Ancestral splice variation is a key substrate for rapid diversification in African cichlids

Pooja Singh[1,2,3], Ehsan P. Ahi[3,4], Anna Duenser[3,5], Marija Durdevic[6,7], Wolfgang Gessl[3], Sylvia Schaeffer[3], Julian Gallaun[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 Centre for Molecular BioMedicine, Institute of Pathology, Medical University of Graz, 8010 Graz, Austria
[7]Core Facility Computational Bioanalytic, Center for Medical Research, Medical University of Graz, Graz, Austria
Corresponding author: P. Singh (pooja.singh09@gmail.com)

Abstract

Adaptive radiation is a major driver of biodiversity. In some of the largest radiations, the pace of morphological change far outstrips the protein-coding mutation rate, suggesting that gene regulation may be important for diversification. Yet, the specific gene regulatory mechanisms shaping adaptive radiation remain poorly understood. Analysing of 200 transcriptomes from three independent but phylogenetically nested African cichlid adaptive radiations we show that alternative splicing (AS) evolved faster than gene expression (GE), playing a pivotal role in shaping novel trophic adaptations in the youngest and most species rich radiations in Lakes Victoria and Malawi. This divergence was largely driven by ancestral alternative isoforms, which, though present at low levels in related lineages that did not form radiations, increased in frequency during adaptive radiation. In addition, novel isoforms of craniofacial genes - some evolving within just a few thousand years contributed further to adaptation. The rapid turnover of AS is consistent with periods of relaxed selection followed by directional selection on alternative isoforms and splice sites, a dynamic that may have preserved a rich cache of isoform variation in these radiations and enabled ecological diversification as adaptive zones became available. We argue that the interplay between splicing and different forms of selection facilitates the generation and maintenance of protein-coding diversity, promoting evolutionary innovation into many ecologically different species at extremely short timescales.

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

Significance statement

Adaptive radiation—when species rapidly diversify to fill new ecological niches—is a key driver of biodiversity but its underlying molecular mechanisms remain unclear. Using comparative transcriptome sequencing across multiple African cichlid fish adaptive radiations, we found that rapid changes in alternative splicing—the process generating different protein-coding isoforms from the same gene—contributed more to early ecological divergence than shifts in gene expression levels. Most adaptive isoforms were present at low levels in the ancestor of these fishes but some new isoforms evolved remarkably fast, shaping diverse ecologies. We show that alternative splice variation, often thought to be biological "noise," can be a powerful and labile source of rapid evolutionary innovation during early stages of adaptive radiation.

Introduction

Adaptive radiations are characterised by bursts of ecological and morphological diversity that have shaped the biodiversity on our planet (1). Rates of molecular evolution in protein-coding genes cannot explain the rate at which morphological evolution happens during adaptive radiation (2–4), which is why gene regulation is hypothesised to be an important mechanism for ecomorphological change in these lineages (5–7). Transcriptional changes in gene expression (GE) fine-tune the amount of messenger RNA (mRNA) and thus proteins produced from a gene (3, 8) and alternative splicing (AS) is a co-/post-transcriptional mechanism that can rapidly diversify the kinds of transcript and protein isoforms coded by a single gene by generating multiple distinct messenger RNAs (mRNAs) (9, 10). While some inroads have been made (7, 11–15), we still understand little about how these transcriptional and post-transcriptional regulatory mechanisms interact and are shaped by evolutionary processes to give rise to species-rich bursts of adaptive radiation (5).

To address this knowledge gap, here we examined the evolutionary dynamics of GE and AS in cichlid fishes from Lake Victoria, Lake Malawi and Lake Tanganyika that represent three of the fastest and ecomorphologically most diverse vertebrate adaptive radiations. Species in these lakes have evolved diverse jaws and feeding habits to adapt to different food sources, forming species-rich communities spanning complex foodwebs (16, 17). Across lakes, similar feeding ecologies and morphologies have evolved repeatedly, suggesting that cichlid lineages solved ecological challenges in similar ways, and that cichlid evolution may be constrained (18–23). We sequenced whole transcriptomes of oral (OJ) and pharyngeal jaws (PJ), the functional decoupling of which is considered a key evolutionary innovation that opened up novel trophic niches for cichlid fishes (24), from 18 phylogenetically nested cichlid species. We selected species with ecologically divergent herbivorous and carnivorous trophic adaptations within lakes and ecologically convergent trophic adaptations across lakes (16, 19, 20) (Figure 1A). This allowed us to address dynamics of repeated evolution as well as divergent evolution. Our sampled lineages represent the evolutionary process of adaptive radiation at different stages – Lake Victoria is the youngest and has the fastest rates of speciation, Lake Malawi is middle-aged, and

Lake Tanganyika is the oldest and most ecomorphologically and genetically distinct (25, 26) – providing a powerful framework to study the temporal role of different gene regulatory mechanisms. We also included data from two non-radiating cichlid species in our study that live in these lakes, and thus has access to ecological opportunity, but did not form radiations. These non-radiating *Astatotilapia*/*Astatoreochromis* lineages are also phylogenetic sister-groups to the radiations, with whom they share a recent common ancestor (25, 27). All our selected species are from the tribe Haplochromini, which is the most species-rich and ecologically diverse cichlid group (28) (Figure 1B for phylogenetic relationships). The 'Out of Tanganyika' theory of East African cichlid evolution suggests that an ancient riverine *Astatoreochromis*-like haplochromine lineage from Lake Tanganyika, colonised Lake Malawi and Lake Victoria, contributing to those radiations, and then re-entered Lake Tanganyika to give rise to the haplochromine radiation of Tropheini cichlids (27). It has been hypothesised that ancestral standing variation may have been an important driver of adaptive radiation in East African cichlids (29–31). Based on this theory, and through the sampling of lineages that formed radiations versus those that did not, we tested whether ancestral variation in GE and AS (1) contributed to divergent and convergent trophic evolution during adaptive radiation, (2) 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, and finally (3) if we can find evidence for directional selection on altenative isoforms of craniofacial genes.

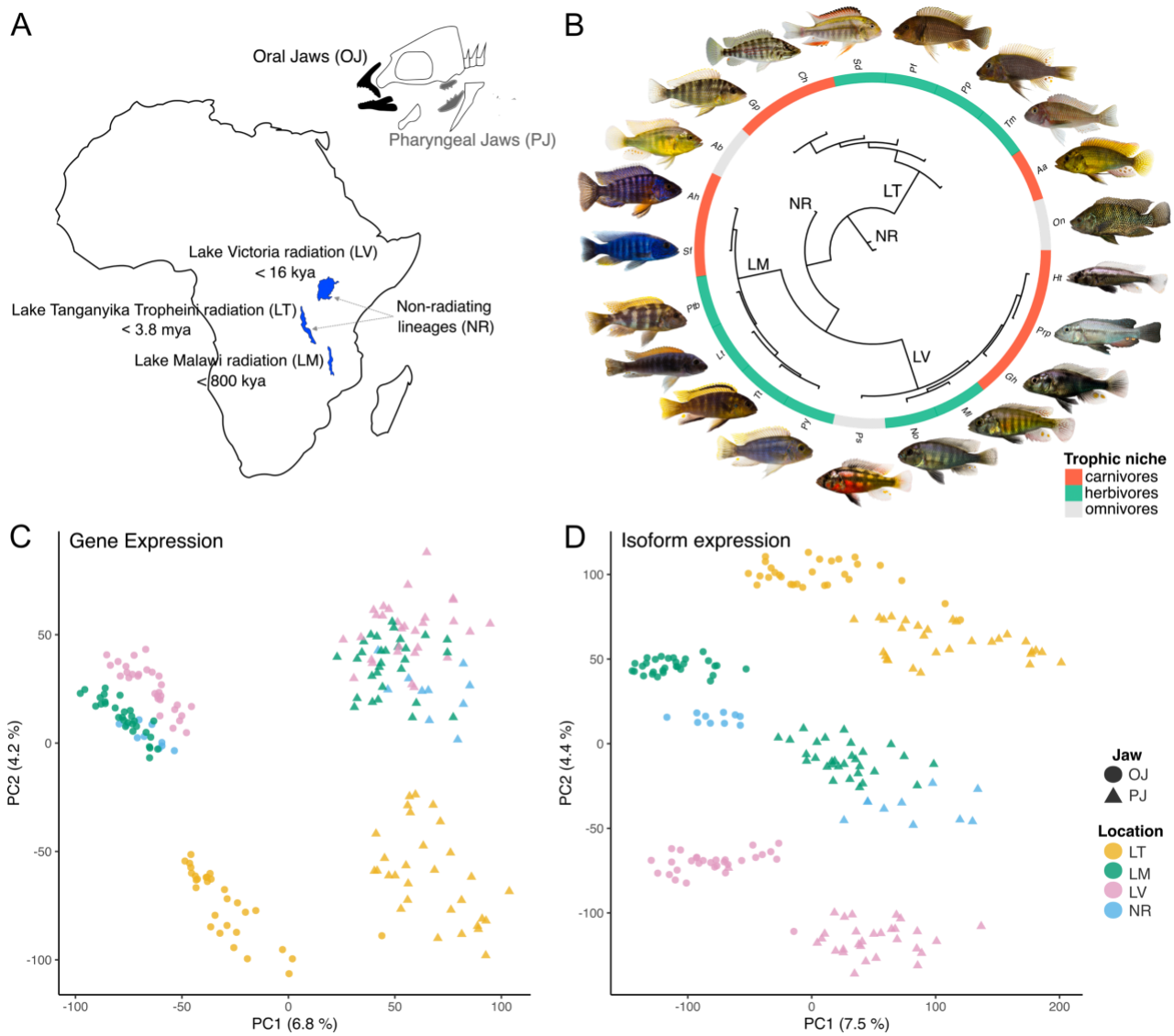


Figure 1 Study overview. (A, B) Biogeographical setting of haplochromine cichlid species sampled in this study (20 species, 2 jaw tissues (oral and pharyngeal), 5 biological replicates per species per tissue, n=200). Phylogenetic relationships reconstructed using transcriptome SNPs. Lake Victoria: *Paralabidochromis sauvagei* (Ps), *Neochromis omnicaeruleus* (No), *Mbipia lutea* (Ml), *Gaurochromis hiatus* (Gh) *Prognathochromis perrieri* (Prp); *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 (NR) haplochromines, *Astatotilapia burtoni* (Ab) and *Astatoreochromis alluaudi* (Aa) are included as riverine lineages that entered lakes but did not radiate, and are also phylogenetic sister lineage to the radiations. *O. niloticus* (On) is the outgroup species which was included for phylogenetic analyses only. **(C, D)** Principal component analysis of transcriptome-wide normalised gene and isoform expression from oral (OJ) and pharyngeal jaws (PJ) of all 20 species.

