A Genomic Cluster Containing Novel and Conserved Genes is Associated with Cichlid Fish Dental Developmental Convergence

Nidal Karagic [ID],[1] Ralf F Schneider,[†,1] Axel Meyer [ID],[1] and C Darrin Hulsey*,[1]

[1]Department for Zoology and Evolutionary Biology, University of Konstanz, Konstanz, Germany
[†]Present address: GEOMAR Helmholtz Institute for Ocean Research Kiel, Kiel, Germany
*Corresponding author: E-mail: darrin.hulsey@uni-konstanz.de.

Associate editor: Emma Teeling

Abstract

The two toothed jaws of cichlid fishes provide textbook examples of convergent evolution. Tooth phenotypes such as enlarged molar-like teeth used to process hard-shelled mollusks have evolved numerous times independently during cichlid diversification. Although the ecological benefit of molar-like teeth to crush prey is known, it is unclear whether the same molecular mechanisms underlie these convergent traits. To identify genes involved in the evolution and development of enlarged cichlid teeth, we performed RNA-seq on the serially homologous-toothed oral and pharyngeal jaws as well as the fourth toothless gill arch of *Astatoreochromis alluaudi*. We identified 27 genes that are highly upregulated on both tooth-bearing jaws compared with the toothless gill arch. Most of these genes have never been reported to play a role in tooth formation. Two of these genes (*unk*, *rpfA*) are not found in other vertebrate genomes but are present in all cichlid genomes. They also cluster genomically with two other highly expressed tooth genes (*odam*, *scpp5*) that exhibit conserved expression during vertebrate odontogenesis. *Unk* and *rpfA* were confirmed via in situ hybridization to be expressed in developing teeth of *Astatotilapia burtoni*. We then examined expression of the cluster's four genes in six evolutionarily independent and phylogenetically disparate cichlid species pairs each with a large- and a small-toothed species. *Odam* and *unk* commonly and *scpp5* and *rpfA* always showed higher expression in larger toothed cichlid jaws. Convergent trophic adaptations across cichlid diversity are associated with the repeated developmental deployment of this genomic cluster containing conserved and novel cichlid-specific genes.

Key words: Cichlidae, pharyngeal jaw, gene birth, odontogenesis, key innovation, evolvability.

Introduction

Convergent adaptations can evolve in response to similar environmental challenges (Losos 2011). However, whether these repeatedly evolved traits develop via novel or conserved genetic mechanisms is often unclear (Shapiro et al. 2004; Kratochwil et al. 2018). The evolution of similar traits could frequently arise through modifications of the expression of different genes (Härer et al. 2018). Alternatively, in some convergent traits such as the loss of stickleback pelvic spines (Shapiro et al. 2006) or acquired toxin resistance in newt-eating snakes (Geffeney et al. 2002), repeated modification of the same genetic loci are known to underlie convergent evolution. Attributes such as whether genes generating convergent traits are highly conserved (Joron et al. 2006), if gene expression shows limited pleiotropy (Papakostas et al. 2014), or how recently convergent species shared a common ancestor (Geffeney et al. 2002; Torres-Dowdall et al. 2017; Kratochwil et al. 2018) are all thought to contribute to whether similar versus novel developmental mechanisms underlie convergence. Yet, few systems with well-characterized genomes exhibit convergence in duplicated structures where

shared expression is highly likely and also display evolutionary replication over different time-scales. However, the phenotypic convergence found in independent adaptive radiations of cichlid fishes provide natural evolutionary experiments that offer exceptional opportunities to study developmental genomic similarities in repeatedly evolved and ecologically relevant traits (Stiassny and Meyer 1999; Torres-Dowdall et al. 2017; Hulsey et al. 2018). For instance, phenotypes such as enlarged molar-like teeth that are used to process hard-shelled mollusks have evolved numerous times independently during cichlid diversification. Using an integrative combination of gene expression and comparative analyses, we examined the developmental genetic role of novel and conserved mechanisms during the convergent evolution of molariform teeth across cichlid fish diversity.

