Article

Drivers and dynamics of a massive adaptive radiation in cichlid fishes

https://doi.org/10.1038/s41586-020-2930-4

Received: 4 February 2020

Accepted: 20 August 2020

Published online: 18 November 2020

Check for updates

Fabrizia Ronco[1]✉, Michael Matschiner[1,2,3], Astrid Böhne[1,4], Anna Boila[1], Heinz H. Büscher[1], Athimed El Taher[1], Adrian Indermaur[1], Milan Malinsky[1], Virginie Ricci[1], Ansgar Kahmen[5], Sissel Jentoft[3] & Walter Salzburger[1,3]✉

Adaptive radiation is the likely source of much of the ecological and morphological diversity of life[1–4]. How adaptive radiations proceed and what determines their extent remains unclear in most cases[1,4]. Here we report the in-depth examination of the spectacular adaptive radiation of cichlid fishes in Lake Tanganyika. On the basis of whole-genome phylogenetic analyses, multivariate morphological measurements of three ecologically relevant trait complexes (body shape, upper oral jaw morphology and lower pharyngeal jaw shape), scoring of pigmentation patterns and approximations of the ecology of nearly all of the approximately 240 cichlid species endemic to Lake Tanganyika, we show that the radiation occurred within the confines of the lake and that morphological diversification proceeded in consecutive trait-specific pulses of rapid morphospace expansion. We provide empirical support for two theoretical predictions of how adaptive radiations proceed, the 'early-burst' scenario[1,5] (for body shape) and the stages model[1,6,7] (for all traits investigated). Through the analysis of two genomes per species and by taking advantage of the uneven distribution of species in subclades of the radiation, we further show that species richness scales positively with per-individual heterozygosity, but is not correlated with transposable element content, number of gene duplications or genome-wide levels of selection in coding sequences.

At the macroevolutionary level, the diversity of life has been shaped mainly by two antagonistic processes: evolutionary radiations increase, and extinction events decrease, organismal diversity over time[5,8,9]. Evolutionary radiations are referred to as adaptive radiations if new lifeforms evolve rapidly through adaptive diversification into a variety of ecological niches, which typically presupposes ecological opportunity[1–3,10]. Whether or not an adaptive radiation occurs depends on a variety of extrinsic and intrinsic factors as well as on contingency, whereas the magnitude of an adaptive radiation is determined by the interplay between its main components, speciation (minus extinction) and adaptation to distinct ecological niches[1,2,4,11]. Despite considerable scientific interest in the phenomenon of adaptive radiation as the cradle of organismal diversity[1,2,10,12,13], many predictions regarding its drivers and dynamics remain untested, particularly in exceptionally species-rich instances. Here, we examine what some consider as the "most outstanding example of adaptive radiation"[14], the species flock of cichlid fishes in Lake Tanganyika. This cichlid assemblage comprises about 240 species[15], which together feature an extraordinary degree of morphological, ecological and behavioural diversity[14–17]. We construct a species tree of Lake Tanganyika's cichlid fauna on the basis of genome-wide data, demonstrate the adaptive nature of the radiation, reconstruct eco-morphological diversification along the species tree, and test general and cichlid-specific predictions related to adaptive radiation.

In situ radiation in Lake Tanganyika

To establish the phylogenetic context of cichlid evolution in Lake Tanganyika, we estimated the age of the radiation through divergence time analyses based on cichlid and other teleost fossils[18], and constructed time-calibrated species trees using 547 newly sequenced cichlid genomes (Supplementary Table 1). Our new phylogenetic hypotheses (Fig. 1, Extended Data Figs. 1–4, Supplementary Figs. 1, 2) support the assignment of the Tanganyikan cichlid fauna into 16 subclades—corresponding to the taxonomic grouping of species into tribes[15]—and confirm that the Tanganyikan representatives of the tribes Coptodonini, Oreochromini and Tylochromini belong to more ancestral and widespread lineages that have colonized the lake secondarily[12,15,19] (Supplementary Discussion). It has been under debate whether all endemic Tanganyikan cichlid tribes evolved within the confines of Lake Tanganyika or whether some of them evolved elsewhere before the formation of the lake[20–22]. Our time calibrations establish that the most recent common ancestor of the cichlid radiation in Lake Tanganyika lived around 9.7 million years ago (Ma) (95% highest-posterior-density

[1]Zoological Institute, Department of Environmental Sciences, University of Basel, Basel, Switzerland. [2]Palaeontological Institute and Museum, University of Zurich, Zurich, Switzerland. [3]Centre for Ecological and Evolutionary Synthesis (CEES), Department of Biosciences, University of Oslo, Oslo, Norway. [4]Centre for Molecular Biodiversity Research (ZMB), Zoological Research Museum Alexander Koenig, Bonn, Germany. [5]Botany, Department of Environmental Sciences, University of Basel, Basel, Switzerland. ✉e-mail: fabrizia.ronco@unibas.ch; walter.salzburger@unibas.ch

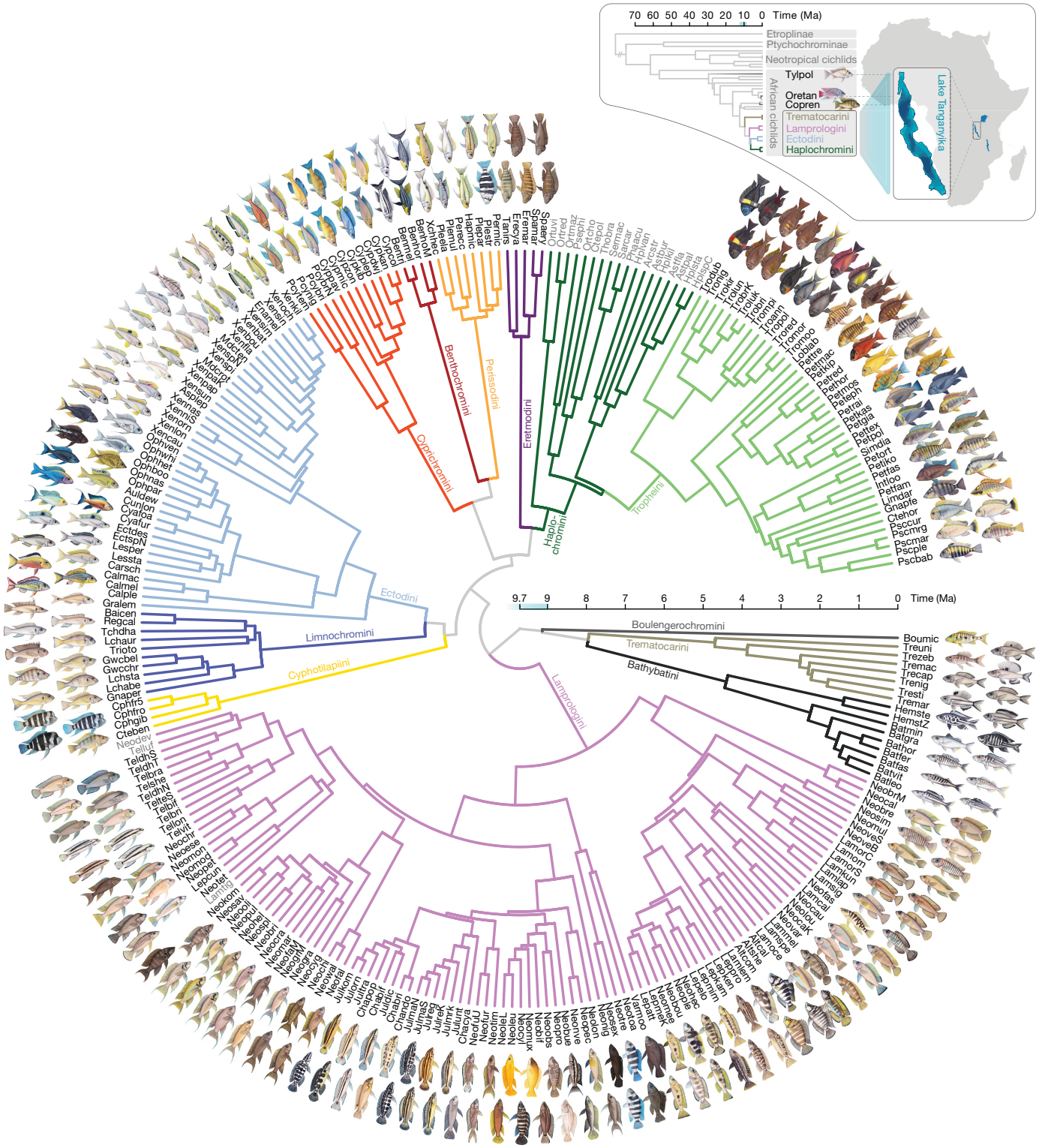


Fig. 1 | Time-calibrated species tree of the cichlid fishes of African Lake Tanganyika. The species tree was time calibrated with a relaxed-clock model and is based on a maximum-likelihood topology inferred from genome-wide SNPs. Species names are abbreviated using a six-letter code, whereby the first three letters represent the genus and the last three letters the species name (Supplementary Table 1; see Extended Data Fig. 2 for the phylogeny with full species names). Branches are coloured according to tribes, and for all lake species an illustration is shown. Representatives of riverine cichlids (grey font) are nested within the radiation. The inset shows the time-calibrated phylogeny of more ancestral cichlid lineages (estimated under the multi-species coalescent model, Extended Data Fig. 1), highlighting the phylogenetic positions of the Tanganyikan representatives of the tribes Coptodonini (*Coptodon rendalli* (Copren)), Oreochromini (*Oreochromis tanganicae* (Oretan)) and Tylochromini (*Tylochromis polylepis* (Tylpol)), which colonized the lake secondarily. The schematic map of the African continent shows the position of the three Great Lakes Victoria, Malawi and Tanganyika, with a magnified section of Lake Tanganyika. The presumed age of Lake Tanganyika[23] (9–12 Myr) is indicated in blue along the time axes. Species trees based on alternative topologies are presented in Extended Data Figs. 2–4, and uncalibrated nuclear and mitochondrial phylogenies on the specimen level are shown in Supplementary Figs. 1, 2.

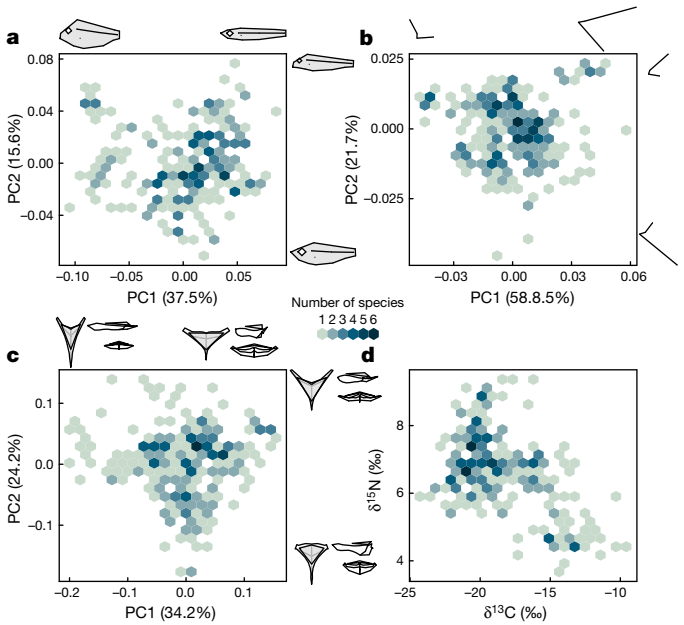


Fig. 2 | Morphospace and ecospace occupation of the cichlid fishes of Lake Tanganyika. a–**c**, PCA of body shape (**a**, *n* = 242 taxa; 2,197 specimens), upper oral jaw morphology (**b**, *n* = 242 taxa; 2,197 specimens) and lower pharyngeal jaw shape (**c**, *n* = 239 taxa, 1,168 specimens) along with the associated shape changes. **d**, Ecospace spanned by the stable C and N isotope composition ($\delta^{13}C$ and $\delta^{15}N$ values; *n* = 236 taxa; 2,259 specimens). The colour scale indicates the number of species in 20 by 20 bins across the trait space (see Extended Data Figs. 6, 7 for PCA and stable-isotope biplots with a focus on morpho- and ecospace occupation per tribe).

age interval: 10.1–9.1 Ma) (Fig. 1), which coincides with the appearance of lacustrine conditions in the Tanganyikan Rift[23]. This suggests that the radiation commenced shortly after the lake had formed and that all endemic cichlid tribes have evolved and diversified in situ, that is, within the temporal and geographical context of Lake Tanganyika.

Phenotypes correlate with environments

Because—in the case of adaptive radiation—diversification occurs via niche specialization, a strong association is expected in the extant fauna between the environment occupied by a species and the specific morphological features used to exploit it[2,3]. To quantify eco-morphological diversification across the radiation, we investigated three trait complexes through landmark-based morphometric analyses. Specifically, we quantified body shape and upper oral jaw morphology using 2D landmarks acquired from X-ray images and the shape of the lower pharyngeal jaw bone based on 3D landmarks derived from micro-computed tomography (μCT) scans (Extended Data Fig. 5). To approximate the ecological niche of each species, we used the carbon and nitrogen stable-isotope composition of muscle tissue, which provides information about the relative position along the benthic–pelagic axis ($\delta^{13}C$ value) and the relative trophic level ($\delta^{15}N$ value), respectively[16,24]—a pattern that we corroborate here for Lake Tanganyika (Extended Data Fig. 6a, Supplementary Discussion). The major axes of shape variation for each trait complex were identified through a principal component analysis (PCA). To test for phenotype–environment correlations and to identify the ecologically most relevant components of each of these trait complexes, we performed a two-block partial least-square analysis (PLS) with the stable-isotope measurements, and applied a phylogenetic generalized least-square analysis (pGLS) to account for phylogenetic dependence.

The quantification of variation in body shape revealed that principal component 1 (PC1) represented mainly differences in aspect ratio, whereas PC2 was loaded with changes in head morphology (Fig. 2a). The changes in aspect ratio (comparable to PC1) were correlated with the $\delta^{13}C$ and $\delta^{15}N$ values (PLS: Pearson's $r = 0.69$, $R^2 = 0.48$, $P = 0.001$; pGLS: $R^2 = 0.12$, $P < 0.001$, $\lambda_{pGLS} = 1.007$). PC1 of upper oral jaw morphology mainly represented changes in the orientation and relative size of the premaxilla, which was also the main correlate to the stable C and N isotope composition (PLS: Pearson's $r = 0.62$, $R^2 = 0.38$, $P = 0.001$; pGLS: $R^2 = 0.09$, $P < 0.001$, $\lambda_{pGLS} = 1.023$), whereas PC2 was defined by changes in the ratio of the rostral versus the lateral part of the bone (Fig. 2b). For lower pharyngeal jaw shape, we found that PC1 reflected mainly changes in the aspect ratio of the jaw bone in combination with an increased posterior thickness, whereas PC2 involved similar shifts in thickness, yet in this case in combination with changes in the lengths of the postero-lateral horns that act as muscle-attachment structures[25] (Fig. 2c). The PLS revealed that shape changes similar to PC2 are best associated with stable-isotope values (PLS: Pearson's $r = 0.67$, $R^2 = 0.45$, $P = 0.001$; pGLS: $R^2 = 0.16$, $P < 0.001$, $\lambda_{pGLS} = 1.018$). The PCAs further revealed that the occupied area of the morphospace and ecospace scales with the number of species in the tribes (Extended Data Figs. 6, 7; ecospace: Pearson's $r = 0.88$, d.f. = 9, $P < 0.001$; body shape: Pearson's $r = 0.91$, d.f. = 9, $P < 0.001$; upper oral jaw morphology: Pearson's $r = 0.88$, d.f. = 9, $P < 0.001$; lower pharyngeal jaw shape: Pearson's $r = 0.83$, d.f. = 9, $P = 0.002$), a pattern that is not driven by sample size only (Supplementary Discussion).

Overall, the significant association between each of the three traits and the stable C and N isotope composition underpins their adaptive value (Extended Data Fig. 8a–c). A joint consideration points out that deep-bodied cichlids with inferior mouths and thick lower pharyngeal jaws with short horns are associated with higher stable-isotope projections (high $\delta^{13}C$ and low $\delta^{15}N$ values), indicating that such fishes occur predominantly in the benthic/littoral zone of the lake and feed on plants and algae, whereas more elongated species with more superior mouths and longer and thinner lower pharyngeal jaws are generally associated with lower stable-isotope projections (low $\delta^{13}C$ and high $\delta^{15}N$ values), suggesting a more pelagic lifestyle and a higher position in the food chain.

Pulses of morphological diversification

Next, we investigated the temporal dynamics of how the observed eco-morphological disparity emerged over the course of the radiation. In addition to the three eco-morphological traits, we also scored male pigmentation patterns to approximate disparity along the signalling axis—another potentially important component of diversification in adaptive radiations[1,6,7,26]. For all four traits, we estimated morphospace expansion through time using ancestral-state reconstructions along the time-calibrated species tree and applying a variable-rates model of trait evolution[27,28] (Extended Data Fig. 8d, e). We calculated morphological disparity as the extent of occupied morphospace in time intervals of 0.15 million years (Myr) in comparison to a null model that assumes Brownian motion. Likewise, evolutionary rates through time were calculated as mean evolutionary rates derived from the variable-rates model, sampled at the same time points along the phylogeny.

