses are available from the Dryad Digital Repository at https://dx.doi.org/10.5061/dryad.5tb2rbp1n [65].

Authors' contributions. A.F.F. scored and analysed colour in both crosses, prepared the RAD sequencing libraries for the red-dorsum × blue cross, processed the sequencing data, generated linkage maps, performed the colour QTL analyses in both crosses, wrote the manuscript together with O.S. M.P.H. reared the fish, took the colour photos, prepared all the red-chest × blue and two of the red-dorsum × blue cross RAD sequencing libraries and provided critical feedback for the manuscript. C.L.P. provided guidance for the analyses and contributed to manuscript writing. O.S. designed and coordinated the study, developed the colour scoring scheme and the colour trait space, collected the species distribution data, wrote the manuscript together with A.F.F.

Competing interests. We declare we have no competing interests.

Funding. This work was funded by Swiss National Science Foundation (SNSF) grant nos. 31003A_163338, 31003A_144046 and 31003A_118293 to O.S.

Acknowledgements. We thank Julia Schwarzer and Philine Feulner for their earlier and preliminary work on the red-chest × blue cross and Oliver Selz for initiating the red-dorsum × blue cross. Many thanks to Salome Mwaiko for assistance in the laboratory, the FishEc team for helpful discussions and feedback throughout the process. Joana Meier for providing the allelic balance filter script. The bioinformatics facilities for sequence data processing were provided by the Genetic Diversity Centre at ETH Zürich. We further thank two anonymous reviewers for their constructive feedback on a previous version of this manuscript.

References

1. Johnson TC, Kelts K, Odada E. 2000 The holocene history of Lake Victoria. *AMBIO A J. Hum. Environ.* **29**, 2–11. (doi:10.1579/0044-7447-29.1.2)
2. Stager JC, Johnson TC. 2008 The late Pleistocene desiccation of Lake Victoria and the origin of its endemic biota. *Hydrobiologia* **596**, 5–16. (doi:10.1007/s10750-007-9158-2)
3. Seehausen O, Witte F, Katunzi EF, Smits J, Bouton N. 1997 Patterns of the remnant cichlid fauna in southern Lake Victoria. *Conserv. Biol.* **11**, 890–904. (doi:10.1046/j.1523-1739.1997.95346.x)
4. Seehausen O. 2002 Patterns in fish radiation are compatible with Pleistocene desiccation of Lake Victoria and 14 600 year history for its cichlid species flock. *Proc. R. Soc. B* **269**, 491–497. (doi:10.1098/rspb.2001.1906)
5. Greenwood PH. 1974 *Cichlid fishes of lake Victoria, east Africa: the biology and evolution of a species flock*. Bristol, UK: John Wright and Sons Ltd, Stonebridge Press.
6. Seehausen O. 1996 *Lake Victoria rock cichlids: taxonomy, ecology, and distribution*. Zevenjuizen, NL: Verduijn Cichlids.
7. Greenwood PH. 1980 Towards a phyletic classification of the 'genus' *Haplochromis* (Pisces, Cichlidae) and related taxa. II. The species from Lakes Victoria, Nabugabo, Edward, George and Kivu. *Bull. Br. Mus. nat. Hist.* **39**, 1–101. (doi:10.5962/bhl.part.13268)
8. Seehausen O, Lippitsch E, Bouton N, Zwennes H. 1998 Mbibi, the rock-dwelling cichlids of Lake Victoria: description of three new genera and fifteen new species (Teleostei). *Ichthyol. Explor. Freshwaters* **9**, 129–228.
9. Seehausen O, Van Alphen JJM, Witte F. 1999 Can ancient colour polymorphisms explain why

some cichlid lineages speciate rapidly under disruptive sexual selection? *Belgian J. Zool.* **129**, 43–60.

10. Seehausen O, van Alphen JJM, Witte F. 1997 Cichlid fish diveristy threatened by eutrophication that curbs sexual selection. *Science* **277**, 1808–1811. (doi:10.1126/science.277.5333.1808)
11. Seehausen O, Schluter D. 2004 Male–male competition and nuptial-colour displacement as a diversifying force in Lake Victoria cichlid fishes. *Proc. R. Soc. B* **271**, 1345–1353. (doi:10.1098/rspb.2004.2737)
12. Seehausen O, van Alphen JJM. 1999 Can sympatric speciation by disruptive sexual selection explain rapid evolution of cichlid diversity in Lake Victoria? *Ecol. Lett.* **2**, 262–271. (doi:10.1046/j.1461-0248.1999.00082.x)
13. Seehausen O. 2009 Progressive levels of trait divergence along a 'speciation transect' in the Lake Victoria cichlid fish *Pundamilia*. In *Speciation and patterns of diversity* (eds RK Butlin, J Bridle, D Schluter), pp. 155–176. Cambridge, UK: Cambridge University Press.
14. Seehausen O. 1997 Distribution of and reproductive isolation among color morphs of a rock-dwelling Lake Victoria cichlid (*Haplochromis nyererei*). *Ecol. Freshw. Fish* **6**, 59–66. (doi:10.1111/j.1600-0633.1997.tb00145.x)
15. Meier JI, Sousa VC, Marques DA, Selz OM, Wagner CE, Excoffier L, Seehausen O. 2017 Demographic modelling with whole-genome data reveals parallel origin of similar *Pundamilia* cichlid species after hybridization. *Mol. Ecol.* **26**, 123–141. (doi:10.1111/mec.13838)
16. Meier JI, Marques DA, Wagner CE, Excoffier L, Seehausen O. 2018 Genomics of parallel ecological speciation in Lake Victoria cichlids. *Mol. Biol. Evol.* **35**, 1489–1506. (doi:10.1093/molbev/msy051)
17. Hamilton SE. 2016 Creation of a bathymetric map of Lake Victoria, Africa. See http://dx.doi.org/10.7910/DVN/ZSWZIA.
18. Seehausen O *et al.* 2008 Speciation through sensory drive in cichlid fish. *Nature* **455**, 620–626. (doi:10.1038/nature07285)
19. Bouton N, Seehausen O, van Alphen JJM. 1997 Resource partitioning among rock-dwelling haplochromines (Pisces: Cichlidae) from Lake Victoria. *Ecol. Freshw. Fish* **6**, 225–240. (doi:10.1111/j.1600-0633.1997.tb00165.x)
20. Dijkstra PD, Seehausen O, Groothuis TGG. 2005 Direct male–male competition can facilitate invasion of new colour types in Lake Victoria cichlids. *Behav. Ecol. Sociobiol.* **58**, 136–143. (doi:10.1007/s00265-005-0919-5)
21. Dijkstra PD, Seehausen O, Pierotti MER, Groothuis TGG. 2007 Male–male competition and speciation: aggression bias towards differently coloured rivals varies between stages of speciation in a Lake Victoria cichlid species complex. *J. Evol. Biol.* **20**, 496–502. (doi:10.1111/j.1420-9101.2006.01266.x)
22. Dijkstra PD, Hemelrijk C, Seehausen O, Groothuis TGG. 2008 Color polymorphism and intrasexual competition in assemblages of cichlid fish. *Behav. Ecol.* **20**, 138–144. (doi:10.1093/beheco/arn125)
23. Seehausen O, Van Alphen JJM. 1998 The effect of male coloration on female mate choice in closely related Lake Victoria cichlids (*Haplochromis nyererei* complex). *Behav. Ecol. Sociobiol.* **42**, 1–8. (doi:10.1007/s002650050405)
24. Selz OM, Pierotti MERR, Maan ME, Schmid C, Seehausen O. 2014 Female preference for male color is necessary and sufficient for assortative mating in 2 cichlid sister species. *Behav. Ecol.* **25**, 612–626. (doi:10.1093/beheco/aru024)
25. Maan ME, Seehausen O, Söderberg L, Johnson L, Ripmeester EAP, Mrosso HDJ, Taylor MI, Van Dooren TJM, Van Alphen JJM. 2004 Intraspecific sexual selection on a speciation trait, male coloration, in the Lake Victoria cichlid *Pundamilia nyererei*. *Proc. R. Soc. B* **271**, 2445–2452. (doi:10.1098/rspb.2004.2911)
26. Selz OM, Thommen R, Pierotti MER, Anaya-Rojas JM, Seehausen O. 2016 Differences in male coloration are predicted by divergent sexual selection between populations of a cichlid fish. *Proc. R. Soc. B* **283**, 20160172. (doi:10.1098/rspb.2016.0172)
27. Haesler MP, Seehausen O. 2005 Inheritance of female mating preference in a sympatric sibling species pair of Lake Victoria cichlids: implications for speciation. *Proc. R. Soc. B* **272**, 237–245. (doi:10.1098/rspb.2004.2946)
28. Svensson O, Woodhouse K, van Oosterhout C, Smith A, Turner GF, Seehausen O. 2017 The genetics of mate preferences in hybrids between two young and sympatric Lake Victoria cichlid species. *Proc. R. Soc. B* **284**, 20162332. (doi:10.1098/rspb.2016.2332)
29. Stelkens RB, Pierotti ME., Joyce DA, Smith AM, van der Sluijs I, Seehausen O. 2008 Disruptive sexual selection on male nuptial coloration in an experimental hybrid population of cichlid fish. *Phil. Trans. R. Soc. B* **363**, 2861–2870. (doi:10.1098/rstb.2008.0049)
30. Felsenstein J. 1981 Skepticism towards Santa Rosalia, or why are there so few kinds of animals? *Evolution* **35**, 124. (doi:10.2307/2407946)
31. Slatkin M. 1987 Gene flow and the geographic structure of natural populations. *Science* **236**, 787–792. (doi:10.1126/science.3576198)
32. Gavrilets S. 2004 *Fitness landscapes and the origin of species*. Princeton, NJ: Princeton University Press.
33. Yeaman S, Otto SP. 2011 Establishment and maintenance of adaptive genetic divergence under migration, selection, and drift. *Evolution* **65**, 2123–2129. (doi:10.1111/j.1558-5646.2011.01277.x)
34. Yeaman S, Whitlock MC. 2011 The genetic architecture of adaptation under migration–selection balance. *Evolution* **65**, 1897–1911. (doi:10.1111/j.1558-5646.2011.01269.x)
35. Griswold CK. 2006 Gene flow's effect on the genetic architecture of a local adaptation and its consequences for QTL analyses. *Heredity* **96**, 445–453. (doi:10.1038/sj.hdy.6800822)
36. Bürger R, Akerman A. 2011 The effects of linkage and gene flow on local adaptation: a two-locus continent–island model. *Theor. Popul. Biol.* **80**, 272–288. (doi:10.1016/j.tpb.2011.07.002)
37. Flaxman SM, Wacholder AC, Feder JL, Nosil P. 2014 Theoretical models of the influence of genomic architecture on the dynamics of speciation. *Mol. Ecol.* **23**, 4074–4088. (doi:10.1111/mec.12750)
38. van Doorn GS, Dieckmann U. 2006 The long-term evolution of multilocus traits under frequency-dependent disruptive selection. *Evolution* **60**, 2226. (doi:10.1554/06-291.1)
39. Turner GF, Burrows MT. 1995 A model of sympatric speciation by sexual selection. *Proc. R. Soc. B* **260**, 287–292. (doi:10.1098/rspb.1995.0093)
40. Higashi M, Takimoto G, Yamamura N. 1999 Sympatric speciation by sexual selection. *Nature* **402**, 523–526. (doi:10.1038/990087)
41. Arnegard ME, Kondrashov AS. 2004 Sympatric spaciation by sexual selection alone is unlikely. *Evolution* **58**, 222–237. (doi:10.1111/j.0014-3820.2004.tb01640.x)
42. Baerends GP, Baerends-Van Roon JT. 1950 *An introduction to the study of the ethology of the cichlid fishes*. Leiden, The Netherlands: E. J. Brill.
43. Sambrook J, Russell DW. 2001 *Molecular cloning: a laboratory manual*, 3rd edn. New York, NY: Cold Spring Harb. Lab. Press.
44. Baird NA, Etter PD, Atwood TS, Currey MC, Shiver AL, Lewis ZA, Selker EU, Cresko WA, Johnson EA. 2008 Rapid SNP discovery and genetic mapping using sequenced RAD markers. *PLoS ONE* **3**, e3376. (doi:10.1371/journal.pone.0003376)
45. Feulner PGD, Schwarzer J, Haesler MP, Meier JI, Seehausen O. 2018 A dense linkage map of Lake Victoria cichlids improved the *Pundamilia* genome assembly and revealed a major QTL for sex-determination. *G3: Genes Genomes Genetics* **8**, 2411–2420. (doi:10.1534/g3.118.200207)
46. Langmead B, Salzberg SL. 2012 Fast gapped-read alignment with Bowtie 2. *Nat. Methods* **9**, 357–359. (doi:10.1038/nmeth.1923)
47. McKenna A *et al.* 2010 The genome analysis toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res.* **20**, 1297–1303. (doi:10.1101/gr.107524.110)
48. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R. 2009 The sequence alignment/map format and SAMtools. *Bioinformatics* **25**, 2078–2079. (doi:10.1093/bioinformatics/btp352)
49. Danecek P *et al.* 2011 The variant call format and VCFtools. *Bioinformatics* **27**, 2156–2158. (doi:10.1093/bioinformatics/btr330)
50. van Ooijen JW. 2006 *Joinmap®4, software for the calculation of genetic linkage maps in experimental populations*. Wageningen, The Netherlands: Kyazma BV.
51. Broman KW, Wu H, Sen S, Churchill GA. 2003 R/qtl: QTL mapping in experimental crosses. *Bioinformatics* **19**, 889–890. (doi:10.1093/bioinformatics/btg112)
52. R Core Development team. 2015 *R: a language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing. See http//www.R-project.org/.

53. Brawand D *et al.* 2014 The genomic substrate for adaptive radiation in African cichlid fish. *Nature* **513**, 375–381. (doi:10.1038/nature13726)
54. Guyon R *et al.* 2012 A high-resolution map of the Nile tilapia genome: a resource for studying cichlids and other percomorphs. *BMC Genomics* **13**, 222. (doi:10.1186/1471-2164-13-222)
55. Beavis WD. 1995 The power and deceit of QTL experiments: lessons from comparative QTL studies. In *Proc. of the 49th Ann. Corn and Sorghum Industry Research Conf., ASTA, Washington, DC*, pp. 252–268. Alexandria, VA: ASTA.
56. Magalhaes IS, Seehausen O. 2010 Genetics of male nuptial colour divergence between sympatric sister species of a Lake Victoria cichlid fish. *J. Evol. Biol.* **23**, 914–924. (doi:10.1111/j.1420-9101.2010.01960.x)
57. van der Sluijs I, Seehausen O, van Dooren TJM, van Alphen JJM. 2010 No evidence for a genetic association between female mating preference and male secondary sexual trait in a Lake Victoria cichlid fish. *Curr. Biol.* **56**, 57–64. (doi:10.1093/czoolo/56.1.57)
58. van der Sluijs I, Dijkstra PD, Lindeyer CM, Visser B, Smith AM, Groothuis TGG, van Alphen JJM, Seehausen O. 2013 A test of genetic association among male nuptial coloration, female mating preference, and male aggression bias within a polymorphic population of cichlid fish. *Curr. Zool.* **59**, 221–229. (doi:10.1093/czoolo/59.2.221)
59. Fruciano C, Franchini P, Kovacova V, Elmer KR, Henning F, Meyer A. 2016 Genetic linkage of distinct adaptive traits in sympatrically speciating crater lake cichlid fish. *Nat. Commun.* **7**, 1–8. (doi:10.1038/ncomms12736)
60. Merrill RM, Rastas P, Martin SH, Melo MC, Barker S, Davey J, McMillan WO, Jiggins CD. 2019 Genetic dissection of assortative mating behavior. *PLoS Biol.* **17**, 1–21. (doi:10.1371/journal.pbio.2005902)
61. Ferris KG, Barnett LL, Blackman BK, Willis JH. 2017 The genetic architecture of local adaptation and reproductive isolation in sympatry within the *Mimulus guttatus* species complex. *Mol. Ecol.* **26**, 208–224. (doi:10.1111/mec.13763)
62. Noor MAF, Grams KL, Bertucci LA, Reiland J. 2001 Chromosomal inversions and the reproductive isolation of species. *Proc. Natl. Acad. Sci. USA* **98**, 12 084–12 088. (doi:10.1073/pnas.221274498)
63. Yeaman S. 2013 Genomic rearrangements and the evolution of clusters of locally adaptive loci. *Proc. Natl. Acad. Sci. USA* **110**, E1743–E1751. (doi:10.1073/pnas.1219381110)
64. Seehausen O, Wagner CE. 2014 Speciation in freshwater fishes. *Annu. Rev. Ecol. Evol. Syst.* **45**, 621–651. (doi:10.1146/annurev-ecolsys-120213-091818)
65. Feller AF, Haesler MP, Peichel CL, Seehausen O. 2020 Data from: Genetic architecture of a key reproductive isolation trait differs between sympatric and non-sympatric sister species of Lake Victoria cichlids. Dryad Digital Repository. (https://dx.doi.org/10.5061/dryad.5tb2rbp1n).

Electronic supplementary material

Genetic architecture of a key reproductive isolation trait differs between sympatric and non-sympatric sister species of Lake Victoria cichlids

DOI: 10.1098/rspb.2020.0270

Appendix S1

We normalized DNA concentration for all samples in one library prior to restriction enzyme digestion. 4-48 individuals carrying TruSeq P1-adaptors and different custom 5bp-8bp barcodes were pooled into one library (see Table S5). For shearing we used a Covaris M220 Focused-ultrasonicator. Sheared fragments between 300-700bp were selected on a SageELF machine. After ligation of the TruSeq P2-adaptors each library was amplified in four aliquots of 50 µl.

Appendix S2

In a first step, PhiX reads were removed using Bowtie2 v2.3.2 [1]. Reads were then demultiplexed and trimmed to 85 bp with process_radtags implemented in stacks v.1.40 [2]. The FASTQ quality filter (http://hannonlab.cshl.edu/ fastx_toolkit/index.html) was used to filter reads for a minimum quality of 10 at all bases and of 30 in at least 95% of the reads.

After alignment to the *Pundamilia nyererei* reference genome [3] with Bowtie2 v2.3.2 [1], base quality recalibration was performed using the BaseRecalibrator and PrintReads modules of the Genome Analysis Toolkit v3.7 [4].

Genotyping (only for uniquely aligned reads) was done with (GATK Unified Genotyper v3.7. [4]). In the red-dorsum x blue cross, this included the 231 focal individuals (225 F2s belonging to two families, of which 44 females, the four F1 parents and the two F0 grandparents), plus 8 individuals (seven F2s and one F1 parent) belonging to a smaller family with the same grandparents and one shared F1 parent, as well as 27 individuals (22 F2s and five F1 parents) with different grandparents of the same two species. The 35 additionally genotyped individuals with different parents/grandparents were not used in the following steps or any analyses. In the red-chest x blue cross, genotyping included a total of 211 individuals (205 F2 individuals belonging to two families, of which 68 females, the four F1 parents, and the two F0 parents).

The vcf file filtering procedure using Bcftools implemented in Samtools v.1.8 [5] and Vcftools v.0.1.14 [6] was the following: after filtering out sites with >50% missing data we removed individuals with a mean depth of <12 and / or >50% missing data. We removed sites within 10 bp of indels and kept only bi-allelic SNPs with a mean genotype depth of less than 1.5 times the interquartile range from the mean (as sites with a greater depth are expected to be enriched for duplicates), and set genotypes with a depth of <10 to missing. Individuals were then checked for PCR duplicates and removed if the heterozygous read balance was heavily skewed. Sites were then again filtered for missing data of max. 50% and for a minor allele frequency of min. 0.05.

For the red-dorsum x blue cross, this resulted in 10,598 SNPs over the remaining 224 individuals (218 F2s, of which 43 females, four F1 parents, and two F0 grandparents). As only 566 of these SNPs appeared to have alternative homozygous genotypes between the two F0 grandparents, we additionally applied a custom allelic balance filter on these two individuals, where heterozygous genotypes failing a binomial test on the genotype distributions were set to homozygous. After this correction, 1,326 SNPs appeared as homozygous fixed in the F0, which is closer to our expectations based on other recent QTL mapping studies on cichlids using RAD sequencing data e.g. [3,7]. 368 of these SNPs were heterozygous as expected in all four F1 parents and were used in linkage map construction.

For the red-chest x blue cross, this resulted in 9,990 SNPs over the remaining 190 individuals (186 F2s, of which 57 females, two F1 parents of one family (the two of the other family were removed in the filtering procedures due to low quality) and two F0 grandparents). 3,473 SNPs were retained as alternative homozygous between the two F0 grandparents after applying the same allelic balance filter as for the red-dorsum x blue cross. 2,358 of these were heterozygous as expected in both F1s and used in linkage map construction.

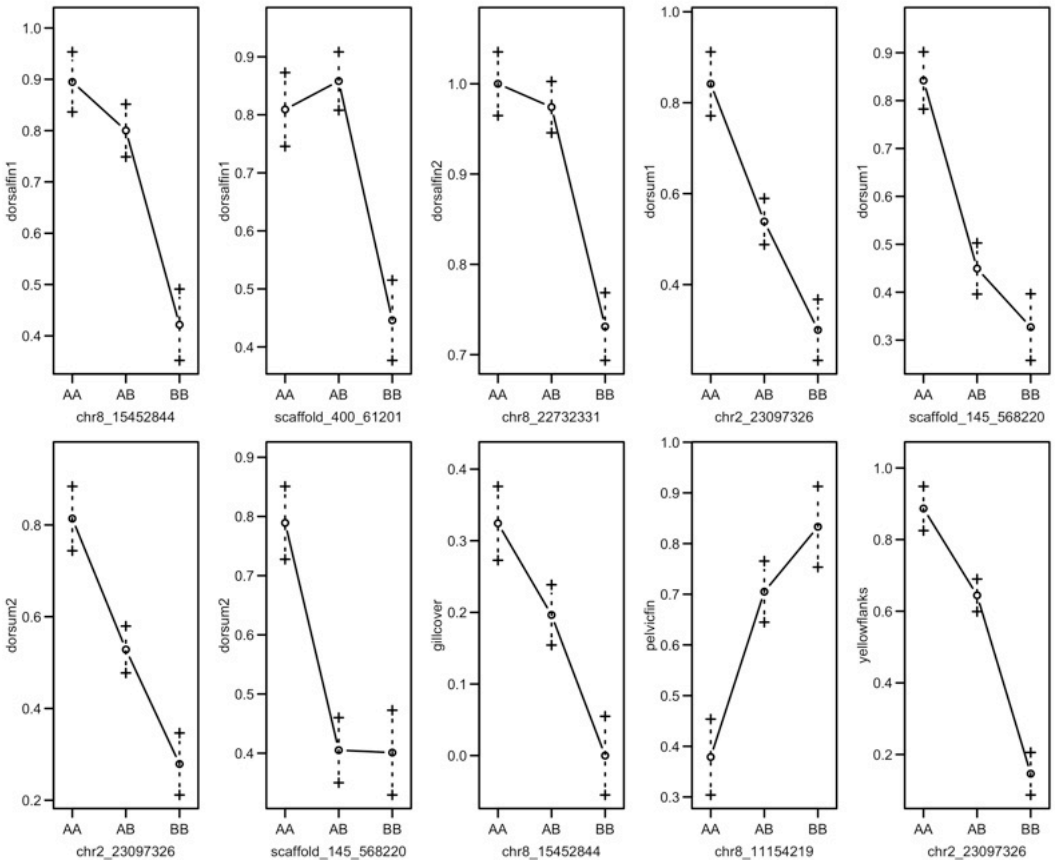


Figure S1. Effect plots at the markers on the QTL peaks for all significant QTLs in the red-dorsum x blue cross. The genotype AA corresponds to homozygous as *P.* sp. "pundamilia-like", BB to homozygous as *P.* sp. "nyererei-like".

Table S1. Locations and effects of QTLs for the presence/absence of red colour on different body parts, for yellow flanks, and for a phenotypic hybrid index.

	trait	# F2s	marker (nearest marker)	Pun-LG	Ore-LG	cM	95% CI (cM)	nearest markers	LOD	P -value	PVE
red-dorsum x blue cross	dorsal fin 1	140	c8.loc38 (chr8_15452844)	8	19	38.0	35-48.3	chr8_14022265 - scaffold_261_382784	5.64	<0.001	16.77
	dorsal fin 1	140	scaffold_400_61201	10	23	12.7	2-25	scaffold_844_38876 - chr10_30070154	5.16	0.001	15.51
	dorsal fin 2	172	c8.loc44 (chr8_22732331)	8	19	44.0	35-48.3	chr8_14022265 - scaffold_261_382784	6.44	<0.001	15.29
	dorsum 1	171	chr2_23097326	2	7	13.2	0-17	chr2_14394351 - chr2_34782227	5.93	0.001	14.82
	dorsum 1	171	scaffold_145_568220	6	11	5.4	0-8	chr6_2497279 - chr6_4731967	7.07	<0.001	17.35
	dorsum 2	171	chr2_23097326	2	7	13.2	0-16	chr2_14394351 - chr2_23097326	5.73	0.002	14.36
	dorsum 2	171	scaffold_145_568220	6	11	5.4	0-8	chr6_249727 - chr6_4731967	4.92	0.005	12.43
	gill cover	174	c8.loc38 (chr8_15452844)	8	19	38.0	15-48.3	chr8_8525018 - scaffold_261_382784	5.08	0.002	12.57
	pelvic fin	126	c8.loc30 (chr8_11154219)	8	19	30.0	24-48.3	chr8_8524751 - scaffold_261_382784	4.22	0.015	14.27
	pelvic fin	126	c12.loc52 (chr12_33255157)	12	17	52	30-58	chr12_25602998 - chr12_33255157	3.18	0.088	10.75
	yellow flanks	174	c2.loc12 (chr2_23097326)	2	7	12.0	8-16	chr2_17806364 - chr2_23097326	13.27	<0.001	29.49
	hybrid index	119	c2.loc12 (chr2_23097326)	2	7	12	0-15	chr2_14394351 - chr2_23097326	7.14	<0.001	17.20
	hybrid index	119	chr6_2497279	6	11	0	0-8	chr6_2497279 - chr6_4731967	3.35	0.084	8.50
	hybrid index	119	scaffold_261_382784	8	19	48.3	33-48.3	chr8_13110496 -scaffold_261_382784	4.41	0.01	10.97
	hybrid index	119	c18.loc9 (chr18_9611117)	18	6	9	3-14	chr18_2214232- chr18_9611117	5.43	0.003	13.28
	pelvic fin (add/int)	126	chr8_19963137	8	19	40.4			5.06	0.022	13.76
	pelvic fin (add/int)	126	chr22_6708694	22	12	24.7			4.33	0.079	6.71
red-chest x blue cross											
	throat (add)	125	chr7_9610352	7	15	9.9			3.77	0.064	8.31
	head *	123	c9.loc21 (chr9_11198659)	9	18	21			3.50	0.057	9.85

F2s, number of F2 males with data; (nearest marker), nearest RAD marker to a QTL at an interpolated marker (where genotypes were inferred by maximum likelihood estimation under a mixture model, using hidden Markov models); Pun-LG, linkage group number corresponding to the anchored *Pundamilia nyererei* reference genome [3]; Ore-LG, linkage group number corresponding to the *Oreochromis niloticus* reference genome [8]; cM, position in centiMorgan; 95% CI , 95% approximate Bayesian credible interval (in cM); PVE, percent variance explained; add, accounting for family as additive covariate; int, accounting for family as additive and as interactive covariate; * found in 2/5 mappings with subsampled map (see also Table S4)

Table S2. Results from two-dimensional two-QTL scans

	trait	chr 1	chr 2	pos 1f	pos 2f	lod full	pval full	lod fv1	pval fv1	lod int	pval int	pos 1a	pos 2a	lod add	pval add	lod av1	pval av1
red-dorsum x blue cross	dorsal fin 1	8	10	48	12	10.86	0.000	5.32	0.576	1.80	1.000	48	12	9.06	0.000	3.52	0.017
	dorsal fin 1	8	12	39	33	11.85	0.000	6.30	0.079	3.87	0.982	39	36	7.98	0.000	2.43	0.347
	dorsal fin 2	1	8	21	42	9.36	0.000	3.15	1.000	-0.04	1.000	18	42	9.40	0.000	3.19	0.036
	upper dorsum 1	2	6	12	6	14.93	0.000	7.90	0.007	0.81	1.000	0	6	14.12	0.000	7.09	0.000
	upper dorsum 1	2	11	12	3	10.29	0.004	4.45	0.994	1.31	1.000	12	3	8.97	0.000	3.14	0.045
	upper dorsum 1	2	18	12	6	9.13	0.028	3.30	1.000	0.20	1.000	12	6	8.93	0.000	3.10	0.050
	upper dorsum 1	6	18	6	9	12.07	0.000	5.04	0.847	0.81	1.000	6	9	11.26	0.000	4.23	0.003
	upper dorsum 2	2	6	12	0	12.48	0.000	6.86	0.066	1.36	1.000	0	0	11.12	0.000	5.50	0.000
	upper dorsum 2	6	18	6	9	8.81	0.041	3.90	1.000	0.94	1.000	6	9	7.87	0.003	2.96	0.083
	gill cover	6	8	0	42	9.15	0.007	4.09	1.000	0.29	1.000	0	42	8.85	0.000	3.80	0.008
	pelvic fin	4	8	24	21	10.81	0.000	6.59	0.086	4.87	0.556	39	39	5.94	0.055	1.71	0.945
	pelvic fin	8	12	30	36	10.32	0.001	6.10	0.216	1.85	1.000	30	36	8.47	0.001	4.25	0.000
	yellow flanks	2	18	12	15	17.93	0.000	4.66	0.976	0.70	1.000	12	15	17.24	0.000	3.96	0.003
red-chest x blue cross																	
	cheek	1	4	15	24	8.78	0.092	7.77	0.016	7.23	0.011	9	27	1.55	1	0.53	1
	throat	1	4	15	30	9.11	0.067	8.03	0.012	7.70	0.006	30	54	1.41	1	0.33	1
	yellow flanks	3	11	54	51	9.37	0.033	7.01	0.088	6.04	0.127	54	60	3.32	0.992	0.97	1

lod.full, comparing the full model (two QTLs with interaction) to the null model (no QTLs); lod.fv1, comparing the full model to the best single QTL model for a respective trait; lod.int, comparing the full model to the additive (two QTLs with no interaction) model i.e. indicating evidence for epistasis; lod.add, comparing the additive model to the null model; lod.av1 comparing the additive model to the best single QTL model.

Table S3. QTL mapping in the red-dorsum x blue cross repeated five times with randomly sampled 125 F2 males. Only traits with scores below a genome-wide significance threshold of 0.1 are shown.

	chr	pos	dorsalfin1	pval	dorsalfin2	pval	dorsum1	pval	dorsum2	pval	gillcover	pval	pelvicfin	pval	yellowflanks	pval
1																
c2.loc1	2	1.0	1.15	1.000	0.92	1.000	4.82	0.007	4.28	0.012	0.63	1.000	0.10	1.000	9.54	0.000
chr2_23097326	2	13.2	0.05	1.000	0.01	1.000	4.20	0.018	4.13	0.018	0.94	1.000	0.16	1.000	11.38	0.000
scaffold_145_568220	6	5.4	0.35	1.000	0.42	1.000	4.43	0.012	3.62	0.042	2.83	0.230	0.77	1.000	0.35	1.000
chr6_4731967	6	8.5	0.07	1.000	0.08	1.000	4.29	0.014	3.88	0.029	2.24	0.604	0.65	1.000	0.81	1.000
chr8_8525018	8	21.0	1.83	0.852	1.14	0.999	1.56	0.953	1.63	0.932	4.69	0.005	3.78	0.039	0.14	1.000
chr8_11154219	8	29.4	2.17	0.656	2.32	0.493	0.92	1.000	0.81	1.000	3.22	0.117	5.16	0.001	0.22	1.000
chr8_19857096	8	42.6	4.04	0.018	3.47	0.022	1.11	1.000	1.15	1.000	4.04	0.023	4.09	0.021	1.11	1.000
chr16_8832104	16	0.0	0.34	1.000	0.62	1.000	2.89	0.193	2.68	0.266	3.33	0.099	1.73	0.887	1.25	0.994
chr18_579662	18	0.0	1.17	1.000	1.23	0.999	3.28	0.091	1.66	0.913	1.07	1.000	1.44	0.977	2.09	0.629
2																
c2.loc12	2	12.0	0.04	1.000	0.22	1.000	4.22	0.015	4.63	0.002	0.96	1.000	1.16	1.000	9.73	0.000
chr2_23097326	2	13.2	0.05	1.000	0.22	1.000	4.31	0.012	4.73	0.000	0.98	1.000	1.15	1.000	9.63	0.000
scaffold_145_568220	6	5.4	0.03	1.000	0.41	1.000	6.42	0.000	4.76	0.000	2.70	0.303	0.38	1.000	0.51	1.000
c8.loc42	8	42.0	2.57	0.364	2.96	0.074	0.28	1.000	0.24	1.000	4.40	0.011	2.36	0.490	1.70	0.896
chr8_19857096	8	42.6	2.67	0.301	3.10	0.053	0.29	1.000	0.25	1.000	4.39	0.011	2.35	0.490	1.76	0.865
3																
c2.loc12	2	12.0	0.04	1.000	0.22	1.000	4.22	0.015	4.63	0.002	0.96	1.000	1.16	1.000	9.73	0.000
chr2_23097326	2	13.2	0.05	1.000	0.22	1.000	4.31	0.012	4.73	0.000	0.98	1.000	1.15	1.000	9.63	0.000

scaffold_145_568220	6	5.4	0.03	1.000	0.41	1.000	6.42	0.000	4.76	0.000	2.70	0.303	0.38	1.000	0.51	1.000
c8.loc42	8	42.0	2.57	0.364	2.96	0.074	0.28	1.000	0.24	1.000	4.40	0.011	2.36	0.490	1.70	0.896
chr8_19857096	8	42.6	2.67	0.301	3.10	0.053	0.29	1.000	0.25	1.000	4.39	0.011	2.35	0.490	1.76	0.865
4																
c2.loc11	2	11.0	0.36	1.000	0.55	1.000	5.03	0.002	4.41	0.011	1.30	0.999	0.02	1.000	9.85	0.000
c2.loc13	2	13.0	0.33	1.000	0.37	1.000	5.06	0.001	4.46	0.011	1.21	1.000	0.01	1.000	9.61	0.000
scaffold_145_568220	6	5.4	0.23	1.000	0.28	1.000	5.77	0.000	3.40	0.076	1.41	0.991	0.01	1.000	0.14	1.000
chr8_11154219	8	29.4	1.73	0.908	2.67	0.257	0.39	1.000	0.31	1.000	2.03	0.742	4.23	0.015	0.28	1.000
chr8_22732331	8	45.1	3.10	0.138	4.04	0.003	0.12	1.000	0.33	1.000	3.42	0.060	3.55	0.047	1.83	0.843
scaffold_261_382784	8	48.3	3.11	0.135	3.97	0.005	0.26	1.000	0.56	1.000	3.55	0.045	3.75	0.033	1.31	0.993
c10.loc13	10	13.0	5.31	0.000	2.91	0.135	0.21	1.000	0.13	1.000	0.13	1.000	1.04	1.000	0.12	1.000
c12.loc53	12	53.0	0.39	1.000	0.07	1.000	0.32	1.000	0.50	1.000	0.47	1.000	*4.99*	*0.000*	1.54	0.960
5																
c2.loc10	2	10.0	0.76	1.000	0.33	1.000	5.62	0.002	6.10	0.002	0.92	1.000	0.03	1.000	9.21	0.000
chr2_23097326	2	13.2	0.76	1.000	0.45	1.000	5.94	0.002	6.44	0.001	1.02	1.000	0.02	1.000	8.91	0.000
scaffold_145_568220	6	5.4	0.09	1.000	0.03	1.000	5.41	0.002	3.52	0.053	1.67	0.928	0.12	1.000	0.05	1.000
c6.loc7	6	7.0	0.06	1.000	0.10	1.000	5.45	0.002	3.51	0.056	1.54	0.963	0.08	1.000	0.16	1.000
c8.loc38	8	38.0	3.95	0.023	3.96	0.006	0.10	1.000	0.01	1.000	3.16	0.151	2.80	0.234	0.17	1.000
c8.loc43	8	43.0	3.82	0.033	4.28	0.005	0.13	1.000	0.01	1.000	2.97	0.205	2.96	0.170	1.07	1.000
chr8_22732331	8	45.1	3.87	0.028	3.76	0.016	0.10	1.000	0.00	1.000	3.18	0.148	3.32	0.089	0.89	1.000
scaffold_261_382784	8	48.3	3.63	0.047	3.45	0.033	0.11	1.000	0.03	1.000	3.46	0.080	2.90	0.193	0.66	1.000

Colour background indicates (marginally) significant LOD scores/p-values. Green: QTLs known from scanone results (darkgreen are the known significant QTLs that were recovered as significant QTLs (i.e. pos within the 95% CI of the known significant QTL)); Blue: QTLs known from scantwo results; Red: not seen before.

Table S4. QTL mapping in the red-chest x blue cross repeated five times with a randomly subsampled linkage map to match the number markers of the red-dorsum x blue cross. Only traits with LOD scores below a genome-wide significance threshold of 0.1 are shown.

	locus	chr	pos	head	pval
1					
	There were no LOD peaks above the threshold.				
2					
	There were no LOD peaks above the threshold.				
3					
	c9.loc21 (chr9_11198659)	9	21	3.50	0.057
4					
	c9.loc20 (chr9_9921245)	9	20	3.23	0.089
5					
	There were no LOD peaks above the threshold.				

Table S5. Overview of RAD sequencing libraries with number of individuals.

library	# total	# F0	# F1	# F2	platform	read length
red-chest x blue cross, see also [3]						
GQI21	47	2	2	43	Bern	100
GQI22	30	2	-	28	Bern	100
GQI26	10	-	-	10	Bern	100
GQI28	40	-	-	40	Bern	100
GQI29	4	4	-	-	Bern	100
GQI66	48*	-	-	36	Bern	100
GQI67	47		-	47	Bern	100
GQI99	37	-	-	37	Lausanne	100
GQI100b	37	-	-	37	Lausanne	100
red-dorsum x blue cross						
GQI101	8	2	6 (2 red-chest x blue)	-	Lausanne	100
GQI102	48	-	-	48	Lausanne	100
GQI123	47	-	-	47	Lausanne	100
GQI124	47	-	3	44	Lausanne	100
GQI134	47	-	-	47	Lausanne	100
GQI140	47	-	2	45	Lausanne	100
GQI144	47	-	1	46	Lausanne	125
GQI146	27**	-	-	15	Lausanne	125
numbers include individuals that were resequenced if they had low quality or low read number in the first run; *contains 12 individuals of a different species; **contains 12 F0/F1 individuals of a different cross						

References for Appendix / Supplementary Materials

1. Langmead B, Salzberg SL. 2012 Fast gapped-read alignment with Bowtie 2. *Nat. Methods* **9**, 357–359. (doi:10.1038/nmeth.1923)
2. Catchen J, Hohenlohe PA, Bassham S, Amores A, Cresko WA. 2013 Stacks: an analysis tool set for population genomics. *Mol. Ecol.* **22**, 3124–3140. (doi:10.1111/mec.12354)
3. Feulner PGD, Schwarzer J, Haesler MP, Meier JI, Seehausen O. 2018 A Dense Linkage Map of Lake Victoria Cichlids Improved the *Pundamilia* Genome Assembly and Revealed a Major QTL for Sex-Determination. *G3; Genes|Genomes|Genetics* **8**, 2411–2420. (doi:10.1534/g3.118.200207)
4. McKenna A *et al.* 2010 The Genome Analysis Toolkit: A MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res.* **20**, 1297–1303. (doi:10.1101/gr.107524.110)
5. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R. 2009 The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25**, 2078–2079. (doi:10.1093/bioinformatics/btp352)
6. Danecek P *et al.* 2011 The variant call format and VCFtools. *Bioinformatics* **27**, 2156–2158. (doi:10.1093/bioinformatics/btr330)
7. Henning F, Machado-Schiaffino G, Baumgarten L, Meyer A. 2017 Genetic dissection of adaptive form and function in rapidly speciating cichlid fishes. *Evolution* **71**, 1297–1312. (doi:10.1111/evo.13206)
8. Brawand D *et al.* 2014 The genomic substrate for adaptive radiation in African cichlid fish. *Nature* **513**, 375–381. (doi:10.1038/nature13726)

Chapter 2 –

Genetic architecture of adaptive radiation across two trophic levels

Genetic architecture of adaptive radiation across two trophic levels

Anna Fiona Feller[1,2], Ole Seehausen[1,2]

1) Division of Aquatic Ecology and Evolution, Institute of Ecology and Evolution, University of Bern, Baltzerstrasse 6, 3012 Bern, Switzerland
2) Department of Fish Ecology and Evolution, Centre of Ecology, Evolution and Biogeochemistry (CEEB), Eawag Swiss Federal Institute of Aquatic Science and Technology, Seestrasse 79, 6047 Kastanienbaum, Switzerland

Published version: https://doi.org/10.1098/rspb.2022.0377

Abstract

Evolution of trophic diversity is a hallmark of adaptive radiation. Yet, transitions between carnivory and herbivory are rare in young adaptive radiations and absent in most. Haplochromine cichlid fish of the African Great Lakes are exceptional in this regard. The cichlid fish radiation in Lake Victoria was colonized by an insectivorous generalist and in less than 20,000 years several different clades of specialized herbivores have evolved, besides many other trophic guilds. Several transitions between carnivory and herbivory have thus occurred in very little time, followed by many speciation events within each trophic level. This is likely to be one explanation for the exceptional species richness that has evolved in these lakes. Carnivorous versus herbivorous lifestyles in cichlid fish require many different adaptations in functional morphology, physiology, and behavior. Transitions in either direction thus require many traits to change in a concerted fashion. This begs the question if this unusually fast trophic radiation might have been facilitated by genetic architectures involving tight linkage or pleiotropy among co-adapted trophic traits. To begin addressing this question, we performed a QTL experiment to investigate the genetic architecture of a suite of traits that distinguish specialized rocky shore algae scrapers from invertivores. We find many different genomic regions to underpin the functional morphological differences between these trophic types but little evidence for linkage or pleiotropy. This implies that genome-wide linkage disequilibrium (LD) between many regions in the genome, i.e. generalized reproductive isolation, had to build to permit the trophic transition. We speculate that this process, which is difficult in the absence of geographical isolation, might have been facilitated by the hybrid origin of the radiation.

Keywords: Lake Victoria, adaptive radiation, cichlids, trophic levels, genetic architecture

Introduction

The causes of the massive variation in the rates at which species diversity evolves and species rich communities assemble in different evolutionary lineages remains one of the most fundamental questions in evolutionary biology. The East African cichlid fish adaptive radiations provide excellent systems to investigate factors contributing to such variation. The three largest lakes in this region, Lakes Victoria, Malawi, and Tanganyika, each harbor several hundred species of cichlids that have evolved within the confines of these lakes (reviewed e.g. in Kocher, 2004; Seehausen, 2006; Henning & Meyer, 2014; Salzburger, 2018). In Lake Victoria alone, approximately 500 haplochromine cichlid species have evolved in less than 20,000 years (Seehausen, 1996; Stager & Johnson, 2008). Lake Victoria haplochromine cichlids can be grouped into at least 16 different trophic groups, all but one of which comprise several to many species (Greenwood, 1974; Witte & Van Oijen, 1990; Seehausen, 1996). This staggering trophic diversity is replicated, with some modifications, across all three major radiations and across multiple smaller radiations in smaller lakes (e.g. reviewed in Seehausen, 2015).

One hallmark of the East African cichlid radiations is the co-existence of many closely related species in full sympatry (Seehausen, 2015). Any one habitat patch in the lakes is shared by dozens of species (Fryer & Iles, 1972; Greenwood, 1974; Goldschmidt *et al.*, 1990; Seehausen, 1996; Bouton *et al.*, 1997, 1998; Turner *et al.*, 2001). While they have retained the ability to feed opportunistically when a profitable food source is abundant, many of these species feature specialized morphologies for one specific source of food (Liem, 1980; McKaye & Marsh, 1983; Bouton *et al.*, 1998, 1999; Golcher-Benavides & Wagner, 2019). These hyper-diverse sympatric communities even make up entire foodwebs that span three trophic levels, including herbivory and piscivory (Fryer & Iles, 1972; Greenwood, 1974; Witte & Van Oijen, 1990; Seehausen, 1996). The radiation across three trophic levels starting from insectivore ancestors (Greenwood, 1974; Joyce *et al.*, 2005; Selz *et al.*, 2014a) was thus key to the evolution of large species richness in these radiations (Seehausen, 2015). Herbivory has evolved from insectivory independently in each of these radiations. Moreover, transitions between insectivory and herbivory may have occurred several times even within the youngest of the radiations (McGee *et al.*, 2020). In other fish radiations, transitions between invertivores and herbivores are unusual. For instance, despite a large literature on adaptive radiation in postglacial fishes, not a single case of evolution of herbivory has been reported from lineages as diverse as stickleback (Bell & Foster, 1994), charr (Jonsson & Jonsson, 2001; Klemetsen, 2010), whitefish (Lu & Bernatchez, 1999; Bernatchez, 2004; Hudson *et al.*, 2007) or sculpins (Sideleva, 2003). Transitions to herbivory are also absent in some old freshwater radiations of other perch-like fish, such as sunfish (Pope *et al.*, 2010; Nakamura & Sato, 2014) or darters (Carlson & Wainwright, 2010).

The evolution of herbivory is somewhat paradoxical since it is considered a nutritionally inefficient feeding strategy compared to carnivory and omnivory, but has evolved many times in both terrestrial and aquatic lineages – albeit generally over long evolutionary time scales (reviewed in Sanchez & Trexler, 2016). Furthermore, while it is a rather rare strategy in teleost fishes (it occurs in <5% of species, and these are within a restricted set of families (Choat & Clements, 1998; Horn, 1998)), clades that include herbivorous-detrivorous fishes are among to most successful and diverse (Nelson, 2006). Several hypotheses have been formulated for the adaptive advantages to herbivory in freshwater fishes, ranging from maximizing intake efficiency by feeding on a non-evasive prey with nearly no search costs to disease avoidance (summarized in Sanchez & Trexler, 2016). The evolution of herbivory requires a complex set of physiological

and morphological innovations, such as a long intestine that breaks down and effectively processes the more resilient and low-nutrient plant tissue (Horn, 1989; Wagner *et al.*, 2009), tooth morphologies and arrangements to efficiently harvest plant matter from the substrate, e.g. in cichlids movably implanted densely spaced bicuspid teeth and many inner tooth rows providing a continuous surface to efficiently scrape algae off rocks (Greenwood, 1981; Witte & Van Oijen, 1990; Streelman & Albertson, 2006), powerful biting abilities (Bouton *et al.*, 1999), and distinct outer morphological features, such as a head profile that allows to get close enough to the food source (Greenwood, 1981). In order to understand why haplochromine cichlids repeatedly and rapidly made evolutionary transitions between invertivory and herbivory, while the same are rare or absent in other fish radiations, we sought to understand the genetic architecture of key traits that need to change in a concerted fashion during these transitions.

Concerted (and repeated) change among several co-adapted traits could be facilitated if these traits are also correlated genetically, i.e. coded by a region containing several physically tightly linked loci or by pleiotropic loci (Kirkpatrick & Barton, 2006; Schwander *et al.*, 2014). For instance, supergenes (tight clusters of two or more loci each responsible for a different trait) can regulate adaptive phenotypes by linking sets of co-adapted alleles or alleles with beneficial epistatic interactions (reviewed in Schwander *et al.*, 2014). Once in place, such supergenes can also spread between populations and species via introgressive hybridization. Examples include e.g. floral pollination strategies in *Primula* (Li *et al.*, 2016) and mimetic color patterns in *Heliconius* (Jay *et al.*, 2018). However, mutations occurring in such a genomic region that affect all correlated traits, may rarely be beneficial, implying that supergenes may be constrained in their further evolution (Fisher, 1930; Muller, 1950). It is for these reasons that some have suggested that an intermediate level of integration and a modular organization may constitute the most evolvable phenotypes and genetic architectures (Hansen, 2003). This is broadly consistent with patterns observed in previous Quantitative Trait Locus (QTL) mapping studies of functional morphology in cichlids. While some studies have found functionally and genetically highly integrated structures within confined anatomical regions such as the lower jaw (e.g. Albertson *et al.*, 2005; Cooper *et al.*, 2011), other studies mapping several functionally integrated, but anatomically separate aspects of foraging morphology (e.g. Henning *et al.*, 2017; Navon *et al.*, 2017), have found minimal overlap of QTLs between the different aspects, suggesting modularity in the genetic architecture underlying these traits. Moreover, distinct genetic architectures have been found for the same traits in different crosses from the same lake (e.g. lower jaw lever in two different Lake Malawi crosses (Albertson *et al.*, 2003, 2005; Parnell *et al.*, 2012)), implying that there are many alleles in these systems that can underlie the same functional traits. This points to a role of genotypic redundancy – when more than one genotype can produce the same phenotype or function (see e.g. review by Láruson *et al.*, 2020). The presence of a high level of segregating redundancy (see Láruson *et al.*, 2020) in cichlids may not be surprising since all major haplochromine radiations have evolved from hybrid populations between distantly related species (Meier *et al.*, 2017a; Irisarri *et al.*, 2018; Svardal et al., 2020), and repeated bouts of admixture have also been demonstrated to have occurred within each radiation (Genner & Turner, 2012; Meier *et al.*, 2017a, 2018). Recombination and sorting of old admixture-derived variation following the colonization of the newly formed Lake Victoria with many new ecological opportunities has provided high levels of segregating functional variation and has facilitated the rapid adaptive radiation within this lake (Meier *et al.*, 2017a).

Here we use an interspecific cross between a representative of one species-rich trophic group of morphologically specialized herbivores (epilithic algae scrapers), and a representative of a closely related lineage of rock-dwelling invertivores, to characterize the genetic architecture

associated with the transition between these trophic groups. We performed QTL mapping on a total of 28 morphological traits including intestine length, several aspects of tooth morphology and dental traits, and shape PC scores. We additionally mapped 12 male nuptial color traits (male nuptial color is a key for pre-zygotic isolation in this system (Seehausen & van Alphen, 1998; Selz *et al.*, 2014b)). We found significant QTLs for nine of the morphological traits and for three of the color traits. There was however little evidence for linkage or pleiotropy among loci for the different traits, and we found many different genomic regions to contribute to the functional morphological differences between the invertivore and the herbivore groups.

Materials and Methods

Experimental cross

Neochromis omnicaeruleus is a representative of a lineage of morphologically specialized epilithic algae scrapers. This trophic group, and the entire genus *Neochromis*, is confined to life on rocky reefs and islands, where they feed by dislodging firmly attached filamentous algae from rocks (Bouton *et al.*, 1997; Bouton, 1999). Currently, approximately 15 species are known in the genus *Neochromis* (Seehausen, 1996), up to five of which can be found in sympatry on a single reef, displaying variation on the common theme of algae scraping. The group as a whole is characterized by a steep dorsal head profile, a short head, a broad lower jaw, closely spaced upright sub-equally to equally bicuspid outer row oral teeth that are movably implanted, many (3-8) rows of inner teeth in the oral jaws barely separated from the outer tooth row by any gap, and a long and coiled intestine (Greenwood, 1981; Seehausen *et al.*, 1998a). Epilithic scrapers with similar trait combinations in the Lake Victoria radiation have additionally been described in the genera *Mbipia*, *Lithochromis* and *Paralabidochromis* (Seehausen, 1996). Similar combinations of traits also define epilithic algal scrapers in Lake Malawi (Fryer & Iles, 1972) and Tanganyika (Brichard, 1978). *N. omnicaeruleus* is widespread in the shallow rocky habitat of islands and lakes in the clear water areas of the open Lake Victoria (it is absent from rocky habitats in turbid water areas of the lake (Seehausen, 1996)).

Pundamilia sp. "nyererei-like" is a representative of a lineage of insectivores and zooplanktivores that are also restricted to rocky reefs and islands, but are characterized by a dorsal head profile that is rather shallow and straight or slightly concave, a rostrally upward inclined mouth gape, a long head, widely spaced unicuspid outer oral jaw teeth that are strongly recurved and firmly implanted, and mostly only two rows of inner teeth separated from the outer tooth row by a distinctive gap (Seehausen, 1996; Seehausen *et al.*, 1998a). *P.* sp. "nyererei-like" is restricted to the Mwanza Gulf in southern Lake Victoria where it likely evolved from a hybrid population of *P. pundamilia* and *P. nyererei* (Meier *et al.*, 2017b), two species that are both widespread and fully sympatric with *N. omnicaeruleus*.

These two species (see Figure 1a) were crossed in the lab to produce second-generation (F2) hybrids. An *N. omnicaeruleus* male from our lab population bred from fishes caught at Makobe island in Lake Victoria in 2010 was used as grandfather, and a *P.* sp. "nyererei-like" female from our lab population bred from fishes caught at Python island in Lake Victoria in 2003 was used as grandmother. Two first-generation (F1) pairs were then mated repeatedly to generate several clutches of F2 hybrids over a time of three years, resulting in a total number of ~210 F2 hybrids (150 males, 33 females, and ~30 juveniles of unknown sex that had died of

natural causes in the aquaria and were not used in any analyses). Clutch size in these fish ranges from 10 to 50 eggs and after spawning, females take on average 3-4 months to mature a new clutch. The rearing protocol was as described in (Feller *et al.*, 2020). All fish were bred and maintained in a large re-circulation aquarium system with a water temperature of 24-26°C and a 12:12h light/dark cycle. All fish were fed an ad libitum diet of flake food once a day, white mosquito larvae once a week and *Mysis* once a month. All F2 hybrids were reared to a minimum age of one year before further processing.

Color photos and specimen processing

To obtain standardized color photos we removed the most dominant males showing full color one by one from their holding tanks and placed them (up to six at a time) individually into one of six adjacent plexiglas cuvettes with grey backgrounds situated inside a separate aquarium. In this set-up the males could see each other, which induced territoriality and made them fully express color. After an acclimatization time of two hours, we took color photos using a Canon EOS 60D camera with a Tokina 35m macro lens (settings: shutter speed 1/100, aperture 2.8, ISO 200, auto white balance, serial mode). Two Walimex pro Daylight 250S lamps set up on either side of the aquarium provided the light source. Once we had obtained good live color photos of each male, we euthanized them (2-5 minutes in 50 mg MS222/l H_2O for sedation and then 20 minutes in 2 g MS222/l H_2O for euthanization) and then put them on melting ice for 10 minutes. This treatment resulted in the relaxation of the melanophores under nervous control (Baerends & Baerends-Van Roon, 1950) and made color better visible. We then pinned the fish down in a dissecting bowl and took a standardized color photo from the left side of the fish using the same camera and lens (settings: shutter speed 1/25, aperture 5.6, ISO 200, auto white balance, two-second delay) and the same two daylight lamps now placed on either side of the bowl. We used Adobe Photoshop CC 2018 to apply a white balance filter to these second photos by setting a white point on the white part of the color bar ruler. For the females we only took this second type of photo.

For each individual we then removed the right pectoral fin and stored it in 98% ethanol for later DNA extraction, and we cut open the ventral body wall to aid fixation of the guts. We fixed the specimens in 10% formalin (for a minimum of three weeks) and then transferred them to 70% ethanol for storage.

External morphology

Using digital calipers we measured 16 linear distances that describe the external morphology of head and body (following the descriptions in Barel et al., 1977) on the F2 hybrid specimens (Figure 1b). The distances were: standard length (SL), body depth (BD), head length (HL), head width (HW), lower jaw length (LJL), lower jaw width (LJW), snout length (SnL), snout width (SnW), eye length (EyL), eye depth (EyD), cheek depth (ChD), pre-orbital depth (POD), inter-orbital width (IOW), pre-orbital width (POW), caudal peduncle length (CPL), and caudal peduncle depth (CPD) (Figure 1b). We measured each trait twice, and if the deviation between the two measurements exceeded 5%, we added a third measurement. We then calculated the mean of the two (or the two closer) measurements and log-transformed them. All 15 measured traits were significantly correlated with SL ('cor.test' function in R (R Core Team, 2020)), and we thus used the residuals of linear regressions ('lm' function in R) of each trait against SL in subsequent analyses. With the exception of CPD we did not detect any significant interaction of sex with SL ('anova' function in R), and so we performed size correction for males and females together.