Teeth are thought to have evolved once in the ancestor of all jawed vertebrates (Smith and Coates 1998) making teeth a useful model for comparisons of vertebrate development. We also have extensive knowledge concerning the developmental genetic basis of odontogenesis from mammalian model systems of tooth development (Thesleff and Hurmerinta 1981;

Chen et al. 1996; Helsinki 1996; Bei and Maas 1998; Dassule et al. 2000; Jernvall et al. 2000; Thesleff et al. 2001; Tucker and Sharpe 2004; Thesleff 2006; Wang et al. 2007; Popa et al. 2019). Importantly, many genes contributing to odontogenesis in mammals also play a role in tooth formation in other vertebrates as evolutionarily divergent as teleost fishes (Wise and Stock 2006; Harris et al. 2008; Vonk et al. 2008; Fraser et al. 2009, 2013; Ellis et al. 2015; Hulsey et al. 2016). But, since mammalian dentition differs substantially in phenotype from those of most other vertebrates, there could also be many novel mechanisms involved in tooth formation that were either lost in mammals or novelly emerged in teleosts. For example, mammals replace their teeth at most a single time (van Nievelt and Smith 2005; Luo 2007). However, cichlids and most vertebrates constantly replace their teeth throughout their lifetimes (Tuisku and Hildebrand 1994; Richman and Handrigan 2011; Fraser et al. 2013; Rasch

et al. 2016). Furthermore, processes such as the teleost-specific genome duplication or gene birth from noncoding elements could have provided modern teleosts with novel odontogenetic loci (Ohno 1970; Kawasaki and Weiss 2003; Wu et al. 2011). Thus, tooth gene diversity in cichlids and other teleosts is likely to consist of both a core set of genes shared across most vertebrates as well as a potentially diverse group of genes specific to particular lineages and dental phenotypes.

Like most teleost fishes, cichlids possess two toothed jaws (Schaeffer and Rosen 1961; fig. 1). Cichlids have oral jaws that are homologous to mammalian jaws and that are primarily used to capture prey. Additionally, they possess toothed pharyngeal jaws that are modified gill arches (GAs) that have lost gills and are used instead to process prey (Liem 1973). In cichlids, the pharyngeal jaws are thought to be an evolutionary novelty (Liem 1973) that has been
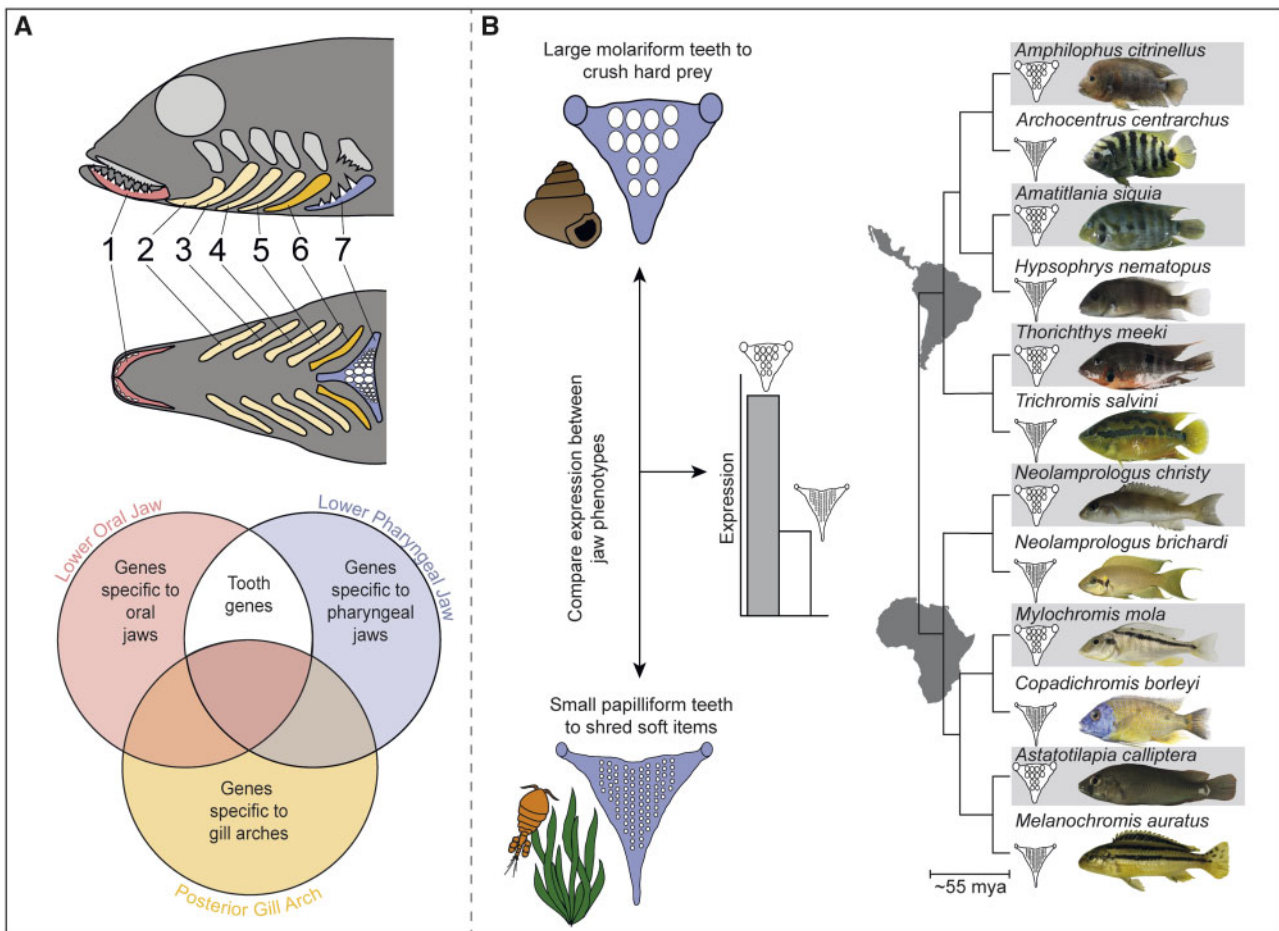

