# Evolutionary dynamics of gene and isoform regulation underlying rapid and parallel adaptive radiations

Pooja Singh[1,2,3], Ehsan P. Ahi[3,4], Anna Duenser[3,5], Marija Durdevic[6,7], Wolfgang Gessl[3], Sylvia Schaeffer[3], Ole Seehausen[1,2] and Christian Sturmbauer[3]

[1]Aquatic Ecology and Evolution, Institute of Ecology and Evolution, University of Bern, 3012 Bern, Switzerland

[2]Swiss Federal Institute of Aquatic Science and Technology (EAWAG), 6048 Kastanienbaum, Switzerland

[3]Department of Biology, University of Graz, 8010 Graz, Austria

[4]Organismal and Evolutionary Biology Research Programme, University of Helsinki, Helsinki, Finland

[5]Federal Agency for Water Management, Institute for Freshwater Ecology and Fisheries Management, Mondsee, Austria

[6]Diagnostic & Research Center for Molecular BioMedicine, Institute of Pathology, Medical University of Graz, 8010 Graz, Austria

[7]Core Facility Computational Bioanalytics, Center for Medical Research, Medical University of Graz, Graz, Austria

Corresponding author: P. Singh (pooja.singh09@gmail.com)

## Abstract

While the contribution of gene expression (GE) to adaptive evolution is widely accepted, the role of alternative splicing (AS) remains less understood. Here, we investigate AS and GE across three iconic adaptive radiations of African cichlid fishes that evolved within <16,000 to 3.8 million years. We show that AS evolves faster than GE, with both sources of variation being 'fine-tuned' over evolutionary time to become species-specific and clade-specific, respectively. Ecologically divergent species from younger radiations exhibit greater differences in splicing than those from older radiations. Most of these differentially spliced isoforms arose from standing variation, which was also present at low levels in non-radiating species, and increased in frequency during the adaptive radiation process. We identified several novel isoforms of craniofacial remodelling genes that emerged within each lake radiation and were differentially incorporated in the jaws of herbivorous vs. carnivorous species. Our findings indicate that a complex temporal interplay of GE and AS underlies adaptive radiation, with ancestral splice variation enabling rapid ecological diversification at early stages of speciation.

Key words: gene regulation, speciation, convergent evolution, trophic adaptation, morphology

## Introduction

Since King & Wilson's[1] landmark paper, the role of gene expression (GE) is widely accepted in adaptive evolution but alternative splicing (AS) has been largely overlooked as a potential substrate for adaptation and speciation[2]. Alternative splicing is a ubiquitous co-/post-transcriptional mechanism in multicellular organisms that can expand the transcriptomic and proteomic repertoire of a cell by generating multiple distinct messenger RNAs (mRNAs) from a single gene[3]. There exists a strong relationship between splicing complexity and organismal complexity across the tree of life[4]. Landmark cross-species comparisons across deeply divergent vertebrates have identified increased rates of AS evolution between species and linked it to some key evolutionary events[5–7]. However, the contribution of AS to the evolution of ecological and species richness, particularly at short evolutionary timescales, is not well understood as large-scale comparative studies are lacking[2].

Adaptive radiations of cichlid fishes are excellent models to study recent and rapid adaptive evolution in response to ecological opportunity[8]. They have evolved diverse jaw and teeth shapes and feeding habits to adapt to different food sources, forming species-rich communities spanning complex foodwebs[9,10]. Although the distinctive features of adaptive radiation are still debated[11,12], it has been proposed that both GE[13–16] and more recently, AS[17,18], may be important for eco-morphological divergence. However, it is not fully known how widespread the contribution of splicing is to adaptive radiation and which factors shape its evolution. Furthermore, the interplay of AS versus GE across varied phylogenetic timescales remains unclear. We propose that these two sources of heritable variation may play distinct roles in biological diversification, such as rapid functional diversfication during/near speciation through splicing versus much slower adaptive fine-tuning after speciation through gene expression. This hypothesis has not been tested. Differently aged radiations of cichlid fishes on the African continent provide a powerful framework to adress this knowledge gap (Figure 1A).

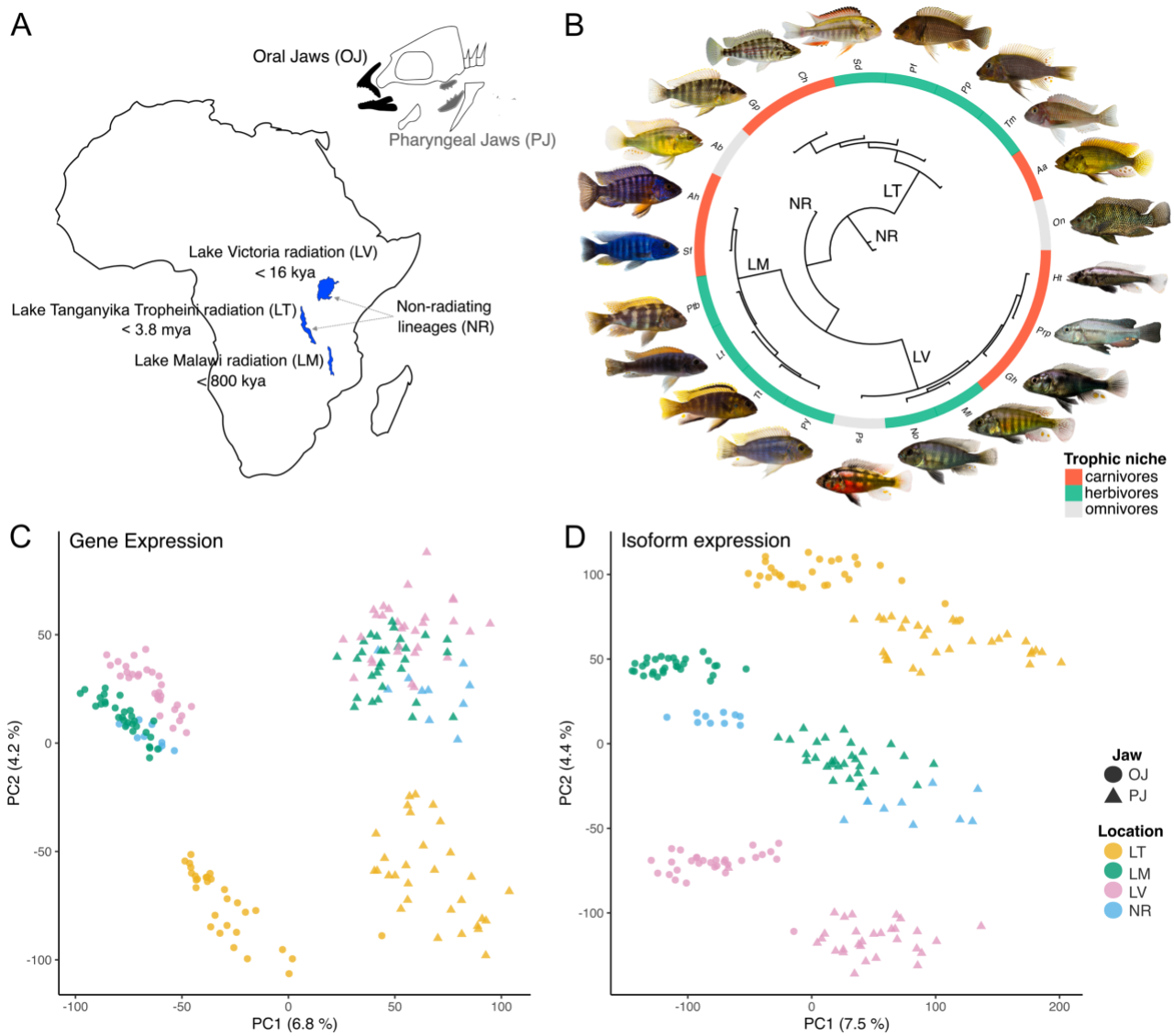


**Figure 1 Study overview. (A, B)** Biogeographical setting of species sampled in this study (20 species, 2 jaw tissues (oral and pharyngeal), 5 biological replicates per species per tissue, n=200). Phylogenetic relationships reconstructed using transcriptome SNPs. Data for *O. niloticus* (On) outgroup is from XX. Lake Victoria: *Paralabidochromis sauvagei* (Ps), *Neochromis omnicaeruleus* (No), *Mbipia lutea* (Ml), *Prognathochromis perrieri* (Pp); *Haplochromis thereuterion* (Ht); Lake Malawi: *Aulonocara hansbaenschi* (Ah), *Sciaenochromis fryeri* (Sf), *Petrotilapia* sp. *'Thick bars'* (Ptb), *Labeotropheus trewavasae* (Lt), *Tropheops tropheops* (Tt), *Petrotilapia* sp. 'yellow-chin' (Py); Lake Tanganyika: *Tropheus moorii* (Tm), *Petrochromis polyodon* (Pp), *Petrochromis famula* (Pf), *Simochromis diagramma* (Sd), *Ctenochromis horei* (Ch), *Gnathochromis pfefferi* (Gp). Non-radiating haplochromines, *Astatotilapia burtoni* (Ab) and *Astatoreochromis alluaudi* (Aa) are included as riverine lineages that entered lakes but did not radiate. **(C, D)** Principal component analysis of transcriptome-wide normalised gene and isoform expression from oral (OJ) and pharyngeal jaws (PJ) of all 20 species.

Here we investigate the evolution of GE and AS in cichlid adaptive radiations from the three African Great Lakes and test if they played a role in trophic diversification. Lake Victoria (LV) hosts the youngest radiation that arose over the past 16,000 years with ~500 species known, Lake Malawi (LM) hosts the second oldest radiation with ~700 species that arose over the past 800,000 years ago, and Tropheini cichlids from Lake Tanganyika (LT) that comprise the oldest radiation in our study with 31 species that arose < 3.8 Ma[19,20]. It is hypothesised that the radiations of Malawi and Victoria, as well as a subclade of the Tanganyika cichlid radiation (Tropheini) were seeded by a common riverine lineage, repeatedly giving rise to the similar eco-morphologies[21]. Thus, these three parallel radiations represent the spread and diversification of the 'modern haplochromines' that are famous for their species richness, morphological, and trophic diversity[21]. With a focus on (within radiations) divergent and (between radiations) convergent trophic adaptation, we sequenced the transcriptomes (mRNA) of oral jaws (OJ) and pharyngeal jaws (PJ) of juvenile larvae (developmental stage 26: yolk-sac absorbed into body cavity[22], an ecomorphologically informative developmental stage for cichlid jaws[17]) from 18 species representing corresponding trophic guilds of herbivory and carnivory from the three African Great Lakes (Figure 1B)[9,10]. Functional decoupling of cichlid oral and phryngeal jaws is thought to be a key innovation that facilitated their rapid diversification[21]. Two non-radiating species (NR) that live in lakes but did not radiate were also included in the analysis (Figure 1A, B).