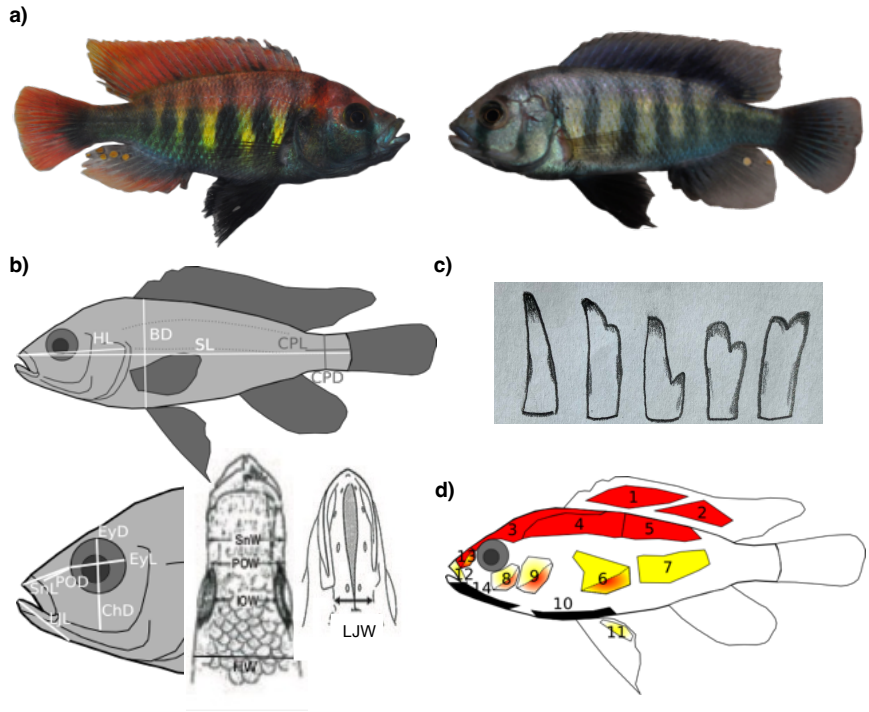


Figure 1 | Cross Overview. a) Male representatives of the two parental species used in the experimental cross. *Pundamilia* sp. "nyererei-like" from Python island (left) and *Neochromis omnicaeruleus* from Makobe island (right). The *N. omnicaeruleus* male is the individual used as grandfather in this cross. Photo credit: Oliver Selz. **b) The 16 measured linear distances,** following the descriptions in Barel *et al.*, 1977. The cartoon fish represents an average F2 hybrid individual based on geometric morphometric analyses (using the 'findMeanSpec' funtcion in R/geomorph). The dorsal and ventral view images are taken and modified from Barel *et al.*, 1977. **c) Tooth shapes.** *Pundamilia* sp. "nyererei-like" adult individuals feature rather widely set unicuspid teeth (on the very left) in the outer row, while *N. omnicaeruleus* have densely set sub-equally to equally bicuspid teeth (on the very right). Drawings adapted from Barel *et al.*, 1977. Intermediate forms were scored as weakly bicuspid (0.25), very unequally bicuspid (0.5), unequally bicuspid (0.75) (from left to right). **d) The presence / absence of red and yellow was scored in 12 different sectors of the body.** The sectors are the same as in (Feller *et al.*, 2020), in which one of the crosses was started with a female of the same parental species (*P.* sp. "nyererei-like") as the cross in the present study. Sectors are filled with red and/or yellow colour if at least a few F2 individuals had either colour in these sectors. Red in sectors 6, 8, 9, 13, and yellow in sectors 8, 11, 12 is not usually seen in either parental species (i.e. they constitute transgressive traits). Sectors 10 and 14 were not scored in this study (all F2 individuals were either black or darkblue).

Tooth morphology

We scored cusp shape of ten teeth in the outer row of each oral jaw (i.e. lower and upper oral jaw), five on each side starting the count from the mid-proximal point of the jaw. Each tooth received a score between 0 (unicuspid) and 1 (sub-equally to equally bicuspid), with three steps of 0.25 in between (weakly bicuspid (0.25), very unequally bicuspid (0.5), unequally bicuspid (0.75), see Figure 1c). We then summed up all scores in one jaw and divided this overall score by the number of scored teeth (up to 10, but excluding broken or missing teeth), resulting in an overall tooth shape score of between 0-1 for each oral jaw. Tricuspid teeth were scored separately (they are rare in both parental species and could thus be considered transgressive in the F2 hybrids). We also scored the density among these front row teeth in both jaws, assigning 0 to a large space between the outer row teeth (i.e. one extra tooth would fit in between two teeth), 1 to no space between the outer row teeth, and 0.5 to an intermediate space. In the lower jaw we additionally counted the number of inner tooth rows, and we scored the extent of the gap

between the outer and first inner tooth row (0: none, 0.5: small, 1: moderate; see (Seehausen *et al.*, 1998a)). In total, we scored six tooth traits.

Intestine length and confirmation of sex

To measure intestine length, we removed the whole alimentary canal from the F2 hybrid specimens, and then unrolled the intestine and laid it out on millimeter paper, taking care not to stretch the tissue. We measured the distance from the posterior end of the stomach to the anus to the nearest 0.1 mm. While dissecting out the alimentary canal, we also confirmed the sex of each individual by visual inspection of the gonads. Size correction for intestine length was performed as described for external morphology.

Trait correlations in F2 hybrid and parental species males

The same set of linear distances (with the exception of CPL and CPD) had also been previously studied on wild-caught *P.* sp. "nyererei-like" and *N. omnicaeruleus* males (van Rijssel *et al.*, 2018). We took the log-transformed raw data of 98 individuals of each species from this study. To compare trait correlations between the two parental species and the F2s we performed size-correction (as described above for the F2s) for both parental species separately, as slopes were non-homogeneous among the three groups in eight out of the 13 traits. To compare trait distributions between the two parental species and the F2s (to also screen for transgressive trait values), however, we performed size correction pooling all three groups to make them directly comparable. Three similar measures of dental trait variation had also been studied in the wild-caught specimens (van Rijssel *et al.*, 2018): the number of oral jaw tooth rows was counted and the gap between the outer and first inner row assigned to three categories (albeit both done in the upper jaw), and the percentage of bicuspid (vs unicuspid) teeth in all teeth of the outer row in the upper jaw was quantified (this is comparable to our tooth shape score). We also included these three aspects of dental and tooth morphology in the comparisons of the correlation matrices. We used the rcorr function in R/Hmisc v.4.4-0 (Harrell, 2020) to produce and the corrplot function in R/corrplot v.0.84 (Wei & Simko, 2017) to plot the correlation matrices. To test for the similarity of the correlation matrices we used Mantel's permutation test (two-sided with 999 permutations; mantel.test function in R/ape (Paradis & Schliep, 2019)).

Color scores

In terms of color, males of *P.* sp. "nyererei-like" are characterized by a crimson red dorsal head surface, dorsum and dorsal fin, and by yellow flanks (see Figure 1a). The most common male morphs of *N. omnicaeruleus* are blue on the head, dorsum, flanks, and dorsal fin. Some other common morphs feature some yellow or orange on the dorsal head surface, flanks, and operculum (and there are also several less common morphs; see (Seehausen *et al.*, 1998a)). The *N. omnicaeruleus* male used as grandfather here was mostly blue, with some yellow on the flanks and hints of yellow on the operculum and nose (see Figure 1a). We used the same color scoring scheme as in Feller *et al.*, 2020, which also features *P.* sp. "nyererei-like" as one of the parental species in a cross (but crossed to a different blue species). In brief, we scored the presence/absence of red and yellow in 12 different sectors on the body and fins (Figure 1d). It has been observed that males of *N. omnicaeruleus* sometimes start out yellow and only turn blue at a later stage, usually at the onset of maturity and territorial behavior (Maan *et al.*, 2006), and one of our F1 hybrid fathers seems to have gone through such a shift. This might mean that yellow is not a stable and mappable color difference between our two species (and we did not

document possible ontogenetic color changes in the F2 hybrid males). Red color in our cross however, is private to *P*. sp. "nyererei-like".

Shape

We used the same standardized color photos to place 20 landmarks as well as 8 semi-landmarks on each male F2 individual using tpsDig2.32 (Rohlf, 2018) (Figure S1). Two additional landmarks were placed on the ruler in the photo for scaling. We then imported the landmark data into R/geomorph v.3.3.1 (Adams *et al.*, 2020), where we performed all further shape analyses. We first scaled the landmarks and then performed a Generalized Procrustes analysis (using bending energy as criterion to optimize the positions of semilandmarks) to obtain shape data. Since we detected a significant effect of size/allometry ('procD.lm' and 'plotAllometry' functions) in this data, we used the residuals of a linear regression of the coordinates against Centroid size in subsequent analyses. We then performed a Principal Coordinate Analysis ('gm.prcomp' function), and extracted the PC scores for use in QTL mapping. To visualize shape changes along each of the six first PC-axes (each explaining a min 5% of variation and together ~70% of the variation), we plotted the overall mean shape against the most extreme shape at both ends of each axis ('plotRefToTarget' function).

Library preparation and sequencing

For DNA extraction from fin clips, we used Qiagen's DNeasy Blood & Tissue Kit. Library preparation for RAD tag sequencing followed Baird et al., 2008 with some modifications. We normalized DNA concentrations for all samples in one library prior to restriction enzyme digestion with SbfI. Each of the four produced libraries contained 48 individuals (42-47 F2s and one to six F1s/F0s). (Seven more F2s were later sequenced in a library for a different project but using the same protocol.) We used TruSeq P1-adaptors and a custom set of 5bp-8bp barcodes with TruSeq P2-adaptors. For shearing we used a Covaris M220 focused-ultrasonicator. For the first size selection after shearing, we used a SageELF machine. Each library was amplified in four aliquots of 50 µl. For the final size selection, we used Agencourt AMPure XP beads. Both size-selection steps targeted fragments of 300-700 bp. The libraries were single-end sequenced (150 bp) on four separate lanes on an Illumina HiSeq 2500 machine. Bacteriophage PhiX genomic DNA was added to each library (approximately 4.7% of reads).

Sequence processing and filtering

We used fastqc/0.11.4 (https://www.bioinformatics.babraham.ac.uk/projects/fastqc/) to assess the overall quality of the libraries and to determine the cutoff for trimming reads in the demultiplexing step. Using bowtie2/2.3.2 (Langmead & Salzberg, 2012) we first removed PhiX reads. For demultiplexing and trimming the reads (to 120 bp) we then used stacks/1.40 (Catchen *et al.*, 2013). We used the FASTQ Quality Filter (http://hannonlab.cshl.edu/fastx_toolkit/index.html) to filter reads for a minimum quality of 10 in all reads and a minimum quality of 30 in at least 95% of the reads. For the alignment to the anchored *Pundamilia nyererei* reference genome (Feulner *et al.*, 2018) we used bowtie2/2.3.2 (Langmead & Salzberg, 2012), using the default end-to-end alignment setting but allowing one mismatch. Base quality recalibration for each library based on its PhiX reads was done using the GATK BaseRecalibrator and PrintReads modules (McKenna *et al.*, 2010). We used GATKs (v3.7) UnifiedGenotyper for variant and genotype calling (minimum base quality score set to 20), only using uniquely aligned reads. Individuals were excluded from genotyping if they had <500,000

reads and /or <15,000 loci with min. 10 reads. The resulting vcf table contained genotypes at 2,045,191 sites for 185 individuals (141 F2 males, 30 F2 females, one F2 of undetermined sex, four F1 parents, seven other F1s, two F0 grandparents).

For filtering the vcf file we used bcftools (samtools/1.9 (Li *et al.*, 2009)) and vcftools/0.1.16 (Danecek *et al.*, 2011). First, we removed sites with >50% missing data, and individuals with a mean depth of <10. We then removed 10 sites around indels and only kept bi-allelic SNPs not exceeding the mean site depth by 1.5 times the interquartile range, and set genotypes with a depth of <10 or with quality <10 to missing. Individuals were checked for signs of PCR duplications (indicated by a skew in heterozygote read balance) using a custom Python script (written by David A. Marques), but none had to be excluded. We then again removed SNPs with >50% missing data and SNPs with a minor allele frequency of <5%. This resulted in a set of 16,604 quality filtered SNPs for 185 individuals (138 F2 males, 29 F2 females, one F2 of indeterminate sex, four F1 parents, seven other F1s, two F0 grandparents). We then identified SNPs with homozygous fixed differences between the two F0 grandparents that were additionally heterozygous in all F1 parents (excluding one F1 mother with poor sequencing quality). This resulted in a set of 2,544 SNPs for 168 F2s (138 males, 29 females, one F2 indeterminate sex) that were used in linkage map construction.

Linkage map construction

We constructed the linkage map in JoinMap 4.0 (van Ooijen, 2006). Seven individuals with unclear family assignments were excluded from linkage map construction and subsequent QTL analyses. We removed loci with extreme segregation distortion ($p<0.001$), loci with more than 20% missing genotypes, loci with high similarity (>0.950; i.e. that map to the same RAD locus and would thus not be informative for QTL mapping), and individuals with >40% missing data (n=5). We identified linkage groups based on an independence LOD of 5. The highest cross-link (SCL) value in the map is 4.9. We then used the Kosambi regression mapping algorithm to build the maps, using a LOD threshold 1.0, a recombination threshold 0.499, goodness-of-fit threshold of 5.0, and no fixed order. We performed two rounds of mapping, with a ripple after addition of each marker to the map. If a marker had a suspiciously high nearest neighbor (NN) fit value (here >20), we excluded it and repeated the mapping round. The final map contained 1,295 markers over 22 linkage groups (LGs) with a total map length of 1,237.1 cM and an average marker spacing of 1.0 cM (35 to 110 +/-16.6 markers per LG. LGs are numbered according to the *P. nyererei* ("Pun") reference and also given the corresponding *Oreochromis niloticus* ("Ore") number (see Feulner *et al.*, 2018).

Due to strong segregation distortion, a large number of markers were missing on the second half of Pun-LG10/Ore-LG23 in this map (see Figure 3). To ensure we were not missing any potential QTLs in this region, we made an alternative map in which we manually forced some markers from the distorted region and from the very end of Pun-LG10/Ore-LG23 into this LG (this led to some inflation in terms of distances on this LG), and we repeated the QTL mapping with this map. However, we found no additional QTLs when using this alternative linkage map.

QTL mapping

161 F2 hybrid individuals of two families were included in QTL mapping analyses. Family 1 contained 49 males and 23 females, family 2 contained 83 males and 6 females. We curently do not know the cause of the distorted sex ratio (but see Results part).

We used R/qtl v1.46-2 (Broman *et al.*, 2003) to perform QTL mapping. Since a few markers still appeared to be very close to each other we first applied the 'jittermap' function to jitter the marker positions in the map. We then calculated genotype probabilities using the 'calc.genoprob' function with a step size of 1 for the single-QTL scans and a step size of 3 for the two-QTL scans, an error probability of 0.05 and the Kosambi map function. For both the single- and two-dimensional QTL scans we used the EM maximum likelihood algorithm with the normal model, except for sex and the color scores, for which we used the binary model. Sex and/or family were included as additive covariates for traits showing a significant effect of either of the two in an anova. Genome-wide significance thresholds were estimated using n=1000 permutations for each trait. For the two-QTL scans (run for all traits but shape) we ran the permutations in four batches of 250 per trait on a high-performance computing cluster (each run with a different seed). Confidence intervals were estimated with the 'bayesint' function (within the 'summary' function in the 'tabByChr' format). We calculated percentage of variance explained (PVE) following Broman & Sen, 2009 using the formula $PVE=1-10^{((-2/n)*LOD)}*100$, where n is the number of individuals.

We additionally used the 'stepwiseqtl' function to fit and compare multiple QTL models for all morphological traits (except for shape). We used the Haley-Knott regression method, normal models, covariates if appropriate as described above, and we considered both additive QTL models and pairwise interactions among QTLs, for a maximum number of five QTLs. We calculated the penalties (using the 'calc.penalties' function) for the penalized LOD scores in these models on the basis of two-QTL scans with permutations (see Broman *et al.*, 2003), which we re-ran as described above but now also using Haley-Knott regression.

Results

Patterns of trait correlations are similar between parental species and F2s

In all trait correlation matrices we found a rather large number of significant trait correlations (Figure 2). As expected, the number of significant correlations was largest and the highest correlation coefficients were observed in the matrix that combines the parental species (Figure S2). The correlation matrices of each of the parental species on its own, and that of the F2s were all highly similar to each other based on Mantel tests: F2s vs *P.* sp. "nyererei-like": z=4.39, p=0.001 (z=4.19, p=0.001 without dental morphology); F2s vs *N. omnicaeruleus*: z=6.54, p=0.001 (z=6.26, p=0.003 without dental morphology); *P.* sp. "nyererei-like" vs *N. omnicaeruleus* z =6.12, p=0.001 (z=6.07, p=0.001 without dental morphology). The number of significant correlations however was smallest in the F2s (Figure 2). Some of the strongest correlations in the F2s were among different aspects of tooth morphology (Figure S2).

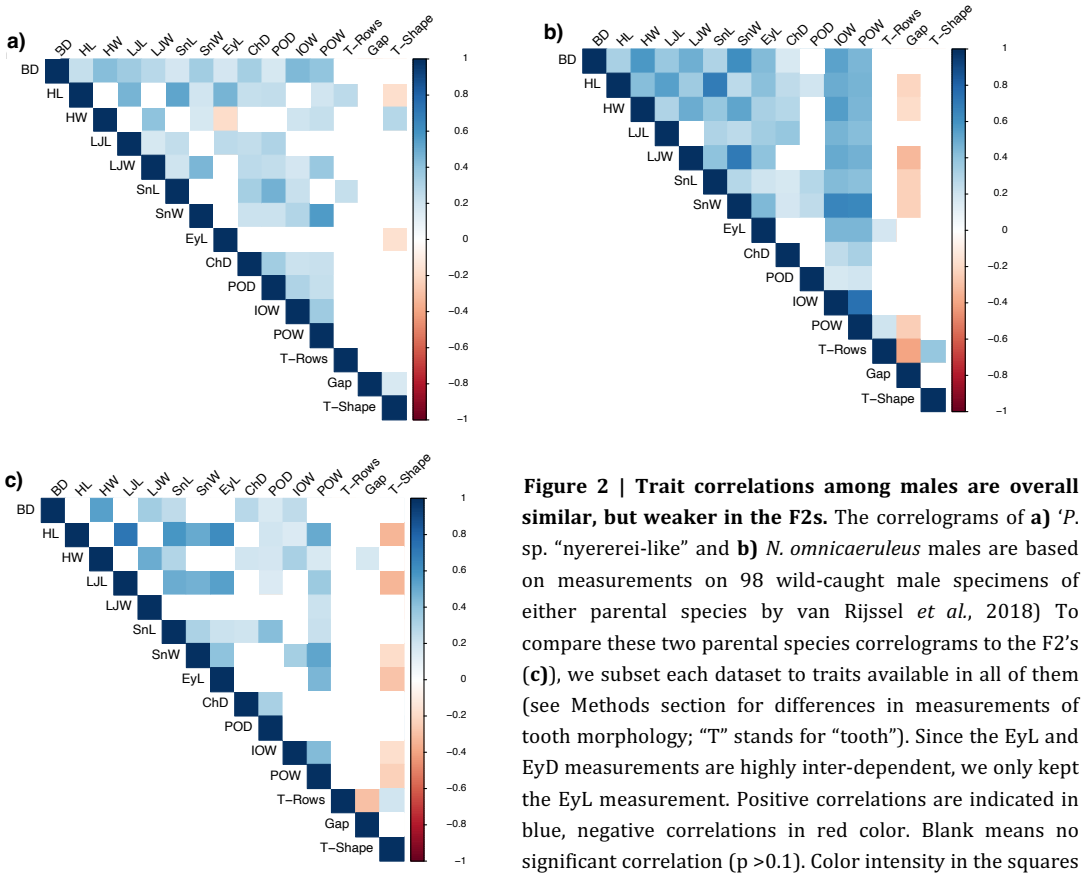


Figure 2 | Trait correlations among males are overall similar, but weaker in the F2s. The correlograms of **a)** '*P.* sp. "nyererei-like" and **b)** *N. omnicaeruleus* males are based on measurements on 98 wild-caught male specimens of either parental species by van Rijssel *et al.*, 2018) To compare these two parental species correlograms to the F2's (**c)**), we subset each dataset to traits available in all of them (see Methods section for differences in measurements of tooth morphology; "T" stands for "tooth"). Since the EyL and EyD measurements are highly inter-dependent, we only kept the EyL measurement. Positive correlations are indicated in blue, negative correlations in red color. Blank means no significant correlation (p >0.1). Color intensity in the squares are proportional to the correlation coefficients, as shown in the bar on the right. See Figure S3 for trait distributions.

QTL mapping indicates a distributed genetic architecture

We found four significant and one marginally significant QTLs for four of the 15 measured external morphology traits, mapping to four different chromosomes (Figure 3, Table 1): one QTL for LJL on Pun-LG4/Ore-LG16-21 (LOD=4.18, p=0.019), one QTL for HL on Pun-LG13/Ore-LG5 (LOD=4.19, p=0.028), an additional marginally significant QTL for HL on Pun-LG15/Ore-LG8-24 (LOD=3.86, p=0.057), one QTL for BD on Pun-LG20/Ore-LG4 (LOD=4.81, p=0.002), and one QTL for POD also on Pun-LG20/Ore-LG4 (LOD=3.99, p=0.033). The peaks for the two latter QTLs mapped to a different position on Pun-LG20/Ore-LG4 but with overlapping confidence intervals (Figure 3, Table 1).

A significant QTL for intestine length was found on Pun-LG16/Ore-LG14 (LOD=5.04, p=0.006). Shape PC6 also mapped to Pun-LG16/Ore-LG14 (LOD=5.51, p=0.001), but to a different region (with non-overlapping confidence intervals; Figure 3, Table 1). Shape on the upper extreme end of PC6 (which explains a modest 5.18% of variation) is mainly characterized by a short head, a small eye, and a steep head profile (overall more *N. omnicaeruleus*-like), while the lower extreme end of this axis is mainly characterized by a long head, a large eye and rather flat head profile (overall more *P.* sp."nyererei-like"-like) (see Figure S4).

Three aspects of tooth morphology mapped to the same chromosome (Pun-LG22/Ore-LG12) with partially overlapping confidence intervals: tooth density in both jaws and the number of tooth rows in the lower jaw (LOD>4, $p<0.05$ for all three; Figure 3, Table 1). While not significant, the highest peaks for tooth shape in both jaws were also seen on this chromosome.

Two-QTL (Table 2) and multiple QTL (Table S1) scans for all these morphological traits revealed only one additional QTL pair: for tooth shape in the lower jaw, two additive QTLs were detected, one on Pun-LG22/Ore-LG12 and one on Pun-LG12/Ore-LG17 (Table 2). The two already known QTLs for HL were confirmed in both analyses and shown to be purely additive (Table 2).

We found QTLs for three of the scored color traits (Figure 3, Table 1): two fully overlapping significant QTLs for red on the two sectors on the dorsal fin on Pun-LG2/Ore-LG7 (both LOD=4.55, $p<0.05$), and one marginally significant QTL for red on the nose on Pun-LG3/Ore-LG22 (LOD=3.52, $p=0.088$). Two-QTL scans revealed some additional and putatively interacting QTLs for red on the first sector of the dorsal fin (Pun-LG14/Ore-LG9 and Pun-LG16/Ore-LG14), for red on the nose (Pun-LGs13/Ore-LG5 and Pun-LG17/Ore-LG20) and for yellow on the cheek (Pun-LG4/Ore-LG16-21 and Pun-LG5/Ore-LG1) (see Table 2).

Finally, sex mapped to Pun-LG6/Ore-LG11 (LOD=5.2, $p<0.001$). Percentage variance explained (PVE) by this QTL for sex was 13.57, and female individuals were either one homozygous type (like *P*. sp. "nyererei-like") or heterozygous, while male individuals could be all three genotypes. Hence, this QTL alone does not explain sex, but our female sample size is too low to detect additional sex determining loci. The fact that all male individuals could be all three genotypes is however consistent with the scenario outlined in (Seehausen *et al.*, 1999), who found that some males had an additional autosomal male determiner that compensated the dominant W female determiner segregating in a population of *N. omnicaeruleus*. Also, the strong segregation distortion on Pun-LG10/Ore-LG23 (known to contain a sex determining region in another *Pundamilia* species (Feulner *et al.*, 2018; Feller *et al.*, 2021) might indicate that there are multiple sex determining systems segregating in this cross. This could also have caused the strong sex bias in our cross.

PVE of QTLs for all traits were between 10.6 and 17.6% (Table 1), i.e. they were all of moderate to large effect. Given our linkage map and our sample size, our statistical power to detect a significant QTL with an effect size of 10% PVE was approximately 64%.

Figure 3 | The QTLs for a suite of traits differentiating a specialised algae scraper from an insectivore are distributed over several and different linkage groups (LGs). Each panel shows one type of trait. **a)** external morphology (15 linear distances) **b)** tooth morphology **c)** intestine length **d)** shape (PC axes 1-6) **e)** colour (presence/absence scores of red and yellow on 12 body sectors) **f)** sex. Within each panel, traits with at least one marginally significant ($p<0.1$) QTL are highlighted in colour (see Table 1), traits with no detectable (marginally) significant QTL are plotted in grey. Dashed lines represent genome-wide significance threshold of $p<0.1$ (coloured by trait). LGs with a (marginally) significant QTL are highlighted with a yellow frame. 'Pun' numbers are given for each linkage group and correspond to the chromosome number of the anchored *P. nyrerei* reference genome (Feulner *et al.*, 2018), 'Ore' numbers are given below and correspond to the *O. niloticus* chromsome numbers (Lee *et al.*, 2005; Brawand *et al.*, 2014). The plus signs below the panels indicate the approximate position of additive QTLs detected in two-QTL scans, the crosses the approximate positions of putatively interacting QTLs (see Table 2).

Table 1. Results of single QTL scans (only (marginally; p<0.1) significant results shown).

trait	# inds	covar	marker	nearest marker	Pun-LG	Ore-LG	pos	ci. low	ci. high	lod	pval	PVE
Red dorsal fin 1	132	family	c2.loc43	chr2_18179005	2	7	43.0	34	56	4.55	0.009	14.68
Red dorsal fin 2	132	family	c2.loc43	chr2_18179005	2	7	43.0	34	63.3	4.55	0.006	14.68
Red nose	132	none	chr3_21045798		3	22	36.4	23	43	3.52	0.088	11.56
Lower Jaw Length	158	both	chr4_4634552		4	16-21	13.7	6	36.4	4.18	0.019	11.47
Head Length	158	both	chr13_11189376		13	5	24.9	14	28	4.19	0.028	11.50
Head Length	158	both	c15.loc10	chr15_7913411	15	8-14	10.0	0	35	3.86	0.057	10.64
Intestine Length	158	family	c16.loc35	chr16_15654960	16	14	35.0	12	41	5.04	0.006	13.66
Shape PC6	131	family	chr16_26688311		16	14	54.4	50	57.9	5.51	0.001	17.61
Body Depth	157	sex	chr20_8108045		20	4	32.2	16	42	4.81	0.002	13.16
Pre-Orbital Depth	158	sex	chr20_3616200		20	4	16.6	5	35.2	3.99	0.033	10.98
Tooth Density Upper Jaw	161	none	chr22_22167437		22	12	43.0	22	50	4.04	0.041	10.91
Number of Inner Tooth Rows	160	both	c22.loc48	chr22_23367185	22	12	48.0	33	53.8	4.97	0.001	13.33
Tooth Density Lower Jaw	161	family	c22.loc28	chr22_7741136	22	12	28.0	0	32	4.83	0.008	12.90

inds , number of individuals; covar, additive covariates used in QTL scan; nearest marker, nearest marker in cases the QTL peak was on an interpolated marker (positions between genotyped markers inferred by the calc.genoprob function in R/qtl); Pun-LG, linkage group numbers corresponding to the anchored *Pundamilia nyrer*ei reference genome; Ore-LG, linkage group number corresponding to the *Oreochromis niloticus* reference genome; pos, position on the linkage map in cM; ci.low and ci.high, lower and upper limit of 95% approximate Bayesian credible interval in cM; pval, p-value determined by 1000 permutation; PVE, percent variance explained.

Table 2. Results of 2-QTL scans (only (marginally) significant results shown.

trait	chr 1	chr 2	pos 1f	pos 2f	lod. full	pval. full	lod. fv1	pval. fv1	lod. int	pval. int	pos 1a	pos 2a	lod. add	pval. add	lod. av1	pval. av1
Head Length	13	15	15	36	9.90	0.017	5.80	0.619	1.78	1	24	9	8.12	0.002	4.02	**0.004**
Tooth Shape Lower Jaw	12	22	27	30	7.50	0.537	4.34	1	0.91	1	27	30	6.59	0.043	3.43	**0.03**
Red dorsal fin 1	14	16	33	21	8.94	0.099	7.12	***0.089***	4.99	0.712	51	0	3.95	0.886	2.13	0.906
Red nose	13	17	21	42	9.03	0.054	8.20	**0.009**	7.59	**0.008**	0	15	1.44	1	0.61	1
Yellow cheek	4	5	0	60	7.54	0.030	5.28	***0.068***	4.08	0.263	3	15	3.46	0.988	1.20	1

lod.full, comparing the full model (two QTLs with interaction) to the null model (no QTLs); lod.fv1, comparing the full model to the best single QTL model; lod.int, comparing the full model to the additive (two QTLs with no interaction) model i.e. indicating evidence for epistasis; lod.add, comparing the additive model to the null model; lod.av1 comparing the additive model to the best single QTL model. Significant p-values are highlighted in bold; marginally significant p-values bold and cursive.

QTL effects and dominance

Most QTLs were in the expected direction. That is, in traits where wild-caught *P.* sp. "nyererei-like" males have larger trait values than *N. omnicaeruleus* males, the largest trait values in the F2s were associated with the *P.* sp. "nyererei-like" (AA) genotype, and the smallest with the *N. omnicaeruleus* (BB) genotype, and vice versa depending on the trait (Figures 4 and S3). Only the QTL for BD did not conform to this pattern. There was no clear expectation for POD, which is one of the few traits in which transgressive trait values were observed in the F2s (here broadly defined as lying outside the combined parental species' distribution; see Figure S3). Most QTLs showed signs of dominance (or even slight overdominance), with the exception of the QTLs for intestine length and red on the dorsal fin (Figure 4). All dominance effects were attributed to the *N. omnicaeruleus* alleles.

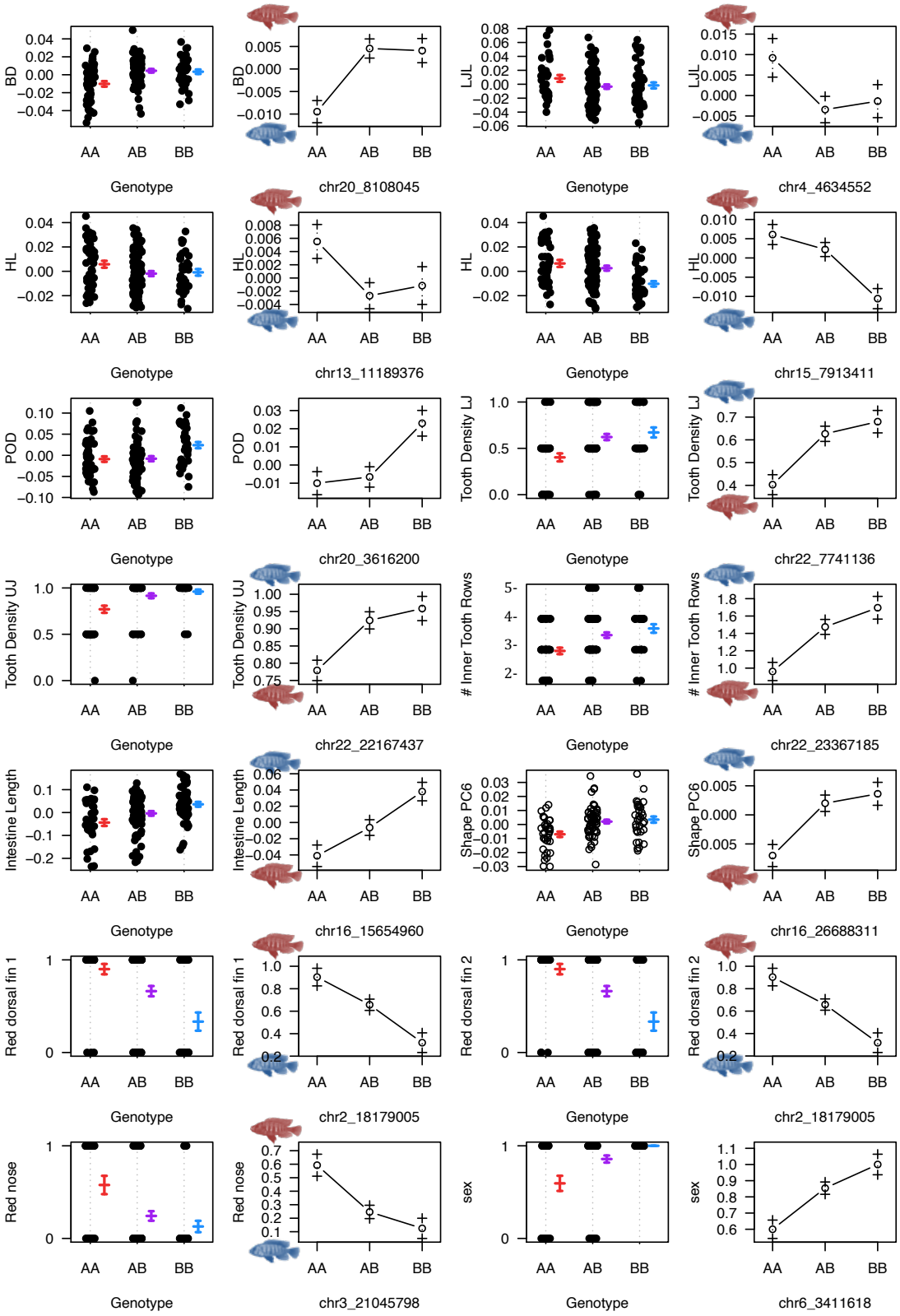


Figure 4 | QTL effects are highly concordant. Shown are trait distributions among the F2s and effect plots for all traits with (marginally) signifcant QTLs (see Table 1). The AA genotype corresponds to *P.* sp. "nyererei-like", the BB genotype to *N. omnicaeruleus*, AB are heterozygotes. If the QTL peak was on an interpolated marker, the nearest flanking marker is shown (see Table 1). The fish cartoons indicate which parental species is associated with the larger values in a given trait (red= *P.* sp. "nyererei-like", blue= *N. omnicaeruleus*; see also Figure S3). Only the QTL for BD is discordant

.

Discussion

We asked if an explosive trophic radiation that involved repeated transitions between trophic levels with considerable functional refinement might have been facilitated by genetic architectures involving tight linkage or pleiotropy among co-adapted trophic traits. However, we found a distributed genetic architecture involving many different chromosomes harboring small to moderate effect loci with little evidence for linkage or pleiotropy to underlie a suite of traits that distinguish specialized herbivores from invertivores.

With our sample size we had limited power to detect QTLs of moderate effect and insufficient power to detect QTLs of small effect. However, the moderate to large effect QTLs we were able to detect clearly show a non-clustered genetic architecture among these traits. This is consistent with other recent QTL mapping studies of functional morphology in haplochromine cichlids. For instance, a large number of small-effect (and mostly additive) loci underlie both lip and head morphology, with overall little overlap, in a cross of the two Lake Victoria invertivores *Paralabidochromis chilotes* and *P. nyererei* (Henning *et al.*, 2017). Similarly, (Navon *et al.*, 2017) found several QTLs each for three types of traits (craniofacial traits, fin and body shape) with little overlap in a cross of the two algivorous Lake Malawi haplochromines *Labeotropheus fuelleborni* and *Tropheops* sp. "red cheek".

We found very little overlap between QTLs for a suite of traits distinguishing our two species from two different trophic levels. The exception to this was oral jaw tooth and dental morphology: the QTLs with the largest effects for several aspects of tooth and dental morphology were located on the same linkage group (Pun-LG22/Ore-LG12), with partially overlapping confidence intervals. This linkage group has not previously been reported to be associated with oral jaw tooth morphology in cichlids. Tooth shape in a cross between the two algivorous Lake Malawi cichlids *L. fuelleborni* and *Metriaclima zebra* has been mapped to Pun-LGs 13, 14, and 7 / Ore-LGs 5, 9, and 15 (Albertson *et al.*, 2003; Streelman & Albertson, 2006), tooth density in another Malawi cichlids cross between the planktivore *Cynotilapia afra* and the algivore *Pseudotropheus elongatus* to several linkage groups with the highest peak on Pun-LG6/Ore-LG11 (Bloomquist *et al.*, 2015). The number of teeth both in the outer row of the oral jaws and on the pharyngeal jaws in the cross between the Lake Victoria cichlids *P. chilotes* and *P. nyererei* also mapped to Pun-LG6/Ore-LG11 (Hulsey *et al.*, 2017). *P. nyererei* is closely related to one of the parental species in our cross (*P.* sp. "nyererei-like"), and the total number of teeth in the outer row of the oral jaws should be related to our measure of tooth density. The finding of a different linkage group may hence seem surprising. However, in both studies the percentage of variance explained by the major QTL was modest (10.9-13.3% in our case, and 7-8% in Hulsey *et al.*, 2017), implying the presence of a number of additional smaller effect loci. Moreover, in our case, the dominant alleles come from the other species (*N. omnicaeruleus*) in the cross. Finally, whereas we crossed a species with many teeth (*N. omnicaeruleus*) and one with moderate number of teeth (*P.* sp. "nyererei-like"), the earlier studies had crossed a species with moderate number of teeth (*P. nyererei*) and one with a very low number of teeth (*P. chilotes*). Therefore, it seems possible that different QTLs are associated with the difference between very low and moderate number of teeth in the oral jaws, and that between the latter and a high number of teeth.

To our knowledge, ours is the first study of genetically mapping intestine length in cichlids. In addition to several morphological traits in the head, both tooth and dental morphology (i.e. number and shape of teeth) and intestine length are strongly correlated with

carnivorous versus herbivorous diet in cichlids (reviewed in (Takahashi & Koblmüller, 2011)) and constitute important diagnostic differences between species of these two trophic levels, including our two crossed species (Greenwood, 1981; Witte & Van Oijen, 1990). A long intestine is a major indicator of adaptation to herbivory not only in fishes (e.g. Ribble & Smith, 1983; Elliott & Bellwood, 2003; Wagner *et al.*, 2009) but also in birds (Ricklefs, 1996), mammals (Korn, 1992) and reptiles (O'Grady *et al.*, 2005). While expensive to maintain, having a longer gut likely serves to better take up nutrients from food that is more resilient to digestion by increasing retention time (Horn, 1989). It has been shown to have a strong plastic component (e.g. Olsson *et al.*, 2007; Naya *et al.*, 2008; Wagner *et al.*, 2009), which has been suggested to provide a mechanism that could facilitate trophic shifts in evolutionary radiations (Wagner *et al.*, 2009). In our study, the QTL for intestine length is among the ones of largest effect, clearly indicating a genetic basis for this trait. We cannot rule out an additional plastic component however, as intestine length in the F2 hybrid males (mean 1.46 +/- 0.27, min 0.84, max 2.1) only exceeded two times their body standard length (SL) in two individuals (and it was shorter than two times their standard length in all others), while it has been described as up to three or four times SL in wild *N. omnicaeruleus* (Seehausen, 1996). The shorter intestines that we observe in general could hence well reflect a plastic response to the common-environment aquarium food, which is made up of both plant and animal components. However, our detection of a QTL in the F2 and the direction of the allelic effects are clear evidence for a genetic basis of differences in intestine length between the invertivorous and the herbivorous species.

Transitioning from insecti-/invertivores to herbivores has likely happened several times in the rapid radiation of the Lake Victoria cichlids, and possibly the reverse has happened too (McGee *et al.*, 2020). Given that such transitions require concerted changes in many different traits, we hypothesized that such concerted changes could have been facilitated by a 'simple' genetic architecture involving pleiotropy and/or physical linkage (e.g. Kirkpatrick & Barton, 2006; Schwander *et al.*, 2014), especially since much of the diversification will have happened in the face of gene flow (Bürger & Akerman, 2011; Yeaman & Otto, 2011; Yeaman & Whitlock, 2011). Contrary to this hypothesis, we found that most of the QTLs for the different morphological traits whose co-adaptation is functionally required were not co-localized. This is comparable to the study of (Arnegard *et al.*, 2014) who found a polygenic and distributed trait architecture to underlie the benthic vs limnetic divergence in freshwater stickleback that has occurred repeatedly in a narrowly confined part of the global stickleback distribution range. Furthermore, while there were significant trait correlations among many traits in our F2 hybrids, they were overall fewer than in either parental species, also implying that physical linkage (or pleiotropy) among these traits is not very strong. The exception to this may be dental traits, in which some of the strongest correlations were observed in the F2s, and which also mapped to the same genomic region. Taken together, this suggests that at the stage where herbivores evolved from the invertivorous ancestors, multilocus LD had to build between many regions in the genome. Theoretical studies have suggested that strong LD between multiple co-selected loci can evolve rapidly even in the presence gene flow, but only after many thousands of generations of a slow build-up of LD (Flaxman *et al.*, 2014; Schilling *et al.*, 2018). This would thus be too slow for speciation events as rapid as in the Lake Victoria radiation. The evolution and maintenance of such genome-wide LD between multiple unlinked genomic regions would not only have required divergent selection but also reproductive isolation (Felsenstein, 1981; Dieckmann & Doebeli, 1999; Kondrashov & Kondrashov, 1999).

Two observations in our study might provide evidence for divergent selection that would have acted during the transition from the invertivorous ancestral to the herbivorous

derived state: the highly concordant effects of all but one of the detected QTLs (only one QTL is in the antagonistic direction) in our study are consistent with the sorting of alleles under divergent ecological selection between the two niches. Furthermore, compared to crosses between ecologically similar species (Stelkens *et al.*, 2009; Selz *et al.*, 2014a) we observe little transgressive variation in our cross, which is also consistent with theoretical expectations of the sorting of alleles under divergent selection (Rieseberg *et al.*, 1999). Moreover, our species belong to lineages that are behaviorally reproductively isolated in sympatry (Seehausen *et al.*, 1998b). Reproductive isolation among many sympatric Lake Victoria cichlids is to a large degree due to assortative mating via behavioral mate choice (Seehausen *et al.*, 1998b). Sexual selection pressures have been shown to be prevalent in this system: Lake Victoria haplochromines are sexually dimorphic (Seehausen, 1996), and male nuptial coloration is under strong sexual selection via both (inter- as well as intraspecific) male-male competition (Seehausen & Schluter, 2004; Dijkstra *et al.*, 2005, 2007) and female mate choice (Haesler & Seehausen, 2005; Stelkens *et al.*, 2008; Selz *et al.*, 2016; Svensson *et al.*, 2017) and is thus a key trait for behavioral reproductive isolation. Theory predicts that the evolution and maintenance of reproductive isolation in a sympatric setting should be facilitated if mate choice traits are linked with traits under disruptive ecological selection (Kirkpatrick & Ravigné, 2002; Gavrilets, 2003; Coyne & Orr, 2004; Van Doorn *et al.*, 2009). But interestingly, we see no physical linkage between mating traits (male color) and functionally relevant morphological traits in our cross.

Rapid radiation across trophic levels in the presence of gene flow thus seems to have required rapid evolution of robust reproductive isolation, and the availability of many genetic variants with at least moderate effects on several co-adapting traits. We argue that combinatorial processes (see Marques *et al.*, 2019) could provide this type of genetic architecture and facilitate the evolution of linkage disequilibrium even with some gene flow. (Meier *et al.*, 2017a) have recently demonstrated that the Lake Victoria radiation was fuelled by ancient hybridization between divergent lineages. Upon colonization of the newly re-filled lake some 15'000 years ago (Johnson *et al.*, 2000; Stager & Johnson, 2008), polymorphisms from ancestral hybridization (predating the re-filling of the lake by several thousands of years (Meier *et al.*, 2017a)) were sorted and recombined into many new combinations by natural and sexual selection. Such old genetic variants have been suggested to provide particularly good substrate for generating many novel combinations and facilitating rapid speciation when open niches are available (Marques *et al.*, 2019). This is because they can constitute blocks of already co-adapted SNPs that have been previously tested by selection and that can have relatively large phenotypic effects, thus providing a genetic architecture that is robust against gene flow and that can allow the crossing of fitness valleys, both of which would be hard to achieve by de novo mutations alone (Gavrilets, 2004; Orr, 2005; Kagawa & Takimoto, 2017; Marques *et al.*, 2019). In the combinatorial view (Marques *et al.*, 2019), an admixed population with initially high frequencies of functionally relevant haplotype polymorphisms could then more easily and rapidly be sorted into many different adaptations (involving concerted change to multiple traits) to a variety of niches (Marques *et al.*, 2019).

In conclusion, we have shown that many different genomic regions contribute to the functional morphological differences between representatives of a carnivorous and a herbivorous lineage within the rapid radiation of Lake Victoria cichlid fish, with little evidence for linkage or pleiotropy. Repeated and rapid transitions between these two feeding modes and trophic levels in the course of the adaptive radiation thus likely required the early rapid evolution of robust reproductive isolation, and the presence of many genetic variants with at least moderate effects on feeding ecology. Follow-up investigations should narrow down the

genomic regions identified by QTL mapping to screen for the causative variants underlying a given trait. We predict that indels might be involved since many of the old haplotypes in the Lake Victoria radiation have been shown to carry indels associated with specific ecologies (McGee *et al.*, 2020)

Acknowledgements

Many thanks to Andreas Taverna for breeding the fish, to Oliver Selz for initiating the cross, to the FishEc team for many valuable discussions, and to Katie Peichel for many useful inputs throughout the project. Sequencing of the RAD libraries was done at the Genomic Technologies Facility of the University of Lausanne, and sequence processing was performed on a cluster managed by the Genetic Diversity Centre (GDC) at ETH Zurich. This research was funded by Swiss National Science Foundation (SNSF) grants 31003A_163338, 31003A_144046 and 31003A_118293 to OS.

Author contributions

OS conceived the study. AFF collected the data and conducted the analyses. AFF and OS wrote the manuscript.

Conflicts of interest statement

We declare that we have no competing interests.

Data accessibility

Raw read (fastq) files for all genotyped individuals will be made available on the sequence read archive (SRA) on GenBank. Phenotype and genotype tables, and R scripts will be made available on the Dryad Digital Repository.

Animal experimentation permit

Fish experimentation and euthanasia followed the Swiss national guidelines for the care and use of laboratory animals and were authorized by the veterinary offices of the cantons of Bern and Lucerne (License number: BE 65/18).

References

Adams, D., Collyer, M. & Kaliontzopoulou, A. 2020. Geomorph: Software for geometric morphometric analyses. R package version 3.2.1.

Albertson, R.C., Streelman, J.T. & Kocher, T.D. 2003. Directional selection has shaped the oral jaws of Lake Malawi cichlid fishes. *Proc. Natl. Acad. Sci.* **100**: 5252–5257.

Albertson, R.C., Streelman, J.T., Kocher, T.D. & Yelick, P.C. 2005. Integration and evolution of the cichlid mandible: The molecular basis of alternate feeding strategies. *Proc. Natl. Acad. Sci. U. S. A.* **102**: 16287–16292.

Arnegard, M.E., McGee, M.D., Matthews, B., Marchinko, K.B., Conte, G.L., Kabir, S., *et al.* 2014. Genetics of ecological divergence during speciation. *Nature* **511**: 307–311.

Baerends, G.P. & Baerends-Van Roon, J.T. 1950. *An Introduction to the Study of the Ethology of the Cichlid Fishes*. E. J. Brill, Leiden, NL.

Baird, N.A., Etter, P.D., Atwood, T.S., Currey, M.C., Shiver, A.L., Lewis, Z.A., *et al.* 2008. Rapid SNP Discovery and Genetic Mapping Using Sequenced RAD Markers. *PLoS One* **3**: e3376.

Barel, C.D.N., van Oijen, M.J.P., Witte, F. & Witte-Maas, E.L.S. 1977. An introduction to the taxonomy and morphology of the haplochromine Cichlidae from Lake Victoria. *Netherlands J. Zool.* **27**: 381–389.

Bell, M.A. & Foster, S.A. 1994. *The Evolutionary Biology of the Threespine Stickleback*. Oxford University Press, Oxford, UK.

Bernatchez, L. 2004. Ecological theory of adaptive radiation: an empirical assessment from coregonine fishes (Salmoniformes). In: *Evolution illuminated: salmon and their relatives* (A. P. Hendry & S. C. Stearns, eds), pp. 175–207. Oxford University Press, Oxford, UK.

Bloomquist, R.F., Parnell, N.F., Phillips, K.A., Fowler, T.E., Yu, T.Y., Sharpe, P.T., *et al.* 2015. Coevolutionary patterning of teeth and taste buds. *Proc. Natl. Acad. Sci. U. S. A.* **112**: E5954–E5962.

Bouton, N. 1999. *Jaws. On the feeding ecology of rock-dwelling cichlid fishes from Lake Victoria*. University of Leiden, Leiden, NL.

Bouton, N., Os, N. & Witte, F. 1998. Feeding performance of Lake Victoria rock cichlids: testing predictions from morphology. *J. Fish Biol.* **53**: 118–127.

Bouton, N., Seehausen, O. & van Alphen, J.J.M. 1997. Resource partitioning among rock-dwelling haplochromines (Pisces: Cichlidae) from Lake Victoria. *Ecol. Freshw. Fish* **6**: 225–240.

Bouton, N., Witte, F., Van Alphen, J.J.M., Schenk, A. & Seehausen, O. 1999. Local adaptations in populations of rock-dwelling haplochromines (Pisces: Cichlidae) from southern Lake Victoria. *Proc. R. Soc. B Biol. Sci.* **266**: 355–360.

Brichard, P. 1978. *Fishes of Lake Tanganyka*. Tropical Fish Hobbyist, Neptune City, NJ.

Broman, K.W. & Sen, S. 2009. *A Guide to QTL Mapping with R/qtl*. Springer New York, New York, NY.

Broman, K.W., Wu, H., Sen, S. & Churchill, G.A. 2003. R/qtl: QTL mapping in experimental crosses. *Bioinformatics* **19**: 889-890.

Bürger, R. & Akerman, A. 2011. The effects of linkage and gene flow on local adaptation: A two-locus continent-island model. *Theor. Popul. Biol.* **80**: 272–288.

Carlson, R.L. & Wainwright, P.C. 2010. The ecological morphology of darter fishes (Percidae: Etheostomatinae). *Biol. J. Linn. Soc.* **100**: 30–45.

Catchen, J., Hohenlohe, P.A., Bassham, S., Amores, A. & Cresko, W.A. 2013. Stacks: an analysis tool set for population genomics. *Mol. Ecol.* **22**: 3124–3140.

Choat, J.H. & Clements, K.D. 1998. Vertebrate herbivores in marine and terrestrial environments: A Nutritional Ecology Perspective. *Annu. Rev. Ecol. Syst.* **29**: 375–403.

Cooper, W.J., Wernle, J., Mann, K. & Albertson, R.C. 2011. Functional and Genetic Integration in the Skulls of Lake Malawi Cichlids. *Evol. Biol.* **38**: 316–334.

Coyne, J.A. & Orr, H.A. 2004. *Speciation*. Sinauer Associates, Sunderland, Massachusets

Danecek, P., Auton, A., Abecasis, G., Albers, C.A., Banks, E., DePristo, M.A., *et al.* 2011. The variant call format and VCFtools. *Bioinformatics* **27**: 2156–2158.

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

Dijkstra, P.D., Seehausen, O. & Groothuis, T.G.G. 2005. Direct male-male competition can facilitate invasion of new colour types in Lake Victoria cichlids. *Behav. Ecol. Sociobiol.* **58**: 136–143.

Dijkstra, P.D., Seehausen, O., Pierotti, M.E.R. & Groothuis, T.G.G. 2007. Male-male competition and speciation: aggression bias towards differently coloured rivals varies between stages of speciation in a Lake Victoria cichlid species complex. *J. Evol. Biol.* **20**: 496–502.

Elliott, J.P. & Bellwood, D.R. 2003. Alimentary tract morphology and diet in three coral reef fish families. *J. Fish Biol.* **63**: 1598–1609.

Feller, A.F., Haesler, M.P., Peichel, C.L. & Seehausen, O. 2020. Genetic architecture of a key reproductive isolation trait differs between sympatric and non-sympatric sister species of Lake Victoria cichlids. *Proc. R. Soc. B Biol. Sci.* **287**.

Feller, A.F., Ogi, V., Seehausen, O. & Meier, J.I. 2021. Identification of a novel sex determining chromosome in cichlid fishes that acts as XY or ZW in different lineages. *Hydrobiologia* **7**.

Felsenstein, J. 1981. Skepticism Towards Santa Rosalia, or Why are There so Few Kinds of Animals? *Evolution.* **35**: 124.

Feulner, P.G.D., Schwarzer, J., Haesler, M.P., Meier, J.I. & Seehausen, O. 2018. A Dense Linkage Map of Lake Victoria Cichlids Improved the Pundamilia Genome Assembly and Revealed a Major QTL for Sex-Determination. *G3; Genes|Genomes|Genetics* **8**: 2411–2420.

Fisher, R. 1930. *The Genetical Theory of Natural Selection*. The Clarendon Ptess, Oxford, UK.

Flaxman, S.M., Wacholder, A.C., Feder, J.L. & Nosil, P. 2014. Theoretical models of the influence of genomic architecture on the dynamics of speciation. *Mol. Ecol.* **23**: 4074–4088.

Fryer, G. & Iles, T.D. 1972. *The cichlid fishes of the great lakes of Africa: their biology and evolution*. Oliver & Boyd, Edinburgh, UK.

Gavrilets, S. 2004. *Fitness landscapes and the origin of species*. Princeton University Press, Princeton, NJ.

Gavrilets, S. 2003. Perspective: Models of speciation - What have we learned in 40 years? *Evolution.* **57**: 2197–2215.

Genner, M.J. & Turner, G.F. 2012. Ancient Hybridization and Phenotypic Novelty within Lake Malawi's Cichlid Fish Radiation. *Mol. Biol. Evol.* **29**: 195–206.

Golcher-Benavides, J. & Wagner, C.E. 2019. Playing out liem's paradox: Opportunistic piscivory across lake tanganyikan cichlids. *Am. Nat.* **194**: 260–267.

Goldschmidt, T., Witte, F. & Visser, J. De. 1990. Ecological Segregation in Zooplanktivorous Haplochromine Species (Pisces : Cichlidae) from Lake Victoria. *Oikos* **58**: 343–355.

Greenwood, P.H. 1974. *Cichlid fishes of Lake Victoria, East Africa: the biology and evolution of a species flock*. John Wright and Sons Ltd., Stonebridge Press, Bristol, UK.

Greenwood, P.H. 1981. *The Haplochromine Fishes of the East African Lakes*. Cornell Unversity Press, Ithaca, NY.

Haesler, M.P. & Seehausen, O. 2005. Inheritance of female mating preference in a sympatric sibling species pair of Lake Victoria cichlids: implications for speciation. *Proc. R. Soc. B Biol. Sci.* **272**: 237–245.

Hansen, T.F. 2003. Is modularity necessary for evolvability? Remarks on the relationship between pleiotropy and evolvability. *BioSystems* **69**: 83–94.

Harrell, F. 2020. R package "Hmisc."

Henning, F., Machado-Schiaffino, G., Baumgarten, L. & Meyer, A. 2017. Genetic dissection of adaptive form and function in rapidly speciating cichlid fishes. *Evolution.* **71**: 1297–1312.

Henning, F. & Meyer, A. 2014. The evolutionary genomics of cichlid fishes: Explosive speciation and adaptation in the postgenomic Era. *Annu. Rev. Genomics Hum. Genet.* **15**: 417–441.

Horn, M. 1989. Biology of marine herbivorous fishes. *Ocean. Mar. Biol. Ann. Rev.* 167–272.

Horn, M. 1998. Feeding and digestion. In: *The physiology of fishes* (D. H. Evans, ed), pp. 43–63. CRC, Boca Raton, FL.

Hudson, A.G., Vonlanthen, P., Müller, R. & Seehausen, O. 2007. Review: The geography of speciation and adaptive radiation in coregonines. *Adv. Limnol.* **60**: 111–146.

Hulsey, C.D., Machado-Schiaffino, G., Keicher, L., Ellis-Soto, D., Henning, F. & Meyer, A. 2017. The Integrated Genomic Architecture and Evolution of Dental Divergence in East African Cichlid Fishes (Haplochromis chilotes x H. nyererei). *G3; Genes|Genomes|Genetics* **7**: 3195–3202.

Irisarri, I., Singh, P., Koblmüller, S., Torres-Dowdall, J., Henning, F., Franchini, P., *et al.* 2018. Phylogenomics uncovers early hybridization and adaptive loci shaping the radiation of Lake Tanganyika cichlid fishes. *Nat. Commun.* **9**: 3159.