FIG. 1. Protocol used to identify conserved and novel genes associated with tooth size convergence: (*A*) We performed RNA-sequencing on the serially homologous lower oral jaw (LOJ, red), posterior-most gill arch (GA, yellow) and lower pharyngeal jaw (LPJ, blue) of the cichlid *Astatoreochromis alluaudi* to identify genes that were highly expressed in forming teeth. Genes that exhibit high expression in the two toothed structures but are not expressed in GA were considered good tooth gene candidates. The localization of expression patterns of several of these genes to teeth were validated using in situ hybridization on cryo-sections of lower oral and pharyngeal jaws. (*B*) On the LPJ, large molariform teeth develop in species (gray) consuming hard prey items like snails, whereas small papilliform teeth develop in species feeding on softer items like most invertebrates or algae. Using qPCR, we examined expression patterns of putative tooth genes identified from (*A*) in the LPJs of six species pairs of cichlids, This replicated evolutionary framework allowed us to examine convergence in gene expression levels in species having larger tooth sizes (Friedman et al. 2013). We predicted that expression of the examined genes would be higher in species with larger teeth since highly expressed genes likely contribute to increasing tooth structure.

modified to be exceptionally efficient at processing food items and allowed the oral jaws to be freed up to diversify independently (Streelman et al. 2003; Hulsey 2006; Hulsey et al. 2008). The independent diversification of these two jaws has likely contributed to the outstanding diversity in shape, size, and number of cichlid teeth (Fryer and Iles 1972; Hulsey et al. 2008; Burress 2015). Therefore, the repeated evolution of similar tooth phenotypes might commonly be due to expression differences in genes that are isolated to a single jaw. Alternatively, since these two jaws are formed from serially homologous pharyngeal arches, adaptations involving their teeth might largely be generated via expression differences in genes that contribute to tooth formation on both jaws (Jernvall et al. 2000; Vonk et al. 2008; Fraser et al. 2009). For instance, the evolution of larger teeth on the pharyngeal jaw might result simply from the increased expression of genes that are highly expressed in teeth found on both jaws. The serially homologous phenotypic duplication that the cichlid pharyngeal arches offer when coupled with the rampant evolutionary replication of tooth phenotypes in cichlids provides a powerful framework for teasing apart which components of tooth development contribute to convergent adaptations.

