

Genetic variation and hybridisation in evolutionary radiations of cichlid fishes

Hannes Svardal¹, Walter Salzburger² and Milan Malinsky²

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

²Zoological Institute, University of Basel, Basel, Switzerland; email: millanek@gmail.com (MM), walter.salzburger@unibas.ch (WS)

Running title: Genomes and hybridization in cichlids

ORCID IDs:

<https://orcid.org/0000-0001-7866-7313> (HS)

<https://orcid.org/0000-0002-9988-1674> (WS)

<https://orcid.org/0000-0002-1462-6317> (MM)

Keywords:

Genetic Variation, Biological Evolution, Molecular Evolution, Hybridization, Cichlids, Population Genetics

Abstract

Evolutionary radiations are responsible for much of the variation in biodiversity. 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 469 cichlid species across multiple radiations to synthesise information about patterns and sharing of genetic variation. Nucleotide diversity within species is low in cichlids, with 93% of surveyed species 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 hybridisation. Population genetics therefore provides a suitable conceptual framework for evolutionary genomic studies of cichlid radiations. We focus in detail on the roles of hybridisation 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 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* [Terms and Definitions 1] (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 the 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 an already large, and rapidly expanding, volume of cichlid genetic and genomic data – we found 105 studies that generated 32.6 trillion base pairs in the NCBI Sequence Read Archive at the time of writing (16). However, it is not straightforward to obtain an overview of genome diversity in cichlids by integrating across studies which (i) used different genetic and sequencing approaches, (ii) were designed to test distinct evolutionary hypotheses, and (iii) 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 synthesising 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 hybridisation 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.

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 3000-4000, which is approximately a tenth of all bony fishes (15). They are distributed across most tropical and some subtropical regions of Africa, the Americas, and on Madagascar. In addition, there are a number of isolated populations and small groups of species, such as in Southern India, southern Iran, and along the eastern coast of the Mediterranean (Figure 1).

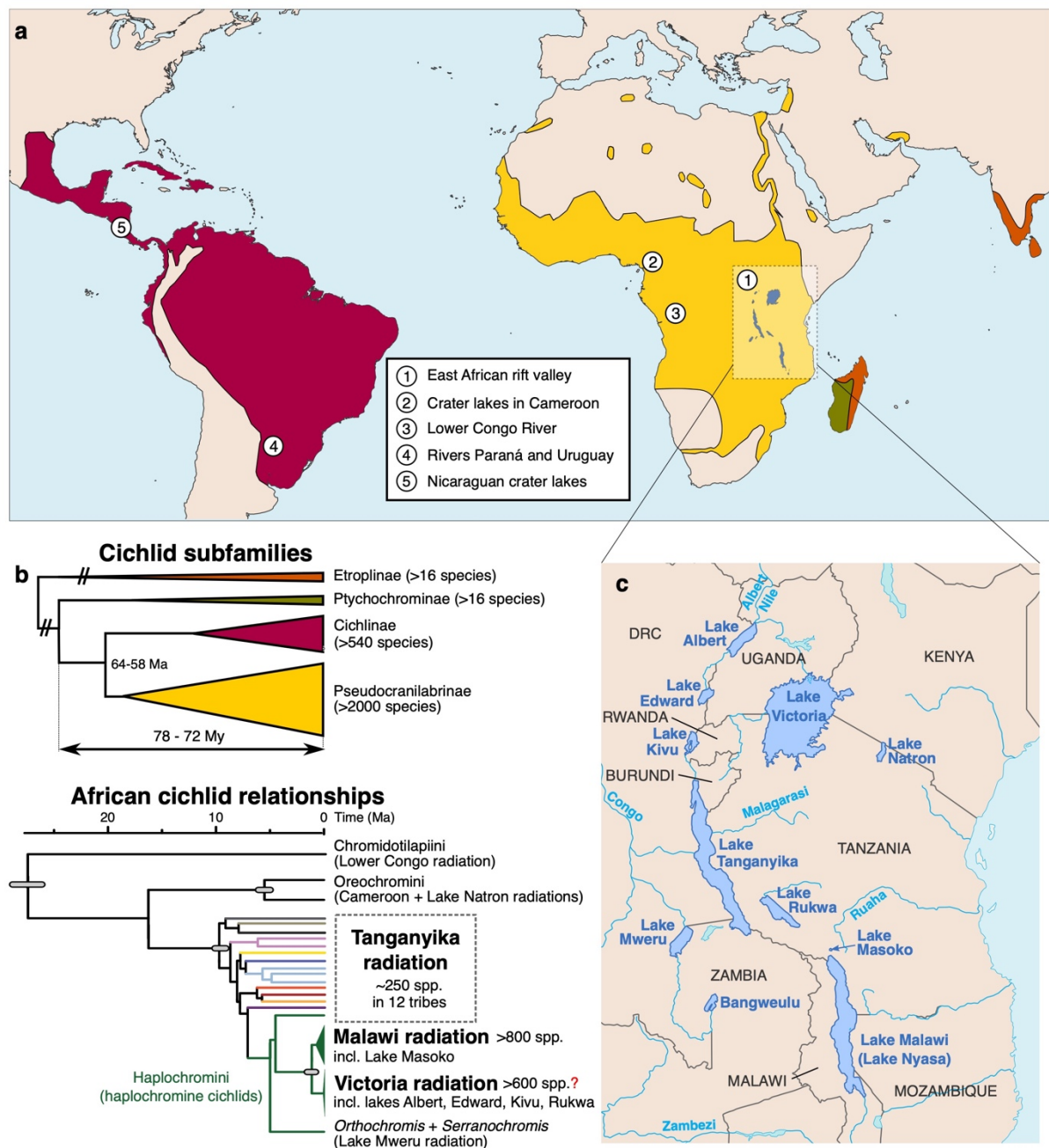


Figure 1: Cichlids and their evolutionary radiations. (a) The global distribution of cichlid fish subfamilies with colours corresponding to panel b. Adapted with permission from (25); copyright 2018, Springer Nature. (b) Relationships among cichlid subfamilies and among African cichlids whose radiations are covered by this review. Timescales are based on (26). A number of other studies have estimated divergence times among some of the same lineages with varying conclusions – this topic is covered in detail in Matschiner (2019) [ref. (25)]. (c) A map of the African Great Lake region showing selected waterbodies which are discussed in this review.

Throughout their distribution area, cichlids have colonized a wide variety of aquatic habitats ranging from brackish waters to small creeks and large rivers to small (crater) lakes to some of

the largest lakes on our planet and including 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 behaviour; as well as in pigmentation patterns (12, 27, 28).

Approximately half of extant diversity of cichlid species is the product of the spectacular adaptive radiations [[Terms and Definitions 2](#)] in three of the Great Lakes of Africa: lakes Tanganyika, Malawi, and Victoria (29). 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, 30). Even though they are sometimes considered as a single phenomenon, e.g. (10, 31), 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 ten million years old and is comprised of multiple highly diverged ($>5\text{My}$; Figure 1b) groups of cichlids that have been assigned to twelve 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 already around five million years ago (Ma), but recent paleoecological studies provide evidence that the lake was largely dry between 1.6Ma and 1Ma and stable deep lake conditions formed only about 800ka (32, 33), 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 15ky old, although its origins are intertwined with the broader Lake Victoria Region Superflock, which also includes cichlids in lakes Albert, Edward, Kivu, and several smaller waterbodies in the region, and is about 100-200ky old (17, 34).

