# ANNUAL REVIEWS

*Annual Review of Animal Biosciences*

# Genetic Variation and Hybridization in Evolutionary Radiations of Cichlid Fishes

Hannes Svardal,[1,2] Walter Salzburger,[3] and Milan Malinsky[3]

[1]Department of Biology, University of Antwerp, 2020 Antwerp, Belgium;
email: hannes.svardal@uantwerpen.be

[2]Naturalis Biodiversity Center, 2333 Leiden, The Netherlands

[3]Zoological Institute, University of Basel, 4051 Basel, Switzerland;
email: walter.salzburger@unibas.ch, millanek@gmail.com

## ANNUAL REVIEWS CONNECT

www.annualreviews.org

- Download figures
- Navigate cited references
- Keyword search
- Explore related articles
- Share via email or social media

## Abstract

Evolutionary radiations are responsible for much of the variation in biodiversity across taxa. Cichlid fishes are well known for spectacular evolutionary radiations, as they have repeatedly evolved into large and phenotypically diverse arrays of species. Cichlid genomes carry signatures of past events and, at the same time, are the substrate for ongoing evolution. We survey genome-wide data and the available literature covering 438 cichlid populations (412 species) across multiple radiations to synthesize information about patterns and sharing of genetic variation. Nucleotide diversity within species is low in cichlids, with 92% of surveyed populations having less diversity than the median value found in other vertebrates. Divergence within radiations is also low, and a large proportion of variation is shared among species due to incomplete lineage sorting and widespread hybridization. Population genetics therefore provides a suitable conceptual framework for evolutionary genomic studies of cichlid radiations. We focus in detail on the roles of hybridization in shaping the patterns of genetic variation and in promoting cichlid diversification.

# 1. INTRODUCTION

Since ancient times, humans have been fascinated by the diversity of life on Earth and sought to explain its origins (1). Today we know that all organismal diversity, including that of our own species, is the product of evolution, whereby DNA sequences carry heritable information across generations, and variation in DNA among individuals provides the substrate for selection to act upon. A substantial body of theory has been developed aiming at describing fundamental evolutionary processes and predicting their signatures on DNA sequence variation (2, 3). Having entered the era of DNA sequencing at the whole-genome level, genomics is now providing large amounts of data well-suited for testing theoretical predictions and advancing our knowledge regarding the interplay between evolutionary forces, genetic variation, and organismal diversity (4–9).

Constituting one of the most species-rich vertebrate families, cichlid fishes (family Cichlidae) serve as a prime model system in evolutionary biology, especially for understanding the mechanisms of organismal diversification in evolutionary radiations (10–15). Dozens of studies have queried cichlid genomes to address some of the most fundamental questions in evolutionary biology, such as,

- What are the genomic signatures and perhaps even determinants of evolutionary radiations?
- What are the links between the amount and distribution of genetic variation (across the genome as well as between genomes within and among species) and the associated potential for evolutionary diversification of populations into species and beyond?
- How is genetic variation generated and maintained, and specifically, what is the role of gene flow in this process?

As a result of these efforts, there is already a large, and rapidly expanding, volume of cichlid genetic and genomic data—under the keyword *cichlid*, we found at the time of writing 116 studies that generated 51.5 trillion base pairs in the National Center for Biotechnology Information (NCBI) Sequence Read Archive (16). However, it is not straightforward to obtain an overview of genome diversity in cichlids by integrating across studies that (*a*) used different genetic and sequencing approaches, (*b*) were designed to test distinct evolutionary hypotheses, and (*c*) probed specific aspects of particular cichlid evolutionary radiations. We believe that this lack of a unified view limits the ability of researchers to draw general and consistent conclusions. Therefore, a major goal of this review is to address this gap in knowledge by synthesizing available information and data about the patterns and sharing of genetic variation across cichlid fishes in the context of the phenomenon of evolutionary radiations.

A clear conclusion from previous studies is that hybridization between species is common in cichlids and provides a mechanism for the generation of (possibly adaptively relevant) genetic variation (e.g., 17–23). Therefore, in this review, we pay particular attention to the role of interspecific hybridization in cichlid radiations and its link to genetic diversity.

> **Evolutionary radiations:** periods of significant acceleration in the rate of species diversification in particular lineages

# 2. CICHLID FISHES AND THEIR EVOLUTIONARY RADIATIONS

Evolutionary radiations explain much of the variation in biological diversity across the tree of life (24). Some groups simply diversify much more and much faster than others. Cichlid fishes fall into this particularly fast category. The number of cichlid species is estimated to be in the range of 3,000–4,000, accounting for approximately one-tenth of all bony fishes (15). They are distributed across most tropical and some subtropical regions of Africa and the Americas and on Madagascar. In addition, there are several isolated populations and small groups of species, for example, in southern India and southern Iran and along the eastern coast of the Mediterranean (**Figure 1**).
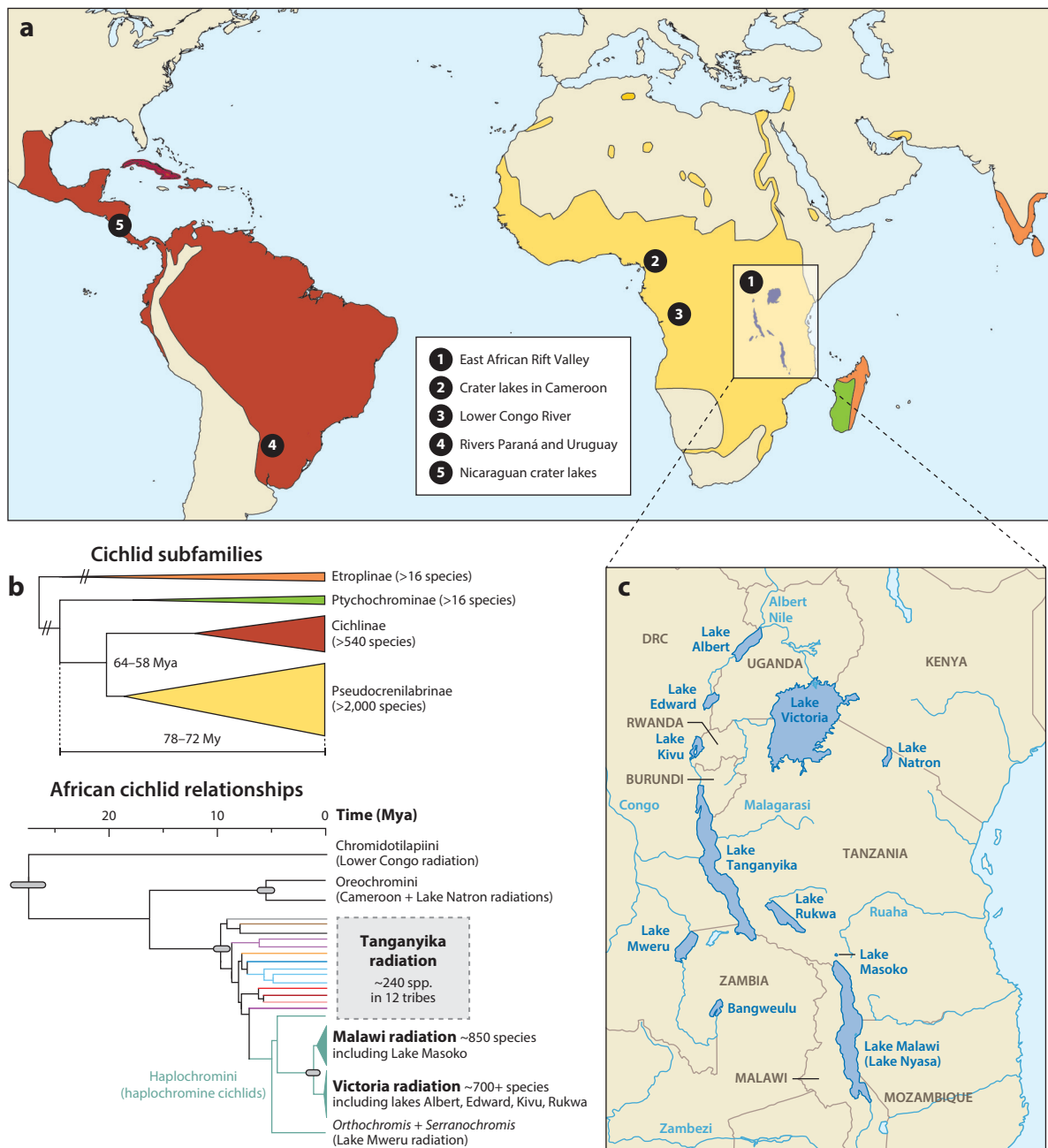
**Figure 1**

Cichlids and their evolutionary radiations. (*a*) The global distribution of cichlid fish subfamilies with colors corresponding to panel *b*. Adapted with permission from Reference 25; copyright 2018 Springer Nature. (*b*) Relationships among cichlid subfamilies and among African cichlids whose radiations are covered by this review. Timescales are based on Reference 26. Several other studies have estimated divergence times among some of the same lineages with varying conclusions; this topic is covered in detail by Matschiner and colleagues (25, 27). (*c*) A map of the African Great Lake region showing selected waterbodies that are discussed in this review.

**Adaptive radiation:**
an evolutionary
radiation in which new
species arise rapidly
through adaptation,
typically accompanied
by specialization, to
distinct ecological
niches

Throughout their distribution area, cichlids have colonized a wide variety of aquatic habitats, ranging from brackish waters to small creeks, large rivers, small (crater) lakes, and some of the largest lakes on our planet, and even extreme habitats, such as alkaline lakes in East Africa or deep and fast-flowing sections of the Congo River, where a blind cichlid occurs. The various cichlid species differ greatly in body size (from 3 cm to almost 1 m) and body proportions (from almost roundish to elongated); in feeding-related traits, such as mouth size and position, jaw morphology, and dentition; in social structures and mating behavior; and in pigmentation patterns (12, 28, 29).

Approximately half of the extant diversity of cichlid species is the product of the spectacular adaptive radiations in three of the Great Lakes of Africa: Tanganyika, Malawi, and Victoria (30). Because of their exceptional phenotypic diversity and speciation rates, the cichlid radiations in these lakes have been the focus of a large proportion of cichlid research (10–15, 26, 31). Although they are sometimes considered as a single phenomenon (e.g., 10, 32), the substantial differences among them call for nuanced consideration. Notably, the ages of the radiations vary by two or even three orders of magnitude. The Lake Tanganyika radiation is approximately 10 My old and is composed of multiple, highly diverged (>5 My; **Figure 1*b***) groups of cichlids that have been assigned to 12 different evolutionary lineages, so-called tribes (26). In contrast, the entirety of the Malawi and Victoria radiations consist of a single tribe, the Haplochromini (haplochromine cichlids). The Lake Malawi basin formed by approximately 5 Mya, but recent paleoecological studies provide evidence that the lake was largely dry between 1.6 and 1 Mya, and stable deep-lake conditions formed only approximately 800 kya (33, 34), which also corresponds to the estimates of the age of the radiations in recent genomic studies (20, 21). Finally, the Lake Victoria radiation itself is only 15,000 years old, although its origins are intertwined with the broader Lake Victoria Region Superflock (LVRS), which also includes cichlids in Lake Albert, Lake Edward, Lake Kivu, and several smaller waterbodies in the region and is approximately 100,000–200,000 years old (17, 35).

Each of the three massive cichlid adaptive radiations in Lake Tanganyika, Lake Malawi, and Lake Victoria brought forth an ecologically and morphologically highly diverse species assemblage characterized by a proportion of endemicity nearing 100% (30). Many of the descendant species of independent radiations nevertheless resemble each other—sometimes even in minute phenotypic detail—suggesting that convergent evolution owing to repeated and independent adaptations to the same environments is common in cichlids (31, 36, 37). The radiations also differ in estimated species number: Lake Tanganyika contains ∼240 cichlid species (38); the LVRS, ∼700 species, of which ∼500 are in Lake Victoria (17, 35); and Lake Malawi ∼850 species (39; see also 40).