Our analyses uncovered a pattern of discrete pulses in morphospace expansion, which were followed, in most cases, by morphospace packing (Fig. 3). The timing of these pulses differed among the traits. For body shape, we found a pulse of rapid morphospace expansion early in the radiation, alongside the first pulse of lower pharyngeal jaw shape diversification (Fig. 3b, c); this early phase of the radiation also features the highest evolutionary rates for body shape (Fig. 3d). The pulse in upper oral jaw diversification occurred in the middle phase of the radiation. Evolutionary rates were increased during this period, and were even higher at a later phase that was dominated by packing of the upper oral jaw morphospace rather than its expansion (Fig. 3b–d). This

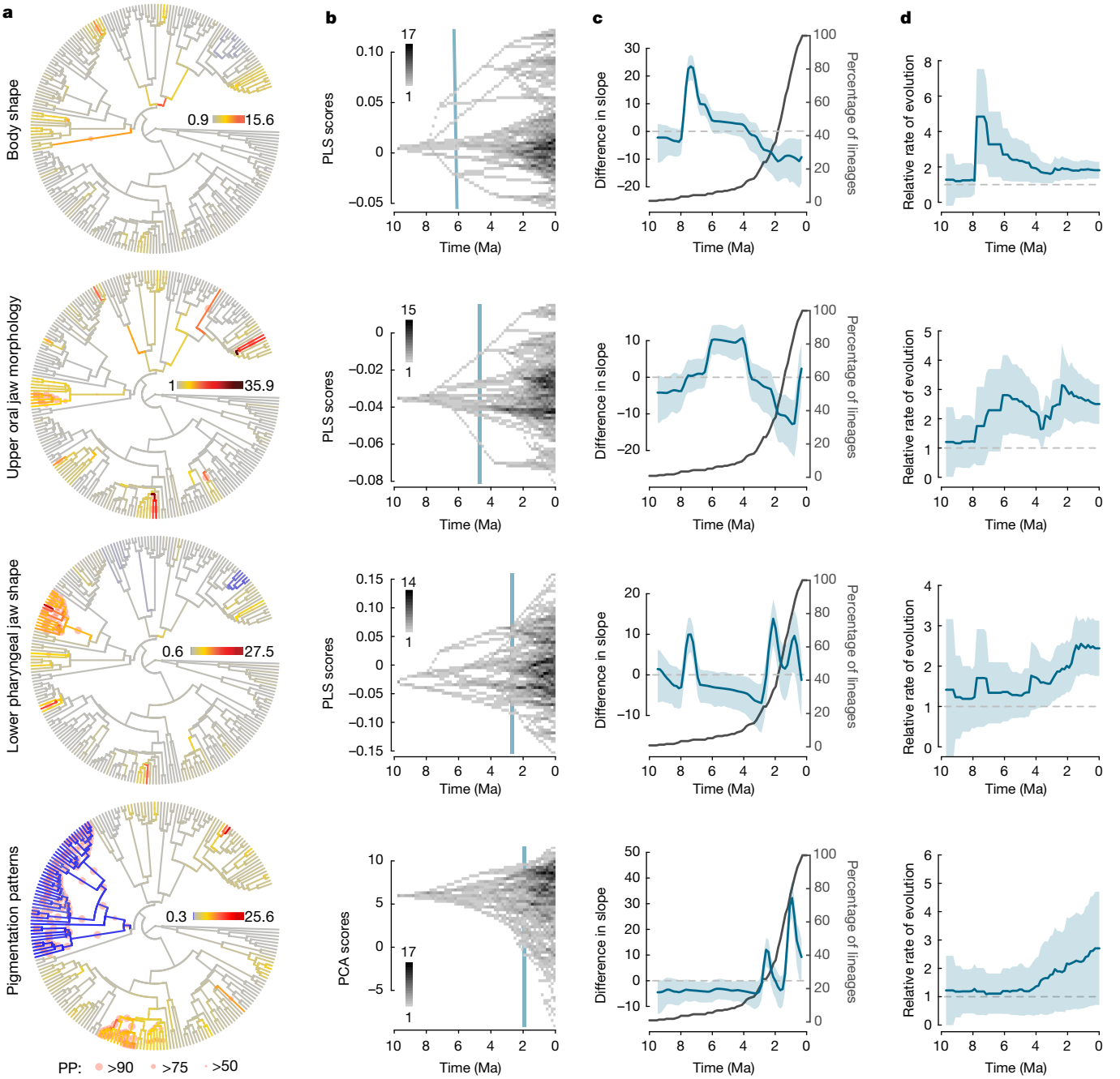


Fig. 3 | Temporal dynamics of morphological diversification in the adaptive radiation of cichlid fishes in Lake Tanganyika. a–**d**, First row: body shape, $n = 232$ taxa, 2,164 specimens; second row: upper oral jaw morphology, $n = 232$ taxa, 2,164 specimens; third row: lower pharyngeal jaw shape, $n = 232$ taxa, 1,148 specimens; fourth row: pigmentation patterns, $n = 218$ taxa, 1,016 specimens. **a**, Species tree (Fig. 1) with branches coloured according to the mean relative rates of trait evolution for each trait. PP, posterior probability for rate shift. **b**, Morphospace densities (number of lineages) through time for each trait. Blue lines indicate the point in time when 50% of the extant morphospace had become occupied. **c**, Comparison of slopes (blue) of morphospace expansion over time between the observed data and the Brownian motion null model of trait evolution (mean across 500 Brownian motion simulations with 95% quantiles). A difference in slopes above zero represents morphospace expansion and values below zero indicate morphospace packing relative to the null model. Lineage accumulation through time derived from the species tree is shown in dark grey. **d**, Mean relative rates of trait evolution over time with standard deviation (blue).

suggests that, in that later phase, rapidly evolving lineages diverged into pre-occupied regions of the morphospace, ultimately resulting in convergent forms[16]. The second pulse in lower pharyngeal jaw morphospace expansion happened late in the radiation when evolutionary rates were also highest for this trait (Fig. 3b–d). Thus, the theoretical prediction that eco-morphological diversification is rapid early in an adaptive radiation and slows down through time as the available niche space becomes filled[1,5] applies only to body shape. Yet, this early burst in body shape diversification was not connected to a substantial increase in lineage accumulation (Fig. 3c).

Pigmentation patterns showed a single pulse of diversification and increased evolutionary rates late in the radiation—a signature unlikely to be caused by a high turnover rate in this trait (Supplementary Discussion). This late pulse of diversification in pigmentation patterns, together with the consecutive pulses of morphospace expansion in the eco-morphological traits, is in agreement with the prediction that

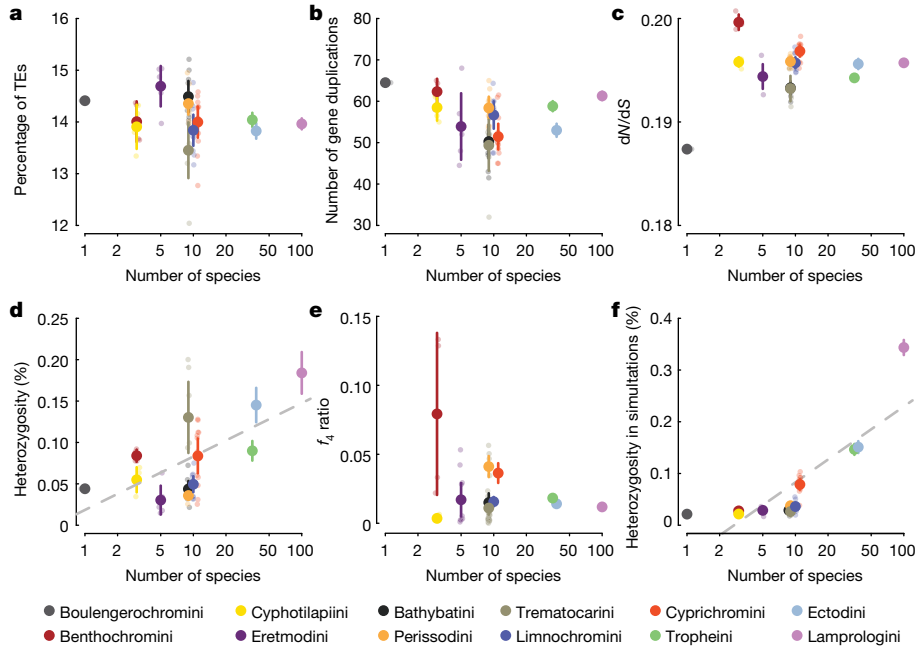


Fig. 4 | Association between genomic features and species richness across the cichlid tribes in Lake Tanganyika. Each genomic summary statistic was tested for a correlation with species richness per tribe (log transformed). To account for phylogenetic structure in the data, we calculated phylogenetic independent contrasts for each variable. Data points are coloured according to tribes; large points are tribe means shown with 95% confidence intervals, small points represent species means and are only shown for group sizes <40. **a**, Percentage of the genome identified as transposable elements (TEs) (Pearson's $r = -0.31$, d.f. = 10, $P = 0.33$; tribe means are based on one genome per species; Extended Data Fig. 9a). **b**, Number of duplicated genes (Pearson's $r = -0.27$, d.f. = 10, $P = 0.40$; tribe means are based on species means). **c**, Genome-wide d*N*/d*S* ratios as a measure of selection on coding sequences (Pearson's $r = 0.26$, d.f. = 10, $P = 0.42$; tribe means are based on species means across a set of 15,294 genes per genome; Extended Data Fig. 9b). **d**, Percentage of heterozygous sites per genome (Pearson's $r = 0.70$, d.f. = 10, $P = 0.012$; tribe means are based on species means). **e**, f_4-ratio statistics as a measure of gene flow among species within each tribe (Pearson's $r = -0.35$, d.f. = 9, $P = 0.29$; tribe means are based on all species triplets within each tribe; see Extended Data Fig. 10 for a summary of the f_4-ratio statistics for all species comparisons). **f**, Mean percentage of heterozygous sites in simulations with within-tribe migration rates sampled from the observed f_4-ratio statistics (Pearson's $r = 0.84$, d.f. = 10, $P = 0.00067$; tribe means are based on species means across 20 simulations; Extended Data Fig. 9c).

diversification in an adaptive radiation proceeds in discrete temporal stages—first in macrohabitat use, then by trophic specialization, followed by a final stage of divergence along the signalling axes[1,6,7]. However, in contrast to the conventional stages model, the most recent stage of the cichlid adaptive radiation in Lake Tanganyika, which coincides with a large number of speciation events (Fig. 3c), is characterized by temporally overlapping pulses of diversification in both a putative signalling trait and in an ecologically relevant trait. The lower pharyngeal jaw shape is the only trait complex showing two discrete pulses of morphospace expansion—one early in the radiation and one late when niche space already became limited. This later pulse suggests that diversification in the pharyngeal jaw apparatus facilitated fine-scaled resource partitioning after body shape and upper oral jaw morphospaces had been explored, resulting in the densely packed niche space observed today (Figs. 2, 3b).

Genomic features and species richness

Finally, we examined whether the diversity patterns arising over the course of the radiation are linked with particular genomic features. It has previously been suggested—on the basis of five reference cichlid genomes—that the radiating African cichlid lineages are characterized by increased transposable element counts, increased levels of gene duplications, and genome-wide accelerated coding-sequence evolution[13]. Because of the phylogenetic substructure of Lake Tanganyika's cichlid fauna and the widely differing species numbers among tribes, our data offered the opportunity to examine genomic features for an association with per-tribe species richness within a large-scale radiation. We did not find evidence that members of species-rich tribes exhibit greater numbers of transposable elements (Fig. 4a) or more duplicated genes in their genomes (Fig. 4b), nor do they feature elevated genome-wide signatures of selection in coding sequences (Fig. 4c) (see also Extended Data Fig. 9). However, we found that a tribe's species richness scales positively with a common measure of genetic diversity: genome-wide heterozygosity (Fig. 4d). That genetic diversity is linked to species richness has been previously suspected, although the nature of this relationship and the determinants of genetic diversity are under debate[29,30].

Elevated levels of heterozygosity could potentially result from hybridization[31], which has itself been suggested as a trigger of cichlid radiations[22,32,33]. In Tanganyikan cichlids, the level of gene flow within tribes (estimated using f_4-ratio values[34]) does not correlate with a tribe's species richness (Fig. 4e, Extended Data Fig. 10). Nevertheless, much of the variation in heterozygosity as well as its correlation with species richness can be explained by the observed levels of gene flow within tribes in combination with the reduced gene flow among them: through coalescent simulations of genome evolution along the species tree we show that variation in migration rates, sampled from the empirical f_4-ratio estimates, can produce levels of heterozygosity that are similar to the ones observed in nature (Fig. 4f). Hence, the correlation between species richness and heterozygosity can be explained by gene flow and phylogenetic structure, which is consistent with the expectation that the effect of gene flow scales positively with the number of hybridizing species and the divergence among these. In the cichlid radiation in Lake Malawi, which is an order of magnitude younger than the one in Lake Tanganyika, heterozygosity levels vary much less among lineages and

do not scale with species richness, which—according to our findings—can be explained by the much lower levels of genetic differentiation between the hybridizing species[33].

Conclusion

On the basis of a comprehensive dataset on cichlid fishes from African Lake Tanganyika, we tested predictions related to the phenomenon of adaptive radiation. We establish that the Tanganyikan cichlid radiation unfolded within the temporal and spatial confines of the lake, giving rise to an endemic fauna consisting of about 240 species in 52 genera and 13 tribes in less than 10 Myr. Although the ancestors of these tribes initially found comparable ecological opportunity, present-day species numbers differ by two orders of magnitude among these phylogenetic sublineages. Our analyses of morphological, ecological and genomic information revealed that, taken as a whole, species-rich tribes occupy larger fractions of the morphospace and ecospace and contain species that are, at the per-genome level, genetically more diverse, which appears to be linked to gene flow. We demonstrate a phenotype–environment association in three trait complexes (body shape, upper oral jaw morphology and lower pharyngeal jaw shape) and pinpoint their most relevant adaptive components. We show that eco-morphological diversification was not gradual over the course of the radiation. Instead, we identified trait-specific pulses of accelerated phenotypic evolution, whereby only diversification in body shape shows an early burst[1,5]. The sequence of the trait-specific pulses essentially follows the pattern postulated in the stages model of adaptive radiation[1,6,7], with the extension that the most recent stage of the cichlid adaptive radiation in Lake Tanganyika, which is characterized by a large number of speciation events, is defined by increased diversification in both an ecological (lower pharyngeal jaw) and a signalling (pigmentation) trait. To what extent the observed diversity and disparity patterns were shaped by past environmental fluctuations and extinction dynamics cannot be answered conclusively through the investigation of the extant fauna alone.

Online content

Any methods, additional references, Nature Research reporting summaries, source data, extended data, supplementary information, acknowledgements, peer review information; details of author contributions and competing interests; and statements of data and code availability are available at https://doi.org/10.1038/s41586-020-2930-4.

1. Gavrilets, S. & Losos, J. B. Adaptive radiation: contrasting theory with data. *Science* **323**, 732–737 (2009).
2. Schluter, D. *The Ecology of Adaptive Radiation* (Oxford Univ. Press, 2000).
3. Simpson, G. G. *The Major Features of Evolution* (Columbia Univ. Press, 1953).
4. Glor, R. E. Phylogenetic insights on adaptive radiation. *Annu. Rev. Ecol. Evol. Syst.* **41**, 251–270 (2010).
5. Foote, M. The evolution of morphological diversity. *Annu. Rev. Ecol. Syst.* **28**, 129–152 (1997).
6. Danley, P. D. & Kocher, T. D. Speciation in rapidly diverging systems: lessons from Lake Malawi. *Mol. Ecol.* **10**, 1075–1086 (2001).
7. Streelman, J. T. & Danley, P. D. The stages of vertebrate evolutionary radiation. *Trends Ecol. Evol.* **18**, 126–131 (2003).
8. Benton, M. J. Diversification and extinction in the history of life. *Science* **268**, 52–58 (1995).
9. Sepkoski, J. J., Jr. Rates of speciation in the fossil record. *Phil. Trans. R. Soc. Lond. B* **353**, 315–326 (1998).
10. Berner, D. & Salzburger, W. The genomics of organismal diversification illuminated by adaptive radiations. *Trends Genet.* **31**, 491–499 (2015).
11. Wagner, C. E., Harmon, L. J. & Seehausen, O. Ecological opportunity and sexual selection together predict adaptive radiation. *Nature* **487**, 366–369 (2012).
12. Salzburger, W. Understanding explosive diversification through cichlid fish genomics. *Nat. Rev. Genet.* **19**, 705–717 (2018).
13. Brawand, D. et al. The genomic substrate for adaptive radiation in African cichlid fish. *Nature* **513**, 375–381 (2014).
14. Fryer, G. & Iles, T. D. *The Cichlid Fishes of the Great Lakes of Africa* (T.F.H. Publications, 1972).
15. Ronco, F., Büscher, H. H., Indermaur, A. & Salzburger, W. The taxonomic diversity of the cichlid fish fauna of ancient Lake Tanganyika, East Africa. *J. Gt. Lakes Res.* **46**, 1067–1078 (2020).
16. Muschick, M., Indermaur, A. & Salzburger, W. Convergent evolution within an adaptive radiation of cichlid fishes. *Curr. Biol.* **22**, 2362–2368 (2012).
17. Salzburger, W., Van Bocxlaer, B. & Cohen, A. S. Ecology and evolution of the African Great Lakes and their faunas. *Annu. Rev. Ecol. Evol. Syst.* **45**, 519–545 (2014).
18. Matschiner, M., Böhne, A., Ronco, F. & Salzburger, W. The genomic timeline of cichlid diversification across continents. *Nat. Commun.* https://doi.org/10.1038/s41467-020-17827-9 (2020).
19. Koch, M. et al. Evolutionary history of the endemic Lake Tanganyika cichlid fish *Tylochromis polylepis*: A recent intruder to a mature adaptive radiation. *J. Zool. Syst. Evol. Res.* **45**, 64–71 (2007).
20. Salzburger, W., Meyer, A., Baric, S., Verheyen, E. & Sturmbauer, C. Phylogeny of the Lake Tanganyika cichlid species flock and its relationship to the Central and East African haplochromine cichlid fish faunas. *Syst. Biol.* **51**, 113–135 (2002).
21. Schedel, F. D. B., Musilova, Z. & Schliewen, U. K. 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 (2019).
22. Irisarri, I. et al. Phylogenomics uncovers early hybridization and adaptive loci shaping the radiation of Lake Tanganyika cichlid fishes. *Nat. Commun.* **9**, 3159 (2018).
23. Cohen, A. S., Soreghan, M. J. & Scholz, C. A. Estimating the age of formation of lakes: an example from Lake Tanganyika, East African Rift system. *Geology* **21**, 511–514 (1993).
24. Post, D. M. Using stable isotopes to estimate trophic position: models, methods, and assumptions. *Ecology* **83**, 703–718 (2002).
25. Liem, K. F. Evolutionary strategies and morphological innovations: cichlid pharyngeal jaws. *Syst. Zool.* **22**, 425–441 (1973).
26. Salzburger, W. The interaction of sexually and naturally selected traits in the adaptive radiations of cichlid fishes. *Mol. Ecol.* **18**, 169–185 (2009).
27. Venditti, C., Meade, A. & Pagel, M. Multiple routes to mammalian diversity. *Nature* **479**, 393–396 (2011).
28. Cooney, C. R. et al. Mega-evolutionary dynamics of the adaptive radiation of birds. *Nature* **542**, 344–347 (2017).
29. Ellegren, H. & Galtier, N. Determinants of genetic diversity. *Nat. Rev. Genet.* **17**, 422–433 (2016).
30. Schluter, D. & Pennell, M. W. Speciation gradients and the distribution of biodiversity. *Nature* **546**, 48–55 (2017).
31. Grant, P. R. & Grant, B. R. *40 Years of Evolution: Darwin's Finches on Daphne Major Island* (Princeton Univ. Press, 2014).
32. Meier, J. I. et al. Ancient hybridization fuels rapid cichlid fish adaptive radiations. *Nat. Commun.* **8**, 14363 (2017).
33. Malinsky, M. et al. Whole-genome sequences of Malawi cichlids reveal multiple radiations interconnected by gene flow. *Nat. Ecol. Evol.* **2**, 1940–1955 (2018).
34. Patterson, N. et al. Ancient admixture in human history. *Genetics* **192**, 1065–1093 (2012).