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 adaptive radiations and two species that entered lakes but did not make radiations (5 individuals per tissue per species, total N = 200). Reads were mapped to the phylogenetically equidistant *Oreochromis niloticus* reference genome(32) (median sequencing depth per tissue: median ~6.4 million 125 base-pair (bp) paired end reads per library; median ~80% uniquely mapped read; Table S1) using the genome-guided transcript assembly function. Gene annotations of species and tissues constructed using StringTie(33) were merged incrementally into one super-annotation(33) that was used for downstream analysis. This strategy helped to reduce false positive while increasing detection of novel and lowly expressed isoforms. Gene and isoform expression was quantified using established methods(33, 34).

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, Lake Tanganyika clustered separately from the younger radiations from Lakes Malawi and Victoria and the non-radiating species, which 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 Lake Malawi samples, particularly for the PJ. Both gene and isoform expression patterns within tissues captures structure among radiations (Figure 1B), with stronger signals emerging from the isoform expression. These patterns suggest that isoform-level variation may capture interspecific differences, even among very young species-flocks, that are not apparent at the gene expression level, although PCA provides only qualitative indication of these relationships.

Splicing and expression have contrasting evolutionary dynamics

Alternative splicing results in the change of relative abundance of isoforms produced from the same gene(9), which may have evolutionary consequences(5, 35). But little is known about how AS evolves at difference timescales during adaptive radiation. 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(36, 37) and computed Spearman's rank correlation coefficients (Spearman's ρ) 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 Spearman's ρ values for GE than AS (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 ρ 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 that transcriptome-wide alternative isoforms are evolving under relaxed selection pressures during adaptive radiation ((38, 39).

Next, we analysed the hierarchical clustering of Spearman's ρ for GE and PSI. We found that GE grouped samples by tissue for younger radiations (Lakes Victoria and Malawi) and nonradiating species (Ab, Aa). However, for the older Lake Tanganyika 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(40). In contrast, Spearman's ρ of PSI clustered samples first by clade/radiation, regardless of tissue, and samples from Lake Tanganyika and nonradiating species further clustered by species, regardless of tissue (Figure 2B, S2). The latter pattern would be expected if changes in splicing were evolutionarily conserved (heritable) in species and clades. 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, non-radiating species clustered with Lake Victoria in the gene expresssion analysis but with Lake Tangayika in the PSI analysis. Interestingly, PJ samples of *H. thereuterion* from Lake Victoria were placed outside of the Lake Victoria clade, sister to the PJ's of the non-radiating *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 a more important role in the evolution of clade and species-specific jaw specialisations compared 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.

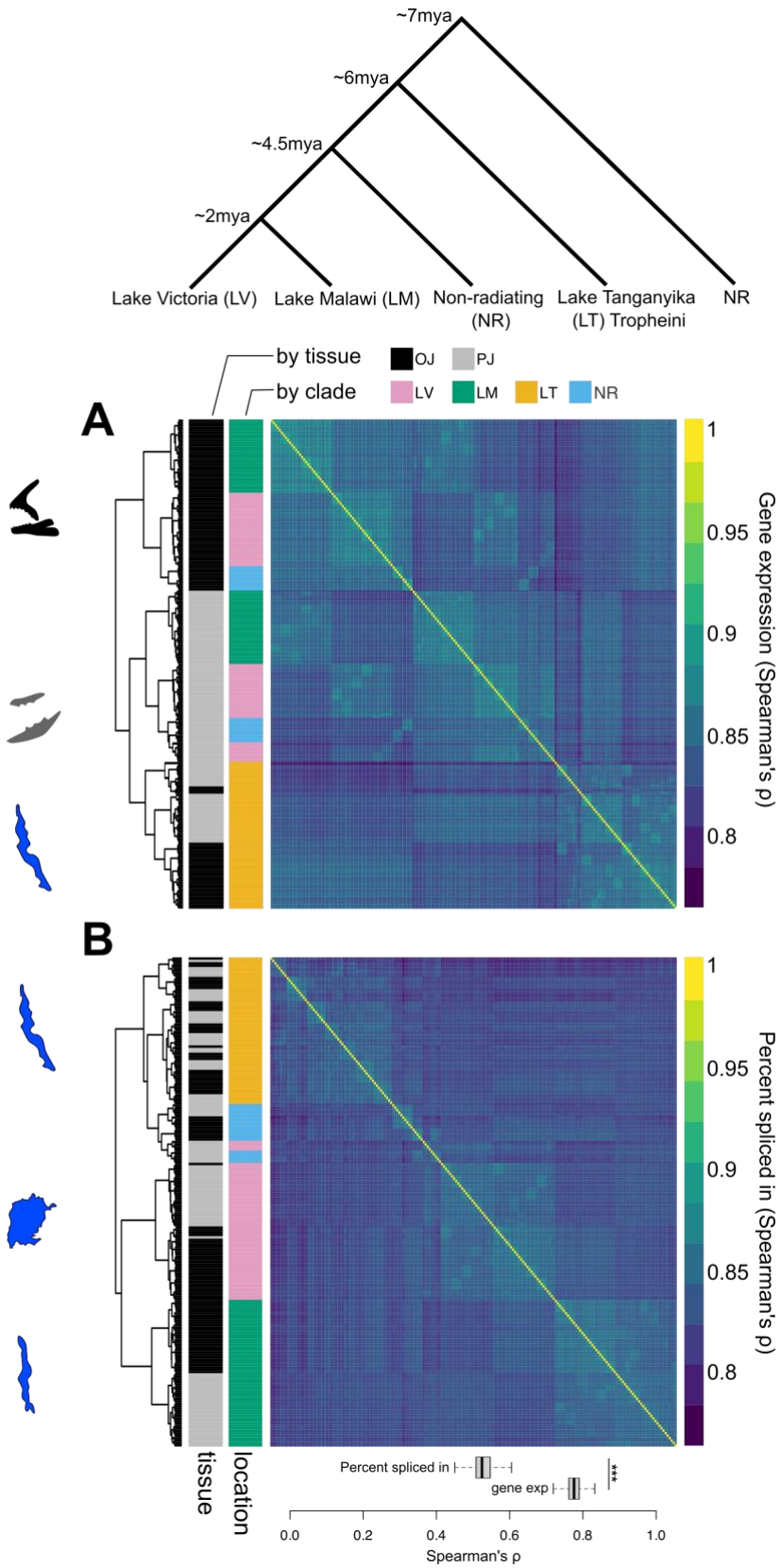

~7mya
~6mya
~4.5mya
~2mya
Lake Victoria (LV)
Lake Malawi (LM)
Non-radiating (NR)
Lake Tanganyika (LT) Tropheini
NR
by tissue
by clade
OJ
PJ
LV
LM
LT
NR
A
B
1
0.95
0.9
0.85
0.8
Gene expression (Spearman's ρ)
Percent spliced in (Spearman's ρ)
tissue
location
Percent spliced in
gene exp
***
0.0
0.2
0.4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

Figure 2 Profiling of gene expression and alternative splicing in East African cichlids. Hierarchical clustering of pairwise Spearman's rank correlation coefficient (ρ) of transcriptome-wide (**A)** normalised gene expression (GE) and (**B)** isoform percent spliced in (PSI) of genes with more than one isoform. Analysis was conducted across all 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 Spearman's ρ of PSI and GE across all samples (two-sides t-test $p < 0.001$). See Figure S1A for boxplots of Spearman's ρ of PSI and GE calculated per lake. Cladogram illustrates simplified cladogram of phylogenetic relationship of species shown in Figure 1B. Divergence times taken from Irisarri*, Singh* et al (2018) ((25). 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(41). Brownian motion models neutral evolution i.e. genetic drift. The distribution of OU variance across the phylogeny, denoted 'evolutionary variance'(42), 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). This has been previously reported for gene expression model species(41) but not for isoform expression. However, in contrast to gene expression, isoform expression had significantly higher evolutionary variance (less constraint) than GE under the OU model for both jaws (Figure S3C). This higher variance could be the result of weaker stabilising selection (or strong relaxed selection or perhaps the result of complex mix of selection) acting on alternative isoform expression than constitutive isoform expression, the latter being under stabilising selection to preserve function. These results reinforce the idea of fast evolving AS during adaptive radiation.