In addition to the African Great Lakes, cichlids have diversified within other African lakes on at least 30 occasions (41). These smaller lakes present evolutionary radiations that are less spectacular in species number and diversity, but the reduced age and complexity mean that they present more tractable systems for addressing a range of evolutionary hypotheses (see, e.g., 42–44). Studies across lakes of varying ages also provide opportunities to investigate different stages of the diversification process, from its onset in as little as 50 years in Lake Chala (45), through the formation of genetically differentiated ecomorphs within a few thousand years in Lake Masoko (44), to mature radiations composed of hundreds of reproductively isolated species in the African Great Lakes (26). In contrast, cichlids colonized African lakes without diversifying on at least 120 occasions, providing opportunities for comparisons between radiating and nonradiating lineages (41).

Interestingly, cichlid diversification and speciation occurred in several cases without any apparent geographical barrier. A major textbook on speciation describes the radiations of cichlids in the Cameroonian crater lakes Barombi Mbo and Bermin as the most convincing examples of sympatric speciation in any group (46), and the above-mentioned ecomorphs of Lake Masoko (44)

and well-documented radiations of at least 13 species within a series of crater lakes in Nicaragua provide further examples of sympatric divergence (47–49). These volcanic crater lakes provide suitable conditions for testing theories and observing patterns of cichlid diversification with and without geographic isolation.

In addition to the lake environments, cichlids are also commonly found in rivers and streams. Some (predominantly) riverine lineages also show high levels of species diversity, especially in South and Central America, where the tribes Geophagini, Cichlasomatini, and Heroini represent hundreds of phenotypically and ecologically diverse species (50), with new species being discovered every year. Compared to their counterparts in lakes, the riverine cichlid faunas have received considerably less scientific attention, possibly because diversification in these lineages generally happened over longer timescales and/or longer ago compared to rapid lacustrine radiations. However, there are exceptions to this rule. Two recently discovered riverine species flocks may be the outcome of recent riverine radiations: At least 13 species from 2 genera originated within the rapids of the lower Congo River (51), and 17 species of the genus *Crenicichla* have evolved in parallel within the Paraná and Uruguay rivers (52).

Whereas the large cichlid radiations in the African Great Lakes have fascinated biologists for more than a century, many of the cases of cichlid diversification described above have been discovered only in the last few decades. New species and cichlid radiations almost certainly await discovery, both within sub-Saharan Africa and in the less accessible parts of South America.

## 3. GENETIC VARIATION IN CICHLIDS

The approximately 1 billion bases (∼1 Gb) of each cichlid's genome are a treasure trove of valuable information. Cichlid genomes inform us about their evolutionary history, and the variation among them is the substrate for ongoing evolution. The genome as a whole is shaped by the fundamental processes of mutation, recombination, and genetic drift, within a framework of demographic history that may also include migration and hybridization (**Figure 2a–c**), and only a relatively small fraction of a typical vertebrate genome is considered to be biologically functional, as assessed by being under long-term direct purifying selection (53, 54). This is also the case for cichlids. For example, Brawand et al.'s (32) analysis of five African cichlid genomes found that protein-coding genes comprised between 32 and 41 Mb, and conserved noncoding elements covered 72 Mb, thus suggesting the proportion of conserved functional DNA in cichlids is in the range of 10–11%. Because most SNPs occur at noncoding and nonfunctional sites, predictions derived from the neutral theory of molecular evolution can help us understand the observed patterns of genetic variation.

The development of high-throughput sequencing technologies allowed for the collection of genome-wide single-nucleotide polymorphism (SNP) data from hundreds of cichlid species. Some of the initial insights from these data sets may appear counterintuitive. Despite their extensive phenotypic diversity, the amounts of genetic diversity tend to be relatively low in cichlid evolutionary radiations (**Figure 3**). Moreover, a large fraction of genetic variation is shared among species (**Figure 4b**), including between those that are ecologically and phenotypically very different (e.g., a large pelagic pursuit predator and a small rock-dwelling algae grazer in Lake Malawi) (20). The rapid succession of speciation events, a characteristic feature of evolutionary radiations, leads to extensive incomplete lineage sorting (ILS), and hybridization provides another source of allele sharing (**Figure 2b,c**). Interestingly, the sharing of genetic variation (**Figures 2d** and **4b**) appears to extend even across independent radiations separated by mountain ranges and millions of years of divergence (55).

SNPs are the most common type of mutations. However, DNA sequences encompass other types of variation that differ in scale and origin and in our ability to interrogate them.

**Genetic drift:** the fluctuation in frequency of a genetic variant due to the randomness of which individuals reproduce

**Neutral theory:** a model of molecular evolution developed under the assumption that most genetic diversity and divergence are selectively neutral

**Incomplete lineage sorting (ILS):** a phenomenon where alleles within a species do not find a common ancestor within the duration of the species
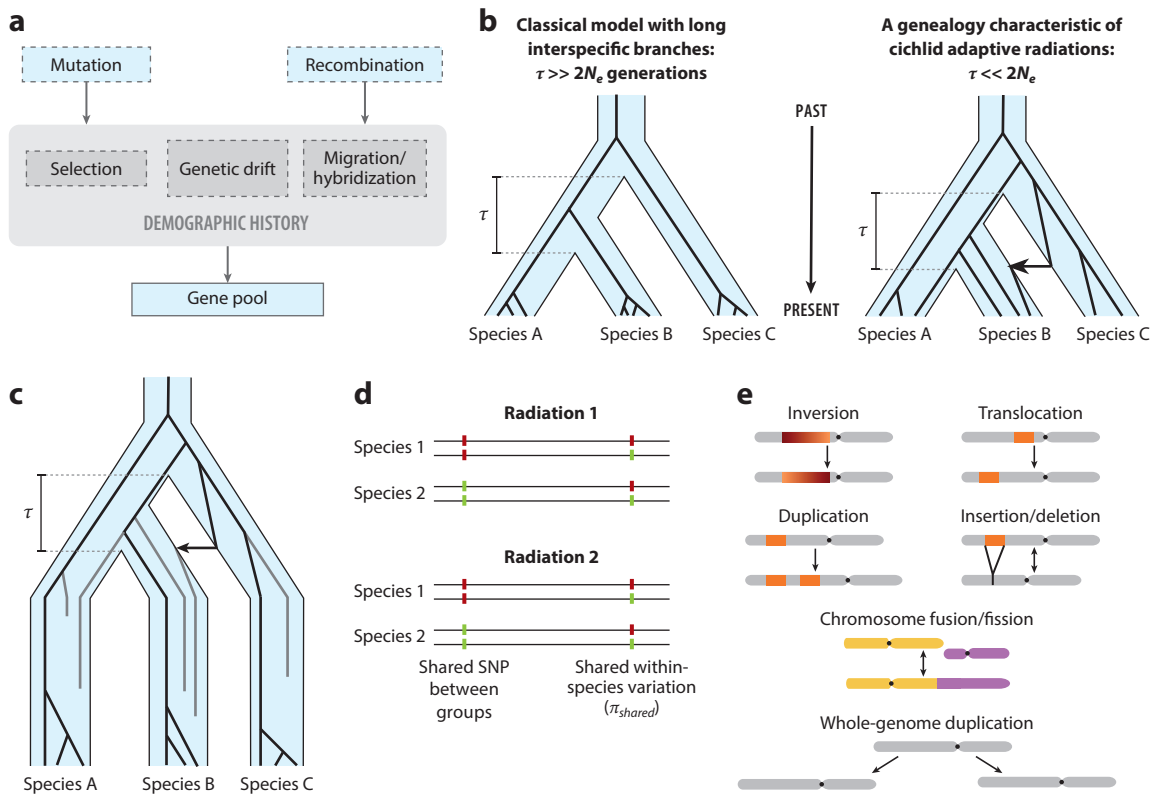
**Figure 2**

Genetic variation—key concepts. (*a*) Factors shaping the genetic composition of natural populations. (*b*) Incomplete lineage sorting (ILS) and hybridization violate the classical phylogenetic model. Genealogies of several sequences sampled from three species are shown for a single genomic locus. When the number of generations ($\tau$) between consecutive speciation events is short relative to the effective population size ($N_e$), this leads to ILS, whereby lineages do not coalesce (have a common ancestor) within the duration of a species. Both ILS and hybridization are extensive within cichlid radiations and in other rapidly speciating lineages, underlining the need for population genetic approaches. (*c*) ILS deep in history persists despite long terminal branches. (*d*) The two concepts of shared genetic variation used in this review. On the left is a single-nucleotide polymorphism (SNP) shared between evolutionary radiations. Despite not being polymorphic in any species, this SNP is counted in this measure, which was used in Reference 55. Shared variation (e.g., a heterozygous site) at the same locus within multiple species and across radiations. This is the basis of the $\pi_{shared}$ measure in **Figure 4b**. (*e*) Categories of structural genetic variation.

Multi-nucleotide changes and larger structural variants (**Figure 2e**), ranging in size from >50 bp to many megabases (Mb), comprise the majority of varying DNA bases among human genomes (56), a phenomenon now known to be taxonomically widespread (57). Therefore, in the final part of this section, we review what is known about structural variation in cichlids from comparisons of genome assemblies and from cytogenetic evidence.

## 3.1. Levels of Genetic Diversity

Genetic diversity is the amount of variation among DNA sequences within a species (or a population). Understanding of how and why genetic diversity levels differ among species is one of the fundamental goals of population genetics, with broad implications for evolutionary biology
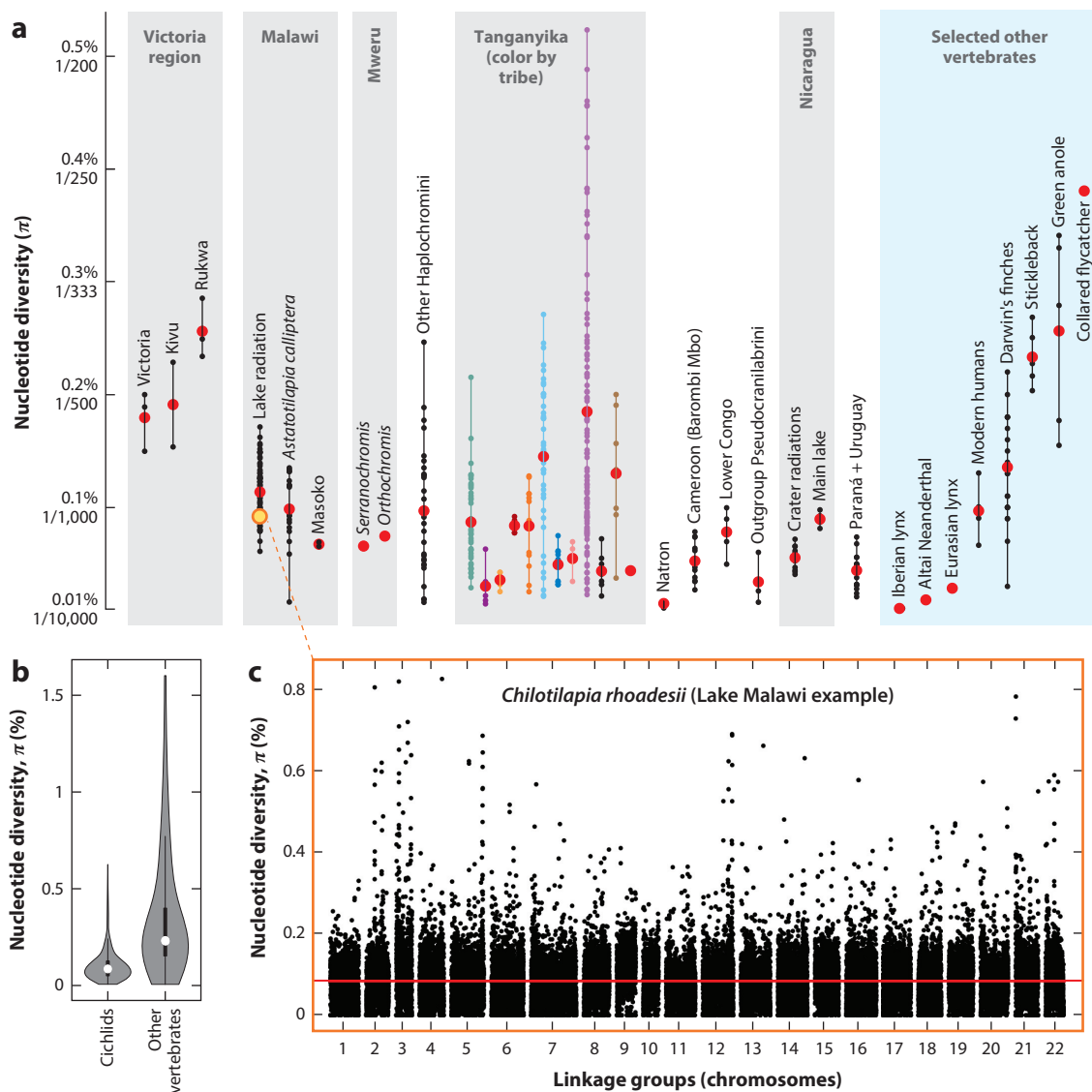
**Figure 3**