Jay, P., Whibley, A., Frézal, L., Rodríguez de Cara, M.Á., Nowell, R.W., Mallet, J., *et al.* 2018. Supergene Evolution Triggered by the Introgression of a Chromosomal Inversion. *Curr. Biol.* **28**: 1839-1845.e3.

Johnson, T.C., Kelts, K. & Odada, E. 2000. The Holocene History of Lake Victoria. *AMBIO A J. Hum. Environ.* **29**: 2–11.

Jonsson, B. & Jonsson, N. 2001. Polymorphism and speciation in Arctic charr. *J. Fish Biol.* **58**: 605–638.

Joyce, D.A., Lunt, D.H., Bills, R., Turner, G.F., Katongo, C., Duftner, N., *et al.* 2005. An extant cichlid fish radiation emerged in an extinct Pleistocene lake. *Nature* **435**: 90–95.

Kagawa, K. & Takimoto, G. 2017. Hybridization can promote adaptive radiation by means of transgressive segregation. *Ecol. Lett.* **21**: 264–274.

Kirkpatrick, M. & Barton, N. 2006. Chromosome inversions, local adaptation and speciation. *Genetics* **173**: 419–434.

Kirkpatrick, M. & Ravigné, V. 2002. Speciation by natural and sexual selection: Models and experiments. *Am. Nat.* **159**.

Klemetsen, A. 2010. The Charr Problem Revisited: Exceptional Phenotypic Plasticity Promotes Ecological Speciation in Postglacial Lakes. *Freshw. Rev.* **3**: 49–74.

Kocher, T.D. 2004. Adaptive evolution and explosive speciation: the cichlid fish model. *Nat. Rev. Genet.* **5**: 288–298.

Kondrashov, A.S. & Kondrashov, F.A. 1999. Interactions among quantitative traits in the course of sympatric speciation. *Nature* **400**: 351–354.

Korn, H. 1992. Intestine lengths of Southern African savanna rodents and insectivores: intra- and interspecific comparisons. *J. Zool.* **228**: 455–460.

Langmead, B. & Salzberg, S.L. 2012. Fast gapped-read alignment with Bowtie 2. *Nat. Methods* **9**: 357–359.

Láruson, Á.J., Yeaman, S. & Lotterhos, K.E. 2020. The Importance of Genetic Redundancy in Evolution. *Trends Ecol. Evol.* **35**: 809–822.

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., *et al.* 2009. The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25**: 2078–2079.

Li, J., Cocker, J.M., Wright, J., Webster, M.A., McMullan, M., Dyer, S., *et al.* 2016. Genetic architecture and evolution of the S locus supergene in Primula vulgaris. *Nat. Plants* **2**: 1–7.

Liem, K.F. 1980. Adaptive significance of intra- and interspecific differences in the feeding repertoires of cichlid fishes. *Integr. Comp. Biol.* **20**: 295–314.

Lu, G. & Bernatchez, L. 1999. Correlated Trophic Specialization and Genetic Divergence in Sympatric Lake Whitefish Ecotypes (Coregonus clupeaformis): Support for the Ecological Speciation Hypothesis. *Society* **53**: 1491–1505.

Maan, M.E., Haesler, M.P., Seehausen, O. & Van Alphen, J.J.M. 2006. Heritability and heterochrony of polychromatism in a Lake Victoria Cichlid fish: stepping stones for speciation? *J. Exp. Zool. Part B Mol. Dev. Evol.* **306B**: 168–176.

Marques, D.A., Meier, J.I. & Seehausen, O. 2019. A Combinatorial View on Speciation and Adaptive Radiation. *Trends Ecol. Evol.* **34**: 531–544.

McGee, M.D., Borstein, S.R., Meier, J.I., Marques, D.A., Mwaiko, S., Taabu, A., *et al.* 2020. The ecological and genomic basis of explosive adaptive radiation. *Nature* **586**: 75–79.

McKaye, K.R. & Marsh, A. 1983. Food switching by two specialized algae-scraping cichlid fishes in Lake Malawi, Africa. *Oecologia* **56**: 245–248.

McKenna, A., Hanna, M., Banks, E., Sivachenko, A., Cibulskis, K., Kernytsky, A., *et al.* 2010. The Genome Analysis Toolkit: A MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res.* **20**: 1297–1303.

Meier, J.I., Marques, D.A., Mwaiko, S., Wagner, C.E., Excoffier, L. & Seehausen, O. 2017a. Ancient hybridization fuels rapid cichlid fish adaptive radiations. *Nat. Commun.* **8**.

Meier, J.I., Marques, D.A., Wagner, C.E., Excoffier, L. & Seehausen, O. 2018. Genomics of parallel ecological speciation in Lake Victoria cichlids. *Mol. Biol. Evol.* **35**: 1489–1506.

Meier, J.I., Sousa, V.C., Marques, D.A., Selz, O.M., Wagner, C.E., Excoffier, L., *et al.* 2017b. Demographic modelling with whole-genome data reveals parallel origin of similar Pundamilia cichlid species after hybridization. *Mol. Ecol.* **26**: 123–141.

Muller, H. 1950. Our load of mutations. *Am. J. Hum. Genet. Vol.* **2**: 111–176.

Nakamura, I. & Sato, K. 2014. Ontogenetic shift in foraging habit of ocean sunfish Mola mola from dietary and behavioral studies. *Mar. Biol.* **161**: 1263–1273.

Navon, D., Olearczyk, N. & Albertson, R.C. 2017. Genetic and developmental basis for fin shape variation in African cichlid fishes. *Mol. Ecol.* **26**: 291–303.

Naya, D.E., Bozinovic, F. & Karasov, W.H. 2008. Latitudinal trends in digestive flexibility: Testing the climatic variability hypothesis with data on the intestinal length of rodents. *Am. Nat.* **172**.

Nelson, J.S. 2006. *Fishes of the world*. Wiley, Hoboken, NJ

O'Grady, S.P., Morando, M., Avila, L. & Dearing, M.D. 2005. Correlating diet and digestive tract specialization: Examples from the lizard family Liolaemidae. *Zoology* **108**: 201–210.

Olsson, J., Quevedo, M., Colson, C. & Svanbäck, R. 2007. Gut length plasticity in perch: Into the bowels of resource polymorphisms. *Biol. J. Linn. Soc.* **90**: 517–523.

Orr, H.A. 2005. The genetic theory of adaptation: A brief history. *Nat. Rev. Genet.* **6**: 119–127.

Paradis, E. & Schliep, K. 2019. ape 5.0: an environment for modern phylogenetics and evolutionary analyses in R. *Bioinformatics* **35**: 526–528.

Parnell, N.F., Hulsey, C.D. & Streelman, J.T. 2012. The Genetic Basis Of A Complex Functional System. *Evolution.* **66**: 3352–3366.

Pope, E.C., Hays, G.C., Thys, T.M., Doyle, T.K., Sims, D.W., Queiroz, N., *et al.* 2010. The biology and ecology of the ocean sunfish Mola mola: A review of current knowledge and future research perspectives. *Rev. Fish Biol. Fish.* **20**: 471–487.

R Core Team. 2020. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, AUT.

Ribble, D. & Smith, M. 1983. Relative intestine length and feeding ecology of freshwater fishes. *Growth* **47**.

Ricklefs, R.E. 1996. Morphometry of the digestive tracts of some passerine birds. *Condor* **98**: 279–292.

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

Rohlf, F.J. 2018. TPSdig Version 2.32. Ecology & Evolution, SUNY at Stony Brook.

Salzburger, W. 2018. Understanding explosive diversification through cichlid fish genomics. *Nat. Rev. Genet.* **19**: 705–717.

Sanchez, J.L. & Trexler, J.C. 2016. The adaptive evolution of herbivory in freshwater systems. *Ecosphere* **7**.

Schilling, M.P., Mullen, S.P., Kronforst, M., Safran, R.J., Nosil, P., Feder, J.L., *et al.* 2018. Transitions from single-to multi-locus processes during speciation with gene flow. *Genes* **9**.

Schwander, T., Libbrecht, R. & Keller, L. 2014. Supergenes and complex phenotypes. *Curr. Biol.* **24**: R288–R294.

Seehausen, O. 2006. African cichlid fish: a model system in adaptive radiation research. *Proc. R. Soc. B Biol. Sci.* **273**: 1987–1998.

Seehausen, O. 1996. *Lake Victoria Rock Cichlids - taxonomy, ecology, and distribution*. Verduijn Cichlids, Zevenjuizen, NL.

Seehausen, O. 2015. Process and pattern in cichlid radiations - inferences for understanding unusually high rates of evolutionary diversification. *New Phytol.* **207**: 304–312.

Seehausen, O., Lippitsch, E., Bouton, N. & Zwennes, H. 1998a. Mbibi, the rock-dwelling cichlids of Lake Victoria: description of three new genera and fifteen new species (Teleostei). *Ichthyol. Explor. Freshwaters* **9**: 129–228.

Seehausen, O. & Schluter, D. 2004. Male-male competition and nuptial-colour displacement as a diversifying force in Lake Victoria cichlid fishes. *Proc. R. Soc. B Biol. Sci.* **271**: 1345–1353.

Seehausen, O. & van Alphen, J.J.M. 1998. The effect of male coloration on female mate choice in closely related Lake Victoria cichlids (Haplochromis nyererei complex). *Behav. Ecol. Sociobiol.* **42**: 1–8.

Seehausen, O., Witte, F., Alphen, J.J.M. & Bouton, N. 1998b. Direct mate choice maintains diversity among sympatric cichlids in Lake Victoria. *J. Fish Biol.* **53**: 37–55.

Seehausen, van Alphen & Lande. 1999. Color polymorphism and sex ratio distortion in a cichlid fish as an incipient stage in sympatric speciation by sexual selection. *Ecol. Lett.* **2**: 367–378.

Selz, O.M., Lucek, K., Young, K.A. & Seehausen, O. 2014a. Relaxed trait covariance in interspecific cichlid hybrids predicts morphological diversity in adaptive radiations. *J. Evol. Biol.* **27**: 11–24.

Selz, O.M., Pierotti, M.E.R.R., Maan, M.E., Schmid, C. & Seehausen, O. 2014b. Female preference for male color is necessary and sufficient for assortative mating in 2 cichlid sister species. *Behav. Ecol.* **25**: 612–626.

Selz, O.M., Thommen, R., Pierotti, M.E.R., Anaya-Rojas, J.M. & Seehausen, O. 2016. Differences in male coloration are predicted by divergent sexual selection between populations of a cichlid fish. *Proc. R. Soc. B Biol. Sci.* **283**.

Sideleva, V.G. 2003. *The endemic fishes of Lake Baikal*. Backhuys Publishers, Leiden, NL.

Stager, J.C. & Johnson, T.C. 2008. The late Pleistocene desiccation of Lake Victoria and the origin of its endemic biota. *Hydrobiologia* **596**: 5–16.

Stelkens, R.B., Pierotti, M.E.., Joyce, D.A., Smith, A.M., van der Sluijs, I. & Seehausen, O. 2008. Disruptive sexual selection on male nuptial coloration in an experimental hybrid population of cichlid fish. *Philos. Trans. R. Soc. B Biol. Sci.* **363**: 2861–2870.

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

Streelman, J.T. & Albertson, R.C. 2006. Evolution of novelty in the cichlid dentition. *J. Exp. Zool. Part B Mol. Dev. Evol.* **306**: 216–226.

Svardal, H., Quah, F.X., Malinsky, M., Ngatunga, B.P., Miska, E.A., Salzburger, W., *et al.* 2020. Ancestral Hybridization Facilitated Species Diversification in the Lake Malawi Cichlid Fish Adaptive Radiation. *Mol. Biol. Evol.* **37**: 1100–1113.

Svensson, O., Woodhouse, K., van Oosterhout, C., Smith, A., Turner, G.F. & Seehausen, O. 2017. The genetics of mate preferences in hybrids between two young and sympatric Lake Victoria cichlid species. *Proc. R. Soc. B Biol. Sci.* **284**.

Takahashi, T. & Koblmüller, S. 2011. The Adaptive Radiation of Cichlid Fish in Lake Tanganyika: A Morphological Perspective. *Int. J. Evol. Biol.* **2011**: 1–14.

Turner, G.F., Seehausen, O., Knight, M.E., Allender, C.J. & Robinson, R.L. 2001. How many species of cichlids are there in African lakes? *Mol. Ecol.* **10**: 793–806.

Van Doorn, G.S., Edelaar, P. & Weissing, F.J. 2009. On the origin of species by natural and sexual selection. *Science* **326**: 1704–1707.

van Ooijen, J.W. 2006. JoinMap®4, software for the calculation of genetic linkage maps in experimental populations. Kyazma BV, Wageningen.

van Rijssel, J.C., Moser, F.N., Frei, D. & Seehausen, O. 2018. Prevalence of disruptive selection predicts extent of species differentiation in Lake Victoria cichlids. *Proc. R. Soc. B Biol. Sci.* **285**: 20172630.

Wagner, C.E., McIntyre, P.B., Buels, K.S., Gilbert, D.M. & Michel, E. 2009. Diet predicts intestine length in Lake Tanganyika's cichlid fishes. *Funct. Ecol.* **23**: 1122–1131.

Wei, T. & Simko, V. 2017. R package “corrplot”: Visualization of a Correlation Matrix (Version 0.84).

Witte, F. & Van Oijen, M.J.P. 1990. *Taxonomy, ecology and fishery of Lake Victoria haplochromine trophic groups*.

Yeaman, S. & Otto, S.P. 2011. Establishment and maintenance of adaptive genetic divergence under migration, selection, and drift. *Evolution* **65**: 2123–2129.

Yeaman, S. & Whitlock, M.C. 2011. The genetic architecture of adaptation under migration-selection balance. *Evolution* **65**: 1897–1911.

Supplementary Materials

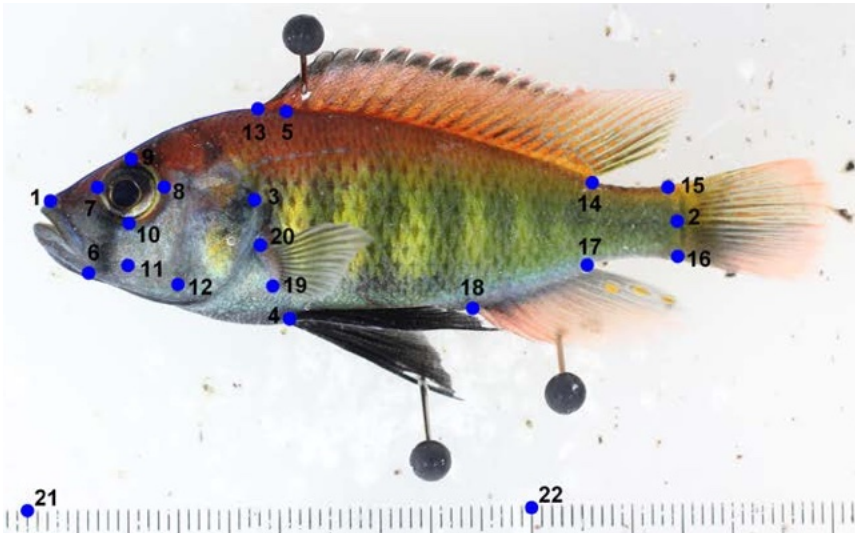


Figure S1 | Landmarks for shape data acquisition. (1) anterior extreme of snout bone; (2) caudal border of hypural plate at the lateral line; (3) posterior end of operculum; (4) insertion point of left pelvic fin; (5) highest point on the body following up a line vertical to body axis from (4); (6) caudal tip of retro-acrticular process; (7) rostral corner of of bony orbit; (8) ventral point of postorbital process; (9) neurocranial border of orbit; (10) orbital margins of the infraorbital; (11) lowest point on muscle following following down a line vertical to body axis from (10); (12) ventral–posterior extreme at the bending point of preoperculum; (13) rostral insertion point of dorsal fin; (14) caudal insertion point of dorsal fin; (15) dorsal insertion point of caudal fin; (16) ventral insertion point of caudal fin; (17) caudal insertion point of anal fin; (18) rostral insertion point of anal fin; (19) ventral insertion point of pectoral fin; (20) dorsal insertion point of pectoral fin; (21) and (22) were placed 5 cm apart on a ruler for scaling. Between (1) and (13) a curve with 8 equally spaced semilandmarks was set.

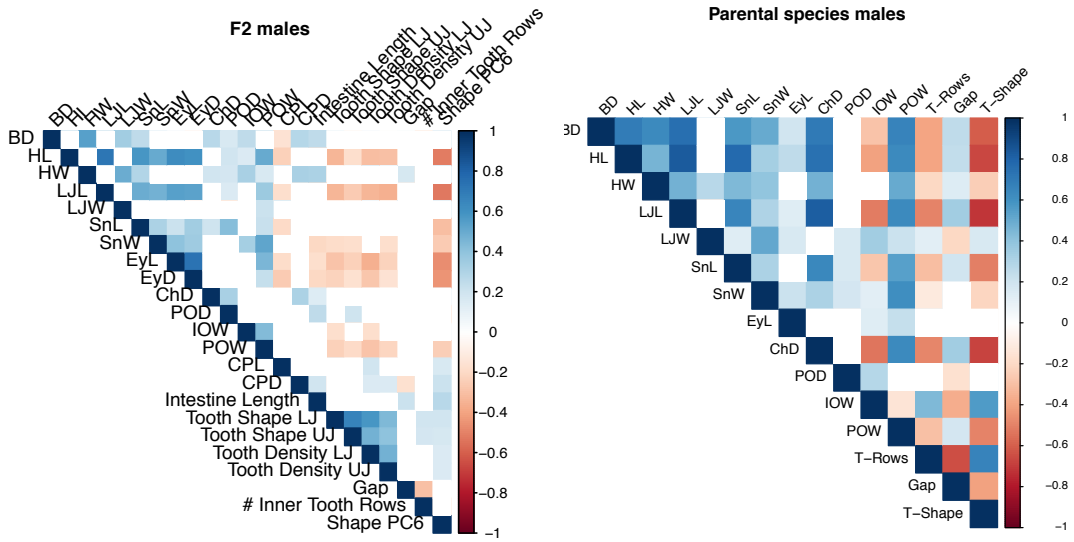


Figure S2 | Trait correlations among males. The correlogram on the left shows correlations between all measured morphological traits in 129 F2 males. (For shape, only PC-axis 6 for which we found a significant QTL is shown). The parental species males' correlogram (on the right) is based on measurements on 98 wild-caught male specimens of either parental species by (van Rijssel *et al.*, 2018); see Methods section for differences in measurements of tooth morphology ("T." stands for "tooth"). Size-correction for this correlogram was performed for both parental species combined. Since the EyL and EyD measurements are highly inter-dependent, we only kept the EyL measurement. Positive correlations are indicated in blue, negative correlations in red color. Blank means no significant correlation ($p > 0.1$). Color intensity in the squares are proportional to the correlation coefficients, as shown in the bar on the right. (See Figure S3 for trait distributions.)

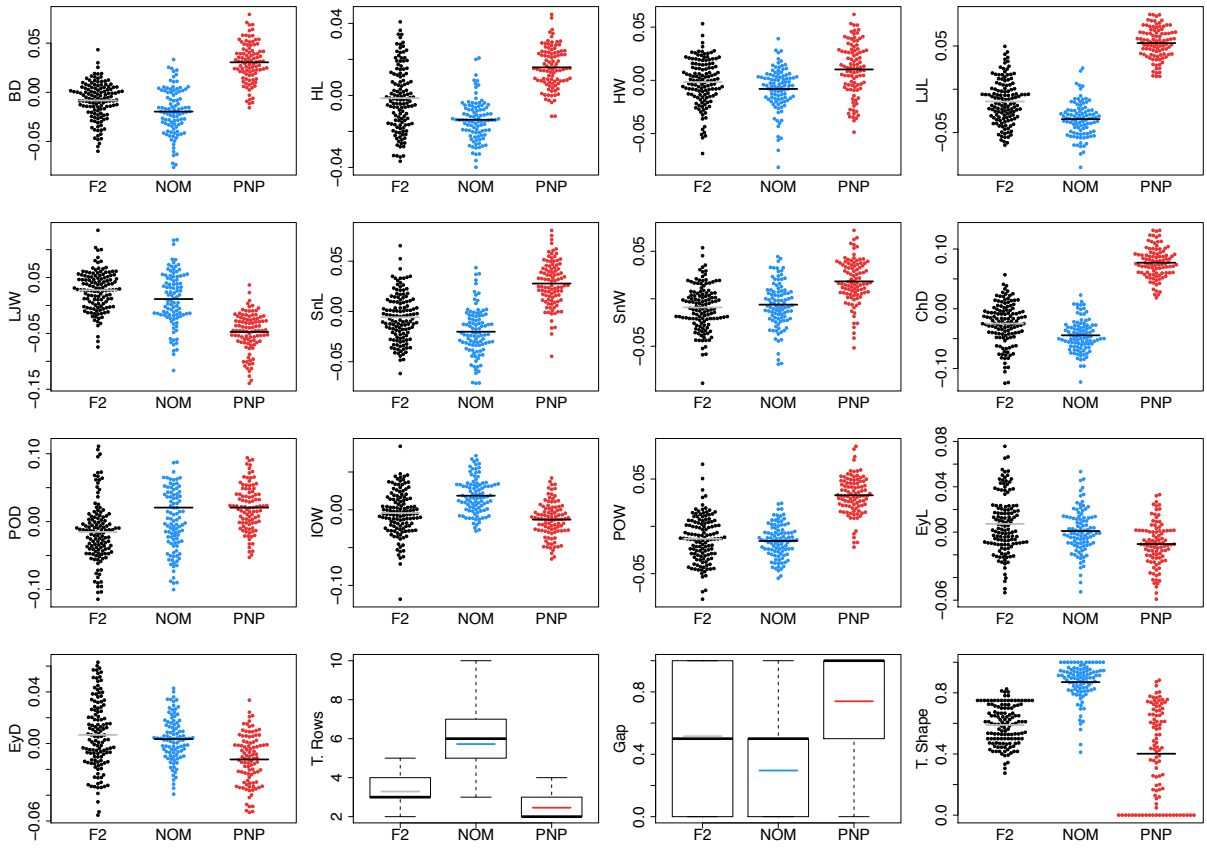


Figure S3 | Trait distributions in F2 males and wild-caught males of both parental species for all measured linear traits and tooth morphology. See Methods for information on size-correction and differences in the tooth measurements "T." stands for "tooth". NOM refers to *N. omnicaeruleus*, PNP to *P.* sp. "nyererei-like".

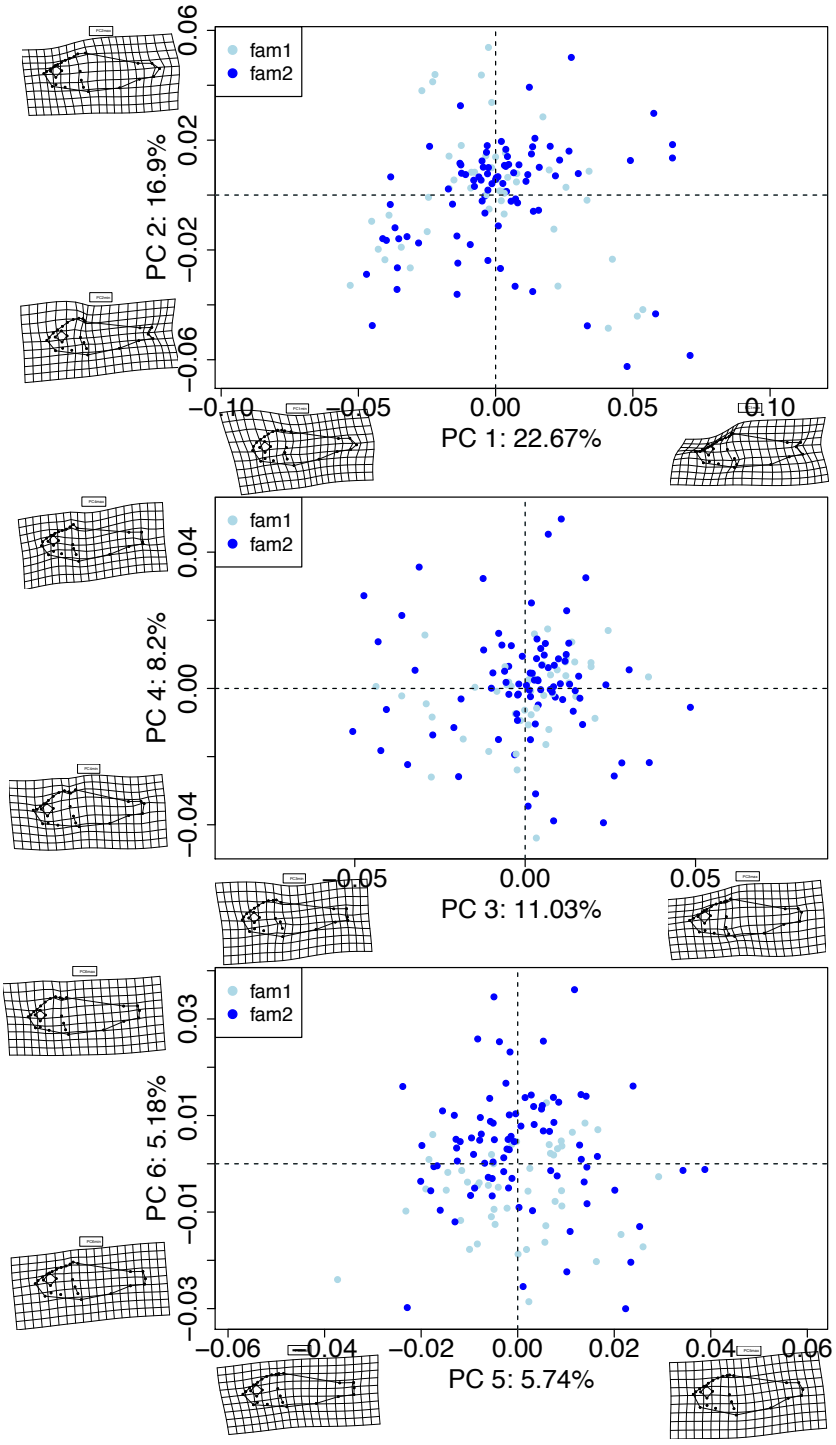


Figure S4 | PCA of shape in F2 males.

Table S1. Results of multiple QTL model comparisons & selection performed with stepwiseqtl. For each morphological trait, the final model with highest penalized LOD score shown. For morphological traits not on this list, no fitting model (i.e. no significant QTL) was found. This approach detected the same (and no additional) QTLs compared to the single QTL scans (see Table 2).

	name	Pun-LG	Ore-LG	pos
Head Length				
pLOD: 0.814	Q1 13@24.7	13	5	24.67
	Q2 15@13.1	15	8-24	13.15
Body Depth				
pLOD: 1.17	Q1 20@32.2	20	4	32.24
Lower Jaw Length				
pLOD: 0.501	Q1 4@13.7	4	16-21	13.65
Pre-Orbital Depth				
pLOD: 0.343	Q1 20@16.0	20	4	16.00
Intestine Length				
pLOD: 1.31	Q1 16@35.0	16	14	35.00
Tooth Density Lower Jaw				
pLOD: 1.08	Q1 22@23.0	22	12	23.00
Tooth Density Upper Jaw				
pLOD: 0.314	Q1 22@43.0	22	12	43.02
Number of Inner Teeth Rows				
pLOD: 1.164	Q1 22@48.0	22	12	48.00

Pun-LG, linkage group numbers corresponding to the anchored *P. nyrerei* reference genome; Ore-LG, linkage group number corresponding to the *O. niloticus* reference genome; pos, position on the linkage map in cM; pLOD, penalized LOD score.

Chapter 3 –

Identification of a novel sex determining chromosome in cichlid fishes that acts as XY or ZW in different lineages

Hydrobiologia
https://doi.org/10.1007/s10750-021-04560-7

ADVANCES IN CICHLID RESEARCH IV

Identification of a novel sex determining chromosome in cichlid fishes that acts as XY or ZW in different lineages

Anna F. Feller · Vera Ogi · Ole Seehausen · Joana I. Meier

Received: 31 March 2020 / Revised: 2 February 2021 / Accepted: 23 February 2021

Abstract Sex determination systems are highly conserved among most vertebrates with genetic sex determination, but can be variable and evolve rapidly in some. Here, we study sex determination in a clade with exceptionally high sex chromosome turnover rates. We identify the sex determining chromosomes in three interspecific crosses of haplochromine cichlid fishes from Lakes Victoria and Malawi. We find evidence for different sex determiners in each cross. In the Malawi cross and one Victoria cross the same chromosome is sex-linked but while females are the heterogametic sex in the Malawi species, males are the heterogametic sex in the Victoria species. This chromosome has not previously been reported to be sex determining in cichlids, increasing the number of different chromosomes shown to be sex determining in cichlids to 12. All Lake Victoria species of our crosses are less than 15,000 years divergent, and we identified different sex determiners among them. Our study provides further evidence for the diversity and evolutionary flexibility of sex determination in cichlids, factors which might contribute to their rapid adaptive radiations.

Keywords Sex determination · Cichlid fishes · Lake Victoria · Lake Malawi

Guest editors: S. Koblmüller, R. C. Albertson, M. J. Genner, K. M. Sefc & T. Takahashi / Advances in Cichlid Research IV: Behavior, Ecology and Evolutionary Biology

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1007/s10750-021-04560-7.

A. F. Feller · V. Ogi · O. Seehausen · J. I. Meier (✉)
Division of Aquatic Ecology & Evolution, Institute of Ecology and Evolution, University of Bern, Baltzerstrasse 6, 3012 Bern, Switzerland
e-mail: jm2276@cam.ac.uk

A. F. Feller · V. Ogi · O. Seehausen · J. I. Meier
Department of Fish Ecology and Evolution, Centre of Ecology, Evolution and Biogeochemistry (CEEB), Eawag Swiss Federal Institute of Aquatic Science and Technology, Seestrasse 79, 6047 Kastanienbaum, Switzerland

J. I. Meier
Department of Zoology, University of Cambridge, Downing Street, Cambridge CB2 3EJ, UK

J. I. Meier
St John's College, University of Cambridge, St John's Street, Cambridge CB2 1TP, UK

Introduction

Most eukaryotes reproduce sexually and individuals are either male producing many small sperm or female producing few large ova (Bell, 1982). One of the most puzzling aspects of sexual reproduction is that while the existence of two sexes is highly conserved, there is a diversity of mechanisms triggering development as either male or female (reviewed in Bachtrog et al., 2014). Among animals with genetic sex determination, genes determining sex have evolved independently many times. Some groups have stable sex chromosomes that remained conserved across millions of years (White, 1977; Ming et al., 2011; Cortez et al., 2014). For example, the mammalian Y chromosome arose approximately 180 million years ago (Cortez et al., 2014), the avian female-determining W chromosome about 140 million years ago (Cortez et al., 2014), and the Lepidopteran W chromosome is more than 180 million years old (Sahara et al., 2012). These ancient sex chromosomes have strongly deteriorated over time and accumulated sexually antagonistic alleles, leading to heteromorphic (morphologically distinct, often degenerated) chromosomes (Bachtrog et al., 2014). In contrast, other taxonomic groups display much faster turnover of sex chromosomes (e.g. frogs Jeffries et al., 2018; fishes Kitano & Peichel, 2012). Only about 10% of fishes have heteromorphic sex chromosomes (Devlin & Nagahama, 2002), and sex determining genes can differ among closely related species or even within a single species (Orzack et al., 1980; Seehausen et al., 1999; Kitano & Peichel 2012; Cheng et al., 2013).

One of the fastest rates of sex determination turnover is found in cichlid fishes (Gammerdinger & Kocher, 2018; Böhne et al., 2019). Many different sex determination systems have been identified involving different chromosomes, including male and female heterogametic systems and polygenic sex determination (Seehausen et al., 1999; Gammerdinger & Kocher 2018; Böhne et al., 2019). In addition, B chromosomes, i.e. accessory chromosomes only found in some individuals of a species, have been suggested to act as female sex determiners in two cichlid species (Yoshida et al., 2011; Clark & Kocher, 2019). The family Cichlidae contains very young species radiations such as those of several hundred species each in Lake Victoria (15,000 years, Bezault et al., 2011; Johnson et al., 2000; McGee et al., 2020) and Lake Malawi (1–5 million years, Genner et al., 2007; Ivory et al., 2016; Malinsky et al., 2018). Even within these young radiations, several different sex determining chromosomes have been identified in the few species whose sex determination systems have been studied to this date. At least five sex determination systems are present in Lake Malawi haplochromine cichlids involving *Oreochromis niloticus* (Linnaeus, 1758) linkage groups (O) 3, 5, 7, and 20 (Gammerdinger & Kocher, 2018; Böhne et al., 2019), and in multiple species, some females have a B chromosome (Clark et al., 2016; Clark & Kocher 2019). Only few species from Lake Victoria have been tested for sex chromosomes. One study identified two QTLs on O2 and O5 in a cross of *Paralabidochromis sauvagei* (Pfeffer, 1896) and *Paralabidochromis chilotes* (Boulenger 1911) (Kudo et al., 2015). A study of a cross between two *Pundamilia* species inferred an XY sex determination system on O23 in a 1.9 Mb region containing the anti-Müllerian hormone gene (*amh*) (Feulner et al., 2018), and a third study discovered the presence of a feminizing B chromosome in *Lithochromis rubripinnis* Seehausen, Lippitsch & Bouton, 1998 (Yoshida et al., 2011).

Theoretical modelling has shown that sex chromosome turnover events can be favoured by deleterious mutation load on the non-recombining chromosome, if the load outweighs the benefits gained by factors favouring maintenance of an initial sex chromosome, including carrying sexually antagonistic genes where males benefit from having one allele, and females another, maintained in linkage disequilibrium (Blaser et al., 2013, 2014). Alternatively, a new sex-determiner can spread if it is physically linked to sexually antagonistic autosomal mutations (van Doorn & Kirkpatrick, 2007; van Doorn & Kirkpatrick, 2010). The invasion of new sex determiners in response to sexually antagonistic selection has been suggested to have occurred in cichlids (Roberts et al., 2009), and this might contribute to rapid and repeated speciation in cichlids (Lande et al., 2001; Kocher, 2004). Finally, novel sex chromosomes can invade if meiotic drive or endoparasites have produced an unequal sex ratio (Kozielska et al., 2010).

Here, we identify the sex determination systems in three crosses of closely related species of a clade with exceptionally fast sex chromosome turnover: haplochromine cichlids. We compare two interspecific crosses from Lake Victoria and one interspecific cross

from Lake Malawi (Fig. 1). Using QTL mapping and the identification of sex differences based on genotype frequencies we aimed to (1) identify the sex determining chromosomes in the three crosses, (2) test if the sex chromosomes show signs of degeneration, (3) test whether they represent male-heterogametic (XY) or female-heterogametic (ZW) systems and (4) trace back the sex determining alleles to the parental species. Our results reveal multiple sex determiners in our crosses, including a new chromosome involved in sex determination that acts as XY or ZW in different species. Overall, our study adds to a growing body of evidence for high flexibility in sex determination in cichlid fishes.

Materials and methods

Experimental crosses

We analysed data from three interspecific second generation (F2) hybrid crosses to infer the sex determination system in each of them. One cross is between the sympatric sister species *Pundamilia* sp. "nyererei-like" and *Pundamilia* sp. "pundamilia-like" ('Victoria1'; see Feller et al., 2020a), one between the non-sympatric species *Pundamilia pundamilia* Seehausen & Bouton, 1998 and *Pundamilia* sp. "red-head" ('Victoria2'; see Feulner et al., 2018; Feller et al., 2020a), and one between distantly related *Astatotilapia calliptera* "Chizumulu" (Günther, 1894; population from Chizumulu island, Konings, 2001) and *Protomelas taeniolatus* (Trewavas, 1935) ('Malawi'; see Stelkens et al., 2009; Selz et al., 2014; Feller et al., 2020b) (Fig. 1).

In both Victoria crosses, the F2 individuals used in our analyses belong to two F1 families (henceforth families A and B). In the Malawi cross the F2 individuals belong to six F1 families, but for most individuals the information to which family they belong was not available. All F2s were reared to an age of at least 1 year in our aquarium system before they were sexed based on colouration and overall morphological appearance, sacrificed (using MS222; 25–50 mg/l for sedation; 300–400 mg/l for euthanization) and fin-clipped. In individuals that showed an inconsistent genetic pattern with phenotypic sex, we additionally inspected the gonads. Furthermore, we performed gonad inspection in ten individuals that were difficult to sex in each cross, which confirmed

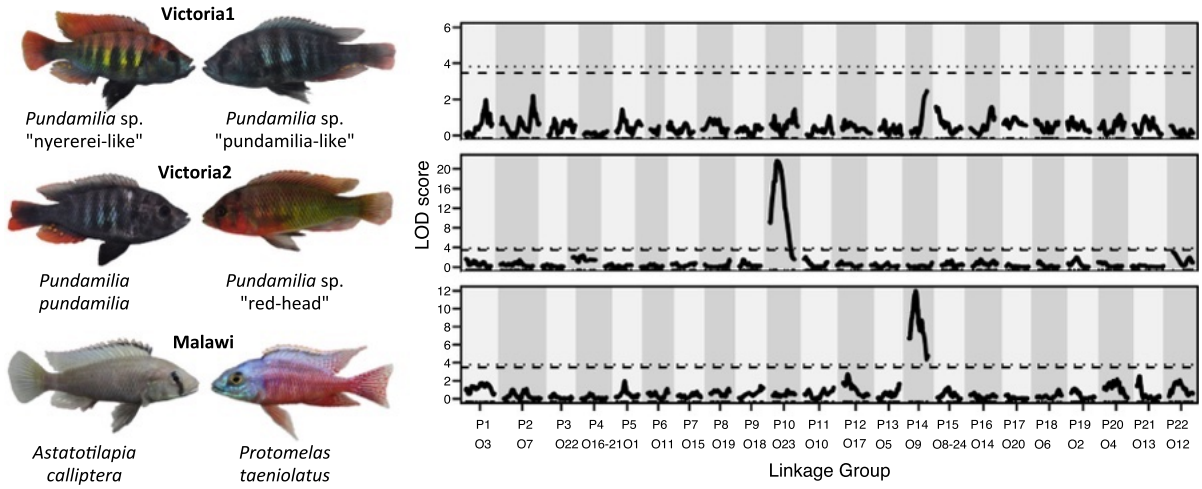


Fig. 1 QTL mapping shows different sex determination in each cichlid cross. On the left, representative male individuals of the six parental species used in the three crosses (parental species used as grandmother in the crossing scheme on the left, parental species used as grandfather on the right). On the right, LOD scores of QTL mapping for each cross separately. The linkage groups (shown with alternating grey background shading) are numbered according to the *Pundamilia nyererei* v2 reference (P, Feulner et al., 2018) and the corresponding *Oreochromis niloticus* reference numbers (O, Brawand et al., 2014) below. The results are from standard interval mapping including all families in a cross (without covariates). The dashed lines represent a genome-wide significance threshold of $P = 0.1$, the dotted lines $P = 0.05$. QTL mapping reveals a clear sex QTL on P10/O23 for the Victoria2 cross and on P14/O9 for the Malawi cross, but does not identify a sex QTL for the Victoria 1 cross (see also Table S1)

that sexing based on external phenotype is reliable. All fish were maintained and bred in a large recirculation facility either at the University of Bern (Victoria1 and Victoria2) or at Eawag (Malawi), with a water temperature of 24–26°C and a 12:12 h light/dark cycle.

We tested if the sex ratio found among all surviving F2 offspring (including those not sequenced) deviates from the expected 0.5 ratio by applying a binomial test with the R function 'binom.test'.

RAD tag sequencing

As described in Feller et al. (2020a) for Victoria1, in Feulner et al. (2018) for Victoria2 and in Feller et al. (2020b) for Malawi, DNA was extracted from fin-clips (stored in 98% ethanol) using phenol–chloroform (Sambrook & Russell, 2001). Restriction-site associated DNA (RAD) sequencing libraries were prepared following Baird et al. (2008) with some modifications. Prior to enzyme digestion with *Sbf*I, DNA concentrations were normalised for all samples in one library. 4–48 individuals carrying custom 5–8 bp barcodes were pooled into one library. This was followed by shearing (on a Covaris M220 focused-ultrasonicator) and size selection of 300–700 bp fragments (on a SageELF machine). Each library was amplified in four 50 ml aliquots reactions, and the size-selection step was repeated with the SageELF machine or magnetic beads. Single end sequencing (100–125 bp) was done on an Illumina HiSeq 2500 platform at the Genomic Technologies Facility of the University of Lausanne or at the Next Generation Sequencing Platform of the University of Bern, using one lane per library. To increase complexity in the first 10 sequenced base pairs, and for base quality recalibration (see below), 4–12.5% PhiX genomic DNA was sequenced together with each library.

Sequence processing

As described in Feller et al. (2020a) for Victoria1 and Victoria2 and in Feller et al. (2020b) for Malawi, PhiX reads were removed with Bowtie2 v2.3.2 (Langmead & Salzberg 2012). Reads were demultiplexed and trimmed to 85–90 bp with process_radtags implemented in stacks v.1.40 (Catchen et al., 2013). Single errors in the barcode were corrected and reads with incomplete restriction sites discarded. This was followed by filtering reads for a minimum quality of 10 at all bases and of 30 in at least 95% of the reads using the FASTQ quality filter (http://hannonlab.cshl.edu/fastx_toolkit/index.html). Bowtie2 v2.3.2 (Langmead & Salzberg 2012) was used for alignment to the anchored version of the reference genome of a male *Pundamilia nyererei* (Witte-Maas & Witte, 1985) (Feulner et al., 2018), allowing one mismatch. Base quality recalibration based on the PhiX reads was done per library using the GATK BaseRecalibrator and PrintReads modules (McKenna et al., 2010). For genotyping, GATK Unified Genotyper v3.7 (McKenna et al., 2010) was used (minimum base quality score set to 20). In Victora1 and Victoria2 only uniquely aligned reads were used.

The resulting vcf files were filtered using bcftools implemented in samtools v.1.8 and v1.9 (Li et al., 2009) and using vcftools v.0.1.14 and v.0.1.16 (Danecek et al., 2011) as described in Feller et al. (2020a) for both Victoria crosses and also applied to the Malawi cross: Sites with $> 50\%$ missing data and individuals with a mean depth of < 12 or $> 50\%$ missing data were excluded. Only bi-allelic SNPs with a mean sequencing depth of less than 1.5 times the interquartile range from the mean were kept and sites within 10 bp of indels were removed. Genotypes with a depth of < 10 were set to missing. Individuals were excluded if the heterozygous read balance was heavily skewed (indicating PCR duplicates). Sites were then again filtered for no more than 50% missing data and for a minor allele frequency of at least 0.05.

This resulted in a dataset with 10,598 SNPs and 224 individuals (218 F2s, 4 F1, 2 F0) for Victoria1, in 9990 SNPs and 192 individuals (186 F2s, 4 F1, 2 F0) for Victoria2, and in 12,187 SNPs and 126 individuals (114 F2s, 12 parental species individuals) for Malawi.

Linkage map construction

As described in Feller et al. (2020a) for both Victoria crosses and also applied to the Malawi cross, linkage maps for all three crosses were constructed in JoinMap 4.0 (Van Ooijen, 2006) after applying an allelic balance correction (https://github.com/joanam/scripts/allelicBalance.py) to the grandparents (F0) and then subsetting the datasets to SNPs that are fixed for alternative alleles (homozygous alternative) between the F0. In Victoria1, 954 SNPs were homozygous in the F0 grandmother and heterozygous

in all 4 parents (F1s) (no missing data allowed). This set differs somewhat from the published map in Feller et al. (2020a) because further analyses (see below) revealed inconsistencies in the F0 grandfather's genotype, and we thus only used the F0 grandmother and the F1s to filter for putative homozygous alternative SNPs, and we applied a more stringent segregation distortion filter ($P < 0.01$) during linkage map construction. In Victoria2, 2358 SNPs were homozygous alternative in the F0 and heterozygous in 2 F1s (no missing data allowed; the other 2 F1s were removed in the filtering process). In the Malawi cross, fin-clips and sequences for the F0 and F1 of this cross were not available. Instead, as described in Feller et al. (2020b), 5–6 individuals of each parental species were sequenced (*A. calliptera* "Chizumulu": 3 males, 2 females, 1 undeterminable (all from our lab population); *P. taeniolatus*: 4 males (two from our lab population and two from two aquarium fish breeders), 1 female (from one of the aquarium fish breeders)). 1775 SNPs were retained as homozygous alternative between the two parental species.

In linkage map construction, SNPs with extreme segregation distortion ($P < 0.001$) or with > 20% missing genotypes, were excluded. Identical SNPs were removed (i.e. SNPs within the same RAD locus; > 0.950). Individuals with > 30% missing data were excluded. The Victoria1 linkage map was generated from 216 F2 individuals (173 males, 43 females), the Victoria2 linkage map from 171 F2 individuals (115 males, 56 females), the Malawi linkage map from 108 F2 individuals (34 males, 66 females, 8 juveniles with undetermined sex). Linkage groups were identified based on a LOD threshold of 5–6, excluding loci with a recombination frequency of above 0.6. The strongest cross-link (SCL) values in the maps are 4.6–5.8. The Kosambi regression mapping algorithm was used to build the linkage maps (LOD threshold 1.0, recombination threshold 0.499, goodness-of-fit threshold 5.0, no fixed order). Two rounds of mapping were performed with a ripple after addition of each marker to the map.

QTL mapping

QTL mapping was performed in R/qtl (Broman et al., 2003). The calc.genoprob function was used to calculate conditional genotypes with a fixed step-size of 1 cM, an assumed genotyping error rate of 0.05, and the Kosambi map function. We performed standard interval mapping with the binary model, and determined significance thresholds by ($n = 1000$) permutations. Bayesian credible intervals (95%) were calculated with the 'bayesint' function and percentage of variance explained (PVE) was calculated as $1 - 10^{-2*LOD/n}$ (Broman & Sen, 2009), where LOD is the highest LOD score, and n the number of individuals.

In both Victoria crosses, we additionally performed the analysis with family as covariate, and for each family separately. Where a significant QTL for sex was found, we repeated the mapping excluding the linkage group containing the QTL to test for additional QTLs.

The number of F2 individuals used in QTL mapping (i.e. with phenotypic and genotypic data) are the following: 217 in Victoria1 (172 males (*138 family A + 34 family B*) and 45 females (*16 family A + 29 family B*)), 186 in Victoria2 (130 males (*76 family A + 45 family B + 9 family unknown*) and 56 females (*47 family A + 9 family B*)), 105 in Malawi (36 males and 69 females).

For markers not mapped to chromosomes, the QTL mapping approach provides only cM positions and LOD scores. For comparison with other statistics, we thus had to infer the bp positions for these markers. We fitted a cubic smoothing spline on cM and bp positions for markers mapped to chromosomes with the smooth.spline R function (stats R package v. 3.6.1) with the smoothing parameter (spar) set to 0.8 and predicted bp positions for markers not mapped to chromosomes with the R function predict.smooth.spline.

Assessing sequencing depth differences between F2 males and F2 females

If the sex chromosomes were heteromorphic due to degeneration of the Y or W chromosome, we would expect differences in sequencing depth between males and females at the sex chromosome. That is, parts of the Y or W chromosome that diverged strongly may no longer map to the corresponding chromosome in the reference genome and some parts may be missing due to deletions in the Y or W chromosome, and this should result in lower sequencing depth in the heterogametic sex (males if XY or females if WZ system). To test this we computed mean sequencing depth in F2 males and females separately with vcftools

v 0.1.15 (Danecek et al., 2011). As sequencing depth is quite variable in general, we visualized sex differences in sequencing depth by averaging across 5 Mb windows with the function 'winScan' in the R package windowscanr (https://github.com/tavareshugo/WindowScanR).

Identifying sex chromosomes with differences in genotypes frequencies between F2 males and F2 females

Next, we computed genotype frequencies to identify the sex determining chromosomes. In order to reduce missing data and correct potential genotyping errors, we phased the dataset with Beagle v. 5.1 (Browning & Browning, 2011). Given that we did not find sex differences in sequencing depth, we assume that reads of both sex chromosomes (X + Y or W + Z) align to the reference genome. We can thus use genotype frequencies to infer sex linkage. Patterns of genotype frequencies are expected to differ between the sexes at the sex determining region. At sites where both F1 parents are heterozygous, both homozygous genotypes should be present at a frequency of 25% in both sexes on autosomes, whereas at sex determining chromosomes, the grandpaternal homozygous genotype should only be found in males and the grandmaternal one only in females (Fig. 2a). Therefore, comparing the proportion of a specific homozygous genotype in male and female offspring should be informative for identifying sex chromosomes. For each cross, we computed the genotype frequencies for the F0 (grandparents), F1 (parents) and the F2 (offspring) separately using vcftools with the option *hardy*. We then extracted bi-allelic sites where the F1 were heterozygous and computed the frequency of the grandparental homozygote genotype among F2 males and females (see Fig. 2a). For the Victoria1 cross, we extracted sites where all four F1s were heterozygous. As we did not have the genotypes of the F1s for the Malawi cross and only low quality F1 genotypes in the Victoria2 cross, we determined sites that are likely heterozygous in both parents by filtering for an allele frequency of 0.45 to 0.55 averaged across both sexes. To avoid including paralogous regions, i.e. reads from duplicated genomic regions that map to a single region on the reference genome, in each cross we excluded all sites with heterozygosity above 75% in the F2 offspring. Because the families of Victoria1 differed both in sex ratio and in the presence of a QTL, we computed the genotypes frequencies of F2 males and females for each family separately.

Single marker regressions

In the Victoria1 cross, we only found evidence for a sex determining region in the genotype frequency analysis (and this was much weaker in family A than in family B), but not in the QTL analyses. One reason for this could be that a large part of P14 (i.e. the second half) is missing in the linkage map due to segregation distortion. Hence, we performed additional single marker regression analyses to screen for associations of markers not included in the linkage map with sex in the two families. For this, we additionally filtered the 10,598 SNPs for HWE proportions with $P > 0.1$ (only within F2s) using bcftools (Li et al., 2009), resulting in 4695 SNPs over the 22 chromosomes. For each family separately, we then performed an ANOVA for each marker in R, and from this calculated LOD scores as $(n/2)*(\log10(F(\mathrm{df}/(n - \mathrm{df} - 1)) + 1)$, where n = number of individuals and df = degrees of freedom (Broman & Sen, 2009) and PVE (see above). This included 4090 markers (of which 590 on scaffolds) in family A (138 males and 16 females), and 4400 markers (of which 620 on scaffolds) in family B (34 males and 29 females). *P*-values of the ANOVAs were corrected for multiple testing using the Bonferroni method.

The single marker regressions were also run for Victoria2 and Malawi to show that the method works and to check for the presence of additional sex determining regions. In Victoria2, this included 5506 markers (of which 647 on unmapped scaffolds) in family A (74 males, 49 females) and 5531 markers (of which 648 on unmapped scaffolds) in family B (46 males, 8 females). In Malawi this included 9311 markers (of which 1567 on unmapped scaffolds) (35 males, 70 females, no family information available).

Assessing heterozygote frequency differences between males and females to infer the heterogametic sex

Next, we wanted to infer whether a putative sex chromosome is male or female heterogametic, i.e. XY or ZW. We would expect that the Y is more divergent from the X than the X chromosomes are from each

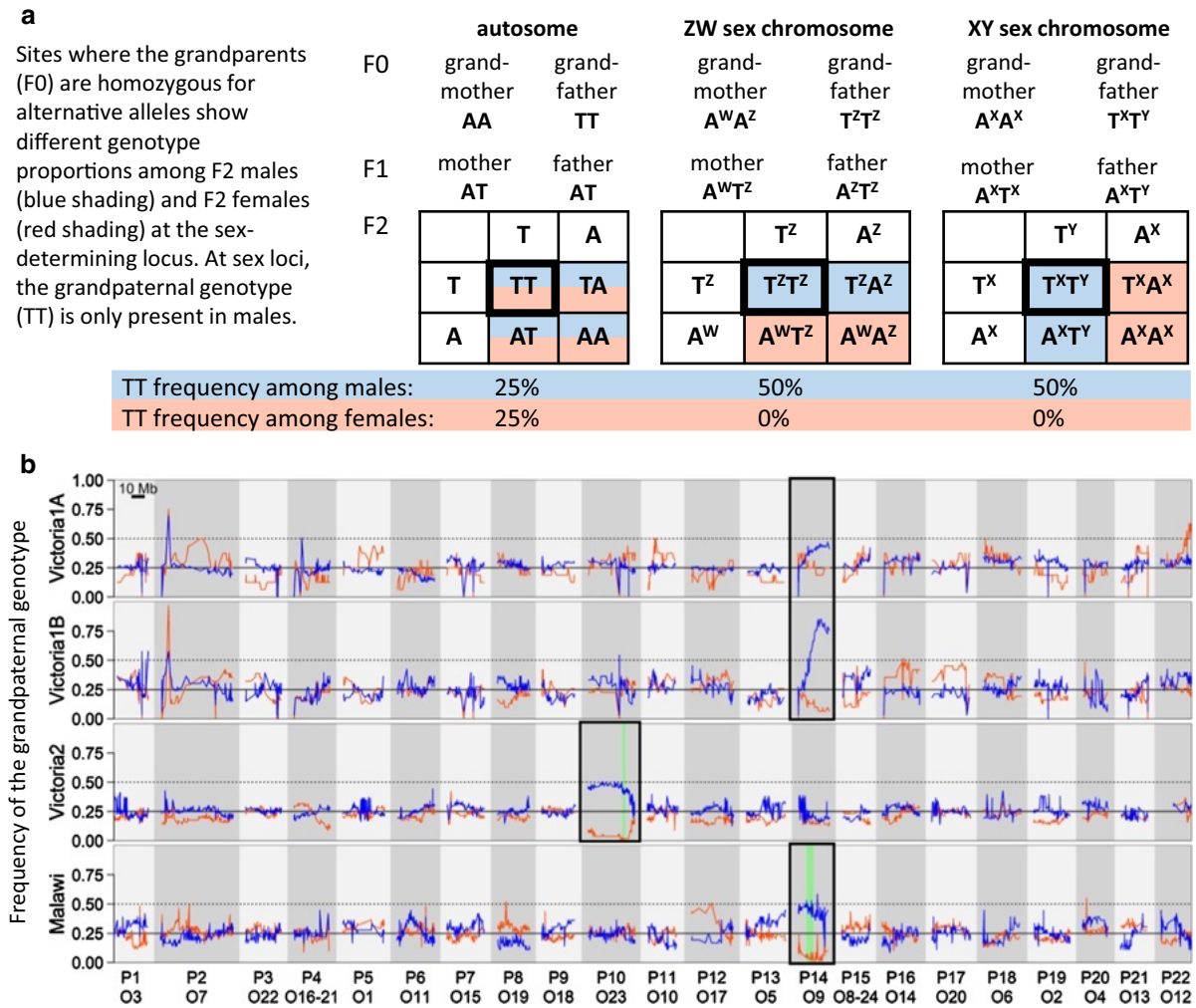


Fig. 2 Genotype frequencies support the sex chromosomes identified with QTL mapping and reveal a sex determiner for Victoria1. **a** Punnet squares showing expected genotype frequencies of F2 males (blue background) and F2 females (red background) at sites where the F0 grandparents are homozygous for alternative alleles. For autosomes, the expected frequency of either homozygous genotype is 25% for both sexes. For sex chromosomes, the grandpaternal genotype (TT) should have a frequency of 50% for male offspring and 0% for female offspring. **b** Frequencies of the grandpaternal genotype among F2 males (blue) and females (red) at sites where the F0 grandparents are homozygous for alternative alleles. Different linkage groups are shown with alternating grey background shading and are numbered according to the *Pundamilia nyererei* reference (P) and the *Oreochromis niloticus* reference (O) below. Chromosomes consistent with sex-determination are framed in black. Green vertical bars indicate the location of the 95% Bayesian confidence intervals obtained in QTL mapping analyses. For Victoria1, the two families are shown separately as they differ in their sex determination patterns. Note that among males of Victoria1B, the grandfather genotype exceeds the expected 50% frequency. This is because heterozygotes with one Z from the grandmother and one Z from the grandfather are strongly underrepresented (see Fig. S3)

other and likewise, the W should be more divergent from the Z than the Z chromosomes are from each other. Given that there is no difference in sequencing depth between the sexes on any chromosome, the heterogametic sex is expected to show a higher heterozygote frequency (i.e. proportion of individuals that are heterozygous) at the sex chromosome than the homogametic sex. This is because any difference between X and Y should increase heterozygote frequency in males compared to females which have two more similar X chromosomes. Similarly, any difference between Z and W should increase heterozygote frequency in females Therefore, higher heterozygote frequency in males compared to females should

indicate an XY system, whereas higher heterozygote frequency in females should indicate a ZW system. We computed heterozygote frequency for male and female offspring separately using the genotype frequencies computed above.