Most splice variation is ancestral, but some may be novel

Ancestral standing variation is proposed to fuel rapid cichlid adaptive radiations(31). One way that this could happen is through the sorting of ancestral variation in gene regulation(43), 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. 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 expressed were significantly higher in samples from the

200 younger radiations (Lakes Malawi and Victoria) and the non-radiating lineages than the older radiation of Lake
201 Tanganyika (Mann-Whitney U test $p < 0.001$; Figure 3A,B; S4 A,B). This overall pattern was also observed
202 when we analysed the number of isoforms per gene (Figure S5) but the difference was only significant between
203 the PJ's of Lake Victoria and Malawi (Mann-Whitney U test $p < 0.05$). We tested the homogeneity of variance
204 among groups using the Barlett's test and found that significant differences for the number of isoforms per gene
205 for PJ ($p = 0.017$) but not the OJ. This suggests that while the absolute number of isoforms expressed per gene
206 is reduced over time, the transcriptional variance among species increases with time during adaptive radiation.
207 To understand how much of this expression variation was ancestral versus putatively *de novo*, we investigated
208 if each gene or isoform was expressed in only one of the clades, or across several radiating and non-radiating
209 clades (Figure 3C). The majority of genes (86.0%) and isoforms (73.0%) were expressed in at least one non-
210 radiating species and one or more radiations – this variation is likely ancestral to all East African
211 haplochromines. These results are consistent with the 'Out of Tanganyika' hypothesis (27) – where similar pool
212 of ancestral variation would be presumably present at the base of all these three radiations, and also in the non-
213 radiating haplochromine lineages. However, in addition to this "Out of Tanganyika" colonisation (27), ancestral
214 hybridisation from different divergent lineages from the Congo and Nile (that are also haplochromine cichlids)
215 have shown to contribute genetic material independently to all of these radiations (25, 44, 45), which would
216 certainly have impacted the transcriptional landscape within each radiations. Furthermore, demographic history
217 of these radiations probably also plays a role, as the Lake Victoria has dried up and refilled (most recently
218 16,000 years ago) and this would impact the shared gene regulatory variation across all three lakes. All these
219 factors probably contribute to why so little gene/isoform expression is shared across all three radiations, but
220 larger sharing is observed in different subsets of radiating/non-radiating species. Given what we know about
221 their phylogenetic relationships (Figure 1B), the most parsimonious explanation is that this variation was
222 ancestral and was lost in some radiations; instead of being independently gained in non-radiating lineages and
223 a subset of radiation. A small subset of isoforms (~12.0%) were unique to one radiation only – pointing towards
224 the evolution of novel splice variation in the course of adaptive radiation – or less likely, repeated loss in all
225 other lineages. Novel isoforms can rapidly evolve through weak purifying selection on alternative splice sites
226 ((39). It is of course possible that these putatively novel isoforms may be ancestral but the pattern is missed by
227 our sample selection. It is also possible that isoform variation arrived through hybridisation, as cycles of
228 hybridisation have fueled cichlid adaptive radiations ((25, 26, 45, 46). Taken together, our data support the
229 hypothesis that substantial transcriptomic variation existed in the common ancestor of all East African
230 haplochromines that was likely shaped by selection and drift during adaptive radiation, with some isoforms
231 being purged and others retained; with some putatively novel isoforms evolving in the course of adaptive
232 radiation. But did these ancestral and novel isoforms contribute to ecological adaptation in cichlid radiations?

233

234

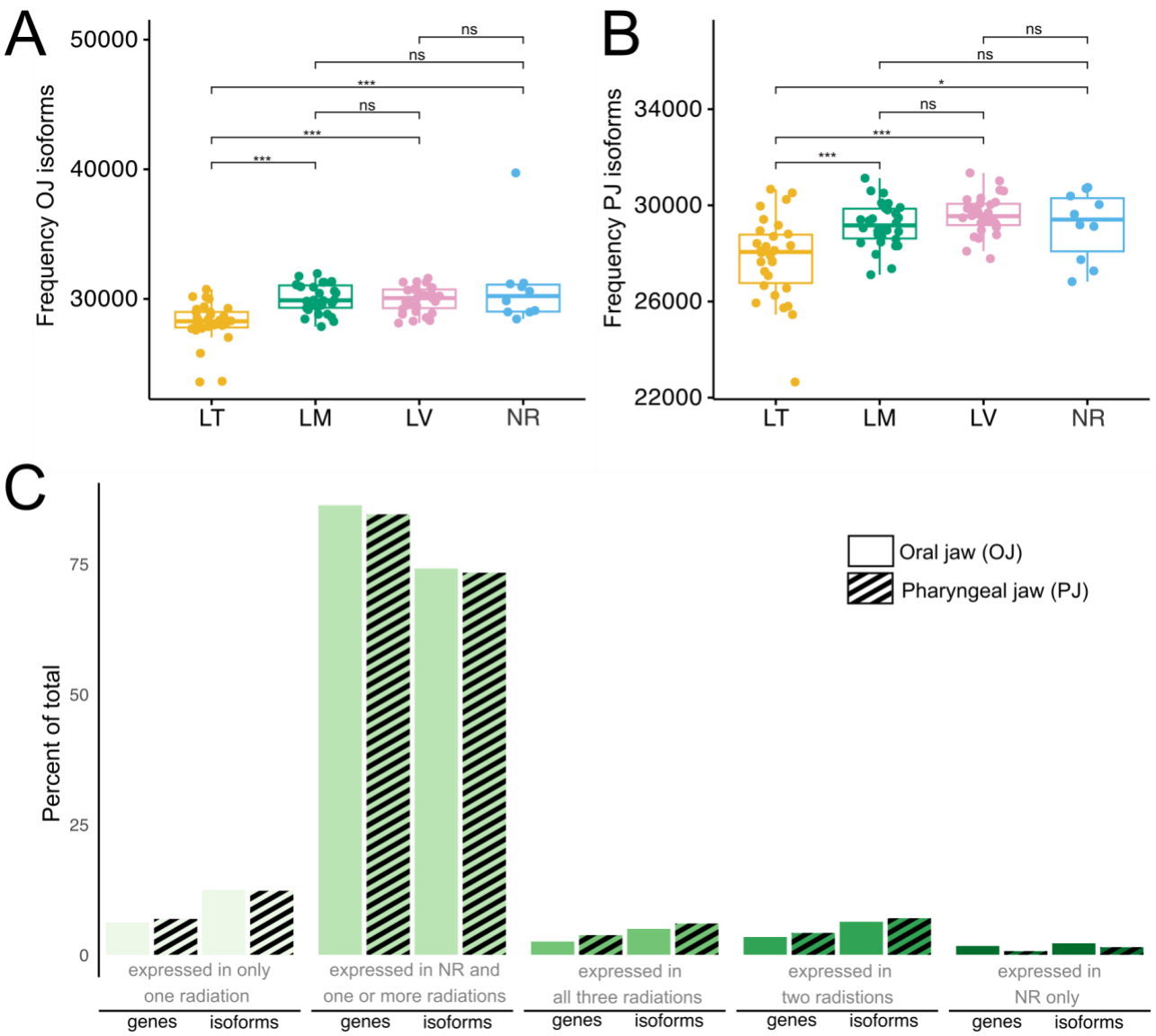
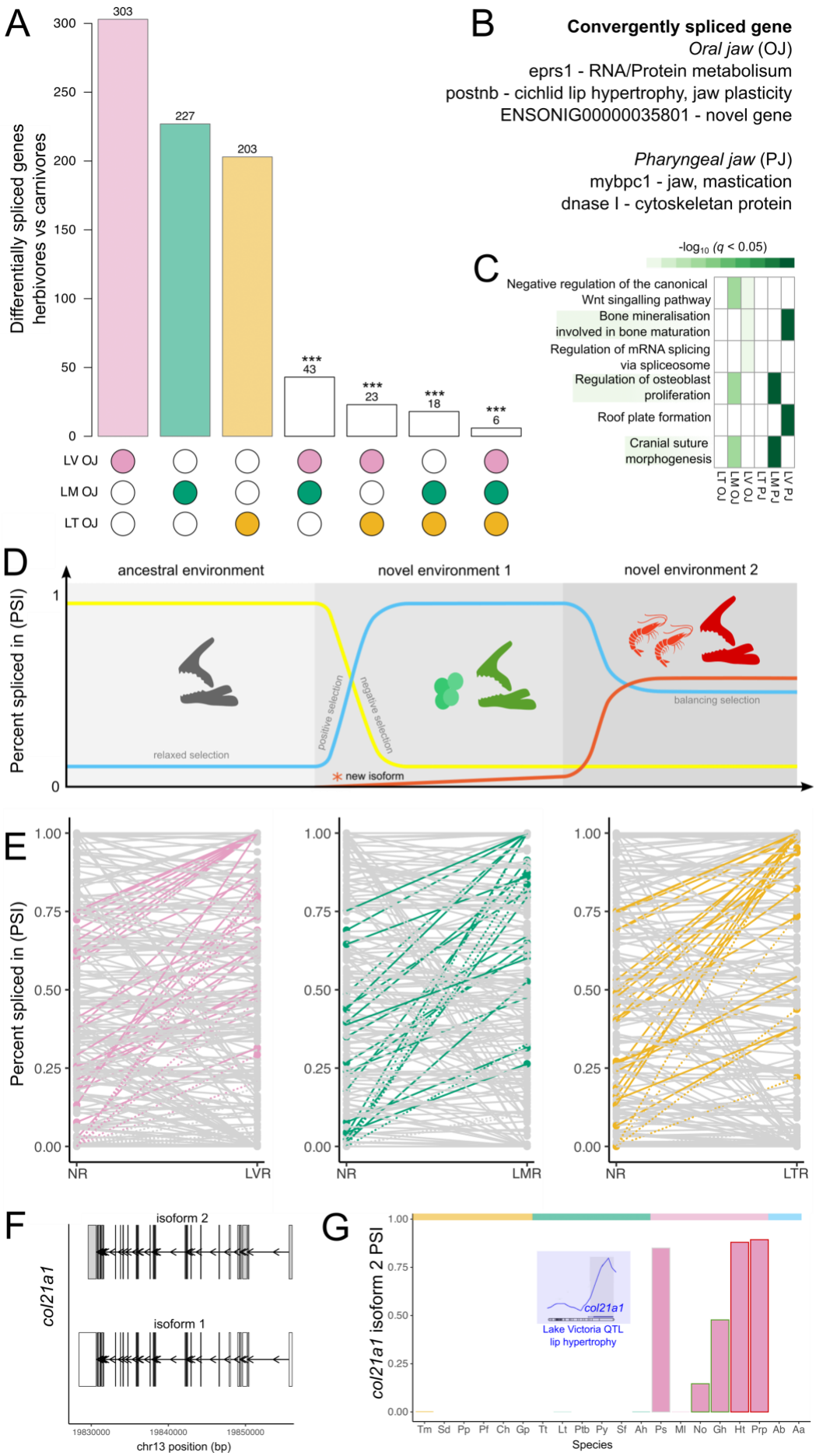


