Syst. Biol. 71(3):741–757, 2022

DOI:10.1093/sysbio/syac006
Advance Access publication February 8, 2022

Between a Rock and a Hard Polytomy: Phylogenomics of the Rock-Dwelling Mbuna Cichlids of Lake Malaŵi

MARK D. SCHERZ[1,2], PAUL MASONICK[1], AXEL MEYER[1,*], AND C. DARRIN HULSEY[1,3,*]

[1]*Department of Biology, University of Konstanz, 78464 Konstanz, Germany;* [2]*Natural History Museum of Denmark, University of Copenhagen, 2100 Copenhagen Ø, Denmark; and* [3]*School of Biology and Environmental Science, University College Dublin, Belfield, Dublin 4, Ireland*

[*]*Correspondence to be sent to: Department of Biology, University of Konstanz, 78464 Konstanz, Germany; E-mail: axel.meyer@uni-konstanz.de; darrin.hulsey1@ucd.ie*

Received 22 March 2021; reviews returned 18 January 2022; accepted 30 January 2022
Associate Editor: Josef Uyeda

Abstract.—Whole genome sequences are beginning to revolutionize our understanding of phylogenetic relationships. Yet, even whole genome sequences can fail to resolve the evolutionary history of the most rapidly radiating lineages, where incomplete lineage sorting, standing genetic variation, introgression, and other factors obscure the phylogenetic history of the group. To overcome such challenges, one emerging strategy is to integrate results across different methods. Most such approaches have been implemented on reduced representation genomic data sets, but whole genomes should provide the maximum possible evidence approach. Here, we test the ability of single nucleotide polymorphisms extracted from whole genome resequencing data, implemented in an integrative genomic approach, to resolve key nodes in the phylogeny of the mbuna, rock-dwelling cichlid fishes of Lake Malaŵi, which epitomize the phylogenetic intractability that often accompanies explosive lineage diversification. This monophyletic radiation has diversified at an unparalleled rate into several hundred species in less than 2 million years. Using an array of phylogenomic methods, we consistently recovered four major clades of mbuna, but a large basal polytomy among them. Although introgression between clades apparently contributed to the challenge of phylogenetic reconstruction, reduction of the data set to nonintrogressed sites still did not help to resolve the basal polytomy. On the other hand, relationships among six congeneric species pairs were resolved without ambiguity, even in one case where existing data led us to predict that resolution would be difficult. We conclude that the bursts of diversification at the earliest stages of the mbuna radiation may be phylogenetically unresolvable, but other regions of the tree are phylogenetically clearly supported. Integration of multiple phylogenomic approaches will continue to increase confidence in relationships inferred from these and other whole-genome data sets. [Incomplete lineage sorting; introgression; linkage disequilibrium; multispecies coalescence; rapid radiation; soft polytomy.]

Complete genomes offer the tantalizing promise of resolving even the most recalcitrant phylogenetic relationships. However, rapid radiations with extensive incomplete lineage sorting (ILS) pose challenges due to their high gene-tree discordance (Edwards 2009). This results in unresolved or poorly resolved nodes in species trees. This is commonplace, and reconstructions of old radiations almost invariably contain one or more nodes that defy resolution, for example, in all major tetrapod groups (Irisarri and Meyer 2016; Suh 2016; Irisarri et al. 2017; Moreira and Schrago 2018; Braun et al. 2019; Hime et al. 2021; Singhal et al. 2021), and various plant groups (Pease et al. 2018; Smith et al. 2020; Cai et al. 2021; Gagnon et al. 2021; Morales-Briones et al. 2021). Although these problems often plague ancient radiations where they can be compounded by other effects such as saturation (e.g., Morales-Briones et al. 2021), polytomies can be just as pervasive in young, explosive radiations, and the problem is likely exacerbated when there is the opportunity for extensive gene flow (Koblmüller et al. 2010; Malinsky et al. 2018).

The cichlid fish radiations of the African Great Lakes epitomize recent and explosive adaptive radiations with gene-flow. These radiations have generated thousands of species in less than 10 million years (Meyer et al. 1990; Meyer 1993; Kornfield and Smith 2000; Seehausen 2006; Brawand et al. 2014; Henning and Meyer 2014; Malinsky et al. 2018) and have a high potential for hybridization (Kocher et al. 1995; Hulsey et al. 2010; Mims et al. 2010; Brawand et al. 2014; Genner et al. 2015; Svardal et al. 2021). As a result of their rapid diversification and frequent opportunity for introgressive hybridization, it is unclear whether the genomes of these fishes contain the information needed to parse their evolutionary history. It may even be impossible to recover a robust phylogenetic hypothesis for much of the diversity of these fishes. Alternatively, it might be that the relationships among the numerous genera and species could readily be placed within a strictly bifurcating and highly supported topology given sufficient genetic data and appropriate phylogenomic methods that take into account the various processes that obfuscate the reconstruction of their evolutionary history. Within this diversity of cichlid fishes, one of the greatest remaining phylogenetic challenges is found in the Lake Malaŵi rock-dwelling cichlids, the mbuna. These consist of 122 currently named species placed in 14 genera (Fricke et al. 2020) and are estimated to include hundreds of undescribed species that all arose in the last 2 million years (Meyer et al. 1990; Kocher et al. 1995; Genner et al. 2015; Schedel et al. 2019).

Previous attempts to reconstruct relationships among Malaŵi cichlids using high-throughput reduced representation sequencing techniques have had limited success. For instance, ultraconserved elements (UCEs) that can reliably isolate the same several hundred

TABLE 1. Known cases of hybridization among mbuna cichlids

Clade Species 1	Species 1	Clade Species 2	Species 2	Reference
A	*Labeotropheus fuelleborni*	A	*Labeotropheus trewavasae*	1[e]
A	*Labeotropheus fuelleborni*	B	*Tropheops* sp. "red cheek"	2[e]
A	*Labeotropheus fuelleborni*	C	*Melanochromis auratus*	3[e]
A	*Labeotropheus fuelleborni*	D	*Maylandia zebra*	4[e], 5[e], 6[e], 7[e], 8[w]
A	*Labeotropheus fuelleborni*	D	*Maylandia zebra*	3[e]
C	*Melanochromis auratus*	D	*Maylandia zebra*	3[e]
D	*Maylandia zebra*	D	*Maylandia mbenji*	9[e]
D	*Maylandia zebra*	D	*Maylandia benetos*	10[n], 11[n]
D	*Cynotilapia afra*	D	*Maylandia zebra*	12[w], 13[w]
D	*Cynotilapia afra*	D	*Chindongo elongatus*	14[e]
D	*Chindongo demasoni*	D	*Pseudotropheus cyaneorhabdos*	15[e]
D	*Maylandia emmiltos*	D	*Maylandia zebra*	16[e]
D	*Maylandia zebra*	D	*Maylandia thapsinogen*	16[e]

[e]Experimental hybridization. [w]Hybridization observed in the wild. [n]*In vitro* hybridization.
Notes: Clade is assigned according to our phylogenomic results, below. We here retain the species assignment of the taxa from the original studies, but note that several may have since undergone taxonomic changes, for example, *Labeotropheus* species.

loci produced incongruent phylogenies among different studies and phylogenetic reconstruction methods (McGee et al. 2016; Hulsey et al. 2017; Irisarri et al. 2017). With a chromosome-level assembly of the Malaŵi mbuna cichlid *Maylandia zebra*, coupled with the ability to resequence genomes with short reads, genome-wide information has now been devoted to the resolution of phylogenetic relationships among Malaŵi cichlids (Brawand et al. 2014; Malinsky et al. 2018; Conte et al. 2019; Urban et al. 2021). Yet, even with these nearly complete genomic data sets, incongruities still remain at several levels.

The power to resolve relationships in an adaptive radiation that has undergone explosive diversification may depend on the stage of the radiation, or rough taxonomic level, at which relationships are examined (Albertson et al. 1999; Streelman and Danley 2003). Many of the mbuna genera and species are readily diagnosable phenotypically (Reinthal and Meyer 1997; Streelman and Danley 2003; Hulsey et al. 2010; Pauers 2010; Hulsey et al. 2013; Kratochwil et al. 2018). Therefore, whole-genome sequencing might be expected to be able to resolve species- and genus-level relationships among much of this radiation despite its rapid timeframe of divergence. However, the earliest components of evolutionary divergence in rapid radiations are often difficult to resolve (Prum et al. 2016; Suh 2016; Moreira and Schrago 2018), and although saturation may play less of a role in younger radiations, other effects may produce similar challenges for their resolution.

As highlighted above, ILS is a major potential source of gene-tree conflict. The ancestral population that gave rise to the mbuna potentially retained substantial genetic polymorphism through multiple early speciation events (Moran and Kornfield 1993). Retained ancestral polymorphisms due to ILS can contribute to substantial gene-tree versus species-tree incongruence in radiations, potentially obscuring or even overriding any molecular phylogenetic signal that arose during their diversification. Methods to efficiently reconstruct species trees in the face of ILS are now well established, making use of coalescent theory by modeling the multispecies coalescent (MSC) process (e.g., Maddison 1997; Rosenberg 2003; Knowles and Carstens 2017; Degnan and Rosenberg 2009; Liu et al. 2010; Bravo et al. 2019; Flouri et al. 2019). At some point, ILS could become insurmountable and render robust phylogenetic reconstructions unachievable, but incorporating ILS into how we model and reconstruct the early branches of explosive radiations like the mbuna will likely help to resolve their phylogeny.

Introgressive hybridization presents another source of gene-tree versus species-tree conflict that can effectively blur phylogenomic signal (Streelman et al. 2004; Mims et al. 2010; Brawand et al. 2014). Hybridization events are departures from strict bifurcation of the species tree, and introduce greater incongruence among gene-trees through the lateral transfer of genes among distinct evolutionary lineages (Braun et al. 2019). Although strict MSC methods account for ILS, they do assume limited introgression among species. Only recently have methods been developed that extend MSC methods to allow for introgression when reconstructing relationships (Degnan 2018; Flouri et al. 2019), but these might not be able to cope with the level of rampant introgression that is assumed to be present in many

cichlid radiations (Irisarri et al. 2018; Malinsky et al. 2018; Salzburger 2018; Svardal et al. 2021).

In the mbuna, introgression is thought to be exceedingly common, because they can often be readily hybridized in the laboratory and have been documented to hybridize in the Lake (Table 1). Opportunities for such introgressive hybridization are plentiful for mbuna cichlids. Many closely related congeneric lineages often coexist in microsympatry (syntopy), in some cases not just locally, but across the entire distribution of the lineages. This is seen frequently in the morphologically unusual genus *Labeotropheus*, which has a uniquely hypertrophied snout and highly derived rectangular jaws (Albertson and Pauers 2019; Conith et al. 2019). At numerous sites around Lake Malaŵi, pairs of deep-bodied and shallow-bodied forms occur that are microendemic to single locations (Ribbink et al. 1983a,b; Konings 2007). In each case, the two forms tend to differ not only in body depth and ecology, but also in male nuptial coloration (Ribbink et al. 1983a,b; Konings 2007; Pauers 2010). For instance, at Thumbi West Island in southern Lake Malaŵi, the shallow-bodied *Labeotropheus trewavasae* and deep-bodied *Labeotropheus artatorostris* (Pauers 2017) occur in syntopy. As these two species are known to produce F1 hybrids in the lab, and have been implicated in hybridization events at Thumbi West Island, they might be particularly difficult to disentangle genetically. Hence, this species pair offers an exemplary system of syntopic lineages in which to test whether whole-genome data can readily differentiate such sympatric, closely related mbuna species.

In addition to elevated ILS and introgressive hybridization, many recent explosive radiations exhibit little genetic differentiation over the majority of the genome, and contain relatively few phylogenetically informative loci (Campbell and Bernatchez 2004; Via and West 2008; Nosil et al. 2009; Brawand et al. 2014; Kautt et al. 2016; Cai et al. 2021). For example, variation in nuclear genes, which tend to be conserved components of the genome, may be relatively uninformative for phylogenetic reconstruction (Mims et al. 2010; Brawand et al. 2014). As a result, only a comparatively small portion of the genome can be expected to hold phylogenetically informative loci. Methods based on sequencing only small fragments of the genome (e.g., RADseq, UCEs) are unlikely to yield many informative loci. Whole-genome resequencing, on the other hand, has the potential to make available all of the phylogenetically informative portions of the genome for analysis.