Methods

No statistical methods were used to predetermine sample size. The experiments were not randomized. The investigators were not blinded to allocation during experiments and outcome assessment.

Sampling

Sampling was conducted between 2014 and 2017 at 130 locations at Lake Tanganyika. To maximise taxon coverage, we included additional specimens from previous expeditions (4.9% of the samples) as well as from other collections (0.8%). The final dataset (301 taxa; $n = 2{,}723$ specimens) contained an almost complete taxon sampling of the cichlid fauna of Lake Tanganyika, as well as 18 representative cichlid species from nearby waterbodies, and 32 outgroup species. All analyses described below are based on the same set of typically 10 specimens per species, or subsets thereof (Supplementary Tables 1, 2, Supplementary Methods).

Whole-genome sequencing

Genomic DNA of typically one male and one female specimen per species ($n = 547$) was extracted from fin clips preserved in ethanol using the E.Z.N.A. Tissue DNA Kit (Omega Bio-Tek) and sheared on a Covaris E220 (60 µl with 10% duty factor, 175 W, 200 cycles for 65 s). Individual libraries were prepared using TruSeq DNA PCR-Free Sample Preparation kit (Illumina; low sample protocol) for 350-bp insert size, pooled (six per lane), and sequenced at 126-bp paired-end on an Illumina HiSeq 2500 (Supplementary Table 1 contains information on read depths).

Assessing genomic variation

After adaptor removal with Trimmomatic[35] (v.0.36), reads of 528 genomes (all species belonging to the cichlid radiation in Lake Tanganyika plus additional species nested within this radiation and some selected outgroup species; Supplementary Table 1) were mapped to the Nile tilapia reference genome (RefSeq accession GCF_001858045.1[36]) using BWA-MEM[37] (v.0.7.12). Variant calling was performed with HaplotypeCaller and GenotypeGVCF tools[38] (v.3.7) (GATK), applying a minimum base quality score of 30. Variant calls were filtered with BCFtools[39] (v.1.6; FS < 20, QD > 2, MQ > 20, DP > 4,000, DP < 8,000, ReadPosRankSum > −0.5, MQRankSum > −0.5). We applied a filter to sites in proximity to indels with a minor allele count greater than 2, depending on the size of the indel. With SNPable (http://lh3lh3.users.sourceforge.net/snpable.shtml), we determined all sites within regions of the Nile tilapia reference genome in which read mapping could be ambiguous and masked these sites. Using VCFtools[40] (v.0.1.14) we further masked, per individual, genotypes with a read depth below 4 or a genotype quality below 20. Sites that were no longer polymorphic after the filtering steps were excluded, resulting in a dataset of 57,751,375 SNPs. Called variants were phased with the software beagle[41] (v.4.1). The phasing of *Neolamprologus cancellatus*, which appeared to be F_1 hybrids, was further improved with a custom script. Further details are provided in the Supplementary Methods.

De novo genome assemblies

De novo genome assemblies were generated from the raw-read data for each individual following an approach described previously[42,43], using CeleraAssembler[44] (v.8.3) and FLASH[45] (v.1.2.11). Eight genomes repeatedly failed to assemble and were therefore excluded from further analyses (specimen vouchers: A188, IRF6, IZC5, JWE7, JWG1, JWG2, LJD3 and LJE8). Assembly quality was assessed with QUAST[46] (v.4.5) and completeness was determined with BUSCO[47] (v.3). Assembly statistics summarized with MultiQC[48] (v.1.7) are available on Dryad.

Determining the age of the radiation

To determine the age of the cichlid radiation in Lake Tanganyika, we applied phylogenomic molecular-clock analyses for representatives of all cichlid subfamilies and the most divergent tribes, together with non-cichlid outgroups (44 species; Extended Data Fig. 1). Following Matschiner et al.[18] we identified and filtered orthologue sequences from genome assemblies and compiled 'strict' and 'permissive' datasets that contained alignments for 510 and 1,161 genes and had total alignment lengths of 542,922 and 1,353,747 bp, respectively. We first analysed the topology of the species with the multi-species coalescent model implemented in ASTRAL[49] (v.5.6.3), based on gene trees that we estimated for both datasets with BEAST2[50] (v.2.5.0). As undetected past introgression can influence divergence-time estimates in molecular clock analyses, we further tested for signals of introgression in the form of asymmetric species relationship in gene trees and excluded five species (*Fundulus heteroclitus*, *Tilapia brevimanus*, *Pelmatolapia mariae*, *Tilapia sparrmanii*, and *Steatocranus* sp. 'ultraslender') potentially affected by introgression from all subsequent molecular-clock analyses. We then estimated divergence times among the most divergent cichlid tribes and the age of the cichlid radiation in Lake Tanganyika with the multi-species coalescent model in StarBEAST2[51] (v.0.15.5), using the 'strict' set of gene alignments (Extended Data Fig. 1). Further details are provided in the Supplementary Methods.

Phylogenetic inference

To infer a complete phylogeny of the cichlid radiation in Lake Tanganyika (the Tanganyikan representatives of the more ancestral tribes Coptodonini, Oreochromini and Tylochromini were excluded) from genome-wide SNPs we applied additional filters, retaining only SNPs with <40% missing data and between-SNP distances of at least 100 bp. The remaining 3,630,997 SNPs were used to infer a maximum-likelihood phylogeny with RAxML[52] (v.8.2.4; Fig. 1, Extended Data Fig. 2, Supplementary Fig. 1). The species-tree topology was further estimated under the multi-species coalescent model from a set of local phylogenies with ASTRAL (Extended Data Fig. 3); these local phylogenies were inferred with IQ-TREE[53] (v.1.7-beta7) from alignments for 1,272 genomic regions determined to be particularly suitable for phylogenetic analysis (see Supplementary Methods). We also applied the multi-species coalescent model implemented in SNAPP[54] (v.1.4.2) to the dataset of genome-wide SNPs (Extended Data Fig. 4). Species-level phylogenies resulting from these different approaches were used as topological constraints in subsequent relaxed-clock analyses of divergence times (see below). In addition, we estimated the mitochondrial phylogeny based on maximum-likelihood with RAxML (Supplementary Fig. 2). Further details are provided in the Supplementary Methods.

Divergence time estimates within the radiation

For relaxed-clock analyses, the 1,272 alignments were further filtered by applying stricter thresholds on the proportion of missing data and the strength of recombination signals. Ten remaining alignments with a length greater than 2,500 bp and less than 130 hemiplasies (total length: 30,738 bp; completeness: 95.8%), were then used jointly to estimate divergence times with the uncorrelated-lognormal relaxed-clock model implemented in BEAST2. To account for phylogenetic uncertainty in downstream phylogenetic comparative analyses, we performed three separate sets of relaxed clock analyses, in which the topology was either fixed to the species-level phylogeny inferred with RAxML (Fig. 1, Extended Data Fig. 2), the species tree inferred with ASTRAL (Extended Data Fig. 3) or the Bayesian species tree inferred with SNAPP (Extended Data Fig. 4). Further details are provided in the Supplementary Methods.

Morphometrics

To quantify body shape and upper oral jaw morphology, we applied a landmark-based geometric morphometric approach to digital X-ray images (for the full set of 10 specimens per species whenever possible; $n = 2{,}197$). We selected 21 landmarks, of which 17 were distributed across the skeleton and four defined the premaxilla (Extended Data Fig. 5a).

Landmark coordinates were digitized using FIJI[55] (v2.0.0-rc-68/1.521i). To extract overall body shape information, we excluded landmark 16, which marks the lateral end of the premaxilla, hence minimizing the impact of the orientation of the upper oral jaw. We then applied a Procrustes superimposition to remove the effect of size, orientation, and translational position of the coordinates.

For upper oral jaw morphology, we used a subset of four landmarks. A crucial feature of the oral jaw morphology is the orientation of the mouth relative to the body axes. However, this component of the upper oral jaw morphology would be lost in a classical geometric morphometric analysis, in which only pure shape information is retained. To overcome this, we extracted the premaxilla-specific landmarks (1, 2, 16 and 21) after Procrustes superimposition of the entire set of landmarks and subsequently recentred the landmarks to align the specimens without rotation. Thus, the resulting landmark coordinates do not represent the pure shape of the premaxilla but additionally contain information on its orientation and size in relation to body axes and body size, respectively.

To quantify lower pharyngeal jaw bone shape in 3D, a landmark-based geometric morphometric approach was applied on μCT scans of the head region of five specimens per species (n = 1,168). To capture all potential functionally important structures of the lower pharyngeal jaw bone, we selected a set of 27 landmarks (10 true landmarks and 17 sliding semi-landmarks) well distributed across the left side of the bone (Extended Data Fig. 5b). Landmark coordinates were acquired using TINA[56] (v.6.0). To retain the lateral symmetric properties of the shape data during superimposition, we reconstructed the right side of the lower pharyngeal jaw bone by mirroring the landmark coordinates across the plane of bilateral symmetry fitted through all landmarks theoretically lying on this plane. We then superimposed the resulting 42 landmarks while sliding the semi-landmarks along the curves by minimizing Procrustes distances and retained the symmetric component only.

To identify the major axes of shape variation across the multivariate datasets we performed a PCA for each trait. We also calculated morphospace size per tribe as the square root of the convex hull area spanned by species means of the PC1 and PC2 scores. We then tested for a correlation between morphospace size and estimated species richness of a tribe[15] (log-transformed to obtain normal distribution). To account for phylogenetic non-independence, we calculated phylogenetic independent contrasts with the R package ape[57] (v.5.2) using the species tree (Fig. 1) pruned to the tribe level. We then calculated Pearson's correlation coefficients for independent contrasts using the function cor.table of the R package picante[58] (v.1.8).

All landmark coordinates for geometric morphometric analyses were processed and analysed in R[59] (v.3.5.2) using the packages geomorph[60] (v.3.0.7) and Morpho[61] (v.2.6). Further details are provided in the Supplementary Methods.

Stable-isotope analysis

To approximate ecology for each species, we measured the stable carbon (C) and nitrogen (N) isotope composition of all available specimens from Lake Tanganyika (n = 2,259). We analysed a small (0.5–1 mg) dried muscle sample of each specimen with a Flash 2000 elemental analyser coupled to a Delta Plus XP continuous-flow isotope ratio mass spectrometer (IRMS) via a Conflo IV interface (Thermo Fisher Scientific). Carbon and nitrogen isotope data were normalized to the VPDB (Vienna Pee Dee Belemnite) and Air-N_2 scales, respectively, using laboratory standards which were calibrated against international standards. Values are reported in standard per-mil notation (‰), and long-term analytical precision was 0.2‰ for $\delta^{13}C$ values and 0.1‰ for $\delta^{15}N$ values. Note that we have used some of these stable-isotope values in a previous study[62].

To confirm interpretability of the $\delta^{13}C$ and $\delta^{15}N$ values, we additionally collected and analysed baseline samples covering several trophic levels from the northern and the southern basin of Lake Tanganyika (Supplementary Methods, Supplementary Discussion).

To test for a correlation of ecospace size with species richness of the tribes, we applied the same approach as described above to the $\delta^{13}C$ and $\delta^{15}N$ values.

Phenotype–environment association

For each trait (body shape, upper oral jaw, lower pharyngeal jaw) we performed a two-block PLS analysis based on species means of the Procrustes aligned landmark coordinates and the stable C and N isotope compositions using the function two.b.pls in geomorph. To account for phylogenetic dependence of the data we applied a pGLS as implemented in the R package caper[63] (v.1.0.1) across the two sets of PLS scores (each morphological axis and the stable-isotope projection) using the time-calibrated species tree based on the maximum-likelihood topology. The strength of phylogenetic signal in the data was accounted for by optimising the branch length transformation parameter lambda using a maximum-likelihood approach.

Scoring pigmentation patterns

To quantify a putative signalling trait in cichlids, we scored the pigmentation patterns in typically five male specimens per species (n = 1,016), on the basis of standardized images taken in the field after capture of the specimens (see Supplementary Methods). Following the strategy described in Seehausen et al.[64], the presence or absence of 20 pigmentation features was recorded, whereby we extended number of scored features to include additional body and fin pigmentation patterns (Extended Data Fig. 5c). We then applied a logistic PCA implemented in the R package logisticPCA[65] (v.0.2) and used the PC1 scores as univariate proxy for differentiation along the signalling axes for further analyses.

Trait evolution modelling and disparity estimates

To investigate the temporal dynamics of morphological diversification over the course of the radiation we essentially followed the strategy of Cooney et al.[28] (which is based on measurements on extant taxa and assumes constant niche space and no or constant extinction over the course of the radiation), using the PLS scores of body shape, upper oral jaw morphology, and lower pharyngeal jaw shape and the PC1 scores of pigmentation patterns as well as the time-calibrated maximum-likelihood species tree topology. For each trait we assessed the phylogenetic signal in the data by calculating Pagel's lambda and Blomberg's K with the R package phytools[66] (v.0.6-60). We then tested the fit of four models of trait evolution for each of the four traits. We applied a white noise model, a Brownian motion model, a single-optimum Ornstein–Uhlenbeck model and an early burst model of trait evolution using the function fitContinuous of the R package geiger[67] (v.2.0.6.1). Additionally, we fitted a variable-rates model (a Brownian motion model which allows for rate shift on branches and nodes) using the software BayesTrait (http://www.evolution.rdg.ac.uk/; v.3) with uniform prior distributions adjusted to our dataset (alpha: −1–1, sigma: 0–0.001 for morphometric traits; alpha: 0–10, sigma: 0–10 for pigmentation pattern) and applying single-chain Markov-chain Monte Carlo runs with one billion iterations. We sampled parameters every 100,000th iteration, after a pre-set burnin of 10,000,000 iterations. We then tested for each trait for convergence of the chain using a Cramer–von Mises statistic implemented in the R package coda[68] (v.0.19-3). The models were compared by calculating their log-likelihood and Akaike information criterion (AIC) difference (Extended Data Fig. 8d). Based on differences in AIC, the variable-rates model was best supported for all traits but body shape, which showed a strong signal of an early burst of trait evolution (Extended Data Fig. 8d, note that the variable-rates model has the highest log-likelihood for body shape as well). We nevertheless focused on the variable-rates model for further analyses of all traits to be able to compare temporal patterns of trait evolution among the traits.

To estimate morphospace expansion through time we used a maximum-likelihood ancestral-state reconstruction implemented in phytools. To account for differences in the rate of trait evolution along the phylogeny, we reconstructed ancestral states using the mean rate-transformed tree derived from the variable-rates model. We then projected the ancestral states onto the original species tree and calculated the morphospace extent (that is, the range of trait values) in time intervals of 0.15 million years (note that this is an arbitrary value; however, differently sized time intervals had no effect on the interpretation of the results). For each time point we extracted the branches existing at that time and predicted the trait value linearly between nodes. We then compared the resulting morphospace expansion over time relative to a null model of trait evolution. We therefore simulated 500 datasets (PLS and PC1 scores) under Brownian motion given the original species tree with parameters derived from the Brownian motion model fit to the original data. For each simulated dataset we produced morphospace-expansion curves using the same approach as described above. We then compared the slopes of our observed data with each of the null models by calculating the difference of slopes through time (Fig. 3) using linear models fitted for each time interval with the two subsequent time intervals. Note that for body shape we also estimated morphospace expansion through time using the early burst model for ancestral-state reconstruction, which resulted in a very similar pattern of trait diversification.