Figure 3 Transcriptomic variation in non-radiating and radiating cichlids. (**A, B)** The absolute number of expressed isoforms identified in the oral and pharyngeal jaw mRNA of all samples. The absolute number of genes expressed is shown in Figure S4. Asterisks denote Mann-Whitney U test significance: $p < 0.001$***, $p < 0.01$**, $p < 0.05$*. 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

248 each lake to identify differentially spliced genes (DSGs) that may play a role in shaping their eco-morphological
249 differences. The largest number of genes were differentially spliced in Lake Victoria (OJ 303, PJ 215) followed
250 by Lake Malawi (OJ 227, PJ 214) and Lake Tanganyika (OJ 203, PJ 126) (Figure 3A for OJ, Figure S6 for PJ,
251 File S1). This indicates that younger cichlid radiations leveraged splice variation relatively more than the older
252 radiation in Lake Tanganyika for trophic adaptation. Differentially spliced genes were significantly enriched
253 (Fisher's exact test $q < 0.05$) for GO terms specifically associated with craniofacial and jaw morphogenesis that
254 are known from literature such as 'regulation of osteoblasts', 'bone mineralisation involved in bone maturation',
255 'roof plate formation', and 'skeletal muscle development' (subset shown in Figure 3C; full data in Figure S7).
256 This was especially true for DSGs in the younger radiations. The Wnt signalling pathway that has been
257 associated with the evolution of new craniofacial phenotypes at the larval stage in Malawi cichlids ((47) was
258 enriched in divergent jaws from Lake Victoria. It is likely that AS of Wnt pathway genes *gpc3* and *lzts2a* (File
259 S1) that control skeletal development and dorsoventral patterning contributed to adaptive jaw shape differences
260 Lake Victoria cichlids ((48, 49). Two genes, *rnfox1* and *rbm5*, involved in 'mRNA splicing via the spliceosome'
261 were differentially spliced in Lake Victoria oral jaws (File S1). This suggests that AS of spliceosome
262 components, which is the molecular machinery that controls splicing, contributed to the evolution of ecologicaly
263 flexibility in the fastest and most species-rich vertebrate adaptive radiation. In Lake Malawi OJ, differential
264 splicing of genes involved in bone maturation (*col1a1*) and cranial morphogenesis (*sp7*) were notable. Overall,
265 twice as many DSGs were shared between Lakes Victoria and Malawi, than either of these lakes shared with
266 Lake Tanganyika (Figure 4A, S6), which is most consistent with these splice variants evolving and/or gaining
267 function after the ancestor of these Lake Tanganyika haplochromines had split from the common ancestor of
268 Lakes Victoria and Malawi, in line with the 'Out of Tanganyika' theory of haplochromine cichlid evolution
269 ((27).


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
Lake Victoria QTL
lip hypertrophy
Species
Tm Sd Pp Pf Ch Gp Tt Lt Ptb Py Sf Ah Ps Ml No Gh Ht Prp Ab Aa

270

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

293

294 Of the genes that were significantly differentially spliced between herbivores and carnivores in each lake (Figure
295 3A, S6A), a dynamic pattern of isoform evolution emerged: many of these were low expressed isoforms in non-
296 radiating species that increased in expression in radiating species (Figure 3E for OJ, S6B for PJ). An increase
297 in frequency would be the expectation if these low expressed ancestral alternative isoforms persisted in
298 populations as standing variation due to relaxed selection pressures (38, 39) and gained adaptive function when
299 these lineages entered lakes and discovered newly available food sources (Figure 3D) as hypothesised by Singh
300 & Ahi ((5). Many highly expressed isoforms in non-radiating species also decreased in frequency or were lost
301 in the lacustrine radiations (presumably as the alternative isoform increased in frequency). We further identified
302 34 potentially novel isoforms that were absent in non-radiating species and only expressed a subset of the three
303 adaptive radiations (Figure 3E dotted lines, File S1). Sixteen of these 34 novel isoforms were found in LV,
304 followed by 12 in Lake Malawi and six in Lake Tanganyika. Novel isoforms of note were of genes *col21a1* and
305 *lzts2a*. *Col21a1* had two isoforms (Figure 3F) of which the main isoform is expressed in the OJ of Lake Victoria
306 herbivores as well as herbivores from Lakes Tanganyika and Malawi (with one exception, File S1). The

307 alternative isoform of *col21a1* is only expressed in Lake Victoria species, with higher expression in carnivores
308 and omnivores (Figure 3G). *Col21a1* is associated with human cleft lip syndrome and maps to a QTL for
309 hypertrophic lips in Lake Victoria cichlids(50). It is conceivable that this isoform may have arisen *de novo* in
310 Lake Victoria cichlids or arrived through hybridisation, which has been shown to contribute genetic variation
311 for rapid adaptive radiation in Lake Victoria(44). In the future, integrating transcriptomic analysis with
312 population genetics and functional genomics (F_{ST} scans, GWAS, spliceQTL) approaches would help to narrow
313 down the genetic variants associated with splicing and the evolutionary processes that shape them. Our results
314 also highlight the conservation of craniofacial development pathways across vertebrates, and the promise of
315 cichlid fishes as a natural model for studying human craniofacial disease(51).

316

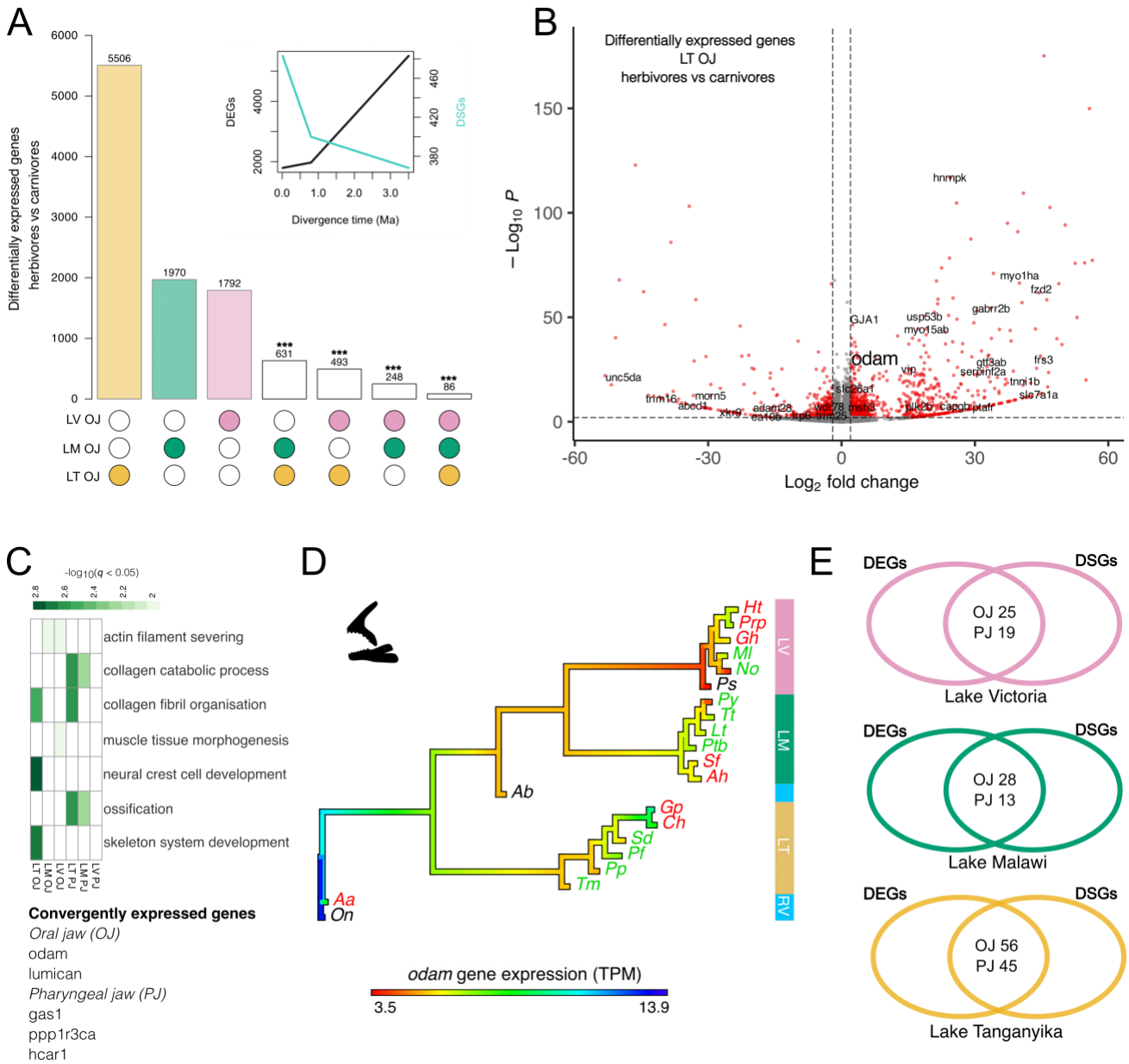
317 **Differential expression of major craniofacial genes contributed to jaw divergence**

