

# Dynamic body size evolution during speciation of predatory cichlid fishes in the Lake Malawi superradiation

Liandong Yang,<sup>1,2,5</sup> Holly Scott,<sup>1,5</sup> Benjamin P. Ngatunga,<sup>3</sup> Asilatu H. Shechonge,<sup>3</sup> George F. Turner,<sup>4</sup> and Martin J. Genner<sup>1</sup>

<sup>1</sup>School of Biological Sciences, University of Bristol, Bristol BS8 1TQ, United Kingdom; <sup>2</sup>State Key Laboratory of Breeding Biotechnology and Sustainable Aquaculture, Institute of Hydrobiology, Chinese Academy of Sciences, Wuhan 430072, China; <sup>3</sup>Tanzania Fisheries Research Institute (TAFIRI), Dar es Salaam, Tanzania; <sup>4</sup>School of Environmental & Natural Sciences, Bangor University, Bangor, Gwynedd LL57 2UW, United Kingdom

Evolutionary divergence in body size is common in animal adaptive radiations and is often associated with differences in key ecological traits, including habitat use and prey consumption. Here we characterize a notable case of body size–associated adaptive radiation in a group of predatory open water cichlid fish species from the Lake Malawi catchment. Using whole-genome sequences, we show that body size differences have evolved multiple times in the focal genus, *Rhamphochromis*, and that the group possesses well-defined signals of ancient interspecific hybridization. We identify genetic variants strongly associated with body size and show that these variants are connected to genes enriched for functions in vertebrate skeletal and nervous system development. We focus our analyses on two species of *Rhamphochromis* endemic to Lake Kingiri, a small (600 m diameter) crater lake geographically isolated from the main body of Lake Malawi but within the catchment. We show that these two ecomorphologically divergent sympatric species—one small-bodied, the other larger-bodied—share a unique common ancestor and diverged from one another ~2000 years ago. We demonstrate strong directional selection focused on the larger-bodied Kingiri species, specifically on genetic variants connected to genes with anatomical development and nervous system function. Collectively, these results are supportive of body size–associated speciation taking place rapidly in the Lake Malawi cichlid fish superradiation. We conclude that body size–associated genetic variants have been important targets of selection during large-scale cichlid fish diversification, including in a crater lake sympatric speciation context.

[Supplemental material is available for this article.]

Body size is intrinsically associated with multiple life history, as well as physiological and ecological traits (Peters 1983; Woodward et al. 2005) and, in fishes, is strongly linked to diet (Romanuk et al. 2011), habitat use (Clarke 2021) and migratory behavior (Burns and Bloom 2020). Furthermore, divergence in body size between sister taxa is seen in many prominent fish adaptive radiations, including three-spined sticklebacks (Smith et al. 2020), Arctic charr (Gudbrandsson et al. 2019), Alpine whitefishes (De-Kayne et al. 2022), notothenioid icefishes (Colombo et al. 2015), gobies (Troyer et al. 2025), and cichlid fishes (Gonzalez-Voyer et al. 2009; Steele and López-Fernández 2014). Therefore, body size evolution is an important component of adaptive radiation in fishes, but the genetic basis of this evolution has been explored in relatively few cases (Troyer et al. 2025). Studies exploring genetic variation associated with growth-related traits in fishes has tended to focus on species of commercial importance (Liu et al. 2016; Yoshida and Yáñez 2021; Yáñez et al. 2023), with studies on other fish species being relatively uncommon (Laine et al. 2013; Takahashi et al. 2021; DeLorenzo et al. 2023). In contrast, genetic variation underpinning body size divergence in other

vertebrates has been more widely studied, and there is now good evidence supporting a polygenic basis to body size variation in humans as well as domestic mammals and birds (Sutter et al. 2007; Bouwman et al. 2018; Zhou et al. 2018).

The superradiations of cichlid fishes in the East African lakes Malawi, Victoria, and Tanganyika are among the most species rich of all vertebrate adaptive radiations. In each of these lakes, there is extreme body size divergence, and body size is at least partly linked to trophic ecology (Takahashi and Koblmüller 2011; Konings 2016), although given complex associations between body size trophic level in fishes (Keppeler et al. 2020), this requires detailed study. Within these radiations, there are cases in which sister taxa have highly divergent body sizes, suggesting body size divergence can associate with speciation. For example, the Lake Tanganyika cichlid *Telmatochromis temporalis* has undergone divergence into a larger rock-living ecotype and a smaller shell-living ecotype (Winkelman et al. 2014). Currently, we have a very limited understanding of the genomic basis of body size divergence in cichlids (Takahashi et al. 2021; DeLorenzo et al. 2023). Nevertheless, because of their high levels of phenotypic disparity, coupled with low levels of genomic divergence (Malinsky et al. 2018), there

<sup>5</sup>These authors contributed equally to this work.

Corresponding author: M.Genner@bristol.ac.uk

Article published online before print. Article, supplemental material, and publication date are at <https://www.genome.org/cgi/doi/10.1101/gr.281193.125>. Freely available online through the *Genome Research* Open Access option.

© 2026 Yang et al. This article, published in *Genome Research*, is available under a Creative Commons License (Attribution 4.0 International), as described at <http://creativecommons.org/licenses/by/4.0/>.

is potential to use these radiations as models to learn more about the drivers of body size evolution in fishes.

The Lake Malawi cichlid superradiation comprises three subradiations, the Pseudotropheina, the Cyrtocarina, and the Rhamphochromina (Oliver 2024), and in each of these, there is considerable size variation associated with ecological and morphological divergence (Konings 2016). Perhaps the most remarkable interspecific size variation is within *Rhamphochromis*, a genus of predators within the Rhamphochromina, with representative adults ranging in standard length from ~7 cm (*Rhamphochromis* sp. “kingiri dwarf,” one of the smallest species in the Malawi radiation) to >40 cm (*Rhamphochromis* sp. “nkhwazi,” the largest species in the entire radiation). Here we focus on this genus *Rhamphochromis*, which in total contains 15 species. Twelve of these species are found in the main body of Lake Malawi, and they typically have broad distributions around the lake, albeit with some species-specific habitat preferences (Turner 1996; Turner and Genner 2025). Beyond Lake Malawi itself, one very small-bodied species has been sampled only from satellite Lake Chilingali (*Rhamphochromis* sp. “chilingali”) (Genner et al. 2007) and is now thought to be extinct in the wild (Turner et al. 2023). The final two species are endemic to the small (~600 m diameter, 34 m deep) crater Lake Kingiri (Turner et al. 2019), namely, *Rhamphochromis* sp. “kingiri large” (20 cm standard length) and *Rhamphochromis* sp. “kingiri dwarf” (7 cm standard length).

In this study, we used the large range in body size in the genus *Rhamphochromis* to explore the genetic basis of body size in cichlids using genome sequence data. We first investigated repeated body size divergence during the evolution of the genus by reconstructing the evolutionary relationships of species. We then identified genomic loci associated with body size and tested if these loci are associated with genes expressed during development. Next, we explored the evidence for body size-associated sympatric speciation in Lake Kingiri, using genome data to date the divergence event, test for gene flow with species found outside the crater lake, and identify suites of genes under selection.

## Results

### Body size variation among *Rhamphochromis* species

From 2004 to 2011, we collected *Rhamphochromis* from three key water bodies: Lake Malawi, Lake Kingiri, and Lake Chilingali (Fig. 1). In total, we collected a total of 877 specimens representing 15 *Rhamphochromis* species that we distinguish based on body size, aspects of head, tooth, and jaw morphology, as well as male breeding colors (Supplemental Tables S1, S2). We quantified multiple morphological traits from these specimens, including dorsal spines, dorsal rays, anal rays, and standard length. Notably, standard length showed more than fourfold variation among the 15 *Rhamphochromis* cichlid fish species (Fig. 1). This exceptional disparity in body size underscores the remarkable evolutionary diversity within this group.

### Phylogeny, population structure, and evolutionary history

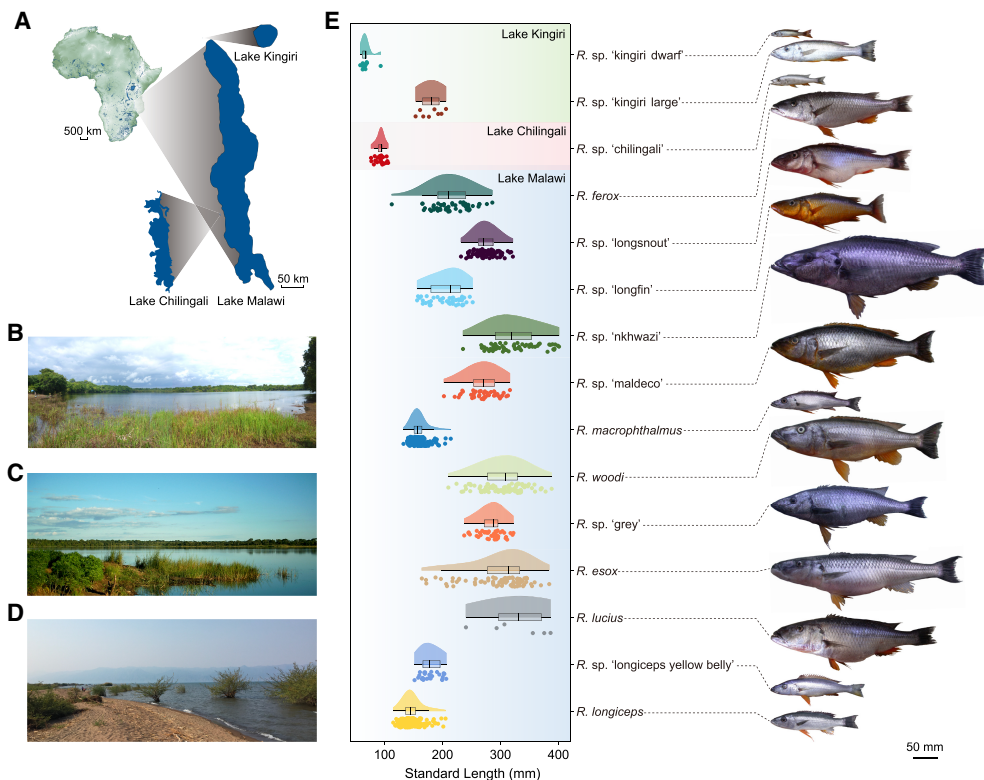
To resolve the evolutionary relationships among *Rhamphochromis* species, we used whole-genome sequence data from 69 individual cichlid fishes, including 67 representatives from the 15 *Rhamphochromis* species, 48 of which were newly sequenced for this study, along with two individuals from outgroup species. The average sequencing depth was 10.3 $\times$ , with an average read mapping rate of 95.81% to the reference genome and a genome

coverage of 97.17% (Supplemental Table S3). One sample, the only representative in our data set of *Rhamphochromis* sp. “longiceps yellow belly” (0.7 $\times$ ), was excluded from the whole-genome phylogenetic analysis owing to low coverage but was retained for mitochondrial genome phylogenetic analysis. In total, we identified 23,039,948 SNPs and 6,157,291 short insertions and deletions (indels). After applying stringent filtering criteria, 2,816,637 high-quality variants were retained for downstream analyses.

To investigate how the *Rhamphochromis* radiation unfolded and the relationships between species with different body sizes, we first constructed a maximum likelihood (ML) phylogenetic tree based on the filtered high-quality SNPs found in the 68 individuals (Fig. 2A; Supplemental Fig. S1). Our phylogeny revealed that sampled Lake Malawi *Rhamphochromis* species formed distinct, reciprocally monophyletic groups with 100% bootstrap support, therefore resolving ambiguities present in a previous study based on mtDNA sequences (Genner et al. 2007). Species with similar body sizes did not cluster together, indicating that body size variation has evolved independently multiple times within the genus (Supplemental Fig. S1). Of particular interest are the two endemic species from Lake Kingiri, *Rhamphochromis* sp. “kingiri large” and *Rhamphochromis* sp. “kingiri dwarf,” that were resolved as sister species. Their closest relative is the small-bodied *Rhamphochromis* sp. “chilingali” from Lake Chilingali. These findings are consistent with the two endemic species in Lake Kingiri representing a case of sympatric speciation, distinguished by substantial phenotypic divergence, particularly in body size. As is typical for Malawi cichlids, the ML phylogenetic tree based on mitochondrial DNA (mtDNA) did not adequately capture the phylogenetic relationships resolved by the whole-nuclear-genome data (Supplemental Figs. S2, S3; Supplemental Table S4). Discrepancies likely reflect extensive incomplete lineage sorting (ILS) and introgression events during the evolutionary history of the radiation (Malinsky et al. 2018).

