RESEARCH ARTICLE

Large inversions in Lake Malawi cichlids are associated with habitat, lineage, and sex determination

Nikesh M Kumar[1], Taylor L Cooper[1], Thomas D Kocher[2], Jeffrey Todd Streelman[1]*, Patrick T McGrath[1,3]*

[1]School of Biological Sciences, Georgia Institute of Technology, Atlanta, United States; [2]Department of Biology, University of Maryland, College Park, United States; [3]School of Physics, Georgia Institute of Technology, Atlanta, United States

eLife Assessment

Using several hundreds of samples and cutting-edge genomic methods, including BioNano, PacBio, HiFi, and advanced bioinformatic pipelines, the authors identify six large chromosomal inversions segregating in over 100 species of Lake Malawi cichlids. This **important** study provides **compelling** evidence for the presence of these six inversions, their differential distribution among populations, and the association of chromosome 10 inversion with a sex-determination locus. This work also provides a starting point for further investigating the role of these inversions with respect to local adaptation, speciation, sex determination, hybridization, and ILS in cichlids, which represent ~5% of the extant vertebrate species and are one of the most prominent examples of adaptive radiations.

***For correspondence:** todd.streelman@biology.gatech.edu (JTS); patrick.mcgrath@biology.gatech.edu (PTM)

Competing interest: The authors declare that no competing interests exist.

Funding: See page 20

Preprint posted 29 October 2024
Sent for Review 05 November 2024
Reviewed preprint v1 31 January 2025
Reviewed preprint v2 21 May 2025
Version of Record 06 November 2026

Reviewing Editor: Virginie Courtier-Orgogozo, CNRS - Universite Paris Cite, France

© Copyright Kumar *et al*. This article is distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use and redistribution provided that the original author and source are credited.

Abstract Chromosomal inversions are an important class of genetic variation that link multiple alleles together into a single inherited block that can have important effects on fitness. To study the role of large inversions in the massive evolutionary radiation of Lake Malawi cichlids, we used long-read technologies to identify four single and two tandem inversions that span half of each respective chromosome, and which together encompass over 10% of the genome. Each inversion is fixed in one of the two states within the seven major ecogroups, suggesting they played a role in the separation of the major lake lineages into specific lake habitats. One exception is within the *benthic* sub-radiation, where both inverted and non-inverted alleles continue to segregate within the group. The evolutionary histories of three of the six inversions suggest they transferred from the pelagic *Diplotaxodon* group into *benthic* ancestors at the time the *benthic* sub-radiation was seeded. The remaining three inversions are found in a subset of *benthic* species living in deep waters. We show that some of these inversions are linked to XY sex-determination systems but are also likely limited to a subset of total lake species. Our work suggests that inversions have been under both sexual and natural selection in Lake Malawi cichlids and that they will be important to understanding how this adaptive radiation evolved.

Introduction

Large genomic inversions are a particularly interesting type of genetic variation in the evolutionary process (***Kirkpatrick, 2010***; ***Wellenreuther and Bernatchez, 2018***; ***Berdan et al., 2023***; ***Hoffmann and Rieseberg, 2008***). These rearrangements of DNA sequence strongly suppress recombination between inverted and non-inverted alleles, enabling the capture and accumulation of genetic variants of smaller molecular size that are linked together on each inversion haplotype (***Andolfatto et al.,***

2001). These alternative haplotypes largely follow independent evolutionary paths within a species, accumulating new mutations that can affect phenotypes under natural and sexual selection. Structural rearrangements can be large, spanning multiple megabases of DNA containing hundreds of genes (*Wellenreuther and Bernatchez, 2018*; *Harringmeyer and Hoekstra, 2022*). They have been shown to play a role in adaptive divergence within a variety of species (*Huang et al., 2020*; *Hager et al., 2022*; *Joron et al., 2011*; *Joron et al., 2006*; *Palmer and Kronforst, 2020*; *Jones et al., 2012*), in creating alternative reproductive strategies and mating types (*Lamichhaney et al., 2016*; *Küpper et al., 2016*; *Wang et al., 2013*; *Tuttle et al., 2016*), in evolution of sex chromosomes (*Kirkpatrick, 2010*; *Hughes et al., 2010*), and in the formation of pre- and post-zygotic barriers between incipient species (*Fishman et al., 2013*; *Noor et al., 2001*).

When comparing the genomes of extant species, chromosomal inversions are often found as fixed differences, consistent with a role in speciation (*Fuller et al., 2019*; *Fuller et al., 2018*; *Kirkpatrick and Barton, 2006*). Inversions have been shown to play a role in adaptation to different environments and the establishment of prezygotic and postzygotic barriers (*Fuller et al., 2018*; *Thompson and Jiggins, 2014*). However, the evolutionary forces responsible for establishing and spreading inversions remain obscure. Natural selection and sexual selection are thought to play a role but neutral forces like genetic drift can also be responsible for their spread. Perhaps most likely, all of these different evolutionary forces contribute to their fixation (*Berdan et al., 2023*; *Faria et al., 2019*).

Few vertebrates offer as much potential for investigating evolution and speciation as the *Cichlidae* family of fish, one of the most species-rich and diverse families of vertebrates, with an estimated 2000–3000 species across the globe (*Kocher, 2004*; *Santos et al., 2023*; *Svardal et al., 2021*). Evolutionary radiations have occurred at least three times in the African Great Lakes (Lake Malawi, Lake Tanganyika, and Lake Victoria). In Lake Malawi, more than 800 species of cichlids evolved in the past 1.2 million years (*Santos et al., 2023*; *Svardal et al., 2021*; *Johnson et al., 2023*). Speciation in cichlid fishes involved extremely high levels of phenotypic divergence, including changes in body shape, jaw structures, and feeding behaviors related to prey in their ecological niches (*Kocher, 2004*; *Streelman and Danley, 2003*). Additionally, traits that create prezygotic barriers between species are also extremely diverse. Color patterns are quite dramatic and diverse among species and include both differences in color and species-specific patterns, including bars, stripes, and blotches that are thought to reinforce barriers between species in sympatry (*Maan and Sefc, 2013*). In some species, mating is seasonal, with gatherings on 'leks' where males build bowers by scooping and spitting out sand over the course of a week or more (*McKaye et al., 2001*). Females choose mates based upon features of these bowers, including shape and size (*Martin and Genner, 2009*). Most of the Lake Malawi species can be induced to interbreed in the lab, providing the opportunity to use genetic approaches to study causality of associated changes. Despite the presence of these prezygotic barriers, substantial gene flow has occurred multiple times within the lake (*Malinsky et al., 2018*; *Loh et al., 2013*).

Our current understanding of how the Lake Malawi radiation created >800 species is a set of three serial diversifications, starting from a single riverine-like ancestor lineage, that created seven major ecogroups that currently live in the lake (*Malinsky et al., 2018*; *Figure 1*). Each ecogroup forms its own clade in a whole-lake phylogeny and occupies distinct habitats within Lake Malawi. A pelagic lineage separated first, further diversifying into a deep-water ecogroup (*Diplotaxodon*) and a mid-water ecogroup (*Rhamphochromis*). A muddy/sandy benthic lineage evolved from the ancestral riverine lineage next, which split into three ecogroups (collectively referred to as *benthics*) living either in the water column (*utaka*), over shallow sandy/muddy shores (*shallow benthics*), or in deep-water habitats (*deep benthics*). Finally, an ecogroup living over rocky habitats evolved (*mbuna*). The riverine generalist *Astatotilapia calliptera* (*AC*) living in the border regions of the lake and surrounding rivers is the seventh ecogroup and is thought to represent the ancestral lineage that seeded the three primary Lake Malawi lineages. Subsequent radiations within these ecogroups based on trophic specializations and sexual selection further diversified the species flock (*Malinsky et al., 2018*; *Streelman and Danley, 2003*).

Here, we investigate the role large inversions played in the Lake Malawi radiation using three technologies. Using the Bionano Saphyr optical mapping system, we identify four single large inversions and two double inversions, ranging from 9.9 to 22.9Mb in size. A fifth single inversion, composed of just one of the two tandem inversions on chromosome 20, was also identified, indicating a serial set of rearrangements created this structural variant. We validate these inversions using new genome

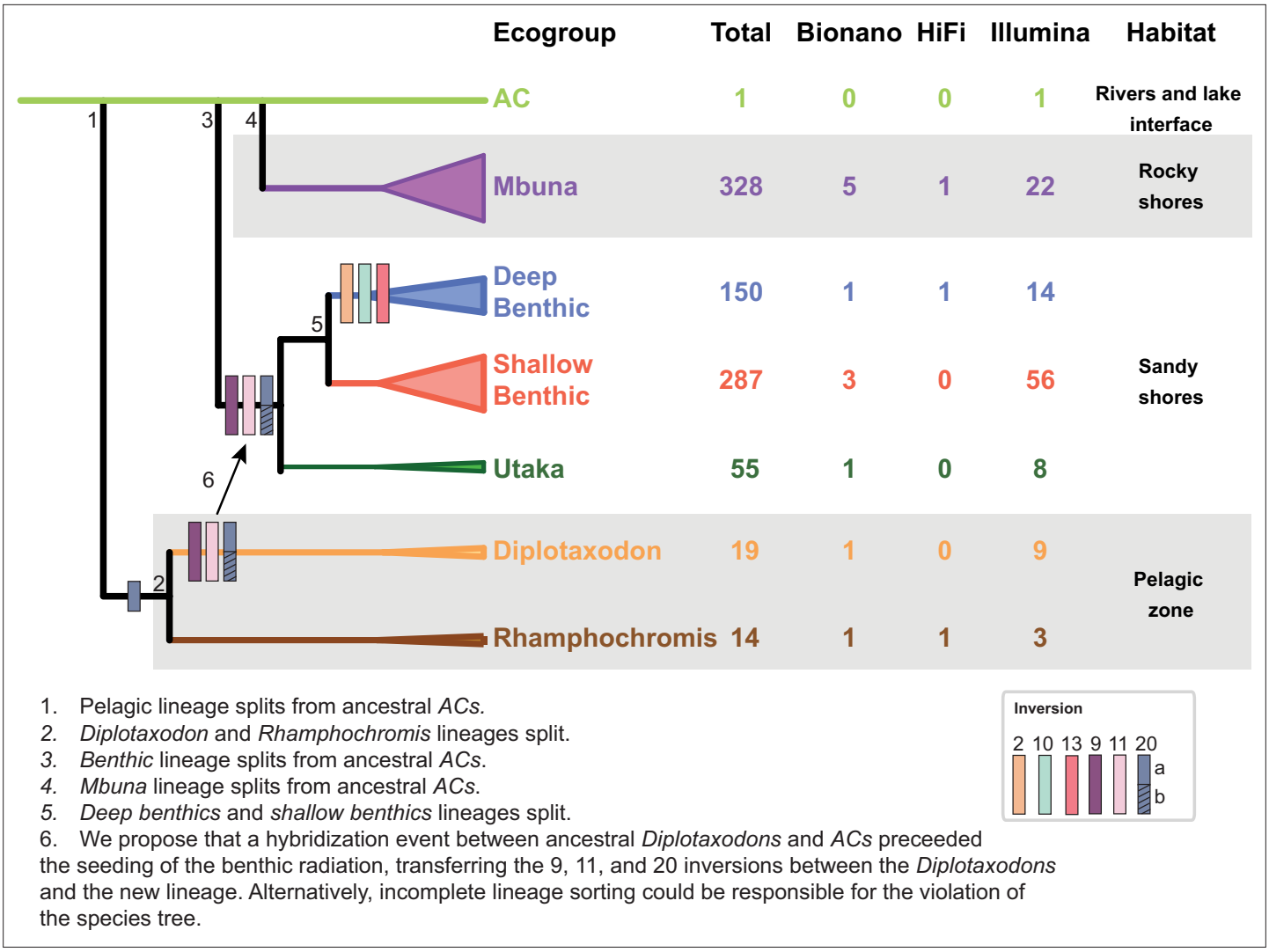


Figure 1. Overview of Lake Malawi cichlids. A summary of the Lake Malawi cichlids and the results from this paper. On the left, a phylogeny showing the seven major lineages that currently live in the lake which are referred to as ecogroups. AC is *Astatotilapia calliptera*, a riverine species that lives in the margins of the river/lake interface. For each ecogroup, we summarize the (1) estimated number of species in the lake, (2) the number of species we have Bionano optical read data for, (3) the number of species we have PacBio based genome assemblies, (4) the number of species with Illumina short-read sequencing data, and (5) the primary habitat these ecogroups live in. Deep and shallow benthics are distinguished by the lake bottom depths where they live while utakas live in the water column but mate over the shore. We also summarize the distribution of six large inversions we have identified, colored by the chromosome they are found on (legend). We propose that the inversions on 9, 11, and 20 originated in the pelagic lineage and spread to the benthics by hybridization very early on in the evolution of the benthic lineage. Alternatively, these inversions distribution could be explained by incomplete lineage sorting.

assemblies created using PacBio HiFi reads. Finally, we use existing and newly generated short-read Illumina sequencing to genotype these inversions in DNA samples from cichlids collected from Lake Malawi. These inversions primarily segregate by ecogroup, with no inversions found in the *mbuna* and *AC* samples, one inversion fixed in *Rhamphochromis*, three inversions fixed in *Diplotaxodon*, and all six present within the *benthic* sub-radiation, five of which remain polymorphic. The evolutionary histories of these inversions are inconsistent with the species phylogeny and suggest that the three *Diplotaxodon* inversions spread into the *benthics* via hybridization around the time the benthic lineage diverged from the ancestral riverine species. The three additional inversions are found primarily in *benthic* species living in deep water habitats. We provide evidence that two of the inversions are linked to a sex determination locus in a subset of *benthic* species. Our work provides a framework for understanding the role of inversions in the adaptive radiation of Lake Malawi cichlids and supports