No single method currently exists to identify and overcome the heterogeneous sources of gene-tree conflict and lack of phylogenetic resolution in rapid radiations. As a result, the emerging approach is to integrate and compare results from multiple methodological approaches, each of which targets specific sources of conflict and disparity (Morales-Briones et al. 2021), often interrogating the data in a node-by-node approach (Singhal et al. 2021). So far, most such approaches have worked with reduced representation sequencing data sets, and not the whole genome of the organisms in question. In this study, we adopt an integrative approach to investigate the power of whole-genome resequencing data to resolve the phylogeny of the exceptionally rapid radiation of Lake Malaŵi rock-dwelling cichlids, taking especially ILS and introgression into account. We demonstrate that the resolving power of even whole-genome level data is limited at basal nodes within such an extremely rapid radiation, but, encouragingly, can provide excellent resolution among genera and even differentiate very closely related syntopic Malaŵi species.

Materials and Methods

Sampling, Genome Resequencing, and Processing

A total of 26 resequenced genomes representing 18 species were included for phylogenomic study (Table S2 of the Supplementary material available on Zenodo at https://doi.org/10.5281/zenodo.5164151). New whole-genome sequences were obtained from 19 adult fish in breeding coloration collected with scuba and barrier nets from Lake Malaŵi in 2010 using permits granted from the Malaŵi Parks Department. High-molecular-weight DNA was extracted from fin or muscle tissue from all individuals using commercial kits (QiaGen Dneasy Blood & Tissue Kit) and included an RNase A treatment step. DNA integrity was manually inspected on agarose gels and concentrations were determined on a QuBit fluorometer. Genomic libraries were prepared using Illumina TruSeq DNA Nano kits (Illumina Inc., San Diego, CA, USA) aiming for 350-bp insert sizes. Genomic libraries were then paired-end sequenced (2×150 bp) on a HiSeq 4000 or HiSeq X-Ten Illumina platform at the Beijing Genomics Institute (BGI, Shenzhen, China). Pooling four to five individuals per lane resulted in an average effective genome coverage (counting only reads with mapping quality $\geq$30, nucleotides with base quality $\geq$20, and no read duplicates) of approximately $20\times$ per individual.

After demultiplexing, we converted raw reads to unmapped BAM files for long-term storage using Picard tools v.2.9.4 (https://broadinstitute.github.io/picard), adding read group information, and marking adaptor sequences in the process (using the FastqToSamMark and Illumina-Adapters modules). Reads were then converted back into fastq format (SamToFastq) for mapping with BWA mem v.0.7.15 to the most recently published *M. zebra* reference genome (GCA_000238955.5: M_zebra_UMD2a of Conte et al. 2019).

New whole-genome assemblies were produced for the following eleven species: *Chindongo* (formerly *Pseudotropheus*) *flavus*, *Labeotropheus artatorostris* (formerly *L. fuelleborni* sensu lato [see Pauers 2017], five individuals), *L. trewavasae* (five individuals), *Labidochromis gigas*, *Labidochromis ianthinus*, *Maylandia xanstomachus*, *M. zebra*, *Melanochromis auratus*, *Melanochromis vermivorus*, *Petrotilapia nigra*, and *Tropheops* sp. aff. *tropheops* "Boadzulu." These new genomes were supplemented by seven previously published genomes (Malinsky et al. 2018) of *Astatotilapia bloyeti* (outgroup),

Haplochromis tweddlei (= *Astatotilapia gigliolii*; Fricke et al. 2020; outgroup), *Cynotilapia axelrodi*, *Genyochromis mento*, *Iodotropheus sprengerae*, *Petrotilapia genalutea*, and *Tropheops tropheops*. New genomes were submitted to NCBI's Short Read Archive and are available under the project number PRJNA783868 at https://www.ncbi.nlm.nih.gov/sra/PRJNA783868. Accession numbers and collection localities are listed for individual samples in Table S2 of the Supplementary material available on Zenodo.

Considering all samples together, we jointly called variants and individual genotypes with freebayes v. 1.1.0 (Garrison and Marth 2012) using default parameters and applying standard quality filters (mapping quality ≥30, base quality ≥20). Subsequently, hard filters were applied using the vcffilter script from the vcflib package (https://github.com/vcflib/vcflib) (-s -f "QUAL > 1 & QUAL/AO > 10 & SAF > 0 & SAR > 0 & RPR > 1 & RPL > 1") to remove low-quality variant sites. Variant representation was normalized using vt norm (Tan et al. 2015). Individual genotype calls based on a read depth smaller than five were set to missing for all downstream analyses. We excluded small scaffolds and restricted analyses to data from the 22 chromosomes only.

After calling single nucleotide polymorphisms (SNPs), VCF files were filtered for the following parameters using vcftools 0.1.15 (Danecek et al. 2011): indels removed, 2 alleles allowed per site, maximally 10% missing data, minimum quality 30, minimum depth 10, and maximum depth 50. After subsequent filtration steps, only variant sites were retained using the -non-ref-ac-any 1 command in vcftools 0.1.15. This filtering resulted in a master VCF file containing 7,421,398 unpruned, genome-wide SNPs. This VCF file was then parsed into separate data sets containing SNPs from noncoding (7,121,933 SNPs) and coding (299,465 SNPs) regions based on CDS annotations in the *M. zebra* reference genome (Conte et al. 2019).

Data Sets

Genomic data sets comprised of two different taxon sampling schemes were analyzed, hereafter referred to as our Species data set and Population data set:

1. **The Species data set** consisted of one genome each of 16 mbuna species (*Chindongo flavus*, *Cynotilapia axelrodi*, *G. mento*, *I. sprengerae*, *L. artatorostris*, *L. trewavasae*, *Labidochromis gigas*, *Labidochromis ianthinus*, *M. xanstomachus*, *M. zebra*, *Melanochromis auratus*, *Melanochromis vermivorus*, *P. genalutea*, *P. nigra*, *Tropheops* sp. aff. *tropheops* "Boadzulu," *Tropheops tropheops*), and the non-Malawi cichlids *A. bloyeti* and *H. tweddlei* as outgroups.
2. **The Population data set** consisted of five genomes each of *L. artatorostris* (formerly the Thumbi Island population of *L. fuelleborni*) and *L. trewavasae*, one *M. zebra*, and one *A. bloyeti* as an outgroup.

These datasets were deposited on Dryad (https://doi.org/10.5061/dryad.ffbg79cvr).

Species Data Set

For some analyses of our Species data set, SNPs were filtered based on linkage disequilibrium (LD) with bcftools 1.11-88 (Li et al. 2009; Danecek et al. 2021). The +prune plugin was used to calculate r^2 values across sites in windows of 500 kb and exclude SNPs in high LD ($r^2 > 0.9$). This reduced the size of the noncoding and coding data sets to 495,087 and 38,104 SNPs, respectively. To maximize signal and reduce potential bias due to selection on coding regions, we focused mostly on the noncoding LD-pruned data set for our analyses.

For some analyses, data sets were phased with Beagle 5.1 (Browning et al. 2018), which implements the method of Browning and Browning (2007). We did not specify a genetic panel, as Beagle does not require one to phase accurately, because it generates its own reference panels during analyses (Browning et al. 2021). Beagle assumed a constant recombination rate of 1 cM per 1 Mb to phase genotypes. VCF files were also converted to nexus and phylip files as needed using the Python script vcf2phylip (Ortiz 2019).

To assess genome-wide LD, we plotted LD decay directly from the master VCF file containing both noncoding and coding SNPs using the program PopLDdecay (Zhang et al. 2019) and observed that the decay of LD (r^2) leveled off to a mean of ∼0.17 at a distance of roughly 25 kb between sites (Fig. S1 of the Supplementary material available on Zenodo). This distance was relevant in the selection of a sliding window size to produce input "gene" trees for our ASTRAL-III analysis (see below).

Phylogenomics.—We compared phylogenomic trees reconstructed using four methods, including one maximum-likelihood-based concatenated approach, and three coalescence-based methods. We considered the consistency of phylogenetic relationships (reproducibility) across independent methods as an indication of phylogenomic hypothesis robustness (Suh 2016). The non-Lake Malawi cichlids *H. tweddlei* and *A. bloyeti* were set as outgroups (Malinsky et al. 2018).

Maximum likelihood concatenated searches were carried out on the noncoding LD-pruned data set in IQ-TREE 1.6.12 (Nguyen et al. 2015) using model finder with ascertainment bias correction (command -m MFP+ASC) (Kalyaanamoorthy et al. 2017), 1000 ultrafast bootstraps (command -bb 1000) (Hoang et al. 2018), and the SH-like approximate likelihood ratio test (SH-aLRT) with 1000 replicates (command -alrt 1000) (Guindon et al. 2010). Site concordance factors were assessed in IQ-TREE 2.0.6 with 100 quartets (command –scf 100) (Minh et al. 2020).

Coalescence species tree inferences were performed using three different approaches:

1. SVDquartets (Chifman and Kubatko 2014) implemented in PAUP* 4.0a (Linux build 166) (Swofford 2002). For this analysis, the noncoding LD-pruned data set was analyzed using exhaustive quartet sampling and 100 bootstrap replicates. The inferred tree was calculated with the QFM algorithm (Reaz et al. 2014) that optimizes the tree based on the maximum quartet consistency.

2. SNAPP (Bryant et al. 2012) implemented in BEAST2 2.6.3 (Bouckaert et al. 2014). As the analysis of large phylogenomic data sets remains computationally prohibitive with SNAPP, we ran analyses across multiple ($n = 10$) SNP data sets that each contained a different sampling of 2200 SNPs ($n = 100$ SNPs randomly sampled from each chromosome of the noncoding LD-pruned data set using bcftools +prune). XML files were generated with the Ruby script snapp_prep.rb (available at https://github.com/mmatschiner/snapp_prep). A starting tree was used as a guide and the ingroup crown age was constrained using a log normal distribution with a mean of 0.5993 and standard deviation of 0.18 based on divergence dating results from Schedel et al. (2019). SNAPP analyses were run in parallel for 5 million generations, sampling every 5000 steps. Log files were examined in Tracer 1.7.1 (Rambaut et al. 2018) and ESS values checked for stationarity across all metrics. After combining all trees sampled from the 10 independent runs in LogCombiner 2.6.3 (discarding a burn-in of 10% from each), a maximum clade credibility tree was produced in TreeAnnotator 2.6.3 assigning median heights at nodes. Maximum clade credibility trees from each of the 10 different runs were also obtained and compared with one another to assess topological convergence between the subsampled data sets.

3. ASTRAL-III 5.6.3 (Zhang et al. 2019) (hereafter "ASTRAL"). For this analysis, we evaluated phased SNPs from the master VCF. Unlike the other three phylogenetic analyses that evaluated the LD-pruned noncoding data set, we here assessed phased SNPs from the genome-wide unpruned VCF file (containing 7,421,398 SNPs). This was done because ASTRAL requires input gene-trees and thus essentially relies on linked data. Maximum likelihood "gene" trees were first generated by conducting RAxML v8.2.12 (Stamatakis 2014) analyses with the GTRCAT model on SNPs that were partitioned into nonoverlapping coordinate windows of 25 kb (–windType coordinate -w 25000) using the Python script raxml_sliding_windows.py (available at https://github.com/simonhmartin/genomics_general). This window size was selected based on the leveling off of the LD decay (r^2) that we observed at approximately this distance (see Fig. S1 of the Supplementary material available on Zenodo). To minimize the generation of suboptimal input trees, maximum likelihood trees were calculated only for windows bearing at least 50 SNPs (-M 50). In total, 26,581 "gene" trees were estimated using this approach. An ASTRAL analysis was then carried out on these trees under default parameters. We also tested for polytomies at deep nodes in the tree with the method of Sayyari and Mirarab (2018), which is based on a Chi-Square test among quartet frequencies for nodes, implemented with the -t 10 command in ASTRAL using these 26,581 trees.

Distance matrices were produced using phased SNPs from the master VCF file and the noncoding and coding LD-pruned data sets with the Python script distMat.py (available at: https://github.com/simonhmartin/genomics_general) (Table S3 of the Supplementary material available on Zenodo). Phylogenetic neighbor networks were then constructed from these three matrices with NeighbourNet in SplitsTree 4.16.1 (Hudson and Bryant 2006) to visualize the general structure of our SNP data (Fig. S2 of the Supplementary material available on Zenodo). We also assessed structure among our ingroup species with genomic principal components analysis (PCA). This was done in plink 1.90 (Chang et al. 2015) using the noncoding LD-pruned data set with the two outgroup taxa removed.

