



Gene expression dynamics during rapid organismal diversification in African cichlid fishes

Athimed El Taher¹✉, Astrid Böhne^{1,2}, Nicolas Boileau¹, Fabrizia Ronco¹, Adrian Indermaur¹, Lukas Widmer¹ and Walter Salzburger¹✉

Changes in gene expression play a fundamental role in phenotypic evolution. Transcriptome evolutionary dynamics have so far mainly been compared among distantly related species and remain largely unexplored during rapid organismal diversification, in which gene regulatory changes have been suggested as particularly effective drivers of phenotypic divergence. Here we studied gene expression evolution in a model system of adaptive radiation, the cichlid fishes of African Lake Tanganyika. By comparing gene expression profiles of 6 different organs in 74 cichlid species representing all subclades of this radiation, we demonstrate that the rate of gene expression evolution varies among organs, transcriptome parts and the subclades of the radiation, indicating different strengths of selection. We found that the noncoding part of the transcriptome evolved more rapidly than the coding part, and that the gonadal transcriptomes evolved more rapidly than the somatic ones, with the exception of liver. We further show that the rate of gene expression change was not constant over the course of the radiation but accelerated at its later phase. Finally, we show that—at the per-gene level—the evolution of expression patterns is dominated by stabilizing selection.

During the second half of the last century, it has become increasingly clear that changes in gene expression play a fundamental role in phenotypic evolution^{1–8}. Previous large-scale studies of gene expression evolution comparing distantly related vertebrate species revealed substantial variation in the rate of transcriptome evolution among organs and evolutionary subclades^{4,9}, between the coding and noncoding part of the transcriptome³, and across developmental time points¹⁰. However, little is known about the dynamics of gene expression evolution during adaptive radiations, which are characterized by the unusually rapid ecological and morphological diversification of an organismal lineage into distinct ecological niches^{11,12}. Yet, precisely for such outbursts of organismal diversity, gene regulatory changes have been proposed as a key mechanism promoting rapid phenotypic divergence^{12–14}, making the study of transcriptome evolution of particular interest in the context of adaptive radiations¹⁵.

Here we examined the dynamics of gene expression evolution in one of the most striking examples of adaptive radiation, the cichlid fishes of African Lake Tanganyika^{15–17}. This species flock comprises about 240 endemic cichlid species that evolved in less than 10 Myr and shows an extraordinary degree of eco-morphological divergence^{16–18}. We sequenced transcriptomes of five organs (brain, gill, liver, ovary and testis) in 3 males and 3 females of 74 cichlid species, representing all phylogenetic subclades—so-called ‘tribes’—and all major ecological guilds of the cichlid fauna of Lake Tanganyika¹⁶ (Fig. 1a and Supplementary Tables 1 and 2). In addition, we sequenced the transcriptomes of the lower pharyngeal jaw bone (LPJ) in the same set of 445 specimens. The LPJ is the central component of the cichlids’ pharyngeal jaw apparatus (that is, a second set of functional jaws in the pharynx used to masticate food¹⁹), and hypothesized to be a key innovation triggering cichlid adaptive radiations^{17,20–23}. These target organs were selected because of their involvement in ecological, physiological and behavioural adaptations during

cichlid adaptive radiations^{15,17,19,24,25} and to enable comparisons to previous studies^{3,10,26}.

Results

Patterns of gene expression. To study gene expression evolution during rapid organismal diversification, we generated a total of 2,131 transcriptome profiles (equivalent to individual RNA-sequencing libraries) from typically 5 organs in 6 adult specimens of 74 species of cichlid fishes from African Lake Tanganyika (median sequencing depth per tissue: 9.6–9.9 million reads per library; 125-base-pair (bp) strand-specific single-end reads; mapped against the phylogenetically equidistant *Oreochromis niloticus* reference genome; median of read mapping ~76%). A time-calibrated species tree based on genome-wide data taken from ref. 17 and pruned to the taxon set of this study is shown in Fig. 1a (details on individual samples including sampling dates and locations are available in Supplementary Table 1; information on sequencing and mapping coverage is provided in Supplementary Fig. 1; and the variance within and between species is shown in Supplementary Figs. 2–4).

As a first step, to explore the global patterns of gene expression differentiation among species and across organs, we performed a principal component analysis (PCA) on the entire dataset. The PCA clearly separated the expression profiles according to organ type (Fig. 1b)—with the exception of gill and LPJ transcriptomes, which showed largely overlapping gene expression profiles. This similarity in gene expression profiles is not surprising, given their common developmental origin (the LPJ is derived from the fusion of the left and right fifth ceratobranchials¹⁹). Within organs, on the other hand, the species-specific transcriptome profiles clustered by tribe (Fig. 2), indicating a strong phylogenetic signal in the data. This in turn suggests that, as in mammals¹, gene expression changes have accumulated over the course of the adaptive radiation of cichlid fishes in Lake Tanganyika, resulting in—overall—more similar gene expression profiles between more closely related species, irrespective

¹Zoological Institute, Department of Environmental Sciences, University of Basel, Basel, Switzerland. ²Present address: Center for Molecular Biodiversity Research, Zoological Research Museum Alexander Koenig, Bonn, Germany. ✉e-mail: athimed.eltaher@unibas.ch; walter.salzburger@unibas.ch

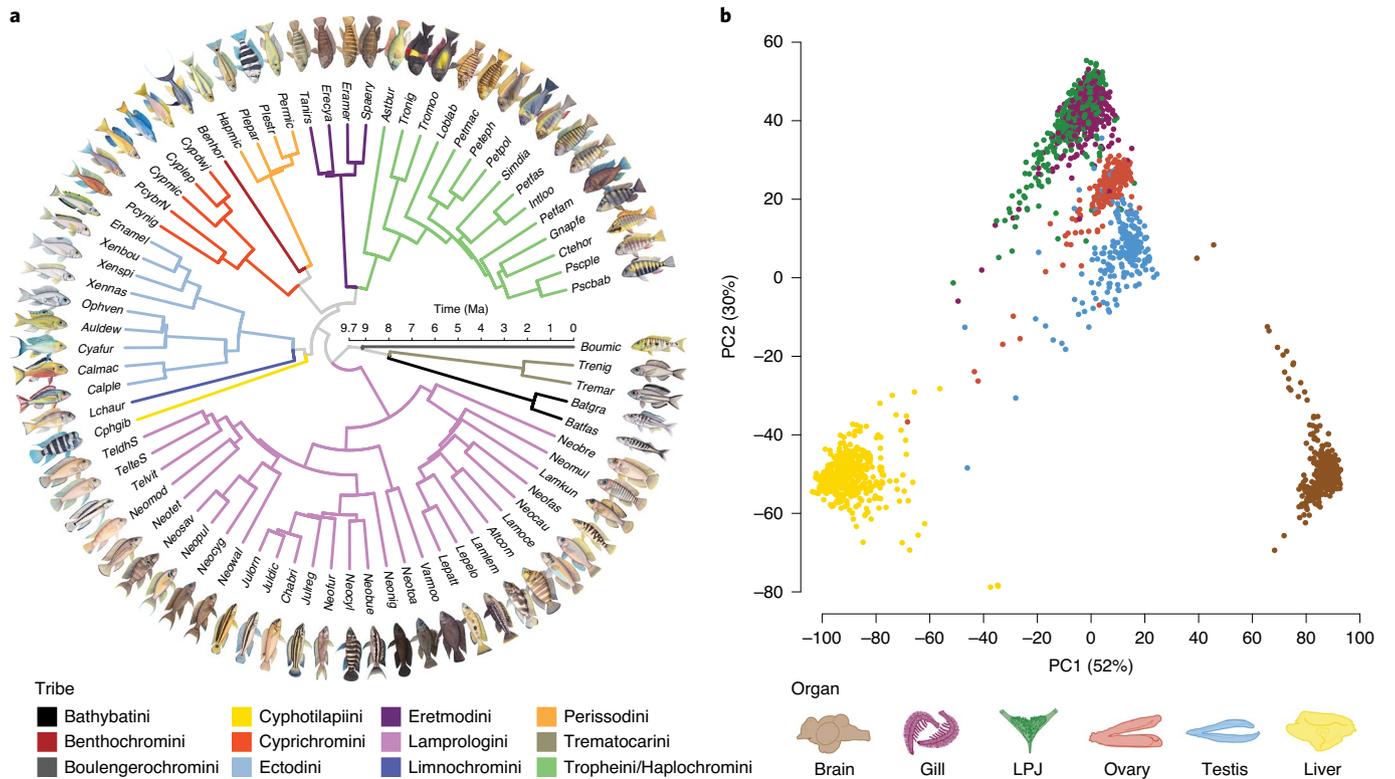


Fig. 1 | Gene expression patterns across the adaptive radiation of cichlid fishes in African Lake Tanganyika. a, A time-calibrated species tree of the adaptive radiation of cichlid fishes in African Lake Tanganyika based on genome-wide data¹⁷, pruned to the 74 taxa used in this study. The species names are abbreviated using a six-letter code, whereby the first three letters represent the genus and the last three letters represent the species name (see Extended Data Fig. 9 and Supplementary Table 2 for the full species names). The branches are colour-coded according to phylogenetic subclades (that is, tribes) as indicated in the lower panel. Ma, million years ago. **b**, PCA of overall gene expression levels. The samples ($n = 2,131$) are coloured according to organ type as indicated in the lower panel. The proportions of variance explained by the first two principal components (PC1 and PC2) are indicated in parenthesis on the x and y axes, respectively.

of their adaptations to particular ecological niches. When making comparisons between the two largest classes of RNAs in the cichlids' transcriptomes—that is, between protein-coding and long noncoding (lnc) RNAs—we found that the correlations in gene expression levels among the 74 cichlid species were significantly higher for protein-coding genes (Fig. 3a) than for lncRNAs (Fig. 3b) (pairwise Spearman's rank correlation coefficients (Spearman's ρ) between gene expression levels measured as transcripts per million (TPM); two-sided t -test: $P < 10^{-8}$; Fig. 3c). Moreover, the hierarchical clustering of gene expression levels revealed a much more pronounced grouping according to organ type in protein-coding genes compared to lncRNAs (Fig. 3a,b). This suggests that, also during the short evolutionary timescale spanned by the cichlid adaptive radiation in Lake Tanganyika (< 10 Myr; ref.¹⁷), the noncoding part of the transcriptome has undergone more rapid turnovers in expression trajectories than the coding part. The observed accelerated evolution of lncRNAs can be interpreted as a sign of more relaxed selection regimes on lncRNAs compared to protein-coding genes³.

When examining the transcriptome profiles of the somatic organs for sex-specific differences, we found a pronounced difference between females and males in liver but not in the other organs (Extended Data Fig. 1). Sex-biased gene expression in liver has been reported before⁴, albeit with a smaller magnitude than observed here. These sex differences in liver gene expression can mainly be attributed to its function as the main metabolic organ secreting hormones and maintaining homeostasis, and being responsive to sex steroids^{27–32} (for an in-depth analysis of sex-specific differences in

Tanganyikan cichlid fishes as well as of sex chromosome evolution, see ref.³³).

Transcriptome evolution. To compare the rate of evolution of the coding and noncoding part of the transcriptome in the different organs, we correlated Spearman's ρ of gene expression levels with divergence times (taken from ref.¹⁷) for all pairs of cichlid species examined (Fig. 4a and Extended Data Fig. 2a). We found that, similar to the pattern observed in tetrapods featuring much deeper phylogenetic splits⁴, the rate of transcriptome evolution (measured as $[1 - \rho]/\text{divergence time}$) differed significantly among organs (Fig. 4b; analysis of variance (ANOVA): $P < 10^{-8}$ for protein-coding genes and lncRNAs), as well as between the transcriptome parts in cichlids (Extended Data Fig. 2; two-sided t -test: $P < 10^{-8}$). The expression levels of protein-coding genes evolved significantly slower in brain, LPJ and gill, than in testis, liver and ovary (left panel in Fig. 4b; ANOVA: $P < 10^{-8}$; Supplementary Table 3a), indicating that—taken as a whole—the organ-specific transcriptomes evolved under different selection regimes during the course of the Tanganyikan cichlid radiation.