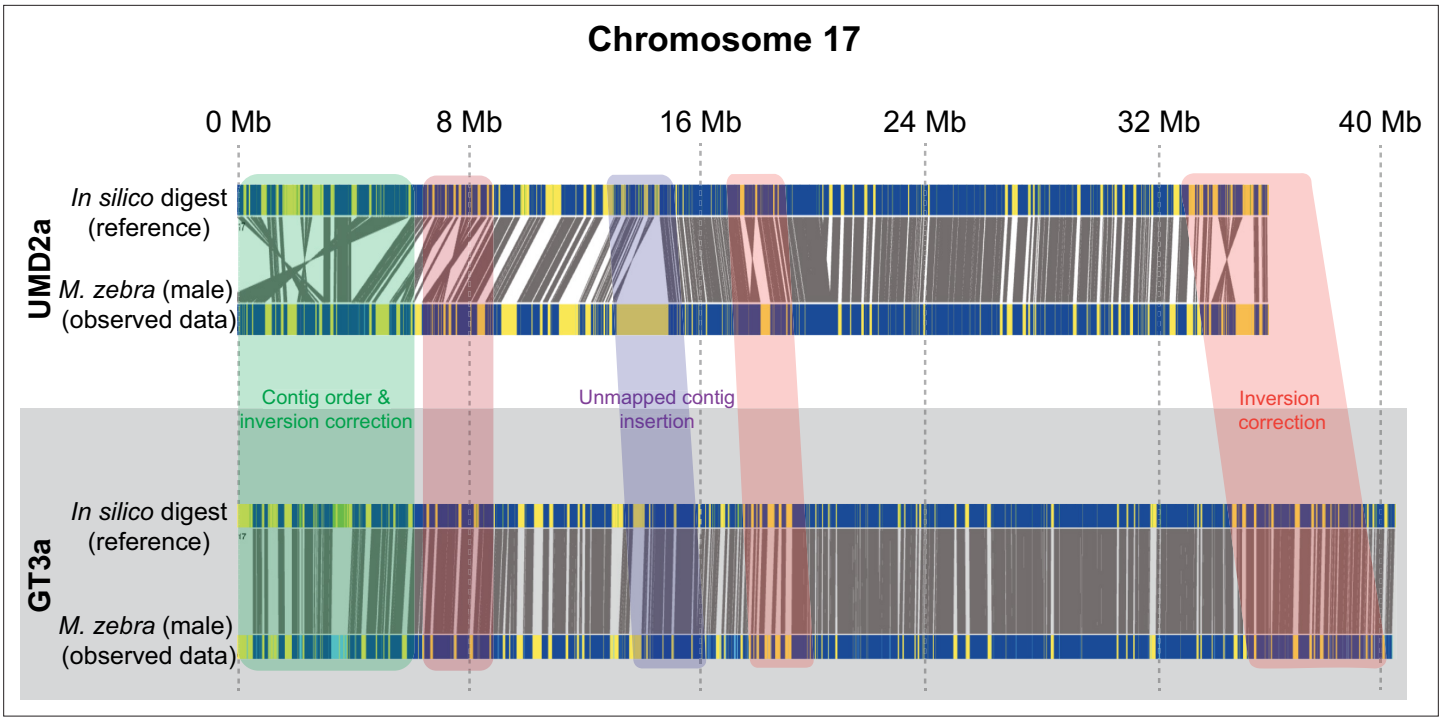


Figure 2. Generation of a new high-quality reference genome for *Metriaclima zebra*. Alignment of Bionano maps derived from a single *Metriaclima zebra* male to two different reference sequences for Chromosome 17 (white and gray, respectively). M_zebra_UMD2a reference (white) is the existing reference genome, and M_zebra_GT3a reference (gray) was created for this manuscript using a combination of PacBio HiFi reads and Bionano molecules. For each reference, the top bar shows the expected location of each CTTAAG motif (blue) created from an in-silico annotation of the reference sequence. These are aligned (gray lines) to observed locations in maps generated from Bionano Saphyr optical molecules (bottom bar). Deviations from co-linearity indicate reference errors (green, blue, and red highlights). Yellow indicates regions of the genome that lack aligned motifs. A summary of the improvements in the new genome is found in ***Supplementary file 1***.

future investigations into their roles in the establishment of ecogroups, adaptation to deep water habitats, and in controlling traits under sexual selection.

Results

Improvement of the *Metriaclima zebra* reference

The *mbuna* cichlid, *Metriaclima zebra*, was the first species sequenced as a reference genome for Lake Malawi cichlids (***Brawand et al., 2014***). The current version of its genome, M_zebra_UMD2a, has been created by assembling PacBio reads into contigs which were anchored to 22 chromosomes using recombinant maps generated using crosses between different Lake Malawi species (***Conte et al., 2019***). While the overall quality of this reference is high compared to other non-model organisms, a total of 20.5% of the DNA was not anchored to a chromosome. Additionally, the orientation of contigs can be ambiguous given how contigs were anchored to linkage groups using recombinant maps. These imperfections could complicate inversion discovery, so we first sought to improve the reference sequence.

The Bionano Saphyr system generates large (150 kb - multiple Mbs) DNA molecules and labels a specific motif (CTTAAG) with a fluorescent probe (***Bocklandt et al., 2019***). While these molecules do not provide sequencing information, the distance between observed fluorescent loci can be used to assemble multiple molecules into large maps, with lengths spanning up to an entire chromosome. Comparison of these maps with an in-silico digest of the reference sequence identifies large structural changes in the observed sample or errors in the reference genome. We processed DNA collected from the blood of three *Metriaclima zebra* individuals using the Bionano pipeline and assembled DNA molecules into optical maps. We identified many discrepancies between the observed optical maps and the M_zebra_UMD2a assembly (***Figure 2***). These differences fall into two primary categories: (1)

Table 1. Inversion genotypes of 31 samples based on Bionano data.

			Inversion						
Species	Ecogroup	# male/female	2	9	10	11	13	20	Notes
Cynotilapia zebroides 'Cobue'	*Mbuna*	1 m/1 f	-	-	-	-	-	-	
Labeotropheus fuelleborni	*Mbuna*	1 m/1 f	-	-	-	-	-	-	
Labeotropheus trewavasae	*Mbuna*	1 m/1 f	-	-	-	-	-	-	
Metriaclima zebra	*Mbuna*	3 m/1 f	-	-	-	-	-	-	Reference species
Pseudotropheus demasoni	*Mbuna*	1 m/1 f	-	-	-	-	-	-	
Copadichromis virginalis	*Utaka*	1 m/2 f	-	X	-	-*	-	X	* One female was het on 11
Protomelas taeniolatus	*Shallow benthic*	1m	-	X*	-	-	-	X	* The male was het on 9
Mchenga conophoros	*Shallow benthic*	1 m/2 f	-	X	-	X	-	X	
Nyassachromis prostoma 'Orange Cap'	*Shallow benthic*	1 m/1 f	-	X	-	X*	-	X	* The male was het on 11
Aulonocara sp. 'chitande type north' *Nkhata* Bay	*Deep benthic*	3 m/1 f	X	X	X*	X	X	X	* All three males were het on 10
Diplotaxodon limnothrissa	*Diplotaxodon*	2 m/0 f	-	X	-	X	-	X	
Rhamphochromis longiceps	*Rhamphochromis*	3 m/1 f	-	-	-	-	-	X*	* All samples had the single left inversion on 20

contigs that were incorrectly ordered or oriented and (2) unplaced contigs that Bionano anchored to specific chromosomes. The first category is consistent with errors in the assembly caused by ambiguity in placing contigs into a linkage group using recombination maps. Our results indicate that 178 of the contigs are incorrectly oriented and 133 of the unanchored contigs could be placed onto specific chromosomes.

To further improve the reference, we used PacBio HiFi reads generated from an additional male *Metriaclima zebra* individual to create a new genome assembly, taking advantage of the low overall error rate (<0.1%) of HiFi reads compared to previous generation PacBio technologies. Contigs generated from these data were combined with Bionano maps to generate a new hybrid genome assembly, which we refer to as the M_zebra_GT3a assembly. The overall quality of the assembly was excellent, with 933 Mb of 962 Mb of DNA assigned to the 22 linkage groups, (***Supplementary file 1***) reducing the amount of unanchored genetic information from 196 to 28 Mb. The N50 length was 32.542 Mb, with many of the scaffolds containing entire chromosomes.

Identification of six large inversions from 11 species of Lake Malawi cichlids

Armed with the improved reference, we used the Bionano system to test 31 samples from 11 different species found in Lake Malawi. Eight of the species were bred in the Streelman or McGrath laboratories (***Table 1*** and ***Supplementary file 2***). These included four *mbuna* species (*Pseudotropheus demasoni*, *Cynotilapia zebroides*, *Labeotropheus fuelleborni, and Labeotropheus trewavasae),* two *shallow benthic* species (*Mchenga conophoros* and *Nyassachromis prostoma* 'Orange Cap'), a single *deep benthic* species (*Aulonocara sp. 'chitande type north' Nkhata Bay*), and a single *utaka* species (*Copadichromis virginalis*). To expand our coverage of the Lake Malawi radiation, we also obtained samples from the pet trade for a single species of the *Diplotaxodon* ecogroup, *Diplotaxodon limnothrissa,* and a single species of the *Rhamphochromis* ecogroup, *Rhamphochromis longiceps.* Finally, we also tested a single male *Protomelas taeniolatus shallow benthic* individual collected from Lake Malawi based on its position in principal component analysis of single nucleotide variants (SNVs) described below. In total, we were able to obtain samples from 6 of the 7 ecogroups, excluding *Astatotilapia calliptera*.

From these 31 individuals, we identified six new inversions ranging from 9.9 to 22.9 Mb in size (***Figure 3***): single inversions on chromosomes 2, 9, 10, and 13, and double inversions on chromosomes 11 and 20. Interestingly, regarding the double inversion on chromosome 20, we also identified four

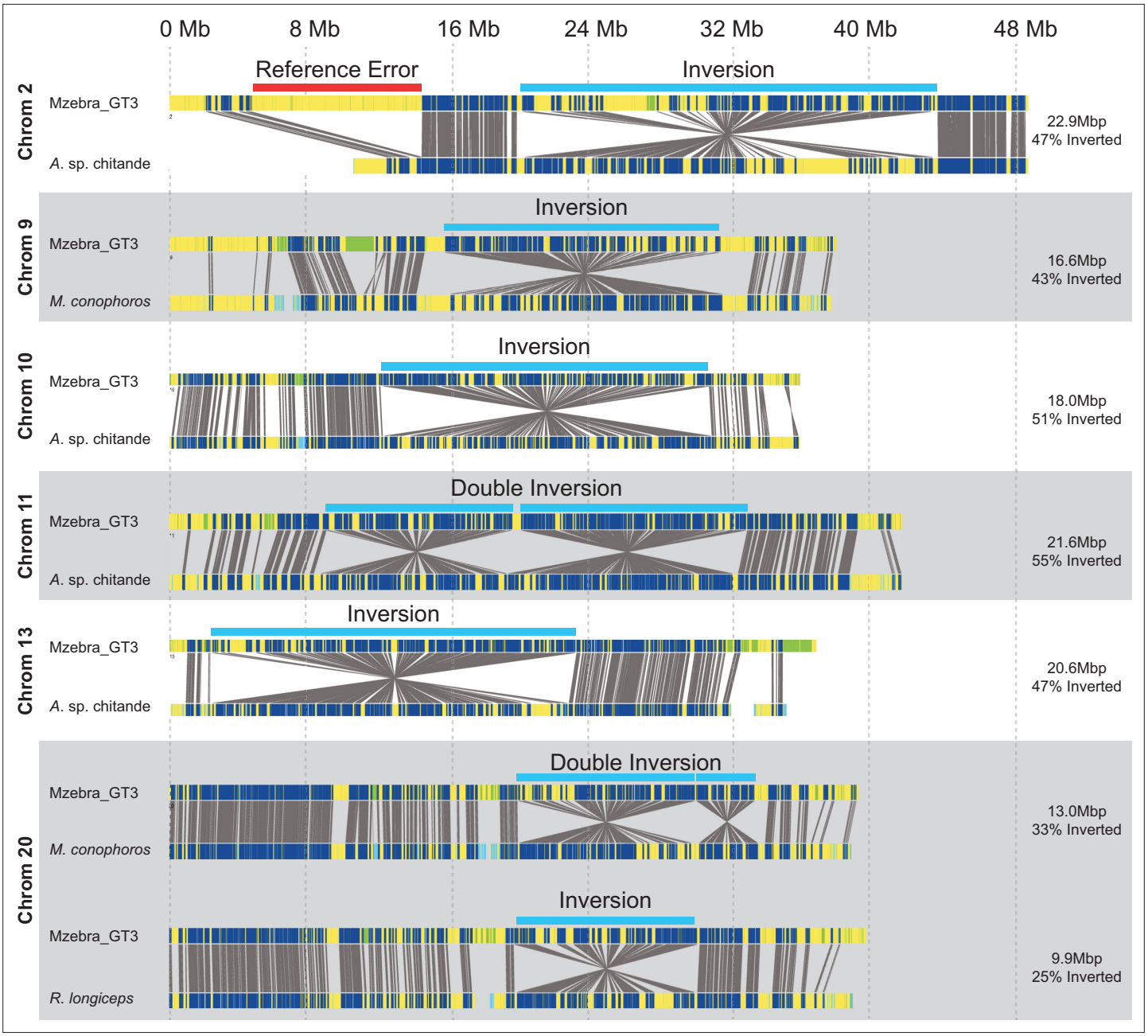


Figure 3. Identification of six large single or double inversions segregating within Lake Malawi cichlids. Representative alignments of inversions or double inversions identified from blood samples obtained from 30 individuals from eleven species. For each inversion, the top shows predicted CTTAAG motifs (dark blue) from an in-silico annotation of the reference genome, the bottom shows motifs identified using Bionano molecules obtained from an individual of the species indicated on the left, and the gray lines indicate matching motifs based upon predicted and observed distances. Motifs in the reference or sample tracks with no matches are shown in green and light blue, respectively. Yellow indicates regions containing no matches between motifs in the reference genome and the mapped sample. Single inversions were identified on chromosomes 2, 9, 10, 13, and 20. Tandem double inversions were identified on chromosomes 11 and 20. The estimated length and percentage of the chromosome spanned by the inversion is shown on the right. The single inversion on 20 identified in *Rhamphochromis longiceps* has the same position as the first inversion of the double inversion on 20 identified from *Mchenga conophoros*. A list of all samples and their inversion genotype is in ***Table 1***. Note that an error or polymorphism in the reference genome in chromosome 2 is indicated in the first panel.