318 In addition to differences in splicing, we investigated GE differences between herbivores and carnivores in each
319 lake radiation to be able to compare the role of GE with AS in trophic adaptation. The largest number of
320 differentially expressed genes (DEGs) were in Lake Tanganyika (OJ 5506, PJ 4094), followed by Lake Malawi
321 (OJ 1970, PJ 1499) and Lake Victoria (OJ 1792, PJ 1166) (Figure 5A for OJ, Figure S8 for PJ, File S1). Several
322 interesting candidate genes for jaw/craniofacial morphogenesis known from literature stood out as the most
323 extreme outliers in the oldest Lake Taganyika where GO terms such as 'neural crest cell development', 'skeletal
324 system development, and 'ossification' were significantly enriched ($q < 0.05$). One such gene for Lake
325 Tanganyika OJ was Frizzled 2 (*fzd2*), a Wnt pathway gene that is highly expressed in the developing face of
326 vertebrates, and mutations of this gene cause cleft palate and wider nose and upper jaw(52). A gene controlling
327 prey crushing teeth morphology in cichlids, *odam (*(53), and a gene linked to craniofacial disease, *gja1,* were
328 also significant for Lake Tanganyika OJ(54). Two interesting outliers in Lake Victoria OJ were *kaznb*, a
329 craniofacial development gene that has been found to be upregulated in the jaws of herbivorous cichlids at larval
330 and adult life stages (55) and *egr1*, that forms part of the regulatory cascade modulating cranial bone
331 morphogenesis through the BMP pathway(56). Interestingly, one of the most extreme outliers for Lake
332 Taganyika OJ was *hnRNPk*, an RNA binding protein that regulates pre-mRNA splicing(57). Overall, the GO
333 terms for DEGs in the youngers lakes were less specific to jaw development than the DSGs. This suggests that
334 shifts AS may be a more flexible route to rapid diversification in explosive adaptive radiations than changes in
335 GE.

336

337

338


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
$-\text{Log}_{10}\,P$
Log_2 fold change
odam
C
$-\log_{10}(q < 0.05)$
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

342 **Figure 5 GE divergence in oral jaws of herbivore and carnivore species. (A)** Number of differentially
343 expressed genes (DEGs, $q < 0.05$) between oral jaws (OJ) of herbivores and carnivores in each lake, with
344 overlaps annotated (hypergeometric test $p < 0.001$). Results for pharyngeal jaws can be found in Figure S8.
345 Inset shows the relationships between number of DEGs and differentially spliced genes (DSGs) in each cichlid
346 radiationnplotted as function of divergence time. (**B)** Volcano plot illustrating DEGs in Lake Tanganyika (LT)
347 oral jaws. Significant DEGs are highlighted in red. Top gene outliers are annotated. (**C)** Gene Ontology
348 enrichment of DEGs between herbivores and carnivores. Only a subset of terms are shown that are known in
349 literature to be associated with jaw/craniofacial divergence. Full list is presented in Figure S9. Some candidate
350 genes for jaw morphogensis/craniofacial development that were convergently expressed in all three radiations
351 in the same direction (higher in herbivores or higher in carnivores). See File S1 for full list. (**D)** Ancestral state
352 reconstruction of GE of the convergently expressed *odam* gene. (**E)** Overlap of DEGs and DSGs between
353 herbivores and carnivores within each lake radiation.
354
355
356 **Contrasting functions for alternative splicing and gene expression**

357 We found dissimilar patterns in the extent of differential GE and AS underlying divergent trophic adaptation
358 across cichlid radiations. Even though the overall number of DEGs was much higher than DSGs across all three
359 radiations (Figures 4A, 5A, S6, S8), the relative number of DEGs across radiations was the highest in the oldest
360 radiation and lowest in the youngest radiation, while the DSGs showed the opposite pattern (Figure 5A inset).
361 Interestingly, more DEGs in each of the younger radiations overlapped with the older radiation, in contrast more
362 DSGs overlapped among the younger radiations (Figures 4A, 5A, S6A, S8). This pattern indicates that a
363 complex temporal interplay of GE and AS underlies adaptive radiation, with ancestral splice variation enabling
364 rapid ecological diversification at early stages of speciation, while GE differences evolve much slower, mostly
365 likely fine-tuning the adaptive morphology long after speciation. This would be consistent with the GO
366 enrichment results for differentially spliced and expressed genes that we found in younger versus the older
367 radiation.

368 It has been proposed that AS and GE perform contrasting biological functions(11, 55, 58–60). However, these
369 studies have lacked a phylogenetic perspective. In our dataset, the overlap of genes that were both differentially
370 spliced and differentially expressed between herbivores and carnivores followed a phylogenetic and
371 evolutionary pattern. The highest overlap was in Lake Tanganyika (OJ 56, PJ 45), followed by Lake Malawi
372 (OJ 28, PJ 13) and Lake Victoria (OJ 25, PJ 19) (Figure 5). Our results reinforce the idea that putatively adaptive
373 AS and GE changes act mostly at different genes – but we show that the extent of gene overlap depends on the
374 evolutionary age of lineages, with fewer genes being regulated by both procesess in more recently diverged
375 populations.

376

Convergent gene expression and splicing underlying convergent trophic evolution

African cichlid radiations are an ideal sytem to study the repeatability of evolution as they have evolved convergent trophic adaptations, particularly in jaw shape, to exploit similar food sources in similar ways in different lakes(19). See results of geometric morphometric analysis in the Supplementary material that illustrates that herbivores and carnivores cluster together in a jaw shape PCA regardless of which lake they are from; and that jaw shape is diverging along parallel axes in the three radiations. We asked if differentially expressed and spliced genes contributing to herbivory vs carnivory showed parallel divergence across all three radiations. We find convergent 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 (61). 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). These results suggest that jaw adaptation in cichlids is a polygenic trait with different genes functioning redundantly across these three radiations. The redundancy is higher in the regulation of genes through AS than GE, most likely due to the species-specific nature of AS.

Discussion

Bursts of diversification that characterise adaptive radiations have challenged the expectations of gradual evolution (62, 63). This divergence is often faster than the rate at which protein-coding mutations accumulate, making gene regulation an important mediator of rapid and diverse phenotypic change (3, 4, 8). However, we still understand little about how different gene regulatory mechanisms shape rapid adaptive radiations (5). Through the analysis of 200 transcriptomes representing three independent but phylogenetically nested cichlid adaptive radiations that evolved in the last <16,000 to 3.8 million years we show that the rich cache of ancestral isoform variation in the ancestor of East African haplochromine cichlids evolved faster than gene expression to rapidly to give rise to novel trophic adaptations. This pattern was most pronoucned in the youngest and most species rich adaptive radiations in Lakes Victoria and Malawi that had releatively higher differences in splicing than Lake Tanganyika of genes are associated with jaw morphogenesis. We further discovered that novel isoforms that most likely evolved in the course of adaptive radiation also contributed to adaptation. The rapid turnover of AS that we observed here 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 (38). Theory

suggests that relaxed selection can increase the rate of evolution of complex adaptations in heterogenous environments, allowing populations to rapidly cross adaptive valleys (64). We propose that relaxed selection maintaining standing splice variation, followed by periods of positive selection on alternative isoforms of craniofacial genes contributed to the rapid evolution of trophic adaptations in cichlids as they entered lakes and were exposed to new and ecologically more stable environments. An alternative explanation for our observation can be the complex influence of diverse forms of selection (perhaps in combination with neutral processes), such as balancing selection, frequency-dependent selection, and diversifying selection, that promote isoform diversity. Taken together, our findings challenge the long-held notion that most alternative isoforms are evolutionary noise (65), because this noise can fuel innovation when ecological opportunity arises, providing one possible mechanistic explanation for how these extraordinary cichlid radiations were able to exploit ecological opportunity as such extremely short timescales.

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 (40, 41, 66, 67). 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(26, 68, 69). They were able to do this by rapidly remodelling their jaw, dentition, and craniofacial morphology to forage foods in new and different ways (68). Our findings may reflect the particular importance of gene regulation for evolutionary transitions in Tropheini jaws during speciation. (74)

We find that signatures of splicing became increasingly more species-specific, and GE became more lineage-specific over evolutionary time. However, specific-specific signatures of splicing do not emerge until ~3.8 million years of evolution, a pattern not observed in previous large-scale studies on older lineages (40, 66). We also 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 was also reflected in the biological functions of these genes: (1) more jaw morphogenesis related gene ontology terms were enriched in differentially spliced genes in Lakes Victoria and Malawi than Tanganyika; (2) in Tanganyika more jaw morphogensis related gene ontology terms were enriched in differentially expressed genes. This pattern indicates 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, while GE differences evolve much slower, mostly likely fine-tuning the adaptive morphology long after 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 (15, 19, 26, 70).