The three massive cichlid adaptive radiations in lakes Tanganyika, Malawi and Victoria each brought forth an ecologically and morphologically highly diverse species assemblage characterized by a proportion of endemism nearing 100% (29). Many of the descendant species of independent radiations nevertheless resemble each other – sometimes even in minute details of their phenotypes – suggesting that convergent evolution due to repeated and independent adaptations to the same environments is common in cichlids (30, 35, 36). The radiations also differ in estimated species number: Lake Tanganyika contains about 240 cichlid species (37), the Lake Victoria Region Superflock about 500-700 species (17, 34), and Lake Malawi ~850 species (38) [also see (39)].

In addition to the African Great Lakes, cichlids have diversified within African lakes on at least 30 occasions (40). These smaller lakes present evolutionary radiations that are less spectacular in the number and diversity of species, but the reduced age and complexity means that they present more tractable systems for addressing a range of evolutionary hypotheses [see e.g. (41)]. 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 (42), through the formation of genetically differentiated ecomorphs within a few thousand years in Lake Masoko (43), to mature radiations comprised of hundreds of reproductively isolated species in the African Great Lakes. On the other hand, there have been at least 120 occasions where cichlids colonised African lakes without diversifying, providing opportunities for comparisons between radiating and non-radiating lineages (40).

Interestingly, cichlid diversification and speciation occurred in a number of 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” (44), and further examples of sympatric divergence are provided by the above-mentioned ecomorphs of Lake Masoko (43) and well-documented radiations of at least 13 species within a series of crater lakes in Nicaragua (45-47). 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 comprise hundreds of phenotypically and ecologically diverse species (48), 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 time scales 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 two genera originated within the rapids of the lower Congo River (49), and 17 species of the genus *Crenicichla* have evolved in parallel within the Parana and Uruguay rivers (50).

While the large cichlid radiations in the African Great Lakes have fascinated biologists for over a century, many of the cases of cichlid diversification described above have only been discovered in the last few decades. It is almost certain that there are new species and new cichlid radiations still waiting to be discovered, both within sub-Saharan Africa and in the less accessible parts of South America.

3. Genetic variation in cichlids

The approximately 1 billion bases (~1Gb) 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. While the genome as a whole is shaped by the fundamental processes of mutation, recombination, and *genetic drift* [Terms and Definitions 3] – within a framework of demographic history which may also include migration and hybridisation (Figure 2a-c) – 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 (51, 52). This is also the case for cichlids. For example, an analysis of five African cichlid genomes by Brawand et al. (31) found that protein coding genes comprised between 32 and 41Mb and conserved non-coding elements covered 72Mb, thus suggesting the proportion of conserved functional DNA in cichlids is in the range of 10% to 11%. Because most SNPs occur at noncoding and non-functioning sites, predictions derived from the *neutral theory* [Terms and Definitions 4] 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 datasets 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) (20). The rapid succession of speciation events, which is a characteristic feature

of evolutionary radiations, leads to extensive *incomplete lineage sorting* (ILS) [Terms and Definitions 5], and hybridisation provides another source of allele sharing (Figure 2b,c). Interestingly, the sharing of genetic variation (Figures 3d, 5b) appears to extend even across independent radiations separated by mountain ranges and millions of years of evolution (53).

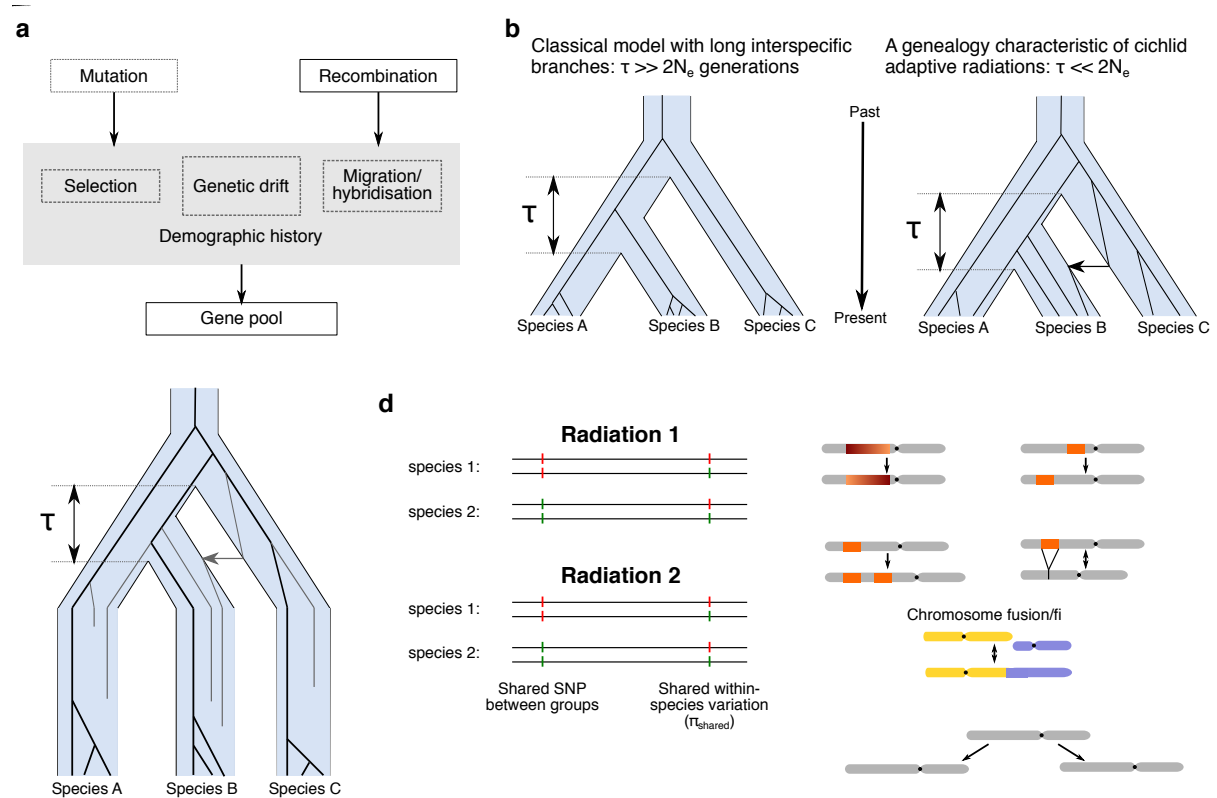


Figure 2: Genetic variation - key concepts. (a) Factors shaping the genetic composition of natural populations. (b) Incomplete lineage sorting (ILS) and hybridisation violate the ‘classical’ phylogenetic model. Genealogies of several sequences sampled from three species are shown for a single genomic locus. When the number of generations (τ) between consecutive speciation events is short relative to the effective population size (N_e), this leads to incomplete lineage sorting (ILS) whereby lineages do not coalesce (have a common ancestor) within the duration of a species. Both ILS and hybridisation are extensive within cichlid radiations and in other rapidly speciating lineages, underlining the need for population genetic approaches. (c) Incomplete lineage sorting 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 SNP shared between evolutionary radiations. Despite not being polymorphic in any species, this SNP is counted in this measure which was used in (53). Shared variation (e.g. a heterozygous site) at the same locus within multiple species and across radiations. This is the basis of the π_{shared} measure in Figure 4b. (e) Categories of structural genetic variation.

Single nucleotide polymorphisms (SNPs) are the most common type of mutations. However, DNA sequences encompass other types of variation which differ in scale, origin, and in our

ability to interrogate them. Multi-nucleotide changes and larger structural variants (Figure 2e), ranging in size from >50bp and up to many Mb, have been found to make up the majority of varying DNA bases among human genomes (54), a phenomenon now known to be taxonomically widespread (55). 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 (56-58). Crucially, genetic variation at functional loci which underlie fitness-related traits provides raw material for natural selection, facilitating adaptation to new and rapidly changing environments (59-61), while a lack of such variation restricts short-term adaptive potential.