Taking advantage of the replicated nature of cichlid diversity, we used a combination of gene expression studies and comparative genomics to identify whether there is a shared developmental genetic basis to their convergence in larger tooth sizes (fig. 1A and B). We asked several questions: 1) Do both known odontogenic genes as well as putatively new tooth genes show increased expression in the toothed oral and pharyngeal jaws as compared with the nontoothed but serially homologous GAs? 2) Are there lineage-specific genes involved in odontogenesis in cichlid fishes?, and 3) Do particular genes that are highly expressed in both toothed jaws contribute to the developmental genetic basis of convergence in pharyngeal jaw tooth size across the evolutionary diversity of cichlids? To identify genes that likely have high expression levels in cichlid teeth, we performed a transcriptomic analysis that took advantage of the serially homologous nature of cichlid tooth-bearing arches (fig. 1A). Using RNA-seq, we identified genes that are upregulated on both jaws as compared with the toothless fourth GA located posterior to the oral jaws but anterior to the pharyngeal jaws. We then examined the role of several select candidate genes during odontogenesis using in situ hybridization to validate whether these candidate genes were expressed in developing teeth. Finally, using evolutionarily disparate species pairs that differ in their pharyngeal jaw tooth size (fig. 1B), we determined if there was convergence in the levels of gene expression of two conserved and two genes present only in cichlids that formed a physical cluster in the cichlid genome.

Results

Expression of Conserved and Novel Tooth Genes in Cichlid Fishes

Based on analyses of gene expression of the toothed jaws and a homologous GA, we identified 212 genes that are differentially expressed between the lower oral jaw (LOJ) and fourth GA, of which 79 were upregulated in the LOJ. In the comparison between the lower pharyngeal jaw (LPJ) and GA, 1,099 genes were differentially expressed and 309 were upregulated in the LPJ (supplementary fig. 1, Supplementary Material online). In the LOJ, genes with functions in calcium binding and structural activity were overrepresented, but not significantly. Genes that were upregulated in LPJ compared with GA were involved in ossification, anatomical structure development, and collagen organization (supplementary fig. 2, Supplementary Material online). In total, 105 genes consistently showed lower expression, whereas 27 genes (supplementary table 1, Supplementary Material online) consistently showed higher expression in both the toothed LOJ and LPJ tissues compared with nontoothed GA (supplementary fig. 1B, Supplementary Material online). From this set of 27 genes, we identified five that were previously described to be involved in tooth development, 15 that were described to have documented functions unrelated to teeth, and seven genes that had no previous known function (supplementary table 1, Supplementary Material online).

Identification of Cichlid-Specific Tooth Genes

From the 27 genes that were upregulated in both toothed jaws, we selected four genes to further investigate based on their high expression levels in both jaws, their high fold change compared with the nontoothed GA, and their genomic location. We found that these four genes formed a compact physically linked cluster confined to a region spanning only ~290 kb of the *Oreochromis niloticus* genome (fig. 3A). We examined: 1) Resuscitation promoting factor A (*rpfA*), 2) odontogenic ameloblast-associated protein precursor (*odam*), 3) secretory calcium-binding phosphoprotein 5 (*scpp5*), and 4) one undescribed gene (*unk*; supplementary table 1, Supplementary Material online). *Odam* and *scpp5* are known to be involved in tooth formation (Kawasaki and Weiss 2003; Kawasaki et al. 2005; Lee et al. 2010, 2012). Neither *unk* nor *rpfA*, although annotated as protein coding in cichlid genomes, have been characterized previously to be involved in tooth formation or any other biological process.

We used BLAST to compare genomic and amino acid sequences of *unk* and *rpfA* from *O. niloticus* to other available teleost genomes and did not detect any orthologous sequences outside of cichlids. However, all cichlid species with publicly available genomes (ensembl 97, Zerbino et al. 2018) did contain orthologs of *unk* as well as *rpfA* in the same orthologous cluster with *scpp5* and *odam*. In figure 3A, general patterns of divergence in synteny of the four genes is depicted for two evolutionarily disparate cichlid species that have well-annotated genomes, *Amphilophus citrinellus* from Nicaragua and *O. niloticus* from Africa, as well as for the damselfish *Acanthochromis polyacanthus* (ASM210954v1, NCBI accession number: GCA_002109545.1) and the Japanese medaka *Oryzias latipes* (ASM223471v1, NCBI accession number: GCA_002234715.1). The genome of the damselfish does not contain orthologous coding regions to either *unk* or *rpfA* and based on reciprocal BLASTs it possesses only small noncoding sequences (<50 bp) in this or any other region that produce hits to the two novel tooth genes (fig. 3A).