Phylogenetic trees associated with this study are available from TreeBASE at http://purl.org/phylo/treebase/phylows/study/TB2:S29285.

Topology weighting. Based on the robustly supported clades recovered using the tree inference methods outlined above (see Results), we conducted five-clade (outgroup included) and four-clade (outgroup excluded) TWISST (Martin and van Belleghem 2017) analyses to examine whether the basal polytomy within the mbuna was due to regionally conflicting signal across the genome. These clades were those recovered robustly in likelihood and coalescent tree building above. We followed the pipeline of Martin and van Belleghem (2017) using the scripts parseVCF.py, raxml_sliding_windows.py, twisst.py, and plot_twisst.R (available at: https://github.com/simonhmartin/genomics_general/). Phased SNPs from the master VCF file were analyzed (7,421,398 SNPs for the five-clade analysis and 3,546,543 SNPs for the four-clade analysis excluding the outgroups). A window size of 250 SNPs and the GTRCAT model implemented in RAxML (Stamatakis 2014) were used. Exploratory analyses using 50- and 100-SNP windows showed very little difference among results (data not shown). *Haplochromis tweddlei* and *A. bloyeti* were assigned as the outgroup clade in the five-clade TWISST. In total, 29,675 250-SNP windows were analyzed in the five-clade TWISST and 14,175 in the four-clade TWISST. Results were visualized in R 4.0.2 (R Core Team 2014) using modifications of the scripts provided alongside the TWISST github page.

Analysis of introgression. We calculated Patterson's *D*-statistics to assess for introgression across the 16 ingroup mbuna using the ABBA–BABA test as implemented in the program Dsuite (Malinsky et al. 2021). This test is applied to biallelic SNPs across four taxa and assumes a pectinate tree topology typically given as ([{P1,P2},P3],O). The outgroup (O) helps to define the ancestral allele (A) from the derived allele (B) and site patterns (BBAA, ABBA, and BABA) are counted. Under the null model where only ILS is present (i.e., no gene flow, $D = 0$), ABBA-BABA patterns are expected to occur in equal frequency, but a significant divergence from this suggests introgression between P3 and either P1 or P2 (Malinsky et al. 2021). Using the 7,421,398 SNPs from the genome-wide unpruned VCF file, our ASTRAL tree as a guide tree, and *A. bloyeti* set as the outgroup, we assessed all possible three taxon combinations (560 in total) with the Dtrios function. Each trio was ordered

according to the tree. Standard jackknife blocks (×20) were used to determine if the resulting *D*-statistic values differed significantly from zero. To account for multiple tests, *P*-values were adjusted in RStudio 4.0.3 by applying the false discovery rate method of Benjamini and Hochberg (1995) with the stats package (command: p.adjust(p_values, method = "fdr")). An α of 0.01 was applied to conservatively identify statistically significant *D*-statistic values. To visualize species pairwise comparisons of *D*-statistic scores, a heatmap was plotted using the Ruby script plot_d.rb (available at: https://github.com/mmatschiner).

After inferring numerous cases of hybridization between the four main clades (A, B, C, and D), we used the Dinvestigate function in Dsuite (Malinsky et al. 2021) to locate specific genomic regions that show elevated patterns of introgression. To examine introgression across the four main mbuna clades, *D*-statistic and related metrics (i.e., f_d,f_{dM}, and *df*) were calculated. These were estimated across sliding windows of 50 informative SNPs at step intervals of 25 SNPs (the program defaults) for trios (240 in total) that yielded significant genome-wide *D*-statistic values in the Dtrios analysis. The statistics were specifically calculated for trios in which P1 and P2 (i.e., of the ABBA-BABA tests) were assigned to taxa of the same main clade. *Astatotilapia bloyeti* was used as the outgroup (O). We then evaluated the resulting f_{dM} scores (described in Malinsky et al. 2015), a modified version of the f_d statistic that was devised by Martin et al. (2015), to detect introgressed loci. Values of f_{dM} are distributed around zero (no introgression), range from −1 to 1, and quantify the proportion of introgression between P3 and P2 (positive values) and also P3 and P1 (negative values) (Malinsky et al. 2021). We inferred genomic regions to contain introgressed loci if they exhibited f_{dM} scores greater than 0.1 or less than −0.1. These arbitrary thresholds were selected after viewing f_{dM} scores plotted along all chromosomes for several of the trios showing the highest levels of introgression and observing that most values fall within this range and with many obvious narrow peaks extending beyond these thresholds (see Fig. S3 of the Supplementary material available on Zenodo). Taken across all 240 trios, the putative introgressed regions based on these thresholds spanned 21.695% of the 22 chromosomes. The genomic coordinates of these regions were concatenated into a single sorted and merged BED file with bedtools 2.26.0 (Quinlan and Hall 2010), which was subsequently called with bcftools to remove introgressed SNPs from the genome-wide unpruned master and noncoding LD-pruned VCF files. This resulted in data sets that included 6,163,040 and 110,402 SNPs, respectively.

To assess the impact of removal of the putative introgressed SNPs on phylogenetic reconstruction and resolution among clades, SVDquartets and IQ-TREE analyses were conducted once more on the "nonintrogressed" noncoding LD-pruned data set with the same parameters as described above in the Phylogenomics subsection. Given the poor performance of SNAPP in our main analysis and the computational constraints it imposes, this program was not used to analyze a "nonintrogressed" data set. An ASTRAL analysis was also performed using the "nonintrogressed" version of the genome-wide unpruned VCF file following the same workflow as outlined before. 21,844 "gene" trees were estimated with RAxML and then analyzed in ASTRAL.

Population Data Set

The population data set was separated from the master VCF file by removing irrelevant individuals, and then removing all invariant sites (–non-ref-ac-any 1) in VCFtools. The resulting data set consisted of 3,731,859 variant SNPs, of which 1,349,540 SNPs were variable within *Labeotropheus* individuals alone.

Gene flow analyses and topology weighting. We analyzed gene flow among syntopic *Labeotropheus* species using ADMIXTURE 1.3.0 (Alexander and Lange 2011), F_{ST}, and genomic PCA. ADMIXTURE was performed with k=1–5 groups, and the optimal grouping assessed by comparing the cross-validation (CV) error. Weir and Cockerham F_{ST} was estimated with 100 kb windows in vcftools 0.1.15 (Danecek et al. 2021), with negative F_{ST} values curtailed to 0. Genomic PCA was calculated using plink 1.90b6.16 (Chang et al. 2015), retaining 38,314 variants for the 10 individuals.

To more closely assess the possibility that gene flow has occurred between *Labeotropheus* and *M. zebra*, as has been suggested in the past (Mims et al. 2010), we calculated *D*-statistics in Dsuite (Malinsky et al. 2021), and performed a four-clade TWISST (Martin and van Belleghem 2017) analysis as outlined above, with the clades being *L. artatorostris*, *L. trewavasae*, *M. zebra*, as well as *A. bloyeti* included as the outgroup, using 250-SNP windows, not thinned for linkage. This window size was selected for consistency with the Species data set TWISST analysis. In total, 14,915 250-SNP windows were analyzed.

Results

Improved but Still Limited Phylogenomic Resolution at the Base of the Mbuna Cichlids

Using SVDquartets on the LD-pruned data sets, we recovered a more robustly supported tree when analyzing the 495,087 SNP noncoding as compared with the 38,104 SNP coding data set (Fig. S4 of the Supplementary material available on Zenodo). Therefore, we focused further analyses on the noncoding SNPs with IQ-TREE and SNAPP and present those results below. The ASTRAL analysis was conducted using "gene" trees generated from the full set of SNPs (coding

+ noncoding) of the master VCF file. Phylogenomic analyses consistently recovered four main clades within the mbuna: (A) *Labeotropheus*, (B) *Petrotilapia* + *Tropheops*, (C) *Melanochromis* + *Labidochromis* + *Iodotropheus*, and (D) *Genyochromis* + *Cynotilapia* + *Maylandia* + *Chindongo* (Fig. 1). Genomic PCA also separated these clades into four distinct clusters (Fig. S5 of the Supplementary material available on Zenodo). All reconstructions agree in placing *Iodotropheus* within *Labidochromis* with full support. They also agree in placing *Melanochromis* as the sister of these taxa with full support. Ignoring the position of the root, SVDquartets and IQ-TREE recovered the same quartet topology among the four clades of A,D|B,C, whereas SNAPP and ASTRAL recovered the quartet topology A,C|B,D, but with different placements of the root (Fig. 1). The third possible quartet, A,B|C,D, was not recovered by these analyses. SVDquartets and IQ-TREE trees yielded almost identical overall topology, differing only in the monophyly of *Tropheops* species in Clade B (monophyletic in IQ-TREE, paraphyletic in SVDquartets), and in the branching order of *Genyochromis*, *Cynotilapia*, and *Chindongo* in Clade D. The latter three taxa branched in that order in all analyses except IQ-TREE, where *Chindongo* fell sister to *Maylandia*, and the other two (*Genyochromis* + *Cynotilapia*) as a sister clade (Fig. 1).

Labeotropheus (Clade A) was found to be sister to all other clades in ASTRAL, SVDquartets, and IQ-TREE, whereas Clade C was recovered as the earliest diverging mbuna lineage in the SNAPP analysis. Three of the 10 SNAPP runs conducted on the independent sets of 2200 SNPs recovered *Labeotropheus* at the base of the radiation (data not shown). Clade C was recovered as monophyletic in only five of these independent SNAPP analyses. The maximum clade credibility tree found by combining these analyses shows Clade C as monophyletic albeit with poor support. We think the poor resolution captured across the SNAPP analyses may be due to the limited number of SNPs used per run. Despite high ESS values, the 10 analyses failed to individually converge on a common topology, and we therefore view the maximum clade credibility tree obtained from evaluating trees taken across all 10 analyses as potentially spurious.

Sources of Poor Resolution at the Base of the Mbuna Cichlids

We identified several factors that likely contributed to the poor resolution of the nodes at the base of the mbuna tree. First, despite moderately high support, a polytomy could not be rejected for one of the deepest nodes within the mbuna (that at the split of Clades B + D: χ^2 test with the method of Sayyari and Mirarab (2018), $P = 0.2048$) in the ASTRAL tree (Fig. 1). All other nodes across that tree remained robust, rejecting the null hypothesis of polytomy at $\alpha = 0.05$. Second, in our IQ-TREE analysis, site concordance factors were low at several of the basal nodes, although SH-aLRT tests produced values of 100 across all nodes (Fig. 1).

Examining quartet frequencies at the nodes of our ASTRAL tree, we were surprised to find that there were no nodes at which the assumption of no ILS was obviously violated. In all cases, the second and third hypotheses had similar frequencies (under introgression, we would expect unequal frequencies of these alternative hypotheses). To investigate this further, we conducted an analysis of *D*-statistics across all taxa using Dsuite (Malinsky et al. 2021) on SNPs from the master VCF file. Significant *D*-statistic values ($P < 0.01$) potentially indicative of low levels of introgression were inferred in 422 of the 560 possible trios and between 96 unique mbuna species pairs (Fig. 2a, Table S4 of the Supplementary material available on Zenodo). Significant *D*-statistic values ranged from 0.0085 to 0.1834 with a mean of 0.0377. The strongest signal was detected between *C. axelrodi* and *G. mento* where ABBA–BABA tests yielded a *D*-statistic of 0.1834 ($P = 0$, BBAA = 42,859.0, ABBA = 48,277.8, BABA = 33,317.0). *Tropheops tropheops* in particular exhibited many patterns of introgression with taxa from Clades A, C, and D. In line with hybridization that has been previously reported (Table 1), *M. zebra* and the two *Labeotropheus* species also show signal of potential introgression with other mbuna.

A neighbor network constructed from the genomic distances among taxa (Fig. 2b, Fig. S2 and Table S3 of the Supplementary material available on Zenodo) also recovers the four clades outlined above, as well as many of the subclades. The position of *Iodotropheus* within *Labidochromis*, the close relationships of the other genera, and the very deep divide between the *Tropheops* species are all notable and help to explain the inconsistency of their relationships among reconstruction methods. The central knot of this network contains little structure, as already indicated by the lack of consistency and support in these nodes in ML and coalescence reconstructions, and inability to reject a basal polytomy in ASTRAL analyses.