Genetic diversity is often estimated by calculating the average fraction of nucleotide differences between any two sequences sampled from the species (or population), a measure called nucleotide diversity (π). We compiled genome-wide average measurements of π from 469 species/populations across nine cichlid evolutionary radiations (Figure 3a). A comparison with data from 61 other vertebrate species surveyed by Leffler et al. (57) revealed that nucleotide diversity levels within cichlids are relatively low (Figure 3b). The median π across cichlids included in this review was 0.09% (one pairwise single nucleotide difference every 1,110bp). The overwhelming majority (437 out of 469) of cichlid species/populations had lower π than

the median nucleotide diversity found in other vertebrates (0.23%), the only exceptions coming from the tribes Ectodini and Lamprologini from Lake Tanganyika (Figure 3a). The lowest π in cichlids approaches 0.01% (one SNP in 10,000bp) which is among the smallest values ever recorded in natural populations of any species, comparable to the highly endangered Iberian lynx or to a Neanderthal sample from the Altai mountains (62). Such extremely low diversity values are not limited to non-radiating 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, thirteen species in seven 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 π within the adaptive radiation of Darwin's finches (63) are comparable to those of cichlids. Threespine stickleback fish (61) – known for rapid adaptation from standing genetic variation – and Green anole lizards (64) have slightly higher π , but still below average when compared with the entire distribution in vertebrates. Overall, these results suggest that low genome-wide nucleotide diversity levels may not 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 (*coalescent time* [Terms and Definitions 6]) and by mutations which introduce 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, non-overlapping generations, and constant population size (so-called Wright-Fisher model), *coalescent rates* [Terms and Definitions 7] are determined by the population size N and $\pi = 4N\mu$. Real populations, including cichlids, depart from these

assumptions in many ways and to account for this we replace N by the *effective population size* [N_e ; see ref. (65)] [Terms and Definitions 8] and the equation becomes $\pi = 4N_e\mu$ (58). In either case, genetic diversity is dependent 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 μ across different cichlid species and evolutionary radiations would confirm whether this is the case.

Demographic bottlenecks, that is periods of low population size, have a disproportionately 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 (65). Harmonic mean is strongly affected by smallest values in the set. Consistent with this result, demographic bottlenecks during the settlement and subsequent diversification in newly formed lakes deplete genetic diversity in the course of evolutionary radiations [e.g. (66)]. Such bottlenecks were found for all the crater lake radiations in Nicaragua (67), and the known source populations there have substantially higher π than the radiations, although still very low in comparison to other vertebrates (Figure 3a,b). Demographic modelling and methods that estimate historical coalescence rates from whole genome SNP data [e.g. (68, 69)] 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.

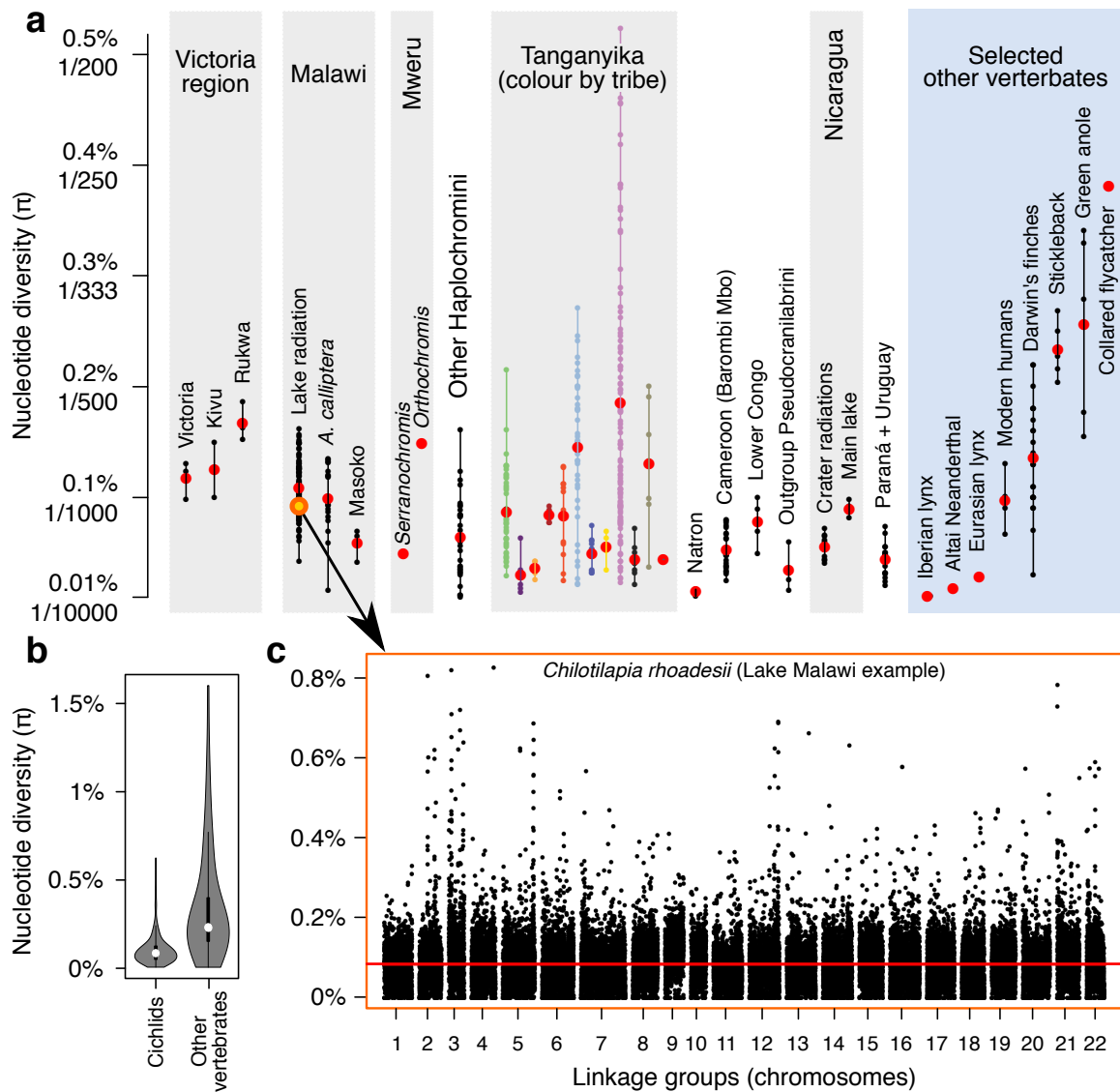


Figure 3: Nucleotide diversity (π). (a) Genome-wide averages values across cichlid radiations, in non-radiating cichlids and in selected other vertebrates. Values for other vertebrates are based on: Lynx, Neanderthal (62); Darwin's finches (63) (Waterson's θ), Green anole (64), and the remainder is from (57). Note that (57) cite a π estimate for the cichlid *Cichla pleiozona*, but the original source article (71) 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 π in cichlids and all (n=61) other vertebrate species surveyed in (57). (c) Nucleotide diversity plotted in 10kb windows along all chromosomes for the Lake Malawi species *Chilotilapia rhoadesii*. Based on data from (20).

It is possible that the founders of cichlid radiations were genetically more diverse than today's species. Both the Malawi and the Victoria cichlid radiations were founded by hybrid populations (17, 70) (see Section 4) and there is also evidence of relaxed purifying selection in ancestral haplochromine cichlids (31). On the other hand, a number of observations speak against this hypothesis. Namely (i) even the youngest radiations have low diversity [e.g. Lake Victoria cichlids have several-fold lower genetic variation than zebrafish populations (57)]; (ii) riverine cichlids, which are potential founders for lacustrine radiations (e.g. 'Other Haplochromini' in Figure 3a), do not have a higher π than radiating cichlids; and (iii) divergence among species within radiations is low (Figure 4a).