## Results

### Global patterns of gene and isoform expression

To study the evolution of GE and AS during adaptive radiation, we generated transcriptomes of two tissues (OJ and PJ) from 100 cichlid fishes belonging to 20 species (Figure 1, Table S1) from three lake radiations and two species that did not make radiations (5 individuals per tissue per species, total N = 200). Reads were mapped to the phylogenetically equidistant *Oreochromis niloticus* reference genome[23] (median sequencing depth per tissue: median ~6.4 million 125 base-pair (bp) paired end reads per library; median ~80% uniquely mapped read; Table S2) using the genome-guided transcript assembly function. Gene annotations of species and tissues constructed using StringTie[24] were merged incrementally into one super-annotation[24] that was used for downstream analysis. This strategy helped to reduce false positive while increasing detection of novel and lowly expressed isoforms. Expression was quantified using established methods[24,25].

To explore global patterns of gene and isoform expression, we performed Principal component analysis (PCA) across all samples for all genes. Gene expression PC1 separated the two tissues, and on PC2, LT clustered separately from the younger radiations of LM, LV and the non-radiating

species, the latter formed an overlapping cluster (Figure 1C). PCA clustering of isoform expression, which is a consequence of AS, was also tissue-specific on PC1, but clearly lake-specific on PC2 (Figure 1D) with non-radiating species clustering closer to LM samples, particularly for the PJ. Both gene and isoform expression patterns within tissues reflect phylogenetic relationships of radiations (Figure 1B), with stronger signals emerging from the isoform expression PCA. This suggests that differences in splicing accumulate more rapidly than GE changes during adaptive radiation, emerging as clades diverge, such that global splicing patterns across species reflects their phylogenetic relationships.

**Splicing and expression have contrasting evolutionary dynamics**

Alternative splicing results in the change of relative abundance of isoforms produced from the same gene[3], which may have evolutionary consequences[2,26]. But little is known about how AS evolves at difference timescales. To adress this, we calculated percent spliced in (PSI), a standard AS metric that represents the relative abundance of an isoform as a proportion of the total GE[27,28] and computed Spearman's rank correlation coefficients (Spearman's $\rho$) across all samples and tissues for GE (all genes) and PSI (only genes with more than one isoform) to compare rate of change among these two processes. Our analysis revealed higher $\rho$ values for GE (Figure 2 boxplot, two-sided t-test $p < 0.001$). This suggets that GE was more conserved (less divergent) across samples than PSI during cichlid evolution; while AS was evolving faster. We next tested if constitutively expressed isoforms were more conserved during cichlid evolution than alternative isoforms. Indeed, we found that Spearmans's $\rho$ of PSI for alternative isoforms was significantly lower than that of constitutive isoforms (Figure S1, two-sided t-test $p < 0.001$). This fits the expectation of alternative isoforms generally evolving under weaker purifying selection[29,30].

Next, we analysed the hierarchical clustering of Spearman's $\rho$ for GE and PSI. We found that GE grouped samples by tissue for younger radiations (LV, LM) and nonradiating species (Ab, Aa). However, for the older LT Tropheini radiation, samples clustered first by clade/radiation and then by tissue within the radiation (Figure 2A), defying well-established expectations of tissue-specific conservation of GE in vertebrates[6]. In contrast, Spearman's $\rho$ of PSI clustered samples first by clade/radiation, regardless of tissue. Samples from Lake Tanganyika and nonradiating species further clustered by species, regardless of tissue (Figure 2B, S2). Evolutionarily conserved (heritable) species-specific and clade-specific changes in splicing would result in samples clustering first by species/clade and then tissue. This suggests that clade-specific divergence in splicing exceeds tissue-specific conserved signatures of splicing at a phylogenetic distance of ~2mya, and species-

specific divergence in splicing exceeds tissue-specific divergence at ~6mya. Hierachical clustering relationships generally reflected phylogenetic relationships of the radiations, however, NR species clustered with LV in the gene expresssion analysis but with LT in the PSI analysis. Interestingly, PJ samples of *H. thereuterion* from LV were placed outside of the LV clade, sister to the PJ's of the NR *A. burtoni*, which is distributed in Lake Tanganyika basin and not found in Lake Victoria. Taken together with the PCA results (Figure 1A,B), our results indicate that AS may have played an important role in the evolution of clade and species-specific jaw specialisations in comparison to GE. It may further suggest that heritable changes in splicing emerge rapidly during/near speciation in cichlid adaptive radiations, while GE changes may fine-tune morphology over longer evolutionary timescales.

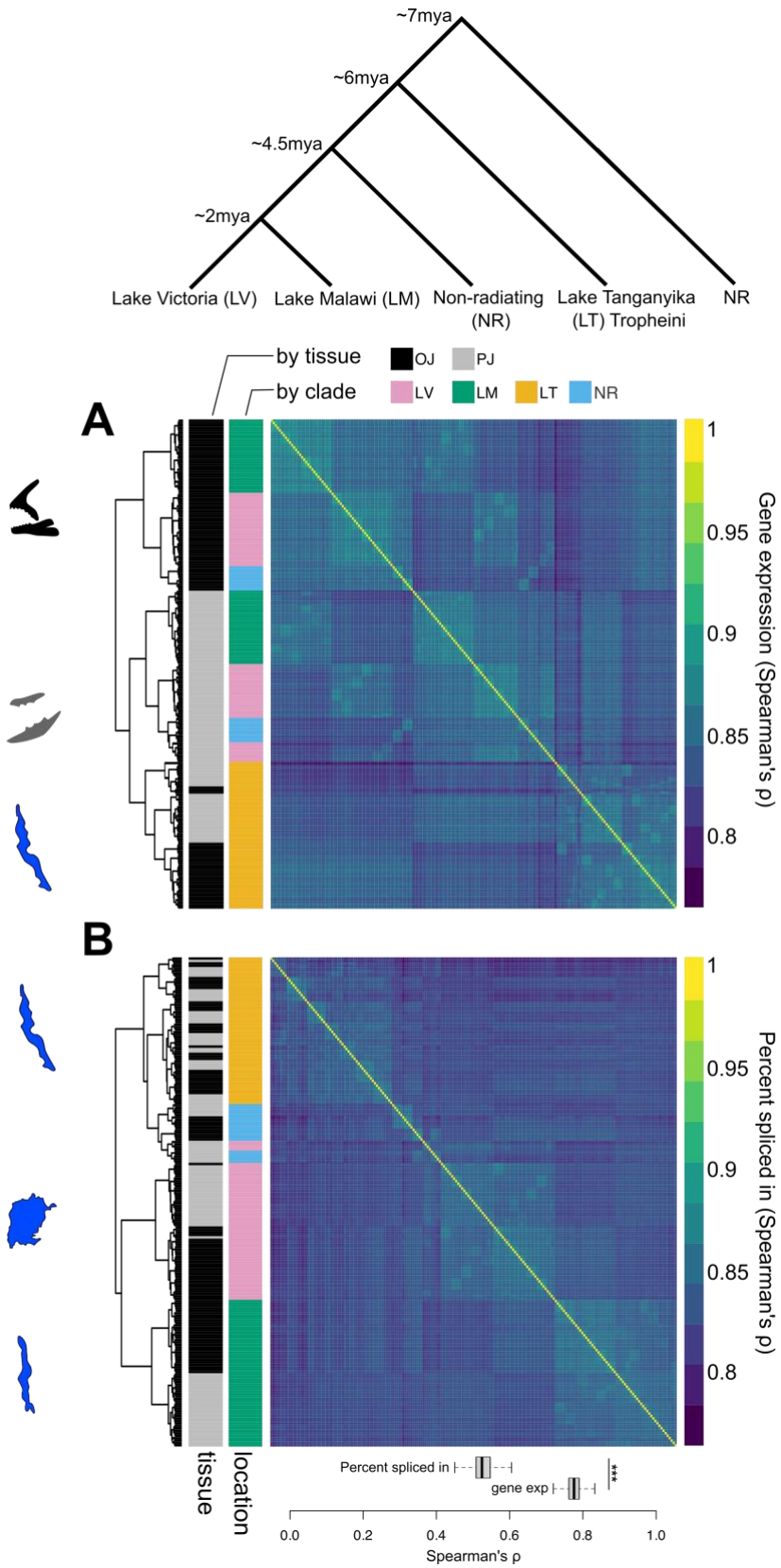
~7mya
~6mya
~4.5mya
~2mya
Lake Victoria (LV)
Lake Malawi (LM)
Non-radiating (NR)
Lake Tanganyika (LT) Tropheini
NR
by tissue
by clade
OJ
PJ
LV
LM
LT
NR
A
B
Gene expression (Spearman's ρ)
Percent spliced in (Spearman's ρ)
1
0.95
0.9
0.85
0.8
tissue
location
Percent spliced in
gene exp
***
0.0
0.2
0.4
0.6
0.8
1.0
Spearman's ρ

**Figure 2** Profiling of alternative splicing and gene expression in cichlid radiations. Hierarchical clustering of Spearman's rank correlation coefficient ($\rho$) of **(A)** normalised GE and **(B)** isoform percent spliced in (PSI) between pairs of samples (n=200). Sample clustering is represented as a tree on the left side of the heatmaps, with tissue and clade (radiations or nonradiating) of origin of species annotated in colours and the dominant cluster signature illustrated as radiation/non-radiation/jaw. Boxplots show significant difference in distribution of PSI and GE (two-sides t-test $p < 0.001$). Cladogram illustrates simplified cladogram of phylogenetic relationship of species shown in Figure S3A. Divergence times taken from Irisarri*, Singh* et al (2018)[19]. See Figure S2 for sample names.

**Genes and isoforms evolve under stabilising selection**

Next, we investigated the evolution of gene and isoform expression along the phylogeny (constructed using SNPs from the RNAseq data, Figure S3A) by fitting Brownian motion (BM) and Ornstein-Uhlenbeck (OU) processes. Ornstein-Uhlenbeck has been applied to model stochastic GE evolution towards a single optimum i.e. stabilising selection[31]. Brownian motion models neutral evolution i.e. genetic drift. The distribution of OU variance across the phylogeny, denoted 'evolutionary variance'[32], depicts how constrained expression evolution is. We found that expression of most genes and isoforms was evolving under stabilising selection (90–92%) and 8-10% of genes were evolving neutrally (chi-squared test $q$-value < 0.05, Figure S3B, File S1), as previously reported in model species[31]. However, isoform expression had significantly higher evolutionary variance (less constraint) than GE under the OU model for both jaws (Figure S3C), which would be the result of weaker stabilising selection acting on isoform expression than gene expression. This further explains the lower cross-species conservation of splicing we found above, even after taking the phylogeny into account.

