Expression variations in Ectodysplasin-A gene (*eda*) may contribute to morphological divergence of scales in Haplochromine cichlids

Authors

Maximilian Wagner[1,2*]
Email: maximilian.wagner@uni-graz.at

Sandra Bračun[1*]
Email: sandra.bracun@uni-graz.at

Anna Duenser[1],
Email: anna.duenser@gmail.com

Christian Sturmbauer[1],
Email: christian.sturmbauer@uni-graz.at

Wolfgang Gessl[1],
Email: wolfgang.gessl@uni-graz.at

Ehsan Pashay Ahi[1,3],
Email: ehsan.pashayahi@helsinki.fi

1. Institute of Biology, University of Graz, Universitätsplatz 2, A-8010 Graz, Austria.

2. Department of Biology, University of Antwerp, Groenenborgerlaan 171, 2020 Antwerp, Belgium.

3. Organismal and Evolutionary Biology Research Programme, University of Helsinki, Viikinkaari 9, 00014, Helsinki, Finland.

Corresponding Authors:

Ehsan Pashay Ahi,
Email: ehsan.pashayahi@helsinki.fi

Christian Sturmbauer,
Email: christian.sturmbauer@uni-graz.at

*** The first two authors contributed equally to this study**

Abstract

Background: Elasmoid scales are one of the most common dermal appendages and can be found in almost all species of bony fish differing greatly in their shape. Whilst the genetic underpinnings behind elasmoid scale development have been investigated, not much is known about the mechanisms involved in the shaping of scales. To investigate the links between gene expression differences and morphological divergence, we inferred shape variation of scales from two different areas of the body (anterior and posterior) stemming from ten haplochromine cichlid species from different origins (Lake Tanganyika, Lake Malawi, Lake Victoria and riverine). Additionally, we investigated transcriptional differences of a set of genes known to be involved in scale development and morphogenesis in fish.

Results: We found that scales from the anterior and posterior part of the body strongly differ in their overall shape, and a separate look on scales from each body part revealed similar trajectories of shape differences considering the lake origin of single investigated species. Above all, nine as well as 11 out of 16 target genes showed expression differences between the lakes for the anterior and posterior dataset, respectively. Whereas in posterior scales four genes (*dlx5*, *eda*, *rankl* and *shh*) revealed significant correlations between expression and morphological differentiation, in anterior scales only one gene (*eda*) showed such a correlation. Furthermore, eda displayed the most significant expression difference between species of Lake Tanganyika and species of the other two younger lakes. Finally, we found genetic differences in downstream regions of *eda* gene (e.g. in the *eda-tnfsf13b* inter-genic region) that are associated with observed expression differences. This is reminiscent of a genetic difference in the *eda-tnfsf13b* inter-genic region which leads to gain or loss of armour plates in stickleback.

Conclusion: These findings provide evidence for cross-species transcriptional differences of an important morphogenetic factor, *eda*, which is involved in formation of ectodermal appendages. These expression differences appeared to be associated with morphological differences observed in the scales of haplochromine cichlids indicating potential role of eda mediated signal in divergent scale morphogenesis in fish.

Keywords

Scale morphology; gene expression; adaptive radiation; East African lakes; Lake Tanganyika; Lake Malawi; Lake Victoria; African cichlids

Background

Cichlids pose a great a model system for evolutionary biology, as they include some of the most striking examples of explosive speciation and adaptive radiation. Many aspects of their life history as well as their behaviour, coloration and feeding morphologies are well studied [1–3]. One of the most striking features is their repeated evolution of parallel eco-morphologies, especially across the radiations of the three East African Great Lakes, Lake Tanganyika (LT), Lake Malawi (LM) and Lake Victoria (LV) [4, 5]. These ecological adaptations are also the focus of many studies, as they promise the opportunity to shed light on different molecular mechanisms underlying repeated evolution and diversification [6, 7]. Regarding skeletal morphogenesis in particular the evolution of their jaws and their phenotypic plasticity are topics of ongoing research [7–12]. However, while the adaptive value of some of the investigated structures (e.g., feeding apparatus) can be more easily connected to certain ecological specializations [5, 13], this is not so obvious in others, such as scales.

Fish scales come in a vast array of different shapes and forms. As a part of the dermal skeleton, which amongst other structures also includes teeth, odontodes, spines and fin rays, these postcranial derivates evolved into morphologically and histologically diverse structures in Actinopterygii [14, 15]. Elasmoid scales, found in most of teleost species, form in the dermal mesenchyme and are mainly used for protection and hypothetically for hydrodynamic modifications [14, 16, 17]. While the elasmoid scales form relatively late in ontogeny and can take diverse forms, they share a composition consisting of three tissues, with elasmodin as the basal component formed in a characteristic plywood-like structure [15, 16]. Scale development, mostly studied in zebrafish, has been found to be orchestrated by several well-known pathways, including Hh, Fgf and Eda [16, 18–20], which are known to be also involved in the appendage formation across several vertebrate groups [21]. Mutations and allele variations in the Eda/Edar pathway, for example, have been linked to fish fin, scale and armour plate development as well as human and mouse hair and teeth growth [19, 22, 23]. Nevertheless, besides a recent extensive comparison of the scale morphology across

Lake Tanganyika cichlids [24], as well as a genetic study of scale shapes in two closely related Lake Malawi cichlids, which tied FgF signalling to scale shape variation [20], not much is known about the molecular mechanisms shaping the elasmoid scale.

In this study, we investigate the morphological differences in the anterior and posterior scales of 10 haplochromine cichlid fish species from three Great East African Lakes, i.e., Lake Tanganyika (LT), Lake Malawi (LM) and Lake Victoria (LV) as well as a riverine haplochromine cichlid species. After identification of a stably expressed reference gene, we also investigate transcriptional differences of a set of genes known to be involved in scale development and morphogenesis in fish. Finally, we tried to find links between the gene expression differences and morphological divergence in both anterior and posterior scales. Our results provide cross-species expression comparisons of scale related genes in haplochromine cichlids and implicate expression differences by which formation of distinct scale morphologies might be determined.

Methods

Fish husbandry and sampling

Ten haplochromine cichlid species; three species from Lake Tanganyika, four species from Lake Malawi, two species from Lake Victoria, and one riverine haplochromine species, were selected for this study (Fig. 1A). The fish were kept and raised in standardized tanks and rearing conditions with the same diet (Spirulina flakes) until they displayed mating behaviour. Between 5 to 11 adult females per species were sampled for morphological analysis and 4 adult females were sampled for gene expression investigation. The sampled fish species were sacrificed by euthanization in with 0.5 g MS-222/litre of water, and 5 anterior and posterior scales from left side of the body were removed for morphological analysis (Fig. 1B), whereas similar numbers of scales were taken from both sides and all anterior or posterior scales from each fish were pooled for gene expression part.

Morphological analysis

To infer shape differences of scales from divergent African cichlids from different lakes a 2D geometric morphometric framework was deployed. Due to major morphological differences of the scales they were separately investigated for the anterior and posterior part of the body (Fig. 1 B and C). Standardized images of scales were taken with a KEYENCE VHX-5000

digital microscope (KEYENCE Germany GmbH). 84 adult specimens from 10 cichlid species inhabiting the three major rift lakes which were reared under standardized aquarium conditions (*Astatotilapia burtoni* = 7; *Neochromis omnicaeruleus* = 10; *Petrochromis famula* = 11, *P. polyodon* = 7; *Paralabidochromis sauvage* = 5*; Simochromis diagramma* = 11; *Sciaenochromis fryeri* = 5; *Tropheops tropheops* = 9; *Labeotropheus trewavasae* = 9; Mz: *Maylandia zebra* = 10) were included for the geometric morphometric analyses. For each individual six scale replicates from the anterior and posterior part of the body were probed (Fig. 1B), leading to a total of 1.008 investigated scales. After randomizing pictures in tpsUtil v.1.6 (available at http://life.bio. sunysb.edu/morph/soft-utility.html), landmark digitization was conducted on a set of 7 fixed landmarks and 14 semi-landmarks (see Figure 1b for positions) in tpsDig v.2.26 (available at http://life.bio.sunysb.edu/ morph/soft-utility.html). To ensure consistency, this step was conducted by a single investigator. Generalized Procrustes superimposition [25] was performed in tpsRelw v.1.65 (available at http://life.bio.sunysb.edu/morph/soft- utility.html) and aligned landmark configurations were exported for further analysis in MorphoJ v.1.06 [26]. In MorphoJ, single observations obtained from the six replicates were averaged to get the mean shape for each landmark. A Principal Component analysis (PCA) was applied to infer variation in morphospace among scale position (anterior vs. posterior), single specimen, and species. Subsequent analyses were based on separated datasets for anterior and posterior scale landmark setting, whereas PC-scores were exported for linear discriminant function analyses (LDA) in PAST v.4.1 [27]. To reduce the number of variables and control for putative over-separation of groups [28], only the first four principal components were used for the LDA. PCA and LDA plots were visualized in R v3.1.2 [29].

RNA isolation and cDNA synthesis

As mentioned in the section above, 10 anterior and posterior scales from each fish were pooled for isolating the total RNA isolation in a single tube containing 0.25 mL of a tissue lysis buffer from Reliaprep RNA tissue miniprep system (Promega, #Z6111, USA) as well as one 1.4 mm ceramic bead to crush the scales. The scales were homogenized using a FastPrep-24 Instrument (MP Biomedicals, Santa Ana, CA, USA) and total RNA was extracted following the instructions provided by the manufacturer (adjusted protocol for small amounts of fibrous tissue). In summary, the instruction follows with mixing of the lysis buffer and homogenized scales with isopropanol and centrifuging the entire mix through a column

provided by the kit, several RNA washing steps and a final DNase treatment step. The RNAs were quantified by a Nanophotometer (IMPLEN GmbH, Munich, Germany) and their quality was checked with RNA ScreenTapes on an Agilent 2200 TapeStation (Agilent Technologies). Next, the RNA samples with a RNA integrity number (RIN) above six were applied to first strand cDNA synthesis using 300ng of RNA and High Capacity cDNA Reverse Transcription kit (Applied Biosystems). The synthesized cDNA from each RNA sample was diluted 1:5 times in nuclease-free water to conduct qPCR.

Gene selection, designing primers and binding site predictions

We selected eight candidate reference genes which have been frequently used in different studies of Haplochromine cichlids and have shown high expression levels in various connective tissues including skeletal tissues [10, 30–35]. Furthermore, we chose 16 target candidate genes, which are implicated in scale development and morphogenesis (Table 1). The primers were designed at conserved sequence of coding regions using the transcriptome data of several East African haplochromine species (*Astatotilapia burtoni*, *Aulonocara baenschi*, *Cyrtocara moorii*, *Pundamilia nyererei*, *Metriaclima zebra*, *Simochromis diagramma*, *Tropheus duboisi*, and *Gnathochromis pfefferi*) as well as two more distant species from different African cichlid tribes (*Oreochromis niloticus* and *Neolamprologus brichardi*) [7, 36–39]. The sequences from all the species were imported to CLC Genomic Workbench, version 7.5 (CLC Bio, Aarhus, Denmark), and after alignment, the exon/exon junctions were specified using the *Astatotilapia burtoni* annotated genome in the Ensembl database (http://www.ensembl.org) [40]. The primers were designed spanning exon junctions and a short amplicon size (<250 bp) as recommended to be optimal for qPCR quantification [41]. The primers were designed and assessed through Primer Express 3.0 (Applied Biosystems, CA, USA) and OligoAnalyzer 3.1 (Integrated DNA Technology) to minimize the occurrence of dimerization and secondary structures.