Interestingly, the range of π is much narrower than expected from differences in population size among species (57, 58). This discrepancy, known since 1970s as 'Lewontin's paradox' (72), was referred to as 'the central problem in population genetics'. The effects of *linked selection* [Terms and Definitions 9] appear to be limited and insufficient to explain the paradox by themselves (73). Intriguingly, a broad scale comparison using 31 families of animals across eight major phyla found that life history traits, especially traits that are related to parental investment, explained 73% of the variance in π (56). A speculative explanation for this surprising result is that taxa with high parental investment are capable of surviving deeper population bottlenecks, which have major long-term effects on π (73). 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 π and life history traits also holds at a smaller scale, within a single family of animals. Variation in the per-generation mutation rate μ could confound such comparison; therefore, this is one area where additional estimates of μ across different cichlid lineages would be beneficial.

There is limited information regarding the variation in diversity along cichlid chromosomes. Therefore, we reanalysed Lake Malawi whole genome SNP data from (20), revealing substantial variation in π . The genome-wide plot of π in 10kb windows for a representative species, *Chilotilapia rhoadesii*, is shown in Figure 3c. Forty-seven of 10kb genomic regions in this species reached $\pi > 0.5\%$ ($>6\times$ the average of 0.08%). Applying the per-generation SNP mutation rate estimate for Lake Malawi cichlids from (20) suggests that the average coalescence time in these regions is $\sim 700k$ generations ago ($\pi/2\mu$), which corresponds to about 2.1Ma – beyond the common ancestor of Malawi and Victoria radiations. What is surprising is not the existence of regions of such deep coalescence, but the size 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 hybridisation or introgression from divergent cichlid lineages (e.g. from *Serranochromis*), could also contribute to this large variance in π along Lake Malawi cichlid genomes, as we will discuss in more detail later when reviewing hybridisation in cichlids. Another possible contributing factor may be the presence of *balanced polymorphisms* [Terms and Definitions 10] which we discuss below. Future studies of recombination and hybridisation dynamics at fine scales along the genome will help to clarify the degree to which these factors contribute to variation in π .

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 π as the average fraction of nucleotide differences 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 (74).

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, suggesting that short reads from entire evolutionary radiations can be aligned to a single reference genome and SNP variants called (75) with reliability comparable to human studies. Nucleotide divergence tends to be greater among species in Lake Tanganyika, reflecting the greater age of this radiation. However the values are still within the range of short read mapping algorithms (76) - d_{xy} is below 1% within tribes, and at most 2.2% between tribes.

One interesting observation that comes from the comparison of d_{xy} with the π values in Figure 3a is that the mean divergence among the three Lake Victoria species is 0.03% higher than their mean π . Given the estimate of $\mu = 3.5 \times 10^{-9}$, this difference which would take ~90 thousand generation to accumulate, while Lake Victoria is only 15ky old. This shows that the genetic divergence among haplotypes in these species started before the establishment of the current lake.

While the fact that incomplete lineage sorting (ILS) leads to sharing of genetic variation in rapid cichlid radiations has been known since at least the early 1990s (77), 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 at were shared with at least one 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 $N_e = \pi/4\mu$) and the mean time to the most recent ancestor of two alleles sampled today is $2N_e$ generations, corresponding to 300ky to 780ky as the average time for two alleles to find a common ancestor, which is close to the age estimate for the entire radiation. Taking into consideration the range of π values across cichlid radiations (Figure 3a), analogous reasoning leads to the conclusion that ILS and allele sharing across species are likely to be ubiquitous.

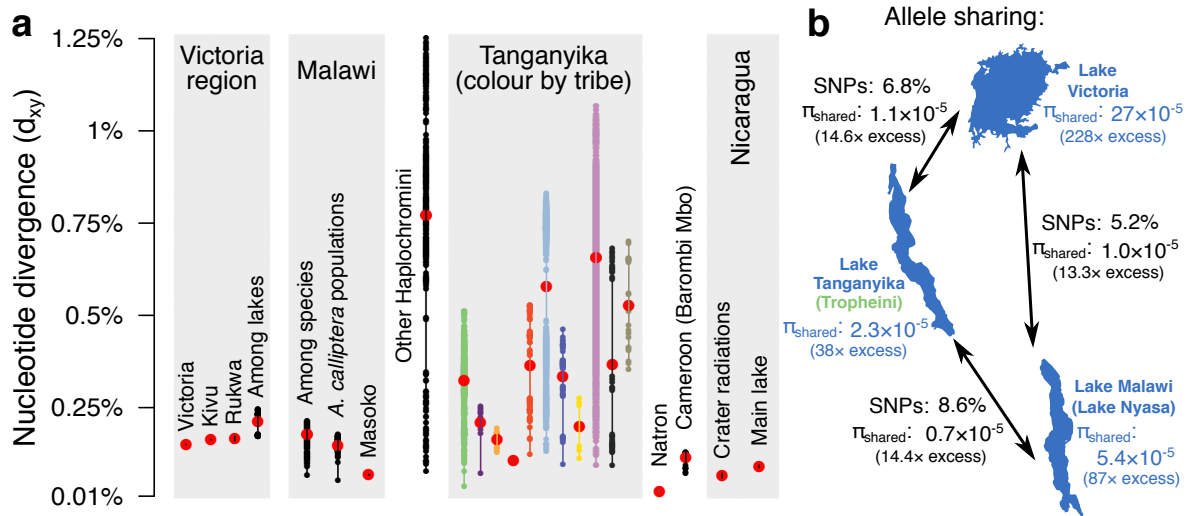


Figure 4: Divergence and sharing of variation. (a) Nucleotide divergence (d_{xy}) within cichlid radiations. (b) Sharing of genetic variation within and between the three large Great Lake cichlid radiations, using the two measures illustrated in Figure 2d – the proportion of shared SNPs and the frequency of shared heterozygous sites between individuals from different species/lakes (π_{shared}). Excess in π_{shared} is relative to random the expectation under which $\pi_{shared} = \pi_{species1} \times \pi_{species2}$.

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 with Victoria, Tanganyika being the outgroup as in Figure 1b), whereas 20.4% of genealogies had Victoria with Tanganyika as sister groups, and 22% of genealogies grouped Malawi and Tanganyika (31). This result reflects the fact that the extent of ILS among species is not determined 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 findings from human-chimpanzee-gorilla comparisons, where for 30% of the genome human and chimpanzee are not sister groups due to ILS (78).

Loh et al. (53) 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 radiation (Figure 2d) is biased by the number and relationships of samples within each group, leading to counterintuitive results such as >10% SNP polymorphisms sharing between Lake Malawi and the cichlid tribe Tilapiini which diverged more than 10Ma (53). Therefore, we introduce a measure of allele sharing that does not suffer from this bias and considers the frequency of heterozygous sites which are shared between pairs of individuals (π_{shared} in Figures 3d and 5b). We find that, on average, heterozygous sites are shared between haplochromine cichlid individuals from different African Great Lake radiations about 13 to 15 times more often than expected at random, with little difference among the different between-lake comparisons (see Figure 4b). On the other hand, the extent of within-lake allele sharing as measured by π_{shared} varies substantially. Compared with what would be expected for completely unrelated species, Lake Victoria species shared on average 228 times more heterozygous sites, Lake Malawi species had an 87× excess, and among species within the

Tanganyika Tropheini group there was only 38× excess of shared heterozygous sites, again reflecting the greater age of the Lake Tanganyika radiation.

Similar to nucleotide diversity, also patterns of genetic divergence 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%), suggesting a large within-genome variance in nucleotide diversity in the ancestor of these species (79).