Nucleotide diversity ($\pi$). (*a*) Genome-wide average values across cichlid radiations, in nonradiating cichlids, and in selected other vertebrates. Cichlid values are from References 20, 26, 72; data analyses from References 73–76; and unpublished data from Z. Musilová (Barombi Mbo), L. Piálek (Paraná and Uruguay), and A. Kautt and A. Meyer (Nicaragua, later published in Reference 148). Values for other vertebrates are based on lynx, Neanderthal (64), Darwin's finches (65) (Waterson's θ), and green anole (66), and the remainder are from Leffler et al. (59). The tribe color scheme for Lake Tanganyika is based on Ronco et al. (26). All cichlid values, together with additional details on the data sources and calculations, are available online in **Supplemental Table 1**. Note that Leffler et al. (59) cite a $\pi$ estimate for the cichlid *Cichla pleiozona*, but the original source article (77) indicates that this was a mistakenly included estimate of mitochondrial diversity, not comparable to the genome-wide estimates reported here. (*b*) The overall distribution of $\pi$ in cichlids compared against all other vertebrate species ($n = 61$) surveyed by Leffler et al. (59). (*c*) Nucleotide diversity plotted in 10-kb windows along all chromosomes for the Lake Malawi species *Chilotilapia rhoadesii*. Based on data from Reference 20.
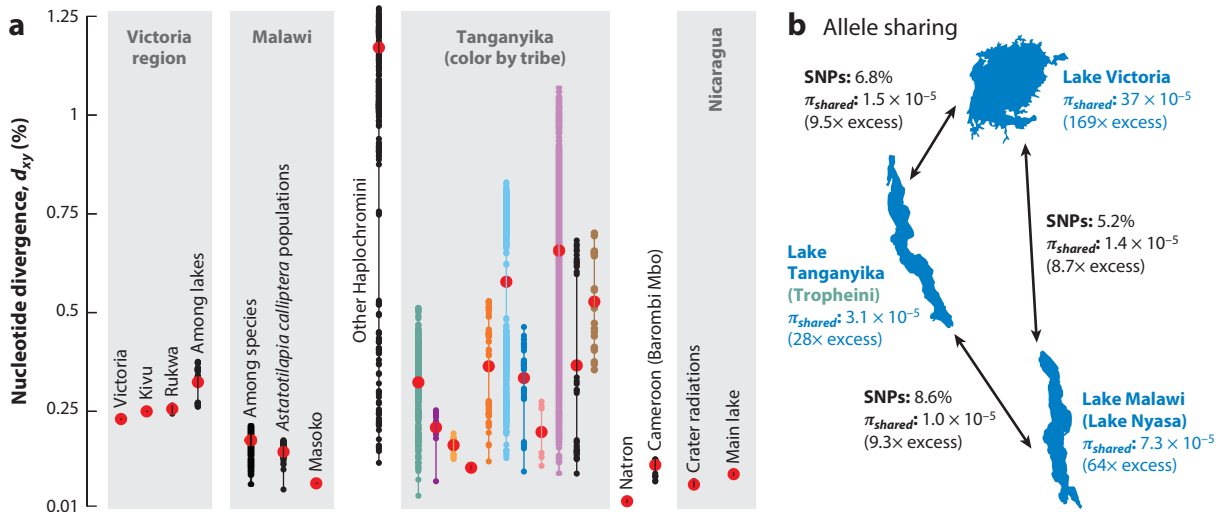
**Supplemental Material** >

**Figure 4**

Divergence and sharing of variation. (*a*) Nucleotide divergence ($d_{xy}$) within cichlid radiations. (*b*) Sharing of genetic variation within and between the three large Great Lake cichlid radiations, using the two measures illustrated in **Figure 2d**—the proportion of shared single-nucleotide polymorphisms (SNPs) and the frequency of shared heterozygous sites between individuals from different species/lakes ($\pi_{shared}$). Excess in $\pi_{shared}$ is relative to the expectations for unrelated species, under which $\pi_{shared} = \pi_{species1} \times \pi_{species2}$.

(58–60). Crucially, genetic variation at functional loci that underlie fitness-related traits provides raw material for natural selection, facilitating adaptation to new and rapidly changing environments (61–63), whereas a lack of such variation restricts short-term adaptive potential.

Genetic diversity is often estimated by calculating the average fraction of nucleotides that differ between any two sequences sampled from a species (or population), a measure called nucleotide diversity ($\pi$). We compiled genome-wide average measurements of $\pi$ from 469 species/populations across 9 cichlid evolutionary radiations (**Figure 3a**). A comparison with data from 61 other vertebrate species surveyed by Leffler et al. (59) revealed that nucleotide diversity levels within cichlids are relatively low (**Figure 3b**). The median $\pi$ across cichlids included in this review was 0.093% (one pairwise single-nucleotide difference in 1,075 bp). The overwhelming majority (403 of 438) of surveyed cichlid species/populations had lower $\pi$ than the median nucleotide diversity found in other vertebrates (0.23%), the only exceptions coming from the tribes Ectodini and Lamprologini from Lake Tanganyika, the haplochromines of Lake Rukwa, and one riverine haplochromine sample (**Figure 3a**). Among cichlids, it is also notable that all the LVRS samples, and the phylogenetically related Lake Rukwa (74), have relatively high $\pi$; the 0.21% mean value is higher than the mean for any other cichlid radiation. It will be interesting to see if this pattern holds as more data from across the LVRS become available. At the other end of the scale, the lowest $\pi$ in cichlids approaches 0.01% (one pairwise difference in 10,000 bp), which is among the smallest values ever recorded in natural populations of any species, comparable to those of the highly endangered Iberian lynx or to a Neanderthal sample from the Altai Mountains (64). Such extremely low diversity values are not limited to nonradiating cichlid lineages from small lakes or rivers, but all of the cichlids of the Lake Natron radiation, one Cameroonian crater lake species, and, most surprisingly, 13 species in 7 tribes of the massive Lake Tanganyika radiation have $\pi < 0.03\%$ (**Figure 3a**). Thus, Lake Tanganyika species span the entire range of nucleotide diversity values we found across cichlids (**Figure 3a**).

Interestingly, the levels of $\pi$ within the adaptive radiation of Darwin's finches (65) are comparable to those of cichlids. Threespine stickleback fish (63)—known for rapid adaptation from standing genetic variation—and green anole lizards (66) have a $\pi$ that is slightly higher but still below average when compared with the entire distribution in vertebrates. Overall, these results suggest that relatively low genome-wide nucleotide diversity levels do not necessarily limit rapid adaptation and speciation.

The amount of genetic diversity is determined by the time it takes for pairs of present-day DNA sequences to find a common ancestor (coalescence time) and by the mutation process introducing new genetic variants. Population genetic theory indicates that under a range of simplifying assumptions, including random mating, equal numbers of males and females, absence of selection, nonoverlapping generations, and constant population size (the so-called Wright–Fisher model), coalescence rates are determined by population size ($N$), and $\pi = 4N\mu$, where $\mu$ is the per-generation mutation rate. Departures from these assumptions are common in real populations, including in cichlids, and to account for this, $N$ is typically replaced with the effective population size ($N_e$; see 67), so that the equation becomes $\pi = 4N_e\mu$ (60). In either case, genetic diversity depends on the per-generation mutation rate. An estimate of mutation rates from whole-genome sequencing of parents and offspring of Lake Malawi cichlids ($\mu = 3.5 \times 10^{-9}$ per bp) is among the lowest reported for any vertebrate (20). Therefore, a low mutation rate may contribute to the observed low diversity levels in cichlids. Additional estimates of $\mu$ across different cichlid species and evolutionary radiations would help to confirm whether this is the case.

Demographic bottlenecks, that is, periods of low population size, have a disproportionally large effect on $N_e$ and therefore on genetic diversity. When population size fluctuations are the only departure from the Wright–Fisher model, $N_e$ is calculated as the harmonic mean of population size over generations (67). The harmonic mean is strongly affected by the smallest values in the set. Consistent with this result, demographic bottlenecks during settlement and subsequent diversification in newly formed lakes are expected to deplete genetic diversity in the course of evolutionary radiations (e.g., 68). Such bottlenecks were found for all of the crater lake radiations in Nicaragua (69), where the known source populations have substantially higher $\pi$ than the radiations, although they are still very low in comparison to other vertebrates (**Figure 3a,b**). Demographic modeling and methods that estimate historical coalescence rates from whole-genome SNP data (e.g., 70, 71) will make it possible to reconstruct past fluctuations in $N_e$ within cichlid radiations and in this way identify any demographic bottlenecks that would have led to losses of genetic diversity.

The founders of cichlid radiations may have been genetically more diverse than today's species. Both the Malawi and the Victoria cichlid radiations were founded by hybrid populations (17, 74) (see Section 4), and there is also evidence of relaxed purifying selection in ancestral haplochromine cichlids (32). However, several observations speak against this hypothesis. Namely, (*a*) even the youngest radiations have low diversity [e.g., the entirety of Lake Victoria cichlids have several-fold-lower genetic variation than zebrafish populations (59)]; (*b*) riverine cichlids, which are potential founders of lacustrine radiations (e.g., see Other Haplochromini in **Figure 3a**), do not have a higher $\pi$ than radiating cichlids; and (*c*) divergence among species within radiations is low (**Figure 4a**).

The range of $\pi$ in animals is much narrower than expected from differences in population size among species (59, 60). This discrepancy, known since the 1970s as Lewontin's paradox (78), was referred to as "the central problem in population genetics" (58, p. 261). Possible explanations include fluctuations in population size and effects of linked selection, but the effects of the latter alone appear to be limited and insufficient to explain the paradox (60). Intriguingly, a broad-scale comparison using 31 families of animals across 8 major phyla found that life-history

**Coalescence time:** the time to the most recent common ancestor

**Coalescence rate:** the rate at which pairs of alleles find their common ancestor backward in time

**Effective population size ($N_e$):** the size of an idealized (Wright-Fisher) population that matches coalescent rates in an actual population

**Linked selection:** the effect of selection acting on genetic variants on other nearby (possibly neutral) genetic variants

traits, especially traits related to parental investment, explained 73% of the variance in $\pi$ (58). A speculative explanation for this surprising result is that taxa with high parental investment can survive deeper population bottlenecks, which have major long-term effects on $\pi$ (60). With a large variation in parental investment strategies and other life-history traits, and a wealth of genetic diversity data, cichlids are well positioned for testing whether the relationship between $\pi$ and life-history traits also holds at a smaller scale, within a single family of animals. Variation in the per-generation mutation rate $\mu$ could confound such comparison; therefore, this is another area in which additional estimates of $\mu$ across different cichlid lineages would be beneficial.