447 The Wnt pathway emerged as a key player in the evolution of new jaw phenotypes at early stages of
448 development, with genes regulated through splicing (*gpc3*, *lzts2a*) and differential expression (*fzd2*). Overall,
449 more convergent differences in GE than AS were found to underlie convergen trophic jaw adaptations across
450 all three cichlid radiations. The lower phylogenetic convergence of AS may underscore its role in shaping
451 functional genetic redundancy - where multiple genetic pathways can produce similar phenotypic outcomes
452 (71). In our case, it is not isoform diversity within a single gene that underpins redundancy as classically
453 discussed (72) but rather the use of different genes regulated by lineage-specific splicing patterns, highlighting
454 a novel form of genetic redundancy mediated by AS. Such redundancy may allow different lineages to evolve
455 comparable traits via distinct molecular routes, increasing evolutionary flexibility and speed.

456

457 It has been hypothesised that ancestral variation plays an important role in the rapid adaptive capacity of African
458 cichlids (31, 73). But, little empirical evidence exists explaining how this ancestral variation mechanistically
459 plays out. We show that ancestral splice variants that were differentially spliced between jaws of herbivores and
460 carnivores are lowly expressed in non-radiating species but increased in frequency in radiating cichlids, most
461 likely enabling trophic divergence. Firstly, this suggests that old genetic variation can fuel new adaptations (31);
462 and that the ancestors of the East African cichlid adaptive radiations, which were most likely a widely dispersing
463 riverine cichlids, acted as reservoirs of genetic variation for diversification on a continental scale (74). And
464 secondly, even if most ancestral alternative isoforms are non-functional (65), lowly expressed, and evolving
465 stochastically (38), they represent a cache of standing variation for adaptive evolution when ecological
466 opportunity arises (5, 75). Standing variation in AS has been shown to be important for the rapid domestication
467 of sunflowers (75), which represents only one ecological transition, while cichlid adaptive radiations involve
468 multiple, often convergent, diverse ecological transitions and speciation events. In addition to ancestral
469 variation, we found evidence of potentially novel isoforms contributing to adaptive jaw differences. This
470 emphasises how fast new protein-coding variation can be generated for evolution through AS.
471
472 Our present study is one of the most comprehensive cross-lake analysis of transcriptomic variation in cichlid
473 fishes to date. However, it stills represents a small subset of the hundreds of cichlid species that are found in the
474 African Great Lakes. In the future, expanding the analyses to include more species would add more confidence
475 to our findings on convergent and divergent gene expression and splicing patterns underlying trophic evolution
476 in these radiations. Furthermore, including more cichlid species that entered lakes and had access to ecological
477 opoortunity but did not form radiations (76), or riverine cichlids that are hypothesised to have contributed
478 genetically to the lake radiations through hybridisation (25, 44, 45), would shed light on how different
479 evolutionary processes promote (or constrain) rapid diversification by re-wiring transcriptional variation.
480 Targeting a single, albeit ecologically and morphologically important, developmental stage is another constraint
481 of our approach. In the future, studying gene regulation across multiple developmental stages, in combination

482 with how jaw shape changes along ontogeny, would really enhance our understanding of gene regulation and
483 the origins of diversity.
484
485 Taken together, our findings suggest that AS may be one piece of the puzzle explaining how ecological and
486 species richness can evolve when sufficient time is not available for new protein-coding mutations to arise, fix,
487 and be replenished. Future studies on gene regulation in evolutionary biology would benefit from moving away
488 from a gene expession-centric view to a more holistic model that incoporates other levels of gene regulation,
489 such as splicing.
490

491 # Materials and Methods

492

493 **Study design and sampling**

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

499 Sample were collected by either through aquarium trade or fieldwork within the framework of collaborations
500 with Zambian and Tanzanian fisheries (TAFIRI). All fish were reared in the fish facility at University of Graz
501 under standardised conditions regarding tank environment, diet, and water parameters to minimise plastic
502 effects. Fertilised eggs were taken from the mouth of mouthbrooding females 24 hours after fertilisation, by
503 applying slight pressure onto their cheeks. The eggs were reared in a separate tank until the yolk sac was
504 completely absorbed, which is also known as the developmental stage 26 – end of larval stage and beginning of
505 juvenile stage, when bony elements have formed but the organism is still growing actively (77–79). Late larval
506 and juvenile developmental stages are ecomorphologically informative stages that have been used to study jaw
507 adaptation and divergence cichlids (11, 80, 81). The external morphology and internal bone and cartilage
508 development at stage 26 is illustrated in Fig. S11. Stage 26 larvae were euthanised and stored in RNAlater in a
509 fridge for seven days and then moved to a -80°C freezer. (11, 15, 72)

510

511 **RNA extraction**

512 Each larva was dissected under a stereo microscope to dissect the oral jaws (OJ: bone, teeth) and the pharyngeal
513 jaws (PJ: bones and teeth) separately (Figure S12). The bones were cleaned as much as possible, but some
514 surrounding tissue (i.e. skin, cartilage, muscle) remained, especially for the OJ, as small size of the larvae

515 (<1cm) makes the dissection challenging. The dissected tissue was submerged in a lysis buffer and 1-
516 thioglycerol according to the protocol of the Promega RNA Reliaprep kit. The sample in the lysis buffer was
517 crushed with a ceramic bullet in a homogenizer (FastPrep-24, MP Biomedicals, Santa Ana, CA, USA) before
518 further processing. Two individuals were pooled per sample to increase the amount of RNA for Lake Victoria
519 species. For Lake Malawi and Tanganyika, one individual yielded sufficient RNA. RNA quality was tested at
520 a TapeStation 2200 with RNA Screen Tapes. It was tried to reach a RINe number of >= 7 for each sample. We
521 extracted RNA for 5 biological replicates per jaw per species, which resulted in 200 mRNA samples in total.

522

523 **Library preparation and sequencing**

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

529

530 **Transcriptome assembly**

531 Each sample has approximately ten million paired-end reads. Raw sequence reads are available at NCBI
532 sequence read archive (Reviewer link:
533 https://dataview.ncbi.nlm.nih.gov/object/PRJNA640176?reviewer=p2dhiiern60seqodpuqbqb1o62). After a
534 quality check with Fastqc (v0.11.8)*(82)* and a trimming step with Trimmomatic (v0.3.9)(83), only reads with a
535 phred > 28 and a minimum length of 70 bp where retained. To enable the assembly of novel isoforms we
536 developed a bioinformatics pipeline with the following steps: Reads were assembled using STAR (v2.7.3.a)(84)
537 in reference guided mode using the *Oreochromis niloticus* reference genome*(85)*. This genome was selected
538 because it is the best assembled and annotated cichlid reference genome and is phylogenetically equidistant to
539 the species sampled in this study. Mapping statistics were generated with samtools idxstats (v1.9)(33, 86).
540 StringTie (v2.0.6) was run first both with a reference genome to assemble the RNA-Seq alignments into
541 potential transcripts and in genome-guided mode. This was done separately for the single files (per individual).
542 The single gtf files for each biological replicate were then incrementally merged according to species and tissue
543 and finally into one super annotation. This repeated merging steps were conducted to reduce the probability of
544 false positives in isoform assembly and accurately identify novel species-specific or lower expressed isoforms.
545 To estimate the accuracy of the produced annotation files we compared them with gffcompare (v0.11.2)*(87)* to
546 our reference *O. niloticus* annotation. We filtered for monoexonic transcripts not contained in the reference and
547 a class code assigned by gffcompare (v0.11.2), indicating 'possible polymerase run-on' fragments. The
548 maximum intron length was reduced to 200,000 bp, which is the maximum intron length found in the *O.niloticus*

549 reference. Based on the super annotation, the expression estimates were generated with StringTie allowing no
550 multimapping. From these expression estimates, count matrices were produced using a perl script from the
551 griffith lab (https://rnabio.org) to extract raw count data from StringTie results. The code used for our RNAseq
552 pipeline in on github(88). SUPPA2 (v2.3)(37) was used to calculate percent spliced in (PSI) index for each
553 isoform for each gene.

554

555 **Differential gene expression analysis**

556 We used DESeq2(34)in R(89) to detect DEGs running comparisons (1) between herbivores and carnivores
557 within each lake radiation for each tissue separately and (2) between oral and pharyngeal jaws within each
558 radiation. DESeq2 uses raw read counts and estimates variance-mean dependence based on a model that utilises
559 the negative binomial distribution. The cutoff for DEGs was chosen at a false discovery rate of ($q < 0.05$).

560

561 **Alternative splicing analysis**

562 Alternative splicing analysis was conducted with SUPPA2 (v2.3)(37) to evaluate differential transcript usage
563 (DTU) between between herbivores and carnivores within each lake radiation for each tissue separately.
564 Visualisation of PSI analysis was conducted in R(89).

565

566 **Modeling gene and isoform expression**

567 We used the methods proposed by Chen et al (2019)(42) to model gene and isoform expression along the
568 phylogeny using Ornstein-Uhlenbeck (OU) and Brownian Motion (BM) processes. The OU process is a
569 modification of a random walk, describing the change in expression (dX_t) across time (dt) by $dX_t = \sigma dB_t + \alpha (\theta$
570 $- X_t) dt$, where dB_t denotes a Brownian motion process. The model elegantly quantifies the contribution of both
571 drift and selective pressure for any given gene: (1) Drift is modeled by Brownian motion with a rate σ while (2)
572 the strength of selective pressure driving expression back to an optimal expression level θ is parameterized by
573 α. The OU process accounts for phylogenetic relationships, thus allowing us to fit individual evolutionary
574 expression trajectories.

575

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

577 We constructed expression trees using the neighbour-joining approach based on pairwise distance matrices
578 between samples using data from all genes and applying functions from the 'ape' package (v5.8.1)(90) in R(89).
579 Distances were computed as $1 - \rho$, where ρ is the Spearman's correlation coefficient, chosen due to its
580 robustness against outliers and any normalisation issues(41). 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)(91) in R. Ancestral state reconstruction was conducted using the 'phytools' package (v2.4.4)(92) in R.