The substantial variation in genetic variation (both in π and d_{xy}) along cichlid genomes is intriguing, because it raises the possibility that genetic diversity at functional loci may at least in some cases be higher than expected from the genome-wide average. Evidence supporting this hypothesis comes from a genomic comparison of a pair of cichlid ecomorphs in the crater lake Massoko (43). A genome scan based on allele frequency differences revealed 98 well demarcated regions of high differentiation (HDRs) which are likely targets of divergent selection between the ecomorphs. Interestingly, a large fraction of these HDRs also had elevated d_{xy} (median d_{xy} within HDRs = 0.08%; outside HDRs d_{xy} = 0.05%). Accumulating this extra 0.03% difference in d_{xy} by mutation would take ~92 thousand generations, which is several times more than the age of the lake. Therefore, it appears that genomic regions with elevated nucleotide diversity in the ancestral population of Lake Masoko were more likely to be under divergent selection upon the ecomorph separation.

Intriguingly, Guerrero and Hahn (74) suggest that such regions may arise due to sorting of ancestral balanced polymorphisms. Balancing selection is generally considered rare (80). However, local adaptation and spatially varying selection are both widespread and easily

capable of generating balanced polymorphisms (81, 82) suggesting 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 (80).

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 (83), 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, 84, 85). At the largest scale, chromosomal fusions and fissions lead to changes in the karyotype (55).

The first five cichlid genome assemblies by Brawand et al. (31), although highly fragmented and covering only ~70% to 80% of the true genome size, represented a landmark in cichlid genomics. Focussing 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 (Metriacrima) 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. While many of the initial results remained descriptive, some suggested functional roles. The genomes revealed hundreds of inversions with an average length of over 50kb and around ten to fifteen thousand long deletions (>50bp) with average lengths of 500-1000bp (86), and thousands of long duplicated sequences (31). Interestingly, a large excess of sequence

duplications which 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). It is possible that these events allowed for subsequent functional innovation by divergence between the sequences or expression patterns of the duplicated genes (31).

Pairwise comparisons of genomes revealed that insertions of transposable elements upstream of genes have on average a positive correlation with gene expression (31), consistent with TE activity having an effect on the evolution of gene regulation as has been shown in other vertebrates (85, 87, 88), for example by introducing a new *promoter sequences* [Terms and Definitions 11]. Specific TE-mediated insertions and deletions have been implicated in evolution of two key traits in cichlids. First, a TE insertion upstream of a pigmentation gene is associated with the formation of egg-spots, which are a key phenotypic innovation in haplochromine cichlids implicated in mating behaviour (89). The insertion functions as a specific *enhancer* [Terms and Definitions 12] in iridophores, a type of pigment cells found in egg spots. Second, a recent study of Lake Malawi cichlids (90) 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* (91) and *M. zebra* (92) are based on long read sequencing with contigs arranged into chromosome-scale assemblies using *genetic maps* [Terms and Definitions 13]. The assemblies reveal that TEs comprise between

35% and 37% of these genomes (92), twice the initial estimate of 16% to 19% from Brawand et al. (31). 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 15kb upstream of genes (1,422 vs. 338), raising the possibility that many TE-mediated regulatory changes may have contributed to diversification in Lake Malawi cichlids (92).

Genetic maps and the chromosome-scale nature of these assemblies also allowed the authors to look for large scale structural differences (92). Both assemblies have 22 *linkage groups* (LGs) [Terms and Definitions 14], and large interchromosomal rearrangements (at Mb scale) appear to be rare. However, there are a number of very large rearrangements within chromosomes, mainly inversions and translocations, of up to 23Mb in size. A comparison with *cytogenetic* evidence [Terms and Definitions 15] (93, 94) 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), while a large ~19Mb 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 (95, 96). 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 vs. absence of *B chromosomes* [Terms and Definitions 16] within the rapid radiations of Lake Victoria (97) and of Lake Malawi (94). Interestingly, B chromosomes are female-specific in at least one Victorian species and in six species of Lake Malawi, suggesting a possibility that these chromosomes may play a role in sex determination.

Even less variation in chromosome numbers was found among American cichlids (96), with 27 species across seven tribes all having $2n = 48$ chromosomes and only two species presenting an exception to this rule: *Laetacara dorsigera* ($2n = 46$) and the Discus fish of the genus *Symphysodon* ($2n = 60$), where the very large number of chromosomes may be the result of ancestral whole genome duplication or of large scale rearrangements possibly linked with ancestral hybridisation (98), and may have contributed to the unusual morphological adaptations of this disk-shaped species.

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 a high-quality cichlid genomes based on long reads continuing to fall, this situation is likely to change over the next few years. Two such new genomes, of the Central American *Archocentrus centrarchus* and of the East African *Astatotilapia calliptera* (associated with lakes Malawi and Massoko), although formally unpublished, are already available to use from public depositories (GenBank accessions: GCA_007364275.2, GCA_900246225.3). The genomes were generated as a part of a community effort to provide high quality assemblies for

10,000 vertebrate species (approximately one sixth of all) (7). More cichlids are in the pipeline, e.g. *Amphilophus citrinellus* of the Nicaraguan crater lakes.

4. Hybridization in cichlids

4.1 Evidence for hybridisation and gene flow in the lab and in the wild

One of the revelations of 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 (99-101). This is particularly true for young evolutionary radiations (63, 102, 103). The effect of this process on organismal diversification is debated (104-106), with the cessation of gene flow between lineages often seen as a prerequisite for speciation [e.g., under the biological species concept (44)]. Hybrid speciation, in which a hybridisation event directly leads to a new species that is reproductively isolated from the parental species occurred for example in the radiation of Darwin's finches (107), but may be relatively rare in animals (106). However, there is a rapidly growing body of literature suggesting that hybridisation can facilitate diversification by fuelling adaptive processes by supplying new (combinations of) functional genetic variants [reviewed in (108)].

One remarkable feature of cichlids is that they are able to hybridise 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 *intrinsic incompatibilities* [Terms and Definitions 17] are orders of magnitude slower than speciation rates (109). The latest time estimates (25, 26) for the split between Malawi and Victoria radiations are in the region of ~1.1My, translating to ~400 thousand generations, roughly

equivalent to the split between modern humans and orangutans (110). Extrapolation from these aquarium experiments suggests that hybridisation 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* sp.) have been shown to hybridize relatively easily in the laboratory (111) 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* (112)].

Indeed, genome sequencing studies found evidence for cross-species genetic exchange in the wild affecting almost all cichlid radiations that have been investigated. These efforts discovered signatures of extensive hybridisation among lineages in the evolutionary radiations of lakes Malawi (20), Victoria (113, 114), Tanganyika (21-23, 26), Mweru (18), Natron (115), Nicaraguan crater lakes Apoyo and Xiloá (116), and the tributaries of the river Paraná (117). Furthermore, there is evidence of hybridisation before or early in an adaptive radiation in the Lake Victoria Region Superflock (17), in the Lake Malawi radiation (70), 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 (47). Finally, hybridisation between the radiating lineages and riverine outgroup species was found for crater lakes Barombi Mbo in Cameroon (118, 119), and Lake Masoko (43).

Because of the extent of incomplete lineage sorting (Figure 2b,c), studies of hybridisation 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 [Sidebar: Detecting gene-flow in cichlids]. In particular, phylogenetic discordance between nuclear and mitochondrial (mtDNA) genetic markers is expected under ILS and it is not safe to

assume that mtDNA evolves as a neutral marker (120). Therefore, studies that suggested hybridisation on the basis of *mito-nuclear discordance* [Terms and Definitions 18] should be interpreted with caution and re-evaluated with genome-wide data where possible. Indeed, a number of hybridisation 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 (121), and substantial hybridisation between the deep benthic and mbuna groups (122) in Lake Malawi, were not supported by genome sequencing data (20). Similarly, mito-nuclear discordances in Iguazú and Paraná rivers (117) and among cichlids in the Congo river (123) are not reflected in genome-wide data in those studies. However, in some other cases [e.g. Lake Tanganyika *Neolamprologus* (124)], early suggestions of hybridisation were later corroborated by whole genome data (26).