There is limited information regarding the variation in diversity along cichlid chromosomes. Therefore, we reanalyzed Lake Malawi whole-genome SNP data from Malinsky et al. (20), revealing substantial variation in $\pi$. **Figure 3c** shows the genome-wide plot of $\pi$ in 10 kb windows for a representative species, *Chilotilapia rhoadesii*, revealing that 47 genomic windows in this species reached $\pi > 0.5\%$, that is, >6× the average of 0.08%. Applying Malinsky et al.'s (20) per-generation SNP mutation rate estimate for Lake Malawi cichlids, we can suggest that the average coalescence time in these regions was ~700,000 generations ago ($\pi/2\mu$), which corresponds to approximately 2.1 Mya when assuming an average generation time of 3 years. What is surprising is not the existence of regions of such deep coalescence but the genomic extent of these highly diverse regions. It is unlikely that variation in mutation rates could lead to regions of elevated diversity of the scale and magnitude observed here; however, strongly suppressed recombination, as is often the case between inverted haplotypes, could be a potential explanation. Ancestral hybridization or introgression from divergent cichlid lineages (e.g., from *Serranochromis*) and balanced polymorphisms could also contribute to this large variance in $\pi$ along Lake Malawi cichlid genomes, as we discuss in more detail below. Future studies of recombination and hybridization dynamics at fine scales along the genome will help to clarify the degree to which these factors contribute to variation in $\pi$.

### 3.2. Divergence and Sharing of Genetic Variation Among Species

**Figure 4a** shows the distributions of genome-wide average nucleotide divergence $d_{xy}$ among species within each cichlid radiation. The $d_{xy}$ measure is defined analogously to $\pi$ as the average fraction of nucleotides that differ between any two sequences sampled from two different species (or populations). In the absence of gene flow, divergence between species reflects the sum of genetic diversity in the ancestral species and additional divergence accumulated due to mutations after the speciation event (79).

The combination of low within-species genetic diversity, a low mutation rate, rapid diversification, and prevalent gene flow means that the genomes of cichlid species within their radiations and beyond are rather similar to each other. We found $d_{xy}$ generally below 0.25% across entire radiations. Among other things, this has practical consequences for experimental design of genomic studies in cichlids—short reads from entire evolutionary radiations can often be aligned to a single reference genome and SNPs called (80), with mapping quality comparable to that of human studies. Nucleotide divergence tends to be greater among species in Lake Tanganyika than elsewhere, reflecting the greater age of this radiation. However, the values are still within the range of short-read mapping algorithms (81): $d_{xy}$ is below 1% within tribes and at most 2.2% between tribes.

One interesting observation from comparing the results in **Figures 3a** and **4a** is that the mean divergence ($d_{xy}$) among the three Lake Victoria species exceeds their mean $\pi$ by a higher margin than expected if the species had diverged from each other only after the formation of the lake ~15,000 years ago ($\overline{d_{xy}} - \bar{\pi} = 0.05\%$). Given the estimate of $\mu = 3.5 \times 10^{-9}$, the difference would

**Balanced polymorphism:** a polymorphism maintained by natural selection, e.g., when selection varies across space or time

take ∼71,000 generations to accumulate (with $d_{xy} - \pi = 2T\mu$, where $T$ is the species divergence time in generations). This suggests that the genetic divergence among haplotypes in these species started before the establishment of the current lake.

Although we have known that ILS leads to sharing of genetic variation in rapid cichlid radiations since at least the early 1990s (82), recent genomic studies have begun to quantify the extent of ILS and its impact on genetic variation among species. As an illustrative example, among 73 representative Lake Malawi species, for 82% of the SNPs heterozygous within individuals, both alleles were found in other species (20). Moreover, the building of 2,543 local phylogenetic trees along the Malawi cichlid's genomes resulted in 2,542 different tree topologies (20), reflecting how local genealogies vary along the genome and do not necessarily reflect the overall species tree (**Figure 2b**). These results are consistent with predictions from population genetic theory. The estimates of long-term $N_e$ in Malawi cichlids range from ∼50,000 to ∼130,000 (with $N_e = \pi / 4\mu$). The mean time to the most recent ancestor of two alleles sampled today is $2N_e$ generations, which corresponds to 300,000–780,000 years, close to the age estimate for the entire radiation. Taking into consideration the range of $\pi$ values across cichlid radiations (**Figure 3a**), analogous reasoning leads to the conclusion that ILS and allele sharing across species are likely to be ubiquitous.

Allele sharing and the effects of ILS are not purely within-radiation phenomena. An initial comparison of haplochromine cichlid genomes from lakes Malawi, Victoria, and Tanganyika revealed that only 56.7% of the genome supported the species tree (grouping Malawi and Victoria, with Tanganyika being the outgroup, as in **Figure 1b**), whereas 20.4% of genealogies had Victoria and Tanganyika as sister groups, and 22% of genealogies grouped Malawi and Tanganyika (32). This result reflects the fact that the extent of ILS among species is determined not by how far back in time the species diverged but by the relative timing of speciation events at the time of divergence (**Figure 2c**) and is consistent, for example, with human–chimpanzee–gorilla comparisons, where for 30% of the genome human and chimpanzee are not sister groups due to ILS (83).

Loh et al. (55) reported that 8.3% of ∼200 SNPs that were polymorphic among Lake Malawi species were also polymorphic in lakes Tanganyika and Victoria. Whole-genome data present a similar picture, with between 5.2% and 8.6% of SNPs shared between haplochromine cichlids in different lakes (**Figure 4b**). However, this measure of shared polymorphism between groups or radiations (**Figure 2d**) is biased by the number and relationships of samples within each group, leading to counterintuitive results, such as >10% SNP sharing between Lake Malawi and the cichlid tribe Tilapiini, which diverged more than 10 Mya (55). Therefore, we introduce a measure of allele sharing that does not suffer from this bias and considers the frequency of heterozygous sites that are shared between pairs of individuals ($\pi_{shared}$ in **Figures 2d** and **4b**). We find that, on average, heterozygous sites are shared between haplochromine cichlid individuals from different African Great Lake radiations approximately 9 times more often than expected for completely unrelated species, with little difference among the between-lake comparisons (see **Figure 4b**). In contrast, the extent of within-lake allele sharing as measured by $\pi_{shared}$ varies substantially. Compared with the expectation for unrelated species, Lake Victoria species shared on average 169 times more heterozygous sites; Lake Malawi species had a 64× excess; and species within the Tanganyika Tropheini group had only 28× excess of shared heterozygous sites, again reflecting the greater age of the Lake Tanganyika radiation.

Similar to nucleotide diversity, patterns of genetic divergence also can vary substantially along cichlid genomes. For example, $d_{xy}$ between closely related species of *Pundamilia* in Lake Victoria shows outliers that are several-fold higher than the mean (mean ∼0.05%, outliers up to 0.3%; windows of 15 SNPs each). Given the close relatedness of these species, this suggests a large within-genome variance in nucleotide diversity in their common ancestor (84).

The finding of substantial differences in the amount of genetic variation (in both $\pi$ and $d_{xy}$) along cichlid genomes is intriguing, because it raises the possibility that genetic diversity at functional loci may in some cases be considerably higher than expected from the genome-wide average. Further evidence supporting this hypothesis comes from a genomic comparison of a pair of cichlid ecomorphs in the crater lake Masoko (44). A genome scan based on allele-frequency differences revealed 98 well-demarcated regions of high differentiation (HDRs) between the ecomorphs. The observed pattern of allele-frequency differentiation and other lines of evidence suggested that these regions are enriched for functional targets of divergent selection (44). Interestingly, a large fraction of these HDRs also had elevated $d_{xy}$, with median $d_{xy}$ within HDRs = 0.08% and outside HDRs $d_{xy}$ = 0.05%. Accumulating this extra 0.03% difference in $d_{xy}$ by mutation would take ~43,000 generations, a timeframe several times longer than the age of the lake. Therefore, it appears that the elevated genetic variation in HDRs predates the origin of the lake. Specifically, genomic regions with elevated nucleotide diversity in the ancestral population of Lake Masoko were more likely to harbor functional targets that came under divergent selection upon the ecomorph separation.

Guerrero & Hahn (79) suggest that such regions may arise owing to sorting of ancestral balanced polymorphisms. Balancing selection is generally considered rare (85). However, both local adaptation and spatially varying selection are widespread and easily capable of generating balanced polymorphisms (86, 87), opening up the possibility that the relative lack of empirical examples of balancing selection may have more to do with the difficulties in detecting it than with its rarity in nature (85).

### 3.3. Cichlid Genome Assemblies and Structural Variation

Structural variants (**Figure 2e**) include deletions, insertions (often duplications), inversions, and translocations of DNA sequence and may result either from mutational processes associated with DNA replication, repair, and recombination (88) or from the activity of transposable elements (TEs)—selfish DNA sequences that move and replicate within the genome using machinery comparable to that found in some viruses (4, 89, 90). At the largest scale, chromosomal fusions and fissions lead to changes in karyotype (57).

The first five cichlid genome assemblies by Brawand et al. (32), although highly fragmented and covering only ~70% to 80% of the actual genome size, represented a landmark in cichlid genomics. Focusing on Africa, the genomes included the riverine Nile tilapia (*Oreochromis niloticus*) and four species from the East African Great Lake region: *Neolamprologus brichardi* of Lake Tanganyika, *Maylandia* (*Metriaclima*) *zebra* of Lake Malawi, *Pundamilia nyererei* of Lake Victoria, and the riverine haplochromine *Astatotilapia burtoni* from the Lake Tanganyika basin.

Comparisons among the genomes facilitated discoveries of structural variation in cichlids. Whereas many of the initial results remained descriptive, some suggested functional roles for structural variation. The genomes revealed hundreds of inversions with an average length of more than 50 kb; ~10,000–15,000 long deletions (>50 bp) with average lengths of 500–1,000 bp (91); and thousands of long duplicated sequences (32). Interestingly, a large excess of sequence duplications that contained genes was assigned to the common ancestors of the lake radiations (4.5- to 6-fold excess and 280 duplicated genes) and to the common ancestor of the rapidly radiating haplochromine cichlids (>6-fold excess and 148 duplicated genes). These events may have allowed for subsequent functional innovation by divergence between the sequences or expression patterns of the duplicated genes (32).

Pairwise comparisons of genomes revealed that TE insertions upstream of genes have on average a positive correlation with gene expression (32). This is consistent with TE activity affecting the evolution of gene regulation, as has been shown in other vertebrates (90, 92, 93), for

example, by introducing new promoter sequences. Specific TE-mediated insertions and deletions have been implicated in the evolution of two key traits in cichlids. First, a TE insertion upstream of a pigmentation gene is associated with formation of egg-spots, which are a key phenotypic innovation in haplochromine cichlids implicated in mating behavior (94). The insertion functions as a specific enhancer in iridophores, a type of pigment cell found in egg-spots. Second, a recent study of Lake Malawi cichlids (75) revealed that specific insertions and deletions caused by TE movement cause differential expression of opsin genes between Lake Malawi species, contributing to regulatory changes underlying visual sensitivity.

Comparisons of the repeat content of individual genome assemblies may be confounded by the differences in their completeness. However, the total sequence length (without gaps) in two recent chromosome-scale assemblies comes close to the full estimated genome size, enabling such comparisons. These greatly improved genomes of *O. niloticus* (95) and *M. zebra* (96) are based on long-read sequencing with contigs arranged into chromosome-scale assemblies using genetic maps. The assemblies reveal that TEs comprise between 35% and 37% of these genomes (96), twice Brawand et al.'s (32) initial estimate of 16% to 19%. Intriguingly, they also show that the Lake Malawi *M. zebra* genome has ~30% more recent TE insertions (sequence divergence <2%) than the riverine *O. niloticus* and an even greater excess of TE insertions in 15 kb upstream of genes (1,422 versus 338), raising the possibility that many TE-mediated regulatory changes may have contributed to diversification in Lake Malawi cichlids (96).

Genetic maps and the chromosome-scale nature of these assemblies also allowed the authors to look for large-scale structural differences (96). Both assemblies have 22 linkage groups (LGs), and large interchromosomal rearrangements (at megabase scale) appear to be rare. However, there are several very large rearrangements within chromosomes, mainly inversions and translocations, of up to 23 Mb in size. A comparison with cytogenetic evidence (97, 98) suggested that at least four (and up to eight) of these large rearrangements are related to metacentric/acrocentric alterations in the position of the centromere. A comparison among four genetic maps from different interspecies crosses between Lake Malawi cichlids also revealed a pattern of minimal interchromosomal changes (only nine events among the four maps), whereas a large ~19-Mb inversion was found on LG11 in the genus *Aulonocara*, and large rearrangements or regions of suppressed recombination were present on three other chromosomes.