Phylogenetic reconstruction (SNPs)

SNP calling on mapped mRNAseq reads was conducted using bcftools(93) across all samples. Filtered high quality SNPs were used to contruct a concatenated alingment with the most common allele at each position for each individual. This alignment was used as input for RAxML (v8)(94) to contstruct a SNP phylogeny.

GO term enrichment

GO terms for *Oreochromis niloticus* were acquired in R via the biomaRt package (v2.46.1)(95) from ensembl(96). Gene set enrichment analysis was conducted using topGO (v2.36.0)*(97)* 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.

(98)(2)(98, 99)(98, 99)

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

References

1. P. Neige, *Events of increased biodiversity* (ISTE Press - Elsevier, 2015).
2. D. Brawand, *et al.*, The genomic substrate for adaptive radiation in African cichlid fish. *Nature* **513**, 375–381 (2014).
3. M. King, A. C. Wilson, Evolution at two levels in humans and chimpanzees. *Science (1979).* **188**, 107–116 (1975).
4. D. L. Stern, Perspective: Evolutionary developmental biology and the problem of variation. *Evolution (N Y).* **54**, 1079–1091 (2000).

612 5. P. Singh, E. P. Ahi, The importance of alternative splicing in adaptive evolution. *Mol. Ecol.* **31**, 1928–
613 1938 (2022).
614 6. D. Berner, W. Salzburger, The genomics of organismal diversification illuminated by adaptive
615 radiations. *Trends in Genetics* **31**, 491–499 (2015).
616 7. M. Barrier, R. H. Robichaux, M. D. Purugganan, Accelerated regulatory gene evolution in an adaptive
617 radiation. *PNAS* **98**, 10208–10213 (2001).
618 8. G. A. Wray, The evolutionary significance of cis-regulatory mutations. *Nat. Rev. Genet.* **8**, 206–216
619 (2007).
620 9. H. Keren, G. Lev-Maor, G. Ast, Alternative splicing and evolution: diversification, exon definition and
621 function. *Nat. Rev. Genet.* **11**, 345–355 (2010).
622 10. W. Gilbert, Why genes in pieces? *Nature* **271** (1978).
623 11. P. Singh, C. Börger, H. More, C. Sturmbauer, The role of alternative splicing and differential gene
624 expression in cichlid adaptive radiation. *Genome Biol. Evol.* **9**, 2764–2781 (2017).
625 12. M. Carruthers, *et al.*, Ecological speciation promoted by divergent regulation of functional genes
626 within African cichlid fishes. *Mol. Biol. Evol.* **39** (2022).
627 13. Y. Terai, N. Morikawa, K. Kawakami, N. Okada, The complexity of alternative splicing of hagoromo
628 mRNAs is increased in an explosively speciated lineage in East African cichlids. *PNAS* **100**, 12798–
629 12803 (2003).
630 14. E. J. Richards, *et al.*, A vertebrate adaptive radiation is assembled from an ancient and disjunct
631 spatiotemporal landscape. *Proc. Natl. Acad. Sci. U. S. A.* **118** (2021).
632 15. P. Singh, E. P. Ahi, C. Sturmbauer, Gene coexpression networks reveal molecular interactions
633 underlying cichlid jaw modularity. *BMC Ecol. Evol.* **21**, 1–17 (2021).
634 16. O. Seehausen, *Lake Victoria rock cichlids* (Verduyn Cichlids, 1996).
635 17. K. Yamaoka, "Trophic ecomorphology of Tanganyikan cichlids." in *Fish Communities in Lake*
636 *Tanganyika*, N. M. Kawanabe H, Hori M, Ed. (Kyoto University Press, 1997), pp. 27–56.
637 18. A. J. Conith, R. C. Albertson, The cichlid oral and pharyngeal jaws are evolutionarily and genetically
638 coupled. *Nat. Commun.* **12**, 5477 (2021).
639 19. R. C. Albertson, T. D. Kocher, Genetic and developmental basis of cichlid trophic diversity. *Heredity*
640 *(Edinb).* **97**, 211–221 (2006).
641 20. G. Fryer, T. D. Iles, *The cichlid fishes of the Great Lakes of Afrika* (T.H.F. Publikations, Inc., Neptune
642 City, NJ, 1972).
643 21. Greenwood PH, Cichlid fishes of Lake Victoria, East Africa: the biology and evolution of a species
644 flock. *Bull Brit Mus (Nat Hist) Zool Suppl* **6**, 1–134 (1974).
645 22. T. D. Kocher, J. A. Conroy, K. R. McKaye, J. R. Stauffer, Similar Morphologies of Cichlid Fish in
646 Lakes Tanganyika and Malawi Are Due to Convergence. *Mol. Phylogenet. Evol.* **2**, 158–165 (1993).
647 23. K. A. Young, J. Snoeks, O. Seehausen, Morphological diversity and the roles of contingency, chance
648 and determinism in African cichlid radiations. *PLoS One* **4**, e4740 (2009).

24. K. F. Liem, Evolutionary Strategies and Morphological Innovations: Cichlid Pharyngeal Jaws. *Syst. Zool.* **22**, 425–441 (1973).
25. I. Irisarri*, *et al.*, Phylogenomics uncovers early hybridization and adaptive loci shaping the radiation of Lake Tanganyika cichlid fishes. *Nat. Commun.* **9**, 3159 (2018).
26. P. Singh, *et al.*, Phylogenomics of trophically diverse cichlids disentangles processes driving adaptive radiation and repeated trophic transitions. *Ecol. Evol.* **12**, e9077 (2022).
27. W. Salzburger, T. Mack, E. Verheyen, A. Meyer, Out of Tanganyika: genesis, explosive speciation, key-innovations and phylogeography of the haplochromine cichlid fishes. *BMC Evol. Biol.* **5**, 17 (2005).
28. S. Koblmüller, *et al.*, Age and spread of the haplochromine cichlid fishes in Africa. *Mol. Phylogenet. Evol.* **49**, 153–69 (2008).
29. H. Nakamura, M. Aibara, M. Nikaido, Ancient standing genetic variation facilitated the adaptive radiation of Lake Victoria cichlids. *Genes Genet. Syst.* **98**, 93–99 (2023).
30. H. Nakamura, *et al.*, Genomic Signatures for Species-Specific Adaptation in Lake Victoria Cichlids Derived from Large-Scale Standing Genetic Variation. *Mol. Biol. Evol.* **38**, 3111 (2021).
31. Y. H. E. Loh, *et al.*, Origins of shared genetic variation in African cichlids. *Mol. Biol. Evol.* **30**, 906–917 (2013).
32. M. A. Conte, W. J. Gammerdinger, K. L. Bartie, D. J. Penman, T. D. Kocher, 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).
33. M. Pertea, *et al.*, StringTie enables improved reconstruction of a transcriptome from RNA-seq reads. *Nat. Biotechnol.* **33**, 290–295 (2015).
34. M. I. Love, W. Huber, S. Anders, Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* **15**, 550 (2014).
35. C. J. Wright, C. W. J. Smith, C. D. Jiggins, Alternative splicing as a source of phenotypic diversity. *Nat. Rev. Genet.* **23**, 697–710 (2022).
36. E. T. Wang, *et al.*, Alternative Isoform Regulation in Human Tissue Transcriptomes. *Nature* **456**, 470 (2008).
37. J. L. Trincado, *et al.*, SUPPA2: Fast, accurate, and uncertainty-aware differential splicing analysis across multiple conditions. *Genome Biol.* **19**, 1–11 (2018).
38. G. Ast, How did alternative splicing evolve? *Nat. Rev. Genet.* **5**, 773–782 (2004).
39. K. Garg, P. Green, Differing patterns of selection in alternative and constitutive splice sites. *Genome Res.* **17**, 1015 (2007).
40. N. L. Barbosa-Morais, M. Irimia, Q. Pan, H. Y. Xiong, S. Gueroussov, The evolutionary landscape of alternative splicing in vertebrate species. *Science (1979).* **338**, 1587–1594 (2012).
41. D. Brawand, *et al.*, The evolution of gene expression levels in mammalian organs. *Nature* (2011). https://doi.org/10.1038/nature10532.