The online version of this article includes the following figure supplement(s) for figure 3:

Figure supplement 1. Bionano maps from individuals that were heterozygous for the inversions on chromosomes 9, 10, or 11.

Figure supplement 2. Genome alignments for chromosome 2.

Figure 3 continued on next page

Figure 3 continued

Figure supplement 3. Genome alignments for chromosome 9.

Figure supplement 4. Genome alignments for chromosome 10.

Figure supplement 5. Genome alignments for chromosome 11.

Figure supplement 6. Genome alignments for chromosome 13.

Figure supplement 7. Genome alignments for chromosome 20.

Figure supplement 8. Genome alignments to Lake Malawi outgroups.

Rhamphochromis individuals that carried only the first inversion, lacking the 3.1 Mb long second part of this rearrangement (***Figure 3***). This indicates that the tandem inversions on 20 were formed by at least a two-part process, most parsimoniously with the left arm of the inversion (20 a) forming first followed by the right inversion (20b) occurring in that genetic background.

The distribution within these species is summarized in ***Table 1***. For simplicity, we will refer to all six rearrangements as inversions, regardless of whether they are single or double inversions. The most common inversions were on chromosomes 9, 11, and 20 which were found in a combination of *benthic* and *Diplotaxodon* individuals. The position of these three inversions was also consistent with regions of low recombination rate previously identified in an intercross between a *mbuna* and *benthic* individual (***Conte et al., 2019***), as expected, since inversions suppress recombination between inverted and non-inverted haplotypes. The remaining three inversions (2, 10, and 13) were only found in the deep benthic individuals (*Aulonocara sp. 'chitande type north' Nkhata Bay*). While for most samples the inversion genotype was homozygous, for six individuals the optical maps supported both inverted and non-inverted haplotypes, indicating the inversions were heterozygous in those samples (***Figure 3—figure supplement 1***).

de novo reference genomes support the presence of these six inversions

To provide further support for these structural rearrangements, we generated de novo hybrid genome assemblies of three *Aulonocara sp. 'chitande type north' Nkhata Bay* individuals, which carry all six of the identified inversions, and two additional *Metriaclima zebra* individuals using a combination of PacBio HiFi reads and Bionano maps. The overall quality of these assemblies is high, with large N50 values and a small number of individual contigs (***Supplementary file 3***). We performed whole-genome alignments between the five new assemblies and the GT3a reference genome (***Figure 3—figure supplements 2–7***). These alignments show that the new assemblies support the presence of all six inversions in the *Aulonocara sp. 'chitande type north' Nkhata Bay* individuals. For most inversions, a single scaffold covered the entire inversion with breakpoints in locations consistent with the Bionano maps. The two new control assemblies were consistent with the M_zebra_GT3a reference and did not show any evidence of inversions in those regions. In other regions of the genome, however, there were differences in the alignment of the M_zebra_GT3a assembly and the *Metriaclima zebra* genomes, most prominently in the left arm of chromosome 2 (***Figure 3***) and in the highly repetitive chromosome 3, which likely indicates some errors in the M_zebra_GT3a assembly or that a large polymorphism segregates within the species.

We refined the location of the endpoints for each inversion using the genomic alignments. However, the presence of a large amount of repetitive DNA in regions near these regions prevented us from defining inversion boundaries at single base pair resolution. The mechanisms that create large inversions often involve repetitive sequence near the breakpoints, making exact localization difficult if not impossible (***Villoutreix et al., 2021***).

We also took advantage of a genome assembly for a *Rhamphochromis sp 'chilingali'* individual (fRhaChi2.1) created by the Darwin Tree of Life Project produced using PacBio data and Arima2 Hi-C data (Accession PRJEB72870). The genome alignment between the fRhaChi2.1 assembly and the GT3a assembly confirmed a single inversion on chromosome 20, consistent with the Bionano data (***Figure 3—figure supplements 2–7***).

For each inversion, we determined the ancestral haplotype using two species as outgroups, *Oreochromis niloticus* and *Pundamilia nyererei*. *Oreochromis niloticus,* or Nile tilapia, is an important food fish that's estimated to have diverged from Lake Malawi cichlids 14.1–30 million years ago while

Pundamilia nyererei is a more closely related haplochromine found in Lake Victoria (**Brawand et al., 2014**; **Ciezarek et al., 2024**). For this analysis, we used a published whole-genome assembly for *Oreochromis niloticus* (UMD_NMBU) (**Conte et al., 2019**) and generated a hybrid genome assembly for *Pundamilia nyererei* using a previously published whole-genome assembly (**Brawand et al., 2014**) combined with Bionano maps we produced from a single individual (**Figure 3—figure supplement 8**). For all six inversions, the reference haplotype found in the GT3a assembly represents the ancestral state.

Inference of the segregation of these inversions in Lake Malawi using short-read sequencing

The distribution of these inversions within the lake can provide important information to their function. Genotyping inversions using short-read sequencing is possible because each haplotype is on an independent evolutionary trajectory due to the suppression of recombination. Population genetics approaches take advantage of the fact that inversions create patterns of inheritance between SNVs that are readily detected by approaches, such as Principal Component Analysis (PCA) (**Nowling et al., 2020**).

We analyzed short-read sequencing data from the 31 samples with Bionano data along with 297 wild individuals published by the Durbin lab (**Malinsky et al., 2018**), 15 wild individuals published by the Streelman lab (**Patil et al., 2021**; **York et al., 2018**), and 22 wild individuals sequenced for this paper, aligning reads and calling small variants against our new M_zebra_GT3a reference genome. PCA analysis was used to analyze the major axes of variation across the whole-genome and within the six inversions. We separately analyzed the left and right inversions on 11 and 20. For the whole-genome, the species within an ecogroup clustered together within the first two principal components, reflecting their evolutionary history. This matched the previously reported results of **Malinsky et al., 2018**. Individuals from four of the ecogroups – *Rhamphochromis*, *Diplotaxodon*, *Astatotilapia calliptera*, and *mbuna* – formed four distinct clusters (**Figure 4A** and **Figure 4—source data 1** for **Figure 4**). A fifth, more diffuse cluster, composed of *shallow benthics*, *deep benthics*, and *utakas*, was also present. The broader distribution of the *benthic* ecogroups in PC space is consistent with both their recent separation as well as large amounts of gene flow thought to have occurred between these different species (**Malinsky et al., 2018**; **Loh et al., 2013**).

The PCA plots using the SNVs within the six inversions often violated the pattern of the whole-genome PCA (**Figure 4A** and **Figure 4—source data 2** to **Figure 4—source data 9**). Using samples that were sequenced with both Illumina and Bionano technologies, we assigned each cluster to specific inversion genotypes (**Supplementary file 2**). We did not have Bionano data for two clusters (one on chromosome 9 and one on 11 – see **Figure 4—source data 3; 5 and 6** for **Figure 4**), so we supplemented our dataset by adding Bionano data from an additional *shallow benthic* individual (*Protomelas taeniolatus*) that fell into both clusters. The PCA clusters varied in their complexities for each inversion depending largely on their distribution within the *benthic* ecogroup (more on this below) and included both homozygous genotypes and heterozygous genotypes. We summarize the overall distribution of the inversions within each of the ecogroups in **Figure 4B**.

All specimens of the *mbuna* and *AC* ecogroup were fixed for the homozygous non-inverted genotype for all six inversions. In *Rhamphochromis*, all the inversions were homozygous non-inverted except for the left arm of 20 which was fixed for the single inversion state. The *Diplotaxodon* individuals all carried the non-inverted alleles for the inversions on 2, 10, and 13 and were fixed for the inversions on 9, 11, and 20.

The distribution of these inversions was the most complicated within the *benthics*. The inversions on 2, 10, and 13, which were identified in *Aulonocara sp. 'chitande type north' Nkhata Bay*, were mostly found together in other *deep benthics*, including the *Alticorpus* (*geoffreyi*, *macrocleithrum*, *peterdaviesi*), *Aulonocara* (*blue chilumba*, *gold* and *minutus*), and *Lethrinops* (*gossei*, *longimanus 'redhead', and sp. olivera*) genera. The inversions on 2 and 13 were also found in the shallow benthic species *Placidochromis longimanus*. However, not all the *deep benthics* carried the three inversions, including five *Aulonocara* species (*baenschi*, *gertrudae*, *steveni*, *stuartgranti*, and *stuartgranti* Maisoni) and a single *Lethrinops* species (*longipinnis*). The distribution of these three inversions in *benthic* species known to inhabit the deepest habitats are suggestive for a role in depth adaptation.

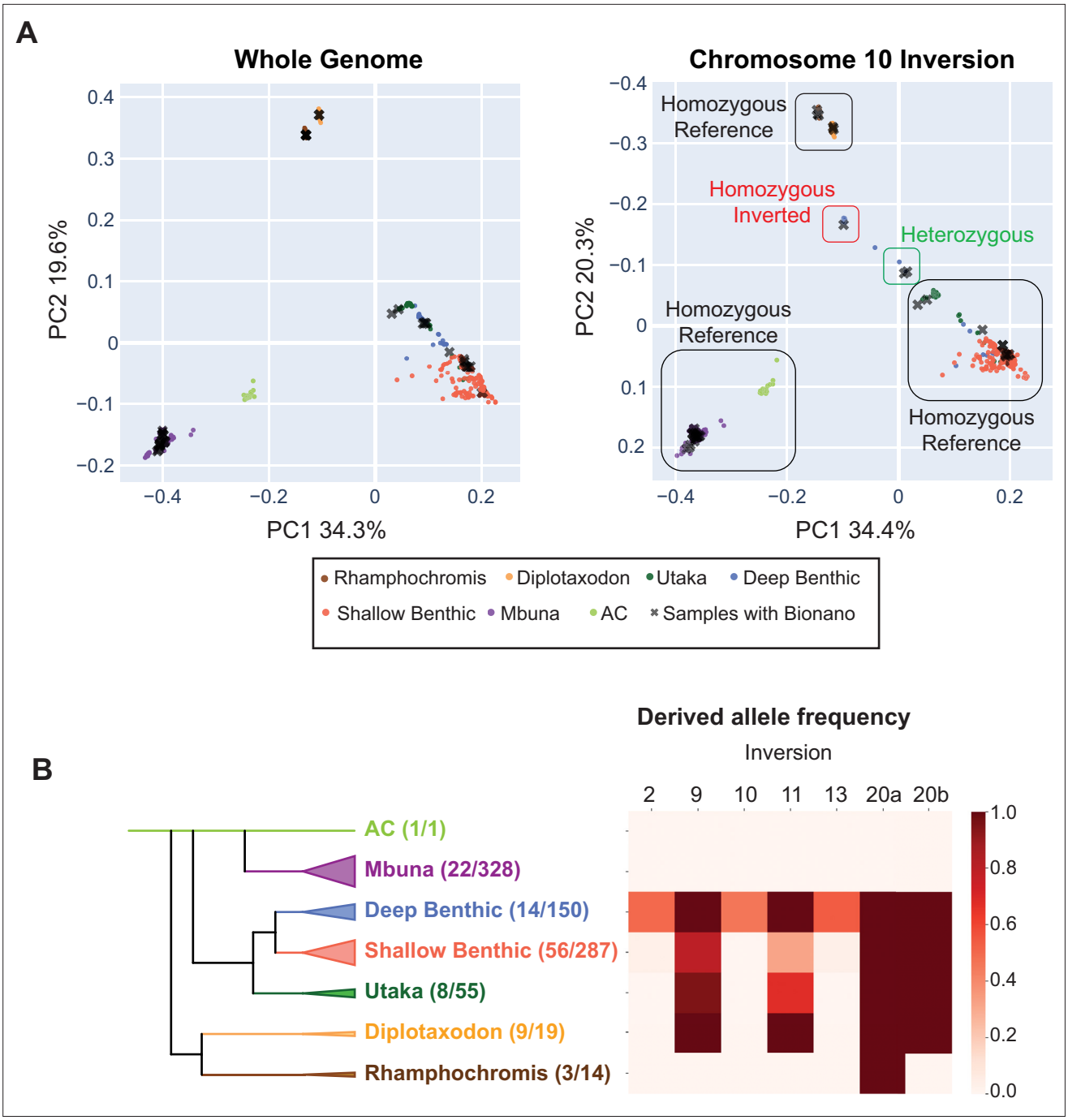


Figure 4. Distribution of inversions within the Lake Malawi ecogroups. Principal component analysis was used to analyze single nucleotide variants (SNVs) distributions in 365 samples within each inversion to genotype the inversion haplotype for all six inversions. (**A**) We illustrate the approach for genotyping the chromosome 10 inversions. The Principal Component Analysis (PCA) plot for the entire genome is shown on the left, with each sample colored by its ecogroup. Samples from each ecogroup cluster together, reflecting their evolutionary history. Individuals with Bionano data are shown as gray Xs. On the right, for the SNVs that fall within the chromosome 10 inversion, the deep benthic/shallow benthic and utaka individuals split into multiple clusters. Using the samples genotyped with Bionano data, we assign each cluster to an inversion genotype, which is shown by the black, green, and red boxes. To make comparisons of the whole-genome PCA and chromosome 10 PCA plots easier, we reversed the y-axis on the right panel. The assigned genotypes of each sample can be found in ***Supplementary file 2***. Interactive PCA plots for each inversion are included in ***Figure 4—source data 1*** to ***Figure 4—source data 9***. (**B**) The derived allele frequency was calculated for each inversion within the seven ecogroups. The number of species that were genotyped and the estimated number of species that live in Lake Malawi are listed in the whole-genome phylogeny (left).