We investigated genome-wide population structure by inferring ancestry proportions for all *Rhamphochromis* species using ADMIXTURE (Supplemental Fig. S4; Alexander et al. 2009). This confirmed the presence of strong population structure among samples, with most species uniquely distinguished when  $K=10$ , with the exception of two closely related species pairs (*Rhamphochromis* sp. “longfin” and *Rhamphochromis* sp. “longsnout”; *Rhamphochromis* sp. “chilingali” and *Rhamphochromis ferox*), one relatively distantly related pair (*Rhamphochromis* sp. “maldeco” and *Rhamphochromis esox*), and the one individual from *R. lucius* that had an unresolved placement (Fig. 2B; Supplemental Fig. S5). We further visualized population structure using principal component analysis (PCA) (Fig. 2C,D), which showed strong clustering by species across the first four axes. Notably, the first component (PC1) clearly separated the two species from Lake Kingiri from the species in Lake Chilingali and Lake Malawi, suggesting that the Lake Kingiri species share a more similar genetic background with one another than other *Rhamphochromis* species. This finding is consistent with the results from the ML phylogenetic analysis (Fig. 2A) and model-based analyses of population admixture ( $K=2$ ) (Fig. 2B).

To explore changes in effective population size ( $N_e$ ) of the ancestral populations for each of the genome-sequenced *Rhamphochromis* species, we used a pairwise sequentially Markovian coalescent analysis (PSMC) (Supplemental Figs. S6, S7; Li and Durbin 2011). The analysis was based on a mutation rate of  $3.5 \times 10^{-9}$  mutations per site per year and a generation time of 3 years (Malinsky et al. 2018). Our results revealed substantial



**Figure 1.** Sampling sites and body size variation in *Rhamphochromis* cichlid fish species. (A) Map of Lake Kingiri, Lake Chilingali, and Lake Malawi. Lake Kingiri has a diameter of ~600 m, and Lake Chilingali has a length of ~5.1 km. (B–D) Pictures from the shoreline of Lake Kingiri (B), Lake Chilingali (C), and Lake Malawi (D). (E) Variation of standard length and images across each of the 15 *Rhamphochromis* cichlid fish species. Relative scaling of fish pictures is to their approximate body sizes. The map of Africa is generated using data from GEBCO (<https://www.gebco.net/>) and Hydrolakes (<https://www.hydrolakes.org/products/hydrolakes>). Lake outlines redrawn from OpenStreetMap (© OpenStreetMap contributors, available under Open Database License ODbL).

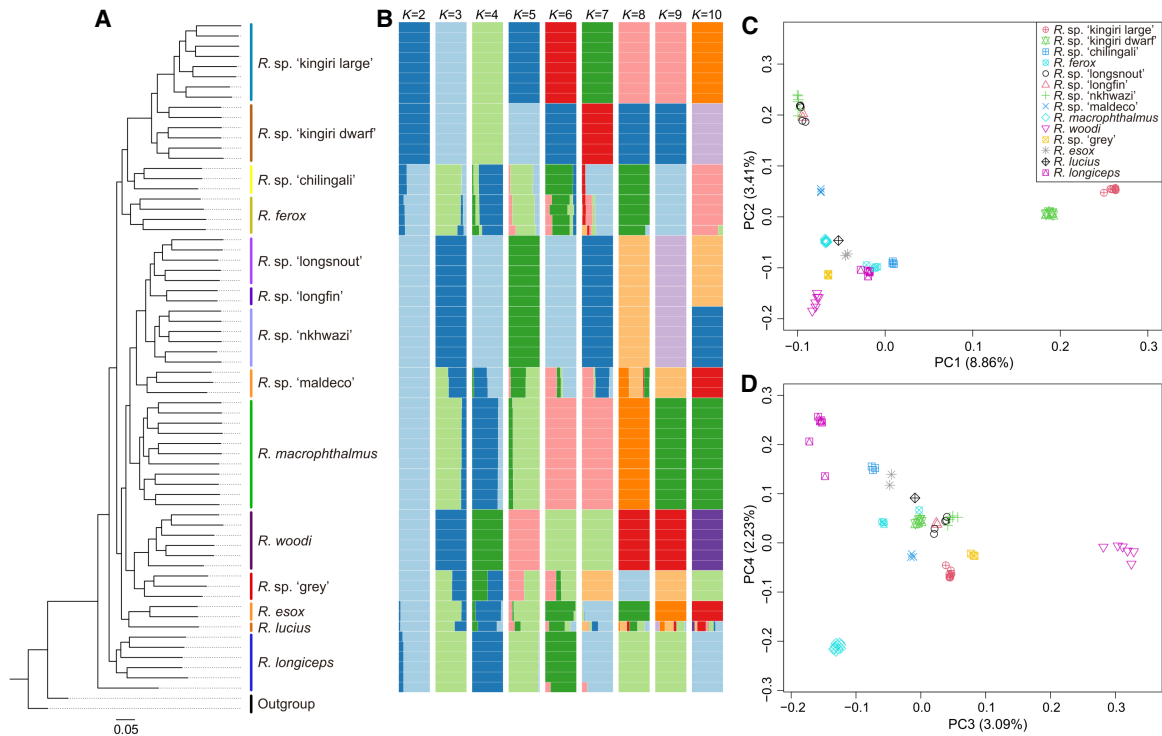
variation in effective population size across the different species, with *Rhamphochromis* sp. “kingiri large” and *Rhamphochromis* sp. “kingiri dwarf” from Lake Kingiri showing the smallest effective population sizes, less than  $1 \times 10^4$ , consistent with their isolation in the small lake (600 m in diameter). Notably, nearly all Lake Malawi species exhibited a period of rapid population expansion that initiated within the past 100,000 years (Supplemental Figs. S6, S7).

### Genome-wide association study of the body size evolution

To examine the genetic basis of body size evolution, we conducted a genome-wide association (GWA) mapping analysis on body size (standard length) across all 66 *Rhamphochromis* individuals, accounting for population structure by incorporating PCA covariates using PLINK v.1.9072 (Purcell et al. 2007). This analysis identified 1347 SNPs ( $P < 0.0001$ ) significantly associated with body size, linked to 570 genes (Fig. 3A; Supplemental Fig. S8; Supplemental Table S5). Notably, several of these genes, including *hoxb4a*, *hoxb3a*, *wnt11*, and *igf1ra*, are involved in developmental processes critical for embryonic growth and body size regulation. These genes are known for their roles in pathways controlling cell differentiation, tissue patterning, and growth, suggesting that body size evolution in *Rhamphochromis* is tightly linked to genetic networks governing early developmental stages and morphogenesis. To validate the robustness of our results using independent methods and data sets, we conducted GWA analyses using linear mixed models

(LMM) implemented using FaSt-LMM in GWASstic (Lück et al. 2024), using an automatically generated kinship matrix to account for population structure. One LMM analysis used standard length as the response variable, whereas a second LMM analysis using  $\log_{10}$ -transformed standard length as the response variable. The two independent analyses identified 2005 and 4481 SNPs ( $P < 0.0001$ ), respectively, significantly associated with body size (Supplemental Tables S6, S7). The overlap between these results and those obtained from PLINK was substantially higher than expected by chance (1347 standard-length PLINK SNPs, 2005 SNPs standard-length LMM SNPs, overlap 766 SNPs, hypergeometric test,  $P < 0.001$ ; 1347 standard-length PLINK SNPs, 4481 SNPs  $\log_{10}$  standard-length LMM SNPs, overlap 340 SNPs, hypergeometric test,  $P < 0.001$ ).

To investigate whether the GWA-identified genes were enriched for specific Gene Ontology (GO) terms, we identified one-to-one orthologs between *Astatotilapia calliptera* and zebrafish, before conducting GO enrichment analyses using DAVID (Fig. 3B; Supplemental Table S8). This analysis revealed that these 570 body size-associated genes are enriched in GO terms related to embryonic development and skeletal morphogenesis. Notable examples include “anterior/posterior pattern specification” (GO:0009952), “embryonic viscerocranium morphogenesis” (GO:0048703), “multicellular organism development” (GO:0007275), and “embryonic cranial skeleton morphogenesis” (GO:0048701). These findings underscore the pivotal roles of these genes in key developmental processes, highlighting their potential



**Figure 2.** Phylogeny and population structure analysis of *Rhamphochromis* species. (A) Maximum likelihood phylogenetic tree reconstructed using whole-genome single-nucleotide polymorphism (SNP) data from all 66 *Rhamphochromis* individuals and two outgroup taxa. Bootstrap values are available in Supplemental Figure S1, and all nodes separating each species have a bootstrap of 100%. (B) Population structure of *Rhamphochromis* estimated using the ADMIXTURE program with admixture proportions of individuals from  $K=2$  to  $K=10$ . Each color represents one population. Each individual is represented by a vertical bar, with the length of each colored segment indicating the proportion of genetic contribution from the respective ancestral population. (C,D) PCA plots of the first four components of *Rhamphochromis* individuals based on whole-genome SNP data.

involvement in the regulation of body size evolution. Moreover, the enrichment of these specific GO terms suggests that body size divergence in *Rhamphochromis* may be driven by modifications to conserved developmental pathways.

To further investigate the functional roles of the GWA-identified genes, we examined their expression patterns during the embryonic development of *Rhamphochromis* sp. “chilingali” using published expression data from whole embryos (Marconi et al. 2024). Notably, we found that body size–associated genes (GWA-identified genes) were significantly more likely to be expressed during development in *Rhamphochromis* sp. “chilingali” than other genes not associated with body size, across embryonic developmental stages (Fig. 3C; Supplemental Table S9). Additionally, the expression levels of GWA-identified genes were higher than those of other genes across embryonic developmental stages (Fig. 3D). These findings suggest that the genes associated with body size in *Rhamphochromis* species may take active roles during early development, which is key period for establishing body plan proportions in vertebrates.

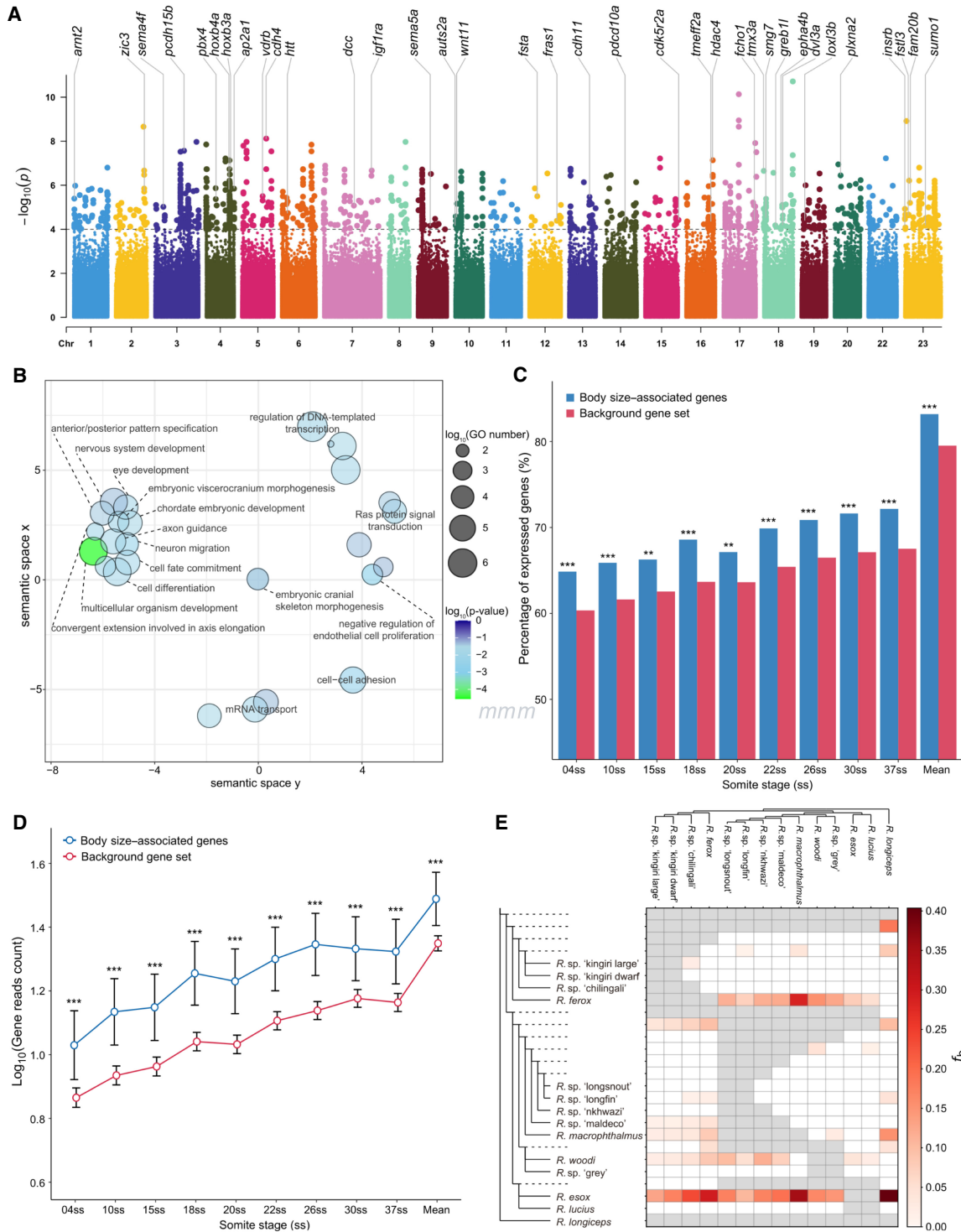
### Hybridization during the evolution of the genus *Rhamphochromis*

