The role of ecology and phenotypic plasticity in lake-stream divergence in an East African haplochromine cichlid fish

Inauguraldissertation

zur

Erlangung der Würde eines Doktors der Philosophie

vorgelegt der

Philosophisch-Naturwissenschaftlichen Fakultät

der Universität Basel

von

Jelena Rajkov

von Serbien

Basel, 2020

Genehmigt von der Philosophisch-Naturwissenschaftlichen Fakultät
auf Antrag von

Prof. Dr. Walter Salzburger
Zoologisches Institut, Universität Basel, Schweiz

und

Prof. Dr. Martine E. Maan
Faculty of Science and Engineering, University of Groningen, Niederlande

Basel, den 22.05.2018

Prof. Dr. Martin Spiess
The Dean of Faculty

Contents

Introduction

Understanding the "*origin of species*" has been one of the major topics in biology ever since Charles Darwin's famous publication (Darwin 1859). One and a half centuries later speciation still remains the "*mystery of mysteries*" (Pennisi 2005).

Relative to the early days, modern speciation research has shifted its focus in several ways. Instead of trying to classify speciation mechanisms by geographical context (into allopatric, parapatric and sympatric), we are now trying to understand the processes that lead to speciation (Butlin et al. 2008). While we have learned a lot from traditional laboratory model organisms, we now realize that many questions in speciation research can only be answered by studying how organisms adapt, behave and diversify in their natural environment (Nosil et al. 2017). Documenting the existence of reproductive barriers in well-defined species taught us about the consequences of speciation. However, to understand which reproductive barriers actually initiate the speciation process, we have to study population pairs across a range of isolation levels (Schemske 2010). The above culminated in an important conceptual shift in speciation research from speciation as an endpoint to speciation as a process (Hendry et al. 2009).

Some of the main questions in speciation research at the moment are: "*Which barriers contribute to reproductive isolation*?", "*What are the genomic patterns of reproductive isolation*?", "*What is the role of plasticity*? ", and "*What is the role of changes in gene expression*" (Butlin et al. 2012). With this thesis I aim to provide part of the answers to these questions by studying one of the best model systems for speciation research: the East African cichlid fishes (Kocher 2004).

Speciation

The process of speciation is gradual and involves the evolution of reproductive isolation between formerly panmictic populations (Dobzhansky 1937; Mayr 1963; Schluter 2000). This isolation is often coupled with genetic and phenotypic differences. As reproductive isolation evolves, gene flow between the populations is reduced, which results in accumulation of genetic differences over time. These differences may ultimately lead to complete reproductive isolation (Coyne and Orr 2004). As differentiation accumulates, populations move along the so-called "speciation continuum" from diverging populations, to incipient species, to completely reproductively isolated species (Hendry et al. 2009; Nosil 2012; Seehausen et al. 2014).

Reproductive isolation is generally defined as the product of all barriers to hybridization or gene flow between divergent populations that are in contact (Mallet 2006). Based on their timing in the life cycle of an organism, reproductive barriers can be classified as prezygotic and postzygotic barriers (Coyne and Orr 2004). Prezygotic barriers include habitat and temporal isolation, immigrant inviability, sexual, behavioural, mechanical isolation and gametic incompatibility (Nosil et

al. 2005). Postzygotic barriers can be divided into extrinsic and intrinsic. Extrinsic or environment-dependent postzygotic barriers occur when hybrids are less efficient at exploiting the dominant parental environment. That is, they fall between the parental niches and thus suffer from reduced fitness (Coyne and Orr 2004; Nosil 2012). On the other hand, intrinsic postzygotic barriers occur when hybrids suffer reduced viability and/or fertility due to intrinsic genetic incompatibilities independent of the environment (e.g. Bateson-Dobzhansky-Müller incompatibilities; Dobzhansky 1936; Coyne and Orr 2004; Seehausen et al. 2014).

Understanding the contributions of different components of reproductive isolation to speciation still leaves open the question of their order of emergence (Butlin et al. 2008; Nosil et al. 2009). This question can be approached by documenting reproductive barriers, and their contribution to isolation, in population pairs across a range of genetic divergence.

When it comes to initiating speciation, prezygotic reproductive barriers are often considered more important than postzygotic barriers. This view stems from: (i) the fact that prezygotic barriers act earlier during ontogeny, thus having higher relative contribution to the total reproductive isolation (Sobel et al. 2010), and (ii) the fact that behavioural mate preference (prezygotic) and hybrid sterility (intrinsic postzygotic), the two most commonly measured reproductive barriers (Butlin et al. 2012), differ in that the former is expected to evolve only later in the speciation process (Bolnick and Near 2005; Mallet 2006; Schemske 2010). As a result, we find a large number of studies that detect prezygotic isolation as compared to a relatively small number of studies that detected intrinsic postzygotic reproductive isolation. However, this view does not properly account for extrinsic postzygotic isolation caused by differential adaptation, which might sometimes be the first step in speciation (Schluter 2009).

In ecological speciation, environmental differences fuel divergent selection and ultimately lead to the evolution of reproductive isolation among groups (Rundle and Nosil 2005; Nosil 2012). Ecological speciation is a gradual process, as evidenced by greatly varying levels of divergence between adjacent populations (Hendry 2000; Funk et al. 2006; Mallet et al. 2007; Seehausen et al. 2008). Systems where such gradient exists are ideal for studying speciation, as they offer a possibility to compare “snapshots” of multiple stages along the speciation process and identify the patterns and mechanisms that operate at each stage (Schemske 2010; Nosil 2012). Distinguishing the ways in which divergent selection has led to reproductive isolation is among the greatest challenges of the empirical study of ecological speciation (Schluter 2001).

Fixed genotypic differences and phenotypic plasticity, i.e. the ability of a single genotype to produce different phenotypes depending on the respective environment, have often been viewed as opposing strategies by which organisms can adapt to different environments (Schlichting and Pigliucci 1998; Kawecki and Ebert 2004). However, there is growing evidence that under certain conditions, both mechanisms can complement each other to jointly facilitate novel phenotypes, divergence, colonisation of new habitats and, eventually, speciation (Price et al. 2003; West-Eberhard 2003; Pfennig et al. 2010; Schneider and Meyer 2017).

The role of phenotypic plasticity in the evolution of reproductive isolation has not been widely studied (Butlin et al. 2012), and is often overlooked in selection/introduction experiments in nature, even though they are among the most powerful ways to scrutinize the role of plasticity in adaptation (Ghalambor et al. 2007).

Divergent selection related to parasite pressure is another potentially important factor in ecological speciation that has not received enough attention. Three prerequisites are necessary for parasite-driven speciation: 1) Different populations experience divergent infection levels, 2) these infection levels cause divergent selection and facilitate adaptive divergence, 3) parasite-driven adaptive divergence facilitates the evolution of reproductive isolation (Karvonen and Seehausen 2012; Raeymaekers et al. 2013). The co-evolutionary dynamics between hosts and their sympatric

parasites are expected to lead to local immunogenetic adaptation in the hosts, as well as local adaptation of parasite infectivity and virulence (Kaltz and Shykoff 1998; Kawecki and Ebert 2004; Eizaguirre and Lenz 2010). Thus, an immune response is another potentially relevant trait in ecological speciation.

Empirical studies describing divergent parasitism among host ecotypes are scarce and the only examples come from a few relatively well-known systems (Karvonen and Seehausen 2012). Long-term data on spatiotemporal consistency of divergent parasite infections in replicated pairs of host populations at different stages of speciation is needed to properly address the role of parasitism in speciation (Karvonen and Seehausen 2012).

East-African cichlid fish

Cichlid fishes are one of the most species-rich vertebrate families, whose natural distribution ranges from Central and South America, across Africa and the Middle East to Madagascar and southern India/Sri Lanka (Turner 2007). With their enormous diversity, cichlids are a role model in speciation and adaptive radiation research (Kornfield and Smith 2000; Kocher 2004; Seehausen 2006).

In cichlids, prezygotic behavioural reproductive barriers, in particular female preference for male colouration, are an important component of reproductive isolation (Knight et al. 1998; Kocher 2004; Maan and Sefc 2013). Premating isolation accumulates fast initially but changes little with increasing genetic distance between species (Stelkens et al. 2010). In contrast, intrinsic postzygotic isolation between closely related species is initially negligible but accumulates rapidly, resulting in complete hybrid inviability after 4.4-18.4 million years (Stelkens et al. 2010).

Cichlid fish provide well-described examples of phenotypic plasticity in key ecological traits, such as pharyngeal jaw anatomy, body shape, gill size and brain mass (Greenwood 1964; Meyer 1987; Wimberger 1992; Smits et al. 1996; Bouton et al. 2002; Crispo and Chapman 2010; Muschick et al. 2011). Adaptive phenotypic plasticity in cichlids is hypothesised to be one of the key contributors to their impressive radiations (Galis and Metz 1998; Muschick et al. 2011; Schneider and Meyer 2017).

The possibility that host-parasite coevolution has made a significant contribution to East African cichlid fish diversity has received limited attention, mostly in Lake Malawi (Ono et al. 1993; Blais et al. 2007) and Lake Victoria (Maan et al. 2006, 2008). Despite the fact that Lake Tanganyika's assemblage of cichlid fishes recently gained momentum as a framework to study parasite ecology and evolution (Kmentová et al. 2016a,b; Vanhove et al. 2016), the influence of parasitism and immune system divergence on cichlid speciation in this lake has hardly been considered (see Raeymaekers et al. 2013; Hablützel et al. 2016 for recent exceptions).

Lake Tanganyika is the deepest and the oldest of the three great African Lakes that are home to cichlid adaptive radiations. It accommodates the genetically, morphologically, and ecologically most diverse cichlid species flock (Salzburger et al. 2002).

Haplochromines are the most species-rich lineage within the family of cichlid fishes. This group includes the entire radiations of Lakes Victoria and Malawi (together ~1200 species) and many riverine and lacustrine species elsewhere in Africa (Turner et al. 2001; Verheyen et al. 2003; Schwarzer et al. 2009, 2012) as well as about 30 haplochromine species endemic to Lake Tanganyika (Salzburger et al. 2005). Compared to the cichlid radiations in the lakes, little is known about speciation in riverine cichlids and the faunal connectivity between rivers and lakes. Haplochromine cichlids are among the only known groups spanning the whole speciation continuum (Nosil et al. 2017) and thus provide an ideal model for studying the plethora of

speciation questions that can only be answered by comparing different "snapshots" along this continuum. However, their distribution in remote areas makes studies of the wild populations more challenging than for most other model systems and logistic difficulties make field-experiments in cichlids scarce.

Astatotilapia burtoni (Günther 1894) is a haplochromine cichlid that inhabits both Lake Tanganyika and affluent rivers (Fernald and Hirata 1977; Vos et al. 2001; Kullander and Roberts 2011). The species is among the five African cichlids to have a complete reference genome (Brawand et al. 2014) and constitutes one of the most important cichlid model species in various fields of research including: developmental biology (Thomas-Chollier and Ledent 2008), neurobiology and behavioural biology (Hofmann 2003; Parikh et al. 2006; Dijkstra et al. 2017), and genetics and genomics (Salzburger et al. 2008; Baldo et al. 2011; Diepeveen et al. 2013). Despite being a research model for almost seven decades (e.g. Leong 1969; Wickler 1962), until recently (Theis et al. 2014, 2017) very little was known about the ecology and behaviour of this species in nature. Now, lake and river populations of *A. burtoni* represent an emerging model system for studying early phases of adaptive divergence. Various lake and river "populations pairs" of A. burtoni show similar adaptations to divergent selection regimes, despite different levels of genetic differentiation (Egger et al. 2017). River fish have a shallower body compared to lake fish, which is associated with different water flow regimes in the two habitat types, whereas lake fish have a superior mouth position, longer gill rakers and more elongated lower pharyngeal jaw bones compared to river fish (Theis et al. 2014). These shifts in trophic structures correspond to different diets: while the lake ecotype feeds predominantly on plant/algae and zooplankton, the river ecotype preys upon snails, insects and plant seeds (Theis et al. 2014). River and lake ecotypes also differ in egg-spot number and relative area, and these parameters correlate with adaptive immune response (Theis et al. 2017). This system represents the first cichlid example that allows direct comparison to currently the best-studied speciation continuum in fish - the stickleback lake-river system (McKinnon and Rundle 2002; Berner et al. 2009; Kaeuffer et al. 2011; Stuart et al. 2017). Although replicated morphological divergence was described, no reproductive barrier was found between these ecotypes so far (Theis et al. 2014).

In my thesis, I investigate the degree of reproductive isolation between different populations and the mechanisms that drive the lake-stream divergence. More specifically, I characterize the reproductive barriers contributing to reproductive isolation and investigate the roles of local adaptation, adaptive phenotypic plasticity, parasites and immune system response in lake-stream divergence. By studying divergent lake and river populations I address two main crucial questions: 1) whether ecology may facilitate speciation via increased levels of reproductive isolation, and 2) whether these levels differ within and between replicates with regards to adaptive divergence and genetic distance. In other words, I evaluate whether the genetically most differentiated populations have already reached the last stage of the speciation continuum.

Thesis outline

The main body of work includes four chapters on East-African haplochromine cichlid fishes as a model for studying early phases of adaptive divergence.

In **Chapter 1**, I investigate the relative contributions of adaptive phenotypic plasticity versus local adaptation to fitness of *A. burtoni* populations from Lake Tanganyika and nearby rivers. Adaptive phenotypic plasticity has previously been proposed to play a key role in the impressive radiations of cichlids. However, mostly due to logistic difficulties, field-experiments are scarce in

this system. By performing two transplant experiments in Lake Tanganyika, using both wild caught and common garden raised F1 individuals, I provide one of the first field-based experiments to assess direct fitness consequences of morphological and genetic divergence between lake and stream cichlids in nature.

In **Chapter 2**, I further investigate the correlation between genetic divergence between populations and the level of reproductive isolation between populations by performing a mesocosm mating experiment in a semi-natural setting at Lake Tanganyika. I assessed reproductive isolation in the presence of male-male competition by analysing survival and growth rates of introduced adults and their reproductive success from genetic parentage of surviving offspring. I provide one of the first field-based experiments to assess multiple environment dependent (extrinsic) components of reproductive isolation in cichlids.

In **Chapter 3**, I investigate the potential role of ecotype-specific parasites and the immune response of hosts in the lake-stream divergence. I performed parasitological screening, immune response measurements, and RNA sequencing of immune-related organs of wild fish from two lake-stream systems and fish from a common garden setup.

In **Chapter 4**, three other haplochromine cichlid species that co-exist with *A. burtoni* in Lake Tanganyika and surrounding drainage systems were used to uncover commonalities and differences in adaptation to the river/lake habitat. We sequenced whole genomes of multiple lake-stream population pairs of *A. burtoni* and *Ctenochromis horei*, *Haplochromis stappersi*, and *Pseudocrenilabrus philander*. These natural replicates offered a unique possibility to test whether the same patterns and processes are involved in the adaptation to the lake-stream transition in four haplochromine species. I performed landmark-based geometric morphometric analysis on digital images of fishes of the four species to capture disparity in body shape and compare it with genomic disparity.

Chapters 5 and 6 include additional work I was involved in during my time as a PhD student.

In **Chapter 5**, we characterised the degree of sex chromosomal differentiation in a population of a basal haplochromine cichlid *P. philander* from lake Chila, by investigation of sex-specific sequence patterns. We compared the XX-XY sex determination system identified on LG7 with another XX-XY system in the same genomic region previously characterized in Lake Malawi haplochromine cichlids.

Chapter 6 raises awareness regarding the threats that currently imperil Africa's Great Lakes. Previously inaccessible remote areas are now being prospected for oil and gas deposits. Virtually unnoticed by the public, such activities are ongoing in the East African Great Lakes region, threatening these ecosystems famed for their hyper-diverse biota.

References

Baldo, L., M. E. Santos, and W. Salzburger. 2011. Comparative transcriptomics of eastern African cichlid fishes shows signs of positive selection and a large contribution of untranslated regions to genetic diversity. Genome Biol. Evol. 3:443–455.

Berner, D., A. C. Grandchamp, and A. P. Hendry. 2009. Variable progress toward ecological speciation in parapatry: Stickleback across eight lake-stream transitions. Evolution 63:1740–1753.

Blais, J., C. Rico, C. van Oosterhout, J. Cable, G. F. Turner, and L. Bernatchez. 2007. MHC adaptive divergence between closely related and sympatric African cichlids. PLoS One 2: e734.

Bolnick, D. I., and T. J. Near. 2005. Tempo of hybrid Inviability in centrarchid fishes (Teleostei : Centrarchidae). Evolution 59:1754–1767.

Bouton, N., F. Witte, and J. J. M. Van Alphen. 2002. Experimental evidence for adaptive phenotypic plasticity in a rock-dwelling cichlid fish from Lake Victoria. Biol. J. Linn. Soc. Biol. J. Linn. Soc. Biol. J. Linn. Soc. 77:185–192.

Butlin, R., A. Debelle, C. Kerth, R. R. Snook, L. W. Beukeboom, R. F. Castillo Cajas, W. Diao, M. E. Maan, S. Paolucci, F. J. Weissing, L. van de Zande, A. Hoikkala, E. Geuverink, J. Jennings, M. Kankare, K. E. Knott, V. I. Tyukmaeva, C. Zoumadakis, M. G. Ritchie, D. Barker, E. Immonen, M. Kirkpatrick, M. Noor, M. C. Garcia, T. Schmitt, and M. Schilthuizen. 2012. What do we need to know about speciation? Trends Ecol. Evol. 27:27–39.

Butlin, R. K., J. Galindo, and J. W. Grahame. 2008. Sympatric, parapatric or allopatric: the most important way to classify speciation? Philos. Trans. R. Soc. B Biol. Sci. 363:2997–3007.

Coyne, J. A., and A. H. Orr. 2004. Speciation. Sinauer, Sunderland, MA.

Crispo, E., and L. J. Chapman. 2010. Geographic variation in phenotypic plasticity in response to dissolved oxygen in an African cichlid fish. J. Evol. Biol. 23:2091–2103.

Darwin, C. 1859. On the origin of species by means of natural selection, or the preservation of favoured races in the struggle for life. John Murray, London.

Diepeveen, E. T., O. Roth, and W. Salzburger. 2013. Immune-related functions of the Hivep gene family in East African cichlid fishes. G3 3:2205–17.

Dijkstra, P. D., S. Maguire, R. Harris, A. Rodriguez, R. DeAngelis, S. Flores, and H. Hofmann. 2017. The melanocortin system regulates body pigmentation and social behaviour in the cichlid fish Astatotilapia burtoni. Proc. R. Soc. B Biol. Sci. 284:20162838.

Dobzhansky, T. 1936. Studies on Hybrid Sterility. II. Localization of Sterility Factors in Drosophila Pseudoobscura Hybrids. Genetics 21:113–135.

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

Eizaguirre, C., and T. L. Lenz. 2010. Major histocompatability complex polymorphism: Dynamics and consequences of parasite-mediated local adaptation in fishes. J. Fish Biol. 77:2023–2047.

Fernald, R. D., and N. R. Hirata. 1977. Field study of *Haplochromis burtoni*: quantitative behavioral observations. Anim. Behav. 64:4.

Funk, D. J., P. Nosil, and W. J. Etges. 2006. Ecological divergence exhibits consistently positive associations with reproductive isolation across disparate taxa. Proc. Natl. Acad. Sci. 103:3209–3213.

Galis, F., and J. A. . Metz. 1998. Why are there so many cichlid species? Trends Ecol. Evol. 13:1–2.

Greenwood, P. H. 1964. Environmental effects on the pharyngeal mill of a cichlid fish, *Astatoreochromis alluandi*, and their taxonomic implications. Proc. Linn. Soc. London 176:1–10.

Hablützel, P. I., A. F. Grégoir, M. P. M. Vanhove, F. A. M. Volckaert, and J. A. M. Raeymaekers. 2016. Weak link between dispersal and parasite community differentiation or immunogenetic divergence in two sympatric cichlid fishes. Mol. Ecol. 25:5451–5466.

Hendry, A. P. 2000. Rapid evolution of reproductive isolation in the wild : evidence from introduced salmon. Science 290:516–9.

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

Hofmann, H. A. 2003. Functional genomics of neural and behavioral plasticity. J. Neurobiol. 54:272–282.

Kaeuffer, R., C. L. Peichel, D. I. Bolnick, and A. P. Hendry. 2011. Parallel and non-parallel aspects of ecological, phenotypic, and genetic divergence across replicate population pairs of lake and stream stickleback. Evolution 402–418.

Kaltz, O., and J. A. Shykoff. 1998. Local adaptation in host-parasite systems. Heredity 81:361–370.

Karvonen, A., and O. Seehausen. 2012. The role of parasitism in adaptive radiations-when might parasites promote and when might they constrain ecological speciation? Int. J. Ecol. 2012: 1–20.

Kawecki, T. J., and D. Ebert. 2004. Conceptual issues in local adaptation. Ecol. Lett. 7:1225–1241.

Kmentová, N., M. Gelnar, S. Koblmüller, and M. P. M. Vanhove. 2016a. First insights into the diversity of gill monogeneans of “Gnathochromis” and Limnochromis (Teleostei, Cichlidae) in Burundi: do the parasites mirror host ecology and phylogenetic history? PeerJ 4:e1629.

Kmentová, N., M. Gelnar, M. Mendlová, M. Van Steenberge, S. Koblmüller, and M. P. M. Vanhove. 2016b. Reduced host-specificity in a parasite infecting non-littoral Lake Tanganyika cichlids evidenced by intraspecific morphological and genetic diversity. Sci. Rep. 6:1–15.

Knight, M. E., G. F. Turner, C. Rico, M. J. H. van Oppen, and G. M. Hewitt. 1998. Microsatellite paternity analysisi on captive Lake Malawi cichlids supports reproductive isolation by indirect 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.

Kornfield, I., and P. F. Smith. 2000. African Cichlid Fishes : Model systems for evolutionary biology. Annu. Rev. Ecol. Syst. 31:163–182.

Kullander, S. O., and T. R. Roberts. 2011. Out of Lake Tanganyika: Endemic Lake fishes inhabit rapids of the Lukuga River. Ichthyol. Explor. Freshwaters 22:355–376.

Leong, C.Y. 1969. Quantitative effect of releasers on attack readiness of the fish *Haplochromis burtoni* (Cichlidae, Pisces). Z. Vgl. Physiol. 65: 29-50.

Maan, M. E., and K. M. Sefc. 2013. Colour variation in cichlid fish: Developmental mechanisms, selective pressures and evolutionary consequences. Semin. Cell Dev. Biol. 24:516–528.

Maan, M. E., M. Van Der Spoel, P. Q. Jimenez, J. J. M. Van Alphen, and O. Seehausen. 2006. Fitness correlates of male coloration in a Lake Victoria cichlid fish. Behav. Ecol. 17:691–699.

Maan, M. E., A. M. C. Van Rooijen, J. J. M. Van Alphen, and O. Seehausen. 2008. Parasite-mediated sexual selection and species divergence in Lake Victoria cichlid fish. Biol. J. Linn. Soc. 94:53–60.

Mallet, J. 2006. What does Drosophila genetics tell us about speciation? Trends Ecol. Evol. 21:386–393.

Mallet, J., M. Beltrán, W. Neukirchen, and M. Linares. 2007. Natural hybridization in heliconiine butterflies: The species boundary as a continuum. BMC Evol. Biol. 7:1–16.

Mayr, E. 1963. Animal Species and Evolution. Cambridge MA, Belknap Press.

McKinnon, J. S., and H. D. Rundle. 2002. Speciation in nature: The threespine stickleback model systems. Trends Ecol. Evol. 17:480–488.

Meyer, A. 1987. Phenotypic plasticity and heterochrony in *Cichlasoma managuense* (Pisces, Chichlidae) and their implications for speciation in cichlid fishes. Evolution 41:1357.

Muschick, M., M. Barluenga, W. Salzburger, and A. Meyer. 2011. Adaptive phenotypic plasticity in the Midas Cichlid fish pharyngeal jaw and its relevance in adaptive radiation. BMC Evol. Biol. 11: 116.

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

Nosil, P., J. L. Feder, S. M. Flaxman, and Z. Gompert. 2017. Tipping points in the dynamics of speciation. Nat. Ecol. Evol. 1:1–8.

Nosil, P., L. J. Harmon, and O. Seehausen. 2009. Ecological explanations for (incomplete) speciation. Trends Ecol. Evol. 24:145–156.

Nosil, P., T. H. Vines, and D. J. Funk. 2005. Reproductive isolation caused by natural selection against immigrants from divergent habitats. Evolution 59:705–719.

Ono, H., C. O'hUigin, H. Tichy, and J. Klein. 1993. Major-histocompatibility-complex variation in two species of cichlid fishes from Lake Malawi. Mol. Biol. Evol. 10:1060–1072.

Parikh, V. N., T. S. Clement, and R. D. Fernald. 2006. Androgen level and male social status in the African cichlid, *Astatotilapia burtoni*. Behav. Brain Res. 166:291–295.

Pennisi, E. 2005. What determines species diversity. Science 309:90.

Pfennig, D. W., M. A. Wund, E. C. Snell-Rood, T. Cruickshank, C. D. Schlichting, and A. P. Moczek. 2010. Phenotypic plasticity's impacts on diversification and speciation. Trends Ecol. Evol. 25:459–467.

Price, T. D., A. Qvarnstrom, and D. E. Irwin. 2003. The role of phenotypic plasticity in driving genetic evolution. Proc. R. Soc. B Biol. Sci. 270:1433–1440.

Raeymaekers, J. A., P. I. Hablützel, A. F. Grégoir, J. Bamps, A. K. Roose, M. P. Vanhove, M. Van Steenberge, A. Pariselle, T. Huyse, J. Snoeks, and F. A. Volckaert. 2013. Contrasting parasite communities among allopatric colour morphs of the Lake Tanganyika cichlid Tropheus. BMC Evol. Biol. 13:41.

Rundle, H. D., and P. Nosil. 2005. Ecological speciation. Ecol. Lett. 8:336–352.

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

Salzburger, W., A. Meyer, S. Baric, E. Verheyen, C. Sturmbauer, W. Salzburger, A. Meyer, S. Baric, E. Verheyen, and C. Sturmbauer. 2002. Phylogeny of the Lake Tanganyika cichlid species flock and its relationship to the central and East African Haplochromine cichlid fish fauna. Syst. Biol. 51:113–135.

Salzburger, W., S. C. P. Renn, D. Steinke, I. Braasch, H. A. Hofmann, and A. Meyer. 2008. Annotation of expressed sequence tags for the East African cichlid fish *Astatotilapia burtoni* and evolutionary analyses of cichlid ORFs. BMC Genomics 9:4–6.

Schemske, D. W. 2010. Adaptation and The Origin of Species. Am. Nat. 176:S4–S25.

Schlichting, C. D., and M. Pigliucci. 1998. Phenotypic Evolution: A Reaction Norm Perspective. Sinauer Associates Inc, Sunderland, Massachusetts.

Schluter, D. 2000. The Ecology of Adaptive Radiation. Oxford, Oxford University Press.

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

Schluter, D. 2009. Evidence for ecological speciation and its alternative. Science 323:737–741.

Schneider, R. F., and A. Meyer. 2017. How plasticity, genetic assimilation and cryptic genetic variation may contribute to adaptive radiations. Mol. Ecol. 26:330–350.

Schwarzer, J., B. Misof, D. Tautz, and U. K. Schliewen. 2009. The root of the East African cichlid radiations. BMC Evol. Biol. 9:186.

Schwarzer, J., E. R. Swartz, E. Vreven, J. Snoeks, F. P. D. Cotterill, B. Misof, and U. K. Schliewen. 2012. Repeated trans-watershed hybridization among haplochromine cichlids (Cichlidae) was triggered by Neogene landscape evolution. Proc. R. Soc. B Biol. Sci. 279:4389–4398.

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

Seehausen, O., R. K. Butlin, I. Keller, C. E. Wagner, J. W. Boughman, P. a Hohenlohe, C. L. Peichel, G.-P. Saetre, C. Bank, A. Brännström, A. Brelsford, C. S. Clarkson, F. Eroukhmanoff, J. L. Feder, M. C. Fischer, A. D. Foote, P. Franchini, C. D. Jiggins, F. C. Jones, A. K. Lindholm, K. Lucek, M. E. Maan, D. a Marques, S. H. Martin, B. Matthews, J. I. Meier, M. Möst, M. W. Nachman, E. Nonaka, D. J. Rennison, J. Schwarzer, E. T. Watson, A. M. Westram, and A. Widmer. 2014. Genomics and the origin of species. Nat. Rev. Genet. 15:176–92.

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

Smits, J. D., F. Witte, and F. G. VanVeen. 1996. Functional changes in the anatomy of the pharyngeal jaw apparatus of *Astatoreochromis alluaudi* (Pisces, Cichlidae), and their effects on adjacent structures. Biol. J. Linn. Soc. 59:389–409.

Sobel, J. M., G. F. Chen, L. R. Watt, and D. W. Schemske. 2010. The biology of speciation. Evolution. 64:295–315.

Stuart, Y. E., T. Veen, J. N. Weber, D. Hanson, M. Ravinet, B. K. Lohman, C. J. Thompson, T. Tasneem, A. Doggett, R. Izen, N. Ahmed, R. D. H. Barrett, A. P. Hendry, C. L. Peichel, and D. I. Bolnick. 2017. Contrasting effects of environment and genetics generate a continuum of parallel evolution. Nat. Ecol. Evol. 1:158.

Theis, A., F. Ronco, A. Indermaur, W. Salzburger, and B. Egger. 2014. Adaptive divergence between lake and stream populations of an East African cichlid fish. Mol. Ecol. 23:5304–5322.

Theis, A., O. Roth, F. Cortesi, F. Ronco, W. Salzburger, and B. Egger. 2017. Variation of anal fin egg-spots along an environmental gradient in a haplochromine cichlid fish. Evolution 71: 766–777.

Thomas-Chollier, M., and V. Ledent. 2008. Comparative phylogenomic analyses of teleost fish Hox gene clusters: Lessons from the cichlid fish *Astatotilapia burtoni:* Comment. BMC Genomics 9:1–16.

Turner, G. F. 2007. Adaptive radiation of cichlid fish. Curr. Biol. 17:827–831.

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

Vanhove, M. P. M., P. I. Hablützel, A. Pariselle, A. Šimková, T. Huyse, and J. A. M. Raeymaekers. 2016. Cichlids: A host of opportunities for evolutionary parasitology. Trends Parasitol. 32:820–832.

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

Vos, L. De, J. Snoeks, and D. T. Van Den Audenaerde. 2001. An annotated checklist of the fishes of Rwanda (East Central Africa), with historical data on introductions of commercially important species. J. East African Nat. Hist. 90:41–68.

West-Eberhard, M. J. 2003. Developmental plasticity and evolution. Oxford University Press, Oxford.

Wickler, W. 1962. 'Egg-dummies' as natural releasers in mouth-breeding cichlids. Nature 194: 1092-1093.

Wimberger, P. H. 1992. Plasticity of fish body shape. The effects of diet, development, family and age in two species of Geophagus (Pisces: Cichlidae). Biol. J. Linn. Soc. 45:197–218.

Main Body of Work

East-African haplochomine cichlid fishes as a model
for studying early phases of adaptive divergence

Chapter 1

Adaptive phenotypic plasticity contributes to divergence between lake and river populations of an East African cichlid fish

J. Rajkov, A.A.T. Weber, W. Salzburger, B. Egger

Ecology and Evolution (2018)

All authors were involved in fieldwork. I conducted the molecular laboratory work, the analysis and wrote the manuscript, with contributions from all authors.

Received: 25 April 2018 | Revised: 3 May 2018 | Accepted: 4 May 2018

DOI: 10.1002/ece3.4241

ORIGINAL RESEARCH

WILEY Ecology and Evolution Open Access

Adaptive phenotypic plasticity contributes to divergence between lake and river populations of an East African cichlid fish

Jelena Rajkov | Alexandra Anh-Thu Weber | Walter Salzburger | Bernd Egger

Zoological Institute, University of Basel, Basel, Switzerland

Correspondence
Jelena Rajkov, Zoological Institute, University of Basel, Vesalgasse 1, CH-4051 Basel, Switzerland.
Email: jelena.rajkov@evobio.eu

Funding information
Swiss Zoological Society; Swiss National Science Foundation, Grant/Award Number: 31003A_156405

Abstract

Adaptive phenotypic plasticity and fixed genotypic differences have long been considered opposing strategies in adaptation. More recently, these mechanisms have been proposed to act complementarily and under certain conditions jointly facilitate evolution, speciation, and even adaptive radiations. Here, we investigate the relative contributions of adaptive phenotypic plasticity *vs.* local adaptation to fitness, using an emerging model system to study early phases of adaptive divergence, the generalist cichlid fish species *Astatotilapia burtoni*. We tested direct fitness consequences of morphological divergence between lake and river populations in nature by performing two transplant experiments in Lake Tanganyika. In the first experiment, we used wild-caught juvenile lake and river individuals, while in the second experiment, we used F1 crosses between lake and river fish bred in a common garden setup. By tracking the survival and growth of translocated individuals in enclosures in the lake over several weeks, we revealed local adaptation evidenced by faster growth of the wild-caught resident population in the first experiment. On the other hand, we did not find difference in growth between different types of F1 crosses in the second experiment, suggesting a substantial contribution of adaptive phenotypic plasticity to increased immigrant fitness. Our findings highlight the value of formally comparing fitness of wild-caught and common garden-reared individuals and emphasize the necessity of considering adaptive phenotypic plasticity in the study of adaptive divergence.

KEYWORDS

adaptive phenotypic plasticity, *Astatotilapia burtoni*, cichlid, lake-stream, local adaptation, transplant experiment

1 | INTRODUCTION

Fixed genotypic differences and phenotypic plasticity, that is, the ability of a single genotype to produce different phenotypes depending on the respective environment, have often been viewed as opposing strategies by which organisms can adapt to different environments (Schlichting & Pigliucci, 1998; Kawecki & Ebert, 2004). However, there is growing evidence that under certain conditions, genotypic variability and phenotypic plasticity are complementary mechanisms that jointly facilitate adaptation, speciation and even

adaptive radiation (for reviews see: Price, Qvarnstrom, & Irwin, 2003; West-Eberhard, 2003; Pfennig et al., 2010; Schneider & Meyer, 2017).

In particular, adaptive phenotypic plasticity—the generation of a phenotype that is better suited for a novel environment (Ghalambor, McKay, Carroll, & Reznick, 2007)—can promote the expansion of populations into new niches (Yeh & Price, 2004; Richards, Bossdorf, Muth, Gurevitch, & Pigliucci, 2006; Thibert-Plante & Hendry, 2011). This is because adaptive phenotypic plasticity can temporarily protect genetic diversity from the direct impact of natural selection, thereby saving time for beneficial mutations to arise and to spread within a population, which may eventually result in genetic differentiation (Schlichting, 2004). Whether adaptive phenotypic plasticity facilitates or constrains adaptive divergence depends on how close the "plastic" phenotype is to the fitness optimum in a given environment.

Theory predicts that if there are no fitness costs associated with plasticity, a close match between the "plastic" phenotype and the fitness optimum would lead to stabilizing selection, so that genetic differentiation is unlikely to build up between populations. On the other hand, any incomplete response relative to a new fitness optimum would lead to directional selection with respect to extreme phenotypes (Price et al., 2003; Ghalambor et al., 2007).

Divergent natural selection between populations exposed to different environments leads to divergence in phenotypic traits that influence survival and reproduction. This adaptive divergence should reduce gene flow between populations because nonadapted migrants will suffer increased costs compared to local residents (Hendry, 2001). To experimentally evaluate whether or not adaptive divergence reduces gene flow in nature, it is necessary to perform manipulative field experiments that mimic secondary contact between divergent populations in a natural habitat (Nosil, 2012). Interestingly, the role of phenotypic plasticity is often overlooked in such experiments, even though selection/introduction experiments in nature are among the most powerful ways to scrutinize the role of plasticity in adaptation (Ghalambor et al., 2007).

Reciprocal transplant experiments provide so far the strongest evidence for divergent selection by demonstrating that ecotypes or incipient species suffer from reduced fitness in each other's environment (reviewed in Hereford, 2009). Such studies are commonly performed in plants (reviewed in Leimu & Fischer, 2008) and are becoming more and more common in insects and fish that inhabit temperate habitats of the northern hemisphere (e.g., Räsänen & Hendry, 2014; Soria-Carrasco et al., 2014; Gosden, Waller, & Svensson, 2015; Moser, Frey, & Berner, 2016; Soudi, Reinhold, & Engqvist, 2016; Kaufmann, Lenz, Kalbe, Milinski, & Eizaguirre, 2017). Very few such studies have, however, been conducted with animals that inhabit remote areas in the tropical climate (e.g., Thorpe, Reardon, & Malhotra, 2005; Schwartz, Weese, Bentzen, Kinnison, & Hendry, 2010; Bongaerts et al., 2011; Kenkel & Matz, 2016).

Cichlid fishes are one of the most species-rich vertebrate families, whose natural distribution ranges from Central and South America, across Africa and the Middle East to Madagascar and southern India/Sri Lanka. Cichlids are an important model system in speciation research (Kornfield & Smith, 2000; Kocher, 2004; Seehausen, 2006) and provide well-described examples of phenotypic plasticity in key ecological traits, such as pharyngeal jaw anatomy, body shape, gill size and brain mass (Greenwood, 1964; Meyer, 1987; Wimberger, 1992; Smits, Witte, & VanVeen, 1996; Bouton, Witte, & Van Alphen, 2002; Crispo & Chapman, 2010; Muschick, Barluenga, Salzburger, & Meyer, 2011). Adaptive phenotypic plasticity in cichlids has been proposed to play a key role in their impressive radiations (Galis & Metz, 1998; Muschick et al., 2011; Schneider & Meyer, 2017).

The Haplochromini represent the most species-rich and ecologically diverse tribe of African cichlids (Turner, 2007). Among them, *Astatotilapia burtoni* (Günther, 1894) is an excellent model system to study early phases of adaptive divergence. This generalist species inhabits Lake Tanganyika and affluent rivers (Figure 1a). Adjacent lake and river environments differ in both abiotic and biotic conditions including water parameters, habitat structure, prey composition, and parasite communities (Theis, Ronco, Indermaur, Salzburger, & Egger, 2014; J. Rajkov, W. Salzburger, B. Egger, unpublished data). Various lake and river "populations pairs" in *A. burtoni* show similar adaptations to divergent selection regimes despite different levels of genetic differentiation (F_{ST}) among them (Egger, Roesti, Böhne, Roth, & Salzburger, 2017).

River fish have a shallower body compared to lake fish, which is associated with different flow regimes in the two habitat types, whereas lake fish have a superior mouth position, longer gill rakers as well as more elongated lower pharyngeal jaw bones compared

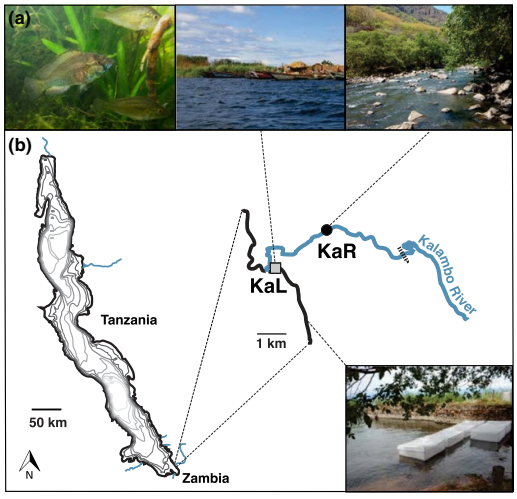


FIGURE 1 *Astatotilapia burtoni* adult male and two females; lake (KaL—Kalambo Lake) and river (KaR—Kalambo River upstream) habitats (a). Lake Tanganyika with inflowing rivers, location of the experimental enclosures and the two populations used in this study (b)

to river fish (Theis et al., 2014). These shifts in trophic structures have been implicated in differential resource use in the two habitat types. Common garden experiments conducted by Theis et al. (2014) demonstrated that differences in body shape and gill raker length have both a plastic and a genetic component and that F1 hybrids are generally intermediate between the parental ecotypes in body shape and gill raker length.

Establishing the link between ecological divergence and fitness differences among populations is crucial to provide evidence that the traits that differ between populations from different habitats are in fact adaptive. In this study, we test the direct fitness consequences of morphological divergence between lake and river *A. burtoni* in nature and evaluate the relative contribution of phenotypic plasticity to fitness and performance (sensu Arnold, 1983). To do so, we performed two independent transplant experiments in Lake Tanganyika, one using wild-caught juvenile lake and river individuals and a second one using different types of F1 crosses between lake and river fish (pure lake, pure river, and hybrids) that were initially bred in ponds filled with lake water.

Our prediction was that, if adaptation to different environment in *A. burtoni* was mainly due to strong local adaptation, resident lake fish would perform better in lake enclosures than foreign river fish in both of our experiments, and that hybrids between lake and river fish would show an intermediate performance between that of the purebred lines. On the other hand, if there was substantial adaptive phenotypic plasticity, we would not expect a difference in performance of the F1 individuals raised in a common habitat.

2 | MATERIAL AND METHODS

2.1 | Study populations and generation of experimental lines

We chose two populations from the Kalambo River system (Figure 1b), a lake population near the estuary (KaL) and an upstream river population (referred to as Ka2 in Theis et al., 2014, 2017; Egger et al., 2017; hereafter referred to as KaR), for the transplant experiments for two main reasons. First, these two populations show the largest difference in body shape and diet composition within any of the lake-stream population pairs examined by Theis et al. (2014): the lake population (KaL) feeds almost exclusively on plant material and algae, whereas the upstream river population (KaR) feeds mostly on macro-invertebrates. Second, the facilities where the experiments could be performed in a sheltered bay protected from waves and fishing activities were adjacent (~3 km) to the location where the KaL population was sampled.

For the first transplant experiment, juvenile fish were collected in October 2015 at the two locations (KaL and KaR), using baited minnow traps. For the second transplant experiment, we generated an F1 cohort by crossing wild-caught adult individuals from the two source populations (KaL and KaR) to create pure lake (KaL × KaL), pure river (KaR × KaR), and hybrid (KaL × KaR) individuals. The parental specimens were caught at the source locations using fishing rods in November 2015. The crosses were raised in concrete ponds supplied with lake water and rocks for shelter between November 2015 and July 2016. Fish were fed with commercial flake food. We used 6–7 ponds for every type of cross, with one male and 3–10 females in each pond to maximize adult and juvenile survival. Hybrid crosses were created in both directions (river female × lake male, river male × lake female). Fish collection in the wild and the transplant experiments were performed under study permits nr. 003376, 004264, 004266 and 004273.

2.2 | Study design

In order to test fitness consequences of the morphological divergence between lake and river *A. burtoni* and to evaluate the relative contribution of phenotypic plasticity to fitness and performance in nature, we performed two transplant experiments, one with wild-caught juveniles from the two populations (KaL and KaR) and one with juveniles from F1 crosses among and between these populations (KaL × KaL, KaR × KaR, and KaL × KaR) (Figure 2). The aim of the first experiment with wild-caught individuals was to mimic the natural situation in the case of migration between environments to test for possible immigration barriers. For example, in this experiment individuals could suffer additive parasite infection resulting from early exposure within their habitat and late exposure to parasites after transplant, just as it would occur for natural migrants (Kaufmann et al., 2017). The use of F1 offspring in the second experiment permitted us to assess the effect of plasticity on fitness in a foreign habitat. To this end, we acclimated individuals from both populations for one generation under common conditions (lake water) similar to environmental conditions under which their fitness was going to be measured as advised in Kawecki and Ebert (2004). In both experiments genetic samples from all individuals were taken during all measurements to enable tracking of individual fish using microsatellite genotyping. Experiments were performed in up to six enclosures (Figure 2, S1) positioned in a sheltered bay ~3 km south of the location where KaL population was sampled. The enclosures were 2 m × 2 m × 1 m and positioned in the lake so that they were filled to less than 1 m. They were built in August 2015 using metal poles and 8 mm square mesh, thus allowing the passage of small organisms across the enclosure walls.

In experiment 1, the enclosures were stocked with 30 individuals each at the end of the dry season (low lake water level). In experiment 2, the enclosures were stocked with 60 individuals each during the mid dry season (higher lake water level) (Figure S1).

In this study, we were limited to one-way transplant experiments in the lake habitat. Although including the reciprocal experimental setup – transplanting lake fish into the river habitat – would have been desirable, this was not feasible given the highly variable river environment and local fishing activities.

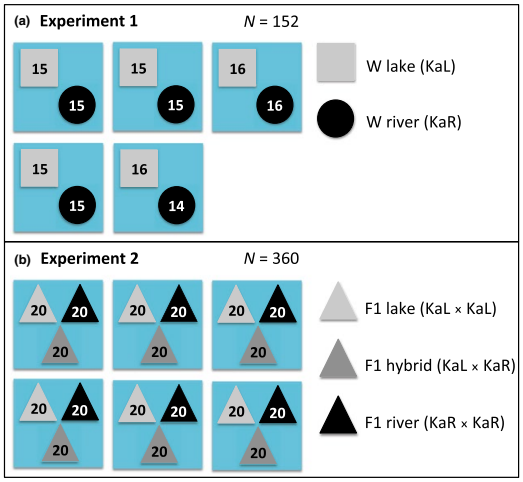


FIGURE 2 Experimental design of the transplant experiments with sample sizes (blue rectangles indicate enclosures). Experiment 1 with wild-caught individuals (a), and experiment 2 with F1 crosses raised in ponds with lake water (b). KaL—Kalambo Lake, KaR—Kalambo River upstream

2.3 | Transplant experiment 1: Wild-caught juvenile performance

Wild-caught juveniles were photographed with a digital camera (Nikon D5000) on their left side, measured with a ruler (±0.5 mm), weighed on an electronic balance (±5 mg), sexed if possible by visual inspection of external coloration and the genital papilla, fin-clipped, and tagged with visible implant elastomer tags (VIE, Northwest Marine Technology) before the start of the experiment. Each individual received a population tag (KaL - front left side of the dorsal fin, KaR - front right) to enable subsequent sorting, size matching and counting of recaptured individuals. Experimental fishes were selected for size and sex to achieve a similar size distribution between the two populations within each enclosure and a ~1:1 sex ratio in each population. After this treatment, all the individuals could recover for 24 hr in concrete tanks filled with lake water (one tank per enclosure) to ensure that fish were all in good shape.

Prior to the release of *A. burtoni*, all enclosures were emptied of wild fish and potential predators by angling and extensive minnow trapping and a fine net skirt was sewn to the inside of each cage and buried to prevent fish from escaping. The enclosures were covered on the top with removable 8 mm mesh lids to prevent bird predation. In October 2015 each of the five enclosures used in this experiment was stocked with 15 individuals from the lake (KaL) and 15 from the river (KaR) population, except for enclosure 3, which was stocked with 16 individuals of each type, and enclosure 5, which was stocked with 16 individuals of KaL and 14 of KaR due to handling errors (n = 152 total) (Figure 2).

The enclosures were checked twice every day and sampled 15 days post-release and again after 24 days, which is when the experiment was terminated. We chose this duration because we wanted to terminate the experiment before the individuals were sexually mature and could start reproducing, to prevent the confounding effect of mouthbrooding and egg laying on female weight gain, as well as possible introduction of non-native populations and hybrid offspring in the wild. To sample the fish in the enclosures, we set 10 minnow traps with inaccessible bait in tea infuser spoons per enclosure one hour before dusk and removed them one hour after dawn on the following day. Recaptured individuals were then assigned to their source population, counted, measured, weighed, sexed if possible, fin-clipped and a photograph was taken, providing survival information as fitness measure and body mass information as performance measure related to fitness (sensu Arnold, 1983). After the first measurement (15 days post-release), all fish were set back into their original enclosures. After the second measurement, all recaptured individuals were euthanized with an overdose of clove oil, dissected to confirm their sex, fin-clipped, and preserved in ethanol.

Genomic DNA from fin clips taken at every time point was extracted using 5% Chelex solution (Casquet, Thebaud, & Gillespie, 2012). The samples were genotyped at five microsatellite loci (Ppun5, Ppun7, Ppun21, UNH130, and Abur82) following the methods described in Theis et al. (2014). Samples from the same individuals taken at different time points were matched using the R package Allelematch (Galpern, Manseau, Hettinga, Smith, & Wilson, 2012) to identify individual fish and to obtain individual-level data for survival and growth.

2.4 | Transplant experiment 2: F1 generation and hybrid juvenile performance

All available F1 offspring were pooled per cross type (KaL × KaL, KaR × KaR or KaL × KaR) before the beginning of the experiment, and experimental individuals were selected from that pool with the aim of achieving a similar size distribution between different types of crosses within each enclosure and a ~1:1 sex ratio in each cross type. Selected individuals were tagged (KaL × KaL – front left side of the dorsal fin, KaR × KaR – front right, KaL × KaR – middle right) with the VIE tags. In July 2016, during the mid dry season, each of the six enclosures was stocked with a total of 60 juvenile *A. burtoni* from our F1 line, whereby 20 juvenile individuals were taken from the pure lake cross (KaL × KaL), 20 from the pure river cross (KaR × KaR) and 20 from the hybrid cross (KaL × KaR), resulting in a total number of experimental fish of n = 240 (Figure 2). The densities used in the experiments are close to those observed at the Kalambo lake location where dozens of fish are typically caught in an empty minnow trap within minutes. Fish were measured after 14 and 28 days as described above. Termination of the experiment, including the microsatellite genotyping, was performed as in experiment 1 (see above).

2.5 | Data analysis

We assessed survival between different experimental populations using generalized linear mixed effect models (GLMMs) with survival

as a dependent variable (coded as 0:dead and 1:alive) and population (lake, river, (hybrid)), initial mass, sex (male, female, immature), size deviation (deviation in initial mass from the mean mass per cage), and their interaction (sex: size deviation) as fixed predictors. The replicated enclosures were set as a random effect. The significance of fixed effect parameters was determined by type II χ^2-based likelihood-ratio tests (based on a binomial distribution with logit function; glmer and drop1 function in R).

We calculated absolute growth rates in mg/day and specific growth rates (SGR = 100*(ln (final mass)-ln (initial mass))/time) for survivors. To correct for individual differences in mass at the beginning of the experiment, specific growth rates were regressed on initial mass. The residual SGR values (rSGR) were used as a measure of relative growth performance (following Scharsack, Kalbe, Harrod, & Rauch, 2007). We assessed growth rates between different experimental populations using linear mixed effect models (LMMs) with growth rate or rSGR as a dependent variable, population (lake, river, (hybrid)), and sex (male, female, immature) as fixed predictors. The replicated enclosures were set as a random effect. The significance of each variable was tested with type II ANOVAs with Kenward-Roger correction for *F*-statistics and *df* in linear mixed models (lmer and ANOVA functions in R).

Some individuals were still immature at the end of the experiment, without visible genital papilla or sex-specific coloration and thus it was not possible to sex them (sexed as immature). We found sex to be the dominant effect in the survival and growth analysis, especially when the immature individuals were included, and therefore, we subsequently conducted the analysis on adults only (immature individuals excluded) and on each sex separately.

Generalized linear mixed effect models and LMMs were calculated with the R package lme4 (Bates, Maechler, Bolker, & Walker, 2015). Significance level for the fixed effects was obtained using the drop1 function of the lme4 package for GLMMs and lmerTest package (Kuznetsova, Brockhoff, & Christensen, 2017) for LMMs. Tukey-Kramer posthoc tests were applied to test for significance of pairwise comparisons between populations using the lsmeans package (Lenth, 2016). All statistical analyses were performed in R version 3.3.2 (R Core Team, 2016).

3 | RESULTS

3.1 | Transplant experiment 1: High overall survival of wild-caught individuals and faster growth of resident population compared to non-residents

The survival was high in this experiment (92%) and did not differ between the lake and river fish (population $\chi^2_{df=1} = 0.340$, $p = 0.560$) (Table 1a, Figure 3a). When only adults were analyzed, there was an effect of size deviation between the experimental fish on survival (size deviation $\chi^2_{df=1} = 4.513$, $p = 0.034$). When male and females were analyzed separately size deviation only had an effect on male survival (size deviation $\chi^2_{df=1} = 6.373$, $p = 0.012$).

Models with relative (rSGR) (Table 2) and absolute growth rate (Table S1) showed comparable results, so we only discuss the results for the relative growth rate here. Absolute growth rate values are shown in Figure S2. Relative growth rate was associated with sex and population of origin (sex $F_{2,131} = 12.229$, $p < 0.001$; population $F_{1,130} = 7.665$, $p = 0.006$) (Table 2a, Figure 4a). When immature individuals were excluded, the effect of sex was comparable to the population effect (sex $F_{1,106.2} = 5.958$, $p = 0.016$; population $F_{1,106.2} = 5.739$, $p = 0.018$). Lake fish grew faster than river fish in their local environment, and males grew faster than females. Relative growth rate was higher in the lake males than in river males (population $F_{1,64.8} = 6.509$, $p = 0.013$), but was not different between the lake and river females (population $F_{1,38.9} = 0.104$, $p = 0.749$).

3.2 | Transplant experiment 2: Low overall survival and no growth differences among F1 individuals

Survival was much lower in this experiment (42%) than in the experiment with wild-caught individuals (Figure 3). The number of survivors per enclosure showed a strong positive correlation with variance in size (standard length at the beginning of the experiment) between individuals within the same enclosure ($r = 0.87$, $R^2 = 0.7$, $p = 0.024$), meaning that survival was higher in enclosures with more variance in size among individuals. There was a sex effect on survival (sex, $\chi^2_{df=1} = 41.91$, $p < 0.001$, Table 1b) due to higher mortality of immature individuals. When only adults were analyzed, there was no apparent difference in survival among different crosses (population $\chi^2_{df=1} = 5.784$, $p = 0.055$), with a tendency of lower survival in river individuals than hybrids (*post hoc* test: KaR x KaR – KaL × KaR, $p = 0.074$, Figure 3b).

Unlike in experiment 1, there was no difference in rSGR among different types of crosses (population $F_{1,142.5} = 1.749$, $p = 0.178$, Table 2, Figure 4b), with a tendency of lower rSGR in lake individuals than hybrids (*post hoc* test: KaL × KaL – KaL × KaR, $p = 0.092$). As in experiment 1, males grew faster than females (sex, $F_{1,141.7} = 36.653$, $p < 0.001$).

4 | DISCUSSION

The goal of this study was to test for local adaptation in divergent lake and river populations of a generalist East African cichlid fish. Using two different setups, one with wild-caught individuals and one with F1 crosses including hybrids raised in a common environment, we were able to examine the contribution of phenotypic plasticity to the adaptation of these populations to different environments. We thus provide the first demonstration of adaptive divergence between lake and river populations of a cichlid species at the level of whole-organism performance, evidenced by higher growth rates in the wild-caught resident population compared to nonresident fish in the first experiment.

TABLE 1 Generalized linear mixed models of survival for *A. burtoni* transferred to lake habitat (*df*: degrees of freedom)

	(a) Experiment 1 - wild-caught fish			(b) Experiment 2 - F1 crosses		
Model 1: whole dataset						
	Residuals *df* = 140			**Residuals *df* = 362**		
Effect	***df***	**χ^2**	***p***	***df***	**χ^2**	***p***
Sex	2	1.036	0.596	**2**	**41.910**	**<0.001**
Population	1	0.340	0.560	**2**	**6.034**	**0.049**
Initial mass	1	1.032	0.310	1	0.300	0.584
Size deviation	1	1.169	0.280	1	0.864	0.353
Sex: size deviation	**2**	**9.968**	**0.007**	**2**	**6.761**	**0.034**
Model 2: immature individuals excluded						
	Residuals *df* = 111			**Residuals *df* = 318**		
Effect	***df***	**χ^2**	***p***	***df***	**χ^2**	***p***
Sex	1	0.106	0.745	1	0.441	0.507
Population	1	0.243	0.622	2	5.784	0.055
Initial mass	1	0.089	0.765	1	0.371	0.542
Size deviation	**1**	**4.513**	**0.034**	1	1.040	0.308
Sex: size deviation	1	1.650	0.199	1	2.435	0.119
Model 3: males only						
	Residuals *df* = 68			**Residuals *df* = 172**		
Effect	***df***	**χ^2**	***p***	***df***	**χ^2**	***p***
Population	1	0.007	0.935	2	2.729	0.256
Initial mass	1	1.803	0.179	1	0.871	0.351
Size deviation	**1**	**6.373**	**0.012**	1	3.131	0.077
Model 4: females only						
	Residuals *df* = 40			**Residuals *df* = 142**		
Effect	***df***	**χ^2**	***p***	***df***	**χ^2**	***p***
Population	1	0.018	0.892	2	3.217	0.200
Initial mass	1	0.269	0.604	1	0.094	0.760
Size deviation	1	0.258	0.611	1	0.033	0.856

Experiment 1—wild-caught fish (a), experiment 2—F1 crosses (b). Significant effects ($p < 0.05$) are highlighted in bold.

On the other hand, we found a strong contribution of adaptive phenotypic plasticity, evidenced by equal growth rates between different types of F1 crosses in the second experiment. In the following, we discuss the findings of this study in the context of adaptive divergence.

4.1 | Higher survival of wild-caught fish and mortality due to male aggression

Survival in experiment 1 using wild-caught individuals was much higher than in experiment 2 using F1 crosses (92 *vs.* 42%). A likely explanation for this result is the lower density of fish in the first experiment (30 *vs.* 60 individuals per enclosure). Alternatively, fitness of wild-born fish could generally be higher. A transplant experiment in trout, for example, revealed that wild-born individuals consistently outperformed both the foreign laboratory-born groups and their laboratory-born locally produced counterparts (Westley, Ward, & Fleming, 2012).

Survival in experiment 2 was higher in enclosures with more variance in size among the individuals. A similar observation was recently reported in threespine stickleback fish, in which survival was lower for average-sized individuals within a cage than for individuals whose initial mass was much larger or smaller than their cage mean (Bolnick & Stutz, 2017).

We further found that male survival in experiment 1 was affected by individual's deviation in size from the mean size per

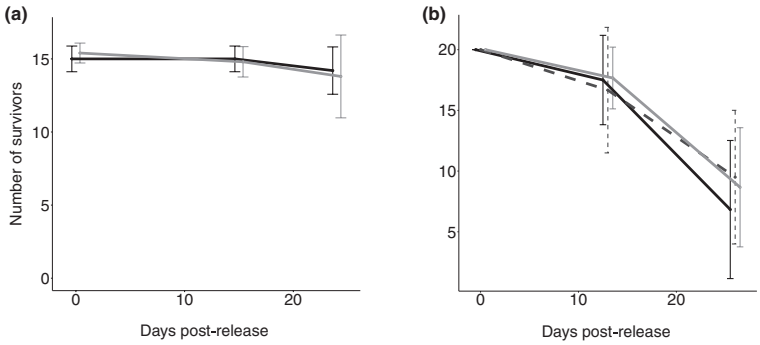


FIGURE 3 Survival (expressed as the average number of surviving fish ±CI 95%) in the lake habitat for wild-caught individuals (a) and F1 crosses (b). Lake (light gray), river (black) and hybrid (dark gray dotted) individuals

enclosure, indicating that male aggression was the most likely causal factor for mortality. This is further substantiated by the inspection of the deceased individuals that we were able to recover on the water surface during the controls of the enclosures; we found that these fish had injuries, likely from fights with conspecifics. *A. burtoni* males are known to be territorial and highly aggressive toward conspecifics (Fernald & Hirata, 1977; Fernald, 1980), and a size difference of less than 10% body length has been shown to provide a significant advantage to the larger opponent in territorial combats (Alcazar, Hilliard, Becker, Bernaba, & Fernald, 2014).

4.2 | Higher growth rate in wild-caught lake fish but not in F1 crosses

As predicted for local adaptation, we found higher growth rates in wild-caught resident lake individuals in their native environment in experiment 1 compared to river fish. Yet, contrary to our prediction for local adaptation, there was no apparent difference in growth of F1 individuals in experiment 2. The lake-river hybrids that were expected to show intermediate performance even grew slightly faster than purebred F1 individuals (Figure 4b). This mirrors results from other systems in which the fitness of some hybrid genotypes equals or exceeds that of purebreds (Rundle, 2002). A large body of research on stickleback provides possible explanations for our results. In a recent stickleback study, Best et al. (2017) found that F1 hybrids performed best in a mesocosm experiment and suggested that this might result from increased heterozygosity in hybrids, helping them overcome the effects of mildly deleterious alleles, or from novel combinations of dominant alleles at different loci. However, because F1s tend to be heterotic, and outbreeding depression is often not expressed until the F2 or later generations, conclusions about the relative fitness of hybrids must be tentative (Lexer, Randell, & Rieseberg, 2009). River stickleback, whether migrants or residents, were found to generally grow faster than lake fish (Scharsack et al., 2007; Kaufmann et al., 2017), suggesting a river-specific trait of faster growth in this species. It has also been suggested that selection on juvenile hybrid stickleback may be weaker than detected in adults (Hatfield & Schluter, 1999). Future studies should aim to investigate fitness of adult hybrids between *A. burtoni* lake and river populations.

In our experiments it was not possible to directly compare the growth rates of wild-caught river fish and F1 generation raised in lake water due to the different seasons in which the two experiments were performed and due to different densities of fish per enclosure. Growth rates were higher in the first experiment with wild-caught individuals, which was, however, performed with lower density and at the beginning of the rainy season when water temperature is higher.

4.3 | Adaptive phenotypic plasticity in *Astatotilapia burtoni*

We found no difference in performance between purebred lake and river crosses in experiment 2 indicating that the juveniles raised in ponds with lake water developed phenotypes with equal fitness as the residents in their non-native environment.

A common garden experiment that examined plasticity *vs.* genetic contribution for body shape and gill raker length in lake and river *A. burtoni* from the same river (Kalambo) found that F1 offspring from between-population crosses display intermediate phenotypes in comparison with within-population crosses (Theis et al., 2014). However, it was also found that the differences between the within-population crosses raised in the ponds were much smaller than the differences observed between wild types. Moreover, offspring of pure river crosses raised in ponds with lake water was closer to lake fish from the wild than to river fish from the wild with respect to body shape and gill raker length. This indicates that the change in the mean trait values is in the same direction favored by selection in the new environment, but below the new adaptive peak, which is one of the conditions for adaptive phenotypic plasticity to facilitate adaptation (Ghalambor et al., 2007). Theory predicts that at intermediate levels of adaptive plasticity the produced phenotype moves into the attractive domain of the higher fitness peak, and a period of constancy of this new environment leads to a peak shift via "genetic assimilation" (Pigliucci, Murren, & Schlichting, 2006). If the resultant phenotypic variation has a fitness effect, then selection takes place;

TABLE 2 Analyses of variance tables of mixed effect models on relative growth (rSGR)

(a) Experiment 1 - wild-caught fish					(b) Experiment 2 - F1 crosses			
Model 1: whole dataset								
Effect	**num.*df***	**den.*df***	***F***	***p***	**num.*df***	**den.*df***	***F***	***p***
Sex	**2**	**131**	**12.229**	**<0.001**	**2**	**141.3**	**18.418**	**<0.001**
Population	**1**	**130**	**7.665**	**0.006**	2	142.5	1.748	0.178
Model 2: immature individuals excluded								
Effect	**num.*df***	**den.*df***	***F***	***p***	**num.*df***	**den.*df***	***F***	***p***
Sex	**1**	**106.2**	**5.958**	**0.016**	**1**	**141.7**	**36.653**	**<0.001**
Population	**1**	**106.1**	**5.739**	**0.018**	2	142.5	1.749	0.178
Model 3: males only								
Effect	**num.*df***	**den.*df***	**F**	***p***	**num.*df***	**den.*df***	***F***	***p***
Population	**1**	**64.8**	**6.509**	**0.013**	2	79.9	1.161	0.319
Model 4: females only								
Effect	**num.*df***	**den.*df***	**F**	***p***	**num.*df***	**den.*df***	***F***	***p***
Population	1	38.9	0.104	0.749	2	58.1	2.837	0.067

F-statistic was corrected with the Kenward–Roger approximation for mixed linear models. Experiment 1—wild-caught fish (a) and experiment 2—F1 crosses (b). Significant effects ($p < 0.05$) are highlighted in bold.

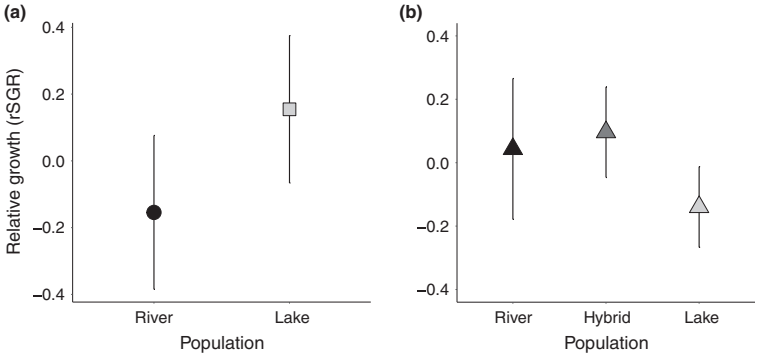


FIGURE 4 Relative growth performance (rSGR) ±CI 95% in the lake habitat for wild-caught individuals (a) and F1 crosses (b)

and if this phenotypic variation has a genetic component, selection leads to "genetic accommodation," that is, adaptive evolution that involves gene-frequency change (West-Eberhard, 2005).

A recent reciprocal transplant experiment in stickleback (Bolnick & Stutz, 2017) detected substantial plastic convergence of immigrant fish toward the gene expression profile of the resident population after translocation (Lohman, Stutz, & Bolnick, 2017). However, stream fish placed in lake cages did not reach the optimum expression in the lake.

Cichlid species that show phenotypic plasticity are often riverine or a part of very recent intralacustrine adaptive radiations (Greenwood, 1964; Meyer, 1989; Smits et al., 1996; Chapman, Galis, & Shinn, 2000) and riverine species show the highest level of adaptive plasticity among the East African cichlids investigated so far, lending support to the 'flexible stem hypothesis' (Schneider & Meyer, 2017). If temporal and/or spatial variation is higher in river than in lake habitat, plasticity would be favored over genetic divergence (Scheiner, 1993; Sultan & Spencer, 2002). East African rivers are prone to strong seasonal and interannual fluctuations in water flow rate (Dettinger & Diaz, 2000). Within the Kalambo River, seasonal fluctuations in environmental parameters associated with a seasonal influx of water during the rainy season are supposedly higher than in the lake (Figure S3); thus, *A. burtoni* likely experiences a high degree of temporal and spatial variation in this river system compared to the lake.

High levels of gene flow among populations should favor the evolution or maintenance of phenotypic plasticity over local adaptation (Sultan & Spencer, 2002). Estimated migration rate for the Kalambo River system is higher from the river to the lake (m~2.02E-04) than *vice versa* (m~6.02E-05) (Egger et al., 2017), which should favor plasticity in the riverine population. In a study

of geographic variation of phenotypic plasticity in another haplochromine cichlid, *Pseudocrenilabrus multicolor*, that inhabits both riverine and swampy areas, high levels of phenotypic plasticity for both gill size and brain mass were observed (Crispo & Chapman, 2010). F1 offspring from populations that are close to the connection between the swamp and river, and thus have the highest potential for dispersal between environments, were shown to have more plastic brains.

Plastic lineages can persist in a new habitat, even if there are no similar niches available, and are therefore expected to have higher potential for adaptive diversification than nonplastic lineages (Schneider & Meyer, 2017). In stickleback, transcriptomic plasticity may play a substantial role in migrants' adaptation to novel environments (Lohman et al., 2017). In this system genetic divergence and plasticity appear to work together in shaping between-ecotype differences in gene expression (Lohman et al., 2017) and parallel adaptive phenotypic divergence between lake and stream populations (Oke et al., 2016). Our results provide support for the same forces working together in a cichlid lake-stream system.

5 | CONCLUSION

This study provides rare empirical data on fitness estimates in a cichlid species in the wild, using both wild-caught and F1 individuals. We found a substantial contribution of plasticity to increased immigrant performance in a foreign environment. This finding highlights the value of formally comparing fitness of wild-caught and common garden-reared individuals in the study of local adaptation. Given that a single lake-stream population pair was studied, it is possible that some of the observed patterns are unique to this system. Future studies should aim to overcome logistical challenges and investigate this and other reproductive barriers in additional lake-stream population pairs including those known to exhibit stronger genomic differentiation in order to achieve a more general understanding of adaptive divergence in this system.

ACKNOWLEDGMENTS

We would like to thank Adrian Indermaur, Fabrizia Ronco, Kolja Smailus, Gaëlle Pauquet, and Daniel Lüscher for help in the field, the crew of the Kalambo Lodge for their logistic support in Zambia, and the Lake Tanganyika Research Unit, Department of Fisheries, Republic of Zambia, for research permits. We thank Dario Moser and Fabrizia Ronco for valuable suggestions on experimental design, Miya Qiaowei Pan for DNA extraction protocol, Nikolas Vellnow for advice on statistical analyses, and Robert Spirig for valuable feedback on the manuscript. This study was supported by grants from the Swiss Zoological Society (SZS) to JR and the Swiss National Science Foundation (SNSF, grant 31003A_156405) to WS.

CONFLICT OF INTERESTS

None declared

AUTHOR CONTRIBUTION

BE and WS conceived and supervised the study, all coauthors contributed to the experimental design, all authors conducted the fieldwork. JR conducted the molecular laboratory work, analyzed the data, and wrote the manuscript, with feedback from all coauthors.

DATA ACCESSIBILITY

Individual measurement data available from the Dryad Digital Repository: https://doi.org/10.5061/dryad.7ns4pk2

ORCID

Jelena Rajkov http://orcid.org/0000-0003-1533-2430
Walter Salzburger http://orcid.org/0000-0002-9988-1674
Bernd Egger http://orcid.org/0000-0002-8928-2301
Alexandra Anh-Thu Weber http://orcid.org/0000-0002-7980-388X

REFERENCES

Alcazar, R. M., Hilliard, A. T., Becker, L., Bernaba, M., & Fernald, R. D. (2014). Brains over brawn: Experience overcomes a size disadvantage in fish social hierarchies. *Journal of Experimental Biology*, *217*, 1462–1468. https://doi.org/10.1242/jeb.097527

Arnold, S. J. (1983). Morphology, performance and fitness. *American Zoologist*, *23*, 347–361. https://doi.org/10.1093/icb/23.2.347

Bates, D. M., Maechler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, *67*, 1–48.

Best, R. J., Anaya-Rojas, J. M., Leal, M. C., Schmid, D. W., Seehausen, O., & Matthews, B. (2017). Transgenerational selection driven by divergent ecological impacts of hybridizing lineages. *Nature Ecology & Evolution*, *1*, 1757–1765. https://doi.org/10.1038/s41559-017-0308-2

Bolnick, D. I., & Stutz, W. E. (2017). Frequency dependence limits divergent evolution by favouring rare immigrants over residents. *Nature*, *546*, 285–288. https://doi.org/10.1038/nature22351

Bongaerts, P., Riginos, C., Hay, K. B., van Oppen, M. J., Hoegh-Guldberg, O., & Dove, S. (2011). Adaptive divergence in a scleractinian coral: Physiological adaptation of Seriatopora hystrix to shallow and deep reef habitats. *BMC Evolutionary Biology*, *11*, 303. https://doi.org/10.1186/1471-2148-11-303

Bouton, N., Witte, F., & Van Alphen, J. J. M. (2002). Experimental evidence for adaptive phenotypic plasticity in a rock-dwelling cichlid fish from Lake Victoria. *Biological Journal of the Linnean Society*, *77*, 185–192. https://doi.org/10.1046/j.1095-8312.2002.00093.x

Casquet, J., Thebaud, C., & Gillespie, R. G. (2012). Chelex without boiling, a rapid and easy technique to obtain stable amplifiable DNA from small amounts of ethanol-stored spiders. *Molecular Ecology Resources*, *12*, 136–141. https://doi.org/10.1111/j.1755-0998.2011.03073.x

Chapman, L. J., Galis, F., & Shinn, J. (2000). Phenotypic plasticity and the possible role of genetic assymilation: Hypoxia-induced trade-offs in the morphological traits of an African cichlid. *Ecology Letters*, *3*, 387–393. https://doi.org/10.1046/j.1461-0248.2000.00160.x

Crispo, E., & Chapman, L. J. (2010). Geographic variation in phenotypic plasticity in response to dissolved oxygen in an African cichlid fish. *Journal of Evolutionary Biology*, *23*, 2091–2103. https://doi.org/10.1111/j.1420-9101.2010.02069.x

Dettinger, M. D., & Diaz, H. F. (2000). Global characteristics of stream flow seasonality and variability. *Journal of Hydrometeorology*, *1*, 289–310.

Egger, B., Roesti, M., Böhne, A., Roth, O., & Salzburger, W. (2017). Demography and genome divergence of lake and stream populations of an East African cichlid fish. *Molecular Ecology*, *23*, 5304–5322.

Fernald, R. D. (1980). Response of male cichlid fish, *Haplochromis burtoni*, reared in isolation to models of conspecifics. *Zeitschrift für Tierpsychologie*, *93*, 85–93.

Fernald, R. D., & Hirata, N. R. (1977). Field study of *Haplochromis burtoni*: Quantitative behavioral observations. *Animal Behavior*, *64*, 4.

Galis, F., & Metz, J. A. (1998). Why are there so many cichlid species?. *Trends in Ecology & Evolution*, *13*, 1–2. https://doi.org/10.1016/S0169-5347(97)01239-1

Galpern, P., Manseau, M., Hettinga, P., Smith, K., & Wilson, P. (2012). Allelematch: An R package for identifying unique multilocus genotypes where genotyping error and missing data may be present. *Molecular Ecology Resources*, *12*, 771–778. https://doi.org/10.1111/j.1755-0998.2012.03137.x

Ghalambor, C. K., McKay, J. K., Carroll, S. P., & Reznick, D. N. (2007). Adaptive versus non-adaptive phenotypic plasticity and the potential for contemporary adaptation in new environments. *Functional Ecology*, *21*, 394–407. https://doi.org/10.1111/j.1365-2435.2007.01283.x

Gosden, T. P., Waller, J. T., & Svensson, E. I. (2015). Asymmetric isolating barriers between different microclimatic environments caused by low immigrant survival. *Proceedings of the Royal Society B*, *282*, 20142459.

Greenwood, P. H. (1964). Environmental effects on the pharyngeal mill of a cichlid fish, Astatoreochromis alluandi, and their taxonomic implications. *Proceedings of the Linnean Society of London*, *176*, 1–10.

Hatfield, T., & Schluter, D. (1999). Ecological speciation in sticklebacks: Environment-dependent hybrid fitness. *Evolution (N. Y).*, *53*, 866–873.

Hendry, A. P. (2001). Adaptive divergence and the evolution of reproductive isolation in the wild: An empirical demonstration using introduced sockeye salmon. *Genetica*, *112–113*, 515–534. https://doi.org/10.1023/A:1013367100865

Hereford, J. (2009). A quantitative survey of local adaptation and fitness trade-offs. *American Naturalist*, *173*, 579–588. https://doi.org/10.1086/597611

Kaufmann, J., Lenz, T. L., Kalbe, M., Milinski, M., & Eizaguirre, C. (2017). A field reciprocal transplant experiment reveals asymmetric costs of migration between lake and river ecotypes of three-spined sticklebacks (*Gasterosteus aculeatus*). *Journal of Evolutionary Biology*, *30*, 938–950. https://doi.org/10.1111/jeb.13057

Kawecki, T. J., & Ebert, D. (2004). Conceptual issues in local adaptation. *Ecology Letters*, *7*, 1225–1241. https://doi.org/10.1111/j.1461-0248.2004.00684.x

Kenkel, C. D., & Matz, M. V. (2016). Gene expression plasticity as a mechanism of coral adaptation to a variable environment. *Nature Ecology & Evolution*, *1*, 14. https://doi.org/10.1038/s41559-016-0014

Kocher, T. D. (2004). Adaptive evolution and explosive speciation: The cichlid fish model. *Nature Reviews Genetics*, *5*, 288–298. https://doi.org/10.1038/nrg1316

Kornfield, I., & Smith, P. F. (2000). African Cichlid Fishes : Model Systems for Evolutionary Biology. *Annual Review of Ecology and Systematics*, *31*, 163–182. https://doi.org/10.1146/annurev.ecolsys.31.1.163

Kuznetsova, A., Brockhoff, P. B., & Christensen, R. H. B. (2017). lmerTest package: Tests in linear mixed effects models. *Journal of Statistical Software*, *82*, 1–26.

Leimu, R., & Fischer, M. (2008). A meta-analysis of local adaptation in plants. *PLoS ONE*, *3*, 1–8.

Lenth, R. V. (2016). Least-squares means: The *R* Package lsmeans. *Journal of Statistical Software*, *69*, 1–33.

Lexer, C., Randell, R. A., Rieseberg, L. H. (2009). Experimental hybridization as a tool for studying selection in the wild. *Ecology*, *84*, 1688–1699.

Lohman, B. K., Stutz, W. E., & Bolnick, D. I. (2017). Gene expression stasis and plasticity following migration into a foreign environment. *Molecular Ecology*, *26*, 4657–4670. https://doi.org/10.1111/mec.14234

Meyer, A. (1987). Phenotypic plasticity and heterochrony in *Cichlasoma managuense* (Pisces, Chichlidae) and their implications for speciation in cichlid fishes. *Evolution (N. Y).*, *41*, 1357.

Meyer, A. (1989). Cost of morphological specialization: feeding performance of the two morphs in the trophically polymorphic cichlid fish, *Cichlasoma citrinellum*. *Oecologia*, *80*, 431–436. https://doi.org/10.1007/BF00379047

Moser, D., Frey, A., & Berner, D. (2016). Fitness differences between parapatric lake and stream stickleback revealed by a field transplant. *Journal of Evolutionary Biology*, *29*, 711–719. https://doi.org/10.1111/jeb.12817

Muschick, M., Barluenga, M., Salzburger, W., & Meyer, A. (2011). Adaptive phenotypic plasticity in the Midas Cichlid fish pharyngeal jaw and its relevance in adaptive radiation. *BMC Evolutionary Biology*, *11*, 116. https://doi.org/10.1186/1471-2148-11-116

Nosil, P. (2012). *Ecological speciation*. Oxford: Oxford University Press. https://doi.org/10.1093/acprof:osobl/9780199587100.001.0001

Oke, K. B., Bukhari, M., Kaeuffer, R., Rolshausen, G., Räsänen, K., Bolnick, D. I., … Hendry, A. P. (2016). Does plasticity enhance or dampen phenotypic parallelism? A test with three lake-stream stickleback pairs. *Journal of Evolutionary Biology*, *29*, 126–143. https://doi.org/10.1111/jeb.12767

Pfennig, D. W., Wund, M. A., Snell-Rood, E. C., Cruickshank, T., Schlichting, C. D., & Moczek, A. P. (2010). Phenotypic plasticity's impacts on diversification and speciation. *Trends in Ecology & Evolution*, *25*, 459–467. https://doi.org/10.1016/j.tree.2010.05.006

Pigliucci, M., Murren, C. J., & Schlichting, C. D. (2006). Phenotypic plasticity and evolution by genetic assimilation. *Journal of Experimental Biology*, *209*, 2362–2367. https://doi.org/10.1242/jeb.02070

Price, T. D., Qvarnstrom, A., & Irwin, D. E. (2003). The role of phenotypic plasticity in driving genetic evolution. *Proceedings of the Royal Society B-Biological Sciences*, *270*, 1433–1440. https://doi.org/10.1098/rspb.2003.2372

R Core Team (2013). *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing. URL http://www.R-project.org/.

Räsänen, K., & Hendry, A. P. (2014). Asymmetric reproductive barriers and mosaic reproductive isolation: Insights from Misty lake-stream stickleback. *Ecology and Evolution*, *4*, 1166–1175. https://doi.org/10.1002/ece3.1012

Richards, C. L., Bossdorf, O., Muth, N. Z., Gurevitch, J., & Pigliucci, M. (2006). Jack of all trades, master of some? On the role of phenotypic plasticity in plant invasions. *Ecology Letters*, *9*, 981–993. https://doi.org/10.1111/j.1461-0248.2006.00950.x

Rundle, H. D. (2002). A test of ecologically dependent postmating isolation between sympatric sticklebacks. *Society for the Study of Evolution*, *56*, 322–329.

Scharsack, J. P., Kalbe, M., Harrod, C., & Rauch, G. (2007). Habitat-specific adaptation of immune responses of stickleback (Gasterosteus aculeatus) lake and river ecotypes. *Proceedings. Biological Sciences*, *274*, 1523–1532. https://doi.org/10.1098/rspb.2007.0210

Scheiner, S. M. (1993). Genetics and evolution of phenotypic plasticity. *Annual Review of Ecology and Systematics*, *24*, 35–68. https://doi.org/10.1146/annurev.es.24.110193.000343

Schlichting, C. D. (2004). The role of phenotypic plasticity in diversification. In T. J. DeWitt, & S. M. Scheiner (Eds.), *Phenotypic Plasticity: Fuctional and Conceptual Approaches* (pp. 191–200). Oxford, UK: Oxford University Press.

Schlichting, C. D., & Pigliucci, M. (1998). *Phenotypic Evolution: A Reaction Norm Perspective*. Sunderland, Massachusetts: Sinauer Associates Inc.

Schneider, R. F., & Meyer, A. (2017). How plasticity, genetic assimilation and cryptic genetic variation may contribute to adaptive radiations. *Molecular Ecology*, *26*, 330–350. https://doi.org/10.1111/mec.13880

Schwartz, A. K., Weese, D. J., Bentzen, P., Kinnison, M. T., & Hendry, A. P. (2010). Both geography and ecology contribute to mating isolation in guppies. *PLoS ONE*, *5*, 1–7.

Seehausen, O. (2006). African cichlid fish: A model system in adaptive radiation research. *Proceedings of the Royal Society B-Biological Sciences*, *273*, 1987–1998. https://doi.org/10.1098/rspb.2006.3539

Smits, J. D., Witte, F., & VanVeen, F. G. (1996). Functional changes in the anatomy of the pharyngeal jaw apparatus of *Astatoreochromis alluaudi* (Pisces, Cichlidae), and their effects on adjacent structures. *Biological Journal of the Linnean Society*, *59*, 389–409.

Soria-Carrasco, V., Gompert, Z., Comeault, A. A., Farkas, T. E., Parchman, T. L., Johnston, J. S., … Nosil, P. (2014). Stick insect genomes reveal natural selection's role in parallel speciation. *Science*, *344*, 738–742. https://doi.org/10.1126/science.1252136

Soudi, S., Reinhold, K., & Engqvist, L. (2016). Ecologically dependent and intrinsic genetic signatures of postzygotic isolation between sympatric host races of the leaf beetle *Lochmaea capreae*. *Evolution*, *70*, 471–479. https://doi.org/10.1111/evo.12846

Sultan, S. E., & Spencer, H. G. (2002). Metapopulation structure favors plasticity over local adaptation. *American Naturalist*, *160*, 271–283. https://doi.org/10.1086/341015

Theis, A., Ronco, F., Indermaur, A., Salzburger, W., & Egger, B. (2014). Adaptive divergence between lake and stream populations of an East African cichlid fish. *Molecular Ecology*, *23*, 5304–5322. https://doi.org/10.1111/mec.12939

Theis, A., Roth, O., Cortesi, F., Ronco, F., Salzburger, W., & Egger, B. (2017). Variation of anal fin egg-spots along an environmental gradient in a haplochromine cichlid fish. *Evolution*, *71*, 1–12.

Thibert-Plante, X., & Hendry, A. P. (2011). The consequences of phenotypic plasticity for ecological speciation. *Journal of Evolutionary Biology*, *24*, 326–342. https://doi.org/10.1111/j.1420-9101.2010.02169.x

Thorpe, R. S., Reardon, J. T., & Malhotra, A. (2005). Common garden and natural selection experiments support ecotypic differentiation in the dominican anole (*Anolis oculatus*). *American Naturalist*, *165*, 495–504. https://doi.org/10.1086/428408

Turner, G. F. (2007). Adaptive radiation of cichlid fish. *Current Biology*, *17*, 827–831. https://doi.org/10.1016/j.cub.2007.07.026

West-Eberhard, M. J. (2003). *Developmental plasticity and evolution*. Oxford: Oxford University Press.