We retrieved downstream sequences (3'UTR and inter-genic region) of *eda* gene for all the species in this study from European Nucleotide Archive (ENA) and Sequence Read Archive (SRA) in order to identify changes in potential binding sites. To do this, we used genomic sequences of the haplochromine species; *A. burtoni* (GCA_000239415.1), *P. famula* (GCA_015108095.1), *N. omnicaeruleus* (SRR12700904), *P. polyodon* (GCA_015103895.1), *S. fryeri* (ERX1818621), *S. diagramma* (GCA_900408965.1), *M. zebra*

(GCA_000238955.1), *T. tropheops* (SAMEA2661272), *L. trewavasae* (SAMN12216683), and *P. sauvagei* (GCA_018403495.1). Next we identify the 3'UTR and inter-genic region of *eda* genes using the annotated genome of *A. burtoni* from Ensembl and aligned them using CLC Genomic Workbench. The different sequence motifs were identified and screened for potential TF binding sites using STAMP [42] and the PWMs obtained from the TRANSFAC database [43].

qPCR and data analysis

The qPCR reactions were generated using Maxima SYBR Green/ROX qPCR Master Mix (2X) (Thermo Fisher Scientific, Germany) and the amplifications were conducted on ABI 7500 real-time PCR System (Applied Biosystems). The qPCR setups followed the recommended optimal sample maximization method [44]. The qPCR program, dissociation step and calculation of primer efficiencies were performed as described in our previous study [45] (Additional file 1).

Three different algorithms were applied to validate the most stable reference genes; BestKeeper [46], NormFinder [47] and geNorm [48]. The Cq value of the most stable reference gene was used as normalization factor ($Cq_{reference}$) to calculate ΔCq of each target gene ($\Delta Cq_{target} = Cq_{target} - Cq_{reference}$). The lowest expressed sample in each expression comparison was used as a calibrator sample and rest of the samples were subtracted from its ΔCq value to calculate ΔΔCq values ($\Delta Cq_{target} - \Delta Cq_{calibrator}$). Relative expression quantities (RQ) were calculated through $E^{-\Delta\Delta Cq}$ [49]. In order to perform statistical analysis, fold differences (FD) were calculated by transformation of RQ values to logarithmic values [50]. The significant expression differences were calculated using ANOVA statistical tests, followed by Tukey's HSD *post hoc* tests. The correlations between gene expression and a morphometric parameter (canonical variate 1) were calculated through Pearson correlation coefficients (r) for each gene using R.

Results

Divergence in scale morphology

The principal component analysis (PCA) revealed a clear separation in overall average individual shape between anterior and posterior scales (Fig. 1D). PC1 and PC2 explained 78.3 % and 8.6 % of the total shape variation, respectively. Generally, on the first axis

anterior scales are anterior-posteriorly more compressed compared to the posterior body part (see deformation grids in Fig. 1D). Along the second PC axis changes can be observed in the shape of the posterior scale field (narrow vs. wide), as well as in the lateral edges of the anterior scale field (edges vs. round).

While comparing different species, large variation in overall shape can be observed in the dataset which is restricted to anterior scales only. *Petrochromis famula*, *Maylandia zebra* and *Sciaenochromis fryeri* occupy large parts of the morphospace and overall, less intraspecific variation can be observed in other species (Fig. 2A). In the anterior dataset changes along the PC1 explain 47.1 % of total variation, and mainly affect the circularity of the overall shape (i.e., that scales get more compressed towards positive values). Changes along the second PC, which explains 19.6 % of the total shape variation, affect the posterior scale field (compression vs. expansion). Compared to anterior scales, less intraspecific shape variation can be observed in posterior scales, whereas PC1 explains 54.8 % and PC2 18.9 % of the total variation, respectively (Fig. 2B). PC1 separates two major clusters (*S. diagramma* + *P. famula* vs. rest) whereas changes along the axis mainly contribute to a dorso-ventrally versus anterior-posterior compression of the scale and the roundness of the anterior scale field. Along PC2 shape changes affect the expansion (or compression) of the anterior and posterior scale fields. Generally, the PCA only poorly resolves the lake (or phylogenetic) origin of the single species for both the anterior and posterior dataset.

The linear discriminant function analysis (LDA) of anterior as well as posterior dataset correctly classified 77.38 % (jackknifed: 67.86 %) and 72.62 % (jackknifed: 65.48%) of species (Fig. 2C and D). The first axis explains 83.66 % and 76.00 % variance of the overall shape variability for the anterior and posterior dataset, respectively. In the anterior dataset, the first LD-axis separates three major clusters made up of samples from Lake Tanganyika, the riverine *Astatotilapia burtoni* and a joint Victoria-Malawi cluster. Similar results were obtained for the posterior dataset, whereas along the first axis the separation between the riverine *A. burtoni* and the Victoria-Malawi cluster is less prominent. Along the second axis, which explains 10.04 % and 13.2 % of the variance of the overall shape variability for the anterior and posterior dataset, respectively, mainly interspecific and intraspecific variation is portrayed. Overall, for the anterior dataset, 83.33% (jackknifed: 75 %) of species were correctly classified according to the lake origin, whereas single classification scores reached values of 100 % (jackknifed: 85.71 %) for *A. burtoni*, as well as 66.67 % (jackknifed: 57.58

%) for Malawi, 86.67 % (jackknifed: 86.67 %) for Victoria and 96.55 % (jackknifed: 86.21 %). In total, for the posterior dataset, 77.38 % (jackknifed: 70.24 %) of the individuals were correctly assigned to the lake origin, with 85.71 % (jackknifed: 71.43 %) for *A. burtoni*, as well as 66.67 % (jackknifed: 54.55 %) for Malawi, 53.33% (jackknifed: 46.67 %) for Victoria and 100 % (jackknifed: 100 %) for individuals from Tanganyika.

Validation of stable reference genes

To quantify the expression levels of the selected target genes, the validation of stable reference gene(s) with least variation in expression across the anterior and posterior scales of different species is a necessary step [51]. The 8 candidates were selected from frequently used reference genes in studies of different tissues in East African cichlids [10, 30–35]. The candidate reference genes showed variable expression levels in the scales, and from highest to lowest expressed were respectively; *actb1*, *hsp90a*, *rps11*, *rps18*, *hprt1*, *gapdh*, *elf1a* and *tbp*. Interestingly, in both anterior and posterior scales, all the three software ranked *actb1* as the most stable reference gene with lowest expression variation across the cichlid species in this study (Table 2). Thus, we used the Cq value of *actb1* as normalization factor (NF) in each sample for quantification of relative expression analyses of the target genes.

Gene expression differences between anterior and posterior scales

The relative expression levels of 16 candidate target genes, *bmp4*, *col1a2*, *ctsk*, *dlx5*, *eda*, *edar*, *fgf20*, *fgfr1*, *mmp2*, *mmp9*, *opg*, *rankl*, *runx2a*, *sema4d*, *shh* and *sp7*, were compared between the anterior and posterior scales in each of the haplochromine species (Fig. 3). Some of these genes such as *bmp4*, *col1a2*, *rankl* and *sp7*, showed almost no expression difference between the anterior and posterior scales. Moreover, none of the target genes showed consistent expression difference across all the species. These indicate potential involvement of various genes in morphological divergence between the anterior and posterior scales. However, two genes, *ctsk* and *shh* exhibited expression difference between the anterior and posterior scales in most of the species (Fig. 3). The directions of expression differences between the anterior and posterior scales for *ctsk* and *shh* were variable depending on the species. Interestingly all the three species form LT showed higher expression in the anterior scale for *shh*, whereas the all the species from LM and LV showed tendency for opposite pattern with increased posterior scale expression. These findings suggest potential role of *ctsk* and *shh* in morphological divergence of the scales along anterior-posterior axis.

Gene expression differences between lakes in anterior and posterior scales

Next, we compared the expression levels of the target genes between the lakes in the anterior or posterior scales by considering all the species from each lake as one group (Fig. 4). In the anterior scales, nine out of 16 target genes showed expression differences between the lakes including *bmp4*, *ctsk*, *eda*, *edar*, *mmp2*, *opg*, *rankl*, *shh* and *sp7*. Most of these differences were between LT and the other lakes, and 4 genes, *ctsk*, *mmp2*, *opg* and *rankl* showed higher expression in LT species, while 3 genes, *bmp4*, *eda* and *shh* showed lower expression LT species. Furthermore, 2 genes, *ctsk* and *eda* displayed the strongest expression differences between the lakes in opposite patterns suggesting their role in morphological divergence of the anterior scales across the lakes (Fig. 4). In the posterior scales, 11 out of 16 target genes showed expression differences between the lakes including *bmp4*, *col1a2*, *ctsk*, *eda*, *edar*, *fgf20*, *fgfr1*, *mmp2*, *opg*, *rankl*, *shh* and *sp7*. Again, most of these differences in the posterior scales were between LT and the other lakes, and 5 genes, *col1a2*, *ctsk*, *mmp2*, *opg* and *rankl* showed higher expression in LT species, while 3 genes, *eda*, *fgf20* and *shh* showed lower expression LT species. In addition, four genes, *fgf20*, *rankl*, *eda* and *shh* displayed the strongest expression differences between the lakes in opposite patterns (*eda* and *rankl* higher in LT, and *fgf20* and *shh* lower in LT) suggesting their role in morphological divergence of the posterior scales across the lakes (Fig. 4). In general, more genes with stronger expression differences between the lakes were observed the posterior scales. Several genes such as *ctsk*, *eda*, *edar*, *mmp2*, *opg*, *rankl* and *shh* appeared to have similar patterns of expression differences between the lakes in both anterior and posterior scales. Importantly, we found only one gene, *eda*, to have strong differences between the lakes in both anterior and posterior scales indicating its potentially crucial role in morphological divergence of the scales across the lakes.