The correlations between Spearman's ρ and divergence times were weaker in lncRNAs than in protein-coding genes (Extended Data Fig. 2b; two-sided t -test: $P < 10^{-8}$), suggesting more relaxed selective constraints in lncRNAs. The lncRNA expression levels of the gonads evolved faster compared to the other organs, followed by liver and then the remaining three organs (brain, gill and LPJ), which showed similar

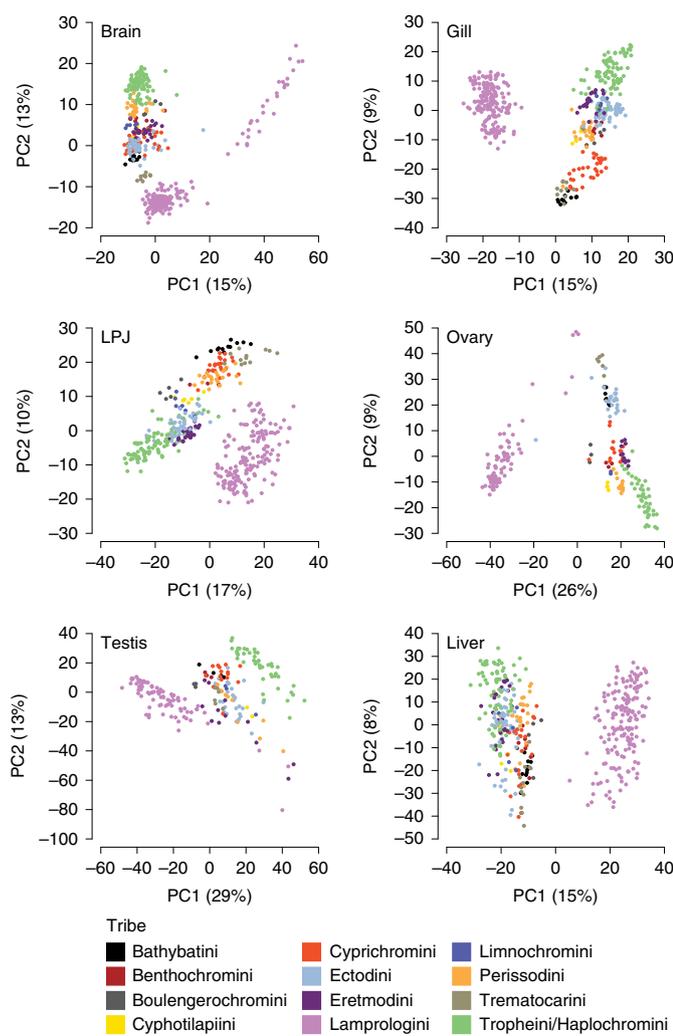


Fig. 2 | PCA of overall gene expression levels per organ. The proportion of variance explained by the first two principal components (PC1 and PC2), indicated in parenthesis on the x and y axes, respectively, for brain, gill, LPJ, ovary, testis and liver. Samples (brain: $n = 428$; gill: $n = 434$; LPJ: $n = 425$; ovary: $n = 219$; testis: $n = 213$; liver: $n = 412$) are coloured according to tribe as shown in the lower panel.

patterns of expression evolution (right panel in Fig. 4b; ANOVA: $P < 10^{-8}$, Supplementary Table 3b). Organ-specific evolutionary rates thus emerge as a transcriptome-wide trend in cichlids, with protein-coding genes being subject to stronger conservation—indicating stronger purifying selection and/or weaker positive selection—in gene expression levels compared to lncRNAs³.

As a complementary approach to the comparison of Spearman's ρ , whole-transcriptome evolutionary trajectories can also be explored with gene expression trees based on pairwise gene expression distances—a strategy less sensitive to outliers and potential inaccuracies in the normalization process^{4,34}. The expression trees based on protein-coding and lncRNA transcriptome profiles of all six organ samples recovered, in most cases, the major phylogenetic clustering of the species into tribes (protein-coding genes: Extended Data Fig. 3; lncRNAs: Extended Data Fig. 4). This once more illustrates that most changes at the level of whole transcriptomes have accumulated over evolutionary time, such that closely related species show more similar gene expression profiles. However, the relative positions of

the tribes to one another were not congruent among the expression trees and differed in all cases from the phylogenetic relationships derived from genome-wide data¹⁷ (Fig. 1a).

The most obvious and consistent difference between the expression trees and the species tree concerns the position of the most species-rich tribe of cichlid fishes in Lake Tanganyika, the Lamprologini, which were placed as a sister clade to all remaining tribes in the expression trees (Extended Data Figs. 3 and 4) but are clearly nested within the cichlid radiation according to the species tree¹⁷ (Fig. 1a). That the Lamprologini show characteristic global transcriptome profiles in all organs that are different from those of all other tribes already became apparent in the organ-specific PCAs (Fig. 2). This pattern is not an artefact of the use of a greater number of Lamprologini species in our analyses (reflecting the de facto greater number of Lamprologini species¹⁶), since repeated reanalyses of the data with a balanced number of representatives per tribe yielded similar results (Supplementary Fig. 5).

Next, to reconstruct gene expression changes along the phylogeny, we projected expression branch lengths on a time-calibrated species tree¹⁷ using the Fitch and Margoliash method³⁵ (protein-coding genes: Extended Data Fig. 5; lncRNAs: Extended Data Fig. 6). The branch lengths of these expression trees correlated positively with the branch lengths of the time-calibrated species tree in all organs and in both transcriptome parts (linear model: $P < 10^{-8}$; Extended Data Fig. 7). We then estimated the rate of expression change for each branch in the species tree (measured as the expression trees' branch lengths divided by the species tree's branch lengths; Extended Data Fig. 7) and quantified rate changes through time (mean rates sampled in steps of 0.15 Myr along the phylogeny as in ref. 17). This revealed that transcriptome evolution was not constant over the course of the cichlid adaptive radiation in Lake Tanganyika, but became accelerated in the late phase of the radiation, coinciding with a high number of speciation events¹⁷ (Fig. 4c and Extended Data Fig. 7). This effect was more pronounced in lncRNAs compared to protein-coding genes (Fig. 4c).

Using the cumulative branch lengths (from root to tip) of the expression trees as a proxy for the rate of transcriptome evolution, irrespective of the temporal signals reported above, we corroborated that gene expression levels evolved differently among organs and between the transcriptome parts (Fig. 4d and Supplementary Table 7). As for the results based on Spearman's ρ (Fig. 4a,b), the organ-specific topology-fixed expression trees of protein-coding genes (Extended Data Fig. 5) showed differences in branch lengths, with liver and testis evolving fastest (Fig. 4d; ANOVA: $P < 10^{-8}$; Supplementary Table 4a). In the lncRNA expression trees (Extended Data Fig. 6), the cumulative branch lengths were similar for brain, gill and LPJ, but significantly longer for ovary, testis and liver (Fig. 4d; ANOVA: $P < 10^{-9}$, Supplementary Table 4b). By calculating Robinson–Foulds distances³⁶ between the trees obtained from the expression data and the time-calibrated species tree¹⁷ (Fig. 4d), we show that the majority of changes in gene expression followed the species tree. However, there is substantial variation among the organs. For example, in agreement with Warnefors and Kaessmann³⁷, brain experienced fewer changes in the overall levels of gene expression than testis.

Finally, by comparing the rates of expression change (represented by the cumulative branch length) among the tribes, we found substantial differences between the radiation's subclades in all organs (ANOVA: $P < 10^{-8}$, Extended Data Fig. 8 and Supplementary Table 5). For example, Trematocarini showed a high rate of expression changes compared to the other tribes in brain, gill, LPJ and ovary, Cyprichromini had a high rate of expression changes in gill and LPJ, while Eretmodini featured a high rate of expression changes in testis. The most species-rich tribe of cichlids in Lake Tanganyika, the Lamprologini, showed intermediate rates of expression change in most organs. Overall, the observed differences in the rate of

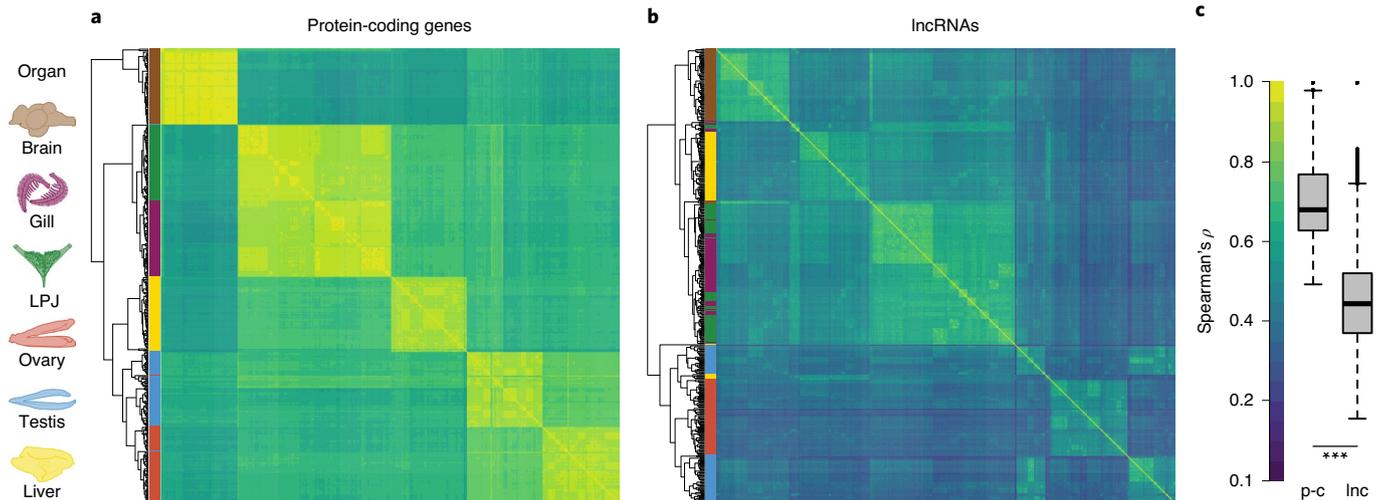


Fig. 3 | Gene expression similarities among species and transcriptome parts. **a, b**, Hierarchical clustering of expression levels estimated as transcripts per million (TPM) for protein-coding genes ($n=27,105$) (**a**) and lncRNAs ($n=4,719$) (**b**). The heatmaps represent Spearman's rank correlation coefficients between pairs of species ($n=442$ samples). Sample clustering is represented as a tree on the left side of each heatmap, with organs colour-coded as indicated in the organ illustrations on the left side. **c**, Spearman's rank correlation coefficients between pairs of species for protein-coding genes (p-c) and lncRNAs (lnc). The box plot centre lines represent the median, the box limits represent the upper and lower quartiles, and the whiskers represent 1.5x the interquartile range; two-sided t -test: $***P < 10^{-8}$.

expression changes among the subclades for the different organs might reflect the organ's lineage-specific involvement in ecological, physiological and behavioural adaptations.

Organ-specific expression patterns. As gene expression evolution might be constrained by core organ functions^{2–4,38}, we determined the degree of organ specificity in gene expression across the different organs using the organ specificity index τ^{39} . The number of organ-specific genes varied substantially, also with respect to the two investigated transcriptome parts (Fig. 5a). In testis, which exhibited the fastest evolving transcriptome at the coding level (Fig. 4), we also found most organ-specific genes (Fig. 5a,b), closely followed by brain (only for protein-coding genes, Fig. 5a), which showed the slowest evolving transcriptome (Fig. 4). We also found substantial differences in the number of organ-specific genes among the different tribes (for protein-coding genes and lncRNAs; Fig. 5b).