Unlike other metrics of disparity (for example, variance or mean pairwise distances) morphospace extent is not sensitive to the density distribution of measurements within the morphospace and captures its full range[69]. Hence, comparing the extent of morphospace between observed data and the null model directly unveils the contribution of morphospace expansion relative to the null model; and because the increase in lineages over time is identical in the observed and the simulated data, this comparison also provides an estimate for morphospace packing.

To summarize evolutionary rates we calculated the mean rate of trait evolution inferred by the variable-rates model in the same 0.15 million years intervals along the phylogeny.

To account for phylogenetic uncertainty in the tree topology we repeated the analyses of trait evolution using the time-calibrated trees based on tree topologies estimated with ASTRAL and SNAPP (Extended Data Figs. 3, 4; Supplementary Methods; Supplementary Discussion). Furthermore, to also account for uncertainty in branch lengths, we repeated the analysis on 100 trees from the Bayesian posterior distribution for each of the three trees (Extended Data Fig. 8d, e, results are provided on Dryad).

Further details can be found in the Supplementary Methods.

Characterization of repeat content

For the repeat content analysis, we randomly selected one de novo genome assembly per species of the radiation ($n = 245$). We performed a de novo identification of repeat families using RepeatModeler (v.1.0.11; http://www.repeatmasker.org). We then combined the RepeatModeler output library with the available cichlid-specific libraries (Dfam and RepBase; v.27.01.2017; http://www.repeatmasker.org; 258 ancestral and ubiquitous sequences, 161 cichlid-specific repeats, and 6 lineage-specific sequences; 65,118, 273,530 and 6,667 bp in total, respectively) and used the software RepeatMasker (v.4.0.7; http://www.repeatmasker.org) (-xsmall -s -e ncbi -lib combined_libraries.fa) to identify and soft-mask interspersed repeats and low complexity DNA sequences in each assembly. The reported summary statistics were obtained using RepeatMasker's buildSummary.pl script (Fig. 4a, Extended Data Fig. 9a, results per genome are provided on Dryad).

Gene duplication estimates

Per genome, gene duplication events were identified with the structural variant identification pipeline smoove (population calling method; https://github.com/brentp/smoove, docker image cloned 20/12/2018), which builds upon lumpy[70], svtyper[71] and svtools (https://github.com/hall-lab/svtools). Variants were called per sample ($n = 488$ genomes, 246 taxa of the Tanganyika radiation) from the initial mapping files against the Nile tilapia reference genome with the function 'call'. The union of sites across all samples was obtained with the function 'merge', then all samples were genotyped at those sites with the function 'genotype', and depth information was added with --duphold. Genotypes were combined with the function 'paste' and annotated with 'annotate' and the reference genome annotation file. The obtained VCF file was filtered with BCFtools to keep only duplications longer than 1 kb and of high quality (MSHQ >3 or MSHQ = −1, FMT/DHFFC[0] > 1.3, QUAL >100). The resulting file was loaded into R (v.3.6.0) with vcfR[72] (v.1.8.0) and filtered to keep only duplications with less than 20% missing genotypes. Next, we removed duplication events with a length outside 1.5 times the interquartile range above the upper quartile of all duplication length, resulting in a final dataset of 476 duplications (Fig. 4b).

Analyses of selection on coding sequence

To predict genes within the de novo genome assemblies, we used AUGUSTUS[73] (v.3.2.3) with default parameters and 'zebrafish' as species parameter ($n = 485$ genomes, 245 taxa). For each prediction we inferred orthology to Nile tilapia genes (GCF_001858045.1_ASM185804v2) with GMAP (GMAP-GSNAP[74]; v.2017-08-15) applying a minimum trimmed coverage of 0.5 and a minimum identity of 0.8. We excluded specimens with less than 18,000 tilapia orthologous genes detected (resulting in $n = 471$ genomes, 243 taxa). Next, we kept only those tilapia protein coding sequences that had at least one of their exons present in at least 80% of the assemblies (260,335 exons were retained, representing 34,793 protein coding sequences). Based on the Nile tilapia reference genome annotation file, we reconstructed for each assembly the orthologous coding sequences. Missing exon sequences were set to Ns. We then kept a single protein coding sequence per gene (the one being present in the maximum number of species with the highest percentage of sequence length), resulting in 15,294 protein coding sequences. Per gene, a multiple sequence alignment was then produced using MACSE[75] (v.2.01). We calculated for each specimen and each gene the number of synonymous (S) and non-synonymous (N) substitutions by pairwise comparison to the orthologue tilapia sequence using codeml with runmode −2 within PAML[76] (v.4.9e). To obtain an estimate of the genome-wide sequence evolution rate that is independent of filtering thresholds, we calculated the genome-wide dN/dS ratio for each specimen based on the sum of dS and dN across all genes (Fig. 4c, Extended Data Fig. 9b).

Signals of past introgression

We used the f_4-ratio statistic[34] to assess genomic evidence for interspecific gene exchange. We calculated the f_4-ratio for all combinations of trios of species on the filtered VCF files using the software Dsuite[77] (v.0.2 r20), with *T. sparrmanii* as outgroup species (we excluded *N. cancellatus* as all specimens of this species appeared to be F_1 hybrids; Supplementary Methods). The f_4-ratio statistic estimates the admixture proportion, that is, the proportion of the genome affected by gene flow. The results presented in this study (Fig. 4e, Extended Data Fig. 10) are based on the 'tree' output of the Dsuite function Dtrios, with each trio arranged according to the species tree on the basis of the maximum-likelihood topology. The per-tribe analyses (Fig. 4e) were based only on comparisons where all species within a trio belong to the same tribe ($n = 243$ taxa).

In addition to the f_4-ratio we also identified signals of past introgression among species using a phylogenetic approach by testing for asymmetry in the relationships of species trios in 1,272 local maximum-likelihood trees generated using IQ-TREE (Supplementary Methods; Extended Data Fig. 10).

Heterozygosity

We calculated the number of heterozygous sites per genome ($n = 488$ genomes, 246 taxa from the Tanganyika radiation) from the VCF files using the BCFtools function stats and then quantified the percentage of heterozygous sites among the number of callable sites per genome (see above) (Fig. 4d).

To explore if the observed levels of heterozygosity per tribe can be explained by the levels of gene flow within tribes we performed coalescent simulations with msprime[78] (v.0.7.4). We simulated genome evolution of all species of the radiation following the time-calibrated species tree (Fig. 1), assuming a generation time of 3 years[79] and a constant effective population size of 20,000 individuals. Species divergences were implemented as mass migration events and introgression within tribes as migration between species pairs with rates set according to their introgression (f_4-ratio) signals inferred with Dsuite. To convert the f_4-ratio values into migration rates, we applied a scaling factor of 5×10^{-6}, which results in a close correspondence in magnitude of the simulated introgression signals to those observed empirically (Fig. 4, Extended Data Fig. 9c). In each of 20 separate simulations, we randomly sampled one pairwise f_4-ratio value for each pair of species (there are many f_4 ratios per species pair—one for each possible third species added to the test trio; the maximum values per pair are shown in Extended Data Fig. 10). The simulated data consisted of one chromosome of 100 kb (mutation rate: 3.5×10^{-9} per bp per generation[33], recombination rate: 2.2×10^{-8} per bp per generation; see Supplementary Methods). Levels of heterozygosity were calculated for all simulated datasets as described for the empirical data.

To account for between-tribe gene flow we further performed simulations in which migration between tribes was also sampled from the empirical f_4-ratio distribution. For simplicity in setting up the simulation model, we assume that gene flow between tribes is ongoing until present day, which is clearly an overestimate (see Supplementary Discussion). Nevertheless, the results of these simulations support our hypothesized scenario, confirming that much of the variation in heterozygosity as well as its correlation with species richness can be explained by the observed levels of gene flow.

Correlation of genome-wide statistics with species richness

We tested for a correlation between tribe means (based on species means) of each genomic summary statistics (transposable element counts, number of gene duplications, genome-wide dN/dS ratio, per-genome heterozygosity, and f_4-ratio, as well as the heterozygosity and f_4-ratio statistics derived from simulated genome evolution) and species richness of the tribes, applying the same approach as described above for tests of correlation between morpho- and ecospace size and species richness.

Reporting summary

Further information on research design is available in the Nature Research Reporting Summary linked to this paper.

Data availability

All newly sequenced genomes for this study and their raw reads are available from NCBI under the BioProject accession number PRJNA550295 (https://www.ncbi.nlm.nih.gov/bioproject/). The VCF file, tree files, summary statistics of the assembled genomes and phenotypic datasets generated and analysed during this study are available as downloadable files on Dryad (https://doi.org/10.5061/dryad.9w0vt4bbf). The Nile tilapia reference genome used is available under RefSeq accession GCF_001858045.1. All X-ray data are available on MorphoSource under the project number P1093. Source data are provided with this paper.

Code availability

Code used to analyse the data is available on GitHub (https://github.com/cichlidx/ronco_et_al), except for analyses where single commands from publicly available software were used and where all settings are fully reported in the Methods and/or Supplementary Methods.

35. Bolger, A. M., Lohse, M. & Usadel, B. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* **30**, 2114–2120 (2014).
36. Conte, M. A., Gammerdinger, W. J., Bartie, K. L., Penman, D. J. & Kocher, T. D. A high quality assembly of the Nile Tilapia (*Oreochromis niloticus*) genome reveals the structure of two sex determination regions. *BMC Genomics* **18**, 341 (2017).
37. Li, H. & Durbin, R. Fast and accurate short read alignment with Burrows–Wheeler transform. *Bioinformatics* **25**, 1754–1760 (2009).
38. McKenna, A. et al. The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res.* **20**, 1297–1303 (2010).
39. Li, H. A statistical framework for SNP calling, mutation discovery, association mapping and population genetical parameter estimation from sequencing data. *Bioinformatics* **27**, 2987–2993 (2011).
40. Danecek, P. et al. The variant call format and VCFtools. *Bioinformatics* **27**, 2156–2158 (2011).
41. Browning, S. R. & Browning, B. L. 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 (2007).
42. Böhne, A. et al. Repeated evolution versus common ancestry: Sex chromosome evolution in the haplochromine *Pseudocrenilabrus philander*. *Genome Biol. Evol.* **11**, 439–458 (2019).
43. Malmstrøm, M., Matschiner, M., Tørresen, O. K., Jakobsen, K. S. & Jentoft, S. Data descriptor: Whole genome sequencing data and de novo draft assemblies for 66 teleost species. *Sci. Data* **4**, 1–13 (2017).
44. Myers, E. W. et al. A whole-genome assembly of *Drosophila*. *Science* **287**, 2196–2204 (2000).
45. Magoč, T. & Salzberg, S. L. FLASH: fast length adjustment of short reads to improve genome assemblies. *Bioinformatics* **27**, 2957–2963 (2011).
46. Gurevich, A., Saveliev, V., Vyahhi, N. & Tesler, G. QUAST: quality assessment tool for genome assemblies. *Bioinformatics* **29**, 1072–1075 (2013).
47. Simão, F. A., Waterhouse, R. M., Ioannidis, P., Kriventseva, E. V. & Zdobnov, E. M. BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. *Bioinformatics* **31**, 3210–3212 (2015).
48. Ewels, P., Magnusson, M., Lundin, S., Käller, M. & Multi, Q. C. MultiQC: summarize analysis results for multiple tools and samples in a single report. *Bioinformatics* **32**, 3047–3048 (2016).
49. Zhang, C., Rabiee, M., Sayyari, E. & Mirarab, S. ASTRAL-III: polynomial time species tree reconstruction from partially resolved gene trees. *BMC Bioinformatics* **19** (Suppl 6), 153 (2018).
50. Bouckaert, R. et al. BEAST 2.5: An advanced software platform for Bayesian evolutionary analysis. *PLOS Comput. Biol.* **15**, e1006650 (2019).
51. Ogilvie, H. A., Bouckaert, R. R. & Drummond, A. J. StarBEAST2 brings faster species tree inference and accurate estimates of substitution rates. *Mol. Biol. Evol.* **34**, 2101–2114 (2017).
52. Stamatakis, A. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* **30**, 1312–1313 (2014).
53. Nguyen, L.-T., Schmidt, H. A., von Haeseler, A. & Minh, B. Q. IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. *Mol. Biol. Evol.* **32**, 268–274 (2015).
54. Bryant, D., Bouckaert, R., Felsenstein, J., Rosenberg, N. A. & RoyChoudhury, A. Inferring species trees directly from biallelic genetic markers: bypassing gene trees in a full coalescent analysis. *Mol. Biol. Evol.* **29**, 1917–1932 (2012).
55. Schindelin, J. et al. Fiji: an open-source platform for biological-image analysis. *Nat. Methods* **9**, 676–682 (2012).
56. Schunke, A. C., Bromiley, P. A., Tautz, D. & Thacker, N. A. TINA manual landmarking tool: software for the precise digitization of 3D landmarks. *Front. Zool.* **9**, 6 (2012).
57. Paradis, E., Claude, J. & Strimmer, K. APE: Analyses of phylogenetics and evolution in R language. *Bioinformatics* **20**, 289–290 (2004).
58. Kembel, S. W. et al. Picante: R tools for integrating phylogenies and ecology. *Bioinformatics* **26**, 1463–1464 (2010).
59. R Development Core Team. R: A language and environment for statistical computing. *R Foundation for Statistical Computing* (2018).
60. Adams, D. C. & Otárola-Castillo, E. Geomorph: An R package for the collection and analysis of geometric morphometric shape data. *Methods Ecol. Evol.* **4**, 393–399 (2013).
61. Schlager, S. in *Statistical Shape and Deformation Analysis* (eds Zheng, G., Li, S. & Szekely, G.) 217–256 (Academic Press, 2017).
62. Ronco, F., Roesti, M. & Salzburger, W. A functional trade-off between trophic adaptation and parental care predicts sexual dimorphism in cichlid fish. *Proc. R. Soc. Lond. B* **286**, 20191050 (2019).
63. Orme, D. *The Caper Package: Comparative Analysis of Phylogenetics and Evolution in R* https://cran.r-project.org/web/packages/caper/vignettes/caper.pdf (2018).
64. Seehausen, O., Mayhew, P. J. & Van Alphen, J. J. M. Evolution of colour patterns in East African cichlid fish. *J. Evol. Biol.* **12**, 514–534 (1999).
65. Landgraf, A. J. & Lee, Y. Dimensionality reduction for binary data through the projection of natural parameters. *J. Multivar. Anal.* 104668 (2020).
66. Revell, L. J. phytools: An R package for phylogenetic comparative biology (and other things). *Methods Ecol. Evol.* **3**, 217–223 (2012).

67. Harmon, L. J., Weir, J. T., Brock, C. D., Glor, R. E. & Challenger, W. GEIGER: investigating evolutionary radiations. *Bioinformatics* **24**, 129–131 (2008).
68. Plummer, M., Best, N., Cowles, K. & Vines, K. CODA: convergence diagnosis and output analysis for MCMC. *R News* **6**, 7–11 (2005).
69. Ciampaglio, C. N., Kemp, M. & McShea, D. W. Detecting changes in morphospace occupation patterns in the fossil record: characterization and analysis of measures of disparity. *Paleobiology* **27**, 695–715 (2001).
70. Layer, R. M., Chiang, C., Quinlan, A. R. & Hall, I. M. LUMPY: a probabilistic framework for structural variant discovery. *Genome Biol.* **15**, R84 (2014).
71. Chiang, C. et al. SpeedSeq: ultra-fast personal genome analysis and interpretation. *Nat. Methods* **12**, 966–968 (2015).
72. Knaus, B. J. & Grünwald, N. J. vcfr: a package to manipulate and visualize variant call format data in R. *Mol. Ecol. Resour.* **17**, 44–53 (2017).
73. Stanke, M., Schöffmann, O., Morgenstern, B. & Waack, S. Gene prediction in eukaryotes with a generalized hidden Markov model that uses hints from external sources. *BMC Bioinformatics* **7**, 62 (2006).
74. Wu, T. D. & Watanabe, C. K. GMAP: a genomic mapping and alignment program for mRNA and EST sequences. *Bioinformatics* **21**, 1859–1875 (2005).
75. Ranwez, V., Douzery, E. J. P., Cambon, C., Chantret, N. & Delsuc, F. MACSE v2: Toolkit for the alignment of coding sequences accounting for frameshifts and stop codons. *Mol. Biol. Evol.* **35**, 2582–2584 (2018).
76. Yang, Z. PAML 4: phylogenetic analysis by maximum likelihood. *Mol. Biol. Evol.* **24**, 1586–1591 (2007).
77. Malinsky, M., Matschiner, M. & Svardal, H. Dsuite–fast *D*-statistics and related admixture evidence from VCF files. *Methods Ecol. Evol*. https://doi.org/10.1111/1755-0998.13265 (2020).
78. Kelleher, J., Etheridge, A. M. & McVean, G. Efficient coalescent simulation and genealogical analysis for large sample sizes. *PLOS Comput. Biol.* **12**, e1004842 (2016).
79. Malinsky, M. et al. Genomic islands of speciation separate cichlid ecomorphs in an East African crater lake. *Science* **350**, 1493–1498 (2015).