To identify significant sex differences in heterozygosity, we performed 10,000 permutations of the sex assignments and reran the heterozygote frequency calculations for each permutated dataset at all sites. This approach accounts for the exact number of samples available for each sex and also accounts for differences in the total number of heterozygotes among sites. We considered a locus to show a significant sex difference in heterozygote frequency if at most one of the 10,000 permutations were as extreme or more extreme than the observed value (one-sided empirical P value < 0.0001).

Tracing sex determining alleles back to the parental species

In order to identify which species in a cross contributed the sex determining allele(s) in genomic regions identified by the methods outlined above, we inspected genotypes in the F2 individuals at sites that are heterozygous in the heterogametic sex (i.e. sites with Y- or W-linked alleles). We extracted bi-allelic sites where the F0 and F1 individuals of the putative heterogametic sex were heterozygous and the F0 and F1 individuals of the putative homogametic sex were homozygous. F2 individuals of the heterogametic sex are expected to be heterozygous at these sites, thus indicating the presence of the W or Y-linked alleles in heterozygous F2 individuals. For the Malawi cross, we had no sequence data of the grandparents and parents, but RAD data of conspecific individuals of the grandparents. Therefore, we extracted sites where the male individuals (the putatively heterogametic sex) of the species used as grandfather, *P. taeniolatus,* were heterozygous and those of the species used as grandmother, *A. calliptera*, were homozygous. For all crosses, we visualized the genotypes for each F2 individual at these sex-linked sites in yellow for homozygotes of the allele derived from the grandfather (Z or Y), orange for heterozygotes (WZ or XY) and red for homozygotes for the allele derived from the grandmother (W or X). To aid visualization, genotypes likely representing genotype errors (e.g. caused by allelic dropout) were replaced by the adjacent genotypes. For this, runs of up to three equal genotypes that differed from the five genotypes on both sides were replaced by the adjacent genotypes.

This analysis revealed that in both families of the Victoria1 cross some males had a W allele and some females did not have a W allele, and in the Malawi cross some females had a Y allele. Therefore, we assessed if the parental lineage origin of the Z or X alleles, respectively, played an additional role in sex determination. For this analysis, we extracted the most strongly sex-linked site with homozygous F0 and heterozygous F1 individuals and the most strongly sex-linked site with heterozygous parents and grandparents of the heterogametic sex. The combination of the genotypes at these two sites allowed us to trace back each of the alleles to the grandparents and to determine the combination of X/Y or Z/W alleles of each individual.

Knowing the combination of sex chromosomes of each individual, we then computed the divergence between the different sex chromosomes as heterozygosity, i.e. the proportion of sites that differ between the two chromosomes among all positions sequenced, for individuals with different combinations of sex chromosomes. For this, we generated a vcf file with all individuals of all three crosses together, calling both monomorphic and polymorphic sites. We removed sites with more than 25% missing data, genotypes with less than 10 reads and SNPs with only one or two allele copy counts (likely sequencing errors). Individuals with more than 25% missing data proportion or recombinants at the sex determination region were removed. We used the Python script by Simon Martin (https://github.com/simonhmartin/genomics_general/popgenWindows.py) to compute heterozygosity for each individual and divergence (d_{xy}) between the Victoria 1 and Victoria 2 cross individuals in windows of 1 Mbp.

Results

Sex ratios among F2 offspring

Of all the Victoria1 (sympatric sister species) F2 offspring that survived to adulthood 37% were females in family A (85 females, 144 males), and 78% were females in family B (114 females, 32 males). This sex ratio strongly differs from the expected 50% with

binomial test P-values of < 0.001 for both families. In Victoria2 (non-sympatric species), both families had even sex ratios (91 males and 100 females in family A, 48 males and 34 females in family B; binomial test $P > 0.15$ in both families). There is a female-bias in the Malawi cross (51 males, 102 females; binomial test $P < 0.001$).

QTL mapping results

We found no significant QTLs for sex in the Victoria1 cross, neither in the analyses including both families with or without accounting for an effect of family by adding it as covariate in the analyses, nor in the analyses for both families separately (Fig. 1).

The presence of a significant QTL for sex on linkage group P10/O23 in the Victoria2 cross previously reported by Feulner et al. (2018) is also supported by our analyses with a slightly modified linkage map and with a different number of individuals (Fig. 1, Table S1; $P < 0.001$, LOD = 21.58, PVE = 41.39), and also when addtionally accounting for an effect of family by adding it as covariate in the analyses (Table S1). The results were also very similar when the mapping was performed for each family separately (Table S1). The location of the QTL is slightly shifted and the association is less strong in the second family, but considering the small number of individuals, these differences may be due to sampling variance. Repeating the QTL mapping excluding P10/O23 revealed no further QTLs.

In the Malawi cross we found a significant QTL for sex on P14/O9 (Fig. 1, Table S1; $P < 0.001$, LOD = 11.96, PVE = 40.84). The Bayesian confidence interval of 7.27 cM covers three markers and spans a region of 5.76 Gb. Pedigree information was not available for this cross. Repeating the QTL mapping excluding P14/O9 revealed no further QTLs.

Sequencing depth differences between males and females

We did not find strong differences in sequencing depth on any chromosomes in any cross (Fig. S1). The ratio of sequencing depth of females divided by males of all 5 Mb window averages range from 0.93–1.10, 0.92–1.09, 0.90–1.08, and 0.92–1.12 in the Victoria1A, Victoria 1B, Victoria2 and Malawi cross, respectively. The 5 Mb window averages at putative sex chromosomes are also very close to 1 (Victoria1A: 0.998–1.086, Victoria1B: 0.959–1.091, Victoria2: 0.954–1.047, Malawi: 0.953–1.046). This indicates that X and Y reads or W and Z reads map to the reference genome, and thus that any potential sex determining chromosomes are not (yet) heteromorphic due to degeneration.

Difference in genotype frequencies between F2 males and F2 females support putative sex chromosomes

In each cross, the frequency of the grandpaternal homozygous genotype shows clear frequency differences between the sexes on a single chromosome, indicating that this chromosome carries a sex determiner (Fig. 2b). The chromosomes P10/O23 and P14/O9 which are identified as sex determining in the Victoria2 and Malawi cross, respectively, are supported by grandpaternal genotype frequencies of close to 50% in males and 0% in females. In the Victoria1 cross, the two families differ in the genotype frequencies. Both families show sex differences in genotype proportions on chromosome P14/O9 but the difference is much stronger in family B. In family B, 28 of the 33 males (85%) have the grandpaternal genotype exceeding the expected 50% frequency (Fig. S3).

Additional single marker regressions confirm the observed patterns

In single-marker regressions performed for each Victoria1 family independently, no marker was associated with sex in family A after correcting for multiple testing (Fig. S2). In family B, several markers on P14/O9 were significantly associated with sex after correcting for multiple testing. The markers that are associated with sex on P14/O9 are located at the end of P14 (between chr14_12466922 - chr14_25950108, which is indeed the region missing markers in the linkage map). The same analyses in Victoria2 and Malawi show the same clear peaks already seen in the standard QTL mapping analyses (compare Fig. 1 with Fig. S2). The two families of Victoria2 show similar sex-associations. The few additional significant markers on P22/O12 in Malawi may be artefacts that could be due to the fact that several families were combined, since we have no other evidence of sex linkage of this chromosome in this cross.

Differences in heterozygote frequencies between the sexes support the putative sex chromosomes and reveal which sex is heterogamous in each cross

The chromosomes previously identified as carrying a sex determiner showed many sites with sex differences in heterozygote frequency (Fig. 3). Family B of the Victoria1 cross showed a higher heterozygote frequency in females than in males, whereas this effect was weaker and not significant in family A. The Victoria2 and Malawi crosses showed a higher heterozygote frequency in males than in females (Fig. 3). We thus conclude that the Victoria1 cross is female heterogametic (ZW) and the Victoria2 and Malawi crosses are male heterogametic (XY) (Figs. 3, 4).

Tracing back the dominant sex determiners to the parental species

In family B of the Victoria1 cross, 24 of the 31 females (77.4%) carry the W allele derived from *P.* sp. "nyererei-like" (W_{NYL}) and only 3 of 33 males (0.09%) carry a W_{NYL} allele (Figs. 5a, S3, S5). Of the 30 ZZ males, 28 (93.3%) carry two Z alleles derived from the *P.* sp. "pundamilia-like" grandfather (Z_{PUL}-Z_{PUL}), and only two males carry one Z allele from the grandfather and one from the grandmother (Z_{NYL}-Z_{PUL}). This difference is not apparent among ZZ females (five $Z_{NYL}Z_{PUL}$, two $Z_{PUL}Z_{PUL}$). Individuals with $Z_{NYL}Z_{PUL}$ are generally underrepresented (12% of all individuals, sex-ratio corrected). It is thus possible that males, but not females, with two different Z alleles ($Z_{NYL}Z_{PUL}$) are less viable (Fig. S3).

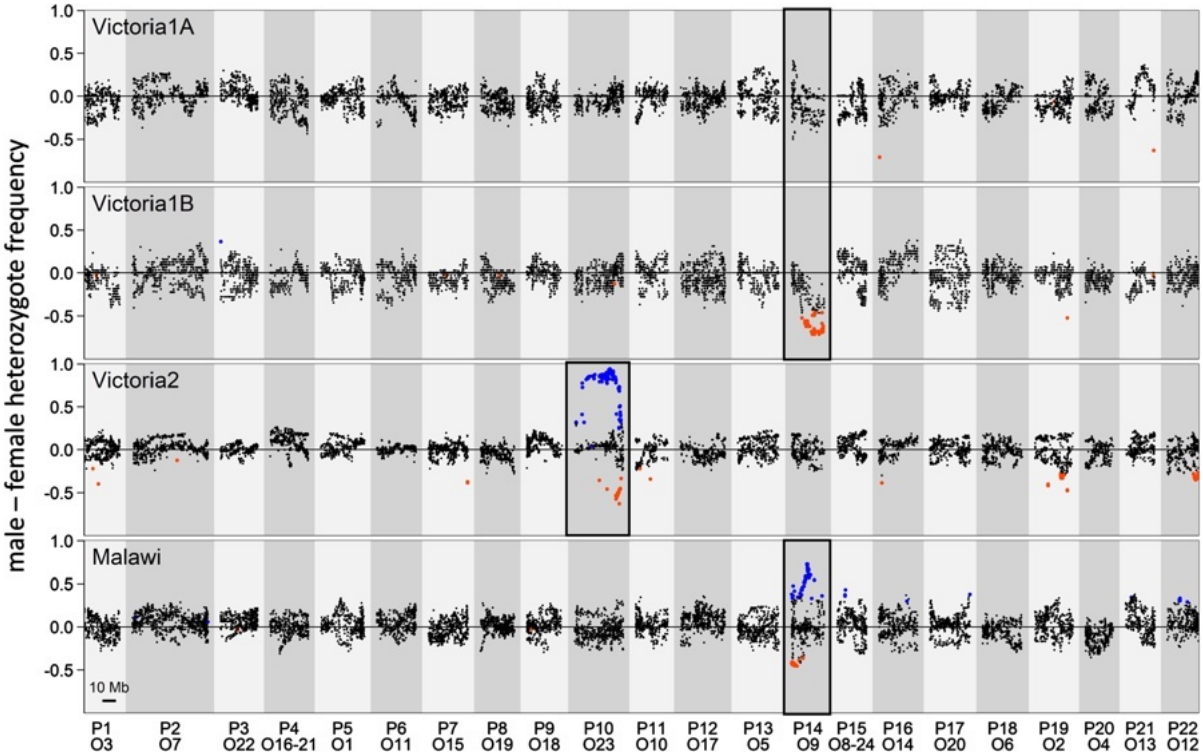
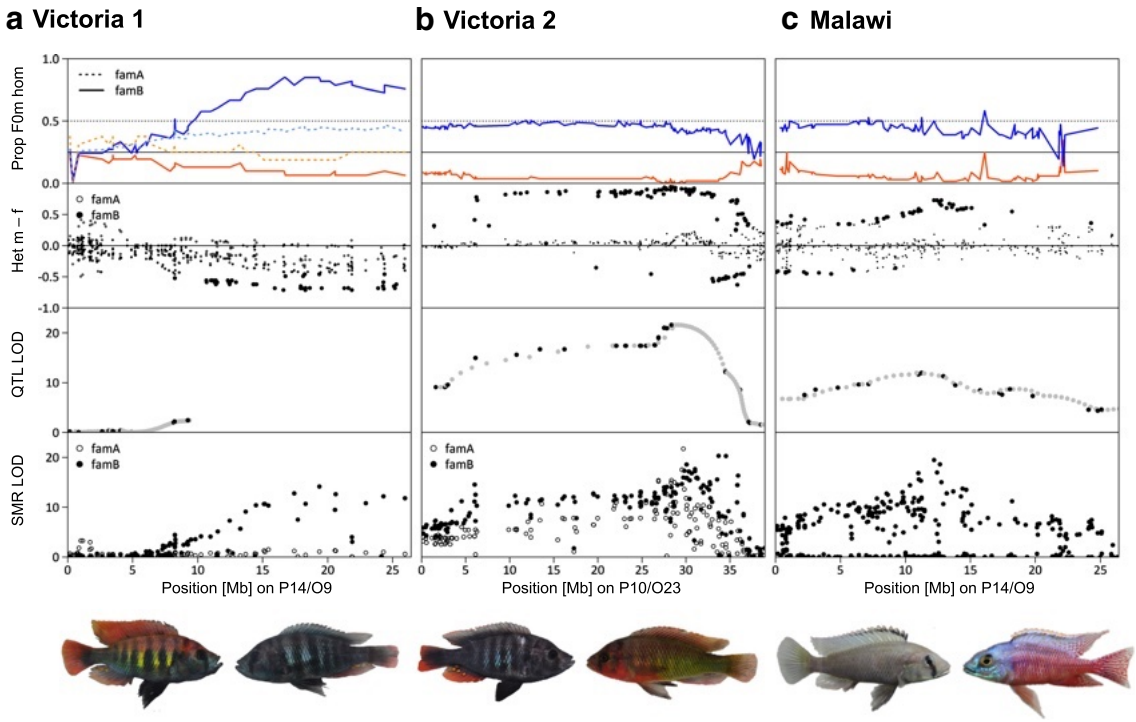


Fig. 3 Sex differences in heterozygote frequency support the previously identified sex determining regions and reveal male and female heterogamety. Proportion of heterozygotes among male minus the proportion of heterozygotes among female F2 individuals shows a clear deviation from 0 at putative sex chromosomes (framed in black). Deviations from 0 greater than at least 1/10,000 permutations (empirical P-value < 0.0001) are shown with larger symbols. Significantly negative heterozygote frequency differences are highlighted in red and indicate that more females than males are heterozygous at these sites. Significantly positive heterozygote frequency differences are highlighted in blue. Chromosomes inferred to be sex-determining in previous analyses are highlighted with a black frame. The heterozygosity frequency differences suggest female heterogamety in the Victoria1 cross and male heterogamety in the Victoria2 and in the Malawi cross



species	*Pundamilia* sp. "nyererei-like"	*Pundamilia* sp. "pundamilia-like"	*Pundamilia pundamilia*	*Pundamilia* sp. "red-head"	*Astatotilapia calliptera*	*Protomelas taeniolatus*
sex-ratio	famA: 1.7x male-biased famB: 2.8x female-biased		even in both families		2x female-biased	
system	ZW P14/O9		XY P10/O23		XY P14/09	
dominant SD	W_{NYL}			Y_{RED}		Y_{TAE}
divergence	W_{NYL} vs Z_{PUL}: 0.24% W_{NYL} vs Z_{NYL}: 0.18% Z_{NYL} vs Z_{PUL}: 0.12%		X_{PUN} vs Y_{RED}: 0.31% X_{RED} vs Y_{RED}: 0.26% X_{PUN} vs X_{RED}: 0.20%		X_{CAL} vs Y_{TAE}: 0.22% X_{TAE} vs Y_{TAE}: 0.11% X_{CAL} vs X_{TAE}: 0.15%	
additional SDs	several		none		female modifier	

Fig. 4 Putative sex chromosomes are supported by multiple lines of evidence. Three lines of evidence support the location of a sex determiner on the putative sex chromosomes in the three crosses: top: Proportion of individuals that are homozygous for the grandfather-derived allele (F0m hom) are shown for females (red) and males (blue). Second row: Sex difference in the proportion of heterozygotes indicates female heterogamety (ZW) in Victoria1 and male heterogamety (XY) in the two other crosses. Sex differences in heterozygote proportions with empirical P-value < 0.0001 are highlighted with larger points. Third row: LOD scores from QTL mapping analyses. Markers on unmapped scaffolds and interpolated markers (where genotypes are inferred as implemented in the calc.genoprob function in R/qtl (Broman & Sen, 2009)) are shown in grey with bp positions predicted from local recombination rates. Note that the Victoria1 QTL LOD scores could not be computed for a large part of the chromosome as the markers in the second half were filtered out due to segregation distortion. This is likely due to the strongly skewed sex ratios in this cross. Bottom: LOD scores of single marker regressions (smr) performed for each family separately. The table below gives an overview of the sex determining systems and sex determiners (SD) found in each cross, and species, respectively. For the divergence estimates between the sex chromosomes see also Fig. S6

In family A of the Victoria1 cross, the W allele derived from *P*. sp. "nyererei-like" (W_{NYL}) is also more common among females than among males but the sex association is much weaker (Fig. 5a). There is likely more than one additional sex determiner or modifier segregating in this family, but with only 16 females (and 138 males), we lack power to identify them. The lack of individuals with two Z chromosomes from different parental species (Z_{NYL}-Z_{PUL}) is not observed in family A. However, it is possible that family A and family B inherited different Z chromosomes from their *P*. sp. "pundamilia-like" grandfather and that only one of them interacts negatively with Z_{NYL}.

In the Victoria2 cross, there are only two of 186 individuals where the phenotypic sex assignment did

not match the genotype at these sites (Fig. 5b). We can thus conclude that there is likely only a single sex determiner on chromosome P10/O23 with a dominant Y-allele derived from the *P.* sp. "red-head" grandfather (Y_{RED}).

In the Malawi cross, all males carry a Y-allele at chromosome P14/O9 derived from the *P. taeniolatus* grandfather (Y_{TAE}; Figs. 5c, S3, S4). Three males even carry two Y-alleles ($Y_{TAE}Y_{TAE}$), indicating that their (F1) mother also had a Y-allele. However, 14 of the 68 females (20%) also carry a Y_{TAE}-allele. The presence of ovaries in these individuals has been confirmed with gonad inspection. Thirteen of these females combine the *P. taeniolatus*-derived Y_{TAE} with an X_{CAL} allele derived from the *A. calliptera* grandmother ($X_{CAL}Y_{TAE}$) and only one has the grandpaternal combination of $X_{TAE}Y_{TAE}$, which could indicate a feminzing effect of the X_{CAL} allele (Figs. S3, sS4). However, many males have the same $X_{CAL}Y_{TAE}$ allele combination. We must therefore assume the presence of an additional female modifier, likely derived from *A. calliptera*, interacting with X_{CAL}. An additional female modifier is also consistent with the finding of a female-biased sex ratio in this cross.

The divergence between the sex chromosomes is greater in heterogametic than in homogametic

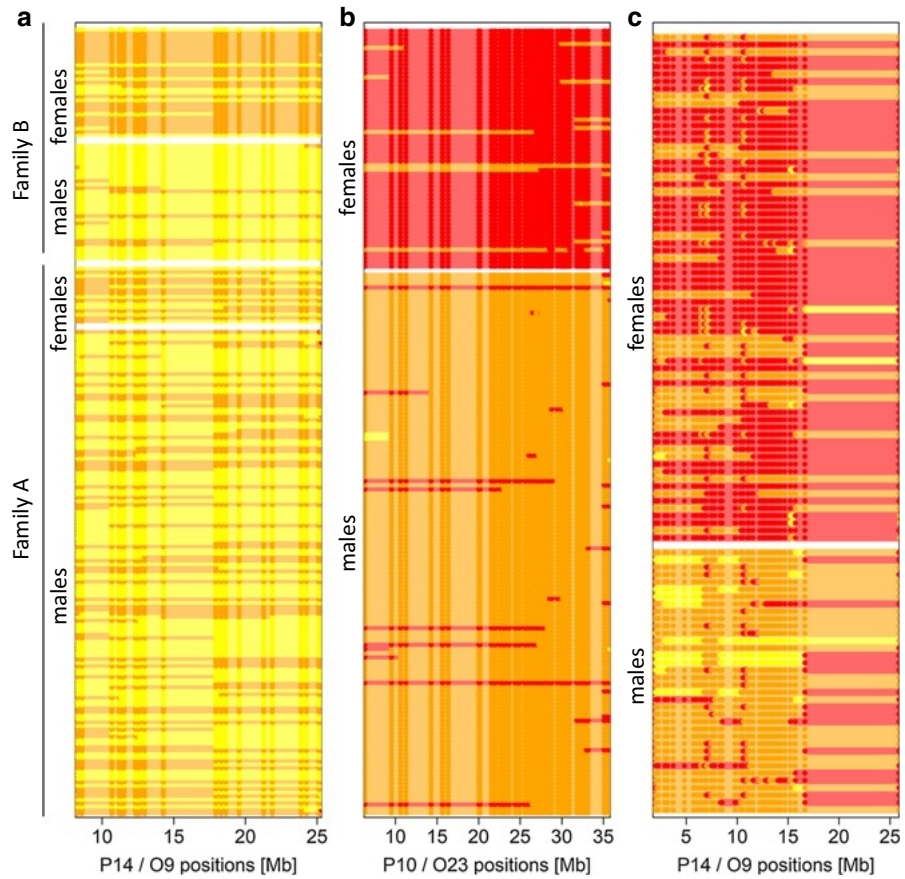


Fig. 5 Genotypes at sites with Y or W-linked alleles. Genotypes of the F2 individuals are visualized as red for homozygotes of the grandmother-derived allele, orange for heterozygotes and yellow for homozygotes of the grandfather-derived allele. **a** Victoria1 sites where the F0 and F1 females are heterozygous. Heterozygotes (orange) carry the W allele from the *P.* sp. "nyererei-like" grandmother (W_{NYL}). Most females of family B have a W allele and most males of family B do not. However, the sex-specific difference is smaller in family A. **b** Victoria2 sites where the male F0 and F1 are heterozygous. Except for three individuals, all males and no females are heterozygous and thus carry the Y-allele from the *P.* sp. "red-head" grandfather (Y_{RED}). **c** Malawi sites where the grandfather surrogates are heterozygous. Heterozygotes carry the Y-allele from the *P. taeniolatus* grandfather (Y_{TAE}). All males carry one or even two Y alleles, and most females have no Y allele. For more details see Figs. S3–S5

individuals in all crosses (Figs. 4, S6). However, the differences are small and the divergence between different sex chromosomes from the same parental species ($W_{NYL}Z_{NYL}$, $X_{RED}Y_{RED}$, $X_{TAE}Y_{TAE}$) is similar to the divergence between the Lake Victoria cichlid species of the different crosses which are less than 15,000 years divergent (Fig. S6). Therefore, we conclude that the sex chromosomes have likely diverged very recently.

Discussion

In our analyses of three interspecific crosses of haplochromine cichlids from Lakes Victoria and Malawi we find different sex determining chromosomes and systems even between the species from Lake Victoria which are all less than 15,000 years divergent. This highlights the high variability in sex determination in haplochromine cichlid fishes, consistent with previous findings (Holzberg, 1978; Seehausen et al., 1999; Lande et al., 2001; Gammerdinger & Kocher 2018; Böhne et al., 2019). Furthermore, we report a chromosome to be involved in sex determination that has not previously been reported as such in East African cichlids (Fig. 6). This chromosome contains a sex determiner in one of the Lake Victoria species and in one of the Lake Malawi species that are $\sim$ 2.5 My divergent, though possibly on different regions of the chromosome and with a different heterogametic sex (Fig. 4).

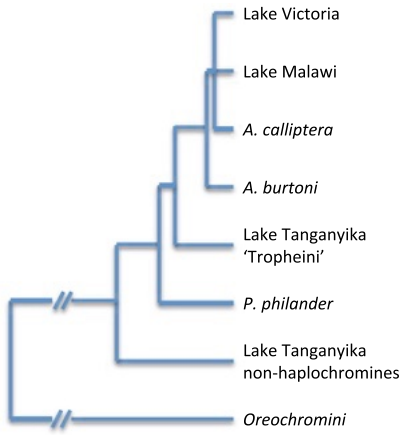
Using interspecific crosses to infer sex determination

Here, we identified sex determiners in interspecific crosses with one set of grandparents. This allowed us to identify dominant sex determiners contributed by one of the grandparents. In the sympatric Victoria cross (Victoria1), we identified a dominant female determiner from the *P.* sp. "nyererei-like" grandmother (W_{NYL}) on chromosome P14/O9 (Fig. 4). Therefore, we can conclude that in *P.* sp. "nyererei-like" sex is determined by a ZW system on P14/09, but we cannot test if it is fixed in *P.* sp. "nyererei-like" or if this sex determiner also exists in the second grandparental species, *P.* sp. "pundamilia-like". In the non-sympatric Victoria cross (Victoria2), we identified a dominant male determiner from the *P.* sp. "red-head" grandfather (Y_{RED}) on chromosome P10/O23. We did not find any evidence for a dominant female sex determiner. Therefore, it is likely that in *P. pundamilia*, the grandmaternal species, sex is also determined by an XY system, but it is unclear if the same chromosome (P10/O23) is sex determining. In the Malawi cross, a dominant male determiner from the *P. taeniolatus* grandfather (Y_{TAE}) is located on chromosome P14/O9. We found evidence for an additional female determiner in this cross which could have been contributed by the grandmaternal species *A. calliptera*. While interspecific crosses allow the detection of dominant sex determiners that can be traced back to one of the grandparental species, additional intra- or interspecific crosses with different sets of grandparents and reciprocal crossings would be needed to infer whether these sex determiners exist in both crossed species.

Multiple sex determiners

In two of our interspecies crosses, the Malawi and the Victoria1 cross, our results suggest the presence of multiple sex determining loci or sex modifiers. However, as the crossed species could have different sex determiners, this does not necessarily mean that multiple sex determiners exist in a single species. A sex determination locus in *A. calliptera*, the species used as grandmother for our Malawi cross, has already been mapped to P2/O7 and identified as a male heterogametic XY system (Peterson et al., 2017). Our *A. calliptera* grandmother should thus have been XX on P2/O7 and should not have contributed a dominant sex determination allele. Consistent with this, we found a different sex determining region on P14/09 for which the Y-allele can be traced back to the *P. taeniolatus* grandfather (Y_{TAE}). However, a simple XY-system cannot explain the female-biased sex ratio in this cross and the fact that 20% of the females carry a Y chromosome (most of which are $X_{CAL}Y_{TAE}$). It is thus likely that a female determiner or modifier interacting with X_{CAL} exists. Alternatively, Y_{TAE} may show incomplete penetrance in the presence of X_{CAL}, comparable to the findings of (Parnell & Streelman, 2013), who found a complex polygenic system that determined sex in a hierarchical fashion in their interspecific cross.

Multiple sex determining loci likely also segregate in our Victoria1 cross of very closely related sympatric



Linkage Group											
O1 P5	O2 P19	O3 P1	O5 P13	O7 P2	O9 P14	O13 P21	O14 P16	O18 P9	O19 P8	O20 P17	O23 P10	B
	??		??		ZW						XY	ZW
		ZW	ZW	XY	XY					XY		ZW
				XY								
			XY			ZW		XY				
									XY			
				XY								
			ZW	ZW								
XY, ZW		XY, ZW					XY				XY	

Fig. 6 Chromosome P14/O9 has not previously been reported to be sex determining in East African cichlids. The table shows an updated overview of sex determination systems in East African cichlids based on information summarised in Böhne et al. (2019). Linkage group numbers according to the *Oreochromis niloticus* (O) and *Pundamilia nyererei* (P) references. XY, ZW and "??" refers to male, female, or unknown heterogamety, respectively. Phylogenetic relationships are shown schematically to the left of the table redrawn from Meyer et al. (2016). Chromosome P14/O9 (highlighted in orange) is sex-determining in the Victoria1 cross and in the Lake Malawi cross. It has not been found to be sex determining in any other East African cichlid species before

species. In this cross, the W-allele in the identified sex determining region on P14/O9 can be traced back to the *P.* sp. "nyererei-like" grandmother (W_{NYL}). However, the two families differ in the strength of this correlation of W_{NYL}-presence with sex, and also feature highly distorted sex ratios (in opposite directions). Both of these are indications for the presence of additional sex determiners. In the female-biased family B, males that combine Z chromosomes from the different grandparental species ($Z_{NYL}Z_{PUL}$) are strongly underrepresented which could indicate reduced survival of these males (see also below). Additional sex determination loci must be present in both families (see Fig. S3). The two species of this cross diverged only a few hundred years ago in sympatry (Meier et al., 2017). Given their recent speciation and continued presence of gene flow during speciation it may seem unlikely that they could have evolved different sex determiners. However, these species are derived from a hybrid population between two older species (*P. nyererei* and *P. pundamilia* (Meier et al., 2017)), both of which may have contributed distinct sex determiners and/or sex modifiers. One of the parental species of the hybrid population from which the two younger species evolved, *P. pundamilia*, was used as the grandmother in our Victoria2 cross. In this cross, we only found a single sex determination locus with a male determiner on P10/O23 that came from the grandfather (Y_{RED}, *P.* sp. "red-head"). It is thus unlikely that our *P. pundamilia* grandmother shared the dominant female determiner on chromosome P14/O9 with *P.* sp. "nyererei-like". Our findings are consistent with different sex determiners in *P. nyererei* and *P. pundamilia* which gave rise to the hybrid swarm from which *P.* sp. "nyererei-like" and *P.* sp. "pundamilia-like" evolved. A next step would be to sequence population samples to test if the sex determiners represent fixed species differences or if multiple sex determiners segregate in some of the species.

While polygenic sex determination can be an evolutionary stable strategy in some cases (Moore & Roberts, 2013), it is more commonly thought to be a transient state between an ancestral and derived sex determination system (Rice, 1986; Van Doorn, 2014). This is because selection against sex ratio distortion or a fitness advantage of one of the sex determiners is expected to lead to the fixation of a single sex

determiner (Orzack et al., 1980; Bull, 1983; Lande et al., 2001; Moore & Roberts 2013; Van Doorn, 2014). Nevertheless, the presence of multiple sex determiners has been demonstrated in the more distantly related haplochromine cichlid *Astatotilapia burtoni* (Günther, 1894), where at least three different chromosomes carry sex determining loci (Böhne et al., 2016; Roberts et al., 2016). Multiple sex-determining loci are likely also segregating in some Lake Malawi cichlid species (e.g. Parnell & Streelman, 2013). If sex determination is indeed commonly polygenic in East African cichlids, multiple crosses with different sets of grandparents are required to identify the different sex determiners. Moreover, larger numbers of individuals would be needed to detect additional weaker sex determining loci. A second sex determiner could be located on a B chromosome and would thus be even harder to detect. The presence of feminizing B chromosomes has been confirmed in other cichlids (Yoshida et al., 2011; Clark & Kocher, 2019). In family B of the Victoria1 cross, the females may either carry a dominant female determiner on linkage group P14/O9 or a B chromosome, explaining the female-biased sex ratio. Similarly, a feminizing B chromosome could explain the female-biased sex ratio in the Malawi cross and the presence of females with a Y chromosome. Whole-genome sequencing or the detection of B-chromosomal markers (as in Clark et al., 2018) would be needed to test this hypothesis.

Same chromosome: different heterogametic sex

In two crosses we have identified the same chromosome (P14/O9) as sex determining but with an XY system in *Protomelas taeniolatus* from Lake Malawi, and a ZW system in *Pundamilia* sp. "nyererei-like" from Lake Victoria. Similar findings have been made in other cichlids, where the same chromosome acts as XY or ZW system in different lineages. Chromosome P13/O5 is an XY sex chromosome in *A. burtoni* (Böhne et al., 2016; Roberts et al., 2016) and ZW in several Lake Malawi species (Ser et al., 2010), and similarly, P2/O7 is an XY sex chromosome in *A. calliptera* and other species from Lake Malawi (Ser et al., 2010; Peterson et al., 2017) and ZW in *Hemibates stenosoma* (Boulenger, 1901) from Lake Tanganyika (Gammerdinger et al., 2018). Theoretical studies suggest that transitions from an XY to a ZW system or vice versa could happen readily in the presence of sex ratio distortion, which in turn can be caused by meiotic drive elements (Scott et al., 2018) or by a new sex determiner that can spread if it is under sexual selection (Lande et al., 2001; Vuilleumier et al., 2007) or if it reduces sexual conflict (van Doorn & Kirkpatrick, 2007, 2010). Sex determination system turnover can even be caused by drift (Veller et al., 2017) or its interaction with sex ratio selection (Vuilleumier et al., 2007). Transitions between male and female heterogamety involving different sex determining chromosomes are widespread in many vertebrate groups (Ezaz et al., 2006; Pennell et al., 2018). However, switches of the heterogametic sex on the same chromosome (homologous transitions sensu van Doorn & Kirkpatrick, 2007) are less common, including only few known cases such as the frog *Rana rugosa* Temminck and Schlegel, 1838 (Miura, 2007; Ogata et al., 2008), *Neochromis* cichlids (Seehausen et al., 1999) and platyfish (Kallman, 1968).

In some species, male and female determiners on the same chromosome are found within a single population. Two different dominant female determiners occur on some X chromosomes in the Makobe Island population of the cichlid fish *Neochromis omnicaeruleus* Seehausen & Bouton, 1998 in Lake Victoria (Seehausen et al., 1999) and in multiple rodents (Veyrunes et al., 2010). Similarly, we find an effect of the recessive allele in both crosses with a sex determiner on chromosome P14/O9. In the Malawi cross, where we identified a dominant male determiner derived from *P. taeniolatus* (Y_{TAE}), some individuals with a Y_{TAE} allele are female. In those XY females, the second allele is mostly derived from *A. calliptera* (X_{CAL}) and only one carries an X_{TAE} allele (Figs. S3, S4). This may indicate that X_{CAL} has a feminizing effect. Similarly, in family B of the Victoria1 cross, we identified a dominant female determiner derived from *P.* sp. "nyererei-like" (W_{NYL}). Most individuals with Z chromosomes derived from the same parental species ($Z_{PUL}Z_{PUL}$) are male (28/30). However, only seven individuals carry Z chromosomes derived from different parental species ($Z_{NYL}Z_{PUL}$) and only two of those are male (Figs. S3, S5). Therefore, Z_{NYL} could have a feminizing effect, but the notable paucity of $Z_{NYL}Z_{PUL}$ individuals in the Victoria1 family B (Fig. S3) more likely indicates reduced survival of males with two different Z chromosomes and thus the presence of incompatibilities.

Even though the Malawi and the Victoria1 crosses revealed the same sex chromosome with different heterogametic sexes, this case is likely not an example of homologous transition in heterogametic sex because the sex determiners do not represent direct sister states. At least four other chromosomes are sex determining in different cichlid species of Lake Malawi and, as our study highlights, other cichlid species from Lake Victoria also feature different sex determining chromosomes. It is thus likely that many switches between sex determining chromosomes occurred in the ancestry of the two species we here found to share the same sex chromosome. Our finding of different sex determiners on the same chromosome may rather support the hypothesis that certain chromosomes are more prone to carry sex determiners than other chromosomes as has been suggested for cichlids (Böhne et al., 2016), Ranidae frogs (Jeffries et al., 2018) and vertebrates in general (Marshall Graves & Peichel, 2010). These chromosomes may be enriched for genes contributing to the sex determination cascade or genes with sexually antagonistic effects (Blaser et al., 2014). Our study contributes to the accumulating information on loci involved in sex determination or sexual conflict in increasingly more species of the East African cichlid fish radiations, which will eventually allow us to test these hypotheses.

Sex chromosome evolution

Male and female versions of sex chromosomes typically diverge over time as they accumulate neutral and deleterious mutations (Charlesworth, 1996; Bachtrog, 2013) and sexually antagonistic mutations (Charlesworth, 2017) on the non-recombining parts of the Y or W chromosome. However, we did not find a sex difference in sequencing depth in any cross, indicating that the sex chromosomes are not (yet) degenerated. The low divergence between the sex chromosomes (Fig. S6) indicates that they are very young, consistent with rapid turnover of sex chromosomes between species in these young radiations. The lack of sex chromosome degeneration may also facilitate further turnover of sex chromosomes as the accumulation of deleterious mutations is expected to stabilize sex determination systems (van Doorn & Kirkpatrick, 2010). This is because the invasion of a novel dominant sex determiner on another chromosome leads to a high proportion of individuals that are homozygous for the ancestral dominant sex determiner (YY or WW individuals) which have reduced fitness if deleterious mutations accumulated on this chromosome. Therefore, the lack of sex chromosome degeneration in the haplochromine cichlids may both be a consequence of rapid sex chromosome turnover and facilitate additional sex chromosome turnover.

Conclusions

Overall, our results give further support to a growing body of evidence that sex determination in African cichlid fish is highly evolvable and may often involve several sex determiners with variable dominance relationships. Furthermore, switches between XY and ZW systems on the same chromosome seem to have occurred repeatedly. All of these factors may play a role in the adaptive radiations of East African cichlids. Studying additional species of the rapid radiations of haplochromine cichlids in Lakes Victoria, Malawi and also the smaller lakes in the region, will eventually allow us to better understand why there is such rapid turnover of sex determination systems in this species group and what the underlying mechanisms are. It will also allow us to find an answer to the question whether the rapid evolution of sex determination is more cause or more consequence of rapid species radiations. Cichlids are thus an ideal system to assess the importance of the evolution of sex determination in species diversification.

Acknowledgments We thank Marcel Haesler for breeding the Lake Victoria cichlids, Andreas Taverna for breeding the Lake Malawi cichlids, Katie Peichel for valuable discussions and feedback. The Genetic Diversity Centre (GDC) at ETH Zürich provided the bioinformatics platform and support for genomic analyses.

Author contributions AFF performed QTL mapping analyses and wrote the manuscript together with JIM. VO performed a part of the genomic analyses. OS planned and started all lab crosses and contributed to study design, interpretation of results and writing. JIM designed the project, performed genomic analyses, and wrote the manuscript together with AFF.

Funding This research was funded by Swiss National Science Foundation (SNSF) Grants 31003A_163338, 31003A_144046 and 31003A_118293 to OS. JIM is supported by a St John's College Cambridge Fellowship and a Branco Weiss Fellowship.

Data availability Raw sequence data (fastq files) for all three crosses are available on the NCBI short read archive (SRA) under the following accession numbers: Victoria1 PRJNA612290 (from Feller et al. (2020a)); Victoria2 PRJNA439430 (from Feulner et al. (2018)); Malawi PRJNA612298 (from Feller et al. (2020b)). All files for QTL mapping and segregation pattern analyses and the code are available on the Dryad Digital Repository at https://doi.org/10.5061/dryad.b5mkkwhc8.

Declarations

Conflict of interest We declare that we have no competing interests.

Research involving animal rights Fish experimentation and euthanasia were authorized by the veterinary offices of the cantons of Lucerne (Licence Number: LU04/07) and Bern (BE 18/15 & BE 65/18) in Switzerland.

References

Bachtrog, D., 2013. Y-chromosome evolution: emerging insights into processes of Y-chromosome degeneration. Nature Reviews Genetics 14: 113–124.

Bachtrog, D., J. E. Mank, C. L. Peichel, M. Kirkpatrick, S. P. Otto, T.-L. Ashman, M. W. Hahn, J. Kitano, I. Mayrose & R. Ming, 2014. Sex determination: why so many ways of doing it? PLoS Biology 12: e1001899.

Baird, N. A., P. D. Etter, T. S. Atwood, M. C. Currey, A. L. Shiver, Z. A. Lewis, E. U. Selker, W. A. Cresko & E. A. Johnson, 2008. Rapid SNP discovery and genetic mapping using sequenced RAD markers. PLoS ONE 3: e3376.

Bell, G., 1982. The Masterpiece of Nature: The Evolution and Genetics of Sexuality. University of California Press, Berkeley.

Bezault, E., S. Mwaiko & O. Seehausen, 2011. Population genomic tests of models of adaptive radiation in Lake Victoria Region cichlid fish. Evolution 65: 3381–3397.

Blaser, O., C. Grossen, S. Neuenschwander & N. Perrin, 2013. Sex-chromosome turnovers induced by deleterious mutation load. Evolution 67: 635–645.

Blaser, O., S. Neuenschwander & N. Perrin, 2014. Sex-chromosome turnovers: the hot-potato model. The American Naturalist 183: 140–146.

Böhne, A., C. A. Wilson, J. H. Postlethwait & W. Salzburger, 2016. Variations on a theme: genomics of sex determination in the cichlid fish *Astatotilapia burtoni*. BMC Genomics 17: 1–12.

Böhne, A., A. A.-T. Weber, J. Rajkov, M. Rechsteiner, A. Riss, B. Egger & W. Salzburger, 2019. Repeated evolution versus common ancestry: sex chromosome evolution in the Haplochromine Cichlid *Pseudocrenilabrus philander*. Genome Biology and Evolution 11: 439–458.

Brawand, D., C. E. Wagner, Y. I. Li, M. Malinsky, I. Keller, S. H. Fan, O. Simakov, A. Y. Ng, Z. W. Lim, E. Bezault, J. Turner-Maier, J. Johnson, R. Alcazar, H. J. Noh, P. Russell, B. Aken, J. Alfoldi, C. Amemiya, N. Azzouzi, J. F. Baroiller, F. Barloy-Hubler, A. Berlin, R. Bloomquist, K. L. Carleton, M. A. Conte, H. D'Cotta, O. Eshel, L. Gaffney, F. Galibert, H. F. Gante, S. Gnerre, L. Greuter, R. Guyon, N. S. Haddad, W. Haerty, R. M. Harris, H. A. Hofmann, T. Hourlier, G. Hulata, D. B. Jaffe, M. Lara, A. P. Lee, I. MacCallum, S. Mwaiko, M. Nikaido, H. Nishihara, C. Ozouf-Costaz, D. J. Penman, D. Przybylski, M. Rakotomanga, S. C. P. Renn, F. J. Ribeiro, M. Ron, W. Salzburger, L. Sanchez-Pulido, M. E. Santos, S. Searle, T. Sharpe, R. Swofford, F. J. Tan, L. Williams, S. Young, S. Y. Yin, N. Okada, T. D. Kocher, E. A. Miska, E. S. Lander, B. Venkatesh, R. D. Fernald, A. Meyer, C. P. Ponting, J. T. Streelman, K. Lindblad-Toh, O. Seehausen & F. Di Palma, 2014. The genomic substrate for adaptive radiation in African cichlid fish. Nature 513: 375–381.

Broman, K. W. & S. Sen, 2009. A Guide to QTL Mapping with R/qtl, Vol. 46. Springer, New York.

Broman, K. W., H. Wu, Ś. Sen & G. A. Churchill, 2003. R/qtl: QTL mapping in experimental crosses. Bioinformatics 19: 889–890.

Browning, S. R. & B. L. Browning, 2011. Haplotype phasing: existing methods and new developments. Nature Reviews Genetics 12: 703–714.

Bull, J. J., 1983. Evolution of Sex Determining Mechanisms. The Benjamin/Cummings Publishing Company, Inc., San Francisco.

Catchen, J., P. A. Hohenlohe, S. Bassham, A. Amores & W. A. Cresko, 2013. Stacks: an analysis tool set for population genomics. Molecular Ecology 22: 3124–3140.

Charlesworth, B., 1996. The evolution of chromosomal sex determination and dosage compensation. Current Biology 6: 149–162.

Charlesworth, D., 2017. Evolution of recombination rates between sex chromosomes. Philosophical Transactions of the Royal Society B: Biological Sciences 372: 20160456.

Cheng, J., T. Czypionka & A. W. Nolte, 2013. The genomics of incompatibility factors and sex determination in hybridizing species of Cottus (Pisces). Heredity 111: 520–529.

Clark, F. E. & T. D. Kocher, 2019. Changing sex for selfish gain: B chromosomes of Lake Malawi cichlid fish. Scientific Reports 9: 20213.

Clark, F. E., M. A. Conte, I. A. Ferreira-Bravo, A. B. Poletto, C. Martins & T. D. Kocher, 2016. Dynamic sequence evolution of a sex-associated B chromosome in Lake Malawi Cichlid fish. Journal of Heredity 108: 53–62.

Clark, F. E., M. A. Conte & T. D. Kocher, 2018. Genomic characterization of a B chromosome in Lake Malawi cichlid fishes. Genes 9: 610.

Cortez, D., R. Marin, D. Toledo-Flores, L. Froidevaux, A. Liechti, P. D. Waters, F. Grützner & H. Kaessmann, 2014. Origins and functional evolution of Y chromosomes across mammals. Nature 508: 488–493.

Danecek, P., A. Auton, G. Abecasis, C. A. Albers, E. Banks, M. A. DePristo, R. E. Handsaker, G. Lunter, G. T. Marth, S. T. Sherry, G. McVean, R. Durbin & G. Genomes Project Analysis, 2011. The variant call format and VCFtools. Bioinformatics 27: 2156–2158.

Devlin, R. H. & Y. Nagahama, 2002. Sex determination and sex differentiation in fish: an overview of genetic, physiological, and environmental influences. Aquaculture 208: 191–364.

Ezaz, T., R. Stiglec, F. Veyrunes & J. A. M. Graves, 2006. Relationships between vertebrate ZW and XY sex chromosome systems. Current Biology 16: R736–R743.

Feller, A. F., M. P. Haesler, C. L. Peichel & O. Seehausen, 2020a. Genetic architecture of a key reproductive isolation trait differs between sympatric and non-sympatric sister species of Lake Victoria cichlids. Proceedings of the Royal Society B-Biological Sciences 287: 20200270.

Feller, A. F., O. M. Selz, M. D. McGee, J. I. Meier, S. Mwaiko & O. Seehausen, 2020b. Rapid generation of ecologically relevant behavioral novelty in experimental cichlid hybrids. Ecology and Evolution 10: 7445–7462.

Feulner, P. G. D., J. Schwarzer, M. P. Haesler, J. I. Meier & O. Seehausen, 2018. A dense linkage map of Lake Victoria cichlids improved the Pundamilia genome assembly and revealed a major QTL for sex-determination. G3: Genes, Genomes, Genetics 8: 2411–2420.

Gammerdinger, W. & T. Kocher, 2018. Unusual diversity of sex chromosomes in African Cichlid fishes. Genes 9: 480.

Gammerdinger, W. J., M. A. Conte, B. A. Sandkam, A. Ziegelbecker, S. Koblmüller & T. D. Kocher, 2018. Novel sex chromosomes in 3 cichlid fishes from Lake Tanganyika. Journal of Heredity 109: 489–500.

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

Holzberg, S., 1978. A field and laboratory study of the behaviour and ecology of Pseudotropheus zebra (Boulenger), an endemic cichlid of Lake Malawi (Pisces; Cichlidae)M1. Journal of Zoological Systematics and Evolutionary Research 16: 171–187.

Ivory, S. J., M. W. Blome, J. W. King, M. M. McGlue, J. E. Cole & 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.

Jeffries, D. L., G. Lavanchy, R. Sermier, M. J. Sredl, I. Miura, A. Borzée, L. N. Barrow, D. Canestrelli, P.-A. Crochet, C. Dufresnes, J. Fu, W.-J. Ma, C. M. Garcia, K. Ghali, A. G. Nicieza, R. P. O'Donnell, N. Rodrigues, A. Romano, Í. Martínez-Solano, I. Stepanyan, S. Zumbach, A. Brelsford & N. Perrin, 2018. A rapid rate of sex-chromosome turnover and non-random transitions in true frogs. Nature Communications 9: 4088–4088.

Johnson, T. C., K. Kelts & E. Odada, 2000. The holocene history of Lake Victoria. Ambio 29: 2–11.

Kallman, K. D., 1968. Evidence for the existence of transformer genes for sex in the teleost Xiphophorus maculatus. Genetics 60: 811–828.

Kitano, J. & C. L. Peichel, 2012. Turnover of sex chromosomes and speciation in fishes. Environmental Biology of Fishes 94: 549–558.

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

Kozielska, M., F. J. Weissing, L. W. Beukeboom & I. Pen, 2010. Segregation distortion and the evolution of sex-determining mechanisms. Heredity 104: 100–112.

Kudo, Y., M. Nikaido, A. Kondo, H. Suzuki, K. Yoshida, K. Kikuchi & N. Okada, 2015. A microsatellite-based genetic linkage map and putative sex-determining genomic regions in Lake Victoria cichlids. Gene 560: 156–164.

Lande, R., O. Seehausen & J. J. M. van Alphen, 2001. Mechanisms of rapid sympatric speciation by sex reversal and sexual selection in cichlid fish. Genetica 112: 435–443.

Langmead, B. & S. L. Salzberg, 2012. Fast gapped-read alignment with Bowtie 2. Nature Methods 9: 357-U54.

Li, H., B. Handsaker, A. Wysoker, T. Fennell, J. Ruan, N. Homer, G. Marth, G. Abecasis & R. Durbin, 2009. The sequence alignment/map format and SAMtools. Bioinformatics 25: 2078–2079.

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

Marshall Graves, J. A. & C. L. Peichel, 2010. Are homologies in vertebrate sex determination due to shared ancestry or to limited options? Genome Biology 11: 205.

McGee, M. D., S. R. Borstein, J. I. Meier, D. A. Marques, S. Mwaiko, A. Taabu, M. A. Kishe, B. O'Meara, R. Bruggmann, L. Excoffier & O. Seehausen, 2020. The ecological and genomic basis of explosive adaptive radiation. Nature 586: 75–79.

McKenna, A., M. Hanna, E. Banks, A. Sivachenko, K. Cibulskis, A. Kernytsky, K. Garimella, D. Altshuler, S. Gabriel, M. Daly & M. A. DePristo, 2010. The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. Genome Research 20: 1297–1303.

Meier, J. I., V. C. Sousa, D. A. Marques, O. M. Selz, C. E. Wagner, L. Excoffier & O. Seehausen, 2017. Demographic modelling with whole-genome data reveals parallel origin of similar Pundamilia cichlid species after hybridization. Molecular Ecology 26: 123–141.

Meyer, B. S., M. Matschiner & W. Salzburger, 2016. Disentangling incomplete lineage sorting and introgression to refine species-tree estimates for Lake Tanganyika cichlid fishes. Systematic Biology 66: 531–550.

Ming, R., A. Bendahmane & S. S. Renner, 2011. Sex chromosomes in land plants. Annual Review of Plant Biology 62: 485–514.

Miura, I., 2007. An evolutionary witness: the frog *Rana rugosa* underwent change of heterogametic sex from XY male to ZW female. Sexual Development 1: 323–331.

Moore, E. C. & R. B. Roberts, 2013. Polygenic sex determination. Current Biology 23: R510–R512.

Ogata, M., Y. Hasegawa, H. Ohtani, M. Mineyama & I. Miura, 2008. The ZZ/ZW sex-determining mechanism originated twice and independently during evolution of the frog, *Rana rugosa*. Heredity 100: 92–99.

Orzack, S. H., J. J. Sohn, K. D. Kallman, S. A. Levin & R. Johnston, 1980. Maintenance of the three sex chromosome polymorphism in the Platyfish, *Xiphophorus maculatus*. Evolution 34: 663–672.

Parnell, N. F. & J. T. Streelman, 2013. Genetic interactions controlling sex and color establish the potential for sexual conflict in Lake Malawi cichlid fishes. Heredity 110: 239–246.

Pennell, M. W., J. E. Mank & C. L. Peichel, 2018. Transitions in sex determination and sex chromosomes across vertebrate species. Molecular Ecology 27: 3950–3963.

Peterson, E. N., M. E. Cline, E. C. Moore, N. B. Roberts & R. B. Roberts, 2017. Genetic sex determination in *Astatotilapia calliptera*, a prototype species for the Lake Malawi cichlid radiation. Naturwissenschaften 104: 41.

Rice, W. R., 1986. On the instability of polygenic sex determination: the effect of sex-specific selection. Evolution 40: 633–639.

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

Roberts, N. B., S. A. Juntti, K. P. Coyle, B. L. Dumont, M. K. Stanley, A. Q. Ryan, R. D. Fernald & R. B. Roberts, 2016. Polygenic sex determination in the cichlid fish *Astatotilapia burtoni*. BMC Genomics 17: 835.

Sahara, K., A. Yoshido & W. Traut, 2012. Sex chromosome evolution in moths and butterflies. Chromosome Research 20: 83–94.

Sambrook, J. & D. W. Russell, 2001. Molecular Cloning: A Laboratory Manual, 3rd ed. Cold Spring Harbor Laboratory Press, Cold Spring Harbor.

Scott, M. F., M. M. Osmond & S. P. Otto, 2018. Haploid selection, sex ratio bias, and transitions between sex-determining systems. PLoS Biology 16: e2005609.

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

Selz, O. M., R. Thommen, M. E. Maan & O. Seehausen, 2014. Behavioural isolation may facilitate homoploid hybrid speciation in cichlid fish. Journal of Evolutionary Biology 27: 275–289.

Ser, J. R., R. B. Roberts & T. D. Kocher, 2010. Multiple interacting loci control sex determination in lake Malawi cichlid fish. Evolution 64: 486–501.

Stelkens, R. B., C. Schmid, O. Selz & O. Seehausen, 2009. Phenotypic novelty in experimental hybrids is predicted by the genetic distance between species of cichlid fish. BMC Evolutionary Biology 9: 283.

Van Doorn, G., 2014. Evolutionary transitions between sex-determining mechanisms: a review of theory. Sexual Development 8: 7–19.

van Doorn, G. S. & M. Kirkpatrick, 2007. Turnover of sex chromosomes induced by sexual conflict. Nature 449: 909–912.

van Doorn, G. S. & M. Kirkpatrick, 2010. Transitions between male and female heterogamety caused by sex-antagonistic selection. Genetics 186: 629–645.

Van Ooijen, J., 2006. JoinMap 4. Software for the calculation of genetic linkage maps in experimental populations Kyazma BV, Wageningen, Netherlands 33

Veller, C., P. Muralidhar, G. W. A. Constable & M. A. Nowak, 2017. Drift-induced selection between male and female heterogamety. Genetics 207: 711–727.

Veyrunes, F., P. Chevret, J. Catalan, R. Castiglia, J. Watson, G. Dobigny, T. J. Robinson & J. Britton-Davidian, 2010. A novel sex determination system in a close relative of the house mouse. Proceedings of the Royal Society B: Biological Sciences 277: 1049–1056.

Vuilleumier, S., R. Lande, J. Van Alphen & O. Seehausen, 2007. Invasion and fixation of sex-reversal genes. Journal of Evolutionary Biology 20: 913–920.

White, M. J. D., 1977. Animal Cytology and Evolution. CUP Archive, Cambridge.

Yoshida, K., Y. Terai, S. Mizoiri, M. Aibara, H. Nishihara, M. Watanabe, A. Kuroiwa, H. Hirai, Y. Hirai, Y. Matsuda & N. Okada, 2011. B chromosomes have a functional effect on female sex determination in Lake Victoria cichlid fishes. PLoS Genetics 7: e1002203.