West-Eberhard, M. J. (2005). Developmental plasticity and the origin of species differences. *Proceedings of the National Academy of Sciences of the United States of America*, *102*(Suppl), 6543–6549. https://doi.org/10.1073/pnas.0501844102

Westley, P. A. H., Ward, E. J., & Fleming, I. A. (2012). Fine-scale local adaptation in an invasive freshwater fish has evolved in contemporary time. *Proceedings of the Royal Society B-Biological Sciences*, *280*, 20122327. https://doi.org/10.1098/rspb.2012.2327

Wimberger, P. H. (1992). Plasticity of fish body shape. The effects of diet, development, family and age in two species of Geophagus (Pisces: Cichlidae). *Biological Journal of the Linnean Society*, *45*, 197–218. https://doi.org/10.1111/j.1095-8312.1992.tb00640.x

Yeh, P. J., & Price, T. D. (2004). Adaptive phenotypic plasticity and the successful colonization of a novel environment. *American Naturalist*, *164*, 531–542. https://doi.org/10.1086/423825

SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section at the end of the article.

How to cite this article: Rajkov J, Weber AAT, Salzburger W, Egger B. Adaptive phenotypic plasticity contributes to divergence between lake and river populations of an East African cichlid fish. *Ecol Evol*. 2018;8:7323–7333. https://doi.org/10.1002/ece3.4241

Supporting information

Adaptive phenotypic plasticity contributes to divergence between lake and river populations of an East African cichlid fish

Jelena Rajkov[1], Alexandra Anh-Thu Weber[1], Walter Salzburger[1], Bernd Egger[1]
[1]Zoological Institute, University of Basel, Vesalgasse 1, CH-4051 Basel, Switzerland
Email: jelena.rajkov@evobio.eu

(a)

(b)

Figure S1. Photographs of enclosures used for the two field transplant experiments in Lake Tanganyika. Experiment 1 conducted during the dry season in 2015 (a), and experiment 2 conducted during the wet season in 2016 (b). For the exact location of the enclosures see Fig. 1 in the main text.

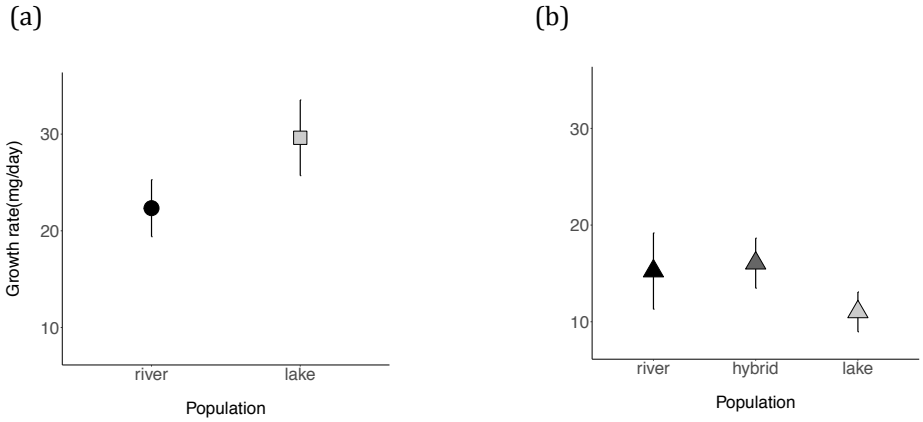


Figure S2. Growth rates (mg/day) ±CI 95% in the lake habitat for wild-caught individuals (a), and F1 crosses (b).

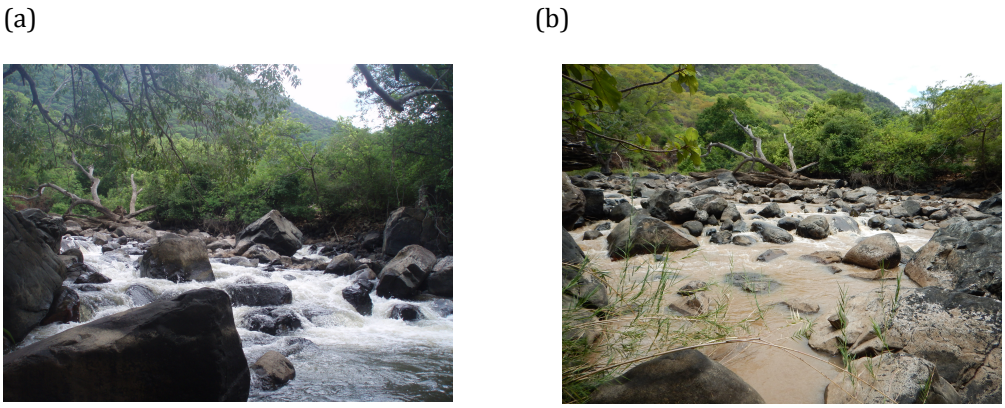


Figure S3. Photographs of Kalambo River upstream (KaR) before (a) and after (b) rain.

Table S1. Analyses of variance tables of mixed effect models on growth rate (mg/day). F-statistic was corrected with the Kenward-Roger approximation for mixed linear models. (a) Experiment 1 – wild-caught fish, (b) experiment 2 – F1 crosses. Significant effects ($P < 0.05$) are highlighted in bold.

	(a) Experiment 1 - wild-caught fish				(b) Experiment 2 - F1 crosses			
model 1: whole dataset								
Effect	num.d.f.	den.d.f.	*F*	*P*	num.d.f.	den.d.f.	*F*	*P*
sex	**2**	**132.3**	**19.170**	**<0.001**	**2**	**141.2**	**17.872**	**<0.001**
population	**1**	**130.1**	**16.330**	**<0.001**	2	142.4	2.230	0.111
model 2: unknown sex excluded								
Effect	num.d.f.	den.d.f.	*F*	*P*	num.d.f.	den.d.f.	*F*	*P*
sex	**1**	**106.5**	**11.037**	**0.001**	**1**	**141.6**	**35.492**	**<0.001**
population	**1**	**106.2**	**13.246**	**<0.001**	2	142.4	2.230	0.111
model 3: males only								
Effect	num.d.f.	den.d.f.	*F*	*P*	num.d.f.	den.d.f.	*F*	*P*
population	**1**	**65.9**	**9.974**	**0.002**	2	78.6	1.538	0.221
model 4: females only								
Effect	num.d.f.	den.d.f.	*F*	*P*	num.d.f.	den.d.f.	*F*	*P*
population	1	39.4	1.517	0.225	2	58.2	2.003	0.144

Chapter 2

Immigrant and extrinsic hybrid inviability contribute to reproductive isolation between lake and river cichlid ecotypes

J. Rajkov, A.A.T. Weber, W. Salzburger, B. Egger

Evolution (2018)

All authors were involved in fieldwork. I conducted the molecular laboratory work, the analysis and wrote the manuscript, with contributions from all authors.

BRIEF COMMUNICATION

doi:10.1111/evo.13612

Immigrant and extrinsic hybrid inviability contribute to reproductive isolation between lake and river cichlid ecotypes

Jelena Rajkov,[1,2] Alexandra Anh-Thu Weber,[1] Walter Salzburger,[1] and Bernd Egger[1]

[1]*Zoological Institute, University of Basel, Vesalgasse 1 CH-4051, Basel, Switzerland*

[2]*E-mail: jelena.rajkov@evobio.eu*

Received April 20, 2018

Accepted September 14, 2018

Understanding how reproductive barriers evolve and which barriers contribute to speciation requires the examination of organismal lineages that are still in the process of diversification and the study of the full range of reproductive barriers acting at different life stages. Lake and river ecotypes of the East African cichlid fish *Astatotilapia burtoni* show habitat-specific adaptations, despite different levels of genetic differentiation, and thus represent an ideal model to study the evolution of reproductive barriers. To evaluate the degree of reproductive isolation between genetically divergent lake and river populations, we performed a mesocosm mating experiment in a semi-natural setting at Lake Tanganyika. We assessed reproductive isolation in the presence of male–male competition by analyzing survival and growth rates of introduced adults and their reproductive success from genetic parentage of surviving offspring. The genetically divergent river population showed reduced fitness in terms of survival, growth rate, and mating success in a lake-like environment. Hybrid offspring between different populations showed intermediate survival consistent with extrinsic postzygotic reproductive barriers. Our results suggest that both prezygotic (immigrant inviability) and postzygotic reproductive barriers contribute to divergence, and highlight the value of assessing multiple reproductive barriers acting at different stages and in natural contexts to understand speciation mechanisms.

KEY WORDS: **Cichlid, hybrid inviability, immigrant inviability, local adaptation, Lake Tanganyika, reproductive isolation.**

Despite continuous interest in the topic (Darwin 1859; Sobel et al. 2010; Seehausen et al. 2014), fundamental questions in speciation research remain open. For example, although one of the main goals of speciation research is to identify the magnitude and order of appearance of isolating barriers that contribute to speciation, there is disagreement as to how this can be accomplished (Sobel et al. 2010). A promising strategy to properly address this question is to examine the full range of potential isolating barriers in incipient species, since only by studying recently diverged taxa is it possible to distinguish the isolating barriers that have actually contributed to speciation from those that have accumulated after speciation is complete (Schemske 2010; Sobel et al. 2010; Nosil 2012). However, while a large body of literature on isolating mechanisms exists, relatively few studies have explored the relative contribution of several potential mechanisms acting at different life stages to total reproductive isolation, and these studies have mostly been conducted in plants (e.g., Ramsey et al. 2003; Kay 2006; Richards and Ortiz-Barrientos 2016). Studies of isolating barriers that use replicate population pairs with different level of genetic divergence (from nascent to young species) are particularly powerful for directly testing which forms of reproductive barriers act at different time-points during speciation, and how rapidly their intensity changes with increasing genetic distance along the so-called "speciation continuum" (Schemske 2010; Nosil et al. 2017).

Reproductive isolation is generally defined as the product of all barriers to gene flow between divergent populations that are in contact (Mallet 2006). Based on their timing in the life cycle of an organism, reproductive barriers can be classified into premating-prezygotic, postmating-prezygotic, and postzygotic barriers (Coyne and Orr 2004). One important, yet long neglected, class of prezygotic reproductive barriers is immigrant

inviability, that is reduced fitness of immigrants relative to locally adapted individuals (Nosil et al. 2005; Ingley and Johnson 2016). Immigrant inviability directly leads to the reduction of interpopulation matings relative to intrapopulation matings, due to reduced survival of immigrants prior to mating in the less suitable (foreign) habitat. Moreover, immigrant females that survive and mate may still perish before they have produced offspring, further reducing genetic exchange via hybrid progeny (Nosil et al. 2005).

Extrinsic or "environment-dependent" postzygotic barriers occur when hybrids suffer reduced fitness in the dominant parental environment, for example because they fall between the parental niches (Coyne and Orr 2004; Nosil 2012). On the other hand, intrinsic postzygotic barriers occur when hybrids suffer reduced viability and/or fertility due to intrinsic genetic incompatibilities independent of the environment (e.g., Bateson-Dobzhansky-Müller incompatibilities) (Dobzhansky 1936; Coyne and Orr 2004; Seehausen et al. 2014). The presence and strength of many potential reproductive barriers, such as natural and sexual selection against migrants and hybrids, can only be experimentally evaluated if divergent populations are brought into secondary contact with each other and/or the divergent environment in natural settings, such as in enclosures in the wild or in mesocosms (Hanson et al. 2016). While some reproductive barriers, such as immigrant inviability, have been assessed in multiple systems, the reduced reproductive success of immigrants has only rarely been estimated in experiments that try to mimic natural conditions (Porter and Benkman 2017; Svensson et al. 2017).

The Haplochromini is the most species-rich and ecologically most diverse tribe of African cichlid fish, well known for multiple adaptive radiations in different lakes in Africa, making it an exceptionally rewarding model system in speciation research (Turner 2007; Maan et al. 2016). Among them, the generalist species *Astatotilapia burtoni* (Günther 1893) that inhabits Lake Tanganyika and affluent rivers (Fig. 1) is an excellent model to study the early phases of adaptive divergence. Lake and river ecotypes of *A. burtoni* show habitat specific adaptations, despite varying levels of genetic differentiation among them (Theis et al. 2014; Egger et al. 2017; Pauquet et al. 2018). Adjacent lake and river environments differ in both abiotic and biotic conditions including water chemistry, habitat structure, and prey composition (Theis et al. 2014). River fish have shallower bodies, associated with the flow regime in the river habitat, whereas lake fish have a superior mouth position, longer gill rakers, and more slender lower pharyngeal jaw bones. The shifts in trophic structures correspond to different diets: while the lake ecotype feeds predominantly on plant/algae and zooplankton, the river ecotype preys upon snails, insects, and plant seeds (Theis et al. 2014).

Previous studies found no or weak reproductive isolation between genetically close lake and river populations ($F_{ST} < 0.06$; Egger et al. 2017) (Theis et al. 2014; Rajkov et al. 2018). Here, we investigate reproductive barriers between populations that exhibit strong genomic differentiation ($F_{ST} > 0.4$; Egger et al. 2017) to achieve a more general understanding of adaptive divergence in this system. We performed a mesocosm experiment in replicate lake-like environments to assess reproductive isolation between genetically divergent lake and river *A. burtoni* populations. Males and females from three different populations (local lake – sampled next to the experimental site, foreign lake, and foreign river – sampled at the opposite coast of Lake Tanganyika, see Fig. 1A) were placed in outdoor mesocosms and their surviving offspring were genotyped to assign parentage. We estimated fitness components and potential reproductive barriers acting at different stages, including survival, growth rate, mating success, F1 offspring survival, and fecundity. Our expectation was that, if there was strong local adaptation, the local lake population would perform the best in the mesocosms with lake-like environment, followed by the foreign lake population, and the foreign river population would show the lowest performance.

Material and Methods

STUDY SYSTEM

Astatotilapia burtoni (Fig. 1B) exhibits a lek-like polygynandrous mating system, where only dominant males gain access to territories and females (Fernald and Hirata 1977). Males are known to be highly aggressive toward conspecifics (Fernald 1980), and a size difference of even less than 10% body length has been shown to provide a significant advantage to the larger opponent in territorial combats (Alcazar et al. 2014). After spawning, females protect a brood of more than 30 developing eggs in their mouth for approximately two weeks and guard the fry for several weeks after releasing them (Fernald and Hirata 1979). Females typically do not feed during the entire period of mouthbrooding (Grone et al. 2012), and mouthbrooding is generally known to cause a loss of body mass in cichlids (Balshine-Earn 1995; Smith and Wooton 1995). Multiple paternity in *A. burtoni* has been detected in mate choice experiments under laboratory conditions in ~7% of genotyped broods (Theis et al. 2012). In the present study, we test for reproductive isolation between two lake populations and one river population. We used a lake population from the estuary of the Kalambo River (KaL) on the east coast of Lake Tanganyika, a lake population from the west coast (NdL), and a genetically divergent river population from the west coast – the Lufubu River (LfR) (referred to as Lf2 in Theis et al. 2014, 2017; Egger et al. 2017; Pauquet et al. 2018) (Fig. 1A).

STUDY DESIGN

The mesocosm experiment was designed to address pre- and postzygotic extrinsic and intrinsic forms of reproductive

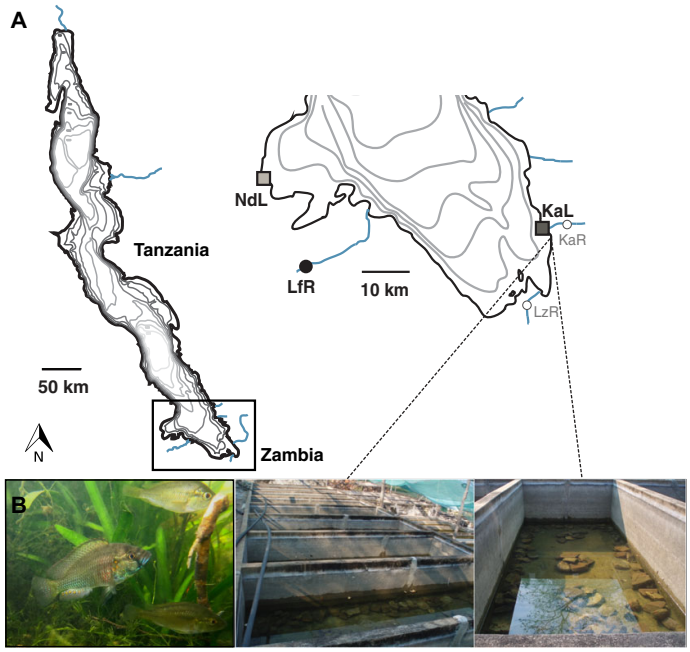


Figure 1. ***Astatotilapia burtoni*** **populations used in this study: KaL – Kalambo lake, NdL – Ndole lake, LfR – Lufubu river, and in previous studies: KaR – Kalambo river, LzR – Lunzua river (A); adult male and two females and location of the experimental mesocosms (B).**

isolation among the three populations. The experiment was carried out between November 2015 and August 2016 in six outdoor concrete ponds (mesocosms) (Fig. 1) at Kalambo Lodge, Zambia, near the location where the KaL population was sampled, under study permits nr. 003376, 004264, and 004266 issued by the Lake Tanganyika Research Unit, Department of Fisheries, Republic of Zambia.

Adult fish were caught at the source locations using hook and line fishing and kept in concrete ponds with lake water for ∼10 days before the start of the experiment. After this acclimation period, wild-caught adults were anaesthetized with clove oil, photographed, measured (±0.5 mm), weighed (±5 mg), sexed by visual inspection of external coloration and the genital papilla, fin-clipped, and tagged with visible implant elastomer tags (VIE, Northwest Marine Technology). Each individual received a population tag (KaL–front left side of the dorsal fin, NdL–front right, LfR–middle right) to enable subsequent sorting, size matching, and counting of individuals. In all cases, individuals returned to normal activity within a few minutes after tagging. Males were selected for size to achieve a similar size distribution between the three populations within each mesocosm. Each mesocosm (dimensions: 3.2 m × 1.4 m × 0.5 m; length × width × water depth) was stocked with three females and three males from each of the three populations ($n_{\text{total}} = 108$). In mesocosm 6, one KaL male was wrongly sexed and one NdL female from mesocosm 4 was accidentally relocated during the experiment. This resulted in mesocosm 6 having two KaL females, four KaL males, and four NdL females at the end of the experiment. Numerous evenly distributed rocks provided territories for males and shelter for females and offspring. Our experimental setup included male–male competition, which represents the natural situation in species that live in social groups with strong dominance hierarchy. Ponds, located in the sun, were supplied with lake water. Algae cover formed on the walls and rocks over the course of the experiment, serving as a food source and mimicking the lake environment. Fish were fed with a supplement of commercial flake food that was not fed ad libitum (∼0.3 g per pond per day) to ensure survival and a successful experiment. After a period of eight months, we emptied the mesocosms and collected all remaining adult fish (72 out of initially 108) and all surviving offspring and eggs. Fish weighing more than 1 g were photographed, measured, fin-clipped, and sexed if possible. Unfortunately, due to (i) logistic constrains imposed by the location in a remote area in Africa with no facilities that would enable construction of an experimental environment with flowing river water; and (ii) the presence of crocodiles and hippos in the riverine environment, no reciprocal control experiment in river environment could be performed. However, we can evaluate the effect of genetic differentiation between populations on reproductive isolation by using the previous experiment with a very similar setup performed in the same ponds with lake water

as a comparison (Theis et al. 2014). In this experiment the same lake population was used (KaL), together with two more closely related river populations from the Lunzua (LzR) and Kalambo (KaR) river.

PARENTAGE ANALYSES

Genomic DNA from incubated eggs and fin clips was extracted using a 5% Chelex solution (Casquet et al. 2012). The samples were genotyped at five microsatellite loci (Ppun5, Ppun7, Ppun21, UNH130, and Abur82) following the methods described in Theis et al. (2014). We genotyped all adults ($n_{\text{adult\_total}} = 108$ introduced + 73 surviving), all free-swimming juveniles and six individuals from each brood collected from mouthbrooding females. Some fry were expelled from the mouth during handling and those individuals were all genotyped (leading to some broods having more than six juveniles per brood genotyped) ($n_{\text{offspring\_total}} = 693$). Samples from all tagged individuals taken at the beginning and at the end of the experiment were matched using the R package Allelematch (Galpern et al. 2012) to identify the introduced adults. Parentage was inferred using the software CERVUS (Kalinowski et al. 2007) with no mismatch allowed. Offspring that were assigned to the same mother and father were counted as a single mating event, except if they belonged to different size classes (free-swimming young *vs.* fry). If more than one father was detected in broods collected from mouthbrooding females, these were treated as separate mating events. Mating patterns were inferred from parentage data by conservatively assuming a single mating event for each batch of full siblings of the same size class assigned to a particular parent pair.

BODY SHAPE

Photographs of the left side of each individual were taken using a Nikon D5000 camera, under standardized lighting conditions with a ruler for scale. To aid in digital landmark placement, we used three metal clips to spread the fins at the anterior insertions of the dorsal and anal fin, and at the insertion of the pectoral fin. We used geometric morphometrics to estimate body shape (Zelditch et al. 2004). In total, the photographs of 271 individuals (Table S1C) were used for geometric morphometric analyses. Using TPSDIG2 (v.2.26; Rohlf 2016) we placed 17 homologous landmarks on the image of each fish (Fig. S1). A tps file containing x and y coordinates was used as input for the program MORPHOJ (v.1.06d; Klingenberg 2011) and superimposed with a Procrustes generalized least squares fit (GLSF) algorithm to remove all nonshape variation (Rohlf and Slice 1990). Principal component analysis (PCA) and Canonical variate analysis (CVA, Mardia et al. 1979) were used to assess shape variation among different populations of introduced adults at the beginning and at the end of the experiment, and all types of offspring crosses. To correct for allometric size effects, the CVA was performed on the residuals of the regression of shape on centroid size. The statistical significance of pairwise differences in mean shape distances of CVA was obtained using permutation tests (10,000 permutations).

DATA ANALYSIS

Adults survival and growth rate

We assessed introduced adult survival among the three experimental populations using binomial generalized linear-mixed effect models (GLMMs) with survival as a binary-dependent variable and population, sex, initial standard length, size deviation (deviation in initial mass from the mean mass per mesocosm) as fixed predictors. The replicated mesocosms were set as a random effect. We calculated specific growth rates as SGR $= \frac{100}{\text{time}} \ln(\frac{m_{\text{final}}}{m_{\text{initial}}})$ for survivors. To correct for individual differences in mass at the beginning of the experiment, specific growth rates were regressed on initial mass. The residual SGR values (rSGR) were used as a measure of relative growth performance (following Barber 2005; Scharsack et al. 2007). We assessed growth rates among the three experimental populations using linear-mixed effect models (LMMs) with rSGR as a dependent variable, population, sex, and sex: population interaction as fixed predictors. The replicated mesocosms were set as a random effect. Since females often show signs of weight loss during mouthbrooding male and female growth rates were analyzed separately.

Reproductive success

We scored reproductive success among the three experimental populations via: male mating status as binary variable (mated *vs.* unmated) (i), proportion of the total number of mating events in each mesocosm per male, that is the proportion of broods completely or partly sired by a male (ii), and number of surviving offspring per mating event (iii). Male mating success was analyzed using binomial GLMMs with either mating status (i), or proportion of mating events (ii) as a dependent variable and population, size deviation (deviation in initial standard length from the mean male length per mesocosm), and male initial size as fixed predictors. The male identity and replicated mesocosms were set as random effects. The models were run using the complete dataset (including males that reproduced during the experiment but did not survive until the end of the experiment) and the survivors (all males that survived until the end of the experiment) only. We assessed offspring survival (iii) using GLMM with the number of surviving free-swimming offspring as a dependent variable and the type of cross (female population × male population), male initial size, and female initial size as fixed predictors. We used Poisson probability distribution for the count variable number of surviving offspring. Female identity, male identity, and replicated mesocosms were set as random effects.

GLMMs and LMMs were calculated using the R package lme4 (Bates et al. 2015). Significance level for the fixed effects

was determined by type II χ^2-based likelihood-ratio tests using the drop1 function of the lme4 package for GLMMs and with type II ANOVAs with Kenward–Roger correction for *F*-statistics and degrees of freedom (d.f.) using lmerTest package (Kuznetsova et al. 2017) for LMMs. We checked all GLMMs for overdispersion and included observation level as a random effect to account for the extravariance in the data, with one random effect level for each observation (male and female identity) (Harrison 2014). Tukey–Kramer post hoc tests were applied to test for significance of pairwise comparisons between populations using the lsmeans package (Lenth 2016). All statistical analyses were performed in R version 3.3.2 (R Core Team 2016). For detailed information on sample sizes see Table S1.

Results

ADULT SURVIVAL

Adult survival depended on population of origin and sex (population $\chi^2_{\text{d.f.}=2} = 31.745$, $P < 0.001$; sex $\chi^2_{\text{d.f.}=1} = 4.123$, $P = 0.042$, Fig. 2A, Table S2) and was higher in lake than in river fish (post hoc test: KaL – LfR, $P < 0.001$, NdL – LfR, $P < 0.001$, KaL – NdL, $P = 0.642$). When males and females were analyzed separately, population had an effect on male (population $\chi^2_{\text{d.f.}=2} = 24.075$, $P < 0.001$) and on female survival (population $\chi^2_{\text{d.f.}=2} = 11.396$, $P = 0.003$). Survival was higher in lake males than in river males (post hoc test: KaL – LfR, $P = 0.005$, NdL – LfR, $P = 0.006$), and in local lake females than in river females (post hoc test: KaL – LfR, $P = 0.010$). There was no difference in survival between lake males (post hoc test: KaL – NdL, $P = 0.996$), lake females (KaL – NdL, $P = 0.500$), and no significant difference between foreign lake and river females (NdL – LfR, $P = 0.096$).

ADULT GROWTH

When the whole dataset was analyzed, relative growth rate was affected by sex and population:sex interaction (sex $F_{1,63.1} = 540.270$, $P < 0.001$, population:sex $F_{1,63} = 18.040$, $P < 0.001$, Fig. 2B, Table S3). The interaction between population of origin and sex resulted from the inverse population growth patterns in males and females due to some lake females losing weight while mouthbrooding (Fig. S2), as they incubated more broods than river females (see below). When the sexes were analyzed separately, population of origin had an effect on growth in females (population $F_{2,24.8} = 25.429$, $P < 0.001$) and in males (population $F_{2,34.7} = 4.789$, $P = 0.015$). Local lake males grew faster than river males (post hoc test: KaL – LfR, $P = 0.009$), whereas river females grew faster than lake females (post hoc test: KaL – LfR, $P < 0.001$, NdL – LfR, $P < 0.001$) and foreign lake females grew faster than local lake females (post hoc test: KaL – NdL, $P < 0.044$). There was no significant difference in growth between foreign lake and river males (post hoc test: NdL – LfR, $P = 0.073$).

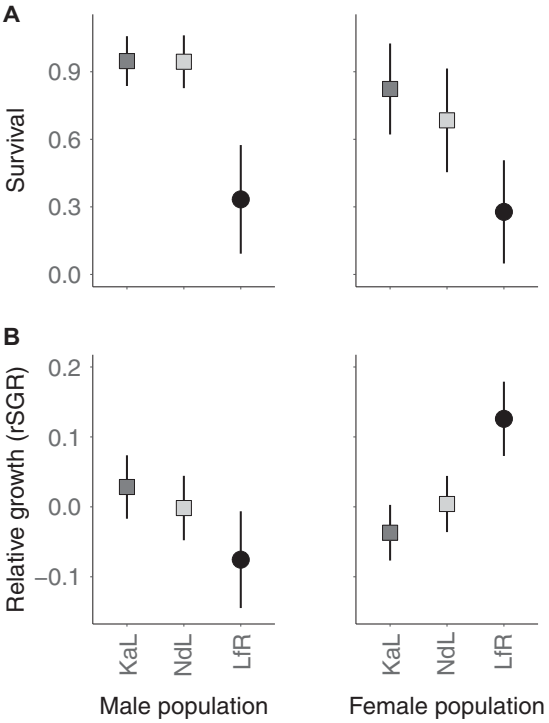


Figure 2. Survival (expressed as the average proportion of surviving individuals ± CI 95%) (A) and relative growth performance (least square means of the full models ± CI 95%) (B) of adults introduced in the mesocosms per population for males and females. Rectangles represent lake populations (KaL – Kalambo lake, NdL – Ndole lake) and circles river population (LfR – Lufubu river).

MATING EVENTS

Of the 693 genotyped offspring 80% (552) could be identified as F1 offspring of the introduced adults and were used to analyze mating patterns. The rest of the genotyped offspring were F2 or backcrosses with introduced adults (125) – collected as eggs from the mouth of F1 females, or unassigned (16). Due to the sharing of alleles between sibling F1 males, it was often not possible to unambiguously assign paternity for F2 offspring to a specific father with the five microsatellite loci used for parentage analyses. Therefore, in the remainder we discuss mating patterns inferred using the F1 offspring only.

Most females had mated with more than one male. When considering free-swimming F1 offspring only, 27 of 42 broods had multiple (up to four) fathers (64%; mean number of fathers per brood 2.0 ± 0.1 across all broods). When only genotyped (F1) eggs were considered, 16 of 25 broods had multiple (up to three)

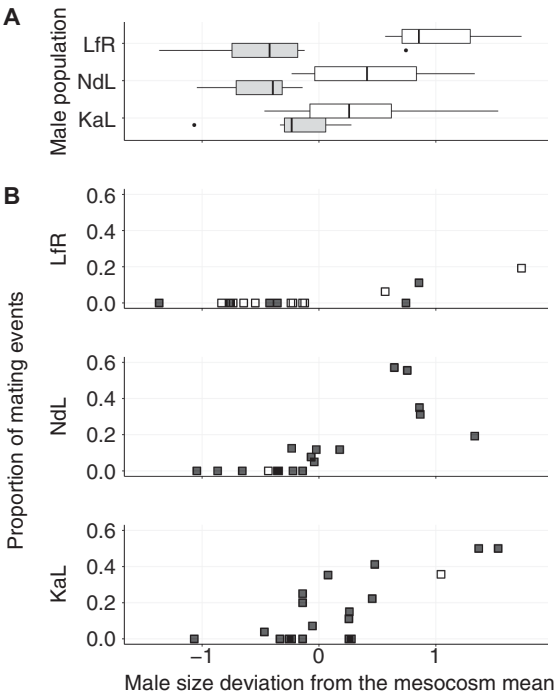


Figure 3. **Male mating success depending on population and deviation in size (initial standard length) from the mesocosm mean. Mating status (white – mated, gray – unmated) (A) and number of mating events (percentage of mesocosm total) (filled signs – survivors, hollow signs – individuals that died) (B).**

fathers (64%; 1.9 ± 0.8). As expected from a previous mesocosm experiment (Theis et al. 2014) there was higher variation in mating success of males than of females, with only 3–5 males mating per pond, and 6–8 females. All surviving females mated, except one river individual (Table S1). KaL females raised the highest number of broods (1.69 ± 0.13), followed by NdL (1.62 ± 0.14) and LfR females (1 ± 0.31). The highest number of mating events was detected between KaL males and KaL females (29), followed by other types of lake-lake crosses (19–21) and lake-river crosses (2–6) (Fig. S3A). We did not detect any mating events between a river male and a river female.

MALE MATING SUCCESS

Male mating success depended on male population and size deviation from the mean male size per mesocosm, both in terms of mating status (population $\chi^2_{\text{d.f.}=2} = 8.968$, $P = 0.011$; size deviation $\chi^2_{\text{d.f.}=1} = 12.477$, $P < 0.001$, Table S4, Fig. 3A) and with regard to the proportion of the total number of mating events (population $\chi^2_{\text{d.f.}=2} = 18.578$, $P < 0.001$; size deviation $\chi^2_{\text{d.f.}=1} = 19.837$, $P < 0.001$, Table S5, Fig. 3B). When survivors were analyzed separately, only size deviation had a significant effect on male mating success and the population effect was marginally significant (model2, Tables S4 and S5). For males that had mated ($n = 25$), population had the largest effect on the proportion of mating events (population $\chi^2_{\text{d.f.}=2} = 15.270$, $P < 0.001$; size deviation $\chi^2_{\text{d.f.}=1} = 11.761$, $P < 0.001$, Table S5).

OFFSPRING SURVIVAL

The number of surviving offspring per parent pair was calculated considering the free-swimming juveniles only. There was no asymmetry in the number of surviving hybrid offspring depending on which population was the mother or the father (model 1, post hoc test, all $P \geq 0.999$, Table S5, Fig. S4). Therefore, reciprocal crosses were subsequently pooled and the number of surviving offspring was analyzed with five levels of cross type instead of eight as an explanatory variable (model 2, Table S7). In both models the number of surviving offspring was explained by cross type (model 1: cross type $\chi^2_{\text{d.f.}=7} = 27.919$, $P < 0.001$; model 2: cross type $\chi^2_{\text{d.f.}=4} = 27.489$, $P < 0.001$, Table S7). The highest number of offspring per parent pair survived when both parents were from KaL (8.3 ± 1.6, max 29) followed by crosses where one parent was KaL and the other NdL (6.3 ± 1.1) and when both parents were from NdL (3.9 ± 1.1) (Fig. 4A). The lowest number of offspring survived when one of the parents was a river fish (LfR.NdL 2.2 ± 0.6, LfR.KaL 1.9 ± 0.4). There was a significant difference in number of surviving offspring between local lake cross and both types of lake-river crosses (post hoc test: KaL.KaL – KaL.LfR, $P < 0.001$; KaL.KaL – NdL.LfR, $P = 0.003$), local lake cross and foreign lake cross (KaL.KaL – NdL.NdL, $P = 0.023$) and lake-lake hybrids and both types of lake-river hybrids (KaL.NdL – KaL.LfR, $P = 0.001$; KaL.NdL – NdL.LfR, $P = 0.016$; Table S6).

BODY SHAPE

The CVA of body shape revealed a significant differentiation between groups (different populations of introduced adults at the beginning and at the end of the experiment, and all types of offspring crosses) and an overlap between both populations of lake adults at the end of the experiment, between both types of lake-river crosses, and between lake-lake hybrids and each of the pure parental lake crosses (Fig. 4B, Table S7). The first two CV axes explained 49% and 35% of the variance in the data, whereas the next eight axes together explained only 16% of the total variance. CV1, which described shape changes in terms of body height (Fig. S5), and CV2, which described shape changes in terms of mouth position (Fig. S5), separated river and lake populations. Whereas mean adult body shape at the end of the experiment overlapped for both lake populations, the river population displayed a distinct shape that did not converge to the lake body shape during the eight-month experiment. PCA showed

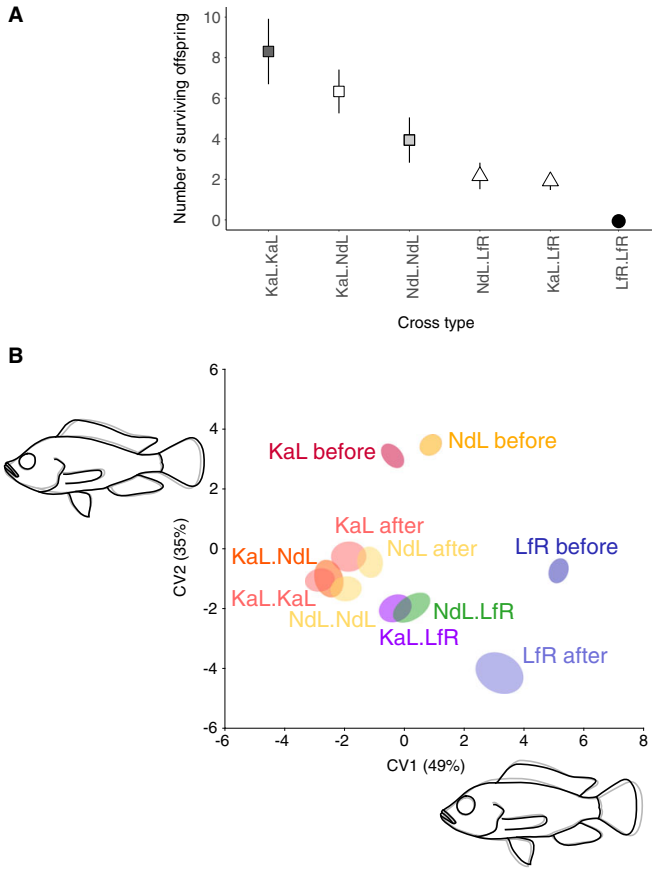


Figure 4. Mean numbers of surviving offspring per brood ± SE depending on the type of cross (A): dark gray – KaL.KaL, light gray – NdL.NdL, black - LfR.LfR, white – hybrids; rectangles represent pure lake crosses, triangles lake-river hybrids, and circle river-river cross. CVA of body shape (B), adult body shape changes during the experiment: red – Kalamabo lake (KaL), yellow – Ndole lake (NdL), blue – Lufubu river (LfR); offspring body shape: red – KaL.KaL, yellow – NdL.NdL, orange – KaL.NdL, purple — LfR.KaL, green — LfR.NdL. CV shape change outlines are for illustration purposes only, from gray to black outlines with increasing values, scaling factor 10 by default.

similar results (Fig. S6), where PC1, that described shape changes in terms of mouth and eye size, separated introduced adults from the offspring; PC2, that described shape changes in terms of head size, separated lake adults from river adults, and the offspring; and PC3, that described shape changes in terms of body height and mouth position, separated lake, and river adults.

Discussion

The goal of this study was to test for reproductive isolation between genetically divergent lake and river populations of *A. burtoni* under seminatural conditions and to evaluate the relative contributions of different reproductive barriers. Estimation of pre- and postzygotic extrinsic barriers revealed substantial reproductive isolation between the divergent Lufubu River population and the two lake populations. River fish suffered from immigrant inviability (in terms of lower adult survival and growth rate) and from lower mating success. Moreover, the complete lack of offspring from river parents and the very small number of surviving lake-river hybrid offspring further support immigrant inviability and extrinsic inviability of lake-river hybrids in the lake environment.

ENVIRONMENT DEPENDENT REPRODUCTIVE ISOLATION (IMMIGRANT INVIABILITY)

The experimental setup used in this study enabled us to simulate lake conditions and to follow the fitness of introduced fish at different stages. The mesocosms were supplied with lake water of the same chemical composition and provided similar food availability as what the local lake (KaL) population is normally exposed to. River and lake habitats substantially differ with

respect to water parameters (e.g., conductivity, pH) (Table S9) and in aquatic environments parasite risk often varies between lake and river environments (Scharsack et al. 2007), as well as between different locations within a lake (Raeymaekers et al. 2013). Moreover, the water parameters also differ between different locations in the lake, and in particular lake water close to the estuary of the Lufubu River has very different chemical properties–closer to the riverine than to the lacustrine environment. The algae cover on the rocks in the mesocosms made the mesocosms an even more suitable environment for the lake populations, which feed mostly on algae and plant material in the wild, unlike the river population that also feeds on macro-invertebrates (Theis et al. 2014). Taken together, all these factors likely contributed to the relatively higher fitness of the local lake population, followed by the foreign lake population and the river population, matching the predictions of the local adaptation hypothesis.

Similar to the results of a previous transplant experiment that compared fitness of wild-caught lake and river fish from the Kalambo river system in cages in the lake (Rajkov et al. 2018), we found lower fitness of river individuals in a lake-like environment. Moreover, the highest number of surviving F1 offspring of local lake parents, followed by the offspring of foreign lake parents, and a lack of offspring of foreign river parents, supports the role of immigrant inviability in reproductive isolation.

Adult growth rates showed opposite patterns in males and females. Local lake males grew the fastest, followed by the foreign lake males and the river males. The reverse pattern in females was likely due to longer total mouthbrooding duration in lake females that raised more broods than river females, leading to higher weight loss in lake females.

MATING PATTERNS

The results of the present study suggest that in general smaller males were outcompeted by larger dominant males (Fig. 3, Tables S4 and S5). In *A. burtoni*, size and dominance are positively correlated (Fernö 1987) and dominant males are much more likely to reproduce. However, some local KaL males succeeded to mate even when they were among the smallest in their mesocosm, and river males only succeeded to mate if they were very large (Fig. 3A), and even then participated only in a small proportion of mating events (Fig. 3B).

The high frequency of multiple paternity observed resembles estimates for other haplochromines from lakes Malawi and Victoria (Kellogg et al. 1995; Parker and Kornfield 1996; Maan et al. 2004; Tyers and Turner 2013) and is very similar to estimates from a similar experimental setup using the Lake Victoria haplochromine species *Pundmilia nyererei* (Maan et al. 2004) (64%; mean number of fathers per brood 2.0 ± 0.1 across all broods in this study *vs.* 68%; 1.8 ± 0.1).

EXTRINSIC POSTZYGOTIC ISOLATION

The only river female that was mouthbrooding when the experiment was terminated incubated 33 larvae with absorbed egg yolk (of which at least six were sired by a NdL male). This suggests that the fecundity of river females—in terms of the number of produced eggs—was probably not reduced. This goes against the argument of lower fecundity of river individuals in a foreign environment but in favor of lower survival of their offspring (extrinsic hybrid inviability). Furthermore, both types of hybrids (KaL.NdL and river.lake) show an intermediate performance in terms of survival, between the pure offspring of their parental types, supporting extrinsic postzygotic isolation (Hatfield and Schluter 1999). Likewise, a recent study in stickleback fish found extremely strong genetic effects on the relative survival and condition of the juveniles in a mesocosm experiment (Best et al. 2017).

BODY SHAPE

Data from this and the previous mesocosm experiment (Theis et al. 2014) demonstrated a clear separation between river and lake ecotypes along the CV that describes shape changes in terms of body height and mouth position (CV3, Fig. S7). However, unlike genetically close river populations that showed a high degree of plasticity (Rajkov et al. 2018) and converged towards the lake body shape in the mesocosms with standing lake water after six months (Theis et al. 2014), the body shape of adults from the genetically divergent LfR population did not change to the lake body shape after eight months under mesocosm conditions. This suggests less plasticity in body shape of LfR individuals compared to previously tested genetically closer river populations (Theis et al. 2014). Furthermore, the body shape of the LfR individuals could be one of the factors contributing to their low survival in a setup with standing lake water and algae as main food source. Other traits that were not investigated in this study, such as gut length, tooth morphology, and immune system in response to parasites could potentially also contribute to the observed differences in performance. Taken together, our data demonstrate that the genetically most divergent and putatively ancestral river population (Pauquet et al. 2018) has low fitness in the lake environment and is less plastic in comparison to other river populations that are more recently derived from adjacent lake populations (Egger et al. 2017).

NO INTRINSIC POSTZYGOTIC BARRIERS DETECTED

We detected F2 offspring from one male of KaL.LfR cross, suggesting that there is not complete hybrid sterility in lake-river hybrids, even between the most divergent populations. Nevertheless, with the data at hand, we cannot completely exclude that some intrinsic incompatibilities could also be reducing hybrid fitness. The apparent lack of intrinsic postzygotic isolation observed here is not surprising, as it is known to accumulate with time, and

very young species display little or no intrinsic postzygotic isolation (Bolnick and Near 2005; Mallet 2006; Schemske 2010). In cichlids, premating isolation accumulates fast initially but then changes little with increasing genetic distance between species (Stelkens et al. 2010). In contrast, intrinsic postzygotic isolation between closely related species is negligible but then accumulates relatively fast, resulting in complete hybrid inviability after 4.4–18.4 million years (Stelkens et al. 2010).

STRENGTH OF REPRODUCTIVE ISOLATION CORRESPONDS TO GENETIC DIVERGENCE

As data on reproductive isolation for *A. burtoni* lake-river pairs begin to accumulate (Theis et al. 2014; Rajkov et al. 2018; this study) it becomes possible to compare the importance of different barriers and the levels of reproductive isolation between lake-river population pairs with different levels of genetic divergence. These experiments used population pairs from different lake-river systems for which estimated divergence times vary from ∼13,000 (Kalambo lake-river pair) to ∼180,000 years (Lufubu lake-river pair) (Egger et al. 2017). While it is admittedly difficult to compare the results of different experiments, it is still possible to draw some general conclusions, as some of the experiments were performed in a similar setup. In a previous mesocosm experiment that used the same mesocosms and tested genetically closer populations from Kalambo and Lunzua rivers, no difference in survival of introduced adults or F1 offspring was detected, and all possible mating combinations occurred (Theis et al. 2014). In addition, the differences in survival and growth rate of introduced adults between the two lake populations and the foreign river population observed in the present study were much higher than the differences observed using wild-caught juvenile fish from two genetically closer populations from the Kalambo River system in a transplant experiment (Rajkov et al. 2018). Furthermore, lake-river F1 hybrids from the Kalambo River system showed equally high fitness as the pure crosses in the lake environment (Rajkov et al. 2018). Taken together, these results suggest that across all investigated population pairs both prezygotic and postzygotic isolation increase with genetic distance between lake and river populations.

RELATIVE IMPORTANCE OF DIFFERENT BARRIERS

Reproductive barriers that first come into play early in life history are of particular importance as subsequent barriers can only prevent gene flow that remains after the effects of earlier-acting barriers (Sobel et al. 2010). As immigrant inviability (prezygotic) acts earlier in the ontogeny than hybrid inviability (postzygotic), it likely has a greater relative contribution to limiting gene flow, ultimately leading to speciation. In speciation driven by divergent selection, extrinsic postzygotic, and prezygotic barriers evolve first and frequently interact to mediate reproductive isolation, and intrinsic postzygotic barriers usually evolve later in the speciation process (Seehausen et al. 2014). In cichlids, prezygotic behavioural reproductive barriers, in particular female preference for male coloration, are known to be an important component of reproductive isolation (Kocher 2004; Kraaijeveld and Pomiankowski 2004; Maan and Sefc 2013; Selz et al. 2014).

A. burtoni lake-river populations represent the first cichlid model that allows direct comparison to what is known from probably the best-studied speciation continuum in fish—the stickleback lake-river system (McKinnon and Rundle 2002; Berner et al. 2009; Kaeuffer et al. 2011; Lucek et al. 2013; Stuart et al. 2017). In stickleback, premating isolation evolves before postmating isolation, and extrinsic isolation is far stronger than intrinsic isolation (Hatfield and Schluter 1999; Vamosi and Schluter 2009; Lackey and Boughman 2016). Here, we detected extrinsic barriers in a cichlid species that act before and after zygote formation and include selection against immigrants and their offspring. Importantly, 111 detected mating events did not result in a single surviving pure river F1 individual. All analyzed fitness parameters suggest that local adaptation to the lake environment contributed to the higher performance of the lake ecotype.

LIMITATIONS

Unlike some laboratory experimental setups such as the "partial partition method" (Turner et al. 2001) that allow for female choice to be investigated independent of male competition, our experimental setup did not allow us to distinguish between female choice and male dominance. However, our setup is closer to the situation in nature, where female *A. burtoni* are surrounded by dominant males, and thus we are confident that this kind of setup provides us with more accurate information with respect to the barriers that are actually important for reproductive isolation in nature. Our experimental design also does not enable detection of mating events with no surviving offspring and thus did not allow us to differentiate whether the lack of offspring from specific individuals resulted from a failure of fish to spawn or through embryo and juvenile mortality. Furthermore, high mortality of the river population decreased the power of our analysis of their mating success. However, studying all the reproductive barriers acting sequentially in a semi-natural setup enabled us to accurately estimate the cumulative effect of all the investigated barriers–on the basis of the total number of surviving free swimming F1 juveniles from each cross type (Fig. S2B).

CONCLUSION

We tested river and lake ecotypes of *A. burtoni* for reproductive barriers that reduce gene flow between them. This is one of the very few studies investigating environment dependent (extrinsic) components of reproductive isolation in cichlids. We found strong prezygotic (selection against immigrants) and postzygotic

barriers (selection against hybrids) in line with local adaptation. As prezygotic barriers are not complete, postzygotic barriers cause a substantial reduction in gene flow between ecotypes. Our results highlight the value of assessing multiple reproductive barriers acting at different stages in natural contexts as well as the importance of postzygotic barriers in addition to prezygotic barriers even during the early stages of speciation. Future studies should try to disentangle the role of female mate choice *vs*. male competition in this system, as well as early inviability *vs.* missed mating opportunities using controlled laboratory setups.

AUTHOR CONTRIBUTIONS

J.R., W.S. and B.E. conceived the study; all coauthors contributed to the experimental design; J.R., A.A.T.W., and B.E. conducted the fieldwork; J.R. conducted the molecular laboratory work, analyzed the data, and wrote the article, with feedback from all coauthors.

ACKNOWLEDGEMENTS

We would like to thank Adrian Indermaur and Fabrizia Ronco for help in the field, the crews of the Kalambo Lodge and the Ndole Bay Lodge for their logistic support in Zambia, and the Lake Tanganyika Research Unit, Department of Fisheries, Republic of Zambia, for research permits. We thank Miya Qiaowei Pan for the DNA extraction protocol and Robert Spirig for valuable feedback on the manuscript. We would also like to thank the associate editor Jeff McKinnon and two anonymous reviewers for valuable comments on a previous version of the manuscript. This study was supported by grants from the Swiss Zoological Society (SZS) to J.R. and the Swiss National Science Foundation (SNSF, grant 31003A_156405) to W.S.

DATA ARCHIVING

The doi for our data is https://doi.org/10.5061/dryad.5fn863j.

CONFLICT OF INTEREST

The authors declare no conflict of interest.

LITERATURE CITED

Alcazar, R. M., A. T. Hilliard, L. Becker, M. Bernaba, and R. D. Fernald. 2014. Brains over brawn: experience overcomes a size disadvantage in fish social hierarchies. J. Exp. Biol. 217:1462–1468.

Balshine-Earn, S. 1995. The costs of parental care in Galilee St Peter's fish, *Sarotherodon galilaeus*. Anim. Behav. 50:1–7.

Barber, I. 2005. Parasites grow larger in faster growing fish hosts. Int. J. Parasitol. 35:137–143.

Bates, D. M., M. Maechler, B. Bolker, and S. Walker. 2015. Fitting linear mixed-effects models using lme4. J. Stat. Softw. 67:1–48.

Berner, D., A. C. Grandchamp, and A. P. Hendry. 2009. Variable progress toward ecological speciation in parapatry: stickleback across eight lake-stream transitions. Evolution 63:1740–1753.

Best, R. J., J. M. Anaya-Rojas, M. C. Leal, D. W. Schmid, O. Seehausen, and B. Matthews. 2017. Transgenerational selection driven by divergent ecological impacts of hybridizing lineages. Nat. Ecol. Evol. 1:1757–1765.

Bolnick, D. I., and T. J. Near. 2005. Tempo of hybrid inviability in centrarchid fishes (Teleostei: Centrarchidae). Evolution 59:1754–1767.

Casquet, J., C. Thebaud, and R. G. Gillespie. 2012. Chelex without boiling, a rapid and easy technique to obtain stable amplifiable DNA from small amounts of ethanol-stored spiders. Mol. Ecol. Resour. 12:136–141.

Coyne, J. A., and A. H. Orr. 2004. Speciation. Sinauer, Sunderland, MA.

Darwin, C. 1859. On the origin of species by means of natural selection, or the preservation of favoured races in the struggle for life. John Murray, London.

Dobzhansky, T. 1936. Studies on hybrid sterility. II. Localization of sterility factors in *Drosophila pseudoobscura* hybrids. Genetics 21:113–135.

Egger, B., M. Roesti, A. Böhne, O. Roth, and W. Salzburger. 2017. Demography and genome divergence of lake and stream populations of an East African cichlid fish. Mol. Ecol. 26:5016–5030.

Fernald, R. D. 1980. Response of male cichlid fish, *Haplochromis burtoni*, reared in isolation to models of conspecifics. Z. Tierpsychol. 93:85–93.

———. and N. R. Hirata. 1977. Field study of *Haplochromis burtoni*: quantitative behavioral observations. Anim. Behav. 64:4.

———. 1979. The ontogeny of social behavior and body coloration int he African cichlid fish *Haplochromis burtoni*. Zeitschrift fuer Tierpsychologie 50:180–187.

Fernö, A. 1987. Aggressive behaviour between territorial cichlids (*Astatotilapia burtoni*) in relation to rank and territorial stability. Behaviour 103:241–258.

Galpern, P., M. Manseau, P. Hettinga, K. Smith, and P. Wilson. 2012. Allelematch: an R package for identifying unique multilocus genotypes where genotyping error and missing data may be present. Mol. Ecol. Resour. 12:771–778.

Grone, B. P., R. E. Carpenter, M. Lee, K. P. Maruska, and R. D. Fernald. 2012. Food deprivation explains effects of mouthbrooding on ovaries and steroid hormones, but not brain neuropeptide and receptor mRNAs, in an African cichlid fish. Horm. Behav. 62:18–26.

Günther, A. (1893). Second report on the reptiles, batrachians, and fishes transmitted by Mr HH Johnston, CB, from British Central Africa. Proceedings of the Zoological Society of London, 1893:616–628.

Hanson, D., J. S. Moore, E. B. Taylor, R. D. H. Barrett, and A. P. Hendry. 2016. Assessing reproductive isolation using a contact zone between parapatric lake-stream stickleback ecotypes. J. Evol. Biol. 29: 2491–2501.

Harrison, X. A. 2014. Using observation-level random effects to model overdispersion in count data in ecology and evolution. PeerJ 2:e616.

Hatfield, T., and D. Schluter. 1999. Ecological speciation in sticklebacks: environment-dependent hybrid fitness. Evolution 53:866–873.

Ingley, S. J., and J. B. Johnson. 2016. Divergent natural selection promotes immigrant inviability at early and late stages of evolutionary divergence. Evolution 70:600–616.

Kaeuffer, R., C. L. Peichel, D. I. Bolnick, and A. P. Hendry. 2011. Parallel and non-parallel aspects of ecological, phenotypic, and genetic divergence across replicate population pairs of lake and stream stickleback. Evolution 66:402–418.

Kalinowski, S. T., M. L. Taper, and T. C. Marshall. 2007. Revising how the computer program CERVUS accommodates genotyping error increases success in paternity assignment. Mol. Ecol. 16:1099–1106.

Kay, K. M. 2006. Reproductive isolation between two closely related hummingbird-pollinated neotropical gingers. Evolution 60:538–552.

Kellogg, K. A., J. A. Markert, J. R. Stauffer, and T. D. Kocher. 1995. Microsatellite variation demonstrates multiple paternity in lekking cichlid fishes from Lake Malawi, Africa. Proc. R Soc. B Biol. Sci. 260:79–84.

Klingenberg, C. P. 2011. MorphoJ: an integrated software package for geometric morphometrics. Mol. Ecol. Resour. 11:353–357.

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

Kraaijeveld, K., and A. Pomiankowski. 2004. Evolution: love thy neighbour. Curr. Biol. 14:419–421.

Kuznetsova, A., P. B. Brockhoff, and R. H. B. Christensen. 2017. lmerTest package: tests in linear mixed effects models. J. Stat. Softw. 82:1–26.

Lackey, A. C. R., and J. W. Boughman. 2016. Evolution of reproductive isolation in stickleback fish. Evolution 102:31–38.

Lenth, R. V. 2016. Least-squares means: the R package lsmeans. J. Stat. Softw. 69:1–33.

Lucek, K., A. Sivasundar, D. Roy, and O. Seehausen. 2013. Repeated and predictable patterns of ecotypic differentiation during a biological invasion: lake-stream divergence in parapatric Swiss stickleback. J. Evol. Biol. 26:2691–2709.

Maan, M. E., O. Seehausen, and T. G. G. Groothuis. 2016. Differential survival between visual environments supports a role of divergent sensory drive in cichlid fish speciation. Am. Nat. 189:78–85.

———. O. Seehausen, L. Soderberg, L. Johnson, E. A. P. Ripmeester, H. D. J. Mrosso, M. I. Taylor, T. J. M. van Dooren, and J. J. M. van Alphen. 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.

———. and K. M. Sefc. 2013. Colour variation in cichlid fish: developmental mechanisms, selective pressures and evolutionary consequences. Semin. Cell Dev. Biol. 24:516–528.

Mallet, J. 2006. What does Drosophila genetics tell us about speciation? Trends Ecol. Evol. 21:386–393.

Mardia, K. V., J. T. Kent, and J. M. Bibby. 1995. Multivariate analysis. Academic Press, New York, New York.

Mardia, K.V., J.T. Kent, J.M. Bibby. 1979. Multivariate Analysis. Academic Press, New York.

McKinnon, J. S., and H. D. Rundle. 2002. Speciation in nature: the threespine stickleback model systems. Trends Ecol. Evol. 17:480–488.

Nosil, P. 2012. Ecological speciation. Oxford Univ. Press, Oxford.

———. J. L. Feder, S. M. Flaxman, and Z. Gompert. 2017. Tipping points in the dynamics of speciation. Nat. Ecol. Evol. 1:1–8.

———. T. H. Vines, and D. J. Funk. 2005. Reproductive isolation caused by natural selection against immigrants from divergent habitats. Evolution 59:705–719.

Parker, A., and I. Kornfield. 1996. Polygynandry in *Pseudotropheus zebra*, a cichlid fish from Lake Malawi. Environ. Biol. Fishes 47:345–352.

Pauquet, G., W. Salzburger, and B. Egger. 2018. The puzzling phylogeography of the haplochromine cichlid fish *Astatotilapia burtoni*. Ecol. Evol. 8:5637–5648.

Porter, C. K., and C. W. Benkman. 2017. Assessing the potential contributions of reduced immigrant viability and fecundity to reproductive isolation. Am. Nat. 189:580–591.

Raeymaekers, J. A., P. I. Hablützel, A. F. Grégoir, J. Bamps, A. K. Roose, M. P. Vanhove, M. Van Steenberge, A. Pariselle, T. Huyse, J. Snoeks, et al. 2013. Contrasting parasite communities among allopatric colour morphs of the Lake Tanganyika cichlid Tropheus. BMC Evol. Biol. 13:41.

Rajkov, J., A. A.-T. Weber, W. Salzburger, and B. Egger. 2018. Adaptive phenotypic plasticity contributes to divergence between lake and river populations of an East African cichlid fish. Ecol. Evol. 8:7323–7333.

Ramsey, J., H. D. Bradshaw, and D. W. Schemske. 2003. Components of reproductive isolation between the monkeyflowers *Mimulus lewisii* and *M. cardinalis* (Phrymaceae). Evolution 57:1520–1534.

R Core Team. 2016. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. https://www.R-project.org/.

Richards, T. J., and D. Ortiz-Barrientos. 2016. Immigrant inviability produces a strong barrier to gene flow between parapatric ecotypes of *Senecio lautus*. Evolution 70:1239–1248.

Rohlf, F. J., and D. Slice. 1990. Extensions of the Procrustes method for the optimal superimposition of landmarks. Syst. Zool. 39:40–59.

Rohlf, F. J. 2016. TPSDIG, Version 2.26. Department of Ecology and Evolution, State University of New York at Stony Brook. Available from http://life.bio.sunysb.edu/morph/.

Scharsack, J. P., M. Kalbe, C. Harrod, and G. Rauch. 2007. Habitat-specific adaptation of immune responses of stickleback (*Gasterosteus aculeatus*) lake and river ecotypes. Proc. R Soc. B Biol. Sci. 274:1523–1532.

Schemske, D. W. 2010. Adaptation and *The Origin of Species*. Am. Nat. 176:S4–S25.

Seehausen, O., R. K. Butlin, I. Keller, C. E. Wagner, J. W. Boughman, P. A. Hohenlohe, C. L. Peichel, G.-P. Saetre, C. Bank, A. Brännström, et al. 2014. Genomics and the origin of species. Nat. Rev. Genet. 15: 176–92.

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

Smith, C., and R. J. Wooton. 1995. The cost of parental care in teleost fishes. Rev. Fish Biol. Fish. 5:7–22.

Sobel, J. M., G. F. Chen, L. R. Watt, and D. W. Schemske. 2010. The biology of speciation. Evolution 64:295–315.

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

Stuart, Y. E., T. Veen, J. N. Weber, D. Hanson, M. Ravinet, B. K. Lohman, C. J. Thompson, T. Tasneem, A. Doggett, R. Izen, et al. 2017. Contrasting effects of environment and genetics generate a continuum of parallel evolution. Nat. Ecol. Evol. 1:0158.

Svensson, O., J. Gräns, M. C. Celander, J. Havenhand, E. H. Leder, K. Lindström, S. Schöld, C. van Oosterhout, and C. Kvarnemo. 2017. Immigrant reproductive dysfunction facilitates ecological speciation. Evolution 71:2510–2521.

Theis, A., F. Ronco, A. Indermaur, W. Salzburger, and B. Egger. 2014. Adaptive divergence between lake and stream populations of an East African cichlid fish. Mol. Ecol. 23:5304–5322.

———. O. Roth, F. Cortesi, F. Ronco, W. Salzburger, and B. Egger. 2017. Variation of anal fin egg-spots along an environmental gradient in a haplochromine cichlid fish. Evolution 71:766–777.

———. W. Salzburger, and B. Egger. 2012. The function of anal fin egg-spots in the cichlid fish *Astatotilapia burtoni*. PLoS One 7:e29878.

Turner, G. F. 2007. Adaptive radiation of cichlid fish. Curr. Biol. 17:827–831.

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

Tyers, A. M., and G. F. Turner. 2013. Signal and preference divergence among populations of the non-endemic basal lake malawi cichlid fish *Astatotilapia calliptera* (Perciformes: Cichlidae). Biol. J. Linn. Soc. 110:180–188.

Vamosi, S. M., and D. Schluter. 2009. Sexual selection against hybrids between sympatric stickleback species: evidence from a field experiment. Evolution 53:874–879.

Zelditch, M. L., D. L. Swiderski, H. D. Sheets, and W. L. Fink. 2004. Geometric morphometrics for biologists. Elsevier, New York.

Associate Editor: J. McKinnon
Handling Editor: M. Servedio

Supporting Information

Additional supporting information may be found online in the Supporting Information section at the end of the article.

Figure S1. Position of 17 landmarks used for body shape analysis.
Figure S2. Absolute growth rates (mg/day) of adults introduced in the mesocosms per population: males (a) and females (b): dark grey – Kalambo lake (KaL), light grey – Ndole lake (NdL), black – Lufubu river (LfR); and photos of typical female from each population at the end of the experiment (c) showing signs of weight loss in mouthbrooding lake females.
Figure S3. Total number of detected mating events (a) and total number of surviving free-swimming F1 juveniles (b) from all replicated mesocosms per type of cross (female population. male population): KaL – Kalambo lake, NdL – Ndole lake, LfR – Lufubu river; dark grey – KaL.KaL, light grey – NdL.NdL, white – hybrid crosses.
Figure S4. Mean numbers of surviving offspring per brood ± SE per type of cross (female population. male population): KaL – Kalambo lake, NdL – Ndole lake, LfR – Lufubu river; dark grey – KaL.KaL, light grey – NdL.NdL, black – LfR.LfR, white – hybrid crosses.
Figure S5. Body shape transformation grids from Canonical variate analysis (CVA) along the CV1 and CV2 axes presented in Figure 4.
Figure S6. PCA of body shape. Adult body shape changes during the experiment: red – Kalamabo lake (KaL), yellow – Ndole lake (NdL), blue – Lufubu river (LfR); offspring body shape: red – KaL.KaL, yellow – NdL.NdL, orange – KaL.NdL, purple - LfR.KaL, green - LfR.NdL.
Figure S7. CVA of body shape from this (2016) and previous experiment (2014) described in Theis et al. 2014 that included the same Kalabo lake population (KaL) as well as two populations from the Kalambo (KaR) and the Lunzua (LzR) River.
Table S1. Sample size details.
Table S2. Generalized linear mixed models of survival of adult *A. burtoni* introduced to the mesocosms.
Table S3. Analyses of variance tables of mixed effect models on relative growth (rSGR).
Table S4. Generalized linear mixed models of mating status (1: mated, 0: unmated) of male *A. burtoni* introduced to the mesocosms.
Table S5. Generalized linear mixed models of mating success (proportion of all mating events in the mesocosm) of male *A. burtoni* introduced to the mesocosms.
Table S6. Results of Tukey-Kramer *post hoc* tests on the least square means of the full models for number of surviving offspring per pair from Table S7.
Table S7. Generalized linear mixed models of reproductive success (number of surviving free-swimming F1 offspring) of *A. burtoni* introduced to the mesocosms.
Table S8. Pairwise body shape differentiation among groups: Procrustes (upper triangular matrix) and Mahalanobis (lower triangular matrix) distances from the CVA (Fig. 4b): b – before the experiment, a – after the experiment.
Table S9. Environmental data for lake and river *A. burtoni* sampling sites.

Supporting Information for

Immigrant and extrinsic hybrid inviability contribute to reproductive isolation between lake and river cichlid ecotypes

Jelena Rajkov, Alexandra Anh-Thu Weber, Walter Salzburger, Bernd Egger

Table S1 Sample size details. Number of introduced adult fish per population and information about survival and reproduction per mesocosm (a), number of mating events, free-swimming F1 individuals used for offspring survival analysis and genotyped eggs per cross (b) and number of individuals used for body shape analysis (c): KaL – Kalambo lake, NdL – Ndole lake, LfR – Lufubu river.

a)

mesocosm	original stock males			original stock females			surviving males			surviving females			non-surviving reproducing males			non-surviving reproducing females			reproducing males			reproducing females			mating events	assigned F1 offspring	
	KaL	NdL	LfR	KaL	NdL	LfR	KaL	NdL	LfR	KaL	NdL	LfR	KaL	NdL	LfR	KaL	NdL	LfR	KaL	NdL	LfR	KaL	NdL	LfR		eggs	free-swimming
1	3	3	3	3	3	3	3	3	1	3	2	0	0	0	1	0	1	2	2	2	1	3	3	2	26	13	68
2	3	3	3	3	3	3	2	3	3	1	1	1	1	0	0	2	2	1	2	1	0	3	3	2	14	6	52
3	3	3	3	3	3	3	3	3	1	3	3	0	0	0	0	0	0	0	2	1	1	3	3	0	18	25	110
4	3	3	3	3	3	3	3	2	0	3	2	1	0	0	1	0	1	0	1	2	1	3	3	1	18	12	100
5	3	3	3	3	3	3	3	3	1	3	3	2	0	0	0	1	0	0	2	2	0	3	3	1	17	6	99
6	4	3	3	2	4	3	4	3	0	1	2	1	0	0	0	1	2	0	3	2	0	2	4	1	18	13	48
total	**19**	**18**	**18**	**17**	**19**	**18**	**18**	**17**	**6**	**14**	**13**	**5**	**1**	**0**	**2**	**4**	**6**	**3**	**12**	**10**	**3**	**17**	**19**	**7**	**111**	**75**	**477**

b)

female population. male population	number of mating events	total number of free-swimming F1	total number of genotyped F1 eggs
KaL.KaL	23	191	20
NdL.KaL	15	104	17
KaL.NdL	15	85	19
NdL.NdL	16	63	13
KaL.LfR	6	12	0
LfR.KaL	5	9	0
LfR.NdL	4	8	6
NdL.LfR	2	5	0
LfR.LfR	0	0	0
total	**86**	**477**	**75**

c)

	N
KaL-before	36
NdL-before	36
LfR-before	36
KaL-after	23
NdL-after	23
LfR-after	11
KaL.KaL	24
KaL.NdL	24
NdL.NdL	24
KaL.LfR	21
NdL.LfR	13
total	271

Table S2 Generalized linear mixed models of survival of adult *A. burtoni* introduced to the mesocosms (d.f.: degrees of freedom). Significant effects ($P < 0.05$) are highlighted in bold.

model 1: whole dataset

	residuals d.f. = 102		
Effect	d.f.	χ^2	*P*
population	**2**	**31.745**	**<0.001**
sex	**1**	**4.123**	**0.042**
initial size	1	0.007	0.935
size deviation	2	0.222	0.638

model 2: males only

	residuals d.f. = 49		
Effect	d.f.	χ^2	*P*
population	**2**	**24.075**	**<0.001**
initial size	1	0.533	0.465
size deviation	1	0.681	0.409

model 3: females only

	residuals d.f. = 48		
Effect	d.f.	χ^2	*P*
population	**2**	**11.396**	**0.003**
initial mass	1	0.090	0.764
size deviation	1	1.706	0.191

Table S3 Analyses of variance tables of mixed effect models on relative growth (rSGR). *F*-statistic was corrected with the Kenward-Roger approximation for mixed linear models. Significant effects ($P < 0.05$) are highlighted in bold. Num. d.f. – numerator degrees of freedom, den. d.f. – denominator degrees of freedom.

model 1: whole dataset

Effect	num.d.f.	den.d.f.	*F*	*P*
population	2	63.3	0.260	0.774
sex	**1**	**63.1**	**540.270**	**<0.001**
population: sex	**1**	**63**	**18.040**	**<0.001**

model 2: males only

Effect	num.d.f.	den.d.f.	*F*	*P*
population	**2**	**34.7**	**4.789**	**0.015**

model 3: females only

Effect	num.d.f.	den.d.f.	*F*	*P*
population	**2**	**24.8**	**25.429**	**<0.001**

Table S4 Generalized linear mixed models of mating status (1: mated, 0: unmated) of male *A. burtoni* introduced to the mesocosms (d.f.: degrees of freedom). Significant effects ($P < 0.05$) are highlighted in bold.

model 1: whole dataset

	residuals d.f. = 48		
Effect	d.f.	χ^2	*P*
population	**2**	**8.968**	**0.011**
male size deviation	**1**	**12.477**	**<0.001**
initial size	1	0.014	0.906

model 2: survivors only

	residuals d.f. = 34		
Effect	d.f.	χ^2	*P*
population	2	5.357	0.069
male size deviation	**1**	**11.761**	**<0.001**
initial size	1	0.012	0.912

Table S5 Generalized linear mixed models of mating success (proportion of all mating events in the mesocosm) of male *A. burtoni* introduced to the mesocosms (d.f.: degrees of freedom). Significant effects ($P < 0.05$) are highlighted in bold.

model 1: whole dataset

	residuals d.f. = 48		
Effect	d.f.	χ^2	*P*
population	**2**	**18.578**	**<0.001**
male size deviation	**1**	**19.837**	**<0.001**
initial size	1	2.692	0.101

model 2: survivors only

	residuals d.f. = 41		
Effect	d.f.	χ^2	*P*
population	2	5.891	0.053
male size deviation	**1**	**18.592**	**<0.001**
initial size	1	2.674	0.102

model 3: mated only

	residuals d.f. = 19		
Effect	d.f.	χ^2	*P*
population	**2**	**15.270**	**<0.001**
male size deviation	**1**	**11.761**	**<0.001**
initial size	1	1.988	0.158

Table S6 Results of Tukey-Kramer post hoc tests on the least square means of the full models for number of surviving offspring per pair from Table S7. Significant comparisons ($P < 0.05$) are highlighted in bold.

model 1:

contrast	estimate	SE	*Z*	*P*	
KaL.KaL - NdL.KaL	0.314	0.255	1.233	0.922	
KaL.KaL - LfR.KaL	1.624	0.461	3.525	**0.010**	
KaL.KaL - KaL.NdL	0.528	0.301	1.750	0.654	
KaL.KaL - NdL.NdL	1.091	0.371	2.942	0.065	
KaL.KaL - LfR.NdL	1.764	0.578	3.049	**0.047**	
KaL.KaL - KaL.LfR	1.937	0.534	3.627	**0.007**	
KaL.KaL - NdL.LfR	1.764	0.754	2.339	0.272	
NdL.KaL - LfR.KaL	1.311	0.463	2.828	0.088	
NdL.KaL - KaL.NdL	0.214	0.373	0.573	0.999	pooled
NdL.KaL - NdL.NdL	0.777	0.346	2.246	0.324	
NdL.KaL - LfR.NdL	1.450	0.589	2.463	0.212	
NdL.KaL - KaL.LfR	1.623	0.580	2.800	0.095	
NdL.KaL - NdL.LfR	1.450	0.736	1.971	0.502	
LfR.KaL - KaL.NdL	-1.097	0.541	-2.027	0.464	
LfR.KaL - NdL.NdL	-0.533	0.545	-0.979	0.977	
LfR.KaL - LfR.NdL	0.139	0.650	0.214	1.000	
LfR.KaL - KaL.LfR	0.313	0.696	0.449	1.000	pooled
LfR.KaL - NdL.LfR	0.139	0.850	0.164	1.000	
KaL.NdL - NdL.NdL	0.564	0.281	2.006	0.478	
KaL.NdL - LfR.NdL	1.236	0.521	2.372	0.255	
KaL.NdL - KaL.LfR	1.410	0.542	2.603	0.155	
KaL.NdL - NdL.LfR	1.236	0.757	1.632	0.731	
NdL.NdL - LfR.NdL	0.673	0.531	1.266	0.911	
NdL.NdL - KaL.LfR	0.846	0.586	1.444	0.837	
NdL.NdL - NdL.LfR	0.673	0.738	0.912	0.985	
LfR.NdL - KaL.LfR	0.174	0.731	0.238	1.000	
LfR.NdL - NdL.LfR	0.000	0.882	0.000	1.000	pooled
KaL.LfR - NdL.LfR	-0.173	0.645	-0.269	1.000	

model 2:

contrast	estimate	SE	*Z*	*P*
KaL.KaL - KaL.LfR	1.759	0.342	5.141	**<0.001**
KaL.KaL - KaL.NdL	0.417	0.202	2.068	0.234
KaL.KaL - NdL.LfR	1.704	0.471	3.620	**0.003**
KaL.KaL - NdL.NdL	1.059	0.354	2.990	**0.023**
KaL.LfR - KaL.NdL	-1.342	0.342	-3.925	**0.001**
KaL.LfR - NdL.LfR	-0.055	0.439	-0.125	1.000
KaL.LfR - NdL.NdL	-0.700	0.424	-1.652	0.464
KaL.NdL - NdL.LfR	1.287	0.415	3.104	**0.016**
KaL.NdL - NdL.NdL	0.642	0.243	2.639	0.064
NdL.LfR - NdL.NdL	-0.645	0.431	-1.497	0.565

Table S7 Generalized linear mixed models of reproductive success (number of surviving free-swimming F1 offspring) of *A. burtoni* introduced to the mesocosms (d.f.: degrees of freedom). Significant effects ($P < 0.05$) are highlighted in bold.

model 1

	residuals d.f. = 73		
Effect	d.f.	χ^2	*P*
cross (pop_f. pop_m)	**7**	**27.919**	**<0.001**
female initial size	1	0.266	0.606
male initial size	1	1.803	0.179

model 2

	residuals d.f. = 76		
Effect	d.f.	χ^2	*P*
cross (pooled)	**4**	**27.489**	**<0.001**
female initial size	1	0.359	0.549
male initial size	1	1.663	0.197

Table S8 Pairwise body shape differentiation among groups: Procrustes (upper triangular matrix) and Mahalanobis (lower triangular matrix) distances from the CVA (Fig. 4b): b – before the experiment, a – after the experiment. Significant body shape differences ($P < 0.05$) are highlighted in bold.

group	KaL_b	NdL_b	LfR_b	KaL.LfR	NdL.LfR	KaL_a	NdL_a	KaL.NdL	KaL.KaL	NdL.NdL	LfR_a
KaL_b		**0.01**	**0.02**	**0.02**	**0.02**	**0.02**	**0.01**	**0.02**	**0.02**	**0.02**	**0.03**
NdL_b	**2.25**		**0.02**	**0.03**	**0.03**	**0.02**	**0.02**	**0.03**	**0.03**	**0.03**	**0.03**
LfR_b	**7.00**	**6.17**		**0.02**	**0.02**	**0.03**	**0.03**	**0.03**	**0.03**	**0.03**	**0.02**
KaL.LfR	**5.44**	**6.00**	**6.32**		0.01	**0.02**	**0.02**	**0.01**	**0.01**	**0.02**	**0.02**
NdL.LfR	**5.66**	**5.92**	**5.93**	**2.41**		**0.02**	**0.02**	**0.02**	**0.02**	**0.02**	**0.02**
KaL_a	**4.11**	**4.95**	**7.33**	**3.32**	**4.12**		0.01	**0.01**	**0.01**	**0.01**	**0.03**
NdL_a	**4.32**	**4.80**	**6.69**	**3.49**	**3.70**	**2.12**		**0.01**	**0.01**	**0.01**	**0.03**
KaL.NdL	**5.06**	**5.83**	**7.95**	**3.32**	**4.12**	**2.14**	**2.78**		**0.01**	0.01	**0.03**
KaL.KaL	**5.14**	**6.19**	**8.22**	**3.94**	**4.41**	**2.31**	**2.91**	**2.27**		**0.01**	**0.03**
NdL.NdL	**5.28**	**5.85**	**7.43**	**3.67**	**3.98**	**2.40**	**2.26**	**2.01**	**2.53**		**0.03**
LfR_a	**8.29**	**8.32**	**5.01**	**5.08**	**5.01**	**6.87**	**6.47**	**7.19**	**7.40**	**6.63**	

Table S9 Environmental data for lake and river *A. burtoni* sampling sites.

location	date	oxygen (mg/l)	oxygen (mbar)	oxygen (%)	temperature (°C)	pH	conductivity (ppm)	conductivity (µS/cm)
Kalambo lake (KaL)	27.07.2015	7.52	186	98.5	24.5	9.3	330	661
Kalambo lake (KaL)	02.08.2015	/	/	/	25.5	9.25	331	660
Kalambo lake (KaL)	17.11.2015	3.12	83.4	41.7	/	8.86	325	652
Kalambo river (KaR)	30.07.2015	7.4	173.5	90.8	21.6	8.6	30	60
Kalambo river (KaR)	07.08.2015	6.6	157	86	/	/	/	/
Kalambo river (KaR)	18.11.2015	6.1	152.7	82.6	25.3	8	23	46
Lunzua lake (LzL)	04.08.2015	7.42	179	104.9	25/23.8	9.29	334	670
Lunzua river (LzR)	25.07.2015	9.2/8.4	204/194.1	111/105/103	21.1/20.9/21.9	8.14/8.03/7.98	9	18
Lufubu lake (LfL)	13.11.2015	5.79	152.9	81.8	27.6	7.6	/	35
Lufubu river (LfR)	10.11.2015	7.54	194	105	28.1	8.55	15	29

Figure S1 Position of 17 landmarks used for body shape analysis.

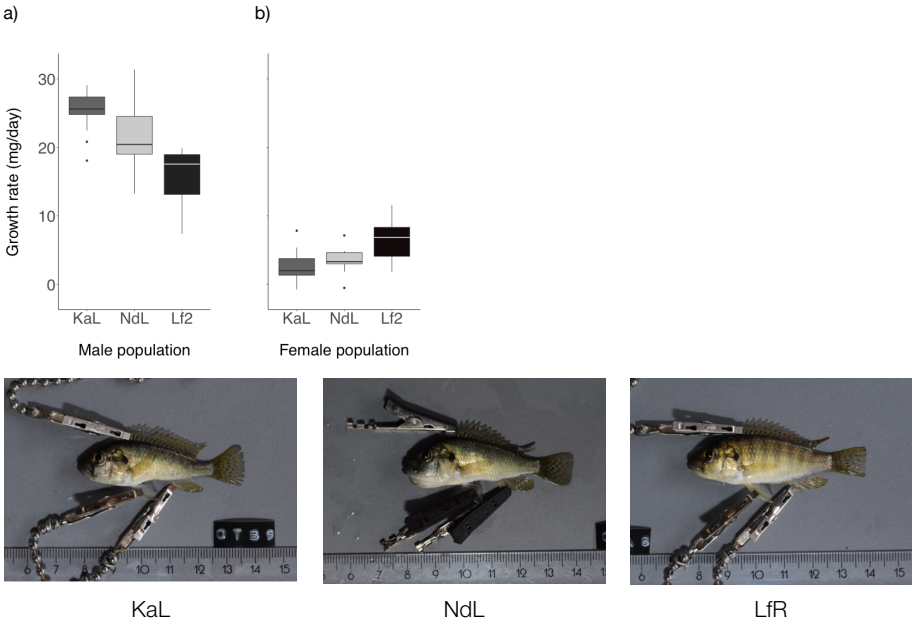


Figure S2 Absolute growth rates (mg/day) of adults introduced in the mesocosms per population: males (a) and females (b): dark grey – Kalambo lake (KaL), light grey – Ndole lake (NdL), black – Lufubu river (LfR); and photos of typical female from each population at the end of the experiment (c) showing signs of weight loss in mouthbrooding lake females.

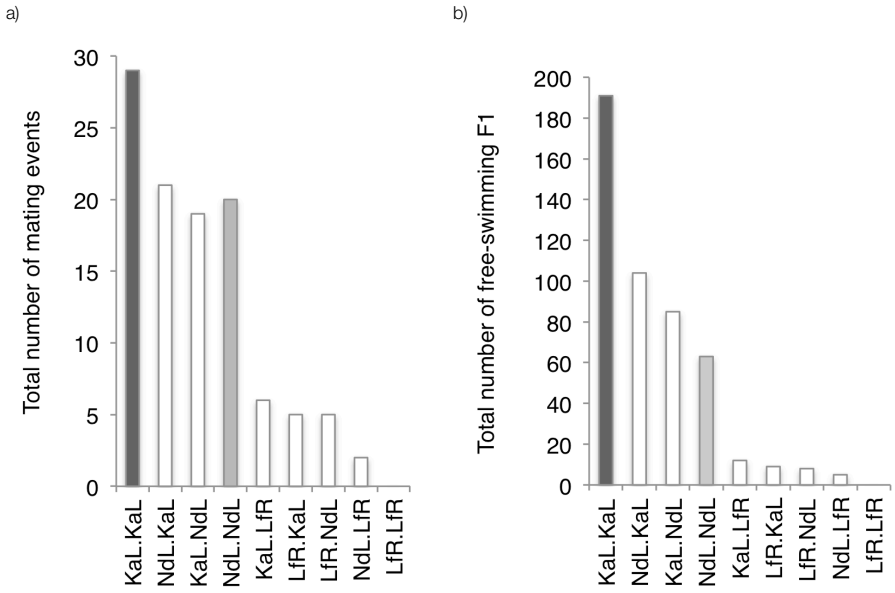


Figure S3 Total number of detected mating events (a) and total number of surviving free-swimming F1 juveniles (b) from all replicated mesocosms per type of cross (female population. male population): KaL – Kalambo lake, NdL – Ndole lake, LfR – Lufubu river; dark grey – KaL.KaL, light grey – NdL.NdL, white – hybrid crosses.

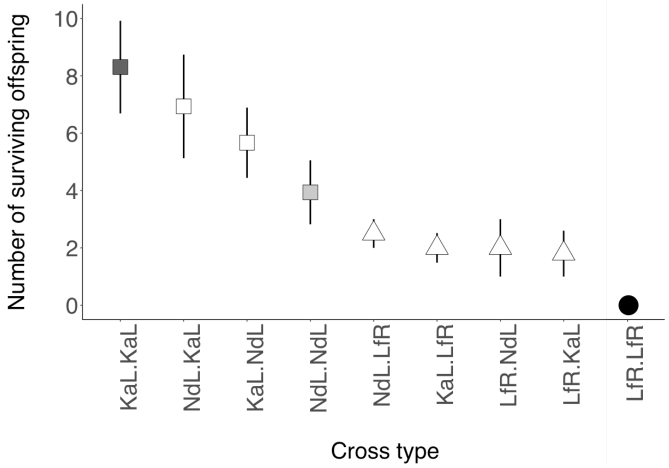
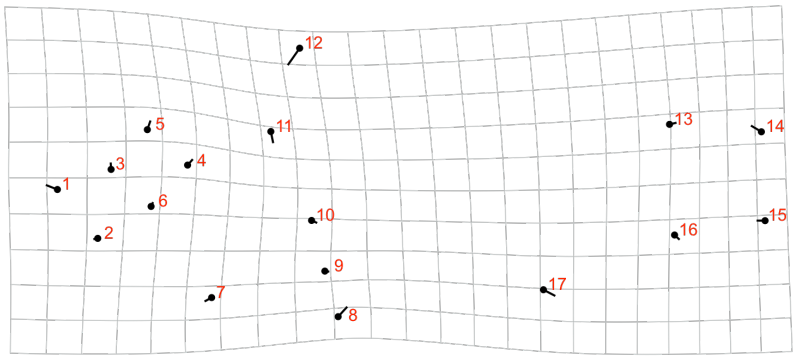
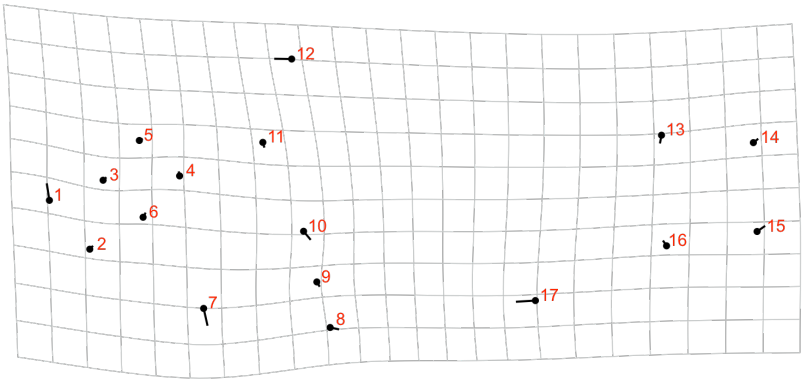


Figure S4 Mean numbers of surviving offspring per brood ± SE per type of cross (female population. male population): KaL – Kalambo lake, NdL – Ndole lake, LfR – Lufubu river; dark grey – KaL.KaL, light grey – NdL.NdL, black – LfR.LfR, white – hybrid crosses.