Acknowledgements We thank the University of Burundi, the Ministère de l'Eau, de l'Environnement, de l'Amenagement du Territoire et de l'Urbanisme, Republic of Burundi, the Centre de Recherche en Hydrobiologie (CRH), Uvira, DR Congo, the Tanzania Commission for Science and Technology (COSTECH), the Tanzania Fisheries Research Institute (TAFIRI), the Tanzania National Parks Authority (TANAPA), the Tanzania Wildlife Research Institute (TAWIRI), the Lake Tanganyika Research Unit, Department of Fisheries, Republic of Zambia, and the Zambian Department for Immigration for research permits; G. Banyankimbona, H. Mwima, G. Hakizimana, N. Muderhwa, P. Masilya, I. Kimirei, M. Mukuli Wa-Teba, G. Moshi, A. Mwakatobe, C. Katongo, T. Banda and L. Makasa for assistance with obtaining research permits; the boat crews of the *Chomba* (D. Mwanakulya, J. Sichilima, H. D. Sichilima Jr and G. Katai) and the *Maji Makubwa II* (G. Kazumbe and family) for navigation, guidance and company; the boat drivers M. Katumba and T. Musisha; the car drivers A. Irakoze and J. Leonard; M. Schreyen-Brichard, M. Mukuli Wa-Teba, G. Kazumbe, I. Kimirei, D. Schlatter, R. Schlatter, M. K. Dominico, H. Sichilima Sr, C. Zytkow, P. Lassen and V. Huwiler for logistic support; G. Banyankimbona, N. Boileau, B. Egger, Y. Fermon, G. Kazumbe, G. Katai, R. Lusoma, K. Smailus, L. Widmer and numerous fishermen at Lake Tanganyika for help during sampling; V. Huwiler, Charity, O. Mangwangwa and the Zytkow family for lodging; people of innumerable villages on the shores of Lake Tanganyika for providing workspace, shelter for night-camps and access to village infrastructure; M. Barluenga, H. Gante, Z. Musilová, F. Schedel, J. Snoeks, M. Stiassny, H. Tanaka, G. Turner and M. Van Steenberge for providing additional samples and/or specimens; M. Sánchez, A. Schweizer and A. Wegmann for assistance with the µCT scanning of large specimens; C. Moes for help with radiographs; V. Evrard for help with stable isotopes; I. Nissen and E. Burcklen for assistance with DNA shearing; M. Conte and T. D. Kocher for sharing the RepeatMasker annotations for Nile tilapia; C. Klingenberg and M. Sánchez for discussions on the morphometric approach; A. Tooming-Klunderud and team at the Norwegian Sequencing Centre and C. Beisel and team at the Genomics Facility Basel at the ETH Zurich Department of Biosystems Science and Engineering (D-BSSE), Basel, for assistance with next-generation sequencing; M. Jacquot, E. Pujades and T. Sengstag for the setup and assistance with the collection database system (LabKey); and J. Johnson and A. Viertler for fish illustrations in Fig. 1 and Extended Data Fig. 5, respectively. Calculations were performed at sciCORE (http://scicore.unibas.ch/) scientific computing centre at University of Basel (with support by the SIB/Swiss Institute of Bioinformatics) and the Abel computer cluster, University of Oslo. This work was funded by the European Research Council (ERC, Consolidator Grant Nr. 617585 'CICHLID-X' jointly hosted by the University of Basel and the University of Oslo) and the Swiss National Science Foundation (SNSF, grants 156405 and 176039) to W.S. A. Böhne was supported by the SNSF (Ambizione grant 161462).

Author contributions F.R., A.I. and W.S. designed this study (with input from H.H.B., A.K. and S.J.). F.R., A.I., H.H.B. and W.S. collected the specimens in the field. F.R. and A. Böhne extracted DNA and prepared the libraries for sequencing. S.J. coordinated sequencing. M. Matschiner performed the mapping, variant calling, phylogenetic analyses and coalescent simulations. M. Malinsky contributed to the variant calling pipelines and performed the f_4-ratio statistics. A. Böhne assembled the genomes and quantified gene duplications, A.E.T. conducted the d*N*/d*S* analyses and V.R. analysed transposable elements. A. Boila assessed stable-isotope compositions, H.H.B. radiographed the specimens and W.S. scored pigmentation patterns. F.R. curated the samples and performed µCT scanning, geometric morphometric analyses, and all analyses incorporating morphological and ecological data as well as correlations with species richness. F.R. and W.S. wrote the manuscript with contributions and/or feedback from all authors. All authors read and approved the final version of the manuscript.

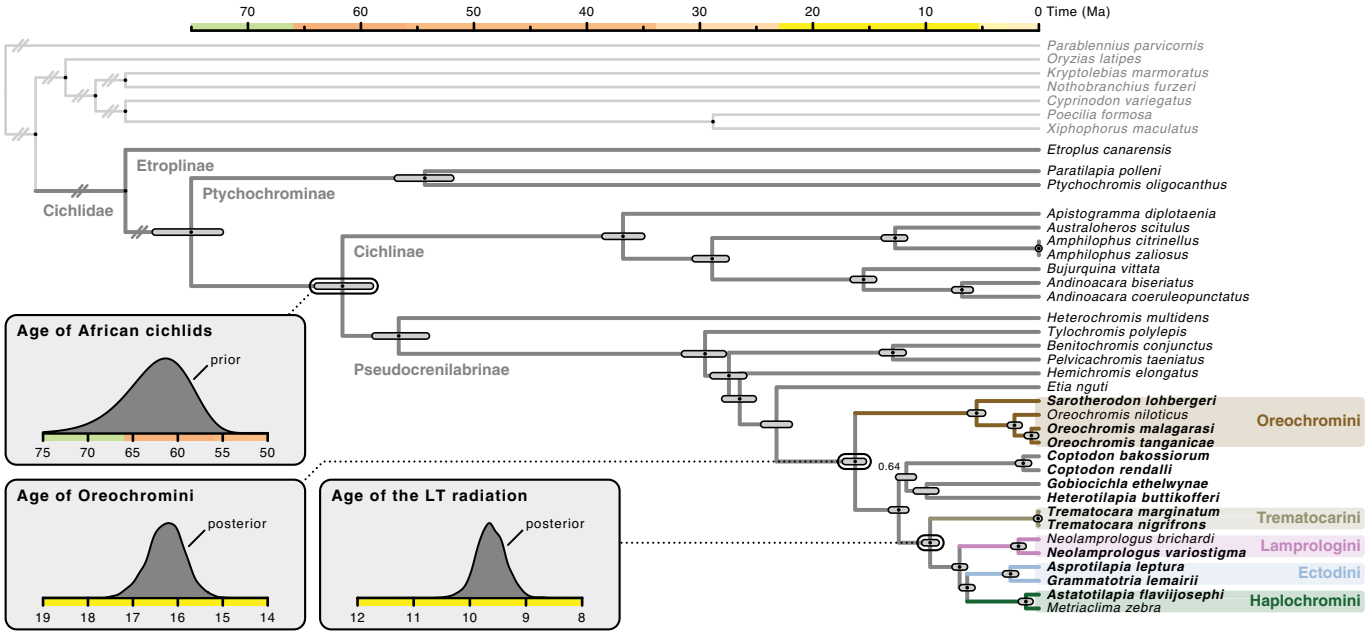
Competing interests The authors declare no competing interests.

Additional information

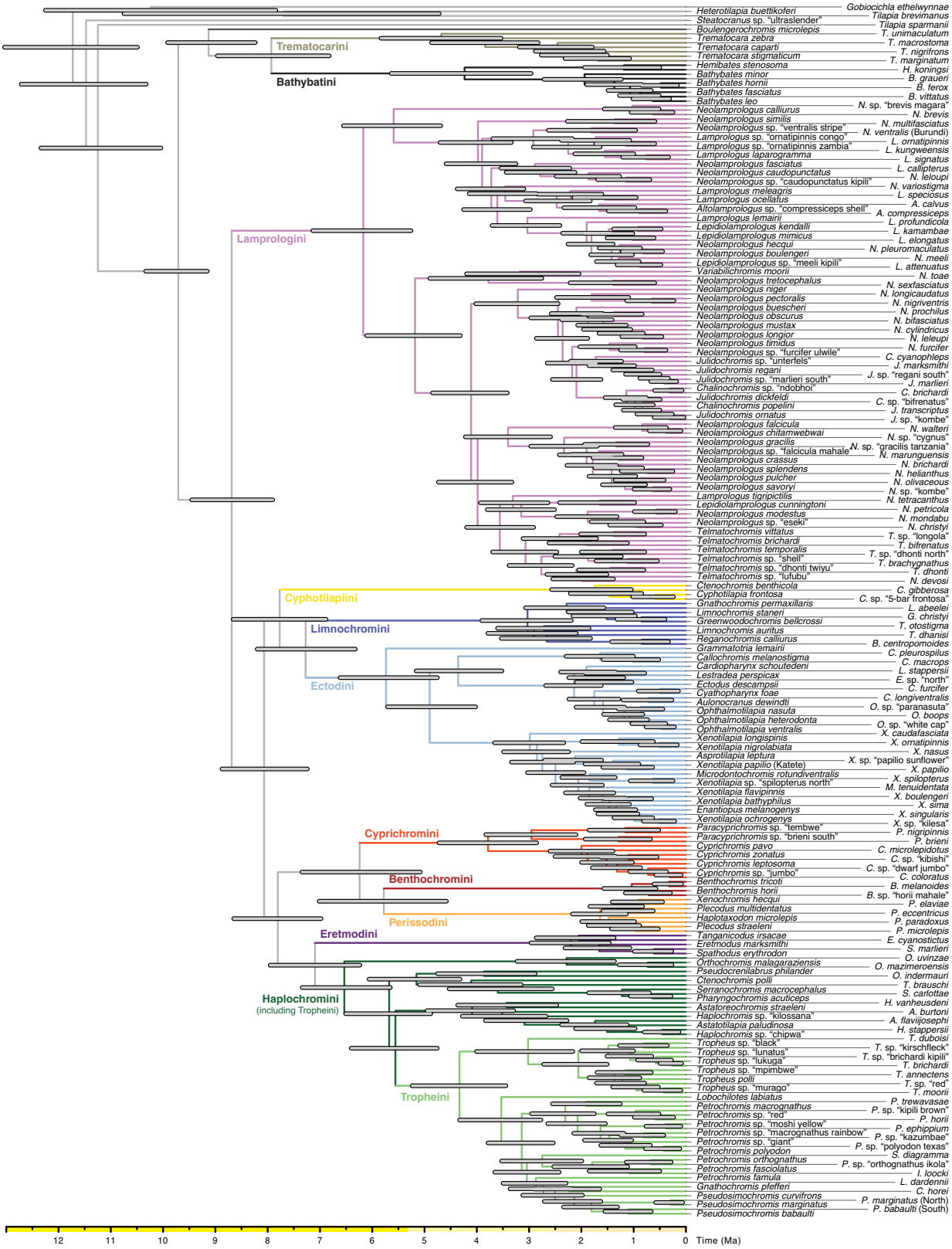
Supplementary information is available for this paper at https://doi.org/10.1038/s41586-020-2930-4.

Correspondence and requests for materials should be addressed to F.R. or W.S.

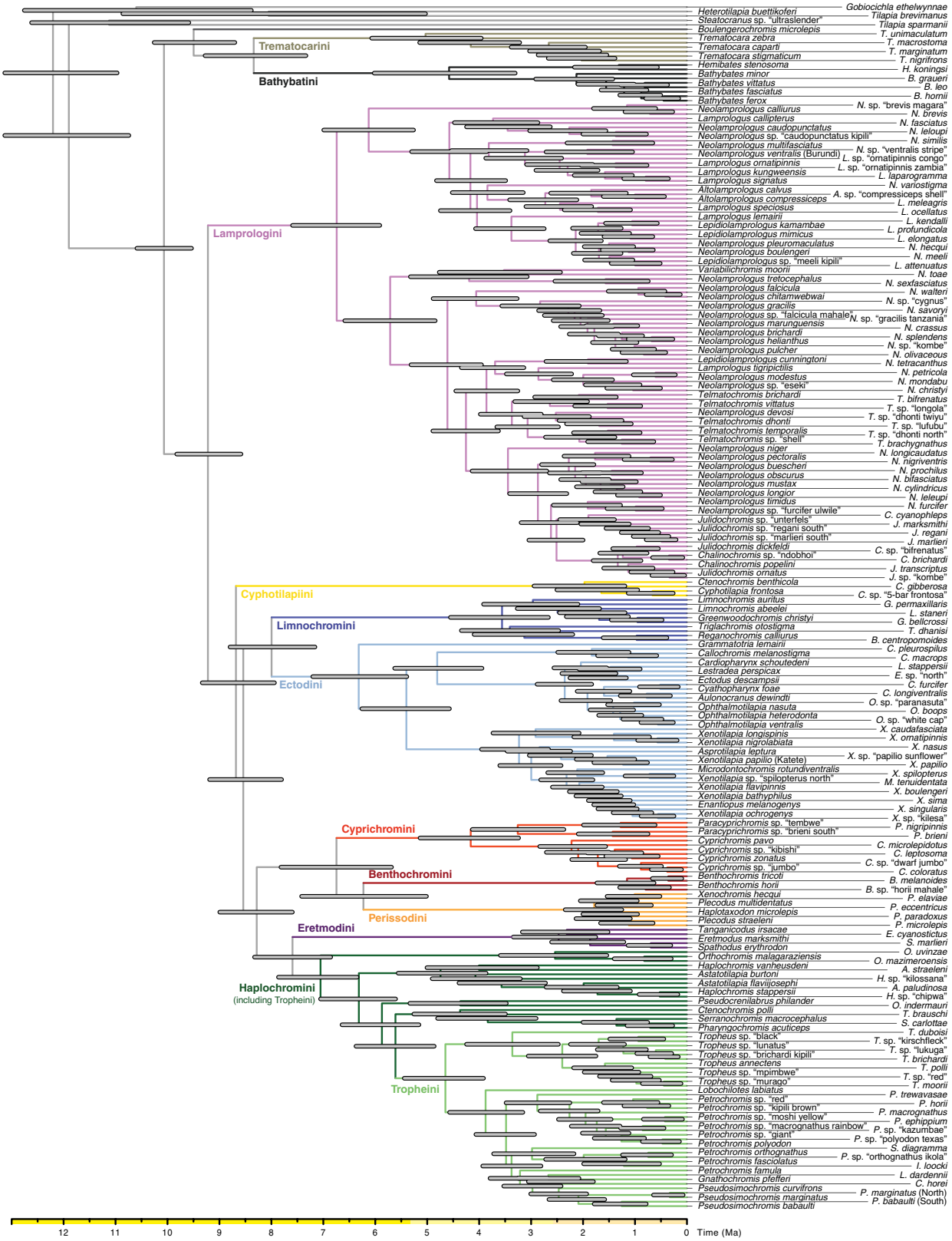
Peer review information *Nature* thanks the anonymous reviewer(s) for their contribution to the peer review of this work.



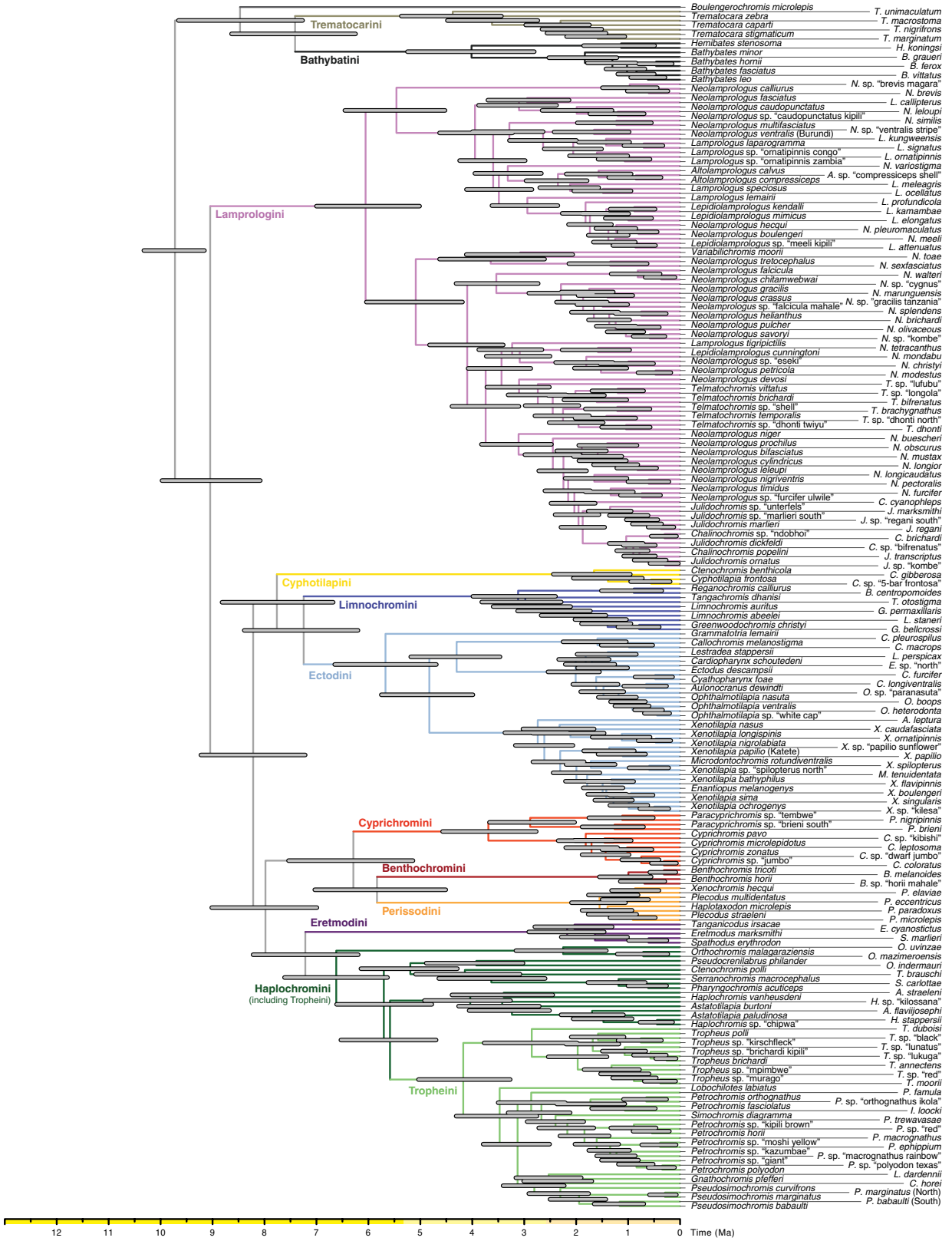
Extended Data Fig. 1 | Age of the adaptive radiation of cichlid fishes in African Lake Tanganyika. Time-calibrated species tree of species representing divergent tribes and subfamilies within cichlids as well as closely-related non-cichlid outgroups, generated with the multi-species coalescent model in StarBEAST2. Nodes marked with a black dot were constrained according to species-tree analyses with ASTRAL. Node bars indicate 95% highest-posterior density age intervals. Outgroup divergence times are not drawn to scale. Insets visualize the prior distribution applied for the age of African cichlids according to Matschiner et al.[18], as well as posterior age estimates for Oreochromini and the cichlid adaptive radiation in Lake Tanganyika (LT).