Supplementary Figures and Tables

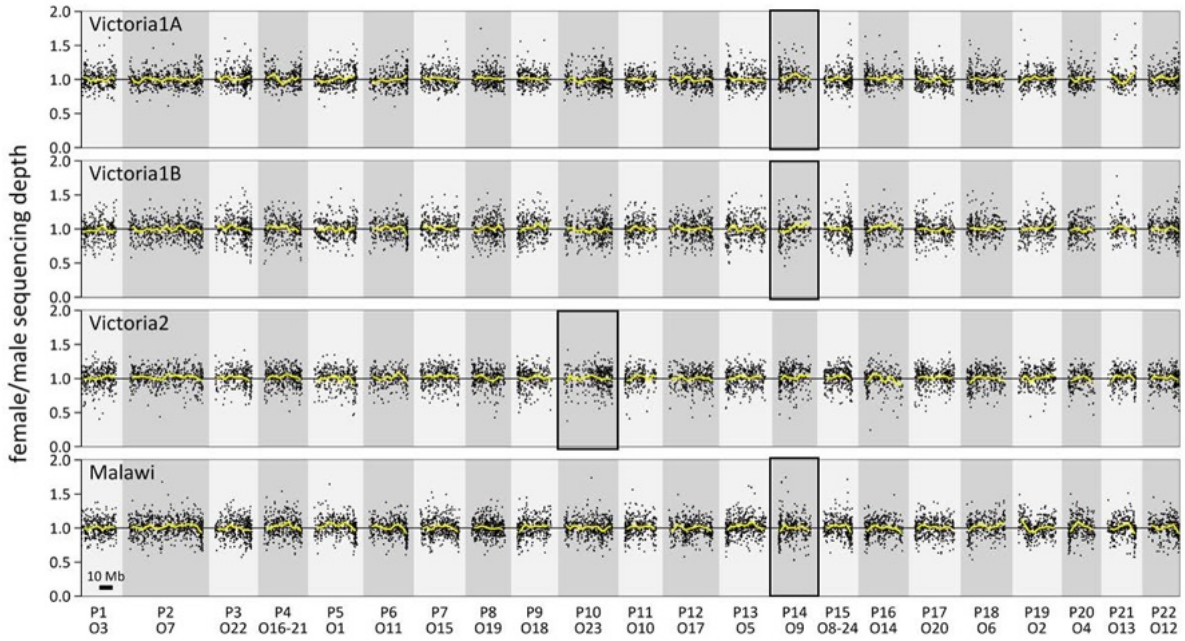


Figure S1 | Sequencing depth ratios between sexes show no evidence for degenerated sex chromosomes. Each dot shows the sequencing depth of females divided by the sequencing depth of males for a single SNP. The values are corrected for mean sequencing depth differences between the sexes to get a genome-wide average of 1. Yellow lines show 5 Mb window averages. Males and females show similar sequencing depths on all chromosomes, including the putative sex chromosomes (framed in black), which is inconsistent with strongly deteriorated Y or W chromosomes that would not map well to the reference genome.

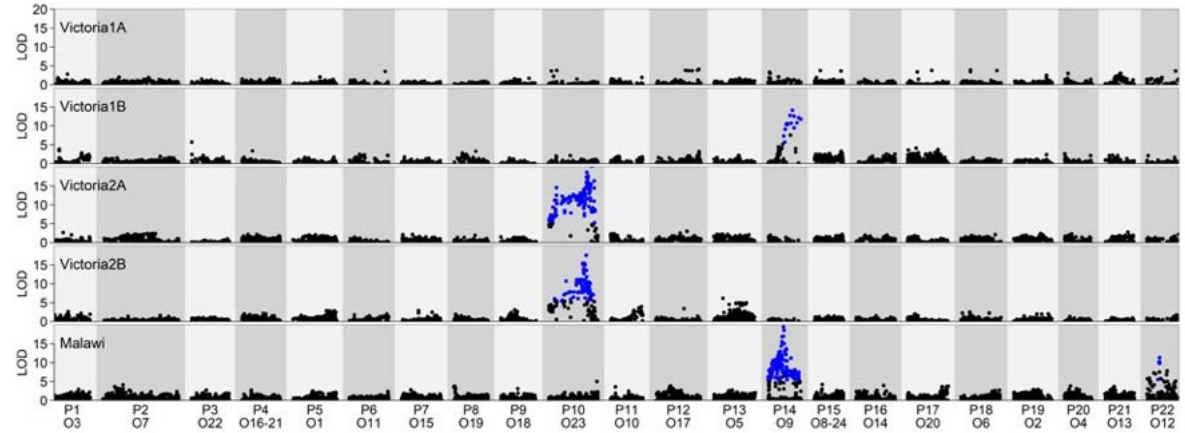


Figure S2 | Single marker regressions. Shown are LOD scores calculated from the F-statistic obtained in ANOVAs performed for each single marker (on x-axis). Highlighted in blue are scores associated with a p-value (from ANOVA) smaller than 0.05 after Bonferroni correction. The two families each of Victoria1 and Victoria2 are shown separately. In Victoria1, there is only a sex-association on chromosome P14/O9 in family A. In Victoria2, both families show similar patterns of a sex-association on chromosome P10/O23.

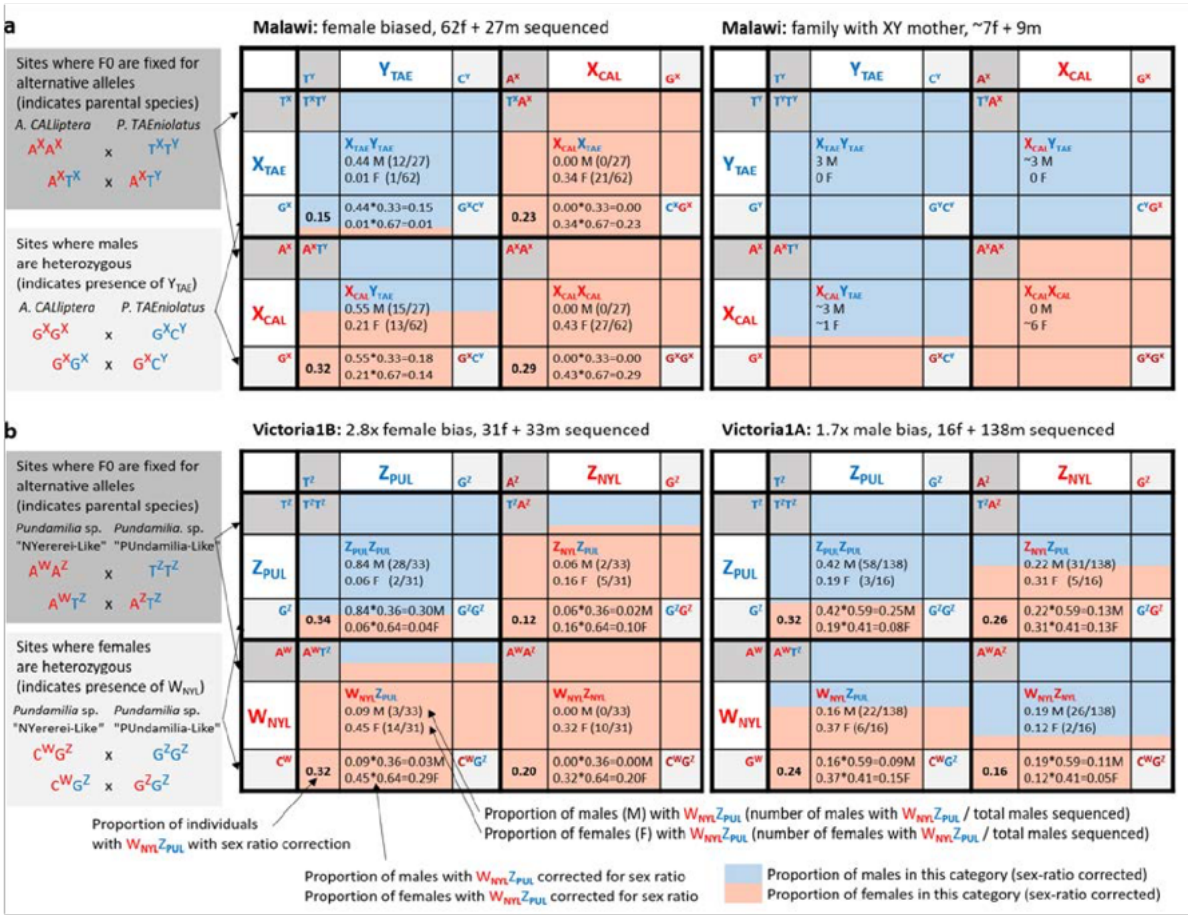


Figure S3 | Genotype proportions of males and females at the sex determining loci of the Malawi and the Victoria1 cross. The genotype at the bi-allelic sites where the grandparents (F0) are fixed for alternative alleles indicate from which parental species the allele is derived. The genotypes at sites where the parents and grandparents of the heterogametic sex are heterozygous indicate the presence of the Y or W allele. These two categories are exemplified with an AT SNP for F0 fixed sites (dark grey) and for a CG SNP for sites heterozygous in the heterogametic sex. The combination of the genotypes at these sites, allows to infer the genotype of each individual at the sex-determining locus. Each of the four genotype categories in the Punnet squares should contain 25% of the individuals. However, given the biased sex ratios in the Malawi and the Victoria1 cross and given that we have sampled males and females at different proportions, we cannot simply add up the number of males and females in each category. Instead, we have computed the proportion of individuals in each genotype category (center of each category) for males and females separately. This proportion was then multiplied by the observed overall proportion of offspring of that sex (including unsequenced individuals) to get the expected genotype proportion if we had sampled the individuals randomly irrespective of sex (bottom of each category). The sum of these sex-ratio corrected proportions for males and for females together gives the corrected proportion of individuals in each category which should be 25%. The background shading roughly indicates the sex-ratio corrected proportion of males (blue) and females (red) in each category, *i.e.* the expected probability that an individual with this genotype will be male or female. **a** Genotype proportions at the most sex associated sites on chromosome P14/O9 (position 12,920,486 as site where F0 are fixed for alternative alleles and position 12,620,027 as site heterozygous in males). We do not know which individuals belong to which families, but three males showing a $Y_{TAE}Y_{TAE}$ genotype and 6 other recombinant males (see Fig. S4) indicate the presence of a family where both F1 parents had an $X_{CAL}Y_{TAE}$ genotype. Given that we have sequenced about twice as many females, we would expect about 6 $X_{CAL}X_{CAL}$ females from this family and one female with a $X_{CAL}Y_{TAE}$ genotype. This family is shown separately on the right. **b** Genotype proportions of Victoria1 family B (left) and family A (right). In family B the association of W_{NYL} with sex is much stronger than in family A. Notably, in family B, there are too few individuals with two different Z chromosomes ($Z_{NYL}Z_{PUL}$; top right category). Individuals in this category are expected to be male, but there are only two $Z_{NYL}Z_{PUL}$ males compared to 28 males with $Z_{PUL}Z_{PUL}$ (binomial test: p-value = 8.7e-7). See also Fig. S5.

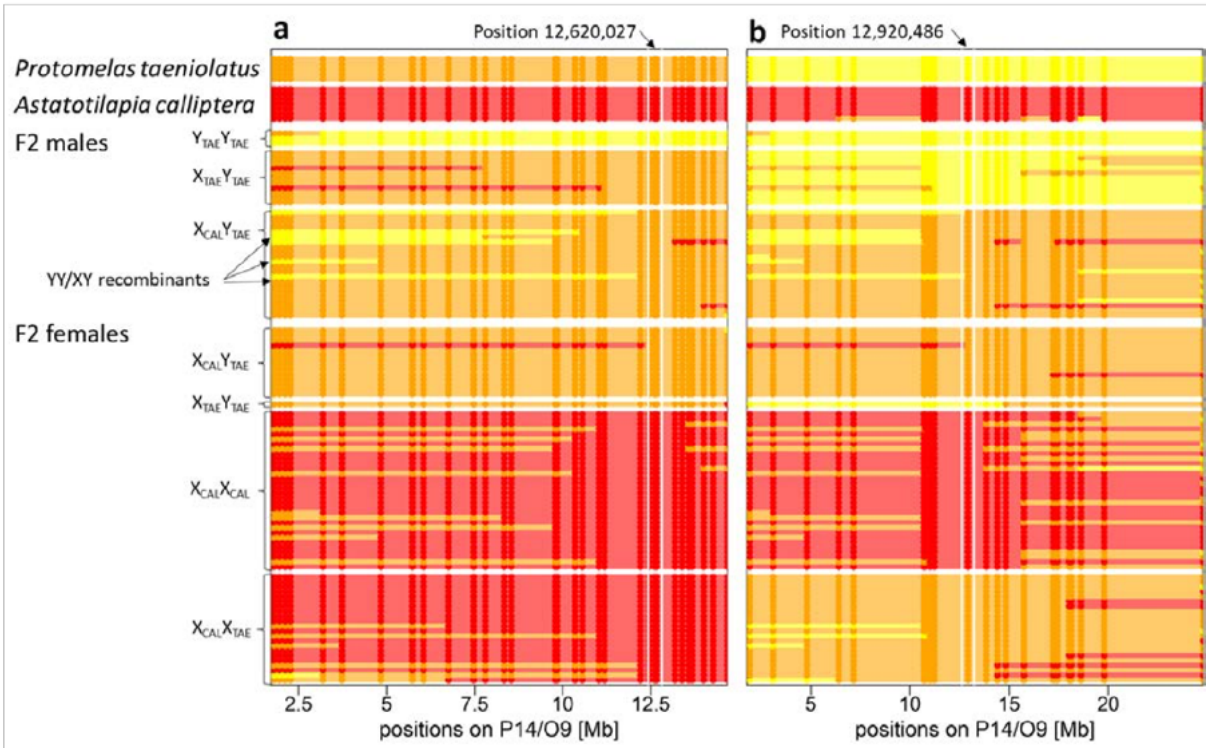


Figure S4 | Visualization of the genotypes along the sex chromosome in the Malawi cross.
a Sites where the grandparental surrogates, male *Protomelas taeniolatus*, are heterozygous and the grandmaternal surrogates, female *Astatotilapia calliptera*, are homozygous. These sites indicate the presence of a Y_{TAE} allele. Each individual is represented on a separate row with genotypes shown as red for XX, orange for XY and yellow for YY. The site used for the genotype proportions in Fig. S3 is indicated with a white frame. **b** Sites where the parental species have an allele frequency difference of at least 0.8. These sites allow to trace back the origin of the alleles to the parental species. The colours indicate parental species ancestry with homozygotes in yellow (TAE TAE) and red (CAL CAL) and heterozygotes (CAL TAE) in orange. The site used for Fig. S3 is framed in white.

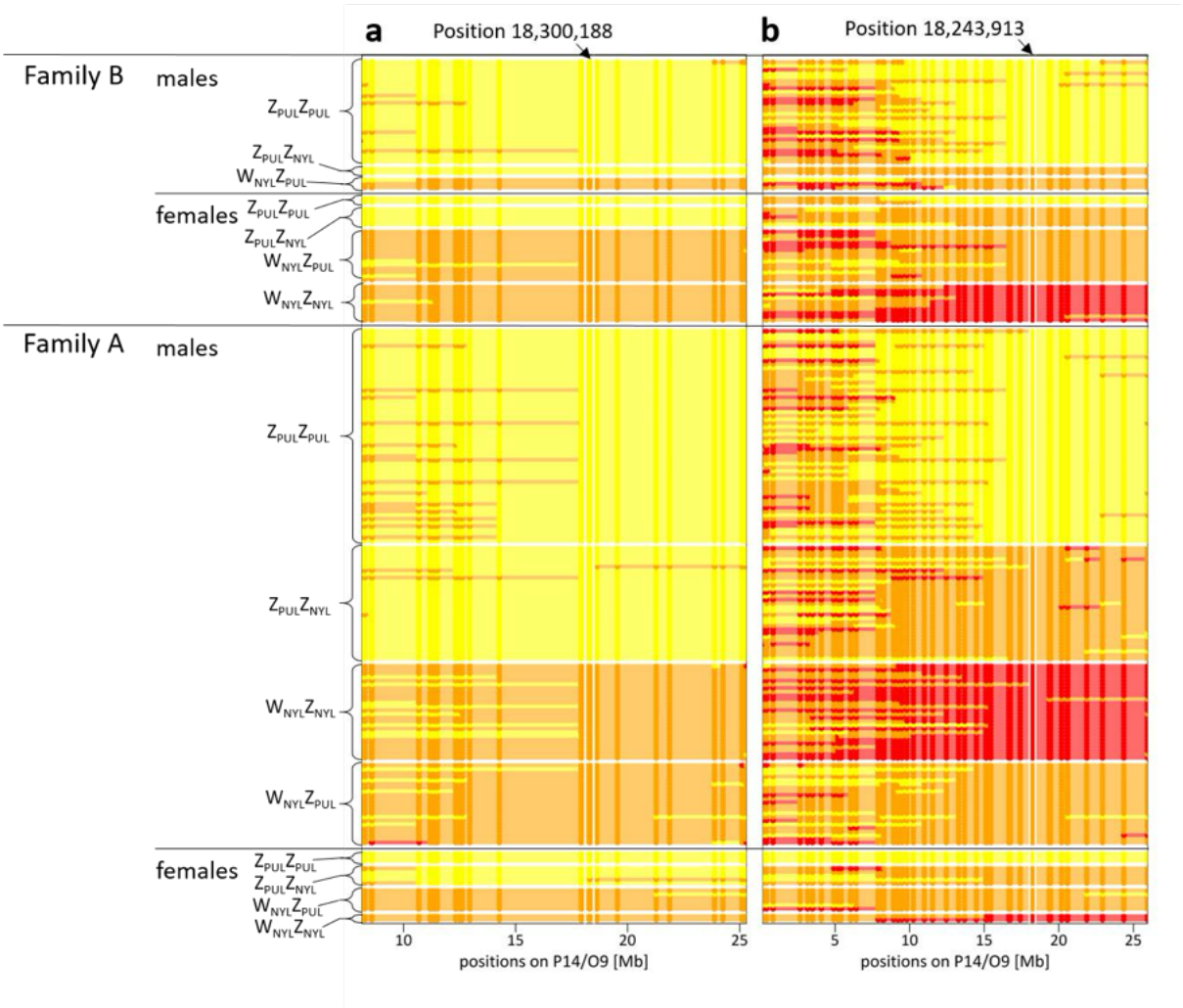


Figure S5 | Visualization of the genotypes along the sex chromosome in the Victoria1 cross.
a Sites where grandmother and mothers are heterozygous and the grandfather and fathers are homozygous. These sites indicate the presence of a W_{NYL} allele. Each individual is represented on a separate row with genotypes shown as yellow for homozygotes of grandmaternal *P.* sp. "nyererei-like" allele ($W_{NYL}W_{NYL}$), orange for heterozygotes ($W_{NYL}Z$) and yellow for ZZ. The site used for the genotype proportions in Fig. S3 is indicated with a white frame. **b** Sites where the grandparents are homozygous for alternative alleles and the parents are heterozygous. These sites allow to trace back the origin of the alleles to the parental species. The colours indicate parental species ancestry with homozygotes in yellow (NYL NYL) and red (PUL PUL) and heterozygotes (NYL PUL) in orange. The site used for Fig. S3 is framed in white.

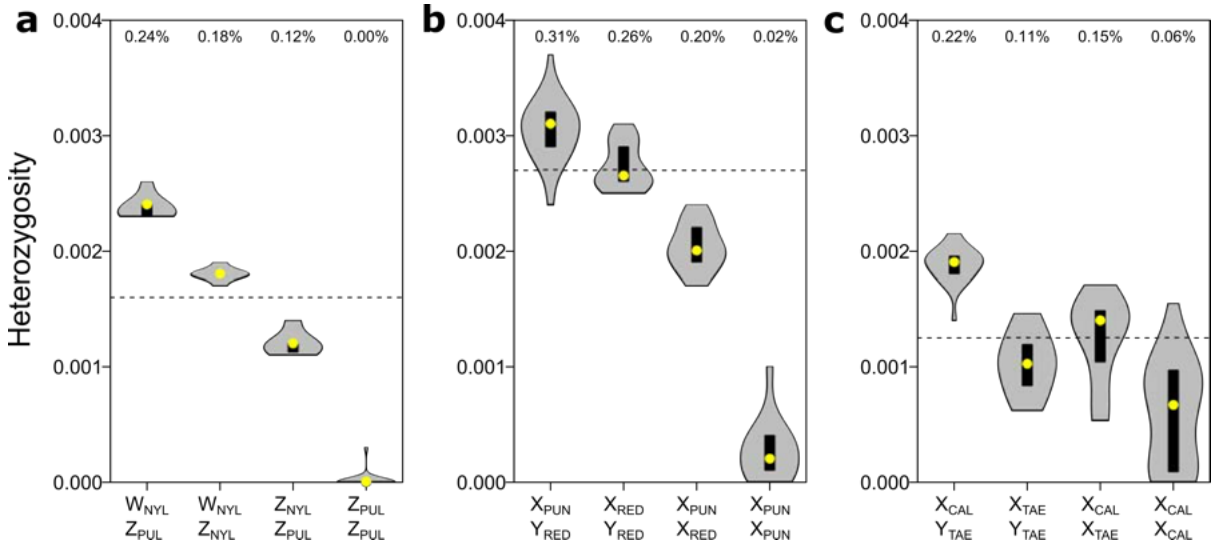


Figure S6 | Low divergence between sex chromosomes in all three crosses
Divergence between sex chromosomes was computed as heterozygosity (number of heterozygous sites among all sites sequenced) at the chromosomal region with strongest sex association. Density kernels (violin plots) show the heterozygosity distributions of all individuals with a specific combination of sex chromosomes indicated below. The median given on top is indicated as a yellow dot and the interquartile range is shown as a black box. The dashed horizontal line corresponds to the divergence (d_{xy}) between Victoria 1 and Victoria 2 individuals and thus represents a divergence time of maximum 15,000 years. **A)** Heterozygosity at the P14/O9 chromosome at 18-19 Mb in the Victoria1 cross, family B. The WZ chromosomes are derived from the *P.* sp. "nyererei-like" (NYL) grandmother or the *P.* sp. "pundamilia-like" (PUL) grandfather. The low heterozygosity in $Z_{PUL}Z_{PUL}$ individuals suggests that the mother and father both inherited the same Z chromosome from the grandfather. **B)** Heterozygosity at the P10/O23 chromosome at 30-31 Mb in the Victoria2 cross. The chromosomes are derived from the *P. pundamilia* (PUN) grandmother or the *P.* sp. "red-head" grandfather (RED). **C)** Heterozygosity at the P14/O9 chromosome at 11-14 Mb in the Malawi cross, with chromosomes derived from the *A. calliptera* (CAL) grandmother or the *P. taeniolatus* (TAE) grandfather. Note that in all three crosses, the heterozygosity is higher in heterogametic (XY/ZW) than in homogametic (XX/ZZ) individuals, and highest in heterogametic individuals with sex chromosomes from different parental species. However, the divergence between different sex chromosomes is not much greater than the divergence between the individuals of the two Lake Victoria cichlid crosses which have diverged maximum 15,000 years ago.

Table S1. QTL mapping results (only results with p-values <0.1 shown)

	#F2 m/f	marker (nearest)	P-LG	pos (cM)	95% CI	nearest flanking markers	O-LG	LOD	p-value	PVE
Victoria2	130/56	c10.loc18 (chr10_28361037)	10	18.00	16.00-25.00	chr10_27536410-chr10_28361037	23	21.58	0.00	41.39
Victoria2; add	121/56	chr10_28361037	10	17.90			23	20.20	0.00	
Victoria2; famA	76/47	c10.loc19 (chr10_28361037)	10	19.00			23	13.80	0.00	
Victoria2; famB	45/9	chr10_26890500	10	15.20			23	6.64	0.00	
Malawi	36/69	chr14_11239278	14	16.97	13.00-20.27	chr14_7158496-chr14_12920486	9	11.96	0.00	40.84

#F2 m/f, number of F2 males/females used in QTL mapping; (nearest), nearest RAD marker to a QTL at an interpolated marker (where genotypes were inferred with the calc.genoprob function in R/qtl); P-LG, linkage group number corresponding to the anchored *Pundamilia nyererei* reference; O-LG, linkage group number corresponding to the *Oreochromis niloticus* reference; pos (cM), position in centiMorgan; 95% CI, 95% approximate Bayesian credible interval (in cM); PVE, percent variance explained; add, accounting for family as additive covariate; famA and famB, separate analyses for the two families.

Chapter 4 –

Testing for a role of post-zygotic incompatibilities in rapidly speciated Lake Victoria cichlids

Testing for a role of postzygotic incompatibilities in rapidly speciated Lake Victoria cichlids

Anna F. Feller[1,2], Catherine L. Peichel[3], Ole Seehausen[1,2]

1) Division of Aquatic Ecology and Evolution, Institute of Ecology and Evolution, University of Bern, Baltzerstrasse 6, 3012 Bern, Switzerland
2) Department of Fish Ecology and Evolution, Centre of Ecology, Evolution and Biogeochemistry (CEEB), Eawag Swiss Federal Institute of Aquatic Science and Technology, Seestrasse 79, 6047 Kastanienbaum, Switzerland
3) Division of Evolutionary Ecology, Institute of Ecology and Evolution, University of Bern, Baltzerstrasse 6, 3012 Bern, Switzerland

Submitted version on bioRxiv: https://doi.org/10.1101/2023.09.26.559612

Abstract

Genetic incompatibilities can importantly contribute to the reduction of gene flow among related species that come into secondary contact before they are fully reproductively isolated. Intrinsic postzygotic incompatibilities that become expressed in hybrids are usually due to negative epistatic interactions between alleles from different parental genomes. While such incompatibilities are thought to be uncommon in classical speciation with gene flow, they could, nonetheless, be important in hybrid speciation or adaptive radiation from a hybrid population. Here we aim to identify possible incompatibilities among the endemic cichlid fish species of Lake Victoria. Hundreds of species have evolved within the lake in <15 k years from a hybrid progenitor. While the importance of prezygotic barriers to gene flow is well established in this system, the possible relevance of intrinsic genetic incompatibilities is currently unknown. Our approach is based on inferring the presence of negative epistatic interactions from systematic patterns of genotype ratio distortions in experimental crosses and wild samples. If intrinsic postzygotic incompatibilities contribute to reproductive isolation among species of the Lake Victoria radiation, unfavorable combinations of alleles should be under negative natural selection and such combinations should thus be underrepresented among different species. If incompatibilities are sufficiently strong to affect viability of early life stages, they should be underrepresented among hybrids in experimental species crosses too. We used RAD sequencing data from three interspecific crosses to scan the second-generation (F2) hybrid genomes for regions with segregation distortion, and we used whole genome sequences from 94 radiation member species to screen each pair of alleles at loci on different chromosomes for high linkage disequilibrium (LD). We then checked the position of putative incompatibility loci inferred by either of these approaches relative to regions of high genetic differentiation between sympatric sister species. We find signatures consistent with the presence of incompatibilities in the cross data as well as very high LD between physically unlinked loci in the whole genome data of wild fish, with some significant overlap between them and with regions of genetic differentiation. Next steps should include identifying genes in the putative incompatibility regions, and exploring methods to account for effects of phylogenetic structure in the LD scans.

Introduction

How reproductive isolation emerges during speciation and how it is maintained in the absence of geographic barriers is a fundamental question in evolutionary biology. Postzygotic incompatibilities -factors that reduce the fitness of hybrid offspring- can constitute a barrier to the exchange of genes (gene flow) between populations or species (Coyne & Orr, 2004). Postzygotic incompatibilities can on the one hand arise as consequence of divergent ecological adaptation, which results in hybrids that are of intermediate fitness in either parental environment and in the absence of an intermediate or alternative environment globally reduces hybrids fitness (extrinsic postzygotic incompatibilities) (Schluter, 2000; Nosil, 2012). On the other hand, postzygotic incompatibilities can arise as by-product of divergence between populations, rendering hybrids less fit independent of the environment they are in (intrinsic postzygotic incompatibilities) (Coyne & Orr, 2004). In the well-understood Bateson-Dobzhansky-Muller (BDM) model, intrinsic postzygotic incompatibilities ("BDMIs") arise due to negative epistatic interactions between alleles from different genomic backgrounds that may have fixed between populations due to drift, parallel selection or divergent selection (Bateson, 1909; Dobzhansky, 1937; Muller, 1942). Often, BDMIs may only become apparent in second or later generations of hybrids, when recessive alleles become fully expressed, referred to as hybrid breakdown (Dobzhansky, 1948; Templeton *et al.*, 1986).

BDMIs are expected to contribute to reproductive isolation between populations when they have diverged in geographic separation (in allopatry) and then come into secondary contact, as drift and selection (parallel or divergent) will over time have led to the accumulation and eventually fixation of genetic differences between the two populations, and these differences can then manifest as BDMIs in the hybrid offspring, preventing or reducing further gene flow between the populations (Orr, 1995; Turelli *et al.*, 2001; Coyne & Orr, 2004). However, BDMIs are unlikely to emerge if species diverge in geographic proximity (in sympatry) and in the presence of some gene flow, as they will be purged by negative selection in both diverging populations (Gavrilets, 2004; Bank *et al.*, 2012).

In the East African cichlid radiations, hundreds of reproductively isolated species have evolved within each of the three largest lakes (Tanganyika, Malawi, and Victoria) in close geographic proximity (reviewed in Seehausen, 2015). In the Lake Victoria haplochromine cichlid adaptive radiation alone, approximately 500 species have evolved within the lake in <15,000 years (Seehausen, 1996; Stager & Johnson, 2008; Marques *et al., in prep.*). The co-existence of up to several dozens of closely related species in any one habitat patch in the lake (Fryer & Iles, 1972; Seehausen, 1996; Turner *et al.*, 2001; Genner *et al.*, 2004) begs the question how they maintain reproductive isolation from one another. Prezygotic reproductive barriers are known to feature importantly in this system: female mate choice on the basis of male nuptial coloration plays a key role in behavioral reproductive isolation (Seehausen *et al.*, 1997; Seehausen & Van Alphen, 1998; Haesler & Seehausen, 2005; Selz *et al.*, 2014), and has likely often let to divergence in association with divergent ecological adaptations –especially along the steep gradient of light-induced ecological conditions along the water depth axis (Seehausen *et al.*, 2008). However, it remains difficult to explain how so many different species can coexist within the same water depth and the same microhabitat. Even if behavioral mate choice is an important feature in many of them (Seehausen *et al.*, 1998), the question remains how it can evolve so rapidly and persist despite some ongoing gene flow (Meier *et al.*, 2018), especially since theoretical studies suggest that in sympatry, disruptive natural selection would have to be unrealistically strong to couple

mate choice genes to ecological adaptation genes in the face on gene flow (Felsenstein, 1981; van Doorn *et al.*, 2004; Butlin *et al.*, 2021).

Recent studies have shown that, while all Lake Victoria cichlid species are very young (Meier *et al.*, 2017a; McGee *et al.*, 2020; Marques et al., *in prep.*), their genomes are mosaics containing much older genetic variants from at least two ancestral lineages (Meier *et al.*, 2017a). The latter authors have shown that the Lake Victoria radiation was seeded by a hybrid population that arose from two quite deeply divergent species, which fuelled the radiation by providing large amounts of genetic variation that could then be sorted into many new combinations. The divergence time between the ancestral hybridizing lineages that seeded the Lake Victoria region superflock (LVRS) falls into the range in which we would expect incompatibilities between them to segregate (Stelkens *et al.*, 2010, 2015; Meier *et al.*, 2017a). Meier *et al.*, 2017a indeed also found patterns consistent with differential sorting of BDMIs in the LVRS: sites that were fixed for alternative alleles in the two ancestral lineages were enriched for outlier loci in Lake Victoria. In the initial hybrid swarm (a mixture of genomes from at least two divergent ancestral lineages) that seeded the Lake Victoria radiation, BDMIs may thus have been present at high frequencies. The hybrid population would then on the one hand have suffered fitness losses from the presence of BDMIs, but on the other hand they carried large amounts of genetic variation at ecologically relevant loci (and at loci that affect mating traits and mating preferences), which allowed diversification into many different niches. If BDMIs then became coupled to polymorphisms maintained by negative frequency-dependent or divergent ecological selection, this could have facilitated rapid, repeated and sympatry-robust speciation (Seehausen, 2013). We thus hypothesised that BDM incompatibilities might play a role for speciation with gene flow from a hybrid population.

Here, we combine three different genomic approaches and datasets to search for signatures consistent with the presence of intrinsic postzygotic incompatibilities (BDMIs) in the Lake Victoria radiation. First, we study segregation patterns in three laboratory F2 experimental crosses between different Lake Victoria cichlid species using restriction site-associated DNA (RAD) marker sequencing to scan hybrid genomes for regions with segregation distortion. If there are alleles in the grandparents that do not work well together and cause early life stage mortality when combined in the hybrids, the absence of the genotypes combining these alleles will result in distorted genotype ratios. We also test if there was selection for increased heterozygosity in the F2 hybrids, as predicted under Fisher's geometric model if the two parental lineages feature co-adapted alleles/genes (Simon *et al.*, 2018). For comparison we also analyzed a fourth cross between two much older allopatric Lake Malawi cichlid species, between which BDM incompatibilities would be strongly expected due to their age and geographic isolation. Second, we use whole genome sequences from 94 species representing nearly all ecological guilds in the Lake Victoria radiation (1 genome per species) to screen each pair of alleles at loci on different chromosomes for signatures of high linkage disequilibrium among these 94 species ("multispecies LD"). Unfavorable allelic combinations should have been removed by natural selection and should thus be underrepresented among the different species (or the favorable combinations overrepresented, respectively). Third, we check the position of putative incompatibility loci relative to highly differentiated regions (based on FST outlier analyses) between sympatric sister species pairs. For the latter we use whole genome sequences of 11 different species pairs with 3-5 individuals per species). Each analysis on its own has significant limitations: 1) limited sample sizes in the F2 hybrids leads to low power for detecting subtle deviations in genotype ratios such as those caused by early-acting BDMIs, and thus, our expectation was not to detect all possible incompatibilities with these analysis; 2), the

multispecies LD analyses may be confounded by drift and phylogenetic inertia. However, if we find regions or loci that appear in both analyses and that overlap with regions of high genetic differentiation between closely related species, this should provide evidence for the presence and relevance in speciation of BDM incompatibilities among these species that evolved from a hybrid swarm in the presence of gene flow.

Materials and methods

RAD sequencing data from second-generation (F2) hybrid crosses

The three crosses between Lake Victoria cichlid species were between *Pundamilia* sp. 'nyererei-like' and *Pundamilia* sp. 'pundamilia-like' (see also Feller *et al.*, 2020a, 2021), between *Pundamilia pundamilia* and *Pundamilia* sp. 'red-head' (see also Feulner *et al.*, 2018; Feller *et al.*, 2020a, 2021), and between *Pundamilia* sp. 'nyererei-like' and *Neochromis omnicaeruleus* (see also Feller & Seehausen (*submitted*)). The Lake Malawi cross was between *Astatotilapia calliptera* (Chizumulu) and *Protomelas taeniolatus* (see also Stelkens *et al.*, 2009; Selz & Seehausen, 2019; Feller *et al.*, 2020b). See the listed references for details on breeding, rearing and processing protocols.

The sequencing and genotyping procedures for all four crosses are described in full in (Feller *et al.*, 2021) and in Feller & Seehausen (*submitted*). In brief, we generated 100-150 bp single end RAD sequencing libraries following Baird *et al.*, 2008, each containing 48-50 barcoded individuals that were then demultiplexed, quality-filtered, aligned to the anchored version of the *Pundamilia nyererei* reference genome (Feulner *et al.*, 2018), and genotyped by cross.

We filtered all four resulting genotype (vcf) tables with the procedure described in Feller & Seehausen (*submitted*). In summary, we used bcftools (samtools/1.9; Li *et al.*, 2009) and vcftools/0.1.16 (Danecek *et al.*, 2011) to filter for bi-allelic SNPs with <50% missing data (genotypes with depth or quality <10 set to missing), a minor allele frequency (MAF) of >0.05, and a maximum sequencing depth of less than 1.5 times the interquartile range from the mean. Individuals with >50% missing data, a mean depth of <10, or indications of high amounts of PCR duplication were excluded in the process. We then subset these sets of quality-filtered SNPs to sites that were alternatively homozygous fixed in the F0 grandparents and heterozygous in the F1 parents ("fixed sites"; note however that some F0 and F1 individuals were excluded or not available; see Table 1, and Feller *et al.*, 2021 for more information). Finally, we used beagle 5.1 (Browning *et al.*, 2018) with standard parameters to impute genotypes for the second-generation hybrids (F2s) at all fixed sites, excluding unmapped scaffolds. Table 1 shows an overview of the numbers of SNPs and individuals per cross (base sequencing depth and quality were high in all four sets (DP >500, QUAL >998)).

Table 1 Overview of the number of individuals and SNPs in each hybrid cross			
Cross	**number of individuals** F0/F1/**F2** (m,f)	**all filtered SNPs**	**"fixed sites" (excluding unmapped scaffolds)**
P. sp. 'nyererei-like 'x *P.* sp. 'pundamilia-like'	2†,/ 4 /**218** (171,47)	10,520	1,272
P. pundamilia x *P.* sp. 'red-head'	2/ 4†/**186** (129,57)	9,907	1,873
P. sp. 'nyererei-like' x *N. omnicaeruleus*	2/11†/**161** (124,27)	16,604	2,187
A. calliptera x *P. taeniolatus*	12†/ 0 /**114** (36,60)	12,024	1,317
†due to uncertainties in some F0/F1's genotypes, only the F0 grandmother was used in *P.* sp. 'nyererei-like' x *P.* sp. 'pundamilia-like'; only two of the four F1s in *P. pundamilia* x *P.* sp. 'red-head'; only three of the four F1s in *P.* sp. 'nyererei-like' x *N. omnicaeruleus*; and in *A. calliptera* x *P. taeniolatus*, the F0s were not available and several parental species individuals were used instead (see Feller *et al.*, 2020b).			

Testing for selection for increased heterozygosity in the F2 hybrids

We used the --het function in vcftools to output tables of homozygous genotype counts and the number of genotyped (among fixed) sites on a per-individual basis. From this, we calculated the proportion of heterozygote genotype counts among genotyped sites for each individual in R (R Core Team, 2020). Following the approach of Simon *et al.*, 2018, we used Wilcoxon's signed rank test (wilcox.test function in R) to test if the distribution of heterozygosity in the F2s was symmetrically distributed around the (Mendelian) null expectation of mu=0.5.

Screening for regions with segregation distortion in the F2 hybrids

To identify and extract regions of segregation distortion in the F2 hybrids we used three complementary approaches. For all three, we subset each cross dataset (the set of fixed sites with imputed genotypes) to an equal number of males and females by randomly selecting individuals of the sex with the higher number of individuals to exclude (using the 'shuf' command in bash).

First, we inspected deviations in allele frequencies across the hybrid genomes from the Mendelian expectation of 0.5. We used the --freq function in vcftools to output per-site allele frequencies. The limited amount of recombination events in an F2 cross make it unlikely that one isolated SNP represents a true deviation rather than a sequencing or genotyping error if its neighboring SNPs do not show the same kind of deviation. Thus, we retained only regions where at least three consecutive SNPs that were each no further than 6.4 Mb away from the following SNP had a major allele frequency of >0.6 of the same grandparental origin. (6.4 Mb is the mean of the average distances at which LD decays to <0.5 within chromosomes in the four crosses.)

Second, we inspected genotype distortions. We used the --hardy function in vcftools to output genotype counts. In R, we tested each SNP for segregation distortion by applying a Chi-square-test. Then, we subset SNPs with significant segregation distortion (Chi-square test-value of >4.605, i.e. a p-value of <0.1) to those where the genotype ratios approximately conformed to the patterns expected if they were involved in a two-locus (recessive) incompatibility with another SNP. That is, one homozygous genotype should be reduced by 1/16 while the other homozygous genotype and the heterozygous genotypes should each be increased by 1/32. We implemented this in R such that the counts of the less frequent homozygous genotype had to be between 0.6-0.8 times that of the more frequent homozygous genotype, and the counts of the more frequent homozygous genotype had to be between 0.3-0.7 times that of the heterozygous genotypes. As in the first approach, we then extracted regions where at least three consecutive SNPs that were each no further than 6.4 Mb away from the following SNP showed this pattern of segregation distortion.

Third, we screened for regions with locally increased heterozygosity. Instead of subsetting the SNPs to those with significant 'two-locus-incompatibility'-patterns as above, we subsetted them to those with excess heterozygous genotype counts -increased by 10% or more- while both homozygous genotype counts were approximately equally reduced (i.e. the count of one homozygous type had to be min. 0.75 or max. 1.25 times that of the other). And again, we extracted regions where at least three consecutive SNPs that were each no further than 6.4 Mb away from the following SNP showed this pattern of heterozygosity distortion.

For analyses of overlap with FST outlier windows and high LD pairs (see below), we compiled all three extracted types of distorted regions into one set of distorted regions for each cross.

Inspection of pairs of segregation distorted SNPs

To check if any two of the SNPs that conformed to the 'two-locus-incompatibility' pattern together behaved like a classic BDM incompatibility' pair (that is, the two underrepresented homozygous genotypes should not occur together) we generated contingency tables in R and inspected the genotype combinations for every SNP that conformed to the 'two-locus-incompatibility' pattern with every other such SNP on a different chromosome.

Whole genome sequencing data

Sample collection and whole genome re-sequencing are described in Meier *et al.*, 2017a and Marques *et al.* (*in prep.*). Marques *et al.* (*in prep.*) aligned the reads to the anchored *P. nyererei* reference genome with bowtie2 v2.2.3 (Langmead & Salzberg, 2012), and used GATK v3.5's HaplotypeCaller (McKenna *et al.*, 2010) to call variants and genotypes. The resulting vcf tables were filtered by Marques *et al.* (*in prep.*) with bcftools, removing genotypes with a depth of <6 and/or quality of <20, only keeping bi-allelic sites with a maximum of 50% missing data, and sites were excluded if they overlapped with one of three masks (regions with more than one 35-kmer self-mapping or high repeatability in the reference genome, and regions in which more than 30 of the 400+ genomes had a depth that exceeded the value of mean depth + 1.5 * inter-quartile range) (see Marques *et al.* (*in prep*) for details and used tools). We then used a subset of the genomes in these SNP tables and further filtered them as described below.

Multispecies Linkage Disequilibrium (LD) analyses across the radiation

To assess non-random allele associations across the species of the Lake Victoria radiation ("multispecies LD") we used one male individual of 94 whole genome sequenced Lake Victoria species (see Table S1). We subset the whole genome SNP tables to these 94 individuals using bcftools, and then filtered out sites with more than 5% missing data, a minor allele frequency of <0.05, and a depth exceeding the mean by 1.5 times the inter-quartile range. This resulted in a set of 2,231,161 SNPs. We used vcftools to generate the input files for plink/1.90 (www.cog-genomics.org/plink/1.9/) (Chang *et al.*, 2015), which we then used to output LD statistics for all locus pairs, only outputting pairs with an r^2-value of >0.2 (settings --r2 'inter-chr' 'with-freqs' --ld-window-r2 0.2). We further removed pairs where both SNPs had a MAF of <0.1 and a highly similar ratio between MAF at locus A and MAF at locus B (i.e. a ratio between 0.8-1.2), because locus pairs in high LD with a very low MAF and equal MAFs are most likely to represent a phylogenetic signal (see also Fig. S1). We performed all subsequent analyses on inter-chromosomal pairs with an r^2-value of >0.9 and with an r^2-value of >0.7.

FST landscapes and outliers

For FST analyses among sister species of Lake Victoria cichlids we subset the whole genome SNP tables to the 28 species listed in Table S2, for each of which we had genome sequences of two to five male individuals (a total of 108 samples; see Table S3 for the list of pairs and numbers of individuals in each). We then removed genotypes with a depth of <10 and quality of <20, and filtered out sites with more than 25% missing data, a depth of <10, and a depth exceeding the mean by 1.5 times the inter-quartile range. This resulted in a set of 47,226,072 SNPs.

Within each species pair we then only kept sites with no missing data and we filtered out sites with a minor allele count of <3, and we calculated Weir and Cockerham's weighted FST as implemented in vcftools in 50 kb non-overlapping windows. Windows with less than 10 SNPs were subsequently removed. We chose this approach of using fixed non-overlapping windows because it makes positions of windows directly comparable between all analyzed species pairs, and avoids pseudoreplication in outlier detection. (The caveat being that the species pairs will differ in the number and location of missing windows). For each species pair we identified the top 5% and top 1% outlier windows in R.

The FST analysis for the Lake Malawi cross species pair was modified slightly because only RAD sequencing data was available (for five individuals each). In the filtering step we increased the missing data threshold to max. 30% (resulting in a set of 138'036 SNPs), we calculated FST in 10 kb windows, and we retained all windows with 3 or more SNPs.

Overlaying the three datasets

Regions with segregation distortion among experimental crosses vs FST outliers

For each experimental cross, we quantified the number of top 5% and top 1% FST outlier windows that were located in a distorted region. To assess if the number of outlier windows found in distorted regions is significantly higher than expected by chance, we randomly shuffled both the outlier state ('top 5%' or 'top 1%' vs not an outlier) and the location state ('in distorted region' vs not) among all FST windows 10,000 times, and each time counted the number of outlier windows that appeared in a distorted region. We considered the observed number of outlier windows in distorted regions to be significant if no more than 500 out of the 10,000 iterations produced the same or a higher count (empirical p-value of <0.05). For each experimental cross we performed this analysis across all chromosomes and also only within chromosomes that carry a distorted region.

Regions with segregation distortion vs high multispecies LD

We checked if the position of some of the loci involved in very high deviations from random associations with another locus (multi-species LD, r^2 >0.8 or r^2 >0.7) across the 94 species genome LD scans coincided with distorted regions observed in the three experimental Lake Victoria species crosses. To test if overlap was significantly more than expected by chance, we randomly shuffled the distortion regions (keeping their size constant, and allowing no overlap between them) among all chromosomes 1,000 times, and every time counted how many of the SNPs that are member of a high multispecies LD pair ended up in a randomly placed distorted region. We considered the observed number of high multispecies LD SNPs in distorted regions as significant if no more than 50 out of the 1,000 iterations produced the same or a higher count (empirical p-value of 0.05).

High multispecies LD vs FST outliers of sympatric species pairs

To assess if SNPs involved in high multi-species LD significantly coincided with FST outlier windows between sympatric sister species, we used the same approach as when comparing overlap of segregation distortion regions among crosses with FST outliers between the same species (see above), shuffling both the outlier state ('top 5%' vs not an outlier) and the long distance LD state ('contains high LD SNP' vs not) of all FST windows 10,000 times, and each time counting the number of outlier windows that contained a high LD SNP. We performed this analysis for 11 species pairs (see Table S3), and for the LD SNP pairs with an r^2-value of >0.8, as well as for the LD SNP pairs with an r^2-value of >0.7. We adjusted the resulting p-values using fdr correction.

Results & Discussion

No evidence of selection for increased heterozygosity in the F2 hybrids

Mean heterozygosity in the F2 hybrids of all four crosses was either non-significantly different from the null expectation of 0.5, or biased towards an overall lower level of heterozygosity (Table 2). The downward biases should probably not be over-interpreted. In the case of *P.* sp. "nyererei-like" x *P.* sp. "pundamilia-like", the downward bias might have been caused by the fact that we could only use the F0 grandmother and the F1s to subset to homozygous fixed SNPs. Hence, not all of the included SNPs may have been truly fixed between the F0 grandparents, and differing genotype combinations would generate different proportions of heterozygosity levels among the F2s.

Selection for increased heterozygosity is predicted for F2 hybrids if the parental lineages feature intrinsic co-adaption among genes (Simon *et al.*, 2018). The absence of increased heterozygosity in the four crosses suggests that there are not as many strong pairwise incompatibilities between these species that they would cause strong selection for genome-wide heterozygosity in the F2s. This might not be too surprising in the three Lake Victoria crosses given the young age of this radiation and that the levels of fixation among populations in this system are rather low (Meier *et al.*, 2017a, 2018). The absence of increased heterozygosity in the Malawi cross is more surprising since the two crossed species are two orders of magnitude older than the Lake Victoria species. We would thus have expected more incompatibilities to be present in this cross since they had more time to accumulate them, and accumulation rates might even increase in a snowball fashion (Orr & Turelli, 2001; Presgraves, 2010) (but see e.g. Gourbière & Mallet, 2010; Stelkens *et al.*, 2010). In fact, Stelkens *et al.*, 2015 found significantly reduced survival in F2 hybrids of this cross.

Table 2 Results of Wilcoxon's signed rank test to test of distribution of heterozygosity in the F2s is symmetrically distributed around the null expectation of mu=0.5	
Cross (approximate divergence age)	**Results of Wilcoxon's signed rank test**
***P.* sp. 'nyererei-like' x *P.* sp. 'pundamilia-like'** (<200 years)	V = 8311.5, p-value = 0.0001*** 95% confidence interval: 0.4579 0.4859 (pseudo) median: 0.4721; *mean 0.4682*
***P. pundamilia* x *P.* sp. 'red-head** (<15k years)	V = 8949.5, p-value = 0.7303 95% confidence interval: 0.4901 0.5144 (pseudo) median: 0.5021; *mean 0.5091*
P.* sp. 'nyererei-like' x *N. omnicaeruleus (<15k years)	V = 6145, p-value = 0.5268 95% confidence interval: 0.4831 0.5087 (pseudo) median: 0.4959; *mean 0.4952*
A. calliptera* x *P. taeniolatus (<800k years)	V = 2514.5, p-value = 0.03109** 95 % confidence interval: 0.4681 0.4985 (pseudo)median :0.4840, *mean 0.4860*
Divergence times based on (Meier *et al.*, 2017a; b) for the Lake Victoria crosses, and on (Malinsky *et al.*, 2018) for the Lake Malawi cross.	

The number of regions with segregation distortion in hybrids increases with divergence time between the parental species

In the youngest cross (*P.* sp. 'nyererei-like' x *P.* sp. 'pundamilia-like') we detected a total of three distorted regions, two on linkage group (LG) 16 and one on LG18 (Fig. 1a). In the second cross within the genus *Pundamilia* (*P. pundamilia* x *P.* sp. 'red-head'), we detected a total of seven distorted regions, one each on LGs 1, 8, 13, 14, 15, 21, 22 (Fig. 1b). In the third Lake Victoria cross (*P.* sp. 'nyererei-like' x *N. omnicaeruleus*) we also detected a total of seven distorted regions, two on LG6, and one each on LGs 3, 4, 5, 10, 21 (Fig. 1c). In the Lake Malawi cross we found a total of 15 distorted regions on 14 out of the 22 LGs (Fig. 1d). See Figs. S2-4 for all three assessed types of distortion plotted separately.

Hence, with increasing divergence time between the crossed species we found more regions with signatures of segregation distortion that could indicate the presence of incompatibilities. This would be consistent with a scenario of sorting of incompatibilities in the course of adaptive radiation from a hybrid population, where most incompatibilities become divergently sorted between the first species that arise, leaving fewer to segregate and become divergently fixed in successively later speciation events (as in the verbal model by Seehausen, 2013).

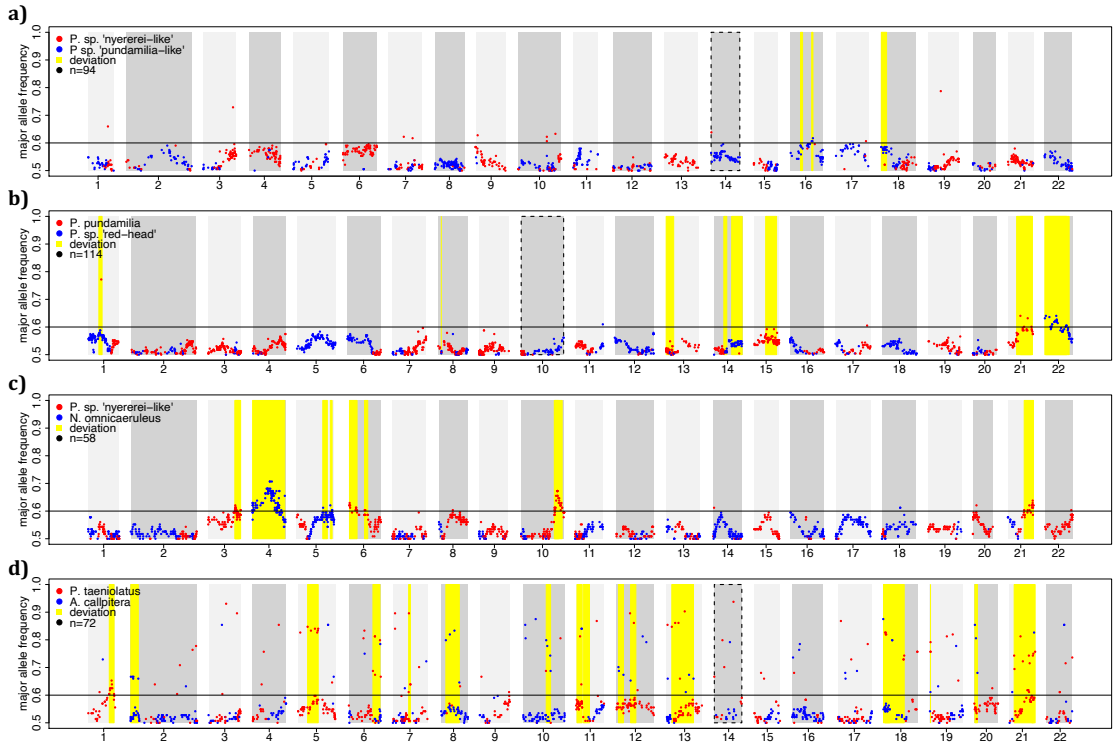


Figure 1 | The number of regions that are distorted (deviating allele frequency, 'classic' segregation distortion, or an excess in heterozygous genotypes) in interspecific cichlid hybrids increases with increasing divergence time of the crossed species (from top to bottom). The three types of distortion are highlighted separately in Figs. S2-4. LGs known to be involved in sex determination (Feller *et al.*, 2021) are framed with a dashed line. Grandparental origin of major allele according to blue/red colour in legend. Equal numbers of males and females were used in each cross (total number of individuals indicated for each cross as "n=" in legend).

Locus pairs with segregation distortion in crosses do not fully conform to classic recessive two-locus incompatibility patterns

None of the inter-chromosomal SNP pairs in any of the experimental crosses showed the exact pattern of a BDM incompatibility. In *P.* sp. 'nyererei-like' x *P.* sp. 'pundamilia-like', none of the 20 inter-chromosomal SNP pairs were lacking either direction of homozygous genotype combinations (aa/bb or bb/aa). In the other three crosses, one of the homozygous genotype combinations was missing but not the other in several pairs (see Tables S4-S6). One issue here is the lack of power to detect the subtle deviations we screened for. Given our modest equal sexes sample sizes, power to detect a deviation of e.g. 1/16 was as low as 0.15-0.27. Hence, we certainly could not detect all deviations that might indicate the presence of such an incompatibility. Moreover, incompatibilities might arise from the combination of incompatible (derived) alleles at more than two loci (Satokangas *et al.*, 2020) and not all of these may have to be fixed in the parental species (Cutter, 2012). Our screens are thus rather conservative in that they only consider the 'classic' two-locus BDMIs.

FST outlier windows are significantly associated with regions with segregation distortion in two experimental crosses

Four top 5% FST outliers between the very young sister species *P*. sp. 'nyererei-like' and P. sp. 'pundamilia-like' overlapped with segregation distorted regions in our experimental *P*. sp. 'nyererei-like' x P. sp. 'pundamilia-like' hybrids (Fig. 2a), but, this was not more than expected by chance based on permutations (empirical p-value = 0.830). In *P. pundamilia* x *P.* sp. 'red-head' a total of 70 top 5% outliers (of which 21 were top 1% outliers) overlapped with segregation distorted regions (Fig. 2b). The top 1% outlier windows were significantly more often associated with a distorted region than by chance across the whole genome (empirical p-value =0.029), and we found significant associations of top 5% outliers with distorted regions within four chromosomes (LG1, LG13, LG14, LG21; empirical p-values <0.05). On LG13 and LG14, top1% outlier windows were also significantly associated with distorted regions (empirical p-values <0.05). In *P*. sp. 'nyererei-like' x *N. omnicaeruleus*, with a total of 45 top 5% outliers in a region with distortion (of which 5 top 1%; Fig. 2c), outlier windows were not significantly more often associated with a distorted region than by chance across the whole genome (empirical p-value =0.99), but on LG21, top 5% outliers were significantly more often associated with a distorted region than by chance (empirical p-value =0.02). In the Malawi cross, the coincidence of 7 top 5% outlier windows (none of which top 1%; Fig. 2d) with a distorted region was also not more than expected by chance (empirical p-value =0.198). This latter result may not be surprising because genomic differentiation between these two much older species is considerably higher and the power to detect any outlier FST and hence also possibly coincidence with a distorted region will thus be decreased.

One complicating factor in this analysis is the difference in resolution between the two datasets. The distorted regions that one can obtain from second generation hybrids (with a limited number of recombination events) are very large (often several Mb), such that they will inherently contain many of the smaller 50 kb outlier windows, which will then often not be more than expected by chance (i.e., power to detect significant associations is low).

We found almost no overlap between the distorted regions that we detect in the three Lake Victoria species crosses (Fig.2). This could on the one hand at least partially be due to the lack of power to detect such regions. Furthermore, the segregating variation we observe in experimental crosses is only ever the differences between two individuals, and might not reflect all the differences between species when these are not completely fixed. On the other hand, in a scenario where incompatibilities that were initially segregating in a hybrid swarm (such as the one from which the Lake Victoria radiation evolved (Meier *et al.*, 2017a)) were then sorted into many new compatible combinations, leading to reproductive isolation between several newly evolving species (Seehausen, 2013; Meier *et al.*, 2017a), we might expect different combinations of BDMIs to act between different new species.