CV1



CV2

Figure S5 Body shape transformation grids from Canonical variate analysis (CVA) along the CV1 and CV2 axes presented in Figure 4.

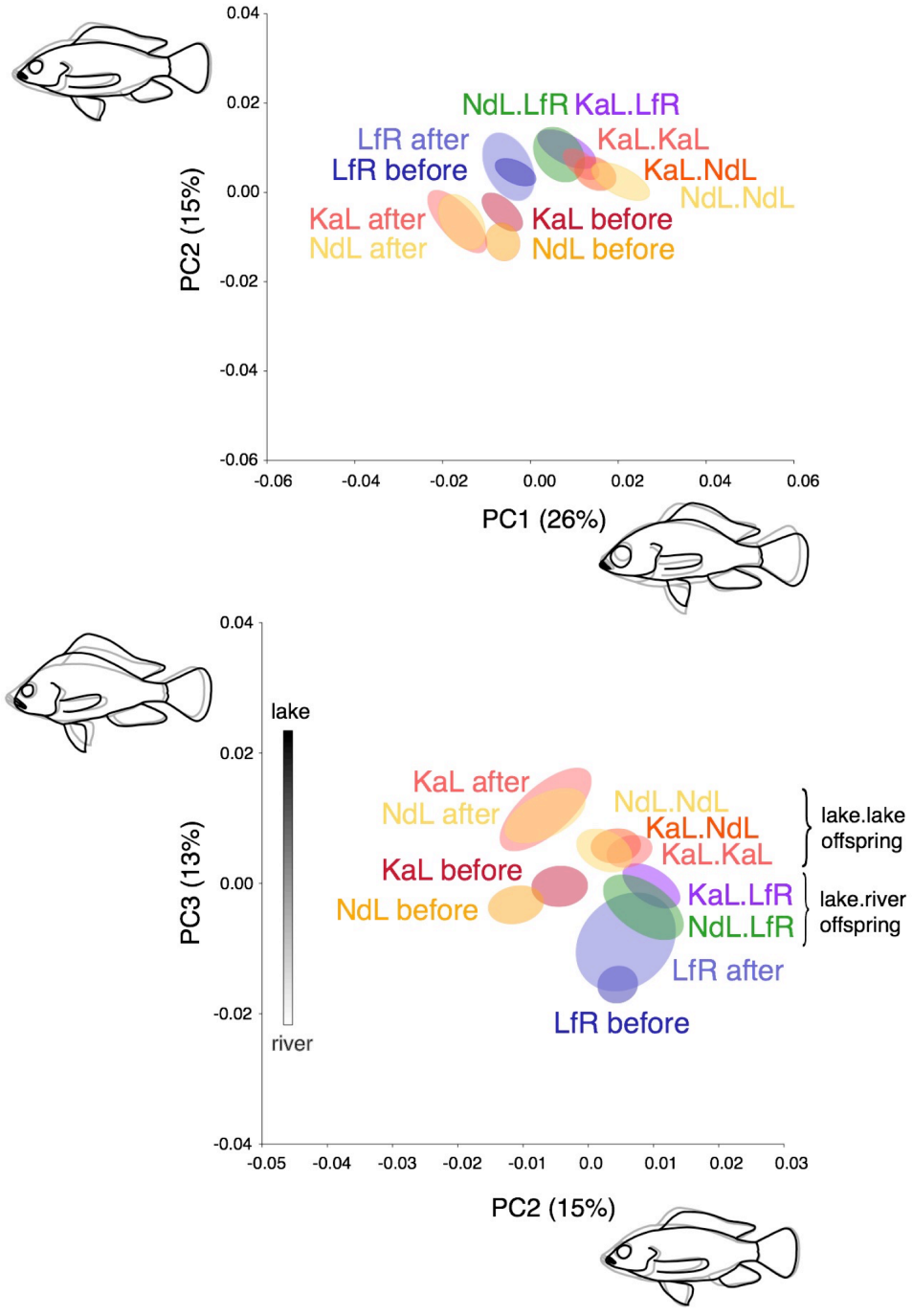


Figure S6 PCA of body shape. Adult body shape changes during the experiment: red – Kalamabo lake (KaL), yellow – Ndole lake (NdL), blue – Lufubu river (LfR); offspring body shape: red – KaL.KaL, yellow – NdL.NdL, orange – KaL.NdL, purple - LfR.KaL, green - LfR.NdL. PC shape change outlines are for illustration purposes only, from light grey to dark outlines with increasing values, scaling factor ten by default.

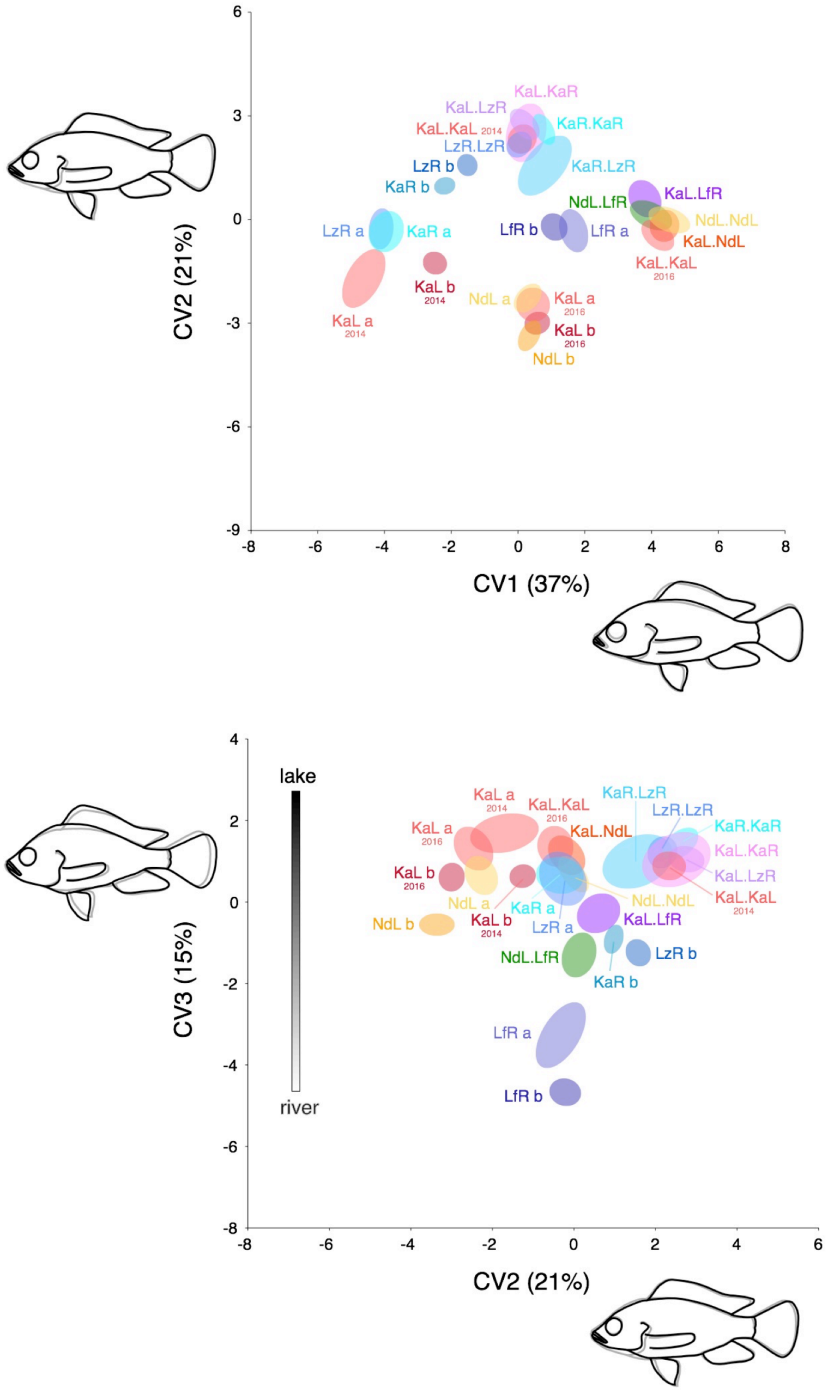


Figure S7 CVA of body shape from this (2016) and previous experiment (2014) described in Theis et al. 2014 that included the same Kalabo lake population (KaL) as well as two populations from the Kalambo (KaR) and the Lunzua (LzR) River. Adult body shape changes during the experiment: b – adult body shape before the experiment, a – adult body shape after the experiment. CV shape change outlines are for illustration purposes only, from light grey to dark outlines with increasing values, scaling factor ten by default.

Chapter 3

Gene expression remodelling and immune response during adaptation to new environments

J. Rajkov, A. El Taher, A. Böhne, W. Salzburger, B. Egger

In preparation

I designed the study together with AET. I performed the sampling and parasite screening together with BE. I performed the wet laboratory work together with AET. I analysed the parasite and microbiota data. I wrote the manuscript with input from all co-authors.

Gene expression remodelling and immune response during adaptation to new environments

Jelena Rajkov[1], Athimed El Taher[1], Astrid Böhne[1], Walter Salzburger[1], Bernd Egger[1]

[1] Zoological Institute, University of Basel, Vesalgasse 1, CH-4051 Basel, Switzerland

Abstract

Variation in gene expression has been suggested to play an important role in ecological speciation by facilitating population persistence in novel environments and by contributing to adaptive genetic divergence. The comparative analysis of gene expression in populations from different environments can furthermore uncover ecologically important phenotypes that are not obvious from morphology alone, and thus lead to a more complete understanding of the role of phenotypic plasticity in adaptive evolution. Here, we investigated the potential contribution of habitat-specific parasites and symbionts as well as the underlying immunological capabilities of ecotype hosts in lake-river population pairs of the haplochromine cichlid fish *Astatotilapia burtoni*. To better understand the role of phenotypic plasticity in adaptive divergence, we compared gene expression patterns, parasite and microbiota communities of fish from natural habitats and a lake-like pond setup. We found that lake populations were more heavily parasitized than river populations, both in terms of parasite taxa composition and infection abundance. Innate immune response in the wild was higher in lake than in river populations and elevated in a river population exposed to lake parasites in the pond setup. Transcriptome analyses provided evidence that environmental differences between lake and river habitat and their distinct parasite communities shape differential gene expression. Finally, by comparing gene expression patterns and bacterial communities between wild-caught individuals and individuals acclimated to (lake-like) pond conditions, we show that shifts in gene expression and microbiota composition contribute to adaptive phenotypic plasticity.

Introduction

Colonization of novel environments is a key step in ecological divergence between populations. Adaptation to new environments can occur through the fixation of pre-existing genotypic differences, novel mutations, via phenotypic plasticity, or a combination thereof (Pfennig et al., 2010; Price, Qvarnstrom, & Irwin, 2003; Schneider & Meyer, 2017; West-Eberhard, 2003).

Adaptive phenotypic plasticity—that is, the ability to generate a phenotype from the same genotype that is better suited for a novel environment (Ghalambor, McKay, Carroll, & Reznick, 2007)—can promote the expansion of populations into new niches (Richards, Bossdorf, Muth, Gurevitch, & Pigliucci, 2006; Thibert-Plante & Hendry, 2011; Yeh & Price, 2004). Adaptive phenotypic plasticity can temporarily shield genetic diversity from the direct impact of natural selection, thereby generating a time-lapse between the origin of beneficial mutations and their spread within a population, which may eventually result in genetic differentiation (Schlichting, 2004). The process by which originally plastic traits can become genetically fixed in a homogeneous environment has been termed ‘genetic assimilation’ (Waddington, 1942, 1953).

However, we are far from understanding how genetic assimilation operates at the molecular level (Ehrenreich & Pfennig, 2016).

Variation in gene expression has been suggested to play an important role in ecological speciation (Pavey, Collin, Nosil, & Rogers, 2010) by (*i*) facilitating population persistence in a new environment and (*ii*) contributing to adaptive genetic divergence which can lead to reproductive isolation and ultimately to speciation. The comparative analysis of gene expression patterns can uncover ecologically important phenotypes that are not immediately evident from the inspection of the overall morphology, such as physiological differences (Pavey et al., 2010). Evidence of parallel trajectories in gene expression i.e. repeated regulation of gene expression in the same direction in one environment compared to another, allows for a more complete evaluation of the degree of parallelism in adaptive divergence than focussing only on parallel genomic divergence (Hanson, Hu, Hendry, & Barrett, 2017). Furthermore, studying gene expression plasticity leads to a more inclusive understanding of the role of phenotypic plasticity in adaptive evolution, moving away from purely theoretical predictions (Ghalambor et al., 2007; Price et al., 2003; West-Eberhard, 2005), in particular when applied to the entire transcriptome level of wild populations (Aubin-Horth & Renn, 2009; Todd, Black, & Gemmell, 2016).

Transplant and common garden experiments are commonly used to characterize the relative contribution of heritability *versus* plasticity for ecologically relevant traits (Schlichting & Pigliucci, 1998; West-Eberhard, 2003). However, most of these studies focused on obvious phenotypic traits such as morphology and size, neglecting traits such as immunity and physiological homeostasis, which may be equally important for performance and fitness (Lohman, Stutz, & Bolnick, 2017). In addition, some drivers of divergent selection via adaptation to different environments, such as predation and resource competition, have received considerable scientific attention (Bolnick, 2004; Ingley & Johnson, 2016; Marchinko, 2009; Nosil, 2004; Nosil & Crespi, 2006; Schluter, 2003; Svanback & Bolnick, 2007; Vamosi & Schluter, 2002; Vamosi, 2003), while local adaptation related to parasite pressure has often been neglected (Rundle & Nosil, 2005). Differences in the abiotic and biotic conditions in different habitats can result in the development and maintenance of contrasting parasite communities (Eizaguirre et al., 2011; Halmetoja, Valtonen, & Koskenniemi, 2000). Selection on immune system diversification can be particularly strong when hosts encounter novel parasites such as after colonization of new habitat (Diepeveen, Roth, & Salzburger, 2013; Jones et al., 2012; Matthews, Harmon, M'Gonigle, Marchinko, & Schaschl, 2010; Scharsack, Kalbe, Harrod, & Rauch, 2007). Thus, immune response is another potentially relevant factor in speciation driven by adaptation to different environments (Eizaguirre & Lenz, 2010). The co-evolutionary dynamics of hosts and their parasites are expected to lead to local immunogenetic adaptation in the hosts as well as local adaptation of parasite infectivity and virulence (Eizaguirre & Lenz, 2010; Kaltz & Shykoff, 1998; Kawecki & Ebert, 2004). The immune system of vertebrates consists of two components: i) innate immunity is the first line of defence and consists of non-specific mechanisms to protect hosts from infection (Medzhitov & Janeway, 2002), and ii) the adaptive immune response is characterized by a specific antigen recognition that drives a secondary pathogen-specific reaction (Castro & Tafalla, 2015).

The vertebrate immune system is also affected by host-associated microorganisms, whose importance and ubiquity (for review see McFall-Ngai et al., 2013) are stimulating work on the role of the microbiome in animal speciation (Brucker & Bordenstein, 2012, 2013; Sharon et al., 2010; Shropshire & Bordenstein, 2016). Plasticity of the gut microbiota has been proposed as an essential factor determining phenomic plasticity of vertebrates, but this hypothesis remains largely unexplored as most studies to date have focused on gut microbiota composition at one time point rather than its plasticity (Alberdi, Aizpurua, Bohmann, Zepeda-Mendoza, & Gilbert, 2016).

The haplochromine cichlid fish *Astatotilapia burtoni* inhabits Lake Tanganyika and affluent rivers (Figure 1, Table S1) - environments that differ in water chemistry, habitat structure and prey availability (Theis, Ronco, Indermaur, Salzburger, & Egger, 2014). *A. burtoni* lake and river ecotypes are an excellent model to study population pairs along a 'speciation continuum', from early divergence to a considerable degree of reproductive isolation (Egger, Roesti, Böhne, Roth, & Salzburger, 2017; Rajkov, Weber, Salzburger, & Egger, 2018a, 2018b; Theis et al., 2017, 2014). The ecotypes show habitat-specific adaptations in body shape and trophic structures that correspond to different diets (Theis et al., 2014). Common garden experiments demonstrated that differences in body shape and gill raker length have both a plastic and a genetic component (Theis et al., 2014). A transplant experiment using F1-lake-river crosses raised in a common garden lake-like setup identified a substantial contribution of adaptive phenotypic plasticity to increased immigrant fitness (Rajkov et al., 2018a). River and lake ecotype also differ in number and relative area of so-called egg-spots (ovoid markings on the anal fins of males) both of which correlate with adaptive immune response (Theis et al., 2017). Genome scans identified candidate genes with immune functions that show differences in expression between lake and river populations (Egger et al., 2017).

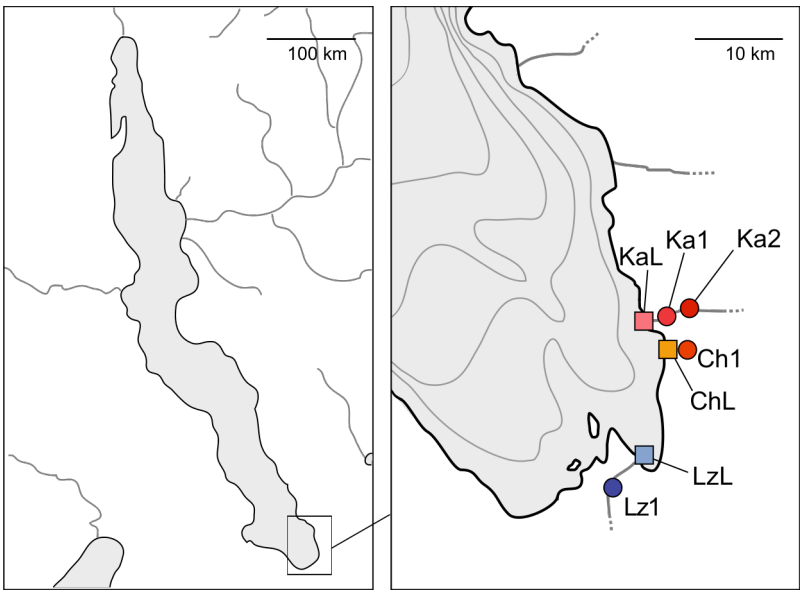


Figure 1 *Astatotilpia burtoni* lake and river populations from the south-east of Lake Tanganyika used in this study. KaL – Kalambo lake, Ka1 – Kalambo river, Ka2 – Kalambo river upstream, ChL – Chitili lake, Ch1 – Chitili river, LzL – Lunzua

Here, we use *A. burtoni* lake-river population pairs to investigate the potential role of habitat-specific parasites and symbionts and the underlying immunological capabilities of ecotype hosts in lake-river divergence. To better understand the role of phenotypic plasticity in adaptive divergence (Rajkov et al., 2018a), we compare gene expression patterns, parasite and microbiota communities of fish from natural habitats and a lake-like pond setup. Previous studies examining gene expression differences between recently diverged ecotypes brought valuable insights but often relied on only one pair of populations (Ghalambor et al., 2015; Jeukens, Renaut, St-Cyr, Nolte, & Bernatchez, 2010; Lenz, Eizaguirre, Rotter, Kalbe, & Milinski, 2013), targeted the expression of only a subset of genes (McCairns & Bernatchez, 2009; Morris et al., 2014), or used wild caught-individuals (Huang, Chain, Panchal, Eizaguirre, & Kalbe, 2016). Here, we use RNA-sequencing (RNA-seq) of replicate populations from two different environments, including their

natural habitat range and a pond lake-like set-up, to evaluate parallelism and plasticity of expression patterns between divergent ecotypes.

Material and methods

Study populations and pond experiment

For parasite screening and RNA-seq, we sampled between 16 and 22 adult *A. burtoni* specimens per locality at a ~1:1 sex ratio from lake and river populations at the Kalambo River (Kalambo lake - KaL, Kalambo river - Ka1 and Kalambo river upstream - Ka2) and the Lunzua River (Lunzua lake - LzL, Lunzua river - Lz1) (Figure 1) in August 2017 (see Table 1 for sampling locations and sample sizes). Additionally, 10 adults per locality at a [~]1:1 sex ratio were sampled for parasite screening from the Chitili River (Chitili lake - ChL, Chitili river - Ch1) in September 2018. For microbiota genotyping, 20 individuals per population were sampled from the same locations at Kalambo and Lunzua River in August 2017. At each location, conductivity and pH of the water were measured with a pH meter (HI 73127; Hanna Instruments, Woonsocket, RI, USA). Fish were caught by hook and line fishing and transported to the field station at Kalambo Lodge in buckets filled with water from sampling locations. Specimens were either processed within a few hours upon capture or kept alive for one night in concrete ponds filled with water from the respective sampling location. To evaluate plasticity in parasite and microbiota communities as well as in gene expression, we included 26 specimens from a previous experiment involving wild-caught adults (November 2015) from Kalambo lake (KaL) and river (Ka2) populations. These fish were kept in two separate concrete ponds supplied with lake water (dimensions: 2m x 1m x 1m; length x width x depth) at high density (~100 individuals per pond) from July 2016 until August 2017 and fed daily with commercial flake food. Fish were euthanized by pithing and immediately photographed, measured (±0.5 mm), weighed (±5 mg), sexed by visual inspection of external colouration and the genital papilla, and fin-clipped. Sampling and parasite screening were performed under study permits number 003376 and 004264 issued by Lake Tanganyika Research Unit, Department of Fisheries, and the Immigration Department, Republic of Zambia.

Table 1 Geographic coordinates of sampling locations, and sample size details for parasite screening and splenosomatic index (SI), RNA-seq and microbiota analysis (16S seq). For Chitili lake and stream population only parasite data was collected.

							sample size									
					GPS coordinates		parasite screening and SI				RNA-seq			16S seq		
site	system	type	habitat	year	latitude	longitude	males	females	juveniles	total	males	females	total	males	females	total
LzL	Lunzua	wild	lake	2017	8°44'57.13″S	31°10'21.86″E	11	8	1	20	4	4	8	4	4	8
Lz1	Lunzua	wild	river	2017	8°47'23.51″S	31°8'14.33″E	11	9	0	20	4	4	8	4	4	8
KaL	Kalambo	wild	lake	2017	8°36'6.27″S	31°11'13.24″E	12	10	0	22	4	4	8	4	4	8
Ka1	Kalambo	wild	river	2017	8°35'35.23″S	31°11'6.18″E	11	5	0	16	4	4	8	5	3	8
Ka2	Kalambo	wild	river	2017	8°35'6.24″S	31°12'29.32″E	13	6	0	19	4	4	8	4	4	8
KaL pond	Kalambo	pond	lake	2017	8°37'25.5"S	31°12'03.7"E	6	6	0	12	4	4	8	5	3	8
Ka2 pond	Kalambo	pond	river	2017	8°37'25.5"S	31°12'03.7"E	8	6	0	14	4	4	8	4	4	8
ChL	Chitili	wild	lake	2018	8°38'18.42″S	31°11'55.34″E	6	5	0	11	NA	NA	NA	NA	NA	NA
Ch1	Chitili	wild	river	2018	8°38'16.91″S	31°12'4.02″E	6	6	0	12	NA	NA	NA	NA	NA	NA

Parasitological survey

Immediately after measuring each fish (see above), we dissected gills, skin, fins, eyes, intestinal tract, heart and liver, and exhaustively screened these organs for metazoan ecto- and endoparasites using a field stereomicroscope (Nikon, SMZ445). We first screened the outer surface of each *A. burtoni* specimen for monogeneans and crustaceans. Next, fins and gills were

dissected and screened separately for ectoparasites. The gastrointestinal tract was excised from oesophagus to anus and placed in saline solution (0.9% NaCl). Livers and hearts were dissected and inspected in a Petri dish filled with lake water. Finally, the intestinal content and teased internal organs were pressed between two glasses of a Petri dish for examination. Endo-parasite specimens were separated according to higher helminth taxa (Acanthocephala, Cestoda, Digenea and Nematoda) following Paperna (1996).

RNA-seq library preparation an sequencing

We focused on immune and homeostasis related organs – spleen and gills, as those are most likely to be involved in adaptation to different osmotic habitats and parasite pressure. Immediately after parasite screening of the gills, one gill from each side as well as the spleen were dissected and preserved in RNAlater (Sigma-Aldrich) and later transferred to Individual TRIzol tubes which were weighed (±0.05 mg) before and after the spleen was added to calculate the spleen weight. After bead-beating for homogenization in TRIzol and chloroform phase separation, samples were transferred to Zymo Direct-zol™-96 RNA plates (Zymo Research). Quality control was done using Nanodrop (Thermo Scientific) and TapeStation 4200 (Agilent) devices. Eight individuals per sampling site were selected for RNA-sequencing based on the RNA-quality while ensuring an equal sex distribution for each population. Sample selection was performed ignoring parasite screening results. Individual libraries from 112 samples (56 specimens, two tissues) were constructed using the TruSeq Stranded Total RNA Library Prep Kit Ribo-Zero Gold (Illumina). Single-end-sequencing to 125-bp reads in five Illumina HiSeq2500 lanes was performed at the Genomics Facility Basel jointly operated by the Department of Biosystem Science and Engineering (D-BSSE), ETH Zürich, and the University of Basel.

Microbiota

Whole intestines and swabs of the buccal mucosa were placed in 99% ethanol for subsequent DNA extraction. The stomach was removed from the intestines before the extraction to remove undigested content. DNA from ethanol preserved cotton swabs was extracted following a modified DNeasy Blood & Tissue Kit (Qiagen) protocol (Keller, Bayer, Salzburger, & Roth, 2017). The modifications included evaporating EtOH at 60°C using vacuum (Concentrator plus, Eppendorf), overnight incubation at 56°C after the addition of ATL buffer and Proteinase *K* into the swab tube, incubation at 56°C for 30 min after addition of AL buffer into the swab tube to increase the recovery of Gram-positive bacteria, and final elution in 2 x 60 µl. DNA from ethanol preserved intestines was extracted using a modified version of repeated bead-beating plus column (RBB+C) protocol following Baldo et al. (2015). For Illumina sequencing of amplicons, DNA was amplified with 16S rRNA gene primers that target the V3-V4 hypervariable region. The library preparation protocol followed Baldo et al. (2017), with minor modifications (see Table S1 for primer sequences). A subset of eight individuals per population with [~]1:1 sex ratio was selected for library preparation after an initial PCR step to test for positive amplification. We used 25 cycles for the first and 15 cycles for the second PCR. Amplicons were cleaned individually with the UltraClean 96 PCR Cleanup Kit (Qiagen). All samples were then quantified with Qubit High Sensitivity Kit (Thermo Scientific) and pooled at equimolar concentrations to create the final library. This library was sequenced on an Illumina MiSeq v.3 (600 cycle cartridge, 300 bp paired-end) with 15% PhiX, at the Genomics Facility Basel.

Data analysis

Parasites and immune response

Infection presence, prevalence (percentage of infected fish per population), abundance (sum of parasite individuals on/in infected fish divided by the number of dissected fish) and median infection intensity (median number of parasite individuals on infected fish) were calculated following Rózsa et al. (2000) for each group of parasites and each population. We calculated three summary variables as estimates of overall parasite infestation rate: (*i*) number of parasite taxa (the number of parasite taxa infecting one fish), (*ii*) total parasite load (the sum of all parasites infecting one fish), and (*iii*) gill parasite load (the sum of all gill parasites infecting one fish).

Differences in (*i*) number of parasite taxa, and (*ii*) total and (*iii*) gill parasite load between different populations were analysed using zero-inflated generalized linear models (GLMs) with negative binomial (number of parasite taxa) or Poisson (parasite load) probability distribution and with population, standard length, and sex as predictors using the zeroinfl function in the pscl package (version 1.5.2) (Zeilies, Kleiber, & Jackman, 2008). Significance of the model terms was determined based on likelihood ratio tests using lrtest in lmtest package (version 0.9.35) (Achim & Hothorn, 2002).

Following an approach similar to that in Karvonen et al. (2015) and Rahn et al. (2016), we estimated pairwise differences in parasitic faunas between the populations using dissimilarity in parasite taxa composition (1- Jaccard similarity). Proportional similarity in parasite taxa composition between all possible population combinations was calculated as Jaccard similarity: a/(a+b+c), where a is the number of parasite taxa found in both populations, and b and c are numbers of taxa found only in the first and the second host population, respectively (Karvonen et al., 2015; Magurran, 1988). We performed Mantel tests (9999 permutations) to test for significant correlations between dissimilarity in parasitic fauna, and absolute differences in pH. Partial Mantel tests were performed to test for significant correlations between dissimilarity in parasitic fauna and pairwise genetic differentiation (F_{ST}) estimated from whole-genome data (Weber et al., in preparation) while accounting for the effect of geographical waterway distance between the sampling sites (Table S2).

We also calculated the splenosomatic index (SI – spleen mass to body mass ratio) to estimate innate immune response to parasite infection (Hadidi, Glenney, Welch, Silverstein, & Wiens, 2008; Kaufmann, Lenz, Kalbe, Milinski, & Eizaguirre, 2017; Lefebvre, Mounaix, Poizat, & Crivelli, 2004). Immune response was analysed in a linear model with the splenosomatic index as a response variable and population, total number of parasites per individual, size and sex as fixed effects.

Total (gill and gut) parasite load, gill parasite load and splenosomatic index were analysed in all wild populations (model 1). Gill parasite load and splenosomatic index were analysed separately in Kalambo lake-river pair (KaL-Ka2) (model 2) to compare these populations in the wild and in the experimental setup. Tukey-Kramer posthoc tests were applied to test for significance of pairwise comparisons between populations using the lsmeans package (version 2.27.61) (Lenth, 2016). Statistical analyses were performed in R version 3.3.2 (R Core Team, 2016).

RNA-seq

Read filtering and mapping

Illumina strand-specific single-end sequences of each library were filtered using Trimmomatic version 0.33 (Bolger, Lohse, & Usadel, 2014) with a four bp window size, required window quality of 15 and a read minimum length of 80 bp (2/3 of the initial read length). Adapters were removed using Trimmomatic. Cleaned reads were mapped against the Nile tilapia (*Oreochromis niloticus*) genome assembly (RefSeq assembly version GCF_001858045.1_ASM185804v2, Brawand et al.,

2014) with STAR version 2.5.2a (Dobin et al., 2013) with the following settings: --outFilterMultimapNmax 1 --outFilterMatchNminOverLread 0.4 --outFilterScoreMinOverLread 0.4. The unique alignments were reported in sorted BAM format and assigned to genes using the HTSeq-count script from the HTSeq framework (version 0.6.1p1) (Anders, Pyl, & Huber, 2015).

Global expression patterns

Prior to all analyses, we excluded genes with very low expression levels, which we considered as noise (present in less than four samples with less than three counts per sample). Expression values were then normalized with the DESeq2 (version 1.24.0.) R (version 3.5.0) package (Love, Huber, & Anders, 2014). We used the tissue type (gills vs. spleen), the habitat type (lake vs. river), the population (Kalambo vs. Lunzua) and the experimental condition (wild vs. pond) in the DESeq2 design. Variance stabilizing transformation (VST) was applied to the normalized expression data to minimize differences between samples when plotting the data. We used principal components analysis to illustrate the global patterns of gene expression differences as implemented in DESeq2. To summarize the data, replicates of each population within each organ were grouped by calculating their median of expression. The summarized expression values were then transformed into transcripts Per Kilobase Million (TPM) values. These values were used for all downstream analyses apart from differential expression analyses.

Rate of gene expression evolution

Following the method of Brawand et al. (2011), 'expression trees' were constructed using the neighbour-joining approach on the pairwise euclidean distance between populations for all protein-coding genes (28,938 genes) and separately for each organ. All neighbour-joining trees were assembled using the R package ape (version 5.3) (Paradis, Claude, & Strimmer, 2004). Branch lengths were represented as the sum of the branch lengths for river samples and the sum of the branch lengths for lake samples.

Differential gene expression analyses

To identify differentially expressed (DE) genes between the two habitats, we compared lake and river samples for each organ (gills and spleen) with DESeq2 by grouping Ka1, Ka2 and Lz1 samples into a river group and KaL and LzL into a lake group. We added the lake-river system information (Kalambo and Lunzua) as interaction factor in our model, following Deseq2 (Love et al., 2014) recommendations, as we identified phylogenetic signal in the expression data. We considered genes as DE with an adjusted p-value (FDR) below 0.05. Next, we intersected all pairwise gene expression comparisons and reported all intersecting genes.

Modules of co-expressed genes and module–trait associations

To identify co-expressed genes associated with habitat type, we constructed signed weighted gene co-expression networks with the R package WGCNA (version 1.66) (Langfelder & Horvath, 2008). We followed the filtering steps recommended in Langfelder & Horvath (2008), which resulted in a matrix of 23,770 genes retaining 39 specimens for the spleen samples and 25'209 genes retaining 38 samples for the gill samples after outlier exclusions. Spleen and gills were analyzed separately and only natural populations (Lz1, LzL, Ka1, Ka2 and KaL) were used for the analysis. Module-trait association analyses were calculated using a weighted Pearson correlation as recommended in Langfelder & Horvath (2008).

Habitat-Specific Variation

We pooled the sequencing reads of the two organs per individual and quality filtered and trimmed them with Trimmomatic 0.33 (Bolger et al., 2014) with a four bp window size, a required window quality of 15 and a minimum read length of 30 bp. We then performed reference-free *de novo* variant calling per population (Kalambo and Lunzua) with KisSplice 2.4.0 (Lopez-Maestre et al., 2016). Only Ka2 samples were used as river samples for the Kalambo system (Ka1 was excluded prior to the analysis as this population is geographically and genetically very close to the lake population (Egger et al., 2017)). KisSplice was run with -s 1 -t 4 -u and –experimental. The so-identified SNPs were placed on the Nile tilapia genome assembly with STAR 2.5.2a (Dobin et al., 2013) (--outFilterMultimapxNmax 1 --outFilterMatchNminOverLread 0.4 --outFilterScoreMinOverLread 0.4). The genome index used for this mapping step was generated with the corresponding STAR parameters: --runMode genomeGenerate, --sjdbOverhang 124, --sjdbGTFfeatureExon exon and the genome annotation file (RefSeq GCF_001858045.1_ASM185804v2). Kiss2Reference (default parameters) was used to classify KisSplice events aligned to the Nile tilapia reference genome and kissDE 1.4.0 (default parameters) was used to determine variants that differ between the two habitats (river and lake). The KisSplice events were filtered according to the following attributes in R 3.5.1: Only SNPs were kept; SNPs placed on mitochondrial DNA or on unplaced scaffolds of the reference genome were removed; only SNPs with significant p-values for an allele difference between habitats (p-value <= 0.05 after adjustment for multiple testing following the Benjamini & Hochberg method) were kept. We then defined candidate genes as genes with SNPs with significant p-values in both populations (73 genes, Table 8).

Plasticity and genetic assimilation

We pooled the sequencing reads of the two organs per individual and quality filtered and trimmed them with Trimmomatic 0.33 (Bolger et al., 2014) with a four bp window size, a required window quality of 15 and a minimum read length of 30 bp. We then performed reference-free *de novo* variant calling per population (Kalambo and Lunzua) with KisSplice 2.4.0 (Lopez-Maestre et al., 2016). Only Ka2 samples were used as river samples for the Kalambo system (Ka1 was excluded prior to the analysis as this population is geographically and genetically very close to the lake population (Egger et al., 2017)). KisSplice was run with -s 1 -t 4 -u and –experimental. The so-identified SNPs were placed on the Nile tilapia genome assembly with STAR 2.5.2a (Dobin et al., 2013) (--outFilterMultimapxNmax 1 --outFilterMatchNminOverLread 0.4 --outFilterScoreMinOverLread 0.4). The genome index used for this mapping step was generated with the corresponding STAR parameters: --runMode genomeGenerate, --sjdbOverhang 124, --sjdbGTFfeatureExon exon and the genome annotation file (RefSeq GCF_001858045.1_ASM185804v2). Kiss2Reference (default parameters) was used to classify KisSplice events aligned to the Nile tilapia reference genome and kissDE 1.4.0 (default parameters) was used to determine variants that differ between the two habitats (river and lake). The KisSplice events were filtered according to the following attributes in R 3.5.1: Only SNPs were kept; SNPs placed on mitochondrial DNA or on unplaced scaffolds of the reference genome were removed; only SNPs with significant p-values for an allele difference between habitats (p-value <= 0.05 after adjustment for multiple testing following the Benjamini & Hochberg method) were kept. We then defined candidate genes as genes with SNPs with significant p-values in both populations (73 genes, Table 8).

Microbiota

PCR negative control samples resulted in very low read coverage (≤ 505) suggesting that contamination was negligible and were hence excluded from the final dataset. Sequence raw reads were quality filtered, assembled into contigs, and classified using Mothur (version 1.39.5) (Schloss et al., 2009) according to the Mothur Illumina MiSeq SOP. Taxonomic classification was performed against the RDP database (Cole et al., 2014) using the classify.seqs function with a bootstrap cut-off of 80%. Chloroplasts, mitochondria, and non-bacterial sequences were removed. To obtain operational taxonomic units (OTUs), sequences were clustered at a 0.03 distance level. The OTU table output of Mothur was imported into R (version 3.3.2) for further processing using the R package phyloseq (version 1.19.1) (McMurdie & Holmes, 2013).

We retained only OTUs shared across two or more samples, and the OTU abundance table was subsampled to even sequencing depth using rarefy_even_depth function in phyloseq. Alpha (inverse Simpson index) and beta (Bray–Curtis distances) diversity measures were calculated in phyloseq. Pairwise Bray–Curtis distances among samples were visualized using non-metric multidimensional scaling (NMDS). The permutational analysis of variance (PERMANOVA) (10000 permutations) was performed with the vegan package (version 2.4-5) in R separately for gut and mouth microbiota to test the habitat effects, the population effects, and their interaction.

Results

Higher parasite diversity and abundance in the lake than in river habitats

The parasite community differed substantially between lake and river fish within lake-river systems (Table 2). Gut parasites (nematodes and Acantocephala), as well as parasitic copepods on the gills, were only present in the lake populations (except for one individual from the Chitili River that could be a recent migrant). River populations Ka1 and Lz1 showed only monogenean infections and no parasites were detected in the Ka2 upstream river population. However, in the specimens kept in the pond setup with lake water, Ka2 fish acquired copepod and monogenean parasites.

Table 2 Sampling site, lake-river system, habitat type, sample size, and prevalence (%)/mean abundance/median intensity of different groups of parasites for five population of *A. burtoni* from the wild and two from a pond setup.

					gill ectoparasites		gut endoparasites	
site	system	type	habitat	N	Monogenea	Copepoda	Acantocephala	Nematoda
LzL	Lunzua	wild	lake	20	55 / 2.1 / 2	90 / 4.45 / 3	5 / 0.05 / 1	65 / 2.3 / 4
Lz1	Lunzua	wild	river	20	50 / 2.2 / 3	0 / 0 / -	0 / 0 / -	0 / 0 / -
KaL	Kalambo	wild	lake	22	59.09 / 3.32 / 4	81.82 / 3.77 / 4	13.64 / 0.09 / 1	9.09 / 0.14 / 1
Ka1	Kalambo	wild	river	16	56.25 / 1 / 1	0 / 0 / -	0 / 0 / -	0 / 0 / -
Ka2	Kalambo	wild	river	19	0 / 0 / -	0 / 0 / -	0 / 0 / -	0 / 0 / -
KaL pond	Kalambo	pond	lake	12	100 / 41.83 / 42.5	16.67 / 0.25 / 1.5	NA	NA
Ka2 pond	Kalambo	pond	river	14	100 / 17.43 / 12.5	57.14 / 1.71 / 2.5	NA	NA
ChL	Chitili	wild	lake	11	100 / 43 / 33	72.73 / 3.18 / 4	36.36 / 0.45 / 1	0 / 0 / -
Ch1	Chitili	wild	river	12	50 / 1.25 / 2	8.33 / 0.08 / 1	0 / 0 / -	0 / 0 / -

The number of parasite taxa depended on population ($\chi^2_{d.f.=6} = 80.68$, $p < 0.001$, supplementary Table S3, Figure 2a). Lunzua lake fish had higher taxon richness compared to river fish from both the Lunzua and the Kalambo River (post hoc test, all $p < 0.05$). Fish from the upstream Kalambo population (Ka2) had lower taxon richness than any lake population (post hoc test, all $p < 0.05$). There was no significant difference in the number of parasite taxa between geographically closer

lake and river populations from the same system with very low genetic divergence (median F_{ST} = 0, Table S2; Egger et al., 2017) (post hoc test: KaL-Ka1, p = 0.131; ChL-Ch1, p = 0.266).

The total parasite load depended on population ($\chi^2_{d.f.=6}$ = 147.03, p < 0.001) and sex of the individual ($\chi^2_{d.f.=2}$ = 6.55, p = 0.038) (Supplementary Table S4, Figure 2b). Lunzua lake fish had a higher parasite load than any river fish (post hoc test, all p < 0.05), and Kalambo lake fish had a higher parasite load than Kalambo river fish (post hoc test: KaL-Ka1, p = 0.049; KaL-Ka2, p = 0.03). The difference in total parasite load between Chitili lake and river fish was not significant (post hoc test: ChL-Ch1, p = 0.075). Males had higher parasite load than females (post hoc test: f-m, p = 0.034).

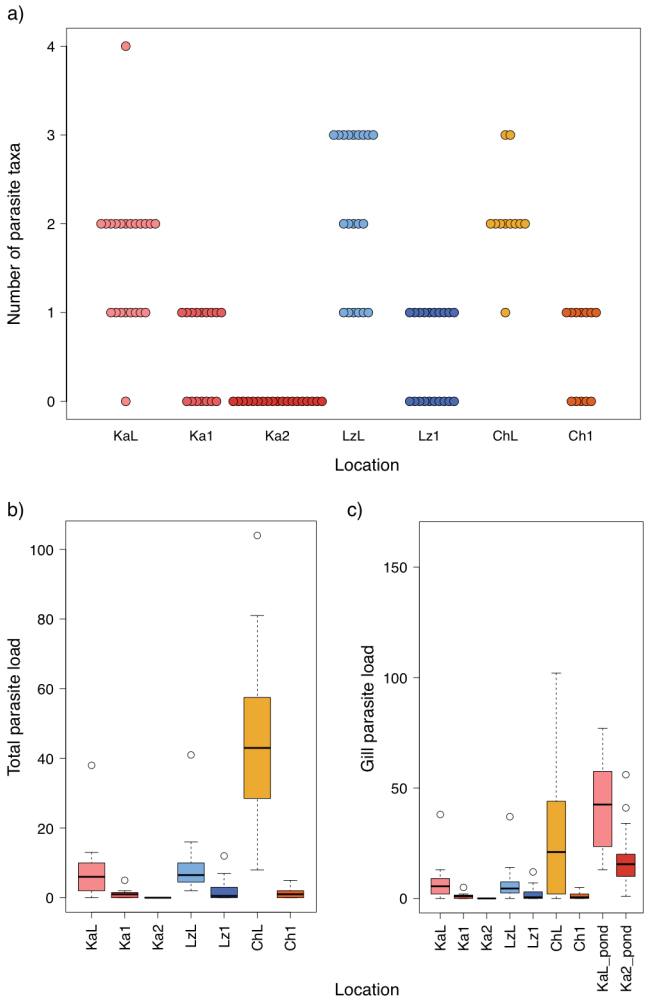


Figure 2 Parasite diversity and abundance per population. (a) Number of parasite taxa in the wild. (b) Total parasite load in the wild. (c) Gill parasite load including pond setup. Colour code according to Figure 1. KaL – Kalambo lake, Ka1 – Kalambo river, Ka2 – Kalambo river upstream, ChL – Chitili lake, Ch1 – Chitili river, LzL – Lunzua lake, Lz1 – Lunzua river.

Gill parasite load in the wild (model 1) depended on population ($\chi^2_{d.f.=6}$ = 1065.50, p < 0.001), fish size ($\chi^2_{d.f.=1}$ = 15.09, p < 0.001), and sex ($\chi^2_{d.f.=2}$ = 9.95, p = 0.007) (model 1; supplementary Table S5, Figure 2c). Lake fish had a significantly higher gill parasite load than river fish (post hoc test: all p < 0.001), except for LzL-Lz1 (p = 0.266) and KaL-Lz1 (p = 0.088). Males had a higher gill

parasite load than females (post hoc test: f-m, p = 0.004), and larger individuals had a higher gill parasite load in the lake habitat and in the ponds (supplementary Figure S1).

In the ponds *versus* wild comparison, gill parasite load depended on population ($\chi^2_{d.f.=3}$ = 102.01, p < 0.001) (model 2; Table S5, Figure 2c). There was a significant difference in gill parasite load between populations in all pairwise comparisons (post hoc test: KaL-Ka2, KaL_pond-Ka2_pond, KaL_pond-KaL, KaL_pond-Ka2, Ka2_pond-Ka2; all p < 0.05) except between KaL in the wild and Ka2 in the pond setup (post hoc test: KaL -Ka2_pond, p = 0.115).

The analysis of dissimilarity in the parasite community revealed that qualitative differences in parasite community composition were significantly associated with genetic differentiation (F_{ST} corrected for geographic distances) (r = 0.64, p = 0.048), but not with the extent of differences in pH (r = 0.31, p = 0.443).

Immune response reflects differences in parasite abundance

The splenosomatic index was analysed in all wild populations (model 1) and separately in the Kalambo lake-river pair (KaL-Ka2) (model 2) to compare these populations in the wild and in the experimental setup. SI depended on population in both models (model 1: $F_{4,83}$= 16.38, p < 0.001; model 2: $F_{3,59}$= 7.00, p < 0.001, supplementary Table S6, Figure 3a). In the populations from the wild, the SI also depended on total parasite load (model 1: $F_{1,83}$= 8.30, p = 0.005, Table S6) with fish having a higher total parasite load also having higher SI (Figure 3b). In the wild (model 1), lake fish had a higher SI than river fish (post hoc test, all pairwise p < 0.05, supplementary Table S6). When fish from the pond setup were compared with their respective source population (model 2), the Ka2 river population had a higher SI in the ponds than in the wild (post hoc test, p = 0.014, supplementary Table S6, Figure 3a).

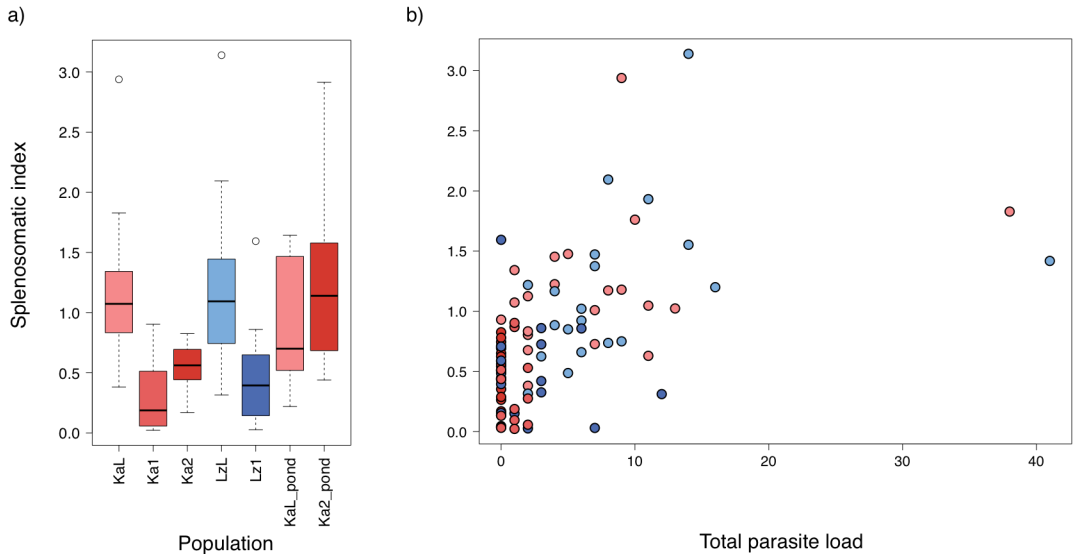


Figure 3 Innate immune response (splenosomatic index) depending on (a) population and on (b) total parasite load in the wild populations. Colour code according to Figure 1. KaL – Kalambo lake, Ka1 – Kalambo river, Ka2 – Kalambo river upstream, LzL – Lunzua lake, Lz1 – Lunzua river.

Global patterns of expression divergence among organs and habitats

A principal component analysis (PCA) separated the two organs along the first principal component (PC1), accounting for 92% of the variance observed in the data set (Figure 4a). There was a strikingly higher variation among individuals in the gene expression of the spleen compared

to the gills. A PCA of gill samples only (Figure 4b) revealed a clear separation of river and lake environment along PC1 (PC1 explained 20% of the variation): The upstream river population (Ka2) clustered separately, followed by other river populations (Ka1 and Lz1), the lake populations (Kal and LzL) and the pond samples (Ka2_pond and KaL_pond). PC2 (9% of variation) separated the upstream Kalambo River population (both from the wild (Ka2) and from the pond (Ka2_pond)) from all the other populations, and the Kalambo system from the Lunzua system.

A PCA of spleen samples only showed no clear clustering by population (Figure 4c). Pond individuals clustered together with the wild-caught lake individuals with high total parasite load along PC2 that explained 15% of the variance (Figure 4d).

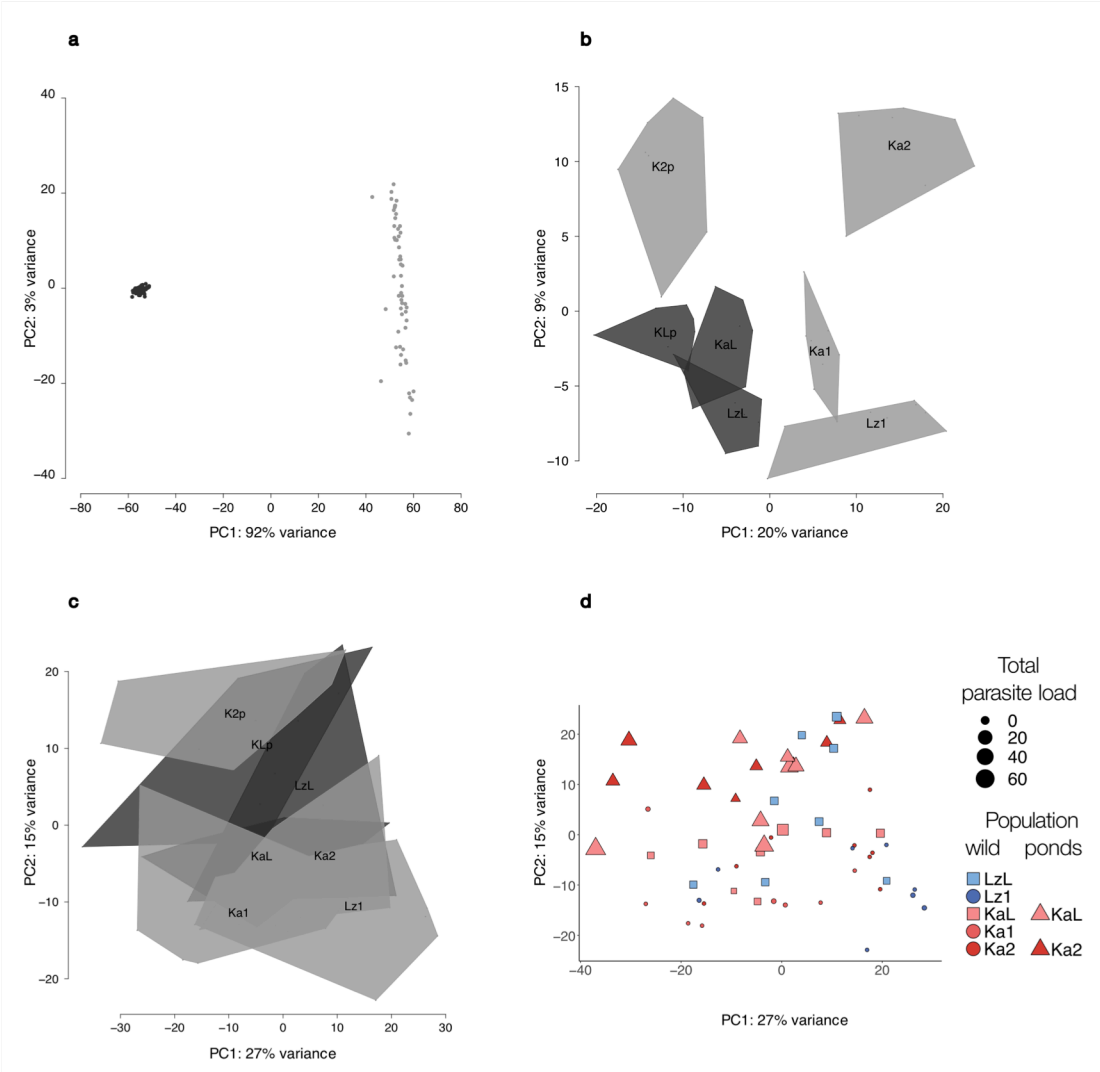


Figure 4 Global patterns of gene expression differences among samples. (a) PCA plot of overall gene expression levels. Samples are coloured according to organs (dark grey: gills, light grey: spleen), proportion of the variance explained by the principal components is indicated next to the axes. (b) PCA plot of gene expression levels in gill samples. (c) PCA plot of gene expression levels in spleen samples. Replicate samples are connected with polygons, polygons are coloured according to habitats (dark grey: lake, light grey: river) and the name of the population is placed in the middle of the respective polygon. KaL – Kalambo lake, Ka1 – Kalambo river, Ka2 – Kalambo river upstream, LzL – Lunzua lake, Lz1 – Lunzua river, KLp – Kalambo lake population in ponds, K2p – Kalambo river upstream population in ponds. (d) PCA plot of gene expression levels in spleen samples. Colours represent populations; squares – lake populations, circles – river populations, triangles – ponds, with symbol size representing total parasite load per specimen.

Rate of expression changes across organs and natural habitats

Overall, in concordance with the PCA, spleen showed a higher rate of expression changes compared to gills (Figure 5a). We further noticed a higher rate of expression changes in river samples compared to lake samples consistently in both organs. The difference in branch lengths suggests that gills and spleen transcriptomes may have experienced different selection regimes throughout the colonization of new habitats (Figure 5b, c). We observed the same difference between the two organs for the two types of habitat. Yet, overall, there was a greater expression divergence in the river samples compared to the lake samples. This suggests once more that the lake samples may have experienced stronger purifying selection or less positive selection than the river samples.

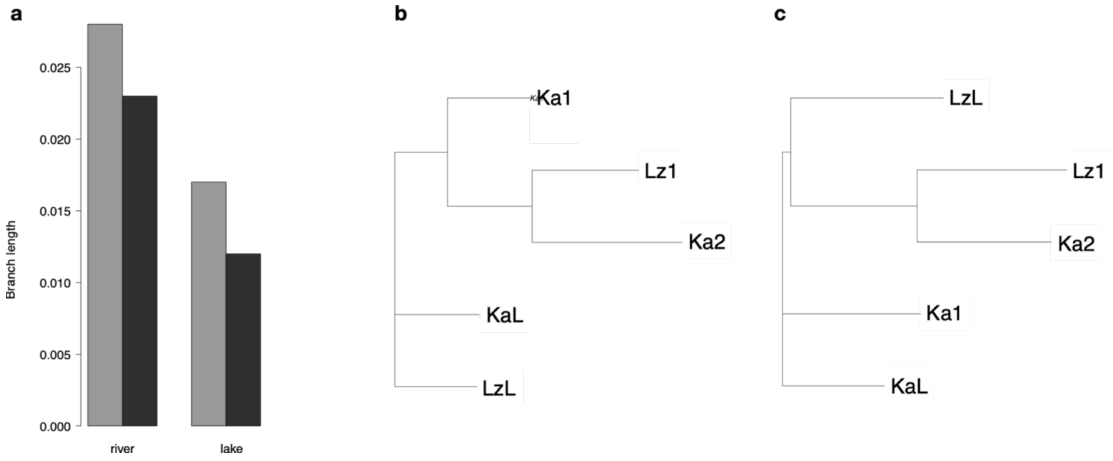


Figure 5 Expression divergence among organs and habitat. (a) Barplots of the branch lengths of protein-coding expression trees of river and lake samples. The barplots are coloured according to the organ (light grey: spleen, dark grey: gills). (b) Protein-coding gene expression trees in the gills. KaL – Kalambo lake, Ka1 – Kalambo river, Ka2 – Kalambo river upstream, LzL – Lunzua lake, Lz1 – Lunzua river, KLp – Kalambo lake population in ponds, K2p – Kalambo river upstream population in ponds. (c) Protein-coding gene expression trees in the spleen.

Differences in gene expression between lake and river samples

A total of 5,082 genes showed significant differential expression (DE; up-regulation and down-regulation) between lake and river habitats (Figure 6a, supplementary Table S7) in the gills and 3,023 in the spleen. From these differentially expressed genes, 1,880 were exclusively down-regulated in the gills, 2167 up-regulated in the gills, 902 down-regulated in the spleen and 1086 up-regulated in the spleen (Figure 6b).

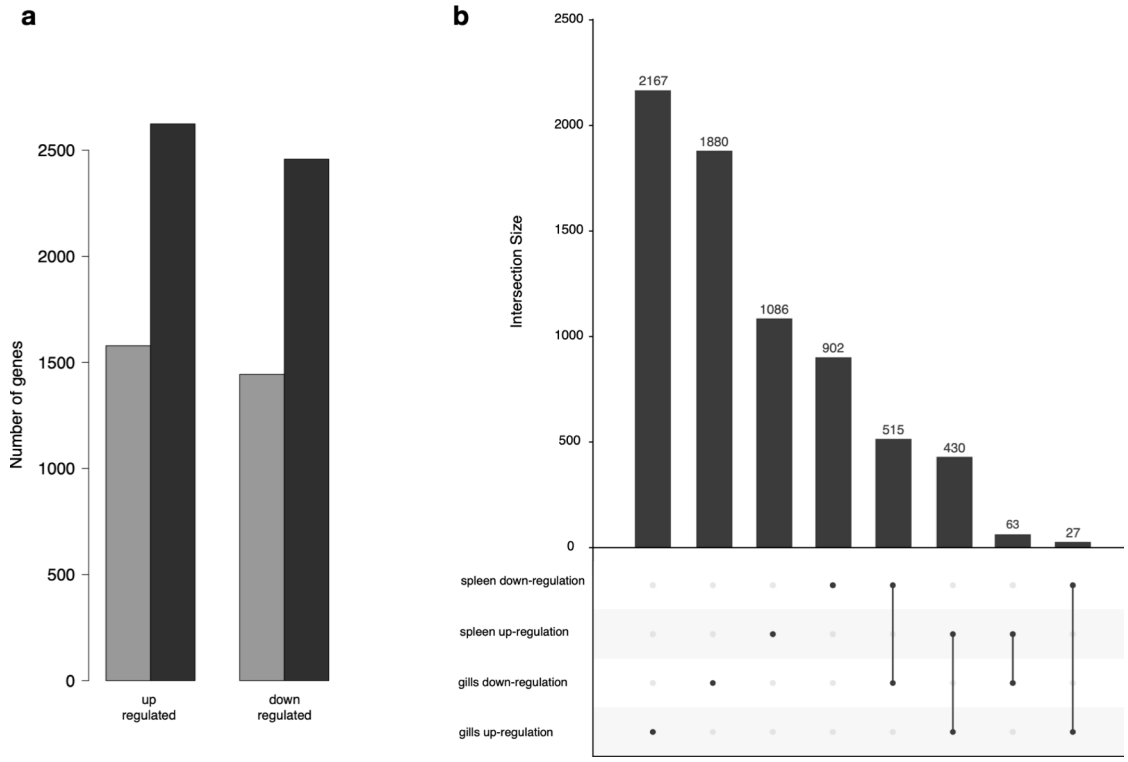


Figure 6 Lake-river differential expression. (a) Barplot of the number of genes up-regulated and down-regulated in the lake - river contrast (Kalmabo and Lunzua population) in the two different organs (dark grey: gills, light grey: spleen). (b) Intersection plot between the up-regulated (up) and down-regulated (down) genes in the two organs.

Gene-expression networks

To further connect gene expression signatures to innate immune response and parasite infection, we constructed signed weighted gene co-expression networks from gill and spleen expression profiles and correlated the obtained networks to various traits: population, SI, river-lake system (Kalambo, Lunzua) and habitat (lake, river) (Figure 7). The gill gene-expression network (Figure 7a) consisted of 28 modules comprising between 27-1692 genes (mean module size = 265, gene module memberships are reported in supplementary Table S8a). Here, the highest correlations were obtained for the greenyellow (223 genes, r=0.86) and the magenta (262 genes, r=-0.86) modules, both with habitat type (lake and river). A gene ontology enrichment analysis (supplementary Figure S2a-b) identified GO terms related to osmoregulation to be enriched in these modules (enrichment for GO terms in greenyellow module (supplementary Figure S2a): e.g. "regulation of pH", "sodium ion transmembrane transport", "sodium:proton antiporter activity", "voltage-gated chloride channel activity"; in magenta module (supplementary Figure S2b): e.g "intracellular signal transduction", "hydrolase activity", "nitrogen utilization").

The spleen network (Figure 7b) consisted of 31 modules (22-3326 genes, mean module size = 365, gene module memberships are reported in supplementary Table S8b). The strongest negative correlation was observed for the red module with SI (463 genes, r=-0.69) and the darkturquoise module with habitat type (64 genes, r=-0.69), which also positively correlated with SI (r=0.65). The orange module correlated with habitat (51 genes, r=0.64). A GO analysis (supplementary Figure S2c-d) clearly linked the red and darkturqoise module to immune response (enrichment for GO terms in red module (supplementary Figure S2c): e.g. "autophagy"," stress-activated MAPK cascade", "extracellular exosome" all of which belong to the broader GO category "defence mechanism"; in darkturquoise module (supplementary Figure S2d): e.g. "antigen processing and

presentation of exogenous protein antigen via MHC class", "positive regulation of antigen processing and presentation of peptide antigen via MHC class").

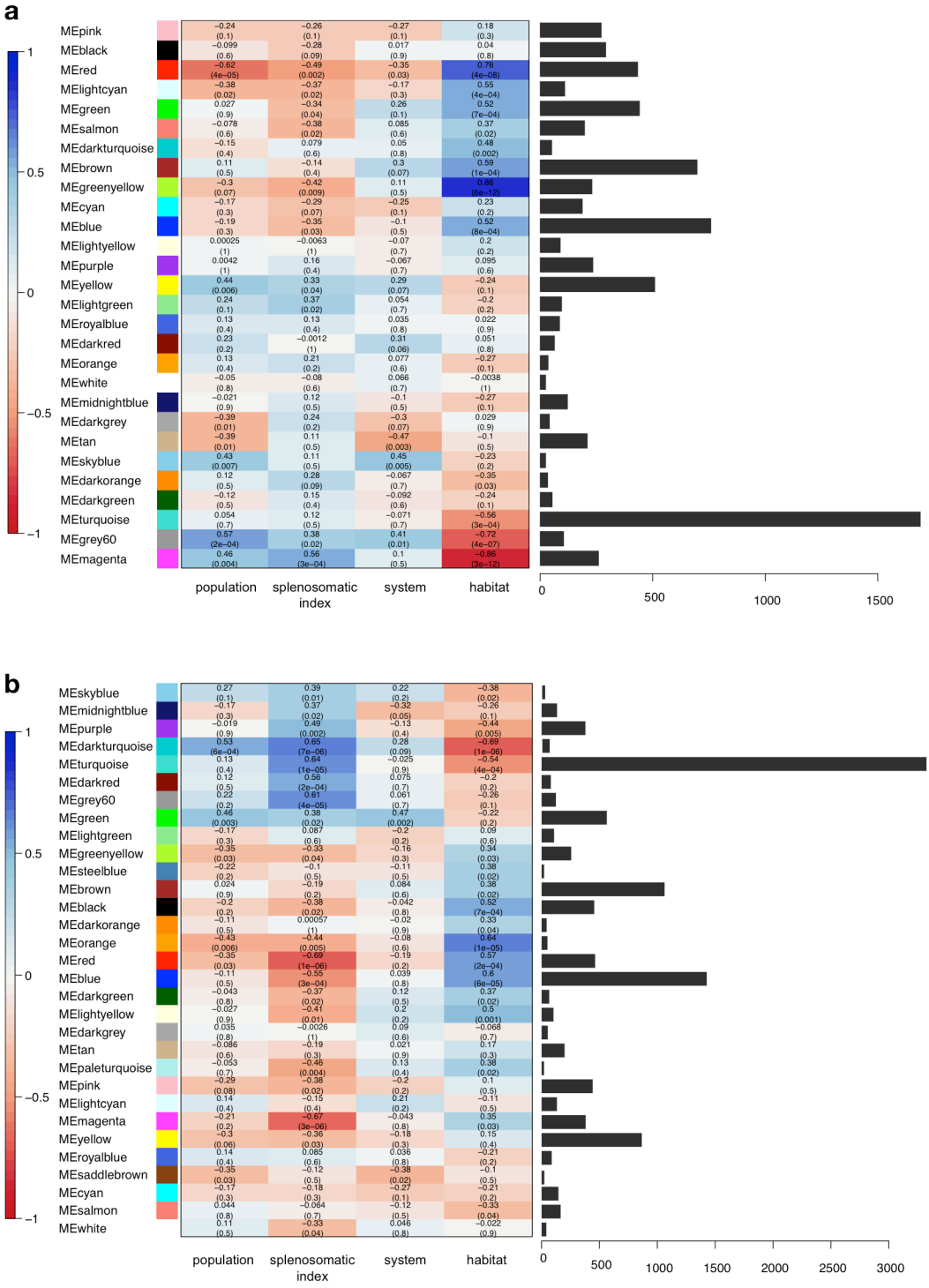


Figure 7 Module-trait correlation. (a) gills (b) spleen.

Habitat-specific genetic variation

We detected 73 significant SNPs (FDR < 0.05) between lake and river habitats (supplementary Table S9). From these, 68 are located in protein-coding genes, two in one lncRNA (LOC109196944) and three in uncharacterized genes. Among the annotated candidate genes, several genes have functions involved in ion transport and homeostasis (sodium/potassium-transporting ATPase subunit alpha-1, sodium/potassium-transporting ATPase subunit alpha-3) and immunity (basigin, ubiquitin carboxyl-terminal hydrolase CYLD, interferon regulatory factor 9, unconventional myosin-Ig, H-2 class I histocompatibility antigen Q9 alpha chain). Sodium/potassium-transporting ATPase subunit alpha-1 (Atp1a1) is a physiological key gene in the adaptation to different osmotic environments in many fish species (e.g. lake whitefish, stickleback, killifish (Scott, 2004), bull shark (Reilly, Cramp, Wilson, Campbell, & Franklin, 2011), brown trout (Larsen et al., 2008). In our samples, it was up-regulated in all river populations in gills and in Lz1 also in spleen. Ubiquitin carboxyl-terminal hydrolase CYLD was up-regulated in lake populations (in comparisons KaL-Ka2 and LzL-Lz1) in spleen. Basigin was up-regulated in river populations (in comparisons KaL-Ka2 and LzL-Lz1) in spleen and gills. Interferon regulatory factor 9 was up-regulated in lake populations (KaL-Ka2 and LzL-Lz1) in spleen. Another gene – cofilin, which was up-regulated in river populations (KaL-Ka2 and LzL-Lz1) in spleen – has previously been found to be involved in changes related to plasticity in fish (Dayan, Crawford, & Oleksiak, 2015; Debes, Normandeau, Fraser, Bernatchez, & Hutchings, 2012). Among the GO terms identified for those candidate genes was “immune system process” (Supplementary Figure S3) illustrating this as an unsupervised method. In addition, in the gills, we identified two candidates (NCBI gene ids: 100691696, and 100699477) included in the highly correlated “greenyellow” module and one candidate (100700200) in the highly correlated magenta module. In the spleen, we identified three candidates (NCBI gene ids: 100694281, 100696602, and 100708925) included in the highly correlated “red” module and one candidate (100701947) in the highly correlated darkturquoise module.

Expression plasticity and genetic assimilation

We identified genes that were differentially expressed (lake-river contrast) in wild and pond samples, for each of the two organs. In agreement with the analysis above, we observed more DE genes in gills (in the wild and in the pond environment) than in spleen (Figure 8a). Generally, more genes were DE between the samples from the wild than the pond ones, which suggests gene expression reprogramming in response to environmental changes (Figure 8a). The majority of these genes was differentially expressed solely in the samples from the wild and only a few were DE in both environments (genetic assimilation) (Figure 8b). We report all genes assigned to either genetic assimilation (Supplementary Table S10) or plastic response (Supplementary Table S11).

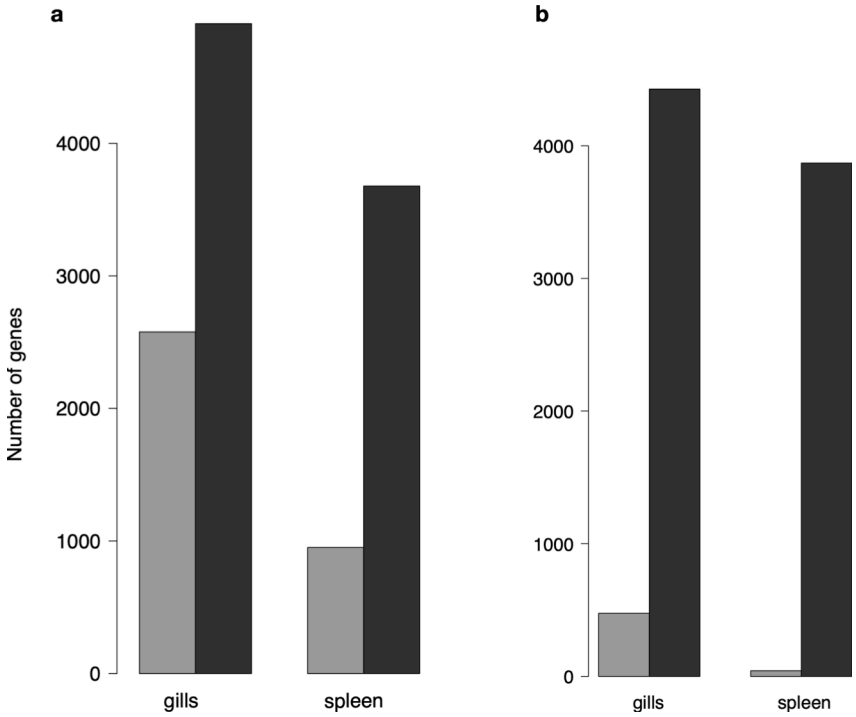


Figure 8 Lake-river differential expression in pond and wild populations. (a) Barplot of the number of genes differentially expressed in lake and river populations sampled in the wild (dark grey) or in the ponds (light grey). Differential expression was assessed for gills and spleen separately. (b) Barplot of the number of genes with plastic expression variation (dark grey) or genetically assimilated (fixed) expression (light grey) for gills and spleen samples.

As a control, we performed a PCA plot of pond and wild samples for the two different categories of genes (genetic assimilation and plastic response) in the two different organs (Figure 9). As expected, in the PCA plots of the genes that show plastic response (Figure 9a and c), PC1 clearly separates samples of the river environment (Ka2) from samples from a lake-like environment (KaL, KLp and K2p) in both organs. The PCA of genes that are likely subject of genetic assimilation (Figure 9b and d), on the other hand, rather separates samples according to source populations along PC1 (phylogenetic signal). This separation was more pronounced in spleen than in gill samples albeit lower numbers of DE genes (Figure 9).

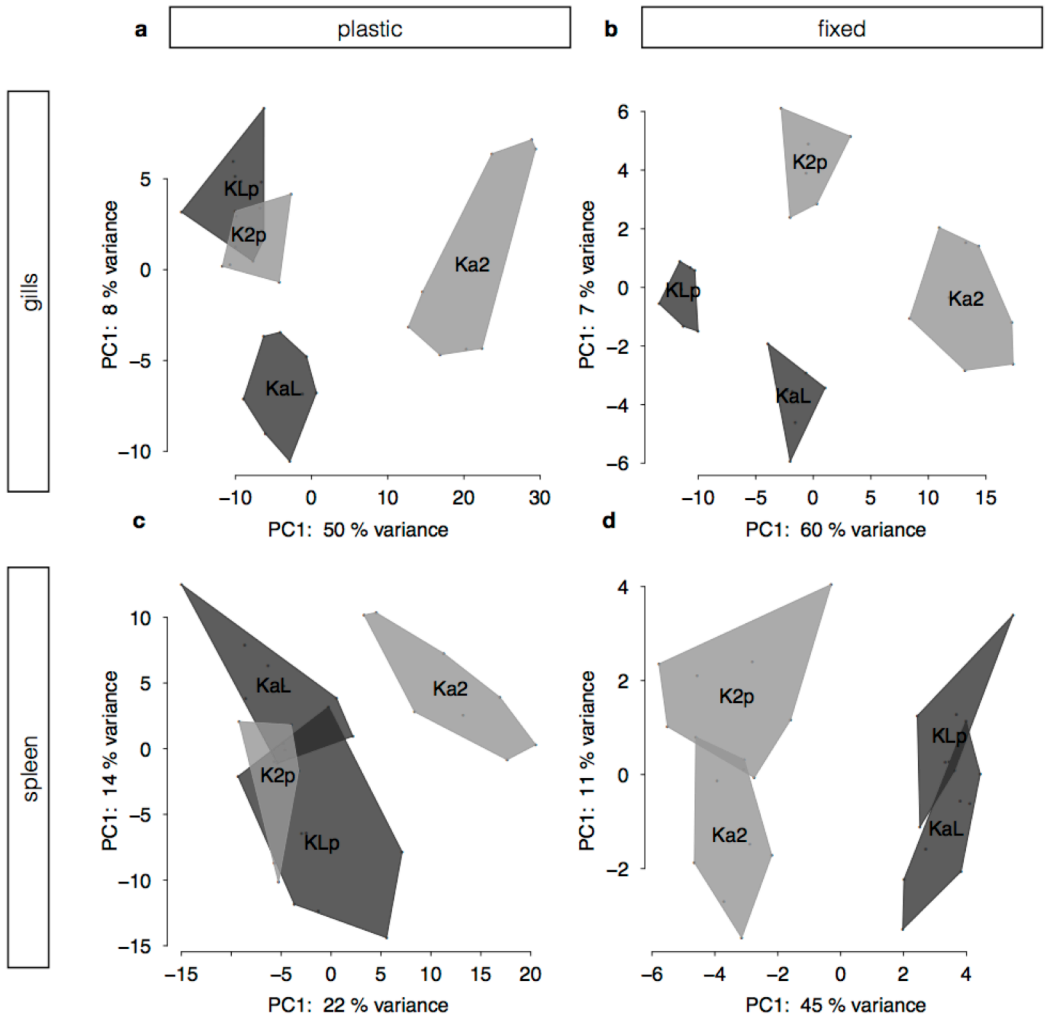


Figure 9 Gene expression variation for different parts of the transcriptome. PCA plots of gills (a, b) and spleen (c, d) samples. The PCA plots were produced using the candidate genes for plastic response to environment changes (a, c) and the genes for which gene expression variation was genetically fixed (b, d). Replicate samples are connected via polygons, polygons are coloured according to habitat (dark grey: lake, light grey: river) and the name of the population is placed in the middle of the respective polygon. KaL – Kalambo lake, Ka1 – Kalambo river, Ka2 – Kalambo river upstream, LzL – Lunzua lake, Lz1 – Lunzua river, KLp – Kalambo lake population in ponds, K2p – Kalambo river upstream population in ponds. The proportion of the variance explained by the principal components is indicated next to the axes.

Microbiota

The microbial community of the buccal and intestinal mucosa over all samples consisted of 77,646 OTUs identified with a 97% identity threshold belonging to 1,000 genera in 35 phyla (22,489 OTUs comprising 903 genera remained after filtering out singletons). The most abundant phyla in the gut mucosa were Fusobacteria (25%), Proteobacteria (16%), and Firmicutes (8%); and in the buccal mucosa were Proteobacteria (33%), Bacteroidetes (11%) and Fusobacteria (4%) (relative abundance percentages given after filtering out low abundance sequences (≤ 0.02), supplementary Figure S4).

The microbiota composition (β-diversity) of all sampled OTUs differed between the habitats (nested PERMANOVA habitat: buccal mucosa $F_{2,53}$= 13.49, $p < 0.001$/intestinal mucosa $F_{2,55}$ = 9.84, $p < 0.001$). Sex had no effect on the microbial community (nested PERMANOVA sex: buccal

mucosa $F_{1,53}$ = 1.15, p = 0.267/intestinal mucosa $F_{1,55}$ = 1.51, p = 0.135). Microbial community richness (α-diversity, inverse Simpson index) was higher in the mouth than in the gut samples, with no differences between sexes. NMDS analysis of bacterial communities based on Bray-Curtis distances showed clustering by tissue and by habitat, as well as a shift in the river communities towards lake-like in the pond environment (Figure 10) paralleling the gene expression patterns.

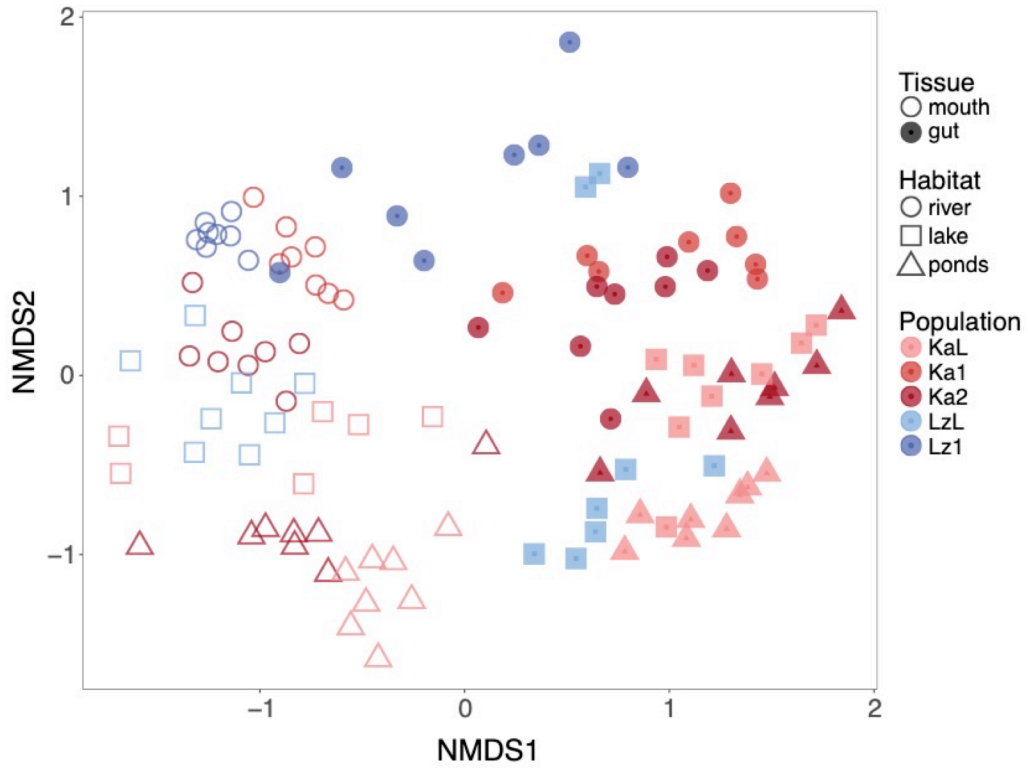


Figure 10 Non-metric Multidimensional Scaling (NMDS) analysis of bacterial communities based on Bray-Curtis distances. Colours represent populations; circles- river habitat, squares –lake habitat, triangles –ponds; filled symbols – intestinal samples, hollow symbols – buccal samples. KaL – Kalambo lake, Ka1 – Kalambo river, Ka2 – Kalambo river upstream, LzL – Lunzua lake, Lz1 – Lunzua river. Stress = 0.171.

Discussion

Higher parasite diversity and abundance in the lake than in rivers

Parasite communities of riverine *A. burtoni* populations were less diverse than those of their adjacent lake populations, which is consistent with findings from stickleback lake–stream population comparisons (Eizaguirre, Lenz, Kalbe, & Milinski, 2012; Eizaguirre et al., 2011; Feulner et al., 2015; Stutz & Bolnick, 2017). Parasitic helminths (nematodes and acantocephalans), as well as parasitic copepods, were only found in the lake habitat. The only parasite group found in both habitats – monogeneans – were more abundant in the KaL lake population than in the river populations. There are several possible explanations for these patterns. Complex life cycles with invertebrate intermediate hosts and mainly vertebrate final hosts are common in nematodes and acanthocephalans (Anderson, 1988; Benesh, Chubb, & Parker, 2014; Chubb, Ball, & Parker, 2010). Lake Tanganyika is home to several adaptive radiations including cichlids, Mastacembelid eels, Synodontis catfish, and crabs (Brown, Rüber, Bills, & Day, 2010; Day, Bills, & Friel, 2009;

Marijnissen et al., 2006), and is thus expected to harbour a higher diversity and abundance of potential hosts and, consequently, also parasites than the affluent rivers. In agreement with this, estimated effective population sizes in *A. burtoni* are much higher for lake than for river populations (Egger et al., 2017). The rate of parasite infection is expected to increase with host population size (Anderson & May, 1979; McCallum, Barlow, & Hone, 2002). Furthermore, a large host population can allow for the persistence of parasite species with low reproductive rates that are unable to persist in smaller populations (Dobson & Carper, 1996; Lindstrom, Foufopoulos, Parn, & Wikelski, 2004).

At four of our study sites (KaL, Ka2, ChL, Ch1) gills of adult individuals caught in November 2015 (wet season) were screened for parasites as a part of a pilot study by experienced parasitologists who found the same pattern: no parasites in Ka2 and 75% prevalence in KaL (Rahmouni et al., in preparation). In the lake-river pair from the Chitili system that shows the lowest genetic divergence (F_{ST} ChL-Ch1= 0.00478), they found 50 *versus* 37% prevalence in lake *versus* river habitat that are separated by a mere 300m (Table S2). This shows that the pattern is consistent between different years and different seasons, opening opportunities for consistent parasite-driven divergent selection and fulfilling one more condition for parasite-driven speciation (Karvonen & Seehausen, 2012).

Differences in parasite taxa composition were positively correlated with the extent of genetic differentiation corrected for the geographic distance. Differentiation in infections was present even in the pairs with the lowest genetic divergence (KaL-Ka1, ChL-Ch1, both $F_{ST} < 0.006$). This is consistent with the idea that differentiation in infections is present as soon as two populations occupy different environments, proceeding genetic differentiation.

Immune response reflects parasite abundance in lake and river

In the wild, lake populations had higher levels of innate immune response estimated by SI. When exposed to lake parasites in the pond setup, the river population showed elevated innate immune response, reflected by both spleen size and spleen gene expression pattern resembling lake samples rather than river samples.

Even though lake fish in the pond experiment had higher gill parasite abundance than river fish, their SI was lower. This suggests that lake fish are well adapted to parasites that they normally encounter in the wild, unlike river fish, whose immune response was elevated in comparison to the wild. Models predict that parasite-host coevolution can facilitate speciation in host populations when they can adapt to the parasite community that infects them (reviewed in Greischar & Koskella, 2007; Kaltz & Shykoff, 1998; Kawecki & Ebert, 2004; Summers et al., 2003). In this case, gene flow from non-adapted host populations could be maladaptive, and assortative mating between host populations may evolve (Karvonen & Seehausen, 2012).