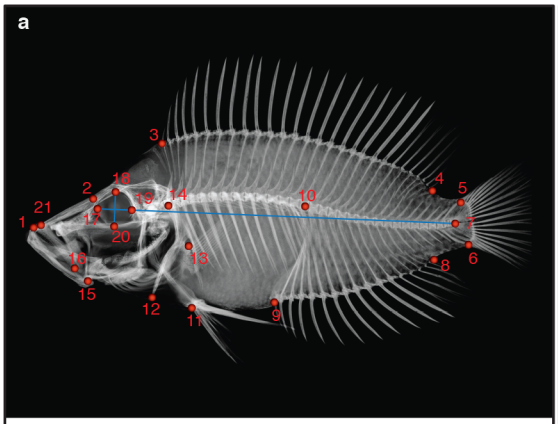
Extended Data Fig. 2 | Time-calibrated species tree of the cichlid adaptive radiation in Lake Tanganyika. The species tree is based on the maximum-likelihood topology estimated with RAxML (Fig. 1) and was time-calibrated using a relaxed-clock model in BEAST2, applied to a selected set of alignments.



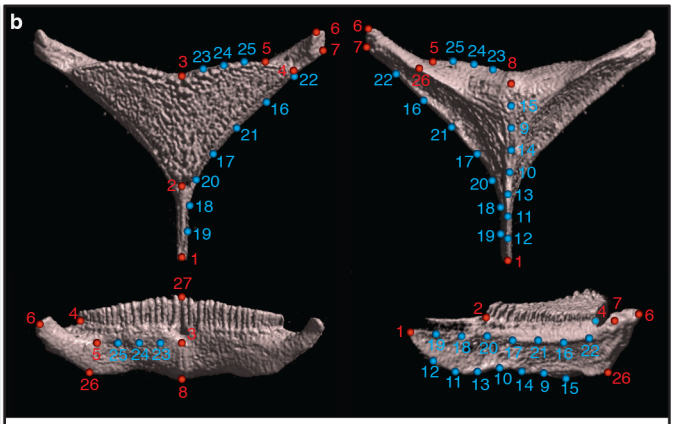
Extended Data Fig. 3 | Alternative time-calibrated species tree of the cichlid adaptive radiation in Lake Tanganyika. The species tree is based on the topology estimated with ASTRAL and was time-calibrated using a relaxed-clock model in BEAST2, applied to a selected set of alignments.



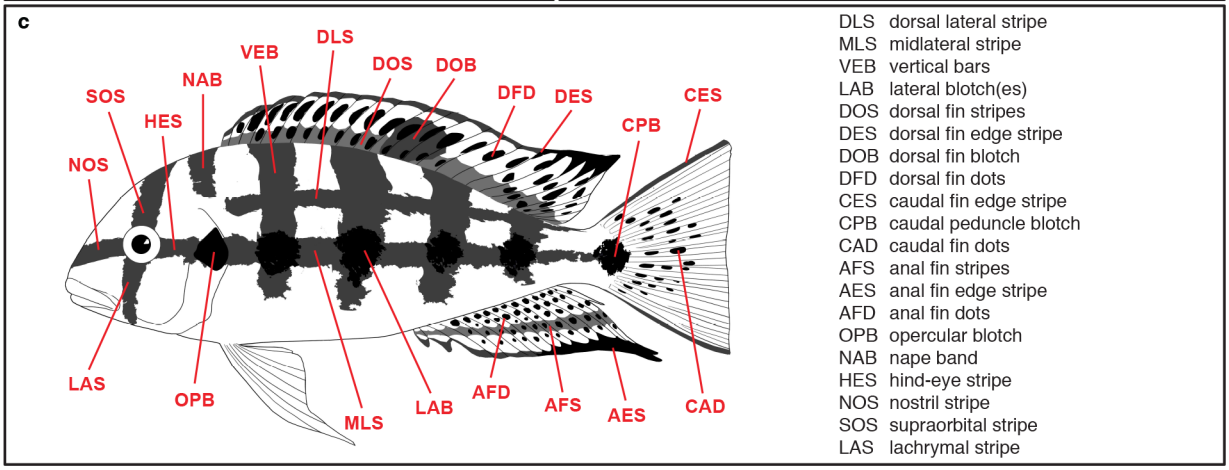
Extended Data Fig. 4 | Alternative time-calibrated species tree of the cichlid adaptive radiation in Lake Tanganyika. The species tree is based on the topology estimated with SNAPP and was time-calibrated using a relaxed-clock model in BEAST2, applied to a selected set of alignments.



1 anterior tip of the premaxilla
2 dorsal tip of the premaxilla
3 basis of the first dorsal fin spine
4 basis of the last dorsal fin ray
5 tip of the last neural spine (branching off the pre-ural centre 3; marking the dorsal basis of the caudal fin)
6 tip of the last haemal spine (branching off the pre-ural centre 3; marking the ventral basis of the caudal fin)
7 centre of the last caudal vertebra (urostyle)
8 basis of the last anal fin ray
9 basis of the first anal fin spine
10 centre of the first caudal vertebra
11 basis of the pelvic fin spine
12 ventral tip of the cleithrum
13 basis of the first (dorsal) pectoral fin ray
14 centre of the first abdominal vertebra
15 joint of the articular and the quadrate
16 lateral end of the premaxilla
17 anterior margin of the eye
18 dorsal margin of the eye
19 posterior margin of the eye
20 ventral margin of the eye
21 angular point on the anterior ridge of the premaxilla

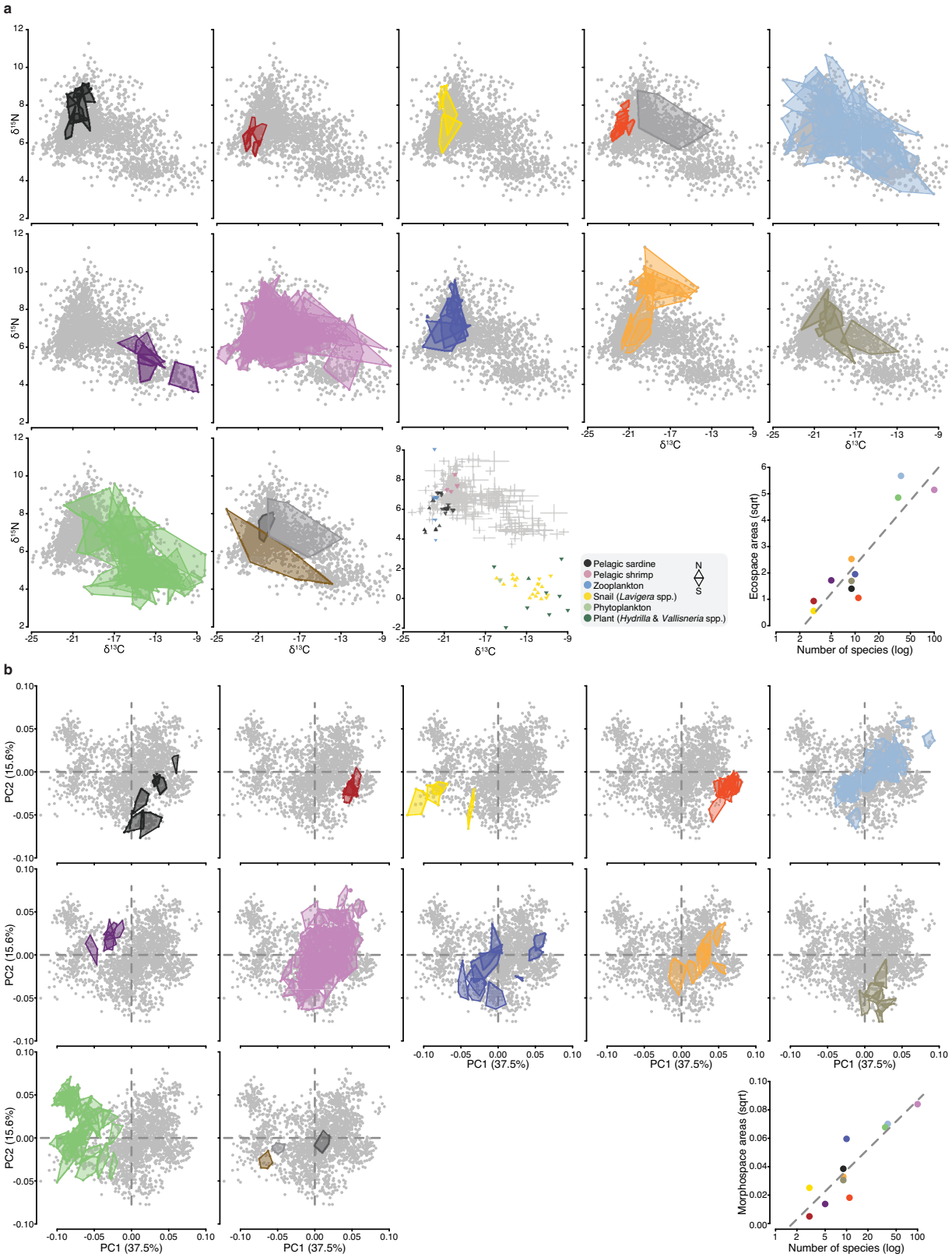


1 anterior tip
2 end of dentition anterior
3 dorsal-posterior centre (~base of first tooth)
4 end of dentition left posterior towards horn
5 angular point centre-to-horn
6 left horn tip sagittal
7 left horn tip lateral
8 ventral-posterior centre
9 ventral sagittal line posterior
10 ventral sagittal line centre
11 ventral sagittal line anterior
12 ventral sagittal line centre in most anterior segment
13 ventral sagittal line centre in centre-anterior segment
14 ventral sagittal line centre in centre-posterior segment
15 ventral sagittal line centre in most posterior segment
16 lateral left line posterior
17 lateral left line centre
18 lateral left line anterior
19 lateral left line centre in most anterior segment
20 lateral left line centre in centre-anterior segment
21 lateral left line centre in centre-posterior segment
22 lateral left line centre in most posterior segment
23 posterior line sagittal
24 posterior line center
25 posterior line lateral
26 angular point ventral-horn-ridge
27 tooth height posterior sagittal

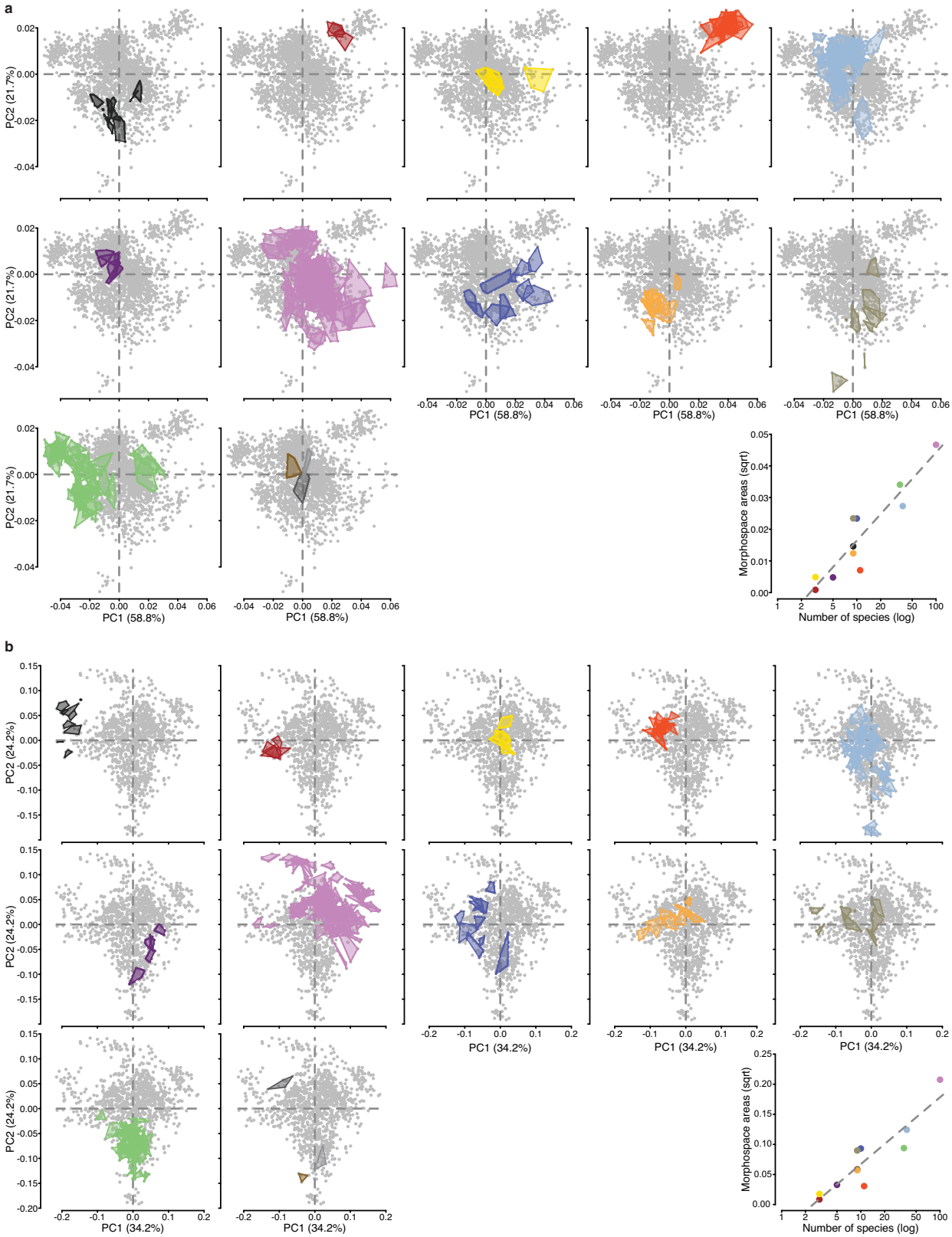


DLS dorsal lateral stripe
MLS midlateral stripe
VEB vertical bars
LAB lateral blotch(es)
DOS dorsal fin stripes
DES dorsal fin edge stripe
DOB dorsal fin blotch
DFD dorsal fin dots
CES caudal fin edge stripe
CPB caudal peduncle blotch
CAD caudal fin dots
AFS anal fin stripes
AES anal fin edge stripe
AFD anal fin dots
OPB opercular blotch
NAB nape band
HES hind-eye stripe
NOS nostril stripe
SOS supraorbital stripe
LAS lachrymal stripe

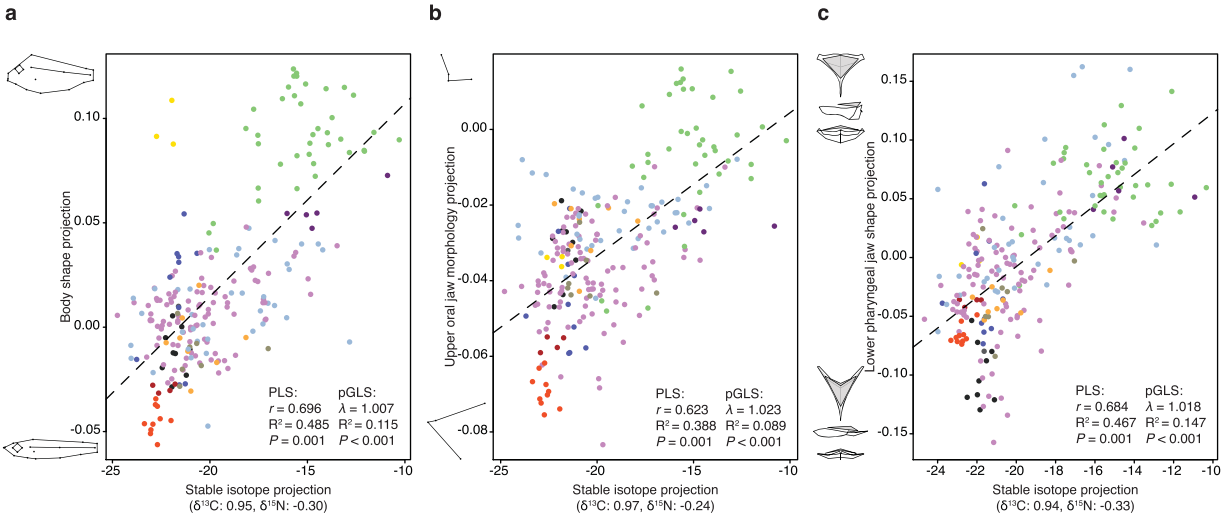
Extended Data Fig. 5 | Phenotyping of the specimens. a, Two-dimensional landmarks placed on X-ray images of the specimens. To quantify overall body shape we excluded landmark 16 (to minimise the effect of the orientation of the oral jaw). To analyse upper oral jaw morphology we used landmarks 1, 2, 16 and 21. **b**, Three-dimensional landmarks used to analyse lower pharyngeal jaw shape on µCT scans of the heads. True landmarks are indicated in red, sliding semi-landmarks are indicated in blue. **c**, Body regions scored for presence/absence of pigmentation patterns.