Correlation analyses between gene expression and morphological divergence in scales

We analysed the correlation between expression of the genes and canonical variate 1 in the anterior or posterior scales across the species. Only one gene, *eda*, showed significant correlation in the anterior scales (Fig. 5), whereas, in the posterior scales 4 genes including *dlx5*, *eda*, *rankl* and *shh* displayed significant correlations between expression and morphological differences (Fig. 6). Among these genes *eda* exhibited the strongest correlation in the posterior scales. However, the correlation patterns differed between the genes in the posterior scales, i.e. eda and shh showed positive while *dlx5* and *rankl* had

346 negative correlations with the morphological changes based canonical variate 1. Therefore,
347 again only one gene, *eda*, showed significant correlation between its expression and the
348 morphological differences in both scales indicating its potential role in divergent scale
349 morphogenesis in the cichlid species. The opposite correlation patterns in the posterior scales
350 might also indicate inhibitory regulatory connections between the genes.
351
352 **Genetic differences in non-coding sequences of *eda* gene**
353 Finally, we were interested to investigate genetic differences in available regulatory
354 sequences of *eda* gene including 5´UTR, 3´UTR and short but conserved inter-genic region
355 between *eda* and *tnfsf13b* (immediate downstream gene) across the species. Interestingly, we
356 found two genetic differences (mutations/deletions) in 3´UTR and one in *eda-tnfsf13b* inter-
357 genic region to differ between LT species versus LM and LV species (Table 3). Next, we
358 parsed the short sequence regions containing the mutations/deletions against transcription
359 factor binding site (TFBS) databases. We found that the two changes in 3´UTR seem to lead
360 to gaining TFBS for transcription factors Mef2 and Tcf1 in the LM and LV species, whereas
361 the changes in the inter-genic region led to gaining TFBS for Lef1 transcription factor in the
362 LT species (Table 3). Importantly, the riverine species A.b appeared to have intermediate
363 genetic changes meaning that for the two changes in 3'UTR it showed a deletion similar to
364 the LT species but a mutation similar to the LV and LM species. Also, for the inter-genic
365 change, A.b showed an intermediate mutation between the LT and the other species from LM
366 and LV, however, this mutation showed no gain of TFBS (similar to the LM and LV species).
367 Taken together, these genetic changes showed similarity with differences in gene expression
368 and scale morphology, where the LT species clustered different from LM and LV species and
369 the riverine species (A.b.) showed intermediate differences. This suggests that the identified
370 genetic changes might be the underlying factors for divergent *eda* expression as well as
371 differences in the scale morphology.
372
373 ## Discussion
374 As river-adapted haplochromine cichlids repeatedly seeded adaptive radiations in several East
375 African lakes, cichlid fishes recurrently adapted to corresponding trophic niches. Thereby, the
376 adaptive value of traits is often mirrored by morphological shape parallelism and concomitant
377 similar lifestyles which result from parallel evolution [4, 5]. Hence, fishes from the cichlid
378 species flocks in various African lakes comprise an exciting model to conduct comparative

morphological and molecular studies. While most previous studies focused on bony elements that can easily be linked particular trophic niches and divergent natural selection as driver of diversification in cichlids (e.g., [52]), other skeletal structures such as scales might show less obvious adaptive trajectories.

Above all, between the three East African Great Lakes, the haplochromine cichlids are especially interesting, as they share common ancestry and comprise the Tropheini at LT and the entire the LV and LM haplochromine radiations [53, 54]. As the lakes have all very different geological histories, with Lake Tanganyika being the oldest [55], LM the intermediate [56] and LV the youngest of the three [57], they also depict three extensive radiations at different time points. Thus, depending on the evolutionary age of the different lakes, species (and their morphologies) had more or less time to diverge, despite sharing parallels. The more time passes, much more elaborated predator-prey and host-parasite relationships can evolve. This is manifested by unique ecological and behavioural features, particularly in the oldest of the three lakes, LT, which contains coocoo-catfish species showing brood parasitism [58], dwarfed gastropod shell breeders [59], putative cleaning behaviour [60], or highly elaborated scale eaters [61] (but also see *Genyochromis mento* from Lake Malawi). Particularly the latter case, scale eating, could have influenced the co-evolution of scale morphology in host species. Lake Tanganyika's scale eaters (i.e., *Perissodus*; Perissodini) show different degrees of specialization, whereas only the shallow water species, *Perrisodus microlepis* and *P. straeleni*, feed almost exclusively on fish scales while other species are not that specialized [4, 61]. The most common prey species of *P. microlepis* are members of the Tropheini and Eretmodini [62]. *P. straeleni* seems to be less specialised to certain prey items, but Tropheini scales still make up a major part of the gut contents [63]. Based on our dataset (Tropheini only) it remains speculative to assume that scale-related gene expression and the concomitant morphology might reflect an adaptation to reduce the risk of scale predation. Future studies, including more early branching (non-modern) haplochromine cichlids (e.g., *Pseudocrenilabrus*, *Thoracochromis*, *Astatoreochromis*) or other Tropheini with different lifestyles (e.g., *Ctenochromis*) will be necessary to establish stronger links between scale morphology and this unique predation pressure.

Nonetheless, understanding which genetic mechanisms underlie the scale morphology might be the key to understand how such similar and/or divergent eco-morphologies evolved.

Perhaps the most striking finding of our study was the highly significant differential expression of *eda* between LT species versus the species from the younger lakes (LM and LV species) in both anterior and posterior scales (Fig. 4). Interestingly, the *eda* expression in the riverine species (Ab), which is believed to be an ancestral species to haplochromine cichlids of the three great African lakes, was at intermediate level between LT and the species from LM and LV in both anterior and posterior scales. Moreover, the expression patterns of *eda* in both scales were highly correlated with morphological divergence across the species in this study. Ectodysplasin A (*eda*) encodes a member of the tumor necrosis factor family and mediates a signal conserved across vertebrates which is essential for morphogenesis of ectodermal appendages, such as scale, hair and feathers [22]. The eda signal is mediated through its receptor (encoded by *edar*) and initiated upon binding of eda to edar on the surface of a target cell [22]. In human, mutations in components of eda signal can cause hypohidrotic ectodermal dysplasia (HED) which is characterized by reduction and abnormal teeth morphology, absence or reduction of sweating glands and hair [64]. Similarly, impaired eda signal in zebrafish and medaka can lead to reduction in the number of scales and teeth [18, 19]. In sculpin (*Cottus*) fishes, genetic changes in the receptor gene (*edar*) has been found to be associated with morphological variations in body prickles (calcified spicules embedded in the skin), which are homologous structures to fish scales [65]. A later study in a highly derived order of teleosts, Tetraodontiformes, which includes ocean sunfishes, triggerfishes and pufferfishes, also showed the importance of eda signaling pathway in developmental formation and morphological variations of dermal spines (an extreme scale derivative) [66]. In stickleback, a mutation within an inter-genic region between *eda* and *tnfsf13b* genes leads to changes in transcriptional responsiveness of *eda* to its upstream Wnt signaling pathway and consequently impairment of armor plate formation [67].

In this study, we also found genetic changes in 3'UTR and the inter-genic region between *eda* and *tnfsf13b* genes that could explain the differences in *eda* expression across the Haplochromine cichlids (Table 3). The genetic changes resulted gain or loss of motifs which were predicted to be binding sites for transcription factors encoded by *mef2*, *tcf1* and *lef1* genes. These changes always discriminated the LT species from the species from LM and LV, whereas the riverine species had changes which could be considered an intermediate to both groups. Interestingly, all of the three predicted transcription factors (*mef2*, *tcf1* and *lef1*) are linked to Wnt signaling pathway. It is already known that mef2 can enhance canonical Wnt signal [68] and it is involved in osteogenesis as well [69–71]. The binding site motif for

mef2 appeared to be deleted in 3'UTR of the LT and riverine species. On the other hand, a binding site motif for tcf1 was gained in 3'UTR of the LM, LV and riverine species. In mice, *tcf1* is demonstrated to be involve in paraxial mesoderm and limb formation and appeared to be act downstream of Wnt signal similar to lef1 transcription factor [72]. Moreover, canonical Wnt signaling has been shown to regulate osteogenesis through tcf1 responsive element on regulatory sequence of runx2 in mammals [73]. The third motif predicted to be a binding site for lef1 and only found within *eda* - *tnfsf13b* inter-genic region of LT species. Lef1 is again a well-known mediator of canonical Wnt signaling pathway which inhibits final stage of osteoblast differentiation [74] but it is essential for osteoblast proliferation and normal skeletal development [75, 76]. During development lef1 function is shown to be essential for scale outgrowth in zebrafish [77], and *eda* expression is known to be regulated by Wnt signal through *lef1* transcriptional activity in mammals [78, 79]. In stickleback, mutation in an inter-genic region between *eda* and *tnfsf13b* genes is suggested to affect a binding site for c-jun transcription factor which its interaction with lef1 is required for *eda* transcriptional response to Wnt signal during armor plate formation [67]. Taken together, these observations, suggest mutations in enhancer sequences required for binding of Wnt signal components as potential underlying reason for the divergent expression of *eda* in both anterior and posterior scales of the cichlid species in this study.

In the posterior scale, in addition to *eda*, three more genes, *dlx5*, *rankl* and *shh*, displayed expression correlation with morphological divergence across the cichlid species (Fig. 6). The first gene, distal less homeobox 5 or *dlx5*, encodes transcription factor stimulating osteoblast differentiation and bone development, and it is also implicated in scale development and regeneration in fish [80–82]. Apart from its role in skeletogenesis, *dlx5* has been found to be involved in divergent development and morphogenesis of other tissues in cichlids such as teeth and nuchal hump [39, 83, 84]. In goldfish, *dlx5* expression appeared to be important at early stages of scale regeneration [80], and in both zebrafish and goldfish, *dlx5* transcription in scale can be affected by environmental clues such as mechanical stimulus [81, 82]. The second gene, *rankl*, encodes a ligand for osteoprotegerin (opg) and play crucial role in osteoclast differentiation and bone remodelling. Changes in *rankl* transcription appeared to be important during scale regeneration in goldfish [80, 85], as well as intercellular communications regulating scale bone remodeling in zebrafish and goldfish [81, 86]. Both *dlx5* and *rankl* have shown expression correlation patterns opposite to *eda* and *shh* in the posterior scales. Although, direct regulatory connections between these factors have not been

investigated in scale but these findings suggest their potential interactions at transcriptional level. Moreover, higher expression of *rankl* in the scales of LT species compared to LM and LV species might indicate higher level of bone remodelling in their scales.

The third gene, sonic hedgehog or *shh,* encodes a ligand of hedgehog signaling pathway which is shown to control scale morphogenesis in relationship with the formation of the epidermal fold in the posterior region of scale in fish [16]. In zebrafish, epidermal expression of *shh* has been shown to regulate scale regeneration through controlling osteoblast population and affecting directional bone growth [87]. We found similar expression pattern between *eda* and *shh* which is more pronounced in the posterior scales. This is consistent with previous findings in other vertebrates, for instance, *eda* has been demonstrated to act upstream of *shh* and induce *shh* expression during ectodermal organogenesis in mammals (e.g. during hair placode formation) [88–92]. Furthermore, it has been shown that the *eda*-dependent regulation of *shh* might be a part of larger molecular cascade in which an upstream signal such as Wnt pathway activates *eda* signal and in turn *eda* induces *shh* transcription [90, 92, 93]. These observations suggest potential role of *Wnt-eda-shh* axis in divergent scale morphogenesis across Haplochromine cichlids, which seems to be more pronounced in the posterior scales.

Conclusions

This is the first attempt to study cross-species association between gene expression and morphological divergence in scales of cichlids from different lakes. Our results provide evidence for potential role of a key signal mediated by *eda* gene to be involved in divergent morphogenesis of scale in closely related cichlid species. We show that *eda* expression has lower level in the scales of species from the older lake (Lake Tanganyika) and correlates with the observed shape variations across species. Our findings shed light on molecular basis of morphological divergence of a less studied skeletal element; however, further investigations are required to understand whether these differences have adaptive relevance in ecological and evolutionary-developmental contexts.

List of abbreviations

LT: Lake Tanganyika, **LM**: Lake Malawi, **LV**: Lake Victoria, ***bmp4***: bone morphogenetic protein 4, ***col1a2***: collagen type I alpha 2 chain, ***ctsk***: cathepsin K, ***dlx5***: distal-less homeobox 5, ***eda***: ectodysplasin A, ***edar***: ectodysplasin A receptor, ***fgf20***: fibroblast growth factor 20,

fgfr1: fibroblast growth factor receptor 1, ***mmp2***: matrix metallopeptidase 2, ***mmp9***: matrix metallopeptidase 9, ***opg***: osteoprotegerin, ***rankl***: receptor activator of nuclear factor kappa B ligand, ***runx2a***: runt-related transcription factor 2 alpha, ***sema4d***: semaphorin 4D, ***shh***: sonic hedgehog signaling molecule, ***sp7***: osterix transcription factor.