Discordance between phylogenetic trees constructed from nuclear and mitochondrial sequence data has been reported in many taxa (Gompert et al. 2008; Yang et al. 2021; Du et al. 2024), which may reflect a history of interspecific gene flow, such as hybridization, or result from processes like ILS (Rivas-González et al. 2023; Steenwyk et al. 2023). With complete genome resources now available, it is feasible to distinguish between the effects

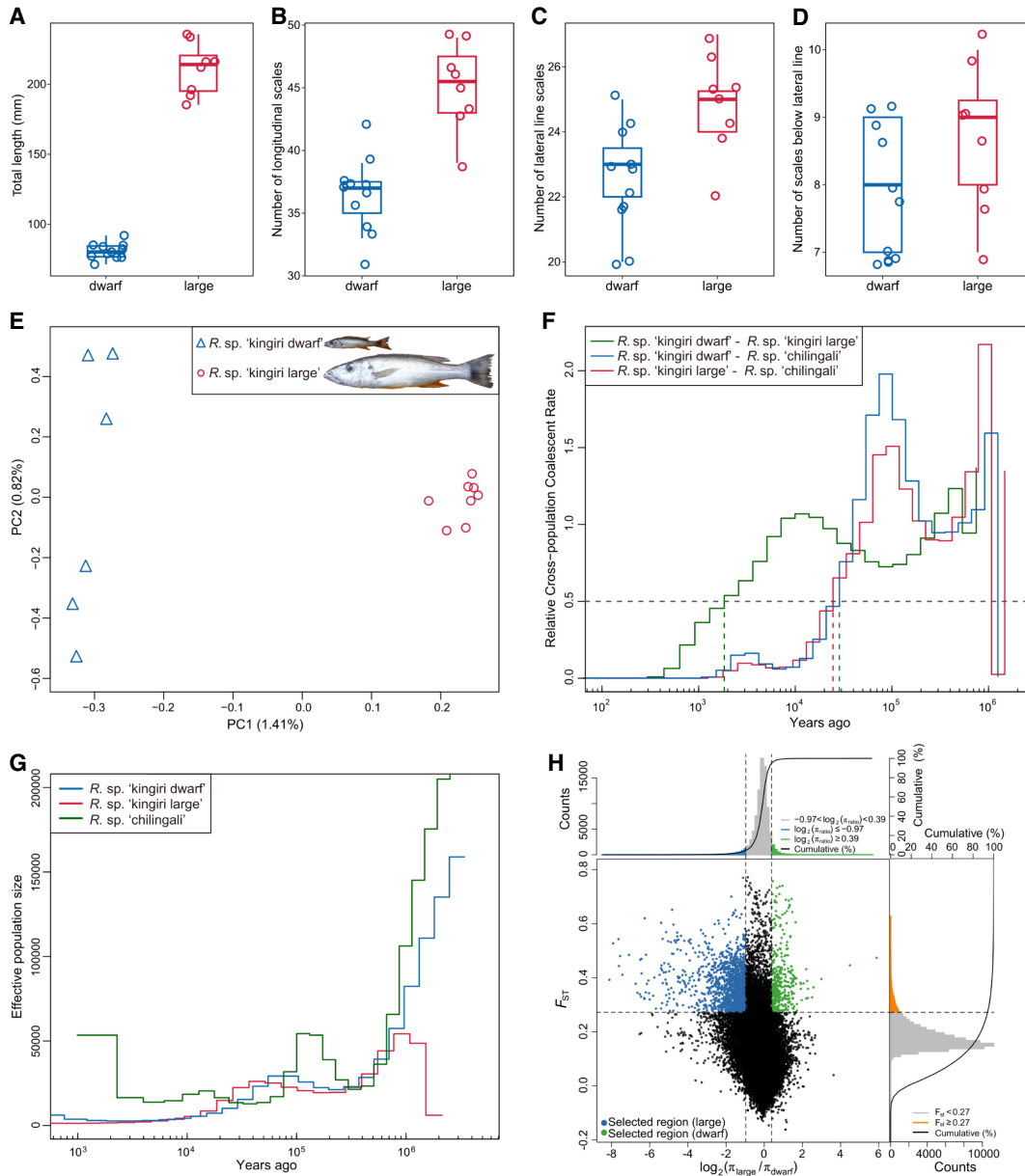
of hybridization and ILS by applying  $f_4$ -ratio statistics using the “*f*-branch” method (Malinsky et al. 2021). In *Rhamphochromis*, *f*-branch statistics revealed several cases of interspecific gene flow, highlighted by the orange-red shading in the heatmaps (Fig. 3E). Most notably, the nuclear genome of *R. esox* contained high admixture proportions with several groups, including *Rhamphochromis longiceps*, an ancestor of the lineage containing *R. ferox*, and an ancestor of the lineage containing *Rhamphochromis macrophthalmus* (Fig. 3E). This pattern was apparent in both sequenced individuals of *R. esox* (Supplemental Fig. S9), consistent with a hybrid origin of *R. esox* from multiple hybridization events. Other apparent hybridization events were apparent between *R. ferox* and *R. macrophthalmus* and between *R. longiceps* and *R. macrophthalmus* (Fig. 3E).

### Sympatric speciation of two species of *Rhamphochromis* in Lake Kingiri

Lake Kingiri is an isolated crater lake situated 21.3 km northwest from the northernmost point of Lake Malawi (~600 m diameter, 34 m deep) (Supplemental Fig. S10; Turner et al. 2019). The two *Rhamphochromis* species present differ in key phenotypic traits (Supplemental Fig. S11; Supplemental Table S2), with *Rhamphochromis* sp. “kingiri dwarf” being significantly smaller than *Rhamphochromis* sp. “kingiri large” (Wilcoxon signed-rank test,  $P=6.17 \times 10^{-8}$ ) (Fig. 4A). Additionally, *Rhamphochromis* sp. “kingiri dwarf” exhibits significantly fewer longitudinal scales, and significantly fewer upper lateral line scales, compared to



**Figure 3.** Association of skeletal and nervous system development with body size evolution. (A) GWA results for the body size for all 2,816,637 polymorphic SNPs within the *Rhamphochromis* radiation across 66 individuals. The  $y$ -axis corresponds to two-tailed  $-\log_{10}(P\text{-values})$ ; the  $x$ -axis corresponds to genomic coordinates (fAstCal1.2). The horizontal dashed line reflects the threshold for calling genome-wide significant SNPs ( $P=1 \times 10^{-4}$ ). (B) Gene Ontology (GO) scatterplot constructed using REVIGO for all the GO terms and genes within the identified significant SNPs for body size. Bubble size indicates the relative frequency of the GO terms, and bubble color represents  $\log_{10} P\text{-values}$  (smaller values indicate lower  $P\text{-values}$ ). (C,D) Transcriptomic analysis of the genes significantly associated with body size evolution during early embryogenesis of *Rhamphochromis* sp. “chilingali,” showing higher proportion of expression (C) and higher expression level (D) for genes associated with body size evolution. (E) Heatmap showing the estimated admixture fractions between pairs of species represented on the  $x$ - and  $y$ -axes. Species are arranged along the axes according to their placement in the species tree. The matrix values ( $f_b$ ) were calculated using Dsuite’s  $f$ -branch statistic, which quantifies excess allele sharing between the branch specified on the expanded tree along the  $y$ -axis (relative to its sister branch) and the species identified on the  $x$ -axis. These values provide insights into the hybrid origin of shared alleles. Gray squares indicate comparisons that cannot be made.



**Figure 4.** Morphological and genomic divergence between sympatric dwarf and large *Rhamphochromis* species in Lake Kingiri. (A–D) Relative to *Rhamphochromis* sp. “kingiri large,” *Rhamphochromis* sp. “kingiri dwarf” has smaller total length (A), fewer longitudinal scales (B), fewer lateral line scales (C), and fewer scales below the lateral line (D). (E) Principal component analysis for all Lake Kingiri *Rhamphochromis* using whole-genome SNP data. (F) MSMC2 cross-coalescence between *Rhamphochromis* sp. “kingiri dwarf” and *Rhamphochromis* sp. “kingiri large” individuals (green line), and between Lake Kingiri *Rhamphochromis* species and *Rhamphochromis* sp. “chilingali” (blue and red lines). Dotted lines indicate the split time for each population pair (relative cross-coalescence rates=0.5). (G) Demographic history and effective population size of the two *Rhamphochromis* species in Lake Kingiri (blue and red lines) and the species from Lake Chilingali (green line) using MSMC2. (H) Distribution of  $\log_2$  ratio of nucleotide diversity ( $\pi$ ;  $\pi_{\text{large}}/\pi_{\text{dwarf}}$ ) and  $F_{ST}$  of 50 kb windows with 10 kb steps. Colored data points indicate selected windows for *Rhamphochromis* sp. “kingiri large” (in blue) and *R. “kingiri dwarf”* (in green). Vertical lines reflect the 5% left and right tails of the  $\log_2(\pi_{\text{large}}/\pi_{\text{dwarf}})$  distribution (thresholds  $-0.97$  and  $0.39$ , respectively). The horizontal dashed line reflects the top 5% of windows in the  $F_{ST}$  distribution (threshold  $F_{ST}=0.27$ ).

*Rhamphochromis* sp. “kingiri large” (Wilcoxon signed-rank test,  $P=4.83 \times 10^{-5}$ ,  $P=0.0068$ , respectively) (Fig. 4B–D).

Analysis of sequences are consistent with the two species being reproductively isolated sympatric species pair (Fig. 4E; for admixture plots on  $K=5$  or higher, see Fig. 2B), with the genetic diversity of *Rhamphochromis* sp. “kingiri large” being slightly lower than that of *Rhamphochromis* sp. “kingiri dwarf” (Supplemental Fig. S12). Cross-coalescence rate analysis using the multiple

sequentially Markovian coalescent (MSMC2) indicated that the divergence between *Rhamphochromis* sp. “kingiri dwarf” and *Rhamphochromis* sp. “kingiri large” occurred <2000 years ago, whereas the divergence between the ancestor of the *Rhamphochromis* from Lake Kingiri and the ancestor of the *Rhamphochromis* from Lake Chilingali took place ~25,000 years ago (Fig. 4F). Effective population size estimates from MSMC2 revealed small population sizes for *Rhamphochromis* sp. “chilingali,”

*Rhamphochromis* sp. “kingiri dwarf,” and *Rhamphochromis* sp. “kingiri large,” with the smallest effective population size in the Lake Kingiri species (Fig. 4G), generally consistent with estimations using PSMC (Supplemental Figs. S6, S7).

### Genomic islands of speciation

Genome-wide divergence for *Rhamphochromis* sp. “kingiri dwarf” and *Rhamphochromis* sp. “kingiri large” was  $F_{ST}=0.067$ , with almost half of variable sites having zero  $F_{ST}$  (48.6%) (Supplemental Table S10). The maximum  $F_{ST}$  for a single SNP site is 6.2 SD above the mean  $F_{ST}$ , with 10,048 SNP sites having  $F_{ST}$  above 4 SD of mean  $F_{ST}$  (Supplemental Figs. S13, S14). Relative divergence ( $ZF_{ST}$ ) values identified 438 nonoverlapping genomic islands, merged into in 88 contiguous islands, between *Rhamphochromis* sp. “kingiri dwarf” and *Rhamphochromis* sp. “kingiri large.” Of these contiguous islands, 21 were  $\geq 100$  kb, and these 21 accounting for 48.84% of total genomic island length (Supplemental Table S11; Supplemental Fig. S14). In comparison to the background, genomic islands exhibited significantly higher levels of absolute divergence ( $D_{xy}$ ; Mann–Whitney  $U$  test;  $P < 2.2 \times 10^{-16}$ ) (Supplemental Figs. S14, S15). This aligns with speciation-with-gene-flow models, which predict that loci involved in speciation should display both high relative divergence ( $F_{ST}$ ) and high absolute sequence divergence ( $D_{xy}$ ) (Noor and Bennett 2009; Cruickshank and Hahn 2014; Malinsky et al. 2015). Similar patterns were observed regardless of whether the  $ZF_{ST}$  cutoff was set at 3.5 or three (Supplemental Table S11; Supplemental Fig. S16). Meanwhile, we found significantly reduced  $\pi$  for both Kingiri taxa in these islands compared with the rest of the genome (Mann–Whitney  $U$  test;  $P < 1.5 \times 10^{-5}$ ) (Supplemental Fig. S17).