Unlike the innate immune response estimated here, the proportion of cells of the adaptive immune system, the lymphocyte ratio, was previously found to be higher in river than in lake populations of *A. burtoni* (Figure S5; Theis et al., 2017). However, as already noted by Theis et al. (2017), a shift in the lymphocyte ratio could also imply that there are more monocytes present, which represent the first line of the immune defence. This would mean that lake fish had a higher monocyte to lymphocyte ratio than river fish, in turn supporting our result that lake fish have higher innate immune response than river fish. Alternatively, a higher adaptive immune response in river compared to lake fish could also reflect investments into alternative immune strategies. The benefit of using different arms of the immune system varies with the abundance of parasites in the environment (Lindstrom et al., 2004; McDade, Georgiev, & Kuzawa, 2016). The innate immune system provides the primary defence system against pathogen invasion, but it is energetically costly (Kraaijeveld & Godfrey, 1997; Lochmiller & Deerenberg, 2000; Moret & Schmid-Hempel,

2000; Sheldon & Verhulst, 1996). An induced response that is only deployed after an invader has been recognized could have evolved to avoid a costly permanent defence capability at the price of a delayed response and the potential risk that the parasite escapes host control (Schmid-Hempel & Ebert, 2003; Shudo & Iwasa, 2001). In lake-river stickleback, the cost of mounting an immune response induces associated fitness costs and might lead to selection for river fish with a relatively low level of innate immune response (Kaufmann et al., 2017), consistent with our findings. Results of the same study (Kaufmann et al., 2017) suggest that varying parasite communities can lead to population-specific immune responses that contribute to varying host-parasite co-evolutionary trajectories, further corroborating that, even though innate immune responses are thought to be rather unspecific, they can contribute to local adaptation (Tschirren et al., 2013).

Comparative gene expression in spleen and gill tissue

Studying gene expression in wild animals permits a view on differential expression responses caused by both genetic and environmental factors (Huang et al., 2016). While the gill expression patterns showed clear clustering by habitat and population (Figure 4b), spleen expression patterns showed clustering by infection status and parasite load (Figure 4c). This pattern was further reflected by higher correlations of the SI with gene expression modules from the spleen network than the gill network and higher correlations of habitat type with gill than spleen modules. Fish gills are a multifunctional organ involved in gas exchange, ion regulation, osmoregulation, acid-base balance, ammonia excretion, hormone production, modification of circulating metabolites and immune defence (Evans, 2005; Secombes & Wang, 2012). Gills are constantly in direct contact with water and are likely to show the greatest transcriptomic response to environmental differences such as water chemistry between different aquatic habitats (Gibbons, Metzger, Healy, & Schulte, 2017; Hughes et al., 2017; Kavembe, Franchini, Irisarri, Machado-Schiaffino, & Meyer, 2015; Lam et al., 2014; Xu et al., 2013). Gill transcriptome remodelling in fish after exposure to elevated salinity has been observed already after one week (Jeffries et al., 2019). Here we find that the gene expression pattern in gills along PC1 (Figure 4b) is associated with different environmental conditions in lake and river habitats. We also found that gill gene expression modules were most strongly correlated with habitat type and grouped genes involved in osmoregulation. Shift of gill gene expression of the Ka2 population along the PC1 axis in the pond setup supports the important role of adaptive phenotypic plasticity in adaptation to different environments. It has previously been shown in a transplant experiment performed in lake habitat, that Ka2 individuals raised in a common garden perform equally well as lake fish, unlike wild-caught Ka2 individuals (Rajkov et al., 2018a). On the other hand, the separation of the upstream Kalambo river population (both from the wild (Ka2) and from the pond (Ka2_pond)) from all the other populations along the PC2 axis, and the separation of the Kalambo from the Lunzua system, supports strong phylogenetic signal (Egger et al., 2017) in gene expression.

As an immune response-generating and pathogen neutralising organ (Press & Evensen, 1999), the spleen is expected to reflect immune system reaction in its transcriptomic response (Huang et al., 2016). Indeed, here spleen expression patterns seemed to be provoked by stress related to parasite exposure: (i) pond individuals with high parasite load clustered together with lake individuals from the wild with high parasite load and (ii) we identified gene expression modules with a function in the innate immune response as being correlated with SI. The number of genes differentially expressed in the river-lake contrast is lower in the spleen than in the gills (Figure 6a and 8). This result was expected, as the general expression pattern observed in the spleen does not show a clear separation between lake and river samples Figure 4c). However, when looking at the genes that showed a plastic response in both organs, the separation of spleen samples is more pronounced than the one of gill samples (Figure 9), which might suggest that some organs

are more plastic than others (Tang, Smith-Caldas, Driscoll, Salhadar, & Shingleton, 2011). As expected, when inspecting the PCA plots of genes that show signatures of adaptive plasticity, PC1 (Figure 9a and c) clearly separates samples from the river environment (Ka2) from samples from a lake-like environment (KaL, KLp and Kp2) in both organs. Phenotypic plasticity in gene expression has been found to contribute to divergence of locally adapted fish populations (Dayan et al., 2015), where phenotypic plasticity and adaptation operate on different suites of genes for the majority of significant differences in gene expression levels.

Nevertheless, previous studies (Egger et al., 2017; Weber et al., in preparation) have identified fixed variation as one important component for long-term adaptation to divergent environment. Some of those fixed variations have been linked to fixed expression divergence (assimilation) while others can be associated with variation in the genomic sequences (e.g. SNPs). Several interesting habitat-specific genetic variations were identified in the context of our study. Among those candidates is the sodium/potassium-transporting ATPase subunit alpha-1 (Atp1a1). Atp1a1 is involved in the maintenance of the ion balance and electrolyte homeostasis in different osmoregulatory epithelia (Evans, 2005). This gene is one of the main candidates for marine to freshwater adaptation in different fish species (DeFaveri, Shikano, Shimada, Goto, & Merilä, 2011; Jones et al., 2012; McCairns & Bernatchez, 2009), and shows clear haplotype clustering by habitat in sticklebacks (Roesti, Gavrilets, Hendry, Salzburger, & Berner, 2014). Atp1a1 and Atp1a3 were also found to be up-regulated in freshwater euryhaline Mozambique tilapia gills in comparison to seawater (Lam et al., 2014). The difference in conductivity between freshwater and marine stickleback habitat is on a similar order of magnitude (e.g. typical conductivity of seawater – 55 000 µS/cm, freshwater 100-200 µS/cm) as the difference in conductivity between river and lake habitat of *A. burtoni* investigated here (Lake Tanganyika – 660-670 µS/cm, river 18-60 µS/cm, Table S9 in (Rajkov et al., 2018b)). Ubiquitin carboxyl-terminal hydrolase CYLD is one of the four genes identified as parallel F_{ST} outliers in lake whitefish (Renaut, Nolte, Rogers, Derome, & Bernatchez, 2011) and is methylated in response to salinity stress in Pacific oyster (Zhang, Li, Kong, & Yu, 2017). Basigin has a role in various physiologic and pathologic contexts (e.g. immune response in tilapia spleen following Streptococcus challenge) (Zhu et al., 2017) and is a receptor essential for erythrocyte invasion by *Plasmodium falciparum* parasite in humans (Crosnier et al., 2011). Interferon regulatory factor 9 has an immune function. The overlap of candidate genes across analyses in our study suggests that genetic and expression changes affect similar biological functions.

Microbiota

The capacity of the gut microorganism community to change its composition or gene-expression pattern in response to the host's physiological changes and variations of the external environment – that is, metagenomic plasticity – is likely an essential factor in host acclimation and adaptation to environmental change (Alberdi et al., 2016). Fish are known to experience complete turnover of their microbiomes during transitions between different aquatic environments (Lokesh & Kiron, 2016; Schmidt, Smith, Melvin, & Amaral-Zettler, 2015) that are expected to cause drastic challenges to their immune system.

While an increasing number of studies is investigating fish gut microbiota and their role in diversification (Baldo et al., 2017; Baldo, Riera, Salzburger, Barluenga, & Costa, 2019; Baldo et al., 2015; Franchini, Fruciano, Frickey, Jones, & Meyer, 2014; Härer et al., 2019; Rennison, Rudman, & Schluter, 2019; M. Sevellec et al., 2014; Maelle Sevellec, Derome, & Bernatchez, 2018; Sullam et al., 2015), the present study is among the first that compared microbiota communities of two different tissues, one in direct contact with the (novel) habitat (mouth), and another that is known to

be affected by food type (gut) (Sullam et al., 2012), and characterized how these communities change upon exposure to a novel habitat in a semi-natural setup.

In a previous study that compared gut microbiota of different cichlid species (Baldo et al., 2015), wild *A. burtoni* carried the most diverse microbiota of all investigated species, being significantly distinct from all other species and from samples of the same species kept in the laboratory, with laboratory individuals displaying a highly reduced microbiota diversity compared to the wild population. While another study found changes in foregut microbiota in lab bread *A. burtoni* females depending on the mouthbrooding stage (Faber-Hammond et al., 2019), we found no differences in mouth or gut microbiota between the sexes.

Host diet and host genotype are the most likely causes of the parallel shifts we observed in microbiome composition across two river systems. Whereas gut microbial communities in our study converged towards lake-like in the ponds with lake water, mouth bacterial communities in the ponds were distinct from the mouth microbial communities of wild-caught fish, indicating that other factors in the pond lake-like setup affect the mouth microbiota communities. Higher microbial community richness observed in the mouth in comparison to the gut indicates that mouth microbiota is, as expected, more strongly influenced by external environment. Overall, our results support the hypothesis that the plasticity of the gut microbiota might be an important factor in phenomic plasticity of vertebrates.

Conclusion

Here we describe, for the first time, parasite and microbiota communities in *A. burtoni* lake-river ecotypes. Lake populations were more heavily parasitized than river populations, both in terms of parasite taxa composition and abundance. Innate immune response in the wild was higher in lake than in river populations. When river population was exposed to lake parasites in a pond lake-like setup the immune response was elevated in comparison to the wild. The present study thus demonstrates the potential for parasite-mediated divergent selection between populations occupying contrasting habitats already at the incipient stages of differentiation. Our RNA-seq data provides evidence that environmental differences between lake and river habitat and their distinct parasite communities shape differential gene expression patterns in *A. burtoni*. By comparing gene expression and bacterial communities between wild-caught individuals and individuals acclimated to lake-like pond conditions, we show that plasticity in gene expression and microbiota composition contribute to previously identified adaptive phenotypic plasticity.

Acknowledgements

We would like to thank our helpers in the field, Alexandra Weber, Adrian Indermaur, Fabrizia Ronco, and Gabriele Pumo, the crew of the Kalambo Lodge for their logistic support in Zambia, and the Lake Tanganyika Research Unit, Dept. of Fisheries, Rep. of Zambia, for research permits. We thank Nikol Kmentova for valuable lessons on cichlid parasitology, Chahrazed Rahmouni for screening gill samples in a pilot study, Christian Owusu for sharing his photos of cichlid parasites, and Maarten Vanhove for discussions on cichlid parasitology. We thank Laura Baldo and Olivia Roth for advice on DNA extraction and library preparation for cichlid gut and mouth microbiota, the support team of sciCORE (center for scientific computing, University of Basel, http://scicore.unibas.ch/) for providing access to computational resources and Alexandra Weber for comments on the manuscript. This study was supported by a grant from the Swiss National Science Foundation (SNSF, grant 31003A_156405) to WS.

Data accessibility

All sequencing data in this paper are available from NCBI under the BioProject accession number PRJNA611922. Data on genetic material contained in this paper are published for non-commercial use only. Utilization by third parties for purposes other than non-commercial scientific research may infringe the conditions under which the genetic resources were originally accessed, and should not be undertaken without obtaining consent from the corresponding author of the paper and obtaining permission from the original provider of the genetic material.

Author contributions

JR and AET designed the study with input from BE and WS. JR and BE performed the sampling and parasite screening. JR, AET and BE performed the wet laboratory work. JR analysed the parasite and the microbiota data. AET analysed the RNA-seq data with the help of AB and drafted corresponding methods and results. JR wrote the manuscript with feedback from all co-authors.

References

Alberdi, A., Aizpurua, O., Bohmann, K., Zepeda-Mendoza, M. L., & Gilbert, M. T. P. (2016). Do vertebrate gut metagenomes confer rapid ecological adaptation? *Trends in Ecology and Evolution*, *31*(9), 689–699. doi:10.1016/j.tree.2016.06.008

Anders, S., Pyl, P. T., & Huber, W. (2015). HTSeq-A Python framework to work with high-throughput sequencing data. *Bioinformatics*, *31*(2), 166–169. doi:10.1093/bioinformatics/btu638

Anderson, R. C. (1988). Nematode transmission patterns. *The Journal of Parasitology*, *74*(1), 30–45.

Anderson, R. M., & May, R. M. (1979). Population biology of infectious diseases: Part I. *Nature*, *280*(5722), 361–367.

Aubin-Horth, N., & Renn, S. C. P. (2009). Genomic reaction norms: Using integrative biology to understand molecular mechanisms of phenotypic plasticity. *Molecular Ecology*, *18*(18), 3763–3780. doi:10.1111/j.1365-294X.2009.04313.x

Baldo, L., Pretus, J. L., Riera, J. L., Musilova, Z., Bitja Nyom, A. R., & Salzburger, W. (2017). Convergence of gut microbiotas in the adaptive radiations of African cichlid fishes. *The ISME Journal*, *11*(9), 1975–1987. doi:10.1038/ismej.2017.62

Baldo, L., Riera, J. L., Salzburger, W., Barluenga, M., & Costa, R. (2019). Phylogeography and Ecological Niche Shape the Cichlid Fish Gut Microbiota in Central American and African Lakes. *Frontiers in Microbiology*, *10*(October), 1–19. doi:10.3389/fmicb.2019.02372

Baldo, L., Riera, J. L., Tooming-Klunderud, A., Albà, M. M., & Salzburger, W. (2015). Gut microbiota dynamics during dietary shift in eastern African cichlid fishes. *PLoS ONE*, *10*(5), 1–23. doi:10.1371/journal.pone.0127462

Benesh, D. P., Chubb, J. C., & Parker, G. A. (2014). The trophic vacuum and the evolution of complex life cycles in trophically transmitted helminths. *Proceedings of the Royal Society B: Biological Sciences*, *281*(1793), 20141462. doi:10.1098/rspb.2014.1462

Bolger, A. M., Lohse, M., & Usadel, B. (2014). Trimmomatic: A flexible trimmer for Illumina sequence data. *Bioinformatics*, *30*(15), 2114–2120. doi:10.1093/bioinformatics/btu170

Bolnick, D. I. (2004). Can intraspecific competition drive disruptive selection? An experimental test in natural populations of sticklebacks. *Evolution; International Journal of Organic Evolution*, *58*(3), 608–18. doi:10.1111/j.0014-3820.2004.tb01683.x

Brawand, D., Wagner, C. E., Li, Y. I., Malinsky, M., Keller, I., Fan, S., … Di Palma, F. (2014). The genomic substrate for adaptive radiation in African cichlid fish. *Nature*, *513*(375–381). doi:10.1038/nature13726

Brawand, David, Soumillon, M., Necsulea, A., Julien, P., Csárdi, G., Harrigan, P., … Kaessmann, H. (2011). The evolution of gene expression levels in mammalian organs. *Nature*, *478*(7369), 343–348. doi:10.1038/nature10532

Brown, K. J., Rüber, L., Bills, R., & Day, J. J. (2010). Mastacembelid eels support Lake Tanganyika as an evolutionary hotspot of diversification. *BMC Evolutionary Biology*, *10*(1). doi:10.1186/1471-2148-10-188

Brucker, R. M., & Bordenstein, S. R. (2012). Speciation by symbiosis. *Trends in Ecology and Evolution*, *27*(8), 443–451. doi:10.1016/j.tree.2012.03.011

Brucker, R. M., & Bordenstein, S. R. (2013). The hologenomic basis of speciation: Gut bacteria cause hybrid lethality in the genus Nasonia. *Science*, *466*(August), 667–669. doi:10.1126/science.1240659

Castro, R., & Tafalla, C. (2015). Overview of fish immunity. In *Mucosal Health in Aquaculture* (pp. 3–54). Elsevier Inc. doi:10.1016/B978-0-12-417186-2.00002-9

Chubb, J. C., Ball, M. A., & Parker, G. A. (2010). Living in intermediate hosts: evolutionary adaptations in larval helminths. *Trends in Parasitology*, *26*(2), 93–102. doi:10.1016/j.pt.2009.11.008

Crosnier, C., Bustamante, L. Y., Bartholdson, S. J., Bei, A. K., Theron, M., Uchikawa, M., … Wright, G. J. (2011). Basigin is a receptor essential for erythrocyte invasion by *Plasmodium falciparum*. *Nature*, *480*(7378), 534–537. doi:10.1038/nature10606

Day, J. J., Bills, R., & Friel, J. P. (2009). Lacustrine radiations in African Synodontis catfish. *Journal of Evolutionary Biology*, *22*(4), 805–817. doi:10.1111/j.1420-9101.2009.01691.x

Dayan, D. I., Crawford, D. L., & Oleksiak, M. F. (2015). Phenotypic plasticity in gene expression contributes to divergence of locally adapted populations of Fundulus heteroclitus. *Molecular Ecology*, *24*(13), 3345–3359. doi:10.1111/mec.13188

Debes, P. V., Normandeau, E., Fraser, D. J., Bernatchez, L., & Hutchings, J. A. (2012). Differences in transcription levels among wild, domesticated, and hybrid Atlantic salmon (*Salmo salar)* from two environments. *Molecular Ecology*, *21*(11), 2574–2587. doi:10.1111/j.1365-294X.2012.05567.x

DeFaveri, J., Shikano, T., Shimada, Y., Goto, A., & Merilä, J. (2011). Global analysis of genes involved in freshwater adaptation in threespine sticklebacks (Gasterosteus aculeatus). *Evolution; International Journal of Organic Evolution*, *65*(6), 1800–1807. doi:10.1111/j.1558-5646.2011.01247.x

Diepeveen, E. T., Roth, O., & Salzburger, W. (2013). Immune-related functions of the Hivep gene family in East African cichlid fishes. *G3*, *3*(December), 2205–17. doi:10.1534/g3.113.008839

Dobin, A., Davis, C. A., Schlesinger, F., Drenkow, J., Zaleski, C., Jha, S., … Gingeras, T. R. (2013). STAR: Ultrafast universal RNA-seq aligner. *Bioinformatics*, *29*(1), 15–21. doi:10.1093/bioinformatics/bts635

Dobson, A. P., & Carper, E. R. (1996). Infectious diseases and human population history. *BioScience*, *46*(2), 115–126. doi:10.2307/1312814

Egger, B., Roesti, M., Böhne, A., Roth, O., & Salzburger, W. (2017). Demography and genome divergence of lake and stream populations of an East African cichlid fish. *Molecular Ecology*, *26*(21), 5016–5030.

Ehrenreich, I. M., & Pfennig, D. W. (2016). Genetic assimilation: A review of its potential proximate causes and evolutionary consequences. *Annals of Botany*, *117*(5), 769–779. doi:10.1093/aob/mcv130

Eizaguirre, C., & Lenz, T. L. (2010). Major histocompatability complex polymorphism: Dynamics and consequences of parasite-mediated local adaptation in fishes. *Journal of Fish Biology*, *77*(9), 2023–2047. doi:10.1111/j.1095-8649.2010.02819.x

Eizaguirre, Christophe, Lenz, T. L., Kalbe, M., & Milinski, M. (2012). Divergent selection on locally adapted major histocompatibility complex immune genes experimentally proven in the field. *Ecology Letters*, *15*(7), 723–731. doi:10.1111/j.1461-0248.2012.01791.x

Eizaguirre, Christophe, Lenz, T. L., Sommerfeld, R. D., Harrod, C., Kalbe, M., & Milinski, M. (2011). Parasite diversity, patterns of MHC II variation and olfactory based mate choice in diverging three-spined stickleback ecotypes. *Evolutionary Ecology*, *25*(3), 605–622. doi:10.1007/s10682-010-9424-z

Evans, D. H. (2005). The multifunctional fish gill: dominant site of gas exchange, osmoregulation, acid-base regulation, and excretion of nitrogenous waste. *Physiological Reviews*, *85*(1), 97–177. doi:10.1152/physrev.00050.2003

Faber-Hammond, J. J., Coyle, K. P., Bacheller, S. K., Roberts, G., Mellies, J. L., Roberts, R. B., & Renn, S. C. P. (2019). The intestinal environment as an evolutionary adaptation to mouthbrooding in the Astatotilapia burtoni cichlid. *FEMS Microbiology Ecology*, *95*(February), 1–14. doi:10.1093/femsec/fiz016

Feulner, P. G. D., Chain, F. J. J., Panchal, M., Huang, Y., Eizaguirre, C., Kalbe, M., … Milinski, M. (2015). Genomics of divergence along a continuum of parapatric population differentiation. *PLOS Genetics*, *11*(2), e1004966. doi:10.1371/journal.pgen.1004966

Franchini, P., Fruciano, C., Frickey, T., Jones, J. C., & Meyer, A. (2014). The gut microbial community of Midas cichlid fish in repeatedly evolved limnetic-benthic species pairs. *PLoS ONE*, *9*(4), e95027. doi:10.1371/journal.pone.0095027

Ghalambor, C. K., McKay, J. K., Carroll, S. P., & Reznick, D. N. (2007). Adaptive versus non-adaptive phenotypic plasticity and the potential for contemporary adaptation in new environments. *Functional Ecology*, *21*(3), 394–407. doi:10.1111/j.1365-2435.2007.01283.x

Ghalambor, Cameron K., Hoke, K. L., Ruell, E. W., Fischer, E. K., Reznick, D. N., & Hughes, K. A. (2015). Non-adaptive plasticity potentiates rapid adaptive evolution of gene expression in nature. *Nature*, *525*(7569), 372–375. doi:10.1038/nature15256

Gibbons, T. C., Metzger, D. C. H., Healy, T. M., & Schulte, P. M. (2017). Gene expression plasticity in response to salinity acclimation in threespine stickleback ecotypes from different salinity habitats. *Molecular Ecology*, *26*(10), 2711–2725. doi:10.1111/mec.14065

Greischar, M. A., & Koskella, B. (2007). A synthesis of experimental work on parasite local adaptation. *Ecology Letters*, *10*(5), 418–434. doi:10.1111/j.1461-0248.2007.01028.x

Hadidi, S., Glenney, G. W., Welch, T. J., Silverstein, J. T., & Wiens, G. D. (2008). Spleen size predicts resistance of rainbow trout to Flavobacterium psychrophilum challenge. *The Journal of Immunology*, *180*(6), 4156–4165. doi:10.4049/jimmunol.180.6.4156

Halmetoja, A., Valtonen, E. T., & Koskenniemi, E. (2000). Perch (Perca fluviatilis L.) parasites reflect ecosystem conditions: A comparison of a natural lake and two acidic reservoirs in Finland. *International Journal for Parasitology*, *30*(14), 1437–1444. doi:10.1016/S0020-7519(00)00115-6

Hanson, D., Hu, J., Hendry, A. P., & Barrett, R. D. H. (2017). Heritable gene expression differences between lake and stream stickleback include both parallel and antiparallel components. *Heredity*, (April), 1–10. doi:10.1038/hdy.2017.50

Härer, A., Torres-Dowdall, J., Rometsch, S., Yohannes, E., Machado-Schiaffino, G., & Meyer, A. (2019). Parallel and non-parallel changes of the gut microbiota during trophic diversification in repeated young adaptive radiations of sympatric cichlid fish. *BioRxiv*, *49*(0), 1–25. Retrieved from https://doi.org/10.1101/793760

Huang, Y., Chain, F. J. J., Panchal, M., Eizaguirre, C., & Kalbe, M. (2016). Transcriptome profiling of immune tissues reveals habitat-specific gene expression between lake and river sticklebacks. *Molecular Ecology*, 943–958. doi:10.1111/mec.13520

Hughes, L. C., Somoza, G. M., Nguyen, B. N., Bernot, J. P., González-Castro, M., Díaz de Astarloa, J. M., & Ortí, G. (2017). Transcriptomic differentiation underlying marine-to-freshwater transitions in the South American silversides *Odontesthes argentinensis* and *O. bonariensis* (Atheriniformes). *Ecology and Evolution*, *7*(14), 5258–5268. doi:10.1002/ece3.3133

Ingley, S. J., & Johnson, J. B. (2016). Divergent natural selection promotes immigrant inviability at early and late stages of evolutionary divergence. *Evolution; International Journal of Organic Evolution*, *70*(3), 600–616. doi:10.1111/evo.12872

Jeffries, K. M., Connon, R. E., Verhille, C. E., Dabruzzi, T. F., Britton, M. T., Durbin-Johnson, B. P., & Fangue, N. A. (2019). Divergent transcriptomic signatures in response to salinity exposure in two populations of an estuarine fish. *Evolutionary Applications*, *12*(6), 1212–1226. doi:10.1111/eva.12799

Jeukens, J., Renaut, S., St-Cyr, J., Nolte, A. W., & Bernatchez, L. (2010). The transcriptomics of sympatric dwarf and normal lake whitefish (Coregonus clupeaformis spp., Salmonidae) divergence as revealed by next-generation sequencing. *Molecular Ecology*, *19*(24), 5389–5403. doi:10.1111/j.1365-294X.2010.04934.x

Jones, F. C., Chan, Y. F., Schmutz, J., Grimwood, J., Brady, S. D., Southwick, A. M., … Kingsley, D. M. (2012). A genome-wide SNP genotyping array reveals patterns of global and repeated species-pair divergence in sticklebacks. *Current Biology*, *22*(1), 83–90. doi:10.1016/j.cub.2011.11.045

Kaltz, O., & Shykoff, J. A. (1998). Local adaptation in host-parasite systems. *Heredity*, *81*(4), 361–370. doi:10.1038/sj.hdy.6884350

Karvonen, A., Lucek, K., Marques, D. a., & Seehausen, O. (2015). Divergent macroparasite infections in parapatric Swiss lake-stream pairs of threespine stickleback (Gasterosteus aculeatus). *Plos One*, *10*(6), e0130579. doi:10.1371/journal.pone.0130579

Karvonen, A., & Seehausen, O. (2012). The role of parasitism in adaptive radiations-when might parasites promote and when might they constrain ecological speciation? *International Journal of Ecology*, *2012*, 1–20. doi:10.1155/2012/280169

Kaufmann, J., Lenz, T. L., Kalbe, M., Milinski, M., & Eizaguirre, C. (2017). A field reciprocal transplant experiment reveals asymmetric costs of migration between lake and river ecotypes of three-spined sticklebacks (Gasterosteus aculeatus). *Journal of Evolutionary Biology*, *30*(5), 938–950. doi:10.1111/jeb.13057

Kavembe, G. D., Franchini, P., Irisarri, I., Machado-Schiaffino, G., & Meyer, A. (2015). Genomics of adaptation to multiple concurrent stresses: insights from comparative transcriptomics of a cichlid fish from one of Earth's most extreme environments, the hypersaline soda Lake Magadi in Kenya, East Africa. *Journal of Molecular Evolution*, *81*(3–4), 90–109. doi:10.1007/s00239-015-9696-6

Kawecki, T. J., & Ebert, D. (2004). Conceptual issues in local adaptation. *Ecology Letters*, *7*(12), 1225–1241. doi:10.1111/j.1461-0248.2004.00684.x

Keller, I. S., Bayer, T., Salzburger, W., & Roth, O. (2017). Effects of parental care on resource allocation into immune defense and buccal microbiota in mouthbrooding cichlid fishes. *Evolution; International Journal of Organic Evolution*. Retrieved from http://onlinelibrary.wiley.com/resolve/doi?DOI=10.1111/evo.13452

Kraaijeveld, A. R., & Godfrey, H. C. J. (1997). Trade-off between parasitoid resistance and larval competitive ability in *Drosophila melanogaster*. *Nature*, *389*(6648), 278–280. doi:10.1038/38483

Lam, S. H., Lui, E. Y., Li, Z., Cai, S., Sung, W. K., Mathavan, S., … Ip, Y. K. (2014). Differential transcriptomic analyses revealed genes and signaling pathways involved in iono-osmoregulation and cellular remodeling in the gills of euryhaline Mozambique tilapia, Oreochromis mossambicus. *BMC Genomics*, *15*(1), 1–23. doi:10.1186/1471-2164-15-921

Langfelder, P., & Horvath, S. (2008). WGCNA: an R package for weighted correlation network analysis. *BMC Bioinformatics*, *9*(2), 559. doi:10.1186/1471-2105-9-559

Larsen, P. F., Nielsen, E. E., Koed, A., Thomsen, D. S., Olsvik, P. A., & Loeschcke, V. (2008). Interpopulation differences in expression of candidate genes for salinity tolerance in winter migrating anadromous brown trout (Salmo trutta L.). *BMC Genetics*, *9*, 1–9. doi:10.1186/1471-2156-9-12

Lefebvre, F., Mounaix, B., Poizat, G., & Crivelli, A. J. (2004). Impacts of the swimbladder nematode Anguillicola crassus on Anguilla anguilla: variations in liver and spleen masses. *Journal of Fish Biology*, *64*(2), 435–447. doi:10.1111/j.0022-1112.2004.00309.x

Lenth, R. V. (2016). Least-squares means: the R package lsmeans. *Journal of Statistical Software*, *69*(1), 1–33. doi:10.18637/jss.v069.i01

Lenz, T. L., Eizaguirre, C., Rotter, B., Kalbe, M., & Milinski, M. (2013). Exploring local immunological adaptation of two stickleback ecotypes by experimental infection and transcriptome-wide digital gene expression analysis. *Molecular Ecology*, *22*(3), 774–786. doi:10.1111/j.1365-294X.2012.05756.x

Lindstrom, K. M., Foufopoulos, J., Parn, H., & Wikelski, M. (2004). Immunological investments reflect parasite abundance in island populations of Darwin's finches. *Proceedings of the Royal Society B: Biological Sciences*, *271*(1547), 1513–1519. doi:10.1098/rspb.2004.2752

Lochmiller, R. L., & Deerenberg, C. (2000). Trade-offs in evolutionary immunology: just what is the cost of immunity? *Oikos*, *88*(1), 87–98. doi:10.1034/j.1600-0706.2000.880110.x

Lohman, B. K., Stutz, W. E., & Bolnick, D. I. (2017). Gene expression stasis and plasticity following migration into a foreign environment. *Molecular Ecology*, *26*(18), 4657–4670. doi:10.1111/mec.14234

Lokesh, J., & Kiron, V. (2016). Transition from freshwater to seawater reshapes the skin-associated microbiota of Atlantic salmon. *Scientific Reports*, *6*(December 2015), 1–10. doi:10.1038/srep19707

Lopez-Maestre, H., Brinza, L., Marchet, C., Kielbassa, J., Bastien, S., Boutigny, M., … Lacroix, V. (2016). SNP calling from RNA-seq data without a reference genome: Identification, quantification, differential analysis and impact on the protein sequence. *Nucleic Acids Research*, *44*(19), 1–13. doi:10.1093/nar/gkw655

Love, M. I., Huber, W., & Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biology*, *15*(12), 1–21. doi:10.1186/s13059-014-0550-8

Magurran, A. E. (1988). *Ecological diversity and its measurement*. Princeton University Press.

Marchinko, K. B. (2009). Predation's role in repeated phenotypic and genetic divergence of armor in threespine stickleback. *Evolution; International Journal of Organic Evolution*, *63*(1), 127–138. doi:10.1111/j.1558-5646.2008.00529.x

Marijnissen, S. A. E., Michel, E., Daniels, S. R., Erpenbeck, D., Menken, S. B. J., & Schram, F. R. (2006). Molecular evidence for recent divergence of Lake Tanganyika endemic crabs (Decapoda: Platythelphusidae). *Molecular Phylogenetics and Evolution*, *40*(2), 628–634. doi:10.1016/j.ympev.2006.03.025

Matthews, B., Harmon, L. J., M'Gonigle, L., Marchinko, K. B., & Schaschl, H. (2010). Sympatric and allopatric divergence of MHC genes in threespine stickleback. *PLoS ONE*, *5*(6), e10948. doi:10.1371/journal.pone.0010948

McCairns, R. J. S., & Bernatchez, L. (2009). Adaptive divergence between freshwater and marine sticklebacks: Insights into the role of phenotypic plasticity from an integrated analysis of candidate gene expression. *Evolution; International Journal of Organic Evolution*, *64*(4), 1029–1047. doi:10.1111/j.1558-5646.2009.00886.x

McCallum, H., Barlow, N., & Hone, J. (2002). Modelling transmission: mass action and beyond. *Trends in Ecology & Evolution*, *17*(2), 64–65. doi:10.1016/S0169-5347(01)02399-0

McDade, T. W., Georgiev, A. V., & Kuzawa, C. W. (2016). Trade-offs between acquired and innate immune defenses in humans. *Evolution, Medicine and Public Health*, *2016*(1), 1–16. doi:10.1093/EMPH/EOV033

McFall-Ngai, M., Hadfield, M. G., Bosch, T. C. G., Carey, H. V., Domazet-Lošo, T., Douglas, A. E., … Wernegreen, J. J. (2013). Animals in a bacterial world, a new imperative for the life sciences. *Proceedings of the National Academy of Sciences*, *110*(9), 3229–3236. doi:10.1073/pnas.1218525110

Medzhitov, R., & Janeway, C. A. (2002). Decoding the pattern of self and nonself by the innate immune system. *Science*, *296*(2002), 298–300. doi:10.1126/science.1068883

Moret, Y., & Schmid-Hempel, P. (2000). Survival for immunity: the price of immune system activation for bumblebee workers. *Science*, *290*(December), 1166–1168. doi:10.1126/science.290.5494.1166

Morris, M. R. J., Richard, R., Leder, E. H., Barrett, R. D. H., Aubin-Horth, N., & Rogers, S. M. (2014). Gene expression plasticity evolves in response to colonization of freshwater lakes in threespine stickleback. *Molecular Ecology*, *23*(13), 3226–3240. doi:10.1111/mec.12820

Nosil, P. (2004). Reproductive isolation caused by visual predation on migrants between divergent environments. *Proceedings of the Royal Society B: Biological Sciences*, *271*(1547), 1521–1528. doi:10.1098/rspb.2004.2751

Nosil, P., & Crespi, B. J. (2006). Experimental evidence that predation promotes divergence in adaptive radiation. *Proceedings of the National Academy of Sciences*, *103*(24), 9090–9095. doi:10.1073/pnas.0601575103

Paperna, I. 1996. Parasites, infections and diseases of fishes in Africa: An update. CIFA Technical Paper (FAO). no. 31.

Paradis, E., Claude, J., & Strimmer, K. (2004). APE: Analyses of phylogenetics and evolution in R language. *Bioinformatics*, *20*(2), 289–290. doi:10.1093/bioinformatics/btg412

Pavey, S. A., Collin, H., Nosil, P., & Rogers, S. M. (2010). The role of gene expression in ecological speciation. *Annals of the New York Academy of Sciences*, *1206*, 110–129. doi:10.1111/j.1749-6632.2010.05765.x

Pfennig, D. W., Wund, M. A., Snell-Rood, E. C., Cruickshank, T., Schlichting, C. D., & Moczek, A. P. (2010). Phenotypic plasticity's impacts on diversification and speciation. *Trends in Ecology and Evolution*, *25*(8), 459–467. doi:10.1016/j.tree.2010.05.006

Press, C. M. L., & Evensen, O. (1999). The morphology of the immune system in teleost fishes. *Fish and Shellfish Immunology*, *9*(4), 309–318. doi:10.1006/fsim.1998.0181

Price, T. D., Qvarnstrom, A., & Irwin, D. E. (2003). The role of phenotypic plasticity in driving genetic evolution. *Proceedings of the Royal Society B: Biological Sciences*, *270*(1523), 1433–1440. doi:10.1098/rspb.2003.2372

Rahn, A. K., Eßer, E., Reher, S., Ihlow, F., MacColl, A. D. C., & Bakker, T. C. M. (2016). Distribution of common stickleback parasites on North Uist, Scotland, in relation to ecology and host traits. *Zoology*, *119*(4), 395–402. doi:10.1016/j.zool.2016.05.009

Rajkov, J., Weber, A. A.-T., Salzburger, W., & Egger, B. (2018a). Adaptive phenotypic plasticity contributes to divergence between lake and river populations of an East African cichlid fish. *Ecology and Evolution*, *8*(15), 7323–7333. doi:10.1002/ece3.4241

Rajkov, J., Weber, A. A.-T., Salzburger, W., & Egger, B. (2018b). Immigrant and extrinsic hybrid inviability contribute to reproductive isolation between lake and river cichlid ecotypes. *Evolution*. doi:10.1111/evo.13612

Reilly, B. D., Cramp, R. L., Wilson, J. M., Campbell, H. A., & Franklin, C. E. (2011). Branchial osmoregulation in the euryhaline bull shark, *Carcharhinus leucas*: a molecular analysis of ion transporters. *Journal of Experimental Biology*, *214*(17), 2883–2895. doi:10.1242/jeb.058156

Renaut, S., Nolte, A. W., Rogers, S. M., Derome, N., & Bernatchez, L. (2011). SNP signatures of selection on standing genetic variation and their association with adaptive phenotypes along gradients of ecological speciation in lake whitefish species pairs (Coregonus spp.). *Molecular Ecology*, *20*(3), 545–559. doi:10.1111/j.1365-294X.2010.04952.x

Rennison, D. J., Rudman, S. M., & Schluter, D. (2019). Parallel changes in gut microbiome composition and function in parallel local adaptation and speciation. *BioRxiv*, 1–29.

Richards, C. L., Bossdorf, O., Muth, N. Z., Gurevitch, J., & Pigliucci, M. (2006). Jack of all trades, master of some? On the role of phenotypic plasticity in plant invasions. *Ecology Letters*, *9*(8), 981–993. doi:10.1111/j.1461-0248.2006.00950.x

Roesti, M., Gavrilets, S., Hendry, A. P., Salzburger, W., & Berner, D. (2014). The genomic signature of parallel adaptation from shared genetic variation. *Molecular Ecology*, *23*(16), 3944–3956. doi:10.1111/mec.12720

Rózsa, L., Reiczigel, J., & Majoros, G. (2000). Quantifying parasites in samples of hosts. *The Journal of Parasitology*, *86*(2), 228–232. doi:10.1645/0022-3395(2000)086[0228:QPISOH]2.0.CO;2

Rundle, H. D., & Nosil, P. (2005). Ecological speciation. *Ecology Letters*, *8*(3), 336–352. doi:10.1111/j.1461-0248.2004.00715.x

Scharsack, J. P., Kalbe, M., Harrod, C., & Rauch, G. (2007). Habitat-specific adaptation of immune responses of stickleback (*Gasterosteus aculeatus*) lake and river ecotypes. *Proceedings of the Royal Society B: Biological Sciences*, *274*(1617), 1523–1532. doi:10.1098/rspb.2007.0210

Schlichting, C. D. (2004). *The role of phenotypic plasticity in diversification in Phenotypic plasticity: fuctional and conceptual approaches*. (T. J. DeWitt & S. M. Scheiner, Eds.). Oxford University Press, Oxford. doi:10.1002/ajhb.20088

Schlichting, C. D., & Pigliucci, M. (1998). *Phenotypic evolution: a reaction norm perspective*. *Phenotypic Evolution: A Reaction Norm Perspective*. Sunderland, Massachusetts: Sinauer Associates Inc. doi:10.2307/2656964

Schluter, D. (2003). Frequency dependent natural selection during character displacement in sticklebacks. *Evolution; International Journal of Organic Evolution*, *57*(5), 1142–1150.

Schmid-Hempel, P., & Ebert, D. (2003). On the evolutionary ecology of specific immune defence. *Trends in Ecology & Evolution*, *18*(1), 27–32. doi:10.1016/S0169-5347(02)00013-7

Schmidt, V. T., Smith, K. F., Melvin, D. W., & Amaral-Zettler, L. A. (2015). Community assembly of a euryhaline fish microbiome during salinity acclimation. *Molecular Ecology*, *24*(10), 2537–2550. doi:10.1111/mec.13177

Schneider, R. F., & Meyer, A. (2017). How plasticity, genetic assimilation and cryptic genetic variation may contribute to adaptive radiations. *Molecular Ecology*, *26*(1), 330–350. doi:10.1111/mec.13880

Scott, G. R. (2004). Intraspecific divergence of ionoregulatory physiology in the euryhaline teleost Fundulus heteroclitus: possible mechanisms of freshwater adaptation. *Journal of Experimental Biology*, *207*(19), 3399–3410. doi:10.1242/jeb.01130

Secombes, C. J., & Wang, T. (2012). *The innate and adaptive immune system of fish*. *Infectious Disease in Aquaculture*. Woodhead Publishing Limited. doi:10.1533/9780857095732.1.3

Sevellec, M., Pavey, S. A., Boutin, S., Filteau, M., Derome, N., & Bernatchez, L. (2014). Microbiome investigation in the ecological speciation context of lake whitefish (Coregonus clupeaformis) using next-generation sequencing. *Journal of Evolutionary Biology*, *27*(6), 1029–1046. doi:10.1111/jeb.12374

Sevellec, Maelle, Derome, N., & Bernatchez, L. (2018). Holobionts and ecological speciation: the intestinal microbiota of lake whitefish species pairs. *Microbiome*, *6*(1), 47. doi:10.1186/s40168-018-0427-2

Sharon, G., SEgal, D., Ringo, J. M., Hefetz, A., Zilber-Rosenberg, I., & Rosenberg, E. (2010). Commensal bacteria play a role in mating preference of Drosophila melanogaster. *Proceedings of the National Academy of Sciences*, *107*, 20051–20056. doi:10.1073/pnas.1302980110

Sheldon, B. C., & Verhulst, S. (1996). Ecological immunology - costly parasite defenses and trade-offs in evolutionary ecology. *Trends in Ecology & Evolution*, *11*(96), 317–321. doi:10.1016/0169-5347(96)10039-2

Shropshire, J. D., & Bordenstein, S. R. (2016). Speciation by symbiosis: The microbiome and behavior. *MBio*, *7*(2), 1–11. doi:10.1128/mBio.01785-15

Shudo, E., & Iwasa, Y. (2001). Inducible defense against pathogens and parasites: Optimal choice among multiple options. *Journal of Theoretical Biology*, *209*(2), 233–247. doi:10.1006/jtbi.2000.2259

Stutz, W. E., & Bolnick, D. I. (2017). Natural selection on MHC IIβ in parapatric lake and stream stickleback: Balancing, divergent, both or neither? *Molecular Ecology*, (April), 4772–4786. doi:10.1111/mec.14158

Sullam, K. E., Essinger, S. D., Lozupone, C. A., O'Connor, M. P., Rosen, G. L., Knight, R., … Russell, J. A. (2012). Environmental and ecological factors that shape the gut bacterial communities of fish: A meta-analysis. *Molecular Ecology*, *21*(13), 3363–3378. doi:10.1111/j.1365-294X.2012.05552.x

Sullam, K. E., Rubin, B. E. R., Dalton, C. M., Kilham, S. S., Flecker, A. S., & Russell, J. A. (2015). Divergence across diet, time and populations rules out parallel evolution in the gut microbiomes of Trinidadian guppies. *ISME Journal*, *9*(7), 1508–1522. doi:10.1038/ismej.2014.231

Summers, K., McKeon, S., Sellars, J., Keusenkothen, M., Morris, J., Gloeckner, D., … Snow, H. (2003). Parasitic exploitation as an engine of diversity. *Biological Reviews of the Cambridge Philosophical Society*, *78*(4), 639–675. doi:10.1017/S146479310300616X

Svanback, R., & Bolnick, D. I. (2007). Intraspecific competition drives increased resource use diversity within a natural population. *Proceedings of the Royal Society B: Biological Sciences*, *274*(1611), 839–844. doi:10.1098/rspb.2006.0198

Tang, H. Y., Smith-Caldas, M. S. B., Driscoll, M. V., Salhadar, S., & Shingleton, A. W. (2011). FOXO regulates organ-specific phenotypic plasticity in Drosophila. *PLoS Genetics*, *7*(11), e1002373. doi:10.1371/journal.pgen.1002373

Theis, A., Ronco, F., Indermaur, A., Salzburger, W., & Egger, B. (2014). Adaptive divergence between lake and stream populations of an East African cichlid fish. *Molecular Ecology*, *23*(21), 5304–5322. doi:10.1111/mec.12939

Theis, A., Roth, O., Cortesi, F., Ronco, F., Salzburger, W., & Egger, B. (2017). Variation of anal fin egg-spots along an environmental gradient in a haplochromine cichlid fish. *Evolution; International Journal of Organic Evolution*, *71*, 766–777. doi:10.1111/evo.13166

Thibert-Plante, X., & Hendry, A. P. (2011). The consequences of phenotypic plasticity for ecological speciation. *Journal of Evolutionary Biology*, *24*(2), 326–342. doi:10.1111/j.1420-9101.2010.02169.x

Todd, E. V., Black, M. A., & Gemmell, N. J. (2016). The power and promise of RNA-seq in ecology and evolution. *Molecular Ecology*, *25*(6), 1224–1241. doi:10.1111/mec.13526

Tschirren, B., Andersson, M., Scherman, K., Westerdahl, H., Mittl, P. R. E., & Raberg, L. (2013). Polymorphisms at the innate immune receptor TLR2 are associated with Borrelia infection in a wild rodent population. *Proceedings of the Royal Society B: Biological Sciences*, *280*(1759), 20130364–20130364. doi:10.1098/rspb.2013.0364

Vamosi, S. M., & Schluter, D. (2002). Impacts of trout predation on fitness of sympatric sticklebacks and their hybrids. *Proceedings of the Royal Society B: Biological Sciences*, *269*(1494), 923–930. doi:10.1098/rspb.2001.1946

Vamosi, Steven M. (2003). The presence of other fish species affects speciation in threespine sticklebacks. *Evolutionary Ecology Research*, *5*(5), 717–730.

Waddington, C. H. (1942). Canalization of development and the inheritance of acquired characters. *Nature*, *150*, 563–565. doi:10.1038/150405d0

Waddington, C. H. (1953). Genetic Assimilation of an Acquired Character Author (s): C . H . Waddington Published by : Society for the Study of Evolution Stable URL : http://www.jstor.org/stable/2405747. *Evolution*, *7*(2), 118–126.

West-Eberhard, M. J. (2003). Developmental plasticity and evolution. *Nature*. Oxford University Press, Oxford. doi:10.2002/ajpa.20219

West-Eberhard, M. J. (2005). Developmental plasticity and the origin of species differences. *Proceedings of the National Academy of Sciences of the United States of America*, *102 Suppl*(2), 6543–6549. doi:10.1073/pnas.0501844102

Xu, J., Li, Q., Xu, L., Wang, S., Jiang, Y., Zhao, Z., … Sun, X. (2013). Gene expression changes leading extreme alkaline tolerance in Amur ide (*Leuciscus waleckii*) inhabiting soda lake. *BMC Genomics*, *14*(1), 682. doi:10.1186/1471-2164-14-682

Yeh, P. J., & Price, T. D. (2004). Adaptive phenotypic plasticity and the successful colonization of a novel environment. *The American Naturalist*, *164*(4), 531–542. doi:10.1086/423825

Zeilies, A., Kleiber, C., & Jackman, S. (2008). Regression models for count data in R. *Journal of Statistical Software*, *27*(8), 1–25. doi:10.18637/jss.v027.i08

Zhang, X., Li, Q., Kong, L., & Yu, H. (2017). DNA methylation changes detected by methylation-sensitive amplified polymorphism in the Pacific oyster (*Crassostrea gigas*) in response to salinity stress. *Genes and Genomics*, *39*(11), 1173–1181. doi:10.1007/s13258-017-0583-y

Zhu, J., Fu, Q., Ao, Q., Tan, Y., Luo, Y., Jiang, H., … Gan, X. (2017). Transcriptomic profiling analysis of tilapia (*Oreochromis niloticus*) following *Streptococcus agalactiae* challenge. *Fish and Shellfish Immunology*, *62*, 202–212. doi:10.1016/j.fsi.2017.01.023

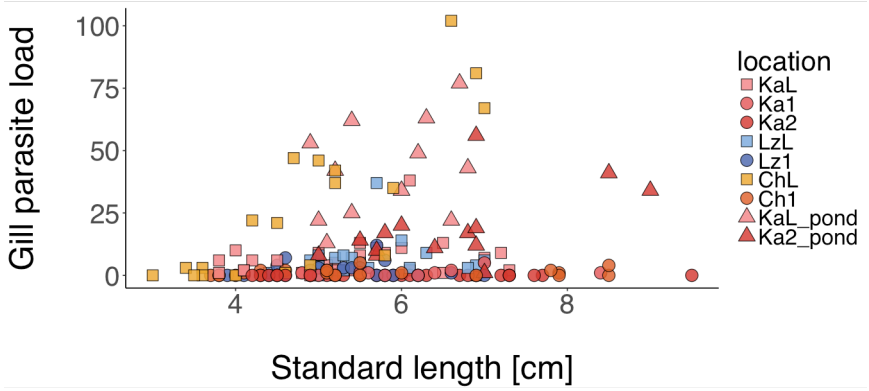


Figure S1 Gill parasite load *versus* fish standard length. Colours represent populations; squares – lake, circles – river, triangles – pond. Colour code according to Figure 1. KaL – Kalambo lake, Ka1 – Kalambo river, Ka2 – Kalambo river upstream, ChL – Chitili lake, Ch1 – Chitili river, LzL – Lunzua lake, Lz1 – Lunzua river.

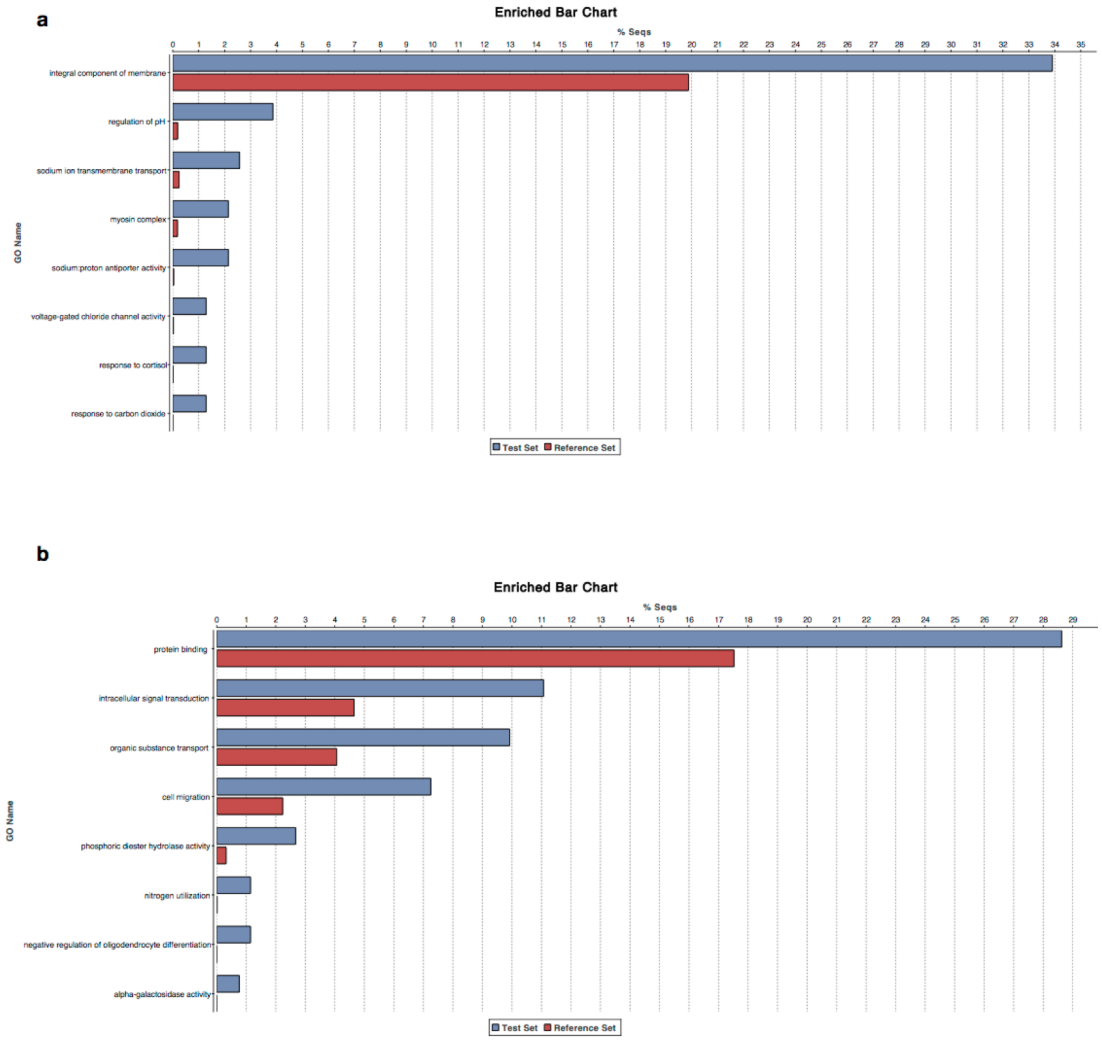


Figure S2 Gene enrichment plots for co-expression networks from gill (greenyellow (a) and magenta (b) modules) and spleen (red (c) and darkturquoise (d) modules) expression profiles.

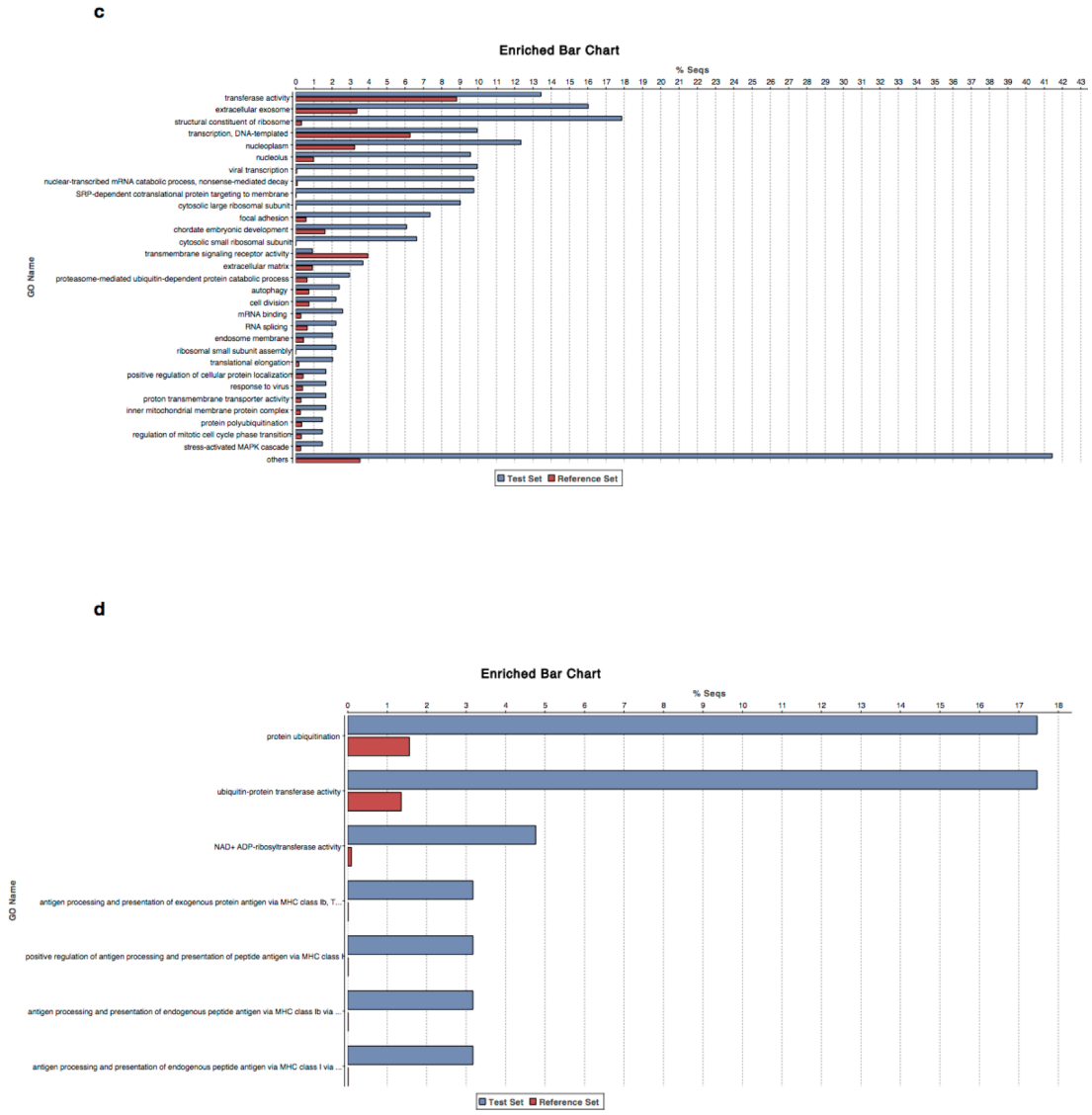


Figure S2 (continued) Gene enrichment plots for co-expression networks from gill (greenyellow (a) and magenta (b) modules) and spleen (red (c) and darkturquoise (d) modules) expression profiles.

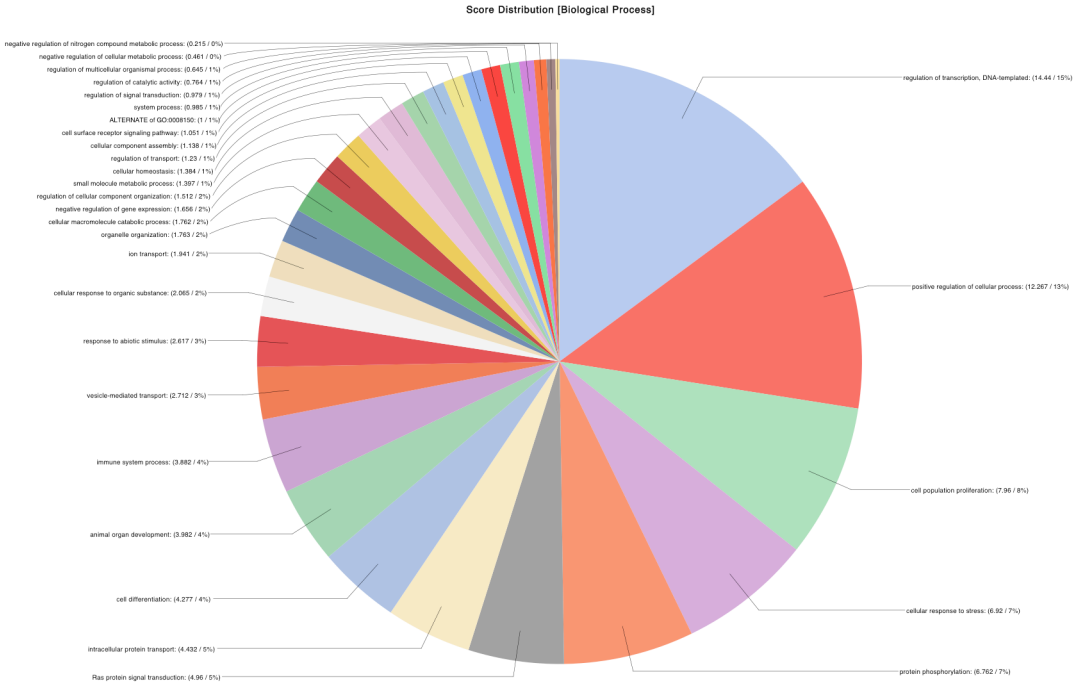


Figure S3 Pie chart of the GO term associated with the genes with significant habitat-specific SNPs (Table S9).

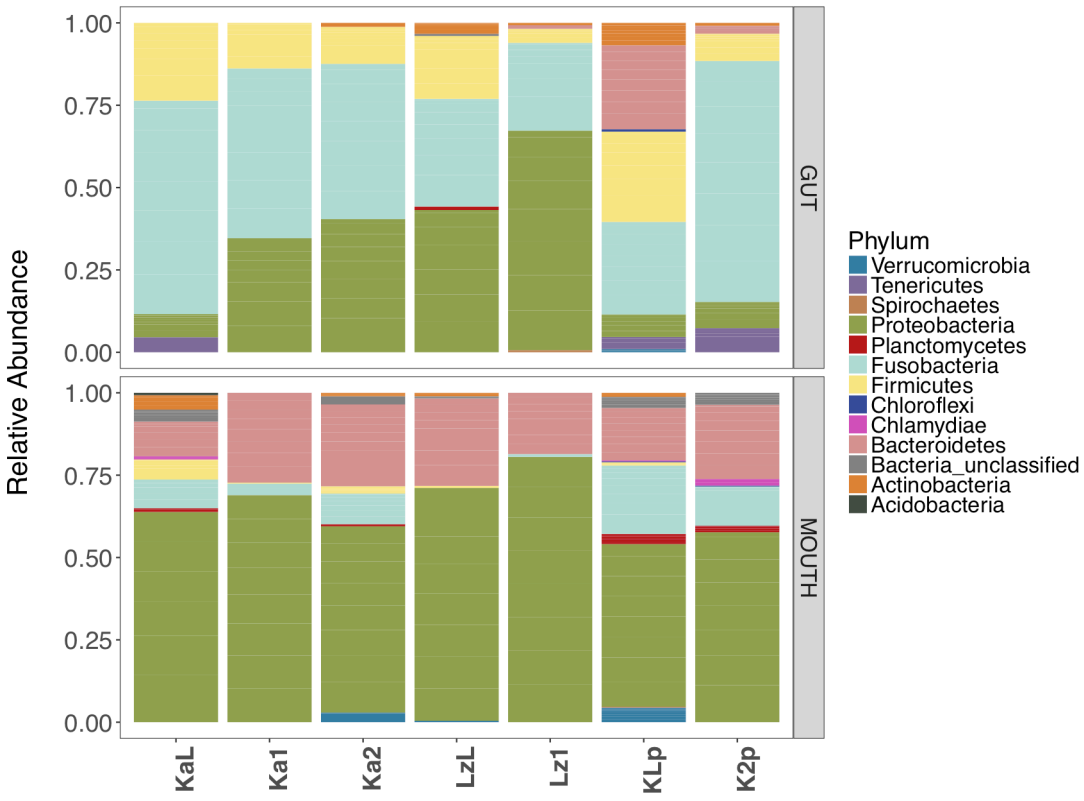


Figure S4 Relative abundance of different microbiota phyla per population given after filtering out low abundance sequences (≤ 0.02). KaL – Kalambo lake, Ka1 – Kalambo river, Ka2 – Kalambo river upstream, LzL – Lunzua lake, Lz1 – Lunzua river, KLp – Kalambo lake ponds, K2p – Kalambo river upstream ponds.

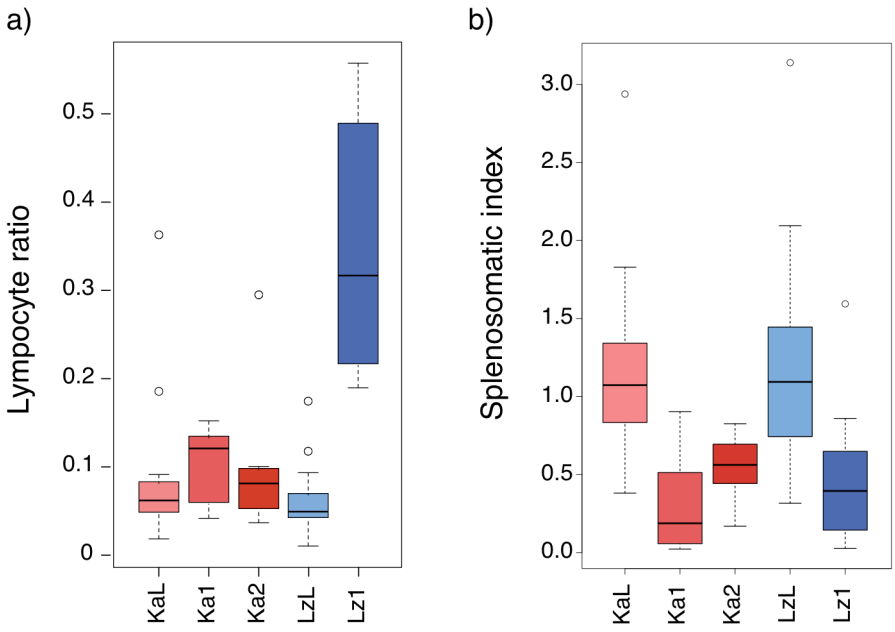


Figure S5 Comparison of adaptive and innate immune response in wild lake and river *A. burtoni* populations**.** Proportion of cells of the adaptive immune system estimated as lymphocyte ratio in the blood (lymphocyte count/(lymphocyte + monocyte counts)) as found in Theis *et al.* (2017) (a) and splenosomatic index as an estimate of the innate immune response in the same populations in this study (b): KaL – Kalambo lake, Ka1 – Kalambo river, Ka2 – Kalambo river upstream, LzL – Lunzua lake, Lz1 – Lunzua river.

Table S1 Primer sequences for the second PCR for 16S Metagenomic Sequencing Library Preparation.

5' PCR primer (length 64 bp):

primer name	sequence	index
TS-HT-D5x-1-f	AAT GAT ACG GCG ACC ACC GAG ATC TAC AC aaccatgc ACA CTC TTT CCC TAC ACG ACG CTC TTC	aaccatgc
TS-HT-D5x-2-f	AAT GAT ACG GCG ACC ACC GAG ATC TAC AC aacctaga ACA CTC TTT CCC TAC ACG ACG CTC TTC	aacctaga
TS-HT-D5x-7-f	AAT GAT ACG GCG ACC ACC GAG ATC TAC AC aacgcgca ACA CTC TTT CCC TAC ACG ACG CTC TTC	aacgcgca
TS-HT-D5x-8-f	AAT GAT ACG GCG ACC ACC GAG ATC TAC AC aacggcgc ACA CTC TTT CCC TAC ACG ACG CTC TTC	aacggcgc
TS-HT-D5x-68-f	AAT GAT ACG GCG ACC ACC GAG ATC TAC AC ggtagaat ACA CTC TTT CCC TAC ACG ACG CTC TTC	ggtagaat
TS-HT-D5x-73-f	AAT GAT ACG GCG ACC ACC GAG ATC TAC AC gttagcgg ACA CTC TTT CCC TAC ACG ACG CTC TTC	gttagcgg
TS-HT-D5x-89-f	AAT GAT ACG GCG ACC ACC GAG ATC TAC AC tggtaatt ACA CTC TTT CCC TAC ACG ACG CTC TTC	tggtaatt
TS-HT-D5x-94-f	AAT GAT ACG GCG ACC ACC GAG ATC TAC AC ttgcatat ACA CTC TTT CCC TAC ACG ACG CTC TTC	ttgcatat

3' PCR primer (length 60 bp):

primer name	sequence	index (rc)
TS-HT-D7x-7-r	CAA GCA GAA GAC GGC ATA CGA GAT aacgaagt GTG ACT GGA GTT CAG ACG TGT GCT CTT C	aacgaagt
TS-HT-D7x-8-r	CAA GCA GAA GAC GGC ATA CGA GAT aacgacta GTG ACT GGA GTT CAG ACG TGT GCT CTT C	aacgacta
TS-HT-D7x-10-r	CAA GCA GAA GAC GGC ATA CGA GAT aagaagac GTG ACT GGA GTT CAG ACG TGT GCT CTT C	aagaagac
TS-HT-D7x-12-r	CAA GCA GAA GAC GGC ATA CGA GAT accaaccg GTG ACT GGA GTT CAG ACG TGT GCT CTT C	accaaccg
TS-HT-D7x-14-r	CAA GCA GAA GAC GGC ATA CGA GAT accaatta GTG ACT GGA GTT CAG ACG TGT GCT CTT C	accaatta
TS-HT-D7x-64-r	CAA GCA GAA GAC GGC ATA CGA GAT ggttagtc GTG ACT GGA GTT CAG ACG TGT GCT CTT C	ggttagtc
TS-HT-D7x-83-r	CAA GCA GAA GAC GGC ATA CGA GAT tgagtcct GTG ACT GGA GTT CAG ACG TGT GCT CTT C	tgagtcct
TS-HT-D7x-95-r	CAA GCA GAA GAC GGC ATA CGA GAT ttggagct GTG ACT GGA GTT CAG ACG TGT GCT CTT C	ttggagct

Table S2 Genome-wide F_{ST} values and pairwise geographic distances.

comparison	type	F_{ST}	distance [m]
Lz1-LzL	lake-river sympatric	0.06381	7246
KaL-LzL	lake-lake	0.05683	24012
Ka1-LzL	lake-river parapatric	0.05449	26197
Ka2-LzL	lake-river parapatric	0.15383	29488
KaL-Lz1	lake-river parapatric	0.12360	31258
Ka1-Lz1	river-river	0.12047	33443
Ka2-Lz1	river-river	0.20970	36734
Ka1-KaL	lake-river sympatric	0.00517	2185
Ka2-KaL	lake-river sympatric	0.11350	5476
Ka2-Ka1	river-river	0.10211	3291
Ch1-ChL	lake-river sympatric	0.00478	317

Table S3 Analyses of variance table and results of Tukey-Kramer *post hoc* pairwise comparisons of means of a zero-inflated Poisson generalized linear model on parasite taxa number for all the wild populations. SL – standard length. Significant effects ($p < 0.05$) are highlighted.

	d.f.	χ^2	p
population	**6**	**80.68**	**<0.001**
SL	1	0.02	0.651
sex	2	1.43	0.490

contrast	estimate	SE	*Z*	p
KaL-Ka1	1.236	0.478	2.583	0.131
KaL-Ka2	1.841	0.552	3.335	**0.015**
KaL-LzL	-0.472	0.479	-0.986	0.957
KaL-Lz1	1.297	0.477	2.721	0.093
KaL-ChL	-0.471	0.592	-0.795	0.986
KaL-Ch1	1.141	0.508	2.246	0.271
Ka1-Ka2	0.605	0.256	2.369	0.212
Ka1-LzL	-1.708	0.533	-3.207	**0.023**
Ka1-Lz1	0.061	0.266	0.229	1.000
Ka1-ChL	-1.706	0.681	-2.504	0.158
Ka1-Ch1	-0.095	0.352	-0.27	1.000
Ka2-LzL	-2.314	0.612	-3.778	**0.003**
Ka2-Lz1	-0.544	0.225	-2.417	0.191
Ka2-ChL	-2.312	0.764	-3.026	**0.040**
Ka2-Ch1	-0.701	0.324	-2.164	0.315
LzL-Lz1	1.769	0.530	3.339	**0.015**
LzL-ChL	0.002	0.605	0.003	1.000
LzL-Ch1	1.613	0.579	2.788	0.078
Lz1-ChL	-1.767	0.681	-2.596	0.127
Lz1-Ch1	-0.156	0.340	-0.459	0.999
ChL-Ch1	1.611	0.714	2.256	0.266

Table S4 Analyses of variance table and results of Tukey-Kramer *post hoc* pairwise comparisons of means of a zero-inflated negative binomial generalized linear model on total parasite load. SL – standard length. Significant effects ($p < 0.05$) are highlighted.

	d.f.	χ^2	p
population	**6**	**147.03**	**<0.001**
SL	1	0.08	0.784
sex	**2**	**6.55**	**0.038**

location

contrast	estimate	SE	*Z*	p
KaL-Ka1	6.041	2.043	2.957	**0.049**
KaL-Ka2	6.907	2.214	3.120	**0.030**
KaL-LzL	-1.228	1.957	-0.628	0.996
KaL-Lz1	4.802	1.835	2.617	0.121
KaL-ChL	-37.988	14.471	-2.625	0.118
KaL-Ch1	5.524	1.985	2.783	0.079
Ka1-Ka2	0.866	0.358	2.421	0.190
Ka1-LzL	-7.270	2.242	-3.242	**0.020**
Ka1-Lz1	-1.239	0.659	-1.880	0.494
Ka1-ChL	-44.029	15.607	-2.821	0.071
Ka1-Ch1	-0.517	0.584	-0.885	0.975
Ka2-LzL	-8.136	2.403	-3.386	**0.013**
Ka2-Lz1	-2.105	0.759	-2.775	0.081
Ka2-ChL	-44.895	15.780	-2.845	0.067
Ka2-Ch1	-1.383	0.611	-2.264	0.262
LzL-Lz1	6.030	2.028	2.973	**0.047**
LzL-ChL	-36.759	14.483	-2.538	0.146
LzL-Ch1	6.753	2.224	3.036	**0.039**
Lz1-ChL	-42.790	15.351	-2.787	0.078
Lz1-Ch1	0.722	0.769	0.939	0.966
ChL-Ch1	43.512	15.519	2.804	0.075

sex

contrast	estimate	SE	*Z*	p
?-f	2.303	7.396	0.311	0.948
?-m	-3.499	7.535	-0.464	0.888
f-m	-5.802	2.332	-2.488	**0.034**

Table S5 Analyses of variance tables and results of Tukey-Kramer *post hoc* pairwise comparisons of means of a zero-inflated negative binomial generalized linear models of gill parasite load for all the wild populations (model 1) and for Kalambo river system comparison between the wild and pond setup (model 2). SL – standard length. Significant effects ($P < 0.05$) are highlighted. Sex: f – females, m-males, ? – juveniles.

model 1: wild

	d.f.	χ^2	p
population	**6**	**1065.50**	**<0.001**
SL	**1**	**15.09**	**<0.001**
sex	**2**	**9.95**	**<0.001**

location

contrast	estimate	SE	*Z*	p
KaL-Ka1	5.696	1.072	5.313	**<0.001**
KaL-Ka2	6.660	1.161	5.739	**<0.001**
KaL-LzL	0.659	0.787	0.837	0.981
KaL-Lz1	2.683	0.978	2.743	0.088
KaL-ChL	-34.595	5.725	-6.043	**<0.001**
KaL-Ch1	5.416	1.078	5.025	**<0.001**
Ka1-Ka2	0.965	0.292	3.303	**0.017**
Ka1-LzL	-5.037	0.904	-5.575	**<0.001**
Ka1-Lz1	-3.013	0.859	-3.506	**0.008**
Ka1-ChL	-40.290	6.554	-6.147	**<0.001**
Ka1-Ch1	-0.280	0.444	-0.631	0.996
Ka2-LzL	-6.002	0.976	-6.148	**<0.001**
Ka2-Lz1	-3.977	0.906	-4.392	**<0.001**
Ka2-ChL	-41.255	6.681	-6.175	**<0.001**
Ka2-Ch1	-1.245	0.402	-3.097	**0.032**
LzL-Lz1	2.024	0.897	2.257	0.266
LzL-ChL	-35.253	5.920	-5.955	**<0.001**
LzL-Ch1	4.757	0.925	5.142	**<0.001**
Lz1-ChL	-37.278	6.152	-6.060	**<0.001**
Lz1-Ch1	2.733	0.886	3.083	**0.034**
ChL-Ch1	40.010	6.534	6.123	**<0.001**

sex

contrast	estimate	SE	*Z*	p
?-f	2.825	3.749	0.754	0.731
?-m	0.218	3.769	0.058	0.998
f-m	**-2.607**	**0.806**	**-3.235**	**0.004**

model 2: wild *vs.* ponds

	d.f.	χ^2	p
population	**3**	**102.01**	**<0.001**
SL	1	1.44	0.231
sex	2	1.17	0.279

location

contrast	estimate	SE	*Z*	p
KaL-Ka2	**7.465**	**1.383**	**5.399**	**<0.001**
KaL-KaL_pond	**-37.999**	**9.357**	**-4.061**	**<0.001**
KaL-Ka2_pond	-8.783	3.937	-2.231	0.115
Ka2-KaL_pond	**-45.464**	**9.350**	**-4.863**	**<0.001**
Ka2-Ka2_pond	**-16.248**	**3.414**	**-4.759**	**<0.001**
KaL_pond-Ka2_pond	**29.216**	**10.207**	**2.862**	**0.022**

Table S6 Analyses of variance tables and results of Tukey-Kramer *post hoc* pairwise comparisons of means of linear models on splenosomatic index for all the wild populations (model 1) and Kalambo river system comparison between the wild and pond setup (model 2). TPL – total parasite load, SL – standard length. Significant effects ($p < 0.05$) are highlighted in bold.

model 1: wild

Effect	d.f.	*F*	p
population	**4**	**16.38**	**<0.001**
TPL	**1**	**8.30**	**0.005**
SL	1	2.95	0.090
sex	2	0.44	0.645
Residuals	83		

contrast	estimate	SE	*t*	p
Ka1-KaL	-0.694	0.1680	-4.136	**<0.001**
Ka2-KaL	-0.535	0.1570	-3.399	**0.009**
LzL-KaL	0.022	0.1410	0.154	1.000
Lz1-KaL	-0.582	0.1500	-3.889	**0.002**
Ka2-Ka1	0.159	0.1710	0.929	0.884
LzL-Ka1	0.716	0.1750	4.095	**<0.001**
Lz1-Ka1	0.112	0.1600	0.703	0.955
LzL-Ka2	0.557	0.1710	3.254	**0.014**
Lz1-Ka2	-0.047	0.1610	-0.290	0.998
Lz1-LzL	-0.603	0.1550	-3.883	**0.002**

model 2: wild *vs.* ponds

Effect	d.f.	*F*	p
location	**3**	**7.00**	**<0.001**
TPL	1	0.67	0.417
SL	1	0.03	0.854
sex	1	0.69	0.409
Residuals	59		

contrast	estimate	SE	*t*	p
KaL-Ka2	0.570	0.187	3.047	**0.016**
Ka2_pond-Ka2	0.691	0.223	3.102	**0.014**
KaL_pond-Ka2	0.153	0.341	0.449	0.967
Ka2_pond-KaL	0.121	0.204	0.592	0.929
KaL_pond-KaL	-0.417	0.289	-1.442	0.461
KaL_pond-Ka2_pond	-0.537	0.267	-2.015	0.183

Online tables

Table S7 Genes differentially expressed between lake and river samples in the gills (a) and in the spleen (b).

Table S8 WGCNA gene module memberships in the gills (a) and in the spleen (b).

Table S9 Habitat-specific genetic variation identified between lake and river samples.

Table S10 Genetically assimilated genes. List of genes that are differentially expressed in lake-river condition, in the wild and in the pond

Table S11 Plastic gene expression. List of genes that are differentially expressed in lake-river condition, in the wild but not in the ponds.

Chapter 4

Genomics of speciation along an ecological gradient in cichlid fishes

A.A.T. Weber, J. Rajkov, K. Smailus, B. Egger, W. Salzburger

In preparation

All authors were involved in fieldwork. I generated and analysed body shape data. I conducted the molecular laboratory work together with AATW and participated in writing the manuscript.

Genomics of speciation along an ecological gradient in cichlid fishes

Alexandra Anh-Thu Weber[1], Jelena Rajkov[1], Bernd Egger[1], Walter Salzburger[1]

[1]Zoological Institute, University of Basel, Vesalgasse 1, CH-4051 Basel, Switzerland

Abstract

The speciation process has occupied scientists for a long time; however, the dynamics of speciation has been little studied so far. Here we investigated the dynamics of diversification in a set of nine population pairs belonging to four East African cichlid species that have all diverged along a lake-stream environmental gradient and that cover the whole speciation continuum. The analysis of 204 genomes revealed a clear separation in genome-wide divergence between the one-species ($F_{ST} \leq 0.3$) and the two-species ($F_{ST} \geq 0.6$) categories, except for two intermediate population pairs from two sympatric species. Contrastingly, the inspection of morphology revealed a gradual pattern of differentiation. The levels of genome-wide divergence between *Astatotilapia burtoni* populations were consistent with previously reported and newly measured degrees of reproductive isolation between the populations. Furthermore, we found that divergent selection accelerates genome-wide differentiation in the early stages of diversification. Finally, the overall levels of parallelism between lake-stream comparisons were low at both the genomic and morphological levels, but were elevated when ancestral populations were genetically and morphologically more similar. Our results suggest to no longer consider speciation as a unidimensional continuum but rather as a complex product of diversification trajectories through multivariate space and time.

Introduction

The formation of new species – speciation – is a fundamental and omnipresent evolutionary process that has attracted much interest since Darwin's seminal book from 1859[1]; still, relatively little is known about the dynamics of this process[2,3]. Speciation can occur 'suddenly' or gradually, along with the evolution of reproductive isolation between diverging populations[2,4,5]. Sudden speciation is possible, for example, via hybridisation[6,7], polyploidization[8], or when a new mutation (e.g. a chromosomal inversion) directly leads to reproductive isolation[9]. Typically, however, speciation has been considered a continuous process, during which genetic and phenotypic differences accumulate gradually between the diverging populations and reproductive barriers become stronger until complete reproductive isolation is reached[5,10]. In the genic view of speciation, a small set of genes under divergent natural (or sexual) selection becomes resistant to gene flow at the initial stages of diversification, creating 'genomic islands' of strong differentiation; as the populations diverge, genetic differentiation expands across the genome, leading to stronger and more genome-wide patterns of differentiation[5,11,12].

Recently, the gradual nature of the speciation process has been challenged[2]. For example, in a simulation study exploring genome-wide differentiation under different scenarios of speciation, Flaxman et al.[3] showed that – in the presence of high levels of migration, divergent selection and

linkage – there can also be a 'sudden' transition from a state of well-intermixed populations to two reproductively isolated entities. This is because of a positive feedback loop between the levels of divergent selection and linkage disequilibrium[2,3,13], a mechanism referred to as 'genome-wide congealing' (GWC). Empirical support for this model-based prediction came from *Timema* stick insects, in which a gap in genome-wide differentiation was found between sympatric host-plant-associated populations of the same species versus pairs of different species[14]. However, further empirical testing of this model has proven difficult for the main reasons that most study systems (*i*) do not have enough internal replicates in nature[2]; that is, they do not contain multiple independent diversification events triggered by similar environmental factors[15]; and/or (*ii*) do not extend across the entire speciation continuum[2]; that is, they do not contain populations of the same species that – when compared in pairs – cover the full spectrum of (genomic) differentiation ranging from the 'one-species category' (fully or partly interbreeding) to the 'two-species category' (reproductively isolated)[16]. Importantly, the dynamics of morphological differentiation along the speciation continuum has rarely been investigated.

Here we examine genomic and morphological diversification along an environmental gradient in haplochromine cichlid fishes inhabiting African Lake Tanganyika and inflowing or adjacent rivers and streams. More specifically, we (*i*) assess the dynamics of genomic and morphological differentiations along the speciation continuum using nine replicated lake-stream population pairs featuring varying degrees of genomic and morphological differentiation; (*ii*) evaluate to what extent genome-wide differentiation scales with reproductive isolation; and (*iii*) quantify genomic and morphological (non-)parallelism[17] among the nine replicated lake-stream populations pairs.

Study system and approach

Burton's cichlid (*Astatotilapia burtoni*, Günther 1893) (Fig. 1a) occurs in African Lake Tanganyika and affluent rivers and is a member of the Haplochromini, the by far most species-rich tribe of cichlids[18,19]. *Astatotilapia burtoni* is an important model organism in behavioural biology[20,21], developmental biology[22], and sex determination research[23–25], and was among the first five cichlid species to have their genomes sequenced[26]. It has previously been established that many tributaries of Lake Tanganyika – be they small creeks or larger rivers – contain *A. burtoni* populations derived from lake fish, thereby forming 'population pairs' consisting of a source (that is, ancestral) population in the lake and a phenotypically distinct river population featuring habitat-specific adaptations in morphology and ecology[27,28]. More specifically, lake fish have a more upward mouth position, longer gill rakers and more elongated lower pharyngeal jaw bones, while stream fish are more elongated and have shorter gill rakers[27]. It has further been shown that different lake-stream population pairs display varying levels of genetic differentiation, and that the populations from the north and south of Lake Tanganyika are genetically clearly distinct[27–30].

To extend the comparative framework of this study, we also included lake-stream population pairs of three additional species of the Haplochromini/Tropheini clade[19], namely *Haplochromis stappersii*, *Ctenochromis horei,* and *Pseudocrenilabrus philander* (Fig. 1a). Two of these species co-occur with *A. burtoni* in two of the largest tributaries to Lake Tanganyika (*H. stappersii* in the Rusizi River in the north and *C. horei* in the Lufubu River in the south), whereas *P. philander* occurs outside the Tanganyika basin[31] (the here investigated population pair was sampled at Lake Chila and the Mbulu creek, located about 25 km south of Lake Tanganyika; Fig. 1b).