The online version of this article includes the following source data for figure 4:

Figure 4 continued on next page

Figure 4 continued

Source data 1. Whole-genome Principal Component Analysis (PCA) of Lake Malawi cichlid samples across all ecogroups.

Source data 2. Principal Component Analysis (PCA) of Lake Malawi cichlids across the chromosome 2 inversion.

Source data 3. Principal Component Analysis (PCA) of Lake Malawi cichlids across the chromosome 9 inversion.

Source data 4. Principal Component Analysis (PCA) of Lake Malawi cichlids across the chromosome 10 inversion.

Source data 5. Principal Component Analysis (PCA) of Lake Malawi cichlids across the first tandem inversion on chromosome 11.

Source data 6. Principal Component Analysis (PCA) of Lake Malawi cichlids across the second tandem inversion on chromosome 11.

Source data 7. Principal Component Analysis (PCA) of Lake Malawi cichlids across the chromosome 13 inversion.

Source data 8. Principal Component Analysis (PCA) of Lake Malawi cichlids across the first tandem inversion on chromosome 20.

Source data 9. Principal Component Analysis (PCA) of Lake Malawi cichlids across the second tandem inversion on chromosome 20.

Both inversions on 20 were homozygous inverted in all benthic individuals. For the single inversion on 9, many *benthics* were homozygous for the inverted haplotype. However, an additional PC cluster was present, composed of 64 *shallow benthic* and *utaka* individuals heterozygous for the non-inverted haplotype (***Supplementary file 2***). Interestingly, no *benthic* individuals were identified as homozygous for the non-inverted state.

Finally, the distribution of the chromosome 11 inversion was also determined. Individuals of all three genotypes (homozygous non-inverted, homozygous inverted, and heterozygous) were observed (***Figure 4B***).

Evolutionary history of 9,11, and 20 suggest introgression from *Diplotaxodon into benthic ancestors*

The distribution of these inversions conflicts with the species tree, suggesting that these inversions could have spread to the *benthics* via hybridization or carried forward by incomplete lineage sorting. To determine their evolutionary history, we built phylogenies for the whole-genome as well as for each inversion using SNVs that fell within each of the six inversions using representative species from each ecogroup (73 total – ***Supplementary file 2***) as well as two previously sequenced *Pundamilia nyererei* individuals as an outgroup (***Figure 5A***, ***Figure 5—figure supplements 1–8***; ***Kratochwil et al., 2019***; ***Feulner et al., 2018***). To avoid issues with haplotype phasing, we focused on individuals that were homozygous for one of the two inverted genotypes.

While the whole-genome phylogeny matched the expected topology based upon our understanding of the Lake Malawi species tree, the topologies of each of the inversions showed key differences between the separation of ecogroups. The *benthic* group often split into separate clades, correlated with their inversion genotype. Prominently, we found that the three inversions on 9, 11, and 20 showed key differences in the relationship between the *Diplotaxodon* and *benthic* ecogroups. While *Diplotaxodon* is a sister group to *Rhamphochromis* at the species level, for these three inversions, the *Diplotaxodon* individuals formed a clade with the *benthics* that also carried the inversion. Additionally, genetic distance between *Diplotaxodon* and *benthics* carrying the inversion was also much smaller than the rest of the genome (***Figure 5B***). Both the phylogenies and genetic distances between ecogroups are consistent with the introgression of the 9, 11, and 20 inversions from the *Diplotaxodon* to *benthic* ancestors at the time of the benthic radiation.

A role for large inversions in sex determination within the benthic ecogroup

The evolution of sex determination is often associated with the presence of inversions to repress recombination between the sex chromosomes (***Johnson and Lachance, 2012***; ***Natri et al., 2019***). We tested whether these inversions could play a role in sex determination in *benthic* species, as we identified individuals that were heterozygous for some of these inversions (***Table 1***). We started with the inversion on 10, as our Bionano data identified three male *Aulonocara sp. 'chitande type north' Nkhata Bay* individuals that were heterozygous and a single female that was homozygous for the inverted haplotype. To test whether this inversion is linked to a genetic locus (or loci) that act as an XY system in this species, we tested two separate broods (Brood 1; 2 parents, 24 offspring and Brood 2: 2 parents, 24 offspring) currently growing in the laboratory, sequencing 12 males and 12 females from

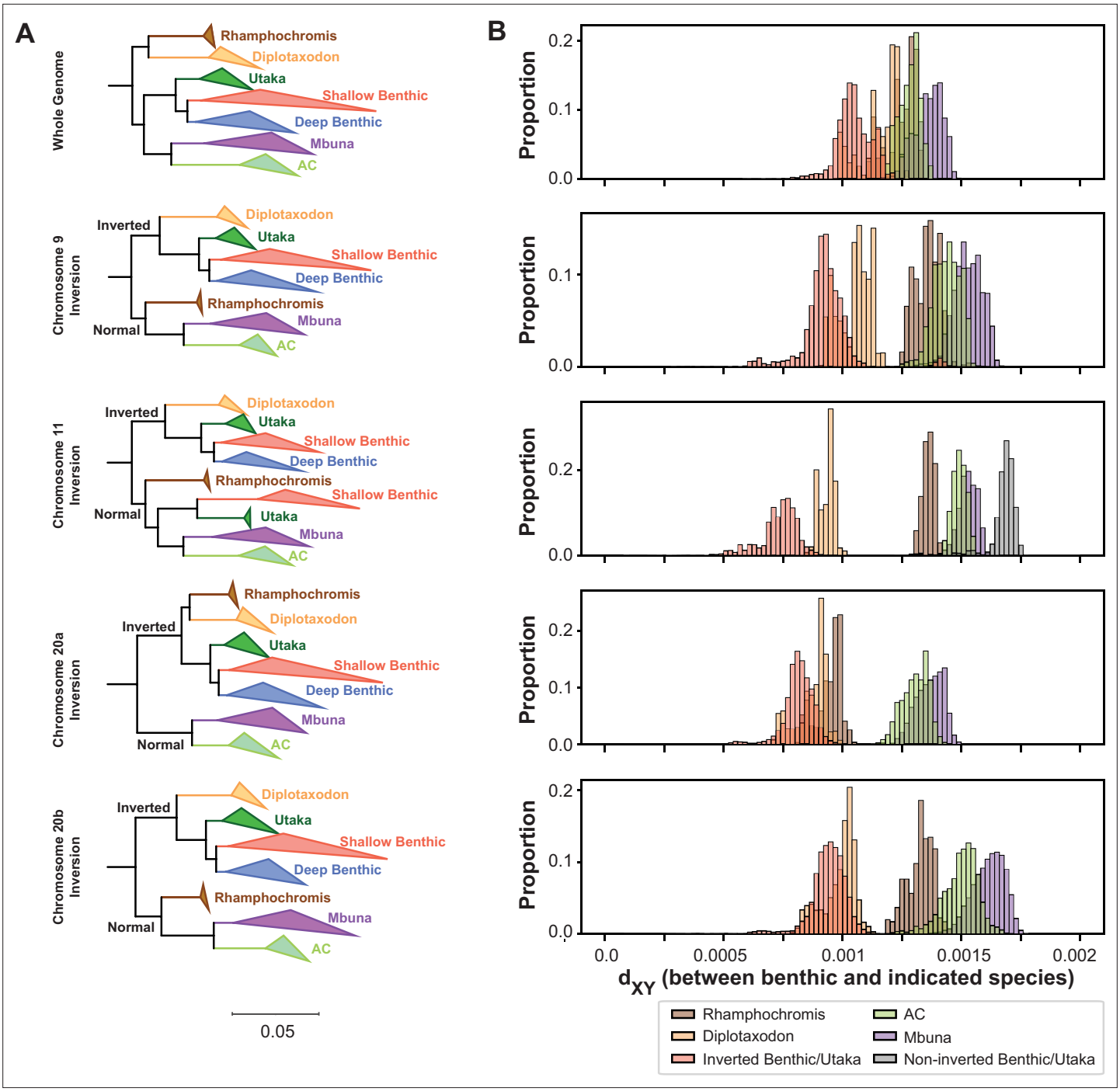


Figure 5. Evolutionary history of large inversions violates species tree. (**A**) Maximum likelihood trees built from either whole-genome single nucleotide variants (SNVs) (top) or within the 9, 11, and 20 inversions. Between 2–25 representative samples were selected for each ecogroup (see ***Supplementary file 2***) and individuals that were heterozygous were excluded from this analysis. Clades were collapsed by ecogroup for visualization purposes. Full phylogenies are presented in ***Figure 5—figure supplements 1–8***. The whole-genome tree indicates *Diplotaxodon* and *Rhamphochromis* species split first and formed their own clade. For each of the displayed inversions, the first split was between samples carrying the inverted vs. non-inverted genotypes. Additionally, the *Diplotaxodon* individuals now formed a clade with the *benthics* that carry the inversion. (**B**) Density plots of d_{XY} values comparing *benthic* species and species of the indicated ecogroup. For the displayed inversions, the *Diplotaxodon* species were much closer in evolutionary distance when compared to the whole-genome plot.

The online version of this article includes the following figure supplement(s) for figure 5:

Figure supplement 1. Whole-genome phylogeny for Lake Malawi cichlids.

Figure 5 continued on next page

Figure 5 continued

Figure supplement 2. Chromosome 2 inversion phylogeny for Lake Malawi cichlids.

Figure supplement 3. Chromosome 9 inversion phylogeny for Lake Malawi cichlids.

Figure supplement 4. Chromosome 10 inversion phylogeny for Lake Malawi cichlids.

Figure supplement 5. Chromosome 11 inversion phylogeny for Lake Malawi cichlids.

Figure supplement 6. Chromosome 13 inversion phylogeny for Lake Malawi cichlids.

Figure supplement 7. Chromosome 20 first tandem inversion phylogeny for Lake Malawi cichlids.

Figure supplement 8. Chromosome 20 second tandem inversion phylogeny for Lake Malawi cichlids.

each brood with Illumina paired-end sequencing. We used PCA to genotype each of the 52 animals (4 parents and 48 offspring) for the chromosome 10 inversion genotype. The association in the offspring between the inversion 10 genotype and sex was almost perfect: 24 females and 1 male were homozygous for the inversion, while 23 males were heterozygous for the inversion (***Figure 6B***). These individuals were sexed by venting, where a visual inspection of the relative sizes of the genital papilla and

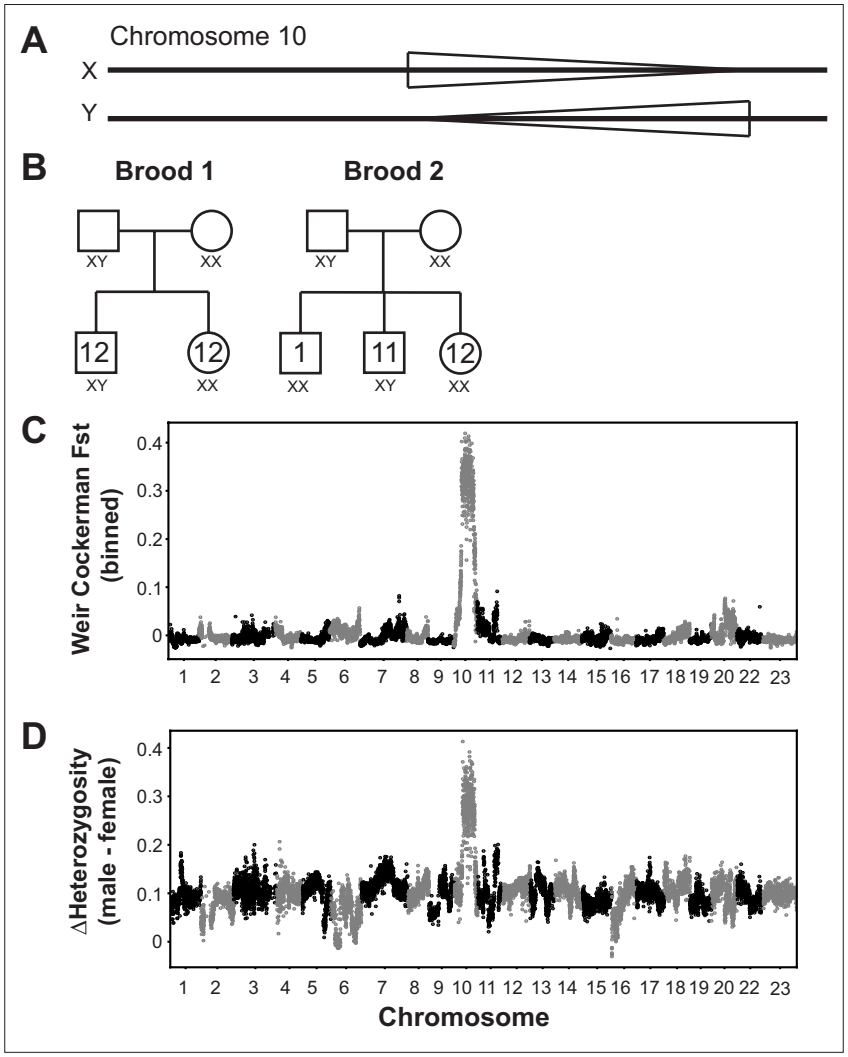


Figure 6. A role for the chromosome 10 inversion in sex determination. (**A**) Schematic of the two inversion states (normal and inverted) and their assignment as X or Y chromosomes. (**B**) 24 offspring (12 male and 12 female) from two separate male/female pairs were genotyped for the inversion. For 47 of 48 individuals, the XY animals were male, and the XX animals were female. One male was XX. (**C**) F_{ST} analysis comparing 24 males and 24 females identified chromosome 10 as the primary genetic region that segregated with sex. (**D**) Comparison of heterozygosity levels of 24 male and 24 female animals also identified chromosome 10 as an outlier.

anal pore are compared, with females typically having a noticeably larger urogenital pore. Because venting is most reliable in reproductively active adults, the imperfect segregation could reflect either an error in sexing or the influence of other modifiers on sex.