**Most splice variation is ancestral, but some may be novel**

Ancestral variation is proposed to fuel rapid adaptive radiations[33]. One way that this could happen is through the adaptive sorting of ancestral gene regulation variation[34], such as splice variation. As a first step to test if ancestral splice variation contributed to trophic divergence, we quantified the number of genes and isoforms expressed uniquely in each radiation versus those shared between radiations and non-radiating species. Firstly, the total number of both genes and isoforms expressed in the OJ was higher than in the PJ; and regardless of tissue, the absolute number of genes and isoforms expressed were significantly higher in samples from the younger radiations (LM, LV) and the non-radiating lineages (NR) than the older radiation of LT (Mann-Whitney U test $p < 0.001$; Figure 3A,B; S4A,B). This pattern was also true when we analysed the number of isoforms per gene (Figure

S5), but the variance in the number of isoforms among species from each radiation increased with age of radiation. However, the difference was only significant between the PJ's of LV and LM (Mann-Whitney U test $p < 0.05$). We tested the homogeneity of variance among groups using the Barlett's test and found that significant differences for the number of isoforms per gene for PJ ($p = 0.017$) but not the OJ. This suggests that while the absolute number of isoforms expressed per gene is reduced over time, the transcriptional variance increases with time during adaptive radiation. Secondly, to understand how much of this expression variation was ancestral versus putatively *de novo*, we investigated if each gene or isoform was expressed in only one of the clades, or across several radiating and non-radiating clades (Figure 3C). The majority of genes (86.0%) and isoforms (73.0%) were expressed in at least one non-radiating species and one or more radiations – this variation is likely ancestral to all East African haplochromines. A small subset of isoforms (~12.0%) were unique to one radiation only – pointing towards the evolution of novel splice variation in the course of adaptive radiation – or less likely, repeated loss in all other lineages. It is of course possible that these putatively novel isoforms may be ancestral but the pattern is missed by our sample selection. Taken together, our data support the hypothesis that substantial transcriptomic variation existed in the common ancestor of all East African haplochromines that was likely shaped by selection and drift during adaptive radiations, with some isoforms being purged and others retained; with some evidence of putatively novel isoforms evolving in the course of adaptive radiation. But did this ancestral and novel isoform variation contribute to ecological adaptation in cichlid radiations?

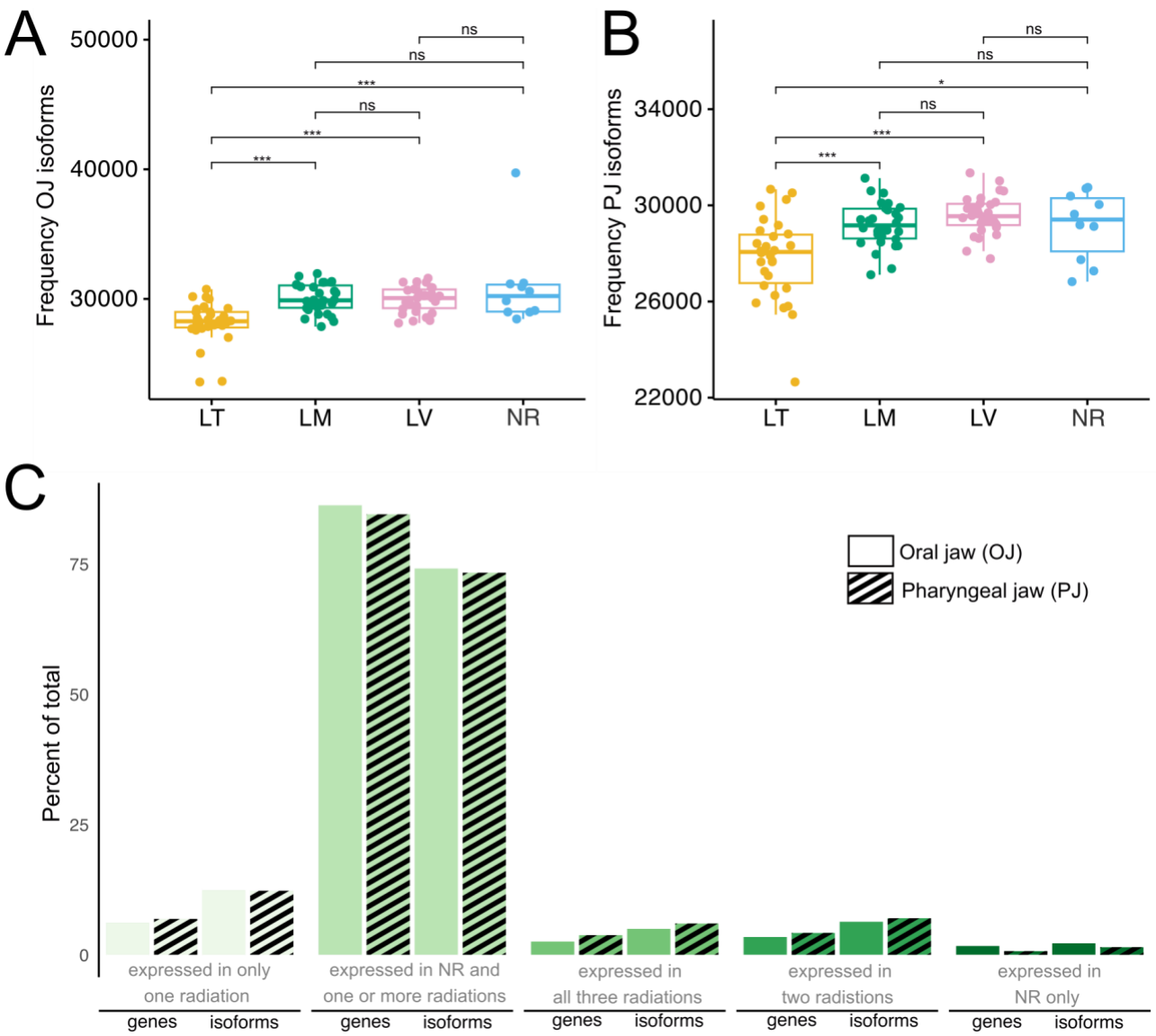


**Figure 3** Transcriptomic variation in non-radiating and radiating cichlids. **(A, B)** The absolute number of expressed isoforms identified in the oral and pharyngeal jaw mRNA of all samples. The absolute number of genes expressed is shown in Figure S4. Asterisks denote Mann-Whitney U test significance: $p < 0.001$***, $p < 0.01$**, $p < 0.05$*. NS: no significance **(C)** Percent of total genes and isoforms expressed in non-radiating species and radiations or unique to one or more of the radiations for both jaws. LV: Lake Victoria, LM: Lake Malawi, LT: Lake Tanganyika, NR: non-radiating.

### Trophic adaptation from old and new splice variants

Since the global patterns of AS revealed that it was evolving rapidly in the jaws of cichlids during adaptive radiation and that much of the splice variation was ancestral, we next asked if ancestral splice variants contributed to trophic divergence of cichlid jaws. We conducted differential PSI analysis between the jaws of species adapted to herbivorous and carnivorous diets (see Figure 1 for

trophic specialisation of species) within each lake to identify differentially spliced genes (DSGs) that may play a role in shaping their eco-morphological differences. The largest number of genes were differentially spliced in LV (OJ 375, PJ 259) followed by LM (OJ 294, PJ 266) and LT (OJ 250, PJ 150) (Figure 3A for OJ, Figure S6 for PJ, File S1). This indicates that younger cichlid radiations leveraged splice variation relatively more than the older radiation in LT for trophic adaptation. Differentially spliced genes were significantly enriched (Fisher's exact test $q < 0.05$) for GO terms associated with craniofacial development such as 'regulation of osteoblasts', 'bone mineralisation involved in bone maturation', 'roof plate formation', and 'skeletal muscle development' (subset shown in Figure 3C; full data in Figure S7). The Wnt signalling pathway that has been associated with the evolution of new craniofacial phenotypes at the larval stage in Malawi cichlids was enriched in divergent OJ and PJ from LV[35]. It is likely that AS of Wnt pathway genes *gpc3* and *lzts2a* (File S1) that control skeletal development and dorsoventral patterning contributed to adaptive jaw shape differences LV cichlids[36,37]. Two genes, *rnfox1* and *rbm5*, involved in 'mRNA splicing via the spliceosome' were differentially spliced in LV oral jaws (File S1). This implicates the spliceosome, the molecular machinery that controls AS, in cichlid adaptation. In LM OJ, differential splicing of genes involved in bone maturation (*col1a1*) and cranial morphogenesis (*sp7*) were notable. Overall, twice as many DSGs were shared between LV and LM, than either of these lakes shared with LT (Figure 4A, S6), which is most consistent with these splice variants evolving and/or gaining function after the ancestor of these LT haplochromines had split from the common ancestor of LV and LM, in line with the 'Out of Tanganyika' theory of haplochromine cichlid evolution[21].