Gene expression dynamics. To examine, in more detail, the dynamics of gene expression evolution on the per-transcript level, we tested, for each organ and for each expressed gene separately, the model fit to three common models of trait evolution along the time-calibrated species tree. More specifically, we asked whether the gene expression levels (TPM) of a particular protein-coding gene or lncRNA are more likely to have evolved under a Brownian motion (BM), a single-optimum Ornstein–Uhlenbeck (OU) or an early burst (EB) model of trait evolution. We found that, for the majority of protein-coding genes (64–88%, depending on the organ; Fig. 5c and Supplementary Table 6a) and lncRNAs (79–88%, depending on the organ; Fig. 5c and Supplementary Table 6b), the gene expression levels evolved according to the OU model of trait evolution, suggesting that (stabilizing) selection has shaped the expression patterns of these genes (this is similar to mammals)⁴⁰. The expression levels of 9–30% of the protein-coding genes and 7–15% of the lncRNAs (depending on the organ) are most compatible with a BM model of trait evolution, suggesting that these transcripts have evolved more or less neutrally. The smallest fraction of transcripts (2–5% for protein-coding genes and 4–6% for lncRNAs) showed expression patterns that fit the EB model of trait evolution, suggesting rapid divergence in gene expression near the

onset of the radiation. When making comparisons between organs and transcriptome parts, we found that the ovary transcriptomes contained a comparatively larger fraction of protein-coding genes with BM-like expression dynamics than other organs along the phylogeny, whereas the transcriptomes of testis and liver featured the largest fractions of protein-coding genes compatible with the OU model (Fig. 5c). For lncRNAs, the liver transcriptomes contained the largest proportion of genes with OU-like expression dynamics (Fig. 5c and Supplementary Table 6).

Discussion

Through the inspection of 2,131 transcriptome profiles from a set of 74 closely related species representative of the adaptive radiation of cichlid fishes in African Lake Tanganyika, we show that the rate of gene expression evolution varies among organs and among the subclades of the radiation, and also between protein-coding genes and lncRNAs. Using several different approaches, we demonstrate that the transcriptomes of brain, gill and LPJ evolve significantly slower than gonadal and liver transcriptomes. This holds true for protein-coding genes as well as for lncRNAs, suggesting that this pattern represents a transcriptome-wide trend in Tanganyikan cichlids.

Our results on gene expression dynamics over the course of the cichlid adaptive radiation in Lake Tanganyika are only partially consistent with previous work on transcriptome evolution at much deeper phylogenetic levels. As in earlier studies^{3,4,26,41,42}, we found that the rate of gene expression evolution (for protein-coding genes and for lncRNAs) was slowest in the brain. The comparatively slow rates of transcriptome evolution in the brain have previously been attributed to the greater degree of specialization in neuronal organs^{3,4,43}. It thus appears plausible that, also in cichlids, organ complexity may explain the differences in transcriptome evolution among brain, gill and LPJ on one side, and the gonadal organs and liver on the other side.

The consistently fastest rates of gene expression evolution as well as the largest number of organ-specific transcripts have so far been reported for testis (in protein-coding genes and lncRNAs)^{3,4,26,42}, and it has been suggested that this is due to sex-related selective forces³ including sperm competition⁴⁴, as well as to the particular

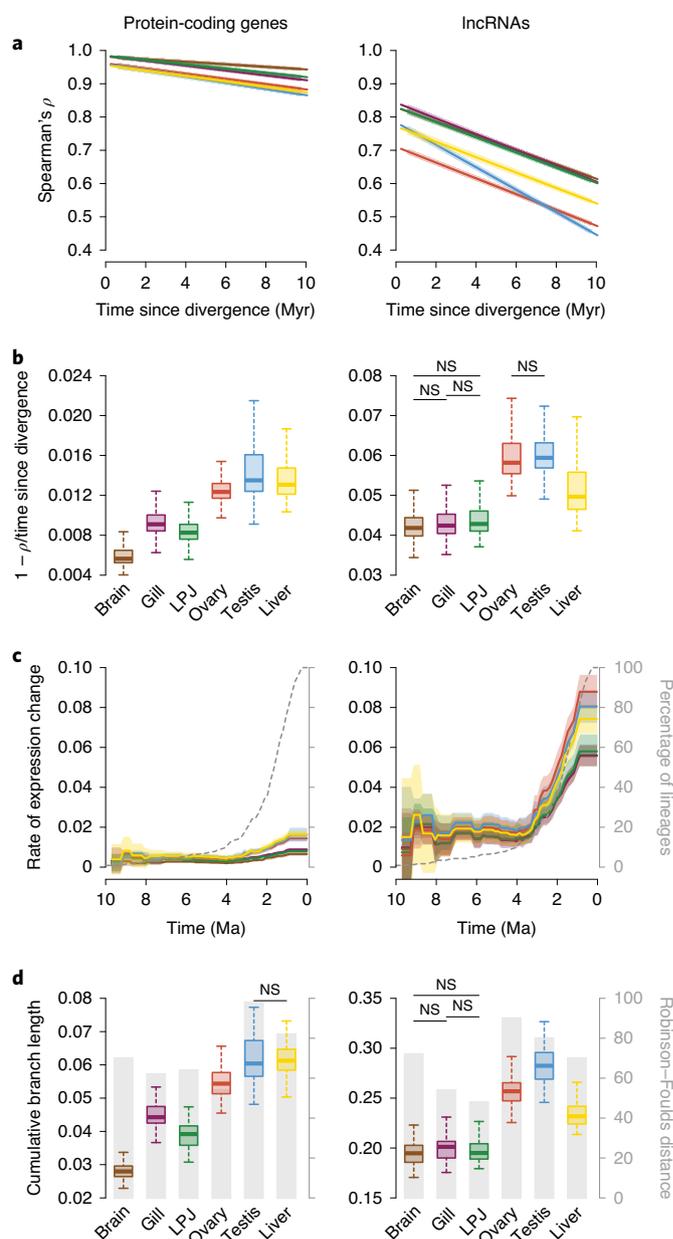
permissive chromatin conformation during spermatogenesis⁴³, leading to greater transcriptional activity and reduced transcriptional constraints, potentially facilitating transcriptional noise^{3,43,45}. We corroborate here that also during rapid adaptive radiation of cichlid fishes in Lake Tanganyika, testis features the single most rapidly evolving transcriptome at the level of protein-coding genes (Fig. 4) and contains the largest number of organ-specific genes (Fig. 5). At the level of lncRNAs, however, the transcriptomes of both gonadal organs, testis and ovary, appear to have evolved equally rapidly (Fig. 4). This argues against transcriptional noise as an explanation for the high rates of gene expression evolution in testis, but rather attests overall high rates of gene expression evolution in gonads in cichlids.

The reconstruction of the mean rate of gene expression change along the time-calibrated species tree (Fig. 4c) revealed that transcriptome evolution was not constant over the course of the radiation, but was accelerated—in all organs and transcriptome parts—in the radiation's late phase. This pattern was particularly evident in lncRNAs (Fig. 4c). We have recently shown that the later phase of the cichlid adaptive radiation in Lake Tanganyika is characterized by an increase in the number of speciation events as well as accelerated phenotypic evolution in the ecologically relevant LPJ and in a signalling trait (body pigmentation)¹⁷. The increase in the rate of gene expression change in this later phase of the radiation is in line with the putative role of gene expression evolution during taxonomic and phenotypic diversification.

A main difference in terms of the rate of transcriptome evolution between our study and previous work on mammals (or tetrapods) concerns the liver. While previous studies reported moderate levels of gene expression evolution in liver^{4,26,42}, we found that the rate of transcriptome evolution in this organ (at the level of protein-coding genes) is nearly as fast as the one observed in testis (Fig. 4). Since some of the most important functions of the liver are connected to the digestive system, it is possible that the accelerated rate of transcriptome evolution in this organ in Tanganyikan cichlids reflects rapid dietary adaptations characteristic for this adaptive radiation^{17,46,47}. On the other hand, the transcriptome of the other feeding-related trait in our study, the LPJ, evolved comparably

slowly, despite being similarly transcriptionally active as other organs. That gonads and liver, which show relatively little morphological variation among species (within each organ type), contain the most rapidly evolving transcriptomes in the cichlid adaptive radiation in Lake Tanganyika, whereas the morphologically highly diverse LPJ¹⁷ features a slowly evolving transcriptome, indicates that the overall rate of gene expression evolution in an adult organ is not related to its rate of morphological evolution. We note, however, that to understand the relationship between transcriptome evolution and varying morphological evolutionary rates, comparative gene expression analyses across different ontogenetic stages are necessary^{3,10}. This developmental perspective is not covered in our study targeting adult transcriptomes and should be in the focus of future investigations.

The patterns of gene expression evolution also differed among the subclades of the cichlid adaptive radiation in Lake Tanganyika (for example, in the number of organ-specific genes; Fig. 5b). The most consistent difference in gene expression pat-



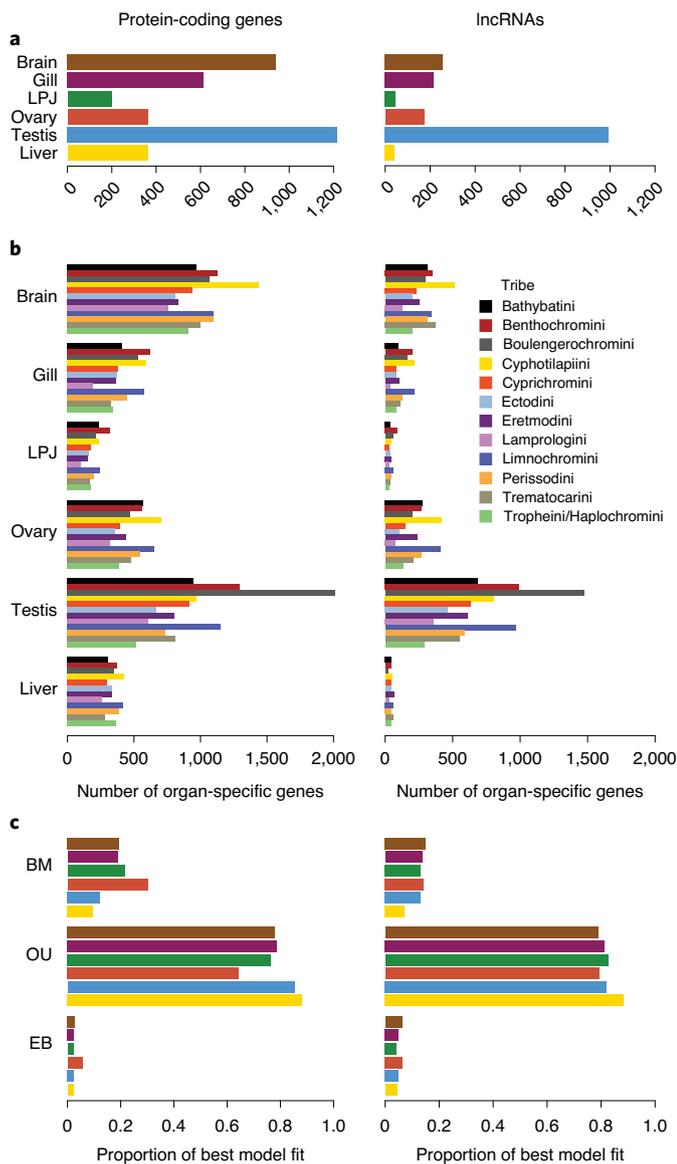


Fig. 5 | Organ-specific expression and expression dynamics of protein-coding genes and lncRNAs. Data for protein-coding genes and lncRNAs are shown in the left and right panels, respectively. **a**, The number of organ-specific genes in each organ. **b**, The number of organ-specific genes in each organ that are shared across species of the same tribe. The bars are colour-coded according to tribes as defined in Fig. 1a. **c**, The proportion of per-gene expression patterns that have evolved under a BM, OU or EB model of trait evolution. Proportions are shown per organ and colour-coded according to organ type as labelled in Fig. 1a.

terns occurred between the most species-rich tribe of cichlids in Lake Tanganyika, the Lamprologini, and the remaining tribes (Fig. 2 and Extended Data Figs. 3 and 4). We can only speculate that the overall deviating gene expression trajectories of the Lamprologini are somehow connected to their unique lifestyle compared to other cichlid tribes in Lake Tanganyika. For example, all Lamprologini are substrate spawners, whereas all but one (*Boulengerochromis microlepis*) of the non-Lamprologini species in Lake Tanganyika are mouth brooders⁴⁸. Alternatively, the distinct gene expression profiles of the Lamprologini could be due to particular features of genome evolution. For example, we have recently shown that the Lamprologini are characterized by the

highest levels of per-genome heterozygosity of all Tanganyikan cichlid tribes¹⁷.

Overall, the observed differences in the rate of gene expression evolution between organs (Figs. 4 and 5), transcriptome parts (Figs. 3 and 5) and the subclades of the radiation (Fig. 5 and Extended Data Fig. 8) suggest that differing strengths of selection have shaped transcriptome evolution in the course of the cichlid adaptive radiation in Lake Tanganyika. This is further supported by the observation that the expression levels of the majority of protein-coding genes and lncRNAs are in line with an OU model of trait evolution with varying strengths of selection (Fig. 5c).