We also used binned Weir Cockerham F_{ST} analysis, grouping male and female individuals to scan the genome for regions that were associated with sex (***Figure 6C***). The inversion on 10 was the region of the genome that was in the strongest linkage disequilibrium with sex. Males also had a much higher heterozygosity in this region than females (***Figure 6D***). Altogether, these results indicate that the inversion on 10 encompasses a region containing an XY sex determiner in the *Aulonocara sp. 'chitande type north' Nkhata Bay* species.

We also tested eight laboratory *Nyassachromis prostoma* 'Orange Cap' individuals, as our Bionano data identified a male that was heterozygous for the inversion on 11 and a female that was homozygous for the inversion. We performed short-read sequencing on three males and five females growing in the lab, genotyping the inversion using PCA analysis. All three males were heterozygous for the inversion while the five females were homozygous, a significant association between genotype and sex (Fisher exact test *p*-value = 0.0179). This suggests a separate XY system segregates in *Nyassachromis prostoma* 'Orange Cap,' linked to the inversion on chromosome 11.

Discussion

Here, we characterize the presence of large inversions found in the Lake Malawi cichlid flock, identifying six inversions segregating within the lake, providing a framework to address the role structural variants played in this lake's adaptive radiation (***Figure 1***). What are the evolutionary forces responsible for their spread and what role do they currently play in the lake? Inversions have multiple, well-characterized roles in adaptive evolution: (1) they can capture beneficial alleles responsible for adaptation to local habitats, (2) they can capture alleles that create prezygotic and postzygotic barriers that prevent gene flow during speciation, and (3) they can capture alleles that influence sex-specific phenotypes that are linked to sex-determination systems (***Kirkpatrick, 2010***; ***Berdan et al., 2023***; ***Hoffmann and Rieseberg, 2008***). These inversions could have also spread through neutral processes like genetic drift or through other adaptive processes not listed here.

This question is much more complicated in the context of the Lake Malawi cichlid radiation with ~800 different species. The role of these inversions must be considered in a species- and time-specific way. The recent history of the lake was quite dynamic, with large cyclical changes in water depth and shoreline in the past 800,000 years that are predicted to create alternating periods of extinctions and radiations (***Malinsky and Salzburger, 2016***; ***Lyons et al., 2015***; ***Ivory et al., 2016***). Similarly, the role the inversions play in each species (such as sex determination) will not be the same. Understanding whether these inversions carry alleles that contribute to the phenotypic diversity found in Lake Malawi will be an important first step towards resolving the evolutionary forces responsible for their spread through the lake. This is possible as most species in the lake can interbreed, allowing for quantitative genetics approaches to link genotype variation to phenotypic variation and determine what traits are influenced by the inversions.

The local adaptation mechanism is one well-characterized method to explain how inversions can spread through adaptive evolution (***Kirkpatrick and Barton, 2006***). Under this hypothesis, two populations undergoing adaptive divergence in the presence of migration and gene flow create conditions that can select for inversions, provided they capture and link adaptive alleles together that increase fitness to their local environments (***Harringmeyer and Hoekstra, 2022***; ***Hager et al., 2022***). Apart from the *benthics*, the inversions are fixed in each ecogroup in the inverted or non-inverted state, each of which live in unique habitats in the lake that place very different evolutionary pressures on the animals that live within them. In this scenario, the inversions would capture adaptive alleles that increase fitness of ancestral species to pelagic and deep-water environments. To illustrate this scenario, we consider the time the pelagic groups split from the ancestral AC fish that is thought to live in rivers and the borders of the lake (***Figure 1***). If the first inversion on 20 captured alleles that allowed these pelagic ancestors to inhabit the lake, it could be selected on provided that these pelagic ancestors continue to interbreed with the AC ancestors. Similarly, the diversification of the *Rhamphochromis* and *Diplotaxodon* groups could have involved the inversions on 9, 11, and right arm of 20, and the adaptation of *benthics* to deeper water habitats could have involved the inversions on 2, 10, and 13. We do note that the existing theory for recruiting inversions to local adaptation does

not account for multiple unlinked inversions (i.e. 9, 11, and right arm of 20). Future experimental and comparative genomics approaches can determine if local adaptation was at play at the time these lineages separated.

We also provide evidence that two inversions span genomic regions involved in sex determination in the *benthics*, with the strongest evidence showing the inversion on 10 is linked to an XY system in a deep benthic species. Additionally, we show that the inversion on 11 is significantly associated with an XY system, in a shallow benthic species. Despite the importance of the maintenance of distinct, reproductively compatible sexes, sex-determination systems can be evolutionarily labile, and Lake Malawi cichlids show exceptional variability, with at least five different genetic sex-determination systems described to date (***Parnell and Streelman, 2013***; ***Behrens et al., 2024***). While we implicated the involvement of two inversions in sex determination, our work also underscores how dynamic sex determination is within the *benthics*. While many *benthic* species carry the different inversions, their genotypic distribution in most species is inconsistent with a role for sex determination. Effectively, the genetic basis of sex determination largely must be determined in a species-by-species way, and individual species may be segregating multiple sex determination systems and/or loci. The selective forces on sex chromosomes are strong, including balancing selection for sex ratio and sexually antagonistic selection on sexually dimorphic phenotypes (***Meisel, 2022***; ***Scott et al., 2018***). Lake Malawi cichlids show amazing sex differences, characterized by diversification in numerous phenotypes, including pigmentation, behavior, and morphology thought to contribute to species barriers (***Santos et al., 2023***). The segregation of sex-determining loci with inversions allows for the accumulation of alleles in linkage with the sex-determination gene which are predicted to have sex-specific effects in the male individuals that carry the Y chromosomes. The presence of these inversions in various ecogroups could simply be remnants of their ancestral linkage to sex determining loci in the pelagic ecogroups and have nothing to do with local adaptation driving their spread. Under this scenario, balancing selection would have maintained both the inverted/non-inverted haplotypes. Sex chromosome turnover would then have ended this balancing selection, leading to subsequent fixation of these inversions. Identification of the causal alleles for sex determination on these inversions will help us to resolve this question.

The presence of non-inverted haplotypes in the *benthics* could have been retained from the time the lineage initially formed, with both haplotypes under balancing selection if they harbored sex-determining genes. A recent publication, however, which also identified these 6 inversions presented here using multiple technologies, also utilized short-read sequencing of 1375 individuals to create an alternative model (***Blumer et al., 2025***). They inferred the non-inverted haplotypes of the 9 and 11 inversions were introgressed back into the *benthics* from riverine species within and outside of the lake. Similarly, the non-inverted haplotypes on chromosomes 2, 10, and 13 in the *deep benthics* were introgressed from *shallow benthics*. In this model, it is unclear what evolutionary forces drove these back introgressions of the non-inverted alleles, but would imply these inversions have not continuously provided the genetic sex determination mechanism since their inception.

The distribution of the inversions in the benthic sub-radiation is uniquely complicated in Lake Malawi. Inverted and non-inverted haplotypes for five of the six inversions are segregating within this group. Additionally, the presence of the 9, 11, and 20 inversions within *benthics* violates the whole-genome phylogeny. We propose these three inversions originated within the *Diplotaxodon* group and spread to a very early *benthic* ancestor via introgression. We favor this over incomplete lineage sorting, which would also lead to violations of the phylogeny, for a few reasons. First, the dynamic history of the lake levels is consistent with times of extinction and hybridization (***Malinsky and Salzburger, 2016***; ***Lyons et al., 2015***; ***Ivory et al., 2016***). Second, genomic analysis of extant cichlids provided evidence for a large amount of hybridization events (***Loh et al., 2013***; ***Malinsky et al., 2018***). Finally, the close genetic distance between *Diplotaxodons* and *benthics* that carry the inverted haplotype (***Figure 5***) is consistent with hybridization. Hybridization has been proposed to fuel adaptive radiations of cichlids in all three great African lakes (Malawi, Tanganyika, and Victoria) (***Santos et al., 2023***; ***Svardal et al., 2021***; ***Meier et al., 2023***) and our work suggests that an ancient hybridization event between early *benthic* and *Diplotaxodon* ancestors occurred prior to the benthic radiation. It will be interesting to study the role of these three inversions in *benthic* species. *Diplotaxodon* are pelagic, deep-water dwellers that feed on plankton or small fish while *benthics* live near the shore over sandy and muddy habitats (***Malinsky et al., 2018***). It is not obvious how the traits of the

Diplotaxodon would benefit *benthic* animals that live in such dissimilar environments. However, the initial habitats that the *benthic* ancestors utilized is not known and potentially their initial habitat was in deeper waters than the *AC* ancestors.

This report provides an initial framework for understanding the role of inversions in the adaptive radiation of Lake Malawi cichlids. Future work identifying and characterizing causal alleles captured within these inversions will enable us to discern the specific mechanisms by which they contribute to speciation using this remarkable evolutionary system as a model for the evolution of all animals.

Methods

Samples

Wild-caught adult cichlid fishes were acquired from Old World Exotic Fish Inc in 2022 and Cichlidenstadel in 2024. Lab-reared species were purchased from Southeast Cichlids, Old World Exotic Fish Inc, and Cichlidenstadel at various points in the past two decades (***Supplementary file 2***). Fish were delivered live to the Georgia Institute of Technology cichlid aquaculture facilities. All samples were collected from adult-stage individuals following anesthetization with tricaine according to procedures approved by the Institutional Animal Care and Use Committee (IACUC protocol numbers A100029 and A100569).

Caudal fins were collected from subjects and flash-frozen on powdered dry ice before storage at –80 °C. Care was taken to minimize freeze-thaw cycles prior to DNA extraction via Bionano or Qiagen protocols. If necessary, fins were sectioned on a sterile aluminum block set in dry ice prior to digestion.

Blood was collected following rapid decapitation of anesthetized subjects. Wide-bore, low-retention pipette tips, 0.5 M EDTA (pH 8.0, Invitrogen Cat# 15575–038), and Bionano Cell Buffer (Part Number 20374) prevented blood from clotting at room temperature. Samples were used immediately for Bionano DNA extraction (see below).

Bionano

Bionano DNA extractions were predominantly performed on fresh whole blood samples (see above). Blood concentration was determined via a manual logical count performed on a hemocytometer and 2 million cells were carried forward for DNA extraction with the Bionano SP-G2 Blood & Cell Culture DNA Isolation Kit (Part Number 80060) as per the Bionano Prep SP-G2 Fresh Human Blood DNA Isolation Protocol (Document Number CG-00005, Rev C). Because nucleated cichlid blood had high concentrations, steps 3 and 4 were skipped and samples were resuspended in Bionano Cell Buffer to a total volume of 200 µL. Following extraction, samples were allowed to homogenize undisturbed at room temperature for at minimum 72 hr prior to quantification using the Qubit dsDNA Quantitation, Broad Range kit (ThermoFisher, Cat # Q32853).

Bionano data for a single *Protomelas taeniolatus* sample (PT_2003_m, see ***Supplementary file 2***) was generated from fin tissue using a combination of the Bionano Prep SP Tissue and Tumor DNA Isolation Protocol (Document Number 30339, Rev A) and the Bionano Prep SP-G2 Fresh Human Blood DNA Isolation Protocol.

750 ng of homogenized DNA was fluorescently labeled using the Bionano Direct Label and Stain-G2 (DLS-G2) Kit (Part Number 80046, Protocol - CG-30553–1, Rev E). Labeled DNA was quantified using the Qubit dsDNA Quantitation, High Sensitivity kit (Thermo Fisher, Cat # Q32854) and loaded onto a flow cell within a Bionano Saphyr Chip G3.3 (Part Number 20440). Samples were imaged on the Bionano Saphyr with at least 500 Gbp of data collected per sample.

Molecules for each sample were assembled using the Bionano de novo assembly pipeline on Bionano Access Version 1.7. The preassembly setting was turned on, and the variant annotation parameters were deselected for every assembly. De novo maps were aligned to the Mzebra_GT3a genome.

PacBio HiFi sequencing and genome assembly

PacBio de novo assemblies were generated from HiFi reads across three sequencing runs. The MZ_GT3.3 and YH_GT1.3 were sequenced following DNA extraction from frozen caudal fin tissue using the Qiagen MagAttract HMW DNA Kit (Cat. No. 67563). In the first sequencing run a *Metriaclima zebra* female was sequenced on a PacBio Sequel II system by the Georgia Genomics and Bioinformatics

core. In the second run, the same *Metriaclima zebra* female and an *Aulonocara sp. 'chitande type north' Nkhata Bay* female was sequenced on a PacBio Revio instrument by the HudsonAlpha Institute for Biotechnology. Reads from both runs were used to assemble the MZ_GT3.3 genome and reads from the second run alone were used to assemble the YH_GT1.3 genome (see details below).

Mzebra_GT3a, MZ_GT3.2, YH_GT1.1, and YH_GT1.2 were assembled using HiFi reads generated with DNA extracted from fresh whole blood and heart tissue. DNA was extracted via the PacBio Nanobind Tissue Kit RT (Part Number 102-208-000) following a modified DNA from animal tissue using the Nanobind PanDNA kit protocol (Part Number 102-574-600). Blood and heart tissue were combined and homogenized together using a Qiagen TissueRuptor (Qiagen 9002755) and centrifugation g-force was halved during the DNA extraction. DNA fragments <25 kb were removed using the PacBio Short-Read Elimination kit (Part Number 102-208-300). Library preparation and sequencing was performed by the University of Maryland Institute for Genome Sciences on a PacBio Revio instrument.

HiFi reads from all samples were assembled using the Mabs de novo assembler (v2.28) (***Schelkunov, 2023***) using the `mabs-hifiasm` algorithm and default parameters. Each assembly from mabs-hifiasm was evaluated using Inspector (`inspector.py` v1.0.1) (***Chen et al., 2021***) run on default parameters. The assemblies were error-corrected with `inspector-correct.py` (v1.0) to resolve large structural errors. These error-corrected contigs from were uploaded to Bionano Access.