Declarations

Authors' contributions

EPA, SB, MW and CS designed the study. SB, EPA, MW, and AD conducted the laboratory experiment, measurements and figure preparations. MW and EPA analysed the data, and EPA, MW, AD and CS wrote the manuscript. WG and AD performed fish breeding and sampling. WG photographed the adult fishes used in Figure 1A. All authors reviewed the manuscript and approved its content.

Acknowledgements

The authors acknowledge Institute of Biology at University of Graz for providing fish breeding and laboratory facilities, and the Austrian Science Fund for the financial support. We also thank Stephan Koblmüller for sharing his precious knowledge on cichlid fishes of the African Lakes.

Competing interests

The authors declare that they have no competing interests.

Availability of data and materials

All data generated or analysed during this study are included in this published article.

Consent for publication

Not applicable.

Ethics approval and consent to participate

Fish keeping and euthanasia were conducted under permit BMWFW-66.007/0004-WF/V/3b/2016 issued by the Federal Ministry of Science, Research and Economy of Austria (BMWFW) in accordance with the ethical guidelines and regulations of the BMWFW. Fish

keeping and sampling was carried out in our certified aquarium facility according to the Austrian animal welfare law.

Funding

This study was funded by the Austrian Science Fund (Grant P29838). The Austrian Science Fund requires clarification of all legal issues concerning animal keeping, animal experiments and sampling design prior to grant submission and evaluation, but does not interfere in writing and data interpretation, but funds open access of the resulting publications.

References

1. Fryer G, Iles TD. The Cichlid Fishes of the Great lakes of Africa: their Biology and Evolution. Edinburgh: Oliver and Boyd.; 1972.

2. Kocher TD. Adaptive evolution and explosive speciation: The cichlid fish model. Nature Reviews Genetics. 2004;5.

3. Salzburger W. Understanding explosive diversification through cichlid fish genomics. Nature Reviews Genetics. 2018;19.

4. Muschick M, Indermaur A, Salzburger W. Convergent evolution within an adaptive radiation of cichlid fishes. Curr Biol. 2012;22:2362–8.

5. Wanek AK, Sturmbauer C. Form, function and phylogeny: comparative morphometrics of Lake Tanganyika's cichlid tribe Tropheini. Zool Scr. 2015;44:362–73.

6. Albertson RC, Kocher TD. Genetic and developmental basis of cichlid trophic diversity. Heredity (Edinb). 2006;97:211–21. doi:10.1038/sj.hdy.6800864.

7. Singh P, Börger C, More H, Sturmbauer C. The role of alternative splicing and differential gene expression in cichlid adaptive radiation. Genome Biol Evol. 2017;9:2764–81.

8. Albertson RC, Streelman JT, Kocher TD, Yelick PC. Integration and evolution of the cichlid mandible: the molecular basis of alternate feeding strategies. Proc Natl Acad Sci U S A. 2005;102:16287–92. doi:10.1073/pnas.0506649102.

9. Muschick M, Barluenga M, Salzburger W, Meyer A. Adaptive phenotypic plasticity in the Midas cichlid fish pharyngeal jaw and its relevance in adaptive radiation. BMC Evol Biol. 2011;11:116. doi:10.1186/1471-2148-11-116.

10. Ahi EP, 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. 2019;19:150. doi:10.1186/s12862-019-1483-3.

11. Singh P, Ahi EP, Sturmbauer C. Gene coexpression networks reveal molecular interactions underlying cichlid jaw modularity. BMC Ecol Evol. 2021;21:1–17. doi:10.1186/s12862-021-01787-9.

12. Parsons KJ, Concannon M, Navon D, Wang J, Ea I, Groveas K, et al. Foraging environment determines the genetic architecture and evolutionary potential of trophic morphology in cichlid fishes. Mol Ecol. 2016;25.

13. van Rijssel JC, Hoogwater ES, Kishe-Machumu MA, Reenen E van, Spits K V., van der Stelt RC, et al. Fast adaptive responses in the oral jaw of Lake Victoria cichlids. Evolution (N Y). 2015;69:179–89. doi:10.1111/evo.12561.

14. Huysseune A, Sire JY. Evolution of patterns and processes in teeth and tooth-related tissues in non-mammalian vertebrates. Eur J Oral Sci. 1998;106 1 SUPPL.

15. Sire JY, Huysseune A. Formation of dermal skeletal and dental tissues in fish: A comparative and evolutionary approach. Biological Reviews of the Cambridge Philosophical Society. 2003;78.

16. Sire JY, Akimenko MA. Scale development in fish: A review, with description of sonic hedgehog (shh) expression in the zebrafish (Danio rerio). International Journal of Developmental Biology. 2004;48:233–47.

17. Wainwright DK, Lauder G V. Mucus Matters: The Slippery and Complex Surfaces of Fish. 2017.

18. Kondo S, Kuwahara Y, Kondo M, Naruse K, Mitani H, Wakamatsu Y, et al. The medaka rs-3 locus required for scale development encodes ectodysplasin-A receptor. Curr Biol. 2001;11:1202–6.

19. Harris MP, Rohner N, Schwarz H, Perathoner S, Konstantinidis P, Nüsslein-Volhard C. Zebrafish eda and edar Mutants Reveal Conserved and Ancestral Roles of Ectodysplasin Signaling in Vertebrates. PLoS Genet. 2008;4:e1000206. doi:10.1371/journal.pgen.1000206.

20. Albertson RC, Kawasaki KC, Tetrault ER, Powder KE. Genetic analyses in Lake Malawi cichlids identify new roles for Fgf signaling in scale shape variation. Commun Biol. 2018;1:1–11.

21. Biggs LC, Mikkola ML. Early inductive events in ectodermal appendage morphogenesis. Seminars in Cell and Developmental Biology. 2014;25–26.

22. Mikkola ML, Thesleff I. Ectodysplasin signaling in development. Cytokine and Growth Factor Reviews. 2003;14:211–24.

23. Colosimo PF, Hosemann KE, Balabhadra S, Villarreal G, Dickson H, Grimwood J, et al. Widespread parallel evolution in sticklebacks by repeated fixation of ectodysplasin alleles.

614 Science (80-). 2005;307.
615 24. Viertler A, Salzburger W, Ronco F. Comparative scale morphology in the adaptive
616 radiation of cichlid fishes (Perciformes: Cichlidae) from Lake Tanganyika. Biol J Linn Soc.
617 2021. doi:10.1093/biolinnean/blab099.
618 25. Rohlf FJ, Slice D. Extensions of the Procrustes Method for the Optimal Superimposition
619 of Landmarks. Syst Zool. 1990;39:40. doi:10.2307/2992207.
620 26. Klingenberg CP. MorphoJ: an integrated software package for geometric morphometrics.
621 Mol Ecol Resour. 2011;11:353–7. doi:10.1111/j.1755-0998.2010.02924.x.
622 27. Hammer Ø, Harper DAT, Ryan PD. Past: Paleontological statistics software package for
623 education and data analysis. Palaeontol Electron. 2001;4.
624 28. Mitteroecker P, Gunz P. Advances in Geometric morphometrics. Evol Biol. 2009;36:235–
625 47.
626 29. R Core Team. R: A Language and Environment for Statistical Computing. 2017.
627 30. Yang CG, Wang XL, Tian J, Liu W, Wu F, Jiang M, et al. Evaluation of reference genes
628 for quantitative real-time RT-PCR analysis of gene expression in Nile tilapia (Oreochromis
629 niloticus). Gene. 2013;527:183–92.
630 31. Gunter HM, Fan S, Xiong F, Franchini P, Fruciano C, Meyer A. Shaping development
631 through mechanical strain: the transcriptional basis of diet-induced phenotypic plasticity in a
632 cichlid fish. Mol Ecol. 2013;22:4516–31.
633 32. Ahi EP, Richter F, Sefc KM. A gene expression study of ornamental fin shape in
634 Neolamprologus brichardi, an African cichlid species. Sci Rep. 2017.
635 33. Ahi EP, Duenser A, Singh P, Gessl W, Sturmbauer C. Appetite regulating genes may
636 contribute to herbivory versus carnivory trophic divergence in haplochromine cichlids. PeerJ.
637 2020;8:e8375. doi:10.7717/peerj.8375.
638 34. Ahi EP, Richter F, Lecaudey LA, Sefc KM. Gene expression profiling suggests
639 differences in molecular mechanisms of fin elongation between cichlid species. Sci Rep.
640 2019;9.
641 35. Lecaudey LA, Sturmbauer C, Singh P, Ahi EP. Molecular mechanisms underlying nuchal
642 hump formation in dolphin cichlid, Cyrtocara moorii. Sci Rep. 2019;9:20296.
643 doi:10.1038/s41598-019-56771-7.
644 36. Brawand D, Wagner CE, Li YI, Malinsky M, Keller I, Fan S, et al. The genomic substrate
645 for adaptive radiation in African cichlid fish. Nature. 2014;513:375–81.
646 37. Ahi EP, Lecaudey LA, Ziegelbecker A, Steiner O, Goessler W, Sefc KM. Expression
647 levels of the tetratricopeptide repeat protein gene ttc39b covary with carotenoid-based skin

colour in cichlid fish . Biol Lett. 2020.

38. Ahi EP, Lecaudey LA, Ziegelbecker A, Steiner O, Glabonjat R, Goessler W, et al. Comparative transcriptomics reveals candidate carotenoid color genes in an East African cichlid fish. BMC Genomics. 2020.

39. Lecaudey LA, Sturmbauer C, Singh P, Ahi EP. Molecular mechanisms underlying nuchal hump formation in dolphin cichlid, Cyrtocara moorii. Sci Rep. 2019.

40. Zerbino DR, Achuthan P, Akanni W, Amode MR, Barrell D, Bhai J, et al. Ensembl 2018. Nucleic Acids Res. 2018;46:D754–61.

41. Fleige S, Pfaffl MW. RNA integrity and the effect on the real-time qRT-PCR performance. Mol Aspects Med. 2006;27:126–39.

42. Mahony S, Benos P V. STAMP: a web tool for exploring DNA-binding motif similarities. Nucleic Acids Res. 2007;35 Web Server issue:W253-8. doi:10.1093/nar/gkm272.

43. Matys V, Fricke E, Geffers R, Gößling E, Haubrock M, Hehl R, et al. TRANSFAC®: Transcriptional regulation, from patterns to profiles. Nucleic Acids Research. 2003;31:374–8.

44. Hellemans J, Mortier G, De Paepe A, Speleman F, Vandesompele J. qBase relative quantification framework and software for management and automated analysis of real-time quantitative PCR data. Genome Biol. 2007;8:R19. doi:10.1186/gb-2007-8-2-r19.

45. Pashay Ahi E, Sefc KM. Towards a gene regulatory network shaping the fins of the Princess cichlid. Sci Rep. 2018;8:9602. doi:10.1038/s41598-018-27977-y.

46. Pfaffl MW, Tichopad A, Prgomet C, Neuvians TP. Determination of stable housekeeping genes, differentially regulated target genes and sample integrity: BestKeeper--Excel-based tool using pair-wise correlations. Biotechnol Lett. 2004;26:509–15. http://www.ncbi.nlm.nih.gov/pubmed/15127793. Accessed 12 Jun 2015.