Given the broad extent of hybridization inferred here across the mbuna, we located genomic regions exhibiting elevated patterns of introgression and masked SNPs from these loci in follow-up phylogenomic analyses to gauge their impact on clade resolution. The putative introgressed SNPs that were identified across analysis of 240 trios spanned 21.695% of the 22 chromosomes (165,098,567 of 760,997,612 bp). After removal of these SNPs, the unrooted quartet branching pattern of the main clades persisted in both the SVDquartets and ASTRAL analyses (A,D|B,C vs. A,C|B,D, respectively), but was inconsistent between the IQ-TREE analyses (Fig. 1, Fig. S6). With IQ-TREE, we recovered the quartet topology A,C|B,D from the introgressed-SNP-pruned data set despite observing A,D|B,C in the original analysis. Although the splitting pattern among the main clades remained unresolved after excluding putative introgressed SNPs, support for the monophyly of Clade D drastically improved in subsequent IQ-TREE, SVDquartet, and ASTRAL analyses. Support for Clade B, on the other hand, was reduced in each of these additional analyses.

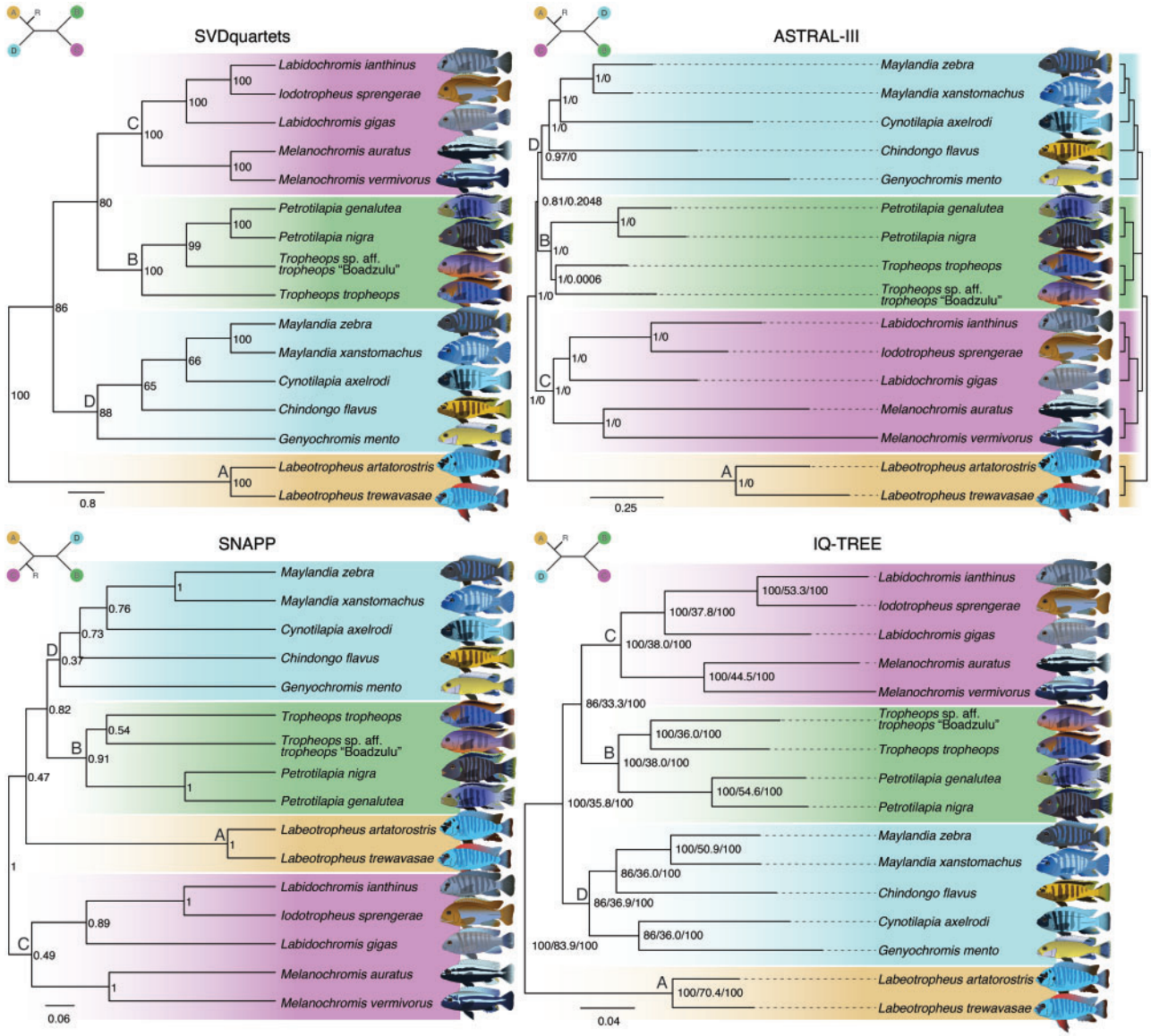

FIGURE 1. Phylogenomic relationships of the mbuna cichlids of Lake Malaŵi, with highly supported, repeatedly recovered clades A, B, C, and D across SVDquartets, SNAPP, ASTRAL-III, and IQ-TREE shaded correspondingly. Inset quartets illustrate the unrooted relationship among mbuna clades recovered by each method, and the position of the root (R). Numbers at nodes correspond to bootstrap support (SVDquartets), posterior probability (SNAPP), local posterior probability/*P*-value of polytomy test (ASTRAL-III), and ultrafast bootstrap support/site concordance factor/SH-aLRT (IQ-TREE). Scale bars correspond to coalescent units (ASTRAL-III), substitutions per site (IQ-TREE), and time in millions of years (SNAPP). To the right of the ASTRAL-III tree is the same phylogeny with the polytomy collapsed based on the polytomy test of Sayyari and Mirarab (2018) (ASTRAL-III). Outgroup removed from all trees for graphical purposes. Figure appears in color in the digital version.

Having recovered four mbuna clades consistently in MSC and concatenated approaches as well as in the distance-based neighbor network, we were interested to know whether specific regions of the mbuna genomes were giving conflicting signals with respect to the deep branches. We investigated this using topology weighting analysis (TWISST) based on these clades and ran the analysis with and without outgroups. Our analysis without outgroups (containing 3,546,543 SNPs) found nearly identical weighting between the two quartets A,C|B,D (the topology found by ASTRAL and SNAPP) and A,B|C,D (the topology not recovered by any of our reconstructions). Both of these were weighted stronger than A,D|B,C (the topology recovered by SVDquartets and IQ-TREE; Fig. 3b; Fig. S7 of the Supplementary material available on Zenodo). There was little indication of particular genomic regionalization of quartet weighting (as exemplified in Fig. 3c). The analysis with outgroups recovered the highest average weighting for the topology R((A,B),(C,D)), slightly over R((A,C),(B,D)) (root here indicated with R) (Fig. 3d; Fig. S8 of the Supplementary material available on Zenodo). These two topologies had substantially higher average weighting than the next best topology (Fig. 3e), but none of

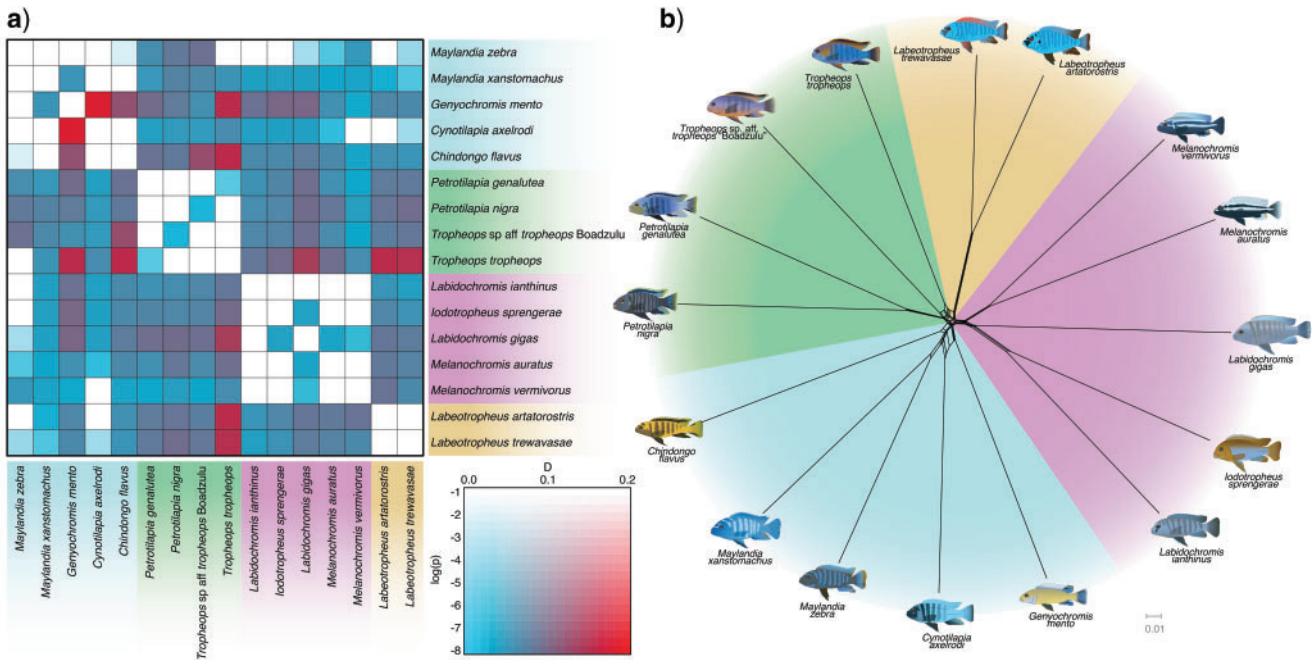

FIGURE 2. *D*-statistics and genomic neighbor network of the mbuna cichlids of Lake Malawi. a) Pairwise *D*-statistics, with *P*-values corrected for multiple testing. Heatmap shading corresponds to *D*-statistic scores. Red blocks indicate a stronger signal of introgression, blue blocks a weaker signal, and whiter, less saturated blocks represent pairwise scores associated with insignificant *P*-values. b) Genomic neighbor network based on the distance matrix of noncoding regions. Clades are shaded consistently with Fig. 1. Figure appears in color in the digital version.

the three highest weighted topologies correspond to those recovered in our phylogenomic reconstructions. Weighting was extremely heterogeneous across the genome based on 250-SNP windows, with two adjacent windows seldom favoring the same topology (Fig. 3f). Together, the results of our TWISST analyses show highly heterogeneous signal across the genomes of these fishes, which corresponds to the lack of clear resolution found in other analyses.

Genomic Distinction of Two Syntopic Labeotropheus Species

We found robust support for the genomic distinction and reciprocal monophyly of the closely related and syntopically occurring *L. artatorostris* and *L. trewavasae*. The two species have a genome-wide weighted Weir and Cockerham F_{ST} of 0.2353 (mean =0.1226; Fig. 4a). In a genomic PCA, the two species were strongly separated in PC1, explaining 35.4% of the genomic variance (Fig. 4c). The remaining PCs each separated a single individual from the rest, as seen in Fig. 4c for PC2 (other PCs not shown). ADMIXTURE with K =2 divided the two species, although K =1 had the lowest CV error, possibly due to our small sample sizes per population (Fig. 4d,e).

We also found little evidence for the introgression between *Labeotropheus* species and *M. zebra* that has previously been hypothesized (Mims et al. 2010). ABBA–BABA tests yielded a *D*-statistic of 0.031 ($P = 5.251 \times 10^{-14}$; BBAA 109,799, ABBA 25,628, BABA 24,090.4), f4-ratio 0.0101. TWISST showed dominant signal for monophyly of *Labeotropheus*, with very little indication of introgression from *M. zebra* into the genome, though marginally more into *L. trewavasae* than into *L. artatorostris* (Fig. 4f–h; Fig. S9 of the Supplementary material available on Zenodo).

DISCUSSION

The integrative and stepwise approach taken here, based on SNP data sets from complete genome sequences of 16 species in 10 genera, was able to provide clear resolution for some, but not all, of the phylogenetic relationships among the Malawi mbuna cichlids. With the most comprehensive phylogenomic sample of mbuna cichlids to date, we were able to recover four main clades within the mbuna with high support. Most previous works have included exemplars of each of these clades but lacked the data to test their interrelationships (Table S1 of the Supplementary material available on Zenodo). Although we ourselves were missing four of the recognized mbuna genera (*Abactochromis* [1 sp.], *Cyathochromis* [1 sp.], *Gephyrochromis* [2 spp.], and *Pseudotropheus* [16 spp.]), these are likely to fall within these four main clades. As phylogenomic data sets frequently yield spuriously high statistical support (Rokas and Carroll 2006; Singhal et al. 2021), the consistent recovery across reconstruction methods (Fig. 1) provides greater confidence than any single method alone in the existence of the four main clades.