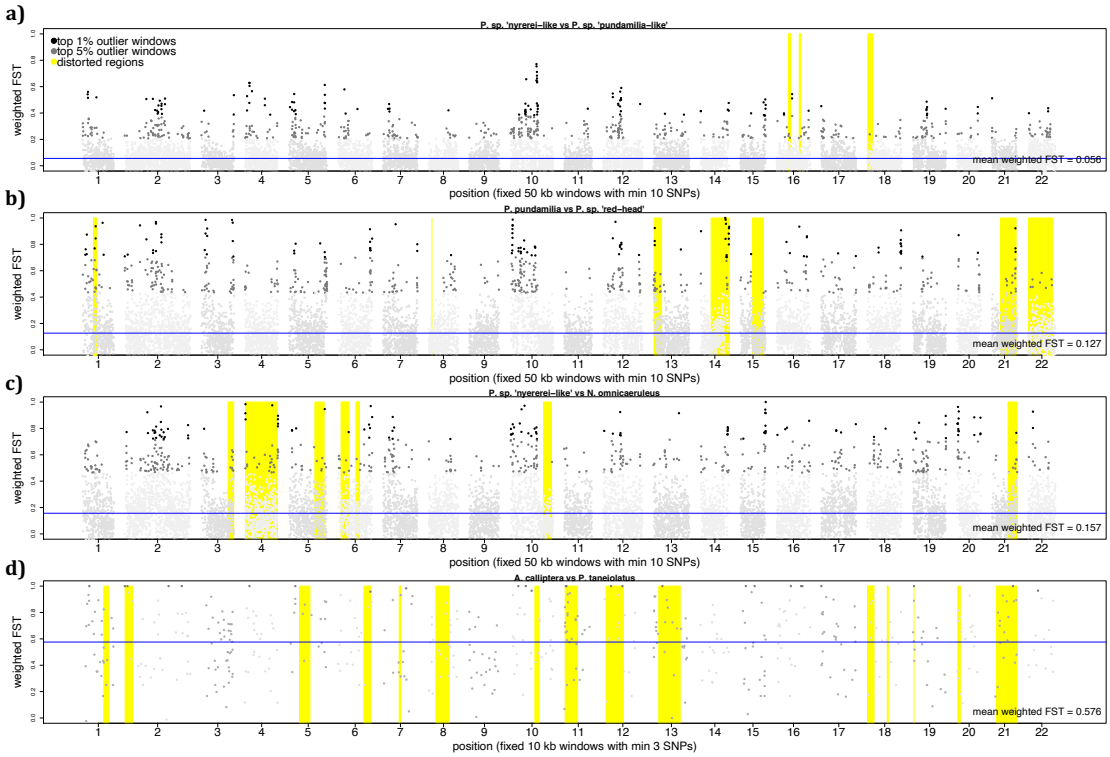


Figure 2 | FST outlier windows are significantly associated with regions with segregation distortion within four LGs in *P. pundamilia* x *P.* sp. 'red-head' (as well as across the whole genome in this cross) and one LG in *P.* sp. "nyererei-like" x *N. omnicaeruleus.*

Strong LD patterns

Of the ~2.37 trillion (!) inter-chromosomal SNP pairs in the dataset consisting of one genome each of 94 Lake Victoria haplochromine cichlid species, 426 million pairs (0.02%) had an r^2-value of >0.2 (Fig. 3 and Table 3). Many thousands of pairs had an even higher r^2-value, with the count falling below one thousand only at an r^2-value of >0.9. These numbers may seem surprising at a first glance, especially when comparing percentages of inter-chromosomal and intra-chromosomal pairs with an r^2-value of 0.2 (see Table 3), and in comparison to other studies that have used inter-chromosomal LD to screen for incompatibilities (e.g. Payseur & Place, 2007; Corbett-Detig *et al.*, 2013; Schumer *et al.*, 2014; but see Hohenlohe *et al.*, 2012). However, the number of SNP pairs analyzed here is unprecedented, and unlike the cited studies, we did not yet include ancestry information in our analyses, so direct comparisons are not trivial to make.

The number of pairs with an r^2-value <0.2 however still make up the vast majority in our case (see Fig. 3a). One explanation for the not much higher percentage of intra-chromosomal pairs with r^2>0.2 compared to the percentage of inter-chromosomal pairs with r^2>0.2 (see Table 3) might be that not much physical linkage within chromosomes has been preserved in this radiation due to high levels of recombination (i.e. effective population size in the hybrid population and the emergent radiation must have been high).

What has to be considered though, is that several distinct scenarios could generate the same or similar patterns of high inter-chromosomal LD, and that the presence of intrinsic postzygotic incompatibilities is only one of them. By using one individual per species we avoided

including effects of population structure and species structure (i.e. speciation) on the LD measure, and thus, any LD we detected in our analysis is due to non-random allele associations shared between multiple species. Effects of shared evolutionary history between species (i.e. 'phylogenetic structure') as well as ecological selection shared between species with similar ecologies are currently not accounted for in our analysis.

Effects of phylogenetic structure could reflect shared BDMIs that were sorted in the speciation event that led to the common ancestor of a clade, but any divergent fixation by drift or divergent selection during the origin of the ancestor would result in shared LD between clade members and species in other clades. Pairs involving SNPs with a MAF <0.1 and with highly similar MAFs are especially likely to represent effects of such shared phylogenetic history, which is why we additionally excluded those kind of pairs for subsequent analyses. Because there is only very mild phylogenetic structure among the cichlid species of the Lake Victoria radiation (Bezault *et al.*, 2011; McGee *et al.*, 2020), we do not expect large effects on our estimates of inter-chromosomal LD. Nonetheless, to fully account for any such effects might require an approach that could specifically account for it, such as the one proposed in Mangin *et al.*, 2012. Effect of shared ecological selection on the other hand, is not unlikely: Lake Victoria cichlid species fall into many distinct ecological guilds (see Table S1), and alleles at adaptation genes that differentially fixed under divergent ecological selection between members of different guilds will then introduce inter-chromosomal LD shared between several species.

One interesting observation is that many of the SNPs are involved in more than one high multispecies LD pair, which could be indicative of higher order epistatic interactions (see Table 3 and Figs. 4 and 5).

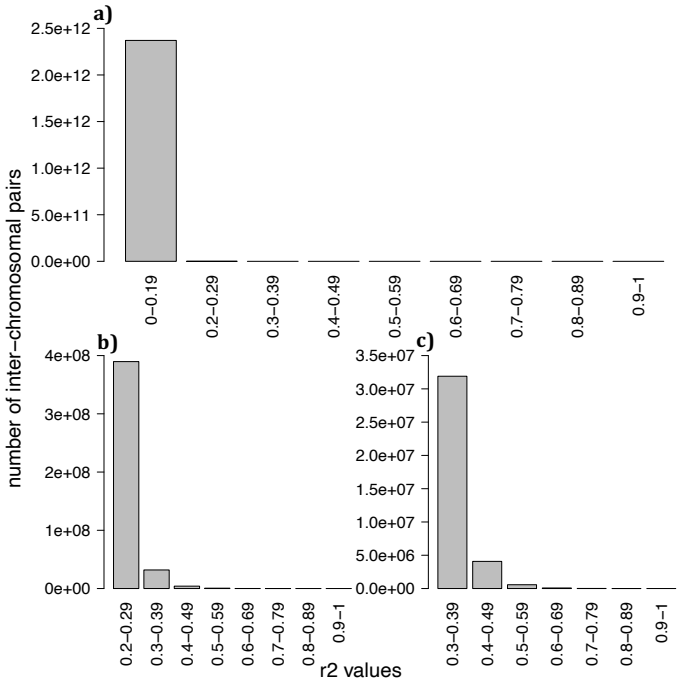


Figure 3 | Distributions of r^2 values among all inter-chromosomal pairs (original set with no additional MAF filters). **b)** and **c)** each show subsets of the full dataset shown in **a)** for better visibility

Table 3 Results of LD scan including one male individual of 94 Lake Victoria species

total number SNPs	2,231,161	
total number of pairs	2,489,038,588,380	
r^2 >0.2 total	462,706,931 *(0.0186 %)*	
number of pairs intra-chr.	117,858,505,472	
r^2 >0.2 intra-chromosomal	36,387,114 *(0.0309 %)*	
number of pairs inter-chr.	2,371,180,082,908	
		with additional MAF filters
r^2 >0.2 inter-chromosomal	426,319,817 *(0.0180 %)*	273'879'253
r^2 >0.3 inter-chromosomal	36,640,000	15'892'766
r^2 >0.4 inter-chromosomal	4,744,073	1'370'345
r^2 >0.5 inter-chromosomal	665,352	146'605
r^2 >0.6 inter-chromosomal	95,036	16'898
r^2 >0.7 inter-chromosomal	14,374	† 3'098
r^2 >0.8 inter-chromosomal	2,241	‡ 963
r^2 >0.9 inter-chromosomal	546	394

† and ‡ were the pairs used in subsequent analyses; † these pairs are made up of 1332 unique SNPs; ‡ these pairs are made up of 430 unique SNPs; i.e. SNPs can be involved in several pairs

Some overlap between the location of loci in high multispecies LD and regions with segregation distortion in one of three experimental crosses

In the *P. pundamilia* x *P.* sp. 'red-head' cross, one or both SNPs belonging to locus pairs in high multispecies LD on both the set with r2>0.8 and r2>0.7 values coincided with a distorted regions more often than expected by chance (empirical p-value <0.05; Fig. 4, Table 4).

Table 4 Results of testing coincidence of loci in high multispecies LD with regions of segregation distortion in hybrid crosses

cross	**r2>0.8**					
	count1	pval1	count2	pval2	count3	pval3
P. sp. 'nyererei-like' x *P.* sp. 'pundamilia-like'	2	0.227	0	1.001	0	1.001
P. pundamilia x *P.* sp. 'red-head'	246	0.203	548	0.01**	242	0.029*
P. sp. 'nyererei-like' x *P. omnicaeruleus*	0	1.001	16	0.645	0	1.001
	r2>0.7					
	count1	pval1	count2	pval2	count3	pval3
P. sp. 'nyererei-like' x *P.* sp. 'pundamilia-like'	78	0.09	83	0.122	0	1.001
P. pundamilia x *P.* sp. 'red-head'	404	0.311	138	0.001***	270	0.025*
P. sp. 'nyererei-like' x *P. omnicaeruleus*	115	0.672	156	0.604	0	1.001

count1 is the number of times the first SNP in a high multispecies LD pair was located in a distorted region; count2 is the number of times the second SNP in a high multispecies LD pair was located in a distorted region; count3 is the number of times both SNPs of a high multispecies LD pair were located in a distorted region.
The corresponding empirical p-values were determined by randomly placing the distorted region in the genome 1000 times and each time assessing the overlap with high multispecies LD SNPs. Significance values (of the here un-corrected p-values) are *p<0.05, **p<0.01, ***p<0.001.

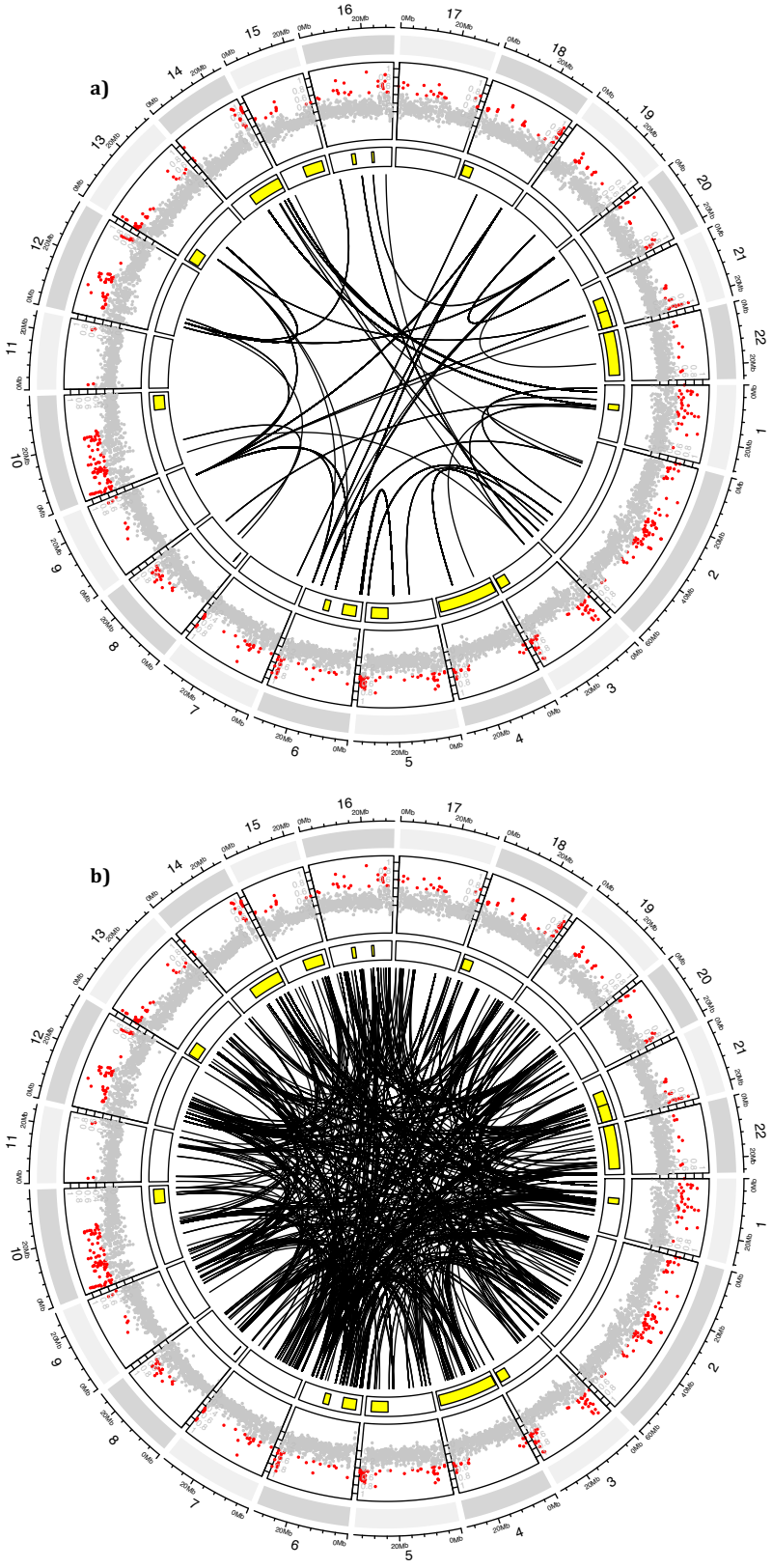


Figure 4 | Circos plot showing the 22 chromosomes (outermost tract), the FST landscape of *P. pundamilia* x *P.* sp. 'red-head' (middle tract; top 5% outliers highlighted in red), regions with distortion (inner tract, highlighted with yellow) of all three Lake Victoria crosses combined, and pairs in high multispecies LD (black connecting lines). **a)** r^2>0.8 LD pairs **b)** r^2>0.7 pairs (both with additional MAF filters applied (see Methods)).

Loci in high multispecies LD and FST outliers significantly coincide in ten out of eleven sympatric sister species pairs

In all 11 species pairs, several of the loci that were in high LD with another locus (430 at r^2>0.8 and 3,098 at r^2>0.7; see Table 3) coincided with a top 5% outlier FST window, and in most cases, this overlap was significant (corrected empirical p-value <0.05) (Fig. 5, Table 5, Fig. S6-15). (See also Fig. S5 for all FST landscapes and mean weighted FST values for all 11 tested species pairs).

Table 5 Results of testing coincidence of loci in high multispecies LD with FST outliers between sister species

Species pair	**overlaps**	**empirical p-value**	**corrected p-value**	**significance**
P. nyererei vs *P. pundamilia*	19	0	0	***
	59	0	0	***
P. sp. 'nyererei-like vs *P.* sp. 'pundamilia-like	7	0.038	0.046	*
	35	0.0008	0.002	**
N. omnicaeruleus vs *N. unicuspid scraper*	3	0.424	0.466	
	26	0.033	0.043	*
Y. pyrrocephalus vs *Y. plumbus*	9	0.002	0.003	**
	36	0.0003	0.001	**
M. mbipi vs *P.* pink anal fin	11	0.0007	0.002	**
	31	0.024	0.033	*
N. gigas vs *P. cyaneus*	1	0.933	0.933	
	19	0.497	0.52	
L. scraper vs. *L.* yellow chin	5	0.218	0.252	
	37	0.0002	0.001	**
EI. cinctus (E) vs *E.II coprologus* (blue F)	9	0.010	0.015	*
	37	0.0009	0.002	**
E. sp. 'new invasive' vs *Y. pyrrocephalus*	11	0.0002	0.001	**
	34	0.001	0.002	**
E.I paropius vs *E.II coprologus*	17	0	0	***
	52	0	0	***
E.I antleter vs *P.* new degeni	10	0.0007	0.002	**
	45	0	0	***

Overlaps is the count of how many top 5% FST outlier windows contained a locus involved in high LD. In each pair, the first row concerns the r^2>0.8 pairs and the second row the r^2>0.7 pairs. P-values were corrected using the FDR method. Significance values (of the corrected p-values) are *p<0.05, **p<0.01, ***p<0.001.

Summary, Conclusions & Outlook

By becoming sorted between subpopulations during adaptive radiation from a hybrid swarm, postzygotic incompatibilities could contribute to reproductive isolation between emerging species (Seehausen, 2013). We experimentally tested this hypothesis in the Lake Victoria haplochromine cichlid radiation by screening for genotype ratio distortions across whole genomes of 94 radiation member species as well as in three experimental hybrid crosses between Lake Victoria cichlid species, bred and raised under laboratory conditions. Our analyses revealed several patterns that could be indicative of the presence of intrinsic postzygotic incompatibilities and that might be consistent with a scenario of sorting of incompatibilities during speciation from a hybrid swarm.

We did not find evidence for the increased genome-wide heterozygosity in the F2 hybrids that would be expected if the parental species feature many intrinsically co-adapted alleles spread out across the genome (Simon *et al.*, 2018), suggesting that incompatibilities between the parental species are not so ubiquitous that they would induce genome-wide selection for increased heterozygosity (Table 2). This is consistent with the lack of evidence for hybrid breakdown among Lake Victoria cichlid species (Stelkens *et al.*, 2015). We did however detect several regions with segregation distortion in each cross, and more so in the crosses between more divergent (i.e. older) species (Fig. 1). Almost all detected regions were private to one cross. The finding of more distortions in crosses between more divergent species is consistent with two scenarios. First, the more divergent species might have had more time to accumulate incompatibilities through fixation of new mutations after speciation. This, however, would probably take longer than the few thousand years since speciation in Lake Victoria started (Stelkens *et al.*, 2010). Second, incompatibilities that would have come to segregate in the ancestral hybrid population might have been sorted during the process of speciation and adaptive radiation from a hybrid swarm, as in the scenario outlined in Seehausen, 2013. Our finding that in each of our three crosses several genomic regions suffer significant segregation distortion and that there is almost no overlap of these regions between the crosses is consistent with this hypothesis (but low power to detect regions may also explain low overlap between crosses).

None of the locus pairs from these distorted regions fully behaved like a classic BDM incompatibility, but this might not be surprising given our low power to detect the subtle deviations in genotype frequencies - and it might also not be expected, given that incompatibility causing epistatic interactions might be contingent on genotype at more than two loci (Satokangas *et al.*, 2020). We have some indications for the latter since many of the SNPs involved in high multispecies LD had high LD with more than one other (physically unlinked) SNP (see Table 3). Moreover, they might not only occur between sites that are reciprocally fixed between the grandparents/parental species, the type of loci to which we restricted our analysis (Cutter, 2012).

In the two more divergent Lake Victoria crosses we found significant overlap between FST outliers and regions with segregation distortion within one and four chromosomes, respectively (Fig. 2), and even across the whole genome in one of these two crosses. This was also the cross that showed the highest positional coincidence of segregation distorted regions with loci found to be in high LD between species across the Lake Victoria cichlid radiation (Fig. 4, Table 4).

In almost all of eleven sister species pairs that we analyzed, the position of high FST outlier regions coincided with loci in high LD between species across the radiation (r^2>0.7 or >0.8) significantly more often than by chance (Fig. 5, Table 5). The interpretation of the LD measure as evidence of between locus incompatibilities has to be done with caution, since it likely still includes effects of phylogenetic structure and effects of ecological adaptation. The first effect could maybe best be accounted for by using an alternative r^2 measure to account for phylogenetic structure in LD scans (Mangin *et al.*, 2012). One possibility to account for the second effect could be to subset the LD analysis to species belonging to the same trophic guild.

Regions that overlap in at least two or in all three analyses (segregation distortion measured under laboratory conditions in experimental species hybrids, high multispecies LD across the radiation, FST outliers) are quite likely to represent incompatibilities that might contribute to reproductive isolation between species in this radiation and will be candidate regions for further investigations. These will include annotating the genes in these regions. It will also be interesting to subset the whole genome dataset to fixed differences between the hybrid swarm ancestors – that is, their closest living relatives.

Future investigations should also consider structural variants such as indels, which was not possible with the current pipeline because it does not call indels reliably (personal comm. D. A. Marques). It is quite likely (see McGee *et al.*, 2020) that this will add to the number of incompatibilities to be discovered

References

Baird, N.A., Etter, P.D., Atwood, T.S., Currey, M.C., Shiver, A.L., Lewis, Z.A., *et al.* 2008. Rapid SNP Discovery and Genetic Mapping Using Sequenced RAD Markers. *PLoS One* **3**: e3376.

Bank, C., Bürger, R. & Hermisson, J. 2012. The limits to parapatric speciation: Dobzhansky-Muller incompatibilities in a continent-Island model. *Genetics* **191**: 845–863.

Bateson, W. 1909. Heredity and variation in modern lights. In: *Darwin and Modern Science* (A. C. Seward, ed), pp. 85–101. Cambridge University Press, Cambridge, UK.

Bezault, E., Mwaiko, S. & Seehausen, O. 2011. Population genomic tests of models of adaptive radiation in Lake Victoria region cichlid fish. *Evolution* **65**: 3381–3397.

Browning, B.L., Zhou, Y. & Browning, S.R. 2018. A One-Penny Imputed Genome from Next-Generation Reference Panels. *Am. J. Hum. Genet.* **103**: 338–348.

Butlin, R.K., Servedio, M.R., Smadja, C.M., Bank, C., Barton, N.H., Flaxman, S.M., *et al.* 2021. Homage to Felsenstein 1981, or why are there so few/many species? *Evolution* **75**: 978-988

Chang, C.C., Chow, C.C., Tellier, L.C.A.M., Vattikuti, S., Purcell, S.M. & Lee, J.J. 2015. Second-generation PLINK: Rising to the challenge of larger and richer datasets. *Gigascience* **4**: 1–16.

Corbett-Detig, R.B., Zhou, J., Clark, A.G., Hartl, D.L. & Ayroles, J.F. 2013. Genetic incompatibilities are widespread within species. *Nature* **504**: 135–137.

Coyne, J.A. & Orr, H.A. 2004. *Speciation*. Sinauer Associates, Sunderland, MA.

Cutter, A.D. 2012. The polymorphic prelude to Bateson-Dobzhansky-Muller incompatibilities. *Trends Ecol. Evol.* **27**: 209–218.

Danecek, P., Auton, A., Abecasis, G., Albers, C.A., Banks, E., DePristo, M.A., *et al.* 2011. The variant call format and VCFtools. *Bioinformatics* **27**: 2156–2158.

Dobzhansky, T. 1937. *Genetics and the Origin of Species*. Columbia University Press, New York, NY.

Dobzhansky, T. 1948. Genetics of natural populations; experiments on chromosomes of Drosophila pseudoobscura from different geographic regions. *Genetics* **33**: 588–602.

Feller, A.F., Haesler, M.P., Peichel, C.L. & Seehausen, O. 2020a. Genetic architecture of a key reproductive isolation trait differs between sympatric and non-sympatric sister species of Lake Victoria cichlids. *Proc. R. Soc. B Biol. Sci.* **287**.

Feller, A.F., Ogi, V., Seehausen, O. & Meier, J.I. 2021. Identification of a novel sex determining chromosome in cichlid fishes that acts as XY or ZW in different lineages. *Hydrobiologia* **7**.

Feller, A. & Seehausen, O. *submitted* Genetic architecture of adaptive radiation across two trophic levels

Feller, A.F., Selz, O.M., McGee, M.D., Meier, J.I., Mwaiko, S. & Seehausen, O. 2020b. Rapid generation of ecologically relevant behavioral novelty in experimental cichlid hybrids. *Ecol. Evol.* **10**: 7445–7462.

Felsenstein, J. 1981. Skepticism Towards Santa Rosalia, or Why are There so Few Kinds of Animals? *Evolution* **35**: 124.

Feulner, P.G.D., Schwarzer, J., Haesler, M.P., Meier, J.I. & Seehausen, O. 2018. A Dense Linkage Map of Lake Victoria Cichlids Improved the Pundamilia Genome Assembly and Revealed a Major QTL for Sex-Determination. *G3; Genes|Genomes|Genetics* **8**: 2411–2420.

Fryer, G. & Iles, T.D. 1972. *The cichlid fishes of the great lakes of Africa: their biology and evolution*. Oliver & Boyd, Edinburgh, UK.

Gavrilets, S. 2004. *Fitness landscapes and the origin of species*. Princeton University Press, Princeton.

Genner, M.J., Seehausen, O., Cleary, D.F.R., Knight, M.E., Michel, E. & Turner, G.F. 2004. How does the taxonomic status of allopatric populations influence species richness within African cichlid fish assemblages? *J. Biogeogr.* **31**: 93–102.

Gourbière, S. & Mallet, J. 2010. Are species real? The shape of the species boundary with exponential failure, reinforcement, and the "missing snowball." *Evolution* **64**: 1–24.

Haesler, M.P. & Seehausen, O. 2005. Inheritance of female mating preference in a sympatric sibling species pair of Lake Victoria cichlids: implications for speciation. *Proc. R. Soc. B Biol. Sci.* **272**: 237–245.

Hohenlohe, P.A., Bassham, S., Currey, M. & Cresko, W.A. 2012. Extensive linkage disequilibrium and parallel adaptive divergence across threespine stickleback genomes. *Philos. Trans. R. Soc. B Biol. Sci.* **367**: 395–409.

Langmead, B. & Salzberg, S.L. 2012. Fast gapped-read alignment with Bowtie 2. *Nat. Methods* **9**: 357–359.

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., *et al.* 2009. The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25**: 2078–2079.

Malinsky, M., Svardal, H., Tyers, A.M., Miska, E.A., Genner, M.J., Turner, G.F., *et al.* 2018. Whole genome sequences of Malawi cichlids reveal multiple radiations interconnected by gene flow. *Nat. Ecol. Evol.* **2**: 1940–1955.

Mangin, B., Siberchicot, A., Nicolas, S., Doligez, A., This, P. & Cierco-Ayrolles, C. 2012. Novel measures of linkage disequilibrium that correct the bias due to population structure and relatedness. *Heredity* **108**: 285–291.

Marques, D.A., Haesler, M.P., Meier, J.I., Kishe,. M.A., McGee, M.D., Mwaiko, S., Wagner, C.E. & Seehausen, O. *in prep*. Whole genomes resolve age of the fastest adaptive radiation.

McGee, M.D., Borstein, S.R., Meier, J.I., Marques, D.A., Mwaiko, S., Taabu, A., *et al.* 2020. The ecological and genomic basis of explosive adaptive radiation. *Nature* **586**: 75–79.

McKenna, A., Hanna, M., Banks, E., Sivachenko, A., Cibulskis, K., Kernytsky, A., *et al.* 2010. The Genome Analysis Toolkit: A MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res.* **20**: 1297–1303.

Meier, J.I., Marques, D.A., Mwaiko, S., Wagner, C.E., Excoffier, L. & Seehausen, O. 2017a. Ancient hybridization fuels rapid cichlid fish adaptive radiations. *Nat. Commun.* **8**.

Meier, J.I., Marques, D.A., Wagner, C.E., Excoffier, L. & Seehausen, O. 2018. Genomics of parallel ecological speciation in Lake Victoria cichlids. *Mol. Biol. Evol.* **35**: 1489–1506.

Meier, J.I., Sousa, V.C., Marques, D.A., Selz, O.M., Wagner, C.E., Excoffier, L., *et al.* 2017b. Demographic modelling with whole-genome data reveals parallel origin of similar Pundamilia cichlid species after hybridization. *Mol. Ecol.* **26**: 123–141.

Muller, H. 1942. Isolating mechanisms, evolution, and temperature. *Biol. Symp.* **6**: 71–125.

Nosil, P. 2012. *Ecological speciation*. Oxford University Press, Oxford, UK.

Orr, H.A. 1995. The population genetics of speciation: The evolution of hybrid incompatibilities. *Genetics* **139**: 1805–1813.

Orr, H.A. & Turelli, M. 2001. The Evolution of Postzygotic Isolation: Accumulating Dobzhansky-Muller Incompatibilities. *Evolution* **55**: 1085–1094.

Payseur, B.A. & Place, M. 2007. Searching the Genomes of Inbred Mouse Strains for Incompatibilities That Reproductively Isolate Their Wild Relatives. *J. Hered.* **98**: 115–122.

Presgraves, D.C. 2010. Speciation genetics: Search for the missing snowball. *Curr. Biol.* **20**: R1073–R1074.

R Core Team. 2020. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, AUT.

Satokangas, I., Martin, S.H., Helanterä, H., Saramäki, J. & Kulmuni, J. 2020. Multi-locus interactions and the build-up of reproductive isolation. *Philos. Trans. R. Soc. B Biol. Sci.* **375**: 20190543.

Schluter, D. 2000. *The Ecology of Adaptive Radiation*. Oxford University Press, New York, NY.

Schumer, M., Cui, R., Powell, D.L., Dresner, R., Rosenthal, G.G. & Andolfatto, P. 2014. High-resolution mapping reveals hundreds of genetic incompatibilities in hybridizing fish species. *Elife* **3**: e02535.

Seehausen, O. 2013. Conditions when hybridization might predispose populations for adaptive radiation. *J. Evol. Biol.* **26**: 279–281.

Seehausen, O. 1996. *Lake Victoria Rock Cichlids - taxonomy, ecology, and distribution*. Verduijn Cichlids, Zevenjuizen, NL.

Seehausen, O. 2015. Process and pattern in cichlid radiations - inferences for understanding unusually high rates of evolutionary diversification. *New Phytol.* **207**: 304–312.

Seehausen, O., Terai, Y., Magalhaes, I.S., Carleton, K.L., Mrosso, H.D.J., Miyagi, R., *et al.* 2008. Speciation through sensory drive in cichlid fish. *Nature* **455**: 620–626.

Seehausen, O. & Van Alphen, J.J.M. 1998. The effect of male coloration on female mate choice in closely related Lake Victoria cichlids (Haplochromis nyererei complex). *Behav. Ecol. Sociobiol.* **42**: 1–8.

Seehausen, O., van Alphen, J.J.M. & Witte, F. 1997. Cichlid fish diveristy threatened by eutrophication that curbs sexual selection. *Science* **277**: 1808–1811.

Seehausen, O., Witte, F., Alphen, J.J.M. & Bouton, N. 1998. Direct mate choice maintains diversity among sympatric cichlids in Lake Victoria. *J. Fish Biol.* **53**: 37–55.

Selz, O.M., Pierotti, M.E.R.R., Maan, M.E., Schmid, C. & Seehausen, O. 2014. Female preference for male color is necessary and sufficient for assortative mating in 2 cichlid sister species. *Behav. Ecol.* **25**: 612–626.

Selz, O.M. & Seehausen, O. 2019. Interspecific hybridization can generate functional novelty in cichlid fish. *Proc. R. Soc. B Biol. Sci.* **286**.

Simon, A., Bierne, N. & Welch, J.J. 2018. Coadapted genomes and selection on hybrids: Fisher's geometric model explains a variety of empirical patterns. *Evol. Lett.* **2**: 472–498.

Stager, J.C. & Johnson, T.C. 2008. The late Pleistocene desiccation of Lake Victoria and the origin of its endemic biota. *Hydrobiologia* **596**: 5–16.

Stelkens, R.B., Schmid, C. & Seehausen, O. 2015. Hybrid Breakdown in Cichlid Fish. *PLoS One* **10**: e0127207.

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

Stelkens, R.B., Young, K.A. & Seehausen, O. 2010. The Accumulation of Reproductive Incompatibilities in African Cichlid Fish. *Evolution* **64**: 617–633.

Templeton, A.R., Hemmer, H., Mace, G., Seal, U.S., Shields, W.M. & Woodruff, D.S. 1986. Local Adaptation , Coadaptation , and Population Boundaries. **125**: 115–125.

Turelli, M., Barton, N.H. & Coyne, J.A. 2001. Theory and speciation. *Trends Ecol. Evol.* **16**: 330–343.

Turner, G.F., Seehausen, O., Knight, M.E., Allender, C.J. & Robinson, R.L. 2001. How many species of cichlids are there in African lakes? *Mol. Ecol.* **10**: 793–806.

van Doorn, G.S., Dieckmann, U. & Weissing, F.J. 2004. Sympatric Speciation by Sexual Selection: A Critical Reevaluation. *Am. Nat.* **163**: 709–725.

Supplementary Materials

Table S1 List of samples used in multispecies LD analysis

Sample	Genus	Species	Location	Lake	Sex	Ecology
103637	Astatotilapia	nubila	Luanso	Victoria	m	insectivore
11536	Astatotilapia	nubila rocks	Python	Victoria	m	insectivore
109432	Astatotilapia	nubila swamp red	Sweya	Victoria	m	insectivore
104621	double-stripe group	tanaos	MG Transect	Victoria	m	zooplanktivore
13819	double-stripe group	thereuterion	Makobe	Victoria	m	insectivore
104013	Enterochromis I	antleter blue (St. E)	MG Transect	Victoria	m	detritivore
104016	Enterochromis I	cinctus (St. E)	MG Transect	Victoria	m	detritivore
108920	Enterochromis I	coprologus (St. E)	MG Transect	Victoria	m	detritivore
103908	Enterochromis I	new invasive	MG Transect	Victoria	m	detritivore
103720	Enterochromis I	paropius	MG Transect	Victoria	m	detritivore
109795	Enterochromis II	coprologus	MG Transect	Victoria	m	detritivore
11447	Enterochromis II	red and blue	Python	Victoria	m	detritivore
103767	Enterochromis II	supramacrops red tail	MG Transect	Victoria	m	detritivore
158683	Platytaeniodus	new degeni	UG3	Victoria	m	oral sheller/detritivore
TSO4	Gaurochromis	hiatus	Sozihe	Victoria	m	insectivore
TMB3	Gaurochromis	iris	Magu	Victoria	m	insectivore
11946	Haplochromis	purple yellow	Makobe	Victoria	m	epiphytic algae scraper
14561	Harpagochromis	cavifrons	Makobe	Victoria	m	piscivore
161543	Harpagochromis	cf. pachycephalus	?	Victoria	m	piscivore
109188	Harpagochromis	howesi	Makobe	Victoria	m?	piscivore
78797	Harpagochromis	odd dupper	Anchor	Victoria	m	piscivore
13135	Harpagochromis	vonlinnei	Makobe	Victoria	m	piscivore
109219	Incertae sedis	large red deepwater	Makobe	Victoria	m	insectivore
13269	Incertae sedis	sp. Lithochrom./Pundam.	Makobe	Victoria	m	insectivore
14298	Incertae sedis	thick skin	Makobe	Victoria	m	insectivore
105005	Incertae sedis	vanoijeni	MG Transect	Victoria	m	detritivore
80054	Labrochromis	cf. theliodon	MG Transect	Victoria	m	insectivore-pharyngeal-crusher
Lishmaeli	Labrochromis	ishmaeli	captive stock	Victoria	m	pharyngeal mollusc crusher
14259	Labrochromis	sp. "stone"	Makobe	Victoria	m	pharyngeal mollusc crusher
106641	Labrochromis	sp. 1	MG Transect	Victoria	m	pharyngeal mollusc crusher
14175	Lipochromis	cryptodon	Makobe	Victoria	m	paedophage
130785	Lipochromis	matumbi hunter	Trawling	Victoria	m	paedophage
11045	Lipochromis	velvet black cryptodon	Makobe	Victoria	m	paedophage
80805	Lithochromis	brown narrow snout	Mabibi	Victoria	m	insectivore
13281	Lithochromis	orange	Makobe	Victoria	m	insectivore
13393	Lithochromis	pseudoblue	Makobe	Victoria	m	insectivore
2801	Lithochromis	rubripinnis	Luanso	Victoria	m	insecti-planktivore
11024	Lithochromis	scraper	Makobe	Victoria	m	epilithic algae scraper
78773	Lithochromis	xanthopteryx	Anchor	Victoria	m	insectivore
109320	Lithochromis	yellow chin	Makobe	Victoria	m	zooplanktivore
11965	Hoplotilapia	retrodens	Makobe	Victoria	m	oral sheller
Ig158	Macropleurodus	bicolor	Igombe	Victoria	m	oral crusher
10448	Ptyochromis	deepwater rock sheller	Python	Victoria	m	oral sheller
79623	Ptyochromis	fischeri	Igombe	Victoria	m	oral sheller
Ig41	Ptyochromis	red rock sheller	Igombe	Victoria	m?	oral sheller
14245	Ptyochromis	striped rock sheller	Makobe	Victoria	m	oral sheller
79609	Ptyochromis	xenognathus	Igombe	Victoria	m	oral sheller
10794	Ptyochromis	xenognathus rocks	Kissenda	Victoria	m	oral sheller
11347	Mbipia	lutea	Makobe	Victoria	m	epilithic algae scraper
11003	Mbipia	mbipi	Makobe	Victoria	m	epilithic algae scraper
10791	Mbipia	red carp	Kissenda	Victoria	m	epilithic algae scraper
11338	Neochromis	gigas	Makobe	Victoria	m	epilithic algae scraper
10897	Neochromis	greenwoodi	Anchor	Victoria	m	epilithic algae scraper
Bi10	Neochromis	long black	Bwiru	Victoria	m	epilithic algae scraper
106816	Neochromis	omnicaeruleus	Makobe	Victoria	m	epilithic algae scraper
10619	Neochromis	rufocaudalis	Makobe	Victoria	m	epilithic algae scraper
11070	Neochromis	unicuspid scraper	Makobe	Victoria	m	generalist-algae scraper
Na15	Neochromis	yellow anal scraper	Nansio	Victoria	m	epilithic algae scraper
158795	Paralabidochromis	victoriae	UG4	Victoria	m?	insect picker
10618	Paralabidochromis	chilotes	Makobe	Victoria	m	insectivore
11807	Paralabidochromis	chromogynos	Makobe	Victoria	m	insectivore
130770	Paralabidochromis	crassilabris	Sesse	Victoria	m	insectivore
104042	Paralabidochromis	plagiodon	MG Transect	Victoria	m	oral sheller/detritivore

11639	Paralabidochromis	plagiodon rocks	Luanso	Victoria	m	oral sheller
11049	Paralabidochromis	sauvagei	Makobe	Victoria	m	insectivore
10628	Paralabidochromis	short snout scraper	Makobe	Victoria	m	epilithic algae scraper
10577	Paralabidochromis	cyaneus	Makobe	Victoria	m	insectivore
80344	Paralabidochromis	flavus	Makobe	Victoria	m	insectivore
Ga2	Paralabidochromis	orange anal rockpicker	Gana	Victoria	m	insectivore
Ki30	Paralabidochromis	pseudorockpicker	Kissenda	Victoria	m	insectivore
Ga33	Paralabidochromis	sky blue picker	Gana	Victoria	m	insectivore
161476	Prognathochromis	dichrourus complex	?	Victoria	m	piscivore
130786	Prognathochromis	perrieri	lab stock	Victoria	m	piscivore
10715	Pundamilia	azurea	Ruti	Victoria	m	zooplanktivore
78887	Pundamilia	orange anal nyererei	Bihiru	Victoria	m	insectivore
11034	Pundamilia	pink anal fin	Makobe	Victoria	m	zooplanktivore
84391	Pundamilia	sp. "red head"	?	Victoria	m?	insecti-planktivore?
E3c	Pundamilia	yellow azurea	UG/Nsimba	Victoria	m	insecti-planktivore
Ju22	Pundamilia	big blue red	Juma	Victoria	m	insectivore
14125	Pundamilia	deepwater giant	Makobe	Victoria	m	insectivore
IG291	Pundamilia	igneopinnis	Igombe	Victoria	m	zooplanktivore
11053	Pundamilia	nyererei	Makobe	Victoria	m	insecti-planktivore
12170	Pundamilia	sp. "nyererei-like"	Python	Victoria	m	insecti-planktivore
10554	Pundamilia	pundamilia	Makobe	Victoria	m	insectivore
11545	Pundamilia	sp. "pundamilia-like"	Python	Victoria	m	insectivore
12849	Pundamilia	sp. Luanso red	Luanso	Victoria	m	insectivore
11154	Pundamilia	blue deepwater	Luanso	Victoria	m	insectivore
13653	Pundamilia	macrocephala	Makobe	Victoria	m	insecti-planktivore
12372	Pundamilia	yellow deepwater	Luanso	Victoria	m	zooplanktivore
103778	Yssichromis	cf. supramacrops	MG Transect	Victoria	m	zooplanktivore
IG104	Yssichromis	laparogramma	Igombe	Victoria	m	zooplanktivore
79560	Yssichromis	plumbus	Igombe	Victoria	m	zooplanktivore
103754	Yssichromis	pyrrhocephalus	MG Transect	Victoria	m	zooplanktivore
Ypiceatus	Yssichromis	piceatus	lab stock	Victoria	m?	zooplanktivore

Table S2 List of species used in FST analyses

Genus	Species
Enterochromis I	paropius
Enterochromis I	antleter
Enterochromis I	cinctus (St. E)
Enterochromis I	coprologus I
Enterochromis I	sp. new invasive
Enterochromis II	coprologus F blue morph
Gaurochromis	hiatus
Harpagochromis	cavifrons
Harpagochromis	vonlinnei
Lithochromis	scraper
Lithochromis	yellow chin
Mbipia	mbipi
Neochromis	gigas
Neochromis	greenwoodi
Neochromis	omnicaeruleus
Neochromis	unicuspid scraper
Paralabidochromis	sauvagei
Paralabidochromis	short snout scraper
Paralabidochromis	cyaneus
Platytaeniodus	new degeni
Pundamilia	nyererei
Pundamilia	pundamilia
Pundamilia	sp. "nyererei-like"
Pundamilia	sp. "pundamilia-like"
Pundamilia	azurea
Pundamilia	sp. "red head"
Yssichromis	pyrrhocephalus
Yssichromis	plumbus

Table S3 List of species pairs and number of individuals used in FST analyses					
Number of individuals	**Genus**	**Species**	**Location**	**Lake**	**Ecology**
Pair1 (PNM and PPM)					
5	Pundamilia	nyererei	Makobe	Victoria	insecti-planktivore
5	Pundamilia	pundamilia	Makobe	Victoria	insectivore
Pair2 (PNP and PPP) †					
5	Pundamilia	sp. "nyererei-like"	Python	Victoria	insecti-planktivore
5	Pundamilia	sp. "pundamilia-like"	Python	Victoria	insectivore
Pair3 (NOM and NUM)					
3	Neochromis	omnicaeruleus	Makobe	Victoria	epilithic algae scraper
3	Neochromis	unicuspid scraper	Makobe	Victoria	generalist-algae scraper
Pair4 (YPyr and Yplu)					
3	Yssichromis	pyrrhocephalus	MG Transect?	Victoria	zooplanktivore
3	Yssichromis	plumbus	?	Victoria	zooplanktivore
Pair5 (MmbipM and PpinkM)					
5	Mbipia	mbipi	Makobe	Victoria	epilithic algae scraper
5	Pundamilia	pink anal fin	Makobe	Victoria	zooplanktivore
Pair6 (NGigM and ParcyM)					
3	Neochromis	gigas	Makobe	Victoria	epilithic algae scraper
3	Paralabidochromis	cyaneus	Makobe	Victoria	insectivore
Pair7 (LitScM and LitYelM)					
3	Lithochromis	scraper	Makobe	Victoria	epilithic algae scraper
3	Lithochromis	yellow chin	Makobe	Victoria	zooplanktivore
Pair8 (Ecinct and Ecopr)					
5	Enterochromis I	cinctus (St. E)	MG Transect	Victoria	detritivore
5	Enterochromis II	coprologus F blue morph	MG Transect	Victoria	detritivore
Pair 9 (Einvasie and Ypyr)					
3	Enterochromis I	sp. new invasive	MG Transect	Victoria	detritivore
3	Yssichromis	pyrrhocephalus	MG Transect	Victoria	zooplanktivore
Pair 10 (Epar and EcoprI)					
4	Enterochromis I	paropius	MG Transect	Victoria	detritivore
4	Enterochromis I	coprologus I	MG Transect	Victoria	detritivore
Pair 11 (Eantl and Platy)					
3	Enterochromis I	antleter	MG Transect	Victoria	detritivore
3	Platytaeniodus	new degeni	MG Transect	Victoria	oral sheller/detritivore
Pair (PNP and NOM) †					
3	Pundamilia	sp. "nyererei-like"	Python	Victoria	insecti-planktivore
3	Neochromis	omnicaeruleus	Makobe	Victoria	epilithic algae scraper
Pair (PPM and PRHZ) †					
3	Pundamilia	pundamilia	Makobe	Victoria	insectivore
3	Pundamilia	sp. "red head"	Zue?	Victoria	?
Pair					
3	Neochromis	gigas	Makobe	Victoria	epilithic algae scraper
3	Neochromis	greenwoodi	Python&Anchor	Victoria	epilithic algae scraper
Pair					
4	Pundamilia	azurea	Ruti	Victoria	zooplanktivore
3	Neochromis	omnicaeruleus	Makobe	Victoria	epilithic algae scraper
Pair					
3	Pundamilia	sp. "red head"	Zue?	Victoria	?
4	Pundamilia	pundamilia	Ruti	Victoria	insectivore
Pair					
3	Enterochromis I	paropius	MG Transect	Victoria	detritivore
2	Gaurochromis	hiatus	stock	Victoria	insectivore
Pair					
2	Paralabidochromis	sauvagei	Makobe	Victoria	insectivore
3	Paralabidochromis	short snout scraper	Makobe	Victoria	epilithic algae scraper
Pair					
2	Harpagochromis	cavifrons	Makobe	Victoria	piscivore
2	Harpagochromis	vonlinnei	Makobe	Victoria	piscivore
grey: not analysed yet (because either from different catch locations, or <3 available individuals) †cross species pairs					

Table S4 Genotype combinations of the F2 hybrids of the cross between *P. pundamilia* x *P. sp.* "red-head" at locus pairs where each locus alone showed a segregation distortion pattern consistent with if it was part of a classic BDMI pair. If the two loci together behaved like a BDMI, b/a and a/b combinations should be missing. Here, only pairs are shown where at least one of these two combinations was missing highlighted in red). In total 1077 pairs were tested in this cross.

locus1	locus2	a/a	b/a	h/a	a/b	b/b	h/b	a/h	b/h	h/h
chr1_9781962	chr15_11239358	4	0	10	6	5	12	17	7	16
chr1_9781962	chr15_11629638	4	0	10	6	6	12	17	7	15
chr1_9781962	chr15_14637876	3	0	9	4	5	9	19	8	19
chr1_9781962	chr15_15673436	4	0	8	5	4	11	18	9	19
chr1_9781962	chr22_1123507	9	5	14	0	0	1	17	8	24
chr1_9781962	chr22_8806451	11	5	14	0	2	3	17	6	22
chr1_9781962	chr22_10535214	8	5	13	0	0	3	18	8	20
chr1_9781962	chr22_10785368	8	4	13	0	1	3	16	7	19
chr1_9781962	chr22_11845172	9	4	14	0	0	3	17	9	22
chr1_9781963	chr15_11239358	4	0	10	6	5	12	17	7	16
chr1_9781963	chr15_11629638	4	0	10	6	6	12	17	7	15
chr1_9781963	chr15_14637876	3	0	9	4	5	9	19	8	19
chr1_9781963	chr15_15673436	4	0	8	5	4	11	18	9	19
chr1_9781963	chr22_1123507	9	5	14	0	0	1	17	8	24
chr1_9781963	chr22_8806451	11	5	14	0	2	3	17	6	22
chr1_9781963	chr22_10535214	8	5	13	0	0	3	18	8	20
chr1_9781963	chr22_10785368	8	4	13	0	1	3	16	7	19
chr1_9781963	chr22_11845172	9	4	14	0	0	3	17	9	22
chr1_10250411	chr15_11239358	4	0	8	7	3	15	19	8	17
chr1_10250411	chr15_11629638	4	0	9	7	4	15	18	9	17
chr1_10250411	chr15_14637876	3	0	9	5	3	12	20	9	19
chr1_10250411	chr15_15673436	4	0	8	6	2	14	20	11	19
chr1_10250411	chr21_13328510	4	0	5	4	3	13	14	3	18
chr1_10415194	chr15_11239358	4	0	11	9	4	16	20	9	17
chr1_10415194	chr15_11629638	4	0	13	8	6	16	20	9	18
chr1_10415194	chr15_15673436	5	0	10	7	4	16	23	12	24
chr1_10415194	chr21_13328510	4	0	6	5	4	13	16	3	19
chr1_11682348	chr15_11239358	4	0	11	8	4	17	19	7	22
chr1_11682348	chr15_11629638	4	0	13	8	6	16	18	7	25
chr1_11682348	chr15_15673436	5	0	10	7	4	17	21	10	29
chr1_11682348	chr21_13328510	4	0	6	4	4	15	14	1	23
chr1_12703004	chr15_11239358	4	0	11	8	4	17	19	8	21
chr1_12703004	chr15_11629638	4	0	13	8	6	16	18	8	24
chr1_12703004	chr15_15673436	5	0	10	7	3	16	21	11	29
chr1_12703004	chr21_13328510	4	0	6	4	3	15	14	2	22
chr1_12703159	chr15_11239358	4	0	11	8	4	16	18	8	22
chr1_12703159	chr15_11629638	4	0	13	8	6	15	17	8	25
chr1_12703159	chr15_15673436	5	0	10	7	4	16	20	11	28
chr1_12703159	chr21_13328510	4	0	6	4	4	15	14	2	23
chr14_16881627	chr22_10535214	6	6	18	0	0	4	12	7	37
chr14_17151603	chr15_11629638	0	0	17	8	3	19	11	9	29
chr14_17151603	chr15_14637876	3	0	14	5	2	19	13	13	38
chr14_17151618	chr15_11629638	0	0	17	8	3	19	11	9	29
chr14_17151618	chr15_14637876	3	0	13	5	2	19	13	13	38
chr14_17228173	chr15_14637876	2	0	13	4	2	17	11	12	36
chr14_17228173	chr15_15673436	1	0	12	5	2	18	14	13	30
chr14_19307011	chr15_11239358	1	0	14	6	4	19	13	9	25
chr14_19307011	chr15_11629638	1	0	16	6	4	20	11	10	29
chr14_19307011	chr15_14637876	3	0	14	5	3	18	13	13	35
chr14_19307047	chr15_11239358	1	0	14	6	4	19	13	8	26
chr14_19307047	chr15_11629638	1	0	16	6	4	20	11	9	30
chr14_19307047	chr15_14637876	3	0	14	5	3	18	13	13	35
chr14_19461879	chr15_11239358	1	0	14	7	4	18	14	8	26
chr14_19461879	chr15_11629638	1	0	16	7	4	19	12	9	29
chr14_19461879	chr15_14637876	3	0	14	5	3	18	15	12	33
chr14_20130495	chr15_11239358	1	0	13	6	4	13	12	5	22
chr14_20130495	chr15_11629638	1	0	14	6	3	14	11	7	22
chr14_20130495	chr15_14637876	2	0	11	4	3	11	12	6	26
chr14_20130495	chr15_15673436	2	0	11	3	3	14	13	7	25
chr14_20130495	chr21_21399950	3	0	9	5	4	12	10	6	24
chr14_20130495	chr21_22104453	4	0	10	4	4	13	11	6	29
chr14_22519314	chr15_11239358	1	0	14	7	5	17	13	7	28
chr14_22519314	chr15_11629638	1	0	16	7	5	18	11	7	32
chr14_22519314	chr15_14637876	2	0	14	5	4	17	14	10	37
chr14_22519314	chr22_10535214	4	8	20	0	0	4	14	5	38
chr14_23161198	chr15_11239358	1	0	14	8	5	16	15	6	27
chr14_23161198	chr15_11629638	1	0	14	7	5	17	13	5	30
chr14_23161198	chr15_14637876	2	0	11	6	4	16	15	7	35
chr14_23161198	chr15_15673436	2	0	10	5	5	17	17	7	33
chr14_23161198	chr21_21399950	3	0	10	8	5	19	11	6	31
chr14_23161198	chr21_22104453	4	0	10	7	5	19	14	7	36
chr14_24976519	chr15_11239358	1	0	14	7	5	17	14	7	28
chr14_24976519	chr15_11629638	1	0	16	7	5	18	12	6	33
chr14_24976519	chr15_14637876	3	0	14	5	4	17	14	10	39
chr14_25185508	chr15_11239358	2	0	13	7	5	15	14	6	27

chr14_25185508	chr15_11629638	2	0	15	7	5	16	12	6	31
chr14_25185508	chr15_14637876	3	0	14	5	4	15	15	8	37
chr14_25185557	chr15_11239358	2	0	13	7	5	17	14	7	27
chr14_25185557	chr15_11629638	2	0	15	7	5	18	12	6	32
chr14_25185557	chr15_14637876	4	0	13	6	4	16	14	10	36
chr14_25916643	chr15_11239358	1	0	14	7	5	17	13	8	28
chr14_25916643	chr15_11629638	1	0	16	7	5	18	12	7	32
chr14_25916643	chr15_14637876	3	0	13	5	4	17	13	10	41
chr14_25950108	chr15_11239358	1	0	14	6	4	16	15	6	26
chr14_25950108	chr15_11629638	1	0	16	6	5	17	13	4	31
chr14_25950108	chr15_14637876	2	0	13	4	4	16	16	7	33
chr14_25950108	chr15_15673436	2	0	12	3	5	18	18	5	33
chr14_26358578	chr15_11239358	2	0	13	7	5	17	14	8	27
chr14_26358578	chr15_11629638	2	0	15	7	5	18	12	7	30
chr14_26358578	chr15_14637876	4	0	12	5	4	17	14	11	39
chr15_15673436	chr22_1123507	5	11	17	0	1	3	10	15	38
chr15_21106438	chr21_13328349	4	0	6	5	7	13	5	10	25
chr21_13328349	chr22_1123507	5	8	14	0	1	1	5	16	24
chr21_13328349	chr22_8806451	3	9	14	0	2	3	7	14	23
chr21_13328349	chr22_10535214	3	9	13	0	2	2	7	12	24
chr21_13328349	chr22_10785368	3	8	11	0	2	3	5	12	22
chr21_13328349	chr22_10947330	4	9	13	0	3	3	5	13	23
chr21_13328349	chr22_11845172	3	9	13	0	2	2	7	14	23
chr21_13328510	chr22_1123507	4	5	15	0	1	1	6	15	22
chr21_13328510	chr22_8806451	4	6	17	0	2	3	6	15	19
chr21_13328510	chr22_10535214	4	5	13	0	2	2	6	12	21
chr21_13328510	chr22_10785368	4	6	12	0	2	3	4	12	18
chr21_13328510	chr22_10947330	5	6	14	0	2	3	5	13	20
chr21_13328510	chr22_11845172	3	7	14	0	2	3	7	13	21
chr21_14162568	chr22_1123507	6	7	18	0	2	1	9	19	31
chr21_14162568	chr22_8806451	4	8	19	0	3	5	11	19	27
chr21_14162568	chr22_10535214	5	8	17	0	2	2	9	14	31
chr21_14162568	chr22_10785368	4	9	15	0	2	4	6	13	29
chr21_14162568	chr22_10947330	5	9	16	0	4	3	8	15	31
chr21_14572420	chr22_1123507	6	8	21	0	2	2	6	20	33
chr21_14572420	chr22_8806451	4	9	20	0	4	6	7	19	31
chr21_14572420	chr22_10535214	4	8	20	0	2	3	7	15	32
chr21_14572420	chr22_10785368	5	9	15	0	2	5	5	15	29
chr21_14572420	chr22_10947330	6	9	18	0	4	4	6	16	33
chr21_14572420	chr22_11845172	3	10	19	0	3	4	8	18	32
chr21_16852421	chr22_1123507	4	7	19	0	2	2	8	16	33
chr21_16852421	chr22_8806451	4	9	18	0	3	6	8	13	31
chr21_16852421	chr22_10535214	4	7	18	0	2	3	7	12	31
chr21_16852421	chr22_10785368	4	8	14	0	2	5	5	11	29
chr21_16852421	chr22_10947330	5	9	16	0	3	5	6	12	31
chr21_16852421	chr22_11845172	3	9	17	0	3	4	8	13	32
chr21_16921911	chr22_1123507	6	9	18	0	2	1	7	16	34
chr21_16921911	chr22_8806451	5	11	17	0	3	5	8	14	32
chr21_16921911	chr22_10535214	5	8	17	0	2	2	8	12	33
chr21_16921911	chr22_10785368	4	9	13	0	2	4	6	10	33
chr21_16921911	chr22_10947330	6	10	15	0	3	4	7	14	32
chr21_16921911	chr22_11845172	4	11	16	0	3	3	9	13	34
chr21_21399950	chr22_1123507	5	9	18	0	5	1	10	19	32
chr21_21399950	chr22_10535214	4	7	17	0	2	3	10	17	24
chr21_21399950	chr22_10947330	4	9	19	0	4	4	10	16	26
chr21_21640091	chr22_1123507	4	8	22	0	4	1	12	13	33
chr21_21640091	chr22_10535214	3	7	21	0	2	3	13	11	26
chr21_21640091	chr22_10947330	4	9	20	0	4	4	12	9	30
chr21_22104453	chr22_1123507	5	9	21	0	5	1	12	19	36
chr21_22104453	chr22_10535214	3	8	21	0	2	3	12	16	29
chr21_22104453	chr22_10947330	5	10	19	0	3	5	11	16	32

a stands for homogyous like one parental species, b like the other parental species, h are heterozygotes