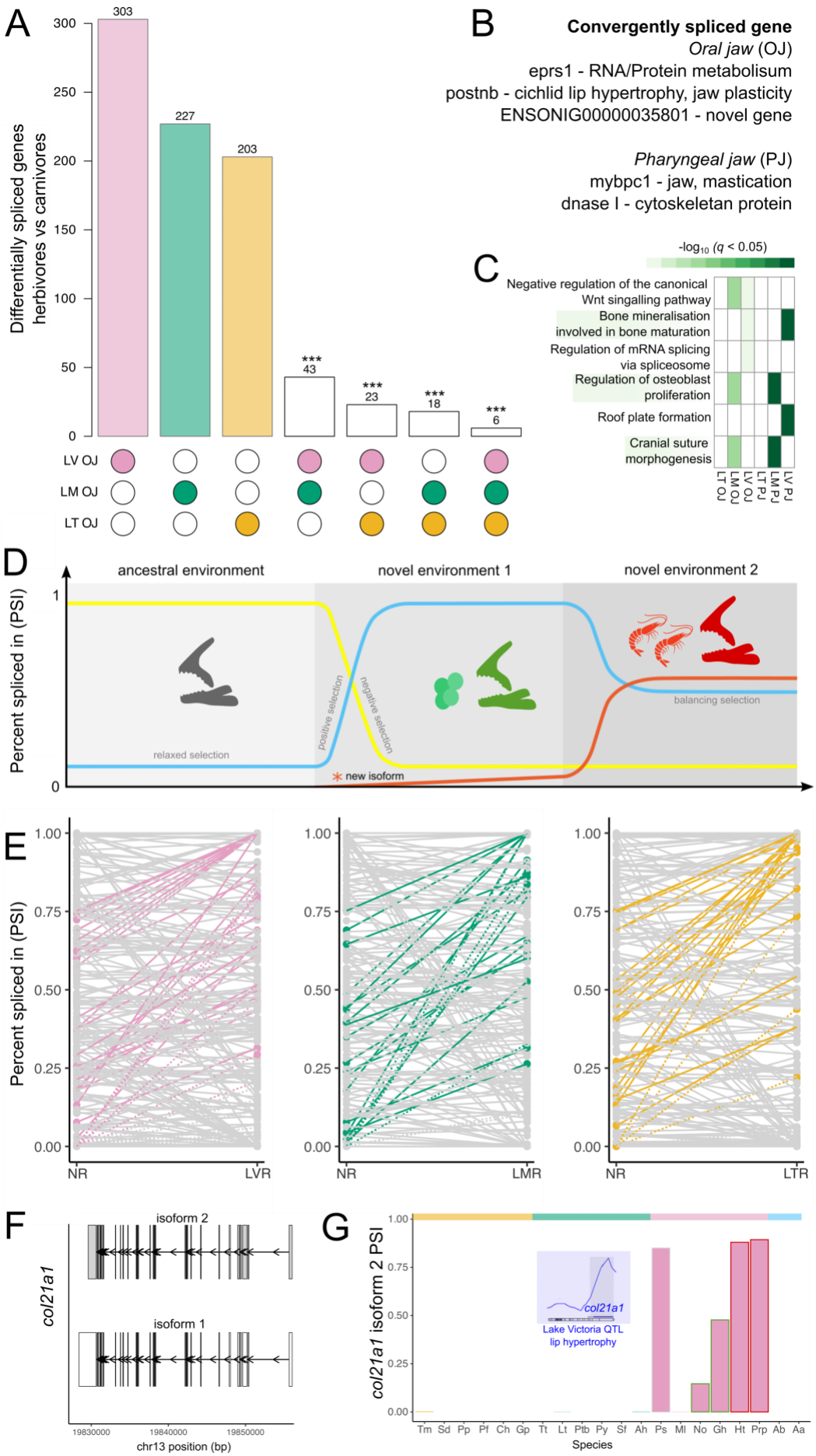
A
Differentially spliced genes herbivores vs carnivores
303
227
203
***
43
***
23
***
18
***
6
LV OJ
LM OJ
LT OJ
B
Convergently spliced gene
Oral jaw (OJ)
eprs1 - RNA/Protein metabolisum
postnb - cichlid lip hypertrophy, jaw plasticity
ENSONIG00000035801 - novel gene
Pharyngeal jaw (PJ)
mybpc1 - jaw, mastication
dnase I - cytoskeletan protein
C
-log10 (q < 0.05)
Negative regulation of the canonical Wnt singalling pathway
Bone mineralisation involved in bone maturation
Regulation of mRNA splicing via spliceosome
Regulation of osteoblast proliferation
Roof plate formation
Cranial suture morphogenesis
LT OJ
LM OJ
LV OJ
LT PJ
LM PJ
LV PJ
D
Percent spliced in (PSI)
ancestral environment
novel environment 1
novel environment 2
relaxed selection
positive selection
negative selection
balancing selection
new isoform
E
Percent spliced in (PSI)
NR
LVR
LMR
LTR
F
col21a1
isoform 2
isoform 1
chr13 position (bp)
G
col21a1 isoform 2 PSI
col21a1
Lake Victoria QTL
lip hypertrophy
Tm Sd Pp Pf Ch Gp Tt Lt Ptb Py Sf Ah Ps Ml No Gh Ht Prp Ab Aa
Species

**Figure 4** Evolutionary dynamics of differentially spliced genes between herbivores and carnivores species. **(A)** Number of significantly differentially spliced genes between oral jaws (OJ) of carnivores and herbivores within each lake radiation ($q < 0.05$) with overlaps shown (hypergeometric test $p < 0.001$). The number of differentially spliced genes between the pharyngeal jaws (PJ) of carnivores and herbivores within each radiation is shown in Figure S6A. **(B)** Genes that were convergently spliced in at least two radiations in the same direction (higher in herbivores or higher in carnivores). No gene were convergently spliced in the same directly in all three radiations. **(C)** Gene Ontology enrichment of differentially spliced genes between herbivores and carnivores. Only a subset of terms are shown. Full dataset is presented in Figure S7. **(D)** Hypothetical expectations for isoform frequency (measured as percent spliced in - PSI) during ecological adaptation. Figure adapted from Singh and Ahi (2022). **(E)** Testing the hypothesis depicted in (D) in our transcriptome data from Lake Victoria radiation (LV), Lake Malawi radiation (LM) and Lake Tanganyika radiation (LT) compared to non-radiating species (NR). Mean PSI was calculated all individuals of all species in each group (NR, LV, LM, LT). All lines represent all isoforms of genes that were signficantly differentially spliced between herbivores and carnivores within each radiation. Solid grey lines represent all isoforms that were expressd at lower or similar levels in the NR species and species from each radiation (PSI differences <= 0.2). Dotted grey lines represent all isoforms that were not expressd in NR species but were expressed species from each radiation (PSI differences <= 0.2). Coloured solid lines indicate isoforms that were expressed at low levels in NR species but were higher expressed in radiating species (PSI differences > 0.2). Coloured dotted lines indicate isoforms that were not expressed in NR species but were expressed in radiating species (PSI differences > 0.2). **(F)** The two isoform models for *col21a1* in the gene annotation. **(G)** Species mean expression of *col21a1* isoform 2 across all species in our dataset. Inset highlights a known QTL lip shape from LV containing *col21a1*.

Of the genes that were significantly differentially spliced between herbivores and carnivores in each lake (Figure 3A, S6A), a dynamic pattern of isoform evolution emerged: many of these were low expressed isoforms in NR species that increased in expression in radiating species (Figure 3E for OJ, S6B for PJ). An increase in frequency would be the expectation if these putatively ancestral isoforms gained adaptive function in response to newly available food sources when the ancestor entered lakes (Figure 3D), as hypothesised by Singh & Ahi[2]. Many highly expressed isoforms in non-radiating species also decreased in frequency or were lost in the lacustrine radiations (presumably as the alternative isoform increased in frequency). We further identified 34 potentially novel isoforms that were absent in non-radiating species and only expressed a subset of the three adaptive radiations (Figure 3E dotted lines, File S1). Sixteen of these 34 novel isoforms were found in LV, followed by 12 in LM and six in LT. Novel isoforms of note were of genes *col21a1* and *lzts2a*. *Col21a1* had two isoforms (Figure 3F) of which the main isoform is expressed in the OJ of LV herbivores as well as

herbivores from LT and LM (with one exception, File S1). The alternative isoform of *col21a1* is only expressed in LV species, with higher expression in carnivores and omnivores (Figure 3G). *Col21a1* is associated with human cleft lip syndrome and maps to a QTL for hypertrophic lips in LV cichlids[38]. It is conceivable that this isoform may have arisen *de novo* in LV cichlids or arrived through hybridisation, which has been shown to contribute genetic variation for rapid adaptive radiation in Lake Victoria[39]. In the future, integrating transcriptomic analysis with population genetics and functional genomics (GWAS, spliceQTL) approaches would help to narrow down the variants associated with splicing and the evolutionary processes that shaped them. Our results also highlight the conservation of craniofacial development pathways across vertebrates, and the promise of cichlid fishes as a natural model for studying human craniofacial disease[40].

**Differential expression of major craniofacial genes contributed to jaw divergence**

In addition to differences in splicing, we investigated GE differences between herbivores and carnivores in each lake radiation to be able to compare the role of GE with AS in trophic adaptation. The largest number of differentially expressed genes (DEGs) were in LT (OJ 6716, PJ 4871), followed by LM (OJ 2935, PJ 2133) and LV (OJ 2619, PJ 1649) (Figure 5A for OJ, Figure S8 for PJ, File S1). Several interesting candidate genes for craniofacial development stood out as the most extreme outliers, particularly in LT where GO terms such as 'neural crest cell development', 'skeletal system development, and 'ossification' were significantly enriched ($q < 0.05$). One such gene for LT OJ was Frizzled 2 (*fzd2*), a Wnt pathway gene that is highly expressed in the developing face of vertebrates, and mutations of this gene cause cleft palate and wider nose and upper jaw[41]. A gene controlling prey crushing teeth morphology in cichlids, *odam*[42], and a gene linked to craniofacial disease, *gja1,* were also significant for LT OJ[43]. Two interesting outliers in LV OJ were *kaznb*, a craniofacial development gene that has been found to be upregulated in the jaws of herbivorous cichlids at larval and adult life stages[44] and *egr1*, that forms part of the regulatory cascade modulating cranial bone morphogenesis through the BMP pathway[45]. Interestingly, one of the most extreme outliers for LT OJ was *hnRNPk*, an RNA binding protein that regulates pre-mRNA splicing[46].

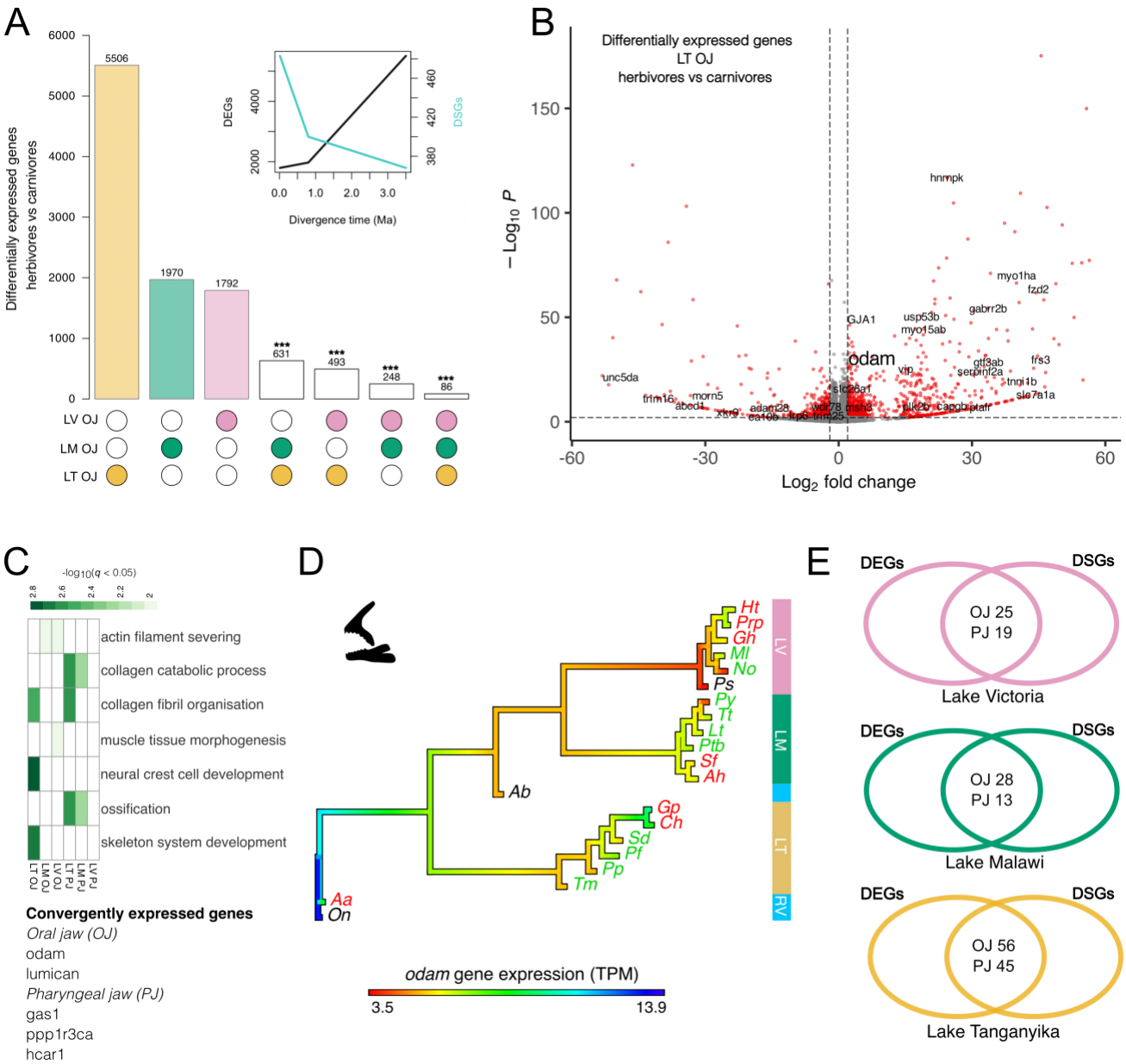
A
Differentially expressed genes
herbivores vs carnivores
5506
1970
1792
631
493
248
86
LV OJ
LM OJ
LT OJ
DEGs
DSGs
Divergence time (Ma)
B
Differentially expressed genes
LT OJ
herbivores vs carnivores
odam
Log2 fold change
C
actin filament severing
collagen catabolic process
collagen fibril organisation
muscle tissue morphogenesis
neural crest cell development
ossification
skeleton system development
Convergently expressed genes
Oral jaw (OJ)
odam
lumican
Pharyngeal jaw (PJ)
gas1
ppp1r3ca
hcar1
D
Ht
Prp
Gh
Ml
No
Ps
Py
Tt
Lt
Ptb
Sf
Ah
Ab
Gp
Ch
Sd
Pf
Pp
Tm
Aa
On
LV
LM
LT
RV
odam gene expression (TPM)
3.5
13.9
E
DEGs
DSGs
OJ 25
PJ 19
Lake Victoria
OJ 28
PJ 13
Lake Malawi
OJ 56
PJ 45
Lake Tanganyika