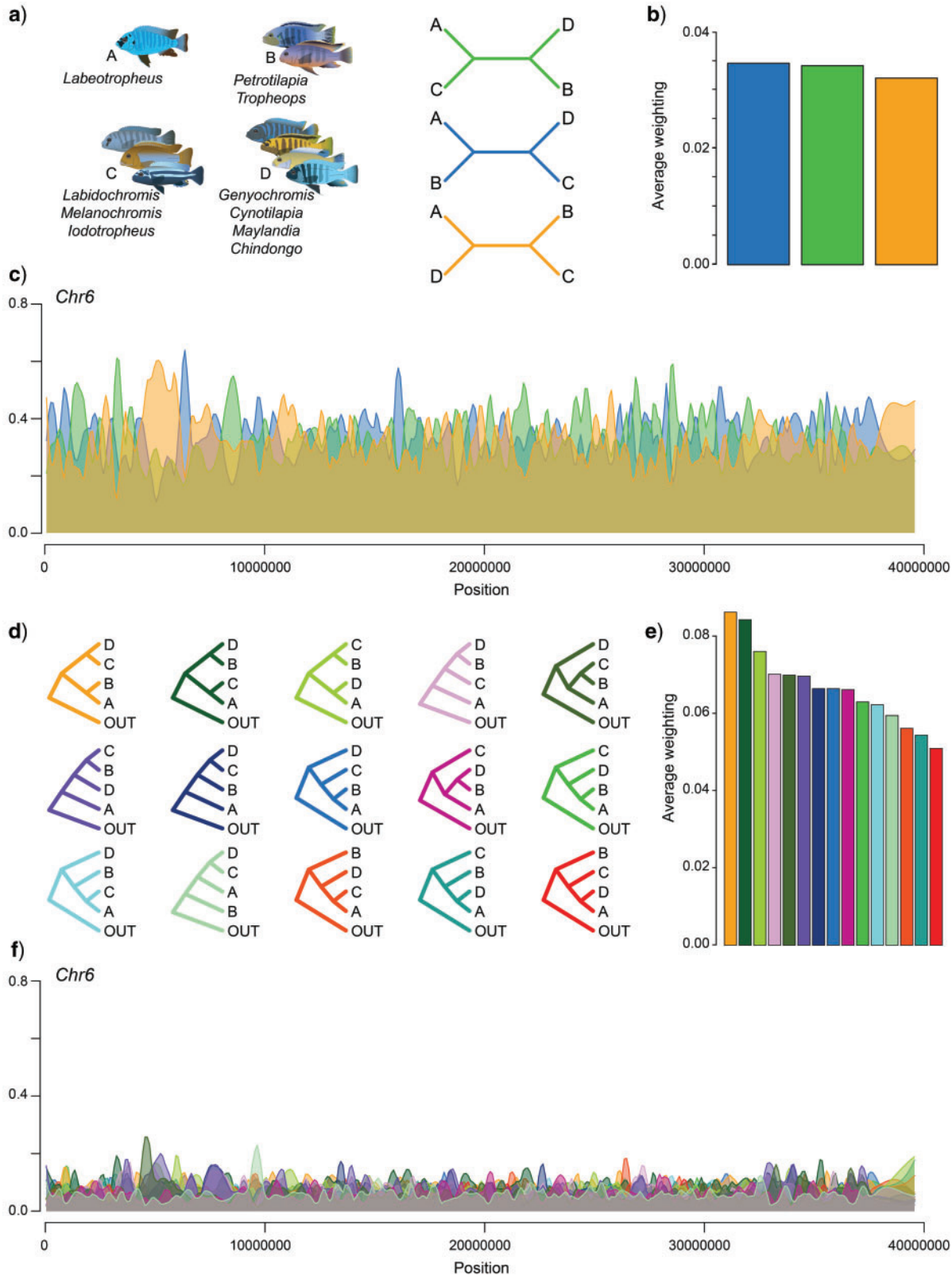

FIGURE 3. Topology weighting (TWISST) analysis of the major clades within the mbuna radiation. a) Three alternative unrooted quartets of the major mbuna clades, and b) their relative weightings across the whole-genome assemblies (ordered by average weighting). c) Smoothed TWISST weights across Chromosome 6 (chosen arbitrarily), illustrating the heterogeneity in local topological weightings; see Fig. S7 of the Supplementary material available on Zenodo for all chromosomes. d) Fifteen alternative rooted topologies including the outgroup, and e) their relative weightings across the whole-genome assemblies. f) Smoothed TWISST weights across Chromosome 6 (chosen arbitrarily), illustrating the heterogeneity in local topological weightings; see Fig. S8 of the Supplementary material available on Zenodo for all chromosomes. Figure appears in color in the digital version.

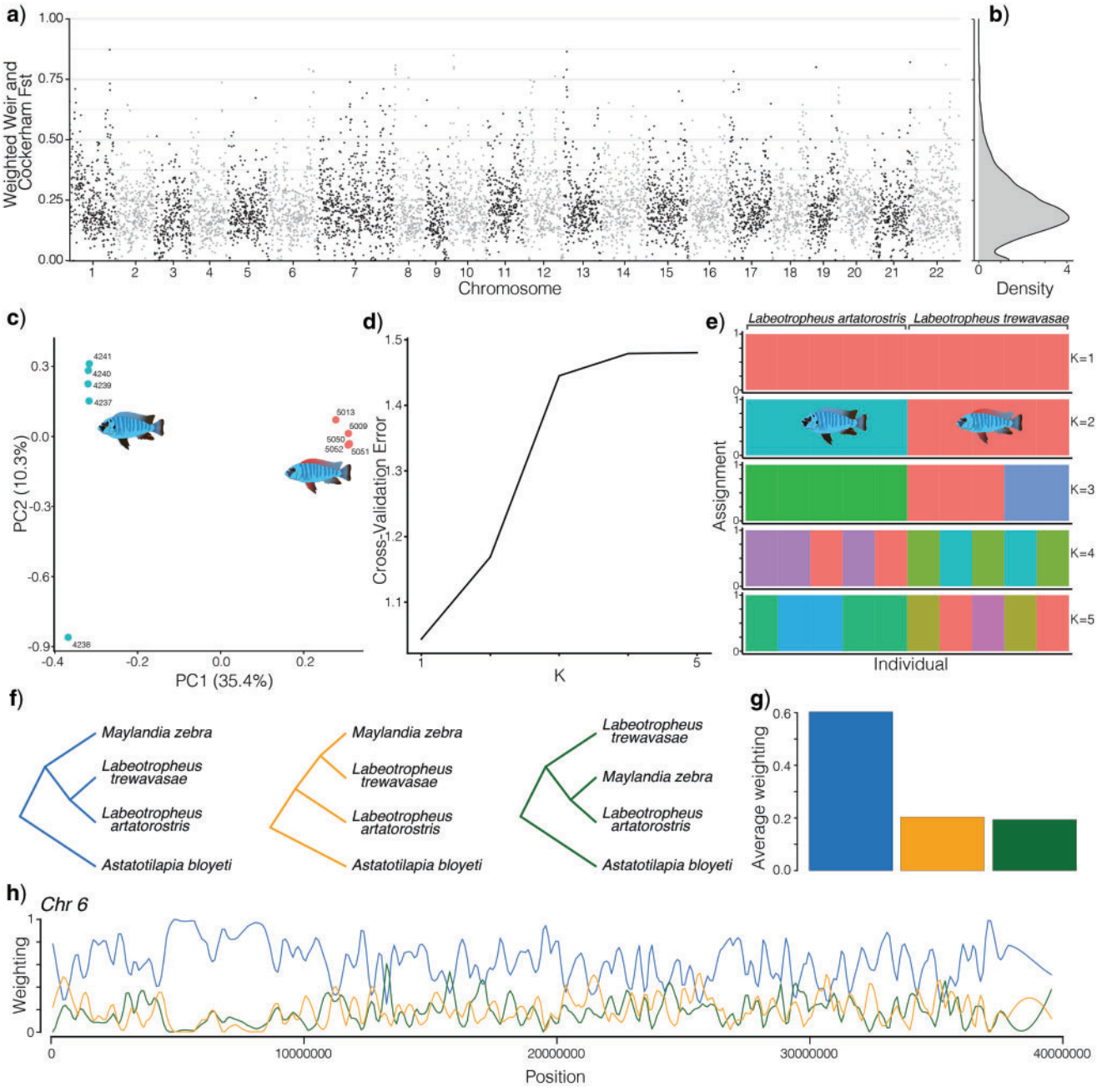

FIGURE 4. Genetic differentiation of *Labeotropheus artatorostris* and *L. trewavasae* from Thumbi West. a) Weighted Weir and Cockerham F_{ST} calculated in 100 kb windows across the genome between *Labeotropheus trewavasae* and *L. artatorostris*. Chromosomes are shaded alternately black and gray. b) Density plot of Weighted F_{ST} values across the genome. c) Genomic principal component analysis projection of the first two components; individual specimen ID numbers are indicated beside the points. d) ADMIXTURE cluster analysis. e) ADMIXTURE assignments across K values. f) Three alternative topologies used in Topology Weighting (TWISST) analysis. g) Average weighting of topologies across the genome. Colors correspond to topologies in f). h) Smoothed weighting plot of Chromosome 6. Colors correspond to topologies in f). Plots of all chromosomes are provided as Fig. S8 of the Supplementary material available on Zenodo. Figure appears in color in the digital version.

However, the resolution among these four clades at the base of the mbuna phylogeny remains uncertain. Our phylogenomic analyses provided moderate to high support for two alternative quartet configurations among the clades, and mostly agreed in placing Clade A, consisting only of *Labeotropheus*, sister to all other sampled mbuna (Fig. 1, Fig. S6 of the Supplementary material available on Zenodo). However, the branches separating Clades B, C, and D were very short, received low support, and failed the polytomy test implemented in ASTRAL. Paradoxically, the one quartet configuration that was not recovered by any of our coalescent or maximum likelihood reconstruction methods, A,B|C,D, was found to be the most frequent local topology

across the genome in our topology weighting analysis (Fig. 3). Hybridization and ILS both likely limit our ability to resolve the evolutionary history of the mbuna to some degree. We were unable to unambiguously resolve the interclade relationships even after excluding putative introgressed SNPs in subsequent phylogenomic analyses. Following the pruning of introgressed loci, SVDquartets, ASTRAL, and IQ-TREE analyses still failed to converge on a common topology (Fig. S6 of the Supplementary material available on Zenodo). Both ASTRAL and IQ-TREE recovered a quartet pattern of A,C|B,D, but SVDquartets reported A,D|B,C, albeit with weak support. Although resolution of the backbone is slightly improved when putatively introgressed loci are excluded with SVDquartets and ASTRAL analyses, support across these deep nodes is diminished in the IQ-TREE phylogeny. Furthermore, the internodes in our ASTRAL tree separating the main four clades remained very short after controlling for introgression. Taking all this into account, it seems that the basal nodes within the mbuna radiation defy resolution even with methods accounting for ILS and introgression from whole-genome data.

On the other hand, despite the polytomy at the base of the radiation and the short timeframe over which the mbuna are thought to have diversified, the entire mbuna phylogeny is not intractable. Within each of the four robust clades, we were able to achieve good phylogenomic resolution, with only little inconsistency among reconstruction methods for most relationships. Although the monophyly of *Tropheops* and *Labidochromis* are called into question (see below), there was substantial phylogenetic structure recovered among the disparate lineages we sampled (Figs. 1 and 2), and our sampling of couplets of species within six of the genera generally provided unambiguous monophyletic relationships. This robust ability to resolve these relationships runs contrary to the expectation that congeneric Malaŵi cichlids would be difficult to distinguish genotypically (Moran and Kornfield 1993; Albertson et al. 1999; Hulsey et al. 2010).