686 42. J. Chen, *et al.*, A quantitative framework for characterizing the evolutionary history of mammalian
687 gene expression. *Genome Res.* **29**, 53–63 (2019).
688 43. P. J. Wittkopp, *et al.*, Evolutionary changes in cis and trans gene regulation. *Nature* **430**, 85–88 (2004).
689 44. J. I. Meier, *et al.*, Ancient hybridization fuels rapid cichlid fish adaptive radiations. *Nat. Commun.* **8**,
690 14363 (2017).
691 45. H. Svardal, *et al.*, Ancestral Hybridization Facilitated Species Diversification in the Lake Malawi
692 Cichlid Fish Adaptive Radiation. *Mol. Biol. Evol.* **37**, 1100–1113 (2020).
693 46. J. I. Meier, *et al.*, Cycles of fusion and fission enabled rapid parallel adaptive radiations in African
694 cichlids. *Science (1979).* **381**, eade2833 (2023).
695 47. K. J. Parsons, A. Trent Taylor, K. E. Powder, R. C. Albertson, Wnt signalling underlies the evolution
696 of new phenotypes and craniofacial variability in Lake Malawi cichlids. *Nature Communications 2014*
697 *5:1* **5**, 1–11 (2014).
698 48. S. Paine-Saunders, B. L. Viviano, J. Zupicich, W. C. Skarnes, S. Saunders, glypican-3 controls cellular
699 responses to Bmp4 in limb patterning and skeletal development. *Dev. Biol.* **225**, 179–187 (2000).
700 49. Y. Li, Q. Li, Y. Longs, Z. Cui, Lzts2 Regulates Embryonic Cell Movements and Dorsoventral
701 Patterning through Interaction with and Export of Nuclear β-Catenin in Zebrafish. *Journal of*
702 *Biological Chemistry* **286**, 45116–45130 (2011).
703 50. P. Masonick, A. Meyer, C. D. Hulsey, A Kiss of Deep Homology: Partial Convergence in the Genomic
704 Basis of Hypertrophied Lips in Cichlid Fish and Human Cleft Lip. *Genome Biol. Evol.* **15**, evad072
705 (2023).
706 51. K. E. Powder, R. C. Albertson, Cichlid fishes as a model to understand normal and clinical craniofacial
707 variation. *Dev. Biol.* **415**, 338–346 (2016).
708 52. H. Yu, *et al.*, Frizzled 1 and frizzled 2 genes function in palate, ventricular septum and neural tube
709 closure: general implications for tissue fusion processes. *Development* **137**, 3707–3717 (2010).
710 53. N. Karagic, R. F. Schneider, A. Meyer, C. D. Hulsey, A Genomic Cluster Containing Novel and
711 Conserved Genes is Associated with Cichlid Fish Dental Developmental Convergence. *Mol. Biol.*
712 *Evol.* **37**, 3165–3174 (2020).
713 54. Y. Hu, *et al.*, A Novel Autosomal Recessive GJA1 Missense Mutation Linked to Craniometaphyseal
714 Dysplasia. *PLoS One* **8**, e73576 (2013).
715 55. A. Duenser*, *et al.*, Developmental tinkering of gene regulation facilitated super rapid adaptive
716 radiation. *bioRxiv* 2024.01.24.577063 (2024). https://doi.org/10.1101/2024.01.24.577063.
717 56. J. Dalcq, *et al.*, RUNX3, EGR1 and SOX9B form a regulatory cascade required to modulate BMP-
718 signaling during cranial cartilage development in zebrafish. *PLoS One* **7** (2012).
719 57. A. Busch, K. J. Hertel, Evolution of SR protein and hnRNP splicing regulatory factors. *Wiley*
720 *Interdiscip. Rev. RNA* **3**, 1–12 (2012).
721 58. C. E. Rodríguez-Ramírez, *et al.*, Molecular mechanisms of Eda-mediated adaptation to freshwater in
722 threespine stickleback. *Mol. Ecol.* **00**, 1–19 (2023).

59. A. M. Jakšić, C. Schlötterer, The interplay of temperature and genotype on patterns of alternative splicing in Drosophila melanogaster. *Genetics* **204**, 315–325 (2016).
60. A. Jacobs, K. R. Elmer, Alternative splicing and gene expression play contrasting roles in the parallel phenotypic evolution of a salmonid fish. *Mol. Ecol.* **30**, 4955–4969 (2021).
61. M. Seppala, *et al.*, Gas1 is a modifier for holoprosencephaly and genetically interacts with sonic hedgehog. *J. Clin. Invest.* **117**, 1575–1584 (2007).
62. G. G. Simpson, *Tempo and mode in evolution* (Columbia University Press, 1944).
63. D. Schluter, *The ecology of adaptive radiation* (Oxford University Press, Oxford, 2000).
64. J. Draghi, C. B. Ogbunugafor, L. Zaman, T. L. Parsons, Relaxed selection can speed the evolution of complex adaptations. *bioRxiv* 2024.07.09.602773 (2024). https://doi.org/10.1101/2024.07.09.602773.
65. M. L. Tress, F. Abascal, A. Valencia, Most alternative isoforms are not functionally important. *Trends Biochem. Sci.* **42**, 408–410 (2017).
66. J. Merkin, C. Russell, P. Chen, C. B. Burge, Evolutionary dynamics of gene and isoform regulation in mammalian tissues. *Science (1979).* **338**, 1593–1599 (2012).
67. A. El Taher, *et al.*, Gene expression dynamics during rapid organismal diversification in African cichlid fishes. *Nat. Ecol. Evol.* **5**, 243–250 (2021).
68. A. K. Wanek, C. Sturmbauer, Form, function and phylogeny: comparative morphometrics of Lake Tanganyika's cichlid tribe Tropheini. *Zool. Scr.* **44**, 362–373 (2015).
69. C. Sturmbauer, U. Hainz, S. Baric, E. Verheyen, W. Salzburger, Evolution of the tribe Tropheini from Lake Tanganyika: synchronized explosive speciation producing multiple evolutionary parallelism. *Hydrobiologia* **500**, 51–64 (2003).
70. E. P. Ahi, Signalling pathways in trophic skeletal development and morphogenesis: insights from studies on teleost fish. *Dev. Biol.* **420**, 11–31 (2016).
71. Á. J. Láruson, S. Yeaman, K. E. Lotterhos, The Importance of Genetic Redundancy in Evolution. (2020). https://doi.org/10.1016/j.tree.2020.04.009.
72. I. Romero, I. Ruvinsky, Y. Gilad, Comparative studies of gene expression and the evolution of gene regulation. *Nat. Rev. Genet.* **13**, 505–516 (2012).
73. D. A. Marques, J. I. Meier, O. Seehausen, A combinatorial view on speciation and adaptive radiation. *Trends Ecol. Evol.* **34**, 531–544 (2019).
74. D. Joyce, *et al.*, An extant cichlid fish radiation emerged in an extinct Pleistocene lake. **435** (2005).
75. C. C. R. Smith, *et al.*, Genetics of alternative splicing evolution during sunflower domestication. *PNAS* **115**, 6768–6773 (2018).
76. C. E. Wagner, L. J. Harmon, O. Seehausen, Ecological opportunity and sexual selection together predict adaptive radiation. (2012). https://doi.org/10.1038/nature11144.
77. K. Fujimura, N. Okada, 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).

760 78. K. Fujimura, N. Okada, Bone development in the jaw of Nile tilapia Oreochromis niloticus (Pisces:
761 Cichlidae). *Dev. Growth Differ.* **50**, 339–355 (2008).
762 79. K. Fujimura, N. Okada, Shaping of the lower jaw bone during growth of Nile tilapia Oreochromis
763 niloticus and a Lake Victoria cichlid Haplochromis chilotes: A geometric morphometric approach.
764 *Dev. Growth Differ.* **50**, 653–663 (2008).
765 80. E. Tetrault, B. Aaronson, M. C. Gilbert, R. C. Albertson, Foraging-induced craniofacial plasticity is
766 associated with an early, robust and dynamic transcriptional response. *Proceedings of the Royal*
767 *Society B: Biological Sciences* **291** (2024).
768 81. P. Le Pabic, W. J. Cooper, T. F. Schilling, Developmental basis of phenotypic integration in two Lake
769 Malawi cichlids. *Evodevo* **7**, 1–26 (2016).
770 82. FastQC: A quality control tool for high throughput sequence data. Available at:
771 www.bioinformatics.babraham.ac.uk/projects/fastqc.
772 83. A. M. Bolger, M. Lohse, B. Usadel, Trimmomatic: A flexible trimmer for Illumina sequence data.
773 *Bioinformatics* **30**, 2114–2120 (2014).
774 84. A. Dobin, *et al.*, STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* **29**, 15–21 (2013).
775 85. M. A. Conte, W. J. Gammerdinger, K. L. Bartie, D. J. Penman, T. D. Kocher, A High Quality
776 Assembly of the Nile Tilapia (Oreochromis niloticus) Genome Reveals the Structure of Two Sex
777 Determination Regions. *bioRxiv* 1–39 (2017). https://doi.org/10.1101/099564.
778 86. H. Li, *et al.*, The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25**, 2078–2079
779 (2009).
780 87. G. Pertea, M. Pertea, GFF Utilities: GffRead and GffCompare [version 1; peer review: 3 approved] .
781 *F1000Res.* **9** (2020).
782 88. P. Singh, Transcriptomics pipeline for accurate gene and isoform analysis . [Preprint] (2025).
783 Available at: https://doi.org/10.5281/zenodo.15393254.
784 89. R Core Team, R: A language and environment for statistical. [Preprint] (2021). Available at:
785 https://www.R-project.org/.
786 90. E. Paradis, K. Schliep, ape 5.0: an environment for modern phylogenetics and evolutionary analyses in
787 R. *Bioinformatics* **35**, 526–528 (2019).
788 91. T. Galili, dendextend: an R package for visualizing, adjusting, and comparing trees of hierarchical
789 clustering. *Bioinformatics* (2015). https://doi.org/10.1093/bioinformatics/btv428.
790 92. L. J. Revell, phytools: an R package for phylogenetic comparative biology (and other things). *Methods*
791 *Ecol. Evol.* **3**, 217–223 (2012).
792 93. P. Danecek, *et al.*, Twelve years of SAMtools and BCFtools. *Gigascience* **10** (2021).
793 94. A. Stamatakis, RAxML version 8: a tool for phylogenetic analysis and post-analysis of large
794 phylogenies. *Bioinformatics* **30**, 1312–1313 (2014).
795 95. D. Smedley, *et al.*, BioMart – biological queries made easy. *BMC Genomics* **10**, 22 (2009).
796 96. D. R. Zerbino, *et al.*, Ensembl 2018. *Nucleic Acids Res.* **46**, D754–D761 (2018).

97. A. Alexa, J. Rahnenfuhrer, topGO: Enrichment analysis for gene ontology. [Preprint] (2019).