Table S5 Genotype combinations of the F2 hybrids of the cross between *P.* sp. "nyererei-like" x *N. omnicaeruleus* at locus pairs where each locus alone showed a segregation distortion pattern consistent with if it was part of a classic BDMI pair. If the two loci together behaved like a BDMI, b/a and a/b combinations should be missing. Here, only pairs are shown where at least one of these two combinations was missing highlighted in red). In total 438 pairs were tested in this cross.

locus1	locus2	a/a	b/a	h/a	a/b	b/b	h/b	a/h	b/h	h/h
chr3_28391306	chr10_33171463	5	3	11	0	0	1	10	2	21
chr5_23450844	chr6_16415805	1	5	9	0	3	2	5	7	21
chr5_23450844	chr6_17119759	1	5	8	0	3	2	5	8	20
chr5_23450844	chr6_17153179	1	5	6	0	4	1	3	6	18
chr5_23450844	chr6_17161041	1	5	8	0	3	2	5	8	21
chr5_23450844	chr6_19354037	1	4	10	0	3	3	5	9	18
chr5_23450844	chr10_33171463	3	4	11	0	1	0	3	11	21
chr5_23585080	chr6_16415805	1	5	10	0	3	2	6	7	20
chr5_23585080	chr6_17119759	1	5	9	0	3	2	6	8	18
chr5_23585080	chr6_17153179	1	5	6	0	4	1	4	6	16
chr5_23585080	chr6_17161041	1	5	9	0	3	2	6	8	19
chr5_23585080	chr6_19354037	1	4	11	0	3	2	6	9	17
chr5_23756418	chr6_16415805	1	7	8	0	3	2	4	8	21
chr5_23756418	chr6_17119759	1	6	8	0	3	2	4	10	20
chr5_23756418	chr6_17153179	1	5	6	0	4	1	3	6	18
chr5_23756418	chr6_17161041	1	6	8	0	3	2	4	10	20
chr5_23756418	chr6_19354037	1	5	10	0	3	3	4	10	18
chr5_23756418	chr10_33171463	2	5	11	0	1	1	3	13	19
chr5_24810585	chr6_16415805	2	4	10	0	3	2	5	6	24
chr5_24810585	chr6_17119759	2	4	9	0	3	2	5	6	23
chr5_24810585	chr6_17153179	2	3	7	0	3	1	3	5	20
chr5_24810585	chr6_17161041	1	4	10	0	3	2	5	6	24
chr5_24810585	chr6_19354037	2	3	11	0	3	3	5	7	21
chr5_24810585	chr10_30107174	3	2	11	0	0	5	4	11	19
chr5_24810585	chr10_30730951	1	2	8	0	0	4	3	11	18
chr5_24810585	chr10_30838246	2	2	10	0	0	3	3	11	21
chr5_24810585	chr10_33171463	4	2	12	0	0	2	3	11	22
chr5_24810585	chr10_34082788	4	2	7	0	1	2	2	9	13
chr5_25488321	chr6_16415805	2	5	9	0	3	2	3	6	25
chr5_25488321	chr6_17119759	2	4	9	0	3	2	3	8	23
chr5_25488321	chr6_17153179	2	3	7	0	4	1	3	4	21
chr5_25488321	chr6_17161041	1	4	10	0	3	2	3	8	24
chr5_25488321	chr6_19354037	2	3	11	0	3	3	3	8	22
chr5_25488321	chr10_30107174	2	3	11	0	1	5	3	11	19
chr5_25488321	chr10_30730951	1	2	8	0	0	4	3	10	19
chr5_25488321	chr10_30838246	2	3	9	0	0	3	3	11	21
chr5_25488321	chr10_33171463	3	4	12	0	0	2	2	11	22
chr5_25488321	chr10_34082788	4	3	7	0	1	2	1	8	15
chr5_25765118	chr6_16415805	2	5	9	0	3	2	4	7	25
chr5_25765118	chr6_17119759	2	4	9	0	3	2	4	9	22
chr5_25765118	chr6_17153179	2	3	7	0	4	1	3	5	20
chr5_25765118	chr6_17161041	1	4	10	0	3	2	4	9	23
chr5_25765118	chr6_19354037	2	3	11	0	3	3	4	9	21
chr5_25765118	chr10_30107174	2	3	11	0	1	5	4	12	18
chr5_25765118	chr10_30730951	1	2	8	0	0	4	3	11	18
chr5_25765118	chr10_30838246	2	3	9	0	0	3	3	12	20
chr5_25765118	chr10_33171463	3	4	13	0	0	2	3	12	21
chr5_25765118	chr10_34082788	4	3	8	0	1	2	1	9	14
chr5_26113404	chr6_16415805	2	5	7	0	4	1	4	6	22
chr5_26113404	chr6_17119759	2	4	7	0	4	1	4	8	20
chr5_26113404	chr6_17153179	2	3	6	0	4	1	3	5	19
chr5_26113404	chr6_17161041	1	4	8	0	4	1	4	8	21
chr5_26113404	chr6_19354037	2	4	8	0	4	2	4	8	20
chr5_26113404	chr10_30107174	2	3	9	0	2	4	4	11	17
chr5_26113404	chr10_30730951	1	2	7	0	1	3	3	11	18
chr5_26113404	chr10_30838246	2	3	7	0	1	2	3	11	18
chr5_26113404	chr10_33171463	3	4	10	0	0	2	3	12	18
chr5_26113404	chr10_34082788	4	3	6	0	1	2	1	8	13
chr5_26509264	chr6_16415805	2	5	9	0	3	2	4	5	24
chr5_26509264	chr6_17119759	2	4	9	0	3	2	4	6	21
chr5_26509264	chr6_17153179	2	3	7	0	3	1	3	5	19
chr5_26509264	chr6_17161041	1	4	10	0	3	2	4	6	22
chr5_26509264	chr6_19354037	2	3	11	0	2	3	4	8	20
chr5_26509264	chr10_30107174	2	3	11	0	0	5	4	10	17
chr5_26509264	chr10_30730951	1	2	8	0	0	4	3	10	18
chr5_26509264	chr10_30838246	2	3	9	0	0	3	3	10	19
chr5_26509264	chr10_33171463	3	3	13	0	0	2	3	10	20
chr5_26509264	chr10_34082788	4	2	8	0	0	2	1	9	13
chr5_26574651	chr6_16415805	1	5	10	0	3	2	3	7	24
chr5_26574651	chr6_17119759	1	4	10	0	3	2	3	9	21
chr5_26574651	chr6_17153179	1	3	8	0	4	1	3	5	19
chr5_26574651	chr6_17161041	1	4	10	0	3	2	3	9	22
chr5_26574651	chr6_19354037	1	3	12	0	3	3	3	9	20

chr5_26574651	chr10_30107174	1	3	12	0	1	5	3	12	17
chr5_26574651	chr10_30730951	1	2	8	0	0	4	3	11	18
chr5_26574651	chr10_30838246	1	3	10	0	0	3	3	11	20
chr5_26574651	chr10_33171463	2	4	14	0	0	2	2	12	20
chr5_26574651	chr10_34082788	3	3	9	0	1	2	1	8	14
chr5_27136684	chr6_16415805	3	5	8	0	3	2	4	7	25
chr5_27136684	chr6_17119759	3	4	8	0	3	2	4	9	22
chr5_27136684	chr6_17153179	3	3	6	0	4	1	3	5	20
chr5_27136684	chr6_17161041	2	4	9	0	3	2	4	9	23
chr5_27136684	chr6_19354037	3	3	10	0	3	3	4	9	21
chr5_27136684	chr10_30107174	2	3	11	0	1	5	5	12	17
chr5_27136684	chr10_30730951	1	2	8	0	0	4	4	11	17
chr5_27136684	chr10_30838246	2	3	9	0	0	3	4	12	19
chr5_27136684	chr10_33171463	3	4	13	0	0	2	4	12	20
chr5_27136684	chr10_34082788	4	3	8	0	1	2	2	9	13
chr5_27136691	chr6_16415805	3	5	8	0	3	2	4	7	25
chr5_27136691	chr6_17119759	3	4	8	0	3	2	4	9	22
chr5_27136691	chr6_17153179	3	3	6	0	4	1	3	5	20
chr5_27136691	chr6_17161041	2	4	9	0	3	2	4	9	23
chr5_27136691	chr6_19354037	3	3	10	0	3	3	4	9	21
chr5_27136691	chr10_30107174	2	3	11	0	1	5	5	12	17
chr5_27136691	chr10_30730951	1	2	8	0	0	4	4	11	17
chr5_27136691	chr10_30838246	2	3	9	0	0	3	4	12	19
chr5_27136691	chr10_33171463	3	4	13	0	0	2	4	12	20
chr5_27136691	chr10_34082788	4	3	8	0	1	2	2	9	13
chr5_27276030	chr6_16415805	3	3	9	0	4	1	4	7	22
chr5_27276030	chr6_17119759	3	2	9	0	4	1	4	9	20
chr5_27276030	chr6_17153179	3	1	7	0	4	1	3	6	19
chr5_27276030	chr6_17161041	2	2	10	0	4	1	4	9	21
chr5_27276030	chr6_19354037	3	1	11	0	4	2	4	10	19
chr5_27276030	chr10_30107174	2	3	11	0	2	4	5	10	17
chr5_27276030	chr10_30730951	1	2	8	0	1	3	4	10	17
chr5_27276030	chr10_30838246	2	3	9	0	0	3	4	11	17
chr5_27276030	chr10_33171463	3	4	12	0	0	2	4	11	18
chr5_27276030	chr10_34082788	4	3	7	0	1	2	2	8	12
chr5_27377399	chr6_16415805	2	4	10	0	3	2	4	6	24
chr5_27377399	chr6_17119759	2	3	10	0	3	2	4	8	21
chr5_27377399	chr6_17153179	2	2	8	0	4	1	3	5	19
chr5_27377399	chr6_17161041	2	3	10	0	3	2	4	8	22
chr5_27377399	chr6_19354037	2	2	12	0	3	3	4	9	20
chr5_27377399	chr10_30107174	1	3	12	0	1	5	5	10	17
chr5_27377399	chr10_30730951	1	2	8	0	0	4	4	10	17
chr5_27377399	chr10_30838246	1	3	10	0	0	3	4	10	19
chr5_27377399	chr10_33171463	2	4	14	0	0	1	4	10	21
chr5_27377399	chr10_34082788	3	3	9	0	1	1	2	8	14
chr5_27560275	chr6_16415805	2	3	10	0	3	2	4	7	21
chr5_27560275	chr6_17119759	2	2	10	0	3	2	4	8	19
chr5_27560275	chr6_17153179	2	2	8	0	3	1	3	5	20
chr5_27560275	chr6_17161041	2	2	10	0	3	2	4	8	20
chr5_27560275	chr6_19354037	2	1	12	0	3	3	4	8	18
chr5_27560275	chr10_30107174	1	3	11	0	0	5	5	10	17
chr5_27560275	chr10_30730951	1	2	8	0	0	4	4	9	17
chr5_27560275	chr10_30838246	1	3	10	0	0	3	4	10	18
chr5_27560275	chr10_33171463	2	3	12	0	0	2	4	10	19
chr5_27560275	chr10_34082788	3	2	8	0	1	2	2	7	12
chr6_16415805	chr10_34082788	5	1	8	0	0	3	6	3	15
chr6_17119759	chr10_34082788	5	1	8	0	0	3	6	3	14
chr6_17153179	chr10_34082788	5	2	7	0	0	2	5	3	12
chr6_17161041	chr10_34082788	4	1	8	0	0	3	6	3	15
chr6_19354037	chr10_34082788	5	2	7	0	1	2	5	2	17

a stands for homogyous like one parental species, b like the other parental species, h are heterozygotes

Table S6 Genotype combinations of the F2 hybrids of the Malawi cross at locus pairs where each locus alone showed a segregation distortion pattern consistent with if it was part of a classic BDMI pair. If the two loci together behaved like a BDMI, b/a and a/b combinations should be missing. Here, only pairs are shown where at least one of these two combinations was missing highlighted in red). In total 223 pairs were tested in this cross.

locus1	locus2	a/a	b/a	h/a	a/b	b/b	h/b	a/h	b/h	h/h
chr18_295529	chr21_20083735	5	0	4	5	4	14	7	5	26
chr18_295529	chr21_20089419	5	0	4	5	4	14	7	6	27
chr18_295529	chr21_21581627	2	0	7	4	4	11	10	6	23
chr18_1787484	chr21_20083735	5	0	4	4	4	14	7	5	26
chr18_1787484	chr21_20089419	5	0	4	4	4	14	7	6	27
chr18_1787484	chr21_21581627	2	0	7	3	4	11	10	6	23
chr18_2547542	chr21_20083735	4	0	5	5	4	14	7	5	26
chr18_2547542	chr21_20089419	4	0	5	5	4	14	7	5	27
chr18_2547542	chr21_21581627	2	0	7	4	4	11	9	5	24
chr18_2818456	chr21_21581627	2	0	7	4	4	11	10	6	22
chr18_4265028	chr21_20083735	5	0	4	5	5	13	7	5	26
chr18_4265028	chr21_20089419	5	0	4	5	5	13	7	5	28
chr18_4265028	chr21_21581627	2	0	7	4	5	10	10	5	24
chr18_4265094	chr21_20083735	5	0	4	6	5	12	7	5	26
chr18_4265094	chr21_20089419	5	0	4	6	5	12	7	5	28
chr18_4265094	chr21_21581627	2	0	7	5	5	9	10	5	24
chr18_5081417	chr21_20083735	6	0	3	6	5	12	7	5	26
chr18_5081417	chr21_20089419	6	0	3	6	5	12	7	5	27
chr18_5081417	chr21_21581627	2	0	7	5	5	9	11	5	22
chr18_5977111	chr21_20083735	5	0	4	5	5	13	7	5	26
chr18_5977111	chr21_20089419	5	0	4	5	5	13	7	5	27
chr18_5977111	chr21_21581627	2	0	7	4	5	10	10	5	23

a stands for homogyous like one parental species, b like the other parental species, h are heterozygotes

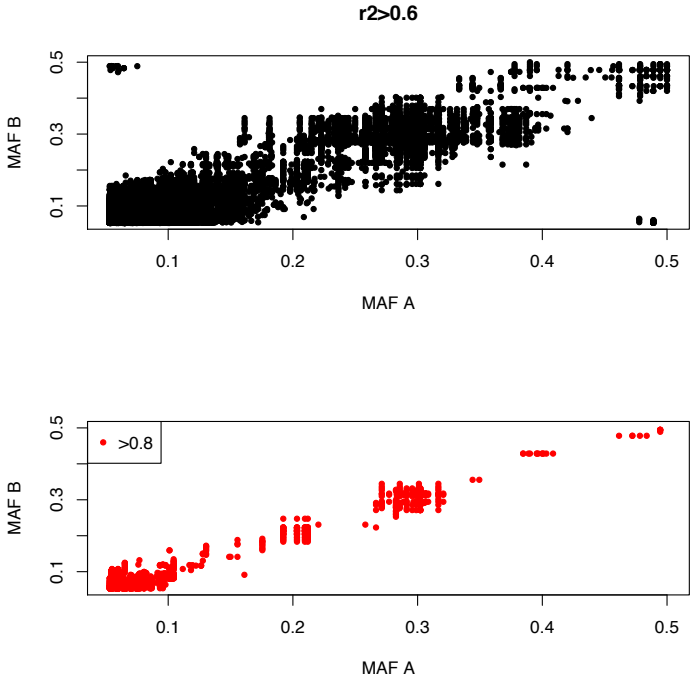


Figure S1 Minor allele frequencies at locus pairs in high LD (r^2>0.6; no additional MAF filters). In the bottom plot, only the subset of pairs with an r^2-value of >0.8 are highlighted.

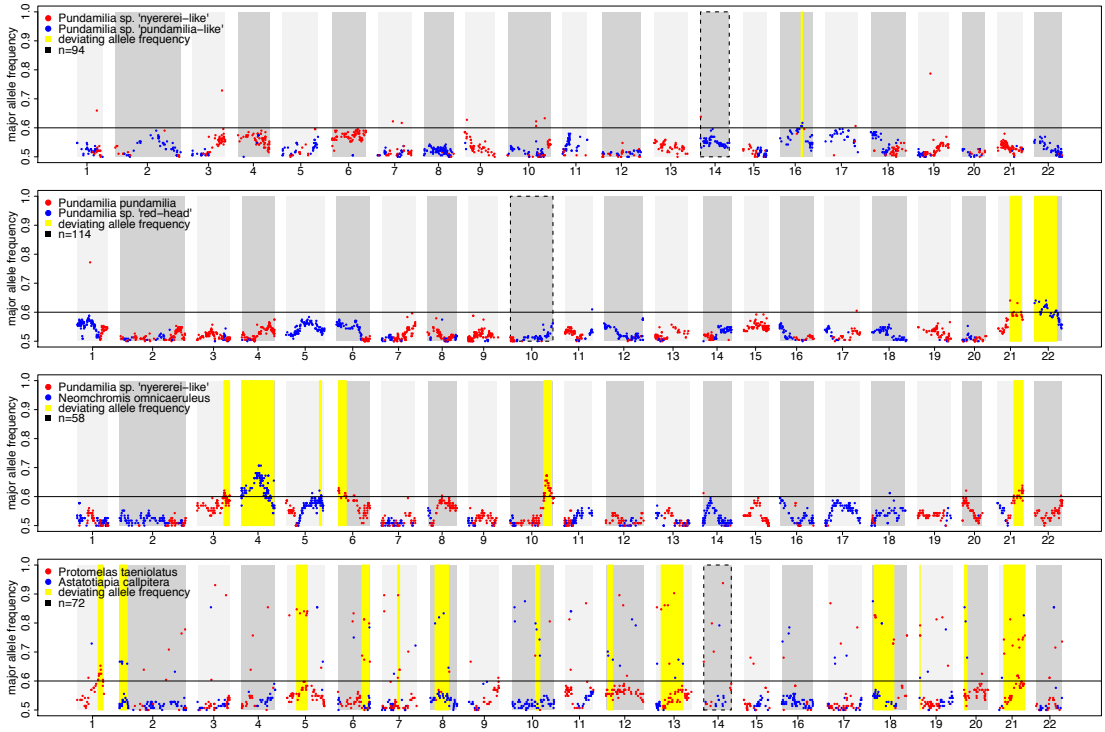


Figure S2 Major allele frequency plots for each cross using an equal number of males and females (total number of individuals indicated for each cross as "n=" in legend), regions with deviating allele frequencies highlighted in yellow. Chromosomes known to be involved in sex determination are framed with a dashed line. Grandparental origin of major allele according to blue/red colour in legend.

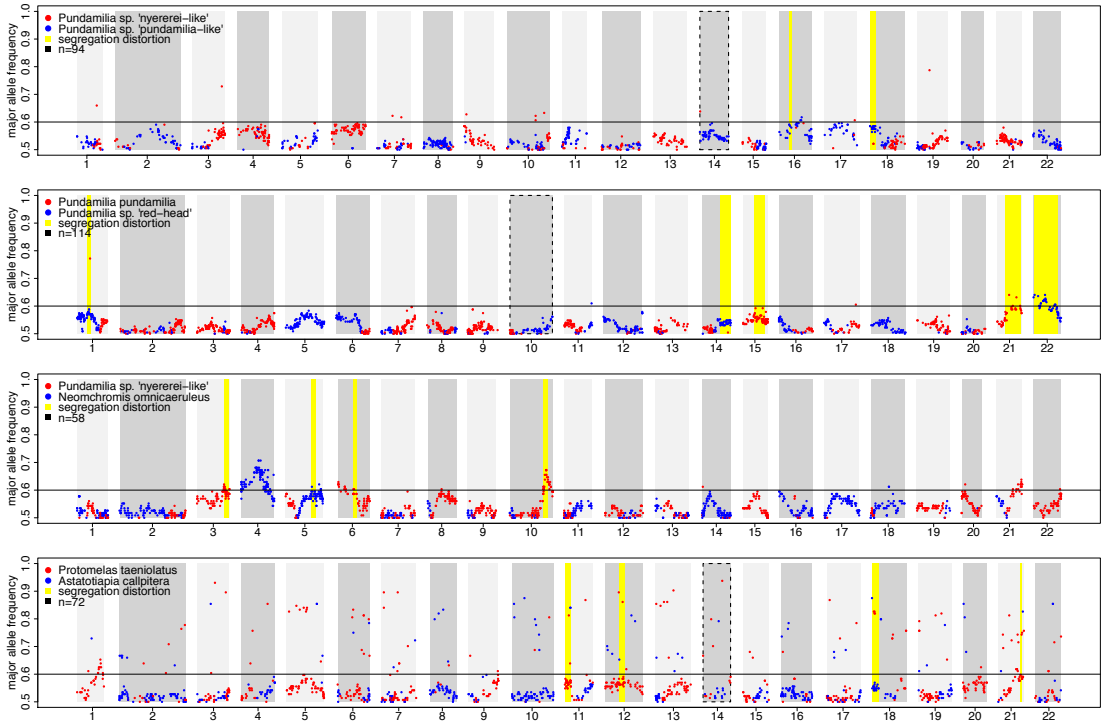


Figure S3 Major allele frequency plots for each cross using an equal number of males and females (total number of individuals indicated for each cross as "n=" in legend), regions with 'classic' segregation distortion highlighted in yellow. Chromosomes known to be involved in sex determination are framed with a dashed line. Grandparental origin of major allele according to blue/red colour in legend.

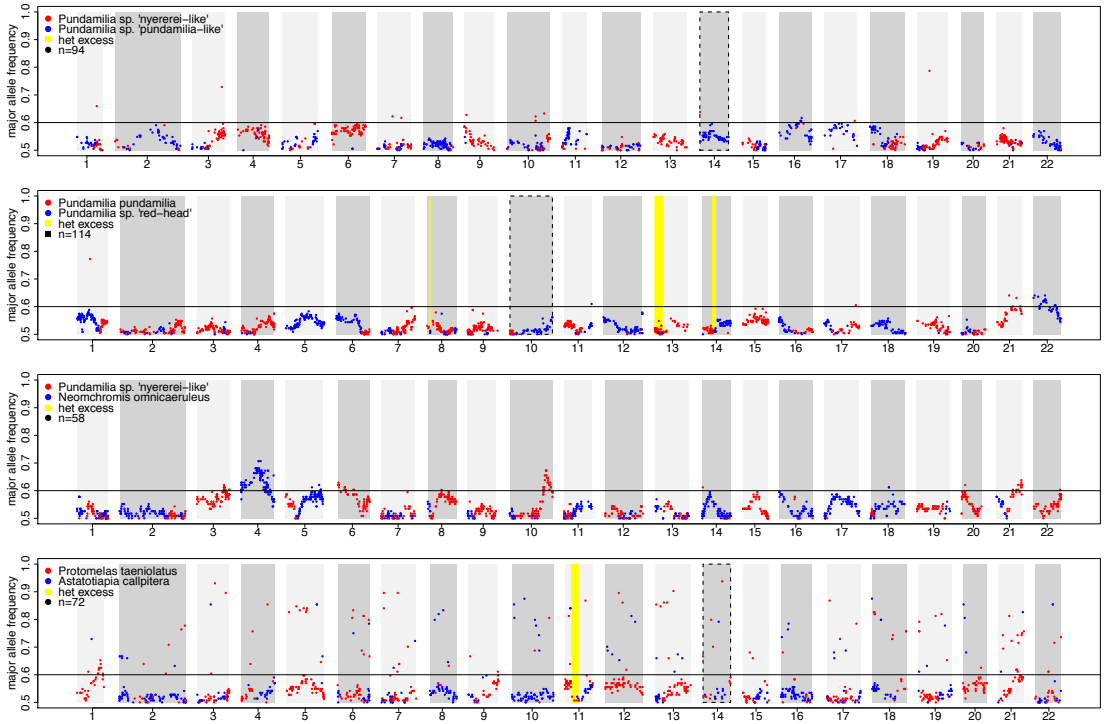


Figure S4 Major allele frequency plots for each cross using an equal number of males and females (total number of individuals indicated for each cross as "n=" in legend), regions with excess of heterozygous genotypes (and both homozygous genotypes equally reduced) highlighted in yellow. Chromosomes known to be involved in sex determination are framed with a dashed line. Grandparental origin of major allele according to blue/red colour in legend.

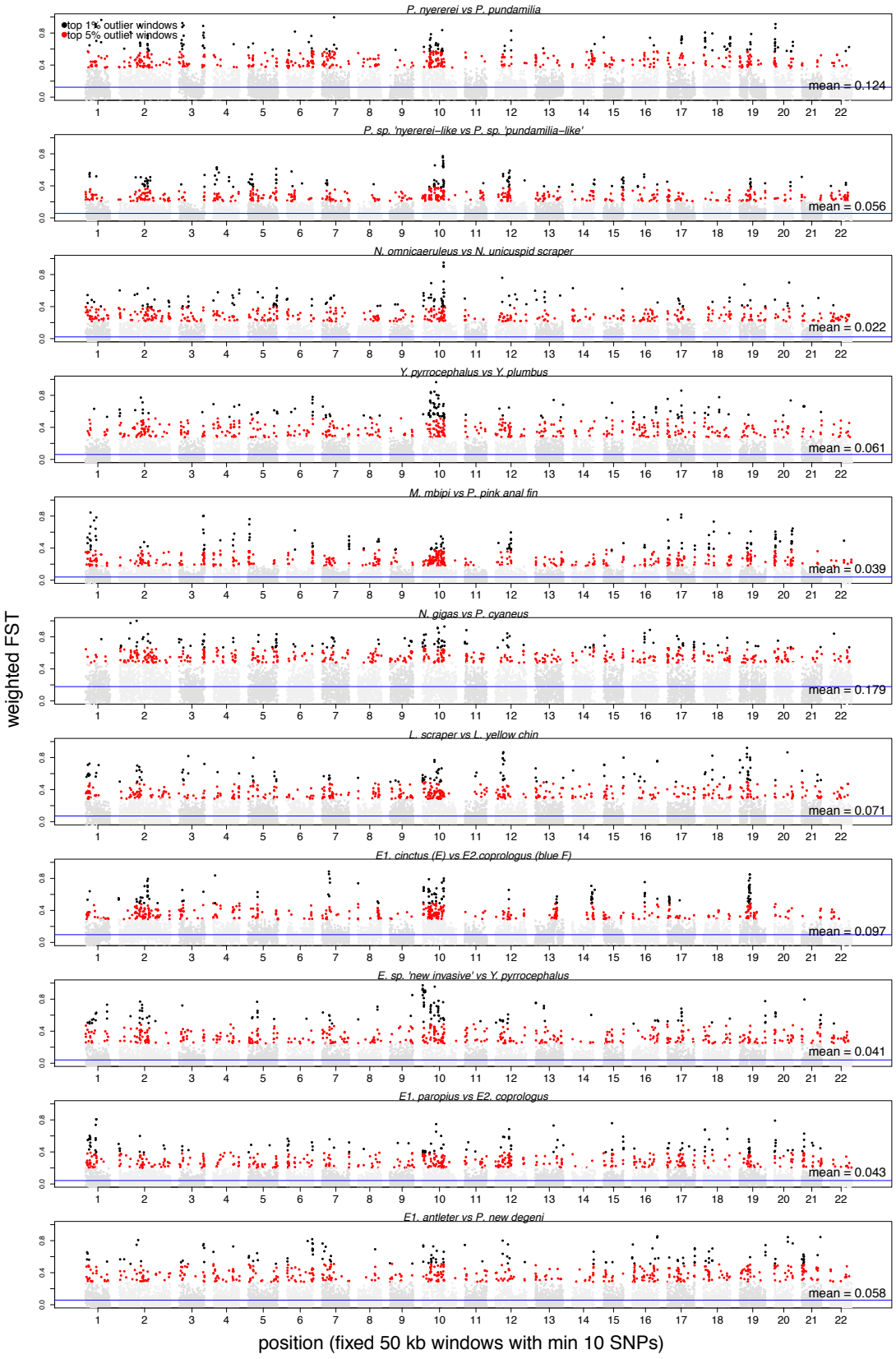


Figure S5 FST landscapes for all 11 tested sister species pairs.

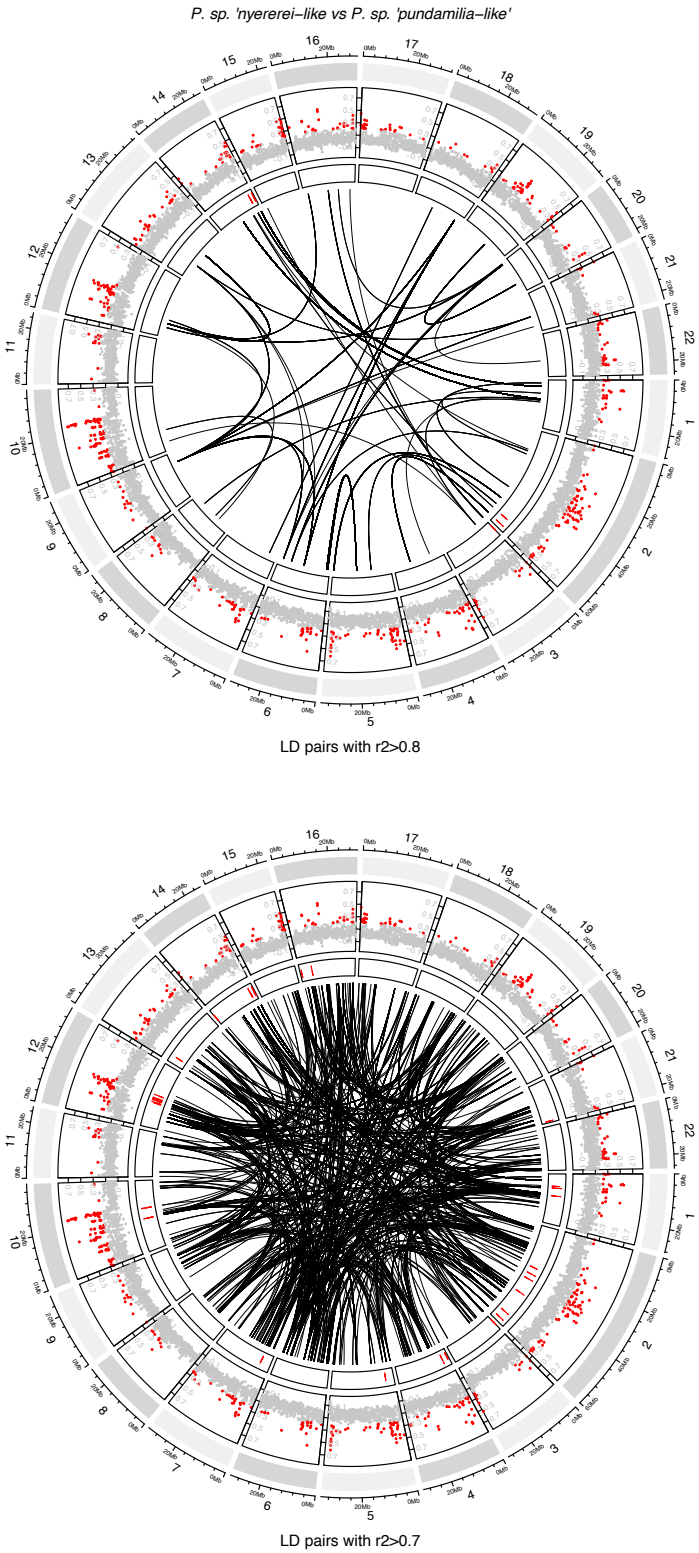
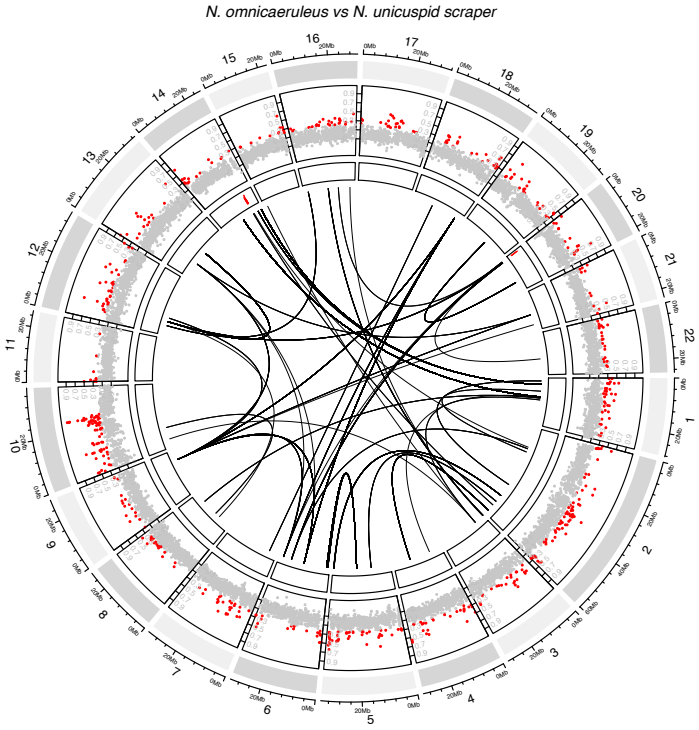
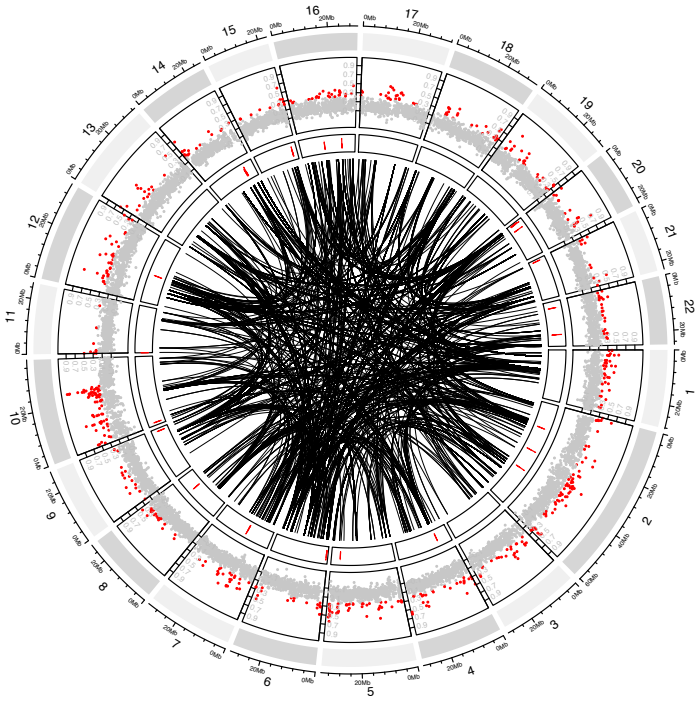

P. sp. 'nyererei−like vs P. sp. 'pundamilia−like'
LD pairs with r2>0.8
LD pairs with r2>0.7

Figure S6

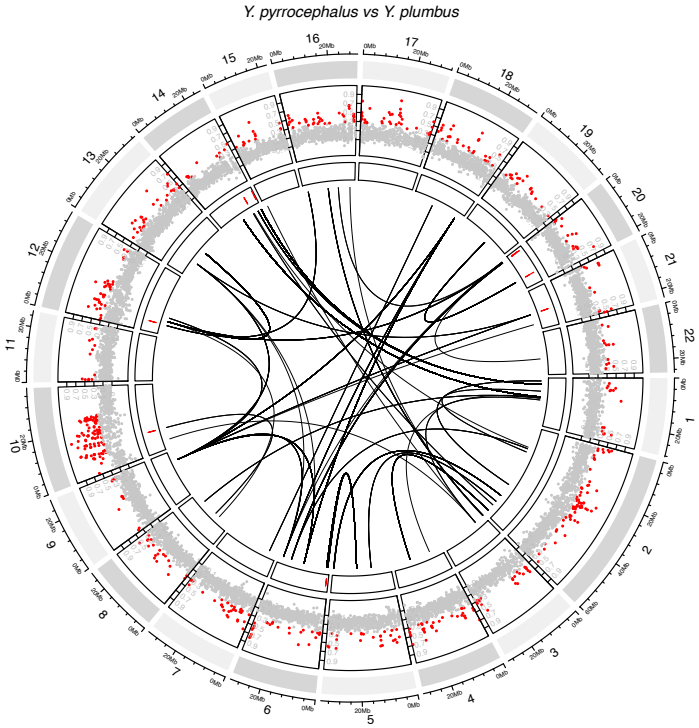


LD pairs with r2>0.8

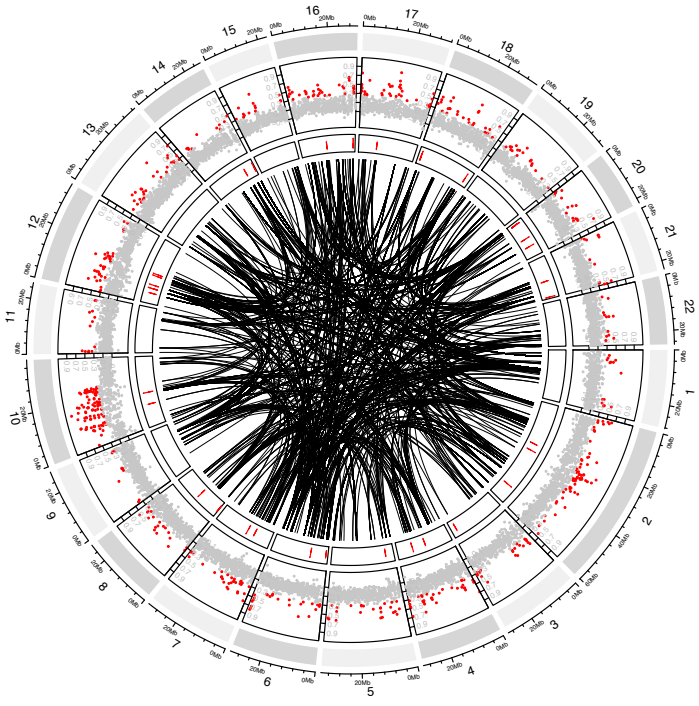


LD pairs with r2>0.7

Figure S7

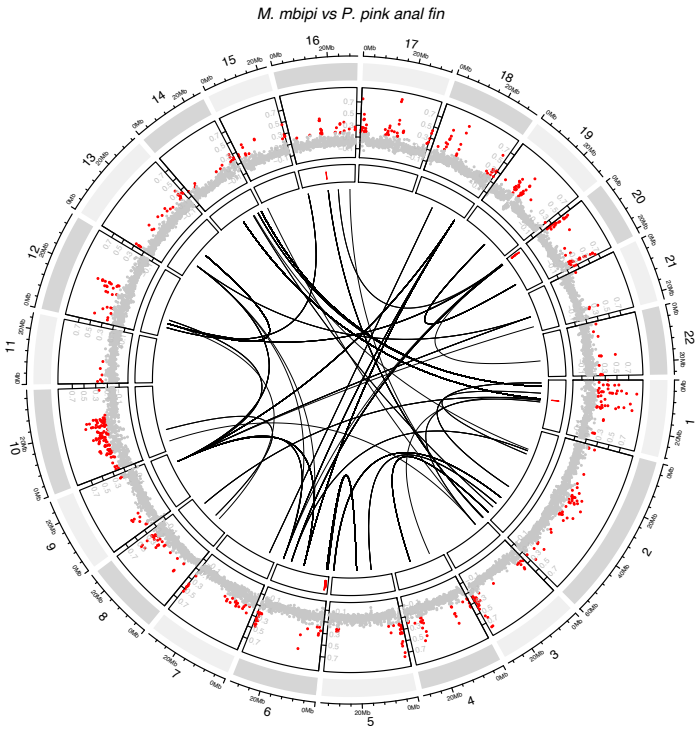


LD pairs with r2>0.8

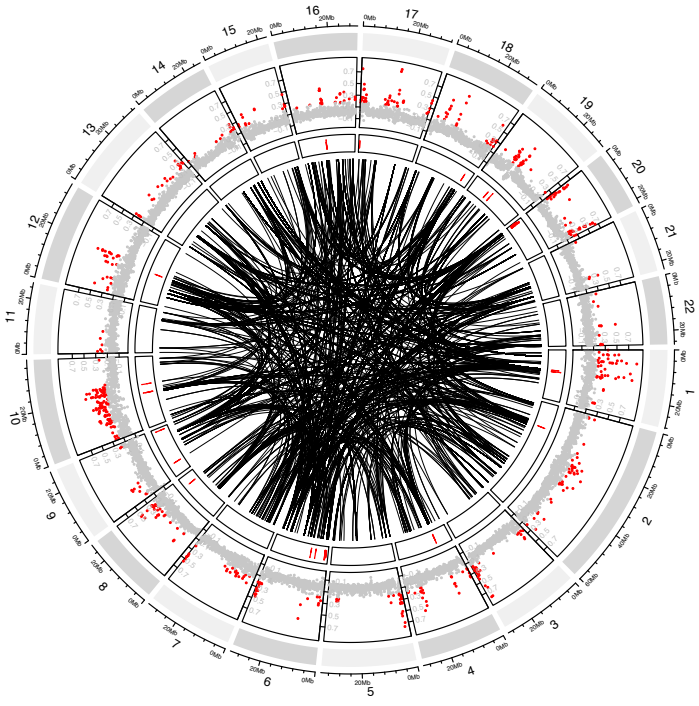


LD pairs with r2>0.7

Figure S8

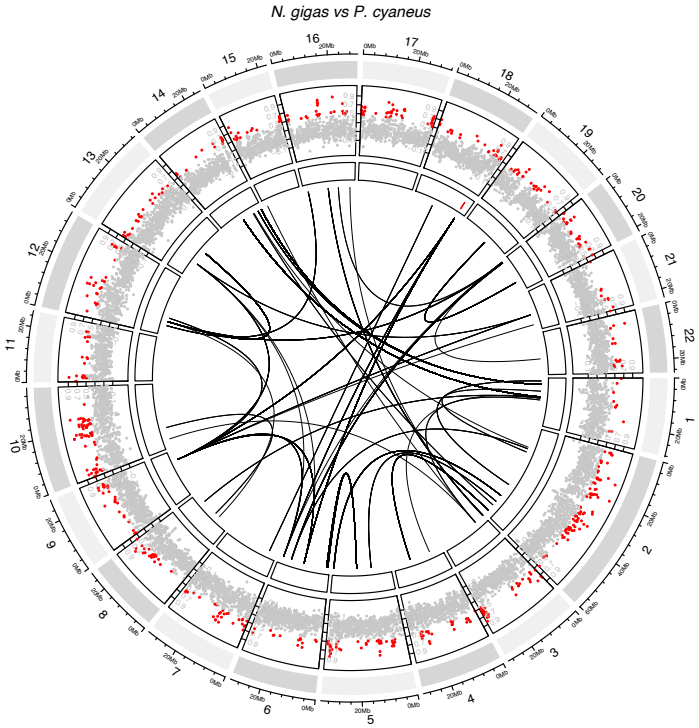


LD pairs with r2>0.8



LD pairs with r2>0.7

Figure S9



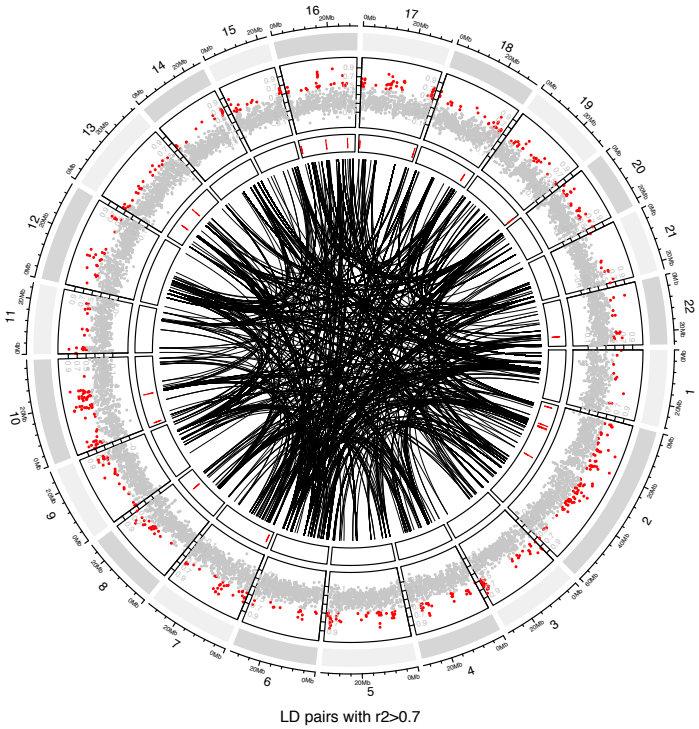
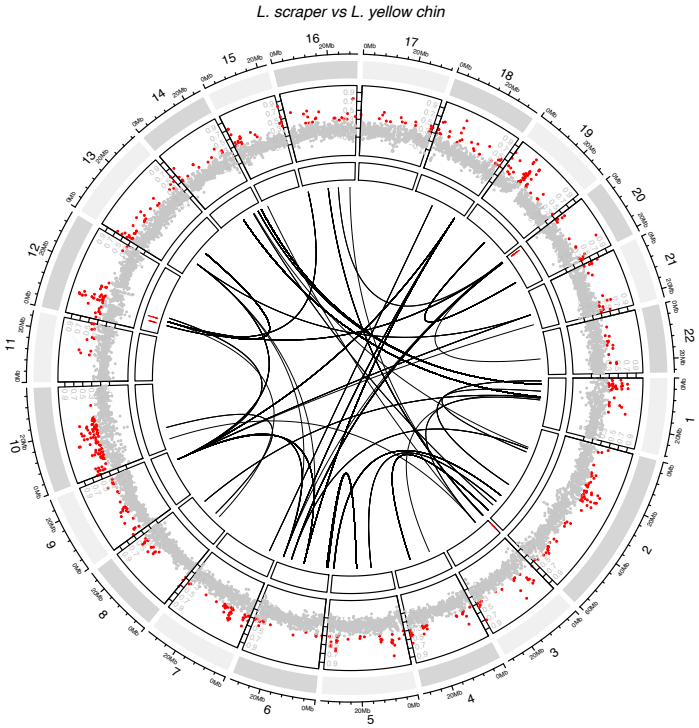
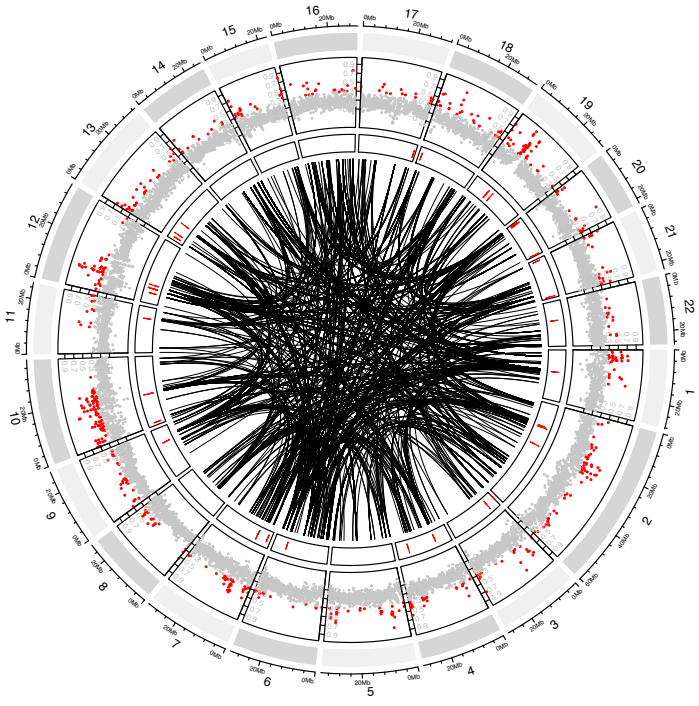


Figure S10

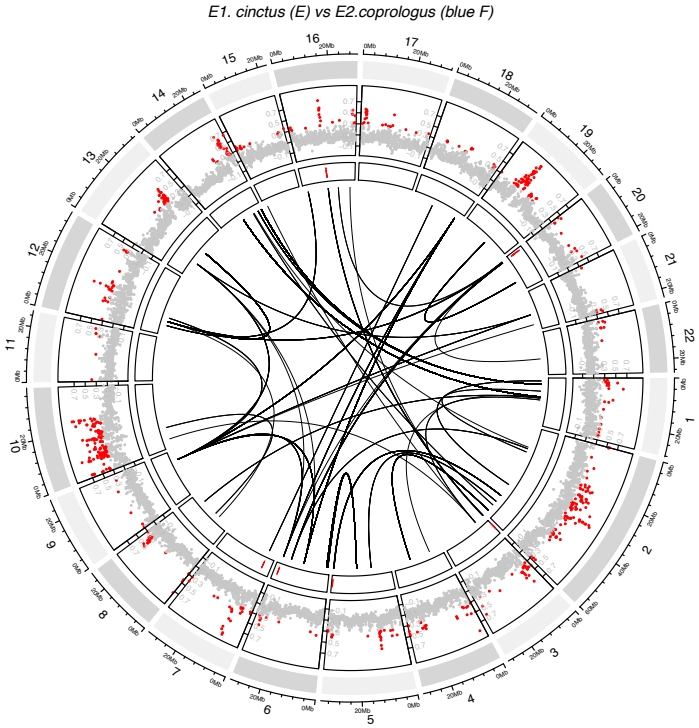


LD pairs with r2>0.8

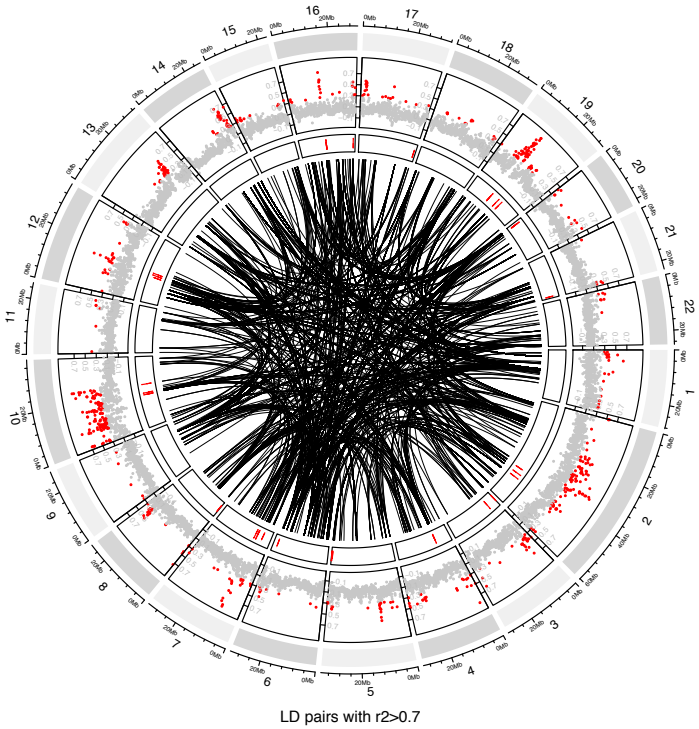


LD pairs with r2>0.7

Figure S11

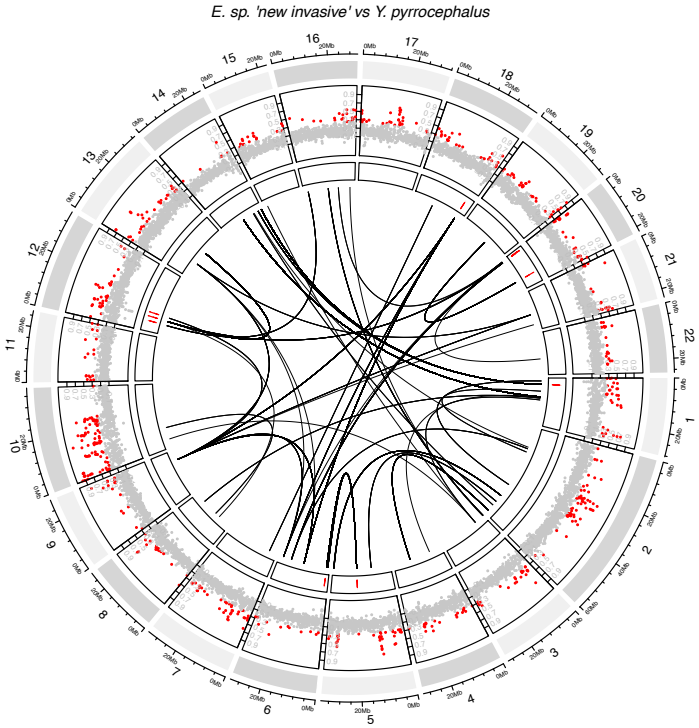


LD pairs with r2>0.8

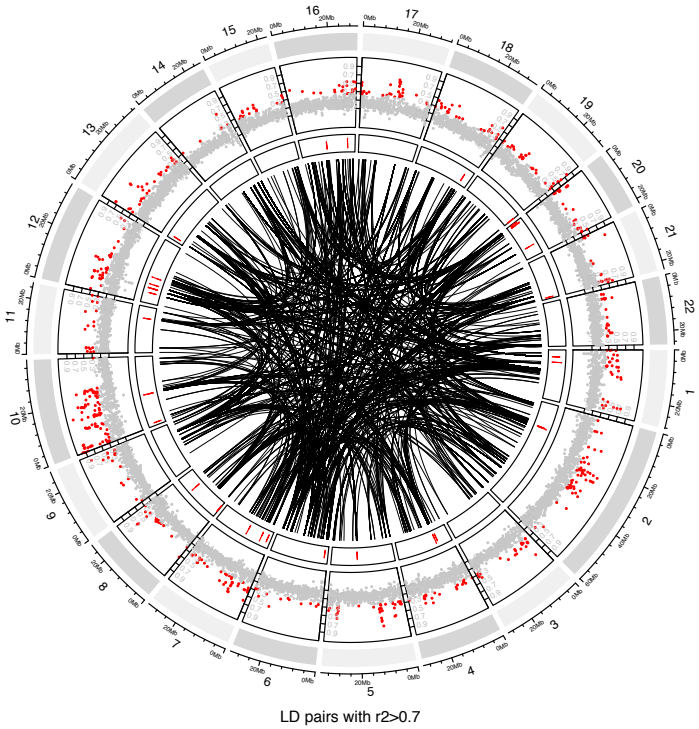


LD pairs with r2>0.7

Figure S12

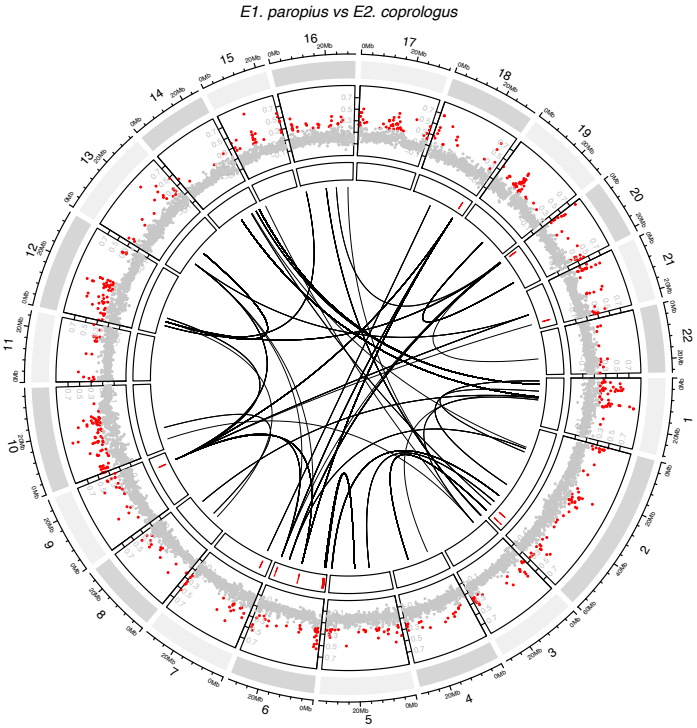


LD pairs with r2>0.8

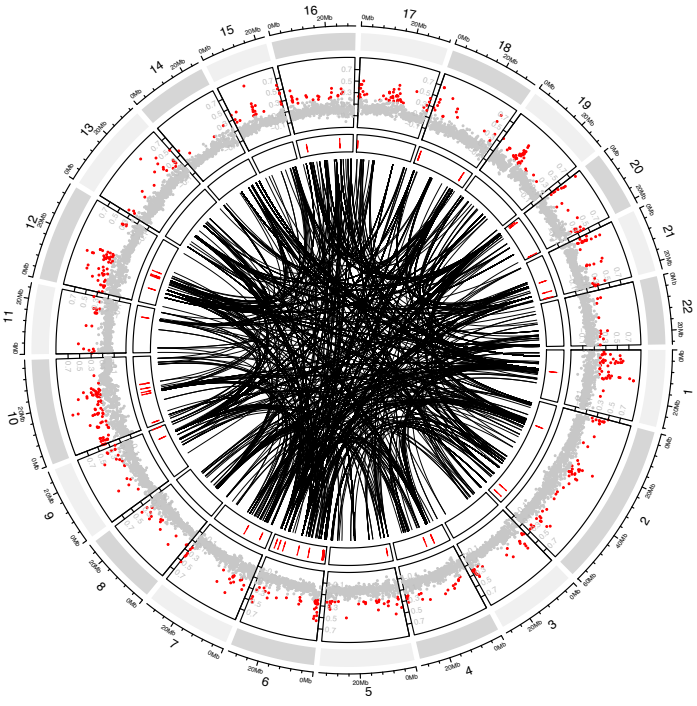


LD pairs with r2>0.7

Figure S13

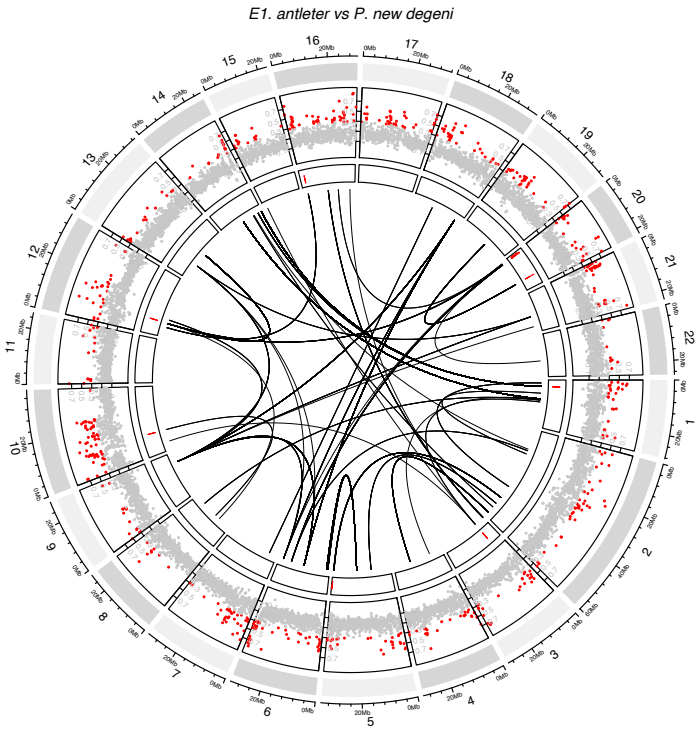


LD pairs with r2>0.8

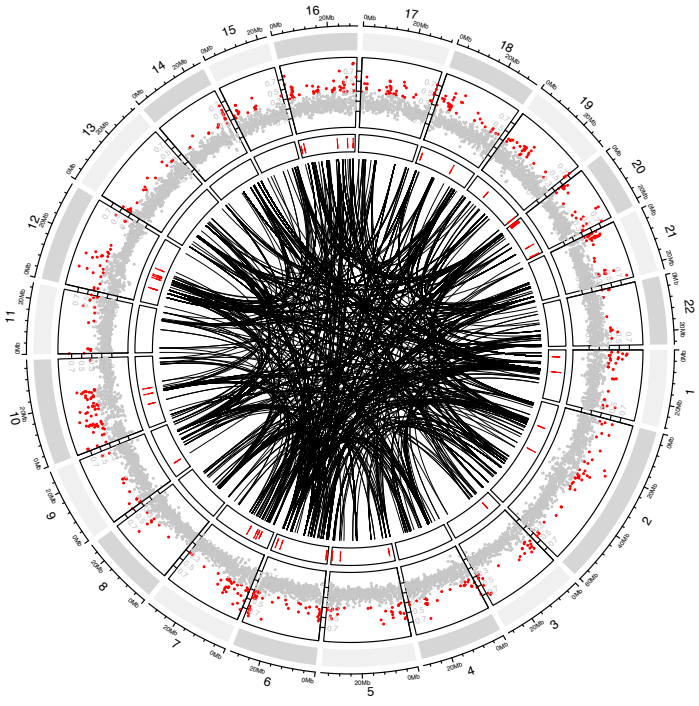


LD pairs with r2>0.7

Figure S14



LD pairs with r2>0.8



LD pairs with r2>0.7

Figure S15

Summary and Synthesis - Towards understanding the genetic architecture of rapid adaptive radiation

Summary and Synthesis - Towards understanding the genetic architecture of rapid adaptive radiation

Adaptive radiations provide intriguing systems to investigate processes and mechanisms that generate high levels of biodiversity. While the evolutionary processes within such radiations operate on local and geographically confined scales, they may well have impacts on global biodiversity (Seehausen, 2015). The African Great Lakes cichlid adaptive radiations are exceptional in terms of the number of species that have evolved within each of them and in light of the short timeframes in which these radiations have unfolded. Especially the Lake Victoria radiation stands out, with over 500 species having evolved within the lake in as few as 15,000 years (Johnson *et al.*, 1996; Seehausen, 1996; Stager & Johnson, 2008; Meier *et al.*, 2017). One hallmark of these cichlid radiations is the coexistence of many closely related species in full sympatry (Seehausen, 2015). How have these species evolved so rapidly? How are they reproductively isolated from one another? How can they persist in geographic immediate proximity, especially if there is ongoing gene flow? These were the broad questions underlying my thesis.