This ability of the whole-genome SNP data to resolve the more recently diverged mbuna relationships was particularly reinforced by the reciprocal monophyly, elevated F_{ST}, and lack of extensive admixture that we found in the two sympatrically occurring *Labeotropheus* species (Fig. 4), despite the fact that they hybridize in the lab (Table 1) and can often be observed to feed side-by-side on the same rock (Hulsey pers. obs.). The data showed that these species have only low levels of introgression, but that other species, especially *Tropheops* and taxa belonging to Clade D, have much higher levels across the mbuna radiation (Fig. 2a). Our results suggest that interclade hybridization is evidently common within the mbuna cichlids, but may not pose a major threat to the reconstruction of a largely bifurcating phylogeny of these fishes, in contrast to previous assumptions (Mims et al. 2010; Malinsky et al. 2018; Svardal et al. 2021). Although the removal of introgressed loci failed to resolve the basal branching order of the four main clades with certainty, it improved resolution among other parts of the tree, particularly for the monophyly of Clade D and relationships within it. This action however, resulted in a significant reduction of data set size (e.g., 110,402 SNPs vs. 495,087 SNPs for the noncoding LD-pruned data set) and as a consequence, the support for some relationships was diminished. As a whole, these results suggest that genome-wide SNP data in the mbuna and likely other recent radiations might commonly contain enough informative markers to account for introgression and still recover underlying phylogenetic history.

Due to their intermediate morphology, it has been suggested that *Tropheops* might be the result of hybrid speciation between *Labeotropheus* and *Maylandia* (Mims et al. 2010). This hypothesis is rejected by our phylogenomic reconstructions, which robustly place *Tropheops* in a clade with *Petrotilapia*, whereas *Maylandia* is more closely related to *Genyochromis*, *Cynotilapia*, and *Chindongo*. However, we did find evidence for considerable gene flow between *Tropheops* and *Labeotropheus*, *Chindongo*, and *Genyochromis* (Fig. 2a). Although this does not directly support a hybrid origin of *Tropheops*, it certainly does warrant further study. SVDquartets recovered *Tropheops* as paraphyletic with respect to *Petrotilapia*, but no other method supported this arrangement. Our phylogenomic reconstructions also reveal that *Iodotropheus* falls within *Labidochromis*. This is strong evidence that the two genera are synonymous (*I. sprengerae* is the type species of *Iodotropheus*). However, formal taxonomic synonymization of these genera should be forestalled until a more comprehensive study is undertaken on the two other valid *Iodotropheus* species and 16 other valid *Labidochromis* species, including the type species of *Labidochromis*, *L. vellicans* (Fricke et al. 2020). Curiously, Malinsky et al. (2018) did find *Maylandia* to be closely associated with *Petrotilapia* and *Tropheops*, differing substantially from our findings. However, their *M. zebra* specimen was a different individual than ours, so it may be that the two sampled individuals assigned to this species were not conspecific. Also, *Labeotropheus* was missing from most of their analyses. In our case, the incorporation of a second *Maylandia* species, *M. xanstomachus*, helps to inform and support the relationships that we recovered. Additional intergeneric sampling could further clarify and reinforce our general understanding of mbuna relationships.

Resolution of the base of the mbuna radiation was often one of the primary foci of earlier phylogenetic studies (Moran and Kornfield 1993; Albertson et al. 1999; Hulsey et al. 2017). However, based on our large genome-wide data set, it seems likely that the lack of resolution reflects biological processes that render this part of the mbuna phylogeny highly intractable to systematic reconstruction. In contrast, reconstruction of more recent bifurcation events in the mbuna radiation were readily resolved, and future phylogenetic studies

should focus primarily on these more recent, tractable, and diagnosable evolutionary relationships. The reconstruction of a phylogeny and subsequent species delimitation of mbuna genera across the whole of Lake Malawi may therefore be possible with whole-genome sequences. Resolving the age-old issue of how many species of cichlids occur in one of the most diverse radiations of fishes in the world may now be feasible.

Between a Rock and a Hard Polytomy

The major clades of mbuna recovered may represent the product of a phylogenetically unresolvable burst of diversification. This type of rapid divergence is thought to characterize many adaptive radiations (Suh 2016; Moreira and Schrago 2018; Pease et al. 2018; Braun et al. 2019; Cai et al. 2021; Gagnon et al. 2021; Hime et al. 2021; Morales-Briones et al. 2021; Singhal et al. 2021). The power to resolve some polytomies may be proportional to the volume of relevant data available (i.e., more loci should better resolve such recalcitrant nodes). In the past, the nodes lacking resolution at the base of the mbuna were treated as soft polytomies that could be resolved with sufficient data (e.g., Hulsey et al. 2017). However, our results show that sufficiently rapid radiations can simply defy phylogenomic resolution, even with whole-genome resequenced data. The relationships among the major clades of mbuna could not be confidently resolved, and the relationships agreed only between two of our four phylogenomic reconstructions (IQ-TREE and SVDquartets) obtained through the main analyses (Fig. 1). Although we would welcome innovative attempts to resolve these relationships, our data call into question whether these early mbuna relationships will ever be resolved with molecular data; this may be a hard polytomy.

Our ability to resolve rapid diversification events is often dictated by the interactions of evolution with time. In some cases, extinction of lineages involved in a burst of diversification increases the remaining internal branch lengths, potentially making early nodes within a radiation easier to resolve as clades age. However, saturation of molecular data, especially when lineages do not undergo substantial extinction, diminishes and eventually extinguishes phylogenetic signal (Xia et al. 2003). The number of orthologous SNPs also decreases over time especially in noncoding regions of the genome, which reduces the usability of SNP-based data sets to resolve diversification bursts of increasing age (Leaché and Oaks 2017). Gene duplications and loss, as well as genomic rearrangements can also pose problems (Bravo et al. 2019). Together, these processes decrease the power of genomic data to resolve bursts of diversification of increasing age and thereby result in decreasing power to resolve polytomies over time, irrespective of whether they are the result of multifurcation or rapid sequential bifurcation.

However, because the mbuna radiation is extremely young (<2 million years; Schedel et al. 2019; Matschiner et al. 2020), especially compared with some other examples of polytomies that are well established in the literature (e.g., that at the base of the Neoaves, ~66 million years; Suh 2016; Braun et al. 2019), it probably has not been affected significantly by time-related signal loss. Both the short branches recovered by ASTRAL and IQ-TREE and the low support values associated with these nodes across all of our reconstructions are consistent with a very rapid radiation at the origin of the mbuna clade. This has likely contributed to extremely high levels of shared standing genetic variation among species and ILS, making its basal branching relationships difficult to disentangle (Svardal et al. 2021). The high levels of standing genetic variation may have increased the level of plasticity in the early radiation, and thereby actually facilitated the diversification of the clade by subsequent "genetic assimilation" (see Meyer 1987; Schneider and Meyer 2017). Similar problems have been identified in other genomic analyses of cichlid relationships (e.g., Browning et al. 2018; Salzburger 2018; Olave and Meyer 2020; Svardal et al. 2021). Additionally, our analysis shows that gene flow has been moderately common among clades, consistent with previous reconstructions of this and other African Great Lake cichlid radiations (Irisarri et al. 2018; Malinsky et al. 2018; Salzburger 2018; Svardal et al. 2021). However, this introgression has not been as obfuscating as we had originally expected. Instead of an overall pattern of high introgression, a few specific lineages appear to be the main drivers of introgressive hybridization (Fig. 2a), and most of these involved interclade, and not intraclade, gene flow. Removing those regions that are particularly strongly introgressed resulted in better support for the placement of some introgressing taxa, but did not improve the interclade relationships. So, although it has likely contributed to uncertainty together with other sources of ILS, introgression does not seem to be the main source of the basal node instability in the mbuna.

Conclusion and Outlook

In the Lake Malawi mbuna cichlid radiation, we have found that integrating results from four different phylogenomic reconstruction methods that utilized SNP data from whole-genome resequencing provided strong phylogenomic resolution at most levels, but not the basal-most nodes. This provided the most robust phylogenetic hypothesis for this lineage to date, with four major clades being resolved that had only partially been recovered before, and *Labeotropheus* emerging as perhaps the sister to all other sampled mbuna. We suggest that the short branches of this rapid radiation may represent an unresolvable polytomy at the base of the mbuna. If this is the case, future work with more extensive taxon sampling will not resolve the basal nodes further (Suh 2016), but molecular examinations of other parts of the tree will likely be more fruitful. This also has implications for the taxonomic efforts to describe mbuna diversity. Taxonomists should work to incorporate molecular data,

especially large genomic data sets, as these efforts should now be seen as finally capable of providing robust and potentially complementary tests of more traditional morphology-based taxonomic diagnoses (e.g., Hanssens 2004; Li et al. 2016; Oliver 2018).

Nevertheless, as we start to accept hard polytomies as describing the evolution of some major clades, it is becoming more important to incorporate these evolutionary patterns into evolutionary studies such as trait evolution. The interest in the evolutionary links between ecology, phenotypes, and genotypes in rapidly diversifying groups is only increasing. Randomization techniques and integration of results across a large set of bootstrapped trees is common practice in current comparative analyses (Losos 1994; Hulsey et al. 2007), but fails to account for the inferential limitations of the underlying data sets (Smith et al. 2020). Moreover, the study of adaptive divergence in explosive radiations has increasingly shown that the evolutionary history of traits themselves sometimes do not correspond to the species tree (Kratochwil et al. 2018; Kautt et al. 2020), so the importance of incorporating ILS into our understanding of trait evolution is likely only going to increase. The use of gene-trees instead of bootstrap trees might make the iterative modeling of evolution over troublesome nodes more informative (Hahn and Nakhleh 2016), but ultimately, what is needed are methods to translate node uncertainty into the models underlying comparative analyses (Singhal et al. 2021). As we continue to improve the methodological toolkit for inference of accurate evolutionary histories, and identify also the limitations of even the most comprehensive genomic data sets, we will be better suited to extract what can and cannot be learned about organismal evolution within the framework of molecular systematics.

Supplementary Material

Supplementary Information submitted to Zenodo, available at https://doi.org/10.5281/zenodo.5164151.

Data Availability

New genomes were submitted to NCBI's Short Read Archive and are available under the project number PRJNA783868 at https://www.ncbi.nlm.nih.gov/sra/PRJNA783868. Nexus files of analysed datasets are available from the Dryad Digital Repository: https://dx.doi.org/10.5061/dryad.ffbg79cvr. Accession numbers are listed for individual samples in Table S2 of the Supplementary material available on Zenodo.

Funding

This work was supported by The University of Konstanz and the Deutsche Forschungsgemeinschaft [grant number DFG ME 1725/21-1 to A.M.].

Acknowledgments

Samples were collected in 2010 by C.D.H. with a permit from the Malaŵi Parks Service to C.D.H. Assembly and SNP-calling was done by P. Xiong. We thank members of the Meyer Lab for thoughtful comments on this manuscript, especially J. Torres-Dowdall and M. Olave. We are also grateful to L. Rancilhac for useful input on MSC methods. We are grateful to B. Carstens, J. Uyeda, and several anonymous reviewers for their insightful comments on our manuscript.

References

Albertson R.C., Kocher T.D., Wainwright P. 2005. Genetic architecture sets limits on transgressive segregation in hybrid cichlid fishes. Evolution 59:680–690.

Albertson R.C., Markert J.A., Danley P.D., Kocher T.D. 1999. Phylogeny of a rapidly evolving clade: the cichlid fishes of Lake Malaŵi, East Africa. Proc. Natl. Acad. Sci. USA 96:5107–5110.

Albertson R.C., Pauers M.J. 2019. Morphological disparity in ecologically diverse versus constrained lineages of Lake Malaŵi rock-dwelling cichlids. Hydrobiologia 832:153–174.

Albertson R.C., Powder K.E., Hu Y., Coyle K.P., Roberts R.B., Parsons K.J. 2014. Genetic basis of continuous variation in the levels and modular inheritance of pigmentation in cichlid fishes. Mol. Ecol. 23:2135–5150.

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

Alexander D.H., Lange K. 2011. Enhancements to the ADMIXTURE algorithm for individual ancestry estimation. BMC Bioinformatics 12:246.

Benjamini Y., Hochberg Y. 1995. Controlling the false discovery rate: a practical and powerful approach to multiple testing. J. R. Stat. Soc. B 57:289–300.

Blais J., Plenderleith M., Rico C., Taylor M.I., Seehausen O., van Oosterhout C., Turner G.F. 2009. Assortative mating among Lake Malaŵi cichlid fish populations is not simply predictable from male nuptial colour. BMC Evol. Biol. 9:53.