Extended Data Fig. 6 | Ecospace and morphospace occupation of the cichlid adaptive radiation in Lake Tanganyika. Scatter plots for each focal tribe (indicated with colours, see Fig. 1 for colour key) against the total eco-and morphospace (grey). Species ranges are indicated with convex hulls. **a**, Stable N and C isotope compositions ($\delta^{15}N$ and $\delta^{13}C$ values). The additional plot shows $\delta^{15}N$ and $\delta^{13}C$ values of a baseline dataset which confirms the interpretability of the stable N and C isotope composition in Lake Tanganyika (see Supplementary Methods and Discussion). **b**, PC1 and PC2 of body shape (for shape changes associated with the PC axes see Fig. 2). The last plot for each trait shows the size of the traitspace per tribe in relation to species numbers (stable isotopes: Pearson's $r = 0.88$, d.f. = 9, $P = 0.0004$; body shape: Pearson's $r = 0.91$, d.f. = 9, $P = 0.0001$). Traitspace size was calculated as the square root of the convex hull area spanned by species means.



Extended Data Fig. 7 | Morphospace occupation of the cichlid adaptive radiation in Lake Tanganyika. a, **b**, Scatter plots of PC1 and PC2 for upper oral jaw morphology (**a**) and lower pharyngeal jaw shape per tribe (**b**) (indicated with colours, see Fig. 1 for colour key) against the total morphospace (grey). Species ranges are indicated with convex hulls. For shape changes associated with the respective PC-axis see Fig. 2. The last plot for each trait shows the size of the morphospace per tribe in relation to species numbers (upper oral jaw morphology: Pearson's $r = 0.88$, d.f. = 9, $P = 0.0003$; lower pharyngeal jaw shape: Pearson's $r = 0.83$, d.f. = 9, $P = 0.0017$). Morphospace size was calculated as the square root of the convex hull area spanned by species means.



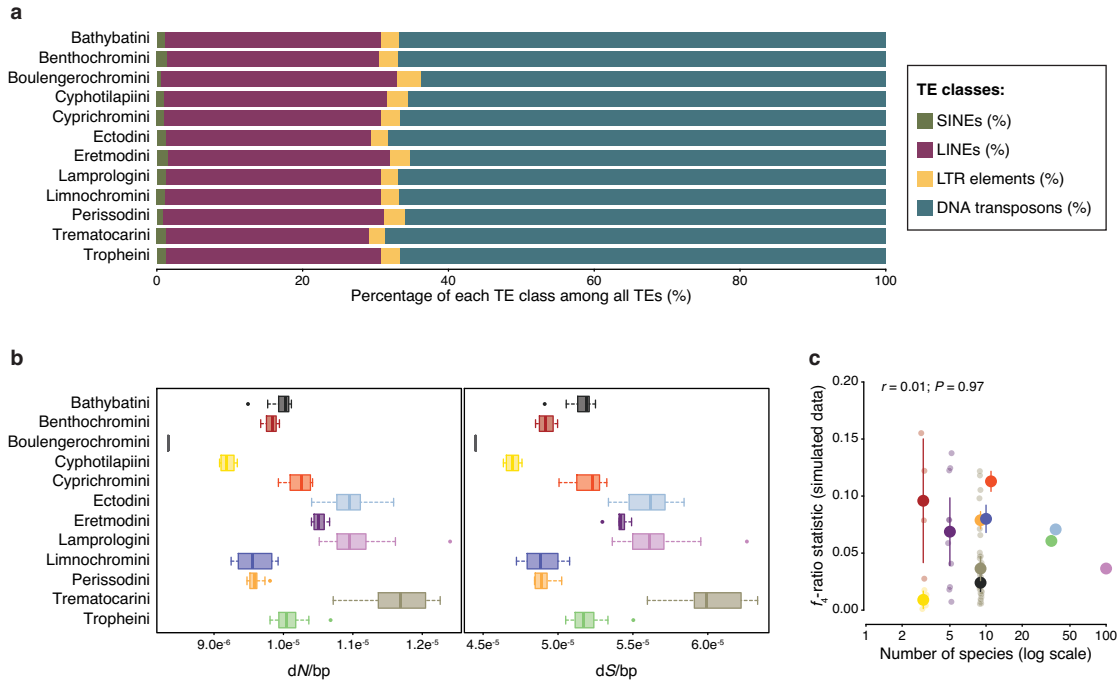
d

Tree topology	Trait	Model comparison log-likelihood difference (AIC difference) White noise	Brownian motion	Ornstein-Uhlenbeck	Early burst	Variable rates	Phylogenetic signal λ	K
Maximum-likelihood (RAxML)	Body shape	256 (477)	21 (8)	21 (10)	16 (**0**)	**0** (22)	*1.01*	*2.14*
	Upper oral jaw morphology	228 (390)	53 (40)	53 (42)	49 (35)	**0** (**0**)	*1.02*	*1.34*
	Lower pharyngeal jaw shape	193 (330)	44 (32)	44 (34)	44 (34)	**0** (**0**)	*1.02*	*1.13*
	Pigmentation pattern	101 (146)	53 (49)	45 (36)	53 (51)	**0** (**0**)	*0.94*	*0.44*
Multi-species coalescent (ASTRAL)	Body shape	258 (483)	23 (11)	23 (13)	16 (**0**)	**0** (21)	*1.02*	*2.28*
	Upper oral jaw morphology	228 (390)	54 (41)	54 (43)	50 (36)	**0** (**0**)	*1.03*	*1.42*
	Lower pharyngeal jaw shape	191 (325)	46 (35)	46 (37)	46 (37)	**0** (**0**)	*1.03*	*1.14*
	Pigmentation pattern	99 (143)	49 (43)	42 (32)	49 (45)	**0** (**0**)	*0.94*	*0.47*
Multi-species coalescent (SNAPP)	Body shape	259 (487)	19 (9)	19 (11)	14 (**0**)	**0** (25)	*1.01*	*2.12*
	Upper oral jaw morphology	225 (386)	48 (31)	48 (33)	44 (26)	**0** (**0**)	*1.02*	*1.32*
	Lower pharyngeal jaw shape	190 (324)	42 (28)	42 (30)	42 (30)	**0** (**0**)	*1.02*	*1.04*
	Pigmentation pattern	101 (147)	50 (44)	43 (32)	50 (46)	**0** (**0**)	*0.94*	*0.43*

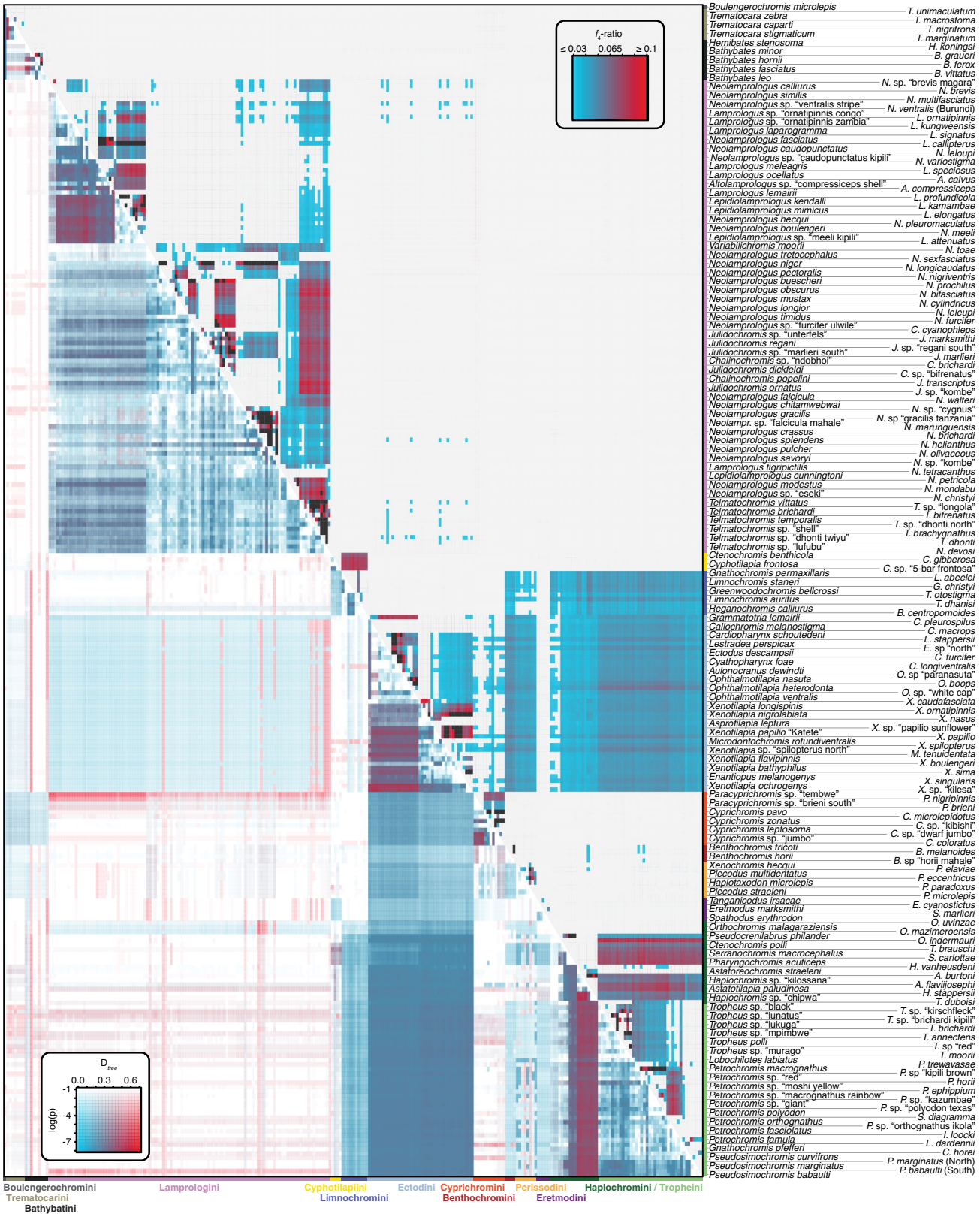
e

Tree topology	Trait	Model comparison (AIC difference) White noise	Brownian motion	Ornstein-Uhlenbeck	Early burst	Variable rates	Phylogenetic signal λ	K	Root age
Maximum-likelihood (RAxML)	Body shape	453-486	2-11	4-13	**0**	5-30	*1-1*	*1.8-2.3*	8.8 - 10.4
	Upper oral jaw morphology	373-398	23-54	25-56	20-55	**0**	*1-1*	*1.1-1.5*	
	Lower pharyngeal jaw shape	319-337	14-65	16-63	16-67	**0**	*0.9-1*	*0.9-1.2*	
	Pigmentation pattern	133-158	37-85	27-59	39-87	**0**	*0.9-1*	*0.3-0.5*	
Multi-species coalescent (ASTRAL)	Body shape	461-494	4-13	6-15	**0**	10-28	*1-1*	*2-2.4*	9.4 - 10.9
	Upper oral jaw morphology	377-397	24-53	26-55	19-52	**0**	*1-1*	*1.2-1.5*	
	Lower pharyngeal jaw shape	316-330	13-60	15-60	15-62	**0**	*1-1*	*0.9-1.3*	
	Pigmentation pattern	130-149	30-83	24-54	32-85	**0**	*0.9-1*	*0.3-0.5*	
Multi-species coalescent (SNAPP)	Body shape	467-498	3-10	5-12	**0**	13-35	*1-1*	*1.9-2.2*	9.0 - 10.5
	Upper oral jaw morphology	371-392	14-43	16-45	10-43	**0**	*1-1*	*1.1-1.4*	
	Lower pharyngeal jaw shape	316-330	10-64	12-61	12-66	**0**	*1-1*	*0.8-1.1*	
	Pigmentation pattern	136-154	35-119	24-71	37-121	**0**	*0.9-1*	*0.3-0.5*	

Extended Data Fig. 8 | PLS fit for each multivariate trait against the stable N and C isotope compositions ($\delta^{15}N$ and $\delta^{13}C$ values) and models of trait evolution. a–**c**, PLS fits for body shape (**a**), upper oral jaw morphology (**b**) and lower pharyngeal jaw shape (**c**). Associated shape changes and loadings of the respective stable isotope projection are indicated next to the axes. Data points represent species means and are coloured according to tribe. **d**, Comparison of model fits for different models of trait evolution and phylogenetic signal for each trait complex using three time-calibrated species trees with alternative topologies. **e**, Overview of the model fits and phylogenetic signal inferred using 100 trees sampled from the posterior distributions of the time calibrations for each of the three alternative tree topologies.



Extended Data Fig. 9 | Genome-wide statistical analyses. a, Proportion of the different classes of transposable elements (TE) among all TE for each tribe (one genome per species, $n = 245$). **b**, Species means of dN (left) and dS (right) values over alignment length for each tribe ($n = 243$ taxa, 471 genomes). The boxes' centre lines show median, box limits show first and third quartiles, and whiskers show the 1.5 × interquartile ranges. **c**, f_4-ratio statistics among species within each tribe in simulated data (tribe means are based on the mean across 20 simulations of each species triplet). Data points are coloured according to tribes; large points are tribe means shown with 95% confidence intervals, small points represent species means and are only shown for group sizes <40 species. To test for a correlation with species richness per tribe (log-transformed), we calculated phylogenetic independent contrasts for each variable and inferred Pearson's r through the origin.



Extended Data Fig. 10 | Signals of introgression among Lake Tanganyika cichlid species. Upper matrix: maximum values of the f_4-ratio statistics between all pairs of species, derived from calculations across all combinations of species trios with *T. sparrmanii* fixed as the outgroup. The f_4-ratio estimates the proportion of the genome affected by gene flow, all presented values are statistically significant (one-sided block-jackknife tests: $P < 5 \times 10^{-5}$ after Benjamini–Hochberg correction for multiple testing). Lower matrix: D_{tree}-statistics (hue) with corresponding P-value (two-tailed binomial test, not adjusted for multiple testing; log-transformed; saturation) based on a phylogenetic approach testing for asymmetry in the relationships of species trios in 1,272 local maximum-likelihood trees (see Supplementary Methods). The two different approaches uncovered little gene flow among the tribes (see Supplementary Discussion).

Corresponding author(s): Fabrizia Ronco
Walter Salzburger

Last updated by author(s): Aug 7, 2020

Reporting Summary

Nature Research wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Research policies, see Authors & Referees and the Editorial Policy Checklist.

Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

n/a	Confirmed
☐	☒ The exact sample size (*n*) for each experimental group/condition, given as a discrete number and unit of measurement
☐	☒ A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
☐	☒ The statistical test(s) used AND whether they are one- or two-sided *Only common tests should be described solely by name; describe more complex techniques in the Methods section.*
☒	☐ A description of all covariates tested
☐	☒ A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
☐	☒ A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
☐	☒ For null hypothesis testing, the test statistic (e.g. *F*, *t*, *r*) with confidence intervals, effect sizes, degrees of freedom and *P* value noted *Give P values as exact values whenever suitable.*
☐	☒ For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
☒	☐ For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
☐	☒ Estimates of effect sizes (e.g. Cohen's *d*, Pearson's *r*), indicating how they were calculated

Our web collection on statistics for biologists contains articles on many of the points above.