47. Andersen CL, Jensen JL, Ørntoft TF. Normalization of real-time quantitative reverse transcription-PCR data: a model-based variance estimation approach to identify genes suited for normalization, applied to bladder and colon cancer data sets. Cancer Res. 2004;64:5245–50. doi:10.1158/0008-5472.CAN-04-0496.

48. Vandesompele J, De Preter K, Pattyn F, Poppe B, Van Roy N, De Paepe A, et al. Accurate normalization of real-time quantitative RT-PCR data by geometric averaging of multiple internal control genes. Genome Biol. 2002;3:RESEARCH0034. http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=126239&tool=pmcentrez&rendertype=abstract. Accessed 27 Jul 2014.

49. Pfaffl MW. A new mathematical model for relative quantification in real-time RT-PCR.

682 Nucleic Acids Res. 2001;29:e45.
683 http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=55695&tool=pmcentrez&rendert
684 ype=abstract. Accessed 31 Jul 2014.
685 50. Bergkvist A, Rusnakova V, Sindelka R, Garda JMA, Sjögreen B, Lindh D, et al. Gene
686 expression profiling--Clusters of possibilities. Methods. 2010;50:323–35.
687 doi:10.1016/j.ymeth.2010.01.009.
688 51. Kubista M, Andrade JM, Bengtsson M, Forootan A, Jonák J, Lind K, et al. The real-time
689 polymerase chain reaction. Mol Aspects Med. 2006;27:95–125.
690 doi:10.1016/j.mam.2005.12.007.
691 52. Ronco F, Matschiner M, Böhne A, Boila A, Büscher HH, El Taher A, et al. Drivers and
692 dynamics of a massive adaptive radiation in cichlid fishes. Nature. 2021;589:76–81.
693 53. Salzburger W, Meyer A, Baric S, Verheyen E, Sturmbauer C. Phylogeny of the Lake
694 Tanganyika cichlid species flock and its relationship to the Central and East African
695 haplochromine cichlid fish faunas. Syst Biol. 2002;51.
696 54. Koblmüller S, Schliewen UK, Duftner N, Sefc KM, Katongo C, Sturmbauer C. Age and
697 spread of the haplochromine cichlid fishes in Africa. Mol Phylogenet Evol. 2008;49:153–69.
698 doi:10.1016/J.YMPEV.2008.05.045.
699 55. Cohen AS, Soreghan MJ, Scholz CA. Estimating the age of formation of lakes: an
700 example from Lake Tanganyika, East African Rift system. Geology. 1993;21:511–4.
701 56. Delvaux D. Age of Lake Malawi (Nyasa) and water level fluctuations. Mus roy Afrr
702 centr, Tervuren (Belg), Dept Geol Min, Rapp ann 1993 1994. 1995;108.
703 57. Johnson TC, Kelts K, Odada E. The Holocene history of Lake Victoria. Ambio. 2000;29.
704 58. Blažek R, Polačik M, Smith C, Honza M, Meyer A, Reichard M. Success of cuckoo
705 catfish brood parasitism reflects coevolutionary history and individual experience of their
706 cichlid hosts. Sci Adv. 2018;4.
707 59. Koblmüller S, Duftner N, Sefc KM, Aibara M, Stipacek M, Blanc M, et al. Reticulate
708 phylogeny of gastropod-shell-breeding cichlids from Lake Tanganyika - The result of
709 repeated introgressive hybridization. BMC Evol Biol. 2007;7.
710 60. Takamura K. Interspecific relationships of aufwuchs-eating fishes in Lake Tanganyika.
711 Environ Biol Fishes. 1984;10:225–41.
712 61. Takahashi R, Watanabe K, Nishida M, Hori M. Evolution of feeding specialization in
713 Tanganyikan scale-eating cichlids: A molecular phylogenetic approach. BMC Evol Biol.
714 2007;7.
715 62. Kovac R, Boileau N, Muschick M, Salzburger W. The diverse prey spectrum of the

716 Tanganyikan scale-eater Perissodus microlepis (Boulenger, 1898). Hydrobiologia. 2019;832.
717 63. Boileau N, Cortesi F, Egger B, Muschick M, Indermaur A, Theis A, et al. A complex
718 mode of aggressive mimicry in a scale-eating cichlid fish. Biol Lett. 2015;11.
719 64. Mikkola ML. Molecular aspects of hypohidrotic ectodermal dysplasia. Am J Med Genet
720 Part A. 2009;149A:2031–6. doi:10.1002/ajmg.a.32855.
721 65. Cheng J, Sedlazek F, Altmüller J, Nolte AW. Ectodysplasin signalling genes and
722 phenotypic evolution in sculpins (*Cottus*). Proc R Soc B Biol Sci. 2015;282:20150746.
723 doi:10.1098/rspb.2015.0746.
724 66. Shono T, Thiery AP, Cooper RL, Kurokawa D, Britz R, Okabe M, et al. Evolution and
725 Developmental Diversity of Skin Spines in Pufferfishes. iScience. 2019;19:1248–59.
726 67. O'brown NM, Summers BR, Jones FC, Brady SD, Kingsley DM. A recurrent regulatory
727 change underlying altered expression and Wnt response of the stickleback armor plates gene
728 EDA. Elife. 2015;2015.
729 68. Ehyai S, Dionyssiou MG, Gordon JW, Williams D, Siu KWM, McDermott JC. A p38
730 MAPK regulated MEF2:β-catenin interaction enhances canonical Wnt signalling. Mol Cell
731 Biol. 2015;36:MCB.00832-15.
732 69. Shen S, Huang D, Feng G, Zhu L, Zhang Y, Cao P, et al. MEF2 Transcription Factor
733 Regulates Osteogenic Differentiation of Dental Pulp Stem Cells. Cell Reprogram.
734 2016;18:237–45. doi:10.1089/cell.2016.0016.
735 70. Kawane T, Komori H, Liu W, Moriishi T, Miyazaki T, Mori M, et al. Dlx5 and Mef2
736 Regulate a Novel Runx2 Enhancer for Osteoblast-Specific Expression. J Bone Miner Res.
737 2014;29:1960–9. doi:10.1002/jbmr.2240.
738 71. Potthoff MJ, Olson EN. MEF2: A central regulator of diverse developmental programs.
739 Development. 2007;134:4131–40.
740 72. Galceran J, Fariñas I, Depew MJ, Clevers H, Grosschedl R. Wnt3a(-/-) -like phenotype
741 and limb deficiency in Lef1(-/-)Tcf1(-/-) mice. Genes Dev. 1999;13:709–17.
742 73. Gaur T, Lengner CJ, Hovhannisyan H, Bhat RA, Bodine PVN, Komm BS, et al.
743 Canonical WNT signaling promotes osteogenesis by directly stimulating Runx2 gene
744 expression. J Biol Chem. 2005;280:33132–40. doi:10.1074/jbc.M500608200.
745 74. Kahler RA, Galindo M, Lian J, Stein GS, van Wijnen AJ, Westendorf JJ. Lymphocyte
746 enhancer-binding factor 1 (Lef1) inhibits terminal differentiation of osteoblasts. J Cell
747 Biochem. 2006;97:969–83. doi:10.1002/jcb.20702.
748 75. Kahler RA, Westendorf JJ. Lymphoid enhancer factor-1 and β-catenin inhibit Runx2-
749 dependent transcriptional activation of the osteocalcin promoter. J Biol Chem.

2003;278:11937–44.

76. Hoeppner LH, Secreto F, Jensen ED, Li X, Kahler RA, Westendorf JJ. Runx2 and bone morphogenic protein 2 regulate the expression of an alternative Lef1 transcript during osteoblast maturation. J Cell Physiol. 2009;221:480–9. doi:10.1002/jcp.21879.

77. Aman AJ, Fulbright AN, Parichy DM. Wnt/β-catenin regulates an ancient signaling network during zebrafish scale development. Elife. 2018;7.

78. Durmowicz MC, Cui CY, Schlessinger D. The EDA gene is a target of, but does not regulate Wnt signaling. Gene. 2002;285:203–11.

79. Lévy J, Capri Y, Rachid M, Dupont C, Vermeesch JR, Devriendt K, et al. *LEF1* haploinsufficiency causes ectodermal dysplasia. Clin Genet. 2020;97:595–600. doi:10.1111/cge.13714.

80. Thamamongood TA, Furuya R, Fukuba S, Nakamura M, Suzuki N, Hattori A. Expression of osteoblastic and osteoclastic genes during spontaneous regeneration and autotransplantation of goldfish scale: A new tool to study intramembranous bone regeneration. Bone. 2012;50:1240–9.

81. Kitamura K ichiro, Takahira K, Inari M, Satoh Y, Hayakawa K, Tabuchi Y, et al. Zebrafish scales respond differently to in vitro dynamic and static acceleration: Analysis of interaction between osteoblasts and osteoclasts. Comp Biochem Physiol - A Mol Integr Physiol. 2013;166:74–80.

82. Suzuki N, Hanmoto T, Yano S, Furusawa Y, Ikegame M, Tabuchi Y, et al. Low-intensity pulsed ultrasound induces apoptosis in osteoclasts: Fish scales are a suitable model for the analysis of bone metabolism by ultrasound. Comp Biochem Physiol -Part A Mol Integr Physiol. 2016;195:26–31.

83. Renz AJ, Gunter HM, Fischer JMF, Qiu H, Meyer A, Kuraku S. Ancestral and derived attributes of the dlx gene repertoire, cluster structure and expression patterns in an African cichlid fish. Evodevo. 2011;2:1. doi:10.1186/2041-9139-2-1.

84. Hulsey CD, Fraser GJ, Meyer A. Biting into the genome to phenome map: Developmental genetic modularity of cichlid fish dentitions. In: Integrative and Comparative Biology. Oxford University Press; 2016. p. 373–88.