Methods

Sampling. Sampling was performed between 2014 and 2017 at 31 locations (see Supplementary Table 1 for sampling locations and Global Positioning System coordinates) around Lake Tanganyika, under research permits issued by the University of Burundi and the Ministère de l'Eau, de l'Environnement, de l'Aménagement du Territoire et de l'Urbanisme, Republic of Burundi; the Tanzania Commission for Science and Technology, the Tanzania National Parks Authority and the Tanzania Wildlife Research Institute, United Republic of Tanzania; the Lake Tanganyika Research Unit, Department of Fisheries, Mpulungu, and the Department of Immigration, Kasama Regional Office, Republic of Zambia. This study included RNA samples from six different sources (brain, gill, liver, ovary, testis and LPJ) of six adult specimens (three males and three females, except for one species for which we had three males and four females; Supplementary Table 2) each of 76 cichlid species. In most cases, the six specimens per species were collected from the same location (Supplementary Table 1). Two species (*Tylochromis polylepis* and *Oreochromis tanganyicae*) were excluded for all downstream analyses, because these are not part of the endemic adaptive radiation of cichlid fishes in Lake Tanganyika but belong to more ancestral lineages that have colonized the lake secondarily¹⁷. One species (*Cyprichromis leptosoma*) was used only for parts of the analyses, because we lacked brain, liver and LPJ samples for this species (see Extended Data Fig. 9). Thus, for the comparative analyses, we used only species that are part of the radiation¹⁷ and for which all organs were collected. Our sampling covers the entire phylogenetic spectrum of the adaptive radiation of cichlids in Lake Tanganyika (tribes Bathybatini, Benthochromini, Boulengerochromini, Cyphotilapiini, Cyprichromini, Ectodini, Eretmodini, Lamprologini, Limnochromini, Perissodini, Trematocarini and Tropheini). The number of species sampled per tribe scales with the tribe's total number of species^{16,17} (Supplementary Tables 1 and 2). Organs were derived from adult wild-caught specimens dissected in the field immediately upon capture. Entire organs (with the exception of liver from very large specimens, which were only partially sampled) and the entire LPJ were stored individually in RNAlater.

Extraction, library preparation and Illumina sequencing. Organs and LPJ were homogenized (FastPrep-24; MP Biomedicals) and total RNA was extracted using the Direct-zol RNA kit (Zymo) according to the manufacturer's protocol. Individual libraries were constructed using the Illumina TruSeq stranded protocol including RiboZero Gold rRNA depletion (Illumina) and sequenced on an Illumina HiSeq 2500 in SE 125-bp mode at approximately 10 million reads per library. Library construction and sequencing were conducted at the Genomics Facility Basel, University of Basel and ETH Zurich Department of Biosystems Science and Engineering. RNA extraction, library preparation and sequencing were randomized with respect to species, organ and sex to avoid batch effects. Library preparation failed for 43 samples (see Supplementary Tables 7 and 8 for more information on the samples).

Quality filtering, mapping and read counting. Illumina strand-specific single-end sequences were quality-filtered using Trimmomatic⁴⁹ (v.0.33) with a 4-bp window size, a required window quality of 15 and a minimum read length of 80 bp ($2/3$ of the initial read length), followed by adapter removal in the same software. In the absence of well-assembled and annotated reference genomes for the vast majority of the cichlid species under investigation, we opted for a read mapping strategy against the Nile tilapia (*O. niloticus*), a closely related and well-annotated cichlid genome. This is also the only cichlid reference genome that has been assembled to the chromosomal level. Mapping of all transcriptomes against a common, phylogenetically equidistant and closely related (see ref.⁵⁰) reference genome has the additional advantage of facilitating orthologue assignments.

Cleaned reads were mapped against the Nile tilapia genome assembly (RefSeq assembly version GCF_001858045.1_ASM185804v2 (ref.⁴)) with STAR⁵¹ (v.2.5.2a), applying the following settings: `--outFilterMultimapNmax 1 --outFilterMatchNminOverLread 0.4 --outFilterScoreMinOverLread 0.4`. Unique alignments were reported in sorted BAM format and assigned to genes using the HTSeq-count script from the HTSeq⁵² framework (v.0.6.1p1; Supplementary Table 7a). Before further analyses and following current recommendations (DESeq2 (ref.⁵³) v.1.24.0.), we excluded 5,829 reference genes from the total of

38,425 annotated Nile tilapia genes (Supplementary Table 7b) on the basis of very low expression levels in our data (five or fewer counts in fewer than three samples). The number of reads per species and library kept at each step of the pipeline is reported in Supplementary Tables 9 and 10. Mapping statistics are reported in Supplementary Fig. 6.

All subsequent analyses were performed on two classes of RNA, protein-coding RNAs and lncRNAs. The latter are the best-represented class of non-protein-coding RNAs annotated in the Nile tilapia genome and have been studied in detail in other organisms^{42,54,55}. Our final gene dataset contained 27,105 protein-coding genes and 4,719 lncRNAs (Supplementary Table 7b), across all organs and species. Outlier samples for each organ were identified via a *k*-mean clustering approach using the function `fviz_cluster` from the R package `factoextra` (v1.0.6) (<https://www.rdocumentation.org/collaborators/name/Alboukadel%20Kassambara>). Samples that did not cluster with any other sample were removed ($n = 52$ samples; Supplementary Table 8). The sample exclusion did not change the number of species included in the analysis (Supplementary Table 7).

Global expression patterns and normalization. Expression counts were normalized with the R (v.3.5.0) package `DESeq2` (ref. ³⁸; v.1.24.0). We used variance-stabilizing transformations to convert the data, resulting in a matrix with values having constant variances along the range of mean values. Multivariate between-group PCA was then used to illustrate global patterns of gene expression differences among samples and across organs with the `DESeq2` `plotPCA` function. The expression values were transformed into TPM values⁵⁶ and the biological replicates of each species and each organ were grouped by calculating the median of TPM values. Variances within species and within sexes are represented in Supplementary Fig. 2. TPM values were then split into two categories (protein-coding genes, $n = 27,105$; and lncRNAs, $n = 4,719$), representing two different parts of the transcriptome (coding versus noncoding). Genes were placed in each group on the basis of the Nile tilapia NCBI annotation file (GCF_001858045.1_ASM185804v2 (ref. ⁴)). The resulting TPM (summarized and split) values were used for all downstream analyses.

Pairwise expression similarities. Similarity of gene expression between pairs of species was estimated (separately for protein-coding genes and lncRNAs) using Spearman's ρ , and the pairwise distances between all pairs of species were computed as $1 - \rho$. Heat maps of expression similarities among samples were produced using the `heatmap` function in the R package `heatmap` (v.1.0.12, <https://cran.r-project.org/web/packages/heatmap/index.html>). To investigate how well the expression patterns reflect the phylogenetic relatedness of the samples, we used a two-sided Student's *t*-test. To test for a possible sample size effect when comparing protein-coding genes and lncRNAs, we performed permutation tests ($n = 10,000$) in which we randomly sampled the same number of protein-coding genes out of the total set of protein-coding genes ($n = 27,105$) as there are annotated lncRNAs ($n = 4,719$), and calculated correlation coefficients on these random subsets.

Expression divergence through time within organs. As in Brawand et al.⁴ and Necsulea and Kaessmann³, we measured the relationship between gene expression and divergence time (separately for protein-coding genes and lncRNAs) with a linear regression between Spearman's ρ (as the *x* variable) and divergence times¹⁷ (as the *y* variable) for all pairs of species. The rate of evolution within each organ was then measured as $((1 - \rho)/\text{divergence time})$ for all pairs of species. A one-way ANOVA was used to determine whether there were any differences between the organs. Whenever significant effects were detected, post hoc evaluations were performed using Tukey's honestly significant difference test.

Gene expression trees. Following Brawand et al.⁴, gene expression phylogenies were constructed using the neighbour-joining approach on the pairwise distance between species (computed as $1 - \rho$, separately for both transcriptome parts and for each organ) with the `NJ` function in the R package `ape`⁵⁷ (v.5.3). Topological dissimilarities (measured as Robinson–Foulds distance⁵⁸) between the expression trees and the time-calibrated species tree based on genome-wide data (taken from Ronco et al.¹⁷) were calculated using the `treedist` function in the R package `phangorn`⁵⁸ (v.2.5.3). To test whether the rate of gene expression change along the species tree was similar among organs, we estimated, for every branch in the species tree, expression distances (computed as $1 - \rho$, separately for both transcriptome parts and each organ), using the Fitch and Margoliash⁵⁹ method as implemented in `PHYMLIP` (v.3.697, <http://evolution.genetics.washington.edu/phymlip.html>). Cumulative branch lengths from root to tips were calculated per species within each organ and then compared across organs. The rate of transcriptome change was then reported as the branch length estimated with expression data divided by the corresponding branch length of the species tree. To illustrate the temporal dynamics of transcriptome evolution, we plotted—per organ—the mean rate of expression change sampled in 0.15-Myr steps (as in ref. ¹⁷) along the time-calibrated species tree. Cumulative branch lengths from root to tips (based on a fixed topology) were calculated per species and organ and then compared across organs and tribes using ANOVA.

Organ-specific expression. Organ-specificity indices (τ) were calculated for the two parts of the transcriptome following a modified version⁵⁹ of the initial τ formula³⁹. The τ of a gene is defined as:

$$\tau_H = \frac{\sum_{j=1}^{n_H} \left(1 - \frac{\left[\log_2 \left(\frac{S_H(i,j)}{S_H(i,\max)} \right) \right]}{\log_2 \left(\frac{S_H(i,\max)}{S_H(i,\max)} \right)} \right)}{n_H - 1}$$

where n_H is the number of organs examined (in our case $n_H = 6$) and $S_H(i,\max)$ is the highest expression signal of gene *i* across the n_H organs. As proposed in Guschanski et al.³⁸, organ-specific indices were calculated using the normalized but not the transformed gene expression matrix. The τ values were then calculated using the median gene expression values per organ and tribe. Any τ indices over 0.8 were considered as indicative of organ-specific expression³⁸. The number of organ-specific genes was reported per organ and tribe.

Gene expression dynamics. To examine the dynamics of gene expression changes along the phylogeny for each gene individually, we fitted models of trait evolution to the TPM gene expression values (summarized as median per species). To do so, we used the `fitContinuous` function within the R (v.3.5.0) package `Geiger`⁶⁰ (v.2.0.6.1). Specifically, we fitted a BM, OU and EB model of trait evolution along the time-calibrated species tree (see above). Note that the EB model was tested because of the prediction that trait evolution should be rapid early in an adaptive radiation and slow down through time as the available niche space becomes filled. We applied 10,000 iterations and default parameter bounds except for the α -parameter in the OU model (attraction strength to central value), which was set to a lower limit of $\exp(-500)$ and an upper limit of 20. The approach was applied for each organ and for each transcriptome part separately (protein-coding genes and lncRNAs). Genes that were not expressed (TPM = 0) within an organ were removed before the analyses. The fraction of expressed genes per organ used for this analysis was between 96% and 99% of all protein-coding genes ($n = 27,105$) and between 81% and 99% of all lncRNAs ($n = 4,719$). We then compared the models by calculating the difference in the Akaike information criterion and reported for each gene the best model fit (the model with the lowest Akaike information criterion). The number of genes per model is reported in Supplementary Table 6.

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

Data availability

The datasets generated during and/or analysed during the current study are available in the NCBI repository under the BioProject accession number [PRJNA550295](https://www.ncbi.nlm.nih.gov/bioproject/PRJNA550295). All related metadata are available on Dryad under the project accession number [fj6q573jsj](https://doi.org/10.7554/736).

Code availability

All custom codes generated during and/or analysed in the current study are available on Dryad under the project accession number [fj6q573jsj](https://doi.org/10.7554/736).