The lack of large interchromosomal rearrangements found in the comparison of genome assemblies is consistent with cytogenetic studies, which suggest that there is little variation in chromosome numbers among cichlids, and especially within evolutionary radiations (99, 100). Chromosome numbers among African cichlids vary from $2n = 40$ to $2n = 48$, but the typical karyotype found in 28 out of 40 examined species was $2n = 44$. Notable examples of variation within the East African radiations include members of the Lamprologini ($2n = 42$), Eretmodini ($2n = 46$), and Tropheini ($2n = 40$) tribes of Lake Tanganyika and *Nimbochromis livingstonii*, a single Lake Malawi species with $2n = 42$. There is also variation in the presence versus absence of B chromosomes within the rapid radiations of Lake Victoria (101) and Lake Malawi (98). Interestingly, B chromosomes are female specific in at least one Victorian species and in six species of Lake Malawi, suggesting the possibility that these chromosomes may play a role in sex determination.

Even less variation in chromosome numbers was found among American cichlids (100). Altogether 27 species across 7 tribes had $2n = 48$ chromosomes, and only two species presented an exception to this rule: *Laetacara dorsigera* ($2n = 46$) and the discus fish of the genus *Symphysodon* ($2n = 60$), in which the very large number of chromosomes may be the result of ancestral whole-genome duplication or of large-scale rearrangements possibly linked with ancestral hybridization (102), and may have contributed to the unusual morphological adaptations of this disk-shaped species.

**Promoter:** the sequence of DNA in which proteins bind to conduct and regulate the transcription of a gene into RNA

**Enhancer:** regulatory DNA sequences where proteins bind, thus enhancing the transcription of a gene

**Genetic map:** shows the order of genetic markers according to how often they tend to be inherited together

**Linkage group (LG):** all of the genetic markers that tend to be inherited together because of physical linkage along the chromosome

**Cytogenetics:** the study of chromosomes and of their number and structure, mainly by means of microscopy

**B chromosome:** an extra nonessential chromosome that is not present in all individuals and does not follow Mendelian inheritance

Unfortunately, in the absence of comparisons among genome assemblies or genetic maps, cytogenetics is currently the only source of information about structural variation in many cichlid groups. Information about structural variation within radiations is lacking almost entirely. With the cost of high-quality cichlid genomes based on long reads continuing to fall, this situation is likely to change over the next few years. Two new genomes were recently generated as part of a community effort to provide high-quality vertebrate genome assemblies (7, 103). These are of the Central American *Archocentrus centrarchus* and of the East African *Astatotilapia calliptera* (associated with lakes Malawi and Masoko) and are now available to use from public depositories (GenBank accessions: GCA_007364275.2, GCA_900246225.3). More cichlids are in the community pipeline, e.g., *Amphilophus citrinellus* of the Nicaraguan crater lakes.

**Genetic incompatibilities:** negative interactions between alleles from parental species leading to decreased fitness, including sterility or inviability, of hybrids

## 4. HYBRIDIZATION IN CICHLIDS

### 4.1. Evidence for Hybridization and Gene Flow in the Lab and in the Wild

One of the revelations brought about by the boom in evolutionary genomics over the last decade has been that hybridization between closely related animal species is the rule rather than the exception (104–106). This is particularly true for young evolutionary radiations (65, 107, 108). The effect of this process on organismal diversification is debated (109–111), with the cessation of gene flow between lineages often seen as a prerequisite for speciation [e.g., under the biological species concept (46)]. Hybrid speciation, in which a hybridization event leads directly to a new species that is reproductively isolated from the parental species, occurred, for example, in the radiation of Darwin's finches (112) but may be relatively rare in animals (111). However, a rapidly growing body of literature suggests that hybridization can facilitate diversification by fueling adaptive processes via supplying new (combinations of) functional genetic variants (reviewed in 113).

One remarkable feature of cichlids is that they are able to hybridize over large evolutionary distances, which may be facilitated by the above-described relative stability of cichlid genome architecture (15). Viable F2 hybrids can be produced between species from the Lake Victoria and Lake Malawi radiations, demonstrating that rates of accumulation of genetic incompatibilities are orders of magnitude slower than speciation rates (114). Mean sequence divergence between Malawi and Victoria radiations is approximately 0.76% (74), and subtracting within-species diversity results in $\overline{d_{xy}} - \bar{\pi} = 0.62\%$, translating to ~885,000 generations, which is roughly equivalent to the number of generations since the split between modern humans and gibbons (115). Extrapolation from these aquarium experiments suggests that hybridization is in principle possible between thousands of haplochromine cichlid species whenever they come into contact. Also, ecologically different species from the Nicaraguan crater lake radiation (*Amphilophus* spp.) have been shown to hybridize relatively easily in the laboratory (116), and some of the most popular cichlids in the aquarium hobby are man-made hybrids [e.g., the flowerhorn is likely a hybrid complex of three genera, including *Amphilophus* (117)].

Indeed, genome sequencing studies found evidence for cross-species genetic exchange in the wild affecting almost all investigated cichlid radiations. These efforts discovered signatures of extensive hybridization among lineages in the evolutionary radiations of lakes Malawi (20), Victoria (118, 119), Tanganyika (21–23, 26), Mweru (18), and Natron (72); Nicaraguan crater lakes Apoyo and Xiloá (120); and the tributaries of the Paraná River (121). Furthermore, there is evidence of hybridization before or early in the adaptive radiation in the LVRS (17), in the Lake Malawi radiation (74), between the early lineages (tribes) of Lake Tanganyika (21, 23; but see 26), in Lake Mweru (18), and in Nicaraguan crater lakes Apoyo and Xiloá (49). Finally, hybridization between the radiating lineages and riverine outgroup species was found for crater lakes Barombi Mbo in Cameroon (122, 123) and Lake Masoko in Tanzania (44).

Because of the extent of ILS (**Figure 2*b*,*c***), studies of hybridization within cichlid radiations require approaches that use data from a large number of neutral markers and take allele sharing among species due to ILS as a part of the null expectation (see the sidebar titled Detecting Gene Flow in Cichlids). In particular, phylogenetic discordance between nuclear and mitochondrial genetic markers is expected under ILS, and it is not safe to assume that mitochondrial DNA evolves as a neutral marker (124). Therefore, studies that suggested hybridization on the basis of mito-nuclear discordance should be interpreted with caution and reevaluated with genome-wide data where possible. Indeed, several hybridization scenarios previously suggested on the basis of mito-nuclear discordance are not supported by whole-genome data. For example, a suggested hybrid origin of the mbuna group (125), and substantial hybridization between the deep benthic and mbuna groups (126) in Lake Malawi, are scenarios that were not supported by genome sequencing data (20). Similarly, mito-nuclear discordances in the Iguazú and Paraná rivers (121) and among cichlids in the Congo River (76) are not reflected in genome-wide data in those studies. However, in some other cases [e.g., Lake Tanganyika *Neolamprologus* (127)], early suggestions of hybridization were later corroborated by whole-genome data (26).

<div style="float:right">

**Mito-nuclear discordance:** disagreement of phylogenies derived from mitochondrial and nuclear genetic markers

</div>

## 4.2. The Direct Effect of Hybridization on Genetic Diversity

Hybridization and introgression can increase within-species genetic diversity, and there are several lines of evidence that this is happening in cichlids. First, a simulation study by Ronco et al. (26) showed that the patterns of introgression among Lake Tanganyika cichlids can by themselves account for the differences in $\pi$ across tribes, including for the correlation of tribe-wise $\pi$ with species richness discovered by the authors. Second, we found that the highest levels of genetic diversity among haplochromine cichlids not belonging to the Great Lake radiations were in populations for which evidence for admixture between divergent lineages was reported by Svardal et al. (74): *Astatotilapia* spp. from the Ruaha River and *A. bloyeti* from Lake Kumba. Finally, it is possible that the relatively elevated $\pi$ within LVRS is linked to the hybridization event before the onset of that radiation (17), although this raises the question of why we do not observe a similar effect following the hybridization which preceded the Lake Malawi radiation (74).

### DETECTING GENE FLOW IN CICHLIDS

The majority of genomic studies of hybridization in cichlids applied tests of imbalance in allele sharing based on the easy-to-compute ABBA-BABA family of statistics (e.g., Patterson's D and the $f_4$-admixture ratio) (141). These tests are robust when applied to genome-wide SNP data sets (142) and are generally considered good indicators of genetic exchange after species or population divergence. Some potential pitfalls remain—notably, ancestral population subdivision and differences in substitution rates among taxa may lead to false positives (143, 144), and pinpointing specific introgression events within a system of many tests in larger radiations can be challenging (20, 145). Several studies, especially within the older Lake Tanganyika radiation (21, 22), but also the Nicaraguan crater lakes (120), applied a phylogenomic framework under the multispecies coalescent. More recent gene flow can be deducted from patterns of nearest-neighbor haplotype sharing (146, 147), a method used for cichlids in Lake Malawi (20), Lake Mweru (18), and the Paraná River (121). Finally, model-based inference can be computationally demanding but is useful for detailed investigations of specific demographic hypotheses, as illustrated for Lake Victoria *Pundamilia* (119), Nicaraguan crater lakes (69), and the Cameroonian Lake Ejagham (123).

## 4.3. The Role of Hybridization in Diversification

Is the prevalence of hybridization in cichlids merely a consequence of incomplete reproductive and limited geographic isolation during rapid diversification, or does hybridization have a positive effect on diversification rates? Experimental work has shown that interspecific cichlid hybrids can produce extreme and novel phenotypes, with the degree of phenotypic novelty increasing with genetic divergence between the hybridizing lineages (128). This phenomenon, known as transgressive segregation, has also been demonstrated in the wild, at secondary contact zones between cichlid populations (129). Moreover, experimental work suggests that the performance of transgressive hybrids, and thus the evolutionary impact of hybridization, may be context dependent—specifically depending on the availability of ecological resources outside of the parental niches (130). The hybrid swarm hypothesis predicts that hybridization at the base of an adaptive radiation, before diversification, could promote adaptive radiation by increasing heritable genetic variation in ecologically relevant traits (109). Furthermore, hybridization among species within a radiation may generate genotype combinations that allow previously unoccupied fitness peaks to be reached (109) by combining alleles from different sources to create novel phenotypes (113). In the following, we review, in the context of cichlid diversification, the evidence for functional roles of the reported occurrences of hybrid swarms, of gene flow between lineages within radiations, and of hybridization with outgroup lineages.

### 4.3.1. Hybrid swarms.

Evidence of hybridization at the base of an adaptive radiation [as documented in lakes Malawi, Mweru, and Tanganyika; in the LVRS; and in the Nicaraguan lakes Apoyo and Xiloá (17, 18, 21, 23, 49, 74)] is a fundamental prediction of the hybrid swarm hypothesis. However, further evidence is required to ascertain whether an early hybridization event facilitated the onset or progression of these radiations. Such evidence is available for the LVRS and Lake Malawi, where patterns of allele segregation indicate that hybridization-derived polymorphisms were under divergent selection during speciation events early in the radiations (17, 74). Moreover, Meier et al. (17) identified that for the long-wavelength sensitive opsin, a gene known to be involved in adaptation and speciation in Lake Victoria (131), the two alleles used in recent adaptive divergence were contributed by the ancestral hybridization event. Additional evidence for hybridization facilitating the Lake Mweru radiation is not genetic but comes in the form of eco-morphological comparisons with the nonradiating lineages of the similar Lake Bangweulu, exemplifying the benefits of integrating different lines of evidence in cichlid research (18). Overall, the cichlid fish family has provided substantial evidence for the involvement of hybrid swarms in evolutionary radiations, but much more work is needed to uncover the mechanisms behind this phenomenon.

Somewhat counterintuitively, overall genetic variation in adaptive radiations with ancestral hybridization is not particularly elevated (**Figure 3a**). Svardal et al. (74) showed that most of the variation derived from the ancient hybridization is no longer segregating across Lake Malawi cichlid species—at most of the loci that originally had hybridization-derived variation, all present-day species carry the same allele. A possible explanation for this is that most hybridization-derived variation was lost via drift or selection before the onset of adaptive radiation, while some functional genetic variants, which were later important for diversification, escaped this process, for example, because of negative frequency-dependent selection across a spatially structured population (87).