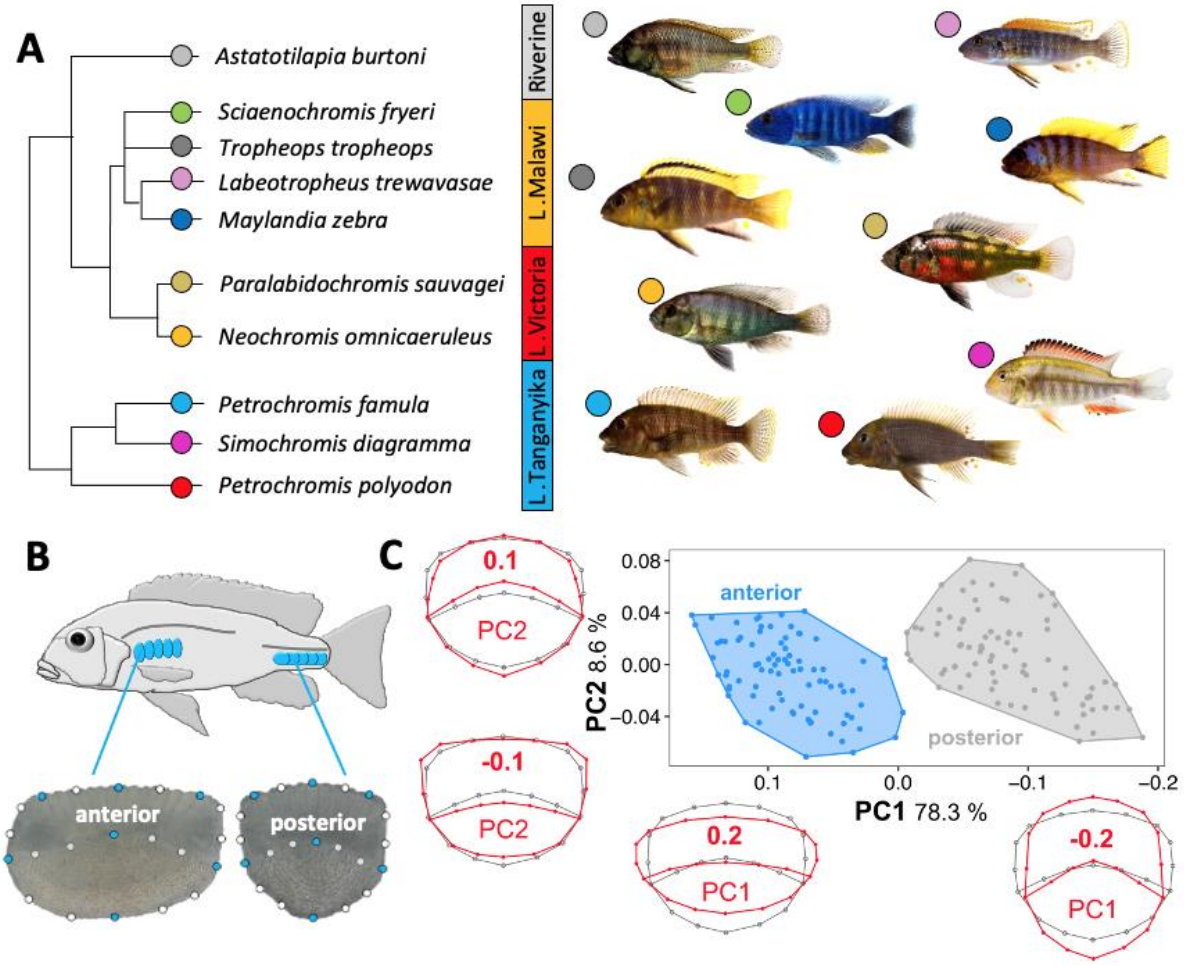
85. Yamamoto T, Ikegame M, Kawago U, Tabuchi Y, Hirayama J, Sekiguchi T, et al. Detection of RANKL-producing cells and osteoclastic activation by the addition of exogenous RANKL in the regenerating scales of goldfish. Biol Sci Sp. 2020;34:34–40.

86. Tazaki Y, Sugitani K, Ogai K, Kobayashi I, Kawasaki H, Aoyama T, et al. RANKL, Ephrin-Eph and Wnt10b are key intercellular communication molecules regulating bone

784 remodeling in autologous transplanted goldfish scales. Comp Biochem Physiol -Part A Mol
785 Integr Physiol. 2018;225:46–58.
786 87. Iwasaki M, Kuroda J, Kawakami K, Wada H. Epidermal regulation of bone
787 morphogenesis through the development and regeneration of osteoblasts in the zebrafish
788 scale. Dev Biol. 2018;437:105–19.
789 88. Schmidt-Ullrich R, Tobin DJ, Lenhard D, Schneider P, Paus R, Scheidereit C. NF-κB
790 transmits Eda A1/EdaR signalling to activate Shh and cyclin D1 expression, and controls
791 post-initiation hair placode down growth. Development. 2006;133:1045–57.
792 89. Fliniaux I, Mikkola ML, Lefebvre S, Thesleff I. Identification of dkk4 as a target of Eda-
793 A1/Edar pathway reveals an unexpected role of ectodysplasin as inhibitor of Wnt signalling
794 in ectodermal placodes. Dev Biol. 2008;320:60–71.
795 90. Cui CY, Yin M, Sima J, Childress V, Michel M, Piao Y, et al. Involvement of Wnt, Eda
796 and Shh at defined stages of sweat gland development. Dev. 2014;141:3752–60.
797 91. Pummila M, Fliniaux I, Jaatinen R, James MJ, Laurikkala J, Schneider P, et al.
798 Ectodysplasin has a dual role in ectodermal organogenesis: Inhibition of Bmp activity and
799 induction of Shh expression. Development. 2007;134:117–25.
800 92. Xiao Y, Thoresen DT, Miao L, Williams JS, Wang C, Atit RP, et al. A Cascade of Wnt,
801 Eda, and Shh Signaling Is Essential for Touch Dome Merkel Cell Development. PLOS
802 Genet. 2016;12:e1006150. doi:10.1371/journal.pgen.1006150.
803 93. Hammerschmidt B, Schlake T. Localization of Shh expression by Wnt and Eda affects
804 axial polarity and shape of hairs. Dev Biol. 2007;305:246–61.
805 94. Hong YJ, Choi YW, Myung KB, Choi HY. The immunohistochemical patterns of
806 calcification-related molecules in the epidermis and dermis of the zebrafish (Danio rerio).
807 Ann Dermatol. 2011;23:299–303.
808 95. Le Guellec D, Morvan-Dubois G, Sire JY. Skin development in bony fish with particular
809 emphasis on collagen deposition in the dermis of the zebrafish (Danio rerio). International
810 Journal of Developmental Biology. 2004;48:217–31.
811 96. Metz JR, Leeuwis RHJ, Zethof J, Flik G. Zebrafish (*Danio rerio*) in calcium-poor water
812 mobilise calcium and phosphorus from scales. J Appl Ichthyol. 2014;30:671–7.
813 97. de Vrieze E, Metz JR, Von den Hoff JW, Flik G. ALP, TRAcP and cathepsin K in
814 elasmoid scales: a role in mineral metabolism? J Appl Ichthyol. 2010;26:210–3.
815 98. Iida Y, Hibiya K, Inohaya K, Kudo A. Eda/Edar signaling guides fin ray formation with
816 preceding osteoblast differentiation, as revealed by analyses of the medaka all-fin less mutant
817 *afl*. Dev Dyn. 2014;243:765–77. doi:10.1002/dvdy.24120.

818 99. Atukorala ADS, Inohaya K, Baba O, Tabata MJ, Ratnayake RAR., Abduweli D, et al.
819 Scale and tooth phenotypes in medaka with a mutated ectodysplasin-A receptor: implications
820 for the evolutionary origin of oral and pharyngeal teeth. Arch Histol Cytol. 2011;73:139–48.
821 doi:10.1679/aohc.73.139.
822 100. Daane JM, Rohner N, Konstantinidis P, Djuranovic S, Harris MP. Parallelism and
823 Epistasis in Skeletal Evolution Identified through Use of Phylogenomic Mapping Strategies.
824 Mol Biol Evol. 2016;33:162–73. doi:10.1093/molbev/msv208.
825 101. Cooper WJ, Wirgau RM, Sweet EM, Albertson RC. Deficiency of zebrafish *fgf20a*
826 results in aberrant skull remodeling that mimics both human cranial disease and
827 evolutionarily important fish skull morphologies. Evol Dev. 2013;15:426–41.
828 doi:10.1111/ede.12052.
829 102. Rohner N, Bercsényi M, Orbán L, Kolanczyk ME, Linke D, Brand M, et al. Duplication
830 of fgfr1 Permits Fgf Signaling to Serve as a Target for Selection during Domestication. Curr
831 Biol. 2009;19:1642–7.
832 103. de Vrieze E, Sharif F, Metz JR, Flik G, Richardson MK. Matrix metalloproteinases in
833 osteoclasts of ontogenetic and regenerating zebrafish scales. Bone. 2011;48:704–12.
834 104. Iimura K, Tohse H, Ura K, Takagi Y. Expression Patterns of runx2, sparc, and bgp
835 During Scale Regeneration in the Goldfish Carassius auratus. J Exp Zool Part B Mol Dev
836 Evol. 2012;318:190–8.
837 105. Zhong Z, Niu P, Wang M, Huang G, Xu S, Sun Y, et al. Targeted disruption of sp7 and
838 myostatin with CRISPR-Cas9 results in severe bone defects and more muscular cells in
839 common carp. Sci Rep. 2016;6:1–14.
840
841

842



843

844 **Figure 1. The haplochromine cichlid species and descriptions of the scale samples.**
845 (a) A simplified phylogenetic relatedness of the East African haplochromine cichlid species
846 used in this study. (b) Positions of the anterior and posterior scales used in this study and
847 landmarks used for the geometric morphometric analyses in both anterior (left) and posterior
848 (right) scales shown as an example for *Petrochromis famula*. Blue dots represent major
849 landmarks, white dots semi-landmarks and 1 mm scale bars are given below the images. (c)
850 Principal component analysis (PCA) plots clearly separate scales from the anterior and
851 posterior part of the body. Additional, warped outline drawings illustrate major shape
852 changes along the axis (red) compared to the overall mean shape (grey).