Received: 14 February 2020; Accepted: 21 October 2020;
Published online: 23 November 2020

References

- King, M. & Wilson, C. A. Evolution at two levels in humans and chimpanzees. *Science* **188**, 107–116 (1975).
- Romero, I. G., Ruvinsky, I. & Gilad, Y. Comparative studies of gene expression and the evolution of gene regulation. *Nat. Rev. Genet.* **13**, 505–516 (2012).
- Necsulea, A. & Kaessmann, H. Evolutionary dynamics of coding and non-coding transcriptomes. *Nat. Rev. Genet.* **15**, 734–748 (2014).
- Brawand, D. et al. The evolution of gene expression levels in mammalian organs. *Nature* **478**, 343–348 (2011).
- Wray, G. A. The evolutionary significance of *cis*-regulatory mutations. *Nat. Rev. Genet.* **8**, 206–216 (2007).
- Barrier, M., Robichaux, R. H. & Purugganan, M. D. Accelerated regulatory gene evolution in an adaptive radiation. *Proc. Natl Acad. Sci. USA* **98**, 10208–10213 (2001).
- Whittall, J. B., Voelckel, C., Kliebenstein, D. J. & Hodges, S. A. Convergence, constraint and the role of gene expression during adaptive radiation: floral anthocyanins in *Aquilegia*. *Mol. Ecol.* **15**, 4645–4657 (2006).
- Stern, D. L. & Orgogozo, V. The loci of evolution: how predictable is genetic evolution? *Evolution* **62**, 2155–2177 (2008).
- Naqvi, S. et al. Conservation, acquisition, and functional impact of sex-biased gene expression in mammals. *Science* **365**, 55–61 (2019).
- Cardoso-Moreira, M. et al. Gene expression across mammalian organ development. *Nature* **571**, 505–509 (2019).
- Schluter, D. *The Ecology of Adaptive Radiation* (Oxford Univ. Press, 2000).
- Berner, D. & Salzburger, W. The genomics of organismal diversification illuminated by adaptive radiations. *Trends Genet.* **31**, 491–499 (2015).

13. Brawand, D. et al. The genomic substrate for adaptive radiation in African cichlid fish. *Nature* **513**, 375–381 (2014).
14. Jones, F. C. et al. The genomic basis of adaptive evolution in threespine sticklebacks. *Nature* **484**, 55–61 (2012).
15. Salzburger, W. Understanding explosive diversification through cichlid fish genomics. *Nat. Rev. Genet.* **19**, 705–717 (2018).
16. Ronco, F., Büscher, H. H., Indermaur, A. & Salzburger, W. The taxonomic diversity of the cichlid fish fauna of Ancient Lake Tanganyika, East Africa. *J. Great Lakes Res.* **46**, 1067–1078 (2020).
17. Ronco, F. et al. Drivers and dynamics of a massive adaptive radiation in African cichlid fishes. *Nature* <https://doi.org/10.1038/s41586-020-2930-4> (2020).
18. Fryer, G. & Iles, T. D. *The Cichlid Fishes of the Great Lakes of Africa: Their Biology and Evolution* (Oliver and Boyd, 1972).
19. Liem, K. F. Evolutionary strategies and morphological innovations: cichlid pharyngeal jaws. *Syst. Zool.* **22**, 425–441 (1973).
20. Hulsey, C. D. Function of a key morphological innovation: fusion of the cichlid pharyngeal jaw. *Proc. Biol. Sci.* **273**, 669–675 (2006).
21. Meyer, A. Ecological and evolutionary consequences of the trophic polymorphism in *Cichlasoma citrinellum* (Pisces: Cichlidae). *Biol. J. Linn. Soc.* **39**, 279–299 (1990).
22. Salzburger, W., Mack, T., Verheyen, E. & Meyer, A. Out of Tanganyika: genesis, explosive speciation, key-innovations and phylogeography of the haplochromine cichlid fishes. *BMC Evol. Biol.* **5**, 17 (2005).
23. Muschick, M., Barluenga, M., Salzburger, W. & Meyer, A. Adaptive phenotypic plasticity in the Midas cichlid fish pharyngeal jaw and its relevance in adaptive radiation. *BMC Evol. Biol.* **11**, 116 (2011).
24. Salzburger, W. The interaction of sexually and naturally selected traits in the adaptive radiations of cichlid fishes. *Mol. Ecol.* **18**, 169–185 (2009).
25. Ronco, F., Roestli, M. & Salzburger, W. A functional trade-off between trophic adaptation and parental care predicts sexual dimorphism in cichlid fish. *Proc. R. Soc. B* **286**, 20191050 (2019).
26. Khaitovich, P., Enard, W., Lachmann, M. & Pääbo, S. Evolution of primate gene expression. *Nat. Rev. Genet.* **7**, 693–702 (2006).
27. Fernández-Pérez, L., Mirecki-Garrido, M., Recio, C. & Guerra, B. in *Chemistry and Biological Activity of Steroids* (eds Ribeiro Salvador, J. A. & Cruz Silva, M. M.) Ch. 4 (IntechOpen, 2019).
28. Van Thiel, D. H. & Gavalier, J. S. in *Modern Concepts in Gastroenterology* (eds Thomson, A. B. R. & Shaffer, E.) 337–353 (Springer, 1992).
29. Shen, M. & Shi, H. Sex hormones and their receptors regulate liver energy homeostasis. *Int. J. Endocrinol.* **2015**, 294278 (2015).
30. Qiao, Q. et al. Deep sexual dimorphism in adult medaka fish liver highlighted by multi-omic approach. *Sci. Rep.* **6**, 32459 (2016).
31. Rose, E., Flanagan, S. P. & Jones, A. D. The effects of synthetic estrogen exposure on the sexually dimorphic liver transcriptome of the sex-role-reversed gulf pipefish. *PLoS ONE* **10**, e0139401 (2015).
32. Zheng, W. et al. Transcriptomic analyses of sexual dimorphism of the zebrafish liver and the effect of sex hormones. *PLoS ONE* **8**, e53562 (2013).
33. El Taher, A., Ronco, F., Matschiner, M. & Salzburger, W. Dynamics of sex chromosome evolution in a rapid radiation of cichlid fishes. Preprint at *bioRxiv* <https://doi.org/10.1101/2020.10.23.335596> (2020).
34. Meisel, R. P., Malone, J. H. & Clark, A. G. Faster-X evolution of gene expression in *Drosophila*. *PLoS Genet.* **8**, e1003013 (2012).
35. Fitch, W. & Margoliash, E. Construction of phylogenetic trees. *Science* **155**, 279–284 (1967).
36. Robinson, D. F. & Foulds, L. R. Comparison of phylogenetic trees. *Math. Biosci.* **53**, 131–147 (1981).
37. Warnefors, M. & Kaessmann, H. Evolution of the correlation between expression divergence and protein divergence in mammals. *Genome Biol. Evol.* **5**, 1324–1335 (2013).
38. Guschanski, K., Warnefors, M. & Kaessmann, H. The evolution of duplicate gene expression in mammalian organs. *Genome Res.* **27**, 1461–1474 (2017).
39. Yanai, I. et al. Genome-wide midrange transcription profiles reveal expression level relationships in human tissue specification. *Bioinformatics* **21**, 650–659 (2005).
40. Chen, J. et al. A quantitative framework for characterizing the evolutionary history of mammalian gene expression. *Genome Res.* **29**, 53–63 (2019).
41. Chan, E. T. et al. Conservation of core gene expression in vertebrate tissues. *J. Biol.* **8**, 33 (2009).
42. Necseulea, A. et al. The evolution of lncRNA repertoires and expression patterns in tetrapods. *Nature* **505**, 635–640 (2014).
43. Soumillon, M. et al. Cellular source and mechanisms of high transcriptome complexity in the mammalian testis. *Cell Rep.* **3**, 2179–2190 (2013).
44. Nielsen, R. et al. A scan for positively selected genes in the genomes of humans and chimpanzees. *PLoS Biol.* **3**, 0976–0985 (2005).
45. Eling, N., Morgan, M. D. & Marioni, J. C. Challenges in measuring and understanding biological noise. *Nat. Rev. Genet.* **20**, 536–548 (2019).
46. Muschick, M., Indermaur, A. & Salzburger, W. Convergent evolution within an adaptive radiation of cichlid fishes. *Curr. Biol.* **22**, 2362–2368 (2012).
47. Wagner, C. E., McIntyre, P. B., Buels, K. S., Gilbert, D. M. & Michel, E. Diet predicts intestine length in Lake Tanganyika's cichlid fishes. *Funct. Ecol.* **23**, 1122–1131 (2009).
48. Konings, A. *Tanganyika Cichlids in Their Natural Habitat* (Cichlid, 2015).
49. Bolger, A. M., Lohse, M. & Usadel, B. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* **30**, 2114–2120 (2014).
50. Matschiner, M., Böhne, A., Ronco, F. & Salzburger, W. The genomic timeline of cichlid fish diversification across continents. *Nat. Commun.* <https://doi.org/10.1038/s41467-020-17827-9> (2020).
51. Dobin, A. et al. STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* **29**, 15–21 (2013).
52. Anders, S., Pyl, P. T. & Huber, W. HTSeq-A Python framework to work with high-throughput sequencing data. *Bioinformatics* **31**, 166–169 (2015).
53. Love, M. I., Huber, W. & Anders, S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* **15**, 550 (2014).
54. Sarropoulos, I., Marin, R., Cardoso-moreira, M. & Kaessmann, H. Developmental dynamics of lncRNAs across mammalian organs and species. *Nature* **571**, 510–514 (2019).
55. Kutter, C. et al. Rapid turnover of long noncoding RNAs and the evolution of gene expression. *PLoS Genet.* **8**, e1002841 (2012).
56. Li, B., Ruotti, V., Stewart, R. M., Thomson, J. A. & Dewey, C. N. RNA-Seq gene expression estimation with read mapping uncertainty. *Bioinformatics* **26**, 493–500 (2009).
57. Paradis, E., Claude, J. & Strimmer, K. APE: analyses of phylogenetics and evolution in R language. *Bioinformatics* **20**, 289–290 (2004).
58. Schliep, K. P. phangorn: phylogenetic analysis in R. *Bioinformatics* **27**, 592–593 (2011).
59. Liao, B. & Zhang, J. Low rates of expression profile divergence in highly expressed genes and tissue-specific genes during mammalian evolution. *Mol. Biol. Evol.* **23**, 1119–1128 (2006).
60. Pennell, M. W. et al. Geiger v2.0: an expanded suite of methods for fitting macroevolutionary models to phylogenetic trees. *Bioinformatics* **30**, 2216–2218 (2014).

Acknowledgements

We thank the University of Burundi, the Ministère de l'Eau, de l'Environnement, de l'Aménagement du Territoire et de l'Urbanisme, Republic of Burundi, the Tanzania Commission for Science and Technology, the Tanzania Fisheries Research Institute, the Tanzania National Parks Authority, the Tanzania Wildlife Research Institute and the Lake Tanganyika Research Unit, Department of Fisheries, Republic of Zambia, for research permits; G. Banyankimbona (University of Burundi), I. Kimirei (TAFIRI, Kigoma, Tanzania) and T. Banda and L. Makasa (Department of Fisheries, Mpulungu, Zambia) for assistance with research permits; the boat crews of the *Chomba* (D. Mwanakulya, J. Sichilima and H. D. Sichilima Jr) and the *Maji Makubwa II* (G. Kazumbe and family) for help during field work; V. Huwiler (Kalambo Lodge, Zambia), 'Charity' (Nkupi Lodge, Zambia) and the Zytkow family (Ndole Bay Lodge, Zambia) for lodging; C. Zytkow (Conservation Lake Tanganyika, Zambia), and P. Lassen and V. Huwiler (Kalambo Lodge, Zambia) for logistic support; C. Beisel and E. Burcklen at the Genomics Facility Basel for library preparation and sequencing; and J. Himes for fish illustrations. We also thank A. Necseulea for discussions and advice on the project. 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). This work was funded by the European Research Council (Consolidator Grant no. 617585 'CICHLID-X') and the Swiss National Science Foundation to W.S.

Author contributions

A.E.T., F.R., A.I., N.B., L.W. and W.S. collected and/or dissected the specimens in the field. A.E.T. and N.B. organized the RNA-sequencing data production. A.E.T. processed and mapped the reads. A.E.T. performed all data analyses except for the temporal dynamics of transcriptome evolution that F.R. performed. A.B. contributed ideas and supervised data analyses. F.R. formatted the final figures. A.E.T. and W.S. wrote the manuscript with input from all authors. The project was originally designed by W.S., with input from A.E.T., F.R., A.B. and A.I.