**Figure 5 GE divergence in oral jaws of herbivore and carnivore species. (A)** Number of differentially expressed genes (DEGs, $q < 0.05$) between oral jaws (OJ) of herbivores and carnivores in each lake, with overlaps annotated (hypergeometric test $p < 0.001$). Results for pharyngeal jaws can be found in Figure S8. Inset shows the relationships between number of DEGs and differentially spliced genes (DSGs) in each cichlid radiationnplotted as function of divergence time. **(B)** Volcano plot illustrating DEGs in Lake Tanganyika (LT) oral jaws. Significant DEGs are highlighted in red. Top gene outliers are annotated. **(C)** Gene Ontology enrichment of DEGs between herbivores and carnivores. Only a subset of terms are shown. Full list is presented in Figure S9. Some candidate genes for craniofacial development that were convergently expressed in all three radiations in the same direction (higher in herbivores or higher in carnivores). See File S1 for full list. **(D)** Ancestral state reconstruction of GE of the convergently expressed *odam* gene. **(E)** Overlap of DEGs and DSGs between herbivores and carnivores within each lake radiation.

**Contrasting functions for AS and GE**

We found dissimilar patterns in the extent of differential GE and AS underlying divergent trophic adaptation across cichlid radiations. Even though the overall number of DEGs was much higher than DSGs across all three radiations (Figures 4A, 5A, S6, S8), the relative number of DEGs across radiations was the highest in the oldest radiation and lowest in the youngest radiation, while the DSGs showed the opposite pattern (Figure 5A inset). Interestingly, more DEGs in each of the younger radiations overlapped with the older radiation, in contrast more DSGs overlapped among the younger radiations (Figures 4A, 5A, S6A, S8). This pattern suggests that splicing changes evolving rapidly to shape jaw divergence during speciation, while GE differences evolve much slower, mostly likely fine-tuning the adaptive morphology long after speciation.

It has been proposed that AS and GE perform contrasting biological functions[17,44,47–49]. However, these studies have lacked a phylogenetic perspective. In our dataset, the overlap of genes that were both differentially spliced and differentially expressed between herbivores and carnivores followed a phylogenetic and evolutionary pattern. The highest overlap was in LT (OJ 56, PJ 45), followed by LM (OJ 28, PJ 13) and LV (OJ 25, PJ 19) (Figure 5, S9). Our results reinforce the idea that putatively adaptive AS and GE changes act mostly at different genes – but we show that the extent of gene overlap depends on the evolutionary age of lineages, with fewer genes being regulated by both procesess in more recently diverged populations.

**Convergent GE and splicing underlying trophic evolutioon**

African cichlid radiations are an ideal sytem to study the repeatability of evolution as they have evolved convergent trophic adaptations, particularly in jaw and teeth shape, to exploit similar food sources in similar ways in different lakes[50]. We asked if differentially expressed and spliced genes contributing to herbivory vs carnivory showed parallel divergence across all three radiations. We find DEGs in both oral (23 upregulated, 12 downregulated in herbivores) and pharyngeal jaws (14 upregulated, 6 downregulated in herbivores) that fit this pattern (Figures 5, S8). Several convergently used OJ genes have been associated with the development of teeth (*odam*), bone (*lum*) and muscle *(myrip*, *myo15aa*) (Figure 5B,C, File S1). *Odam* expression has been associated with larger vs smaller teeth in cichlid fishes and our ancestral state reconstruction suggests that this gene increased in expression repeatedly in the oreal jaws of carnivores in each lake radiation (Figure 5D). Convergent genes for PJ were *hcar1*, *ppp1r3c*, myl13, and *gas1*. *Gas1* interacts with *shh* is early face development[51]. We validated the convergent expression of *odam, gas1, hcar1,* and *ppp1r3ca* using qPCR (Figure S10). No DSGs were convergently spliced in the same direction in all three radiations. We only found evidence of convergent splicing in the same direction in two of the three radiations in the following noteworthy genes: OJ: *eprs1*, *postnb* and PJ: *mybpc1*, *Dnase1* (Figure 4A, B).

**ALX and PAX transcription factors are key drivers of cichlid jaw modularity**

Modularity of cichlid oral and pharygeal jaws is considered a key innovation that decoupled the task of prey capture (OJ) and processing (PJ)[52]. Little is known about the gene regulatory basis of this functional decoupling[16]. Differential GE analysis between the two jaw tissues within each lake revealed that the majority of DEGs (5167) contribute to jaw divergence repeatedly in all three lakes, with 2555 being upregulated and 2591 downregulated in the PJ (Figure 6, File S1). These parallel expressed genes are associated with 'skeletal muscle fibre/tissue development' and 'sarcomere organisation' GO terms ($q < 0.05$, File S1). Notable genes with parallel expression were ALX skeletogenesis transcription factors (*alx1*, *alx4a*, *alx4b*) that were upregulated in OJ versus PJ. PAX transcription factor family genes also had parallel expression patterns (upregulated OJ: *pax3a*, *pax3b*, *pax6b*; upregulated PJ: *pax1a*, *pax5*) (Figure 6, File S1). PAX genes are master regulators of early tissue development throught neural crest induction[53] and pax3a/b are expressed in a modular fashion in some cichlid species[16]. Many HOX (*hoxa4a*, *hoxa5a*, *hoxb2a*, *hoxb3a*, *hoxb4a*, *hoxb5b*, *hoxb6b*, *hoxb8*, *hoxc5a*, *hoxc6a*, *hoxc9a*, *hoxd10a*, *hoxd3a*, *hoxd9*) and Troponin (*tnni1b*, *tnni2a*, *tnnt1*, *tnnt2c*, *tnnt2d*, *tnnt3a*) genes also showed signatures of parallel expression. Hox genes are implicated in cichlid dentition[54] and bone development[55]. Furthermore, many OJ vs PJ DEGs were unique to each lake radiation (LV 1900, LM 1730, LT 1176).

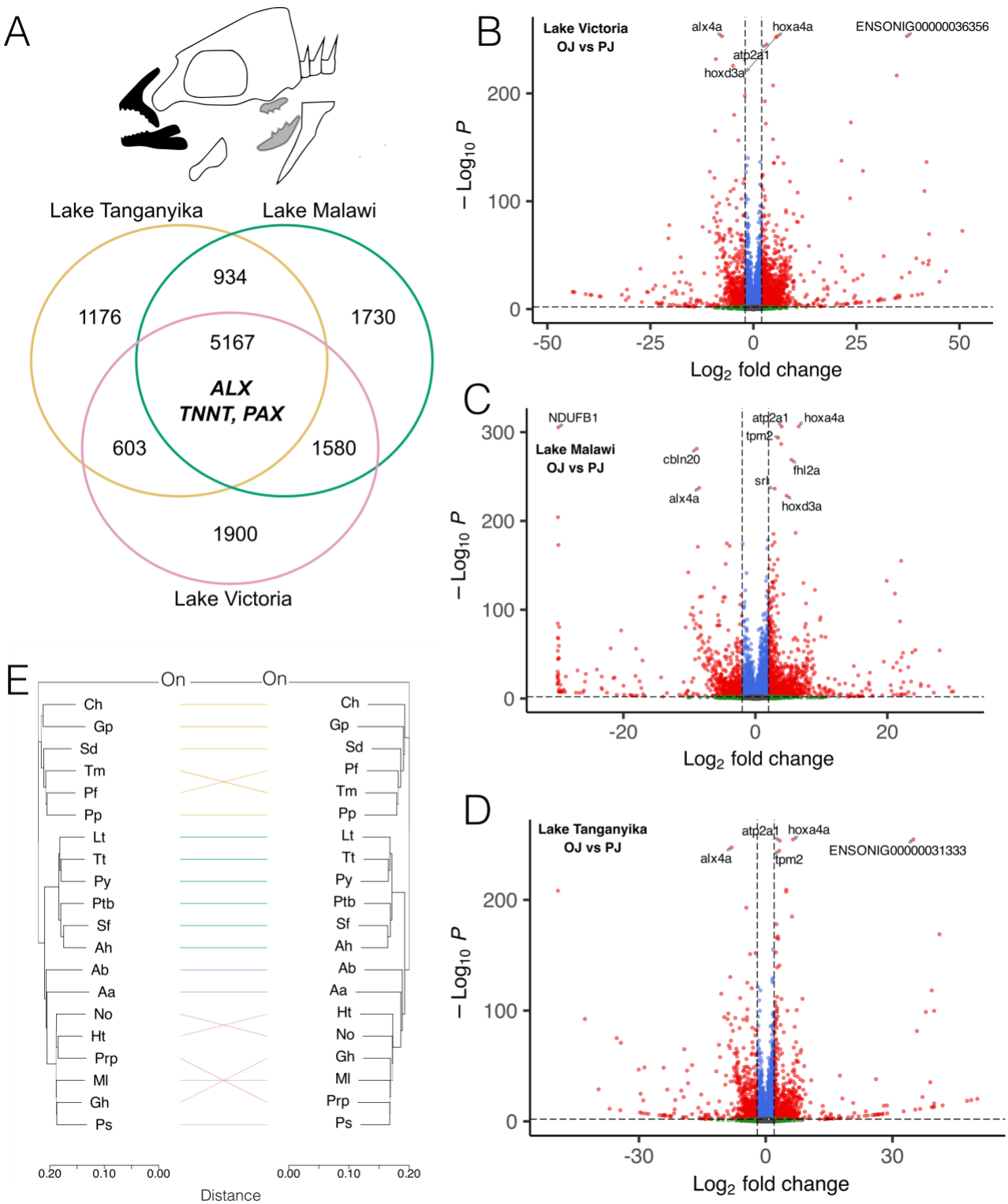
A
Lake Tanganyika
Lake Malawi
934
1176
1730
5167
ALX
TNNT, PAX
603
1580
1900
Lake Victoria
B
Lake Victoria
OJ vs PJ
alx4a
hoxa4a
ENSONIG00000036356
atp2a1
hoxd3a
– Log10 P
Log2 fold change
C
NDUFB1
Lake Malawi
OJ vs PJ
cbln20
alx4a
atp2a1
hoxa4a
tpm2
fhl2a
sri
hoxd3a
D
Lake Tanganyika
OJ vs PJ
atp2a1
hoxa4a
alx4a
tpm2
ENSONIG00000031333
E
On
On
Ch
Gp
Sd
Tm
Pf
Pp
Lt
Tt
Py
Ptb
Sf
Ah
Ab
Aa
No
Ht
Prp
Ml
Gh
Ps
Distance