### 4.3.2. Adaptive introgression within radiations.

Despite the extensive reports of hybridization in cichlid evolutionary radiations (18, 20–23, 26, 72, 118–121), so far we know only little about

specific genes exchanged across species and even less about possible adaptive advantages conferred by it. Despite the difficulty in demonstrating an adaptive function for introgressed regions, several examples have recently been reported in other organisms (reviewed in 9, 132, 133), including, for example, wing-pattern mimicry loci in *Heliconius* butterflies (134) and altitude adaptation alleles introgressed into modern humans from archaic hominins (135). Possibly the most convincing evidence for adaptive introgression in cichlids to date comes from Lake Malawi (20), suggesting that opsin genes have introgressed between relatively distantly related clades facilitating visual system adaptation to the deep-water environment. Meier et al. (84, 119) provide evidence consistent with introgression facilitating parallel ecological speciation into deep- and shallow-water specialists in *Pundamilia* cichlids from Lake Victoria. On the whole, the study suggests that parallel selection acted on introgression-derived variants, which may also include variants in opsins relevant to adaptation to different light regimes.

### 4.3.3. Introgression with outgroups.

Evidence for introgression with outgroups comes predominantly from studies of crater lake cichlid radiations, where analyses focused on detection of secondary gene flow that could have interfered with sympatric diversification within the lakes. Such signatures of secondary gene flow were reported in all cases, whereas estimates of admixture fraction tend to be variable: $\sim10^{-6}$ in Barombi Mbo in Cameroon (122), 1.2% in Lake Masoko (44), 4.3% in Nicaraguan Lake Apoyo (69), 4.7% in Cameroonian Lake Ejagham (123), and up to 32% in the Nicaraguan crater lake Asososca Managua (69). However, a key question is whether the alleles obtained by secondary gene flow contribute to the sympatric divergence within the crater lakes. This appears to be the case in Lake Ejagham (123), where introgressed blocks of olfactory receptor genes may have contributed to subsequent speciation events. In contrast, no correlation between secondary introgression and signatures of species divergence was found in lakes Barombi Mbo (122) and Masoko (44). For the Nicaraguan crater lakes, this question remains to be explored.

Loh et al. (55) hypothesized that the unexpectedly high sharing of polymorphisms among the three East African Great Lake radiations (despite their geographic separation) may be mediated by riverine transport of alleles. This is supported by the observation that major catchment boundaries are permeable; e.g., *A. calliptera*, which is a part of the Lake Malawi radiation, is also found in the Indian Ocean catchment (20), and a Lake Victoria–related lineage is found in Lake Tanganyika (136). Furthermore, there are signatures of gene flow between the Tanganyika endemic haplochromine cichlids (Tropheini) and riverine haplochromines, which are nested in the Tanganyika radiation (26).

## 5. CONCLUSIONS

A large body of literature aims to explain why cichlids diversify and what determines their diversification rates (10–15), including discussions of genomic features (10, 14, 15). However, an integrated overview of genetic variation, e.g., of the type provided by the 1000 Genomes (137) or Human Genome Diversity (138) projects for humans, is currently missing in cichlids. In this review, we made a step toward filling this knowledge gap by surveying levels of genetic diversity from publications and genome-wide data covering hundreds of cichlid species across a range of evolutionary radiations on two continents.

We found that cichlids have low levels of nucleotide diversity within species when compared to other vertebrates. This is true across different radiations, as well as nonradiating riverine cichlids, making this a general feature of the cichlid family. It appears that the rapid diversification and massive and phenotypically diverse evolutionary radiations could unfold in cichlids despite these relatively low levels of genetic variation at the single-nucleotide level, which is intriguing.

In this context, it is important to keep in mind the distinction between genome-wide (and therefore largely neutral) genetic diversity and genetic diversity at functional loci, which underlies variation in fitness among individuals and therefore is the basis for adaptation and organismal differentiation (139). Initial evidence suggests that balanced polymorphisms may contribute to the maintenance of genetic diversity at these "loci of evolution" in cichlids (79).

Sharing of genetic variation across species boundaries is clearly ubiquitous, due to both ILS and hybridization. This means that genomic analyses of cichlid species flocks—though they consist of many species and are phenotypically diverse—generally fall within a population genetic framework. However, the differences in the extent of (population) genetic structure within radiations are substantial. Notably, the Lake Tanganyika radiation is by far the most structured, offering perhaps the best opportunity to study both micro- and macroevolutionary processes within a single evolutionary radiation.

The roles of structural variation in adaptation and speciation are gaining increasing recognition (57). Initial findings suggest that gene duplications and TE activity have been important in cichlid evolution. The role of TEs is interesting for two main reasons. First, these selfish elements provide an alternative route to rapidly generating functional genetic diversity, which may be important for adaptation when standing genetic variation is low. Second, repression of TE activity can break down in hybrids (e.g., 140), perhaps because of mismatches between the TE sequences and the host mechanisms repressing TE activity in the germ line. Therefore, hybridization in cichlids may contribute to increased TE activity (75).

Hybridization has been widespread within cichlid radiations and contributed substantially to patterns of genetic diversity. Moreover, cichlids provide some of the strongest evidence available for the hybrid swarm hypothesis. However, specific examples of adaptive introgression within cichlid radiations are so far limited.

Despite the already large body of work and amount of data collected, our understanding of the roles of genetic diversity in cichlid evolutionary radiations is still only scratching the surface. With the boom in evolutionary genomics set to continue and likely accelerate, we look forward to many new discoveries in the coming years.

## FUTURE ISSUES

1. To understand the origin of genetic variation, we need to know if per-generation mutation rates vary substantially among cichlid species or among radiations. We also need to know more about how much the average generation time varies among cichlid species.

2. To appreciate the role of the random process of genetic drift in reducing genetic variation, we must reconstruct demographic histories of cichlid evolutionary radiations at a broad taxonomic scale.

3. To make progress toward pinpointing the functional elements of genetic variation, we need more work on detecting signatures of selection and genotype-trait association within and among cichlid species. This will benefit from collections of whole-genome SNP data from multiple individuals per species.

4. Detecting selection within admixed species will further our understanding of the adaptive role of gene flow in cichlids.

5. Additional insights into functional genetic variation will come from an expansion of transcriptomic approaches. Gene editing is likely to become common for functional validation of putative major effect loci.

6. We need more long-read whole-genome assemblies to discover structural variation within evolutionary radiations, especially gene duplications and transposable element insertions within gene promoters.

7. Obtaining fine-scale genetic maps along chromosomes will provide insights into the potential roles of suppressed recombination in maintaining favorable combinations of functional alleles.

8. To better understand the feasibility of rapid adaptation while lacking available functional standing genetic variation, we need studies assessing phenotypic plasticity and epigenetic changes in cichlids.

## DISCLOSURE STATEMENT

## ACKNOWLEDGMENTS

## LITERATURE CITED

1. Aristotle. 1965. *History of Animals*, Vol. I: Books 1–3. Cambridge, MA: Harvard Univ. Press
2. Barton NH. 2007. *Evolution*. Cold Spring Harbor, NY: Cold Spring Harb. Lab. Press
3. Charlesworth B, Charlesworth D. 2010. *Elements of Evolutionary Genetics*. Greenwood Village, CO: Roberts & Co.
4. Zerbino DR, Paten B, Haussler D. 2012. Integrating genomes. *Science* 336(6078):179–82
5. Seehausen O, Butlin RK, Keller I, Wagner CE, Boughman JW, et al. 2014. Genomics and the origin of species. *Nat. Rev. Genet.* 15(3):176–92
6. Berner D, Salzburger W. 2015. The genomics of organismal diversification illuminated by adaptive radiations. *Trends Genet.* 31(9):491–99
7. Koepfli K-P, Paten B, Genome 10K Community Sci., O'Brien SJ. 2015. The Genome 10K Project: a way forward. *Annu. Rev. Anim. Biosci.* 3:57–111
8. Lewin HA, Robinson GE, Kress WJ, Baker WJ, Coddington J, et al. 2018. Earth BioGenome Project: sequencing life for the future of life. *PNAS* 115(17):4325–33
9. Taylor SA, Larson EL. 2019. Insights from genomes into the evolutionary importance and prevalence of hybridization in nature. *Nat. Ecol. Evol.* 3(2):170–77
10. Kocher TD. 2004. Adaptive evolution and explosive speciation: the cichlid fish model. *Nat. Rev. Genet.* 5(4):288–98
11. Seehausen O. 2007. Evolution and ecological theory: Chance, historical contingency and ecological determinism jointly determine the rate of adaptive radiation. *Heredity* 99:361–63
12. Turner GF. 2007. Adaptive radiation of cichlid fish. *Curr. Biol.* 17(19):R827–31
13. Santos ME, Salzburger W. 2012. Evolution. How cichlids diversify. *Science* 338(6107):619–21
14. Seehausen O. 2015. Process and pattern in cichlid radiations—inferences for understanding unusually high rates of evolutionary diversification. *New Phytol.* 207(2):304–12

15. Salzburger W. 2018. Understanding explosive diversification through cichlid fish genomics. *Nat. Rev. Genet.* 19(11):705–17

16. Leinonen R, Sugawara H, Shumway M, Int. Nucleotide Seq. Database Collab. 2011. The sequence read archive. *Nucleic Acids Res.* 39(Database Issue):D19–D21

17. Meier JI, Marques DA, Mwaiko S, Wagner CE, Excoffier L, Seehausen O. 2017. Ancient hybridization fuels rapid cichlid fish adaptive radiations. *Nat. Commun.* 8:14363

18. Meier JI, Stelkens RB, Joyce DA, Mwaiko S, Phiri N, et al. 2019. The coincidence of ecological opportunity with hybridization explains rapid adaptive radiation in Lake Mweru cichlid fishes. *Nat. Commun.* 10:5391

19. Willis SC, Macrander J, Farias IP, Ortí G. 2012. Simultaneous delimitation of species and quantification of interspecific hybridization in Amazonian peacock cichlids (genus *Cichla*) using multi-locus data. *BMC Evol. Biol.* 12:96

20. Malinsky M, Svardal H, Tyers AM, Miska EA, Genner MJ, et al. 2018. Whole-genome sequences of Malawi cichlids reveal multiple radiations interconnected by gene flow. *Nat. Ecol. Evol.* 2(12):1940–55

21. Meyer BS, Matschiner M, Salzburger W. 2016. Disentangling incomplete lineage sorting and introgression to refine species-tree estimates for Lake Tanganyika cichlid fishes. *Syst. Biol.* 66(4):531–50

22. Gante HF, Matschiner M, Malmstrøm M, Jakobsen KS, Jentoft S, Salzburger W. 2016. Genomics of speciation and introgression in Princess cichlid fishes from Lake Tanganyika. *Mol. Ecol.* 25(24):6143–61

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

24. Wiens JJ. 2017. What explains patterns of biodiversity across the Tree of Life? *BioEssays* 39(3):1600128

25. Matschiner M. 2019. Gondwanan vicariance or trans-Atlantic dispersal of cichlid fishes: a review of the molecular evidence. *Hydrobiologia* 832:9–37

26. Ronco F, Matschiner M, Böhne A, Boila A, Büscher H, et al. 2020. Drivers and dynamics of a massive adaptive radiation in African cichlid fishes. *Nature*. In press

27. Matschiner M, Böhne A, Ronco F, Salzburger W. 2020. The genomic timeline of cichlid fish diversification across continents. *Nat. Commun.* 11(1):5895

28. Barlow G. 2008. *The Cichlid Fishes: Nature's Grand Experiment in Evolution*. New York: Basic Books

29. Salzburger W. 2009. The interaction of sexually and naturally selected traits in the adaptive radiations of cichlid fishes. *Mol. Ecol.* 18(2):169–85

30. Salzburger W, Van Bocxlaer B, Cohen AS. 2014. Ecology and evolution of the African Great Lakes and their faunas. *Annu. Rev. Ecol. Evol. Syst.* 45:519–45