Competing interests

The authors declare no competing interests.

Additional information

Extended data is available for this paper at <https://doi.org/10.1038/s41559-020-01354-3>.

Supplementary information is available for this paper at <https://doi.org/10.1038/s41559-020-01354-3>.

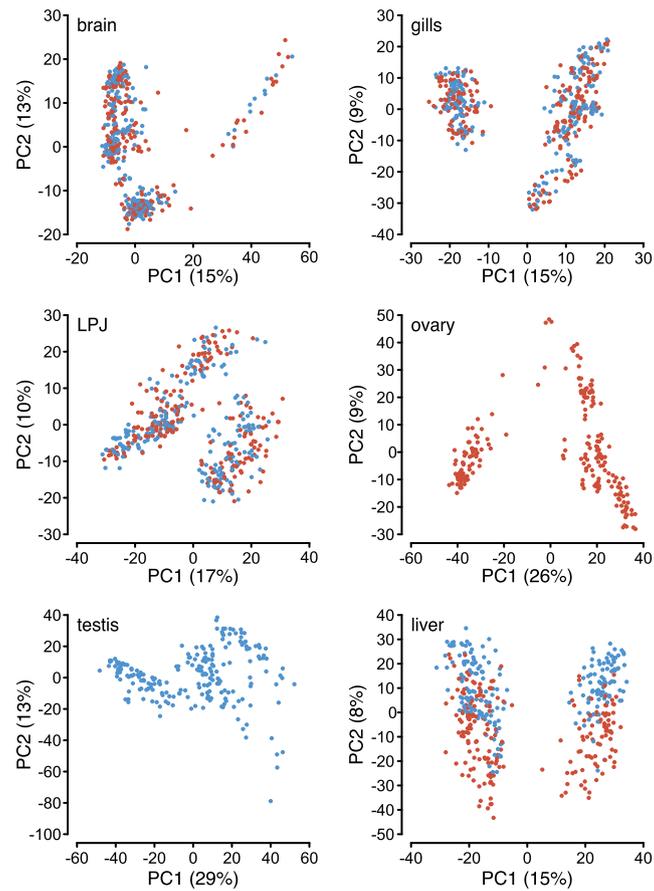
Correspondence and requests for materials should be addressed to A.E. or W.S.

Peer review information *Nature Ecology & Evolution* thanks Emily Wong, Marie Sémon and the other, anonymous, reviewer(s) for their contribution to the peer review of this work.

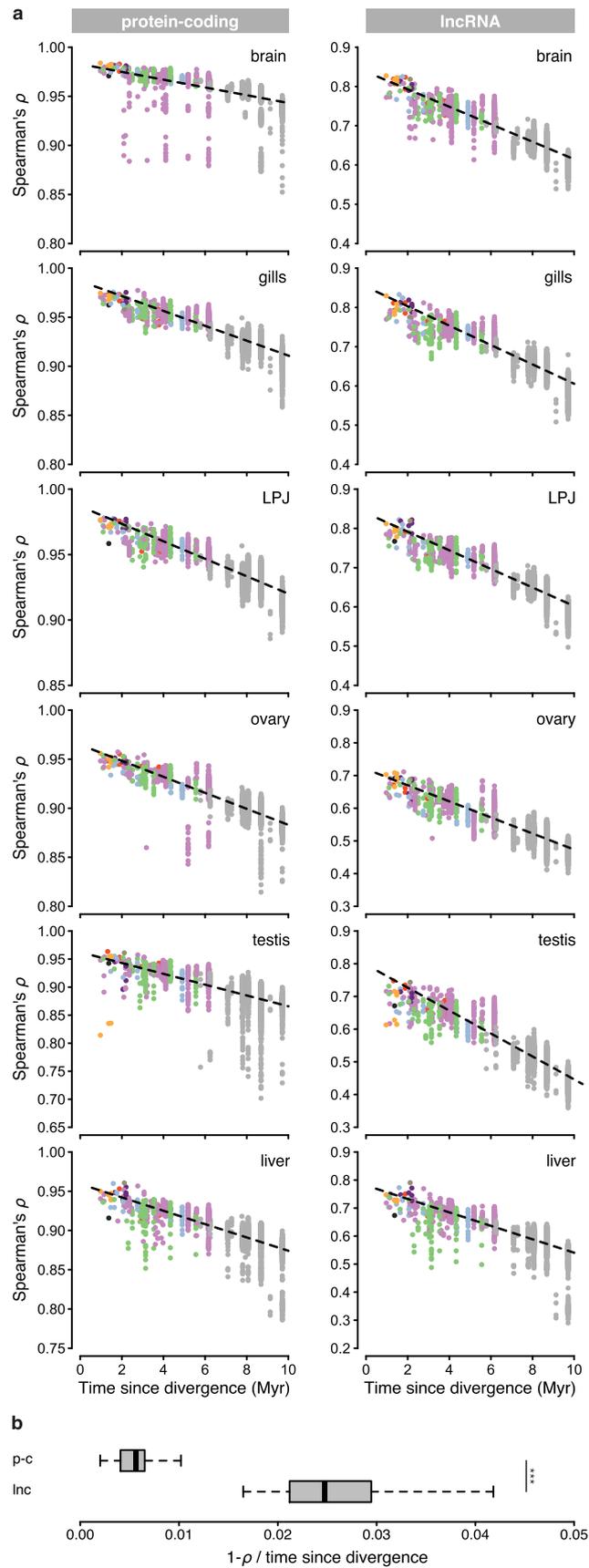
Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

© The Author(s), under exclusive licence to Springer Nature Limited 2020

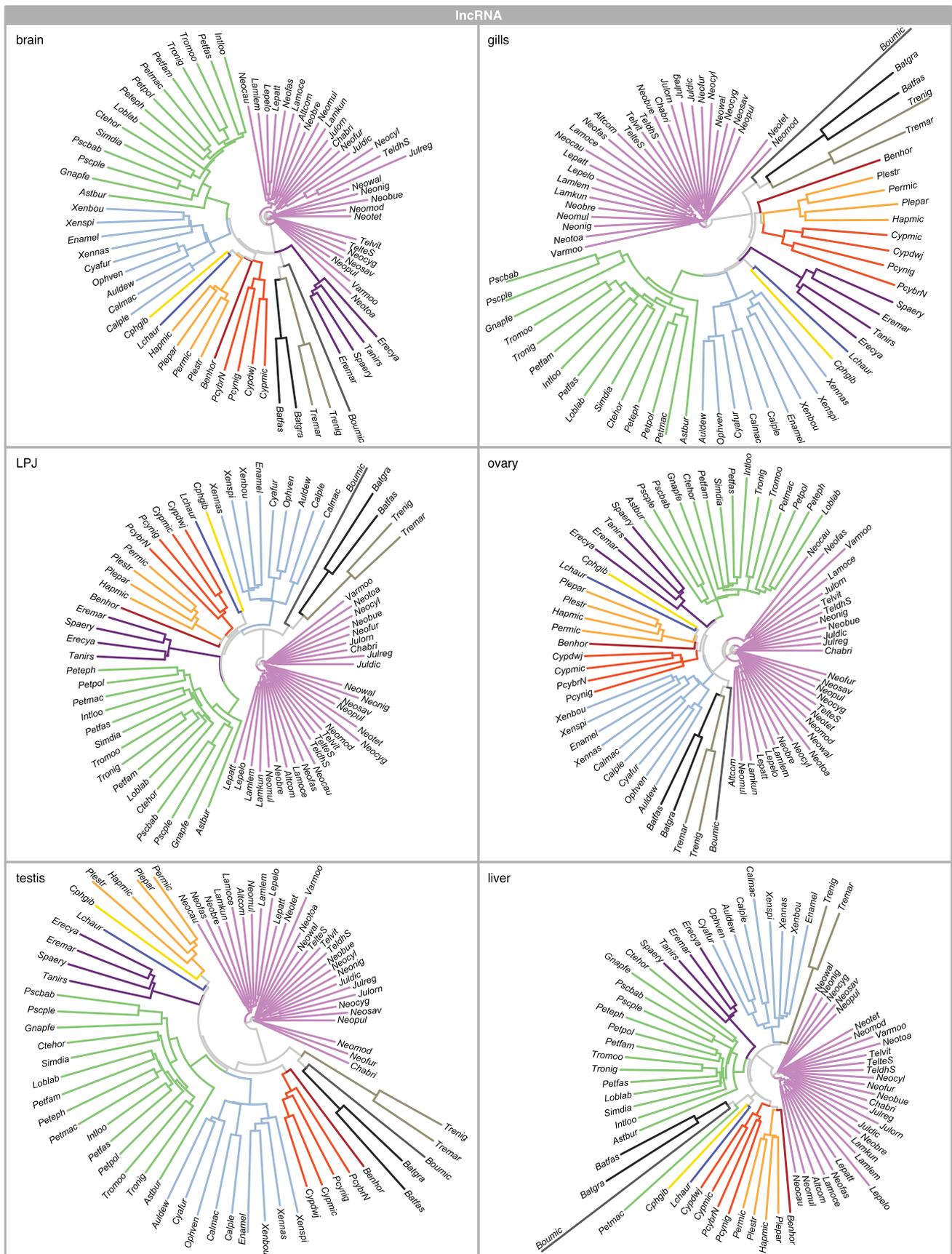


Extended Data Fig. 1 | Gene expression patterns per organ and sex. Principal component analyses of overall gene expression levels in brain, gill, lower pharyngeal jaw bone (LPJ), ovary, testis, and liver. Samples (brain: $n = 428$; gill: $n = 434$; LPJ: $n = 425$; ovary: $n = 219$; testis: $n = 213$; liver: $n = 412$) are coloured according to sex (red: female, blue: male). The proportion of variance explained by the first two principal components (PC1 and PC2) for each organ are indicated in parenthesis at x and y axes, respectively.

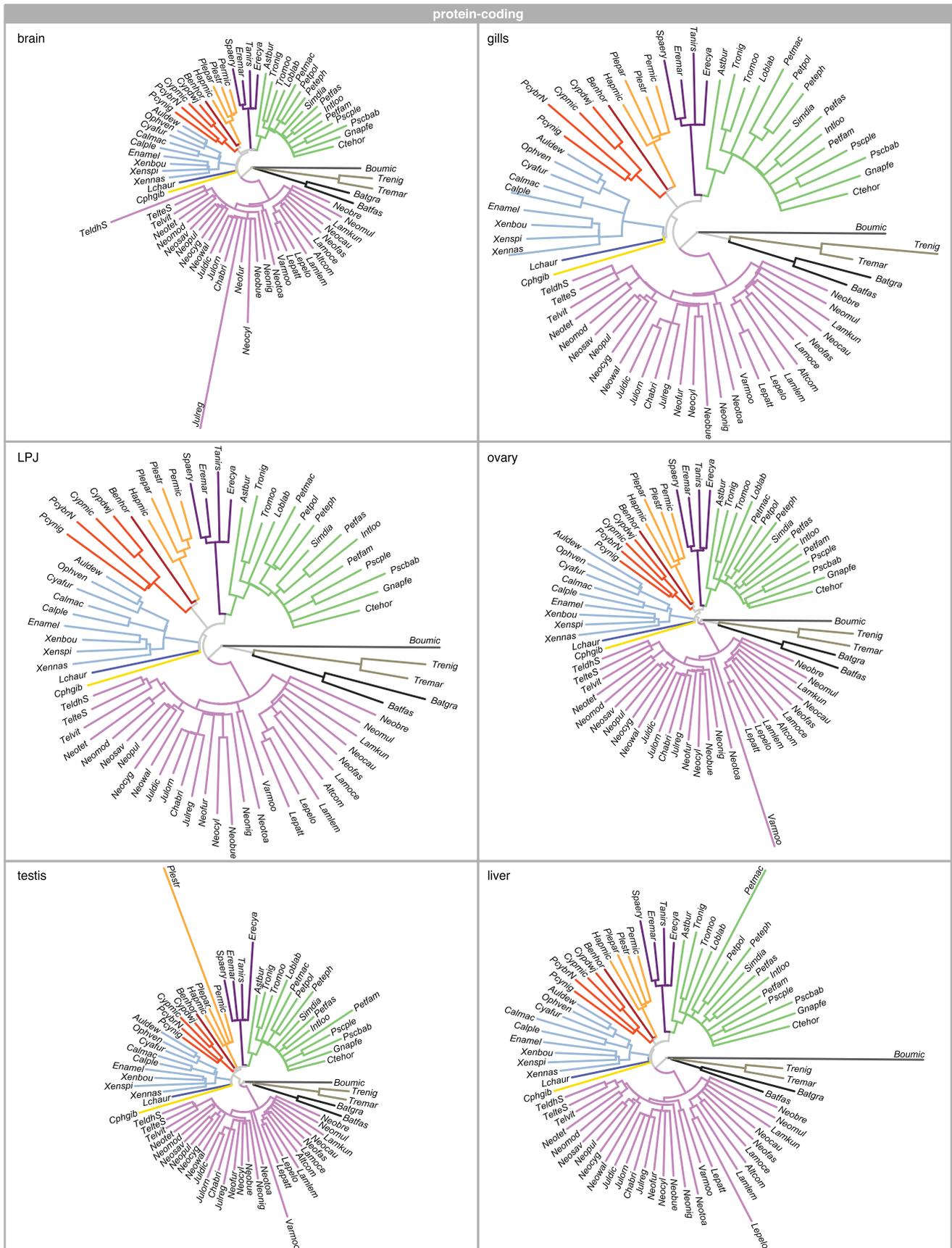


Extended Data Fig. 2 | See next page for caption.