We inspected the genomes of 204 specimens (132 *A. burtoni* and 24 of each of the three additional haplochromine species) from nine lake-stream population pairs (Fig. 1, Extended Data Fig. 1, Supplementary Tables 1 and 2; see Methods). Each population pair consisted of one lake and one stream population, and was named after the respective river, except for the Lake Chila

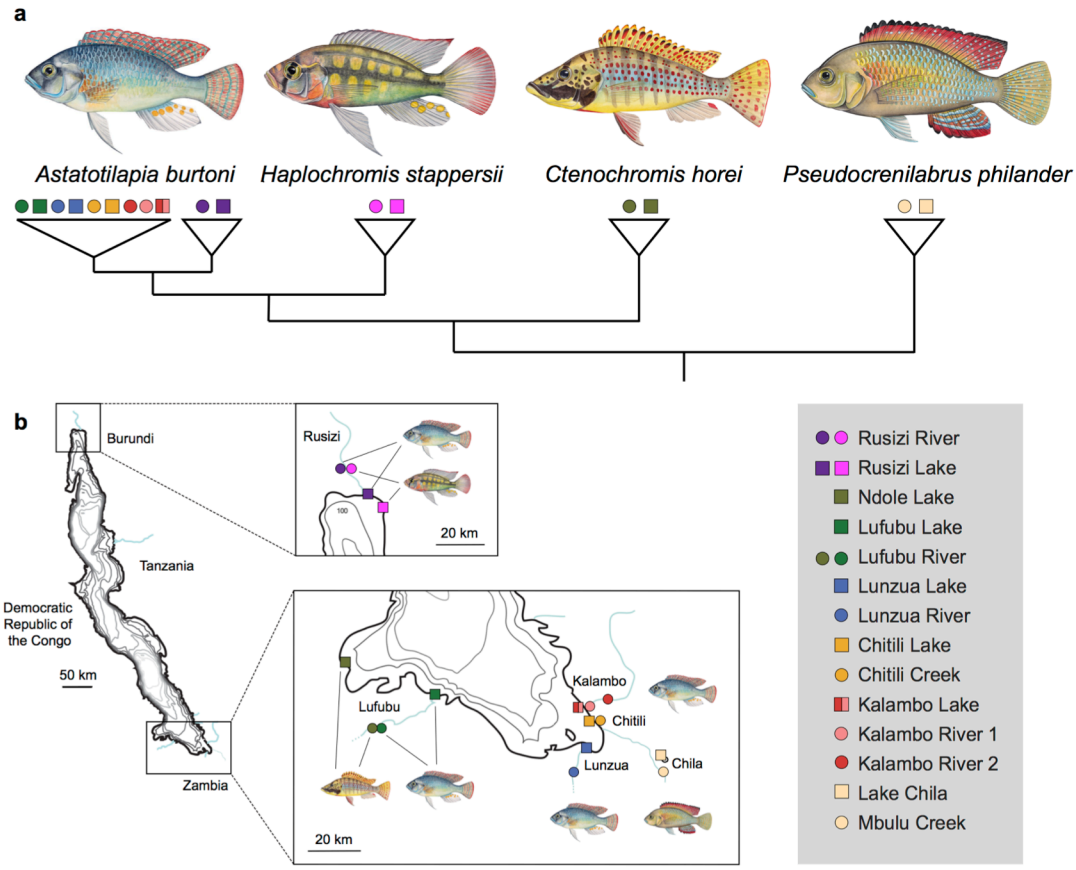


Figure 1 The study system comprising nine lake-stream population pairs in four cichlid fish species. **a**, Illustrations of the four species used in this study and a schematic representation of their phylogenetic relationships (see Extended Data Fig. 1 and [19]). **b**, Map of sampling localities and names of the different lake-stream population pairs, referred to as "systems" here.

system (Fig. 1b). Note that from the Kalambo drainage we sampled two ecologically distinct river populations – one from a locality near the estuary where the river is deep and flows slowly (Kalambo 1) and the other one from ~6 km upstream in a white-water environment (Kalambo 2) – resulting in two population pairs for this system[27]. In addition, we quantified body shape of 468 specimens covering all populations (Supplementary Table 2). To evaluate the degree of premating reproductive isolation between the genetically most distinct *A. burtoni* populations, we furthermore performed mate-choice experiments in the laboratory.

Results and discussion

The dynamics of diversification along an ecological gradient

When investigating speciation continuum dynamics, the relative distribution of differentiation across the diverging populations is informative[2,14]. Accordingly, we assessed the levels of genomic (F_{ST}) and morphological (Mahalanobis distance, D_M) differentiation between the six *A. burtoni* lake-stream population pairs and those of the three other haplochromine species. Genome-wide divergence was heterogeneous among the lake-stream population pairs in *A. burtoni*, ranging from F_{ST} = 0.005 (Chitili) to F_{ST} = 0.465 (Lufubu) (Fig. 2). Divergence between lake-stream population

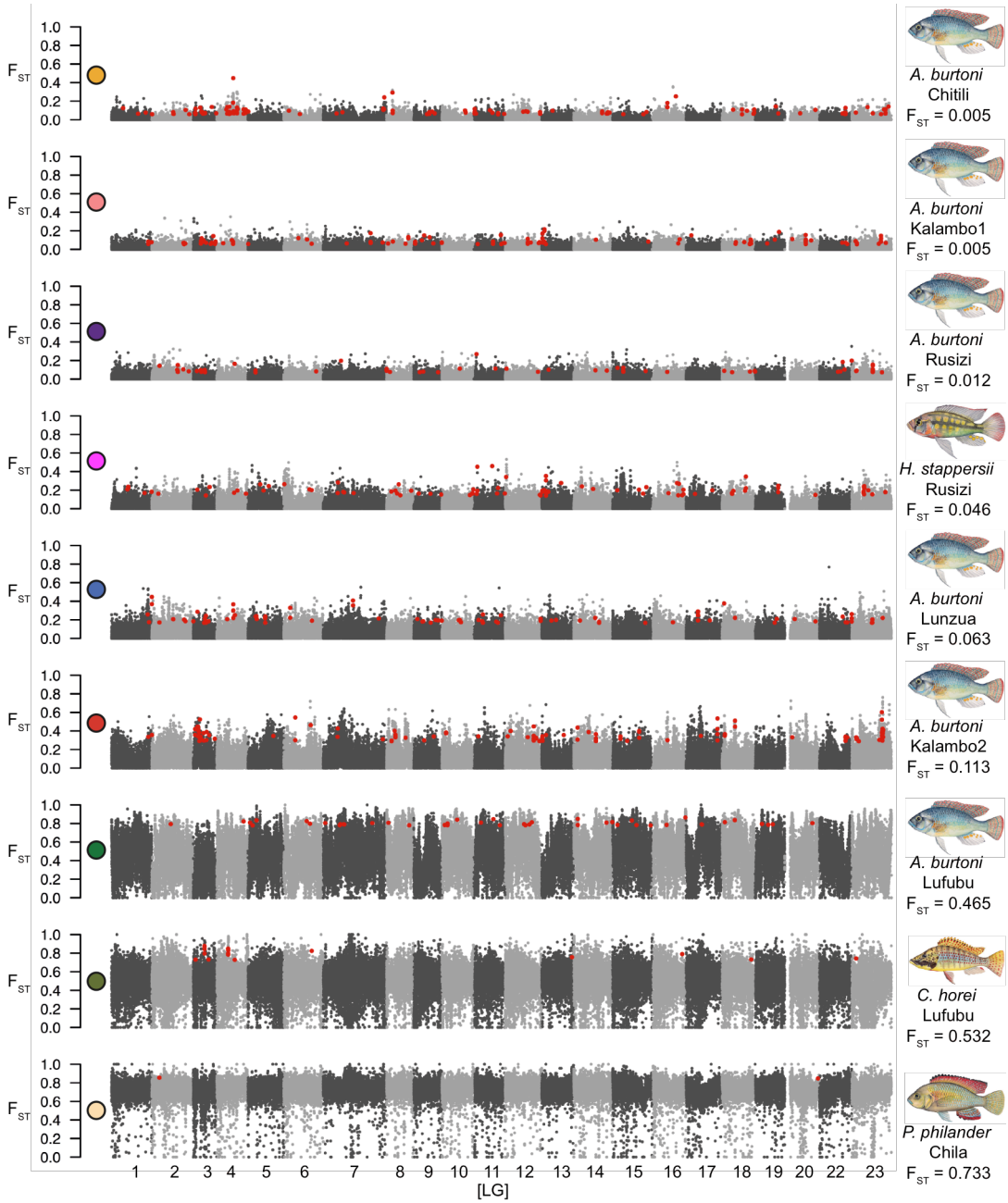


Figure 2 Distribution of genome-wide F_{ST} for the nine lake-stream systems sorted by increasing genome-wide F_{ST}-value. Each dot represents an F_{ST}-value calculated in a 10-kb window along each linkage group. Linkage groups are highlighted in different shades of grey. Regions of differentiation (overlap of the top 5% values of F_{ST}, d_{XY} and π) are highlighted in red.

pairs in the other three haplochromines ranged from low (*H. stappersii*; F_{ST} = 0.046) to intermediate (*C. horei*; F_{ST} = 0.532) to high (*P. philander*; F_{ST} = 0.733), corroborating that the *P. philander* populations may actually represent two distinct species[31]. In *A. burtoni*, differentiation in body shape between lake and stream population pairs ranged from D_M = 2.8 (Chitili) to D_M = 5.5 (Kalambo 2). In agreement with the measured levels of genomic differentiation, morphological differentiation was high in the *P. philander* population pair (D_M = 6.0).

When ranking all nine lake-stream population pairs as well as the sympatric between-species comparisons according to genome-wide F_{ST}-values (Fig. 3a), the within-species lake-stream

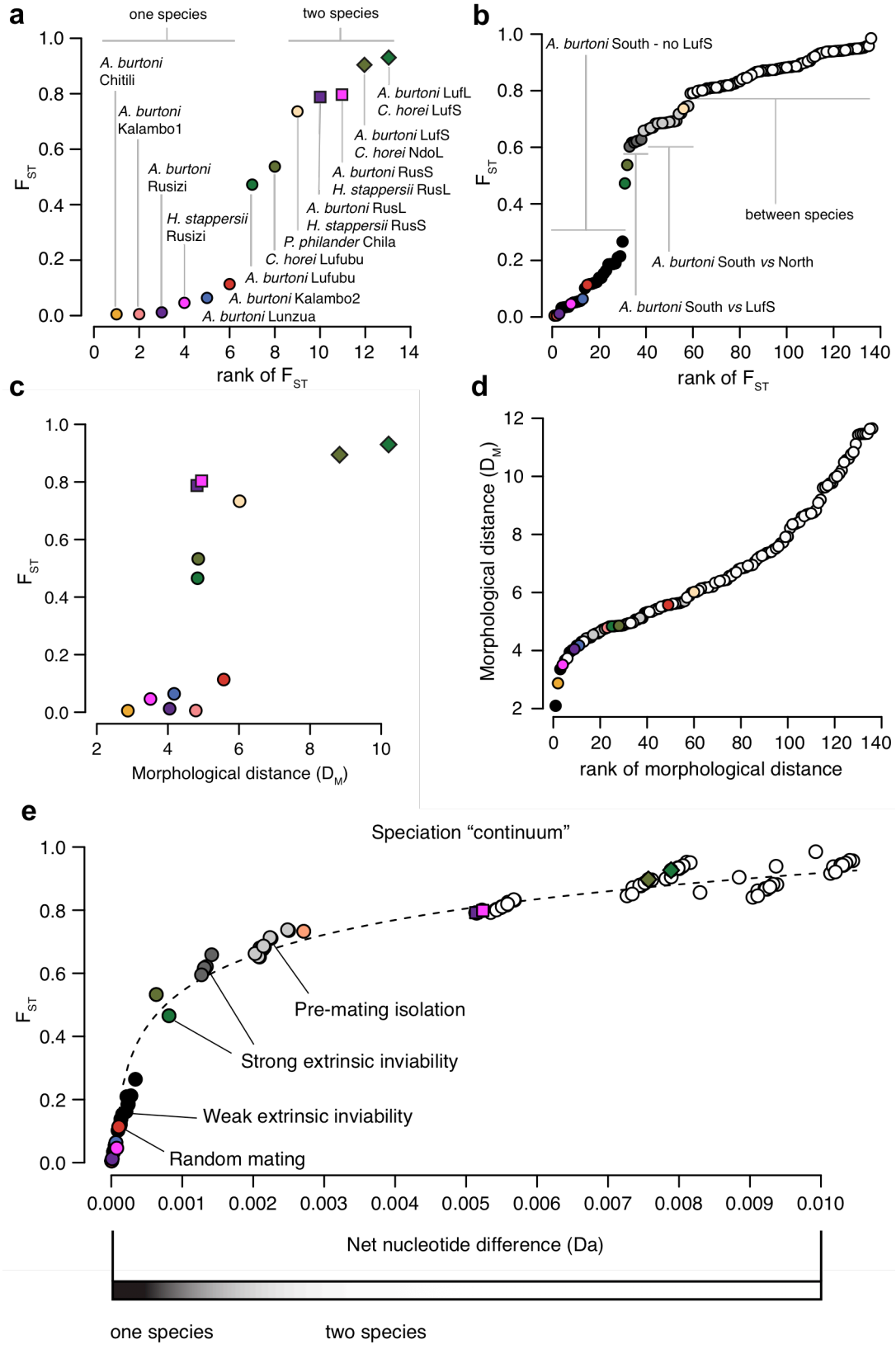


Figure 3 Legend on next page.

Figure 3 Dynamics of the speciation continuum in haplochromine cichlids. **a**, Genomic differentiation based on genome-wide F_{ST} for nine lake-stream population pairs (circles) and four sympatric species pairs (squares and diamonds) sorted by increasing value. The colour coding for lake-stream systems is the same as in Fig. 1. RusR: Rusizi River; RusL: Rusizi Lake; NdoL: Ndole Lake; LufR: Lufubu River; LufL: Lufubu Lake. **b**, Genomic differentiation based on genome-wide F_{ST} for 136 lake-stream population and species pairs including all possible pairwise comparisons, sorted by increasing values, showing a tipping point between the one-species (within *A. burtoni* South – no Lufubu River) and the two-species (*A. burtoni* South versus Lufubu River; *A. burtoni* South versus *A. burtoni* North) categories. **c**, Morphological distances (Mahalanobis distances calculated from body shape) between lake and stream populations of each system are plotted against genetic distances (genome-wide F_{ST}). Morphological distances increase gradually whereas a tipping point is observed for genetic distances. **d**, Morphological differentiation based on Mahalanobis distances for 136 lake-stream population and species pairs including all possible pairwise comparisons, sorted by increasing values. Morphological distances, contrary to genomic distances, increase gradually. **e**, Genome-wide F_{ST} plotted against the net nucleotide difference ($D_a = d_{XY} - (\pi_1+\pi_2)/2$, a proxy for time since divergence, in 136 lake-stream population and species pairs, representing the speciation continuum. A logarithmic regression was fitted to the data (dashed line; $0.162 \times \ln(x)+1.66$; $R^2 = 0.96$). Genomic differentiation accumulates fast during early stages of divergence ($F_{ST} < 0.3$; that is, the one-species category; indicated in black) but then slows down as D_a increases ($F_{ST} > 0.6$; that is, the two-species category, in grey and white). The sympatric population pairs from the Lufubu system are intermediate (F_{ST} *A. burtoni*: 0.46; F_{ST} *C. horei*: 0.53). Different levels of reproductive isolation between populations are reported, when available (see Table 1 for details on the experiments and populations used).

population pairs (the 'one-species category', Fig. 3a) are clearly separated from the between-species comparisons (the 'two-species category', Fig. 3a, including *P. philander*), with two exceptions: The sympatric population pairs from the Lufubu River (*A. burtoni* and *C. horei*) are right in the middle of the distribution, suggesting that these occupy an intermediate position between the one-species and the two-species categories in the speciation continuum. A comparison across all population and species pairs (136 pairwise comparisons) revealed a tipping point in genome-wide differentiation, with two clearly separated categories in F_{ST}-values (≤ 0.3 and ≥ 0.6; again with the exception of the population pairs of the two Lufubu River species; Fig. 3b). Our results in cichlids are, thus, similar to what has been observed in *Timema* stick insects, in which comparable levels of genomic differentiation were found along an ecologically-driven speciation continuum (F_{ST} < 0.3 in within-species population comparisons versus $F_{ST} > 0.7$ in between-species comparisons), except that no intermediates were found in *Timema*[14].

The extent of body shape differentiation of the different lake-stream population pairs was only partially in agreement with the respective levels of genome-wide divergence (Fig. 3c). For example, the population pair from the Kalambo River involving the upstream population (*A. burtoni* Kalambo 2: $D_M = 5.5$) was morphologically more distinct than the comparisons in the Rusizi River (*A. burtoni* versus *H. stappersii*: $D_M = 3.7$) and the sympatric population pairs from the Lufubu River (*A. burtoni*: $D_M = 4.8$; *C. horei*: $D_M = 4.9$). Nevertheless, the two Lufubu River population pairs also occupy intermediate positions with respect to body shape differentiation. In contrast to the levels of genome-wide differentiation, the ranking of all investigated population and species pairs according to D_M did not reveal a tipping point (Fig. 3d). Together, this illustrates that body shape differentiation along the lake-stream environmental gradient is somewhat decoupled from genome-wide differentiation, suggesting that it is more the environment and less so the genetic background that drives overall body shape differentiation in haplochromine cichlids (Fig. 3d, Extended Data Fig. 2).

Next, we investigated the increase in genomic differentiation against a proxy of time since population divergence (net nucleotide difference, D_a). We found a rapid increase in F_{ST} at low levels of D_a, which slowed down as D_a increased further. Given this trend, we fitted a logarithmic model to the data (Fig. 3e) and calculated the slope of the tangent line of the logarithmic regression at different intervals of D_a (0.001) to describe how fast genome-wide differentiation accumulates over time. For D_a-values lower than 0.001 (that is, within the one-species category), there was a sharp increase in F_{ST} as revealed by a ~90% slope decrease (D_a [0.0001-0.001]). Then, in the D_a interval

0.001-0.002, pairwise F_{ST}-values accumulated more slowly (~50% slope decrease), which became even slower for the following intervals of D_a (0.002-0.003: ~33%; 0.003-0.004: ~25%; 0.004-0.005: ~20%), which corresponded to the two-species category (Fig. 3e). Similar results were obtained when using *AFD* (allele frequency difference)[32] instead of F_{ST}, indicating that F_{ST} and *AFD* perform equally well to represent genome-wide differentiation in our dataset (Extended Data Fig. 3).

Overall, our results in cichlids are in agreement with the genome-wide congealing theory[2,3]. Nevertheless, the existence of intermediate stages of genomic differentiation (the two population pairs from the Lufubu system) was not predicted by the model of Flaxman et al.[3], nor was this found in stick insects[14]. To test if the lake-stream population pairs follow a model of primary divergence with gene flow (the divergence model used by Flaxman et al.[3]), we compared eight demographic scenarios of divergence for the nine lake-stream population pairs (models with and without gene flow; with and without periods of isolation; Extended Data Fig. 4a). In eight of the nine population pairs, including the two pairs with intermediate levels of divergence, the most likely demographic scenarios included a past secondary contact event (Extended Data Fig. 4b, Supplementary Table 3), while refuting the model of primary divergence with gene flow (for the remaining population pair a demographic scenario of ancient migration was supported). In the light of the history of Lake Tanganyika, which is characterized by lake level fluctuations[33,34], deviations from a simple model of primary divergence with gene flow might be expected. However, the persistence over time of these intermediate stages from the Lufubu River is unknown, as disentangling between transient stages (that is, a collapse in one species or a split in two species) and actual dynamic equilibria (that is, the presence of a hybrid zone) is unfeasible at present.

Reproductive isolation begins establishing at low levels of genome-wide differentiation

As a next step, we assessed to what extent the observed levels of genome-wide differentiation scale with the degree of reproductive isolation between populations in *A. burtoni*. To this end, we revisited available data from previous experiments[27,35,36] and conducted mate-choice experiments in the laboratory between the genetically most distinct populations of *A. burtoni* in the north and the south of Lake Tanganyika[30] (Extended Data Fig. 7, Table 1).

Previous experiments involving populations that feature low levels of genomic divergence (Kalambo Lake versus Kalambo River 2[35]; Kalambo Lake versus Lunzua River[27]) revealed random mating patterns with respect to source population, suggesting a lack of reproductive isolation. In a mesocosm experiment with populations at a slightly higher level of genome-wide differentiation, yet still within the one-species category (Kalambo Lake versus Ndole Lake), weak levels of extrinsic hybrid inviability were found[36]. In contrast, a mesocosm experiment with populations at intermediate levels of genome-wide divergence (Ndole Lake versus Lufubu River, Kalambo Lake versus Lufubu River) – that is, positioned between the one-species and the two-species categories – provided evidence for strong levels of extrinsic hybrid inviability[36]. Finally, our new laboratory-based mate-choice experiments targeting two populations that feature one of the highest genome-wide F_{ST}-values (Kalambo Lake versus Rusizi Lake) revealed signatures of assortative mating with respect to source population – at least in a multi-sensory laboratory setting allowing for a combination of mating cues (Extended Data Fig. 7 and Appendix 1). This suggests that premating reproductive isolation mechanisms are at play between the genetically most distinct *A. burtoni* clades – the northern and southern lineages[30] – at a level of genomic differentiation that is similar to the one typically observed *between* other haplochromine species (Fig. 3e). Taken together, the *A. burtoni* populations from the north and from the south of Lake Tanganyika 'behave' like distinct

Table 1 Summary of experimental testing of reproductive isolation in *A. burtoni.*

Populations	F_{ST}	Genetic markers	Experimental set-up	Main results	Reproductive isolation	Reference
Kalambo Lake x Kalambo River 2	0.113	Whole-genome	Transplant experiment in the lake environment	adaptive phenotypic plasticity	weak for wild-caught adults; not observed for F1	35
Kalambo Lake x Lunzua River	0.123	Whole-genome	Common garden experiment in mesocosms	random mating	not observed	27
Kalambo Lake x Ndole Lake	0.189[1]	Whole-genome[1]	Common garden experiment in mesocosms	Immigrant and extrinsic hybrid inviability (weak)	Immigrant and extrinsic hybrid inviability (weak)	36
Ndole Lake x Lufubu River	0.472[1]	Whole-genome[1]	Common garden experiment in mesocosms	Immigrant and extrinsic hybrid inviability (strong)	Extrinsic postzygotic isolation (strong)	36
Kalambo Lake x Lufubu river	0.624	Whole-genome	Common garden experiment in mesocosms	Immigrant and extrinsic hybrid inviability (strong)	Extrinsic postzygotic isolation (strong)	36
Kalambo Lake x Rusizi Lake	0.693	Whole-genome	Laboratory mate-choice experiment	Partial assortative mating	Premating isolation	This study

[1]Genetic distances inferred from the Lufubu Lake population rather than Ndole Lake population (both populations are geographically close and belong to the same genetic cluster[27]

species, once more illustrating that our sample of lake-stream population pairs in *A. burtoni* covers the entire speciation continuum from the one-species to the two-species category.

We are aware that a comparison across such different kinds of experiments, initially intended to address independent questions[27,35,36], can be problematic, for example because of the different population pairs investigated, the variable durations of the experiments, and the diverging experimental settings (semi-natural versus laboratory). Nevertheless, the joint consideration of earlier and new results allows some general conclusions: Apparently, in *A. burtoni* the building up of reproductive isolation begins at low levels of genome-wide differentiation, probably triggered by extrinsic barriers[36]. This is consistent with a scenario of speciation by divergent selection, in which extrinsic and sexual barriers act before intrinsic barriers evolve[5]. In support of this, intrinsic postzygotic barriers do not seem to play a strong role at the early stages of speciation, as exemplified by the discovery of a natural hybrid between *A. burtoni* and *Astatoreochromis alluaudi* in our dataset (Extended Data Fig. 1). Taken together, previous data and our new experiment suggest that the levels of reproductive isolation scale with the levels of genome-wide divergence in *A. burtoni.*

Divergent selection accelerates genome-wide differentiation

We then assessed the impact of the environment on genome-wide differentiation trajectories. To do so, we performed pairwise comparisons among the more closely-related southern populations of *A. burtoni* (within the one-species category) focusing on within habitat (lake-lake) versus between habitats (lake-stream) comparisons. In both cases, we found an increase in F_{ST} over geographic distance (Fig. 4a), which is compatible with an isolation-by-distance scenario. However, genomic differentiation increased much stronger with respect to geographic distance when populations were compared that inhabit contrasting environments (lake-stream population comparisons; that is, in the presence of divergent selection and drift) than when they inhabit the same environment (lake-lake comparisons; that is, in the presence of drift only) (Fig. 4a).

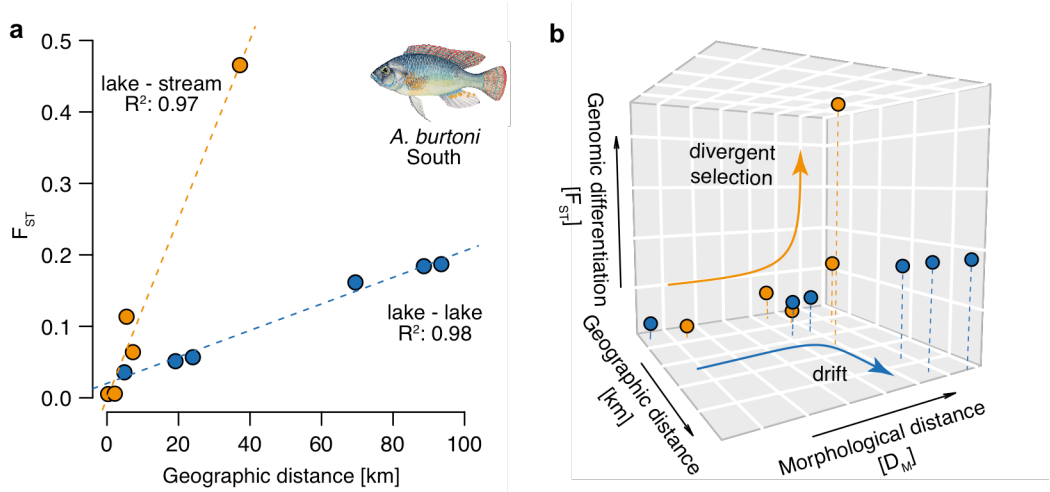


Figure 4 Effect of ecology on genome-wide differentiation trajectories. **a**, Isolation-by-distance in the two comparison categories 'lake-stream' (in orange) and 'lake-lake' (in blue) in the southern populations of *A. burtoni* (that is, within the one-species category). R^2: Pearson's correlation coefficient. **b**, Trajectories for three differentiation axes: morphology, geography and genetics. In "lake-stream" comparisons, morphological differentiation builds up first, then genomic differentiation increases sharply, probably due to the effect of divergent selection at small geographic distance. In contrast, in the absence of divergent selection (that is, in the 'lake-lake' comparisons), morphological differentiation builds up first, then genomic differentiation accumulates only moderately, which is likely due to non-adaptive processes (drift).

Interestingly, the diversification trajectories were similar between lake-stream and lake-lake comparisons at low levels of genetic, morphological and geographic distances (Fig. 4b), but diverged as genomic differentiation built up rapidly compared to geographic distance in the presence of divergent selection (lake-stream comparisons) (Fig. 4b). This corroborates that the environment (via divergent selection) plays a crucial role in differentiation trajectories in *A. burtoni* with the result that population pairs in similar ecological settings display similar levels of genome-wide differentiation. That the environment indeed predicts differentiation trajectories is further substantiated by the finding of very similar levels and patterns of genome-wide and morphological differentiation between population pairs of different haplochromine cichlid species that co-occur in the same river systems and, hence, diversified along the very same environmental gradient (*A. burtoni* and *H. stappersii* in the Rusizi system; *A. burtoni* and *C. horei* in the Lufubu system; Figs. 2, 3), whereby we can rule out gene flow between species as the reason for these similarities between species (Extended Data Figs. 5, 6).

Little overlap between differentiation regions among independent lake-stream systems

We then turned our attention to genomic regions of high differentiation between lake-stream population pairs and to the question to which extent such outlier regions (defined here as the intersection between the top 5% 10-kb windows with respect to F_{ST}, d_{XY}, and absolute value of π difference) are shared between population pairs and species. Our analyses revealed between 2 and 101 outlier regions of high differentiation per lake-stream population pair and that these regions were between 10-70 kb in length (red dots in Fig. 2, Extended Data Fig. 8, Supplementary Table 4). The number of outlier regions reported here is relatively small compared to other studies in cichlids[37,38] or threespine sticklebacks[39], which, however, can be explained by the more stringent definition of such regions in our study (the intersection between three metrics). It has recently been shown that heterogeneity in crossover rates can produce contrasting patterns of genomic differentiation between diverging populations that are not due to divergent selection, manifested, for example, in greater levels of differentiation near chromosome centres (where crossover rates

are low) compared to the peripheries (where crossover rates are high)[40]. We did not find evidence for an accumulation of differentiation regions in the chromosome centres in any of the lake-stream population pairs nor when all 525 outlier regions were considered jointly (Extended Data Fig. 8), suggesting that our results reflect true signatures of divergent selection.

We then evaluated to what extent differentiation regions were shared among lake-stream population pairs and species, and found that there was little overlap (Fig. 5a) and that no such region of high differentiation was shared between more than two systems (Fig. 5a). The regions of high differentiation were distributed across all linkage groups, although there seems to be an overrepresentation on LG3 (64 out of 525), which remained when chromosome length was taken into account (Supplementary Table 4). The 525 differentiation regions included a total of 637 genes. However, there was no obvious overrepresentation in functional enrichment with respect to Gene Ontology categories. The only exception was the Lunzua lake-stream comparison, in which outlier genes were significantly enriched for the molecular function "binding" (Supplementary Table 5). The 25 genes located in the 19 differentiation regions shared between two systems also showed no functional enrichment (Supplementary Table 6).

In situations of a shared evolutionary history of the populations in question, such as in our set-up, frequency-based outlier detection methods may not be the most appropriate way to detect regions under selection[41]. Thus, we also performed Bayesian analyses of selection at the SNP level that take into account population relatedness[42]. Due to their high levels of genome-wide differentiation, the population pairs of *C. horei* and *P. philander* were excluded from these analyses, as signatures of selection and drift cannot easily be disentangled in such cases. For the same reason, we treated the northern and southern populations of *A. burtoni* as separate units. We identified 1,704 shared 10-kb windows of differentiation between *A. burtoni* from the south and *H. stappersii*, 1,683 shared windows between *A. burtoni* from the north and *H. stappersii*, and 1,542 shared windows between *A. burtoni* from the north and from the south (Fig. 5b). In total, 373 windows were shared among the three core sets, containing 366 genes (Fig. 5b; Supplementary Table 7). Interestingly, some genes involved in sensory perception (sound and light) were overrepresented in the common set of outliers between *H. stappersii* and the southern *A. burtoni* populations. Taken together, these results highlight that, although a large majority of outliers are not shared, some functions are important for riverine adaptation and may repeatedly be the target of natural selection. Consistent with this, a recent study[38] identified common highly differentiated genomic regions between a young and an old cichlid species pairs diverging along a depth gradient in Lake Victoria, and found that two thirds of the differentiation regions were private in each species pair. This highlights that adaptive divergence often encompasses both parallel and system-specific (non-parallel) components.

The overall low levels of gene and function sharing between lake-stream systems reported here may be due to cryptic environmental heterogeneity in the stream environment, which encompasses diverse ecological niches, perhaps requiring different mechanisms of adaptation (e.g.[16]). Alternatively, differentiation regions may not be due to divergent selection but due to genetic drift in allopatry and background selection[41], at least for population pairs with elevated levels of genome-wide differentiation[43]. Finally, adaptive phenotypic plasticity[35,44] as well as epigenetic factors[45,46] might play a complementary role in adaptive divergence. Yet, their investigation was beyond the scope of the present study.

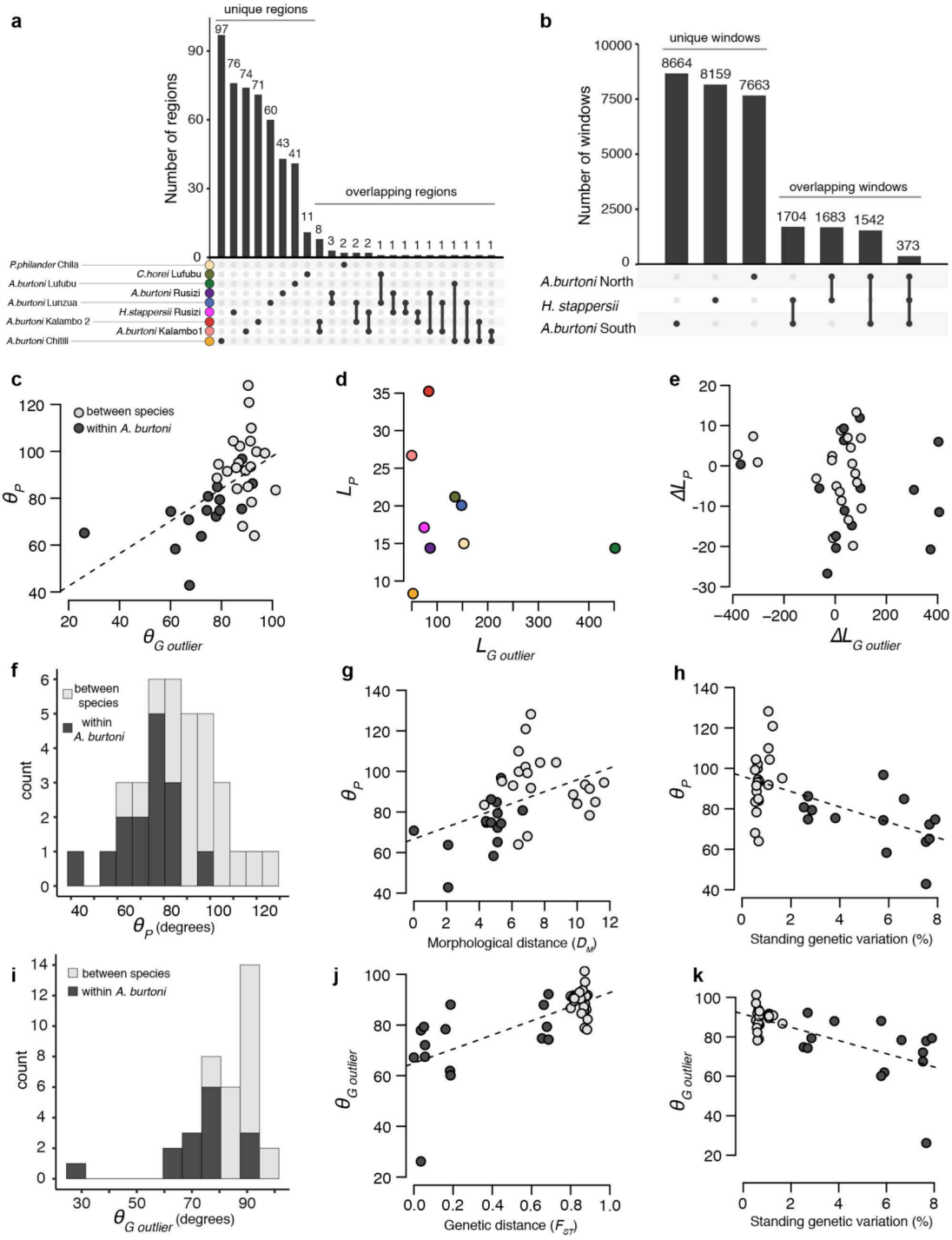


Figure 5 Legend on next page.

Figure 5 (Non-)parallel evolution among nine lake-stream cichlid population pairs. **a**, Number of unique and overlapping differentiation regions (overlap of top 5% values of F_{ST}, d_{XY} and π) in the nine lake-stream population pairs. **b**, Number of unique and overlapping differentiation windows in the three sets of 'core' outliers (core model of differentiation from BayPass) from *A. burtoni* northern populations, *A. burtoni* southern populations and *H. stappersii*. **c**, The angles of phenotypic (θ_P) and genetic outlier ($\theta_{G\ outlier}$) lake-stream divergence vectors are positively correlated (Mantel test: P = 0.0075; Pearson correlation coefficient: R^2 = 0.27). The dashed line indicates significant correlation at the 0.05 level. **d**, The lengths of phenotypic (L_P) and genetic outlier ($L_{G\ outlier}$) lake-stream divergence vectors are not correlated (Linear regression model: P = 0.55). The colour scheme is the same as in panel a. **e**, The differences between phenotypic (ΔL_P) and genetic outlier ($\Delta L_{G\_outlier}$) vector length are not correlated (Mantel test: P = 0.26). **f**, Histogram of the 36 (pairwise) angles between lake-stream phenotypic divergence vectors (θ_P) in degrees. Within *A. burtoni* and between species comparisons are highlighted in different colours. **g**, The angles of phenotypic divergence vectors (θ_P) and morphological distances (D_M) between lake (ancestral) populations are positively correlated (Mantel test: P = 0.045; R^2 = 0.17). For example, θ_P between the Lunzua and Kalambo2 lake-stream systems is plotted against the Mahalanobis distance between Lunzua Lake and Kalambo Lake fish. In other words, if lake populations of two lake-stream systems are more similar morphologically, their direction of phenotypic divergence tends to be more parallel (these have a small θ_P). **h**, The angles of phenotypic divergence vectors (θ_P) and the amount of standing genetic variation between lake populations are negatively correlated (Mantel test P = 0.0061; R^2 = 0.32). **i**, Histogram of the 36 (pairwise) angles between lake-stream genetic outlier divergence vectors ($\theta_{G\ outlier}$) in degrees. **j**, The angles of genetic outlier divergence vectors ($\theta_{G\ outlier}$) and genetic (F_{ST}) distances between lake populations are positively correlated (Mantel test: P = 0.0039; R^2 = 0.46). **k**, The angles of genetic outlier divergence vectors ($\theta_{G\ outlier}$) and the amount of standing genetic variation between lake populations are negatively correlated (Mantel test P = 0.0117; R^2 = 0.43).

Morphological and genomic parallelisms decrease when ancestral populations are more divergent

We finally aimed to assess the extent of phenotypic and genomic (non-)parallelism among the lake-stream population pairs. It has recently been proposed to consider 'parallel' evolution as continuum ranging from parallel to divergent evolution[17]. To examine this in our set of haplochromine cichlids, we performed vector analyses[17,47–49] using the morphological (37 traits representing body shape) and genomic (outlier and non-outlier SNPs) data at hand (outlier SNPs: BayPass outliers, based on 78 principal components from genomic PCA, potentially impacted by natural selection; non-outlier SNPs: all SNPs but excluding BayPass outliers). Following Stuart et al.[16] we quantified variation in lake-stream divergence by calculating vectors of phenotypic and genomic differentiation for each lake-stream population pair, whereby the length of a vector (L) represents the magnitude of lake-stream divergence, and the angle between two vectors (θ) informs about the directionality of divergence. Accordingly, two lake-stream systems are more 'parallel' if θ is small (similar direction of divergence) and ΔL (difference in length between two vectors) is close to zero (similar magnitude of divergence)[17].

Most comparisons between independent lake-stream population pairs fell into the category 'non-parallel', with close to orthogonal (~90°) angles of differentiation in both phenotype (θ_P) and genotype ($\theta_{G\_outlier}$ and θ_G) (Fig. 5f,i, Extended Data Fig. 9d). However, parallelism was higher when only the within *A. burtoni* comparisons were considered, with values of θ_P and $\theta_{G\_outlier}$ between 70° and 80° in many cases (Fig. 5f, i), but not for θ_G (Extended Data Fig. 9d). These values are similar to what has been previously reported in lake-stream population pairs of threespine sticklebacks[16,50]. More clear signatures of parallelism were found in closely related *A. burtoni* lake-stream systems, with the Lunzua-Kalambo1 pair being the most 'parallel' system at the phenotype level (θ_P: 42°) and the Chitili-Kalambo1 pair the most 'parallel' at the genotype level ($\theta_{G\_outlier}$: 26°). Overall, the directions of phenotypic (θ_P) and genetic ($\theta_{G\_outlier}$) differentiations were significantly correlated (Mantel test: P = 0.0075; Pearson correlation coefficient: R^2 = 0.27; Fig. 5c), whereas their magnitudes were not (Mantel test ΔL_P and $\Delta L_{G\_outlier}$: P = 0.26; Linear regression model L_P and $L_{G\_outlier}$: P = 0.55; Fig. 5d, e). In contrast, none of the non-outlier genetic vectors were correlated to phenotypic vectors (θ_P versus θ_G and ΔL_P versus ΔL_G: Mantel tests: P = 0.34; P = 0.35; L_P and L_G:

Linear regression: P = 0.93; Extended Data Fig. 9a-c). This strongly suggests that adaptive phenotypic divergence has a genetic basis in haplochromine cichlids, which is similar to what has been reported for lake-stream population pairs of threespine stickleback fish[16].

It has previously been suggested that the probability of parallelisms at the molecular level decreases with time since divergence[51–53]. We thus examined whether the degree of similarity in the ancestral lake populations correlates with the extent of genetic and morphological parallelisms. We found a significant correlation in both datasets with a stronger effect in the genetic than in the morphological data (Mantel tests: $\theta_{G\_outlier}$ versus $F_{ST:}$ P = 0.0039; θ_P versus D_M: P = 0.045; R^2 = 0.46 and 0.17, respectively) (Fig. 5g, j), indicating that phenotypic and genetic parallelisms decrease when ancestral populations are phenotypically and genetically more divergent. The genetic non-outliers (that is, the neutral markers) did not reveal such a correlation (Mantel test: θ_G versus $F_{ST:}$ P = 0.48; Extended Data Fig. 9e). Finally, we found that the proportion of standing genetic variation between ancestral populations was negatively correlated with θ_P and $\theta_{G\_outlier}$ (Mantel tests: P = 0.0061 and 0.0117; R^2 = 0.32 and 0.43, respectively) (Fig. 5h, k) but not with θ_G (Mantel test: P = 0.19; Extended Data Fig. 9f). In other words, lake-stream population pairs sharing small amounts of standing genetic variation display less parallelism at the level of both the phenotype and the genotype. Parallelism is, thus, likely constrained by the amount of standing genetic variation upon which natural selection can act, as the effect of *de novo* beneficial mutations on parallel evolution is much less likely to play an important role in adaptive divergence in recently diverged population pairs[54]. Previous theoretical work has further shown that even small differences in the directionality of selection can greatly reduce genetic parallelism, especially in the case of complex organisms with many traits[55]. This suggests that, besides time since divergence, also cryptic habitat heterogeneity (leading to small differences in the directionality of selection) can decrease the likelihood of parallelism. In support of this, a recent comparison of regional (within Vancouver Island) versus global (North America versus Europe) lake-stream population pairs of sticklebacks showed that parallelism decreases at increased spatial scales[50].

Parallel and non-parallel components in adaptive divergence have been reported in other systems, such as threespine sticklebacks[16,39,50], Bahamas mosquitofish[56], European whitefish[57] and Trinidadian guppies[58] (see[59] for a review). The comparably low levels of parallelism detected among the nine lake-stream population pairs of haplochromine cichlid fishes might be due to the large amount of standing genetic variation present in cichlids[26], the involvement of complex traits in cichlid adaptive divergence, and/or a large (cryptic) complexity of the habitats to which the diverging cichlid populations adapt to in replicate settings. Perhaps it is the ability of cichlids to always ‘come up’ with what is locally the best adaptive solution (resulting in many cases in morphological and genetic non-parallelism) that ultimately led to their unparalleled species-richness.

Conclusion

We examined nine cichlid population pairs that have diverged along a lake-stream environmental gradient at and around African Lake Tanganyika. We found that, as predicted by simulated data, differentiation at the level of genomes is sudden during adaptive divergence. However, we also uncovered intermediate stages between the one-species and the two-species categories. These intermediates may be explained by periods of allopatry, violating the primary divergence model. Furthermore, we showed that while there was little parallelism at the gene level among lake-stream systems, the directions of phenotypic and genetic divergences were correlated, suggesting that population pairs have adapted to cryptic environmental heterogeneity. A binary ‘lake-stream’

categorisation might thus be an oversimplification. Overall, our examination of lake-stream population pairs revealed that adaptive divergence is a complex process in cichlids.

Methods

Sampling, DNA extraction and sequencing

Individuals of *Astatotilapia burtoni* (N = 132), *Haplochromis stappersii* (N = 24), *Ctenochromis horei* (N = 24) and *Pseudocrenilabrus philander* (N = 24) were collected in Zambia and Burundi between January and November 2015 (Tables S1, S2). Fish were sampled in six different river systems, whereby each system comprises a riverine population (N = 10-12) and a lake population (N = 12-14) and was named after the river, except for the Lake Chila system that was sampled outside of the Lake Tanganyika basin. *H. stappersii* were sampled at the Rusizi River, in the north of Lake Tanganyika, along with sympatric *A. burtoni* populations (Fig. 1; Supplementary Table 1). All other populations were sampled in the south of Lake Tanganyika. *A. burtoni* and *C. horei* were sampled at the Lufubu River; *A. burtoni* was further sampled in the Lunzua, Chitili and Kalambo rivers (Fig. 1; Supplementary Table 1). As two river populations were sampled in the Kalambo River, two lake-stream population comparisons were used for this river, namely Kalambo1 (comparison Kalambo Lake versus Kalambo1) and Kalambo2 (comparison Kalambo Lake versus Kalambo2). Finally, *P. philander* were sampled in small Lake Chila and in Mbulu creek (Fig. 1; Supplementary Table 1). All fish were caught with fishing rods or minnow traps and anaesthetised using clove oil. Photographs of the left lateral side were taken using a Nikon D5000 digital camera, under standardised lighting conditions, and with a ruler for scale. To aid in digital landmark placement, three metal clips were used to spread the fins at the anterior insertions of the dorsal and anal fin, and at the insertion of the pectoral fin (Extended Data Fig. 2a). To increase the sample size for morphological analyses, additional individuals were sampled and photographed at the same locations and time points as the individuals whose genomes were sequenced. Standard length, total length, and weight were measured. A piece of fin clip was preserved in 99% ethanol for DNA extraction. Whole specimens were preserved in 70% ethanol. DNA was extracted from fin clips using the EZNA Tissue DNA Kit (Omega Bio-Tek) following the manufacturer's instructions. Individual genomic libraries were prepared using TruSeq DNA PCR-free Low Sample Kit (Illumina) and subsequently sequenced (150 bp paired-end) on an Illumina HiSeq3000 sequencer at the Genomics Facility Basel (GFB) operated jointly by the ETH Zurich Department of Biosystems Science and Engineering (D-BSSE) and the University of Basel.

Raw data processing, read alignment, variant calling, filtering and phasing:

For each library, the quality of raw reads was visually inspected using FastQC[60] (v0.11.3) and Illumina adapters were trimmed using Trimmomatic[61] (v0.36). Filtered reads of each individual were aligned separately against the *Metriaclima zebra* reference genome (assembly M_zebra_UMD1[62]). We chose this reference genome rather than the *Astatotilapia burtoni* reference genome[26] to avoid any reference bias when comparing *A. burtoni* with the other species. We also chose *M. zebra* rather than *Oreochromis niloticus* as reference genome to maximise the number of reads mapped, as *M. zebra* is phylogenetically closer to *A. burtoni, C. horei, H. stappersii*, and *P. philander* than *O. niloticus*[63].

The *M. zebra* reference genome was indexed using BWA[64] (v.0.7.12) and alignments were performed using BWA-mem with default parameters. Obtained alignments in SAM format were coordinate-sorted and indexed using SAMtools[65] (v.1.3.1). The average coverage per individual ranged from 9.8× to 24.5× (Supplementary Table 1). We performed an indel realignment using

RealignerTargetCreator and IndelRealigner of the Genome Analysis Tool Kit (GATK)[66] (v3.4.0). Variants were called using the GATK functions HaplotypeCaller (per individual and per scaffold), GenotypeGVCFs (per scaffold), and CatVariants (to merge all VCF files). The VCF file corresponding to the mitochondrial genome (scaffold CM003499.1) was then isolated from the VCF file corresponding to the nuclear genome (that is, all other scaffolds). The latter was annotated with the features ExcessHet (that is, the Phred-scaled p-value for an exact test of excess heterozygosity) and FisherStrand (that is, the Strand bias estimated using Fisher's exact test) using the GATK function VariantAnnotator. To filter the VCF file, empirical distributions of depth (DP) and quality (QUAL) were examined. The VCF file was filtered using the GATK function VariantFiltration with the following values (variants meeting the criteria were excluded): DP<2000; DP>4000; QUAL<600; FisherStrand>20; ExcessHet>20.

In addition, variants were called using SAMtools mpileup (per scaffold) with the following options: -C50 -pm2 -F0.2 -q 10 -Q 15. Files per scaffold were then converted to VCF format, concatenated (except the mitochondrial genome) and indexed using BCFtools (v.1.3.1). The VCF file was annotated for ExcessHet and FisherStrand, and the distribution of depth and quality were visually assessed as described above. The VCF file was filtered using the GATK function VariantFiltration with the following values: DP<1500; DP>4000; QUAL<210; FisherStrand>20; ExcessHet>20. Filtered GATK and SAMtools datasets were then combined using bcftools isec. The final VCF file contained variants present in both datasets. Genotypes were then imputed and phased per scaffold using beagle[67] (v.4.0). In total, the final VCF file contained 26,704,097 variants. Chomonomer (v.1.05; http://catchenlab.life.illinois.edu/chromonomer/) was used to place the 3,555 *M. zebra* scaffolds in 22 linkage groups using two linkage maps[68,69]. For BayPass selection analyses and allele frequency calculations (see below), we excluded indels and non-biallelic sites, resulting in a VCF file containing 20,343,366 SNPs.

Genetic structure and phylogenetic relationships

Population genetic structure was examined using principal components analyses implemented in the smartPCA module of Eigensoft[70] (v.6.1.1). To reconstruct a whole-genome nuclear phylogeny, a sequence corresponding to the first haplotype of each scaffold was extracted using bcftools consensus --haplotype 1 of BCFtools v1.5 (https://github.com/samtools/bcftools). Individual whole genome sequences were then concatenated and a maximum-likelihood (ML) analysis was performed in RAxML[71] (v.8.2.11) using the GTRGAMMA sequence evolution model and 20 fast bootstrap replicates. The option *–f a* was used to report the best-scoring ML tree with branch lengths. KBC4, a putative *A. burtoni* individual sampled in the Rusizi River, did not cluster with other *A. burtoni* individuals in the phylogeny (labelled “hybrid” in Extended Data Fig. 1a). This specimen results most likely from a hybridisation event with *Astatoreochromis alluaudi,* as its mitochondrial genome is closely related to *A. alluaudi* (data not shown). Therefore, this individual was excluded from further analyses. In order to test for introgression or retention of ancestral polymorphism between sympatric species (that is, *A. burtoni* and *H. stappersii* in the Rusizi system, and *A. burtoni* and *C. horei* in the Lufubu system), a topology weighting analysis reconstructing fixed-length 5-kb-windows phylogenies was performed using *Twisst* (topology weighting by iterative sampling of subtrees[72]).

Demographic modeling

For each of the nine lake-stream population pairs, demographic simulations based on the joint site frequency spectrum (SFS) were performed in order to estimate the most likely model of population divergence as well as the best values of demographic parameters (effective population sizes, divergence times, migration rates). Simulated SFS were obtained using diffusion approximation

implemented in ∂a∂i[73] (v.1.7.0). A modified version of the program including additional predefined models and the calculation of the Akaike Information Criterion (AIC) for model selection was used for the simulations[74]. Eight demographic models of population divergence were tested (Extended Data Fig. 4a): Bottle-Growth (BG), Strict Isolation (SI), Isolation with Migration (IM), Ancient Migration (AM), Secondary Contact (SC), as well as versions of these models including two categories of migration rates (IM2M; AM2M; SC2M). These two categories of migration rates can separate, for example, selected versus neutral loci. For the population pairs with a genome-wide F_{ST} > 0.47 (*A. burtoni* Lufubu, *C. horei* Lufubu, and *P. philander*), the model BG was not tested as it was obvious that the populations are separated. Each model was fitted to the observed joint SFS using three successive optimisation steps: "hot" simulated annealing, "cold" simulated annealing, and BFGS[74]. For each lake-stream population pair, 20 replicate analyses comparing seven or eight models were run, using different parameter starting values to ensure convergence. After these 20 runs, the model displaying the lowest AIC and the least variance among the replicates was chosen as the best model. For parameter estimation, additional runs were performed so that the total number of runs was 20 for the best model. To calculate the divergence times in years, a generation time of one year was used. As the scaled mutation rate parameter θ is estimated, we used the relation $\theta = 4 \times Ne \times \mu \times L$ to infer the ancestral effective population size (*Ne*). The mutation rate (μ) (3.51×10^{-9} mutation per generation per year) of Lake Malawi cichlids[75] and the length of the genome assembly (L) of *M. zebra* (UMD1: 859,842,111 bp) were used.

Regions of differentiation

For each lake-stream system, genome-wide F_{ST} (Hudson's estimator of F_{ST} [76]), $|\ \pi_{lake} - \pi_{stream}\ |$ (absolute value of the difference in nucleotide diversity between the lake and the river populations), and d_{XY} (absolute divergence) were calculated on 10 kb non-overlapping sliding windows using evo (https://github.com/millanek/evo). We define as window of differentiation each window that is contained in the overlap of the top five percent values of these three metrics. Adjacent differentiation windows and windows separated by 10 kb were then merged in differentiation regions. To test if the differentiation regions of each system are affected by chromosome centre-biased differentiation (CCBD[40]), each region was placed either in the "centre" or in the "periphery" categories. These categories were defined by splitting each chromosome into four parts of equal length, where the "centre" category encompasses the two central parts of the chromosome and the "periphery" category encompasses the two external parts of the chromosome.

Bayesian selection and association analyses

To detect signatures of selection at the SNP level, we used the Bayesian method BayPass[42]. The core model performs a genome scan for differentiation by estimating a population covariance matrix of allele frequencies. It allows determining outlier SNPs based on the top 1% of simulated *XtX* values, where *XtX* is a differentiation measure corresponding to a SNP-specific F_{ST} corrected for the scaled covariance of population allele frequencies[42]. As the northern and southern populations of *A. burtoni* are highly divergent (see[30] and Extended Data Fig. 1), only the southern *A. burtoni* populations were analysed jointly. We thus compared outlier sets for the southern populations of *A. burtoni*, the northern populations of *A. burtoni,* and *H. stappersii*. For the southern *A. burtoni* populations, an additional association analysis using one categorical covariate (lake population versus stream population) was performed using the auxiliary variable covariate model (AUX). Five replicate runs were performed using different starting seed values and default search parameters, except for the number of pilot runs (25). The final correlated SNP set contained the overlap of SNPs for which the Bayes Factor (BF) was higher than 10 and which were in the top

five percent of δ values (the posterior probability of association of the SNP with the covariable) in the five replicate runs.

GO annotation and enrichment analyses

To infer if candidate genes in differentiation regions were enriched for a particular function, all genes included in differentiation regions of each lake-stream system were extracted. In addition, genes including overlapping SNPs between the southern *A. burtoni,* the northern *A. burtoni*, and *H. stappersii* core outliers were reported, as well as genes including overlapping SNPs between *A. burtoni* core outliers and SNPs significantly correlated with lake versus stream environment and morphology. All candidate genes were blasted (blastx) against the NR database (version 12.10.2017) using BLAST+ v.2.6.0 and the first 50 hits were reported. To obtain a reference gene set, all *M. zebra* genes were blasted against NR in the same way. Gene Ontology and InterProScan annotations were retrieved from Blast2GO PRO[77] (v.4.1.9). Enrichment analyses were performed using Fisher's exact test for each differentiated gene set (one set per system) against the reference gene set (significance level: 0.001). For the genes located in the overlap of differentiation regions among systems, an additional step was performed by manually retrieving the annotations from *Homo sapiens* dataset in Uniprot (accessed online 20.11.2017).

Speciation dynamics and genomic divergence in similar versus contrasted environments

To infer the dynamics of genomic differentiation along the lake-stream axis, genome-wide pairwise F_{ST} and the net nucleotide difference D_a (proxy of time since differentiation; $d_{XY} - (\pi_1+\pi_2 / 2)$[78]) were calculated for the nine lake-stream population pairs (within system) as well as for four between-species comparisons from the same river system (*A. burtoni* Rusizi Lake versus *H. stappersii* Rusizi River; *A. burtoni* Rusizi River versus *H. stappersii* Rusizi Lake; *A. burtoni* Lufubu Lake versus *C. horei* Lufubu River; *A. burtoni* Lufubu River versus *C. horei* Ndole Lake) resulting in 13 within-system comparisons. Furthermore, genome-wide pairwise F_{ST}- and D_a-values were calculated for all possible population pair combinations, resulting in 136 within and between systems comparisons. A logarithmic regression was fitted to the data using the lm function ($lm(F_{ST} \sim ln(D_a)$ implemented in R[79] (v.3.4.2). For comparative purposes, SNP-based *AFD* (allele frequency difference; $| AF_1 - AF_2 |$) was calculated for each possible population pair using allele frequencies reported by the command --freq of VCFtools (v.0.1.14), with a filter for minor allele frequency of 0.2.

To estimate the influence of divergent selection at early stages of genomic differentiation in sympatry/parapatry, population pairwise F_{ST} of the southern populations of *A. burtoni* were used ('lake-lake': 6 comparisons; 'lake-stream': 5 comparisons), as well as the pairwise Mahalanobis distances (D_M; see below). The northern *A. burtoni* populations were not used due to the high levels of genomic divergence compared to the southern populations. For each comparison, the geographic coastline distance between populations was measured using Google Earth (https://www.google.com/intl/en/earth/). Then, a linear model was fitted for each category ('lake-lake' and 'lake-stream') and the adjusted coefficient of determination R^2 was reported in R.

Morphological analyses

Geometric morphometrics was used to compare adult body shape between populations[80]. Three juvenile individuals of *P. philander* from Mbulu creek whose genomes had been sequenced were excluded from the morphological analyses. In total, the photographs of 468 individuals (Supplementary Table 2) were used for geometric morphometric analyses (289 *A. burtoni*, 81 *H. stappersii*, 67 *C. horei*, and 31 *P. philander)*. Using TPSDIG2[81] (v.2.26) we placed 17 homologous landmarks (Extended Data Fig. 2a) on the lateral image of each fish. The tps file with x and y

coordinates was used as an input for the program MORPHOJ[82] (v.1.06d) and superimposed with a Procrustes generalized least squares fit (GLSF) algorithm to remove all non-shape variation[83]. Canonical variate analysis (CVA)[84] was used to assess shape variation among *A. burtoni* populations (Extended Data Fig. 2 b, d) and among all populations of the four species (Extended Data Fig. 2 c, e). The mean shape distances of CVA were obtained using permutation tests (10,000 permutations). Mahalanobis distances (D_M) among groups from the CVA were calculated and sorted by increasing value. Morphological distances were then plotted against the rank of these values.

Vectors of phenotypic and genomic divergence

We followed the method first developed by Adams and Collyer, and Collyer et al.[47–49] and described in detail in Stuart et al.[16] and Bolnick et al.[17] to calculate multivariate vectors of phenotypic and genomic divergence. For vectors of morphological divergence, 37 traits were used: centroid size (17 landmarks), standard length (landmarks 1-14), body depth (landmarks 8-12) corrected by standard length (ratio BD/SL), and the x and y coordinates of each of the 17 landmarks. We then calculated two types of vectors of genomic divergence. First, 'outlier' vectors were calculated using the first 78 principal components (88% variance; same as for non-outliers) of a genomic principal components analysis (PCA) based on outliers SNPs present in at least one set of BayPass outliers (*H. stappersii* core model; *A. burtoni* North core model; *A. burtoni* South core model; *A. burtoni* South auxiliary model). Then, 'neutral' vectors were calculated using the first 78 principal components of a genomic PCA based on all remaining SNPs (that is, non-outliers) (88% variance; all PC summarising between-population variation). For each lake-stream pair and separately for phenotypic and genomic outlier and non-outlier data, we calculated vector length (L), the difference in length between two vectors (ΔL) and the angle in degrees between two vectors (θ). For each morphological trait (respectively each genomic PC), we ran t-tests and used the t-statistic as an estimate of lake-stream divergence for each trait. Thus, vectors of phenotypic divergence were represented as a data frame of 37 t-statistics for each trait × 9 lake-stream pairs, and vectors of genomic divergence were represented as a data frame of 78 t-statistics for each PC × 9 lake-stream pairs. We calculated 9 L_P (phenotype), 9 L_G (genotype), 9 $L_{G\_outlier}$ values, all lake-stream pairwise comparisons ΔL_P, θ_P, ΔL_G, θ_G, $\Delta L_{G\_outlier}$ and $\theta_{G\_outlier}$. We used Mantel tests (mantel.test function in the ape[85] R package v.5.3; 9,999 permutations) to test the correlation between: θ_P and θ_G / $\theta_{G\_outlier}$; ΔL_P and ΔL_G / $\Delta L_{G\_outlier}$ and linear regression models (lm function in R) to test the correlation between L_P and L_G / $L_{G\_outlier}$.

Factors influencing the direction of (non-)parallelism

We then investigated the influence of similarity of ancestral (that is, lake) populations on parallel evolution. We used Mahalanobis distances and F_{ST} between lake populations as proxies of morphological and genetic similarities, respectively. We performed Mantel tests between lake-lake Mahalanobis distances and θ_P, and between lake-lake F_{ST} and θ_G / $\theta_{G\_outlier}$ to infer if similarity of ancestral populations between systems was correlated to the direction of (non-)parallelism. Finally, we assessed if the proportion of standing genetic variation is correlated with the extent of morphological or genetic parallelism. For each of the 136 population pairwise comparison, we extracted allele frequencies of biallelic SNPs using VCFtools --freq command. We then categorised SNPs in four different groups: identical sites (fixed in both populations for the same allele); differentially fixed sites (fixed in both populations for different alleles); fixed and variable sites (fixed in one population and variable in the second population) and standing genetic variation (SGV; variable in both populations). We performed Mantel tests between the proportion of SGV SNPs between lake populations and θ_P, and between the proportion of SGV SNPs and θ_G / $\theta_{G\_outlier}$.

Mate-choice experiments

We used two mate-choice experiments to test for reproductive isolation between the two geographically most distant and genetically most divergent *A. burtoni* populations, Rusizi Lake and Kalambo Lake. Detailed methods and results of the experiments are provided in Appendix 1. Briefly, in the first experiment, a two-way female choice set up was used to test whether females preferred males of their own population over others when only visual cues are available (Extended Data Fig. 7a). Following the method described in[86,87], we placed a gravid female of either population (n = 44) in a central tank and allowed visual contact and interaction with two sized-matched males from Rusizi Lake and Kalambo Lake presented in two outer tanks. Within a period of up to 12 days, we assessed if the female had laid the eggs in front of the Rusizi Lake male, the Kalambo Lake male, in front of both, or in the central section. The position of the laid eggs was used as a measure for female preference (conspecific choice coded as 1; heterospecific and no choice coded as 0). The binomial data were then analysed with a generalized linear mixed model, which tested if the probability of the females spawning with the conspecific male was significantly different from 0.5.

In the second experiment, female spawning decisions were determined in a multi-sensory setting with free contact between females and males. As in[88], a single tank was subdivided into three equally sized compartments by plastic grids (Extended Data Fig. 7b). The middle compartment offered a resting and hiding place for the females whereas the two outer compartments served as male territories. The grid size was chosen to allow the smaller females to migrate between the three compartments, and to prevent direct contact between the larger males to exclude male-male competition. We conducted eight trials, each time using two males (one of each population) and six females from both populations (n_{males} = 16; $n_{females}$ = 46). Mouthbrooding females were caught and ten larvae from each were collected for paternity analyses based on five microsatellite markers. All adult males and females were also genotyped for these five markers. The percentage of fertilised offspring by con- or heterospecific males in each replicate was then used to infer if the females spawned more frequently with the conspecific males. These experiments were performed under the cantonal veterinary permit nr. 2356.

Data availability

The raw sequence reads were deposited on SRA and are available with accession numbers SRP156808 (*A. burtoni, C. horei*, and *H. stappersii*) and SRP148476 (*P. philander*)[89].

Code availability

Key R codes and bash scripts used in this study are available at https://github.com/Anh-Thu-Weber/Speciation_dynamics_cichlids

Acknowledgements

We would like to thank Nicolas Boileau, Adrian Indermaur, Gaëlle Pauquet, and Fabrizia Ronco for their help during fieldwork, the crews of the Kalambo Lodge and the Ndole Bay Lodge for their logistic support in Zambia, and the Lake Tanganyika Research Unit, Department of Fisheries, Republic of Zambia, and the University of Burundi and the Ministère de l'Eau, de l'Environnement, de l'Amenagement du Territoire et de l'Urbanisme, Republic of Burundi, for assistance with the research permits. We thank Athimed El Taher and Milan Malinsky for their help with data analyses, Matthew Conte for sharing the chromonomer results, Astrid Böhne and Michael Matschiner for helpful comments on the manuscript, and Julie Himes for fish illustrations. Finally, we are grateful to the support team of sciCORE (center for scientific computing, University of Basel,

http://scicore.unibas.ch/) for providing access to computational resources, especially to Pablo Escobar Lopez. This study was supported by grants from the Swiss Zoological Society (SZS) to J.R., and the Swiss National Science Foundation (SNSF, grant 31003A_156405) and the European Research Council (ERC, CoG 617585 'CICHLID~X') to W.S. A.A.-T.W. is currently supported by an Endeavour Postdoctoral Fellowship awarded by the Australian Department of Education and Training (Grant Agreement: 6534_2018) and a Marie Sklodowska-Curie Global Fellowship awarded by the Research Executive Agency of the European Commission (Grant Agreement: 797326).

Author contributions

W.S. and B.E. conceived and supervised the study; all co-authors conducted the fieldwork; A.A.-T.W. and J.R. conducted the molecular laboratory work; J.R. generated and analysed the morphometric data; K.S. and B.E. designed, conducted and analysed the mate-choice experiments; A.A.-T.W. analysed the genomic data and drafted the manuscript, with feedback from all co-authors.

References

1. Darwin, C. On the origins of species by means of natural selection. *Lond. Murray* (1859).
2. Nosil, P., Feder, J. L., Flaxman, S. M. & Gompert, Z. Tipping points in the dynamics of speciation. *Nat. Ecol. Evol.* 1, 1–8 (2017).
3. Flaxman, S. M., Wacholder, A. C., Feder, J. L. & Nosil, P. Theoretical models of the influence of genomic architecture on the dynamics of speciation. *Mol. Ecol.* 23, 4074–4088 (2014).
4. Coyne, J. A. & Orr, H. A. *Speciation. Sunderland, MA*. (Sinauer Associates, Inc, 2004).
5. Seehausen, O. et al. Genomics and the origin of species. *Nat. Rev. Genet.* 15, 176–92 (2014).
6. Marques, D. A., Meier, J. I. & Seehausen, O. A combinatorial view on speciation and adaptive radiation. *Trends Ecol. Evol.* (2019).
7. Mallet, J. Hybrid speciation. *Nature* 446, 279 (2007).
8. Wood, T. E. *et al.* The frequency of polyploid speciation in vascular plants. *Proc. Natl. Acad. Sci. USA* 106, 13875–13879 (2009).
9. Kirkpatrick, M. & Barton, N. Chromosome inversions, local adaptation and speciation. *Genetics* 173, 419–434 (2006).
10. De Queiroz, K. Species concepts and species delimitation. *Syst. Biol.* 56, 879–886 (2007).
11. Wu, C.-I. The genic view of the process of speciation. *J. Evol. Biol.* 14, 851–865 (2001).
12. Wolf, J. B. & Ellegren, H. Making sense of genomic islands of differentiation in light of speciation. *Nat. Rev. Genet.* 18, 87–100 (2017).
13. Tittes, S. & Kane, N. C. The genomics of adaptation, divergence and speciation: a congealing theory. *Mol. Ecol.* 23, 3938–3940 (2014).
14. Riesch, R. *et al.* Transitions between phases of genomic differentiation during stick-insect speciation. *Nat. Ecol. Evol.* 1, 0082 (2017).
15. Roux, C. *et al.* Shedding light on the grey zone of speciation along a continuum of genomic divergence. *PLoS Biol.* 14, e2000234 (2016).
16. Stuart, Y. E. *et al.* Contrasting effects of environment and genetics generate a continuum of parallel evolution. *Nat. Ecol. Evol.* 1, 0158 (2017).
17. Bolnick, D. I., Barrett, R. D., Oke, K. B., Rennison, D. J. & Stuart, Y. E. (Non) parallel evolution. *Annu. Rev. Ecol. Evol. Syst.* 49, 303–330 (2018).
18. Salzburger, W. Understanding explosive diversification through cichlid fish genomics. *Nat. Rev. Genet.* 19, 705–717 (2018).
19. Ronco, F. *et al.* Drivers, dynamics, and progression of a massive adaptive radiation in African cichlid fishes. co-*submitted*.
20. Theis, A., Salzburger, W. & Egger, B. The function of anal fin egg-spots in the cichlid fish *Astatotilapia burtoni. PLoS ONE* 7, (2012).
21. Fernald, R. D. & Hirata, N. R. Field study of *Haplochromis burtoni*: Quantitative behavioural observations. *Anim. Behav.* 25, 964–975 (1977).
22. Santos, M. E. *et al.* The evolution of cichlid fish egg-spots is linked with a cis-regulatory change. *Nat. Commun.* 5, 5149 (2014).
23. Böhne, A., Wilson, C. A., Postlethwait, J. H. & Salzburger, W. Variations on a theme: Genomics of sex determination in the cichlid fish *Astatotilapia burtoni. BMC Genomics* 17, 883 (2016).

24. Roberts, N. B. *et al.* Polygenic sex determination in the cichlid fish *Astatotilapia burtoni. BMC Genomics* 17, 835 (2016).

25. El Taher, A., Ronco, F., Matschiner, M., Salzburger, W. & Böhne, A. Dynamics of sex chromosome evolution in a rapid radiation of cichlid fish. *co-submitted*.

26. Brawand, D. *et al.* The genomic substrate for adaptive radiation in African cichlid fish. *Nature* 513, (2014).

27. Theis, A., Ronco, F., Indermaur, A., Salzburger, W. & Egger, B. Adaptive divergence between lake and stream populations of an East African cichlid fish. *Mol. Ecol.* 23, 5304–5322 (2014).

28. Egger, B., Roesti, M., Bohne, A., Roth, O. & Salzburger, W. Demography and genome divergence of lake and stream populations of an East African cichlid fish. *Mol. Ecol.* 26, 5016–5030 (2017).

29. Theis, A. *et al.* Variation of anal fin egg-spots along an environmental gradient in a haplochromine cichlid fish. *Evolution* 71, 766–777 (2017).

30. Pauquet, G., Salzburger, W. & Egger, B. The puzzling phylogeography of the haplochromine cichlid fish *Astatotilapia burtoni*. *Ecol. Evol.* (2018).

31. Egger, B. *et al.* Phylogeographic and phenotypic assessment of a basal haplochromine cichlid fish from Lake Chila, Zambia. *Hydrobiologia* 748, 171–184 (2015).

32. Berner, D. Allele frequency difference AFD–an intuitive alternative to F_{ST} for quantifying genetic population differentiation. *Genes* 10, 308 (2019).

33. Scholz, C. A. *et al.* East African megadroughts between 135 and 75 thousand years ago and bearing on early-modern human origins. *Proc. Natl. Acad. Sci.* 104, 16416–16421 (2007).

34. Salzburger, W., Van Bocxlaer, B. & Cohen, A. S. Ecology and evolution of the African great lakes and their faunas. *Annu. Rev. Ecol. Evol. Syst.* 45, 519–545 (2014).

35. Rajkov, J., Weber, A. A.-T., Salzburger, W. & Egger, B. Adaptive phenotypic plasticity contributes to divergence between lake and river populations of an East African cichlid fish. *Ecol. Evol.* 8, 7323–7333 (2018).

36. Rajkov, J., Weber, A. A.-T., Salzburger, W. & Egger, B. Immigrant and extrinsic hybrid inviability contribute to reproductive isolation between lake and river cichlid ecotypes. *Evolution* 72, 2553–2564 (2018).

37. Malinsky, M. *et al.* Genomic islands of speciation separate cichlid ecomorphs in an East African crater lake. *Science* 350, 1493–1498 (2015).

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

39. Rennison, D. J., Stuart, Y. E., Bolnick, D. I. & Peichel, C. L. Ecological factors and morphological traits are associated with repeated genomic differentiation between lake and stream stickleback. *Philos. Trans. R. Soc. B* 374, 20180241 (2019).

40. Berner, D. & Roesti, M. Genomics of adaptive divergence with chromosome-scale heterogeneity in crossover rate. *Mol. Ecol.* 26, 6351–6369 (2017).

41. Ravinet, M. *et al.* Interpreting the genomic landscape of speciation: a road map for finding barriers to gene flow. *J. Evol. Biol.* 30, 1450–1477 (2017).

42. Gautier, M. Genome-wide scan for adaptive divergence and association with population-specific covariates. *Genetics* 201, 1555–1579 (2015).

43. Burri, R. Interpreting differentiation landscapes in the light of long-term linked selection. *Evol. Lett.* 1, 118–131 (2017).

44. Oke, K. B. *et al.* Does plasticity enhance or dampen phenotypic parallelism? A test with three lake–stream stickleback pairs. *J. Evol. Biol.* 29, 126–143 (2016).

45. Artemov, A. V. *et al.* Genome-wide DNA methylation profiling reveals epigenetic adaptation of stickleback to marine and freshwater conditions. *Mol. Biol. Evol.* 34, 2203–2213 (2017).

46. Hu, Y. & Albertson, R. C. Baby fish working out: an epigenetic source of adaptive variation in the cichlid jaw. *Proc. R. Soc. B Biol. Sci.* 284, 20171018 (2017).

47. Collyer, M. L. & Adams, D. C. Analysis of two-state multivariate phenotypic change in ecological studies. *Ecology* 88, 683–692 (2007).

48. Collyer, M. L., Sekora, D. J. & Adams, D. C. A method for analysis of phenotypic change for phenotypes described by high-dimensional data. *Heredity* 115, 357 (2015).

49. Adams, D. C. & Collyer, M. L. A general framework for the analysis of phenotypic trajectories in evolutionary studies. *Evol. Int. J. Org. Evol.* 63, 1143–1154 (2009).

50. Paccard, A. *et al.* Repeatability of adaptive radiation depends on spatial scale: regional versus global replicates of stickleback in lake versus stream habitats. *J. Hered.* doi:10.1093/jhered/esz056. early online version.

51. Storz, J. F. Causes of molecular convergence and parallelism in protein evolution. *Nat. Rev. Genet.* 17, 239–250 (2016).

52. Bollback, J. P. & Huelsenbeck, J. P. Parallel genetic evolution within and between bacteriophage species of varying degrees of divergence. *Genetics* 181, 225–234 (2009).

53. Conte Gina L., Arnegard Matthew E., Peichel Catherine L. & Schluter Dolph. The probability of genetic parallelism and convergence in natural populations. *Proc. R. Soc. B Biol. Sci.* 279, 5039–5047 (2012).

54. Barrett, R. D. & Schluter, D. Adaptation from standing genetic variation. *Trends Ecol. Evol.* 23, 38–44 (2008).

55. Thompson, K. A., Osmond, M. M. & Schluter, D. Parallel genetic evolution and speciation from standing variation. *Evol. Lett.* 3, 129–141 (2019).

56. Langerhans, R. B. Predictability and parallelism of multitrait adaptation. *J. Hered.* 109, 59–70 (2018).

57. Siwertsson, A., Knudsen, R., Adams, C. E., Præbel, K. & Amundsen, P.-A. Parallel and non-parallel morphological divergence among foraging specialists in European whitefish (*Coregonus lavaretus*). *Ecol. Evol.* 3, 1590–1602 (2013).

58. Fitzpatrick, S. W., Torres-Dowdall, J., Reznick, D. N., Ghalambor, C. K. & Chris Funk, W. Parallelism isn't perfect: could disease and flooding drive a life-history anomaly in Trinidadian guppies? *Am. Nat.* 183, 290–300 (2014).

59. Oke, K. B., Rolshausen, G., LeBlond, C. & Hendry, A. P. How parallel is parallel evolution? A comparative analysis in fishes. *Am. Nat.* 190, 1–16 (2017).

60. Andrews, S. FastQC: a quality control tool for high throughput sequence data. (2010).

61. Bolger, A. M., Lohse, M. & Usadel, B. Trimmomatic: A flexible trimmer for Illumina sequence data. *Bioinformatics* 30, 2114–2120 (2014).

62. Conte, M. A. & Kocher, T. D. An improved genome reference for the African cichlid, *Metriaclima zebra*. *BMC Genomics* 16, 724 (2015).

63. Meier, J. I. *et al.* Ancient hybridization fuels rapid cichlid fish adaptive radiations. *Nat. Commun.* 8, 14363 (2017).

64. Li, H. & Durbin, R. Fast and accurate short read alignment with Burrows–Wheeler transform. *Bioinformatics* 25, 1754–1760 (2009).

65. Li, H. *et al.* The sequence alignment/map format and SAMtools. *Bioinformatics* 25, 2078–2079 (2009).

66. McKenna, A. *et al.* The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res.* 20, 1297–1303 (2010).

67. Browning, S. R. & Browning, B. L. Rapid and accurate haplotype phasing and missing-data inference for whole-genome association studies by use of localized haplotype clustering. *Am. J. Hum. Genet.* 81, 1084–1097 (2007).

68. Albertson, R. C. *et al.* Genetic basis of continuous variation in the levels and modular inheritance of pigmentation in cichlid fishes. *Mol. Ecol.* 23, 5135–5150 (2014).

69. T O'Quin, C., Drilea, A. C., Conte, M. A. & Kocher, T. D. Mapping of pigmentation QTL on an anchored genome assembly of the cichlid fish, *Metriaclima zebra*. *BMC Genomics* 14, 287 (2013).

70. Price, A. L. *et al.* Principal components analysis corrects for stratification in genome-wide association studies. *Nat. Genet.* 38, 904 (2006).

71. Stamatakis, A. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* 30, 1312–1313 (2014).

72. Martin, S. H. & Van Belleghem, S. M. Exploring evolutionary relationships across the genome using topology weighting. *Genetics* 206, 429–438 (2017).

73. Gutenkunst, R. N., Hernandez, R. D., Williamson, S. H. & Bustamante, C. D. Inferring the joint demographic history of multiple populations from multidimensional SNP frequency data. *PLoS Genet.* 5, e1000695 (2009).

74. Tine, M. *et al.* European sea bass genome and its variation provide insights into adaptation to euryhalinity and speciation. *Nat. Commun.* 5, 5770 (2014).

75. Malinsky, M. *et al.* Whole-genome sequences of Malawi cichlids reveal multiple radiations interconnected by gene flow. *Nat. Ecol. Evol.* 2, 1940–1955 (2018).

76. Bhatia, G., Patterson, N., Sankararaman, S. & Price, A. L. Estimating and interpreting FST: the impact of rare variants. *Genome Res.* 23, 1514–1521 (2013).

77. Conesa, A. *et al.* Blast2GO: a universal tool for annotation, visualization and analysis in functional genomics research. *Bioinformatics* 21, 3674–3676 (2005).

78. Nei, M. & Li, W. H. Mathematical model for studying genetic variation in terms of restriction endonucleases. *Proc. Natl. Acad. Sci. USA* 76, 5269–5273 (1979).

79. Team, R. C. *R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. 2016*. (2017).

80. Zelditch, M. L., Swiderski, D. L., Sheets, H. D. & Fink, W. L. *Geometric morphometrics for biologists*. *Elsevier* vol. 59 (Elsevier, 2004).

81. Rohlf, F. J. *Program tpsDIG2 for Windows version 2.12. Departament of Ecology and Evolution, State University of New York, Stony Book*. (2008).

82. Klingenberg, C. P. MorphoJ: an integrated software package for geometric morphometrics. *Mol. Ecol. Res.* 11, 353–357 (2011).

83. Rohlf, F. J. & Slice, D. Extensions of the Procrustes method for the optimal superimposition of landmarks. *Syst. Zool.* 39, 40–59 (1990).

84. Mardia, K. V., Kent, J. T. & Bibby, J. M. *Multivariate analysis*. vol. 23 (Academic Press, New York, 1995).

85. Paradis, E. & Schliep, K. ape 5.0: an environment for modern phylogenetics and evolutionary analyses in R. *Bioinformatics* 35, 526–528 (2018).

86. Theis, A., Salzburger, W. & Egger, B. The function of anal fin egg-spots in the cichlid fish *Astatotilapia burtoni*. *PLoS One* 7, e29878 (2012).

87. Theis, A., Bosia, T., Roth, T., Salzburger, W. & Egger, B. Egg-spot pattern and body size asymmetries influence male aggression in haplochromine cichlid fishes. *Behav. Ecol.* 26, 1512–1519 (2015).

88. Plenderleith, M., Van Oosterhout, C., Robinson, R. L. & Turner, G. F. Female preference for conspecific males based on olfactory cues in a Lake Malawi cichlid fish. *Biol. Lett.* 1, 411–414 (2005).

89. Böhne, A. *et al.* Repeated evolution versus common ancestry: Sex chromosome evolution in the haplochromine cichlid *Pseudocrenilabrus philander*. *Genome Biol. Evol.* 11, 439–458 (2019).

Appendix 1: Detailed methods and results of the mate-choice experiments

Extended methods

We tested for reproductive isolation between two geographically distant and genetically divergent populations: Rusizi Lake and Kalambo Lake. Fish from the south of Lake Tanaganyika, Zambia were imported in 2011 and from the northern part of Lake Tanaganyika, near Bujumbura, Burundi in 2014. Males from each population were kept in separate aquaria (100x50x50 cm), in individual mesh cylinders provided with a half clay pot as a territorial hideout. Females from each population were kept separately in female only tanks (100x50x50 cm). Aquaria were maintained at standardized conditions (24°C and a 12:12 h light:dark cycle). Fish were fed twice a day with flake food and once a day with Artemia. All laboratory mate choice experiments were performed at the Zoological Institute of the University of Basel under the permission of the Cantonal Veterinary Office, Basel, Switzerland (permit numbers: 2536, 26037).

Experiment 1: Visual cues only

A two-way female choice set up[1,2] was used to test whether females preferred males of their own population when only visual cues are available (Fig. S7a). In each experimental round, we placed a gravid female of either population (n = 44) in a central tank (60 x 30 x 30 cm) and allowed visual contact with two males from Rusizi Lake and Kalambo Lake presented in two outer tanks (40 x 25 x 25 cm). The paired males were size matched in standard body length (SL) as precisely as possible (number of male pairs = 8; mean SL difference ± standard deviation [SD] = 0.51 ± 0.82 mm; range 0.0 – 3.41 mm) and introduced at least 24 hours before the start of each experimental round to allow for acclimation and territorial behaviour to develop. In each experimental round, the female was able to see and interact with both males of the stimulus pair and laid eggs within a period of few hours up to 12 days. The experiment was terminated if the female did not lay eggs within this time period. Because of the grid placed in the aquaria, eggs laid by the female would fall into this "egg-trap" before the females were able to take them into their mouth for incubation. The "egg-trap", which completely covered the floor of the female tank, made it possible to assess if the female laid the eggs in front of the Rusizi Lake male, the Kalambo Lake male, or in front of both. The position of the laid eggs was used as a measure for female preference.

After the introduction, female behaviour was recorded using a Sony handicam (HDR-X550VE, 12.0 mega pixels) for one hour and male behaviour was recorded for the same amount of time with GoPro cameras (HERO3 Silver Edition HD3.02.03.00; one per outer tank). The videos were later analysed with QuickTime player (v. 10.0) to assess the amount of time a female spent, and the time of direct activity (interaction time), with each male. For males, the time of activity, which we defined as the time males were not resting or hiding, the frequency of different behaviours i.e. lateral display, quivering, charge, facial bar, clicking[3–5] and the coloration (yellow or blue) were recorded. The side allocated to the male of each population was swapped after every second experimental run to control for directional bias; and females were tested randomly regarding the source population. Because many of the females (89%) laid their eggs exclusively next to one of the males, the data was coded into 1 and 0 to circumvent the problem of zero inflation in statistical analyses. The data was coded as 1 if the conspecific male received more than 50% of the eggs and as 0 if the conspecific male received fewer than 50% of the eggs (there was no case were both males received exactly 50% of the eggs).