### Selection signals in the lake Kingiri sympatric species pair

We combined  $F_{ST}$  and  $\pi$  analyses between the two sympatric taxa to investigate selection underlying the phenotypic divergence (Fig. 4H). In total, we identified 1081 genes from 1648 putatively selected regions in *Rhamphochromis* sp. “kingiri large” (Supplemental Table S12), considerably higher than the 242 genes from 401 putatively selected regions in *Rhamphochromis* sp. “kingiri dwarf” (Supplemental Table S13). This indicates that *Rhamphochromis* sp. “kingiri large” has experienced stronger divergent selection compared with *Rhamphochromis* sp. “kingiri dwarf,” an interpretation consistent with the strongly reduced  $\pi$  in *Rhamphochromis* sp. “kingiri large” (Supplemental Fig. S12). We also found that the overlap between these selected genes within the Lake Kingiri populations and the GWA-identified body size-associated genes was significantly greater than expected by chance. This was the case for both populations (1081 *Rhamphochromis* sp. “kingiri large” selected genes, 570 GWAS selected genes, overlap 47 genes, hypergeometric test,  $P = 2.76 \times 10^{-6}$ ; 242 *Rhamphochromis* sp. “kingiri dwarf” selected genes, 570 GWAS selected genes, overlap 16 genes, hypergeometric test,  $P = 6.31 \times 10^{-5}$ ).

Many genes with functional roles in tissue development and organ morphogenesis were identified as selected genes in *Rhamphochromis* sp. “kingiri large,” whereas only a few such genes were estimated to be under recent selection in *Rhamphochromis* sp. “kingiri dwarf” (Fig. 5A). For example, the genes, *fgfr11a*, *fgf21*, *wnt11*, and *fgfr10p*, which are all involved in organ development, were shown as under recent selection in *Rhamphochromis* sp. “kingiri large.” Consistent with this, GO analysis of these 1081 selected genes in *Rhamphochromis* sp. “kingiri large” indicated that

they were significantly enriched for regulation of cell migration, animal organ morphogenesis, axon guidance, and tissue development (Fig. 5B; Supplemental Table S14). However, only three GO terms were significantly enriched for the 242 selected genes in *Rhamphochromis* sp. “kingiri dwarf,” including nucleotide-excision repair, cell differentiation, and DNA repair (Fig. 5C; Supplemental Table S15). Among the selected genes, we observed that both  $F_{ST}$  and  $D_{xy}$  are particularly high in the *VAV3* genic region. Here, six missense single-nucleotide polymorphisms (SNPs) were also identified, exhibiting  $F_{ST}$ -values greater than 0.5, and XP-EHH scores below  $-4.8$  (Fig. 5D; Supplemental Fig. S18). The genetic diversity estimator  $\pi$  is significantly lower in *Rhamphochromis* sp. “kingiri large” for this gene, indicating it has recently been under selection in this species. Another gene with a strong signal of selection is *shank3a*.  $F_{ST}$ ,  $D_{xy}$ , and XP-EHH are all markedly elevated in the *shank3a* region, suggesting strong positive selection at this locus. Additionally, a significantly reduced nucleotide diversity ( $\pi$ ) is observed in *Rhamphochromis* sp. “kingiri dwarf,” indicates that the *shank3a* developmental gene is under strong selection in this species.

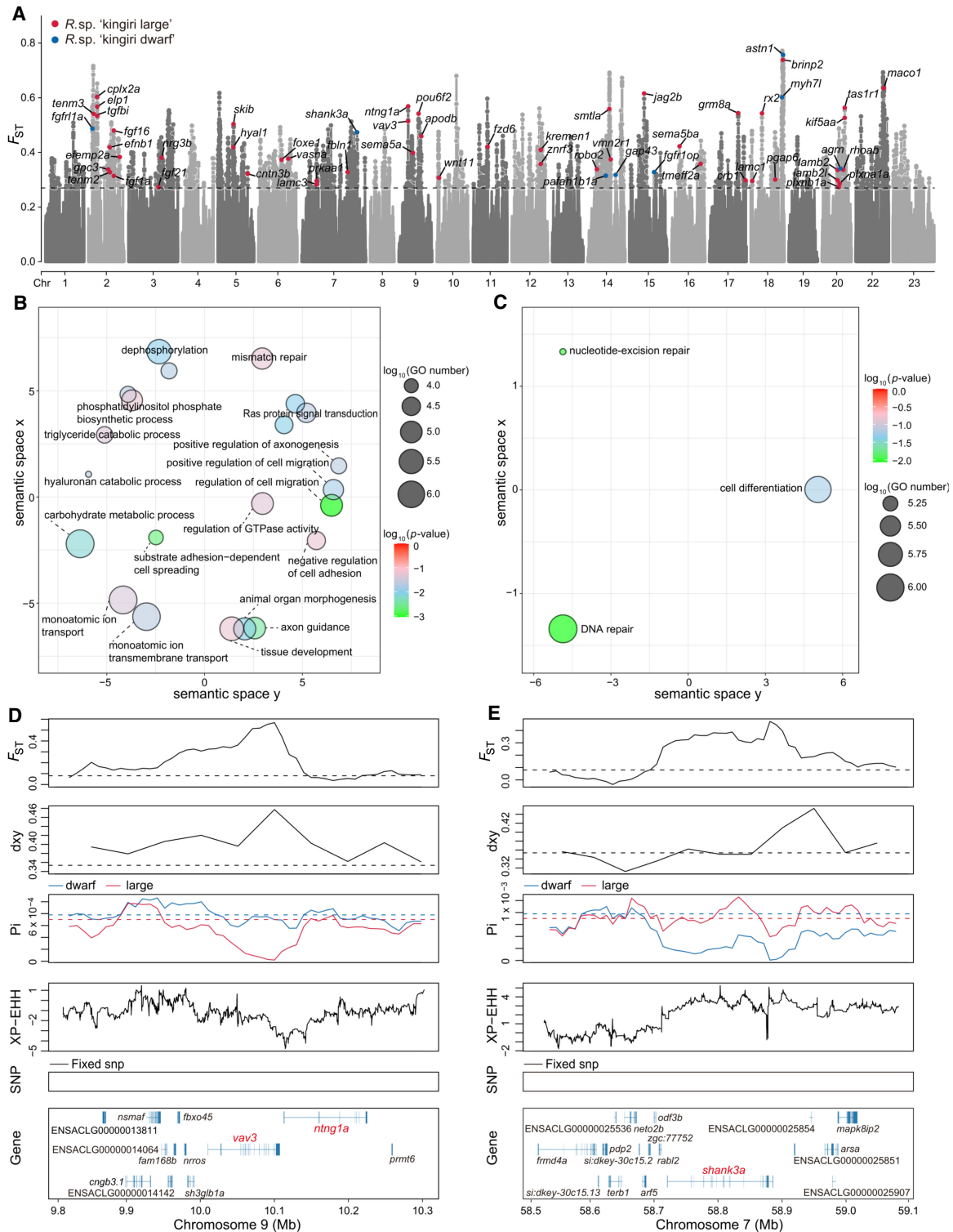
## Discussion

### Genetic structure in *Rhamphochromis*

To investigate how the radiation of this genus of pelagic predators unfolded, we generated the first comprehensive phylogenetic reconstruction using whole-genome data. Our analyses of nuclear genomes span 14 of the 15 species that comprise the radiation, the exception being the relatively rare *Rhamphochromis* sp. “longiceps yellow belly” (Genner et al. 2007; Konings 2016), from which we were unable to obtain sufficient nuclear genome data for inclusion. Our results clarify the species richness of the genus and are supported by combination of morphological characteristics, as well male breeding colors (Genner et al. 2007), which is meaningful given that *Rhamphochromis* species as a whole are considered one of the most challenging groups within the Lake Malawi cichlid radiation to reliably identify to species. In total, nine of the 15 species in the radiation remain undescribed, including the largest cichlid species in the Lake Malawi radiation, here referred to as *Rhamphochromis* sp. “nkhwazi,” and the one of the smallest species in the entire radiation, *Rhamphochromis* sp. “kingiri dwarf.” Our clear evidence of the genetic distinction of these nine species will inform taxonomic descriptions of the group, beneficial as it would enable IUCN conservation assessments. This is important as all *Rhamphochromis* in Lake Malawi are targeted species in commercial fisheries. Meanwhile, the satellite lake species are also highly threatened owing to their restricted ranges, including *Rhamphochromis* sp. “chilingali,” which is now believed to be extinct in the wild following the desiccation of Lake Chilingali in 2012 (Turner et al. 2023), and *Rhamphochromis* sp. “kingiri large,” which was the target of an artisanal gillnet fishery when last sampled in 2011 (Turner et al. 2019).

### Genetic basis of body size diversity in *Rhamphochromis*

Our phylogenetic analyses based on nuclear genomes reveal the evolution of body size diversity in *Rhamphochromis*. Specifically, we were able to reconstruct repeated shifts in body size, as indicated by the small-bodied species including *R. longiceps*, *R. macrophthalmus*, and *Rhamphochromis* sp. “chilingali” being dispersed across the phylogeny. This dynamic body size evolution is a broader characteristic of fish radiations in cichlids and is strongly



**Figure 5.** Selection analyses of the sympatric large and dwarf species of *Rhamphochromis* in Lake Kingiri. (A) Manhattan plot of the pairwise fixation index ( $F_{ST}$ ) between *Rhamphochromis* sp. “kingiri large” and *Rhamphochromis* sp. “kingiri dwarf,” with selected genes in *Rhamphochromis* sp. “kingiri large” in red and selected genes in *Rhamphochromis* sp. “kingiri dwarf” in blue. The dashed gray line represents the up 5%  $F_{ST}$  threshold of 0.27. (B, C) GO scatterplots constructed using REVIGO for all the GO terms and genes identified as selected genes in *Rhamphochromis* sp. “kingiri large” (B) and genes identified as selected genes in *Rhamphochromis* sp. “kingiri dwarf” (C). Bubble size indicates the relative frequency of the GO terms and bubble color represents  $\log_{10} P$ -values (smaller values indicate lower  $P$ -values). (D, E) Selective sweep on Chromosome 9 (9.8 to 10.3 Mb; D) and selective sweep on Chromosome 7 (58.5 to 59.1 Mb; E). The putative sweep region is validated by  $F_{ST}$ ,  $D_{xy}$ ,  $P_i$ , and XP-EHH test. Horizontal dashed lines represent the genome-wide means for the corresponding parameters. Gene annotations in the sweep region and SNPs nearly fixed for derived alleles are indicated at the bottom.

associated with diet, including in *Rhamphochromis*. The smallest *Rhamphochromis*, including *Rhamphochromis* sp. “kingiri dwarf” and *Rhamphochromis* sp. “chilingali” feed upon zooplankton and very small fish. Meanwhile, medium-size species such as *R. longiceps* forage on zooplankton and shoaling cyprinids (Allison et al. 1996). The largest of all *Rhamphochromis* are specialist cichlid predators, feeding primarily on deep water species such as *Diplotaxodon* (Allison et al. 1996). The strong link between body size and diet in *Rhamphochromis* is therefore reflective of other fish radiations with notable body size diversity, such as Arctic charr (Malmquist et al. 1992) and coregonid whitefishes (Østbye et al. 2006).

The extensive shared genetic variation among species within African cichlid radiations makes them useful for the identification of genetic variants associated with phenotypic traits (Sommer-Trembo et al. 2024). Our GWA analysis identified numerous genes that were strongly associated with body size in *Rhamphochromis*, consistent with a highly polygenic basis to body size divergence in cichlids, as observed in other vertebrate groups, including mammals (Bouwman et al. 2018) and birds (Luo et al. 2025). It is plausible that many of the genes we identified are only indirectly associated with body size, for example, being associated with co-varying environmental variables including water depth or diet. Nevertheless, because our analyses identified suites of genes associated with morphogenesis and embryonic development and because those genes tended to exhibit elevated levels of expression during larval development, then we have confidence that our analysis has identified relevant candidate genes underpinning body size evolution. Importantly, multiple candidate genes identified in this study have previously been identified in body size regulation in other studies of vertebrates. For example, *hoxb3a* is expressed in skeletal development of three-spined stickleback *Gasterosteus aculeatus* (Ahn and Gibson 1999) and has been invoked as a potential contributor to the development of a giant phenotype of the Siamese fighting fish *Betta splendens* (Wang et al. 2022). Meanwhile *igf1ra*, an insulin-like growth factor receptor orthologous to *igf1r*, is a key regulator of morphological development in zebrafish (Schlueter et al. 2007) and has been found to be under selection in the giant ocean sunfish *Mola mola* (Pan et al. 2016), as well as being linked to growth variants in the striped catfish *Pangasianodon hypophthalmus* (Tran et al. 2021). Moreover, the insulin-like growth factor 1, which activates *igf1r*, is known to be a primary regular of growth hormone, as well as a mediator of body size in multiple vertebrate groups (Sutter et al. 2007; Kappeler et al. 2008). Collectively, this GWA analysis had indicated considerable potential for further investigation of the body size evolution using *Rhamphochromis*, as well as potentially other haplochromine cichlid radiations, that could be validated using QTL-mapping approaches and genetic manipulation methods (Santos et al. 2023).