853

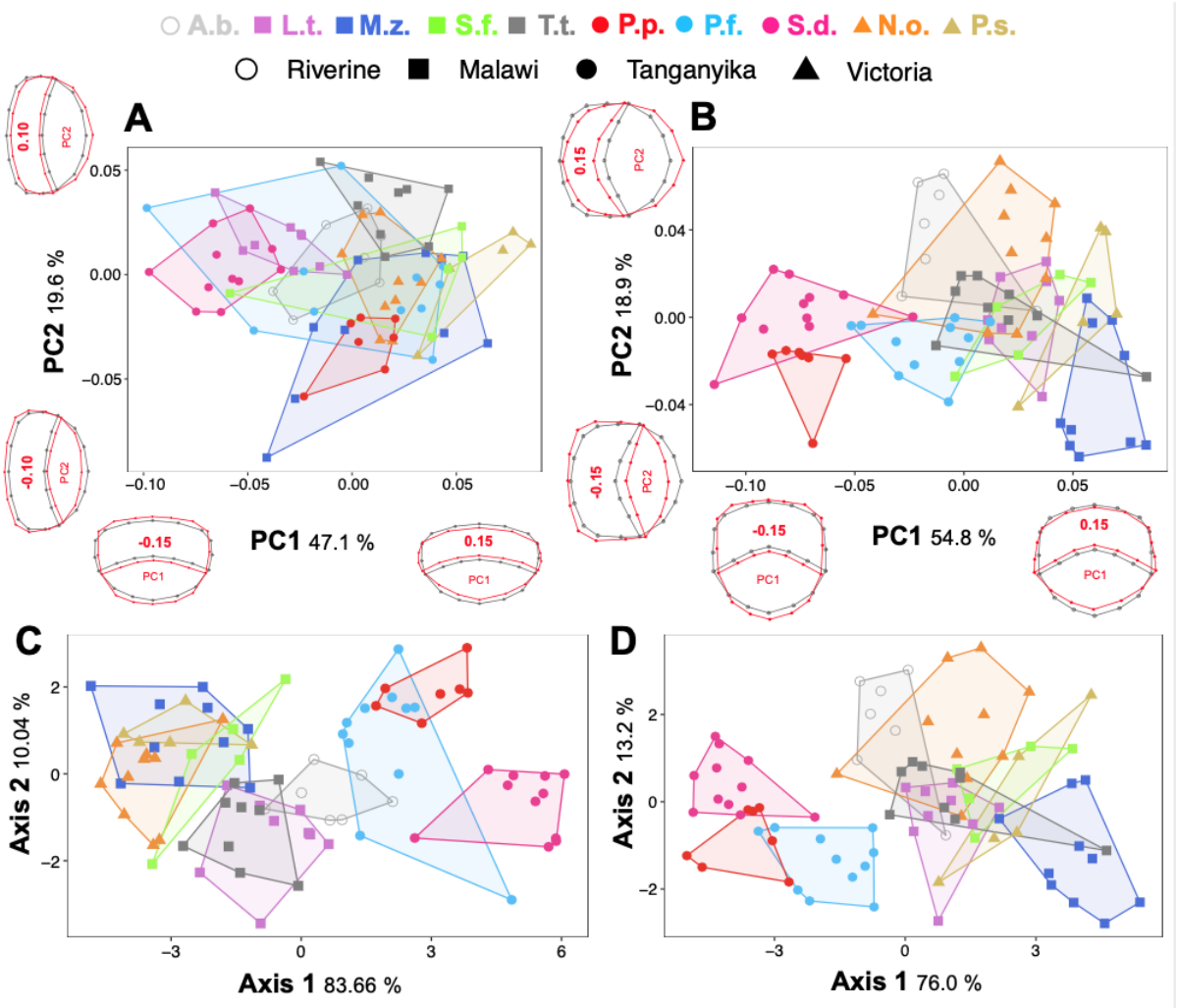


Figure 2. Morphospace of investigated scales from different species and body parts. Principal component analysis based on average shape of scales collected for the anterior (a) and posterior (b) part of the body and respective shape differences along the axis (grey: overall mean shape; red: shape change). Linear discriminant function analysis based on the first four PC-scores for anterior (c) and posterior (d) scales. All data points represent mean shapes obtained from 6 individually collected scales and shapes represent different lake origins. Abbreviations: A.b. Astatotilapia burtoni; N.o.: Neochromis omnicaeruleus; P.f.: Petrochromis famula, P.p.: P. polyodon; P.s.: Paralabidochromis sauvage S.d.: Simochromis diagramma; S.f.: Sciaenochromis fryeri; T.t.: Tropheops tropheops; L.t.: Labeotropheus trewavasae; M.z.: Maylandia zebra.

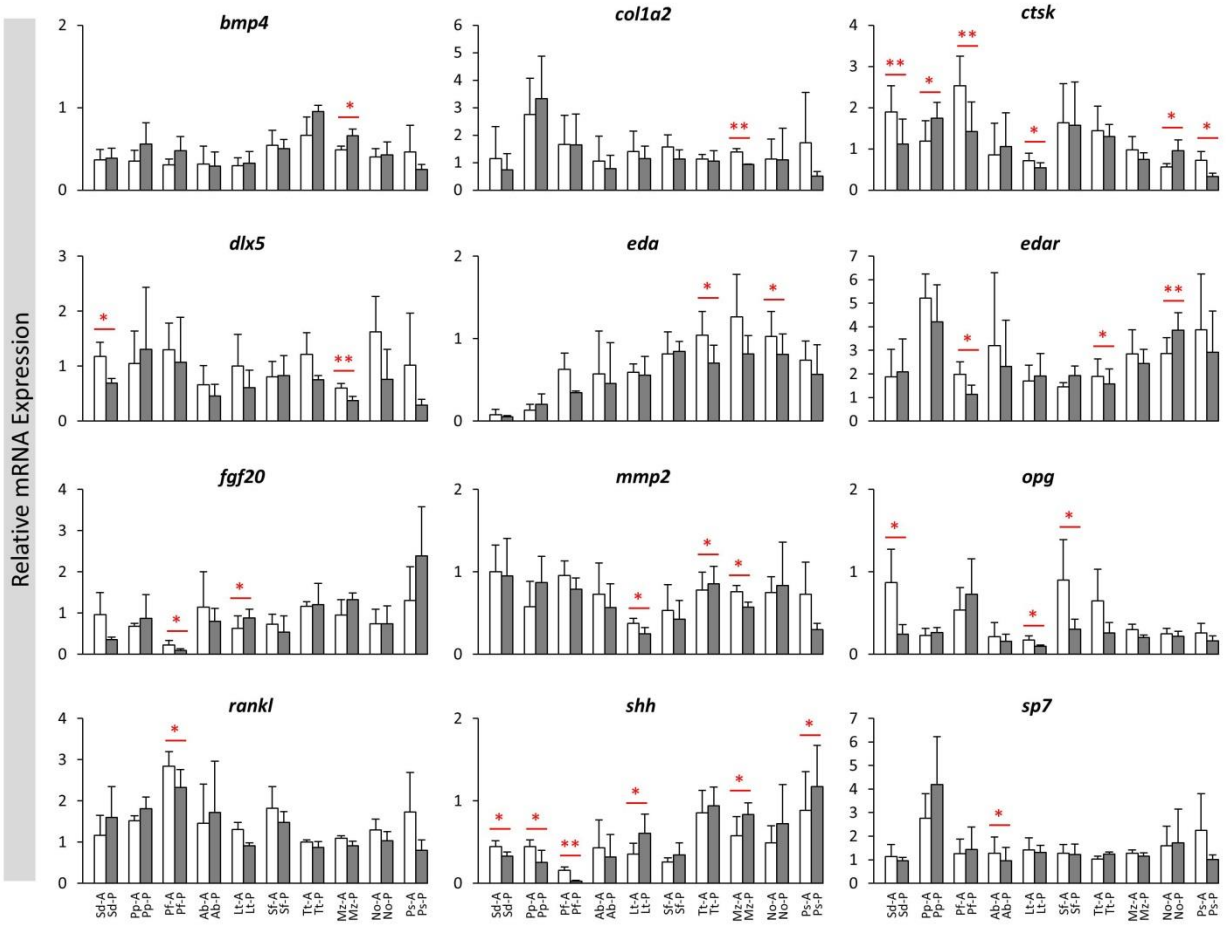


Figure 3. The anterior versus posterior scales expression differences of the candidate target genes in haplochromine cichlids from three East African lakes.

Comparisons of relative expression levels between anterior versus posterior scales for 16 candidate target genes in different lakes in East Africa at young adult stage. Significant differences between the are indicated by red asterisks (*$P < 0.05$; **$P < 0.01$). See Figure 1A for corresponding species abbreviations.

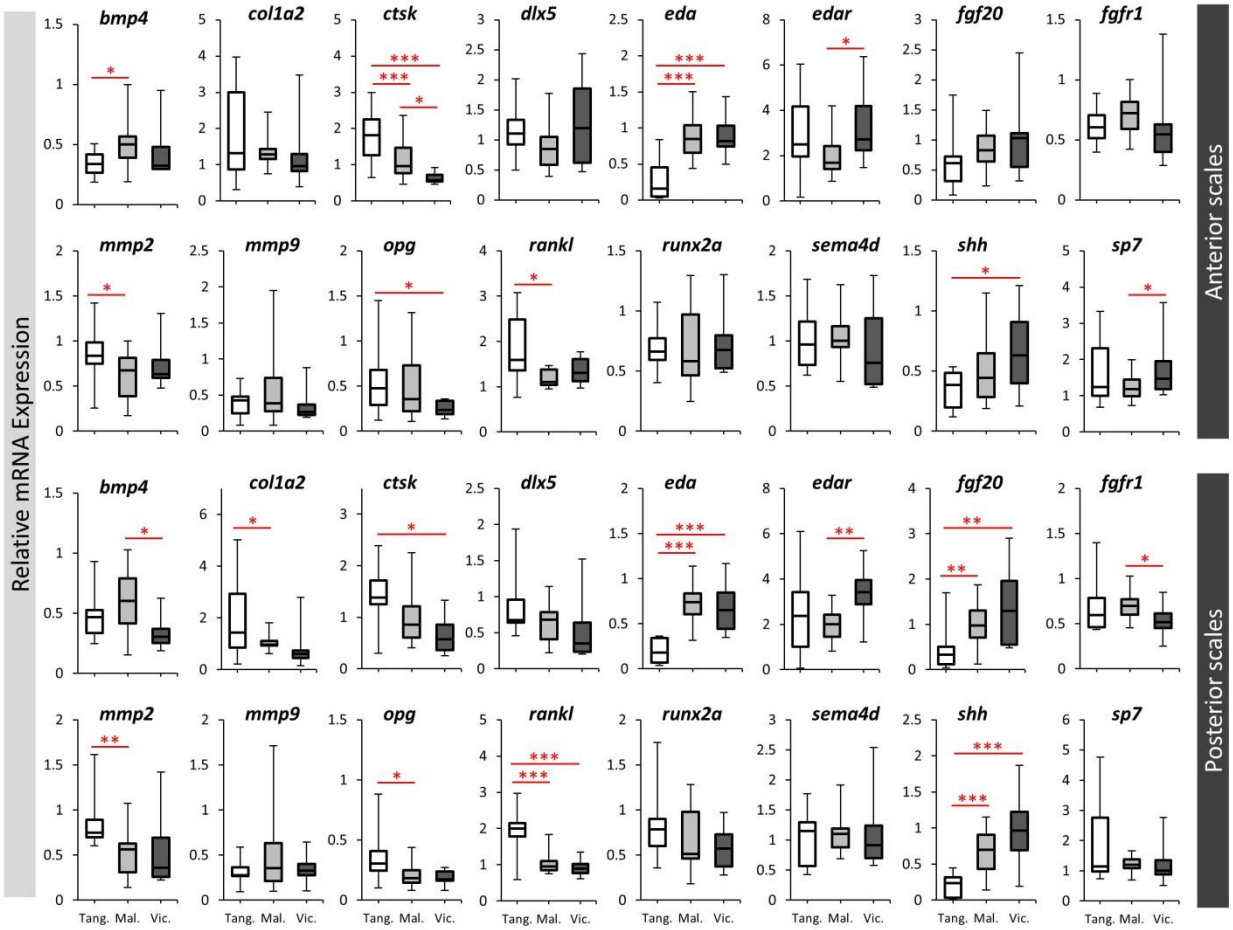


Figure 4. The lake-based expression differences of the candidate target genes in haplochromine cichlids in this study.

Comparisons of relative expression levels between the lakes, when all species of each lake were combined, within anterior or posterior scales for 16 candidate target genes. Significant differences between the are indicated by red asterisks (*P < 0.05; **P < 0.01; ***P < 0.001). See Figure 1A for corresponding species abbreviations.