**Figure 6 Transcriptional modularity of cichlid oral (OJ) and pharyngeal jaws (PJ). (A)** Venn diagram showing overlap of differentially expressed genes ($q < 0.05$) between oral and pharyngeal jaws of species within each lake radiation. **(B, C, D)** Volcano plots illustrating DEGs in Lake Tanganyika (LT) and Lake Victoria (LV) between oral and pharyngeal jaws. Significant DEGs are highlighted in red. Top gene outliers are annotated. **(E)** Tanglegram of GE phylogenies of OJ (left) and PJ (right). Discordant tips are annotated with lines.

To further understand gene regulation evolutions of the jaws, we contrasted their transcriptome-wide GE phylogenies of OJ and PJ for all species (Figure 6). Topological discordance in the phylogenies were interpreted as evidence of decoupled GE evolution. Applying this interpretation, evidence of decoupled GE evolution between OJ and PJ was found only in LV and LT cichlids, but not in LM. In LV, OJ of the two piscivores (Ht and Prp) clustered together, but the PJ of Prp clustered with the omnivore (Ps). Similarly, in LT the OJ of the two algae grazers (Pp, Pf) clustered together, but the PJ of Pp clustered with the algae browser Tm. These patterns are consistent with prey capturing being functionally decoupled to prey prcoessing.

## Discussion

The role of AS in ecological adaptation has not been explored across broad phylogenetic timescales, in the context of adaptive radiation. Through the analysis of 200 transcriptomes representing three independent but phylogenetically nested adaptive radiations that evolved in the last <16,000 to 3.8 million years we show that the cache of ancestral transcriptomic variation in East African haplochromine cichlids was shaped over time presumably through selection[56] and stochastic processes[57] as lineages diversified. These evolutionary processes most likely led to the faster evolution of AS than GE that we observe here. Signatures of splicing became increasingly more species-specific, and GE became more lineage-specific over evolutionary time. However, specific-specific signatures of splicing did not emerge until ~3.8 million years of evolution, a pattern not observed in previous large-scale studies on older lineages[5,6]. This rapid turnover of AS is consistent with relaxed selection pressures on splice-sites, as alternative isoforms can escape the effects of purifying selection as long as the major isoform is expressed[29]. Theory suggests that relaxed selection can increase the rate of evolution of complex adaptations in heterogenous environments, allowing populations to rapidly cross adaptive valleys[58]. Relaxed selection maintaining standing variation, followed by periods of positive selection on isoforms of craniofacial genes may have contributed to the rapid evolution of trophic adaptations in cichlids as they entered lakes and were exposed to new and ecologically more stable environments.

Our results on the evolutionary dynamics of GE are partially consistent with previous work on transcriptome evolution at deeper timescales. In contrast to AS, GE is largely conserved at the tissue-level during vertebrate evolution[5,6,13,31]. Remarkably, we found lineage-specific clustering of GE in Tropheini jaws. Tropheini are named after their diverse trophic adaptations that allowed them to enter the littoral zones of Lake Tanganyika at a much later stage and successfully diversify alongside other established cichlid radiations[20,59,60]. They were able to do this by rapidly remodelling their jaw, dentition, and craniofacial ecomorphology to forage foods in new and different ways[59]. Our findings may reflect the particular importance of gene regulation for evolutionary transitions in Tropheini jaws during speciation. Furthermore, we find that the parallel modular expression of thousands of genes may underlie the differences between the functionally decoupled OJ and PJ across all three cichlid radiations. Of note were genes from the *alx* craniofacial skeletogenesis transcription factor family (*alx1*, *alx4*), of which *alx1* is associated with rapid diversification of beak shape in the adaptive radiation of Darwin's finches[61]. Bird beaks are a specialised modification of the upper and lower oral jaws found in all jawed vertebrates. It is conceivable that *alx* transcription factors are particularly responsive to environmental change and mediate large effect changes in jaw bone shape.

We found that both rapidly evolving AS and GE contributed to jaw adaptations in herbivorous and carnivorous cichlids, but their contribution was inversely predicted by evolutionary age. Relatively more differences in AS were found to underlie trophic differences in jaws of cichlids from the younger radiations, while GE differences dominated in the older radiation. This would suggest that rapidly evolving AS from standing variation is a substrate for rapid adaptation[17] as species diverge and may be one piece of the puzzle explaining how ecological and species richness can evolve when suffifient time is not available for new mutations to arise, fix, and be replenished. In contrast, ecologically important gene expression changes emerge at later stages of speciation. Consistent with previous studies, we found many genes implicated in jaw divergence suggesting that jaw remodelling is a complex phenotype governed by myriad genes[16,20,50,62]. The Wnt pathway emerged as a key player in the evolution of new jaw phenotypes at early stages of development, with genes regulated through splicing (*gpc3*, *lzts2a*) and differential expression (*fzd2*). Overall, more convergent differences in GE than AS were found across all three cichlid radiations. We propose that the lower cross-species and -radiation conservation of AS may underscore its functional role in shaping lineage and species-specific adaptations at short timescales.

It has been hypothesised that ancestral variation plays an important role in the rapid adaptive capacity of African cichlids[33,63]. But, little empirical evidence exists explaining how this ancestral variation mechanistically plays out. We show that ancestral splice variants (also lowly expressed in non-

radiating species) that were differentially spliced between jaws of herbivores and carnivores increased in frequency in radiating cichlids, most likely enabling trophic divergence. First, this suggests that old genetic variation can fuel new adaptations[33]; and that the acestors of the East African cichlid adaptive radiations, which were most likely a widely dispersing riverine cichlids, acted as reservoirs of genetic variation for diversification on a continental scale[64]. And even if most ancestral alternative isoforms are non-functional[65], lowly expressed, and evolving stochastically[29], they represent a cache of standing variation for adaptive evolution when ecological opportunity arises[2,66]. Standing variation in AS has been shown to be important for the rapid domestication of sunflowers[66], which represents only one ecological transition, while cichlid adaptive radiations involve multiple, often convergent, diverse ecological transitions. In addition to ancestral variation, we found evidence of potentially novel isoforms contributing to adaptive jaw differences. This suggests that AS itself can evolve during adaptive radiation, generating new genetic material for evolutionary change at short timescales. Taken together, our findings may support the flexible-stem model of adaptive radiation, whereby standing variation in the ancestral lineage facilitates plastic responses in new environments, that become genetically assimilated and heritable over time[67].

## Materials and Methods

### Study design and sampling

In this study we focused on six cichlid species from Lake Tanganyika, six cichlid species from Lake Malawi, six cichlid species from Lake Victoria as well as two cichlid species from surrounding rivers, all of which are part of the 'modern haplochromine' adaptive radiations (Figure 1). The 18 species from the three lakes represent eco-morphological replicates of herbivore and carnivore trophic adaptations. We focused on the functionally decoupled oral (OJ) and pharyngeal jaw (PJ) tissues, that are involved in food capture and processing.

Sample were collected by either through aquarium trade or fieldwork within the framework of collaborations with Zambian and Tanzanian fisheries (TAFIRI). All fish were reared in the fish facility at University of Graz under standardised conditions regarding tank environment, diet, and water parameters to minimise plastic effects. Fertilised eggs were taken from the mouth of mouthbrooding females 24 hours after fertilisation, by applying slight pressure onto their cheeks. The eggs were reared in a separate tank until the yolk sac was completely absorbed, which is also known as the developmental stage 26[22]. Stage 26 larvae were euthanised and stored in RNAlater in a fridge for seven days and then moved to a -80°C freezer. Stage 26 was selected as it is an ecomorphologically informative developmental stage for cichlid with different trophic specialisations[16,17,59].

**RNA extraction**

Each larva was dissected under a stereo microscope to dissect the oral jaws (OJ: bone, teeth) and the pharyngeal jaws (PJ: bones and teeth) separately (Figure 1). The bones were cleaned as much as possible, but some surrounding tissue (i.e. skin, cartilage, muscle) remained, especially for the OJ, as small size of the larvae (<1cm) makes the dissection challenging. The dissected tissue was submerged in a lysis buffer and 1-thioglycerol according to the protocol of the Promega RNA Reliaprep kit. The sample in the lysis buffer was crushed with a ceramic bullet in a homogenizer (FastPrep-24, MP Biomedicals, Santa Ana, CA, USA) before further processing. Two individuals were pooled per sample to increase the amount of RNA. For Lake Malawi and Tanganyika, one individual yielded sufficient RNA. RNA quality was tested at a TapeStation 2200 with RNA Screen Tapes. It was tried to reach a RINe number of >= 7 for each sample. We extracted RNA for 5 biological replicates per jaw per species, which resulted in 200 mRNA samples in total.

**Library preparation and sequencing**

Library preparation was done with the TruSeq Stranded mRNA Sample Prep Kit using 24 indexing adapters and following the standard protocol. For the input RNA we tried to reach 1000ng for each sample. The quality of the cDNA was tested on a TapeStation 2200 with D1000 Screen Tapes. Sequencing was executed at the Vienna BioCenter Core Facilities on a HiSeq 2500 with paired end 125 cycles per read (2 x 125 cycles). Demultiplexing was conducted by the same facility.