31. Fryer G, Iles TD. 1972. *The Cichlid Fishes of the Great Lakes of Africa: Their Biology and Evolution*. Edinburgh, UK: Oliver & Boyd

32. Brawand D, Wagner CE, Li YI, Malinsky M, Keller I, et al. 2014. The genomic substrate for adaptive radiation in African cichlid fish. *Nature* 513:375–81

33. Ivory SJ, Blome MW, King JW, McGlue MM, Cole JE, Cohen AS. 2016. Environmental change explains cichlid adaptive radiation at Lake Malawi over the past 1.2 million years. *PNAS* 113(42):11895–900

34. Malinsky M, Salzburger W. 2016. Environmental context for understanding the iconic adaptive radiation of cichlid fishes in Lake Malawi. *PNAS* 113(42):11654–56

35. Verheyen E. 2003. Origin of the superflock of cichlid fishes from Lake Victoria, East Africa. *Science* 300(5617):325–29

36. Kocher TD, Conroy JA, McKaye KR, Stauffer JR. 1993. Similar morphologies of cichlid fish in Lakes Tanganyika and Malawi are due to convergence. *Mol. Phylogenet. Evol.* 2(2):158–65

37. Muschick M, Indermaur A, Salzburger W. 2012. Convergent evolution within an adaptive radiation of cichlid fishes. *Curr. Biol.* 22(24):2362–68

38. Ronco F, Büscher HH, Indermaur A, Salzburger W. 2019. The taxonomic diversity of the cichlid fish fauna of ancient Lake Tanganyika, East Africa. *J. Great Lakes Res.* 46:1067–78

39. Konings A. 2007. *Malaŵi Cichlids in Their Natural Habitat*. El Paso, TX: Cichlid. 4th ed.

40. 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

41. Wagner CE, Harmon LJ, Seehausen O. 2012. Ecological opportunity and sexual selection together predict adaptive radiation. *Nature* 487(7407):366–69

42. Fan S, Elmer KR, Meyer A. 2012. Genomics of adaptation and speciation in cichlid fishes: recent advances and analyses in African and Neotropical lineages. *Philos. Trans. R. Soc. B Biol. Sci.* 367(1587):385–94

43. Musilová Z, Indermaur A, Bitja-Nyom AR, Omelchenko D, Kłodawska M, et al. 2019. Evolution of the visual sensory system in cichlid fishes from crater lake Barombi Mbo in Cameroon. *Mol. Ecol.* 28(23):5010–31

44. Malinsky M, Challis RJ, Tyers AM, Schiffels S, Terai Y, et al. 2015. Genomic islands of speciation separate cichlid ecomorphs in an East African crater lake. *Science* 350(6267):1493–98

45. Moser FN, van Rijssel JC, Mwaiko S, Meier JI, Ngatunga B, Seehausen O. 2018. The onset of ecological diversification 50 years after colonization of a crater lake by haplochromine cichlid fishes. *Proc. R. Soc. B Biol. Sci.* 285(1884):20180171

46. Coyne JA, Orr HA. 2004. *Speciation*. Sunderland, MA: Sinauer Assoc.

47. Barluenga M, Stölting KN, Salzburger W, Muschick M, Meyer A. 2006. Sympatric speciation in Nicaraguan crater lake cichlid fish. *Nature* 439(7077):719–23

48. Elmer KR, Fan S, Kusche H, Spreitzer ML, Kautt AF, et al. 2014. Parallel evolution of Nicaraguan crater lake cichlid fishes via non-parallel routes. *Nat. Commun.* 5:5168

49. Kautt AF, Machado-Schiaffino G, Meyer A. 2016. Multispecies outcomes of sympatric speciation after admixture with the source population in two radiations of Nicaraguan Crater Lake cichlids. *PLOS Genet.* 12(6):e1006157

50. López-Fernández H, Arbour JH, Winemiller KO, Honeycutt RL. 2013. Testing for ancient adaptive radiations in Neotropical cichlid fishes. *Evolution* 67(5):1321–37

51. Schwarzer J, Misof B, Ifuta SN, Schliewen UK. 2011. Time and origin of cichlid colonization of the lower Congo rapids. *PLOS ONE* 6(7):e22380

52. Burress ED, Piálek L, Casciotta JR, Almirón A, Tan M, et al. 2018. Island- and lake-like parallel adaptive radiations replicated in rivers. *Proc. R. Soc. B Biol. Sci.* 285(1870):20171762

53. Siepel A, Bejerano G, Pedersen JS, Hinrichs AS, Hou M, et al. 2005. Evolutionarily conserved elements in vertebrate, insect, worm, and yeast genomes. *Genome Res.* 15(8):1034–50

54. Lander ES. 2011. Initial impact of the sequencing of the human genome. *Nature* 470(7333):187–97

55. Loh Y-HE, Bezault E, Muenzel FM, Roberts RB, Swofford R, et al. 2013. Origins of shared genetic variation in African cichlids. *Mol. Biol. Evol.* 30(4):906–17

56. Sudmant PH, Rausch T, Gardner EJ, Handsaker RE, Abyzov A, et al. 2015. An integrated map of structural variation in 2,504 human genomes. *Nature* 526(7571):75–81

57. Mérot C, Oomen RA, Tigano A, Wellenreuther M. 2020. A roadmap for understanding the evolutionary significance of structural genomic variation. *Trends Ecol. Evol.* 35(7):561–72

58. Romiguier J, Gayral P, Ballenghien M, Bernard A, Cahais V, et al. 2014. Comparative population genomics in animals uncovers the determinants of genetic diversity. *Nature* 515(7526):261–63

59. Leffler EM, Bullaughey K, Matute DR, Meyer WK, Ségurel L, et al. 2012. Revisiting an old riddle: What determines genetic diversity levels within species? *PLOS Biol.* 10(9):e1001388

60. Ellegren H, Galtier N. 2016. Determinants of genetic diversity. *Nat. Rev. Genet.* 17(7):422–33

61. Barton NH, Keightley PD. 2002. Understanding quantitative genetic variation. *Nat. Rev. Genet.* 3:11–21

62. Barrett RDH, Schluter D. 2008. Adaptation from standing genetic variation. *Trends Ecol. Evol.* 23(1):38–44

63. Jones FC, Grabherr MG, Chan YF, Russell P, Mauceli E, et al. 2012. The genomic basis of adaptive evolution in threespine sticklebacks. *Nature* 484(7392):55–61

64. Abascal F, Corvelo A, Cruz F, Villanueva-Cañas JL, Vlasova A, et al. 2016. Extreme genomic erosion after recurrent demographic bottlenecks in the highly endangered Iberian lynx. *Genome Biol.* 17:251

65. Lamichhaney S, Berglund J, Almén MS, Maqbool K, Grabherr M, et al. 2015. Evolution of Darwin's finches and their beaks revealed by genome sequencing. *Nature* 518(7539):371–75

66. Bourgeois Y, Ruggiero RP, Manthey JD, Boissinot S. 2019. Recent secondary contacts, linked selection, and variable recombination rates shape genomic diversity in the model species *Anolis carolinensis*. *Genome Biol. Evol.* 11(7):2009–22

67. Charlesworth B. 2009. Fundamental concepts in genetics: effective population size and patterns of molecular evolution and variation. *Nat. Rev. Genet.* 10(3):195–205

68. Naciri Y, Linder HP. 2020. The genetics of evolutionary radiations. *Biol. Rev.* 95:1055–72

69. Kautt AF, Machado-Schiaffino G, Meyer A. 2018. Lessons from a natural experiment: allopatric morphological divergence and sympatric diversification in the Midas cichlid species complex are largely influenced by ecology in a deterministic way. *Evol. Lett.* 2(4):323–40

70. Schiffels S, Durbin R. 2014. Inferring human population size and separation history from multiple genome sequences. *Nat. Genet.* 46(8):919–25

71. Speidel L, Forest M, Shi S, Myers SR. 2019. A method for genome-wide genealogy estimation for thousands of samples. *Nat. Genet.* 51(9):1321–29

72. Ford AGP, Dasmahapatra KK, Rüber L, Gharbi K, Cezard T, Day JJ. 2015. High levels of interspecific gene flow in an endemic cichlid fish adaptive radiation from an extreme lake environment. *Mol. Ecol.* 24(13):3421–40

73. McGee MD, Neches RY, Seehausen O. 2016. Evaluating genomic divergence and parallelism in replicate ecomorphs from young and old cichlid adaptive radiations. *Mol. Ecol.* 25(1):260–68

74. Svardal H, Quah FX, Malinsky M, Ngatunga BP, Miska EA, et al. 2020. Ancestral hybridization facilitated species diversification in the Lake Malawi cichlid fish adaptive radiation. *Mol. Biol. Evol.* 37(4):1100–13

75. Carleton KL, Conte M, Malinsky M, Nandamuri SP, Sandkam BA, et al. 2020. Movement of transposable elements contributes to cichlid diversity. *Mol. Ecol.* 29:49

76. Alter SE, Munshi-South J, Stiassny MLJ. 2017. Genomewide SNP data reveal cryptic phylogeographic structure and microallopatric divergence in a rapids-adapted clade of cichlids from the Congo River. *Mol. Ecol.* 26(5):1401–19

77. Carvajal-Vallejos FM, Duponchelle F, Ballivian JPT, Hubert N, Rodríguez JN, et al. 2010. Population genetic structure of *Cichla pleiozona* (Perciformes: Cichlidae) in the Upper Madera basin (Bolivian Amazon): Sex-biased dispersal? *Mol. Phylogenet. Evol.* 57(3):1334–40

78. Lewontin RC. 1974. *The Genetic Basis of Evolutionary Change*. New York: Columbia Univ. Press

79. Guerrero RF, Hahn MW. 2017. Speciation as a sieve for ancestral polymorphism. *Mol. Ecol.* 26(20):5362–68

80. Paul JS, Albrechtsen A, Song YS. 2011. Genotype and SNP calling from next-generation sequencing data. *Nat. Rev. Genet.* 12(6):443–51

81. Lunter G, Goodson M. 2011. Stampy: a statistical algorithm for sensitive and fast mapping of Illumina sequence reads. *Genome Res*. 21(6):936–39

82. Moran P, Kornfield I. 1993. Retention of an ancestral polymorphism in the mbuna species flock (Teleostei: Cichlidae) of Lake Malawi. *Mol. Biol. Evol.* 10(5):1015–29

83. Scally A, Dutheil JY, Hillier LW, Jordan GE, Goodhead I, et al. 2012. Insights into hominid evolution from the gorilla genome sequence. *Nature* 483(7388):169–75

84. Meier JI, Marques DA, Wagner CE, Excoffier L, Seehausen O. 2018. Genomics of parallel ecological speciation in Lake Victoria cichlids. *Mol. Biol. Evol.* 35(6):1489–506

85. Fijarczyk A, Babik W. 2015. Detecting balancing selection in genomes: limits and prospects. *Mol. Ecol.* 24(14):3529–45

86. Savolainen O, Lascoux M, Merilä J. 2013. Ecological genomics of local adaptation. *Nat. Rev. Genet.* 14(11):807–20

87. Svardal H, Rueffler C, Hermisson J. 2015. A general condition for adaptive genetic polymorphism in temporally and spatially heterogeneous environments. *Theor. Popul. Biol.* 99:76–97

88. Carvalho CMB, Lupski JR. 2016. Mechanisms underlying structural variant formation in genomic disorders. *Nat. Rev. Genet.* 17(4):224–38

89. Brookfield JFY. 2005. The ecology of the genome—mobile DNA elements and their hosts. *Nat. Rev. Genet.* 6(2):128–36

90. Feschotte C. 2008. Transposable elements and the evolution of regulatory networks. *Nat. Rev. Genet.* 9(5):397–405

91. Fan S, Meyer A. 2014. Evolution of genomic structural variation and genomic architecture in the adaptive radiations of African cichlid fishes. *Front. Genet.* 5:163

92. Simonti CN, Pavlicev M, Capra JA. 2017. Transposable element exaptation into regulatory regions is rare, influenced by evolutionary age, and subject to pleiotropic constraints. *Mol. Biol. Evol.* 34(11):2856–69