The MZ_GT3.2, MZ_GT3.3, YH_GT1.1, YH_GT1.2 and YH_GT1.3 scaffold-level assemblies were generated by bridging gaps in their contigs using Bionano maps via the single-enzyme Bionano Hybrid Scaffold pipeline using default parameters. A unique set of Bionano maps was used to scaffold the contigs from each de novo assembly. The resulting hybrid scaffold `NCBI.fasta` file (which denotes gaps filled by Bionano maps with stretches of N nucleotides) was concatenated with any unscaffolded contigs from the error-corrected mabs assembly. These five scaffold-level assemblies have been deposited at DDBJ/ENA/GenBank with the following BioSample and WGS accessions: MZ_GT3.2 (SAMN44293007; JBRIJA000000000), MZ_GT3.3 (SAMN44293008; JBRIIZ000000000), YH_GT1.1 (SAMN44293009; JBRIIY000000000), YH_GT1.2 (SAMN44293010; JBRIIX000000000), and YH_GT1.3 (SAMN44293011; JBRIIW000000000).

A scaffold-level genome was identically generated for Mzebra_GT3a. These scaffolds were further aligned to the corrected UMD2a genome (see below) and anchored to linkage groups using D-GENIES (***Cabanettes and Klopp, 2018***) with the Minimap2 v2.28 aligner and 'Many repeats' options. An anchored genome was output using the Query assembled as reference export option.

Note that the mitochondrial DNA sequence in Mzebra_GT3a is the same as the mitochondria assembled in UMD2a. This Whole-Genome Shotgun project has been deposited at DDBJ/ENA/GenBank under the accession JBEVYI000000000. The version described in this paper is version JBEVYI010000000.

Correcting errors in M_zebra_UMD2a

We generated Bionano molecules for three male and one female *Metriaclima zebra*, assembled Bionano maps, and aligned them to the UMD2a reference genome. We reordered or reoriented contigs that were revealed as misassemblies or inversions, respectively, in all four Bionano maps and corresponded to breakpoints between contigs in UMD2a. By filtering interchromosomal translocation calls in the Bionano Access software, we were able to insert unmapped contigs into the established linkage groups.

Illumina extractions and sequencing

DNA was extracted from fresh or frozen fin tissues for short-read sequencing using the Qiagen MagAttract HMW DNA Kit (Cat. No. 67563) or the Qiagen DNeasy Blood & Tissue Kit (Cat No. 69504) using the Manual Purification of High-Molecular Weight Genomic DNA from Fresh or Frozen Tissue and Purification of Total DNA from Animal Tissues (Spin-Column) protocols respectively. DNA was delivered to the Georgia Tech Molecular Evolution Core where libraries were prepared with the NEBNext Ultra II FS DNA Library Prep Kit for Illumina (NEB #E6177) using the Protocol for FS DNA Library Prep Kit (E7805, E6177) with Inputs ≥100 ng. A small subset of samples was sent to an external collaborator by the Molecular Evolution Core where library preparation was performed using the KAPA HyperPrep Kit (Roche, Material Number: 07962363001). Samples across all runs were sequenced on a NovaSeq 6000 instrument using v1.5 chemistry.

Variant calling and PCA analysis

We provide a link to a GitHub repository (https://github.com/ptmcgrat/CichlidSRSequencing/tree/Kumar_eLife, copy archived at ***McGrath, 2026***) containing the scripts used for the major analysis in the paper. Because our data is behind a secure Dropbox account, readers will be limited to seeing the exact programs, filters, and parameters used for this manuscript embedded within each script.

Fastq files from Illumina sequencing were converted to UBAM format using the `gatk` (***Poplin et al., 2018***) `FastqToSam` algorithm. UBAM files were used for alignment to Mzebra_GT3 using `bwa-mem` (***Li and Durbin, 2009***); the `-M` and `-p` flags were used. To carry forward metadata from the unaligned BAMs, `gatk MergeBamAlignment` was called on the alignments. BAM files were converted to GVCF format using `gatk HaplotypeCaller`. Variant calling was performed using the GVCF files for the analysis cohort. The `gatk GenomicsDBImport` and `gatk GenotypeGVCFs` algorithms were used to generate a master vcf file which was subsequently filtered with `gatk VariantFiltration` (parameters below). The variants that passed filtering were used for PCA analysis.

PCA was performed with `plink2` (***Chang et al., 2015***). To avoid overrepresenting species from the cohort, a core subset of the samples was used for PC 1 and 2 calculations (***Supplementary file 2***, column S). Eigenvectors from the whole cohort were then plotted on this PC space and visualized using `plotly`.

Note that GATK ≥v4.3.0.0 and Python ≥3.7 were used for these analyses. Custom Python scripts were written to automate and parallelize processing of the samples in the cohorts.

The commands used for variant calling and PCA are the following:

UBAM generation

```
>gatk FastqToSam --FASTQ<fq1> --FASTQ2<fq2> --READ_GROUP_NAME <RUNID >
--TMP_DIR <Temp_dir> --OUTPUT<temp_bam_file> --SAMPLE_NAME <sample_name>
--LIBRARY_NAME <library_name> --PLATFORM<platform >
```

Alignment to GT3

```
>bwa mem -t<threads > -M -p<GT3_FASTA><UBAM_file>
```

Merge metadata from UBAM to BAM

```
>gatk MergeBamAlignment -R<GT3_FASTA> --UNMAPPED_BAM <UBAM_file> --ALIGNED_
BAM <BAM_file>
```

Generate GVCFS

```
>gatk HaplotypeCaller -R<GT3_FASTA> -I<BAM_file> -ERC GVCF -O<GVCF_file>
```

`-ERC GVCF` is used to generate GVCF files that can be used for subsequent joint genotyping

Read depth information

```
>gatk CountReads -I<BAM_file>
```

Generate genomicsDB

```
>gatk --java-options -Xmx<memory > GenomicsDBImport --genomicsdb-workspace-
path<database_output_path> --intervals<interval > --sample-name-map<cohort_
samples.txt>
```

Joint genotyping

```
>gatk --java-options -Xmx<memory > GenotypeGVCFs -R<GT3_FASTA> -V<path_to_
database> -O<VCF_file> --heterozygosity 0.00175 -A DepthPerAlleleBySample
-A Coverage -A GenotypeSummaries -A TandemRepeat -A StrandBiasBySample
-A ReadPosRankSumTest -A AS_ReadPosRankSumTest -A AS_QualByDepth -A
AS_StrandOddsRatio -A AS_MappingQualityRankSumTest -A FisherStrand -A
QualByDepth RMSMappingQuality -A DepthPerSampleHC -G StandardAnnotation AS_
StandardAnnotation -G StandardHCAnnotation
```

Note that an average measure of heterozygosity, 0.00175 was used from what was reported in ***Malinsky et al., 2018***.

Filtering variants

```
>gatk VariantFiltration -R<GT3_FASTA> -V<VCF_file> -O<Filtered_VCF> --filter-
name 'allele_freq' --filter-expression 'AF <0.05' --filter-name 'inbreeding_
test' --filter-expression 'InbreedingCoeff <-0.6' --filter-name 'depth_Qual'
--filter-expression 'QD <2.0' --filter-name 'max_DP' --filter-expression 'DP
>11000' --filter-name 'min_DP' --filter-expression 'DP <7600' --filter-name
'strand_bias' --filter-expression 'FS >40.0' --filter-name 'mapping_quality'
--filter-expression 'MQ <50.0' --filter-name 'no_calls' --filter-expression
'NCC >119' --verbosity ERROR"))
```

Variants were filtered by allele frequency, inbreeding coefficient, quality by depth, depth, fisher's exact test for strand bias, mapping quality, and by excess missingness according to methods published in ***Malinsky et al., 2018***. We filtered by depth to include variants with a mean depth per sample between 10% and 95% of the distribution of all variant depths.

```
>gatk SelectVariants -V<Filtered_VCF> --exclude-filtered -O<pass_VCF_file>
```

PCA

For PCA, first a subset of the pass_VCF_file containing the samples we used to create the principal components was generated using `bcftools` (***Li, 2011***) (see ***Supplementary file 2***, Column S). Next, the relevant pfiles were generated for both the pass_VCF_file and the subset_samples_VCF. Linkage pruning was only performed for the whole-genome and whole chromosome PCAs using the –indep-pairwise 50 5 0.1 flag and parameters. Since inversions inherently link together variants, linkage pruning was not used when restricting the analysis to within inverted regions. A linear scoring system is generated using the subset_samples_VCF pfiles. This scoring system is applied to the genotype matrix for the whole cohort's samples to scale each sample consistently. The resulting first two eigenvectors per sample from the .sscore file are plotted with `plotly`.

```
>plink2 --vcf<pass_VCF_file> --out<whole_pfiles>
>plink2 --vcf<subset_samples_VCF_file> --out<subset_pfiles>
>plink2 --pfile<whole_pfiles> --set-missing-var-ids @:# --make-pgen
--out<whole_corrected>
```

LD Pruning for Whole-genome and whole-chromosome PCAs

```
>plink2 --pfile <subset_pfiles> --out<subset_samples_ld_pruning_intermediate>
--allow-extra-chr --set-missing-var-ids @:#' --indep-pairwise 50 5 0.1
>plink2 --pfile <subset_pfiles> --freq counts -pca allele-wts -out <sample_
subset_yes_ld_pruning_pca> --allow-extra-chr --set-missing-var-ids @:#
--max-alleles 2 --extract <subset_samples_ld_pruning_intermediate.prune.in
```

```
>plink2 --pfile <whole_corrected> --read-freq<sample_subset_yes_ld_pruning_
pca.acount> --score<sample_subset_yes_ld_pruning_pca.eigenvec.allele>2 5
header-read no-mean-imputation variance-standardize --score-col-nums 6-15
--out <projected_pca_yes_ld> --allow-extra-chr
```

No LD Pruning for inverted region PCAs

```
>plink2 --pfile<subset_pfiles> --freq counts --pca allele-wts -out <subset_
samples_no_ld_pruning_pca> --allow-extra-chr --set-missing-var-ids @:#
--max-alleles 2
>plink2 --pfile <whole_corrected> --read-freq<subset_samples_no_ld_pruning_
pca.acount> --score<subset_samples_no_ld_pruning_pca.eigenvec.allele>2 5
header-read no-mean-imputation variance-standardize --score-col-nums 6-15
--out<projected_PCA_no_ld> --allow-extra-chr
```

Genome alignment

Genomes were aligned using `minimap2` (***Li, 2018***) (v2.22) to align the query genome to the M_zebra_GT3 reference using default settings. Custom scripts were used to convert the output paf files to dataframe objects, filtering removed alignments that were (1) secondary matches, (2) had less than 30% identity, and (3) were less than 10,000 base pairs in length. Furthermore, we excluded any alignments from a contig containing less than 2,000,000 total bp of alignments, to account for repetitive DNA. Alignments were plotted using `seaborn`. We estimated the intervals containing the breakpoints for the inversions using genome alignments of *Aulonocara sp. 'chitande type north' Nkhata Bay* and *M_zebra_GT3* (***Supplementary file 4***).

Phylogenies

To create phylogenies, `bcftools` was used to filter out non-SNV variants from the master vcf file. `Bcftools` was also used to create individual vcf files for each inversion, filtering out SNVs that fell outside of the inverted region. `vcf2phylip` (***Ortiz, 2019***) was used to convert the to the phylip format, using the -m 50 option to filter out SNVs with less than 50 genotyped samples. `Iqtree2` (***Minh et al., 2020***) was used to create trees and estimate confidence values for each node using the following options: `-nt`, `'24'`, `'-mem'`, `'96G'`, `'-v'`, `'--seqtype'`, `'DNA'`, `'-m'`, `'GTR +I + G'`, `'-B'`, `'1000'`. Trees were visualized with `iTol` (***Letunic and Bork, 2024***) or `ete3` (***Huerta-Cepas et al., 2016***).

Nucleotide diversity analysis

To analyze the Pi_{XY} between individuals of different species, we used `scikit-allel` (v1.3.8) (***Miles, 2024***) to calculate the genetic difference between each sample using the `allel.pairwise_distance` function and the `cityblock` for the metric. We restricted the analysis to benthic animals that carried the inversion and compared these to animals of various ecogroups. We used the `seaborn` function `histplot` to create density histograms for each of the inversions.

F_{ST} & pedigree

A vcf file was created for the *Aulonocara* samples listed in ***Supplementary file 2*** (Column I). The PCA approach (described above) was used to genotype each of the samples for the inversion on 10. For the F_{ST} analysis, we used `scikit-allel` (v1.3.8) to perform `allel.average_weir_cockerham_fst` with a window size of 100. For the heterozygosity analysis, we used the `allel.heterozygosity_observed` function on both the male and female population, subtracting the average heterozygosity from the female offspring from the male offspring.

Acknowledgements

We thank Krish Roy for the acquisition of the Bionano Saphyr system, Manasi Pimpley for assistance in adapting existing Bionano kits to cichlid tissue, and the Molecular Evolution Core Laboratory at the

Parker H Petit Institute for Bioengineering and Bioscience at the Georgia Institute of Technology for the use of their shared equipment, services, and expertise. This work was supported in part by NIH R35 GM139594 and the Nelson and Bennie Abell Professorship to PTM, R01GM144560 to JTS, and U.S. National Science Foundation (DEB-1830753) to TDK.

Additional information

Funding

Funder	Grant reference number	Author
National Institutes of Health	R35GM139594	Patrick T McGrath
National Institutes of Health	R01GM144560	Jeffrey Todd Streelman
U.S. National Science Foundation	DEB-1830753	Thomas D Kocher

The funders had no role in study design, data collection and interpretation, or the decision to submit the work for publication.