### Genomic insights into the evolution of the Lake Malawi *Rhamphochromis* radiation

The genus *Rhamphochromis* is part of a distinct pelagic-dwelling subtribe within the Lake Malawi radiation. Most of the species-specific signals of population expansion, as inferred by PSMC analyses, began ~400,000 years ago, potentially indicative of the start of diversification within the subtribe. This time line broadly fits with the upper limits of the estimated start of the radiation dating to 860,000–990,000 years ago (Malinsky et al. 2018). Notably most of the main lake populations show expansions over the past

100,000 years, coinciding with a major rise in the level of Lake Malawi to present levels (Lyons et al. 2015), potentially reflecting pelagic habitat expansion and novel ecological opportunity after a severe lowstand.

Our study has revealed several instances of admixture within *Rhamphochromis*, for example, between *R. esox* and *R. longiceps* (also seen in Malinsky et al. 2018) and between *R. ferox* and *R. macropthalmus*. These events have plausibly led to sharing of adaptive alleles, potentially promoting radiation in the lineage, as has been proposed in other examples of ancestral gene flow in Lake Victoria (Meier et al. 2023), as well as Lake Malawi (Malinsky et al. 2018). Although cases of admixture between major lineages of cichlid fishes in Lake Malawi have been revealed through whole-genome sequencing (Malinsky et al. 2018; Blumer et al. 2025; Kumar et al. 2025), there is currently no evidence of horizontal gene flow between *Rhamphochromis* and other Lake Malawi cichlid lineages (Malinsky et al. 2018), perhaps reflective of the distinctive pelagic predator ecology of *Rhamphochromis* and their ability to breed in open water, which is uncommon in Malawi cichlids.

### Sympatric speciation in Lake Kingiri

The two sympatric *Rhamphochromis* species in Lake Kingiri, *Rhamphochromis* sp. “kingiri large” and *Rhamphochromis* sp. “kingiri dwarf,” were found to be morphologically and genetically divergent sister species. Both species had very small effective population sizes, including evidence of historic reductions in the population size of the ancestral species which may be expected following containment in the crater lake. Notably, we estimated these have diverged in the past 2000 years from common ancestry. The precise age of Lake Kingiri is not known, but it is one of several maar crater lakes in the region, and the geographically proximate Lake Masoko has been dated to 50,000 years old based on carbon dating of drill core material (Williamson et al. 1999), and it is plausible Lake Kingiri was formed during the same period of volcanic activity.

The mechanism of colonization of Lake Kingiri is unclear, but it was likely colonized directly from the proximate River Mguwesi. Lake Kingiri currently has a linear distance of 16 km from the current shoreline of Lake Malawi and, with a surface elevation of 523 m, is currently ~50 m above the current Lake Malawi level of ~473 m. Lacustrine sediments (raised beaches) provide evidence of Lake Malawi stands ~40 m higher than present-day levels in the upper Pleistocene (Pike 1958; Clark et al. 1966). Hence, Lake Malawi was historically in very close proximity to Lake Kingiri, which may have facilitated colonization of the crater lake by the ancestor of the *Rhamphochromis* species pair during or after the separation from *Rhamphochromis* sp. “chilingali” ~25,000 years ago.

Although it is possible that the Lake Kingiri was colonized on multiple occasions, each time leading to a speciation event from a common Lake Malawi ancestor, most parsimoniously, the divergence was a result from one common ancestor, as this would require only one extinction event of a main lake ancestor since separation from *Rhamphochromis* sp. “chilingali.” Demographic modeling may help to clarify the most likely evolutionary scenario (Kautt et al. 2016). Importantly, neither of the Lake Kingiri species shows any clear signs of admixture with Lake Malawi or Lake Chilingali species. The slightly closer position of *Rhamphochromis* sp. “kingiri dwarf” compared with *Rhamphochromis* sp. “kingiri large” to the most genetically similar species *Rhamphochromis* sp. “chilingali” in the PCA plot is consistent with differential influence of genetic drift or selection on the populations after divergence from

*Rhamphochromis* sp. “chilingali.” Notably *R. “kingiri dwarf”* has a higher effective population size, and from a phylogenetic perspective, the nested position of *Rhamphochromis* sp. “kingiri large” within *Rhamphochromis “kingiri dwarf”* is indicative that *Rhamphochromis* sp. “kingiri large” may be the derived species.

The Lake Kingiri speciation event is remarkable for the divergence in body size when contrasted with other cichlid fish crater lake speciation events. In nearby Lake Masoko, the two ecotypes of *A. calliptera* have diverged most patently in trophic morphology and male breeding colors rather than size (Malinsky et al. 2015). Similarly, species from the crater lake cichlid radiations in Cameroon and Nicaragua tend to vary primarily in body shape, as well as tooth and jaw morphology, with conspicuous contrasts in adult body size less common (Trewavas et al. 1972; Stauffer et al. 2008; Dunz and Schliewen 2010). We speculate that the speciation event in Lake Kingiri has followed similar selection pressures to the size-based diversification of predatory ecotypes seen in coregonid whitefish (Lu and Bernatchez 1999) and Arctic charr (Doenz et al. 2019). In those systems, the availability of distinct prey fields, coupled with strong competition, is proposed to lead to selection favoring multiple distinct size-differentiated ecotypes (Salisbury and Ruzzante 2022). Speciation may further be promoted by targeted selection, including from predators (Öhlund et al. 2020), leading ecotypes to occupy different breeding habitats or have asynchronous breeding schedules. In Lake Kingiri, adult *Rhamphochromis* sp. “kingiri dwarf” are commonplace in shallow littoral habitat, whereas adult *Rhamphochromis* sp. “kingiri large” have only been captured offshore deeper water, suggestive of different patterns of habitat occupancy of breeding fish. We also speculate that this reproductive isolation may be reinforced if larger individuals are perceived as potential predators by breeding *Rhamphochromis* sp. “kingiri dwarf” and if smaller individuals are perceived as potential prey by breeding *Rhamphochromis* sp. “kingiri large.”

We observed islands of genomic differentiation between *Rhamphochromis* sp. “kingiri dwarf” and *Rhamphochromis* sp. “kingiri large,” in a pattern consistent with a relative recent divergence driven by selection (Malinsky et al. 2015). Our subsequent analyses showed stronger selection operating on *Rhamphochromis* sp. “kingiri large” relative to *Rhamphochromis* sp. “kingiri dwarf.” This is compatible with the concept that *Rhamphochromis* sp. “kingiri large” has evolved from a small-bodied ancestor, which perhaps was phenotypically similar to the closely related dwarf species *Rhamphochromis* sp. “chilingali.” One interpretation of the evidence is that an ancestral small-bodied species once had a much wider distribution around Lake Malawi and was specialized for shallow marginal habitats but was ultimately extirpated from the main lake body. However, this ancestral lineage persisted in satellite Lake Chilingali and Lake Kingiri, finding ecological opportunity to speciate only in the deeper Lake Kingiri.

Our analyses revealed large numbers of genes experiencing selection in Lake Kingiri. Given overlap with the genes identified in GWA analyses, many are likely to be associated with divergence in body size, whereas others may be associated with other characteristics linked to their speciation. Genes under selection were linked to skeletal, organ, and neurological development. A number of genes with evidence of selection were associated with fibroblast growth factors, including *fgfr11a*, a gene linked to gill cartilage development in zebrafish (Hall et al. 2006), and *fgf2*, a gene known to determine body size in Asian seabass *Lates calcarifer* (Wang et al. 2014). A further gene highlighted by the analysis was *Wnt11*, a regulator anterior-posterior axis elongation in mammals (Andre et al.

2015). We also found strong evidence for *VAV3* and *shank3a* being under positive selection in *Rhamphochromis* sp. “kingiri large” and *Rhamphochromis* sp. “kingiri dwarf,” respectively. *VAV3* encodes a guanine nucleotide exchange factor (GEF) for Rho family GTPases, which regulate key cellular processes such as cell growth, transformation, and development. In neurons, it plays a crucial role in axon elongation and dendrite formation (Sauzeau et al. 2010). Meanwhile, *shank3a* has a crucial role in brain development and function, particularly in synapse formation and maturation. Mutations in this gene have been associated with autism spectrum disorder (ASD) and Phelan–McDermid syndrome (Liu et al. 2022). Collectively, these analyses have provided multiple candidate genes suitable for further research into links between functional genetic variation and divergent phenotypic development.

Although limited in the number of sequences studied for each species, the use of only one reference genome for mapping, and our focus on body size, our study provides unprecedented insight into the evolution of a radiation of pelagic cichlid fishes. In particular, our study has highlighted how rapid speciation in the lineage associates with body size divergence. To date, the genetic basis of adaptive evolution in body size of cichlid fishes has been largely overlooked, yet this appears to be a key axis of ecological diversification across radiations in Lake Malawi, Lake Victoria, and Lake Tanganyika, as well as teleost fishes more broadly. There remains a considerable amount to learn about the ecology and adaptive radiation of pelagic cichlids, including their biogeography, habitat preferences, breeding behavior, and the genetic basis of other phenotypic traits, both in Lakes Malawi and Kingiri. Nonetheless, our results provide fundamental insight into their diversity, which will support future research into this ecologically, economically, and culturally important group of Lake Malawi cichlids.

## Methods

### Field sampling

*Rhamphochromis* samples from the Lake Malawi catchment area, including Lake Kingiri, Lake Chilingali, and Lake Malawi, were collected during multiple expeditions between 2004 and 2011. Fish were captured using fixed gill nets and scuba or obtained from local fishermen. Sampling was focused on individuals that were identifiable to species, based on anatomical and color traits. Upon retrieval, the fish were euthanized with an overdose of anesthetic (MS-222). For genetic analysis, fin clips were collected from each fish and preserved in ethanol. Sequence data were obtained at the level of the individual fish. For morphological measurements, whole specimens were fixed in ethanol or ~4% formalin.

### Morphological analysis

For *Rhamphochromis* samples, standard length was measured, and the number of dorsal spines, dorsal rays, and anal rays were counted (Supplemental Table S1). For *Rhamphochromis* sp. “kingiri dwarf” and *Rhamphochromis* sp. “kingiri large,” additional meristic counts and length measurements were taken for 19 specimens: 11 *Rhamphochromis* sp. “kingiri dwarf” and eight *Rhamphochromis* sp. “kingiri large,” following the method of Snoeks (2004). These were total length, standard length, number of lateral line scales (longitudinal scales), scales above lateral line (upper transversal line scales), scales below lateral line (lower transversal line scales), lower jaw total outer tooth count, dorsal spine count, dorsal soft ray count, and anal spine count (Supplemental Table S2; following the method of Snoeks 2004). Length measurements were taken using digital calipers.

## DNA extraction and sequencing

Standard Qiagen DNeasy blood and tissue protocol was used to extract DNA from fin clips preserved in ethanol between 2004 and 2011. To ensure high-molecular-weight DNA was present in the final extraction products, samples were analyzed using Qubit, NanoDrop, and finally gel electrophoresis. High-quality samples were then chosen for whole-genome sequencing, ensuring an adequate scope across the *Rhamphochromis* genus. Final extracted products were then sequenced by Novogene using their standard whole-genome library prep and protocol. Sequencing was done using Illumina technology (150 bp paired-end) to a mean coverage of 10.3× (Supplemental Table S3). As the coverage of the sample of *Rhamphochromis* sp. “longiceps yellow belly” was only about 0.7× coverage, it was removed for whole-genome phylogenetic analysis and only used for mitochondrial genome analysis.