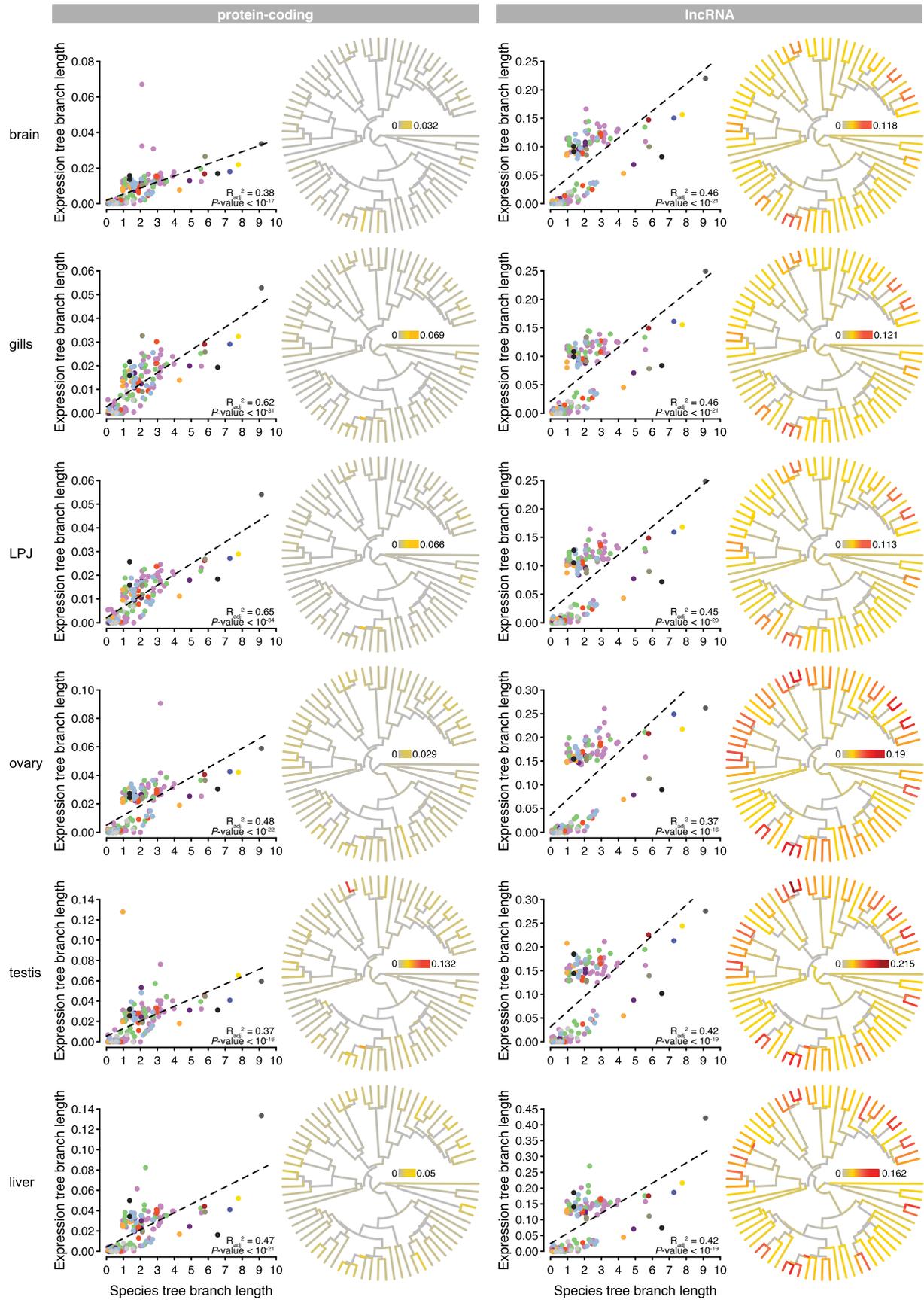
Extended Data Fig. 2 | Expression variation through time within organs and transcriptome parts. **a**, Pairwise Spearman's rank correlation coefficient (ρ) of per species (brain, ovary, gill and testis: $n = 74$ taxa; LPJ and liver: $n = 73$ taxa) as a function of divergence time¹⁷ for protein-coding genes (left panel) and lncRNAs (right panel) in brain, gill, LPJ, ovary, testis, and liver. Samples are colour-coded according to tribe as defined in Fig. 1a; pairs of species belonging to two different tribes are indicated in grey. The regression line is represented with a dashed black line. **b**, Comparison of rate of expression change (measured as $[1 - \rho] / \text{divergence time}^{17}$) between protein-coding genes (p-c) and lncRNAs (lnc) (two-sided t-test: $***P < 10^{-16}$). The box plot centre lines represent the median, box limits the upper and lower quartiles, and whiskers the 1.5x interquartile range. Outliers are not shown.



Extended Data Fig. 4 | IncRNAs expression trajectories. Neighbour-joining trees based on pairwise distance matrices of gene expression between pairs of species ($n = 73$ taxa) for IncRNAs for brain, gill, LPJ, ovary, testis, and liver. All branches are colour-coded according to tribe as defined in Fig. 1a (see Extended Data Fig. 9 and Supplementary Table 2 for full species names).

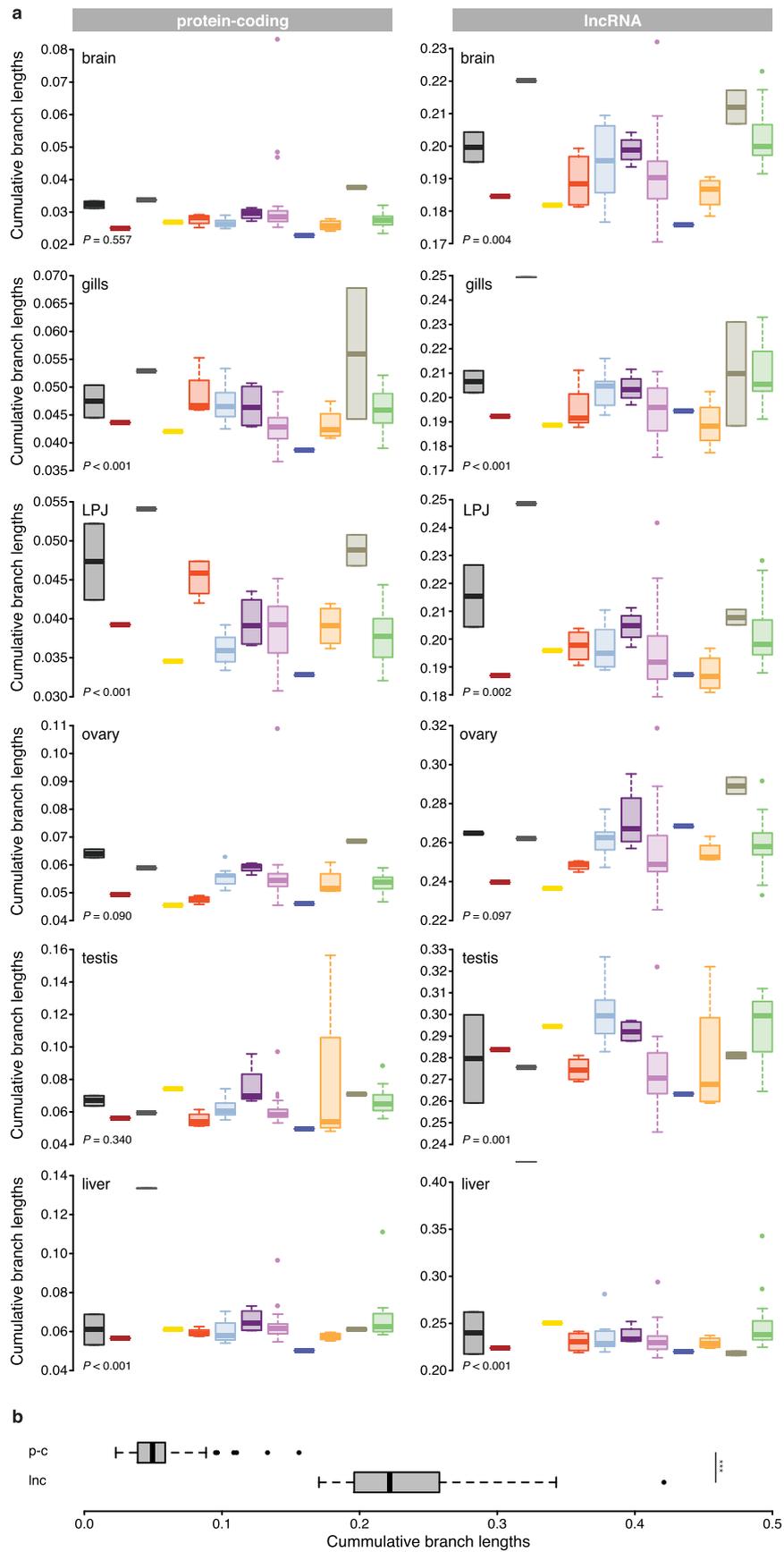


Extended Data Fig. 5 | Rate of protein-coding gene expression evolution along the species tree. Species tree with branch lengths estimated along the fixed species tree topology³⁵ ($n = 73$ taxa) based on pairwise correlations of gene expression of protein-coding genes in brain, gill, LPJ, ovary, testis, and liver. All branches are colour-coded according to tribe as defined in Fig. 1a (see Extended Data Fig. 9 and Supplementary Table 2 for full species names).



Extended Data Fig. 7 | See next page for caption.

Extended Data Fig. 7 | Rate of transcriptome evolution within organs for protein-coding genes (left panel) and lncRNAs (right panel). Linear regression of the expression tree branch length (calculated along the fixed species tree ($n = 73$ taxa) topology, Extended Data Fig. 3c, d) as a function of species tree branch lengths (Fig. 1a) for brain, gill, LPJ, ovary, testis, and liver. Data points representing branches within tribes are colour-coded corresponding to the tribe as defined in Fig. 1a, and data points representing branches that link species from different tribes are coloured in grey. Dashed lines represent linear model fits. Next to each plot, a time-calibrated species tree is shown, with branches coloured according to the rate of transcriptome evolution (measured as expression tree branch length divided by species tree branch length).



Extended Data Fig. 8 | See next page for caption.

Extended Data Fig. 8 | Level of expression variation within organs. a, Cumulative branch lengths (from root to tip of expression tree branch length calculated along the fixed species tree ($n = 73$ taxa) topology; Extended Data Fig. 3c, d) for protein-coding genes (left panel) and lncRNAs (right panel) in brain, gill, LPJ, ovary, testis, and liver calculated per species and summarised per tribe ($n = 12$ tribes). Boxplots are colour-coded according to tribe as defined in Fig. 1a; boxplot centre lines represent the median, box limits the upper and lower quartiles, and whiskers the 1.5x interquartile range. Differences among the tribes were assessed using an ANOVA (see Supplementary Table 5 for the P -values for all pairwise comparisons). **b,** Comparison of cumulative branch lengths between protein-coding genes (p-c) and lncRNAs (lnc) (two-sided t-test: $***P < 10^{-8}$). Boxplot centre lines represent the median, box limits the upper and lower quartiles, and whiskers the 1.5x interquartile range.