Author contributions

Nikesh M Kumar, Conceptualization, Data curation, Software, Formal analysis, Validation, Investigation, Visualization, Methodology, Writing – original draft, Writing – review and editing; Taylor L Cooper, Investigation, Methodology, Writing – original draft, Writing – review and editing; Thomas D Kocher, Data curation, Investigation, Writing – original draft, Writing – review and editing, Sequencing Funding; Jeffrey Todd Streelman, Conceptualization, Writing – original draft, Writing – review and editing; Patrick T McGrath, Conceptualization, Resources, Data curation, Software, Formal analysis, Supervision, Funding acquisition, Validation, Investigation, Visualization, Methodology, Writing – original draft, Project administration, Writing – review and editing

Author ORCIDs

Nikesh M Kumar https://orcid.org/0000-0003-1932-3910
Taylor L Cooper https://orcid.org/0000-0002-7735-6545
Thomas D Kocher https://orcid.org/0000-0002-7547-0133
Jeffrey Todd Streelman https://orcid.org/0000-0003-0057-5069
Patrick T McGrath https://orcid.org/0000-0002-1598-3746

Ethics

Animal handling and sample collection for this study were conducted under approved Institutional Animal Care and Use Committee (IACUC) protocols (A100029 and A100569) at the Georgia Institute of Technology. Fish were sampled under anesthesia with tricaine (MS-222) according to these approved procedures. Every effort was made to minimize discomfort and maximize ease of recovery following procedures. Wild-caught individuals used in the study were acquired through the aquarium trade via the Georgia Tech Department of Animal Resources approved vendors.

Peer review material

Version of Record:
Reviewer #2 (Public review): https://doi.org/10.7554/eLife.104923.3.sa1
Author response https://doi.org/10.7554/eLife.104923.3.sa2

Reviewed preprint v2
Reviewer #2 (Public review): https://doi.org/10.7554/eLife.104923.2.sa1
Author response: https://doi.org/10.7554/eLife.104923.2.sa0
eLife Assessment https://doi.org/10.7554/eLife.104923.2.sa2

Reviewed preprint v1
Reviewer #1 (Public review): https://doi.org/10.7554/eLife.104923.1.sa3

Reviewer #2 (Public review): https://doi.org/10.7554/eLife.104923.1.sa2
Reviewer #3 (Public review): https://doi.org/10.7554/eLife.104923.1.sa1
Author response: https://doi.org/10.7554/eLife.104923.1.sa0
eLife Assessment https://doi.org/10.7554/eLife.104923.1.sa4

Additional files

Supplementary files

Supplementary file 1. Genome statistics for the M_zebra_UMD2a and M_zebra_GT3a reference genomes. Per Linkage Group (LG; chromosome) anchored nucleotide content is compared between the M_zebra_UMD2a and M_zebra_GT3a genome assemblies.

Supplementary file 2. Sample Metadata. Metadata for all samples used in this study. Where known, sample sex, parents, ecogroup, species name, whether the sample was lab-reared, and average whole-genome sequencing coverage are given. Columns F–I indicate whether a sample was included in the analysis that produced the referenced figure. Columns L–R give inversion haplotypes for each of the six inversions, with the tandem inversions on chromosome 20 separated into the first (20 a) and second (20b) regions. Column S indicates which samples were used to generate the initial principal components onto which the remaining samples were projected.

Supplementary file 3. Summary statistics for additional de novo genome assemblies for *Metriaclima zebra* and *Aulonocara sp. 'chitande type north' Nkhata Bay*. Assembly statistics for the five scaffold-level de novo assemblies generated from PacBio HiFi reads and Bionano optical maps. Two are from *Metriaclima zebra* (MZ_GT3.2, MZ_GT3.3) and three from *Aulonocara sp. 'chitande type north' Nkhata Bay* (YH_GT1.1, YH_GT1.2, YH_GT1.3). The number of scaffolds, total assembly size (bp), and N50 (bp) are given for each. Assembly GenBank accessions are listed in the Data Availability section.

Supplementary file 4. Estimates of breakpoint locations in M_zebra_GT3a. Breakpoint intervals for the six large inversions are given in M_zebra_GT3a coordinates (bp). Interval estimates were calculated using whole genome alignments between *Aulonocara sp. 'chitande type north' Nkhata Bay* and the M_zebra_GT3a genome. The breakpoint intervals for the tandem inversions on chromosomes 11 and 20 are separated by the first (11 a/20 a) and second (11b/20b) inversions.

MDAR checklist

Data availability

All Illumina and PacBio sequencing reads have been deposited to the NCBI Sequence Read Archive at BioProject PRJNA1112855. The master VCF file used in these analyses is available through the Dryad Digital Repository (https://doi.org/10.5061/dryad.z08kprrx1). The M_Zebra_GT3a assembly is available via GenBank Accession JBEVYI000000000 (version JBEVYI010000000). A RefSeq genome annotation for the M_Zebra_GT3a assembly is available via NCBI Datasets accession GCF_041146795.1. The five additional scaffold-level assemblies have been deposited at DDBJ/ENA/GenBank with the following BioSample and WGS accessions: MZ_GT3.2 (SAMN44293007; JBRIJA000000000), MZ_GT3.3 (SAMN44293008; JBRIIZ000000000), YH_GT1.1 (SAMN44293009; JBRIIY000000000), YH_GT1.2 (SAMN44293010; JBRIIX000000000), and YH_GT1.3 (SAMN44293011; JBRIIW000000000). The scripts used for variant calling, PCA, phylogenetic, F_{ST}, and nucleotide-diversity analyses are available at https://github.com/ptmcgrat/CichlidSRSequencing/tree/Kumar_eLife, copy archived at ***McGrath, 2026***.

Continued on next page

The following datasets were generated:

Author(s)	Year	Dataset title	Dataset URL	Database and Identifier
Kumar NM, Cooper TL, Kocher TD, Streelman JT, McGrath PT	2026	Large inversions in Lake Malawi cichlids are associated with habitat, lineage, and sex determination	https://doi.org/10.5061/dryad.z08kprrx1	Dryad Digital Repository, 10.5061/dryad.z08kprrx1
Kumar NM, Cooper TL, Kocher TD, Streelman JT, McGrath PT	2024	Understanding the role of inversions in Lake Malawi Cichlids	https://www.ncbi.nlm.nih.gov/bioproject/PRJNA1112855	NCBI BioProject, PRJNA1112855

References

Andolfatto P, Depaulis F, Navarro A. 2001. Inversion polymorphisms and nucleotide variability in Drosophila. *Genetical Research* **77**:1–8. DOI: https://doi.org/10.1017/s0016672301004955, PMID: 11279826

Behrens KA, Koblmueller S, Kocher TD. 2024. Diversity of sex chromosomes in vertebrates: six novel sex chromosomes in basal haplochromines (Teleostei: Cichlidae). *Genome Biology and Evolution* **16**:evae152. DOI: https://doi.org/10.1093/gbe/evae152, PMID: 39073759

Berdan EL, Barton NH, Butlin R, Charlesworth B, Faria R, Fragata I, Gilbert KJ, Jay P, Kapun M, Lotterhos KE, Mérot C, Durmaz Mitchell E, Pascual M, Peichel CL, Rafajlović M, Westram AM, Schaeffer SW, Johannesson K, Flatt T. 2023. How chromosomal inversions reorient the evolutionary process. *Journal of Evolutionary Biology* **36**:1761–1782. DOI: https://doi.org/10.1111/jeb.14242, PMID: 37942504

Blumer LM, Burskaia V, Artiushin I, Saha J, Garcia JC, Jiménez FC, Hooft van Huysdynen A, Elkin J, Fischer B, Van Houtte N, Zhou C, Gresham S, Malinsky M, Linderoth T, Sawasawa W, Vernaz G, Bista I, Hickey A, Kucka M, Louzada S, et al. 2025. Introgression dynamics of sex-linked chromosomal inversions shape the Malawi cichlid radiation. *Science* **388**:eadr9961. DOI: https://doi.org/10.1126/science.adr9961, PMID: 40504893

Bocklandt S, Hastie A, Cao H. 2019. Bionano genome mapping: high-throughput, ultra-long molecule genome analysis system for precision genome assembly and haploid-resolved structural variation discovery. In: Suzuki Y (Ed). *Single Molecule and Single Cell Sequencing*. Singapore: Springer. p. 97–118. DOI: https://doi.org/10.1007/978-981-13-6037-4_7

Brawand D, Wagner CE, Li YI, Malinsky M, Keller I, Fan S, Simakov O, Ng AY, Lim ZW, Bezault E, Turner-Maier J, Johnson J, Alcazar R, Noh HJ, Russell P, Aken B, Alföldi J, Amemiya C, Azzouzi N, Baroiller J-F, et al. 2014. The genomic substrate for adaptive radiation in African cichlid fish. *Nature* **513**:375–381. DOI: https://doi.org/10.1038/nature13726, PMID: 25186727

Cabanettes F, Klopp C. 2018. D-GENIES: dot plot large genomes in an interactive, efficient and simple way. *PeerJ* **6**:e4958. DOI: https://doi.org/10.7717/peerj.4958, PMID: 29888139

Chang CC, Chow CC, Tellier LC, Vattikuti S, Purcell SM, Lee JJ. 2015. Second-generation PLINK: rising to the challenge of larger and richer datasets. *GigaScience* **4**:7. DOI: https://doi.org/10.1186/s13742-015-0047-8, PMID: 25722852

Chen Y, Zhang Y, Wang AY, Gao M, Chong Z. 2021. Accurate long-read de novo assembly evaluation with Inspector. *Genome Biology* **22**:312. DOI: https://doi.org/10.1186/s13059-021-02527-4, PMID: 34775997

Ciezarek AG, Mehta TK, Man A, Ford AGP, Kavembe GD, Kasozi N, Ngatunga BP, Shechonge AH, Tamatamah R, Nyingi DW, Cnaani A, Ndiwa TC, Di Palma F, Turner GF, Genner MJ, Haerty W. 2024. Ancient and recent hybridization in the oreochromis cichlid fishes. *Molecular Biology and Evolution* **41**:msae116. DOI: https://doi.org/10.1093/molbev/msae116, PMID: 38865496

Conte MA, Joshi R, Moore EC, Nandamuri SP, Gammerdinger WJ, Roberts RB, Carleton KL, Lien S, Kocher TD. 2019. Chromosome-scale assemblies reveal the structural evolution of African cichlid genomes. *GigaScience* **8**:giz030. DOI: https://doi.org/10.1093/gigascience/giz030, PMID: 30942871

Faria R, Johannesson K, Butlin RK, Westram AM. 2019. Evolving inversions. *Trends in Ecology & Evolution* **34**:239–248. DOI: https://doi.org/10.1016/j.tree.2018.12.005, PMID: 30691998

Feulner PGD, Schwarzer J, Haesler MP, Meier JI, Seehausen O. 2018. A dense linkage map of lake victoria cichlids improved the *Pundamilia* genome assembly and revealed a major QTL for sex-determination. *G3: Genes, Genomes, Genetics* **8**:2411–2420. DOI: https://doi.org/10.1534/g3.118.200207, PMID: 29760203

Fishman L, Stathos A, Beardsley PM, Williams CF, Hill JP. 2013. Chromosomal rearrangements and the genetics of reproductive barriers in mimulus (monkey flowers). *Evolution; International Journal of Organic Evolution* **67**:2547–2560. DOI: https://doi.org/10.1111/evo.12154, PMID: 24033166

Fuller ZL, Leonard CJ, Young RE, Schaeffer SW, Phadnis N. 2018. Ancestral polymorphisms explain the role of chromosomal inversions in speciation. *PLOS Genetics* **14**:e1007526. DOI: https://doi.org/10.1371/journal.pgen.1007526, PMID: 30059505

Fuller ZL, Koury SA, Phadnis N, Schaeffer SW. 2019. How chromosomal rearrangements shape adaptation and speciation: Case studies in *Drosophila* pseudoobscura and its sibling species *Drosophila* persimilis. *Molecular Ecology* **28**:1283–1301. DOI: https://doi.org/10.1111/mec.14923, PMID: 30402909

Hager ER, Harringmeyer OS, Wooldridge TB, Theingi S, Gable JT, McFadden S, Neugeboren B, Turner KM, Jensen JD, Hoekstra HE. 2022. A chromosomal inversion contributes to divergence in multiple traits between deer mouse ecotypes. *Science* **377**:399–405. DOI: https://doi.org/10.1126/science.abg0718, PMID: 35862520

Harringmeyer OS, Hoekstra HE. 2022. Chromosomal inversion polymorphisms shape the genomic landscape of deer mice. *Nature Ecology & Evolution* **6**:1965–1979. DOI: https://doi.org/10.1038/s41559-022-01890-0, PMID: 36253543

Hoffmann AA, Rieseberg LH. 2008. Revisiting the impact of inversions in evolution: from population genetic markers to drivers of adaptive shifts and speciation? *Annual Review of Ecology, Evolution, and Systematics* **39**:21–42. DOI: https://doi.org/10.1146/annurev.ecolsys.39.110707.173532, PMID: 20419035

Huang K, Andrew RL, Owens GL, Ostevik KL, Rieseberg LH. 2020. Multiple chromosomal inversions contribute to adaptive divergence of a dune sunflower ecotype. *Molecular Ecology* **29**:2535–2549. DOI: https://doi.org/10.1111/mec.15428, PMID: 32246540

Huerta-Cepas J, Serra F, Bork P. 2016. ETE 3: reconstruction, analysis, and visualization of phylogenomic data. *Molecular Biology and Evolution* **33**:1635–1638. DOI: https://doi.org/10.1093/molbev/msw046, PMID: 26921390

Hughes JF, Skaletsky H, Pyntikova T, Graves TA, van Daalen SKM, Minx PJ, Fulton RS, McGrath SD, Locke DP, Friedman C, Trask BJ, Mardis ER, Warren WC, Repping S, Rozen S, Wilson RK, Page DC. 2010. Chimpanzee and human Y chromosomes are remarkably divergent in structure and gene content. *Nature* **463**:536–539. DOI: https://doi.org/10.1038/nature08700, PMID: 20072128

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**:11895–11900. DOI: https://doi.org/10.1073/pnas.1611028113, PMID: 27698127

Johnson NA, Lachance J. 2012. The genetics of sex chromosomes: evolution and implications for hybrid incompatibility. *Annals of the New York Academy of Sciences* **1256**:E1–E22. DOI: https://doi.org/10.1111/j.1749-6632.2012.06748.x, PMID: 23025408