The binomial data were then analysed with generalized linear mixed models (GLMMs) with a logistic link function using the package lme4[6] in R[7] (v. 3.1.3). To correct for multiple testing of male pairs, pairs were used as a random factor. Three different models were used: model 1 tested whether the probability of the females spawning with the conspecific male was significantly

different from 0.5, which was indicated by an intercept on the logit scale different from 0. Female spawning decisions were used as a response variable. Model 2 tested first (1) if females a) spent more time, or b) interacted more with the conspecific male relative to the heterospecific male; second (2) if females a) spent more time, or b) interacted more with the male they later chose in comparison to the rejected male, and third (3) if females a) spent more time, or b) interacted more with the more active male. An observation level was included as a random effect to account for the extra variance in the data. Model 3 tested if the final choice of a female depended on a) the time females spent with a respective male, b) the interaction time of females with a respective male, c) the time males were active, d) the frequency of behaviours of males (pooled and separately), e) coloration of males. We further analysed if there were differences in activity of males between populations, or between chosen and non-chosen males (as in Model 2).

Experiment 2: Direct contact

The partial partition method[8] was used to infer female spawning decisions in a multi-sensory setting with free contact between females and males. A single tank (150x50x50 cm) was subdivided into 3 equally sized compartments (30x50x50 cm) by plastic grids (Fig. S7b). The middle compartment offered a resting and hiding place for the females whereas the two outer compartments served as male territories. The grid size was chosen to allow the smaller females to migrate between the three compartments, and to prevent direct contact between the larger males to exclude male-male competition. Additionally, we positioned opaque plates on each half of the grid to prevent visual contact between males. We conducted eight trials, each time using two males and six females (number of males = 16; number of females = 46; in one replicate only 4 females were available due to sample size restrictions). In every trial two males, one of each population, were placed in the opposite compartments, where both males were size matched and acclimation time was at least 24 hours, as in experiment 1. Then a total of 4-6 females (2-3 females of each population) were introduced in the middle compartment. After each time two females had spawned, we switched the position of the males to avoid compartment effects.

Mouthbrooding females were caught and the fry was removed and anesthetized using clove oil before being transferred into ethanol for later DNA extraction. A fin clip of the caudal fin of the female was taken, and after an experimental round was terminated (all females spawned and were removed) a fin clip of the potential males was also taken for later DNA extraction. We recorded fertilization rate by counting the number of fertilized vs. unfertilized eggs. DNA of ten larvae per clutch, their corresponding mothers and the putative fathers was used for paternity testing with microsatellite markers Ppun5, Ppun7, Ppun21[9], UNH130[10] and Abur82[11]. The amplified DNA samples were genotyped on an Applied Biosystems (ABI) 3130xl genetic analyser and sized in comparison to LIZ 500(-250) (ABI) internal size standard. The genotypes were determined manually with the software Peakscanner (v. 1.0) and paternity was determined using the software Cervus[12] (v.3.0) with no mismatch allowed. For statistical analyses we coded the data as in experiment 1 and analysed each of the eight replicates separately, and then pooled together, using binomial tests with a probability of 0.5 and a confidence interval of 0.95, to check if the females spawned more with the conspecific males.

Results

Experiment 1: Visual cues only

In our female two-way choice experiment pairs of size-matched males, one from Rusizi Lake and one from Kalambo Lake, were presented to a focal female in two outer tanks arranged on both sides of the central female tank (Fig. S7a). Out of 44 females, 24 females laid more than 50% of

the eggs with the conspecific male (11 Rusizi Lake females, 13 Kalambo Lake females). Males from neither population were more likely to receive more eggs from females (model 1; GLMM, number of females = 44, number of male pairs = 8, z = 0.06, p = 0.55). Disentangling the results for the two populations also indicated random choice for both Rusizi and Kalambo Lake females (GLMM, number of male pairs = 8; number of females Rusizi Lake = 21, z Rusizi Lake = -0.63, p Rusizi Lake = 0.53; number of females Kalambo Lake = 23, z Kalambo Lake = 1.34, p Kalambo Lake = 0.18; Fig. S7c).

For model 2 we used the time a female spent with either male within the first 30 minutes of the experiment. During this time, females did not spend more time with the conspecific nor with the heterospecific male when analysed pooled or separately (model 2, (1a) GLMM, number of females = 44, number of male pairs = 8, z = 0.84, p = 0.40). Females did not interact more with neither the conspecific nor with the heterospecific male (model 2, (1b) GLMM, number of females = 44, number of male pairs = 8, z = 1.24, p = 0.22). Females showed a trend towards interacting more with the male they later spawned with (model 2 (2b) GLMM, number of females = 44, number of male pairs = 8, z = 1.68, p = 0.09) and females of the Kalambo Lake population interacted more with the chosen male (GLMM, number of females = 23, number of male pairs = 8, z = 2.604, p = 0.009). Nevertheless, there was no significant interaction between the choice of egg-laying and interaction time of females with the chosen male, as shown by the results of Model 3b (GLMM, z = 0.998, p = 0.32). Analyses (3) of Model 2, with all females pooled (GLMM, number of females = 44, number of male pairs = 8, z = 1.94, p = 0.05) suggested that females tended to interact more with males that were more active. When females were analysed separately for each population, Kalambo Lake females interacted more with males that were more active (GLMM, number of females = 23, number of male pairs = 8, z= 3.09, p = 0.002).

The results from model 3 showed that Rusizi Lake females tended to lay their eggs more in front of more active males (GLMM, z = 1.47, p = 0.08); and tended to spawn more often with yellow males (GLMM, z = 1.77, p = 0.08). Kalambo Lake females tended to prefer to lay their eggs with males that exhibit a higher frequency of lateral display (GLLM, z = 1.72, p = 0.08). Kalambo Lake males tended to be more active compared to Rusizi Lake males during the experimental runs with Rusizi Lake females (GLMM, z = 1.83, p = 0.07), and with Kalambo Lake females (GLMM, z = 1.657, p = 0.09). Rusizi Lake males showed yellow coloration more often in comparison to Kalambo Lake males (z = 1.87, p = 0.06). When testing for differences in activity between chosen and non-chosen males, we found that the males chosen by Kalambo Lake females tended to have higher frequencies in overall behaviours (GLMM, z = 1.97, p = 0.05) and higher frequencies in lateral display (GLMM, z = 2.165, p = 0.03).

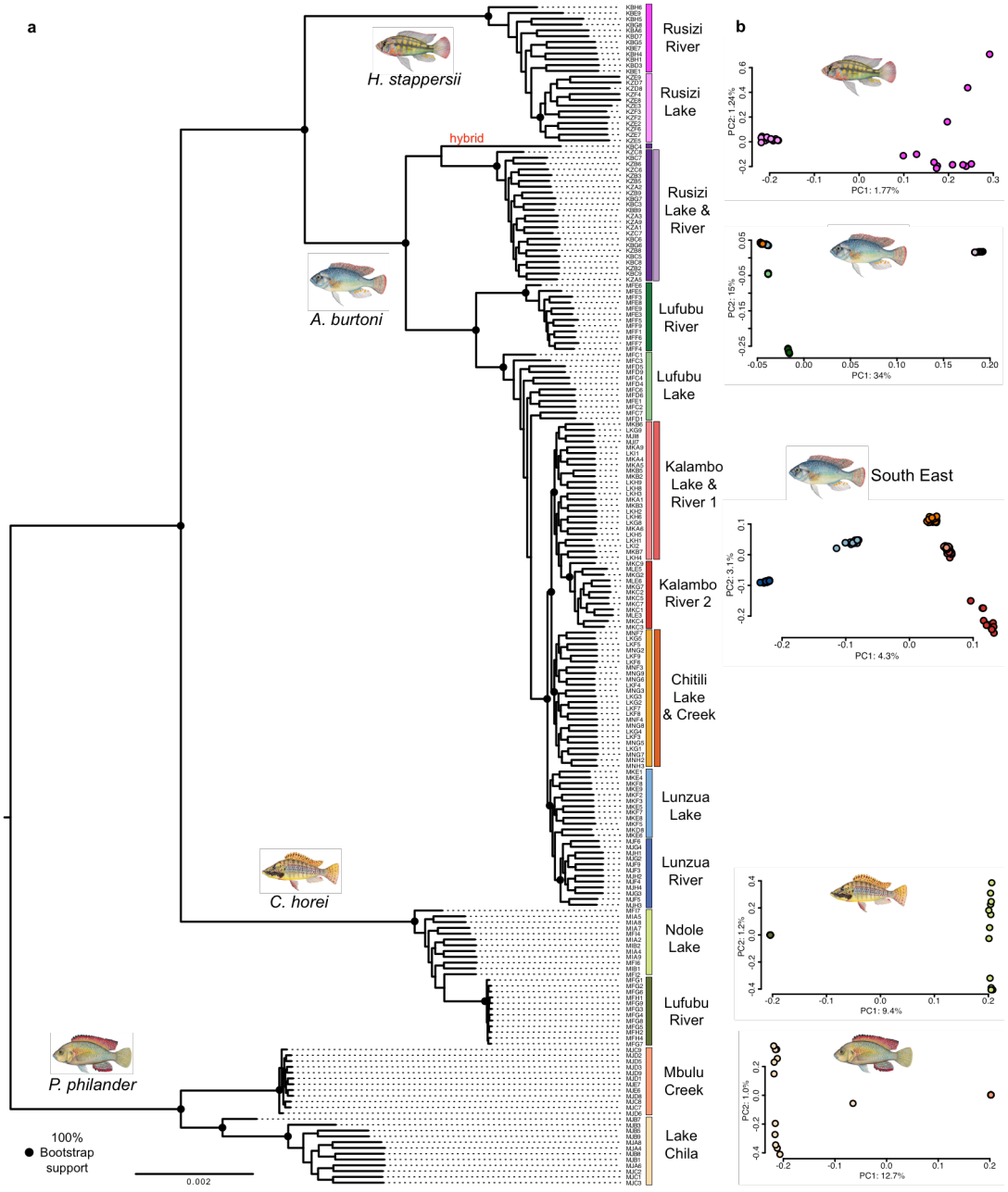
Experiment 2: Direct contact

In the partial partition experimental set up with multiple cues available, females mated significantly more often with a conspecific male (binomial test: n = 46, p < 0.001, Fig. S7d). This result was consistent when each population was analysed separately (binomial test Rusizi Lake: n = 23, p = 0.01; Binomial test Kalambo Lake: n = 23, p = 0.03). In six out of eight replicates, the probabilities of females to spawn with their conspecific male ranged between 0.66 and 1; in one replicate the probability to spawn with the conspecific male was 0.33, and in the replicate where only four females were tested the probability to spawn with the conspecific male was 0.5. Fertilization rate was 100% and each brood was fertilized by a single male. Due to the fact that size matched male pairs were tested with several (4-6) females, we additionally applied a general linear mixed model with pairs as a random factor to control for multiple testing. The random effects were not significant and the results of this test were consistent with the previous ones (GLMM, number of

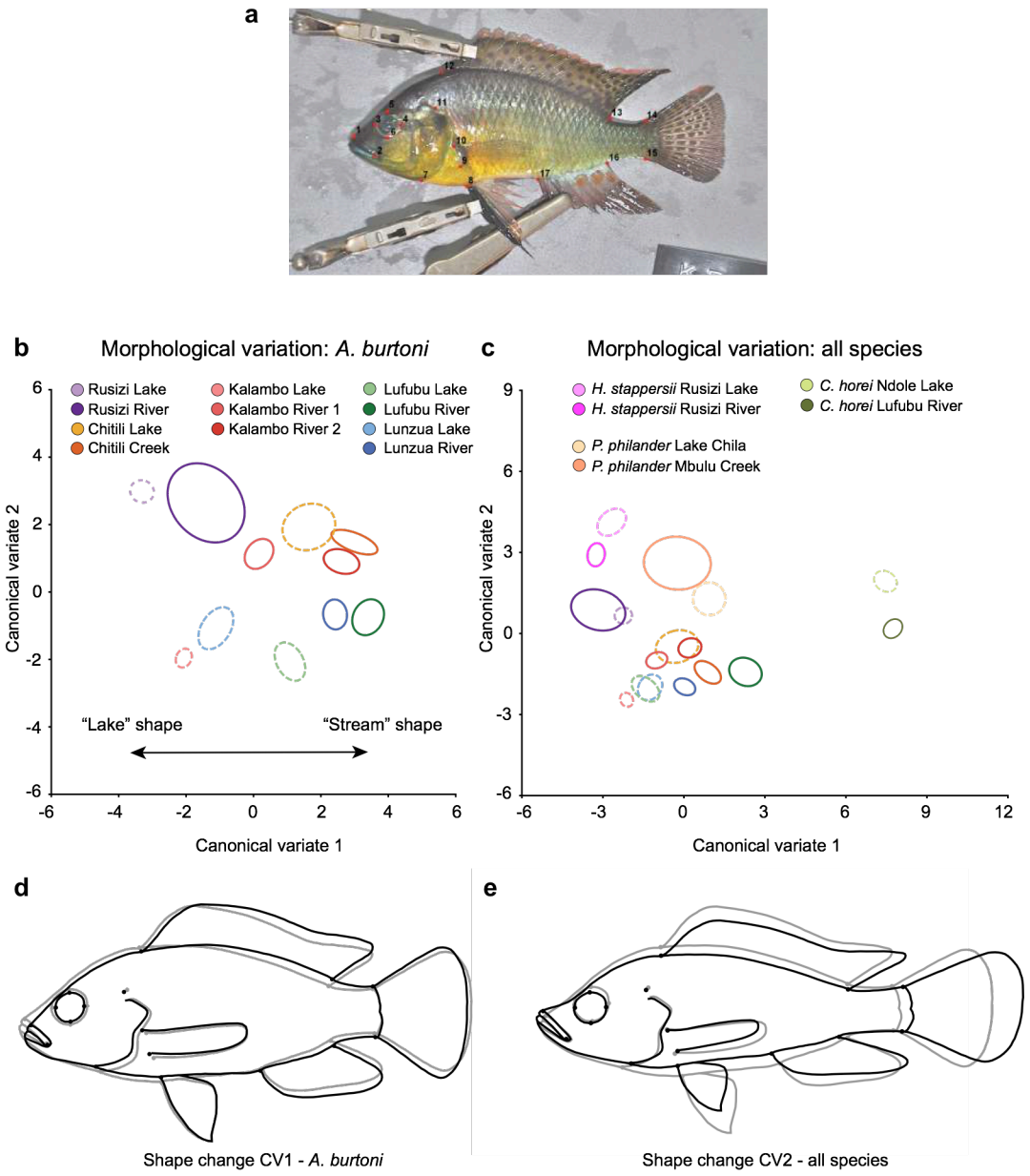
females = 46, number of male pairs = 8, z = 2.617, p = 0.008; Rusizi Lake: number of females = 23, z = 2.534, p = 0.01; Kalambo Lake: number of females = 23, z = 2.212, p = 0.02).

References

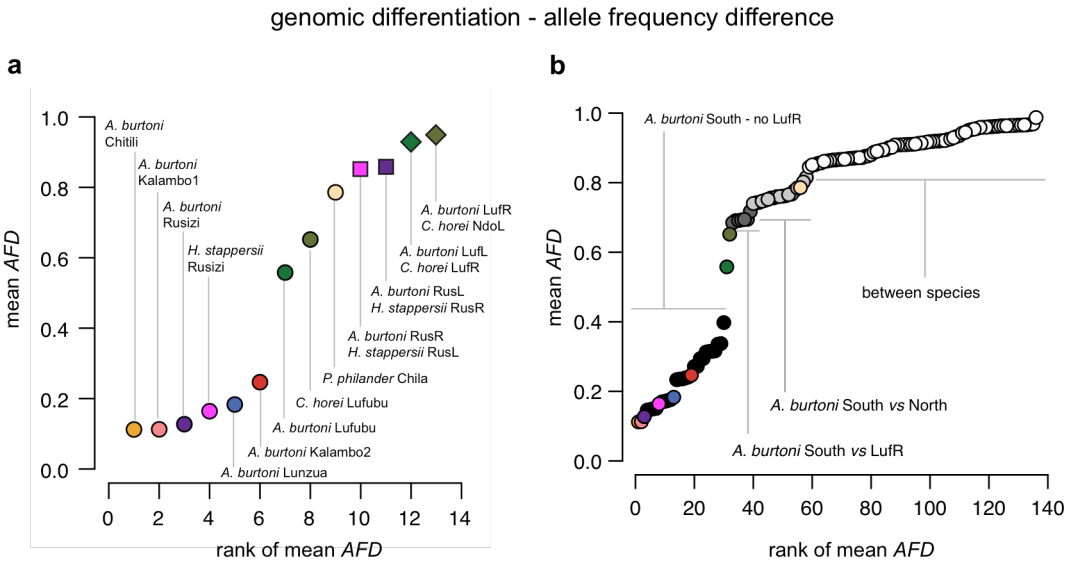
1. Theis, A., Salzburger, W. & Egger, B. The function of anal fin egg-spots in the cichlid fish *Astatotilapia burtoni*. *PLoS One* 7, e29878 (2012).
2. Theis, A., Bosia, T., Roth, T., Salzburger, W. & Egger, B. Egg-spot pattern and body size asymmetries influence male aggression in haplochromine cichlid fishes. *Behav. Ecol.* 26, 1512–1519 (2015).
3. Baerends, G. P. & Baerends-van Roon, J. M. An introduction to the study of the ethology of the cichlid fishes. *Behav. Suppl.* III–243 (1950).
4. Dijkstra, P. D. *et al.* The melanocortin system regulates body pigmentation and social behaviour in a colour polymorphic cichlid fish†. *Proc. R. Soc. B Biol. Sci.* 284, 20162838 (2017).
5. Maruska, K. P., Ung, U. S. & Fernald, R. D. The African cichlid fish Astatotilapia burtoni uses acoustic communication for reproduction: sound production, hearing, and behavioral significance. *PLoS One* 7, e37612 (2012).
6. Bates, D., Maechler, M., Bolker, B. & Walker, S. lme4: Linear mixed-effects models using Eigen and S4. R package version 1.1–7. 2014. (2015).
7. R Core Team, R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. 2016. (2017).
8. Plenderleith, M., Van Oosterhout, C., Robinson, R. L. & Turner, G. F. Female preference for conspecific males based on olfactory cues in a Lake Malawi cichlid fish. *Biol. Lett.* 1, 411–414 (2005).
9. Taylor, M. I. *et al.* Characterization of tetranucleotide microsatellite loci in a Lake Victorian, haplochromine cichlid fish: a *Pundamilia pundamilia* x *Pundamilia nyererei* hybrid. *Mol. Ecol. Resour.* 2, 443–445 (2002).
10. Lee, W.-J. & Kocher, T. D. Microsatellite DNA markers for genetic mapping in Oreochromis niloticus. *J. Fish Biol.* 49, 169–171 (1996).
11. Sanetra, M., Henning, F., Fukamachi, S. & Meyer, A. A microsatellite-based genetic linkage map of the cichlid fish, *Astatotilapia burtoni* (Teleostei): a comparison of genomic architectures among rapidly speciating cichlids. *Genetics* 182, 387–397 (2009).
12. Kalinowski, S. T., Taper, M. L. & Marshall, T. C. Revising how the computer program CERVUS accommodates genotyping error increases success in paternity assignment. *Mol. Ecol.* 16, 1099–1106 (2007).



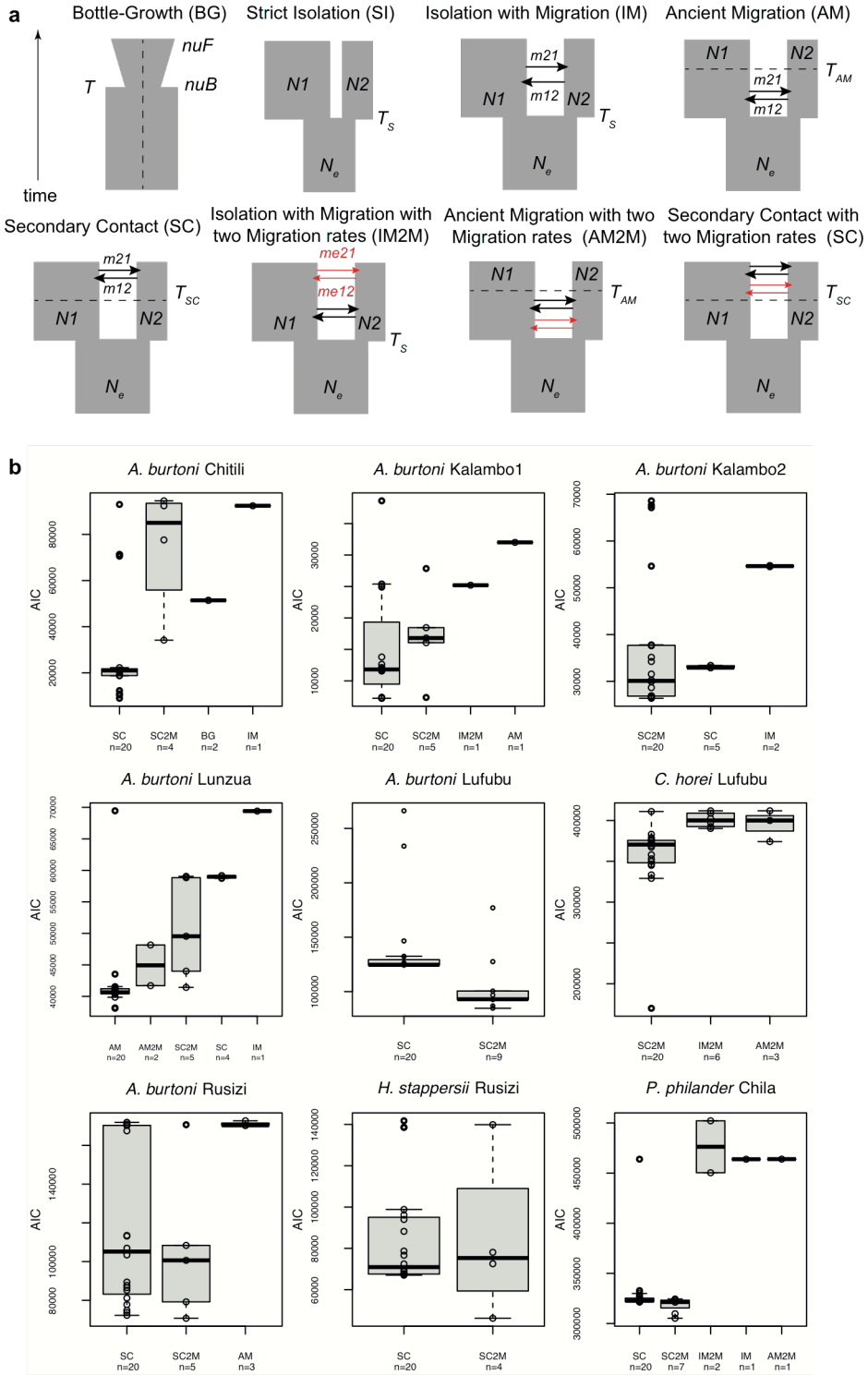
Extended Data Figure 1 Phylogeny and genetic structure of all individuals used in this study**. a**, RAxML phylogenetic reconstruction based on a concatenated whole-genome dataset. A deep split was found between the northern and the southern populations of *A. burtoni*, confirming the findings of a recent phylogeographic study[29]. KBC4, a putative hybrid individual excluded from the analyses, is highlighted in red. The riverine populations of *C. horei* and *P. philander* display extremely low levels of genetic diversity, suggesting that these populations are currently experiencing or have experienced a strong bottleneck. **b**, First and second axes of whole-genome principal component analyses (PCA), split per species.



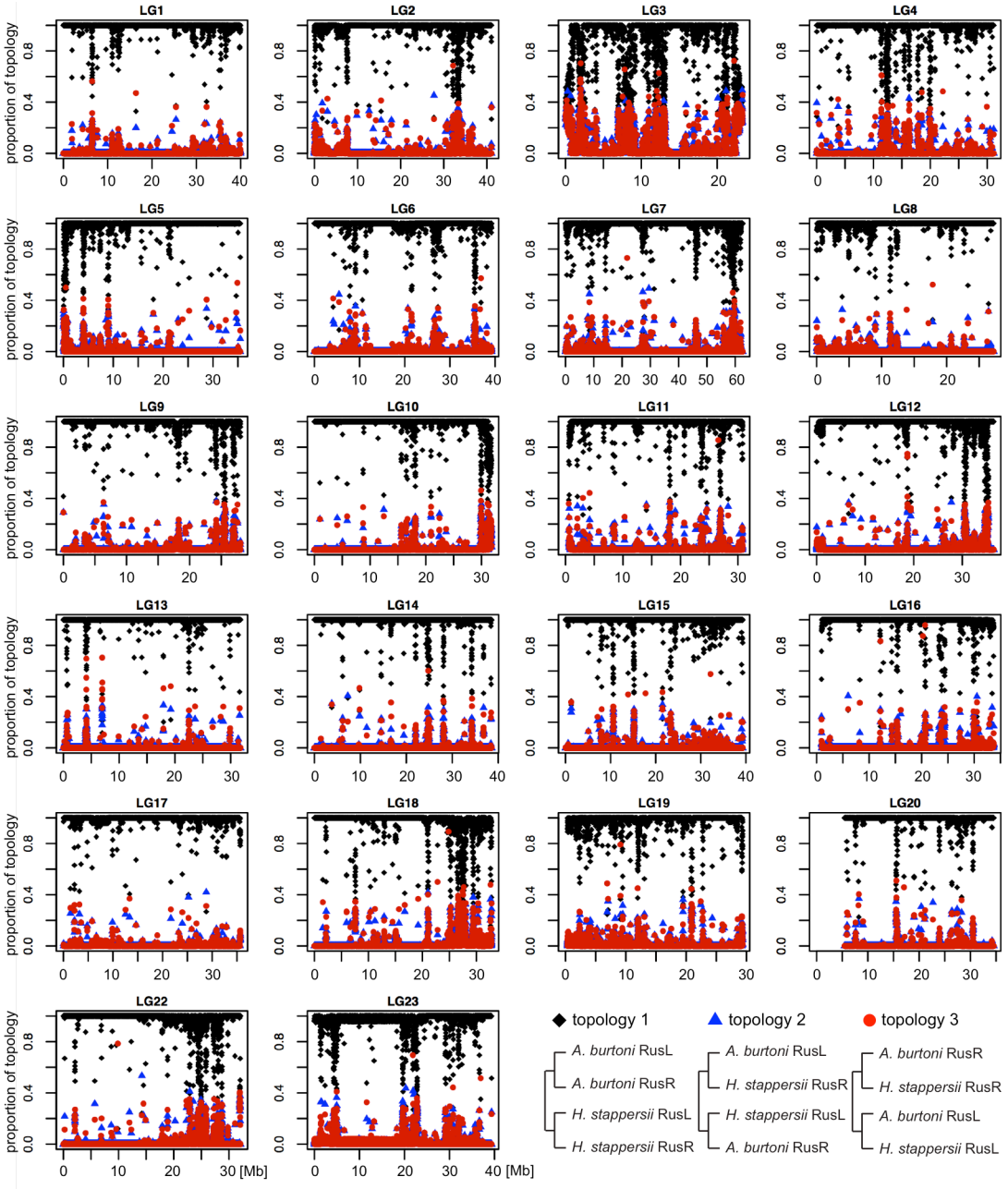
Extended Data Figure 2 Geometric morphometric analyses of *A. burtoni* (N=289), *H. stappersii* (N=81), *C. horei* (N=67) and *P. philander* (N=31). **a**, Position of 17 landmarks used for geometric morphometrics. **b**, Canonical variate analysis (CVA) of body shape for *A. burtoni* only. Lake population outlines are shown in dashed lines. **c**, Canonical variate analysis (CVA) of body shape for all species. Lake population outlines are shown in dashed lines. **d**, CV1 shape change for *A. burtoni* only (scaling factor: 10; outlines are for illustration purposes only, light grey outlines: positive values; black outlines: negative values). **e**, CV2 shape change for all species (scaling factor: 10; outlines are for illustration purposes only, from light grey to dark outlines with increasing values).



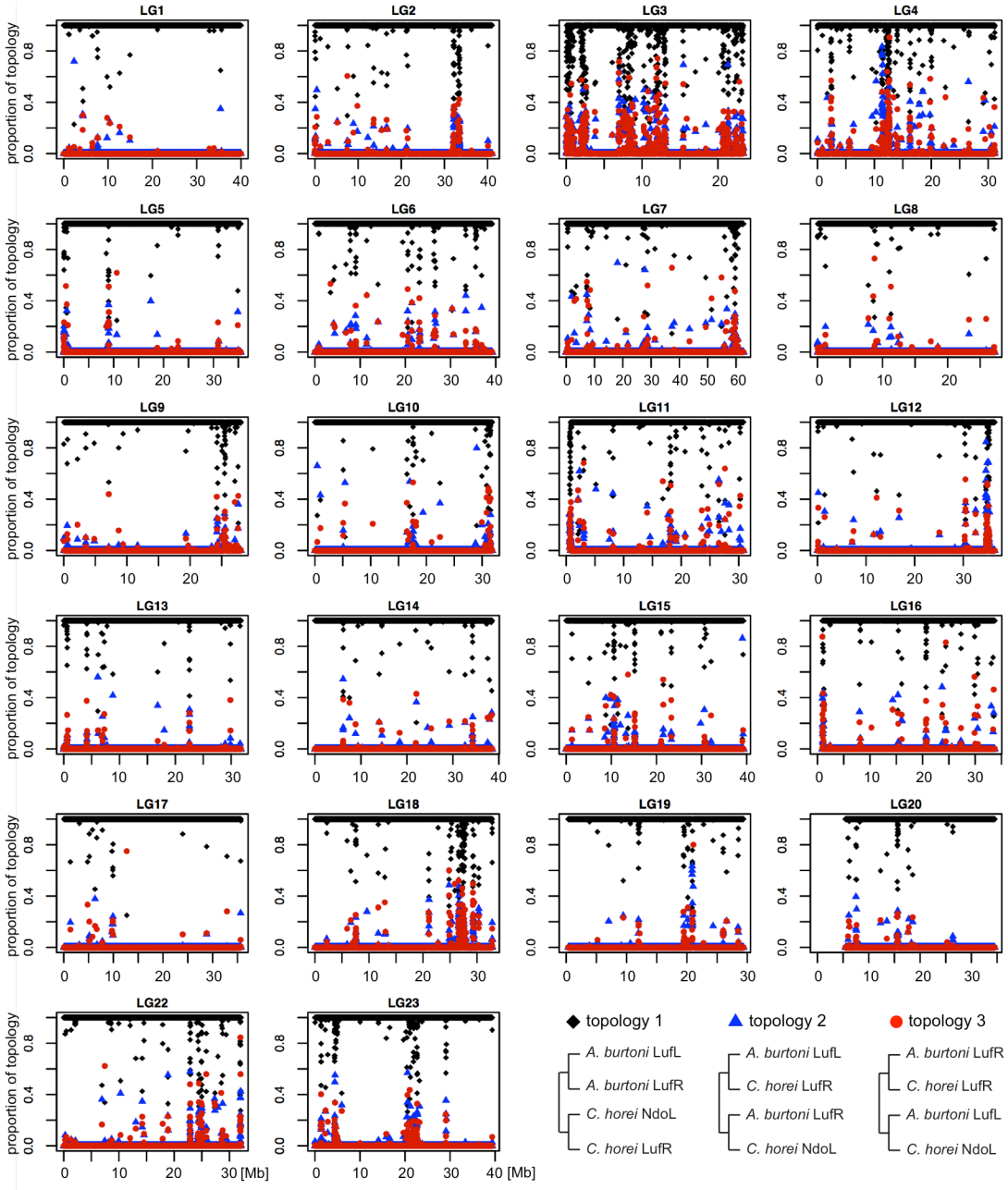
Extended Data Figure 3 *AFD*-based genomic differentiation **a**, Genomic differentiation based on average allele frequency difference (*AFD*) for nine lake-stream population pairs (circles) and four sympatric species pairs (squares and diamonds) sorted by increasing value. The colour scheme for lake-stream systems is the same as in Fig. 1. RusR: Rusizi River; RusL: Rusizi Lake; NdoL: Ndole Lake; LufR: Lufubu River; LufL: Lufubu Lake. **b**, Genomic differentiation based on average *AFD* for 136 lake-stream population and species pairs including all possible pairwise comparisons, sorted by increasing values, showing a tipping point between the one-species (within *A. burtoni* South – no Lufubu River) and the two-species (*A. burtoni* South versus Lufubu River; *A. burtoni* South versus *A. burtoni* North) category.



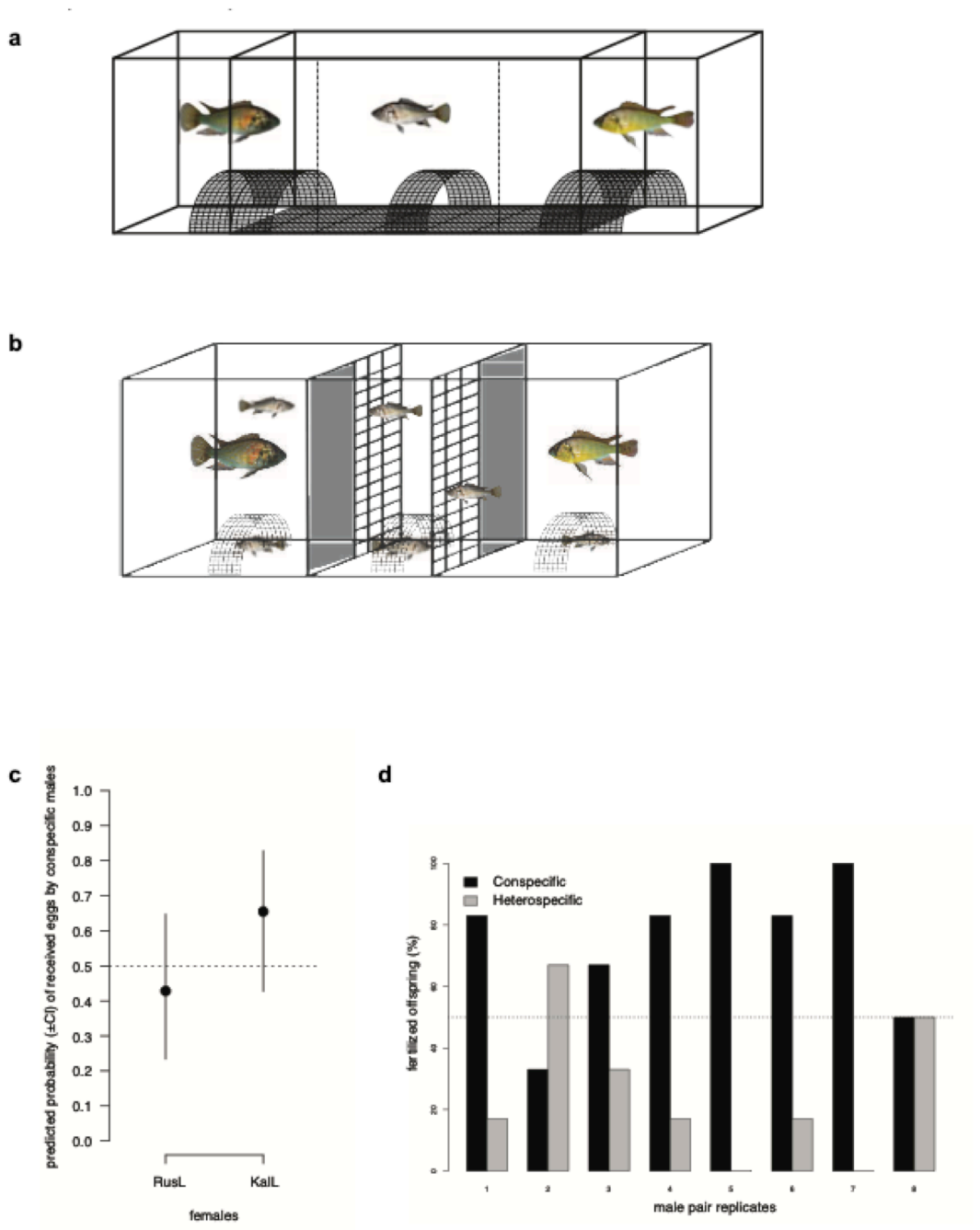
Extended Data Figure 4 Models and results of demographic modelling based on ∂a∂i**. a**, Schematic representation of the eight demographic models tested. Bottle-Growth (BG) model: T, Time of population size change; nuB, population size at the time of the change; nuF, current population size. Strict Isolation (SI) model: Ts, splitting time; Ne, ancestral population size; N1, current size of population 1; N2, current size of population 2. Isolation with Migration (IM) model: m21, migration rate from population 1 to population 2; m12, migration rate from population 2 to population 1. Ancient Migration (AM) model: T_{AM}, Time at which migration stopped. Secondary Contact (SC) model: T_{SC}, Time of the secondary contact. Isolation with Migration with two Migration rates (IM2M): me21, second migration rate from population 1 to population 2. me12, second migration rate from population 2 to population 1. **b**, Most appropriate demographic model for each system in 24-32 replicates. A low AIC represents a good fit between a model and the data.



Extended Data Figure 5 Topology weighting results per linkage group from *A. burtoni* and *H. stappersii* from the Rusizi system**.** Topology weighting analysis (*Twisst*) reconstructed fixed-length 5 kb windows phylogenies. The species topology (topology 1) is recovered in all cases and highlights that no introgression was detected between *A. burtoni* and *H. stappersii* in sympatry. RusL: Rusizi Lake. RusR: Rusizi River.

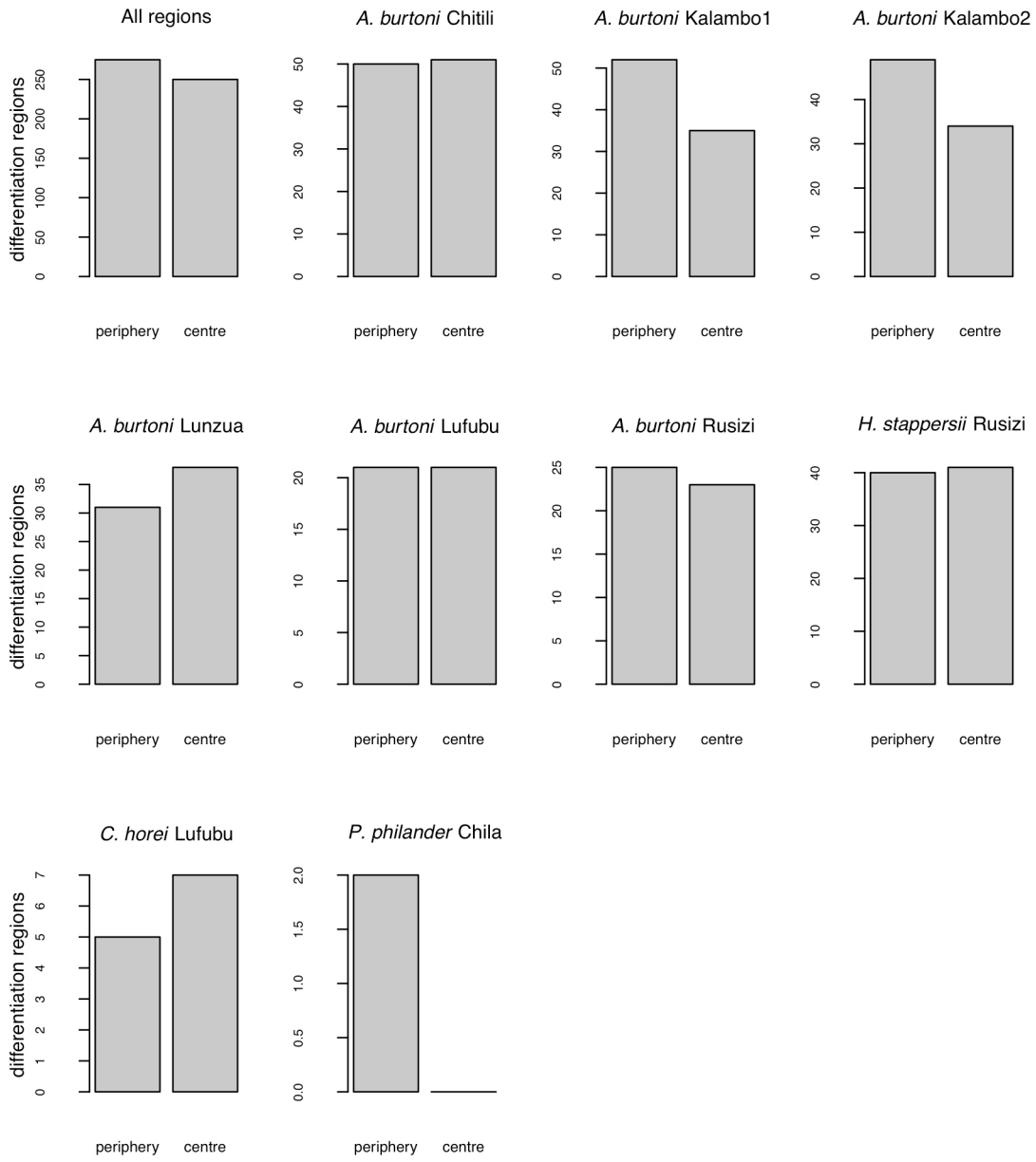


Extended Data Figure 6 Topology weighting results per linkage group from *A. burtoni* and *C. horei* from the Lufubu system. Topology weighting analysis (*Twisst*) reconstructed fixed-length 5 kb windows phylogenies. The species topology (topology 1) is recovered in all cases and highlights that no introgression was detected between *A. burtoni* and *C. horei* in sympatry. LufL: Lufubu Lake. LufR: Lufubu River. NdoL: Ndole Lake.

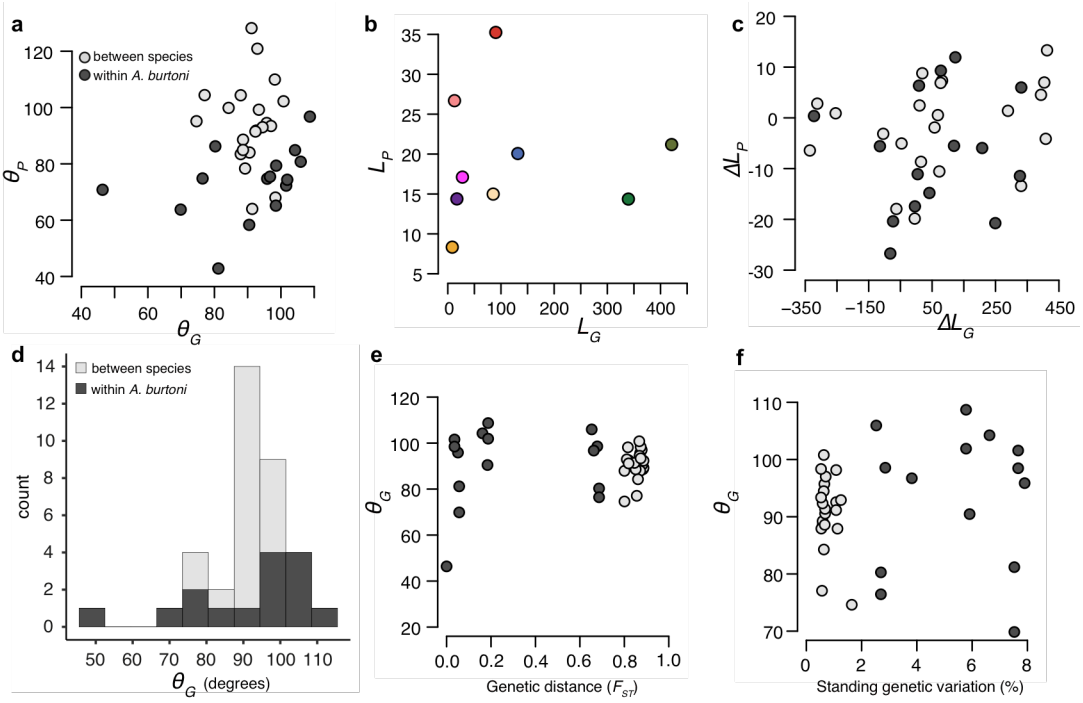


Extended Data Figure 7 Set-up and results of mate-choice experiments**. a,** Set-up of mate-choice experiment 1, a two-way female choice experiment based on visual cues only[86,87]. A female is placed into the central aquarium equipped with an egg-trap; two stimulus males are placed in the flanking aquaria. The dashed line indicates the choice zone. **b**, Set-up for mate-choice experiment 2, a two-way female-choice experiment with direct contact between territorial males (in the outer compartments) and the freely swimming females. **c**, Results from the mate-choice experiment 1 (model 1). The probability of females from both populations laying eggs with their conspecific male is not different from random. RusL: Rusizi Lake; KalL: Kalambo Lake **d**, Results from the mate-choice experiment 2 showing the percentage of fertilized offspring for each replicate. The females mated significantly more often with a conspecific male. Additional details on methods and results of mate-choice experiments are available in Appendix 1.

ß



Extended Data Figure 8 Number and position of the differentiation windows on chromosomes (chromosome centre versus periphery) shown for all regions combined and for each system separately.



Extended Data Figure 9 Phenotypic and non-outlier genetic divergence vectors are not correlated**. a**, The angles of phenotypic (θ_P) and genetic non-outlier (θ_G) lake-stream divergence vectors are not correlated (Mantel test: P = 0.34). **b**, The lengths of phenotypic (L_P) and genetic non-outlier (L_G) lake-stream divergence vectors are not correlated (Linear regression model: P = 0.93). The colour scheme is the same as in Figure 2. **c**, The differences between phenotypic (ΔL_P) and genetic non-outlier (ΔL_G) vector length are not correlated (Mantel test: P = 0.35). **d**, Histogram of the 36 (pairwise) angles between lake-stream non-outlier genetic divergence vectors (θ_G) in degrees. **e**, The angles of genetic non-outlier divergence vectors (θ_G) and genetic (F_{ST}) distances between lake populations are not correlated (Mantel test: P = 0.48). **f**, The angles of genetic non-outlier divergence vectors (θ_G) and the amount of standing genetic variation between lake populations are not correlated (Mantel test P = 0.19).

Supplementary tables

Supplementary Table 1: Individual measurements and genome statistics.

Supplementary Table 2: Details on sampling localities, sample sizes and genome-wide population statistics.

Supplementary Table 3: Population parameters inferred with ∂a∂i simulations.

Supplementary Table 4: Number and localization of differentiation regions per system and per linkage group

Supplementary Table 5: List of 637 outlier candidate genes from the differentiation regions of each system and their respective GO annotations

Supplementary Table 6: List of 25 outlier candidate genes found in the overlap of differentiation regions among systems.

Supplementary Table 7: List of 367 outlier candidate genes from the overlap of the three outlier core sets from *A. burtoni* northern populations, *A. burtoni* southern populations and *H. stappersii* populations.

Side Projects

Sex chromosome evolution in cichlids and
conservation of biodiversity in the East African Great Lakes

Chapter 5

Repeated evolution versus common ancestry: Sex chromosome evolution in the haplochromine cichlid *Pseudocrenilabrus philander*

A. Böhne, A.A.T. Weber, J. Rajkov, M. Rechsteiner, A. Riss, B. Egger, W. Salzburger

Genome Biology and Evolution (2019)

I conducted the fieldwork together with AATW and BE. I conducted the molecular laboratory work for genome sequencing together with AATW. All authors contributed to manuscript finalization.

GBE

Repeated Evolution Versus Common Ancestry: Sex Chromosome Evolution in the Haplochromine Cichlid *Pseudocrenilabrus philander*

Astrid Böhne[1,*,†], Alexandra Anh-Thu Weber[1,2,†], Jelena Rajkov[1], Michael Rechsteiner[1], Andrin Riss[1], Bernd Egger[1,3], and Walter Salzburger[1]

[1]Department of Environmental Sciences, Zoological Institute, University of Basel, Switzerland

[2]Museums Victoria, Melbourne, Victoria, Australia

[3]Program Man Society Environment, University of Basel, Switzerland

[†]These authors contributed equally to this work.

*Corresponding author: E-mail: astrid.boehne@unibas.ch.

Accepted: January 7, 2019

Data deposition: The raw sequencing reads were deposited at SRA and are available under the accession SRP148476.

Abstract

Why sex chromosomes turn over and remain undifferentiated in some taxa, whereas they degenerate in others, is still an area of ongoing research. The recurrent occurrence of homologous and homomorphic sex chromosomes in distantly related taxa suggests their independent evolution or continued recombination since their first emergence. Fishes display a great diversity of sex-determining systems. Here, we focus on sex chromosome evolution in haplochromines, the most species-rich lineage of cichlid fishes. We investigate sex-specific signatures in the *Pseudocrenilabrus philander* species complex, which belongs to a haplochromine genus found in many river systems and ichthyogeographic regions in northern, eastern, central, and southern Africa. Using whole-genome sequencing and population genetic, phylogenetic, and read-coverage analyses, we show that one population of *P. philander* has an XX–XY sex-determining system on LG7 with a large region of suppressed recombination. However, in a second bottlenecked population, we did not find any sign of a sex chromosome. Interestingly, LG7 also carries an XX–XY system in the phylogenetically more derived Lake Malawi haplochromine cichlids. Although the genomic regions determining sex are the same in Lake Malawi cichlids and *P. philander*, we did not find evidence for shared ancestry, suggesting that LG7 evolved as sex chromosome at least twice in haplochromine cichlids. Hence, our work provides further evidence for the labile nature of sex determination in fishes and supports the hypothesis that the same genomic regions can repeatedly and rapidly be recruited as sex chromosomes in more distantly related lineages.

Key words: sex determination, species complex, teleost fish, speciation, genome sequencing, population genetics.

Introduction

Sexual reproduction is nearly universal across eukaryotes (Speijer et al. 2015; Garg and Martin 2016). One of the most puzzling aspects of this ancient trait is the remarkable contrast between ultra-conserved features (e.g., meiosis, ploidy changes, and cell fusion) and plastic components (e.g., sex determination and modes of reproduction) (Lode 2012; Heule et al. 2014; Capel 2017; Pannell 2017). In particular, the great diversity of sex-determining (SD) mechanisms suggests their repeated and continuous evolution throughout the eukaryotic tree of life (reviewed by Heitman [2015], Blackmon et al. [2017], Capel [2017], and Pannell [2017]), supporting the view of sex as a threshold phenotype that can be canalized into either one of two discrete states by a variety of extrinsic or intrinsic factors as well as a combination thereof (Perrin 2016; Capel 2017). The involvement of extrinsic factors in SD is summarized under the term environmental sex determination (ESD). Intrinsic factors, commonly referred to as genetic sex determination (GSD), comprise systems ranging from single base pair differences between the

sexes (Kamiya et al. 2012) to highly differentiated sex chromosomes as in mammals or birds (Graves 2006, 2008, 2014) and including polyfactorial SD (Moore and Roberts 2013) and even SD via RNA instead of protein-coding genes (Akagi et al. 2014; Kiuchi et al. 2014). Sex chromosomes originate from autosomes when one locus acquires a mutation such that heterozygous individuals develop into one sex, whereas homozygous ones develop into the other sex. If sex chromosomes evolve within an ancestrally hermaphroditic (or monoecious) species, at least two mutations are necessary to induce the evolution of GSD (Muller 1932; Westergaard 1958; Charlesworth and Charlesworth 1978) and hence of sex chromosomes.

The canonical model of sex chromosome evolution predicts that suppression of recombination between such proto-sex chromosomes is favored (Muller 1918) and adjacent sexually antagonistic mutations may cause the spread of reduced recombination along the chromosome (Charlesworth [2017] but see also Cavoto et al. [2018]). Suppressed recombination will lead to a reduced effective population size of the sex-limited chromosome (Y in male-heterogametic species; W in female-heterogametic species) and an increase of Hill–Robertson interferences (Charlesworth et al. 1987; Charlesworth and Charlesworth 2000; Charlesworth 2017). Deleterious mutations on the Y/W can no longer be purged and, consequently, accumulate under the impact of Muller's ratchet, background selection, and selective sweeps (Charlesworth and Charlesworth 2000; Charlesworth et al. 2005). This can lead to chromosomal decay, as exemplified by the mammalian Y chromosome (reviewed by Graves [2006], Bellott and Page [2009], and Schartl et al. [2016]). One escape route to this "evolutionary trap" can be sex chromosome turnover suggested to be induced by deleterious mutation load (Blaser et al. 2013) or sex-antagonistic mutations occurring on autosomes (van Doorn and Kirkpatrick 2007; van Doorn and Kirkpatrick 2010) driving the evolution of a new sex chromosome pair. Sex chromosome turnovers have indeed been shown in, for example, fishes and amphibians (Miura 2007; Volff et al. 2007; Kitano and Peichel 2012; Sessions et al. 2016; Jeffries et al. 2018), with cichlids illustrating the role of sexual antagonism as a driving force in this process (Roberts et al. 2009).

Alternatively, low levels of recombination might be maintained between the two sex chromosomes that are sufficient to allow the purging of deleterious mutations from Y or W chromosomes (Guerrero et al. 2012; Dufresnes et al. 2014). The loci that pave the way for sex chromosome evolution are often unknown. Still, comparisons across different animal taxa revealed the recurrent evolution of certain genes as master SD genes. This has led to the proposition that there are "limited options" for SD genes or even sex chromosomes (Marshall Graves and Peichel 2010).

With ~3,000–4,000 species, cichlid fishes are one of the largest vertebrate families (Salzburger and Meyer 2004). Because of their taxonomic richness, their phenotypic and ecologic diversity, and their propensity to diversify, cichlids are an important model system in evolutionary biology (Kornfield and Smith 2000; Henning and Meyer 2014; Seehausen 2015; Salzburger 2018). The most species-rich lineage within Cichlidae is Haplochromini, which includes the members of the adaptive radiations in Lakes Victoria and Malawi (together ~1,200 species), many riverine and lacustrine species elsewhere in Africa (Turner et al. 2001; Verheyen et al. 2003; Schwarzer et al. 2009; Schwarzer et al. 2012), as well as ~30 species endemic to Lake Tanganyika (the "Tropheini") (Salzburger et al. 2005).

Cichlid fishes perfectly exemplify the plastic components of sexual reproduction in that closely related species feature various breeding systems and a variety of SD mechanisms including ESD and GSD systems (Römer and Beisenherz 1996; Cnaani et al. 2008; Ser et al. 2010; Yoshida et al. 2011; Parnell et al. 2012; Parnell and Streelman 2013; Reddon and Hurd 2013; Kudo et al. 2015; Böhne et al. 2016; Roberts et al. 2016; Peterson et al. 2017; Feulner et al. 2018; Gammerdinger et al. 2018a, b) (fig. 1*A*). Cichlids are, thus, an excellent model to study the dynamics of SD system evolution. Previous research on the evolution of SD systems in African cichlids lends some support to the "limited options" hypothesis. Two particular chromosomes (corresponding to LG5 and LG7 in the Nile tilapia genome, an outgroup species to the East African Great Lakes, often used as reference) have repeatedly been recruited as sex chromosomes in different species of the East African Great Lakes (Parnell et al. 2012; Kudo et al. 2015; Böhne et al. 2016; Roberts et al. 2016; Peterson et al. 2017; Gammerdinger et al. 2018a; Ser et al. 2010).

In this study, we approach cichlid sex determination from a phylogenetic perspective by investigating sex chromosome signatures in the *Pseudocrenilabrus philander* species complex, a member of a sister-clade to the modern haplochromines of Lakes Victoria, Malawi, and Tanganyika. We sampled two populations for whole-genome sequencing in northern Zambia: Mbulu creek and Lake Chila, a small lake 20 km south of Lake Tanganyika, which is connected to the Mbulu creek via its outflow (fig. 1*B* and *C*). The *P. philander* species complex (Katongo et al. 2005; Koblmuller et al. 2012) comprises two major mitochondrial lineages, one representing the Zambezi–Kafue drainage and one lineage of mainly Congolese origin (Egger et al. 2015). Both lineages occur in Lake Chila, with the Zambezi–Kafue lineage being far more frequent (Egger et al. 2015). Population assignment tests based on microsatellite data suggest that the two lineages represent a single panmictic population. The Mbulu creek population belongs to the Zambezi–Kafue lineage and experienced genetic bottlenecks probably induced by strong seasonal variation in water volume (Egger et al. 2015). Upon the inspection of 24 newly sequenced *P. philander* genomes and a marker-based approach in a larger set of individuals, we provide strong evidence for an XX–XY SD system on LG7 in

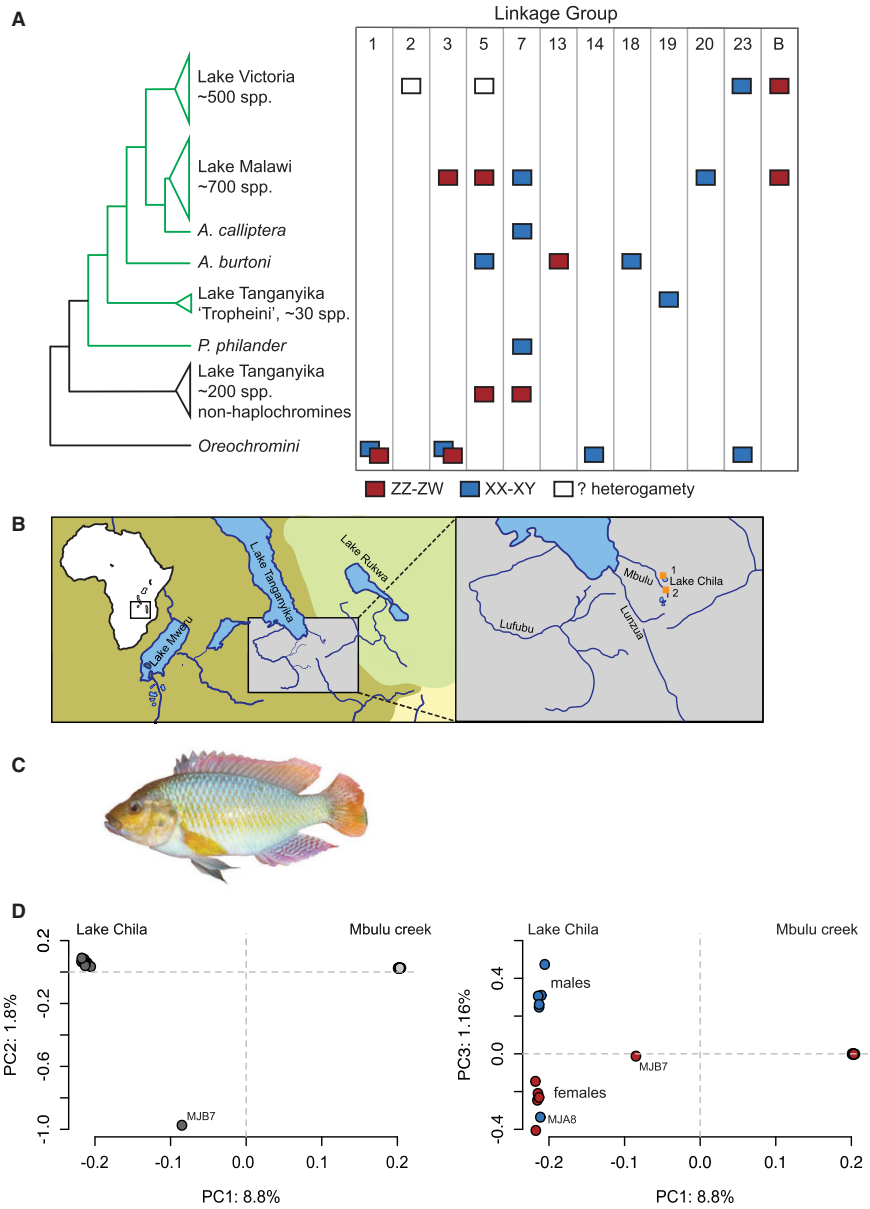


Fig. 1.—Phylogenetic relationships and sex determination in East African cichlids. (*A*) Schematic phylogenetic relationships of East African cichlids. Information on sex determination systems based on Böhne et al. (2016), Cnaani et al. (2008), Feulner et al. (2018), Gammerdinger et al. (2018a, b), Kudo et al. (2015), Parnell and Streelman (2013), Peterson et al. (2017), Roberts et al. (2016), Ser et al. (2010), and Yoshida et al. (2011). Haplochromine lineages are depicted in green. (*B*) Map of East Africa and a zoom on the sampling locations: 1, Lake Chila and 2, Mbulu creek. (*C*) Male specimen of *Pseudocrenilabrus philander*. (*D*) PCA on genome-wide variant data of all *P. philander* individuals of this study. PC1 separates the lake individuals from the creek population. PC3 separates males from females. The outlier MJB7 and the potential sex-reversed individual MJA8 are highlighted: dark gray: Lake Chila, light gray: Mbulu creek, red: females, and blue: males.

the lake population. We could not detect this or any other GSD system in the genomes of the creek population. We compare our results to an XX–XY system in the same genomic region of cichlids from Lake Malawi (Ser et al. 2010; Parnell and Streelman 2013; Peterson et al. 2017). Finally, we show that the XX–XY SD system on LG7 in *P. philander* possibly evolved within Lake Chila, because it seems absent in other populations of the *P. philander* species complex.

Materials and Methods

Sampling, DNA Extraction, and Sequencing

For this study, we sampled six males and six females of *P. philander* from Lake Chila and 12 individuals (4 males, 3 females, and 5 juveniles) from the adjacent Mbulu creek for whole-genome sequencing (fig. 1). In addition, we included 78 specimens sampled for a previous study (Egger et al. 2015) for polymerase chain reaction (PCR) genotyping (see below). Fin clips and whole specimens were preserved in ethanol. Individuals were sexed by visual inspection of the gonads and body coloration. Five specimens from Mbulu creek did not show distinguishable gonads and were defined as juveniles. DNA was extracted from fin clips with EZNA Tissue DNA Kit (Omega Bio-Tek). Individual genomic libraries were prepared with TruSeq DNA PCR-free Low Sample Kit (Illumina), pooled per population and subsequently sequenced (150 bp paired-end) on four lanes of an Illumina HiSeq3000 by the genomics facility of the D-BSSE (Basel, Switzerland; supplementary table S1, Supplementary Material online). Sequencing data were deposited in the SRA (SRP148476). Research involving animals was performed with approval of the Swiss authorities under a research permit issued by the Lake Tanganyika Research Unit, Department of Fisheries, Mpulungu, Zambia.

Raw Data Processing, Read Alignment, Variant Calling, and Filtering

Raw reads were inspected with FastQC (0.11.3, https://www.bioinformatics.babraham.ac.uk/projects/fastqc/; last accessed January 23, 2019) and adapters trimmed with Trimmomatic 0.36 (ILLUMINACLIP:TruSeq3-PE-3.fa:2:30:10:2:true) (Bolger et al. 2014). We used the Nile tilapia (*Oreochromis niloticus*) genome assembly version 2 (refseq accession number GCF_001858045.1_ASM185804v2) as reference. Unplaced scaffolds were concatenated lexicographically into an "UNPLACED" super chromosome. This reference was indexed with BWA 0.7.13 and alignments of each individual performed using bwa-mem with default parameters (Li and Durbin 2009) (table 1 and supplementary table S1, Supplementary Material online). Alignments were coordinate sorted and indexed with SAMtools 1.3.1 (Li et al. 2009). We performed an indel realignment (RealignerTargetCreator and IndelRealigner, GATK 3.4.0) (McKenna et al. 2010). Variants were called with GATK's HaplotypeCaller (per individual and per chromosome), GenotypeGVCFs (per chromosome), and CatVariants (to merge all obtained VCF files). The VCF file was filtered with DP < 100; DP > 800; MQ < 20; FS > 60; SOR > 10; MQRankSum < −10; ReadPosRankSum < −10; and QD < 2. Variants with >50% missing data were excluded using –max-missing 0.5 in VCFtools 0.1.14 (Danecek et al. 2011).

Population Structure and Phylogeny

To assess population structure, between-population genome-wide F_{ST}, average d_{xy} (absolute divergence), and average π (nucleotide diversity) were calculated in 10 kb windows on the filtered VCF file including single nucleotide polymorphisms (SNPs) and indels using evo (https://github.com/millanek/evo/; last accessed January 23, 2019). Average d_a (net divergence) was calculated using Nei and Li's formula: $d_a = d_{xy} - (\pi_x + \pi_y)/2$ (Nei and Li 1979). Tajima's *D* was calculated for each population in 10 kb windows in VCFtools 0.1.14 (Danecek et al. 2011). Population structure was examined on the whole-genome VCF data set with a principal component analysis (PCA) using smartPCA (Eigensoft 6.1.4) (Patterson et al. 2006). Alignments to the mitochondrial reference scaffold NC_013663.1 were extracted from individual BAM files, sorted with SAMtools 1.3.1. (Li et al. 2009) and converted to fastq format using Picard 2.8.0 SamToFastq (http://broadinstitute.github.io/picard; last accessed January 23, 2019). Mitochondrial genomes were reconstructed from these reads with MIRA 4 (Chevreux et al. 1999). The regions corresponding to the control region (D-Loop) were subsequently extracted and aligned with additional public sequences from the *P. philander* species complex (sequences from Egger et al. [2015]); using MAFFT online service 7 (Katoh et al. 2017) under the FFT-NS-i option, that is, with fast construction of an initial alignment followed by iterative refinement until convergence. Identical sequences were collapsed into haplotypes using DNA collapser (FaBox) (Villesen 2007). Bayesian inference of phylogeny was done in MrBayes 3.2.2 (Ronquist et al. 2012). Posterior probabilities were obtained from Markov chain Monte Carlo simulations in two independent runs (10 chains with 10 Mio generations each, chain temperature: 0.25, trees sampled every 1,000 generations) using the best-fit model of molecular evolution as suggested by jModelTest (Posada 2008). A 50% majority-rule consensus tree was constructed after a 1 million generation burn-in (chain stationarity and run parameter convergence were checked with Tracer 1.6, http://tree.bio.ed.ac.uk/software/tracer/; last accessed January 23, 2019, using posterior probability as a measure of clade support). A whole nuclear genome phylogeny was built by reconstructing for each individual a sequence corresponding to the first haplotype of each linkage group using samtools faidx (LG) (Li et al. 2009) and bcftools consensus –haplotype 1 (BCFtools 1.5,

Table 1
Detailed Information and Genome Statistics for the Individuals Used in This Study

Sample Name	Population	Mitochondrial Haplotype	Phenotypic Sex	Genotypic Sex	Mean Seq. Coverage	Genome Wide F_{IS}	All but LG7 F_{IS}	LG7 F_{IS}
MJA4	LC	Ht31	F	F	15.32	0.183	0.162	0.450
MJA6	LC	Ht31	F	F	15.15	0.163	0.149	0.341
MJC1	LC	Ht31	F	F	13.56	0.226	0.219	0.312
MJC2	LC	Ht31	F	F	13.49	0.229	0.219	0.360
MJC3	LC	Ht18	F	F	14.14	0.223	0.211	0.381
MJB7	LC	Ht32	F	F	14.34	−0.345	−0.338	−0.433
MJA8	LC	Ht18	M	F	13.14	0.239	0.226	0.420
MJB1	LC	Ht31	M	M	12.95	0.219	0.239	−0.038
MJB3	LC	Ht31	M	M	12.86	0.201	0.219	−0.038
MJB5	LC	Ht18	M	M	13.38	0.185	0.209	−0.121
MJB8	LC	Ht31	M	M	13.44	0.242	0.263	−0.019
MJB9	LC	Ht31	M	M	14.10	0.169	0.189	−0.081
MJC7	MC	Ht13	M	U	13.66	0.166	0.165	0.187
MJC8	MC	Ht13	M	U	15.00	0.097	0.095	0.139
MJC9	MC	Ht13	M	U	12.71	0.208	0.205	0.266
MJE7	MC	Ht13	M	U	14.73	0.095	0.087	0.237
MJD1	MC	Ht13	F	U	16.39	0.048	0.045	0.101
MJD2	MC	Ht13	F	U	13.47	0.181	0.172	0.320
MJD3	MC	Ht13	J	U	13.15	0.219	0.212	0.337
MJD5	MC	Ht13	J	U	13.59	0.182	0.182	0.188
MJD6	MC	Ht13	J	U	14.09	0.178	0.167	0.359
MJD8	MC	Ht13	J	U	14.60	0.153	0.147	0.241
MJD9	MC	Ht13	J	U	15.26	0.108	0.100	0.236
MJE6	MC	Ht13	F	U	13.45	0.179	0.175	0.250

NOTE.—Mitochondrial haplotypes correspond to naming in Egger et al. (2015). Mean sequencing coverage was calculated on the final VCF file. F_{IS} was calculated on the final VCF file subset per population. LC, Lake Chila; MC, Mbulu creek; F, female; M, male; J, juvenile; U, undifferentiated; F_{IS}, inbreeding coefficient.

https://samtools.github.io/bcftools/; last accessed January 23, 2019). The sequences of each linkage group were then concatenated and merged into one sequence per individual using EMBOSS union (Rice et al. 2000). Maximum likelihood inference was done with RAxML 8.2.11 (*–k*, *−# 100*, *−f a*) (Stamatakis 2014). Branch length estimation (*−k*) is given in number of mutations per bp per generation. In order to obtain divergence times in number of generations, we used the Lake Malawi cichlid mutation rate estimation of 3.5×10^{-9} per bp per generation (95% CI: 1.6×10^{-9} to 4.6×10^{-9}) from Malinsky et al. 2018. The VCF file was phased and genotypes were imputed with Beagle 4.1 (Browning and Browning 2007, 2016). For topology weighting, we used *Twisst* (Martin and Van Belleghem 2017) with 1, 5, and 10 kb windows to infer if Chila and Mbulu males were more closely related to each other than to the females of their respective population in a specific region of LG7.

Sex Chromosome Identification and Characterization of the Type of SD System on LG7

Male–female F_{ST} and difference in nucleotide diversity ($\pi_{diff} = \pi_{males} - \pi_{females}$) were calculated in 10 kb windows on the filtered VCF file including SNPs and indels with evo (https://github.com/millanek/evo/; last accessed January 23, 2019). We tested for a difference in nucleotide diversity between males and females of each population with a Welch two sample *t*-test in R 3.4.2. (R Core Team 2017). We calculated male–female F_{ST} per population (five males vs. five females for Lake Chila; four males vs. three females for Mbulu creek) as well as for both populations combined. A maximum likelihood phylogeny was reconstructed as described above on LG7 only and on all chromosomes excluding LG7. A relatedness statistic (unadjusted Ajk statistic) (Yang et al. 2010), of all individuals was calculated separately for LG7 and for all the remaining chromosomes in VCFtools (–relatedness and –chr LG7 or –not-chr LG7). F_{IS} (inbreeding coefficient) was calculated separately for LG7 and all LGs excluding LG7 in VCFtools (–het and –chr LG7 or –not-chr LG7) for each individual within its respective population. The inbreeding coefficient F_{IS} was also calculated in 10 kb windows per sex within each population along LG7 and correlated to male–female F_{ST} following the method described by Rodrigues and Dufresnes (2017). To obtain the average normalized F_{IS} value per sex for each 10 kb window, the per individual genome-wide F_{IS} value excluding LG7 was subtracted from the LG7 individual F_{IS} value. Then, the individual normalized F_{IS} values were averaged per sex. Next, we selected biallelic sites from the initial filtered, unphased VCF file for five males and five females from the lake population of the same mitochondrial linage resulting in

a total of 30,811,926 sites. We selected sites for which all females were homozygous and all five males heterozygous (XY-sites) as proposed by Brelsford et al. (2017). XY-sites on LG7 were annotated using SnpEff 4.3 (Cingolani et al. 2012).

De Novo Genome Assemblies and Alignment

We followed the pipeline described in Malmstrøm et al. (2017) to generate a female and male draft genome de novo assembly for Lake Chila *P. philander* using CeleraAssembler 8.3 (Myers et al. 2000) and FLASh 1.2.11 (Magoc and Salzberg 2011), pooling the raw reads of three females and three males. Assembly quality was assessed with QUAST 4.5 (Gurevich et al. 2013) and assembly completeness with BUSCO 3 (Simao et al. 2015) (supplementary table S2, Supplementary Material online). To anchor contigs onto the *O. niloticus* reference genome, we used LAST 861 (lastdb –uNEAR–cR11; lastal–m75–E0.05) (Kiełbasa et al. 2011). MAF alignment output was converted into tabular format with LAST. Female alignments to LG7 were extracted from the tabular output and filtered to keep scaffolds of >2 kb length and alignment sequence coverage of 50% resulting in 3,340 contigs representing the X chromosome. Scaffolds were ordered based on the start position of their longest alignment. For comparative purposes, we extracted the female scaffolds aligning to LG6 with the same settings (2,048 contigs).

Sequence Coverage Analysis

Coverage was calculated for each sex from mapping against the de novo assembled genomes. Quality filtered reads of the five male and five female individuals of Lake Chila were mapped against the female and male draft genome using bwa-mem of BWA (Li and Durbin 2009). Alignments were converted to BAM format, sorted, and indexed with SAMtools 1.3.1 (Li et al. 2009). Coverage per individual per site was calculated with samtools depth –aa (SAMtools 1.3.1) (Li et al. 2009). The median coverage against the female de novo assembly over all sites and all individuals per sex for each population was calculated in R 3.3.1, resulting in 17 for Lake Chila males and 18 for Lake Chila females. We did the same analysis keeping only alignments with zero mismatches resulting in a median coverage of 3 in Lake Chila males and 4 in females. Next, we calculated median coverage per site and sex for the scaffolds anchored to LG7 and LG6 (for comparative reasons) and normalized it by the sex-specific median. From these values, we calculated averages of 10 kb windows, which were log 2 transformed for plotting. These steps were run in R using the packages reshape 0.8.7 (Wickham 2007), miscTools 0.6-22 (Henningsen and Toomet 2016), zoo 1.8 (Zeileis and Grothendieck 2005), and ggplot2 2.2.1 (Wickham 2009). From the mapping against the male de novo assembly, we identified regions of "male-only-coverage" (potential Y-specific sequences) as regions in which consecutive positions of 1 kb length had coverage in at least four out of the five males with a total coverage >5 and a coverage over all females <3.

K-mer Analysis and Assemblies

To assemble Y chromosome–specific sequences, we followed a method described by Akagi et al. (2014). We identified Y-specific reads over their difference in k-mer composition compared with female reads. Raw reads were filtered with Trimmomatic 0.36 (Bolger et al. 2014) (PE mode, adapters.fasta.2:30:10LEADING:3TRAILING:3SLIDINGWINDOW:4:15MINLEN:5). From the trimmed reads, we generated k-mer tables for all 37 k-mers starting with the trigger sequence "AG" and having at least 5 counts reducing the k-mer complexity and computational cost as established by Akagi et al. (2014) using a Python script provided by the Comai lab. Using "CT" as the trigger sequence yielded similar results (data not shown). For comparative reasons, we applied the same method to a human data set of Great Britain ancestry from the 1000 Genomes Project Consortium (http://www.internationalgenome.org/; last accessed January 23, 2019, samples ERR020230, ERR050089, SRR189815, ERR050086, SRR068180, and SRR190845) that had already been used in a k-mer assembly for Y chromosomes (Carvalho and Clark 2013).

Resulting male and female k-mer counts were compared and potential Y-k-mers identified as k-mers that had >9 counts in males but <5 counts in females resulting in 3,612,202 unique Y-k-mers (out of 130,094,951 total unique k-mers). We extracted male reads matching these Y-k-mers and their mate with bbduk (BBTools 37.57, https://jgi.doe.gov/data-and-tools/bbtools/; last accessed January 23, 2019).

The resulting 55,627,673 read pairs were de novo assembled with MEGAHIT 1.1.1 (Li et al. 2015) with stepsize 10, kmin 21, kmax 121, and minimum length 1 kb. Male and female reads were back-mapped on the so-obtained 122,977 contigs. We removed contigs that had over 50 reads coverage at a single position in at least one male (likely individual specific repetitive elements) and those with a 5 read coverage in females. The resulting 233 contigs were blasted against the male and female draft genomes (Blast+ 2.6.0, BlastN with -qcov_hsp_perc 70 and -num_alignments 10, all other settings in default) (Camacho et al. 2009), and discarded if they had a match to the female genome with ≥95% sequence identity. From these remaining 138 contigs, 35 were also present in the full male genome assembly. The 138 contigs were loaded into Blast2GO (Conesa et al. 2005) and scanned for coding sequences with the integrated version of AUGUSTUS (Hoff and Stanke 2013) and *Danio rerio* as reference organism. Obtained genes were blasted against nr (BlastX), searched against Interpro, mapped, and annotated with default settings within Blast2GO. We calculated male and female coverage for these contigs following the same

method as described for the stringent method of X-chromosomal coverage.

K-mer Composition of the X Chromosome

To extract k-mers from the X chromosome, we selected k-mers that had a female/male count ratio between 1.75 and 2.25. The obtained 7,227,218 k-mers were blasted against the reconstructed X chromosome of *P. philander* with BlastN-short allowing only for perfect matches and maximum 10 alignments per query (Blast+ 2.6.0) (Camacho et al. 2009) resulting in 424,156 k-mers placed on the X chromosome.

Comparison of LG7 in Other Cichlids

LG7 carries an XY system in cichlids from Lake Malawi. WGS sequences for *Astatotilapia calliptera*, *Aulonocara stuartgranti*, and *Lethrinops lethrinus* were downloaded from the SRA (accession numbers in supplementary table S3, Supplementary Material online), transformed to fastq, trimmed, quality filtered, and mapped to the Nile tilapia genome as described above. Variant calling, filtering, and phasing were also performed as described above. For each individual (24 *P. philander* and 6 from Lake Malawi), a sequence corresponding to the first haplotype of LG7 was extracted using bcftools consensus –haplotype 1 BCFtools 1.5 (https://github.com/samtools/bcftools; last accessed January 23, 2019). Maximum likelihood inference and subsequent divergence time estimation were performed as described above. To infer if *P. philander* males and Lake Malawi males were more closely related in a specific region of LG7 than to their respective females, fixed-length phylogenies were calculated with *Twisst* (Martin and Van Belleghem 2017) using 1, 5, and 10 kb window sizes. For each window size, the support for each topology was quantified by counting the number of windows supporting strongly (100% data; >75% data) or moderately (>66% data) each topology. For comparative purposes, the same topology weighting analysis was also performed on LG6.

We extracted reads aligning to the genomic region of *gsdf* plus 2 kb up- and down-stream (*O. niloticus*: NC_031972.1: 17,568,814–17,579,211). Alleles per individual of the *gsdf* region were de novo assembled using SeCaPr (Andermann et al. 2018) and maximum likelihood phylogenies were conducted as described above. In the same way, we constructed phylogenies for eight candidate genes of sex determination (supplementary table S7, Supplementary Material online, candidate genes are marked in yellow and genomic coordinates from the reference genome are indicated).

Sex-specific variant sites for Lake Malawi cichlids were retrieved from O'Quin (2014) and visually inspected. Sequences for two XX–XY loci described by Parnell et al. (2012) and Parnell and Streelman (2013) were downloaded from SNPdb and placed on the Nile tilapia genome using Blast (Camacho et al. 2009). Marker 27028 (SNP: rs267732628) is located on scaffold NW_017615339.1: 59,608–59,966. Marker 45045 (SNP: rs267732730) is located on NC_031972.1: 1,010,601–1,010,981. We extracted raw reads corresponding to these regions with SAMtools (Li et al. 2009). BAM files were sorted and indexed using SAMtools (Li et al. 2009). Genotypes of the two SNPs for each individual (24 *P. philander* and 6 Lake Malawi cichlids) were visually inspected using SAMtools tview (Li et al. 2009).

PCR Genotyping of Lake Chila Y-Chromosomal Markers

DNA was extracted from fin clips preserved in ethanol applying a proteinase K digestion followed by a high-salt extraction (Bruford et al. 1998), or already extracted DNA from Egger et al. (2015) was used. Two potential Y-chromosomal markers (*herc3* and *K02A2.6-like*) were coamplified with the autosomal control gene *rpl7*. Primers for *herc3* (GCAAGAAAAGGCTTGTGAACC, TGACAGATACTGGGAG TGAGA), *K02A2.6-like* (GAAACTGACCTCACAGCCCA, GCCAGAAGTTTGTTTGGCGA), and *rpl7* (TGCGGGATAAA AGCGTTAGGA, ATTCCTTGCAGCAGTCATAGA) were constructed on the Lake Chila male de novo genome assembly using Primer-Blast (https://www.ncbi.nlm.nih.gov/tools/primer-blast/; last accessed January 23, 2019). PCR was performed on 5 ng of DNA in a final volume of 12.5 μl using REDTaq DNA polymerase (Sigma-Aldrich) following the manufacturer's instructions (annealing temperature 58 °C, 35 PCR cycles) Each PCR was done twice. Amplification was verified on 1.5% Tris-acetate-ethylenediaminetetraacetic-acid agarose gels with SYBR Green (ThermoFisher).

Results

Genome-Wide Statistics, Population Structure, and Demography

The 24 individuals sequenced in this study could all be assigned to previously identified mitochondrial DNA (mtDNA) haplotypes and fell into clades described by Egger et al. (2015). All specimens from Mbulu creek and eight Lake Chila specimens featured mtDNA haplotype Ht13 (supplementary fig. S1*A*, Supplementary Material online) (Egger et al. 2015), three Lake Chila specimens had Ht18 of the Kafue–Zambezi lineage; and one Lake Chila sample, MJB7, displayed Ht32 (table 1 and supplementary fig. S1*A*, Supplementary Material online).

Aligning the *P. philander* genome sequences to the *O. niloticus* reference genome resulted in 38,260,972 variant sites (SNPs and indels, table 1). The mean sequencing coverage per individual ranged from 12.7× to 16.4× (table 1) being in a range that allows accurate genotyping of heterozygous sites with the GATK multisample caller (Cheng et al. 2014; Meynert et al. 2014). A whole-genome nuclear phylogeny showed that Lake Chila and Mbulu creek populations are reciprocally monophyletic, with an estimated coalescence time of about 620,000 generations for Mbulu

Table 2
Genome-Wide Population Statistics for *Pseudocrenilabrus philander* from Lake Chila and Mbulu Creek

Statistic	Population Analysis		
	Lake Chila Versus Mbulu Creek	Within Lake Chila	Within Mbulu Creek
F_{ST}	0.538	—	—
Mean d_{xy}	0.00764	—	—
Mean d_a	0.00411		
Mean π	—	0.00512	0.00193
Mean Tajima's *D*	—	0.0515	0.4273

NOTE.—F_{ST}, relative divergence; d_{xy}, absolute divergence; d_a, net divergence; π, nucleotide diversity.

creek (95% CI: 472,000–1,357,000) and 912,000 generations for Lake Chila (95% CI: 694,000–1,996,000, supplementary table S4 and supplementary fig. S1*B*, Supplementary Material online). Genome-wide F_{ST} between the two populations was 0.538, average d_{xy} (absolute genetic divergence) was 0.00764, and average d_a (net genetic divergence) was 0.00411 (table 2). Lake and creek individuals were clearly separated on PC1 in a genome-wide PCA (fig. 1*D*). The Mbulu creek population displayed low levels of within population nucleotide diversity π (0.00193; ~2.6-fold smaller than Lake Chila) and a highly positive Tajima's *D* (0.42; ~8-fold larger than Lake Chila, table 2), indicative of an excess of haplotypes compared with the number of segregating sites, compatible with an ongoing population contraction event, that is, a bottleneck. This reduction in effective population size is further supported by the short branch lengths of the Mbulu individuals in the whole-genome phylogeny (supplementary fig. S1*B*, Supplementary Material online). Finally, the bottleneck scenario for the Mbulu creek population is corroborated by the genome-wide PCA, where all Mbulu individuals are strongly overlapping on the first and second PC axes, as well as by their identical mtDNA haplotypes (table 1, fig. 1*D*, and supplementary fig. S1*A*, Supplementary Material online).