Software and code

Policy information about availability of computer code

Data collection	FIJI (v2.0.0-rc-68/1.521i), TINA (v.6.0)
Data analysis	ASTRAL (v.5.6.3), aTRAM (v.2.0.alpha.5), AUGUSTUS (v.3.2.3), BayesTrait (http://www.evolution.rdg.ac.uk/, v.3), bcftools (v.1.6), beagle (v.4.1), BEAST 2 (v.2.5.0; packages: bModelTest (v.1.1.2), SNAPP (v.1.4.2), StarBEAST2 (v.0.15.5), TreeAnnotator (v.2.5.0)), BEDtools (v.2.21.0), BMGE (v.1.1), BUSCO (v.3), BWA-MEM (v.0.7.12), CeleraAssembler (v.8.3), Concaterpillar (v.1.7.2), Dsuite (v.0.2 r20), CT Pro 3D (V5.1.6054.18526), FLASH (v.1.2.11), GATK (v.3.6 and v.3.7), GMAP (GMAP-GSNAP, v.2017-08-15), IQ-TREE (v.1.6.8 and v.1.7-beta7), Kollector (v.1.0.1), MACSE (v.2.01, MAFFT (v.7.300), MIRA (v.4.0.2), MITObim (v.1.8), MultiQC (v.1.7), NRecon (v.1.6.10.2), msprime (v.0.7.4), PAML (v.4.9e and v.4.6), Picard-tools (v.2.7.1), PartitionFinder (v.2.1.1), PAUP* (v.4.0a164, v.4.0a163, and v.4.0a161), Python (v.2.7.10; packages: ete3 (v.3.1.1)), QUAST (v.4.5), R (v.3.5.2 and v.3.6.0; packages: ape (v.5.2), coda (v.0.19-3), Geiger (v.2.0.6.1), Geomorph (v.3.0.7), logisticPCA (v.0.2), phytools (v.0.6-60), vcfR (v.1.8.0), picante (v.1.8), ape (v.5.2), caper (v.1.0.1)), RAxML (v.8.2.4), RepeatModeler (v.1.0.11), RepeatMasker (v.4.0.7), samtools (v.1.3.1), smoove (https://github.com/brentp/smoove, docker image cloned 20/12/2018), lumpy (https://github.com/brentp/smoove, docker image cloned 20/12/2018), svtyper (https://github.com/brentp/smoove, docker image cloned 20/12/2018), svtools (https://github.com/hall-lab/svtools as part of https://github.com/brentp/smoove, docker image cloned 20/12/2018), "snapp_prep.rb" (github.com/mmatschiner/snapp_prep), SNPable (http://lh3lh3.users.sourceforge.net/snpable.shtml), TBLASTN, Tracer (v.1.7.1), Trimmomatic (v.0.36), vcftools (v.0.1.14), Identification of first-generation hybrid samples following github.com/mmatschiner/tutorials/tree/master/analysis_of_introgression_with_snp_data Code used to analyse the data is available on GitHub (https://github.com/cichlidx/ronco_et_al), except for analyses where single commands from publicly available software were used and where all settings are fully reported in the Methods and/or Supplementary Methods sections.

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors/reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Research guidelines for submitting code & software for further information.

Data

Policy information about availability of data

All manuscripts must include a data availability statement. This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A list of figures that have associated raw data
- A description of any restrictions on data availability

All newly sequenced genomes for this study and their raw reads are available from NCBI under the BioProject accession number PRJNA550295 (https://www.ncbi.nlm.nih.gov/bioproject/). The VCF file, tree files, summary statistics of the assembled genomes, and phenotypic datasets generated and analysed during this study are available as downloadable files on Dryad (https://doi.org/10.5061/dryad.9w0vt4bbf). The Nile tilapia reference genome used is available under RefSeq accession GCF_001858045.1. All X-ray data are available on MorphoSource under the project number P1093.

Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

☐ Life sciences ☐ Behavioural & social sciences ☒ Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see nature.com/documents/nr-reporting-summary-flat.pdf

Ecological, evolutionary & environmental sciences study design

All studies must disclose on these points even when the disclosure is negative.

Study description	For the purpose of a comprehensive exploration of the evolution of cichlid fishes in Lake Tanganyika, we collected ten specimens of nearly all cichlid species occurring in that lake; sequenced the genome of one male and one female specimen per species (plus one genome of some outgroups and riverine sister taxa); assessed eco-morphological divergence by quantifying body shape (10 per species), oral jaw morphology (10 per species), lower pharyngeal jaw shape (5 per species) and stable carbon and nitrogen isotope compositions (10 per species); and quantified divergence in pigmentation patterns (5 per species).
Research sample	Our set of samples consists of ten specimens of nearly all cichlid fish species occurring in Lake Tanganyika, a set of selected outgroup species and a set of riverine species nested within the radiation. This sample was selected to maximally represent the cichlid fauna in the Lake Tanganyika drainage and the phylogenetic spectrum of East African cichlids. A comprehensive list of taxa (n=297) and specimens (n= 2'723; typically 5 males and 5 females per species) including information on the sex of the specimens is provided as Supplementary Tables 1 and 2. The ages are unknown for all specimens, but all specimens were adults. No manipulations were performed.
Sampling strategy	We collected specimens of cichlid fishes at African Lake Tanganyika that were either caught with barrier nets while snorkeling or Scuba diving, or purchased from local fishermen. After euthanasia with clove oil, we measured, weighted and photographed each specimen and took a fin clip for later DNA extraction. Specimen were formalin fixed and in a standardized way. Sampling was performed under research permits issued by the relevant authorities in the Republic of Burundi, the United Republic of Tanzania, and the Republic of Zambia. To maximize taxon sampling we included additional specimens from previous expeditions (4.9% of the samples) as well as from other collections (0.8%). The final dataset (297 taxa; n = 2'723 specimens) contained an almost complete taxon sampling of the cichlid fauna of Lake Tanganyika, 18 non-Tanganyikan cichlids nested within the radiation, and 28 outgroup species (see Supplementary Tables 1 and 2 for details). No sample size calculations were performed a priori. We sampled 10 adult specimens per species, which is sufficient to quantify eco-morphological disparity and estimate representative species means for comparative analyses. For genome sequencing we selected, whenever available, one male and one female individual per species to have both sexes represented.
Data collection	Digitalisation of Landmarks for body shape and upper oral jaw morphology: Data recorded by Fabrizia Ronco using the Software FIJI (v2.0.0-rc-68/1.521i) based on X-ray Images of the specimens. Digitalisation of Landmarks for lower pharyngeal jaw morphology: Data recorded by Fabrizia Ronco using the Software TINA (v.6.0) based on CT-scans of the specimens. Scoring pigmentation pattern: Data recorded by Walter Salzburger, scored by eye based on photographs of the specimens. Genome sequencing: DNA extraction and library preparation was performed by Fabrizia Ronco and Astrid Böhne. Sequencing was performed at the Norwegian Sequencing Centre (NSC), Oslo, and the Genomics Facility Basel (GFB) at the ETH Zurich Department of Biosystems Science and Engineering (D-BSSE), Basel, on Illumina HiSeq 2500 devices. Stable carbon (C) and nitrogen (N) isotope composition: Sample preparation was performed by Anna Boila with the help of Fabrizia Ronco. Samples were analysed by Anna Boila and Ansgar Kahmen on a Flash 2000 elemental analyser coupled to a Delta Plus XP continuous-flow isotope ratio mass spectrometer (IRMS) via a Conflo IV interface (Thermo Fisher Scientific, Bremen, Germany).
Timing and spatial scale	Sampling was conducted between 2014 and 2017 at 130 locations around Lake Tanganyika, followed by sample processing in Basel and Oslo, which required the following work packages and durations: DNA extraction and genome sequencing: April 2014 – February 2017 (Basel and Oslo).

Digitalisation of landmarks for body shape and upper oral jaw morphology: January 2018 – March 2018 (Basel).
Digitalisation of landmarks for lower pharyngeal jaw morphology: June 2016 – November 2017 (Basel).
Scoring pigmentation pattern: September 2019 (Basel).
Stable carbon (C) and nitrogen (N) isotope composition: March 2016 – October 2017 (Basel).

Data exclusions

Based on preestablished exclusion criteria for morphological analyses, specimens with broken jaws or bended bodies were excluded.
We excluded one of the sequenced genomes based on signs of contamination or DNA degradation.

Reproducibility

Due to the rather long time period for the digitalisation of landmarks for lower pharyngeal jaw morphology, we repeated (at the end of the data collection period) the data collection for the first 100 specimens scored. Biological interpretation remained unchanged.
No other data collection process was repeated, replicated or performed independently. All data collection steps which are potentially influenced by the observer (landmark digitalization, pigmentation scoring) were performed by a single person (experienced in the task) each to avoid investigator bias.

Randomization

Library pooling for Illumina sequencing was not specifically randomised, however, samples were allocated to pools based on suitable adapter combinations (according to the Illumina pooling guidelines).
Sample allocation into experimental groups is not relevant to this study, as no experimental groups were used.

Blinding

The possibility of blinding of the specimens was very limited. Although we labeled images and CT-scans only with specimen voucher IDs, species identification based on the image itself cannot be ruled out.
For all other data collection steps (data sets: stable isotope analyses and genome sequencing) blinding was not relevant as an investigator biased can be ruled out.
No blinding was applied for data analyses as taxonomic information was relevant for the analyses.

Did the study involve field work? ☒ Yes ☐ No

Field work, collection and transport

Field conditions

No field conditions are relevant to this study because we were exclusively interested in the biological specimens.

Location

Specimens were collected at Lake Tanganyika between 2014 and 2017 at 130 locations in the Republic of Burundi, the United Republic of Tanzania and the Republic of Zambia. GPS coordinates of the sampling location for each specimen are provided as downloadable file on dryad; https://doi.org/10.5061/dryad.9w0vt4bbf.

Access and import/export

All samples were collected and exported in agreement with local authorities with the following permits issued:

Republic of Burundi:

Sampling Permit, issued by the Ministère de l'Eau, de l'Environnement, de l'Amenagement du Territoire et de l'Urbanisme, Republic of Burundi
770 06/62710, issued 27/12/2014

Research permit issued by the Universite du Burundi (Cabinet du Recteur and Directeur de la Recherche et de l'Innovation)
2014/R991/Invitation (Heinz Büscher, Adrian Indermaur, Fabrizia Ronco, Walter Salzburger), issued 17/12/2014
Order de mission 35/2015 (Heinz Büscher, Adrian Indermaur, Fabrizia Ronco, Walter Salzburger), issued 19/01/2015

Work permit (Mission de travail), issued by the Permanent Mission of the Republic of Burundi to the United Nations, Geneva:
544/GE/2014/N.M.A (Heinz Büscher), valid 29/12/2014 to 28/01/2015
545/GE/2014/N.M.A (Fabrizia Ronco), valid 29/12/2014 to 28/01/2015
546/GE/2014/N.M.A (Adrian Indermaur), valid 29/12/2014 to 28/01/2015
547/GE/2014/N.M.A (Walter Salzburger), valid 29/12/2014 to 28/01/2015

Export permits, issued by the Universite du Burundi (Cabinet du Recteur and Directeur de la Recherche et de l'Innovation) and the Ministère de l'Eau, de l'Environnement, de l'Amenagement du Territoire et de l'Urbanisme:
Export/transport permit, issued 21/01/2105

The United Republic of Tanzania:

Research permits, issued by the Tanzania Commission for Science and Technology (COSTECH):
2015-173-NA-2015-96 (Adrian Indermaur), valid 29/05/2015 to 18/05/2016
2015-174-NA-2015-96 (Fabrizia Ronco), valid 29/05/2015 to 18/05/2016
2015-175-NA-2015-96 (Heinz Büscher), valid 29/05/2015 to 18/05/2016
2015-176-NA-2015-96 (Walter Salzburger), valid 29/05/2015 to 18/05/2016
2016-373-NA-2015-96 (Walter Salzburger), valid 12/12/2016 to 11/12/2017
2016-376-NA-2015-96 (Fabrizia Ronco), valid 12/12/2016 to 11/12/2017
2016-377-NA-2015-96 (Adrian Indermaur), valid 12/12/2016 to 11/12/2017
2016-378-NA-2015-96 (Heinz Büscher), valid 12/12/2016 to 11/12/2017

Research permits, issued by the Tanzania National Parks Authority (TANAPA):
TNP/HQ/C.10/13/2015 (Heinz Büscher, Adrian Indermaur, Fabrizia Ronco, Walter Salzburger), valid 30/6/15 to 29/09/16
TNP/HQ/C.10/13/2017 (Heinz Büscher, Adrian Indermaur, Fabrizia Ronco, Walter Salzburger), valid 12/12/16 to 11/12/17

Research Clearance, issued by the Tanzania Wildlife Research Institute (TAWIRI):
13300 (Heinz Büscher, Adrian Indermaur, Fabrizia Ronco, Walter Salzburger), dated 09/01/2017

Residence permits, issued by the Department of Immigration:
CTA0329015 (Heinz Büscher), valid 22/06/2015 to 21/08/2016
CTA0329016 (Walter Salzburger), valid 22/06/2015 to 21/08/2016
CTA0329017 (Adrian Indermaur), valid 22/06/2015 to 21/08/2016
CTA0329018 (Fabrizia Ronco), valid 22/06/2015 to 21/08/2016
RPC11100834 (Walter Salzburger), valid 11/12/2016 to 10/12/2017
RPC11100835 (Fabrizia Ronco), valid 11/12/2016 to 10/12/2017
RPC11100836 (Heinz Büscher), valid 11/12/2016 to 10/12/2017
RPC11100836 (Adrian Indermaur), valid 11/12/2016 to 10/12/2017

Sample export and transport permits, issued by the Tanzanian Fisheries Research Institute (TAFIRI), Ministry of Livestock and Fisheries Development:
TAF/KGM/R/VOL.V/236, issued 16/07/2015
TAF/KGM/R.1/VOL.V/121, issued 10/02/2017

Republic of Zambia:

Study permits (including residence permits), issued by the Department of Immigration and the Department of Fisheries, Ministry of Agriculture and Livestock, based on a Memorandum of Understanding (MOU)
SP000627 (Fabrizia Ronco), valid 13/07/2012 to 08/08/2016
SP000710 (Adrian Indermaur), valid 13/07/2012 to 30/10/2015
SP001995 (Walter Salzburger), valid 05/07/2013 to 05/07/2015
SP002417 (Heinz Büscher), valid 05/08/2015 to 12/11/16
SP004273 (Walter Salzburger), valid 30/07/2015 to 13/07/2020
SP005937 (Fabrizia Ronco), valid 29/07/2016 to 28/07/2018
SP005943 (Adrian Indermaur), valid 27/07/2016 to 28/07/2018

Export permits, issued by the Department of Fisheries, Ministry of Agriculture and Livestock:
Export/transport permit, issued 02/08/2013
Export/transport permit, issued 23/01/2014
Export/transport permit, issued 26/08/2015
Export/transport permit, issued 13/09/2016
Export/transport permit, issued 29/08/2017
Export/transport permit, issued 10/09/2018

Schweizerische Eidgenossenschaft/Confoederatio Helvetica (CH):

CITES Approval, issued by the Bundesamt für Veterinärwesen, Eidgenössisches Departement für Inneres:
CH018 (Adrian Indermaur, Walter Salzburger, Zoological Institute, University of Basel), valid 23/01/2013 to 31/12/2020

Recognition as Scientific Institution (according to EU-directive 92/65/EWG, Annex C), issued by the Cantonal Veterinary Office Basel Stadt:
CH-I-BS017 (Walter Salzburger), valid 11/06/2012 to 31/12/2017
CH-I-BS003h (Walter Salzburger), valid 19/02/2015 to 31/12/2019

Permit for an animal facility for cichlid fishes, issued by the Cantonal Veterinary Office Basel Stadt:
1010H (Walter Salzburger), valid 01/11/2013 to 31/10/2023

Permit to conduct and supervise animal experiments, issued by the Cantonal Veterinary Office Basel Stadt:
A2015 (Walter Salzburger), issued 19/01/2010

Permit to take tissue samples from cichlid fishes, issued by the Cantonal Veterinary Office Basel Stadt:
2317_22449 (Walter Salzburger), valid 01/12/2011 to 31/12/2014
2317_25931 (Walter Salzburger), valid 01/01/2015 to 01/01/2018
2317_29387 (Walter Salzburger), valid 02/01/2018 to 31/12/2020

Disturbance

We collected specimens primarily during snorkelling and scuba diving which allows to target individual specimens with minimum bycatch.

Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

Materials & experimental systems

n/a	Involved in the study
☒	☐ Antibodies
☒	☐ Eukaryotic cell lines
☒	☐ Palaeontology
☐	☒ Animals and other organisms
☒	☐ Human research participants
☒	☐ Clinical data

Methods

n/a	Involved in the study
☒	☐ ChIP-seq
☒	☐ Flow cytometry
☒	☐ MRI-based neuroimaging

Animals and other organisms

Policy information about studies involving animals; ARRIVE guidelines recommended for reporting animal research

Laboratory animals	This study did not involved laboratory animals.
Wild animals	We collected specimens of cichlid fishes at African Lake Tanganyika that were either caught with barrier nets while snorkelling or Scuba diving, or purchased from local fishermen. After euthanasia with clove oil, we measured, weighted and photographed each specimen and took a fin clip for later DNA extraction. Specimen were formalin fixed and in a standardized way. Sampling was performed under research permits issued by the relevant authorities in the Republic of Burundi, the United Republic of Tanzania, and the Republic of Zambia. A comprehensive list of taxa (n=297) and specimens (n= 2'723; typically 5 males and 5 females per species) including information on the sex of the specimens is provided as Supplementary Tables 1 and 2. No animals were transported or kept alive.
Field-collected samples	No experiments were conducted in the field as only terminal samples were collected (see above).
Ethics oversight	Republic of Burundi: Ministère de l'Eau, de l'Environnement, de l'Amenagement du Territoire et de l'Urbanisme, Universite du Burundi (Cabinet du Recteur and Directeur de la Recherche et de l'Innovation) The United Republic of Tanzania: Tanzania Commission for Science and Technology (COSTECH): Tanzanian Fisheries Research Institute (TAFIRI), Ministry of Livestock and Fisheries Development: Republic of Zambia: Department of Fisheries, Ministry of Agriculture and Livestock Schweizerische Eidgenossenschaft/Confoederatio Helvetica (CH): Cantonal Veterinary Office Basel Stadt:

Note that full information on the approval of the study protocol must also be provided in the manuscript.