## Read mapping and variant calling

A total of 48 *Rhamphochromis* individuals were sequenced in this study, with an additional 21 obtained from NCBI: 19 *Rhamphochromis* and two outgroups (Supplemental Table S3). Low-quality reads were first filtered using fastp with default parameters (Chen et al. 2018). Then, the high-quality reads were aligned to the reference genome of *A. calliptera* (fAstCal1.2) from Ensembl (Release 111) using Burrows–Wheeler aligner (BWA) v0.7.16a-r1181 (Li and Durbin 2009), with the parameter “bwa mem -t 10 R.” This reference species is another member of the Lake Malawi radiation and should be approximately genetically equidistant to all representatives of the focal genus *Rhamphochromis* limiting reference bias. SAMtools v1.9 (Li and Durbin 2011) was used to sort and merge files, as well as to convert mapping outputs to binary alignment map (BAM) format. Subsequent steps, including local realignment and duplicate read marking, were carried out using Picard tools (version 2.1.1; <http://broadinstitute.github.io/picard/>). Variant calling for each sample was conducted with the Genome Analysis Toolkit (GATK) v4.1.1.0 (McKenna et al. 2010). Specifically, GATK’s HaplotypeCaller was employed to identify variants for each accession. Additionally, a joint genotyping process was performed on genome variant call format (gVCF) files to produce a unified set of variations. To minimize the occurrence of false-positive SNPs, variants were filtered using GATK v4.1.1.0 (McKenna et al. 2010) and VCFtools v1.13 (Danecek et al. 2011) with the cutoff of “QUAL < 30.0 || QD < 2.0 || FS > 60.0 || MQ < 40.0 || SOR > 4.0” to discard site types probably caused by mapping bias and following cutoffs: minimum depth > 20, minor allele frequency (MAF) > 0.05, missing rate < 0.5, and only keep biallelic sites. Finally, a total of 2,816,637 sites that passed all these filtering criteria were used in further phylogenetic analyses.

## Phylogenetics, admixture, and PCA analyses

A phylogenetic approach was first used to investigate the relationships between each of the *Rhamphochromis* species. The filtered VCF variant file was converted to a PHYLIP file using vcf2phylip v.277 (<https://zenodo.org/records/2540861>), and the ML tree was constructed using iqtree2 (Minh et al. 2020) with default parameters. FigTree v.1.4.3 (<http://tree.bio.ed.ac.uk/software/figtree/>) was used to display the ML tree. Admixture analysis was assessed using the default setting in the ADMIXTURE v.1.3.0 (Alexander et al. 2009) with the number of assumed genetic clusters  $K$  ranging from two to 12, specifying 20 cross-validations ( $-cv = 20$ ). PCA for all 66 individuals was performed by PLINK v.1.9072 (Purcell et al. 2007) with the parameter “-pca.”

## Phylogenetic relationships of mitochondrial genomes

The mitochondrial genome of each individual was de novo assembled using MitoFinder (Allio et al. 2020), with *A. calliptera* (obtained from the NCBI Nucleotide database [<https://www.ncbi.nlm.nih.gov/nucleotide/>] under accession NC\_018560) set as a starting reference. For information on the completeness of the mitochondrial genome assemblies for each lineage, see Supplemental Table S4. Mitochondrial genomes were annotated using MitoAnnotator (Iwasaki et al. 2013). Each of the 37 mitochondrial genes was aligned separately using MUSCLE (Edgar 2004). For ML phylogenetic reconstruction, we used IQ-TREE 2 (Minh et al. 2020) with 1000 bootstrap replicates.

## Demographic history

The PSMC model (Li and Durbin 2011) was employed to infer the history of the ancestral population dynamics of *Rhamphochromis*. PSMC reconstructed the historical changes in effective population size over time by analyzing the distribution of the most recent common ancestor between two alleles of an individual, with the parameter of “-u  $3.5 \times 10^{-9}$  -g 3” (Malinsky et al. 2018). For *Rhamphochromis* sp. “kingiri dwarf” and *Rhamphochromis* sp. “kingiri large,” additionally, the multiple sequential coalescent Markovian model (MSMC) method (Schiffels and Durbin 2014) was used to estimate historical changes in effective population sizes. The input files for MSMC were generated using msmc-tools (<https://github.com/stschiff/msmc-tools>). For each of the 14 individuals (six *Rhamphochromis* sp. “kingiri large” and eight *Rhamphochromis* sp. “kingiri dwarf”), only sites with uniquely mapped reads and sites with coverage depths  $\geq 5\times$  were used for the analyses. All sites were phased using Beagle v5.2 (Browning and Browning 2007). For effective population size inference, three individuals (six phased haplotypes) from each species were used. We also used the MSMC2 to calculate the relative cross-population coalescent rate (rCCR) with default parameter. rCCR of 0.5 was considered as an indicator of population split.

## GWA mapping

To investigate the genetic basis of body size variation across all *Rhamphochromis*, an association test was performed using PLINK v1.90 (Purcell et al. 2007). Briefly, the VCF file was first pruned to remove SNPs in strong linkage disequilibrium ( $r^2 > 0.3$ ) in every 2 kb window to keep only one SNP. Then PCA was performed for the pruned VCF file, and PC1 was extracted to represent the main axis of population genetic structure. Finally, an association test was calculated with the median value of length of body size (standard length) for each *Rhamphochromis* species as the response variable, as well as genomic PC1 as a covariate to account for population genetic structure.

## Association between SNPs and genes

We associated genes with SNPs from the GWA analysis if the SNPs were within the gene body or within  $\pm 5$  kb from the transcription start and end sites. If no genic region overlapped the lead SNP, then we associated them with the closest genes within 10 kb.

## $f$ -Branch statistics

To calculate excess allele-sharing across the data set and assess whether different species have evolved in the presence of interspecific gene flow, we computed the  $f$ -branch statistic  $f_b(C)$  using Dsuite v.0.386 (Malinsky et al. 2021). First, we simplified the full RAXML phylogenetic tree with the R package ape (Paradis et al. 2004) to account for multiple individuals per species. We then

applied Dsuite to calculate branch-specific estimates of gene flow, referred to as the “ $f$ -branch statistic.” Next, we calculated  $f_4$ -ratios based on the relationships from the phylogenetic tree for all populations using the Dtrios program within Dsuite. The  $f$ -branch statistic for each phylogenetic branch was then estimated with the Fbranch program within Dsuite, using the parameter “-p 0.05.” Finally, we visualized the resulting  $f$ -branch statistics using the dtools.py script, again part of Dsuite.

### RNA sequencing data analysis

Whole-embryo bulk RNA sequencing data from early stages of *Rhamphochromis* sp. “chilingali” development were obtained from a recent study (Marconi et al. 2024). We first removed the adapter sequences and filtered the low-quality sequences (Phred < 20) using Trim Galore! v0.6.6 (<https://github.com/FelixKrueger/TrimGalore>). Then high-quality RNA-seq reads were mapped to the *A. calliptera* genome assembly (fAstCall.2 in Ensembl 111), using STAR v.2.7.1a (Dobin et al. 2013) and only uniquely mapped reads retained for subsequent analysis. Finally, the number of reads mapped to each gene in the *A. calliptera* reference genome was counted in STAR using the built-in HTSeq-count option (Anders et al. 2015).

### Population genetics analysis

To compare estimates of genetic diversity of the two Lake Kingiri species, VCFtools v0.1.17 (Danecek et al. 2011) was used to calculate the average pairwise nucleotide diversity ( $\pi$ ) and Tajima’s D statistics within 50 kb sliding windows of 10 kb steps. To estimate the LD patterns within species, we calculated the mean  $r^2$ -values for pairwise markers with a MAF greater than 0.05 using PopLDdecay (Zhang et al. 2019). These parameters were set to “-MaxDist 500 -MAF 0.05 -Het 0.88 -Miss 0.2.” The population scaled recombination rate ( $\rho = 4Ner$ ) was calculated using FastEPFR (Gao et al. 2016).

### Genome-wide patterns of divergence for genomic islands

Genomic differentiation between the Kingiri fish pair was assessed using relative divergence ( $F_{ST}$ ) and absolute divergence ( $D_{xy}$ ) across the genome, calculated in 50 kb nonoverlapping windows. This analysis was performed using pixy v.1.2.11. beta1 (Korunes and Samuk 2021). To enable comparisons of genomic divergence landscapes across pairs with varying divergence times, per-window  $F_{ST}$ -values were standardized to Z-transformed values ( $ZF_{ST}$ ) using the formula:  $ZF_{ST} = (F_{ST} - \mu F_{ST}) / \sigma F_{ST}$ , where  $F_{ST}$  represents the window-specific value,  $\mu F_{ST}$  is the mean  $F_{ST}$  across all windows, and  $\sigma F_{ST}$  is the standard deviation of  $F_{ST}$ -values across the genome (Karlsson et al. 2007). Genomic regions with  $ZF_{ST} \geq 4$  were classified as genomic islands, with thresholds of  $ZF_{ST} \geq 3.5$  and  $ZF_{ST} \geq 3$  also being tested. Adjacent significant windows were merged into contiguous regions. The significance of  $D_{xy}$ ,  $\pi$ , and  $\rho$  within genomic islands was evaluated relative to the genomic background using the Wilcoxon rank-sum test.

### Identification of selected genomic regions

To identify the potential regions under selection for each of the Kingiri taxa, nucleotide diversity ( $\pi$ ) and population fixation statistics ( $F_{ST}$ ) were calculated using VCFtools v0.1.17 (Danecek et al. 2011) in a 50 kb sliding window with a step size of 10 kb. We filtered the window with the top 5% of  $F_{ST}$ -values ( $F_{ST} \geq 0.27$ ) as the outlier windows. On this basis,  $\log_2$  ratio values were used ( $\log_2$  ratio of  $\pi$ -values:  $\log_2(\pi[\text{large}]/\pi[\text{dwarf}]) \geq 0.39$ , top 5%;  $\log_2(\pi[\text{GS}]/\pi[\text{GE}]) \leq -0.97$ , bottom 5%) to identify selected

genomic regions in *Rhamphochromis* sp. “kingiri dwarf” and *Rhamphochromis* sp. “kingiri large,” respectively.

We also employed the integrated haplotype score (iHS) statistic (Voight et al. 2006) to detect genomic regions characterized by exceptionally long haplotypes in comparison to the rest of the genome to detect selective sweeps. Strongly negative iHS values signified increased haplotype homozygosity driven by a rise in the frequency of the derived allele, whereas positive values indicated a high frequency of the ancestral allele. These analyses were performed using the R package rehh v.3.2.2 (Gautier and Vitalis 2012).

We identified outlier  $F_{ST}$  windows across a 50 kb sliding window with a step size of 10 kb. We then identified genes overlapping with the outlier windows as outlier  $F_{ST}$ -associated genes. We note this approach will identify genes under selection and also those genes in linkage disequilibrium with those genes under selection.

### GO enrichment analysis

We used the ortholog between *A. calliptera* and zebrafish to perform GO enrichment analysis, as zebrafish has the most extensive functional gene annotation. One-to-one orthologs between *A. calliptera* and zebrafish were downloaded from Ensembl (release 111) using BioMart database. In total, there were 15,158 one-to-one orthologs of the two species, from 28,001 *A. calliptera* genes and 37,241 zebrafish genes. The GO and KEGG enrichment analysis were carried out using the online tool DAVID (Sherman et al. 2022; <https://david.ncifcrf.gov/>) using the default statistical parameters. Significantly enriched GO terms were further imported to the online tool REVIGO to identify the most connected and represented GO terms (Supek et al. 2011).

### Data access

The sequence data generated in this study have been submitted to the NCBI BioProject database (<https://www.ncbi.nlm.nih.gov/bioproject/>) under accession number PRJNA1241418.

### Competing interest statement

The authors declare no competing interests.

### Acknowledgments

Sample collections were supported by Royal Society-Leverhulme Trust Africa awards (AA100023 and AA130107, to M.J.G., B.P.N., and G.F.T.) and the Natural Environment Research Council (NER/A/S/2003/00362, to G.F.T.). During analysis and writing, L.Y. was supported by the China Scholarship Council (202104910171), National Natural Science Foundation of China (32422010), and Youth Innovation Promotion Association, Chinese Academy of Sciences (<http://www.yicas.cn>), and M.J.G. was supported by the Natural Environment Research Council (NE/S001794/1). We thank researchers from the Tanzania Fisheries Research Institute (TAFIRI) and the Fisheries Research Unit of the Government of Malawi for their considerable support. We thank Alan Smith for field support and photographs of the Lake Kingiri species.

**Author contributions:** M.J.G. conceived and designed the project. M.J.G., B.P.N., A.H.S., and G.F.T. collected the fish samples. H.S. prepared samples for genome sequencing, measured morphological traits, and undertook preliminary analyses. L.Y. performed all data processing and analysis shown in the manuscript. L.Y., G.F.T., and M.J.G. interpreted the results, and L.Y. and M.J.G. wrote the manuscript.