4.2 The role of hybridisation in diversification

Is the prevalence of hybridisation in cichlids merely a consequence of incomplete reproductive and limited geographic isolation during rapid diversification, or does hybridisation have a positive effect on diversification rates? Experimental work has shown that inter-specific cichlid hybrids can produce extreme and novel phenotypes, with the degree of phenotypic novelty increasing with genetic divergence between the hybridising lineages (125). This phenomenon, known as transgressive segregation, has also been demonstrated in the wild, at secondary contact zones between cichlid populations (126). Moreover, experimental work suggests that the performance of transgressive hybrids, and thus the evolutionary impact of hybridisation, may be context-dependent – specifically depending on availability of ecological resources outside of the parental niches (127). The *hybrid swarm* [Terms and Definitions 19] hypothesis predicts that hybridisation at the base of an adaptive radiation, before diversification, could

promote adaptive radiation by increasing heritable genetic variation in ecologically relevant traits (104). Furthermore, it has been suggested that hybridisation among species within a radiation can generate genotype combinations that allow previously unoccupied fitness peaks to be reached (104) by combining alleles from different sources to create novel phenotypes (108). In the following we review the evidence for functional roles of the reported occurrences of hybrid swarms, of gene flow between lineages within radiations, and of hybridisation with outgroup lineages, in cichlid diversification.

4.2.1 Hybrid swarms

Evidence of hybridisation at the base of an adaptive radiation [as documented in lakes Malawi, Mweru, Tanganyika, the Victoria Region Superflock (LVRS), and in the Nicaraguan lakes Apoyo and Xiloá (17, 18, 21, 23, 47, 70)] is a fundamental prediction of the hybrid swarm hypothesis. However, further evidence is required to ascertain if an early hybridisation 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 hybridisation-derived polymorphisms were under divergent selection during speciation events early in the radiations (17, 70). 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 (128), the two alleles used in recent adaptive divergence were contributed by the ancestral hybridization event. Additional evidence for hybridisation facilitating the Lake Mweru radiation is not genetic but comes in the form of eco-morphological comparisons with the non-radiating 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 hybridisation is not particularly elevated (Figure 3a). Svandal et al. (70) showed that most of the variation derived from the ancient hybridisation is no longer segregating within Lake Malawi cichlid species – they carry only one of the two alleles at most loci. A possible explanation for this is that most hybridisation-derived variation was lost by drift or selection before the onset of adaptive radiation, and that a number of 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 (82).

4.2.2 Adaptive introgression within radiations

Despite the extensive reports of hybridisation in cichlid evolutionary radiations (18, 20-23, 26, 114-117, 129), so far there is only very limited knowledge on 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, a number of examples have recently been reported in other organisms [reviewed in (9, 130, 131)], including for example wing-pattern mimicry loci in *Heliconius* butterflies (132) and altitude adaptation alleles introgressed into modern humans from archaic hominins (133). 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. (79) 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. However, the authors were not able to distinguish between introgression and parallel selection on standing genetic variants for specific alleles involved in the species divergence.

4.2.3 Introgression with outgroups

Evidence for introgression with outgroups comes predominantly from studies of crater lake cichlid radiations, where analyses focussed 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 while estimates of admixture fraction tend to be variable – ranging from $\sim 10^{-6}$ in Barombi Mbo in Cameroon (118), 1.2% in Lake Masoko (43), 4.3% in Nicaraguan Lake Apoyo (67), 4.7% in Cameroonian Lake Ejagham (119), and up to 32% in the Nicaraguan crater lake Asososca Managua (67). 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 (119), where introgressed blocks of olfactory receptor genes may have contributed to subsequent speciation events. On the other hand, no correlation between secondary introgression and signatures of species divergence were found in lakes Barombi Mbo (118) and Masoko (43). For the Nicaraguan crater lakes this question remains unexplored.

Loh et al. (53) hypothesised 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. *Astatotilapia 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 (134)]. 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 at explaining 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, of the type provided e.g. by the 1000 Genomes (135) or Human Genome Diversity (136) projects for humans, is currently missing in cichlids. In this review we made a step towards 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 non-radiating 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 low levels of genetic variation at the SNP 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 the genetic diversity at functional loci, which underlies variation in fitness among individuals and therefore is the basis for adaptation and organismal differentiation (137). Initial evidence suggest that balanced polymorphisms may contribute to the maintenance of genetic diversity at these ‘loci of evolution’ in cichlids (74).

It is clear that sharing of genetic variation across species boundaries is ubiquitous, both due to incomplete lineage sorting and due to hybridisation. 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 (55). Initial findings suggest that gene duplications and transposable element (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. (138)], perhaps because of mismatches between the TE sequences and the host mechanisms repressing TE activity in the germ line. Therefore, hybridisation in cichlids may contribute to increased TE activity (90).

Hybridisation 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. On the other hand, specific examples of adaptive introgression within cichlid radiations are so far limited.

Future issues list

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 need to reconstruct demographic histories of cichlid evolutionary radiations at a broad taxonomic scale.
3. To make progress towards pinpointing the elements of genetic variation that are functional, we need more work on detecting signatures of selection 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. We need more long-read whole genome assemblies to discover structural variation within evolutionary radiations, especially gene duplications and transposable element insertion within gene promoters
6. To understand the feasibility of rapid adaptation without available functional standing genetic variation, we need studies assessing phenotypic plasticity and epigenetic changes and possible inheritance in cichlids

Some alternatives:

Mechanisms of maintenance of old alleles and whether these particularly contribute to adaptation.

The interplay of genome architecture (i.e. the arrangement of functional elements and especially of adaptive loci along chromosomes) and recombination in creating new or maintaining favourable combinations of alleles

To what degree do genetic variants shared among species underlie parallel and convergent evolution of cichlid phenotypes? Riverine transport of alleles?

[Sidebar: Detecting gene-flow in cichlids](#)

The majority of genomic studies of hybridisation 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) (139). These tests are robust when applied to genome-wide SNP datasets (140), 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 (141, 142), and pinpointing specific introgression events within a system of many tests in larger radiations can be challenging (20, 143). A number of studies – especially within the older Lake Tanganyika radiation (21, 22), but also the Nicaraguan crater lakes (116) – applied a phylogenomic framework under the multispecies coalescent [**Terms and Definitions**]. More recent gene-flow can be deduced from patterns of nearest-neighbour haplotype sharing (144, 145), a method used for cichlids in Lake Malawi (20), Lake Mweru (18) and in the Paraná river (117). Finally, model-based inference can be computationally demanding, but is useful for detailed investigations of specific demographic hypotheses, as illustrated for Lake Victoria *Pundamilia* (114), Nicaraguan crater lakes (67) and for the Cameroonian Lake Ejagham (119).