Tribe	Abbreviation	Species Name
Bathybatini	Batfas	<i>Bathybates fasciatus</i>
Bathybatini	Batgra	<i>Bathybates graueri</i>
Benthochromini	Benhor	<i>Benthochromis horii</i>
Boulengerochromini	Boumic	<i>Boulengerochromis microlepis</i>
Cyphotilapiini	Ophgib	<i>Cyphotilapia gibberosa</i>
Cyprichromini	PcybrN	<i>Paracyprichromis brienii</i>
Cyprichromini	Pcynig	<i>Paracyprichromis nigripinnis</i>
Cyprichromini	Cypdwj	<i>Cyprichromis</i> sp. "dwarf jumbo"
Cyprichromini	Cypmic	<i>Cyprichromis microlepidotus</i>
Ectodini	Ophven	<i>Ophthalmotilapia ventralis</i>
Ectodini	Auldew	<i>Aulonocranus dewindti</i>
Ectodini	Calmac	<i>Callochromis macrops</i>
Ectodini	Calple	<i>Callochromis pleurospilus</i>
Ectodini	Cyafur	<i>Cyathopharynx furcifer</i>
Ectodini	Enamel	<i>Enantiopus melanogenys</i>
Ectodini	Xenbou	<i>Xenotilapia boulengeri</i>
Ectodini	Xennas	<i>Xenotilapia nasus</i>
Ectodini	Xenspi	<i>Xenotilapia spiloptera</i>
Eretmodini	Erecya	<i>Eretmodus cyanostictus</i>
Eretmodini	Eremar	<i>Eretmodus marksmithi</i>
Eretmodini	Tanirs	<i>Tanganicodus irsacae</i>
Eretmodini	Spaery	<i>Spathodus erythrodon</i>
Lamprologini	Altcom	<i>Altalamprologus compressiceps</i>
Lamprologini	Neofas	<i>Neolamprologus fasciatus</i>
Lamprologini	Chabri	<i>Chalinochromis brichardi</i>
Lamprologini	Juldic	<i>Julidochromis dickfeldi</i>
Lamprologini	Julorn	<i>Julidochromis ornatus</i>
Lamprologini	Julreg	<i>Julidochromis regani</i>
Lamprologini	Lamkun	<i>Lamprologus kungweensis</i>
Lamprologini	Lamlem	<i>Lamprologus lemairii</i>
Lamprologini	Lamoce	<i>Lamprologus ocellatus</i>
Lamprologini	Lepatt	<i>Lepidolamprologus attenuatus</i>
Lamprologini	Lepelo	<i>Lepidolamprologus elongatus</i>
Lamprologini	Neobre	<i>Neolamprologus brevis</i>
Lamprologini	Neobue	<i>Neolamprologus buescheri</i>
Lamprologini	Neocau	<i>Neolamprologus caudopunctatus</i>
Lamprologini	Neocyg	<i>Neolamprologus</i> sp. "cygnus"
Lamprologini	Neocyl	<i>Neolamprologus cylindricus</i>
Lamprologini	Neofur	<i>Neolamprologus furcifer</i>
Lamprologini	Neomod	<i>Neolamprologus modestus</i>
Lamprologini	Neomul	<i>Neolamprologus multifasciatus</i>
Lamprologini	Neonig	<i>Neolamprologus niger</i>
Lamprologini	Neopul	<i>Neolamprologus pulcher</i>
Lamprologini	Neosav	<i>Neolamprologus savoryi</i>
Lamprologini	Neotet	<i>Neolamprologus tetracanthus</i>
Lamprologini	Neotoa	<i>Neolamprologus toae</i>
Lamprologini	Neowal	<i>Neolamprologus walteri</i>
Lamprologini	TeldhS	<i>Telmatochromis dhonti</i>
Lamprologini	TelteS	<i>Telmatochromis temporalis</i>
Lamprologini	Telvit	<i>Telmatochromis vittatus</i>
Lamprologini	Varmoo	<i>Variabilichromis moorii</i>
Limnochromini	Lchaur	<i>Limnochromis auritus</i>
Perissodini	Permic	<i>Perissodus microlepis</i>
Perissodini	Plepar	<i>Plecodus paradoxus</i>
Perissodini	Plestr	<i>Plecodus straeleni</i>
Perissodini	Hapmic	<i>Haplotaxodon microlepis</i>
Trematocarini	Tremer	<i>Trematocara marginatum</i>
Trematocarini	Trenig	<i>Trematocara nigrifrons</i>
Tropheini / Haplochromini	Peteph	<i>Petrochromis ephippium</i>
Tropheini / Haplochromini	Petfam	<i>Petrochromis famula</i>
Tropheini / Haplochromini	Petfas	<i>Petrochromis fasciolatus</i>
Tropheini / Haplochromini	Petmac	<i>Petrochromis macrognathus</i>
Tropheini / Haplochromini	Petpol	<i>Petrochromis polyodon</i>
Tropheini / Haplochromini	Astbur	<i>Astatotilapia burtoni</i>
Tropheini / Haplochromini	Ctehor	<i>Ctenochromis horei</i>
Tropheini / Haplochromini	Gnapfe	<i>Gnathochromis pfefferi</i>
Tropheini / Haplochromini	Intloo	<i>Interochromis loocki</i>
Tropheini / Haplochromini	Loblab	<i>Lobochilotes labiatus</i>
Tropheini / Haplochromini	Pscbab	<i>Pseudosimochromis babaulti</i>
Tropheini / Haplochromini	Simdia	<i>Simochromis diagramma</i>
Tropheini / Haplochromini	Pscple	<i>Pseudosimochromis babaulti</i> (South)
Tropheini / Haplochromini	Tromoo	<i>Tropheus moorii</i>
Tropheini / Haplochromini	Tronig	<i>Tropheus</i> sp. "black"

Extended Data Fig. 9 | Species information. List of species used in this experiment with abbreviation code, full species name and tribe information.

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 our [Editorial Policies](#) 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 No software was used.

Data analysis Data were analyzed using the following publicly available software:
Trimomatic (v.0.33), STAR (v.2.5.2a), HTSeq(v.0.6.1p1), Deseq2(v.1.24.0), R (v.3.5.0), pheatmap (v.1.0.12), ape (v.5.3), phangorn (v. 2.5.3), PHYLIP (v.3.697), UpSetR(v.1.4.0), Geiger60 (v.2.0.6.1)

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 and 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

Data used in this study are available from NCBI under the BioProject accession numbers PRJNA552202 and PRJNA550295. All custom codes and all metadata are available on Dryad under the project accession number <https://doi.org/10.5061/dryad.fj6q573sj>.

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](https://www.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	We examine patterns of gene expression evolution across 73 species representing all radiating lineages of Lake Tanganyika cichlids. We sequenced the transcriptomes of five organs (brain, gill, liver, gonad, lower pharyngeal jaw) of six individuals per species (3 males, 3 females). A comprehensive list of species is provided in Extended Data Table 1.
Research sample	We chose endemic cichlid species representing all lineages of the Lake Tanganyika radiation.
Sampling strategy	Fish were either caught with barrier nets while snorkeling or Scuba diving, or purchased from local fishermen. 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. Following the current standards in the field, we collected six specimen per species (three males and three females). The selected set of 76 species represent the diversity among the 15 subclades of the radiation.
Data collection	Organ samples were taken during the field work described below, samples were immediately placed in RNAlater after organ dissection. RNA extraction was performed at the Zoological Institute of the University of Basel. Sequencing libraries were prepared at the Genomics Facility Basel (GFB), University of Basel and ETH Zurich Department of Biosystems Science and Engineering (D-BSSE) in Basel, which is also where sequencing was done.
Timing and spatial scale	Fish were sampled during several field expeditions at 31 locations at Lake Tanganyika from 2014 till 2017 in the Republic of Burundi, the United Republic of Tanzania and Republic of Zambia. RNA was extracted between August 2016 and May 2018. RNA was sequenced between October 2016 and July 2018.
Data exclusions	In the analysis, outlier samples (individual RNAseq libraries that did not cluster with their respective biological replicates) were excluded as described in the methods section, which did not change the number of included species. 5,829 reference genes with very low expression levels in our dataset (five or less counts in less than three samples) were excluded from the analyses based on current standards in the field.
Reproducibility	Data analyses are reproducible with the information given in the methods and supplementary information.
Randomization	This study is based on multi-tissue single species transcriptome data, randomisation does not apply.
Blinding	Blinding is not relevant to this study, analyses required an understanding of the source organisms and their phylogenetic relationships.
Did the study involve field work?	<input checked="" type="checkbox"/> Yes <input type="checkbox"/> No

Field work, collection and transport

Field conditions	No field conditions are relevant to this study.
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 in Supplementary Table 1.
Access & 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'Aménagement du Territoire et de l'Urbanisme, Republic of Burundi 770 06/62710, issued 27/12/2014 Research permit issued by the Université 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 Université 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'Aménagement 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):

2016-373-NA-2015-96 (Walter Salzburger), valid 12/12/2016 to 11/12/2017
2016-374-NA-2015-96 (Athimed El Taher), valid 12/12/2016 to 11/12/2017
2016-375-NA-2015-96 (Lukas Widmer), 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 (Adrian Indermaur), valid 12/12/2016 to 11/12/2017

Research permits, issued by the Tanzania National Parks Authority (TANAPA):

TNP/HQ/C.10/13/2017 (Heinz Büscher, Athimed El Taher, Adrian Indermaur, Fabrizia Ronco, Lukas Widmer, Walter Salzburger), valid 12/12/16 to 11/12/17

Research Clearance, issued by the Tanzania Wildlife Research Institute (TAWIRI):

13300 (Heinz Büscher, Athimed El Taher, Adrian Indermaur, Fabrizia Ronco, Walter Salzburger), dated 09/01/2017

Residence permits, issued by the Department of Immigration:

RPC11100834 (Walter Salzburger), valid 11/12/2017 to 10/12/2017
RPC11100835 (Fabrizia Ronco), valid 11/12/2017 to 10/12/2017
RPC11100836 (Heinz Büscher), valid 11/12/2017 to 10/12/2017
RPC11100837 (Adrian Indermaur), valid 11/12/2017 to 10/12/2017
RPC11100839 (Lukas Widmer), valid 11/12/2017 to 10/12/2017
RPC11100840 (Athimed El Taher), valid 11/12/2017 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.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
SP005940 (Athimed El Taher), 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 13/09/2016
Export/transport permit, issued 29/08/2017

Schweizerische Eidgenossenschaft/Confœderatio 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_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	Involvement	Item
<input checked="" type="checkbox"/>	<input type="checkbox"/>	Antibodies
<input checked="" type="checkbox"/>	<input type="checkbox"/>	Eukaryotic cell lines
<input checked="" type="checkbox"/>	<input type="checkbox"/>	Palaeontology and archaeology
<input type="checkbox"/>	<input checked="" type="checkbox"/>	Animals and other organisms
<input checked="" type="checkbox"/>	<input type="checkbox"/>	Human research participants
<input checked="" type="checkbox"/>	<input type="checkbox"/>	Clinical data
<input checked="" type="checkbox"/>	<input type="checkbox"/>	Dual use research of concern

Methods

n/a	Involvement	Item
<input checked="" type="checkbox"/>	<input type="checkbox"/>	ChIP-seq
<input checked="" type="checkbox"/>	<input type="checkbox"/>	Flow cytometry
<input checked="" type="checkbox"/>	<input type="checkbox"/>	MRI-based neuroimaging

Animals and other organisms

Policy information about [studies involving animals](#); [ARRIVE guidelines](#) recommended for reporting animal research

Laboratory animals	The study did not involve laboratory animals.
Wild animals	Information on animal samples is provided in Supplementary Table 1.
Field-collected samples	Samples were collected as described under Sampling Strategy and Data Collection.
Ethics oversight	The study was approved by the Ministère de l'Eau, de l'Environnement, de l'Aménagement du Territoire et de l'Urbanisme, Republic of Burundi, 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 Cantonal Veterinary Service Basel Stadt.

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