Similarly, in the Japanese medaka, we could only find short noncoding orthologous sequences (fig. 3A). Additionally, in many other teleosts, this putative cichlid tooth cluster appears to show rearrangements as the region spanning *unk* and *odam* is inverted in the damselfish and many other species when compared with the genomes found in medaka and cichlids (fig. 3A). Extensive genomic comparisons therefore suggest that this cluster contains two cichlid-specific genes in a region exhibiting structural rearrangements compared with many other teleosts (see supplementary fig. 6, Supplementary Material online, for details on all investigated species).

Confirmation That Candidates Are Involved in Tooth Formation

Our in situ hybridization results confirmed expression of our candidate genes in developing cichlid teeth on both the LOJ and the LPJ (fig. 2C). The previously described tooth genes *scpp5* and *odam* were expressed in the pulp and dental epithelium of forming teeth on both jaws. A similar pattern was observed for *unk* and *rpfA*, where both were found to be expressed in dental epithelium either lateral to the forming tooth (*rpfA*) or at its tip (*unk*; fig. 2C). Expression of these four genes was not detected in any other part of either jaw,
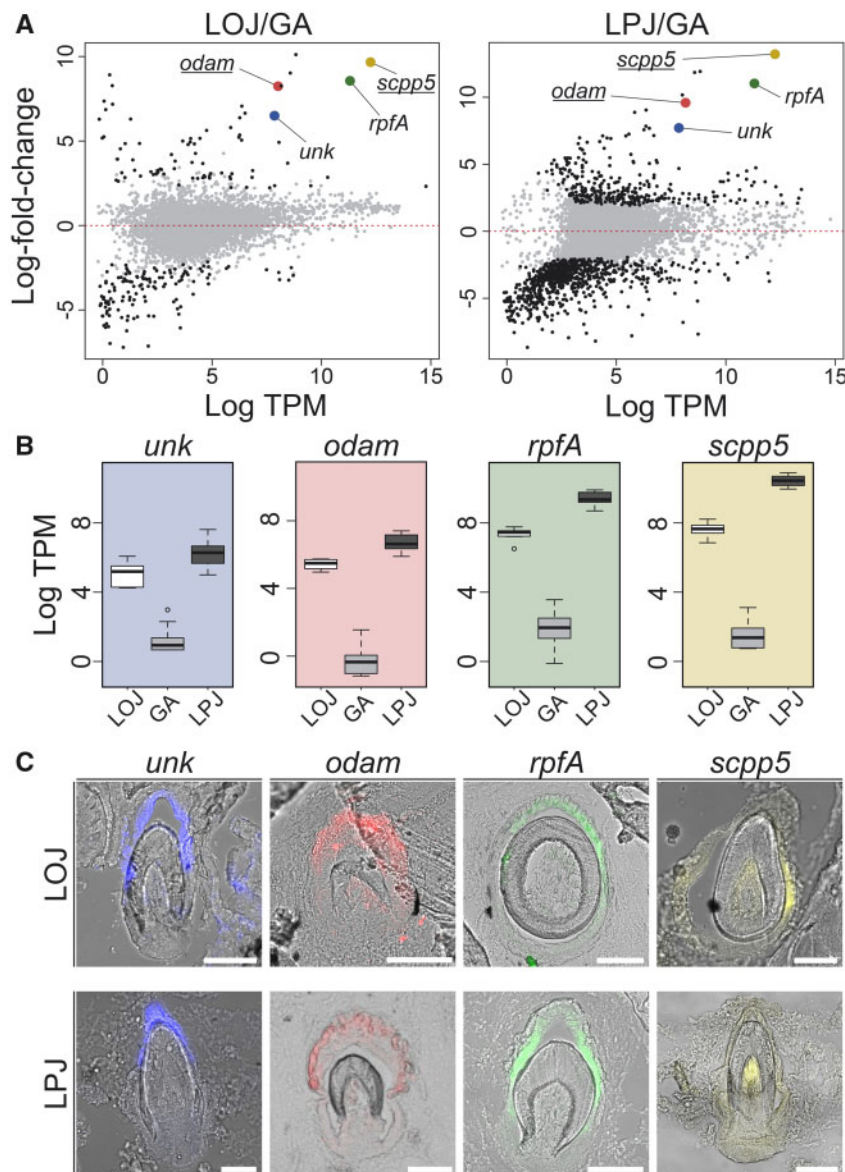