Terms and Definitions list:

(max. 20 terms, max. 20 words)

1. **Evolutionary radiations** – periods of significant acceleration in the rate of species diversification in particular lineages
2. **Adaptive radiations** -
3. **Genetic drift** – the fluctuation in frequency of a genetic variant due to the randomness of which individuals reproduce
4. **Neutral theory** -
5. **Incomplete lineage sorting** – a phenomenon where alleles within a species do not find a common ancestor within the duration of the species
6. **Coalescence time** – the time to the most recent common ancestor
7. **Coalescence rate** – the rate at which pairs of alleles find their common ancestor backward in time
8. **Effective population size (N_e)** – The size of an idealised Wright-Fisher population that matches coalescent rates in the real population

9. Linked selection -

10. Balanced polymorphisms –

11. Promoter sequences –

12. Enhancer –

13. Genetic map –

14. Linkage groups –

15. Cytogenetic –

16. B chromosomes –

17. Intrinsic incompatibilities –

19. Mito-nuclear discordance – conflict between phylogenies derived from mitochondrial and nuclear DNA markers.

Related Resources

Bouillabase: <http://cichlid.umd.edu/cichlidlabs/kocherlab/bouillabase.html>

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

Bibliography of Lake Malawi Biology (with emphasis on fish systematics, ecology, and evolution). Compiled by **Michael K. Oliver**, Ph.D. <https://malawicichlids.com/mw14001.htm>

Bibliography:

1. Aristotle. 1965. *History of Animals, Volume I: Books 1-3*. Cambridge, MA: Harvard University Press
2. Barton NH. 2007. *Evolution*. CSHL Press. 1 p.
3. Charlesworth B. 2010. *Elements of Evolutionary Genetics*. Roberts Publishers. 1 p.
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 of Scientists, O'Brien SJ. 2015. The Genome 10K Project: a way forward. *Annu Rev Anim Biosci.* 3(1):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. *Proc. Natl. Acad. Sci. U.S.A.* 115(17):4325–33
9. Taylor SA, Larson EL. 2019. Insights from genomes into the evolutionary importance and prevalence of hybridization in nature. *Nature Ecology & Evolution*. 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. Chance, historical contingency and ecological determinism jointly determine the rate of adaptive radiation. *Heredity*, Oct., pp. 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, International Nucleotide Sequence Database Collaboration. 2011. The sequence read archive. *Nucleic Acids Res.* 39(Database issue):D19–D21
17. Meier JJ, 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 JJ, 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(1):5391–11
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(1):96–24
20. Malinsky M, Svoldal H, Tyers AM, Miska EA, Genner MJ, et al. 2018. Whole-genome sequences of Malawi cichlids reveal multiple radiations interconnected by gene flow. *Nature Ecology & Evolution*. 457:830

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.*
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*
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(1):1105
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(1):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.** *Under review*
27. Barlow G. 2008. *The Cichlid Fishes*. Basic Books. 1 p.
28. Salzburger W. 2009. The interaction of sexually and naturally selected traits in the adaptive radiations of cichlid fishes. *Mol Ecol.* 18(2):169–85
29. 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(1):519–45
30. Fryer G, Iles TD. 1972. *The cichlid fishes of the great lakes of Africa: their biology and evolution*. Oliver and Boyd
31. 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(7518):375–81
32. 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. *Proceedings of the National Academy of Sciences.* 113(42):11895–900
33. Malinsky M, Salzburger W. 2016. Environmental context for understanding the iconic adaptive radiation of cichlid fishes in Lake Malawi. *Proceedings of the National Academy of Sciences.* 113(42):11654–56
34. Verheyen E. 2003. Origin of the Superflock of Cichlid Fishes from Lake Victoria, East Africa. *Science.* 300(5617):325–29
35. 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
36. Muschick M, Indermaur A, Salzburger W. 2012. Convergent evolution within an adaptive radiation of cichlid fishes. 22(24):2362–68
37. Ronco F, Büscher HH, Indermaur A, Salzburger W. 2019. The taxonomic diversity of the cichlid fish fauna of ancient Lake Tanganyika, East Africa. *Journal of Great Lakes Research*
38. Konings A. 2007. *Malawi Cichlids in Their Natural Habitat*. Cichlid Press. 1 p. 4 ed.
39. 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
40. Wagner CE, Harmon LJ, Seehausen O. 2012. Ecological opportunity and sexual selection together predict adaptive radiation. *Nature.* 487(7407):366–69

41. Fan S, Elmer KR, Meyer A. 2012. Genomics of adaptation and speciation in cichlid fishes: recent advances and analyses in African and Neotropical lineages. *Philosophical Transactions of the Royal Society B: Biological Sciences*. 367(1587):385–94
42. Moser FN, van Rijssel JC, Mwaiko S, Meier JJ, Ngatunga B, Seehausen O. 2018. The onset of ecological diversification 50 years after colonization of a crater lake by haplochromine cichlid fishes. *Proceedings of the Royal Society B: Biological Sciences*. 285(1884):20180171
43. 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
44. Coyne JA, Orr HA. 2004. Speciation
45. Barluenga M, Stölting KN, Salzburger W, Muschick M, Meyer A. 2006. Sympatric speciation in Nicaraguan crater lake cichlid fish. *Nature*. 439(7077):719–23
46. 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(1):5168–68
47. 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
48. López-Fernández H, Arbour JH, Winemiller KO, Honeycutt RL. 2013. Testing for ancient adaptive radiations in neotropical cichlid fishes. *Evolution*. 67(5):1321–37
49. 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
50. Burress ED, Piálek L, Casciotta JR, Almirón A, Tan M, et al. 2018. Island- and lake-like parallel adaptive radiations replicated in rivers. *Proceedings of the Royal Society B: Biological Sciences*. 285(1870):20171762
51. 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
52. Lander ES. 2011. Initial impact of the sequencing of the human genome. *Nature*. 470(7333):187–97
53. Loh Y-HE, Bezaul 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
54. 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
55. Mérot C, Oomen RA, Tigano A, Wellenreuther M. 2020. A Roadmap for Understanding the Evolutionary Significance of Structural Genomic Variation. *Trends Ecol. Evol. (Amst.)*
56. 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
57. 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
58. Ellegren H, Galtier N. 2016. Determinants of genetic diversity. *Nat. Rev. Genet*. 17(7):422–33

59. Barton NH, Keightley PD. 2002. Understanding quantitative genetic variation. *Nat. Rev. Genet.* 3(1):11–21
60. Barrett RDH, Schluter D. 2008. Adaptation from standing genetic variation. *Trends Ecol. Evol. (Amst.)*. 23(1):38–44
61. 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
62. 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(1):251–19
63. 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
64. 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 Biology and Evolution*. 11(7):2009–22
65. Charlesworth B. 2009. Fundamental concepts in genetics: Effective population size and patterns of molecular evolution and variation. *Nat. Rev. Genet.* 10(3):195–205
66. Naciri Y, Linder HP. 2020. The genetics of evolutionary radiations. *Biol Rev Camb Philos Soc.* 82(3):467
67. 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. *Evolution Letters*. 2(4):323–40
68. Schiffels S, Durbin R. 2014. Inferring human population size and separation history from multiple genome sequences. *Nat. Genet.* 46(8):919–25
69. 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
70. Svoldal 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):1100–1113
71. 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
72. Lewontin RC. 1974. *The Genetic Basis of Evolutionary Change*. 1 p.
73. Ellegren H, Ellegren N. 2016. Determinants of genetic diversity. *Nat. Rev. Genet.*, pp. 1–12
74. Guerrero RF, Hahn MW. 2017. Speciation as a sieve for ancestral polymorphism. *Mol Ecol*, Oct., pp. 5362–68
75. Paul JS, Albrechtsen A, Song YS. 2011. Genotype and SNP calling from next-generation sequencing data. *Nat. Rev. Genet.* 12(6):443–51
76. Lunter G, Goodson M. 2011. Stampy: a statistical algorithm for sensitive and fast mapping of Illumina sequence reads. *Genome Res.* 21(6):936–39
77. Moran P, Kornfield I. 1993. Retention of an ancestral polymorphism in the Mbuna species flock (Teleostei: Cichlidae) of Lake Malawi. *Mol. Biol. Evol.*
78. 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