93. Zeng L, Pederson SM, Kortschak RD, Adelson DL. 2018. Transposable elements and gene expression during the evolution of amniotes. *Mob. DNA* 9:17

94. Santos ME, Braasch I, Boileau N, Meyer BS, Sauteur L, et al. 2014. The evolution of cichlid fish egg-spots is linked with a *cis*-regulatory change. *Nat. Commun.* 5:5149

95. 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 Genom.* 18:341

96. Conte MA, Joshi R, Moore EC, Nandamuri SP, Gammerdinger WJ, et al. 2019. Chromosome-scale assemblies reveal the structural evolution of African cichlid genomes. *GigaScience* 8(4):giz030

97. Ferreira IA, Poletto AB, Kocher TD, Mota-Velasco JC, Penman DJ, Martins C. 2010. Chromosome evolution in African cichlid fish: contributions from the physical mapping of repeated DNAs. *Cytogenet. Genome Res.* 129(4):314–22

98. Clark FE, Conte MA, Ferreira-Bravo IA, Poletto AB, Martins C, Kocher TD. 2017. Dynamic sequence evolution of a sex-associated B chromosome in Lake Malawi cichlid fish. *J. Hered.* 108(1):53–62

99. Ozouf-Costaz C, Coutanceau JP, Bonillo C, Mercot H, Fermon Y, Guidi-Rotani C. 2017. New insights into the chromosomal differentiation patterns among cichlids from Africa and Madagascar. *Cybium* 41(1):35–43

100. Poletto AB, Ferreira IA, Cabral-de-Mello DC, Nakajima RT, Mazzuchelli J, et al. 2010. Chromosome differentiation patterns during cichlid fish evolution. *BMC Genet.* 11:50

101. Yoshida K, Terai Y, Mizoiri S, Aibara M, Nishihara H, et al. 2011. B chromosomes have a functional effect on female sex determination in Lake Victoria cichlid fishes. *PLOS Genet.* 7(8):e1002203

102. Gross MC, Feldberg E, Cella DM, Schneider MC, Schneider CH, et al. 2009. Intriguing evidence of translocations in Discus fish (*Symphysodon*, Cichlidae) and a report of the largest meiotic chromosomal chain observed in vertebrates. *Heredity* 102(5):435–41

103. Rhie A, McCarthy SA, Fedrigo O, Damas J, Formenti G, et al. 2020. Towards complete and error-free genome assemblies of all vertebrate species. bioRxiv. **https://doi.org/10.1101/2020.05.22.110833**

104. Mallet J, Besansky N, Hahn MW. 2016. How reticulated are species? *BioEssays* 38(2):140–49

105. Novikova PY, Hohmann N, Nizhynska V, Tsuchimatsu T, Ali J, et al. 2016. Sequencing of the genus *Arabidopsis* identifies a complex history of nonbifurcating speciation and abundant trans-specific polymorphism. *Nat. Genet.* 48(9):1077–82

106. Svardal H, Jasinska AJ, Apetrei C, Coppola G, Huang Y, et al. 2017. Ancient hybridization and strong adaptation to viruses across African vervet monkey populations. *Nat. Genet.* 49(12):1705–13

107. Stryjewski KF, Sorenson MD. 2017. Mosaic genome evolution in a recent and rapid avian radiation. *Nat. Ecol. Evol.* 1(12):1912–22

108. Edelman NB, Frandsen PB, Miyagi M, Clavijo B, Davey J, et al. 2019. Genomic architecture and introgression shape a butterfly radiation. *Science* 366(6465):594–99

109. Seehausen O. 2004. Hybridization and adaptive radiation. *Trends Ecol. Evol.* 19(4):198–207

110. Abbott R, Albach D, Ansell S, Arntzen JW, Baird SJE, et al. 2013. Hybridization and speciation. *J. Evol. Biol.* 26(2):229–46

111. Schumer M, Rosenthal GG, Andolfatto P. 2014. How common is homoploid hybrid speciation? 68(6):1553–60

112. Lamichhaney S. 2018. Rapid hybrid speciation in Darwin's finches. *Science* 359(6372):224–28

113. Marques DA, Meier JI, Seehausen O. 2019. A combinatorial view on speciation and adaptive radiation. *Trends Ecol. Evol.* 34(6):531–44

114. Stelkens RB, Schmid C, Seehausen O. 2015. Hybrid breakdown in cichlid fish. *PLOS ONE* 10(5):e0127207

115. Locke DP, Hillier LW, Warren WC, Worley KC, Nazareth LV, et al. 2011. Comparative and demographic analysis of orang-utan genomes. *Nature* 469(7331):529–33

116. Franchini P, Fruciano C, Spreitzer ML, Jones JC, Elmer KR, et al. 2014. Genomic architecture of ecologically divergent body shape in a pair of sympatric crater lake cichlid fishes. *Mol. Ecol.* 23(7):1828–45

117. Nico LG, Beamish WH, Musikasinthorn P. 2007. Discovery of the invasive Mayan Cichlid fish "*Cichlasoma*" *urophthalmus* (Günther 1862) in Thailand, with comments on other introductions and potential impacts. *Aquat. Invas.* 2(3):197–214

118. Keller I, Wagner CE, Greuter L, Mwaiko S, Selz OM, et al. 2013. Population genomic signatures of divergent adaptation, gene flow and hybrid speciation in the rapid radiation of Lake Victoria cichlid fishes. *Mol. Ecol.* 22(11):2848–63

119. Meier JI, Sousa VC, Marques DA, Selz OM, Wagner CE, et al. 2017. Demographic modelling with whole-genome data reveals parallel origin of similar *Pundamilia* cichlid species after hybridization. *Mol. Ecol.* 26(1):123–41

120. Olave M, Meyer A. 2020. Implementing large genomic single nucleotide polymorphism data sets in phylogenetic network reconstructions: a case study of particularly rapid radiations of cichlid fish. *Syst. Biol.* 69(5):848–62

121. Piálek L, Burress E, Dragová K, Almirón A, Casciotta J, Říčan O. 2019. Phylogenomics of pike cichlids (Cichlidae: *Crenicichla*) of the *C. mandelburgeri* species complex: rapid ecological speciation in the Iguazú River and high endemism in the Middle Paraná basin. *Hydrobiologia* 832:355–75

122. Richards EJ, Poelstra JW, Martin CH. 2018. Don't throw out the sympatric speciation with the crater lake water: fine-scale investigation of introgression provides equivocal support for causal role of secondary gene flow in one of the clearest examples of sympatric speciation. *Evol. Lett.* 2(5):524–40

123. Poelstra JW, Richards EJ, Martin CH. 2018. Speciation in sympatry with ongoing secondary gene flow and a potential olfactory trigger in a radiation of Cameroon cichlids. *Mol. Ecol.* 27(21):4270–88

124. Ballard JWO, Whitlock MC. 2004. The incomplete natural history of mitochondria. *Mol. Ecol.* 13(4):729–44

125. Joyce DA, Lunt DH, Genner MJ, Turner GF, Bills R, Seehausen O. 2011. Repeated colonization and hybridization in Lake Malawi cichlids. *Curr. Biol.* 21(3):R108–9

126. Genner MJ, Turner GF. 2012. Ancient hybridization and phenotypic novelty within Lake Malawi's cichlid fish radiation. *Mol. Biol. Evol.* 29(1):195–206

127. Salzburger W, Baric S, Sturmbauer C. 2002. Speciation via introgressive hybridization in East African cichlids? *Mol. Ecol.* 11(3):619–25

128. Stelkens RB, Schmid C, Selz O, Seehausen O. 2009. Phenotypic novelty in experimental hybrids is predicted by the genetic distance between species of cichlid fish. *BMC Evol. Biol.* 9:283

129. Nichols P, Genner MJ, van Oosterhout C, Smith A, Parsons P, et al. 2015. Secondary contact seeds phenotypic novelty in cichlid fishes. *Proc. R. Soc. B Biol. Sci.* 282(1798):20142272

130. Selz OM, Seehausen O. 2019. Interspecific hybridization can generate functional novelty in cichlid fish. *Proc. R. Soc. B Biol. Sci.* 286:20191621

131. Seehausen O, Terai Y, Magalhaes IS, Carleton KL, Mrosso HDJ, et al. 2008. Speciation through sensory drive in cichlid fish. *Nature* 455(7213):620–26

132. Suarez-Gonzalez A, Lexer C, Cronk QCB. 2018. Adaptive introgression: a plant perspective. *Biol. Lett.* 14(3):20170688

133. Arnold ML, Kunte K. 2017. Adaptive genetic exchange: a tangled history of admixture and evolutionary innovation. *Trends Ecol. Evol.* 32(8):601–11

134. Heliconius Genome Consort. 2012. Butterfly genome reveals promiscuous exchange of mimicry adaptations among species. *Nature* 487(7405):94–98

135. Huerta-Sánchez E, Jin X, Asan, Bianba Z, Peter BM, et al. 2014. Altitude adaptation in Tibetans caused by introgression of Denisovan-like DNA. *Nature* 512(7513):194–97

136. Meyer BS, Indermaur A, Ehrensperger X, Egger B, Banyankimbona G, et al. 2015. Back to Tanganyika: a case of recent trans-species-flock dispersal in East African haplochromine cichlid fishes. *R. Soc. Open Sci.* 2(3):140498

137. 1000 Genomes Proj. Consort., Auton A, Brooks LD, Durbin RM, Garrison EP, et al. 2015. A global reference for human genetic variation. *Nature* 526(7571):68–74

138. Bergström A, McCarthy SA, Hui R, Almarri MA, Ayub Q, et al. 2020. Insights into human genetic variation and population history from 929 diverse genomes. *Science* 367(6484):eaay5012

139. Ellegren H, Sheldon BC. 2008. Genetic basis of fitness differences in natural populations. *Nature* 452(7184):169–75

140. Dion-Côté A-M, Renaut S, Normandeau E, Bernatchez L. 2014. RNA-seq reveals transcriptomic shock involving transposable elements reactivation in hybrids of young lake whitefish species. *Mol. Biol. Evol.* 31(5):1188–99

141. Patterson N, Moorjani P, Luo Y, Mallick S, Rohland N, et al. 2012. Ancient admixture in human history. *Genetics* 192(3):1065–93

142. Zheng Y, Janke A. 2018. Gene flow analysis method, the D-statistic, is robust in a wide parameter space. *BMC Bioinform.* 19:10

143. Eriksson A, Manica A. 2012. Effect of ancient population structure on the degree of polymorphism shared between modern human populations and ancient hominins. *PNAS* 109(35):13956–60

144. Pease JB, Hahn MW. 2015. Detection and polarization of introgression in a five-taxon phylogeny. *Syst. Biol.* 64(4):651–62

145. Malinsky M, Matschiner M, Svardal H. 2020. Dsuite - fast D-statistics and related admixture evidence from VCF files. *Mol. Ecol. Res.* **https://doi.org/10.1111/1755-0998.13265**

146. Lawson DJ, Hellenthal G, Myers S, Falush D. 2012. Inference of population structure using dense haplotype data. *PLOS Genet.* 8(1):e1002453

147. Malinsky M, Trucchi E, Lawson DJ, Falush D. 2018. RADpainter and fineRADstructure: population inference from RADseq data. *Mol. Biol. Evol.* 35(5):1284–90

148. Kautt A, Kratochwil C, Nater A, Machado-Schiaffino G, Olave M, et al. 2020. Contrasting signatures of genomic divergence during sympatric speciation. *Nature* 588:106–11

## RELATED RESOURCES

BouillaBase: **http://cichlid.umd.edu/cichlidlabs/kocherlab/bouillabase.html**

Cambridge Cichlid Browser: **http://cichlid.gurdon.cam.ac.uk**

Oliver MK. 2019. *Bibliography of Lake Malawi biology (with emphasis on fish systematics, ecology, and evolution).* **https://malawicichlids.com/mw14001.htm**

# Contents

**Errata**

An online log of corrections to *Annual Review of Animal Biosciences* articles may be
found at http://www.annualreviews.org/errata/animal