Bouckaert R., Heled J., Kühnert D., Vaughan T., Wu C.-H., Xie D., Suchard M.A., Rambaut A., Drummond A.J. 2014. BEAST 2: a software platform for Bayesian evolutionary analysis. PLoS Comput. Biol. 10:e1003537.

Braun E.L., Cracraft J., Houde P. 2019. Resolving the avian tree of life from top to bottom: the promise and potential boundaries of the phylogenomic era. In: Kraus R.H.S., editor. Avian genomics in ecology and evolution: from the lab into the wild. Cham (Switzerland): Springer International Publishing. p. 151–210.

Bravo G.A., Antonelli A., Bacon C.D., Bartoszek K., Blom M.P.K., Huynh S., Jones G., Knowles L.L., Lamichhaney S., Marcussen T., Morlon H., Nakhleh L.K., Oxelman B., Pfeil B., Schliep A., Wahlberg N., Werneck F.P., Wiedenhoeft J., Willows-Munro S., Edwards S.V. 2019. Embracing heterogeneity: coalescing the Tree of Life and the future of phylogenomics. PeerJ 7:e6399.

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

The genomic substrate for adaptive radiation in African cichlid fish. Nature 513:375–381.
Browning B.L., Tian X., Zhou Y., Browning S.R. 2021. Fast two-stage phasing of large-scale sequence data. Am. J. Hum. Genet. 108:1880–1890.
Browning B.L., Zhou Y., Browning S.R. 2018. A one-penny imputed genome from next generation reference panels. Am. J. Hum. Genet. 103:338–348.
Browning S.R., Browning B.L. 2007. Rapid and accurate haplotype phasing and missing data inference for whole genome association studies by use of localized haplotype clustering. Am. J. Hum. Genet. 81:1084–1097.
Bryant D., Bouckaert R., Felsenstein J., Rosenberg N.A., RoyChoudhury A. 2012. Inferring species trees directly from biallelic genetic markers: bypassing gene trees in a full coalescent analysis. Mol. Biol. Evol. 29:1917–1932.
Cai L., Xi Z., Lemmon E.M., Lemmon A.R., Mast A., Buddenhagen C.E., Liu L., Davis C.C. 2021. The perfect storm: gene tree estimation error, incomplete lineage sorting, and ancient gene flow explain the most recalcitrant ancient angiosperm clade, Malpighiales. Syst. Biol. 70:491–507.
Campbell D., Bernatchez L. 2004. Generic scan using AFLP markers as a means to assess the role of directional selection in the divergence of sympatric whitefish ecotypes. Mol. Biol. Evol. 21: 945–956.
Chang C.C., Chow C.C., Tellier L.C.A.M., Vattikuti S., Purcell S.M., Lee J.J. 2015. Second-generation PLINK: rising to the challenge of larger and richer datasets. GigaScience 4:7.
Chifman J., Kubatko L. 2014. Quartet inference from SNP data under the coalescent model. Bioinformatics 30:3317–3324.
Conith M.R., Conith A.J., Albertson R.C. 2019. Evolution of a soft-tissue foraging adaptation in African cichlids: roles for novelty, convergence, and constraint. Evolution 73:2072–2084.
Conte M.A., Joshi R., Moore E.C., Nandamuri S.P., Gammerdinger W.J., Roberts R.B., Carleton K.L., Lien S., Kocher T.D. 2019. Chromosome-scale assemblies reveal the structural evolution of African cichlid genomes. GigaScience 8:giz030.
Cooper W.J., Wernle J., Mann K., Albertson R.C. 2011. Functional and genetic integration in the skulls of Lake Malaŵi cichlids. Evol. Biol. 38:316–334.
Danecek P., Auton A., Abecasis G., Albers C.A., Banks E., DePristo M.A., Handsaker R., Lunter G., Marth G., Sherry S.T., McVean G., Durbin R.; 1000 Genomes Project Analysis Group. 2011. The variant call format and VCFtools. Bioinformatics 27:2156–2158.
Danecek P., Bonfiield J.K., Liddle J., Marshall J., Ohan V., Pollyard M.O., Whitwham A., Keane T., McCarthy S.A., Davies R.M., Li H. 2021. Twelve years of SAMtools and BCFtools. GigaScience 10:giab008.
Degnan J.H. 2018. Modeling hybridization under the network multispecies coalescent. Syst. Biol. 67:786–799.
Degnan J.H., Rosenberg N.A. 2009. Gene tree discordance, phylogenetic inference and the multispecies coalescent. Trends Ecol. Evol. 24:332–340.
Ding B., Daugherty D.W., Husemann M., Chen M., Howe A.E., Danley P.D. 2014. Quantitative genetic analyses of male color pattern and female mate choice in a pair of cichlid fishes of Lake Malaŵi, East Africa. PLoS One 9:e114798.
Edwards S.V. 2009. Is a new and general theory of molecular systematics emerging? Evolution 63:1–19.
Fishelson L. 2003. Comparison of testes structure, spermatogenesis, and spermatocytogenesis in young, aging, and hybrid cichlid fish (Cichlidae, Teleostei). J. Morphol. 256:285–300.
Flouri T., Jiao X., Rannala B., Yang Z. 2019. A Bayesian implementation of the multispecies coalescent model with introgression for phylogenomic analysis. Mol. Biol. Evol. 37:1211–1223.
Fricke R., Eschmeyer W.N., van der Laan R. 2020. Eschmeyer's catalog of fishes: genera, species, references. San Francisco (CA): California Academy of Sciences. Available from: http://researcharchive.calacademy.org/research/ichthyology/catalog/fishcatmain.asp.
Gagnon E., Hilgenhof R., Orejuela A., McDonnell A., Sablok G., Aubriot X., Giacomin L., Gouvêa Y., Bohs L., Dodsworth S., Martine C., Poczai P., Knapp S., Särkinen T. 2021. Phylogenomic data reveal hard polytomies across the backbone of the large genus *Solanum* (Solanaceae). bioRxiv. 2021.03.25.436973.
Garrison E., Marth G. 2012. Haplotype-based variant detection from short-read sequencing. arXiv. 1207.3907.
Genner M.J., Ngatunga B.P., Mzighani S., Smith A., Turner G.F. 2015. Geographical ancestry of Lake Malaŵi's cichlid fish diversity. Biol. Lett. 11:20150232.
Guindon S., Dufayard J.-F., Lefort V., Anisimova M., Hordijk W., Gascuel O. 2010. New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. Syst. Biol. 59:307–321.
Hahn M.W., Nakhleh L. 2016. Irrational exuberance for resolved species trees. Evolution 70:7–17.
Hanssens M. 2004. The deep-water *Placidochromis* species. In; Snoeks J., Konings A., editors. The cichlid diversity of Lake Malaŵi/Nyasa/Niassa: identification, distribution and taxonomy. El Paso (TX): Cichlid Press. p. 104–197.
Henning F., Meyer A. 2014. The evolutionary genomics of cichlid fishes: explosive speciation and adaptation in the postgenomic era. Annu. Rev. Genom. Hum. G. 15:417–441.
Hime P.M., Lemmon A.R., Lemmon E.C.M., Prendini E., Brown J.M., Thomson R.C., Kratovil J.D., Noonan B.P., Pyron R.A., Peloso P.L.V., Kortyna M.L., Keogh J.S., Donnellan S.C., Mueller R.L., Raxworthy C.J., Kunte K., Ron S.R., Das S., Gaitonde N., Green D.M., Labisko J., Che J., Weisrock D.W. 2021. Phylogenomics reveals ancient gene tree discordance in the amphibian tree of life. Syst. Biol. 70: 49–66.
Hoang D.T., Chernomor O., von Haeseler A., Minh B.Q., Vinh L.S. 2018. UFBoot2: improving the ultrafast bootstrap approximation. Mol. Biol. Evol. 35:518–522.
Hudson D.H., Bryant D. 2006. Application of phylogenetic networks in evolutionary studies. Mol. Biol. Evol. 23:254–267.
Hulsey C.D., Alfaro M.E., Zheng J., Meyer A., Holzman R. 2019. Pleiotropic jaw morphology links the evolution of mechanical modularity and functional feeding convergence in Lake Malaŵi cichlids. Proc. R. Soc. B-Biol. Sci. 286:20182358.
Hulsey C.D., Mims M.C., Parnell N.F., Streelman J.T. 2010. Comparative rates of lower jaw diversification in cichlid adaptive radiations. J. Evol. Biol. 23:1456–1467.
Hulsey C.D., Mims M.C., Streelman J.T. 2007. Do constructional constraints influence cichlid craniofacial diversification? Proc. R. Soc. B-Biol. Sci. 274:1867–1875.
Hulsey C.D., Roberts R.J., Loh Y.-H.E., Rupp M.F., Streelman J.T. 2013. Lake Malaŵi cichlid evolution along a benthic/limnetic axis. Ecol. Evol. 3:2262–2272.
Hulsey C.D., Zheng J., Faircloth B.C., Meyer A., Alfaro M.E. 2017. Phylogenomic analysis of Lake Malaŵi cichlid fishes: further evidence that the three-stage model of diversification does not fit. Mol. Phlogenet. Evol. 114:40–48.
Husemann M., Tobler M., McCauley C., Ding B., Danley P.D. 2017. Body shape differences in a pair of closely related Malaŵi cichlids and their hybrids: effects of genetic variation, phenotypic plasticity, and transgressive segregation. Ecol. Evol. 7:4336–4346.
Irisarri I., Baurain D., Brinkmann H., Delsuc F., Sire J.-Y., Kupfer A., Petersen J., Jarek M., Meyer A., Vences M., Philippe H. 2017. Phylotranscriptomic consolidation of the jawed vertebrate timetree. Nat. Ecol. Evol. 1:1370–1378.
Irisarri I., Meyer A. 2016. The identification of the closest living relative(s) of tetrapods: phylogenomic lessons for resolving short ancient internodes. Syst. Biol. 65:1057–1075.
Irisarri I., Singh P., Koblmüller S., Torres-Dowdall J., Henning F., Franchini P., Fischer C., Lemmon A.R., Lemmon E.M., Thallinger G.G., Sturmbauer C., Meyer A. 2018. Phylogenomics uncovers early hybridization and adaptive loci shaping the radiation of Lake Tanganyika cichlid fishes. Nat. Commun. 9:3159.
Kalyaanamoorthy S., Minh B.Q., Wong T.K., von Haeseler A., Jermiin L.S. 2017. ModelFinder: fast model selection for accurate phylogenetic estimates. Nat. Methods 14:587.
Kautt A.F., Kratochwil C.F., Nater A., Machado-Schiaffino G., Olave M., Henning F., Torres-Dowdall J., Härer A., Hulsey C.D., Franchini P., Pippel M., Myers E.W., Meyer A. 2020. Contrasting signatures of genomic divergence during sympatric speciation. Nature 588:106–111.
Kautt A.F., Machado-Schiaffino G., Meyer A. 2016. Multispecies outcomes of sympatric speciation after admixture with the source