**Transcriptome assembly**

Each sample has approximately ten million paired-end reads. Raw sequence reads are available at NCBI sequence read archive (Reviewer link: https://dataview.ncbi.nlm.nih.gov/object/PRJNA640176?reviewer=p2dhiiern60seqodpuqbqb1o62). After a quality check with Fastqc (v0.11.8)[68] and a trimming step with Trimmomatic (v0.3.9)[69], only reads with a phred > 28 and a minimum length of 70 bp where retained. To enable the assembly of novel isoforms we developed a bioinformatics pipeline with the following steps: Reads were assembled using STAR (v2.7.3.a)[70] in reference guided mode using the *Oreochromis niloticus* reference genome[71]. This genome was selected because it is the best assembled and annotated cichlid reference genome and is phylogenetically equidistant to the species sampled in this study. Mapping statistics were generated with samtools idxstats (v1.9)[24,72]. StringTie (v2.0.6) was run first both with a reference genome to assemble the RNA-Seq alignments into potential transcripts and in

genome-guided mode. This was done separately for the single files (per individual). The single gtf files for each biological replicate were then incrementally merged according to species and tissue and finally into one super annotation. This repeated merging steps were conducted to reduce the probability of false positives in isoform assembly and accurately identify novel species-specific or lower expressed isoforms. To estimate the accuracy of the produced annotation files we compared them with gffcompare (v0.11.2)[73] to our reference *O. niloticus* annotation. We filtered for monoexonic transcripts not contained in the reference and a class code assigned by gffcompare (v0.11.2), indicating 'possible polymerase run-on' fragments. The maximum intron length was reduced to 200,000 bp, which is the maximum intron length found in the *O.niloticus* reference. Based on the super annotation, the expression estimates were generated with StringTie allowing no multimapping. From these expression estimates, count matrices were produced using a perl script from the griffith lab (https://rnabio.org) to extract raw count data from StringTie results. The code used for our RNAseq pipeline in on github[74]. SUPPA2 (v2.3)[28] was used to calculate percent spliced in (PSI) index for each isoform for each gene.

**Differential GE analysis**

We used DESeq2[25]in R[75] to detect DEGs running comparisons between all species and tissues as well as comparisons between the tissues within a species. DESeq2 uses raw read counts and estimates variance-mean dependence based on a model that utilises the negative binomial distribution. The cutoff for DEGs was chosen at a false discovery rate of ($q < 0.05$).

**Alternative splicing analysis**

Alternative splicing analysis was conducted with SUPPA2 (v2.3)[28] to evaluate differential transcript usage (DTU) between the samples as well as differential exon usage (DEU) to detect the mode of splicing. Visualization of the analysis was conducted in R[75].

**Modeling gene and isoform expression**

We used the methods proposed by Chen et al (2019)[32] to model gene and isoform expression along the phylogeny using Ornstein-Uhlenbeck (OU) and Brownian Motion (BM) processes. The OU process is a modification of a random walk, describing the change in expression ($dX_t$) across time (dt) by $dX_t = \sigma dB_t + \alpha(\theta - X_t)\, dt$, where $dB_t$ denotes a Brownian motion process. The model elegantly quantifies the contribution of both drift and selective pressure for any given gene: (1) Drift is modeled by Brownian motion with a rate σ while (2) the strength of selective pressure driving expression back

to an optimal expression level θ is parameterized by α. The OU process accounts for phylogenetic relationships, thus allowing us to fit individual evolutionary expression trajectories.

**Phylogenetic analysis and ancestral state reconstruction of gene expression**

We constructed expression trees using the neighbour-joining approach based on pairwise distance matrices between samples using data from all genes and applying functions from the 'ape' package (v5.8.1)[76] in R[75]. Distances were computed as 1 − ρ, where ρ is the Spearman's correlation coefficient, chosen due to its robustness against outliers and any normalisation issues[31]. To visually evaluate the differences in expression evolution, we constructed a tanglegram of the expression phylogenies of the two jaw tissues using the 'dendextend' package (v1.19.0)[77] in R. Ancestral state reconstruction was conducted using the 'phytools' package (v2.4.4)[78] in R.

**Phylogenetic reconstruction (SNPs)**

SNP calling on mapped mRNAseq reads was conducted using bcftools[79] across all samples. Filtered high quality SNPs were used to contruct a concatenated SNP phylogeny using RAxML (v8) [80]

**GO term enrichment**

GO terms for *Oreochromis niloticus* were acquired in R via the biomaRt package (v2.46.1)[81] from ensembl[82]. Gene set enrichment analysis was conducted using topGO (v2.36.0)[83] using the method weight to account for GO topology and Fisher's exact test to correct for multiple testing for the enrichment analysis. We used our super annotation to build the gene universe.

**qPCR analysis**

From 4 replicates of RNA samples of each jaw, stage and species, we used 500ng of RNA to synthesize first strand cDNA using High Capacity cDNA Reverse Transcription kit (Applied Biosystems). The cDNA from each sample was diluted 1:10 times in nuclease-free water to be used for qPCR. We used *rpl18* as a reference gene, which we have already validated as a stably expressed gene in both cichlid jaws across different trophic niches[14] to normalize our expression data for qPCR. The primers were designed at conserved sequence regions of the coding sequences from all species using CLC Genomic Workbench (v7.5 CLC Bio, Aarhus, Denmark) for alignment and delineation of exon boundaries using *A. burtoni* genome[12]. Primer Express 3.0 (Applied Biosystems, CA, USA) and OligoAnalyzer 3.1 (Integrated DNA Technology) were used to design the primers as described in[14,84].

Maxima SYBR Green/ROX qPCR Master Mix (2X) (Thermo Fisher Scientific, Germany) was used to generate qPCR reactions. The amplification steps were conducted in 96 well-PCR plates on ABI 7500 real-time PCR System (Applied Biosystems). The primer efficiency calculation, qPCR program, plate set-up and a dissociation step were performed as described in[14,84]. To normalize the expression levels, the Cq value of *rpl18* was used as normalization factor (Cq reference), and ΔCq of each target gene was calculated accordingly (ΔCq target = Cq target – Cq reference). Relative expression quantities (RQ) were calculated for the normalized values using E−ΔΔCq and then fold difference values were calculated by transformation of RQ values to logarithmic values in order to conduct further statistical analysis. The statistical expression differences were identified using ANOVA statistical tests, followed by Tukey's HSD post hoc tests.

## Data Availability

All data has been deposited to the NCBI SRA BioProject (PRJNA640176). Reviewer's link:
https://dataview.ncbi.nlm.nih.gov/object/PRJNA640176?reviewer=p2dhiiern60seqodpuqbqb1o62
All bioinformatics code can be found on github:
https://github.com/poojasingh09/2024_singh_transcriptomeanalysis_pipeline

## Author Contributions

Pooja Singh: Conceptualisation, study design, sampling, bioinformatics analyses and interpretation, and writing manuscript
Ehsan Pashay Ahi: Dissections, RNA extraction & library preparation, qPCR
Anna Duenser: Sampling, RNA extraction & library preparation
Marija Durdevic: Bioinformatics resources
Wolfgang Gessl: Fish husbandry and sampling
Sylvia Schaeffer: Library preparation
Ole Seehausen: Sampling, study design
Christian Sturmbauer: Conceptualisation, study design, sampling

## Acknowledgements

We would like to thank Ursula Winkler from the high-performance computing server (sauron) of University of Graz for her patience and support. We are very grateful to the Zambian and Tanzanian fisheries for their long-term support in establishing cichlid species collections at Uni. Bern, EAWAG, and Uni Graz.

## References


1. King, M. & Wilson, A. C. Evolution at two levels in humans and chimpanzees. *Science (1979)* **188**, 107–116 (1975).
2. Singh, P. & Ahi, E. P. The importance of alternative splicing in adaptive evolution. *Mol Ecol* **31**, 1928–1938 (2022).
3. Keren, H., Lev-Maor, G. & Ast, G. Alternative splicing and evolution: diversification, exon definition and function. *Nat Rev Genet* **11**, 345–355 (2010).
4. Chen, L., Bush, S. J., Tovar-Corona, J. M., Castillo-Morales, A. & Urrutia, A. O. Correcting for differential transcript coverage reveals a strong relationship between alternative splicing and organism complexity. *Mol Biol Evol* **31**, 1402–1413 (2014).
5. Merkin, J., Russell, C., Chen, P. & Burge, C. B. Evolutionary dynamics of gene and isoform regulation in mammalian tissues. *Science (1979)* **338**, 1593–1599 (2012).
6. Barbosa-Morais, N. L., Irimia, M., Pan, Q., Xiong, H. Y. & Gueroussov, S. The evolutionary landscape of alternative splicing in vertebrate species. *Science (1979)* **338**, 1587–1594 (2012).
7. Mazin, P. V, Khaitovich, P., Cardoso-Moreira, M. & Kaessmann, H. Alternative splicing during mammalian organ development. *Nat Genet* **53**, 925–934 (2021).
8. Schluter, D. *The Ecology of Adaptive Radiation*. (Oxford University Press, New York, 2000).
9. Seehausen, O. *Lake Victoria Rock Cichlids*. (Verduyn Cichlids, 1996).
10. Yamaoka, K. Trophic ecomorphology of Tanganyikan cichlids. in *Fish Communities in Lake Tanganyika* (ed. Kawanabe H, Hori M, N. M.) 27–56 (Kyoto University Press, Kyoto, 1997).
11. Seehausen, O. Process and pattern in cichlid radiations-inferences for understanding unusually high rates of evolutionary diversification. *New Phytologist* **207**, 304–312 (2015).
12. Brawand, D. *et al.* The genomic substrate for adaptive radiation in African cichlid fish. *Nature* **513**, 375–381 (2014).
13. El Taher, A. *et al.* Gene expression dynamics during rapid organismal diversification in African cichlid fishes. *Nat Ecol Evol* **5**, 243–250 (2021).
14. Ahi, E. P., Singh, P., Duenser, A., Gessl, W. & Sturmbauer, C. Divergence in larval jaw gene expression reflects differential trophic adaptation in haplochromine cichlids prior to foraging. *BMC Evol Biol* **19**, 150 (2019).
15. Duenser, A. *et al.* Conserved molecular players involved in human nose morphogenesis underlie evolution of the exaggerated snout phenotype in cichlids. *Genome Biol. Evol* **15**, (2023).