79. Meier JJ, Marques DA, Wagner CE, Excoffier L, Seehausen O. 2018. Genomics of Parallel Ecological Speciation in Lake Victoria Cichlids. *Mol. Biol. Evol.* 35(6):1489–1506
80. Fijarczyk A, Babik W. 2015. Detecting balancing selection in genomes: limits and prospects. *Mol Ecol.* 24(14):3529–45
81. Savolainen O, Lascoux M, Merilä J. 2013. Ecological genomics of local adaptation. *Nat. Rev. Genet.* 14(11):807–20
82. 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
83. Carvalho CMB, Lupski JR. 2016. Mechanisms underlying structural variant formation in genomic disorders. *Nat. Rev. Genet.* 17(4):224–38
84. Brookfield JFY. 2005. The ecology of the genome - mobile DNA elements and their hosts. *Nat. Rev. Genet.* 6(2):128–36
85. Feschotte C. 2008. Transposable elements and the evolution of regulatory networks. *Nat. Rev. Genet.* 9(5):397–405
86. Fan S, Meyer A. 2014. Evolution of genomic structural variation and genomic architecture in the adaptive radiations of African cichlid fishes. *Front. Genet.* 5:
87. 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
88. Zeng L, Pederson SM, Kortschak RD, Adelson DL. 2018. Transposable elements and gene expression during the evolution of amniotes. *Mob DNA.* 9(1):17–19
89. 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
90. Carleton KL, Conte M, Malinsky M, Nandamuri SP, Ben Sandkam, et al. 2020. Movement of transposable elements contributes to cichlid diversity. *bioRxiv.* 4(1):2020.02.26.961987
91. Conte MA, Gammerdinger WJ, Bartie KL, Penman DJ, Kocher TD. 2017. A high quality assembly of the Nile Tilapia (*Oreochromis niloticus*) genome reveals the structure of two sex determination regions. *BMC Genomics.* 18(1):341–19
92. 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):288
93. 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
94. 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
95. Ozouf-Costaz C, Coutanceau JP, J CBCI, 2017. New insights into the chromosomal differentiation patterns among cichlids from Africa and Madagascar. *researchgate.net*
96. 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

97. 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
98. 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
99. Mallet J, Besansky N, Hahn MW. 2016. How reticulated are species? *Bioessays.* 38(2):140–49
100. 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
101. Svoldal 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
102. Stryjewski KF, Sorenson MD. 2017. Mosaic genome evolution in a recent and rapid avian radiation. *Nature Ecology & Evolution.* 1(12):1912–22
103. 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
104. Seehausen O. 2004. Hybridization and adaptive radiation. *Trends Ecol. Evol. (Amst.).* 19(4):198–207
105. Abbott R, Albach D, Ansell S, Arntzen JW, Baird SJE, et al. 2013. Hybridization and speciation. *J. Evol. Biol.* 26(2):229–46
106. Schumer M, Rosenthal GG, Andolfatto P. 2014. How common is homoploid hybrid speciation? 68(6):1553–60
107. Lamichhaney S. Rapid hybrid speciation in Darwin's finches. *Science*, p. eaao4593
108. Marques DA, Meier JJ, Seehausen O. 2019. A Combinatorial View on Speciation and Adaptive Radiation. *Trends Ecol. Evol. (Amst.).* 34(6):531–44
109. Stelkens RB, Schmid C, Seehausen O. 2015. Hybrid breakdown in cichlid fish. *PLoS ONE.* 10(5):e0127207
110. 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
111. 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
112. Nico LG, Beamish WH, Invasions PMA, 2007. Discovery of the invasive Mayan Cichlid fish “*Cichlasoma*” urophthalmus (Günther 1862) in Thailand, with comments on other introductions and potential *ciclidox-mexico.com*
113. 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
114. Meier JJ, 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

115. 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
116. Olave M, Biology AMS, 2020. Implementing Large Genomic Single Nucleotide Polymorphism Data Sets in Phylogenetic Network Reconstructions: A Case Study of Particularly Rapid Radiations of *academic.oup.com*
117. Piálek L, Burrell E, Dragová K, Almirón A, Casciotta J, Řičan 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(1):355–75
118. 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. *Evolution Letters.* 2(5):524–40
119. 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
120. Ballard JWO, Whitlock MC. 2004. The incomplete natural history of mitochondria. *Mol Ecol.* 13(4):729–44
121. Joyce DA, Lunt DH, Genner MJ, Turner GF, Bills R, Seehausen O. 2011. Repeated colonization and hybridization in Lake Malawi cichlids. 21(3):R108–9
122. Genner MJ, Turner GF. 2012. Ancient hybridization and phenotypic novelty within Lake Malawi's cichlid fish radiation. *Mol. Biol. Evol.* 29(1):195–206
123. 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
124. Salzburger W, Baric S, Sturmbauer C. 2002. Speciation via introgressive hybridization in East African cichlids? *Mol Ecol.* 11(3):619–25
125. 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(1):283–13
126. Nichols P, Genner MJ, van Oosterhout C, Smith A, Parsons P, et al. 2015. Secondary contact seeds phenotypic novelty in cichlid fishes. *Proceedings of the Royal Society B: Biological Sciences.* 282(1798):20142272
127. Selz OM, Seehausen O. 2019. Interspecific hybridization can generate functional novelty in cichlid fish. *Proceedings of the Royal Society B: Biological Sciences.* 286(1913):20191621
128. 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
129. 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
130. Suarez-Gonzalez A, Lexer C, Cronk QCB. 2018. Adaptive introgression: a plant perspective. *Biol. Lett.* 14(3):20170688
131. Arnold ML, Kunte K. 2017. Adaptive Genetic Exchange: A Tangled History of Admixture and Evolutionary Innovation. *Trends Ecol. Evol. (Amst.).* 32(8):601–11

132. Heliconius Genome Consortium. 2012. Butterfly genome reveals promiscuous exchange of mimicry adaptations among species. *Nature*. 487(7405):94–98
133. Huerta-Sanchez 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
134. 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
135. 1000 Genomes Project Consortium, Auton A, Brooks LD, Durbin RM, Garrison EP, et al. 2015. A global reference for human genetic variation. *Nature*. 526(7571):68–74
136. 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
137. Ellegren H, Sheldon BC. 2008. Genetic basis of fitness differences in natural populations. *Nature*. 452(7184):169–75
138. 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
139. Patterson N, Moorjani P, Luo Y, Mallick S, Rohland N, et al. 2012. Ancient admixture in human history. *Genetics*. 192(3):1065–93
140. Zheng Y, Janke A. 2018. Gene flow analysis method, the D-statistic, is robust in a wide parameter space. *BMC Bioinformatics*. 19(1):10–19
141. Eriksson A, Manica A. 2012. Effect of ancient population structure on the degree of polymorphism shared between modern human populations and ancient hominins. *Proc. Natl. Acad. Sci. U.S.A.* 109(35):13956–60
142. Pease JB, Hahn MW. 2015. Detection and polarization of Introgression in a five-taxon phylogeny. *Syst. Biol.* 64(4):651–62
143. Malinsky M, Matschiner M, Svardal H. 2020. Dsuite - fast D-statistics and related admixture evidence from VCF files. *bioRxiv*. 9:634477
144. Lawson DJ, Hellenthal G, Myers S, Falush D. 2012. Inference of Population Structure using Dense Haplotype Data. *PLoS Genet.* 8(1):e1002453
145. Malinsky M, Trucchi E, Lawson DJ, Falush D. 2018. RADpainter and fineRADstructure: Population Inference from RADseq Data. *Mol. Biol. Evol.* 35(5):1284–90