## References

- Ahn DG, Gibson G. 1999. Expression patterns of threespine stickleback *Hox* genes and insights into the evolution of the vertebrate body axis. *Dev Genes Evol* **209**: 482–494. doi:10.1007/s004270050281
- Alexander DH, Novembre J, Lange K. 2009. Fast model-based estimation of ancestry in unrelated individuals. *Genome Res* **19**: 1655–1664. doi:10.1101/gr.094052.109
- Allio R, Schomaker-Bastos A, Romiguier J, Prosdociimi F, Nabholz B, Delsuc F. 2020. MitoFinder: efficient automated large-scale extraction of mitochondrial data in target enrichment phylogenomics. *Mol Ecol Resour* **20**: 892–905. doi:10.1111/1755-0998.13160
- Allison EH, Irvine K, Thompson AB, Ngatunga BP. 1996. Diets and food consumption rates of pelagic fish in Lake Malawi, Africa. *Freshw Biol* **35**: 489–515. doi:10.1111/j.1365-2427.1996.tb01764.x
- Anders S, Pyl PT, Huber W. 2015. HTSeq—a Python framework to work with high-throughput sequencing data. *Bioinformatics* **31**: 166–169. doi:10.1093/bioinformatics/btu638
- Andre P, Song H, Kim W, Kispert A, Yang Y. 2015. Wnt5a and Wnt11 regulate mammalian anterior-posterior axis elongation. *Development* **142**: 1516–1527. doi:10.1242/dev.119065
- Blumer M, Burskaia V, Artiushin I, Saha J, Garcia JC, Elkin J, Fischer B, Zhou C, Gresham S, Malinsky M, et al. 2025. Introgression dynamics of sex-linked chromosomal inversions shape the Malawi cichlid adaptive radiation. *Science* **388**: eadr9961. doi:10.1126/science.adr9961
- Bouwman AC, Daetwyler HD, Chamberlain AJ, Ponce CH, Sargolzaei M, Schenkel FS, Sahana G, Govignon-Gion A, Boitard S, Dolezal M, et al. 2018. Meta-analysis of genome-wide association studies for cattle stature identifies common genes that regulate body size in mammals. *Nat Genet* **50**: 362–367. doi:10.1038/s41588-018-0056-5
- Browning SR, Browning BL. 2007. Rapid and accurate haplotype phasing and missing-data inference for whole-genome association studies by use of localized haplotype clustering. *Am J Hum Genet* **81**: 1084–1097. doi:10.1086/521987
- Burns MD, Bloom DD. 2020. Migratory lineages rapidly evolve larger body sizes than non-migratory relatives in ray-finned fishes. *Proc Biol Sci* **287**: 20192615. doi:10.1098/rspb.2019.2615
- Chen S, Zhou Y, Chen Y, Gu J. 2018. fastp: an ultra-fast all-in-one FASTQ preprocessor. *Bioinformatics* **34**: i884–i890. doi:10.1093/bioinformatics/bty560
- Clark JD, Stephens EA, Coryndon SC. 1966. Pleistocene fossiliferous lake beds of the Malawi (Nyasa) rift: a preliminary report. *Am Anthropol* **68**: 46–87. doi:10.1525/aa.1966.68.2.02a00960
- Clarke JT. 2021. Evidence for general size-by-habitat rules in actinopterygian fishes across nine scales of observation. *Ecol Lett* **24**: 1569–1581. doi:10.1111/ele.13768
- Colombo M, Damerou M, Hanel R, Salzburger W, Matschiner M. 2015. Diversity and disparity through time in the adaptive radiation of Antarctic notothenioid fishes. *J Evol Biol* **28**: 376–394. doi:10.1111/jeb.12570
- Cruickshank TE, Hahn MW. 2014. Reanalysis suggests that genomic islands of speciation are due to reduced diversity, not reduced gene flow. *Mol Ecol* **23**: 3133–3157. doi:10.1111/mec.12796
- Danecek P, Auton A, Abecasis G, Albers CA, Banks E, DePristo MA, Handsaker RE, Lunter G, et al. 2011. The variant call format and VCFtools. *Bioinformatics* **27**: 2156–2158. doi:10.1093/bioinformatics/btr330
- De-Kayne R, Selz OM, Marques DA, Frei D, Seehausen O, Feulner PGD. 2022. Genomic architecture of adaptive radiation and hybridization in Alpine whitefish. *Nat Commun* **13**: 4479. doi:10.1038/s41467-022-32181-8
- DeLorenzo L, Mathews D, Brandon AA, Joglekar M, Carmona Baez A, Moore EC, Cicotto PJ, Roberts NB, Roberts RB, Powder KE. 2023. Genetic basis of ecologically relevant body shape variation among four genera of cichlid fishes. *Mol Ecol* **32**: 3975–3988. doi:10.1111/mec.16977
- Dobin A, Davis CA, Schlesinger F, Drenkow J, Zaleski C, Jha S, Batut P, Chaisson M, Gingeras TR. 2013. STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* **29**: 15–21. doi:10.1093/bioinformatics/bts635
- Doenz CJ, Krähenbühl AK, Walker J, Seehausen O, Brodersen J. 2019. Ecological opportunity shapes a large Arctic charr species radiation. *Proc Biol Sci* **286**: 20191992. doi:10.1098/rspb.2019.1992
- Du K, Ricci JMB, Lu Y, Garcia-Olazabal M, Walter RB, Warren WC, Dodge TO, Schumer M, Park H, Meyer A, et al. 2024. Phylogenomic analyses of all species of swordtail fishes (genus *Xiphophorus*) show that hybridization preceded speciation. *Nat Commun* **15**: 6609. doi:10.1038/s41467-024-50852-6
- Dunz AR, Schliwien UK. 2010. Description of a *Tilapia* (*Coptodon*) species flock of Lake Ejagham (Cameroun), including a redescription of *Tilapia deckerti* Thys van den Audenaerde, 1967. *Spixiana* **33**: 251–280.
- Edgar RC. 2004. MUSCLE: a multiple sequence alignment method with reduced time and space complexity. *BMC Bioinformatics* **5**: 113. doi:10.1186/1471-2105-5-113
- Gao F, Ming C, Hu WJ, Li HP. 2016. New software for the fast estimation of population recombination rates (FastEPRR) in the genomic era. *G3 (Bethesda)* **6**: 1563–1571. doi:10.1534/g3.116.028233
- Gautier M, Vitalis R. 2012. *Rehh*: an R package to detect footprints of selection in genome-wide SNP data from haplotype structure. *Bioinformatics* **28**: 1176–1177. doi:10.1093/bioinformatics/bts115
- Genner MJ, Nichols P, Carvalho G, Robinson RL, Shaw PW, Smith A, Turner GF. 2007. Evolution of a cichlid fish in a Lake Malawi satellite lake. *Proc Biol Sci* **274**: 2249–2257. doi:10.1098/rspb.2007.0619
- Gompert Z, Forister ML, Fordyce JA, Nice CC. 2008. Widespread mito-nuclear discordance with evidence for introgressive hybridization and selective sweeps in *Lycaeides*. *Mol Ecol* **17**: 5231–5244. doi:10.1111/j.1365-294X.2008.03988.x
- Gonzalez-Voyer A, Winberg S, Kolm M. 2009. Distinct evolutionary patterns of brain and body size during adaptive radiation. *Evolution* **63**: 2266–2274. doi:10.1111/j.1558-5646.2009.00705.x
- Gudbrandsson J, Kapralova KH, Franzdóttir SR, Bergsveinsclóttir PM, Hafstad V, Jónsson ZO, Snorrason SS, Pálsson A. 2019. Extensive genetic differentiation between recently evolved sympatric Arctic charr morphs. *Ecol Evol* **9**: 10964–10983. doi:10.1002/ece3.5516
- Hall C, Flores MV, Murison G, Crosier K, Crosier P. 2006. An essential role for zebrafish *Fgf11* during gill cartilage development. *Mech Dev* **123**: 925–940. doi:10.1016/j.mod.2006.08.006
- Iwasaki W, Fukunaga T, Isagozawa R, Yamada K, Maeda Y, Satoh TP, Sado T, Mabuchi K, Takeshima H, Miya M, et al. 2013. MitoFish and MitoAnnotator: a mitochondrial genome database of fish with an accurate and automatic annotation pipeline. *Mol Biol Evol* **30**: 2531–2540. doi:10.1093/molbev/mst141
- Kappeler L, Filho CDM, Dupont J, Leneuve P, Cervera P, Périn L, Loudes C, Blaise A, Klein R, Epelbaum J, et al. 2008. Brain IGF-1 receptors control mammalian growth and lifespan through a neuroendocrine mechanism. *PLoS Biol* **6**: e254. doi:10.1371/journal.pbio.0060254
- Karlsson EK, Baranowska I, Wade CM, Salmon Hillbertz NH, Zody MC, Anderson N, Biagi TM, Patterson N, Pielberg GR, Kulbokas EJ, et al. 2007. Efficient mapping of mendelian traits in dogs through genome-wide association. *Nat Genet* **39**: 1321–1328. doi:10.1038/ng.2007.10
- Kautt AF, Machado-Schiaffino G, Meyer A. 2016. Multispecies outcomes of sympatric speciation after admixture with the source population in two radiations of Nicaraguan crater lake cichlids. *PLoS Genet* **12**: e1006157. doi:10.1371/journal.pgen.1006157
- Keppeler FW, Montaña CG, Winemiller KO. 2020. The relationship between trophic level and body size in fishes depends on functional traits. *Ecol Monogr* **90**: e01415. doi:10.1002/ecm.1415
- Konings A. 2016. *Malawi cichlids in their natural habitat*, 5th ed. Cichlid Press, El Paso, TX.
- Korunes KL, Samuk K. 2021. paxy: unbiased estimation of nucleotide diversity and divergence in the presence of missing data. *Mol Ecol Resour* **21**: 1359–1368. doi:10.1111/1755-0998.13326
- Kumar NM, Cooper TL, Kocher TD, Strelman JT, McGrath PT. 2025. Large inversions in Lake Malawi cichlids are associated with habitat preference, lineage, and sex determination. *eLife* **14**: RP104923. doi:10.7554/eLife.104923
- Laine VN, Shikano T, Herczeg G, Vilkkilä J, Merilä J. 2013. Quantitative trait loci for growth and body size in the nine-spined stickleback *Pungitius pungitius* L. *Mol Ecol* **22**: 5861–5876. doi:10.1111/mec.12526
- Li H, Durbin R. 2009. Fast and accurate short read alignment with Burrows–Wheeler transform. *Bioinformatics* **25**: 1754–1760. doi:10.1093/bioinformatics/btp324
- Li H, Durbin R. 2011. Inference of human population history from individual whole-genome sequences. *Nature* **475**: 493–496. doi:10.1038/nature10231
- Liu H, Fu B, Pang M, Feng X, Wang X, Yu X, Tong J. 2016. QTL fine mapping and identification of candidate genes for growth-related traits in big-head carp (*Hypophthalmichthys nobilis*). *Aquaculture* **465**: 134–143. doi:10.1016/j.aquaculture.2016.08.039
- Liu X, Yuan M, Lau BWM, Li Y. 2022. SHANK family on stem cell fate and development. *Cell Death Dis* **13**: 880. doi:10.1038/s41419-022-05325-3
- Lu G, Bernatchez L. 1999. Correlated trophic specialization and genetic divergence in sympatric lake whitefish ecotypes (*Coregonus clupeaformis*): support for the ecological speciation hypothesis. *Evolution* **53**: 1491–1505. doi:10.1111/j.1558-5646.1999.tb05413.x
- Lück S, Scholz U, Douchkov D. 2024. Introducing GWAS<sup>t</sup>: a user-friendly, cross-platform solution for genome-wide association studies and genomic prediction. *Bioinform Adv* **4**: vbae177. doi:10.1093/bioadv/vbae177
- Luo C, Xu X, Zhao C, Wang Q, Wang R, Lang D, Zhang J, Hu W, Mu Y. 2025. Insight into body size evolution in Aves: based on some body size-related genes. *Integr Zool* **20**: 1124–1135. doi:10.1111/1749-4877.12927
- Lyons RP, Scholz CA, Cohen AS, King JW, Brown ET, Ivory SJ, Johnson TC, Deino AL, Reinthal PN, McGlue MM, et al. 2015. Continuous 1.3-million-year record of east African hydroclimate, and implications for