Johnson ZV, Hegarty BE, Gruenhagen GW, Lancaster TJ, McGrath PT, Streelman JT. 2023. Cellular profiling of a recently-evolved social behavior in cichlid fishes. *Nature Communications* **14**:4891. DOI: https://doi.org/10.1038/s41467-023-40331-9, PMID: 37580322

Jones FC, Grabherr MG, Chan YF, Russell P, Mauceli E, Johnson J, Swofford R, Pirun M, Zody MC, White S, Birney E, Searle S, Schmutz J, Grimwood J, Dickson MC, Myers RM, Miller CT, Summers BR, Knecht AK, Brady SD, et al. 2012. The genomic basis of adaptive evolution in threespine sticklebacks. *Nature* **484**:55–61. DOI: https://doi.org/10.1038/nature10944, PMID: 22481358

Joron M, Papa R, Beltrán M, Chamberlain N, Mavárez J, Baxter S, Abanto M, Bermingham E, Humphray SJ, Rogers J, Beasley H, Barlow K, ffrench-Constant RH, Mallet J, McMillan WO, Jiggins CD. 2006. A conserved supergene locus controls colour pattern diversity in Heliconius butterflies. *PLOS Biology* **4**:e303. DOI: https://doi.org/10.1371/journal.pbio.0040303, PMID: 17002517

Joron M, Frezal L, Jones RT, Chamberlain NL, Lee SF, Haag CR, Whibley A, Becuwe M, Baxter SW, Ferguson L, Wilkinson PA, Salazar C, Davidson C, Clark R, Quail MA, Beasley H, Glithero R, Lloyd C, Sims S, Jones MC, et al. 2011. Chromosomal rearrangements maintain a polymorphic supergene controlling butterfly mimicry. *Nature* **477**:203–206. DOI: https://doi.org/10.1038/nature10341, PMID: 21841803

Kirkpatrick M, Barton N. 2006. Chromosome inversions, local adaptation and speciation. *Genetics* **173**:419–434. DOI: https://doi.org/10.1534/genetics.105.047985, PMID: 16204214

Kirkpatrick M. 2010. How and why chromosome inversions evolve. *PLOS Biology* **8**:e1000501. DOI: https://doi.org/10.1371/journal.pbio.1000501, PMID: 20927412

Kocher TD. 2004. Adaptive evolution and explosive speciation: the cichlid fish model. *Nature Reviews. Genetics* **5**:288–298. DOI: https://doi.org/10.1038/nrg1316, PMID: 15131652

Kratochwil CF, Liang Y, Urban S, Torres-Dowdall J, Meyer A. 2019. Evolutionary dynamics of structural variation at a key locus for color pattern diversification in cichlid fishes. *Genome Biology and Evolution* **11**:3452–3465. DOI: https://doi.org/10.1093/gbe/evz261, PMID: 31821504

Küpper C, Stocks M, Risse JE, Dos Remedios N, Farrell LL, McRae SB, Morgan TC, Karlionova N, Pinchuk P, Verkuil YI, Kitaysky AS, Wingfield JC, Piersma T, Zeng K, Slate J, Blaxter M, Lank DB, Burke T. 2016. A supergene determines highly divergent male reproductive morphs in the ruff. *Nature Genetics* **48**:79–83. DOI: https://doi.org/10.1038/ng.3443, PMID: 26569125

Lamichhaney S, Fan G, Widemo F, Gunnarsson U, Thalmann DS, Hoeppner MP, Kerje S, Gustafson U, Shi C, Zhang H, Chen W, Liang X, Huang L, Wang J, Liang E, Wu Q, Lee SM-Y, Xu X, Höglund J, Liu X, et al. 2016. Structural genomic changes underlie alternative reproductive strategies in the ruff (Philomachus pugnax). *Nature Genetics* **48**:84–88. DOI: https://doi.org/10.1038/ng.3430, PMID: 26569123

Letunic I, Bork P. 2024. Interactive Tree of Life (iTOL) v6: recent updates to the phylogenetic tree display and annotation tool. *Nucleic Acids Research* **52**:W78–W82. DOI: https://doi.org/10.1093/nar/gkae268, PMID: 38613393

Li H, Durbin R. 2009. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* **25**:1754–1760. DOI: https://doi.org/10.1093/bioinformatics/btp324, PMID: 19451168

Li H. 2011. A statistical framework for SNP calling, mutation discovery, association mapping and population genetical parameter estimation from sequencing data. *Bioinformatics* **27**:2987–2993. DOI: https://doi.org/10.1093/bioinformatics/btr509, PMID: 21903627

Li H. 2018. Minimap2: pairwise alignment for nucleotide sequences. *Bioinformatics* **34**:3094–3100. DOI: https://doi.org/10.1093/bioinformatics/bty191, PMID: 29750242

Loh Y-HE, Bezault E, Muenzel FM, Roberts RB, Swofford R, Barluenga M, Kidd CE, Howe AE, Di Palma F, Lindblad-Toh K, Hey J, Seehausen O, Salzburger W, Kocher TD, Streelman JT. 2013. Origins of shared genetic variation in African cichlids. *Molecular Biology and Evolution* **30**:906–917. DOI: https://doi.org/10.1093/molbev/mss326, PMID: 23275489

Lyons RP, Scholz CA, Cohen AS, King JW, Brown ET, Ivory SJ, Johnson TC, Deino AL, Reinthal PN, McGlue MM, Blome MW. 2015. Continuous 1.3-million-year record of East African hydroclimate, and implications for patterns of evolution and biodiversity. *PNAS* **112**:15568–15573. DOI: https://doi.org/10.1073/pnas.1512864112, PMID: 26644580

Maan ME, Sefc KM. 2013. Colour variation in cichlid fish: developmental mechanisms, selective pressures and evolutionary consequences. *Seminars in Cell & Developmental Biology* **24**:516–528. DOI: https://doi.org/10.1016/j.semcdb.2013.05.003, PMID: 23665150

Malinsky M, Salzburger W. 2016. Environmental context for understanding the iconic adaptive radiation of cichlid fishes in Lake Malawi. *PNAS* **113**:11654–11656. DOI: https://doi.org/10.1073/pnas.1614272113, PMID: 27791049

Malinsky M, Svardal H, Tyers AM, Miska EA, Genner MJ, Turner GF, Durbin R. 2018. Whole-genome sequences of Malawi cichlids reveal multiple radiations interconnected by gene flow. *Nature Ecology & Evolution* **2**:1940–1955. DOI: https://doi.org/10.1038/s41559-018-0717-x, PMID: 30455444

Martin CH, Genner MJ. 2009. A role for male bower size as an intrasexual signal in a Lake Malawi cichlid fish. *Behaviour* **146**:963–978. DOI: https://doi.org/10.1163/156853908X396836

McGrath PT. 2026. CichlidSRSequencing. swh:1:rev:7d542e44c797a8b0c0343033ea4775ee13a427d5. Software Heritage. https://archive.softwareheritage.org/swh:1:dir:979bec32618f9541faa43b6e68b82008a2048000;origin=https://github.com/ptmcgrat/CichlidSRSequencing;visit=swh:1:snp:c797eac806ceb2e8957b5b7233b904087d0db364;anchor=swh:1:rev:7d542e44c797a8b0c0343033ea4775ee13a427d5

McKaye KR, Stauffer JR, Turner GF, Konings A, Sato T. 2001. Fishes, as well as birds, build bowers. *Journal of Aquariculture and Aquatic Sciences* **9**:121–128.

Meier JI, McGee MD, Marques DA, Mwaiko S, Kishe M, Wandera S, Neumann D, Mrosso H, Chapman LJ, Chapman CA, Kaufman L, Taabu-Munyaho A, Wagner CE, Bruggmann R, Excoffier L, Seehausen O. 2023. Cycles of fusion and fission enabled rapid parallel adaptive radiations in African cichlids. *Science* **381**:eade2833. DOI: https://doi.org/10.1126/science.ade2833, PMID: 37769075

Meisel RP. 2022. Ecology and the evolution of sex chromosomes. *Journal of Evolutionary Biology* **35**:1601–1618. DOI: https://doi.org/10.1111/jeb.14074, PMID: 35950939

Miles A. 2024. Cggh/scikit-allel. v1.3.13. Zenodo. https://doi.org/10.5281/zenodo.13772087

Minh BQ, Schmidt HA, Chernomor O, Schrempf D, Woodhams MD, von Haeseler A, Lanfear R. 2020. IQ-TREE 2: new models and efficient methods for phylogenetic inference in the genomic era. *Molecular Biology and Evolution* **37**:1530–1534. DOI: https://doi.org/10.1093/molbev/msaa015, PMID: 32011700

Natri HM, Merilä J, Shikano T. 2019. The evolution of sex determination associated with a chromosomal inversion. *Nature Communications* **10**:145. DOI: https://doi.org/10.1038/s41467-018-08014-y, PMID: 30635564

Noor MA, Grams KL, Bertucci LA, Reiland J. 2001. Chromosomal inversions and the reproductive isolation of species. *PNAS* **98**:12084–12088. DOI: https://doi.org/10.1073/pnas.221274498, PMID: 11593019

Nowling RJ, Manke KR, Emrich SJ. 2020. Detecting inversions with PCA in the presence of population structure. *PLOS ONE* **15**:e0240429. DOI: https://doi.org/10.1371/journal.pone.0240429, PMID: 33119626

Ortiz EM. 2019. Vcf2phylip v2.0: convert a VCF matrix into several matrix formats for phylogenetic analysis. Zenodo. https://doi.org/10.5281/zenodo.2540861

Palmer DH, Kronforst MR. 2020. A shared genetic basis of mimicry across swallowtail butterflies points to ancestral co-option of doublesex. *Nature Communications* **11**:6. DOI: https://doi.org/10.1038/s41467-019-13859-y, PMID: 31900419

Parnell NF, Streelman JT. 2013. Genetic interactions controlling sex and color establish the potential for sexual conflict in Lake Malawi cichlid fishes. *Heredity* **110**:239–246. DOI: https://doi.org/10.1038/hdy.2012.73, PMID: 23092997

Patil C, Sylvester JB, Abdilleh K, Norsworthy MW, Pottin K, Malinsky M, Bloomquist RF, Johnson ZV, McGrath PT, Streelman JT. 2021. Genome-enabled discovery of evolutionary divergence in brains and behavior. *Scientific Reports* **11**:13016. DOI: https://doi.org/10.1038/s41598-021-92385-8, PMID: 34155279

Poplin R, Ruano-Rubio V, DePristo MA, Fennell TJ, Carneiro MO, Van der Auwera GA, Kling DE, Gauthier LD, Levy-Moonshine A, Roazen D, Shakir K, Thibault J, Chandran S, Whelan C, Lek M, Gabriel S, Daly MJ, Neale B, MacArthur DG, Banks E. 2018. Scaling accurate genetic variant discovery to tens of thousands of samples. *bioRxiv*. DOI: https://doi.org/10.1101/201178

Santos ME, Lopes JF, Kratochwil CF. 2023. East African cichlid fishes. *EvoDevo* **14**:1. DOI: https://doi.org/10.1186/s13227-022-00205-5, PMID: 36604760

Schelkunov MI. 2023. Mabs, a suite of tools for gene-informed genome assembly. *BMC Bioinformatics* **24**:377. DOI: https://doi.org/10.1186/s12859-023-05499-3, PMID: 37794322

Scott MF, Osmond MM, Otto SP. 2018. Haploid selection, sex ratio bias, and transitions between sex-determining systems. *PLOS Biology* **16**:e2005609. DOI: https://doi.org/10.1371/journal.pbio.2005609, PMID: 29940019

Streelman JT, Danley PD. 2003. The stages of vertebrate evolutionary radiation. *Trends in Ecology & Evolution* **18**:126–131. DOI: https://doi.org/10.1016/S0169-5347(02)00036-8

Svardal H, Salzburger W, Malinsky M. 2021. Genetic variation and hybridization in evolutionary radiations of cichlid fishes. *Annual Review of Animal Biosciences* **9**:55–79. DOI: https://doi.org/10.1146/annurev-animal-061220-023129, PMID: 33197206

Thompson MJ, Jiggins CD. 2014. Supergenes and their role in evolution. *Heredity* **113**:1–8. DOI: https://doi.org/10.1038/hdy.2014.20, PMID: 24642887

Tuttle EM, Bergland AO, Korody ML, Brewer MS, Newhouse DJ, Minx P, Stager M, Betuel A, Cheviron ZA, Warren WC, Gonser RA, Balakrishnan CN. 2016. Divergence and functional degradation of a sex chromosome-like supergene. *Current Biology* **26**:344–350. DOI: https://doi.org/10.1016/j.cub.2015.11.069, PMID: 26804558

Villoutreix R, Ayala D, Joron M, Gompert Z, Feder JL, Nosil P. 2021. Inversion breakpoints and the evolution of supergenes. *Molecular Ecology* **30**:2738–2755. DOI: https://doi.org/10.1111/mec.15907, PMID: 33786937

Wang J, Wurm Y, Nipitwattanaphon M, Riba-Grognuz O, Huang Y-C, Shoemaker D, Keller L. 2013. A Y-like social chromosome causes alternative colony organization in fire ants. *Nature* **493**:664–668. DOI: https://doi.org/10.1038/nature11832, PMID: 23334415

Wellenreuther M, Bernatchez L. 2018. Eco-evolutionary genomics of chromosomal inversions. *Trends in Ecology & Evolution* **33**:427–440. DOI: https://doi.org/10.1016/j.tree.2018.04.002, PMID: 29731154

York RA, Patil C, Abdilleh K, Johnson ZV, Conte MA, Genner MJ, McGrath PT, Fraser HB, Fernald RD, Streelman JT. 2018. Behavior-dependent *cis* regulation reveals genes and pathways associated with bower building in cichlid fishes. *PNAS* **115**:E11081–E11090. DOI: https://doi.org/10.1073/pnas.1810140115, PMID: 30397142