population in two radiations of Nicaraguan crater lake cichlids. PLoS Genet. 12:e1006157.
Knowles L.L., Carstens B.C. 2007. Delimiting species without monophyletic gene trees. Syst. Biol. 56:887–895.
Koblmüller S., Egger B., Sturmbauer C., Sefc K.M. 2010. Rapid radiation, ancient incomplete lineage sorting and ancient hybridization in the endemic Lake Tanganyika cichlid tribe Tropheini. Mol. Phylogenet. Evol. 55:318–334.
Kocher T.D., Conroy J.A., McKaye K.R., Stauffer J.R., Lockwood S.F. 1995. Evolution of NADH dehydrogenase subunit 2 in East African cichlid fish. Mol. Phylogenet. Evol. 4:420–432.
Konings A. 2007. Malaŵi cichlids in their natural habitat. 4th ed. El Paso (TX): Cichlid Press.
Kornfield I., Smith P.F. 2000. African cichlid fishes: model systems for evolutionary biology. Annu. Rev. Ecol. Syst. 31:163–196.
Kratochwil C.F., Liang Y., Gerwin J., Urban S., Henning F., Machado-Schiaffino G., Woltering J.M., Hulsey C.D., Meyer A. 2018. Agouti related peptide 2 facilitates convergent evolution of stripe patterns across cichlid fish radiations. Science 362:457–460.
Leaché A.D., Oaks J.R. 2017. The utility of single nucleotide polymorphism (SNP) data in phylogenetics. Annu. Rev. Ecol. Evol. S. 48:69–84.
Li H., Handsaker B., Wysoker A., Fennell T., Ruan J., Homer N., Marth G., Abecasis G., Durbin R.; 1000 Genome Project Data Processing Subgroup. 2009. The sequence alignment/map (SAM) format and SAMtools. Bioinformatics 25:2078–2079.
Li S., Konings A.F., Stauffer J.R.J. 2016. A revision of the *Pseudotropheus elongatus* species group (Teleostei: Cichlidae) with description of a new genus and seven new species. Zootaxa 4168:353–381.
Liu L., Yu L., Edwards S.V. 2010. A maximum pseudo-likelihood approach for estimating species trees under the coalescent model. BMC Evol. Biol. 10:302.
Losos J.B. 1994. An approach to the analysis of comparative data when a phylogeny is unavailable or incomplete. Syst. Biol. 43:117–123.
Maddison W.P. 1997. Gene trees in species trees. Syst. Biol. 46:523–536.
Malinsky M., Challis R.J., Tyers A.M., Schiffels S., Terai Y., Ngatunga B.P., Miska E.A., Durbin R., Genner M.J., Turner G.F. 2015. Genomic islands of speciation separate cichlid ecomorphs in an East African crater lake. Science 350:1493.
Malinsky M., Matschiner M., Svardal H. 2021. Dsuite - fast *D*-statistics and related admixture evidence from VCF files. Mol. Ecol. Resour. 21:584–595.
Malinsky M., Svardal H., Tyers A.M., Miska E.A., Genner M.J., Turner G.F., Durbin R. 2018. Whole-genome sequences of Malaŵi cichlids reveal multiple radiations interconnected by gene flow. Nat. Ecol. Evol. 2:1940–1955.
Martin S.H., Davey J.W., Jiggins C.D. 2015. Evaluating the use of ABBA-BABA statistics to locate introgressed loci. Mol. Biol. Evol. 32:244–257.
Martin S.H., van Belleghem S.M. 2017. Exploring evolutionary relationships across the genome using topology weighting. Genetics 206:429–438.
Matschiner M., Böhne A., Ronco F., Salzburger W. 2020. The genomic timeline of cichlid fish diversification across continents. Nat. Commun. 11:5895.
McElroy D.M., Kornfield I. 1993. Novel jaw morphology in hybrids between *Pseudotropheus zebra* and *Labeotropheus fuelleborni* (Teleostei: Cichlidae) from Lake Malaŵi, Africa. Copeia 1993:933–945.
McGee M.D., Faircloth B.C., Bornstein S.R., Zheng J., Hulsey C.D., Wainwright P.C., Alfaro M.E.. 2016. Replicated divergence in cichlid radiations mirrors a major vertebrate innovation. Proc. R. Soc. B-Biol. Sci. 283:20151413.
Meyer A. 1987. Phenotypic plasticity and heterochrony in *Cichlasoma managuense* (Pisces, Cichlidae) and their implications for speciation in cichlid fishes. Evolution 41:1357–1369.
Meyer A. 1993. Phylogenetic relationships and evolutionary processes in East African cichlid fishes. Trends Ecol. Evol. 8:279–284.
Meyer A., Kocher T., Basasibwakii P., Wilson A.C.. 1990. Monophyletic origin of Lake Victoria cichlid fishes suggested by mitochondrial DNA sequences. Nature 347:550–553.
Mims M.C., Hulsey C.D., Fitzpatrick B.M., Streelman J.T. 2010. Geography disentangles introgression from ancestral polymorphism in Lake Malaŵi cichlids. Mol. Ecol. 19:940–951.
Minh B.Q., Schmidt H.A., Chernomor O., Schrempf D., Woodhams M.D., von Haeseler A., Lanfear R. 2020. IQ-TREE 2: new models and efficient methods for phylogenetic inference in the genomic era. Mol. Biol. Evol. 37:1530–1534.
Morales-Briones D.F., Kadereit G., Tefarikis D.T., Moore M.J., Smith S.A., Brockington S.F., Timoneda A., Yim W.C., Cushman J.C., Yang Y. 2021. Disentangling sources of gene tree discordance in phylogenomic data sets: testing ancient hybridizations in Amaranthaceae s.l. Syst. Biol. 70:219–235.
Moran P., Kornfield I. 1993. Retention of an ancestral polymorphism in the Mbuna species flock (Teleostei: Cichlidae) of Lake Malaŵi. Mol. Biol. Evol. 10:1015–1029.
Moreira F.R.R., Schrago C.G. 2018. Coalescent-based phylogenetic inference from genes with unequivocal historical signal suggests a polytomy at the root of the placental mammal tree of life. bioRxiv:423996.
Nguyen L.-T., Schmidt H.A., von Haeseler A., Minh B.Q. 2015. IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. Mol. Biol. Evol. 32:268–274.
Nosil P., Funk D.J., Ortiz-Barrientos D. 2009. Divergent selection and heterogeneous genomic divergence. Mol. Ecol. 18:375–402.
Olave M., Meyer A. 2020. Implementing large genomic SNP datasets in phylogenetic network reconstructions: a case study of particularly rapid radiations of cichlid fish. Syst. Biol. 69:848–862.
Oliver M.K. 2018. Six new species of the cichlid genus *Otopharynx* from Lake Malaŵi (Teleostei: Cichlidae). B. Peabody Mus. Nat. Hi. 59:59–197.
O'Quin C.T., Drilea A.C., Roberts R.B., Kocher T.D. 2012. A small number of genes underlie male pigmentation traits in Lake Malaŵi cichlid fishes. J. Exp. Zool. 318B:199–208.
Ortiz E.M. 2019. vcf2phylip v2.0: convert a VCF matrix into several matrix formats for phylogenetic analysis. Available from: https://github.com/edgardomortiz/vcf2phylip.
Parnell N.F., Hulsey C.D., Streelman J.T. 2012. The genetic basis of a complex functional system. Evolution 66:3352–3366.
Pauers M.J. 2010. Species concepts, speciation, and taxonomic change in the Lake Malaŵi mbuna, with special reference to the genus *Labeotropheus* Ahl 1927 (Perciformes: Cichlidae). Rev. Fish Biol. Fisher. 20:187–202.
Pauers M.J. 2017. A new species of *Labeotropheus* (Perciformes: Cichlidae) from southern Lake Malaŵi, Africa. Copeia 105:399–414.
Pauers M.J., Fox K.R., Hall R.A., Patel K. 2018. Selection, hybridization, and the evolution of morphology in the Lake Malaŵi endemic cichlids of the genus *Labeotropheus*. Sci. Rep. 8:15842.
Pease J.B., Brown J.W., Walker J.F., Hinchliff C.E., Smith S.A. 2018. Quartet sampling distinguishes lack of support from conflicting support in the green plant tree of life. Am. J. Bot. 105:385–403.
Prum R.O., Brev J.S., Dornburg A., Field D.J., Townsend J.P., Lemmon E.M., Lemmon A.R. 2016. A comprehensive phylogeny of birds (Aves) using targeted next-generation DNA sequencing. Nature 534:S7–S8.
Quinlan A.R., Hall I.M. 2010. BEDTools: a flexible suite of utilities for comparing genomic features. Bioinformatics 26:841–842.
R Core Team. 2014. R: a language and environment for statistical computing. Vienna (Austria): R foundation for statistical computing. Available from: http://www.R-project.org/.
Rambaut A., Drummond A.J., Xie D., Baele G., Suchard M.A. 2018. Posterior summarisation in Bayesian phylogenetics using Tracer 1.7. Syst. Biol. 67:901–904.
Reaz R., Bayzid M.S., Rahman M.S. 2014. Accurate phylogenetic tree reconstruction from quartets: a heuristic approach. PLoS One 9:e104008.
Reinthal P.N., Meyer A. 1997. Molecular phylogenetic tests of speciation models in African cichlid fishes. In: Givnish T.J., Sytsma K.J., editors. Molecular evolution and adaptive radiations. Cambridge (UK): Cambridge University Press. p. 375–390.
Ribbink A.J., Marsh B.A., Marsh A.C., Ribbink A.C., Sharp B.J. 1983a. A preliminary survey of the cichlid fishes of rocky habitats in Lake Malaŵi. S. Afr. J. Zool. 18:149–310.
Ribbink A.J., Marsh A.C., Marsh B.A., Sharp B.J. 1983b. The zoogeography, ecology, taxonomy of the genus *Labeotropheus* Ahl, 1927 of Lake Malaŵi (Pisces: Cichlidae). Zool. J. Linn. Soc.-Lond. 79:223–243.

Rokas A., Carroll S.B.. 2006. Bushes in the tree of life. PLoS Biol. 4:e352.
Rosenberg N.A. 2003. The shapes of neutral gene genealogies in two species: probabilities of monophyly, paraphyly, and polyphyly in a coalescent model. Evolution 57:1465–1477.
Salzburger W. 2018. Understanding explosive diversification through cichlid fish genomics. Nat. Rev. Genet. 19:705–717.
Sayyari E., Mirarab S. 2018. Testing for poyltomies in phylogenetic species tree using quartet frequencies. Genes 9:132.
Schedel F.D.B., Musilova Z., Schliewen U.K. 2019. East African cichlid lineages (Teleostei: Cichlidae) might be older than their ancient host lakes: new divergence estimates for the east African cichlid radiation. BMC Evol. Biol. 19:94.
Schneider R.F., Meyer A. 2017. How plasticity, genetic assimilation and cryptic genetic variation may contribute to adaptive radiations. Mol. Ecol. 26:330–350.
Seehausen O. 2006. African cichlid fish: a model system in adaptive radiation research. Proc. R. Soc. B-Biol. Sci. 273:1987–1998.
Singhal S., Colston T.J., Grundler M.R., Smith S.A., Costa G.C., Colli G.R., Moritz C., Pyron R.A., Rabosky D.L. 2021. Congruence and conflict in the higher-level phylogenetics of squamate reptiles: an expanded phylogenomic perspective. Syst. Biol. 70:542–557.
Smith S.A., Walker-Hale N., Walker J.F., Brown J.W. 2020. Phylogenetic conflicts, combinability, and deep phylogenomics in plants. Syst. Biol. 69:579–592.
Stamatakis A. 2014. RAxML Version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. Bioinformatics 30:1312–1313.
Stauffer J.R., Bowers N.J., Kocher T.D., McKaye K.R. 1996. Evidence of hybridization between *Cynotilapia afra* and *Pseudotropheus zebra* (Teleostei: Cichlidae) following an intralacustrine translocation in Lake Malaŵi. Copeia 1996:203–208.
Streelman J.T., Danley P.D. 2003. The stages of vertebrate evolutionary radiation. Trends Ecol. Evol. 18:126–131.
Streelman J.T., Gmyrek S.L., Kidd M.R., Robinson R.L., Hert E., Ambali A.J., Kocher T.D. 2004. Hybridization and contemporary evolution in an introduced cichlid fish from Lake Malaŵi National Park. Mol. Ecol. 13:2471–2479.
Suh A. 2016. The phylogenomic forest of bird trees contains a hard polytomy at the root of Neoaves. Zool. Scr. 45:50–62.
Svardal H., Salzburger W., Malinsky M. 2021. Genetic variation and hybridization in evolutionary radiations of cichlid fishes. Annu. Rev. Anim. Biosci. 9:55–79.
Swofford D.L. 2002. PAUP*. Phylogenetic Analysis Using Parsimony (*and Other Methods). Version 4. Sunderland (MA): Sinauer Associates.
Tan A., Abecasis G.R., Kang H.M. 2015. Unified representation of genetic variants. Bioinformatics 31:2202–2204.
Urban S., Nater A., Meyer A., Kratochwil C.F. 2021. Different sources of allelic variation drove repeated color pattern divergence in cichlid fishes. Mol. Biol. Evol. 38:465–477.
Via S., West J. 2008. The genetic mosaic suggests a new role for hitchhiking in ecological speciation. Mol. Ecol. 17:4334–4345.
Xia X., Xie Z., Salemi M., Chen L., Wang Y. 2003. An index of substitution saturation and its application. Mol. Phylogenet. Evol. 26:1–7.
Zhang C., Dong S.S., Xu J.Y., He W.M., Yang T.L. 2019. PopLDdecay: a fast and effective tool for linkage disequilibrium decay analysis based on variant call format files. Bioinformatics 35:1786–1788.
Zhang C., Rabiee M., Sayyari E., Mirarab S. 2018. ASTRAL-III: polynomial time species tree reconstruction from partially resolved gene trees. BMC Bioinformatics 19:153.