FIG. 2. Identification of putatively novel tooth genes: Using RNA-seq data, we identified genes putatively involved in cichlid tooth formation. We compared the average log-fold change expression between toothed tissues and the nontoothed gill arch (*A*). Genes that are highly expressed in LPJ and LOJ and whose expression is investigated further are highlighted with gene-specific color in both comparisons. The genes *odam* and *scpp5* (underlined) are genes known to be involved in tooth formation, whereas *rpfA* and a previously undescribed gene *unk* are not known to be involved in tooth development. Expression levels for the four genes (*B*) highlight their enhanced expression in the toothed arches (LPJ, dark gray; GA, gray; LOJ, white). (*C*) The localization of expression to developing teeth in LOJ and LPJ cryo-sections is shown via fluorescent in situ hybridization for the four genes clustered in the genome. Colors of fluorescent signal are pseudocolors and were digitally replaced using ImageJ (Schindelin et al. 2012).

supporting our inference that these genes are involved in tooth formation (fig. 2C).

Convergent Molecular Basis of Tooth Size

Using quantitative real-time PCR (qPCR), we analyzed expression levels of the four clustered genes (unk, odam, rpfA, scpp5) in the LPJs of 12 species that were divergent in pharyngeal tooth size. In each of the six species pairs, one species with enlarged molar-like teeth and one species with smaller pointed teeth were examined (fig. 3B), and their pharyngeal tooth sizes were also confirmed to be significantly different (fig. 3B). Expression levels of the four genes commonly showed significant differences ($P < 0.01$) among members of the species pairs (supplementary table 4, Supplementary Material online). For unk, expression differences occurred in the opposite directions for some species pairs. But, we found a pattern for odam in which large-toothed species either showed statistically indistinguishable or higher expression compared with smaller toothed species. However, across all six species pairs, rpfA and scpp5 showed higher expression in the larger toothed species (fig. 3B). This gene cluster contributes to a convergent developmental genetic basis for increased tooth size across cichlid diversity.
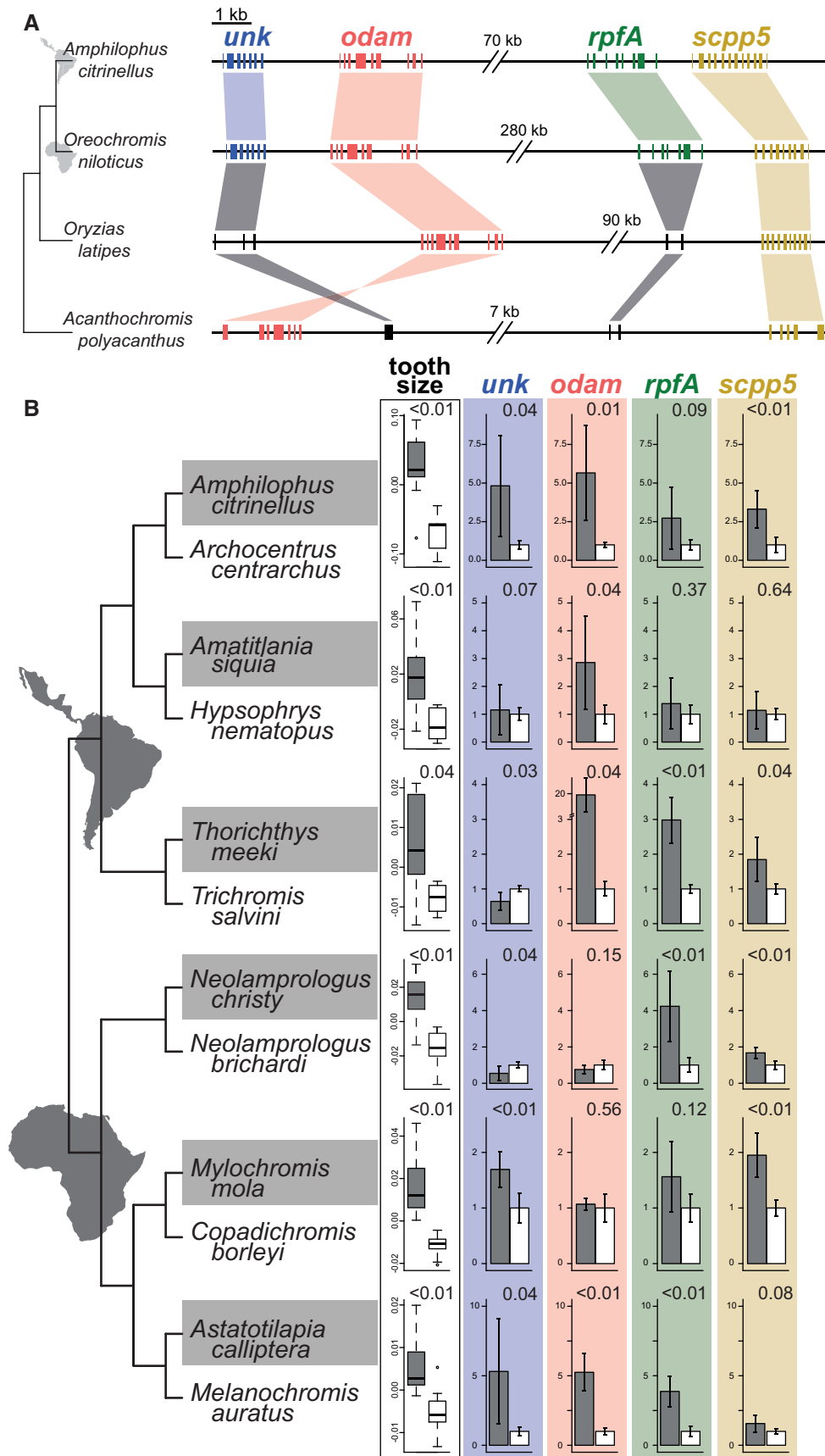
Discussion