One female of the Lake Chila population (MJB7) displayed a negative genome-wide F_{IS} (table 1 and supplementary table S5, Supplementary Material online), indicating that its heterozygosity is higher than expected under Hardy–Weinberg equilibrium; furthermore, it belongs to a different mtDNA lineage and is clearly separated from all other individuals in PC2 of the genome-wide PCA (fig. 1*D*). Taken together, this is suggestive of MJB7 being a hybrid between a Lake Chila *Pseudocrenilabrus* individual and an unknown second parent. To avoid any bias potentially induced by the high levels of heterozygosity, MJB7 was excluded from further analyses.

Interestingly, PC3 of the genome-wide PCA separated males and females from Lake Chila (fig. 1*D*). This signal cannot be explained by intralake genetic structure, as males and females share mtDNA haplotypes (table 1 and supplementary fig. S1, Supplementary Material online) and form a dense cluster on the first two PCA axes (fig. 1*D*). One phenotypic male from Lake Chila (MJA8) clustered with the Lake Chila females, suggesting that it is a sex-reversed individual (fig. 1*D*). Therefore, MJA8 was also excluded from further analyses.

LG7 Functions as a Sex Chromosome in the Lake Population of *P. philander*

Given the clear-cut separation of males and females in PC3 of the genome-wide PCA (fig. 1*D*), we next aimed to identify the genomic region responsible for the differentiation between the sexes. We first calculated genome-wide F_{ST} between males and females within each population and F_{ST} per chromosome. The average genome-wide male–female F_{ST} within the lake population was 0.04 (average male–female F_{ST} excluding LG7: 0.032), whereas the average F_{ST} for LG7 was 0.18 indicating a large region of male–female differentiation on this chromosome (fig. 2*A* and supplementary fig. S2, Supplementary Material online). Next, males and females formed distinct clades in a phylogeny on variant data of LG7 only, whereas no such grouping was found when all LGs excluding LG7 were considered (fig. 3), nor in phylogenies built from any other individual LG (supplementary fig. S3, Supplementary Material online). Furthermore, relatedness analyses (Yang et al. 2010) showed that males and females formed two distinct groups on LG7 but did not do so when all linkage groups except LG7 were considered (supplementary fig. S4, Supplementary Material online). Finally, a PCA based on LG7 only, clearly separated males and females from Lake Chila on PC2 (supplementary fig. S5*A*, Supplementary Material online), whereas the first three principal components did not separate the sexes in a PCA based on sequence information from all LGs but LG7 (supplementary fig. S5*B* and *C*, Supplementary Material online).

Contrastingly, in the bottlenecked Mbulu creek fish, the male–female F_{ST} of LG7 alone was similar to the genome-wide level (LG7: $F_{ST\ (male–female)} = 0.12$, genome-wide: $F_{ST\ (male–female)} = 0.09$, and genome-wide excluding LG7: $F_{ST\ (male–female)} = 0.077$). However, these values should be taken with caution due to the low sample size of the Mbulu creek population. Furthermore, individuals did not cluster by sex in any of the phylogenies reconstructed from individual LGs (supplementary fig. S3, Supplementary Material online), nor in a LG7 relatedness analysis (supplementary fig. S4, Supplementary Material online). Finally, we performed a topology weighting analysis, using four different "populations": Lake Chila males, Lake Chila females, Mbulu creek males, and Mbulu creek females. This analysis did not reveal any region where Chila and Mbulu males were more closely related to each other than they were to females (supplementary fig. S6, Supplementary Material online). Therefore, we did not find any evidence for a common sex locus between the two populations.

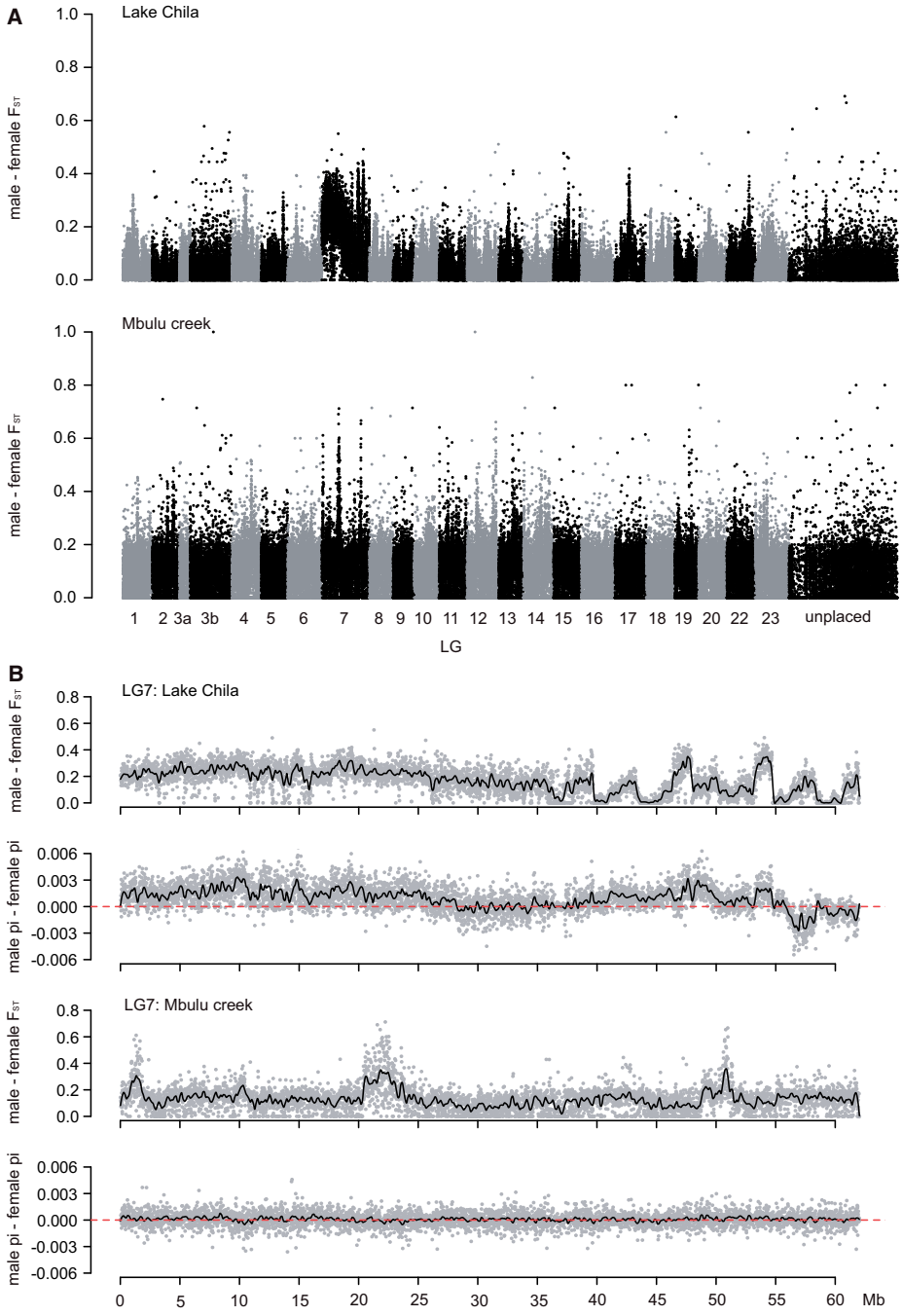


Fig. 2.—Genomic signatures of male–female differentiation in *Pseudocrenilabrus philander*. (*A*) Male–female F_{ST} for individuals from Lake Chila (upper panel) and Mbulu creek (lower panel) along the reference genome of *Oreochromis niloticus*. Each dot represents a single F_{ST} value per 10 kb window. (*B*) Male–female F_{ST} and difference in nucleotide diversity between sexes ($\pi_{diff} = \pi_{males} - \pi_{females}$) along LG7. Each gray dot represents a single value per 10 kb window. Black line: smoothed value (loess parameter = 0.01) and red line: no difference in nucleotide diversity between males and females.

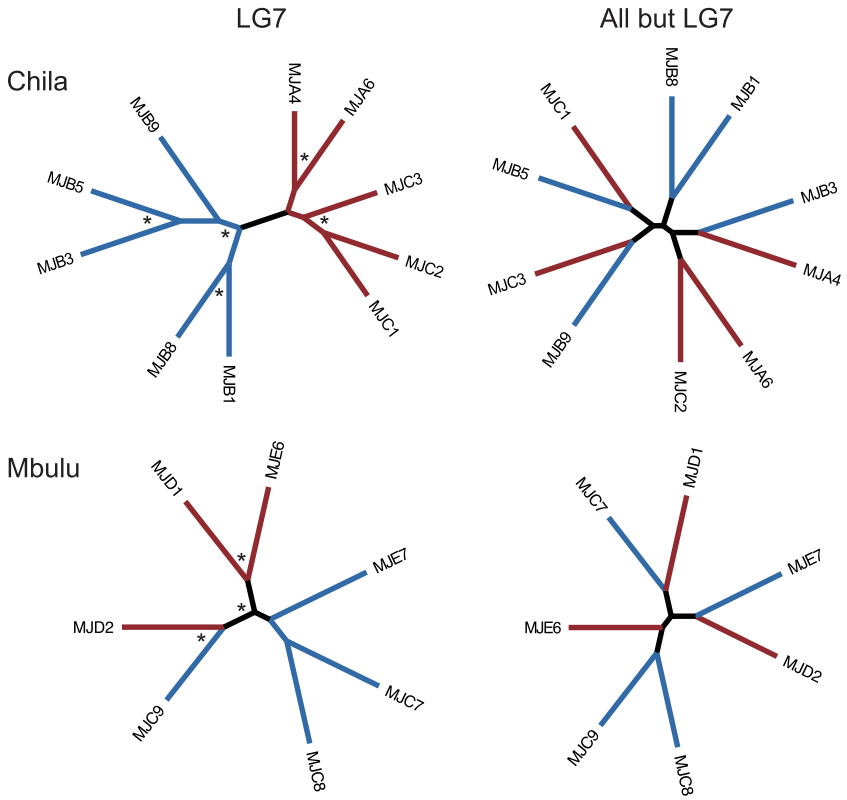


Fig. 3.—Phylogenetic analysis within the two *Pseudocrenilabrus philander* populations based on markers on LG7 and using genome-wide variants on all LGs but LG7. Maximum likelihood phylogeny of LG7 and all other LGs except LG7 for Lake Chila (upper panel) and Mbulu creek (lower panel); blue: males, red: females, and asterisks: 100% bootstrap support.

LG7 Harbors an XY System in the Lake Population

In a simple sex-chromosomal system, the heterogametic sex shares half of its sex-chromosomal alleles with the homogametic sex (e.g., X alleles in an XX–XY system and Z alleles in a ZZ–ZW system), whereas Y/W alleles are specific to the heterogametic sex. This results in an expected maximum male–female F_{ST} of 0.5 for completely sex-differentiated sites (i.e., if the allele frequency for the heterogametic sex is 0.5 and the allele frequency for the homogametic sex is 1, then the expected F_{ST} is 0.5 in an infinite population) (Brelsford et al. 2017; Fontaine et al. 2017; Rodrigues and Dufresnes 2017). Furthermore, the heterogametic sex (XY or ZW) shows an excess of heterozygous sites compared with the homogametic sex, reflected by negative F_{IS} values. Consequently, F_{ST} and F_{IS} show a negative correlation in the heterogametic sex (Rodrigues and Dufresnes 2017).

In Lake Chila *P. philander*, males had negative F_{IS} values on LG7 (table 1 and supplementary table S5, Supplementary Material online), indicating higher levels of heterozygosity in males. Furthermore, females had higher F_{IS} values on LG7 (0.31–0.45) compared with the rest of the genome excluding LG7 (0.15–0.22), denoting low levels of heterozygosity on LG7 (table 1). Males of the lake population also showed significantly higher nucleotide diversity (π) compared with females (fig. 2*B* and supplementary fig. S7, Supplementary Material online) and a negative correlation between F_{IS} and male–female F_{ST} on LG7 (supplementary fig. S8*A*, Supplementary Material online), strongly suggesting that males are the heterogametic sex and that LG7 functions as an XX–XY system.

In the Mbulu creek population, males also displayed higher π compared with females (supplementary fig. S7, Supplementary Material online), yet the male–female difference in mean π was much smaller than for males and females of Lake Chila (mean π Lake Chila males: 0.0039; mean π Lake Chila females: 0.0030; mean π Mbulu creek males: 0.00103; and mean π Mbulu creek females: 0.00094). Moreover, individual F_{IS} values did not differ between males and females on LG7. They were higher in both sexes than their corresponding

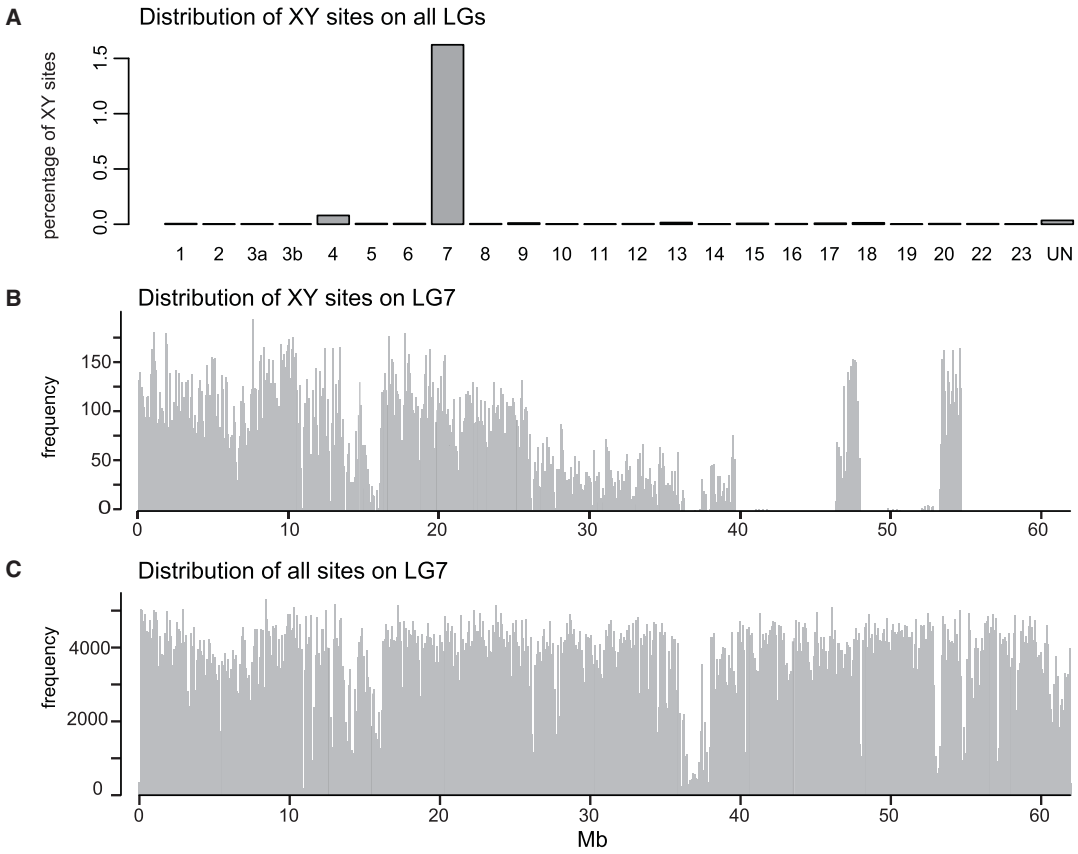


FIG. 4.—XY-sites in *Pseudocrenilabrus philander* from Lake Chila. (*A*) Distribution of potential XY sex-patterned sites across all LGs in the Lake Chila population normalized by total number of sites per LG. (*B*) Distribution of XY-sites along LG7 in 10 kb bins. (*C*) Distribution of all variant sites called on LG7 in 10 kb bins.

genome-wide estimates, however, overall lower than the female values of the Lake Chila individuals (table 1). Male–female F_{ST} on LG7 in the Mbulu population did not indicate a large region with an expected F_{ST} for sex chromosomes of 0.5 but several peaks along the chromosome of F_{ST} values above 0.5 (fig. 2*A*), which are likely false positives arising from the low sample size. Also, F_{ST} and F_{IS} along LG7 showed a positive correlation in both sexes (supplementary fig. S8*B*, Supplementary Material online). Hence, there is no indication for an XX–XY system or other sex-specific signals in the Mbulu creek population on any of the LGs.

Sex Chromosome Differentiation and the SD Region in *P. philander* from Lake Chila

To further delimit the SD region in Lake Chila fish, we identified sites that showed an XY sex-specific pattern, that is, sites for which all females are homozygous and all males heterozygous. We identified a total of 41,309 XY-patterned sites across the genome, of which the great majority (38,429; 93%) is placed on LG7 (fig. 4*A*). The XY-sites of *P. philander* were distributed along the entire chromosome with a slightly higher frequency at ~7–12 Mb, (fig. 4*B*) and less to no sites between 27 and 60 Mb with the exceptions of two peaks between 45 and 55 Mb. This block-like distribution of XY-sites might indicate regions of suppressed recombination (e.g., sex chromosome strata) (Lahn and Page 1999), probably caused by chromosomal rearrangements. Alternatively, and probably more likely, these blocks indicate chromosomal rearrangements between *P. philander* and the used reference genome *O. niloticus* and hence a difference in sequence order. The distribution of XY-patterned sites suggests that the SD region is located in the first 25 Mb of LG7. Again, we did not observe such a pattern in the Mbulu creek population (only 2013 potential XY-sites genome-wide, of which 91 are on LG7, supplementary fig. S9, Supplementary Material online).

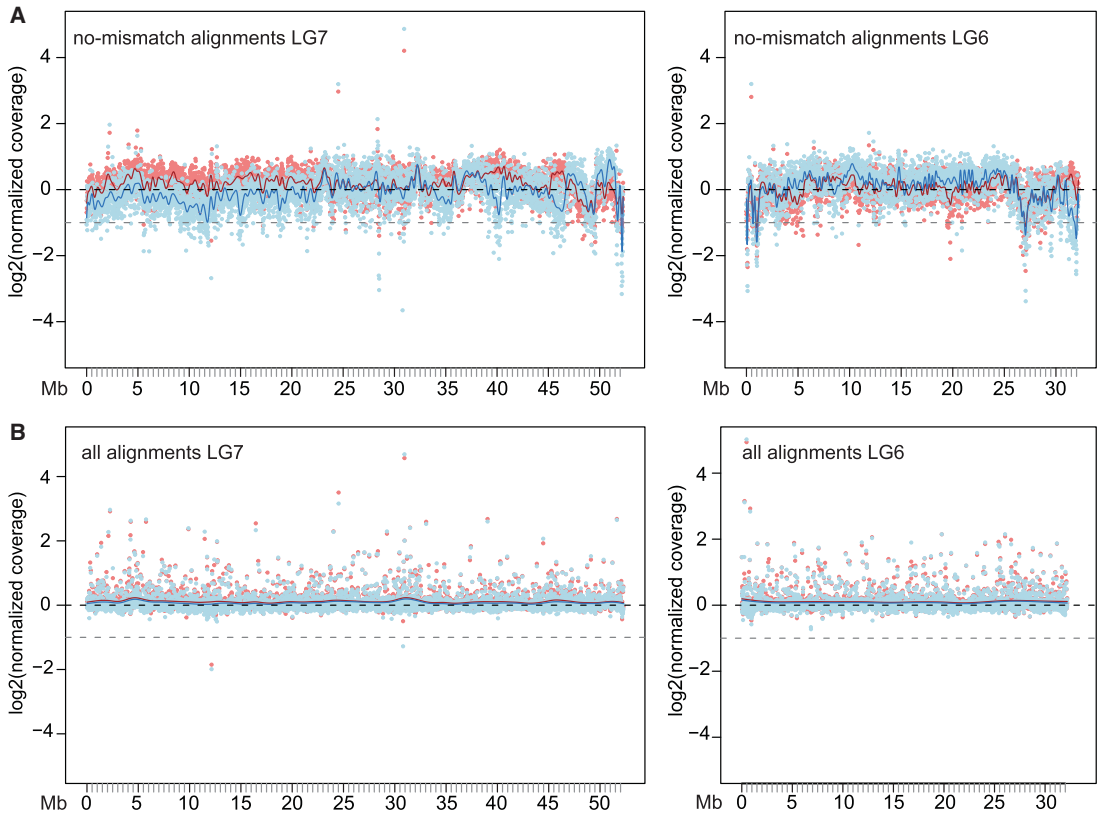


Fig. 5.—Sex chromosome coverage in *Pseudocrenilabrus philander* from Lake Chila. (*A*) Coverage of perfect alignments of males (blue) and females (red) along the de novo assembled Lake Chila female X-chromosome (left) and for comparison along the de novo assembled LG6 (right). (*B*) Coverage of all alignments of males and females along the de novo assembled Lake Chila female X-chromosome (left) and for comparison along the de novo assembled LG6 (right); red and blue lines: smoothing spline, black dotted lines: normalized coverage of 1, and gray dotted line: normalized coverage of 0.5.

The number of XY-sites in Lake Chila *P. philander* exceeded that reported for other cichlid sex-chromosomal system. In the Nile tilapia (*O. niloticus*), for example, LG1 has a 9-Mb large XY SD region, which contains 12,225 such sites (out of 38,718 total sex-differentiated sites) (Conte et al. 2017). In the blue tilapia, *Oreochromis aureus*, LG3 carries a ZZ–ZW SD system, which shows 24,983 sex-differentiated sites (total differentiated sites in the genome 103,406) (Conte et al. 2017).

As a next step, we functionally annotated the Lake Chila XY-sites to investigate the effect of variants on coding sequences. The highest density of nonsynonymous sites with "moderate" or "high effect" (i.e., coding sequence variant, frameshift, missense mutation, insertions, deletions, and inversions) was detected between 22 and 23 Mb of LG7 (supplementary fig. S10 and supplementary table S6, Supplementary Material online). The 43 "high effect" variants were located in 33 genes. As expected, Lake Chila males were heterozygous and females homozygous for these SNPs and, all Mbulu creek individuals were homozygous, matching the female lake genotype.

To further investigate the extent of sex-chromosomal differentiation and delimit the SD region, we analyzed male–female differences in sequence coverage along the sex chromosomes. To avoid any potential bias introduced by using the *O. niloticus* reference genome, we generated a male and a female draft genome assembly for the lake population. In species with heteromorphic XY sex chromosomes, the X chromosome is present in a hemizygous state in males, resulting in ~50% reduced sequencing coverage for the X in males compared with the X in females or any autosome. When all read alignments with default mapping parameters in the two sexes were considered, which is the standard approach (e.g., Vicoso et al. 2013), no difference in sequence coverage was visible along the X chromosome (fig. 5*B*, coverage follows the expected black line). This indicates that X and Y in *P. philander* from Lake Chila are at early stages of sex chromosome differentiation. However, when considering only perfect alignments (excluding alignments that contain any mismatch), a drop in male sequence coverage became

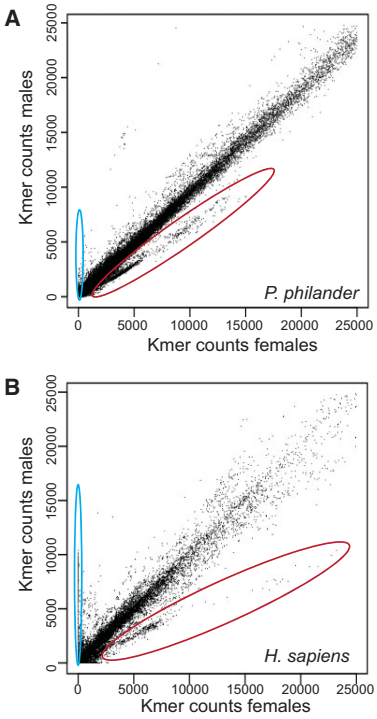


Fig. 6.—K-mer comparison in males and females. (*A*) Counts of 37 bp k-mers in male and female Lake Chila *Pseudocrenilabrus philander*. (*B*) Counts of 37 bp k-mers in human males and females. Humans have strongly differentiated sex chromosomes. K-mers derived from the Y chromosome are expected to have zero counts in females; k-mers derived from the X chromosome should have half the count in males than in females. Potential Y-k-mers are highlighted with a blue circle, X-mers with a red circle.

evident, especially in the first 20 Mb of the X chromosome (fig. 5*A*).

To further investigate this pattern of sex chromosome differentiation, we built a catalog of 37-bp-long subsequences (k-mers) and counted their presence in the male and female reads (fig. 6). Although the sex chromosomes of *P. philander* are certainly much younger and much less differentiated than the one in humans, the k-mer comparison between males and females is similar in these two species (fig. 6). X-linked k-mers are clearly visible in both species as the second largest cloud with higher counts in females than in males (fig. 6, red circle). We investigated the location of potential X-linked k-mers in the female X chromosome assembly which revealed their highest frequency at ~12.5 Mb (corresponding to ~15.4 Mb on LG7 in the reference genome, supplementary fig. S11, Supplementary Material online). Combining the analyses of XY-sites, coverage and X-linked k-mers, the SD region *of P. philander* Lake Chila is likely located at 0.3–16 Mb on LG7.

This region has 518 protein-coding gene annotations in the reference genome assembly. A full overview of these genes with corresponding gene ontologies is provided in supplementary table S7, Supplementary Material online, and genes with a potential role in SD are highlighted in yellow. These include two HMG-domain genes, a protein domain also encoded by the mammalian SD gene *Sry* (Sinclair et al. 1990), and *foxl1* and *foxd1*, belonging to the forkhead box family of transcription factors, which play a role in ovarian development and function (Ottolenghi et al. 2005; Uhlenhaut and Treier 2011). They further include *wt1*, which regulates early gonad development in mammals (Wilhelm and Englert 2002).

Two Reference-Free Approaches to Detect Y-Chromosomal Candidates in the Lake Population

In an XX–XY system, Y chromosome–specific sequences are not present in females resulting in zero sequencing coverage of such regions by female sequencing reads. We searched the male de novo genome assembly for regions of male-only coverage of at least 1 kb in length and detected 12 such regions located on 11 different scaffolds. The longest region was 2,124 bp long. When compared with the reference genome, ten of these scaffolds were placed on LG7 (eight within the first 10 Mb of LG7, supporting the analyses above that this is the SD region) and one on the unplaced scaffold NW_017613955.1. A BlastX search of the candidate regions revealed similarities to five coding sequences (the ubiquitin-protein ligase *herc3* in the 2,124-bp region, two transposable element related sequences, two uncharacterized proteins) and two ncRNAs (supplementary table S8, Supplementary Material online). In the creek population, all but three of these regions showed sequence coverage in both sexes. These three remaining regions, which included the one with *herc3*, do apparently not exist in the creek population genomes (supplementary fig. S12, Supplementary Material online).

Although X and Y are clearly differentiating in Lake Chila *P. philander*, (most of) our analyses revealed a substantial degree of sequence similarity between X and Y and also could not delimit the SD region further than to the first ~16 Mb of LG7. Our male de novo genome assembly likely contains a consensus assembly for X/Y haplotypes of LG7. When sequencing a male genome of a diploid XY species, Y-specific sequences will have reduced coverage in comparison to autosomal regions. Also, differentiating Y chromosomes typically accumulate repetitive sequences (Chalopin et al. 2015). These two factors may hamper the reconstruction of Y chromosomes using standard assembly tools (Tomaszkiewicz et al. 2017). To identify sequence information derived from Y-specific male-only regions also potentially missing in the reference genome, we applied a method described by Akagi et al. (2014) that makes use of k-mers. We extracted male-specific k-mers from the above-mentioned k-

mer catalog and used reads containing them for a targeted assembly of putative Y-chromosomal contigs. We obtained 138 Y-contigs containing 48 potential genes (supplementary table S9, Supplementary Material online), of which 38 could be functionally annotated. Strikingly, 15 of these genes (~30%) showed strong similarities to transposable elements, suggesting a higher transposable element content on the *P. philander* Y chromosome than the genome-wide average for cichlids of 16–19% (Brawand et al. 2014), a characteristic feature of sex chromosomes (Chalopin et al. 2015).

Among the other genes, we detected two genes involved in spermatogenesis, *psmb2* (Gupta 2005) and *kelch10* (Yan et al. 2004). We also recovered one of the uncharacterized proteins that we previously identified in the full de novo male assembly in a region with zero female coverage (uncharacterized protein *K02A2.6-like*), which functions in nucleic acid and zinc ion binding (supplementary tables S8 and S9, Supplementary Material online). This gene contains a *retropepsinlike* domain of invertebrate retrotransposons (DeMarco et al. 2005).

LG7 Probably Evolved Twice as a Sex Chromosome in Haplochromine Cichlids

LG7 is known to function as XX–XY system in many haplochromine species endemic to Lake Malawi (Ser et al. 2010; Parnell and Streelman 2013; Peterson et al. 2017) and likely represents the ancestral sex chromosome state of the radiation in this lake (Peterson et al. 2017). We therefore aimed to examine whether or not the Lake Malawi SD system corresponds to the one we identified in *P. philander* of Lake Chila. To this aim, we performed a topology weighting analysis on LG7 to infer if Lake Chila and Lake Malawi males were more closely related to each other compared with the females of their respective population/species in specific genomic regions. If the XX–XY system was ancestral and shared between Lake Malawi cichlids and *P. philander*, one would expect that the SD locus and closely linked loci that do not recombine between X and Y cluster by sex and not by species in a phylogeny (Stock et al. 2011). We included three species (*A. calliptera* XX–XY on LG7 [Peterson et al. 2017]; *A. stuartgranti* and *L. lethrinus*), as these represent, to the best of our knowledge, the only currently available full-genome data covering both sexes per species in Lake Malawi cichlids. Our analyses indicated no strongly supported region in which Lake Malawi and Lake Chila males were more closely related to each other than to the females of their respective species (fig. 7 and supplementary table S10, Supplementary Material online). Rather, the species topology was strongly supported for each window size on LG7, as well as for the non sex-linked LG6 (supplementary fig. S13 and supplementary table S10, Supplementary Material online).

An outstanding candidate gene for the SD locus on LG7 is *gsdf* (*gonadal soma-derived factor*), which has been described as a master SD gene in several fish species (e.g., Myosho et al. 2012; Rondeau et al. 2013). In agreement with this, Peterson et al. (2017) proposed *gsdf* as the SD gene of Lake Malawi cichlids. In another study on Lake Malawi cichlids, focusing on *Metriaclima zebra* and *M. mbenji*, O'Quin (2014) also reported sex-patterned sites in *gsdf*. We thus reconstructed a phylogeny for the *gsdf* locus in the set of Lake Malawi cichlids and *P. philander*. Again, male sequences of the different species did not group together (supplementary fig. S14, Supplementary Material online). When examining the sequences for individual sites, none of them supported a shared sex pattern (supplementary table S11, Supplementary Material online).

In a previous study on the Lake Malawi cichlids *Cynotilapia afra* and *Pseudotropheus elongates*, Parnell and Streelman (2013) detected two distinct XX–XY loci on LG7. We also inspected these XX–XY markers (RAD-tags 27028 and 45045, see figure 4 in Parnell et al. [2012] and Parnell and Streelman [2013]) in *P. philander*. The marker 45045 was homozygous (C/C) in all individuals of all species and all *P. philander*, *A. stuartgranti*, and *L. lethrinus* individuals were homozygous (C/C) for the marker 27028. The *A. calliptera* male was heterozygous (C/T) at this site and the *A. calliptera* female was homozygous (C/C), supporting an XX–XY pattern for this marker only in this species.

Finally, a full LG7 phylogeny including all 24 *P. philander* individuals and male and female individuals from Lake Malawi provides further support for a young age of *P. philander's* sex chromosomes, with a divergence time of ~423,000 generations for females and ~455,000 generations for males in LG7 (supplementary fig. S15 and supplementary table S12, Supplementary Material online). Assuming one generation per year, it is reasonable to conclude that X and Y of *P. philander* in Lake Chila diverged less than a million years ago, as the 95% confidence interval did not reach 1 Myr.

We also investigated eight additional single gene phylogenies for genes on LG7 with a potential role in sex determination identified as candidate genes in the SD region of *P. philander* from Lake Chila in this study (supplementary table S7, Supplementary Material online). Similar to *gsdf* and the topology weighting analysis, these gene trees mostly recovered the species tree (supplementary fig. S16, Supplementary Material online). Two genes showed differing topologies, however, with overall low support and not indicative of a shared sex locus.

LG7 Likely Evolved as a Sex Chromosome within Lake Chila

The LG7 system detected in the Lake Chila population likely evolved independently from the one in Lake Malawi cichlids. Furthermore, we could not detect this system in the adjacent and closely related Mbulu creek population. Given the size of our data set used for full-genome sequencing and the question the origin of this XX–XY system, we aimed to test for the presence/absence of the LG7

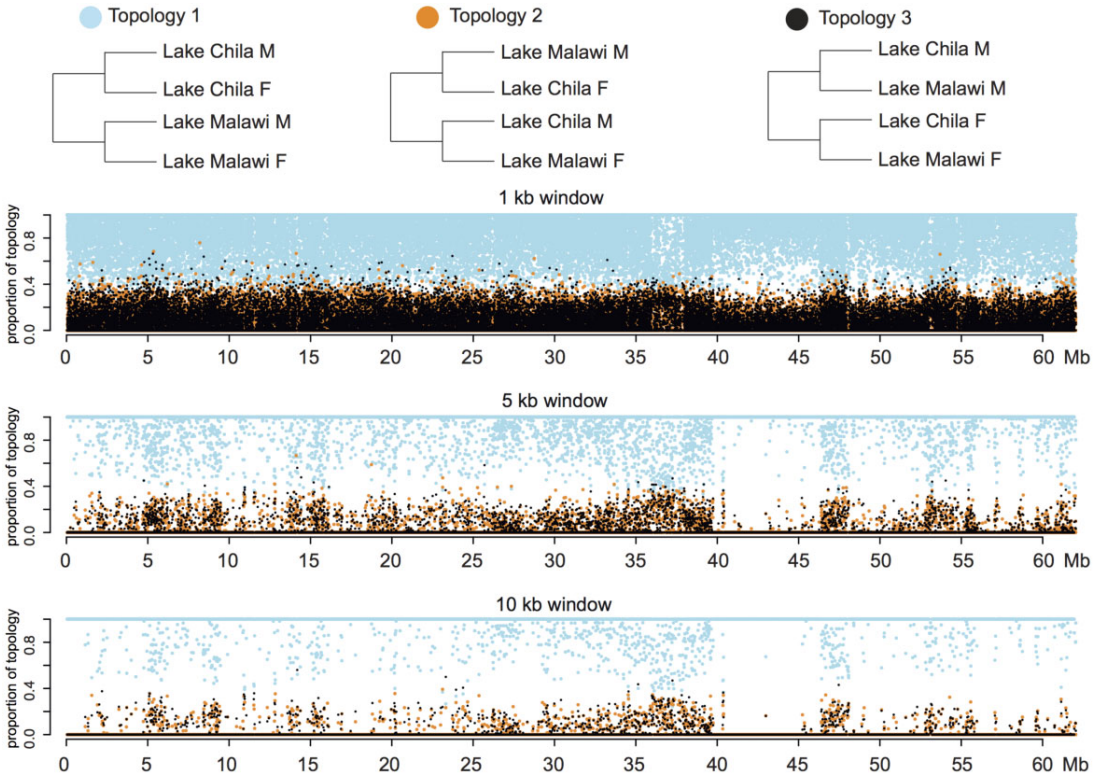


Fig. 7.—Topology weighting analysis of LG7. Topology weighting analysis using 1-, 5-, and 10-kb windows between the four "populations" Lake Chila males, Lake Chila females, Lake Malawi males, and Lake Malawi females.

XX–XY system in additional individuals of the *P. philander* species complex. We tested 78 individuals belonging to five clades of the *P. philander* species complex and *P. nicholsi* (Egger et al. 2015) by PCR for two markers which were Y chromosome linked in *P. philander* from Lake Chila, namely *herc3* (identified as the largest region absent from the female genomes) and *K02A2.6-like* (also identified as a region with zero female coverage in Lake Chila and over Y-k-mer specific assembly). Within the Lake Chila samples (additional $n = 34$), *herc3* was present in all tested males (15), and 12 males were positive for *K02A2.6-like*. All but two phenotypic females were negative for the two markers. We can thus largely confirm male sex-linkage of the two markers within Lake Chila and hence the presence of an XX–XY SD system in this population. Our PCR assay also included individuals of the two divergent mtDNA haplotype lineages. However, all populations other than Lake Chila did not show sex linkage for the two markers, which were either present or absent in both sexes (supplementary table S13 and supplementary fig. S17, Supplementary Material online).

Discussion

Cichlid fishes display a breathtaking diversity in basically every phenotypic trait investigated so far including, coloration, morphology, habitat use, breeding systems, or diet (Albertson and Kocher 2006; Sefc 2011; Muschick et al. 2012; Miyagi and Terai 2013; Salzburger 2018) and sex determination is likely another flexible property of this astonishing group of fish. Here, we investigated sex chromosome evolution in a phylogeographically complex species, the haplochromine cichlid *P. philander* (Egger et al. 2015). We detected an XX–XY system in the Lake Chila *P. philander* population, whereas this signature was not detectable in the genomes of an adjacent riverine stock. The creek population likely underwent a genetic bottleneck, so it is possible that the apparent absence of any detectable SD system in this population may be due to demographic events. The creek population may have been founded by XX individuals only, or XY recombination resumed in the creek population. However, markers that were male specific in Lake Chila did not show a sex-specific pattern in specimen from six other *P. philander* populations nor in *P. nicholsi*. Given the nested placement of the Lake Chila population

within the *P. philander* species complex (supplementary fig. S1, Supplementary Material online), the most parsimonious explanation for this pattern is that the XX–XY system evolved or at least differentiated within the Lake Chila population.

In agreement with this scenario, we could also not find support for a shared (ancestral) XX–XY LG7 system between Lake Malawi cichlids, and *P. philander* from Lake Chila neither in our divergence time estimates nor in a topology weighting analysis nor in single gene phylogenies for candidate genes of sex determination. We therefore propose that LG7 evolved repeatedly (convergently) as a sex chromosome in different lineages of haplochromine cichlids. This would lend further support to the limited options theory, that is, that certain chromosomes are particularly well suited to become sex chromosomes and evolve as such more often than other chromosomes (Marshall Graves and Peichel 2010). Marshall Graves and Peichel proposed that a likely candidate for a "limited option" is the ancestral teleost chromosome TEL6 (Marshall Graves and Peichel 2010). The sex chromosomes of several fish species are derived from TEL6 including those of the medaka *Oryzias luzonensis*, the sablefish *Anoplomba fimbria* as well as the guppy *Poecilia reticulata* (Marshall Graves and Peichel 2010; Myosho et al. 2012; Rondeau et al. 2013). The marker *SLC45A2* that Marshall Graves and Peichel (2010) used to identify these sex chromosomes as being syntenic to TEL6 is indeed located on LG7 of cichlids (*O. niloticus* LG7: 17,405,957–17,428,885). Together with our study, LG7/TEL6 has been described three times as a sex chromosome in cichlids, suggesting that TEL6 evolved to become a sex chromosome in at least five lineages of teleost fish (Lake Malawi cichlids Ser et al. 2010, Lake Tanganyika cichlid *Hemibates stenosoma* Gammerdinger et al. 2018a, *P. philander* Lake Chila, medaka, guppy, and sablefish) supporting the "limited options" theory. However, other genes on cichlid LG7 are syntenic to TEL 7 (Marshall Graves and Peichel 2010), indicating additional rearrangements of LG7 in cichlids or in the lineage leading towards them. Certainly more data on cichlid sex chromosomes is needed to properly test the "limited options" theory within cichlids.

With our limited data set, we cannot exclude the presence of yet another SD system in the other *P. philander* populations or that the LG7 XX–XY system was present also in the creek but has secondarily been lost or started to recombine again. In addition, in the genome-wide data of the creek population we failed to detect any other sex-chromosomal system. It is possible that a SD region in this population is too small to be detected with our limited sample size. We can also not exclude that this population might rely on an ESD system or a multifactorial combination of environmental and genetic factors. We also report a sex-reversed individual in Lake Chila, as well as mismatches between phenotypic sex and Y chromosome markers in a PCR genotypic assay. If we exclude any sexing errors, this could mean an occurrence of 6–10% of individuals that also do not underlie the XX–XY system within Lake Chila. It could also mean that the markers we tested by PCR genotyping still recombine and are hence not fully Y-linked. Note that the two markers also differed in their presence–absence pattern with *K02A2.6-like* showing only two genotype–phenotype mismatches. It might thus be closer to the actual SD locus than *herc3*. Further (genome-wide) data would be needed to support either of these scenarios.

Still, specimens from the Lake Chila population showed clear signs of sex chromosome differentiation along large sections of LG7, especially in the first 16 Mb. Yet, there were also peaks in male–female F_{ST}, XY sex–patterned sites as well as male-reduced coverage in other regions along LG7. This block-like distribution of signatures of differentiation (especially visible at 45–50 and 53–55 Mb; figs. 2 and 4) might reflect "sex-chromosome strata," which are parts of a chromosome that stopped recombining at different points of time in the past (Lahn and Page 1999). This strata formation can result from chromosomal rearrangements such as inversions, which immediately cause suppression of recombination (Sturtevant 1921). Alternatively and probably more likely, these blocks result from genome rearrangements between the reference genome *O. niloticus* and *P. philander*.

We identified several candidate genes for the SD locus in *P. philander* from Lake Chila based on male-specific sequence features. Among these, the most promising ones were *herc3* and the uncharacterized gene *K02A2.6-like*. We would like to point out that *herc3* is also located on the sex chromosomes in another fish, the medaka (Kondo et al. 2006), but is not the master SD gene in this species. We could not find any support for the previously known SD gene *gsdf* as the master SD locus in *P. philander*. The dating of the split between X and Y chromosomes in *P. philander* from Lake Chila to $<\sim$1 Myr suggests a similar age as the one proposed for the origin of the XX–XY system on LG7 in Lake Malawi cichlids (Peterson et al. 2017). When compared with the ZZ–ZW sex determination system on LG3 of another cichlid, the blue tilapia *O. aureus* (Conte et al. 2017), we found that there are more sex-patterned sites in *P. philander* than in *O. aureus*. This suggests a higher level of sex chromosome differentiation in *P. philander*. The LG3 sex chromosome system is ancestral in the Oreochromini lineage (Lee et al. 2004; Cnaani et al. 2008; Cnaani 2013), dating back to the split before *O. aureaus* and *O. niloticus*, estimated to $\sim$3 Ma (Xiao et al. 2015). Also, our comparison of male–female k-mer compositions in *P. philander* and humans points to a remarkable level of differentiation of the *P. philander* sex chromosomes despite their probably young age.

Autosomes are recruited as sex chromosomes and subsequently follow the path of sex chromosome differentiation as in *P. philander* or the *Oreochromini* lineage. Demographic events such as lake colonizations or population size fluctuations might impact the patterns of differentiation. Under which conditions a differentiated sex chromosome system represents a selective advantage remains an open question,

at least for cichlids. Elegant work on sticklebacks demonstrated that newly evolving sex chromosomes contribute to phenotypic divergence and reproductive isolation between sympatric species, probably facilitating speciation (Yoshida et al. 2014). Whether or not the XX–XY system of the Lake Chila individuals causes this population to be reproductively incompatible with other populations remains to be tested. Taken together, our study highlights the contrast between genomic signatures that fit the canonical view on sex chromosome evolution (recombination suppression and sequence differentiation) and the instability that such systems nevertheless face. Remarkably, we show that sex-chromosomal systems can differ within a single cichlid species, at the level of geographically separated populations (see also Böhne et al. [2016]), suggesting that demographic events can impact sex chromosome evolution and, vice versa, that changes in SD systems might contribute to diversification.

Supplementary Material

Supplementary data are available at *Genome Biology and Evolution* online.

Acknowledgments

Calculations were performed at sciCORE (http://scicore.unibas.ch/) scientific computing core facility at University of Basel, with support by the SIB Swiss Institute of Bioinformatics. We especially thank P. Escobar Lopez for bioinformatics support. We thank G. Pauquet and K. Smailus for help with fieldwork and the Lake Tanganyika Research Unit, Department of Fisheries, Republic of Zambia, for research permits and help with fieldwork, and V. Huwiler, P. Lassen, and their crew for logistical support in Zambia. We thank F. Ronco for help with library preparation and the Comai lab and especially M. Lieberman for sharing scripts with us. We thank P. Cingolani for extending the SNPeff database for us and A. El Taher and O.K. Tørresen for advice on scripting. A.B. was funded by the SNSF Ambizione grant PZ00P3_161462. A.A.T.W., J.R. and B.E. were funded through the SNSF grant 1003A_156405 to W.S.; W.S. further received funding from the European Research Council (CoG 617585 "CICHLID∼X"). These latter two grants funded field trips and research expenses. A.A.T.W. further received funding by the Australian Department of Education and Training (Endeavour Research Fellowship 6534_2018). We thank three anonymous reviewers and the editorial team for valuable comments on previous versions of this article.

Author Contributions

A.B., A.A.T.W., and W.S. designed the study. A.A.T.W., J.R., and B.E. carried out fieldwork. A.A.T.W. and J.R. constructed sequencing libraries. A.B., B.E., M.R., and A.R. performed PCR genotyping. A.B., A.A.T.W., and B.E. analyzed the data. A.B. finalized the manuscript. All authors contributed to manuscript drafting.

Literature Cited

Akagi T, Henry IM, Tao R, Comai L. 2014. A Y-chromosome-encoded small RNA acts as a sex determinant in persimmons. Science 346(6209):646–650.
Albertson RC, Kocher TD. 2006. Genetic and developmental basis of cichlid trophic diversity. Heredity 97(3):211–221.
Andermann T, Cano A, Zizka A, Bacon C, Antonelli A. 2018. SECAPR-a bioinformatics pipeline for the rapid and user-friendly processing of targeted enriched Illumina sequences, from raw reads to alignments. PeerJ 6:e5175.
Bellott DW, Page DC. 2009. Reconstructing the evolution of vertebrate sex chromosomes. Cold Spring Harb Symp Quant Biol. 74:345–353.
Blackmon H, Ross L, Bachtrog D. 2017. Sex determination, sex chromosomes, and karyotype evolution in insects. J Hered. 108(1):78–93.
Blaser O, Grossen C, Neuenschwander S, Perrin N. 2013. Sex-chromosome turnovers induced by deleterious mutation load. Evolution 67(3):635–645.
Böhne A, Wilson CA, Postlethwait JH, Salzburger W. 2016. Variations on a theme: genomics of sex determination in *Astatotilapia burtoni*. BMC Genomics. 17(1):883.
Bolger AM, Lohse M, Usadel B. 2014. Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics 30(15):2114–2120.
Brawand D, et al. 2014. The genomic substrate for adaptive radiation in African cichlid fish. Nature 513(7518):375–381.
Brelsford A, Lavanchy G, Sermier R, Rausch A, Perrin N. 2017. Identifying homomorphic sex chromosomes from wild-caught adults with limited genomic resources. Mol Ecol Resour. 17(4):752–759.
Browning BL, Browning SR. 2016. Genotype imputation with millions of reference samples. Am J Hum Genet. 98(1):116–126.
Browning SR, Browning BL. 2007. Rapid and accurate haplotype phasing and missing-data inference for whole-genome association studies by use of localized haplotype clustering. Am J Hum Genet. 81(5):1084–1097.
Bruford MWO, Hanotte O, Brookfield JFY, Burke T. 1998. Multilocus and single-locus DNA fingerprinting; molecular genetic analysis of populations, a practical approach. Oxford: Oxford University Press.
Camacho C, et al. 2009. BLAST+: architecture and applications. BMC Bioinformatics 10:421.
Capel B. 2017. Vertebrate sex determination: evolutionary plasticity of a fundamental switch. Nat Rev Genet. 18(11):675–689.
Carvalho AB, Clark AG. 2013. Efficient identification of Y chromosome sequences in the human and *Drosophila* genomes. Genome Res. 23(11):1894–1907.
Cavoto E, Neuenschwander S, Goudet J, Perrin N. 2018. Sex-antagonistic genes, XY recombination, and feminized Y chromosomes. J Evol Biol. 31(3):416–427.
Chalopin D, Volff JN, Galiana D, Anderson JL, Schartl M. 2015. Transposable elements and early evolution of sex chromosomes in fish. Chromosome Res. 23(3):545–560.
Charlesworth B, Charlesworth D. 1978. A model for the evolution of dioecy and gynodioecy. Am Nat. 112(988):975–997.
Charlesworth B, Charlesworth D. 2000. The degeneration of Y chromosomes. Philos Trans R Soc Lond B Biol Sci. 355(1403):1563–1572.
Charlesworth B, Coyne JA, Barton NH. 1987. The relative rates of evolution of sex chromosomes and autosomes. Am Nat. 130(1):113–146.
Charlesworth D. 2017. 2017. Evolution of recombination rates between sex chromosomes. Philos Trans R Soc Lond B Biol Sci. 372(1736):20160456.

Charlesworth D, Charlesworth B, Marais G. 2005. Steps in the evolution of heteromorphic sex chromosomes. Heredity 95(2):118–128.
Cheng AY, Teo YY, Ong RT. 2014. Assessing single nucleotide variant detection and genotype calling on whole-genome sequenced individuals. Bioinformatics 30(12):1707–1713.
Chevreux B, Wetter T, Suhai S. 1999. Genome sequence assembly using trace signals and additional sequence information. Proceedings of the German Conference on Bioinformatics. 99(1):45–56.
Cingolani P, et al. 2012. A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff: SNPs in the genome of *Drosophila melanogaster* strain w1118; iso-2; iso-3. Fly 6(2):80–92.
Cnaani A. 2013. The tilapias' chromosomes influencing sex determination. Cytogenet Genome Res. 141(2–3):195–205.
Cnaani A, et al. 2008. Genetics of sex determination in tilapiine species. Sex Dev. 2(1):43–54.
Conesa A, et al. 2005. Blast2GO: a universal tool for annotation, visualization and analysis in functional genomics research. Bioinformatics 21(18):3674–3676.
Conte MA, Gammerdinger WJ, Bartie KL, Penman DJ, Kocher TD. 2017. A high quality assembly of the Nile tilapia (*Oreochromis niloticus*) genome reveals the structure of two sex determination regions. BMC Genomics. 18(1):341.
Danecek P, et al. 2011. The variant call format and VCFtools. Bioinformatics 27(15):2156–2158.
DeMarco R, Machado AA, Bisson AW, Verjovski-Almeida S. 2005. Identification of 18 new transcribed retrotransposons in *Schistosoma mansoni*. Biochem Biophys Res Commun. 333(1):230–240.
Dufresnes C, Stock M, Brelsford A, Perrin N. 2014. Range-wide sex-chromosome sequence similarity supports occasional XY recombination in European tree frogs (*Hyla arborea*). PLoS One 9(6):e97959.
Egger B, et al. 2015. Phylogeographic and phenotypic assessment of a basal haplochromine cichlid fish from Lake Chila, Zambia. Hydrobiologia 748(1):171–184.
Feulner PGD, Schwarzer J, Haesler MP, Meier JI, Seehausen O. 2018. A dense linkage lap of Lake Victoria cichlids improved the *Pundamilia* genome assembly and revealed a major QTL for sex-setermination. G3 (Bethesda) 8:2411–2420.
Fontaine A, et al. 2017. Extensive genetic differentiation between homomorphic sex chromosomes in the mosquito vector, *Aedes aegypti*. Genome Biol Evol. 9(9):2322–2335.
Gammerdinger W, et al. 2018a. Novel sex chromosomes in 3 cichlid fishes from Lake Tanganyika. J Hered. 109(5):489–500.
Gammerdinger WJ, Conte MA, Sandkam BA, Penman DJ, Kocher TD. 2018b. Characterization of sex chromosomes in three deeply diverged species of Pseudocrinlabrinae (Teleostei: Cichlidae). Hydrobiologia Advance Access published September 26, 2018. https://doi.org/10.1007/s10750-018-3778-6.
Garg SG, Martin WF. 2016. Mitochondria, the cell cycle, and the origin of sex via a syncytial eukaryote common ancestor. Genome Biol Evol. 8(6):1950–1970.
Graves JAM. 2006. Sex chromosome specialization and degeneration in mammals. Cell 124(5):901–914.
Graves JAM. 2008. Weird animal genomes and the evolution of vertebrate sex and sex chromosomes. Annu Rev Genet. 42(1):565–586.
Graves JAM. 2014. Avian sex, sex chromosomes, and dosage compensation in the age of genomics. Chromosome Res. 22(1):45–47.
Guerrero RF, Kirkpatrick M, Perrin N. 2012. Cryptic recombination in the ever-young sex chromosomes of Hylid frogs. J Evol Biol. 25(10):1947–1954.
Gupta GS. 2005. Proteomics of spermatogenesis. New York: Springer.
Gurevich A, Saveliev V, Vyahhi N, Tesler G. 2013. QUAST: quality assessment tool for genome assemblies. Bioinformatics 29(8):1072–1075.
Heitman J. 2015. Evolution of sexual reproduction: a view from the fungal kingdom supports an evolutionary epoch with sex before sexes. Fungal Biol Rev. 29(3–4):108–117.
Henning F, Meyer A. 2014. The evolutionary genomics of cichlid fishes: explosive speciation and adaptation in the postgenomic era. Annu Rev Genomics Hum Genet. 15:417–414.
Henningsen A, Toomet O. 2016. miscTools: miscellaneous tools and utilities. Available from: https://cran.r-project.org/web/packages/miscTools/index.html; last accessed January 23, 2019.
Heule C, Salzburger W, Böhne A. 2014. Genetics of sexual development—an evolutionary playground for fish. Genetics 196(3):579–591.
Hoff KJ, Stanke M. 2013. WebAUGUSTUS—a web service for training AUGUSTUS and predicting genes in eukaryotes. Nucleic Acids Res. 41(Web Server issue):W123–W128.
Jeffries DL, et al. 2018. A rapid rate of sex-chromosome turnover and non-random transitions in true frogs. Nat Commun. 9(1):4088.
Kamiya T, et al. 2012. A trans-species missense SNP in *amhr2* is associated with sex determination in the tiger pufferfish, *Takifugu rubripes* (Fugu). PLoS Genet. 8(7):e1002798.
Katoh K, Rozewicki J, Yamada KD. 2017. MAFFT online service: multiple sequence alignment, interactive sequence choice and visualization. Brief Bioinformatics. bbx108.
Katongo C, Koblmüller S, Duftner N, Makasa L, Sturmbauer C. 2005. Phylogeography and speciation in the *Pseudocrenilabrus philander* species complex in Zambian Rivers. Hydrobiologia 542(1):221–233.
Kiełbasa SM, Wan R, Sato K, Horton P, Frith MC. 2011. Adaptive seeds tame genomic sequence comparison. Genome Res. 21(3):487–493.
Kitano J, Peichel C. 2012. Turnover of sex chromosomes and speciation in fishes. Environ Biol Fishes 94(3):549–558.
Kiuchi T, et al. 2014. A single female-specific piRNA is the primary determiner of sex in the silkworm. Nature 509(7502):633–636.
Koblmuller S, Katongo C, Phiri H, Sturmbauer C. 2012. Past connection of the upper reaches of a Lake Tanganyika tributary with the upper Congo drainage suggested by genetic data of riverine cichlid fishes. Afr Zool. 47:182–186.
Kondo M, et al. 2006. Genomic organization of the sex-determining and adjacent regions of the sex chromosomes of medaka. Genome Res. 16(7):815–826.
Kornfield I, Smith P. 2000. African cichlid fishes: model system for evolutionary biology. Annu Rev Ecol Syst. 31(1):163–196.
Kudo Y, et al. 2015. A microsatellite-based genetic linkage map and putative sex-determining genomic regions in Lake Victoria cichlids. Gene 560(2):156–164.
Lahn BT, Page DC. 1999. Four evolutionary strata on the human X chromosome. Science 286(5441):964–967.
Lee BY, Hulata G, Kocher TD. 2004. Two unlinked loci controlling the sex of blue tilapia (*Oreochromis aureus*). Heredity 92(6):543–549.
Li D, Liu C-M, Luo R, Sadakane K, Lam T-W. 2015. MEGAHIT: an ultra-fast single-node solution for large and complex metagenomics assembly via succinct de Bruijn graph. Bioinformatics 31(10):1674–1676.
Li H, Durbin R. 2009. Fast and accurate short read alignment with Burrows–Wheeler transform. Bioinformatics 25(14):1754–1760.
Li H, et al. 2009. The sequence alignment/map format and SAMtools. Bioinformatics 25(16):2078–2079.
Lode T. 2012. Oviparity or viviparity? That is the question … Reprod Biol. 12:259–264.
Magoc T, Salzberg SL. 2011. FLASH: fast length adjustment of short reads to improve genome assemblies. Bioinformatics 27(21):2957–2963.
Malmstrøm M, Matschiner M, Tørresen OK, Jakobsen KS, Jentoft S. 2017. Whole genome sequencing data and de novo draft assemblies for 66 teleost species. Sci Data 4:160132.

Marshall Graves JA, Peichel CL. 2010. Are homologies in vertebrate sex determination due to shared ancestry or to limited options? Genome Biol. 11(4):205.
Martin SH, Van Belleghem SM. 2017. Exploring evolutionary relationships across the genome using topology weighting. Genetics 206(1):429–438.
McKenna A, et al. 2010. The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. Genome Res. 20(9):1297–1303.
Milansky et al. 2018. Whole-genome sequences of Malawi cichlids reveal multiple radiations interconnected by gene flow. Nat Ecol Evol. 2(12):1940–1955.
Meynert AM, Ansari M, FitzPatrick DR, Taylor MS. 2014. Variant detection sensitivity and biases in whole genome and exome sequencing. BMC Bioinformatics 15:247.
Miura I. 2007. An evolutionary witness: the frog *Rana rugosa* underwent change of heterogametic sex from XY male to ZW female. Sex Dev. 1(6):323–331.
Miyagi R, Terai Y. 2013. The diversity of male nuptial coloration leads to species diversity in Lake Victoria cichlids. Genes Genet Syst. 88(3):145–153.
Moore EC, Roberts RB. 2013. Polygenic sex determination. Curr Biol. 23(12):R510–R512.
Muller HJ. 1918. Genetic variability, twin hybrids and constant hybrids in a case of balanced lethal factors. Genetics 3(5):422–499.
Muller HJ. 1932. Some genetic aspects of sex. Am Nat. 66(703):118–138.
Muschick M, Indermaur A, Salzburger W. 2012. Convergent evolution within an adaptive radiation of cichlid fishes. Curr Biol. 22(24):2362–2368.
Myers EW, et al. 2000. A whole-genome assembly of *Drosophila*. Science 287(5461):2196–2204.
Myosho T, et al. 2012. Tracing the emergence of a novel sex-determining gene in medaka, *Oryzias luzonensis*. Genetics 191(1):163–170.
Nei M, Li WH. 1979. Mathematical model for studying genetic variation in terms of restriction endonucleases. Proc Natl Acad Sci U S A. 76(10):5269–5273.
O'Quin CT. 2014. The genetic basis of pigment pattern differentiation in Lake Malawi African cichlids. College Park (MD): University of Maryland.
Ottolenghi C, et al. 2005. Foxl2 is required for commitment to ovary differentiation. Hum Mol Genet. 14(14):2053–2062.
Pannell JR. 2017. Plant sex determination. Curr Biol. 27(5):R191–R197.
Parnell NF, Hulsey CD, Streelman JT. 2012. The genetic basis of a complex functional system. Evolution 66(11):3352–3366.
Parnell NF, Streelman JT. 2013. Genetic interactions controlling sex and color establish the potential for sexual conflict in Lake Malawi cichlid fishes. Heredity 110(3):239–246.
Patterson N, Price AL, Reich D. 2006. Population structure and eigenanalysis. PLoS Genet. 2:2074–2093.
Perrin N. 2016. Random sex determination: when developmental noise tips the sex balance. Bioessays 38(12):1218–1226.
Peterson EN, Cline ME, Moore EC, Roberts NB, Roberts RB. 2017. Genetic sex determination in *Astatotilapia calliptera*, a prototype species for the Lake Malawi cichlid radiation. Naturwissenschaften 104(5–6):41.
Posada D. 2008. jModelTest: phylogenetic model averaging. Mol Biol Evol. 25(7):1253–1256.
R Core Team. 2017. R: A language and environment for statistical computing. Vienna (Austria): R Foundation for Statistical Computing. Available from: https://www.r-project.org/; last accessed January 23, 2019.
Reddon AR, Hurd PL. 2013. Water pH during early development influences sex ratio and male morph in a West African cichlid fish, *Pelvicachromis pulcher*. Zoology (Jena) 116(3):139–143.
Rice P, Longden I, Bleasby A. 2000. EMBOSS: the European molecular biology open software suite. Trends Genet. 16(6):276–277.
Roberts NB, et al. 2016. Polygenic sex determination in the cichlid fish *Astatotilapia burtoni*. BMC Genomics. 17(1):835.
Roberts RB, Ser JR, Kocher TD. 2009. Sexual conflict resolved by invasion of a novel sex determiner in Lake Malawi cichlid fishes. Science 326(5955):998–1001.
Rodrigues N, Dufresnes C. 2017. Using conventional F-statistics to study unconventional sex-chromosome differentiation. PeerJ 5:e3207.
Römer U, Beisenherz W. 1996. Environmental determination of sex in Apistogrammai (Cichlidae) and two other freshwater fishes (Teleostei). J Fish Biol. 48:714–725.
Rondeau E, et al. 2013. Genomics of sablefish (*Anoplopoma fimbria*): expressed genes, mitochondrial phylogeny, linkage map and identification of a putative sex gene. BMC Genomics. 14:452.
Ronquist F, et al. 2012. MrBayes 3.2: efficient Bayesian phylogenetic inference and model choice across a large model space. Syst Biol. 61(3):539–542.
Salzburger W. 2018. Understanding explosive diversification through cichlid fish genomics. Nat Rev Genet. 19(11):705–717.
Salzburger W, Mack T, Verheyen E, Meyer A. 2005. Out of Tanganyika: genesis, explosive speciation, key-innovations and phylogeography of the haplochromine cichlid fishes. BMC Evol Biol. 5:17.
Salzburger W, Meyer A. 2004. The species flocks of East African cichlid fishes: recent advances in molecular phylogenetics and population genetics. Naturwissenschaften 91(6):277–290.
Schartl M, Schmid M, Nanda I. 2016. Dynamics of vertebrate sex chromosome evolution: from equal size to giants and dwarfs. Chromosoma 125(3):553–571.
Schwarzer J, Misof B, Tautz D, Schliewen UK. 2009. The root of the East African cichlid radiations. BMC Evol Biol. 9:186.
Schwarzer J, et al. 2012. Repeated trans-watershed hybridization among haplochromine cichlids (Cichlidae) was triggered by Neogene landscape evolution. Proc Biol Sci. 279(1746):4389–4398.
Seehausen O. 2015. Process and pattern in cichlid radiations—inferences for understanding unusually high rates of evolutionary diversification. New Phytol. 207(2):304–312.
Sefc KM. 2011. Mating and parental care in Lake Tanganyika's cichlids. Int J Evol Biol. 2011:470875.
Ser JR, Roberts RB, Kocher TD. 2010. Multiple interacting loci control sex determination in Lake Malawi cichlid fish. Evolution 64(2):486–501.
Sessions SK, Bizjak Mali L, Green DM, Trifonov V, Ferguson-Smith M. 2016. Evidence for sex chromosome turnover in proteid salamanders. Cytogenet Genome Res. 148(4):305–313.
Simao FA, Waterhouse RM, Ioannidis P, Kriventseva EV, Zdobnov EM. 2015. BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. Bioinformatics 31(19):3210–3212.
Sinclair AH, et al. 1990. A gene from the human sex-determining region encodes a protein with homology to a conserved DNA-binding motif. Nature 346(6281):240–244.
Speijer D, Lukes J, Elias M. 2015. Sex is a ubiquitous, ancient, and inherent attribute of eukaryotic life. Proc Natl Acad Sci U S A. 112(29):8827–8834.
Stamatakis A. 2014. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. Bioinformatics 30(9):1312–1313.
Stock M, et al. 2011. Ever-young sex chromosomes in European tree frogs. PLoS Biol. 9:e1001062.
Sturtevant AH. 1921. A case of rearrangement of genes in *Drosophila*. Proc Natl Acad Sci U S A. 7(8):235–237.
Tomaszkiewicz M, Medvedev P, Makova KD. 2017. Y and W chromosome assemblies: approaches and discoveries. Trends Genet. 33(4):266–282.

Turner GF, Seehausen O, Knight ME, Allender CJ, Robinson RL. 2001. How many species of cichlid fishes are there in African lakes? Mol Ecol. 10(3):793–806.
Uhlenhaut NH, Treier M. 2011. Forkhead transcription factors in ovarian function. Reproduction 142(4):489–495.
van Doorn GS, Kirkpatrick M. 2007. Turnover of sex chromosomes induced by sexual conflict. Nature 449(7164):909–912.
van Doorn GS, Kirkpatrick M. 2010. Transitions between male and female heterogamety caused by sex-antagonistic selection. Genetics 186(2):629–645.
Verheyen E, Salzburger W, Snoeks J, Meyer A. 2003. Origin of the superflock of cichlid fishes from Lake Victoria, East Africa. Science 300(5617):325–329.
Vicoso B, Emerson JJ, Zektser Y, Mahajan S, Bachtrog D. 2013. Comparative sex chromosome genomics in snakes: differentiation, evolutionary strata, and lack of global dosage compensation. PLoS Biol 11(8):e1001643.
Villesen P. 2007. FaBox: an online toolbox for FASTA sequences. Mol Ecol Notes 7(6):965–968.
Volff J-N, Nanda I, Schmid M, Schartl M. 2007. Governing sex determination in fish: regulatory putsches and ephemeral dictators. Sex Dev. 1(2):85–99.
Westergaard M. 1958. The mechanism of sex determination in dioecious flowering plants. Adv Genet. 9:217–281.
Wickham H. 2007. Reshaping data with the reshape package. J Stat Softw. 21(12):1–20.
Wickham H. 2009. ggplot2: elegant graphics for data analysis. New York: Springer Verlag.
Wilhelm D, Englert C. 2002. The Wilms tumor suppressor WT1 regulates early gonad development by activation of Sf1. Genes Dev. 16(14):1839–1851.
Xiao J, et al. 2015. Transcriptome analysis revealed positive selection of immune-related genes in tilapia. Fish Shellfish Immunol. 44(1):60–65.
Yan W, Ma L, Burns KH, Matzuk MM. 2004. Haploinsufficiency of kelch-like protein homolog 10 causes infertility in male mice. Proc Natl Acad Sci U S A. 101(20):7793–7798.
Yang JA, et al. 2010. Common SNPs explain a large proportion of the heritability for human height. Nat Genet. 42(7):565–569.
Yoshida K, et al. 2011. B chromosomes have a functional effect on female sex determination in Lake Victoria cichlid fishes. PLoS Genet. 7(8):e1002203.
Yoshida K, et al. 2014. Sex chromosome turnover contributes to genomic divergence between incipient stickleback species. PLoS Genet. 10(3):e1004223.
Zeileis A, Grothendieck G. 2005. zoo: S3 infrastructure for regular and irregular time series. J Stat Softw. 14:1–27.

Associate editor: Laura A. Katz

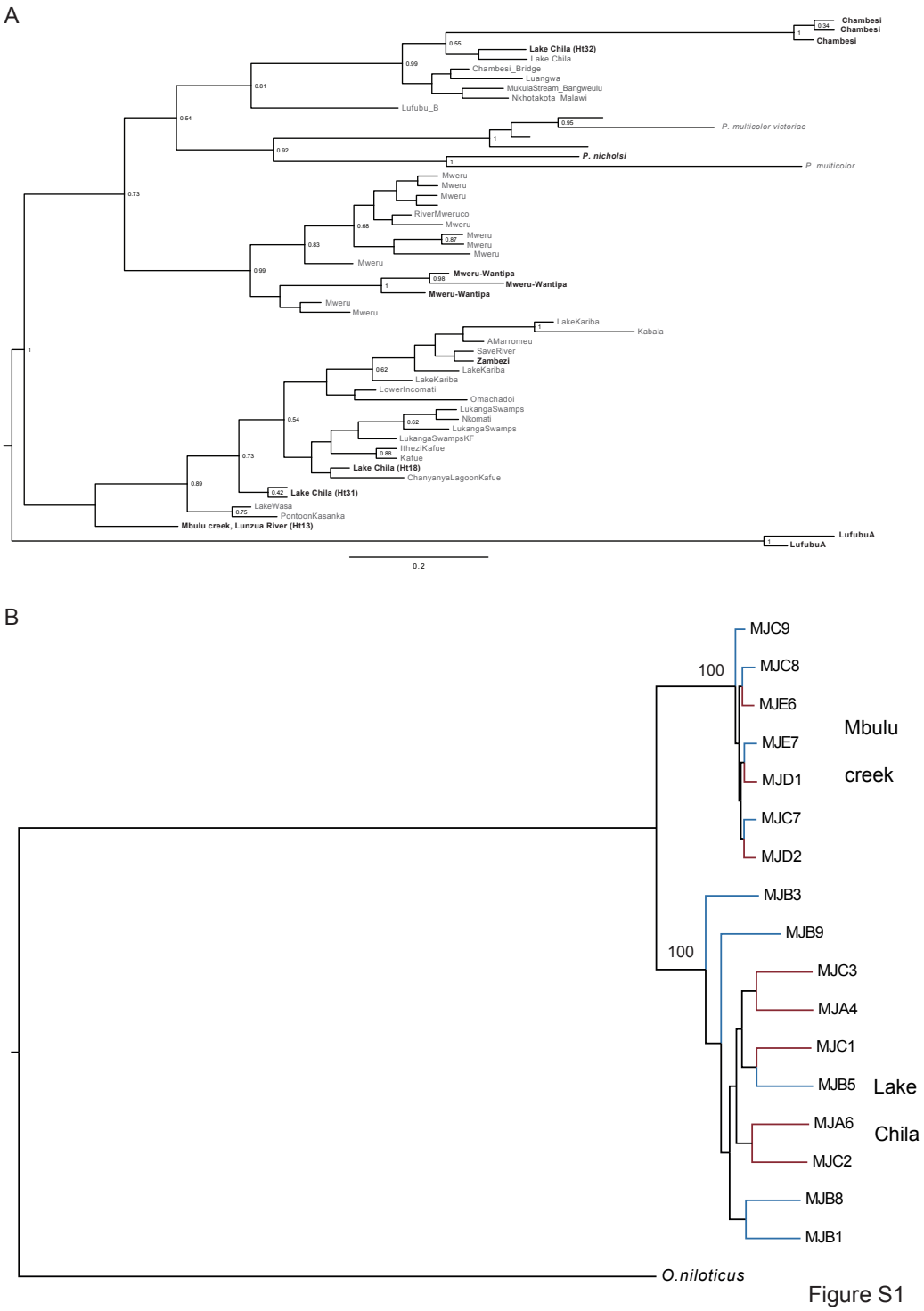

A
Chambesi
Chambesi
Chambesi
Lake Chila (Ht32)
Lake Chila
Chambesi_Bridge
Luangwa
MukulaStream_Bangweulu
Nkhotakota_Malawi
Lufubu_B
P. multicolor victoriae
P. nicholsi
P. multicolor
Mweru
Mweru
Mweru
RiverMweruco
Mweru
Mweru
Mweru
Mweru
Mweru
Mweru-Wantipa
Mweru-Wantipa
Mweru-Wantipa
Mweru
Mweru
LakeKariba
Kabala
AMarromeu
SaveRiver
Zambezi
LakeKariba
LakeKariba
LowerIncomati
Omachadoi
LukangaSwamps
Nkomati
LukangaSwamps
LukangaSwampsKF
ItheziKafue
Kafue
Lake Chila (Ht18)
ChanyanyaLagoonKafue
Lake Chila (Ht31)
LakeWasa
PontoonKasanka
Mbulu creek, Lunzua River (Ht13)
LufubuA
LufubuA
0.2
B
MJC9
MJC8
MJE6
MJE7
MJD1
MJC7
MJD2
MJB3
MJB9
MJC3
MJA4
MJC1
MJB5
MJA6
MJC2
MJB8
MJB1
O.niloticus
100
100
Mbulu creek
Lake Chila

Figure S1

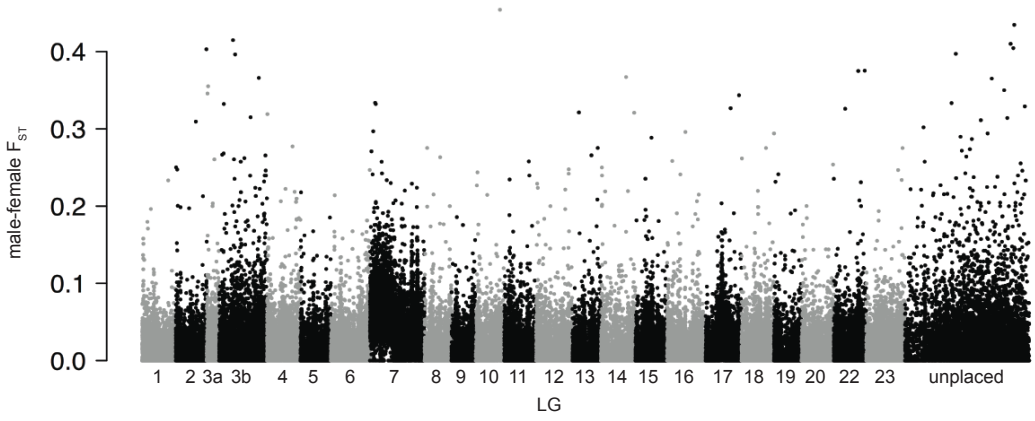

male-female F_{ST}
0.4
0.3
0.2
0.1
0.0
1
2
3a
3b
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
22
23
unplaced
LG

Figure S2

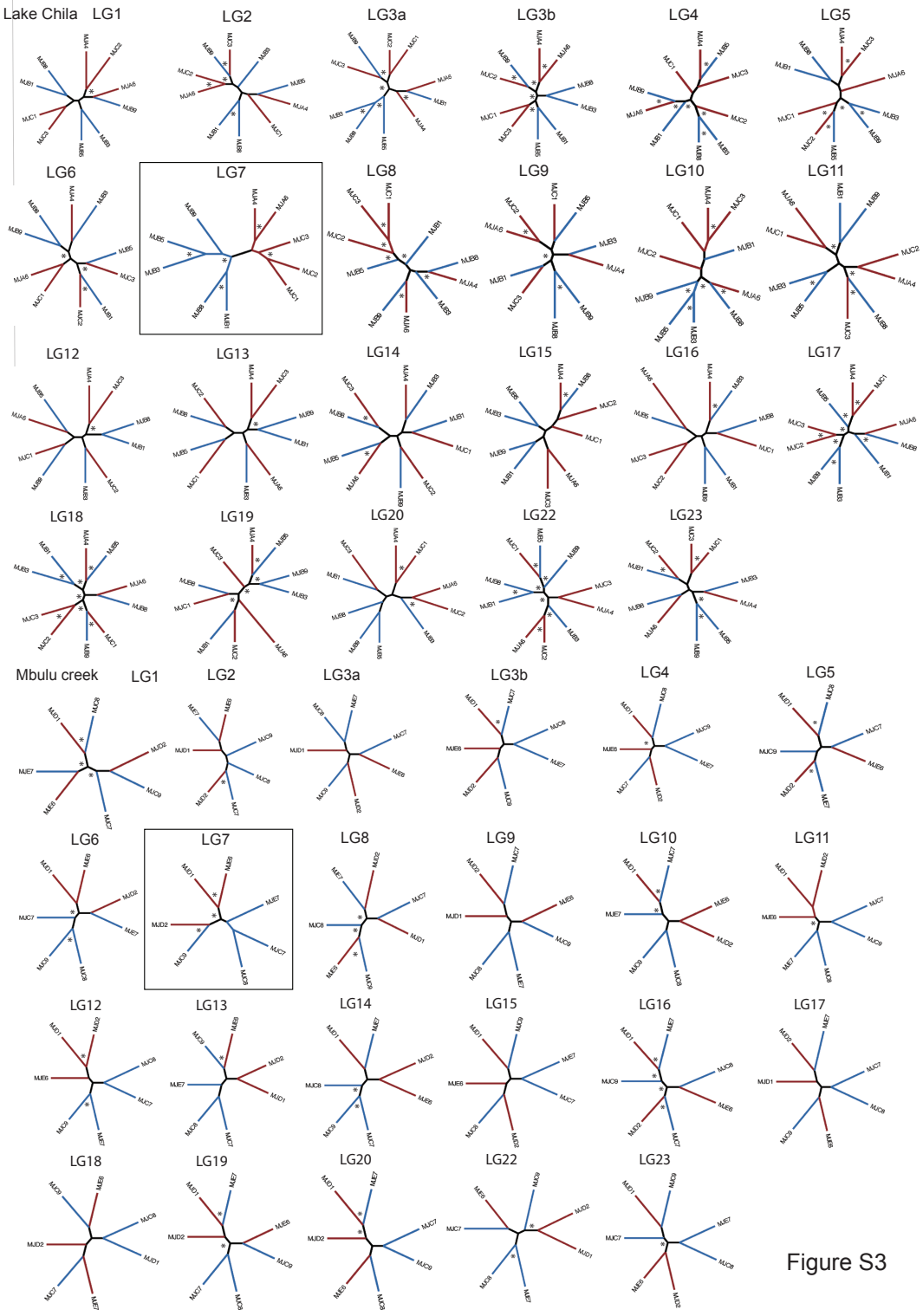

Lake Chila
LG1
LG2
LG3a
LG3b
LG4
LG5
LG6
LG7
LG8
LG9
LG10
LG11
LG12
LG13
LG14
LG15
LG16
LG17
LG18
LG19
LG20
LG22
LG23
Mbulu creek
LG1
LG2
LG3a
LG3b
LG4
LG5
LG6
LG7
LG8
LG9
LG10
LG11
LG12
LG13
LG14
LG15
LG16
LG17
LG18
LG19
LG20
LG22
LG23

Figure S3

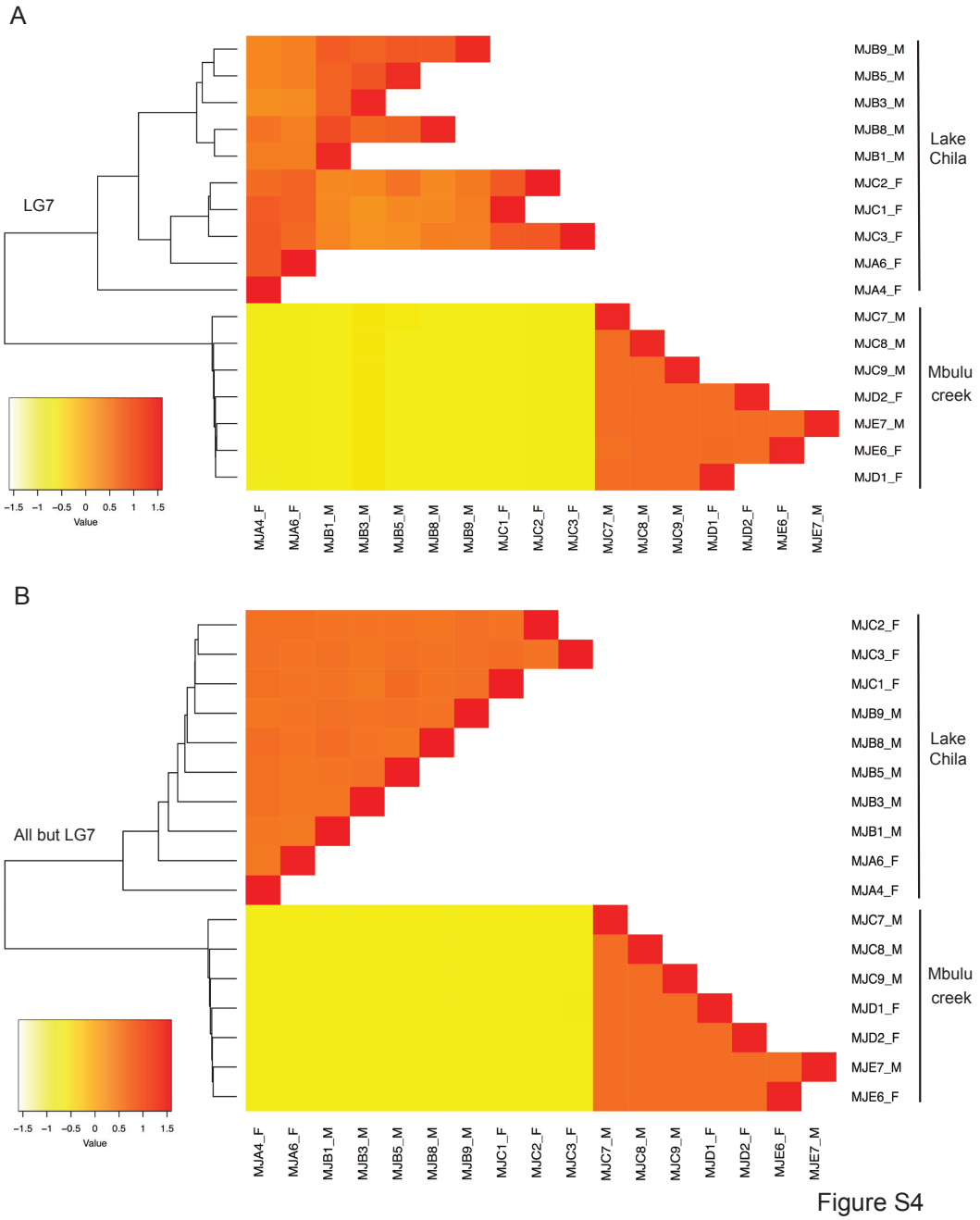

A
LG7
MJB9_M
MJB5_M
MJB3_M
MJB8_M
MJB1_M
MJC2_F
MJC1_F
MJC3_F
MJA6_F
MJA4_F
Lake Chila
MJC7_M
MJC8_M
MJC9_M
MJD2_F
MJE7_M
MJE6_F
MJD1_F
Mbulu creek
−1.5 −1 −0.5 0 0.5 1 1.5
Value
MJA4_F
MJA6_F
MJB1_M
MJB3_M
MJB5_M
MJB8_M
MJB9_M
MJC1_F
MJC2_F
MJC3_F
MJC7_M
MJC8_M
MJC9_M
MJD1_F
MJD2_F
MJE6_F
MJE7_M
B
All but LG7
MJC2_F
MJC3_F
MJC1_F
MJB9_M
MJB8_M
MJB5_M
MJB3_M
MJB1_M
MJA6_F
MJA4_F
Lake Chila
MJC7_M
MJC8_M
MJC9_M
MJD1_F
MJD2_F
MJE7_M
MJE6_F
Mbulu creek
−1.5 −1 −0.5 0 0.5 1 1.5
Value
MJA4_F
MJA6_F
MJB1_M
MJB3_M
MJB5_M
MJB8_M
MJB9_M
MJC1_F
MJC2_F
MJC3_F
MJC7_M
MJC8_M
MJC9_M
MJD1_F
MJD2_F
MJE6_F
MJE7_M

Figure S4

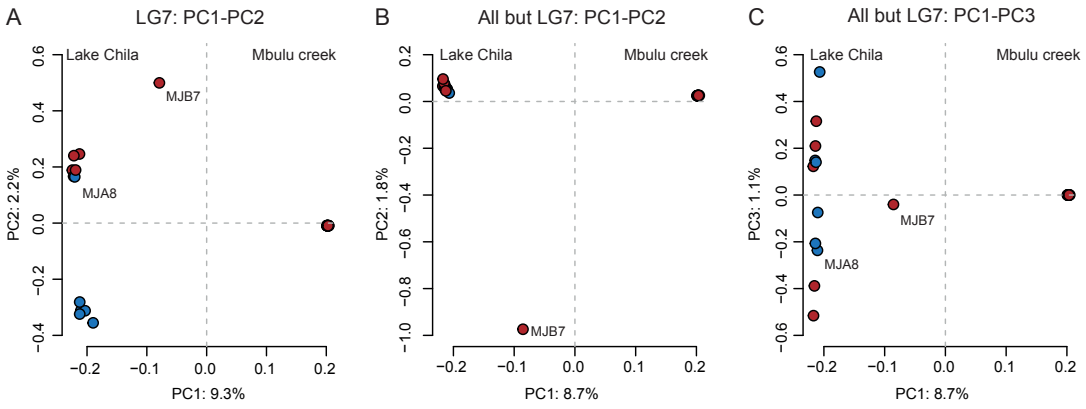

A
LG7: PC1-PC2
Lake Chila
Mbulu creek
MJB7
MJA8
PC2: 2.2%
PC1: 9.3%
B
All but LG7: PC1-PC2
Lake Chila
Mbulu creek
MJB7
PC2: 1.8%
PC1: 8.7%
C
All but LG7: PC1-PC3
Lake Chila
Mbulu creek
MJB7
MJA8
PC3: 1.1%
PC1: 8.7%