Over the many years of research on these systems it has become clear that several coinciding factors were likely key to the emergence of the spectacular cichlid radiations (Kocher, 2004; Genner & Turner, 2005; Wagner *et al.*, 2012; Brawand *et al.*, 2014; Seehausen, 2015; Meier *et al.*, 2017, 2018; Salzburger, 2018; McGee *et al.*, 2020; Svardal *et al.*, 2020). Among them are ecological opportunity, sexual selection with rapidly evolving mate choice, and a number of genomic features (Wagner *et al.*, 2012; Brawand *et al.*, 2014; Seehausen, 2015; McGee *et al.*, 2020). Especially the latter have received much attention over the past decade. The most recent studies (Brawand *et al.*, 2014; Meier *et al.*, 2017; McGee *et al.*, 2020) show that the most rapidly radiating cichlid lineages feature large amounts of standing genetic variation and extraordinary genomic potential (e.g., a high density of ancient indel polymorphisms), for which a key contributing factor was ancient hybridisation. It has been suggested that genetic variation derived from ancient hybridisation could provide a particularly good substrate for adaptive radiation and rapid speciation because it contains variants and co-adapted haplotypes that have previously been tested by selection, and that (unlike haplotypes that gradually build through new mutations) are at initially high frequencies, which can then be recombined and sorted into many new combinations (Marques *et al.*, 2019).

In my thesis, I investigated the genetic architecture underlying some of the potential key features to rapid diversification and to the emergence and/or maintenance of reproductive isolation and ecological coexistence among Lake Victoria cichlids: a trait under sexual selection which affects mate choice (male nuptial colour; Chapter 1); repeated and rapid transitions between trophic levels (Chapter 2); a variety of sex determining systems (Chapter 3); and intrinsic postzyotic incompatibilities (Chapter 4). I will now briefly summarise and discuss the findings of each chapter, highlighting the novel aspects they provided and suggesting ideas for future work. I will then outline how I think my findings fit in with our current knowledge of this system. Finally, I will reflect on what we have and can still learn from QTL mapping experiments, one of the main tools I used in my thesis.

In **Chapter 1** I compared the genetic architecture of similar red and yellow male nuptial colour motifs between a sympatric species pair and a non-sympatric species pair. Male nuptial coloration is a key cue for female mate choice and male competition, and thus an important trait for reproductive isolation in many species in the Lake Victoria radiation. The pattern where a species with a red dorsum co-occurs with a blue sister species in sympatry is highly repeated across the radiation and occurs in many sympatric species pairs (Seehausen, 1996, 1997, 2009; Seehausen *et al.*, 1998a). In contrast, species with a similar red colour motif, but where red is on the chest instead of the dorsum, are almost never seen in sympatry with a blue sister species. Theory predicts that loci of large effect on a single trait, physical linkage or pleiotropy could provide genetic architectures that allow the persistence of distinct phenotypes in the face of gene flow (e.g. Gavrilets, 2004; Yeaman & Whitlock, 2011; Akerman & Bürger, 2014). In line with this, I found QTLs of moderate to large effects for red and yellow in the representative sympatric species pair, but not in the representative non-sympatric species pair. The absence of any moderate or large effect QTLs for red and yellow in the latter pair likely indicates that a larger number of small effect loci is involved, which could be more easily broken up by recombination if there is gene flow, and would thus quickly lead to the loss of the two distinct phenotypes.

The novel component in my study is the direct comparison of a sympatric vs a non-sympatric pair, which each are representatives of colour motifs that are generally associated with many sympatric (or non-sympatric pairs, respectively) across the radiation. Other recent studies of the genetic architecture of traits associated with ecological variation or mate choice that diverged or persist in the face of gene flow have also found 'simple' genetic architectures involving large effect QTLs and/or tight linkage or pleiotropy, such as for flowering time and flower size traits in *Mimulus* (Ferris *et al.*, 2017), body shape and pharyngeal jaw morphology in Midas cichlids (Fruciano *et al.*, 2016), morphology and optical sensitivity in Lake Massoko cichlids (Malinsky *et al.*, 2015), adaptation in the visual system in *Pundamilia* (Carleton *et al.*, 2010), and colour patterns and mate preference in *Heliconius* butterflies (Merrill *et al.*, 2019). None of these studies however include a direct comparison with a comparable non-sympatric species pair, which I would argue makes it difficult to know if the finding of a certain type of genetic architecture is in fact unique to the investigated scenario with much gene flow. Due to the high replication of similar phenotypes in different geographic settings (see e.g. Seehausen, 1996), the Lake Victoria radiation is a great system to make such comparisons. Many more comparisons involving other species pairs could be made on colour traits on the one hand to consolidate our findings (see also the paragraph on the benefits and limitations of QTL mapping below), and on many ecologically relevant traits on the other hand. Work is already ongoing to map mate choice, which is also predicted to be underpinned by a rather simple architecture (Haesler & Seehausen, 2005; Svensson *et al.*, 2017) but probably not physically linked to colour genes (van der Sluijs *et al.*, 2010).

The focus in **Chapter 2** was on characterising the genetic architecture underpinning a whole suite of traits distinguishing species specialized to feeding at two different trophic levels. Radiation across three trophic levels including herbivory and piscivory starting from invertivorous ancestors (Greenwood, 1974; Joyce *et al.*, 2005; Selz *et al.*, 2014) was one key to the generation of the incredible species richness in the haplochromine cichlid radiations. Transitions between (and even back and forth between) trophic levels have likely occurred multiple times even in the young Lake Victoria radiation (McGee *et al.*, 2020). We hypothesised that a supergene-like architecture might have facilitated rapid transitions between trophic levels.

But contrary to that, I found little evidence for physical linkage or pleiotropy among the regions coding for the different but co-adapted ecologically relevant traits. Instead, I found a distributed genetic architecture with some QTLs of moderate (to large) effect and likely many more of minor effect which we did not have the power to detect. We concluded that multilocus linkage disequilibrium (LD) had to build up between many unlinked regions in the genome at the stage where herbivores evolved from the invertivorous ancestors. In the face of some gene flow, this would have required both divergent selection and the emergence of robust reproductive isolation early on in the process. There is evidence for both in this system (Chapter2; Seehausen *et al.*, 1998b; Magalhaes *et al.*, 2009).

To my knowledge, my study is the first to characterise the genetic architecture of many of the major aspects of ecologically relevant morphology (external morphology, teeth, intestine length) in a cross between African cichlid species specialized for feeding at two different trophic levels (invertivory vs herbivory). Knowledge of the genetic architecture is a key part in understanding how these repeated and rapid transitions between invertivory and herbivory may have occurred. To complete the picture of the genetic architecture underlying traits that need to change in a coordinated fashion in such transitions, it would be highly interesting to also incorporate physiological and behavioural aspects of feeding in these two different trophic niches into this analysis. A preliminary study that I conducted on long chain fatty acid composition in muscle tissue of (lab-bred) individuals of the two parental species did not reveal any differences between them (Feller *et al., unpublished data*). But given that most African cichlid species have likely retained opportunistic capabilities (McKaye & Marsh, 1983; Bouton *et al.*, 1998, 1999; Golcher-Benavides & Wagner, 2019), this may not be too surprising. Other physiological aspects to be investigated could for instance be digestive enzyme composition, or growth rates in relation to phosphorus content (for the latter see e.g. Elser *et al.*, 2000). Selz & Seehausen, 2019 demonstrated differences in feeding efficiency between the two parental species of this cross: they were each more efficient when feeding on the resource they are specialised on, and less efficient when feeding on the resource the other species is specialised on. This is probably affected by both aspects of functional morphology and behavioural components (Seehausen, 2006; Selz & Seehausen, 2019).

The findings of **Chapter 3** add to a growing body of evidence for the high evolvability of sex determination systems in the cichlid radiations. We found sex determining regions on a chromosome that had not previously been known to be sex determining in East African cichlids. One region acts as XY (male heterogametic) system in *Protomelas taeniolatus* from Lake Malawi and the other as ZW (female heterogametic) system in *Pundamilia* sp. 'nyererei-like' from Lake Victoria. Our divergence estimates suggest the sex chromosomes identified in this study have diverged recently (i.e. within the past 20,000 years). In addition to the identified dominant sex determiners, we found evidence for additional sex determiners in two out of the three crosses (*P.* sp. 'nyererei-like' x *P.* sp. 'pundamilia-like' and *Astatotilapia calliptera* and *P. taeniolatus*).

The highly distorted sex ratio in the cross between *P.* sp. 'nyererei-like' and *Neochromis omnicaeruleus* (the cross studied for other traits in chapter 2) suggests that multiple sex determiners were also segregating in this cross, and this is also supported by several other lines of evidence. First, based on our results in Chapter 3, *P.* sp. 'nyererei-like' should have contributed a dominant female sex determiner on Pun-LG14/Ore-LG09, but the only QTL for sex I found (in Chapter 2) was on Pun-LG6/Ore-LG11 (another chromosome that has not previously been found to be sex determining in cichlids!). Second, this QTL only explained a modest 13.5% of the

variation, and males could be all three genotypes at this locus. Previous crossing experiments with individuals from a polymorphic population of *N. omnicaeruleus* suggest that multiple sex determining loci are segregating in these populations, including dominant colour-linked female determiners (W) linked to the original recessive X female determiner and autosomal (recessive) male determiners (Seehausen *et al.*, 1999).

These latter observations have led to the formulation of models of speciation by selection on sex reversal and sexual selection (Seehausen *et al.*, 1999; Lande *et al.*, 2001) and genetic conflict (Kocher, 2004). Gammerdinger & Kocher, 2018 hypothesise in their review that the rapid evolution of sex chromosomes in cichlids might be due to presence of many sexually antagonistic loci across the genome, and that frequent transitions among sex determining regions could be an important factor to rapid speciation in these radiations. Future studies may want to characterise sex determination in many more species to get a better estimate of how many sex determining chromosomes there are in the Lake Victoria radiation, perhaps taking advantage of whole genome sequencing data (where enough male and female individuals of each of several species are available), and by genotyping (intra-specific) laboratory bred families (which would remove the difficult step of tracing back the sex determiners to the parental species in intra-specific crosses; see Chapter 3). A more complete picture of the present sex determining chromosomes might eventually also allow to determine if they arose by 'classic' turnovers (see e.g. Van Doorn & Kirkpatrick, 2007; Blaser *et al.*, 2013, 2014), how fast this would have happened, and if some chromosomes are more prone to be recruited as sex chromosome, as suggested by several studies on cichlid sex determination (Böhne *et al.*, 2016, 2019). Or if alternatively, all these sex determiners were already present in the hybrid lineage that gave rise to the Lake Victoria radiation and were then recombined and sorted into an array of distinct systems (see also the discussion of genetic architectures in the context of adaptive radiation and speciation from a hybrid swarm below). Given how rapidly different sex determining systems have evolved in the Lake Victoria radiation, I would argue that the latter scenario seems more likely.

In **Chapter 4** I combined three datasets and approaches to look for signature of intrinsic postzygotic incompatibilities. Intrinsic postzygotic incompatibilities (BDMIs for short) were long thought not to play an appreciable role in the cichlid radiations because for one, prezgyotic barriers feature so importantly, second, the radiations are so young that the accumulation of BDMIs seems unlikely, and third, BDMIs are deemed unlikely to accumulate during speciation with gene flow (Gavrilets, 1997; Kondrashov, 2003; Bank *et al.*, 2012) (but see Agrawal *et al.*, 2011; Kulmuni & Westram, 2017). However, it has been suggested that postzygotic incompatibilities could play a role for speciation with gene flow from a hybrid swarm (Seehausen, 2013). In this model, sorting of incompatibilities occurs between subpopulations of a hybrid swarm, and by becoming coupled to loci that are under divergent ecological selection they then contribute to reproductive isolation between the divergently adapting emerging species, as well as (indirectly) to adaptive divergence by reducing gene flow between them (Seehausen, 2013). The divergent lineages that seeded the hybrid swarm from which the Lake Victoria radiation evolved (Meier *et al.*, 2017) are of an age which suggests that incompatibilities were very likely to have been segregating between them (Stelkens *et al.*, 2010; Meier *et al.*, 2017), and Meier *et al.*, 2017 indeed also found signatures that might indicate the presence of BDMIs in the radiation, in line with the predictions of the verbal model mentioned above (Seehausen, 2013).

Chapter 4 is the first experimental test of these predictions. The screens I performed in this Chapter revealed patterns that could be consistent with incompatibilities segregating among 94 species of the radiation as well as some signatures of early acting incompatibilities in F2 hybrid crosses. I found that especially the former putative incompatibility loci significantly coincided with regions of high genomic differentiation between most sympatric sister species. More in-depth analyses are required to rule out effects of shared phylogenetic history or shared adaptations among species, and to explore the regions where putative incompatibilities coincide with regions of high genomic differentiation in more detail.

A role for postzygotic barriers?

One thing that emerges from the third and fourth chapter is that postzygotic reproductive isolating mechanisms might indeed also play non-neglectable roles in the cichlid adaptive radiations. Recent reviews recognise that they may play a role (Salzburger, 2018; Rometsch *et al.*, 2020), but to date, this has been largely unexplored (but see van der Slujis *et al.*, 2008; Stelkens *et al.*, 2010, 2015; Meier *et al.*, 2017), and much of the focus of research on these systems has been on premating behavioural and ecological barriers, which undoubtedly play major roles. It remains difficult however, especially in the very young Lake Victoria radiation, to explain how so many closely related species can persist in sympatry with occasional gene flow (where recombination counteracts selection for coupling of adaptation and mate choice loci) if they are only isolated by behavioural mate choice. Also, without any additional mechanisms, disruptive/divergent natural selection alone may rarely be strong enough to cause reproductive isolation (van Rijssel *et al.*, 2018). The additional presence of postzygotic barriers could help explain how so many closely related species can persist in sympatry. The verbal model by (Seehausen, 2013) suggests that selection for coupling of extrinsic ecological barrier loci to intrinsic barrier loci in subpopulations of a hybrid swarm might be a driver of speciation. The results of Chapter 4 hint at the existence of intrinsic incompatibilities. In addition, combining insights from Chapters 3 and 2 shows how diverse sex determining mechanisms are even just among a handful of tested species in the Lake Victoria radiation. The rapid evolution or turnover of sex determining systems might generate postzygotic barriers or create genetic conflicts that could accelerate species divergence (Crespi & Nosil, 2013; Gammerdinger & Kocher, 2018).

It is becoming clearer and clearer that a single mechanism alone is unlikely to explain the rapid cichlid adaptive radiations (see also reviews by Seehausen, 2015; Gammerdinger & Kocher, 2018; Salzburger, 2018; Rometsch *et al.*, 2020), and most species are probably isolated by multiple barriers despite their extreme evolutionary youth. Future work should thus focus on assessing not just one barrier at a time but as many as possible in a given system. The Lake Victoria radiation is especially interesting because it features so many species that are in early stages of differentiation yet are already fully sympatric (see e.g. Meier *et al.*, 2018; van Rijssel *et al.*, 2018).

Genetic architectures in the context of adaptive radiation and speciation from a hybrid swarm

I argue that both genetic architectures found in Chapters 1 and 2 are consistent with a scenario of speciation from a hybrid swarm, or more generally, the recently proposed combinatorial view on speciation and adaptive radiation (Marques *et al.*, 2019). The latter authors propose that the reassembly of ancient genetic variation into new combinations could facilitate rapid speciation and adaptive radiation. Such old genetic variation was for instance

present in the hybrid lineage(s) that then gave rise to the Lake Victoria radiation (and other smaller radiations in the region; (Meier *et al.*, 2017)). These old genetic variants would have contained functionally relevant haplotype polymorphisms (i.e. blocks of co-adapted SNPs that have previously been tested by selection in the divergent lineages before they hybridised) with relatively large phenotypic effects. This would make them a good substrate for the generation of many new combinations -without the need for the emergence of de-novo mutations- that due to their large effects would also have provided genetic architectures robust to gene flow (Gavrilets, 2004; Orr, 2005), and could facilitate the evolution of linkage disequilibrium even in the face of gene flow (Marques *et al.*, 2019).

Overall consistent with this scenario, I found genetic architectures involving at least some moderate to large effect loci for both male nuptial colour traits and for trophic traits in species that broadly overlap geographically. There was little evidence for linkage/pleiotropy in the trophic traits. While I found some evidence for physical linkage/pleiotropy among the QTLs for colour, at least four different chromosomes were involved even for these traits. Thus, both kinds of traits showed a genetic architecture distributed over several chromosomes. More complex polygenic architectures underlying barrier traits under divergent selection may be more conducive to generating and maintaining genome-wide differentiation in sympatry with some gene flow than 'simple' oligogenic architectures, but such a polygenic architecture may not easily evolve in the face of gene flow (Feder & Nosil, 2009; Nosil *et al.*, 2009; Kautt *et al.*, 2020). On the other hand, a 'simple' oligogenic architecture may facilitate the retention of adaptive differentiation against gene flow, but it may not allow the build up of genome-wide differentiation in sympatry with some gene (Feder & Nosil, 2009; Nosil *et al.*, 2009; Kautt *et al.*, 2020). In a scenario of adaptive radiation and speciation from a combinatorial perspective (Marques *et al.*, 2019) however, a polygenic/distributed genetic architecture would not have to evolve de-novo, but would come together by recombining co-adapted haplotype blocks. As these are predicted to have relatively large phenotypic effects, they would simultaneously also provide a genetic architecture that is robust to gene flow and could facilitate genome-wide differentiation.

One prediction emerging from the combinatorial scenario would be that admixture-derived genetic variation might be enriched for loci with large phenotypic effects, which could be tested in Lake Victoria cichlids by comparing the genetic architectures of key traits in radiating lineages derived from the hybrid swarm and non-admixed non-radiating lineages (such as the genus *Astatoreochromis* or haplochromines of Lakes Babati and Chala; (Moser *et al.*, 2018)).

The benefits and limitations of QTL mapping for studying genetic architectures

One major problem that QTL mapping studies pose to adaptive radiation research is that they require sample sizes that are much larger than can be obtained by breeding members of most species of the 'classic' adaptive radiations (Martin & Richards, 2019), such as *Anolis* lizards, Darwin's finches, or Hawaiian silverwords. Knowledge of the genetic architecture of trait diversity has thus remained relatively limited in these radiations. The East African haplochromine cichlids provide a system where -with considerable breeding effort- it is feasible to obtain the minimum required sample sizes for QTL mapping, allowing to map critical traits in species of 'classic' adaptive radiations.

The two QTL mapping-based studies in my thesis have shown that QTL mapping can provide a useful first overview of the genetic architecture of traits of interest even with modest sample sizes. There are limitations, however, to the details that can be inferred from such results.

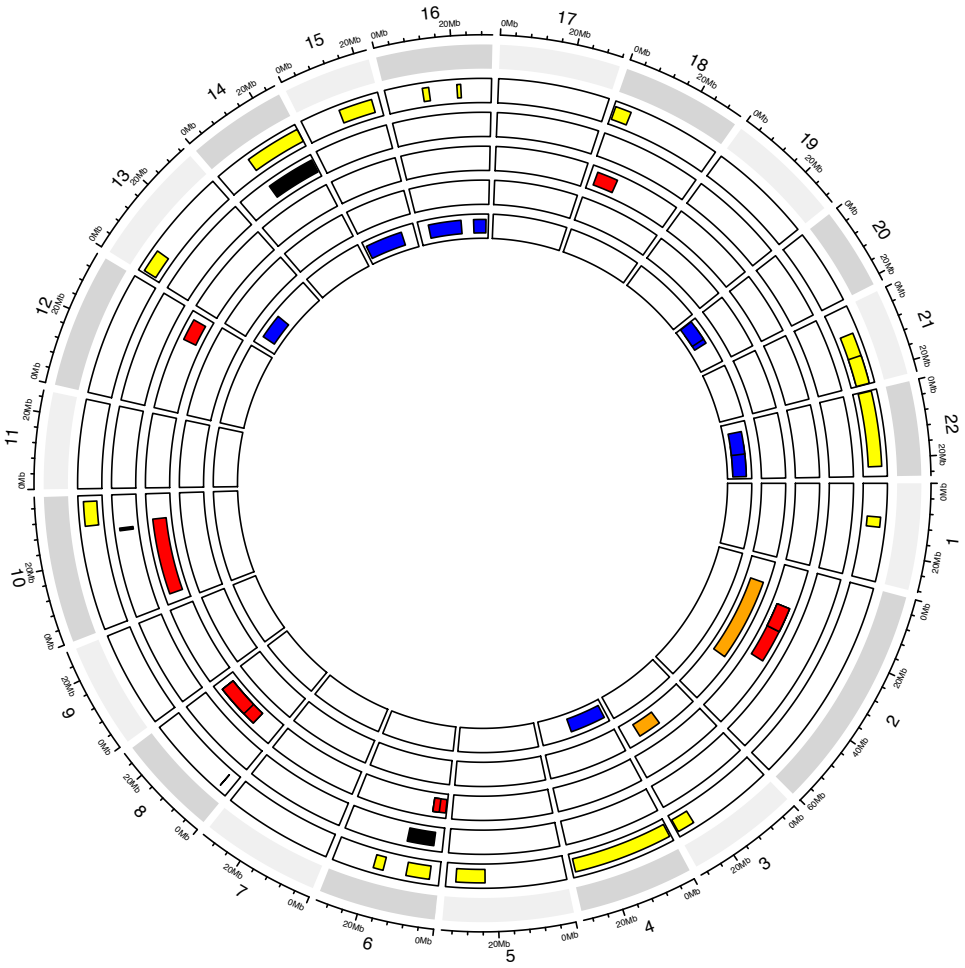
For one, small effect QTLs are almost impossible to detect with small to modest sample sizes of e.g. 100-200 individuals per cross. Hence, the presence of a larger number of small effect QTLs can only be inferred from their absence, or by the often large percentage of variance that remains unexplained by the detected moderate to large effect loci. The location of small effect QTLs and their exact effects on the phenotype remain unresolved. With small to modest sample sizes it is also difficult to get reliable estimates for the effect size of a detected QTL – particularly the effect sizes of large effect QTLs tend to be overestimated (Beavis, 1994; Göring *et al.*, 2001). While knowledge of the exact effect sizes was not a crucial component to the general conclusions on the genetic architectures described in Chapter 2, and did not affect the relative comparisons made in Chapter 1, the distribution and effects of small effect QTLs remains unknown. Furthermore, as the recombination events in a second-generation hybrid cross are still rather limited, the confidence intervals around identified QTLs are typically very large (they can sometimes span almost a whole chromosome), making the task to screen for candidate genes in such a region daunting. However, QTL mapping studies such as the ones preformed in this thesis provide a solid basis to start more in-depth investigations of the exact regions and genes underlying a given trait, if this is of interest.

A promising approach to investigate genetic architectures of traits in more depth may be to combine GWAS and QTL mapping ((Korte & Farlow, 2013) and references therein). However, there are two major (among other) challenges that limit the use of a GWAS approach in wild populations (reviewed in Santure & Garant, 2018): first, it requires very large sample sizes (i.e. many hundreds of individuals) that may not be reasonable (or ethically sound) to obtain in many systems; second, the presence of population and family structure can lead to false positives, but this information may be difficult to obtain. For further QTL mapping studies aimed at investigating the genetic architecture of key traits to rapid diversification in Lake Victoria cichlids, I would thus rather suggest to produce reciprocal crosses with even larger families of more than just two F1 parental pairs.

One major question resulting from such QTL mapping experiments is how repeatable the results from one single cross are. The variation segregating in one cross only ever represents the variation present in one individual per crossed species (the two F0 grandparents), and may not always be representative of all the variation segregating between species (see e.g. Korte & Farlow, 2013). Ideally, one would make several crosses as well as crosses in both directions to get a sense of how much the results of one cross can be generalised. In non-model organisms such as East African haplochromine cichlids however, this quickly reaches limitations in space and time as to generate 200 adult F2 offspring from one cross and two F1-parental couples takes at least three years in cichlids. Despite these limitations, we could still learn a lot more about genetic architectures of key reproductive isolation traits in cichlids by performing QTL mapping studies. For instance, East African haplochromine cichlids are relatively easily inter-breedable that it would even be possible to cross predators with invertivores or herbivores (at least with some additional behavioural tricks), which would provide additional insights on the genetic architectures underpinning traits that have to change in a coordinated way in transitions between all three major trophic levels. Another interesting approach could be to investigate the same sets of several types of traits in crosses among species of non-radiating lineages vs radiating lineages to compare the genetic architectures between them.

Concluding remarks

My thesis highlights two themes that contribute to our understanding of rapid diversification and the evolution of barriers to gene flow in an adaptive radiation: 1) the genetic architectures I found are overall consistent with patterns that might be expected in a scenario of speciation with gene flow from a hybrid swarm (see Synthesis Figure below) 2) and my results provide a promising outlook for the investigation of the less considered, but likely also important, postzygotic barriers in the Lake Victoria cichlid radiation. The cichlid adaptive radiations remain to be an intriguing system to study processes of diversification and speciation. We are still only beginning to understand the processes and mechanisms that generate biodiversity, but every even so small puzzle piece brings us closer to understanding the stunning world around us and hopefully in one way or another contributes to our efforts in protecting it.



Synthesis Figure | Towards understanding the genetic architecture of the fastest known animal adaptive radiation. The genetic architecture of traits that are key to rapid diversification and speciation in the Lake Victoria haplochromine cichlid radiation is highly distributed across the genome and includes some QTLs with moderate to large effects (see Chapters 1+2 for details on the latter). The outermost tract shows the 22 chromosomes in alternating grey tones. The second tract (yellow) shows the locations of segregation distorted regions (i.e. containing putative intrinsic postzygotic incompatibilities) in three experimental hybrid crosses. The third tract (black) shows the 95% confidence intervals of QTLs for sex in three experimental hybrid crosses. The fourth tract (red) shows the 95% confidence intervals of QTLs for red/yellow male nuptial coloration (a trait under sexual selection) in the experimental hybrid cross between the young sympatric species pair *Pundamilia* sp. 'nyererei-like' and *P.* sp. 'pundamilia-like'. The fifth tract (orange) shows the 95% confidence intervals of QTLs for red male nuptial coloration in the experimental hybrid cross between *P.* sp. 'nyererei-like' x *Neochromis omnicaeruleus*. The sixth and innermost tract (blue) shows the 95% confidence intervals of QTLs for a whole suite of trophic traits distinguishing the representative specialised invertivorous species *P.* sp. 'nyererei-like' from the representative specialised herbivorous species *N. omnicaeruleus.* This combined picture of the genetic architecture of multiple key traits is overall consistent with patterns that might be expected by the recombining and sorting of genetic variation derived from ancient hybridization, a process which could facilitate rapid speciation and the evolution of linkage disequilibrium between multiple genomic regions even in the face of gene flow (Marques *et al.*, 2019). In such speciation with gene flow from a hybrid swarm, incompatibilities might be sorted between emerging divergently adapting species and by becoming coupled to regions under ecological selection could contribute to reproductive isolation between them (Seehausen, 2013). It may need further inspection and quantification, but it seems that much of the present overlap between highlighted regions seems to involve regions coding for ecologically relevant traits (blue) and regions containing putative incompatibilities (yellow).

References

Agrawal, A.F., Feder, J.L. & Nosil, P. 2011. Ecological divergence and the origins of intrinsic postmating isolation with gene flow. *Int. J. Ecol.* **2011** 1-15.

Akerman, A. & Bürger, R. 2014. The consequences of gene flow for local adaptation and differentiation: A two-locus two-deme model. *Theor. Pop. Biol.* **80**: 272-288.

Bank, C., Bürger, R. & Hermisson, J. 2012. The limits to parapatric speciation: Dobzhansky-Muller incompatibilities in a continent-Island model. *Genetics* **191**: 845–863.

Beavis, W.D. 1994. The power and deceit of QTL experiments: Lessons from comparative QTL studies. In: *49th Annual Corn & Sorghum Research Conference American Seed Trade Association* (D. B. Wilkinson, ed), pp. 250–266. Chicago, IL.

Blaser, O., Grossen, C., Neuenschwander, S. & Perrin, N. 2013. Sex-chromosome turnovers induced by deleterious mutation load. *Evolution* **67**: 635–645.

Blaser, O., Neuenschwander, S. & Perrin, N. 2014. Sex-chromosome turnovers: The hot-potato model. *Am. Nat.* **183**: 140–146.

Böhne, A., Weber, A.A., Rajkov, J., Rechsteiner, M., Riss, A., Egger, B., *et al.* 2019. Repeated Evolution Versus Common Ancestry : Sex Chromosome Evolution in the Haplochromine Cichlid Pseudocrenilabrus philander. *GBE* **11**: 439–458.

Böhne, A., Wilson, C.A., Postlethwait, J.H. & Salzburger, W. 2016. Variations on a theme: Genomics of sex determination in the cichlid fish Astatotilapia burtoni. *BMC Genomics* **17**: 1–12.

Bouton, N., Os, N. & Witte, F. 1998. Feeding performance of Lake Victoria rock cichlids: testing predictions from morphology. *J. Fish Biol.* **53**: 118–127.

Bouton, N., Witte, F., Van Alphen, J.J.M., Schenk, A. & Seehausen, O. 1999. Local adaptations in populations of rock-dwelling haplochromines (Pisces: Cichlidae) from southern Lake Victoria. *Proc. R. Soc. B Biol. Sci.* **266**: 355–360.

Brawand, D., Wagner, C.E., Li, Y.I., Malinsky, M., Keller, I., Fan, S., *et al.* 2014. The genomic substrate for adaptive radiation in African cichlid fish. *Nature* **513**: 375–381.

Carleton, K.L., Hofmann, C.M., Klisz, C., Patel, Z., Chircus, L.M., Simenauer, L.H., *et al.* 2010. Genetic basis of differential opsin gene expression in cichlid fishes. *J. Evol. Biol.* **23**: 840–853.

Crespi, B. & Nosil, P. 2013. Conflictual speciation: Species formation via genomic conflict. *Trends Ecol. Evol.* **28**: 48–57.

Elser, J.J., Sterner, R.W., Gorokhova, E., Fagan, W.F., Markow, T.A., Cotner, J.B., *et al.* 2000. Biological stoichiometry from genes to ecosystems. *Ecol. Lett.* **3**: 540–550.

Feder, J.L. & Nosil, P. 2009. The efficacy of divergence hitchhiking in generating genomic islands during ecological speciation. *Evolution* **64**: 1729–1747.

Ferris, K.G., Barnett, L.L., Blackman, B.K. & Willis, J.H. 2017. The genetic architecture of local adaptation and reproductive isolation in sympatry within the Mimulus guttatus species complex. *Mol. Ecol.* **26**: 208–224.

Fruciano, C., Franchini, P., Kovacova, V., Elmer, K.R., Henning, F. & Meyer, A. 2016. Genetic linkage of distinct adaptive traits in sympatrically speciating crater lake cichlid fish. *Nat. Commun.* **7**: 1–8.

Gammerdinger, W.J. & Kocher, T.D. 2018. Unusual Diversity of Sex Chromosomes in African Cichlid Fishes. *Genes* **9**: 480.

Gavrilets, S. 2004. *Fitness landscapes and the origin of species*. Princeton University Press, Princeton, NJ.

Gavrilets, S. 1997. Hybrid Zones With Dobzhansky-Type Epistatic. *Evolution* **51**: 1027–1035.

Genner, M.J. & Turner, G.F. 2005. The mbuna cichlids of Lake Malawi: A model for rapid speciation and adaptive radiation. *Fish Fish.* **6**: 1–34.

Golcher-Benavides, J. & Wagner, C.E. 2019. Playing out liem's paradox: Opportunistic piscivory across lake tanganyikan cichlids. *Am. Nat.* **194**: 260–267.

Göring, H.H.H., Terwilliger, J.D. & Blangero, J. 2001. Large upward bias in estimation of locus-specific effects from genomewide scans. *Am. J. Hum. Genet.* **69**: 1357–1369.

Greenwood, P.H. 1974. *Cichlid fishes of Lake Victoria, East Africa: the biology and evolution of a species flock*. John Wright and Sons Ltd., Stonebridge Press, Bristol, UK.

Haesler, M.P. & Seehausen, O. 2005. Inheritance of female mating preference in a sympatric sibling species pair of Lake Victoria cichlids: implications for speciation. *Proc. R. Soc. B Biol. Sci.* **272**: 237–245.

Johnson, T.C., Scholz, C.A., Talbot, M.R., Kelts, K., Ricketts, R.D., Ngobi, G., *et al.* 1996. Late pleistocene desiccation of Lake Victoria and rapid evolution of cichlid fishes. *Science* **273**: 1091–1093.

Joyce, D.A., Lunt, D.H., Bills, R., Turner, G.F., Katongo, C., Duftner, N., *et al.* 2005. An extant cichlid fish radiation emerged in an extinct Pleistocene lake. *Nature* **435**: 90–95.

Kautt, A.F., Kratochwil, C.F., Nater, A., Machado-Schiaffino, G., Olave, M., Henning, F., *et al.* 2020. Contrasting signatures of genomic divergence during sympatric speciation. *Nature* **588**: 106–111.

Kocher, T.D. 2004. Adaptive evolution and explosive speciation: the cichlid fish model. *Nat. Rev. Genet.* **5**: 288–298.

Kondrashov, A.S. 2003. Accumulation of Dobzhansky-Muller incompatibilities within a spatially structured population. *Evolution* **57**: 151–153.

Korte, A. & Farlow, A. 2013. The advantages and limitations of trait analysis with GWAS: A review. *Plant Methods* **9**.

Kulmuni, J. & Westram, A.M. 2017. Intrinsic incompatibilities evolving as a by-product of divergent ecological selection: Considering them in empirical studies on divergence with gene flow. *Mol. Ecol.* **26**: 3093–3103.

Lande, R., Seehausen, O. & Alphen, J.J.M.V. 2001. Mechanisms of rapid sympatric speciation by sex reversal and sexual selection in cichlid fish. *Genetica* **112–113**: 435–443.

Magalhaes, I.S., Mwaiko, S., Schneider, M. V. & Seehausen, O. 2009. Divergent selection and phenotypic plasticity during incipient speciation in Lake Victoria cichlid fish. *J. Evol. Biol.* **22**: 260–274.

Malinsky, M., Challis, R.J., Tyers, A.M., Schiffels, S., Terai, Y., Ngatunga, B.P., *et al.* 2015. Genomic islands of speciation separate cichlid ecomorphs in an East African crater lake. *Science* **350**: 1493–1498.

Marques, D.A., Meier, J.I. & Seehausen, O. 2019. A Combinatorial View on Speciation and Adaptive Radiation. *Trends Ecol. Evol.* **34**: 531–544.

Martin, C.H. & Richards, E.J. 2019. The Paradox Behind the Pattern of Rapid Adaptive Radiation: How Can the Speciation Process Sustain Itself Through an Early Burst? *Annu. Rev. Ecol. Evol. Syst.* **50**: 569–593.

McGee, M.D., Borstein, S.R., Meier, J.I., Marques, D.A., Mwaiko, S., Taabu, A., *et al.* 2020. The ecological and genomic basis of explosive adaptive radiation. *Nature* **586**: 75–79.

McKaye, K.R. & Marsh, A. 1983. Food switching by two specialized algae-scraping cichlid fishes in Lake Malawi, Africa. *Oecologia* **56**: 245–248.

Meier, J.I., Marques, D.A., Mwaiko, S., Wagner, C.E., Excoffier, L. & Seehausen, O. 2017. Ancient hybridization fuels rapid cichlid fish adaptive radiations. *Nat. Commun.* **8**.

Meier, J.I., Marques, D.A., Wagner, C.E., Excoffier, L. & Seehausen, O. 2018. Genomics of parallel ecological speciation in Lake Victoria cichlids. *Mol. Biol. Evol.* **35**: 1489–1506.

Merrill, R.M., Rastas, P., Martin, S.H., Melo, M.C., Barker, S., Davey, J., *et al.* 2019. Genetic dissection of assortative mating behavior. *PLoS Biol.* **17**: 1–21.

Moser, F.N., van Rijssel, J.C., Mwaiko, S., Meier, J.I., Ngatunga, B. & Seehausen, O. 2018. The onset of ecological diversification 50 years after colonization of a crater lake by haplochromine cichlid fishes. *Proc. R. Soc. B Biol. Sci.* **285**: 20180171.

Nosil, P., Harmon, L.J. & Seehausen, O. 2009. Ecological explanations for (incomplete) speciation. *Trends Ecol. Evol.* **24**: 145–156.

Orr, H.A. 2005. The genetic theory of adaptation: A brief history. *Nat. Rev. Genet.* **6**: 119–127.

Rometsch, S.J., Torres-Dowdall, J. & Meyer, A. 2020. Evolutionary dynamics of pre- and postzygotic reproductive isolation in cichlid fishes. *Philos. Trans. R. Soc. Lond. B. Biol. Sci.* **375**: 20190535.

Salzburger, W. 2018. Understanding explosive diversification through cichlid fish genomics. *Nat. Rev. Genet.* **19**: 705–717.

Santure, A.W. & Garant, D. 2018. Wild GWAS—association mapping in natural populations. *Mol. Ecol. Resour.* **18**: 729–738.

Seehausen, O. 2006. African cichlid fish: a model system in adaptive radiation research. *Proc. R. Soc. B Biol. Sci.* **273**: 1987–1998.

Seehausen, O. 2013. Conditions when hybridization might predispose populations for adaptive radiation. *J. Evol. Biol.* **26**: 279–281.

Seehausen, O. 1997. Distribution of and reproductive isolation among color morphs of a rock-dwelling Lake Victoria cichlid (Haplochromis nyererei). *Ecol. Freshw. Fish* **6**: 59–66.

Seehausen, O. 1996. *Lake Victoria Rock Cichlids - taxonomy, ecology, and distribution*. Verduijn Cichlids, Zevenjuizen (NL).

Seehausen, O. 2015. Process and pattern in cichlid radiations - inferences for understanding unusually high rates of evolutionary diversification. *New Phytol.* **207**: 304–312.

Seehausen, O. 2009. Progressive levels of trait divergence along a ' speciation transect ' in the Lake Victoria cichlid fish Pundamilia. In: *Speciation and patterns of diversity* (R. K. Butlin, J. Bridle, & D. Schluter, eds), pp. 155–176. Cambridge University Press, Cambridge, UK.

Seehausen, O., Lippitsch, E., Bouton, N. & Zwennes, H. 1998a. Mbibi, the rock-dwelling cichlids of Lake Victoria: description of three new genera and fifteen new species (Teleostei). *Ichthyol. Explor. Freshwaters* **9**: 129–228.

Seehausen, O., Witte, F., Alphen, J.J.M. & Bouton, N. 1998b. Direct mate choice maintains diversity among sympatric cichlids in Lake Victoria. *J. Fish Biol.* **53**: 37–55.

Seehausen, van Alphen & Lande. 1999. Color polymorphism and sex ratio distortion in a cichlid fish as an incipient stage in sympatric speciation by sexual selection. *Ecol. Lett.* **2**: 367–378.

Selz, O.M., Lucek, K., Young, K.A. & Seehausen, O. 2014. Relaxed trait covariance in interspecific cichlid hybrids predicts morphological diversity in adaptive radiations. *J. Evol. Biol.* **27**: 11–24.

Selz, O.M. & Seehausen, O. 2019. Interspecific hybridization can generate functional novelty in cichlid fish. *Proc. R. Soc. B Biol. Sci.* **286**.

Stager, J.C. & Johnson, T.C. 2008. The late Pleistocene desiccation of Lake Victoria and the origin of its endemic biota. *Hydrobiologia* **596**: 5–16.

Stelkens, R.B., Schmid, C. & Seehausen, O. 2015. Hybrid Breakdown in Cichlid Fish. *PLoS One* **10**: e0127207.

Stelkens, R.B., Young, K.A. & Seehausen, O. 2010. The Accumulation of Reproductive Incompatibilities in African Cichlid Fish. *Evolution* **64**: 617–633.

Svardal, H., Quah, F.X., Malinsky, M., Ngatunga, B.P., Miska, E.A., Salzburger, W., *et al.* 2020. Ancestral Hybridization Facilitated Species Diversification in the Lake Malawi Cichlid Fish Adaptive Radiation. *Mol. Biol. Evol.* **37**: 1100–1113.

Svensson, O., Woodhouse, K., van Oosterhout, C., Smith, A., Turner, G.F. & Seehausen, O. 2017. The genetics of mate preferences in hybrids between two young and sympatric Lake Victoria cichlid species. *Proc. R. Soc. B Biol. Sci.* **284**.

van der Sluijs, I., Seehausen, O., Dooren, T.J.M. Van & Alphen, J.J.M. Van. 2010. No evidence for a genetic association between female mating preference and male secondary sexual trait in a Lake Victoria cichlid fish. *Curr. Biol.* **56**: 57–64.

van der Slujis, I., van Dooren, T.J.M., Seehausen, O. & van Alphen, J.J.M. 2008. A test of fitness consequences of hybridization in sibling species of Lake Victoria cichlid fish. *J. Evol. Biol.* **21**: 480–491.

Van Doorn, G.S. & Kirkpatrick, M. 2007. Turnover of sex chromosomes induced by sexual conflict. *Nature* **449**: 909–912.

van Rijssel, J.C., Moser, F.N., Frei, D. & Seehausen, O. 2018. Prevalence of disruptive selection predicts extent of species differentiation in Lake Victoria cichlids. *Proc. R. Soc. B Biol. Sci.* **285**: 20172630.

Wagner, C.E., Harmon, L.J. & Seehausen, O. 2012. Ecological opportunity and sexual selection together predict adaptive radiation. *Nature* **487**: 366–369.

Yeaman, S. & Whitlock, M.C. 2011. The genetic architecture of adaptation under migration-selection balance. *Evolution* **65**: 1897–1911.

Acknowlegdements

Acknowledgements

To become a biologist has always been my dream. A dream that took a while for me to find the courage to pursue. When after a few years of working as a primary school teacher I finally started my first year of undergrad biology, I had no idea if I would even make it through this first year, and I certainly never imagined I would eventually end up doing a PhD (also because I had no idea what that even meant). Many people have shaped my academic journey since then, in all kinds of different ways, and here I would like to thank each of them.

First, I would like to thank my PhD advisor Ole Seehausen. His second and third year courses on evolutionary biodiversity dynamics and speciation captured my attention like no other courses and fundamentally impacted the trajectory of my subsequent studies. I feel grateful and privileged to have been given the opportunity to study and work under his supervision, in a highly stimulating research environment shaped by his infectious enthusiasm for science.

Second, I thank my PhD co-advisor Katie Peichel for her steady support over these four years. Discussing science and progress and whatever else was up in -over time more and more- regular meetings was incredibly helpful, insightful, and also fun.

Now I will go back in time to highlight some of the experiences that had big impacts along the way: to work as a student assistant in Cris Kuhlemeier's group was a great experience, and a special thanks goes out to Hester Sheehan for having made this happen, and to Korinna Esfeld for teaching me a whole bunch of lab skills. Then, I was lucky enough to have been supervised by David Marques for my Bachelor's project, which was a fantastic experience and which I consider my real introduction into academia. During my Master's project I was supervised by Oli Selz, Matt McGee, and Joana Meier, who all taught me a lot about very different things. In particular, I will never forget the moment when Joana introduced me to the terminal on my Macbook. With great patience she then went on to teach me the basics of sequence processing, which laid the foundation to almost everything that followed. Our later collaboration on the sex determination work has been insightful and thoroughly enjoyable. Also during my Master's thesis time, Salome Mwaiko introduced me to RAD library preparation. In my opinion, Salome is a star in the lab, and FishEc is very fortunate to have her! Andreas Taverna was my go-to person for anything cichlid aquarium-related during this phase, as well as later during my PhD. During the second year of my PhD, I had the opportunity to co-supervise Vera Ogi's Bachelor thesis. Thanks Vera for the courage to take on a difficult project and for sticking it out with us.

Throughout the first three years of my PhD, working at the IEE in Bern once or twice a week was a given. Thanks to everyone in the 'original' Bern crew (Marcel, David, Joana, Florian, Carmela, Tim, Sara) and members of the Peichel group (Sophie, Hilary, Matt J., Diana, Matt Z., Marius) who always made this a good experience.

I feel incredibly fortunate to have been given the opportunity to work at Eawag in Kastanienbaum. In addition to providing a diverse and stimulating research environment in an amazing geographic setting, KB always felt like a family to me. A place like KB would not function without the hard work of people like Beat, Nadja, Michelle, and many others. I will always look back fondly at being part of KB-cookers (with Typhaine, Bene, Kathrin, Cas, Nico, Carmela, Jaime, Matt, Miguel, Jacco, and later on also several 'Rata' members) and at the social events, from cheese nights to Badehüsli events. This past year of pandemic, when all of these things could no longer take place, to me has highlighted even more how just important they are for our general well-being and social and scientific connections.

There were too many people in the FishEc group over the years to mention all of them by name, but I always highly appreciated our Wednesday seminars and the regular scientific exchange. Special thanks go out to: my KB office buddy Cas Retel – it's been a pleasure, and the little coding puzzles solved by you were immensely helpful and also helped improve my own coding. Philine Feulner, who was always there for me next door when I was stuck with some problem, and who assisted me in my very first QTL mapping attempts. Rishi De-Kayne, for being my little big PhD brother and a great friend.

I am grateful to Carmela Doenz, Hanna ten Brink, Moritz Muschick, and Rishi for giving me feedback on parts of my thesis.

Many thanks to Nicola Nadeau for taking the time to evaluate my thesis and acting as external examiner, and to Eva Ringler for chairing my defense.

Finally, my family has provided the stable backbone throughout everything. My parents looked at me with wonder but also pride, when I announced I was going to quit my job and study biology. I think my mum had always envisioned me on a kayak studying whales in the arctic, or at least something along these lines, so she was not surprised. Even if what I ended up doing for my PhD was much more abstract than that, I could always count on their unwavering support and love, and without their financial support I would probably not have made it through my undergrad studies. My grandmother and my uncle Hans provided additional and much appreciated financial support during these early years. My wonderful siblings, Eva and Tobias, were always there when I needed them. Lastly, I thank my partner Kathy – for supporting my in any way she could, for thinking I am superhero, and for always reminding me that ultimately, nothing in life makes sense except in the light of love.

The last clause alludes to T. Dobzhansky's 1973
"Nothing in Biology Makes Sense Except in the Light of Evolution"

Erklärung

gemäss Art. 18 PromR Phil.-nat. 2019

Name/Vorname:

Matrikelnummer:

Studiengang:

Bachelor ☐ Master ☐ Dissertation ☐

Titel der Arbeit:

LeiterIn der Arbeit:

Ich erkläre hiermit, dass ich diese Arbeit selbständig verfasst und keine anderen als die angegebenen Quel-len benutzt habe. Alle Stellen, die wörtlich oder sinn-gemäss aus Quellen entnommen wurden, habe ich als solche gekennzeichnet. Mir ist bekannt, dass andern-falls der Senat gemäss Artikel 36 Absatz 1 Buchstabe r des Gesetzes über die Universität vom 5. September 1996 und Artikel 69 des Universitätssta-tuts vom 7. Juni 2011 zum Entzug des Doktortitels be-rechtigt ist.
Für die Zwecke der Begutachtung und der Überprüfung der Einhaltung der Selbständigkeitserklärung bzw. der Reglemente betreffend Plagiate erteile ich der Univer-sität Bern das Recht, die dazu erforderlichen Perso-nendaten zu bearbeiten und Nutzungshandlungen vor-zunehmen, insbesondere die Doktorarbeit zu vervielfäl-tigen und dauerhaft in einer Datenbank zu speichern sowie diese zur Überprüfung von Arbeiten Dritter zu verwenden oder hierzu zur Verfügung zu stellen.

Ort/Datum

AFelber

Unterschrift

ACADEMIC CURRICULUM VITAE - ANNA F. FELLER

EDUCATION

05/2017 – 06/2021	**Doctoral studies** in Ecology and Evolution, University of Bern, Switzerland. Thesis title*: "On the genetic architecture of an adaptive radiation".* Advisor: Prof. Dr. Ole Seehausen; co-advisor: Prof. Dr. Catherine Peichel
09/2015 – 04/2017	**MSc in Ecology and Evolution** with special qualification in Evolution, University of Bern, Switzerland. Thesis title: "*Genetics of functional novelty in interspecific cichlid hybrids"*. Advisors: Dr. Oliver M. Selz, Dr. Matthew D. McGee, Dr. Joana I. Meier & Prof. Dr. Ole Seehausen
09/2012 – 08/2015	**BSc in Biology** with special qualification in Ecology and Evolution, University of Bern, Switzerland. Thesis title: "*Habitat choice and female preference in a sympatric ecotype pair of sticklebacks"*. Advisors: Dr. David A. Marques & Prof. Dr. Ole Seehausen
08/2005 – 07/2008	**BA in Pre-Primary and Primary Education** and Teaching Diploma, Pädagogische Hochschule (School of Education) Bern, Switzerland

PUBLICATIONS IN PEER-REVIEWED SCIENTIFIC JOURNALS

Feller AFF, Ogi V, Seehausen O, Meier JI (2021): Identification of a novel sex determining chromosome in cichlid fishes that acts as XY or ZW in different lineages. *Hydrobiologia*, https://doi.org/10.1007/s10750-021-04560-7

Feller AFF, Haesler MP, Peichel CL, Seehausen O. (2020) Genetic architecture of a key reproductive isolation trait differs between sympatric and non-sympatric sister species of Lake Victoria cichlids. *Proceedings of the Royal Society London B,* 287: 20200270. http://dx.doi.org/10.1098/rspb.2020.0270

Feller AFF, Selz OM, McGee MD, Meier JI, Mwaiko S, Seehausen O. (2020) Rapid generation of ecologically relevant behavioral novelty in experimental cichlid hybrids. *Ecology and Evolution*, 10:7445–7462. https://doi.org/10.1002/ece3.6471

Marques DA, Lucek K, Haesler MP, **Feller AFF**, Meier JI, Wagner CE, Excoffier L, Seehausen O (2017) Genomic landscape of early ecological speciation initiated by selection on nuptial colour. *Molecular Ecology*, 26, 7–24. https://doi.org/10.1111/mec.13774

Feller AFF, Seehausen O, Lucek K, Marques DA (2016) Habitat choice and female preference in a polymorphic stickleback population. *Evolutionary Ecology Research*, 17: 419–435. https://www.dora.lib4ri.ch/eawag/islandora/object/eawag:10459