Developmental diversity of teeth in cichlids and likely other nonmammalian vertebrates is probably greater than previously appreciated. Our comparisons of the genes in the serially homologous pharyngeal elements of cichlids allowed us to identify a substantial number of known odontogenic and putatively new tooth genes that could be involved in forming cichlid teeth. Despite being expressed in both toothed jaws, some of these novel genes appear to be exclusively found in cichlid genomes and were confirmed to be deployed extensively during odontogenesis. A genomic cluster containing a combination of two of these novel genes as well as highly conserved odontogenic genes were found to have gene expression changes that are associated with convergent evolution of cichlid tooth-size diversity. The evolutionary coupling of the cichlid pharyngeal jaw's anatomical novelties with this "tooth size" genomic cluster could have provided a key mechanism for cichlid's remarkable evolutionary success.

The genes we identified as being highly expressed in toothed jaws compared with nontoothed GAs provide evidence that many previously unidentified genes could be significantly involved in cichlid tooth formation (supplementary table 1, Supplementary Material online). Although they are embedded within much less dynamic and mineralized jaw bones, formation of replacement teeth in cichlids involves production of various tissues that are highly transcriptionally active (Rasch et al. 2016). However, not all of these genes could be involved in odontogenesis as there are structures other than teeth that these highly expressed genes could contribute to in the jaws (Alappat et al. 2003). Nevertheless, our stringent filtering probably led to an underestimation of the number of genes involved in odontogenesis. If these expression results are as generalizable to other toothed vertebrate structures as we suspect, then our

transcriptomic analyses suggest that evolutionarily and phenotypically distinct mammalian models of tooth development could represent only a small fraction of the developmental genetic diversity of vertebrate teeth. This finding is likely to be true for many traits such as paired limbs (Stopper and Wagner 2005), chambered hearts (Olson 2006), and teeth (Kawasaki et al. 2004) that despite having diversified extensively likely had a single origin during vertebrate evolution. In general, the diversity of genes expressed during organogenesis for many characteristically vertebrate structures might be far greater than a biased focus on mostly mammalian models of development has led us to believe.