- patterns of evolution and biodiversity. *Proc Natl Acad Sci* **112**: 15568–15573. doi:10.1073/pnas.1512864112
- Malinsky M, Challis RJ, Tyers AM, Schifffels S, Terai Y, Ngatunga BP, Miska EA, Durbin R, Genner MJ, Turner GF. 2015. Genomic islands of speciation separate cichlid ecomorphs in an east African crater lake. *Science* **350**: 1493–1498. doi:10.1126/science.aac9927
- 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. *Nat Ecol Evol* **2**: 1940–1955. doi:10.1038/s41559-018-0717-x
- Malinsky M, Matschiner M, Svardal H. 2021. Dsuite: fast *D*-statistics and related admixture evidence from VCF files. *Mol Ecol Resour* **21**: 584–595. doi:10.1111/1755-0998.13265
- Malmquist HJ, Snorrason SS, Skúlason S, Jonsson B, Sandlund OT, Jónasson PM. 1992. Diet differentiation in polymorphic Arctic charr in Thingvallavatn, Iceland. *J Animal Ecol* **61**: 21–35. doi:10.2307/5505
- Marconi A, Vernaz G, Karunaratna A, Ngochera MJ, Durbin R, Santos ME. 2024. Genetic and developmental divergence in the neural crest program between cichlid fish species. *Mol Biol Evol* **41**: msae217. doi:10.1093/molbev/msae217
- McKenna A, Hanna M, Banks E, Sivachenko A, Cibulskis K, Kernytzky A, Garimella K, Altshuler D, Gabriel S, Daly M, et al. 2010. The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res* **20**: 1297–1303. doi:10.1101/gr.107524.110
- Meier JJ, McGee MD, Marques DA, Mwaiko S, Kiske M, Wandera S, Neumann D, Mrosso H, Chapman LJ, Chapman CA, et al. 2023. Cycles of fusion and fission enabled rapid parallel adaptive radiations in African cichlids. *Science* **381**: eade2833. doi:10.1126/science.ade2833
- 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. *Mol Biol Evol* **37**: 1530–1534. doi:10.1093/molbev/msaa015
- Noor MA, Bennett SM. 2009. Islands of speciation or mirages in the desert? examining the role of restricted recombination in maintaining species. *Heredity (Edinb)* **103**: 439–444. doi:10.1038/hdy.2009.151
- Öhlund G, Bodin M, Nilsson KA, Öhlund SO, Mobley KB, Hudson AG, Peedu M, Brännström Å, Bartels P, Præbel K, et al. 2020. Ecological speciation in European whitefish is driven by a large-gaped predator. *Evol Lett* **4**: 243–256. doi:10.1002/evl3.167
- Oliver MK. 2024. African cichlid fishes: morphological data and taxonomic insights from a genus-level survey of supraneurals, pterygiophores, and vertebral counts (Ovalentaria, Blenniiformes, Cichlidae, Pseudocrenilabrinae). *Biodivers Data J* **12**: e130707. doi:10.3897/BDJ.12.e130707
- Østbye K, Amundsen PA, Bernatchez L, Klemetsen A, Knudsen R, Kristoffersen R, Naesje TF, Hindar K. 2006. Parallel evolution of ecomorphological traits in the European whitefish *Coregonus lavaretus* (L.) species complex during postglacial times. *Mol Ecol* **15**: 3983–4001. doi:10.1111/j.1365-294X.2006.03062.x
- Pan H, Yu H, Ravi V, Li C, Lee AP, Lian MM, Tay BH, Brenner S, Wang J, Yang H, et al. 2016. The genome of the largest bony fish, ocean sunfish (*Mola mola*), provides insights into its fast growth rate. *GigaScience* **5**: s13742–s13016. doi:10.1186/s13742-016-0144-3
- Paradis E, Claude J, Strimmer K. 2004. APE: analyses of Phylogenetics and evolution in R language. *Bioinformatics* **20**: 289–290. doi:10.1093/bioinformatics/btg412
- Peters RH. 1983. *The ecological implications of body size*. Cambridge University Press, Cambridge, UK.
- Pike JG. 1958. A brief note on the upper Pleistocene raised beach of the North Karonga shoreline. *Nyasaland J* **11**: 57–59.
- Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MA, Bender D, Maller J, Sklar P, de Bakker PI, Daly MJ, et al. 2007. PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am J Hum Genet* **81**: 559–575. doi:10.1086/519795
- Rivas-González I, Rousselle M, Li F, Zhou L, Dutheil JY, Munch K, Shao Y, Wu D, Schierup MH, Zhang G. 2023. Pervasive incomplete lineage sorting illuminates speciation and selection in primates. *Science* **380**: eabn4409. doi:10.1126/science.abn4409
- Romanuk TN, Hayward A, Hutchings JA. 2011. Trophic level scales positively with body size in fishes. *Glob Ecol Biogeogr* **20**: 231–240. doi:10.1111/j.1466-8238.2010.00579.x
- Salisbury SJ, Ruzzante DE. 2022. Genetic causes and consequences of sympatric morph divergence in Salmonidae: a search for mechanisms. *Annu Rev Anim Biosci* **10**: 81–106. doi:10.1146/annurev-animal-051021-080709
- Santos ME, Lopes JF, Kratochwil CF. 2023. East African cichlid fishes. *Evodevo* **14**: 1. doi:10.1186/s13227-022-00205-5
- Sauzeau V, Horta-Junior JA, Rioloobos AS, Fernández G, Sevilla MA, López DE, Montero MJ, Rico B, Bustelo XR. 2010. Vav3 is involved in GABAergic axon guidance events important for the proper function of brainstem neurons controlling cardiovascular, respiratory, and renal parameters. *Mol Biol Cell* **21**: 4251–4263. doi:10.1091/mbc.e10-07-0639
- Schifffels S, Durbin R. 2014. Inferring human population size and separation history from multiple genome sequences. *Nat Genet* **46**: 919–925. doi:10.1038/ng.3015
- Schlueter PJ, Sang X, Duan C, Wood AW. 2007. Insulin-like growth factor receptor 1b is required for zebrafish primordial germ cell migration and survival. *Dev Biol* **305**: 377–387. doi:10.1016/j.ydbio.2007.02.015
- Sherman BT, Hao M, Qiu J, Jiao XL, Baseler MW, Lane HC, Imamichi T, Chang WZ. 2022. DAVID: a web server for functional enrichment analysis and functional annotation of gene lists (2021 update). *Nucleic Acids Res* **50**: W216–W221. doi:10.1093/nar/gkac194
- Smith C, Zięba G, Spence R, Klepaker T, Przybylski M. 2020. Three-spined stickleback armour predicted by body size, minimum winter temperature and pH. *J Zool* **311**: 13–22. doi:10.1111/jzo.12766
- Snoeks J. 2004. *The cichlid diversity of Lake Malawi/Nyasa/Niassa: identification, distribution and taxonomy*. Cichlid Press, El Paso, TX.
- Sommer-Trembo C, Santos ME, Clark B, Werner M, Fages A, Matschiner M, Hornung S, Ronco F, Oliver C, Garcia C, et al. 2024. The genetics of niche-specific behavioral tendencies in an adaptive radiation of cichlid fishes. *Science* **384**: 470–475. doi:10.1126/science.adj9228
- Stauffer JR Jr, McCrary JK, Black KE. 2008. Three new species of cichlid fishes (Teleostei: Cichlidae) from lake Apoyo, Nicaragua. *Proc Biol Soc Wash* **121**: 117–129. doi:10.2988/06-37.1
- Steele SE, López-Fernández H. 2014. Body size diversity and frequency distributions of Neotropical cichlid fishes (Cichliformes: Cichlidae: Cichlinae). *PLoS One* **9**: e106336. doi:10.1371/journal.pone.0106336
- Steenwyk JL, Li YN, Zhou XF, Shen XX, Rokas A. 2023. Incongruence in the phylogenomics era. *Nat Rev Genet* **24**: 834–850. doi:10.1038/s41576-023-00620-x
- Supek F, Bošnjak M, Škunca N, Šmuc T. 2011. REVIGO summarizes and visualizes long lists of gene ontology terms. *PLoS One* **6**: e21800. doi:10.1371/journal.pone.0021800
- Sutter NB, Bustamante CD, Chase K, Gray MM, Zhao K, Zhu L, Padhukasahasram B, Karlins E, Davis S, Jones PG, et al. 2007. A single *IGF1* allele is a major determinant of small size in dogs. *Science* **316**: 112–115. doi:10.1126/science.1137045
- Takahashi T, Koblmüller S. 2011. The adaptive radiation of cichlid fish in lake Tanganyika: a morphological perspective. *Int J Evol Biol* **2011**: 620754. doi:10.4061/2011/620754
- Takahashi T, Nagano AJ, Sota T. 2021. Mapping of quantitative trait loci underlying a magic trait in ongoing ecological speciation. *BMC Genomics* **22**: 615. doi:10.1186/s12864-021-07908-4
- Tran TTH, Nguyen HT, Le BTN, Tran PH, Van Nguyen S, Kim OTP. 2021. Characterization of single nucleotide polymorphism in *IGF1* and *IGF1R* genes associated with growth traits in striped catfish (*Pangasianodon hypophthalmus* sauvaige, 1878). *Aquaculture* **538**: 736542. doi:10.1016/j.aquaculture.2021.736542
- Trewavas E, Green J, Corbet SA. 1972. Ecological studies on crater lakes in west Cameroon fishes of Barombi Mbo. *J Zool* **167**: 41–95. doi:10.1111/j.1469-7998.1972.tb01722.x
- Troyer EM, White WT, Betancur-R R, Arcila D. 2025. Parallel shifts in differential gene expression reveal convergent miniaturization in fishes. *Proc Natl Acad Sci* **122**: e2512299122. doi:10.1073/pnas.2512299122
- Turner GF. 1996. *Offshore cichlids of Lake Malawi*. Cichlid Press, Lauenau, Germany.
- Turner GF, Genner MJ. 2025. Identification of the cichlid fishes of Lake Malawi/Nyasa part 3: Rhamphochromina and others. *EcoEvoRxiv* doi:10.32942/X2807R
- Turner GF, Ngatunga BP, Genner MJ. 2019. The natural history of the satellite lakes of Lake Malawi. *EcoEvoRxiv* doi:10.32942/osf.io/sehhdq
- Turner GF, Crampton DA, Genner MJ. 2023. A new species of *Lethrinops* (Cichliformes: Cichlidae) from a Lake Malawi satellite lake, believed to be extinct in the wild. *Zootaxa* **5318**: 515–530. doi:10.11646/zootaxa.5318.4.5
- Voight BF, Kudaravalli S, Wen XQ, Pritchard JK. 2006. A map of recent positive selection in the human genome. *PLoS Biol* **4**: 446–458. doi:10.1371/journal.pbio.0040072
- Wang L, Xia JH, Liu XJ, Liu P, Wan ZY, Yue GH. 2014. Molecular characterization and mapping of *Fgf21* gene in a foodfish species Asian seabass. *PLoS One* **9**: e90172. doi:10.1371/journal.pone.0090172
- Wang L, Sun F, Lee M, Yue GH. 2022. Whole-genome resequencing infers genomic basis of giant phenotype in Siamese fighting fish (*Betta splendens*). *Zool Res* **43**: 78–80. doi:10.24272/j.issn.2095-8137.2021.266
- Williamson D, Jackson MJ, Banerjee SK, Marvin J, Merdaci O, Thouveny N, Decobert M, Gibert-Massault E, Massault M, Mazaudier D, et al. 1999. Magnetic signatures of hydrological change in a tropical maar-lake (Lake Massoko, Tanzania): preliminary results. *Phys Chem Earth Part A* **24**: 799–803. doi:10.1016/S1464-1895(99)00117-9

- Winkelmann K, Genner MJ, Takahashi T, Rüber L. 2014. Competition-driven speciation in cichlid fish. *Nat Commun* **5**: 3. doi:10.1038/ncomms4412
- Woodward G, Ebenman B, Emmerson M, Montoya JM, Olesen JM, Valido A, Warren PH. 2005. Body size in ecological networks. *Trends Ecol Evol* **20**: 402–409. doi:10.1016/j.tree.2005.04.005
- Yáñez JM, Barria A, López ME, Moen T, Garcia BF, Yoshida GM, Xu P. 2023. Genome-wide association and genomic selection in aquaculture. *Rev Aquacult* **15**: 645–675. doi:10.1111/raq.12750
- Yang W, Feiner N, Pinho C, While GM, Kaliontzopoulou A, Harris DJ, Salvi D, Uller T. 2021. Extensive introgression and mosaic genomes of Mediterranean endemic lizards. *Nat Commun* **12**: 2762. doi:10.1038/s41467-021-22949-9
- Yoshida GM, Yáñez JM. 2021. Multi-trait GWAS using imputed high-density genotypes from whole-genome sequencing identifies genes associated with body traits in Nile tilapia. *BMC Genomics* **22**: 57. doi:10.1186/s12864-020-07341-z
- Zhang C, Dong SS, Xu JY, He WM, Yang TL. 2019. PopLDdecay: a fast and effective tool for linkage disequilibrium decay analysis based on variant call format files. *Bioinformatics* **35**: 1786–1788. doi:10.1093/bioinformatics/bty875
- Zhou Z, Li M, Cheng H, Fan W, Yuan Z, Gao Q, Xu Y, Guo Z, Zhang Y, Hu J, et al. 2018. An intercross population study reveals genes associated with body size and plumage color in ducks. *Nat Commun* **9**: 3974. doi:10.1038/s41467-018-06521-6

Received July 14, 2025; accepted in revised form March 29, 2026.