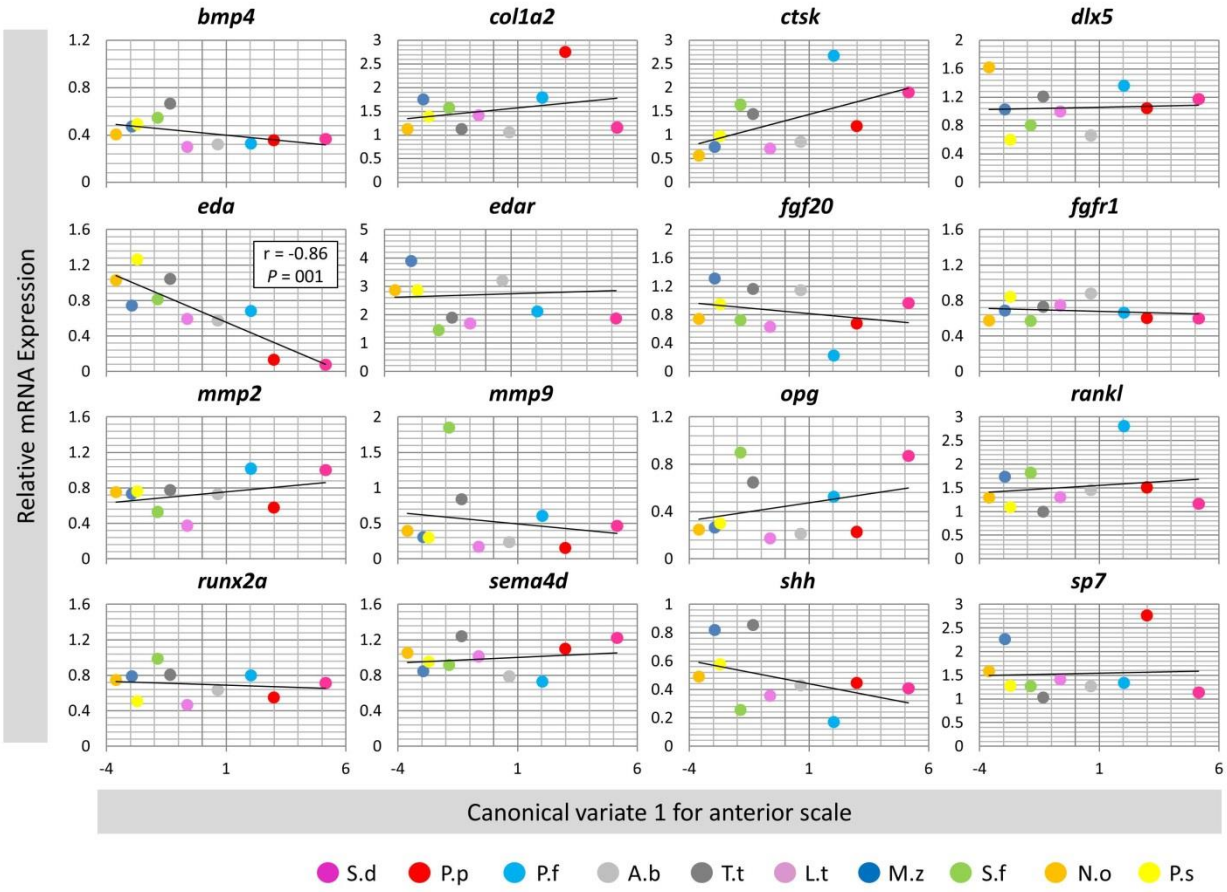


Figure 5. Correlation analyses of candidate target gene expressions and the anterior scale morphololgical divergence across the haplochromine species.

A Pearson correlation coefficient (r) was used to assess the similarity between differences in expression level of the target genes and the major canical variate in the anterior scales across all species. See Figure 1A for corresponding species abbreviations.

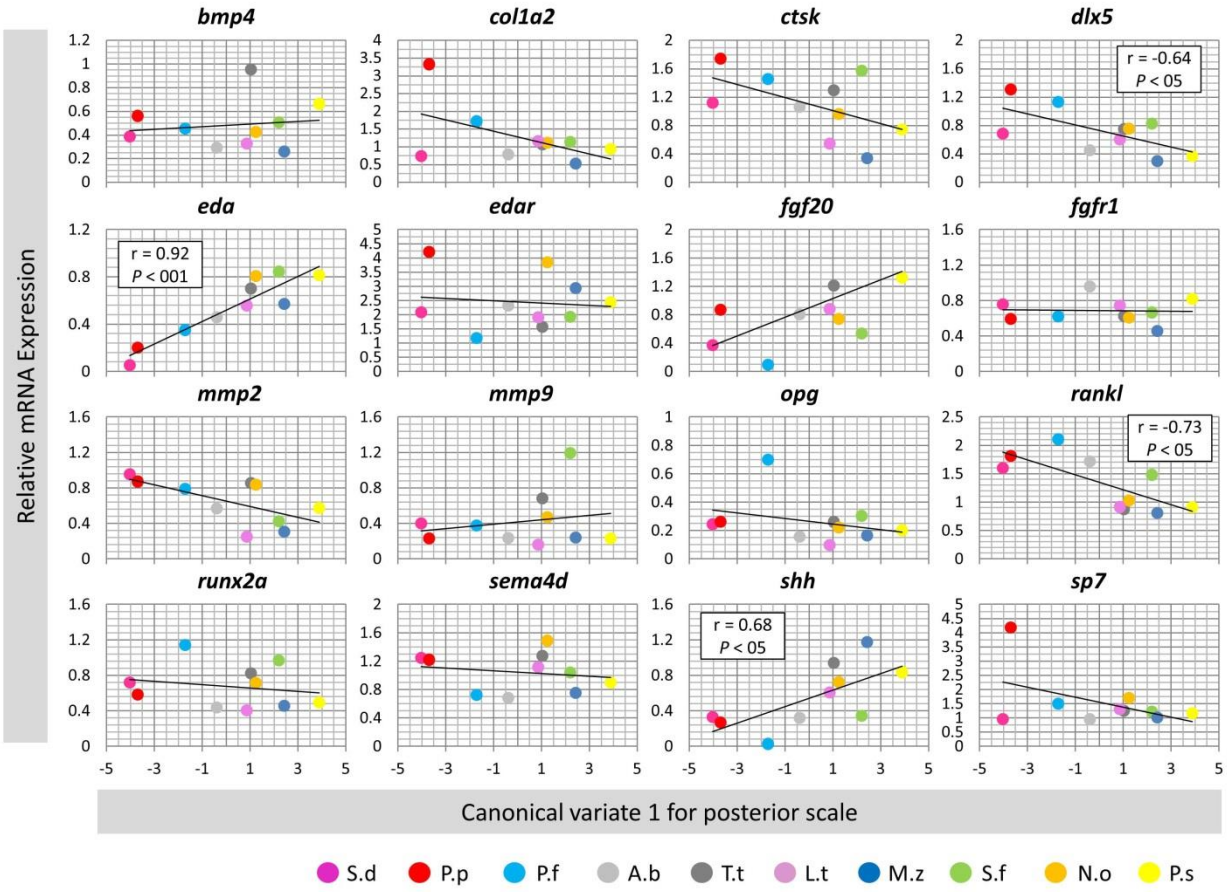


Figure 6. Correlation analyses of candidate target gene expressions and the posterior scale morphololgical divergence across the haplochromine species.

A Pearson correlation coefficient (r) was used to assess the similarity between differences in expression level of the target genes and the major canical variate in the posterior scales across all species. See Figure 1A for corresponding species abbreviations.

Table 1. Selected target genes involved in the development and/or morphogenesis of scales in teleost fish.

Gene	Related functions	Species	References
bmp4	A ligand of the TGF-β superfamily implicated in formation and calcification of elasmoid scale	zebrafish	[94]
col1a2	A member of collagen family highly expressed in both developing and adult elasmoid scales and responsive to environmental changes	zebrafish	[95, 96]
ctsk	A lysosomal cysteine proteinase involved in bone remodeling/resorption and expressed in in both developing and adult elasmoid scale and responsive to environmental changes	zebrafish	[81, 82, 96, 97]
dlx5	A homeobox transcription factor involved in bone development and scale formation and regeneration and responsive to environmental changes	zebrafish goldfish	[80–82]
eda *edar*	A tumor necrosis factor and its receptor mediating a signal involved in development of ectodermal organs and playing role in scale formation and morphogenesis	zebrafish medaka sculpin stickleback	[19, 65, 67, 98, 99]
fgf20	A fibroblast growth factor involved in formation of scale development and morphogenesis	zebrafish	[100, 101]
fgfr1	A conserved receptor of fibroblast growth factor involved in formation of scales during juvenile development and morphological changes of scales in adult	zebrafish carp cichlid	[20, 102]
mmp2 *mmp9*	Members of matrix metalloproteinases involved in development, regeneration and tissue remodeling of scale	zebrafish	[103]
opg	An osteoblast-secreted decoy receptor that functions as a negative regulator of bone resorption involved in scale formation and regeneration	zebrafish goldfish	[80, 81]
rankl	A ligand for opg and functions as a key factor for osteoclast differentiation bone remodeling involved in scale formation and regeneration and responsive to environmental changes	zebrafish goldfish	[80, 81, 85, 86]
runx2a	A transcription factors essential for osteoblastic differentiation and skeletal morphogenesis involved in scale formation and regeneration and responsive to environmental changes	zebrafish goldfish	[80, 82, 104]
sema4d	A cell surface receptor involved in cell-cell signaling and scale formation and responsive to environmental changes	zebrafish	[81]
shh	A ligand of Hedgehog signaling pathway involved in the control of scale morphogenesis in relationship with the formation of the epidermal fold in the posterior region	zebrafish	[16, 87]
sp7	A bone specific transcription factor required for osteoblast differentiation and scale formation and regeneration	zebrafish carp goldfish	[80, 87, 105]

Table 2. Ranking of reference genes in anterior and posterior scales across all of the haplochromine species used in this study.

SD indicates a ranking calculation based on standard deviation generated by BestKeeper, whereas SV, stability value, and M, mean expression stability value, are calculated by geNorm and NormFinder, respectively.

	Anterior scales						Posterior scales					
	BestKeeper		geNorm		NormFinder		BestKeeper		geNorm		NormFinder	
	Ranks	SD	Ranks	M	Ranks	SV	Ranks	SD	Ranks	M	Ranks	SV
1	*actb1*	0.398	*actb1*	1.110	*actb1*	0.478	*actb1*	0.421	*actb1*	1.133	*actb1*	0.435
2	*rps11*	0.532	*rps11*	1.168	*rps11*	0.589	*rps11*	0.571	*rps11*	1.174	*rps11*	0.471
3	*rps18*	0.745	*tbp*	1.256	*tbp*	0.669	*tbp*	0.804	*tbp*	1.276	*tbp*	0.705
4	*tbp*	0.766	*hsp90a*	1.379	*rps18*	0.708	*hsp90a*	0.829	*hsp90a*	1.346	*hsp90a*	0.763
5	*gapdh*	0.869	*rps18*	1.396	*hprt1*	0.871	*rps18*	0.869	*rps18*	1.536	*rps18*	0.812
6	*hsp90a*	0.882	*hprt1*	1.590	*hsp90a*	0.879	*gapdh*	1.003	*gapdh*	1.703	*hprt1*	0.918
7	*hprt1*	0.929	*gapdh*	1.631	*gapdh*	1.227	*hprt1*	1.089	*hprt1*	1.729	*gapdh*	1.215
8	*elf1a*	1.507	*elf1a*	1.827	*elf1a*	1.420	*elf1a*	1.362	*elf1a*	1.754	*elf1a*	1.310

Table 3. Identified genetic differences in regulatory sequences of *eda* gene and predicted binding sites for potential upstream regulators.

PWD ID indicates positional weight matrix ID of a predicted binding site and E-values refer to matching similarity between the predicted motif sequences and the PWD IDs.

Region	Sequence	Species	TFBS	PWM ID	E-value
eda 3'UTR	A - - - - - - - - A	Ab, Sd, Pp, Pf	-	-	-
	AAAAATAGCTA	All the others	Mef2	M00941	2.1296e-12
eda 3'UTR	GAATA**G**ATTAAC	Sd, Pp, Pf	-	-	-
	GAATA**T**ATTAAC	All the others	Tcf1	MA0046.1	2.6516e-06
eda / *tnfsf13b* Intergenic	ACTT**T** - - - - GCGAG	Sd	Lef1	M00805	2.4053e-06
	ACTT**TG**ACTGCGAG	Pp, Pf	Lef1	M00805	5.1356e-05
	ACTT**CG**ACTGCGAG	Ab	-	-	-
	ACTT**CA**ACTGCGAG	All the others	-	-	-

Additional file 1.xls. Information about qPCR primers used in this study.