16. Singh, P., Ahi, E. P. & Sturmbauer, C. Gene coexpression networks reveal molecular interactions underlying cichlid jaw modularity. *BMC Ecol Evol* **21**, 1–17 (2021).
17. Singh, P., Börger, C., More, H. & Sturmbauer, C. The role of alternative splicing and differential gene expression in cichlid adaptive radiation. *Genome Biol Evol* **9**, 2764–2781 (2017).
18. Terai, Y., Morikawa, N., Kawakami, K. & Okada, N. The complexity of alternative splicing of hagoromo mRNAs is increased in an explosively speciated lineage in East African cichlids. *PNAS* **100**, 12798–803 (2003).
19. Irisarri*, I. *et al.* Phylogenomics uncovers early hybridization and adaptive loci shaping the radiation of Lake Tanganyika cichlid fishes. *Nat Commun* **9**, 3159 (2018).
20. Singh, P. *et al.* Phylogenomics of trophically diverse cichlids disentangles processes driving adaptive radiation and repeated trophic transitions. *Ecol Evol* **12**, e9077 (2022).
21. 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).
22. Fujimura, K. & Okada, N. Development of the embryo, larva and early juvenile of Nile tilapia Oreochromis niloticus (Pisces: Cichlidae). Developmental staging system. *Dev Growth Differ* **49**, 301–324 (2007).
23. 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**, 1–19 (2017).
24. Pertea, M. *et al.* StringTie enables improved reconstruction of a transcriptome from RNA-seq reads. *Nat Biotechnol* **33**, 290–295 (2015).
25. 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).
26. Wright, C. J., Smith, C. W. J. & Jiggins, C. D. Alternative splicing as a source of phenotypic diversity. *Nat Rev Genet* **23**, 697–710 (2022).
27. Wang, E. T. *et al.* Alternative Isoform Regulation in Human Tissue Transcriptomes. *Nature* **456**, 470 (2008).
28. Trincado, J. L. *et al.* SUPPA2: Fast, accurate, and uncertainty-aware differential splicing analysis across multiple conditions. *Genome Biol* **19**, 1–11 (2018).
29. Ast, G. How did alternative splicing evolve? *Nat Rev Genet* **5**, 773–782 (2004).
30. Garg, K. & Green, P. Differing patterns of selection in alternative and constitutive splice sites. *Genome Res* **17**, 1015 (2007).
31. Brawand, D. *et al.* The evolution of gene expression levels in mammalian organs. *Nature* (2011) doi:10.1038/nature10532.

32. Chen, J. *et al.* A quantitative framework for characterizing the evolutionary history of mammalian gene expression. *Genome Res* **29**, 53–63 (2019).
33. Loh, Y. H. E. *et al.* Origins of shared genetic variation in African cichlids. *Mol Biol Evol* **30**, 906–917 (2013).
34. Wittkopp, P. J. *et al.* Evolutionary changes in cis and trans gene regulation. *Nature* **430**, 85–88 (2004).
35. Parsons, K. J., Trent Taylor, A., Powder, K. E. & Albertson, R. C. Wnt signalling underlies the evolution of new phenotypes and craniofacial variability in Lake Malawi cichlids. *Nature Communications 2014 5:1* **5**, 1–11 (2014).
36. Paine-Saunders, S., Viviano, B. L., Zupicich, J., Skarnes, W. C. & Saunders, S. glypican-3 controls cellular responses to Bmp4 in limb patterning and skeletal development. *Dev Biol* **225**, 179–187 (2000).
37. Li, Y., Li, Q., Longs, Y. & Cui, Z. Lzts2 Regulates Embryonic Cell Movements and Dorsoventral Patterning through Interaction with and Export of Nuclear β-Catenin in Zebrafish. *Journal of Biological Chemistry* **286**, 45116–45130 (2011).
38. Masonick, P., Meyer, A. & Hulsey, C. D. A Kiss of Deep Homology: Partial Convergence in the Genomic Basis of Hypertrophied Lips in Cichlid Fish and Human Cleft Lip. *Genome Biol Evol* **15**, evad072 (2023).
39. Meier, J. I. *et al.* Ancient hybridization fuels rapid cichlid fish adaptive radiations. *Nat Commun* **8**, 14363 (2017).
40. Powder, K. E. & Albertson, R. C. Cichlid fishes as a model to understand normal and clinical craniofacial variation. *Dev Biol* **415**, 338–346 (2016).
41. Yu, H. *et al.* Frizzled 1 and frizzled 2 genes function in palate, ventricular septum and neural tube closure: general implications for tissue fusion processes. *Development* **137**, 3707–3717 (2010).
42. Karagic, N., Schneider, R. F., Meyer, A. & Hulsey, C. D. A Genomic Cluster Containing Novel and Conserved Genes is Associated with Cichlid Fish Dental Developmental Convergence. *Mol Biol Evol* **37**, 3165–3174 (2020).
43. Hu, Y. *et al.* A Novel Autosomal Recessive GJA1 Missense Mutation Linked to Craniometaphyseal Dysplasia. *PLoS One* **8**, e73576 (2013).
44. Duenser*, A. *et al.* Developmental tinkering of gene regulation facilitated super rapid adaptive radiation. *bioRxiv* 2024.01.24.577063 (2024) doi:10.1101/2024.01.24.577063.
45. Dalcq, J. *et al.* RUNX3, EGR1 and SOX9B form a regulatory cascade required to modulate BMP-signaling during cranial cartilage development in zebrafish. *PLoS One* **7**, (2012).
46. Busch, A. & Hertel, K. J. Evolution of SR protein and hnRNP splicing regulatory factors. *Wiley Interdiscip Rev RNA* **3**, 1–12 (2012).

47. Rodríguez-Ramírez, C. E. *et al.* Molecular mechanisms of Eda-mediated adaptation to freshwater in threespine stickleback. *Mol Ecol* **00**, 1–19 (2023).
48. Jakšić, A. M. & Schlötterer, C. The interplay of temperature and genotype on patterns of alternative splicing in Drosophila melanogaster. *Genetics* **204**, 315–325 (2016).
49. Jacobs, A. & Elmer, K. R. Alternative splicing and gene expression play contrasting roles in the parallel phenotypic evolution of a salmonid fish. *Mol Ecol* **30**, 4955–4969 (2021).
50. Albertson, R. C. & Kocher, T. D. Genetic and developmental basis of cichlid trophic diversity. *Heredity (Edinb)* **97**, 211–221 (2006).
51. Seppala, M. *et al.* Gas1 is a modifier for holoprosencephaly and genetically interacts with sonic hedgehog. *J Clin Invest* **117**, 1575–1584 (2007).
52. Wainwright, P. C. *et al.* The evolution of pharyngognathy: a phylogenetic and functional appraisal of the pharyngeal jaw key innovation in labroid fishes and beyond. *Syst Biol* **61**, 1001–1027 (2012).
53. Monsoro-Burq, A. H. PAX transcription factors in neural crest development. *Semin Cell Dev Biol* **44**, 87–96 (2015).
54. Fraser, G. J. *et al.* An ancient gene network is co-opted for teeth on old and new jaws. *PLoS Biol* **7**, e31 (2009).
55. Le Pabic, P., Stellwag, E. J., Scemama, J. & Carolina, N. Embryonic development and skeletogenesis of the pharyngeal jaw apparatus in the cichlid Nile tilapia (Oreochromis niloticus). *Anat Rec* **292**, 1780–1800 (2009).
56. Fay, J. C. & Wittkopp, P. J. Evaluating the role of natural selection in the evolution of gene regulation. *Heredity (Edinb)* **100**, 191–9 (2008).
57. Khaitovich, P. *et al.* A Neutral Model of Transcriptome Evolution. *PLoS Biol* **2**, e132 (2004).
58. Draghi, J., Ogbunugafor, C. B., Zaman, L. & Parsons, T. L. Relaxed selection can speed the evolution of complex adaptations. *bioRxiv* 2024.07.09.602773 (2024) doi:10.1101/2024.07.09.602773.
59. Wanek, A. K. & Sturmbauer, C. Form, function and phylogeny: comparative morphometrics of Lake Tanganyika's cichlid tribe Tropheini. *Zool Scr* **44**, 362–373 (2015).
60. Sturmbauer, C., Hainz, U., Baric, S., Verheyen, E. & Salzburger, W. Evolution of the tribe Tropheini from Lake Tanganyika: synchronized explosive speciation producing multiple evolutionary parallelism. *Hydrobiologia* **500**, 51–64 (2003).
61. Lamichhaney, S. *et al.* Evolution of Darwin's finches and their beaks revealed by genome sequencing. *Nature* (2015) doi:10.1038/nature14181.
62. Ahi, E. P. Signalling pathways in trophic skeletal development and morphogenesis: insights from studies on teleost fish. *Dev Biol* **420**, 11–31 (2016).
63. Marques, D. A., Meier, J. I. & Seehausen, O. A combinatorial view on speciation and adaptive radiation. *Trends Ecol Evol* **34**, 531–544 (2019).

64. Joyce, D. *et al.* An extant cichlid fish radiation emerged in an extinct Pleistocene lake. **435**, (2005).
65. Tress, M. L., Abascal, F. & Valencia, A. Most alternative isoforms are not functionally important. *Trends Biochem Sci* **42**, 408–410 (2017).
66. Smith, C. C. R. *et al.* Genetics of alternative splicing evolution during sunflower domestication. *PNAS* **115**, 6768–6773 (2018).
67. West-Eberhard, M. J. *Developmental Plasticity and Evolution*. (Oxford University Press, 2003).
68. FastQC: A quality control tool for high throughput sequence data. www.bioinformatics.babraham.ac.uk/projects/fastqc.
69. Bolger, A. M., Lohse, M. & Usadel, B. Trimmomatic: A flexible trimmer for Illumina sequence data. *Bioinformatics* **30**, 2114–2120 (2014).
70. Dobin, A. *et al.* STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* **29**, 15–21 (2013).
71. 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. *bioRxiv* 1–39 (2017) doi:10.1101/099564.
72. Li, H. *et al.* The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25**, 2078–2079 (2009).
73. Pertea, G. & Pertea, M. GFF Utilities: GffRead and GffCompare [version 1; peer review: 3 approved] . *F1000Res* **9**, (2020).
74. Singh, P. Transcriptomics pipeline for accurate gene and isoform analysis . Preprint at https://doi.org/10.5281/zenodo.15393254 (2025).
75. R Core Team. R: A language and environment for statistical. Preprint at https://www.R-project.org/ (2021).
76. Paradis, E. & Schliep, K. ape 5.0: an environment for modern phylogenetics and evolutionary analyses in R. *Bioinformatics* **35**, 526–528 (2019).
77. Galili, T. dendextend: an R package for visualizing, adjusting, and comparing trees of hierarchical clustering. *Bioinformatics* (2015) doi:10.1093/bioinformatics/btv428.
78. Revell, L. J. phytools: an R package for phylogenetic comparative biology (and other things). *Methods Ecol Evol* **3**, 217–223 (2012).
79. Danecek, P. *et al.* Twelve years of SAMtools and BCFtools. *Gigascience* **10**, (2021).
80. Stamatakis, A. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* **30**, 1312–1313 (2014).
81. Smedley, D. *et al.* BioMart – biological queries made easy. *BMC Genomics* **10**, 22 (2009).
82. Zerbino, D. R. *et al.* Ensembl 2018. *Nucleic Acids Res* **46**, D754–D761 (2018).
83. Alexa, A. & Rahnenfuhrer, J. topGO: Enrichment analysis for gene ontology. Preprint at (2019).

84. Ahi, E. P., Singh, P., Lecaudey, L. A., Gessl, W. & Sturmbauer, C. Maternal mRNA input of growth and stress-response-related genes in cichlids in relation to egg size and trophic specialization. *Evodevo* **9**, 1–17 (2018).