Figure S5

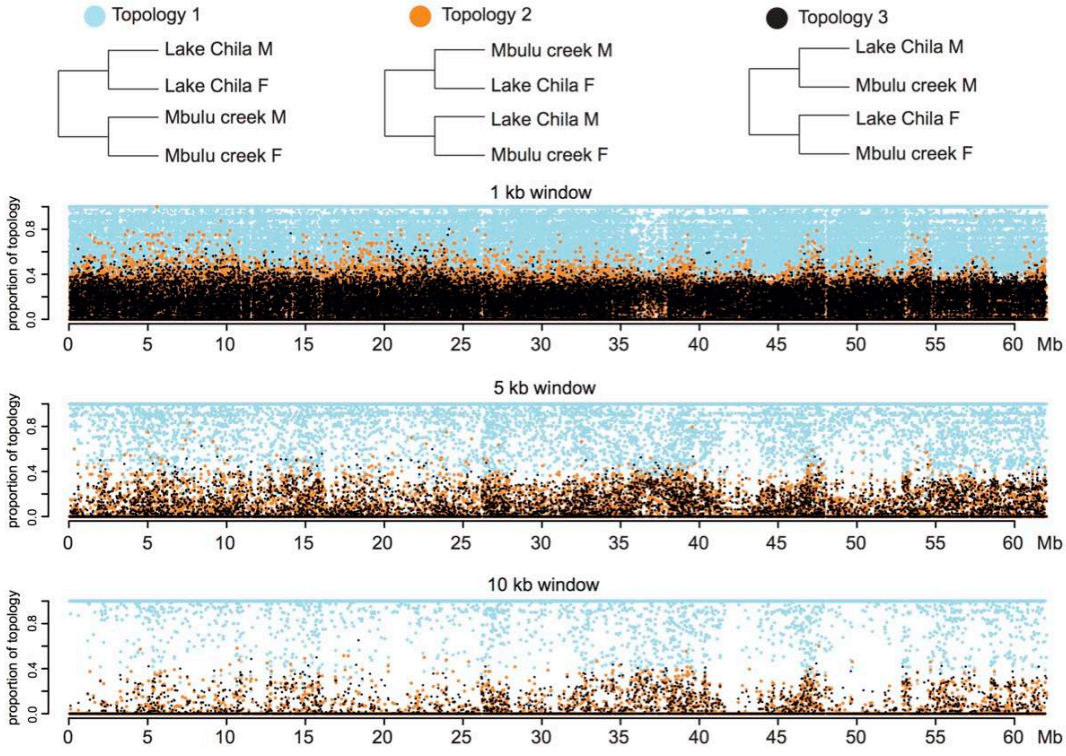

Topology 1
Lake Chila M
Lake Chila F
Mbulu creek M
Mbulu creek F
Topology 2
Mbulu creek M
Lake Chila F
Lake Chila M
Mbulu creek F
Topology 3
Lake Chila M
Mbulu creek M
Lake Chila F
Mbulu creek F
1 kb window
5 kb window
10 kb window
proportion of topology
0.0
0.4
0.8
0
5
10
15
20
25
30
35
40
45
50
55
60
Mb

Figure S6

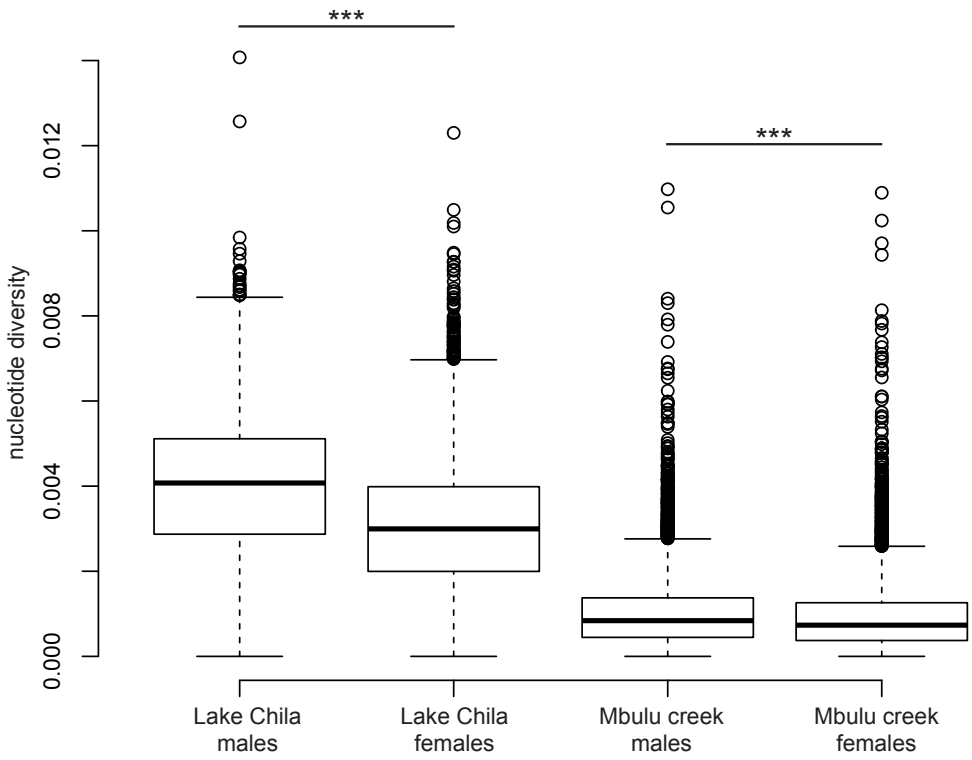

***
***
nucleotide diversity
0.000
0.004
0.008
0.012
Lake Chila males
Lake Chila females
Mbulu creek males
Mbulu creek females

Figure S7

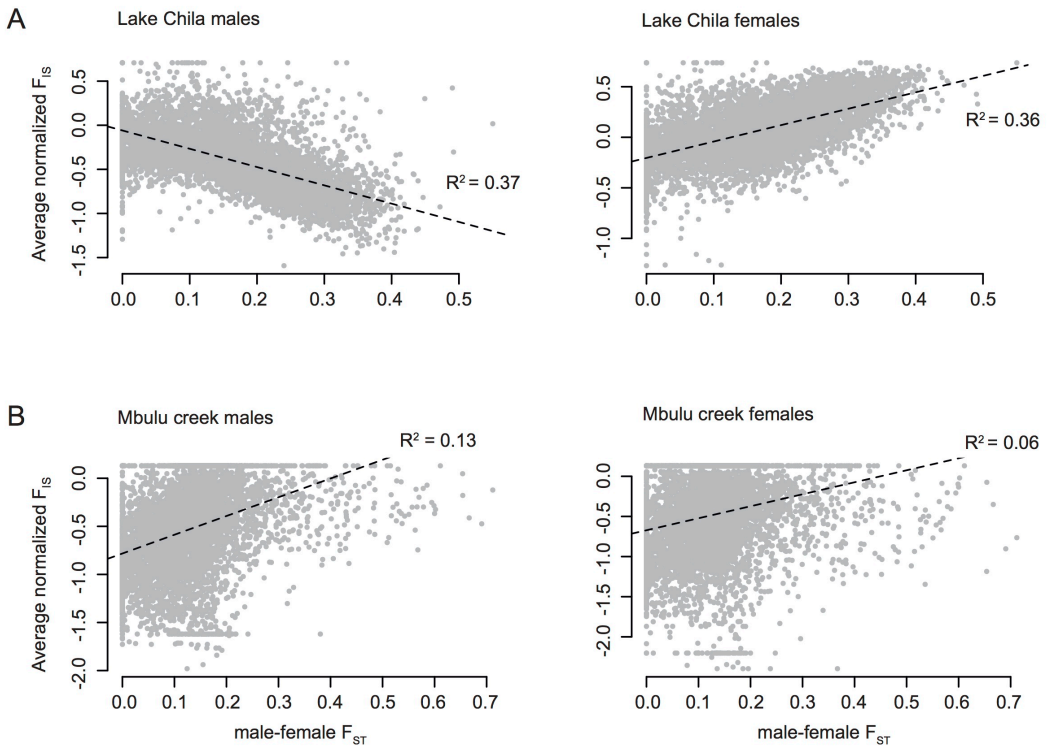

A
Lake Chila males
Average normalized F_{IS}
R^2 = 0.37
Lake Chila females
R^2 = 0.36
B
Mbulu creek males
R^2 = 0.13
Average normalized F_{IS}
male-female F_{ST}
Mbulu creek females
R^2 = 0.06
male-female F_{ST}

Figure S8

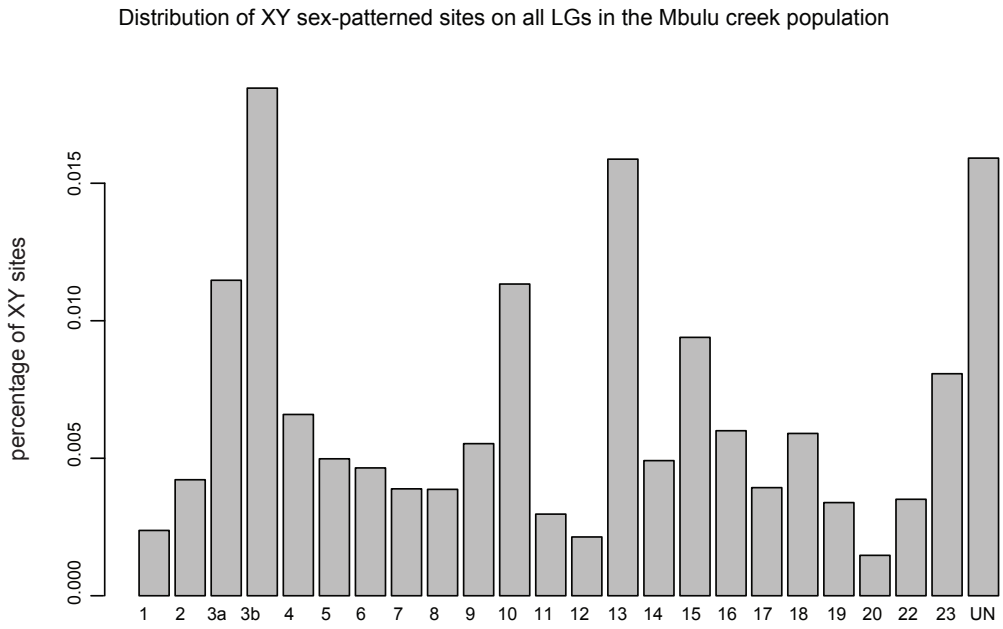

Distribution of XY sex-patterned sites on all LGs in the Mbulu creek population
percentage of XY sites
0.000
0.005
0.010
0.015
1
2
3a
3b
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
22
23
UN

Figure S9

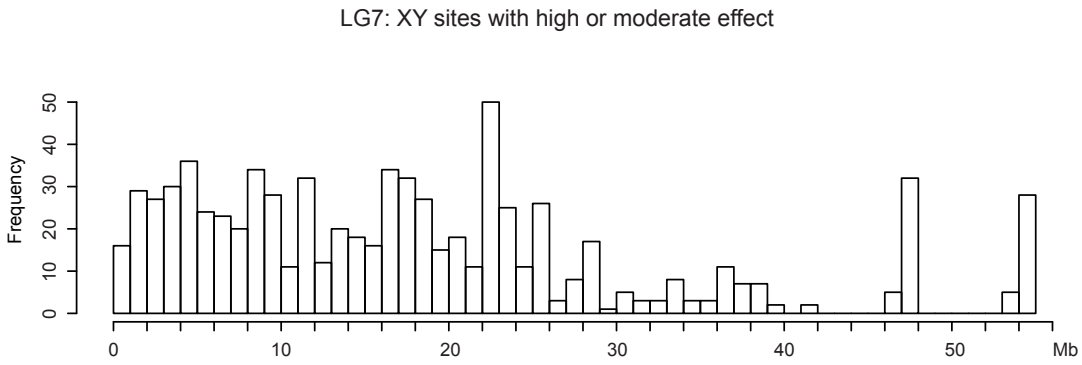

LG7: XY sites with high or moderate effect
Frequency
0
10
20
30
40
50
0
10
20
30
40
50
Mb

Figure S10

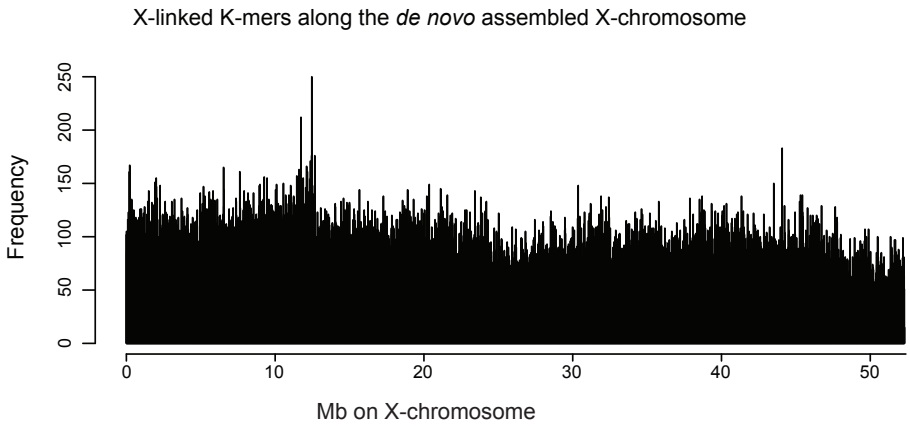

X-linked K-mers along the *de novo* assembled X-chromosome
Frequency
0
50
100
150
200
250
0
10
20
30
40
50
Mb on X-chromosome

Figure S11

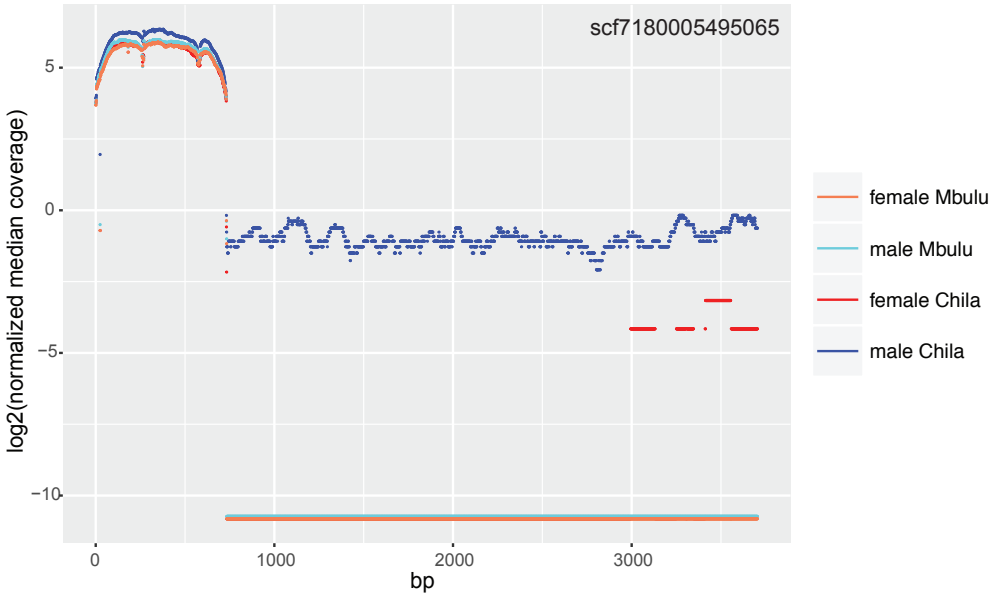

scf7180005495065
log2(normalized median coverage)
5
0
−5
−10
0
1000
2000
3000
bp
female Mbulu
male Mbulu
female Chila
male Chila

Figure S12

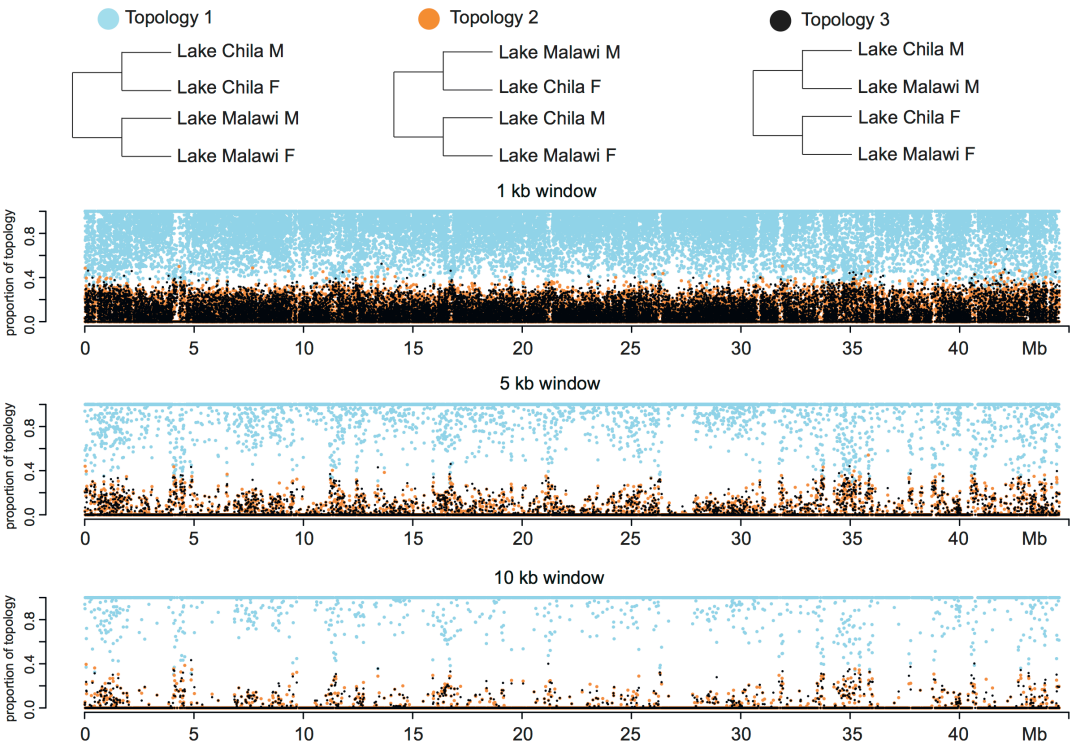

Topology 1
Lake Chila M
Lake Chila F
Lake Malawi M
Lake Malawi F
Topology 2
Lake Malawi M
Lake Chila F
Lake Chila M
Lake Malawi F
Topology 3
Lake Chila M
Lake Malawi M
Lake Chila F
Lake Malawi F
1 kb window
5 kb window
10 kb window
proportion of topology
0.0
0.4
0.8
0
5
10
15
20
25
30
35
40
Mb

Figure S13

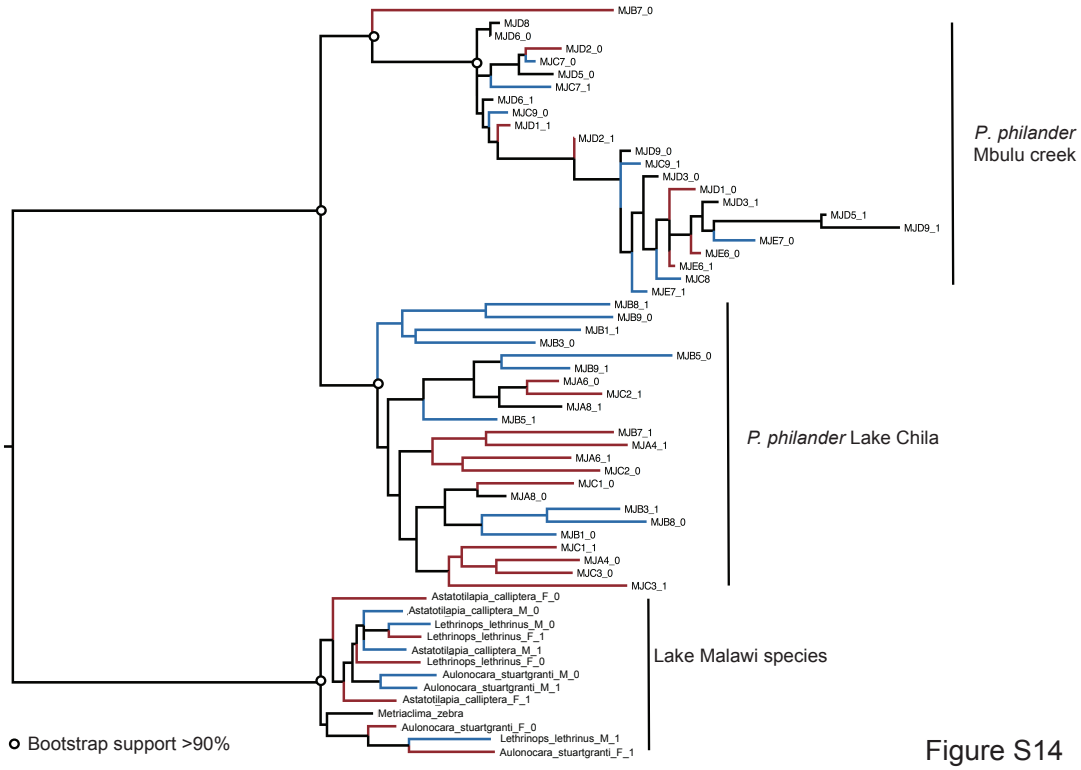

MJB7_0
MJD8
MJD6_0
MJD2_0
MJC7_0
MJD5_0
MJC7_1
MJD6_1
MJC9_0
MJD1_1
MJD2_1
MJD9_0
MJC9_1
MJD3_0
MJD1_0
MJD3_1
MJD5_1
MJD9_1
MJE7_0
MJE6_0
MJE6_1
MJC8
MJE7_1
P. philander
Mbulu creek
MJB8_1
MJB9_0
MJB1_1
MJB3_0
MJB5_0
MJB9_1
MJA6_0
MJC2_1
MJA8_1
MJB5_1
MJB7_1
MJA4_1
MJA6_1
MJC2_0
MJC1_0
MJA8_0
MJB3_1
MJB8_0
MJB1_0
MJC1_1
MJA4_0
MJC3_0
MJC3_1
P. philander Lake Chila
Astatotilapia_calliptera_F_0
Astatotilapia_calliptera_M_0
Lethrinops_lethrinus_M_0
Lethrinops_lethrinus_F_1
Astatotilapia_calliptera_M_1
Lethrinops_lethrinus_F_0
Aulonocara_stuartgranti_M_0
Aulonocara_stuartgranti_M_1
Astatotilapia_calliptera_F_1
Metriaclima_zebra
Aulonocara_stuartgranti_F_0
Lethrinops_lethrinus_M_1
Aulonocara_stuartgranti_F_1
Lake Malawi species
Bootstrap support >90%

Figure S14

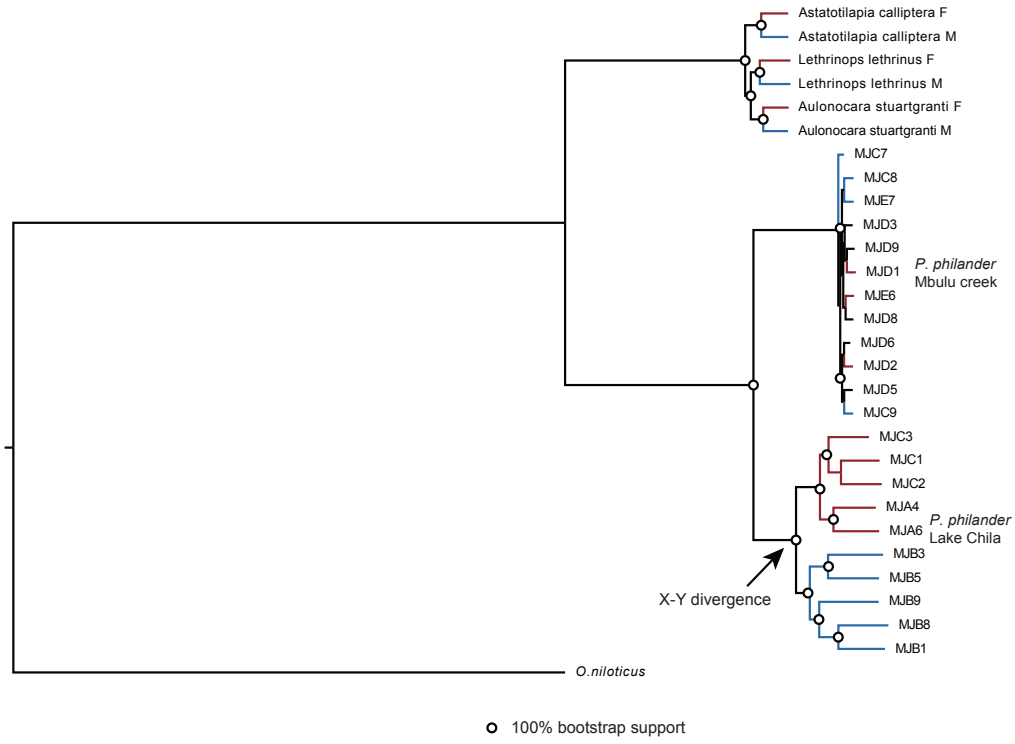
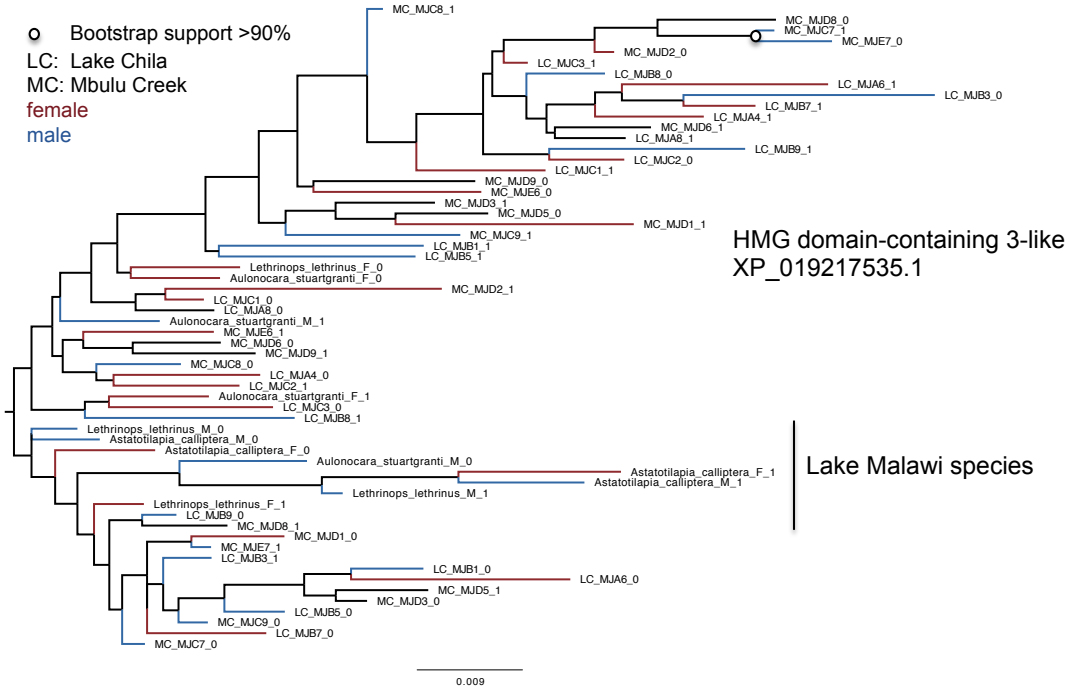

Astatotilapia calliptera F
Astatotilapia calliptera M
Lethrinops lethrinus F
Lethrinops lethrinus M
Aulonocara stuartgranti F
Aulonocara stuartgranti M
MJC7
MJC8
MJE7
MJD3
MJD9
MJD1
MJE6
MJD8
MJD6
MJD2
MJD5
MJC9
P. philander
Mbulu creek
MJC3
MJC1
MJC2
MJA4
MJA6
P. philander
Lake Chila
MJB3
MJB5
MJB9
MJB8
MJB1
X-Y divergence
O.niloticus
100% bootstrap support

Figure S15



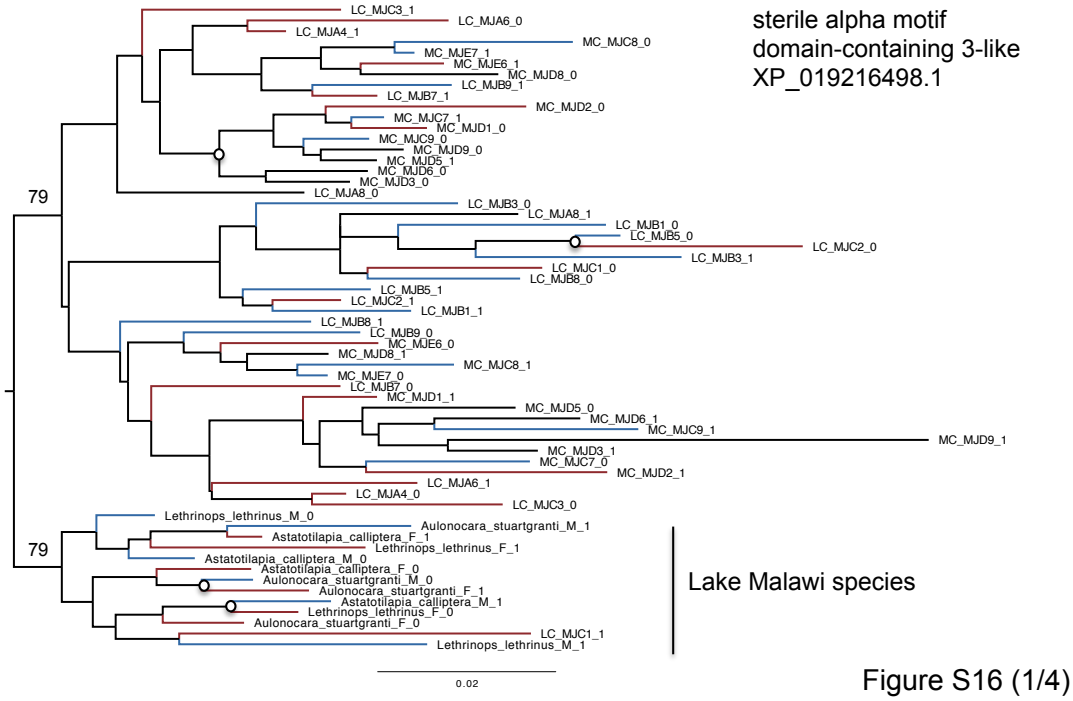
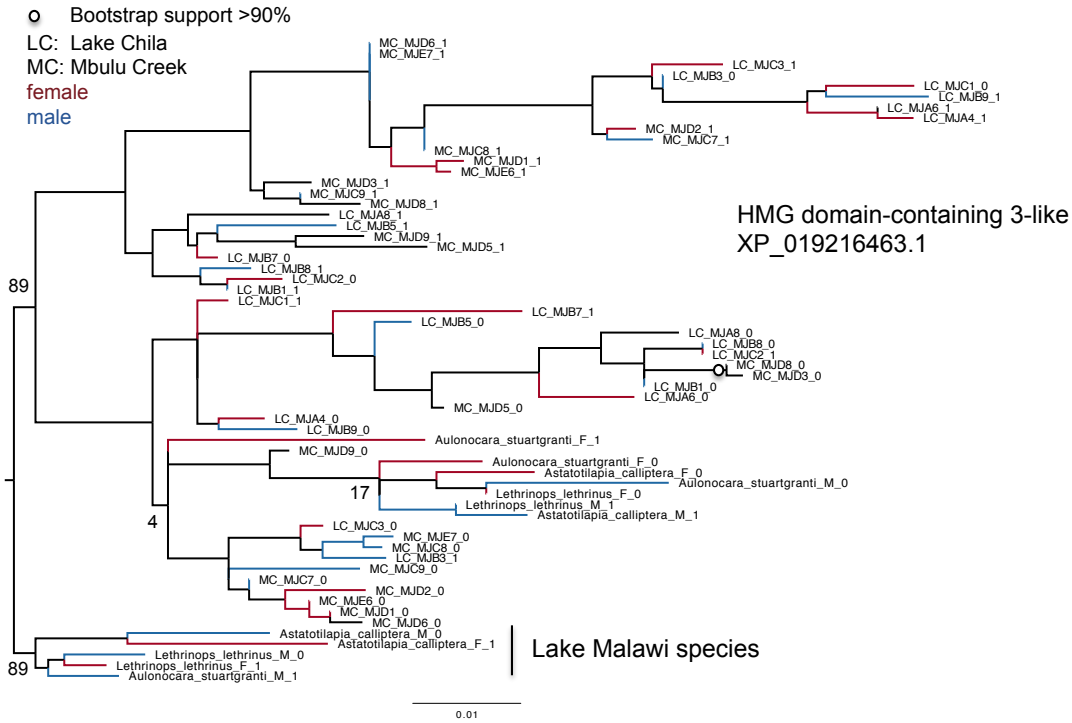


Figure S16 (1/4)



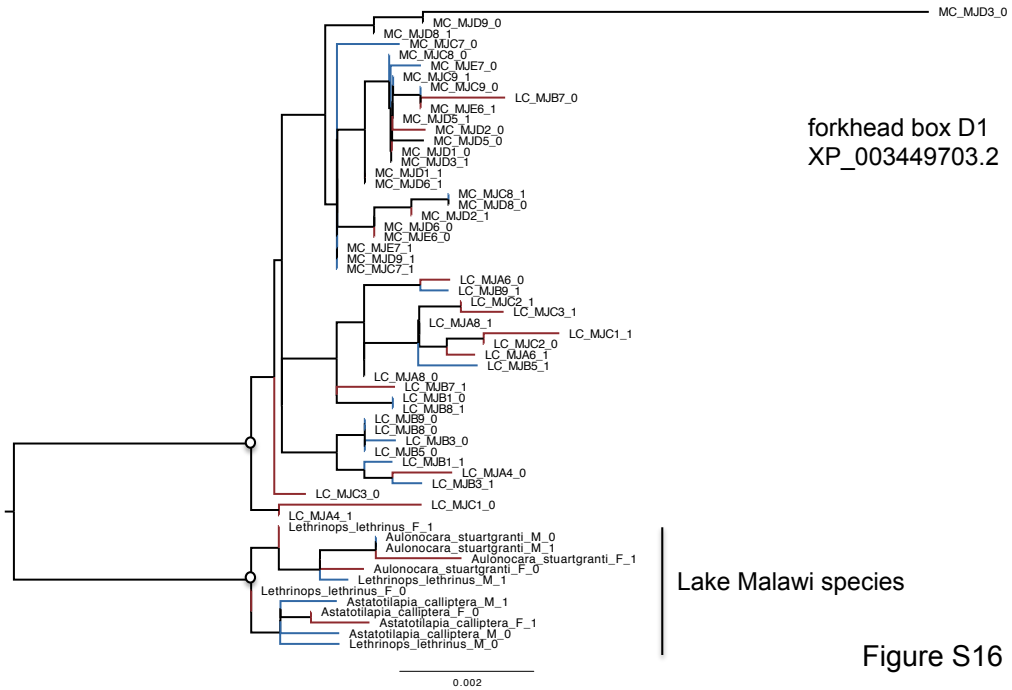
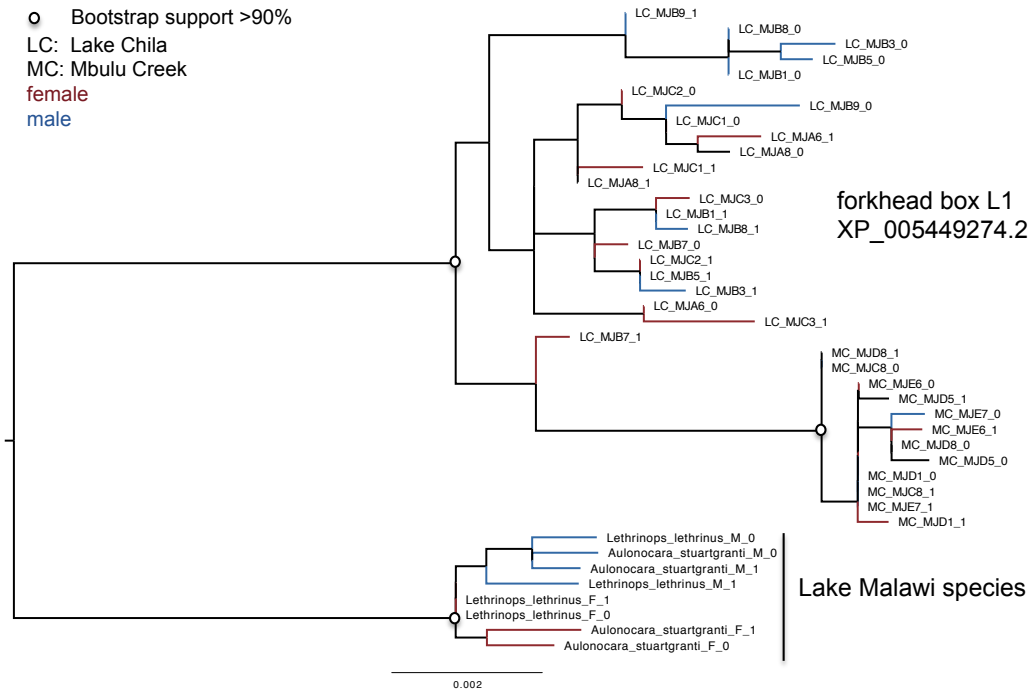


Figure S16 (2/4)



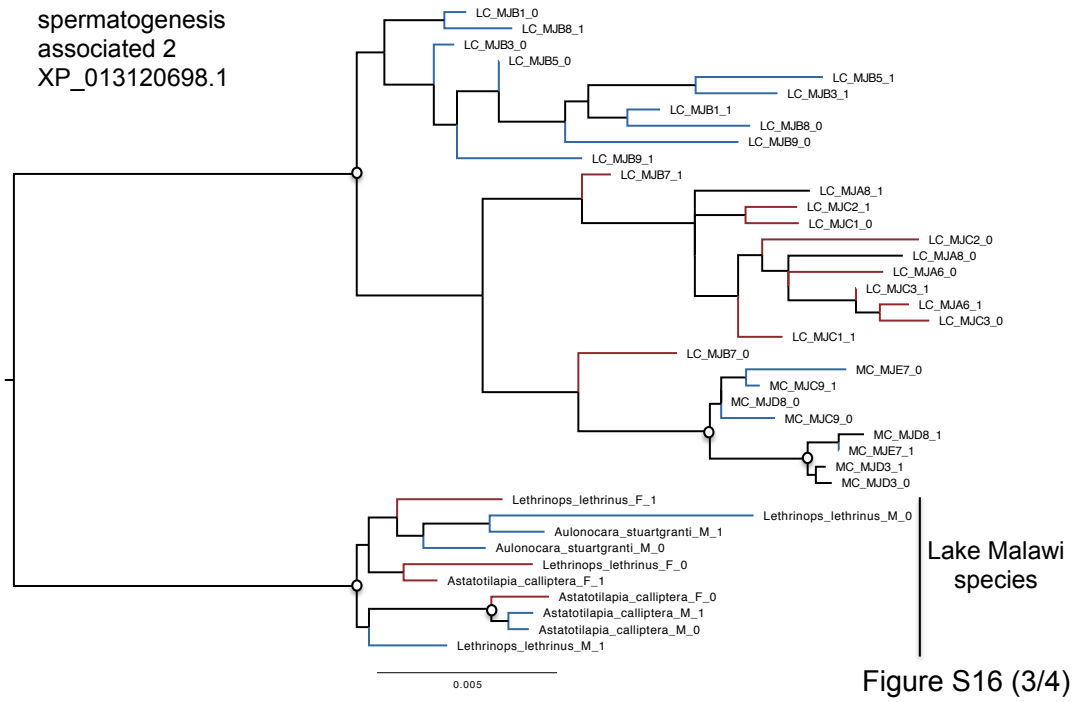
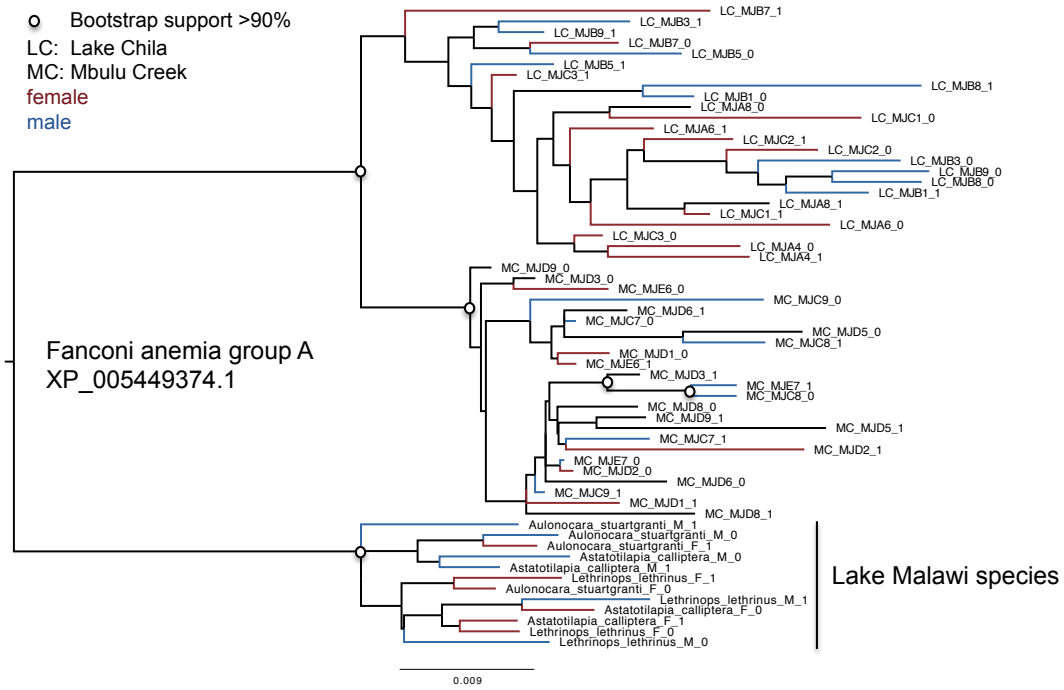


Figure S16 (3/4)



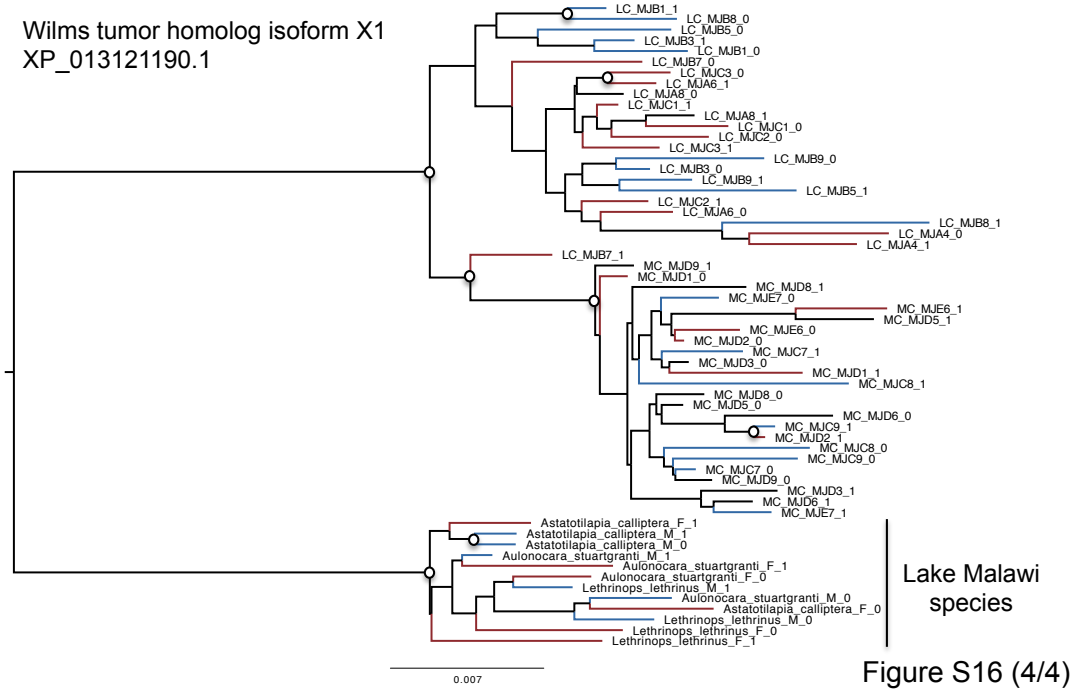
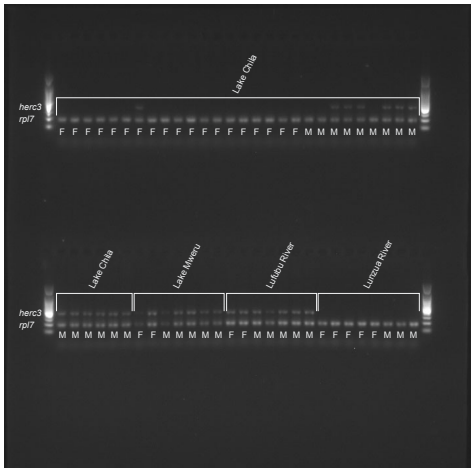
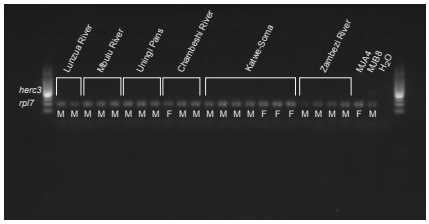


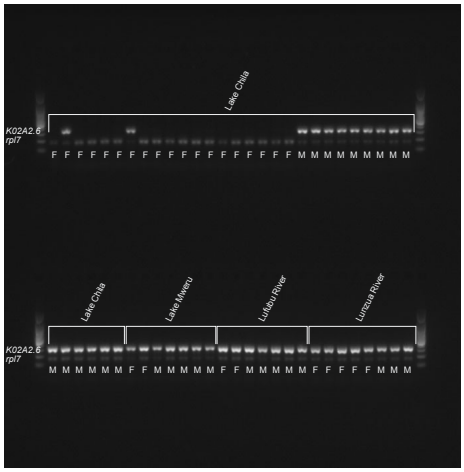
Figure S16 (4/4)

herc3





K02A2.6-like



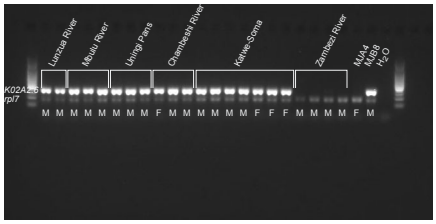


Figure S17

Chapter 6

Oil extraction imperils Africa's Great Lakes

R. Abila, P. Akoll, C. Albertson, D. Antunes, T. Banda, R. Bills, A. Bulirani, M.A. Chocha, A.S. Cohen, F. Cunha-Saraiva, S. Derycke, I. Donohue, M. Du, A.M. Dudu, B. Egger, K. Fritzsche, J.G. Frommen, H.F. Gante, M.J. Genner, A. Harer, H. Hata, K. Irvine, M.P. Isumbisho, L.J. de Bisthoven, A. Jungwirth, P. Kaleme, C. Katongo, L. Kever, S. Koblmuller, A. Konings, A. Lamboj, F. Lemmel-Schaedelin, G.M. Schiaffino, K. Martens, P.M. Mulungula, M.P. Masilya. A. Meyer, H.L. More, Z. Musilova, B.F. Muterezi, R. Muzumani, G. Ntakimazi, W. Okello, H. Phiri, L. Pialek, P.D. Plisnier, J.A.M. Raeymaekers, J. Rajkov, O. Rican, R. Roberts, W. Salzburger, I. Schoen, K.M. Sefc, P. Singh, P. Skelton, J. Snoeks, K. Schneider, C. Sturmbauer, H. Svardal, O. Svensson, J.T. Dowdall, G.F. Turner, A. Tyers, J.C. van Rijssel, M. Van Steenberge, M.P.M. Vanhove, E. Verheyen, A.A.T. Weber, O. Weyl, A. Ziegelbecker, H. Zimmermann

Science (2016)

I participated in the workshop on Oil extraction in Africa's Great Lakes during Cichlid Science 2015 Meeting in Graz that lead to the publication of the manuscript.

Children play on fishing boats at Lake Tanganyika, one of the East African Great Lakes threatened by oil exploitation.

PHOTO: BLICKWINKEL/ALAMY STOCK PHOTO

Oil extraction imperils Africa's Great Lakes

AS THE WORLD'S demands for hydrocarbons increase (*1*), remote areas previously made inaccessible by technological limitations are now being prospected for oil and gas deposits. Virtually unnoticed by the public, such activities are ongoing in the East African Great Lakes region, threatening these ecosystems famed for their hyper-diverse biota, including the unique adaptive radiations of cichlid fishes (*2*). Countries in the region see exploitation of hydrocarbon reserves as a vital economic opportunity. In the Lake Albert region of Uganda, for example, the government foresees a $3.6 billion oil profit per year starting in 2018—a sum almost as high as the country's current annual budget (*3*). However, oil extraction in the East African Great Lakes region poses grave risks to the environment and local communities.

The thousands of oil spills reported in Nigeria (*4*) demonstrate that the extraction and transport of oil are prone to accidents. This is especially bad news for the African Great Lakes because they are virtually closed ecosystems. For example, for Lake Tanganyika, which contains about one-fifth of the world's surface freshwater (*5*), the flushing time is ~7000 years (*5*). This time frame implies that the recovery from an oil spill could take millennia. To make matters worse, the lakes' location in a remote part of the world would impede a quick and effective reaction to an oil spill. Appropriate infrastructures are currently unavailable at the lakes, and bringing in heavy equipment at the time of a spill would be cumbersome, logistically impossible, or prohibitively expensive.

An oil spill would markedly affect the health, water supply, and food security of local communities (*6*). More than 10 million people depend on Lake Tanganyika alone for fisheries and water resources, and many more along the Congo River, into which the lake drains, are highly dependent on the lakes' ecosystem (*7*). In addition to the toll on humans, an oil spill in these lakes would be a global catastrophe for biodiversity. Combined, these lakes are home to thousands of species, almost all of them endemic (*2*). An accident might deal a final blow to these ecosystems, which have already been rendered fragile by anthropogenic stressors such as overfishing, deforestation, and global warming (*8*).

Finally, large parts of the East African region still lack political stability and security (*9*). In addition to the possibility of accidents, competition for hydrocarbon resources could lead to sabotage, as has unfortunately been frequently observed in the Niger delta (*4*).

We are concerned that the risks associated with the intended exploitation of fossil hydrocarbons in the East African Great Lakes region are seriously underestimated. We urge the countries involved in these undertakings to engage with the scientific and lake management communities to identify strong mitigation and control measures that could be put in place before hydrocarbon production begins. Local governments should foster alternative, sustainable plans to develop the region in accordance with the United Nations Sustainable Development Goals (*10*). To this end, the local population, regional stakeholders, governments, nongovernmental organizations, and scientists must cooperate to develop economically and ecologically viable strategies for the region, as is currently being attempted for the Virunga National Park in the DR Congo (*11*).

Erik Verheyen,* on behalf of the Cichlid Science 2015 Meeting and concerned scientists

*Corresponding author.
Email: erik.verheyen@naturalsciences.be

The full list of authors is available online.

REFERENCES

1. International Energy Agency, Oil Market Report (2016); www.iea.org/oilmarketreport/omrpublic/.
2. W. Salzburger, B. van Bocxlaer, A.S. Cohen, *Annu. Rev. Ecol. Evol. Syst.* **45**, 519 (2014).
3. M. L. Oketch, "Uganda to earn Shs9 trillion from oil annually—Tullow," *Daily Monitor* (2014); www.monitor.co.ug/News/National/Uganda-to-earn-Shs9-trillion-from-oil-annually---Tullow/688334-2451478-rag0a9/index.html.
4. Nigerian Oil Spill Monitor (https://oilspillmonitor.ng/).
5. R. H. Spigel, G. W. Coulter, 1996, in *The Limnology, Climatology and Paleoclimatology of the East African Lakes*, T. C. Johnson, E. O. Odada, Eds. (Gordon and Breach Science Publishers, Amsterdam, 1966), pp. 103–140.
6. NCF/WWF/IUCN, "Niger Delta Natural resources damage assessment and restoration project scoping report, May 2006" [Nigerian Conservation Foundation, UK World Wildlife Federation, and International Union for

Conservation of Nature, Commission on Environmental, Economic, and Social Policy from the Federal Ministry of Environment (Abuja), 2006].
7. H. Mölsä, J. E. Reynolds, E. J. Coenen, O. V. Lindqvist, *Hydrobiologia* **407**, 1 (1999).
8. A. S. Cohen *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **113**, 9563 (2016)
9. A. Adusei, *African Security Rev.* **24**, 332 (2015).
10. United Nations, "17 goals to transform our world" (2015); www.un.org/sustainabledevelopment/sustainable-development-goals/.
11. Global Witness, "Protecting Virunga National Park from oil companies" (2016); www.globalwitness.org/en/campaigns/oil-gas-and-mining/protecting-virunga-national-park-oil-companies/.

SUPPLEMENTARY MATERIALS

www.sciencemag.org/content/354/6312/561.1/suppl/DC1
Full author list

10.1126/science.aal1722

Time for responsible peatland agriculture

THE 15TH INTERNATIONAL Peat Congress, held in Asia for the first time, brought together industry, policy-makers, and academia to discuss responsible peatland management. In Southeast Asia, peatland management is largely driven by the palm oil industry. After the Congress, misleading reports were published by leading Asian newspapers. They claimed that oil palm plantations on peatland can be viewed as sustainable [e.g., (*1*)] and supported the continuation of business-as-usual peatland agriculture. This is contrary to the opinion of an overwhelming number of tropical peatland scientists (*2*) and the vast majority of science published in the past two decades.

Deep, carbon-rich peat deposits are maintained by a combination of steady organic matter inputs and high water tables, which inhibit microbial decomposition (*3*). Conversion of peat swamp forest (the natural vegetation of Southeast Asian peatlands) to agricultural land requires removing vegetation and lowering groundwater tables. The combination of slash and burn techniques and drainage used to prepare peat for agriculture promotes smoldering fires and rapid peat oxidation. Peat fires are globally significant for their greenhouse gas emissions and threats to human health and regional economies (*4*). Peat oxidation leads to high CO_2 emissions and land subsidence. As the land surface falls toward river and sea levels, it will be subject to periodic and eventually permanent flooding, limiting future agricultural production (*5*). Agricultural use of peatlands cannot, therefore, be considered sustainable from either environmental or socioeconomic perspectives.

Industry and academia are working together to develop peatland agricultural systems (*6*) that minimize negative environmental and commercial impacts. In the interim, steps should be taken to improve hydrological management of peatlands under agriculture and to implement landscape-scale management planning. Denial of known issues slows progress toward responsible solutions, which are urgently needed to prevent avoidable losses of Southeast Asia's peatlands, as well as global consequences.

Lahiru S. Wijedasa,[1,2,3]* Susan E. Page,[4] Christopher D. Evans,[5] Mitsuru Osaki[6]

[1]Department of Biological Sciences, National University of Singapore, 117543, Singapore. [2]ConservationLinks, 120433, Singapore. [3]Rimba, Malaysia, Selangor, MY 43650, Malaysia. [4]Centre for Landscape & Climate Research, Department of Geography, University of Leicester, Leicester, LE1 7RH, UK. [5]Centre for Ecology and Hydrology, Bangor LL57 2UW, UK. [6]Research Faculty of Agriculture, Hokkaido University, Hokkaido 060-0808, Japan.

*Corresponding author. Email: lahirux@gmail.com

REFERENCES

1. B. Nurbianto, "Malaysia challenges the world over palm oil on peatland," *The Jakarta Post* (2016); www.thejakartapost.com/news/2016/08/24/malaysia-challenges-the-world-over-palm-oil-on-peatland.html.
2. L. S. Wijedasa *et al.*, *Glob. Chang. Biol.* 10.1111/gcb.13516 (2016).
3. T. Hirano *et al.*, *Glob. Chang. Biol.* **18**, 3410 (2012).
4. R. A. Chisholm, L. S. Wijedasa, T. Swinfield, *Conserv. Biol.* **30**, 5 (2016).
5. E. Sumarga, L. Hein, A. Hooijer, R. Vernimmen, *Ecol. Soc.* **21**, 52 (2016).
6. International Peat Society, "Statement regarding the Jakarta Post article of 18th August" (2016); www.peatlands.org/news/ips-statement-congress-may-change-views-cultivation-peatland-jakarta-post.

10.1126/science.aal1794

Protecting China's soil by law

AFTER SEVERE PROBLEMS with air and water pollution, China is getting serious about its soil (*1*, *2*). On 31 May, China's State Council released an action plan for soil pollution prevention and remediation, aiming to make 90% of polluted, arable land safe for human use by 2020 and increase this to 95% by 2030 (*3*). This ambitious action plan calls for the support of strong environmental laws to monitor, prevent, and remediate serious levels of soil contamination. However, national legislation protecting soil quality has lagged behind that of air and water for more than a decade, while Chinese lawmakers debate the focus and purpose of a soil protection law.

On 3 September, China released a draft of its first environmental tax law, which designates four taxable types of pollution: airborne and water pollutants, solid waste, and noise (*4*). Soil is conspicuously absent. A soil protection law could close the current environmental legislation system's loopholes, make China's new environmental tax system more comprehensive, and protect China's soil.

To make a pragmatic soil protection law, the central government must clearly identify local government's liability and responsibility for soil pollution, as ambiguous responsibility has been one of the major problems in soil management in the past. The law must stipulate the division of duties between government agencies, establish a surveying and monitoring system, and introduce funding mechanisms. Remediation of contaminated soil is extremely costly, and China needs to create a national soil fund by allocating an adequate percentage of its land revenues and environmental tax revenues. Some members of the soil pollution plan panel initially suggested that 10% of land revenues be designated to the fund (*5*). Because China lacks comprehensive risk assessment systems for contaminated land management (*6*), the law should stipulate risk management and control approaches for contaminated sites.

China's soil pollution has become a critical issue that affects public health and creates social unrest and instability (*1*, *7*, *8*). China should not repeat its past mistakes of focusing on economic growth at the expense of the environment.

Jinnan Wang,[1] Qing Hu,[2] Xiahui Wang,[1] Xiaoliang Li,[1,3] Xiao Jin Yang[3]*

[1]Chinese Academy of Environmental Planning, Beijing, 100012, China. [2]South University of Science and Technology of China, Shenzhen, Guangdong, 518055, China. [3]Beijing University of Chemical Technology, Beijing, 100029, China.

*Corresponding author. Email: yangxj@mail.buct.edu.cn

REFERENCES

1. C. Larson, *Science* **343**, 1415 (2014).
2. Ministries of Environmental Protection and Land and Resources of the People's Republic of China, National Soil Pollution Report 2013 (2014); www.mlr.gov.cn/xwdt/jrxw/201404/t20140417_1312998.htm [in Chinese].
3. Central Government of the People's Republic of China, "The State Council issues an action plan for soil pollution prevention and control" (2016); www.gov.cn/zhengce/content/2016-05/31/content_5078377.htm [in Chinese].
4. The National Congress of the People's Republic of China, Environmental Protection Tax Law (Draft). (2016); www.npc.gov.cn/npc/flcazqyj/2016-09/02/content_1996531.htm [in Chinese].
5. E. Wang, "A suggestion of allocating 10% of land revenues for soil pollution remediation fund," *21st Century Business Herald* (2016); http://epaper.21jingji.com/html/2016-03/08/content_33885.htm [in Chinese].
6. F. Coulon *et al.*, *Environ. Int.* **91**, 196 (2016).
7. G. He, "Special report: The victims of China's soil pollution crisis," *China Dialogue* (2014); www.chinadialogue.net/article/show/single/en/7073-Special-report-The-victims-of-China-s-soil-pollution-crisis.
8. Xinhua News, "No sure cure for China's soil pollution," *China Daily* (2014); www.chinadaily.com.cn/china/2014-04/29/content_17474094.htm.

10.1126/science.aal1847

Science
AAAS

Oil extraction imperils Africa's Great Lakes

Erik Verheyen (November 3, 2016)

Science **354** (6312), 561-562. [doi: 10.1126/science.aal1722]

Editor's Summary

Science (print ISSN 0036-8075; online ISSN 1095-9203) is published weekly, except the last week in December, by the American Association for the Advancement of Science, 1200 New York Avenue NW, Washington, DC 20005.

www.sciencemag.org/content/354/6312/561/suppl/DC1

Supplementary Materials for

Oil extraction imperils Africa's Great Lakes

Erik Verheyen,* on behalf of the Cichlid Science 2015 Meeting and concerned scientists

*Corresponding author.
Email: erik.verheyen@naturalsciences.be

Published 4 November 2016, *Science* **354**, 561 (2016)
DOI: 10.1126/science.aal1722

This PDF file includes:

Full author list

Supplementary Materials

Members of the 'Cichlid Science 2015 Meeting and concerned scientists' listed alphabetically:

R. Abila, Maasai Mara University, Narok, Kenya
P. Akoll, College of Natural Sciences, Makerere University, Kampala, Uganda
C. Albertson, Department of Biology, University of Massachusetts, Amherst MA, USA
D. Antunes, Institute of Ecology and Evolution, University of Bern, Hinterkappelen, Switzerland
T. Banda, Ministry of Fisheries and Livestock, Chilanga, Zambia
R. Bills, South African Institute for Aquatic biodiversity, Grahamstown, South Africa
A. Bulirani, Department of Fisheries, Ministry of Energy, Mines and Natural Resources, Lilongwe, Malawi
A. Chocha Manda, Université de Lubumbashi & Université de Kalemie, D.R.Congo
A.S. Cohen, Department of Geosciences, University of Arizona, Tucson, AZ , USA
F. Cunha-Saraiva, Konrad Lorenz Institute of Ethology, Vienna, Austria
S. Derycke, Royal Belgian Institute of Natural Sciences, Brussels, Belgium, Ghent University, Ghent, Belgium
I. Donohue, Trinity College Dublin, Ireland
M. Du, Gurdon Institute, University of Cambridge, Cambridge, United Kingdom
A.M. Dudu, Université de Kisangani & Centre de Surveillance de la Biodiversité, Kisangani, D.R.Congo
B. Egger, Zoologisches Institut, Universität Basel, Basel, Switzerland
K. Fritzsche, Institute of Zoology, University of Graz, Graz, Austria
J.G. Frommen, Institute of Ecology and Evolution, University of Bern, Hinterkappelen, Switzerland
H.F. Gante, Zoologisches Institut, Universität Basel, Basel, Switzerland
M.J. Genner, University of Bristol, Bristol, United Kingdom
A. Härer, Department of Biology, University of Konstanz, Germany
H. Hata, Graduate School of Science and Engineering, Ehime University, Ehime, Japan
K. Irvine, UNESCO- Institute for Water Education (IHE), Delft, The Netherlands
P. Isumbisho Mwapu, Centre de Recherches Universitaires du Kivu & ISP-Bukavu, Bukavu, D.R.Congo
L. Janssens de Bisthoven, Royal Belgian Institute of Natural Sciences, Brussels, Belgium
A. Jungwirth, Hughes Hall & Department of Zoology, University of Cambridge, United Kingdom
P. Kaleme, Centre de Recherche en Sciences Naturelles, Lwiro, D.R.Congo
C. Katongo, University of Zambia, Lusaka, Zambia
L. Kéver, University of Liège, Liège, Belgium
S. Koblmüller, Institute of Zoology, University of Graz, Graz, Austria
A. Konings, Cichlid Press, El Paso, TX, USA
A. Lamboj, Department for Integrative Zoology, University of Vienna, Vienna, Austria
F. Lemmel-Schaedelin, Konrad-Lorenz-Institut für Vergleichende Verhaltensforschung , Veterinärmedizinische Universität Wien, Vienna, Austria
G. Machado Schiaffino, Department of Biology, University of Konstanz, Konstanz, Germany
K. Martens, Royal Belgian Institute of Natural Sciences, Brussels, Belgium
P. Masilya Mulungula, Centre de Recherche en Hydrobiologie, Uvira & ISP-Bukavu, D.R.Congo
A. Meyer, Department of Biology, University of Konstanz, Konstanz, Germany
H.L. More, University of Graz, Graz, Austria, Simon Fraser University, Burnaby, Canada
Z. Musilova, Department of Zoology, Faculty of Science, Charles University in Prague, Prague, Czech Republic

F. Muterezi Bukinga, Centre de Recherche en Hydrobiologie, Uvira, D.R.Congo
R. Muzumani, Centre de Recherche en Hydrobiologie, Uvira, D.R.Congo
G. Ntakimazi, Université du Burundi, Bujumbura, Burundi
W. Okello, The National Fisheries Resources Research Institute, Jinja, Uganda
H. Phiri, Ministry of Fisheries and Livestock, Chilanga, Zambia
L. Pialek, Department of Zoology, University of South Bohemia, České Budějovice, Czech Republic
P. D. Plisnier, Royal Museum for Central Africa, Tervuren, Belgium
J.A.M. Raeymaekers, Centre for Biodiversity Dynamics, Norwegian University of Science and Technology, Trondheim, Norway
J. Rajkov, Zoologisches Institut, Universität Basel, Basel, Switzerland
O. Rican, Faculty of Science, University of South Bohemia, Czech Republic
R. Roberts, North Carolina State University, Raleigh, NC, USA
W. Salzburger, Zoologisches Institut, Universität Basel, Basel, Switzerland
I. Schoen, Royal Belgian Institute of Natural Sciences, Brussels, Belgium
K.M. Sefc, Institute of Zoology, University of Graz, Graz, Austria
P. Singh, Institute of Zoology, University of Graz, Graz, Austria
P. Skelton, South African Institute for Aquatic biodiversity, Grahamstown, South Africa
J. Snoeks, Royal Museum for Central Africa, Tervuren; KU Leuven, Belgium
K. Schneider, Institute of Biodiversity Animal Health and Comparative Medicine, Glasgow, United Kingdom
C. Sturmbauer, Institute of Zoology, University of Graz, Graz, Austria
H. Svardal, Wellcome Trust Sanger Institute, Hinxton, Cambridge, CB10 1SA, United Kingdom
O. Svensson, Department of Biological and Environmental Sciences, Gothenburg, Sweden
J. Torres Dowdall, Department of Biology, University of Konstanz, Konstanz, Germany
G.F. Turner, Bangor University, School of Biological Sciences, Bangor, Gwynedd, Wales, United Kingdom
A. Tyers, Bangor University, School of Biological Sciences, Bangor, Gwynedd, Wales, United Kingdom
J.C. van Rijssel, EAWAG Center of Ecology, Evolution and Biogeochemistry, Kastanienbaum, Switzerland
M. Van Steenberge, Royal Museum for Central Africa, Tervuren, Belgium
M.P.M. Vanhove, Royal Belgian Institute of Natural Sciences, Brussels, Belgium; KU Leuven, Leuven, Belgium; Masaryk University, Brno, Czech Republic; Hasselt University, Diepenbeek, Belgium.
*E. Verheyen, Royal Belgian Institute of Natural Sciences, Brussels; University of Antwerp, Belgium**
A-T. Weber, Zoologisches Institut, Universität Basel, Basel, Switzerland
O. Weyl, South African Institute for Aquatic biodiversity, Grahamstown, South Africa
A. Ziegelbecker, Institute of Zoology, University of Graz, Graz, Austria
H. Zimmermann, Institute of Zoology, University of Graz, Graz, Austria

* Corresponding author: Email: erik.verheyen@naturalsciences.be

Discussion

The work presented in my doctoral thesis focuses on the role of ecology and phenotypic plasticity in lake-stream divergence in an East African haplochromine cichlid fish, thus contributing to the elucidation of mechanisms responsible for this replicated divergence. The only previous study that investigated reproductive isolation in lake-stream cichlid pairs found no evidence of isolation between the closely related pairs of *A. burtoni* (Theis et al. 2014). In my thesis, I largely expanded this work by investigating several components of reproductive isolation in population pairs ranging from intermediate to high genetic divergence. As expected in cases of ecological speciation (Nosil et al. 2009), genetically and ecologically divergent lake-stream pairs exhibited greater levels of reproductive isolation than genetically divergent but ecologically similar lake-lake pairs. Thus, my findings provide support for the important role of ecology in speciation in haplochromine cichlids.

In **Chapter 1** of my thesis I investigated reproductive isolation between a lake-stream pair with intermediate genetic divergence, with particular emphasis on the role of local adaptation and phenotypic plasticity. Using both wild-caught individuals and common garden-raised F1 individuals in field transplant experiments, with survival and growth rate as fitness measures, I provide the first demonstration of adaptive divergence between lake and stream cichlids at the level of whole-organism performance. I found that the wild lake population performed better than the wild river population in the lake environment, thereby confirming the presence of local adaptation and reproductive isolation. However, there was no difference in performance between different types of F1 crosses raised in common garden conditions, demonstrating an important role of phenotypic plasticity in the adaptation to different habitats. These experiments provide rare empirical data on fitness estimates in a cichlid species in the wild. My findings highlight the value of formally comparing the fitness of wild-caught and common-garden-reared individuals and emphasize the necessity of considering adaptive phenotypic plasticity in the study of adaptive divergence. Future work should establish whether the same holds true for other haplochromine species that show lake-stream divergence.

In **Chapter 2,** I investigated reproductive isolation between populations with higher genetic divergence, using more inclusive fitness estimates, such as reproductive success and offspring survival. The highly divergent and likely ancestral river population showed a low degree of plasticity in comparison to the river population used in the experiments described in Chapter 1. In Chapter 2, I presented evidence for strong prezygotic (selection against immigrants) and postzygotic barriers (selection against hybrids) in cichlids. My findings highlight the value of assessing multiple reproductive barriers over an inclusive portion of lifespan in natural contexts, as well as the importance of postzygotic barriers in addition to prezygotic barriers even during the early stages of

speciation. Together, the results of the first two chapters provide some of the first demonstrations of extrinsic reproductive isolation in cichlids.

The mechanisms of mate choice that would explain why river individuals had lower mating success remain to be investigated. Future laboratory experiments could investigate the importance of mate preference and related physiological mechanisms, using experiments similar to those described in Chapter 4 between distantly related lake *A. burtoni* populations from the north and from the south of Lake Tanganyika.

In **Chapter 3**, I described parasite infections in *A. burtoni* lake-stream pairs for the first time. I found that lake fish suffered from higher parasite load and lake specific parasite taxa, followed by an elevated innate immune response in nature. Furthermore, when river fish were exposed to lake parasites in the common garden setup, their immune response was elevated compared to the wild fish. Since these divergent infections are consistent between different years, present even in the pairs with the lowest divergence, and correlate with the degree of genetic divergence, I proposed that divergent selection related to parasite pressure could be a mechanism that is acting already at the early stages of speciation. Furthermore, I suggest that this kind of differential parasite pressure between ecotypes may initiate species divergence. Parasite infection in a non-native environment has been associated with reduced growth in sticklebacks (MacColl and Chapman 2010; Kaufmann et al. 2017) and is a likely explanation for the reduced growth of the wild river population in the transplant experiment described in Chapter 1.

By studying RNA-expression patterns in the common garden experiment, I revealed one mechanism for the plasticity of the river population raised in common garden, as observed in Chapter 1. Gill expression pattern of river individuals showed a shift in the common garden in comparison to the wild, clustering together with the lake population, thus providing an explanation for their equally high performance in the lake. A follow-up study on RNA expression will focus on identifying differentially expressed genes between different habitats, which might be directly related to habitat adaptation. Furthermore, we will compare differentially expressed genes with those identified in genome scans described in Chapter 4, as well as with those identified in similar studies in stickleback fish lake-stream system, to establish the generality of the patterns.

In Chapter 3, I focused exclusively on readily visible macroparasites. However, immune system response could plausibly also depend on readily overlooked symbionts, including, but not limited to, gut microbiota (Bolnick et al. 2014). The newly discovered importance and ubiquity of host-associated microorganisms (for review see McFall-Ngai et al., 2013) is stimulating research on the role of the microbiome in animal speciation (Sharon et al. 2010; Brucker and Bordenstein 2012, 2013; Shropshire and Bordenstein 2016). Although the microbiome influences numerous aspects of organismal fitness, its role in animal evolution and the origin of new species is largely unknown and requires careful consideration on a case-by-case basis (Chandler and Turelli 2014; Leftwich et al. 2018).

The potential role of gut microbiota in fish diversification has been investigated in South American crater lake cichlid species-pairs (Franchini et al. 2014) and in low- and high-predation guppy ecotypes (Sullam et al. 2015). Fish are known to experience complete turnover of their microbiomes during transitions between different aquatic environments (Schmidt et al. 2015; Lokesh and Kiron 2016) that are expected to cause drastic challenges to their immune system. A recent study on cichlid gut microbiota suggests a fundamental role of gut bacteria in cichlid niche expansion and adaptation (Baldo et al. 2017). Thus, it will be important to investigate cichlid gut and mouth microbiota communities in different habitats and their influence on lake-stream divergence. The question of whether gut and mouth habitat-specific microbiota could also play a role in lake-stream divergence is part of an ongoing study, in which we use 16S rRNA amplicon

sequencing to characterize gut and mouth bacteria of lake and stream ecotypes collected from two different systems and the common garden setup described in Chapter 3.

In **Chapter 4** nine cichlid population pairs diverging along the same ecological gradient were examined to infer the dynamics of genome differentiation during adaptive divergence, patterns of natural selection, and the link between genome divergence and reproductive isolation. Genomic differentiation has been found to be sudden, in contrast with a gradual morphological differentiation. The divergent lake-stream *A. burtoni* pair investigated in Chapter 2 and a sympatric lake-stream population pair of *C. horei* fell in the intermediate stage between the one- and two-species categories. The results of this chapter highlight the strong role that ecology plays in genome differentiation, and that population pairs from a similar ecological settings display parallel global levels of genomic differentiation. The degree of reproductive isolation and the relative contribution of adaptive phenotypic plasticity in pairs of other haplochromine species with lake-stream divergence remain to be investigated. Such research would enable uncovering generality of the patterns observed in *A. burtoni*.

The results of **Chapter 5** provide a strong evidence for a differentiating XY sex chromosomal system in a population of haplochromine cichlid *P. philander*. Furthermore, this sex chromosomal system, located on LG7, was lost in another population of the same species from the Mbulu stream, likely due to demographic events. There was no signature of shared ancestry with another XX-XY system previously characterized in the same genomic region in Lake Malawi haplochromine cichlids, which indicates that the two systems evolved independently. The same chromosome has also been recruited as sex chromosome in even more distantly related species, supporting the "limited options" theory (Marshall Graves and Peichel 2010), which states that certain chromosomes are particularly well suited to become sex chromosomes at different phylogenetic levels.

Chapter 6 raises a concern that the risks associated with the intended exploitation of fossil hydrocarbons in the East African Great Lakes region are gravely underestimated and endanger any future research possibilities. To develop economically and ecologically viable strategies for the region, all stakeholders - the local population, regional stakeholders, governments, nongovernmental organizations, and scientists - must cooperate, which is currently being done for the Virunga National Park in the DR Congo. Unless we manage to protect these biodiversity generators and reservoirs, we will lose the chance to further our understanding of some of the most interesting examples of a rare biological phenomenon: explosive adaptive radiations.

In conclusion, I believe that the results of the experiments presented here are a valuable addition to a growing knowledge about early stages of speciation. I hope that the results presented in my thesis will contribute to the further establishment of *A. burtoni* as a model system for linking ecology and genome evolution, and finally shed more light on the speciation process.

References

Baldo, L., J. L. Pretus, J. L. Riera, Z. Musilova, A. R. Bitja Nyom, and W. Salzburger. 2017. Convergence of gut microbiotas in the adaptive radiations of African cichlid fishes. ISME J. 11:1975–1987.

Bolnick, D. I., L. K. Snowberg, J. G. Caporaso, C. Lauber, R. Knight, and W. E. Stutz. 2014. Major Histocompatibility Complex class IIb polymorphism influences gut microbiota composition and diversity. Mol. Ecol. 23:4831–4845.

Brucker, R. M., and S. R. Bordenstein. 2012. Speciation by symbiosis. Trends Ecol. Evol. 27:443–451.

Brucker, R. M., and S. R. Bordenstein. 2013. The hologenomic basis of speciation: gut bacteria cause hybrid lethality in the genus Nasonia. Science 466:667–669.

Chandler, J. A., and M. Turelli. 2014. Comment on "The hologenomic basis of speciation: Gut bacteria cause hybrid lethality in the genus Nasonia." Science 345:1011.

Franchini, P., C. Fruciano, T. Frickey, J. C. Jones, and A. Meyer. 2014. The gut microbial community of Midas cichlid fish in repeatedly evolved limnetic-benthic species Pairs. PLoS One 9: e95027.

Kaufmann, J., T. L. Lenz, M. Kalbe, M. Milinski, and C. Eizaguirre. 2017. A field reciprocal transplant experiment reveals asymmetric costs of migration between lake and river ecotypes of three-spined sticklebacks (*Gasterosteus aculeatus*). J. Evol. Biol. 30:938–950.

Leftwich, P. T., N. V. E. Clarke, M. I. Hutchings, and T. Chapman. 2018. Gut microbiomes and reproductive isolation in Drosophila. Proc. Natl. Acad. Sci. 115: E2487.

Lokesh, J., and V. Kiron. 2016. Transition from freshwater to seawater reshapes the skin-associated microbiota of Atlantic salmon. Sci. Rep. 6:1–10.

MacColl, A. D. C., and S. M. Chapman. 2010. Parasites can cause selection against migrants following dispersal between environments. Funct. Ecol. 24:847–856.

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

McFall-Ngai, M., M. G. Hadfield, T. C. G. Bosch, H. V. Carey, T. Domazet-Lošo, A. E. Douglas, N. Dubilier, G. Eberl, T. Fukami, S. F. Gilbert, U. Hentschel, N. King, S. Kjelleberg, A. H. Knoll, N. Kremer, S. K. Mazmanian, J. L. Metcalf, K. Nealson, N. E. Pierce, J. F. Rawls, A. Reid, E. G. Ruby, M. Rumpho, J. G. Sanders, D. Tautz, and J. J. Wernegreen. 2013. Animals in a bacterial world, a new imperative for the life sciences. Proc. Natl. Acad. Sci. 110:3229–3236.

Nosil, P., L. J. Harmon, and O. Seehausen. 2009. Ecological explanations for (incomplete) speciation. Trends Ecol. Evol. 24:145–156.

Schmidt, V. T., K. F. Smith, D. W. Melvin, and L. A. Amaral-Zettler. 2015. Community assembly of a euryhaline fish microbiome during salinity acclimation. Mol. Ecol. 24:2537–2550.

Sharon, G., D. SEgal, J. M. Ringo, A. Hefetz, I. Zilber-Rosenberg, and E. Rosenberg. 2010. Commensal bacteria play a role in mating preference of *Drosophila melanogaster*. Proc. Natl. Acad. Sci. 107:20051–20056.

Shropshire, J. D., and S. R. Bordenstein. 2016. Speciation by symbiosis: The microbiome and behavior. MBio 7:1–11.

Sullam, K. E., B. E. R. Rubin, C. M. Dalton, S. S. Kilham, A. S. Flecker, and J. A. Russell. 2015. Divergence across diet, time and populations rules out parallel evolution in the gut microbiomes of Trinidadian guppies. ISME J. 9:1508–1522.

Theis, A., F. Ronco, A. Indermaur, W. Salzburger, and B. Egger. 2014. Adaptive divergence between lake and stream populations of an East African cichlid fish. Mol. Ecol. 23:5304–5322.

Acknowledgement

First and foremost I am grateful to **Walter Salzburger** for providing me with an opportunity to conduct my PhD thesis in his lab and to take part in several field trips to Zambia, for introducing me to the exciting world of cichlids and for always being optimistic.

In particular I would like to thank **Bernd Egger** for introducing me to his "baby", the *Astatotilapia burtoni* lake-stream system, for supervision, support, and exciting fishing trips.

I am very thankful to **Martine Maan** and **Olivia Roth** for agreeing to be part of my defence committee.

I would like to acknowledge the **Swiss Zoological Society (SZG)** for a travel grant for the fieldwork that enabled me to produce the third chapter.

A big thank you goes to my co-authors, **Alexandra Weber** for being a great fieldwork and lab work companion, **Athimed El Taher**, for developing an idea for the third chapter and for introducing me to the exciting world of RNA, to **Astrid Böhne** for using our existing dataset to develop another interesting topic, and to all the people who participated in the **workshop on Oil extraction in Africa's Great Lakes** during Cichlid Science 2015 Meeting in Graz that lead to the publication of the manuscript in Chapter 6.

For an introduction to geometric morphometrics my thanks go to **Anya Theis**, for help with the statistics to **Nikolas Vellnow**, and for all the help in the field to **Fabrizia Ronco** and **Adrian Indermaur**.

A special thank you goes to **Nikol Kmentová** for introducing me to the world of fish parasites.

A big thanks goes to all the current and past members of the **Salzburger lab** for field trips, parties and friendly atmosphere. In particular I would like to thank my officemates **Dario**, **Marco**, **Nico**, **Athimed**, and **Telma**, for being a great and encouraging company, and especially for tolerating me these last few months; **Zuza** and **Fabio** for being my first friends in the lab and for being so enthusiastic about science; **Hugo** and **Atilla** for interesting discussions; **Kolja** and **Gaëlle** for being a great company in the field and in Basel; and My 'Basel ladies' **Alex**, **Athi** and **Telma** for all the fun.

For the logistic support in the field I am grateful to **Lorenz Makasa** and **Taylor Banda** from the Fisheries Department in Mplungu, **Vicky Huwiler** and **Peter Larsson** from the Kalambo Lodge who provided us with luxurious accommodation, and **Charity** and **Marino** who always made me feel welcome in Mplungu.

I owe another thank you to the **kids from Chipwa Zambia village** for being amazing *burtoni* fisherman and showing the mzungu how it's done.

I owe a special thank you to the crew of the Kalambo lodge without whom my experiments would have been impossible to conduct: **Philemon**, for taking care of the ponds, **Alfred**, for being the best *burtoni* fisherman, **Bernard**, for making every meal my favourite part of the day, and **Thomas** for being the best coxswain, translator and a guide to the local culture in Zambia.
For help during the last few days of writing a huge thanks goes to **Robert**, **Telma**, **Flora**, **Marija** and **Marinela** for proofreading, and to **Robert**, **Lukas**, **Barbara** and **Fabrizia** for feeding me.

Marinela, **Niko** and **Roberto**, thank you for adopting me and being my first family in Basel.

Niko, thank you for introducing me to Marinela and saving me from being homeless.

Marinela, thanks for being the best Basel party guide, a friend and a neighbour I could always count on.

Flora, thank you for being my friend and making the first days in Switzerland bearable.

Paulina, **Damian**, **Charlotte**, **Sara**, **Maryam**, and **Angelica** thanks for being my Mexican-Dutch-Columbian MEME crew in Europe.

Bibi, thanks for being the most fun friend in Switzerland, for introducing me to the Rhine swimming and being my 'Cupid'.

Zuza, thanks for being a friend and introducing me to the river snorkelling and fish adventures.

Fabio, thanks for the beers by the Rhine when I needed them the most.

Telma, thanks for being an inspiration.

Athi and **Alex**, thanks for all the fun climbing and bouldering sessions and teaching me how to push when I think I reached my limit.

Marija, thanks for always being there and proofreading my drafts in the operation room.

The **Spirigs**, thank you for making me feel like I have another family in Switzerland.

Robert, thank you for being the reason why I could be happy to live in one place for so long and for being my most critical manuscript reviewer.

Last but definitely not least, my mom **Duška**, thank you for believing in me and supporting me no matter what.