Our results provide a putative example where lineage-specific developmental mechanisms were integrated into otherwise highly conserved developmental programs. Based on their physical linkage and shared high expression levels in transcriptomes from both cichlid jaws, unk and rpfA might be inferred to be coregulated with the more conserved and more widely deployed tooth genes, scpp5 and odam (Kawasaki et al. 2005; Lee et al. 2010). Further, one might infer from the function and location of expression of numerous other tooth genes (Kawasaki et al. 2005) that unk and rpfA could be similarly involved in mineralization of enameloid or production of the enameloid organic base (Kawasaki et al. 2005; Assaraf-Weill et al. 2014). However, protein-coding regions of both genes lack calcium-binding motifs characteristic of genes involved in tooth mineralization (Kawasaki and Weiss 2003). In addition, our in situ hybridization results do show that the novel tooth genes exhibit somewhat unique patterns of expression in developing teeth that are distinct from scpp5 and odam (fig. 2C). Additionally, their minor comparative discrepancies in relative expression across species pairs in comparisons with the two more conserved genes suggests the two novel genes likely are subjected to some independent regulation (fig. 3B). The relative independence of gene regulation is known to be critical to whether gene clusters promote adaptation (Singer et al. 2005; Lotterhos et al. 2018; Zeng et al. 2018). However, our results provide evidence that genomic clustering as observed for these four physically linked genes could be a common mechanism through which novel genes are incorporated into conserved but adaptive developmental programs like those for tooth morphogenesis.

How novel genes like unk and rpfA originate is often unclear, and we can envision multiple possible scenarios. Gene duplication with subsequent neofunctionalization is a frequently observed means of creating novel genes (Ohno 1970). But, we did not identify any paralogous sequences of unk or rpfA in cichlid genomes or any orthologous sequences in published genomes outside of the cichlid lineage (fig. 3A). It is also possible that unk and rpfA could have originated from noncoding genomic elements in the ancestor of all cichlids (Wu et al. 2011; Neme and Tautz 2013), but the existence of several exons in each gene would seem to argue against this. As the genomes of additional teleosts that are more closely related to cichlids become available, we might be able to clarify the origin of these genes. Additionally, these are not likely the only novel genes that contribute to cichlid diversification. The birth of new genes was once regarded as a rare

FIG. 3. Differential expression of the four genes in the tooth cluster: (A) Genomic regions of the putative cichlid tooth gene cluster in the genome of a Neotropical (*Amphilophus citrinellus*) and an African (*Oreochromis niloticus*) cichlid as well as in the genomes of the closely related medaka *Oryzias latipes* and the damselfish *Acanthochromis polyacanthus*. The cichlid tooth cluster contains the four genes *unk*, *odam*, *rpfA*, and *scpp5*.

event and generally was inferred to have little influence on how phenotypes evolve (Jacob 1977). But, in recent years evidence has accumulated that gene birth occurs more frequently than previously thought (Wu and Zhang 2013) and these recently originated genes can rapidly acquire functions that contribute to adaptation (Chen et al. 2010; Wu et al. 2011; McLysaght and Guerzoni 2015). The *unk* and *rpfA* genes appear to commonly contribute to the repeated enlargement of teeth that permit the processing of more durable prey items. Thus, these types of novel lineage-specific genes might generally play an important role in organismal diversification and in convergently evolved adaptations.

The pharyngeal jaw that bears the repeatedly evolved enlarged teeth examined here is often considered to be a key innovation that facilitated the immense trophic diversity and even speciation of cichlids (Liem 1973; Kocher 2004; Hulsey et al. 2008). Expression of genes in the cluster is associated with tooth size divergence and could have evolutionarily augmented the macroscopic anatomical specializations that constitute this key innovation. The novel insertion of the pharyngeal muscles, unique upper jaw joints, and sutured medial fusion of the LPJ all contribute to the ability of cichlids to crush durable prey, an ability that is rare in teleosts (Liem 1973). The "tooth size" genomic cluster investigated here therefore might have repeatedly and synergistically interacted with these other characteristic cichlid traits and facilitated the evolutionary likelihood that these fishes will readily re-evolve enlarged teeth that are commonly associated with a crushing trophic habit. The cichlid pharyngeal jaw has been proposed to enhance the evolvability, or capacity for adaptive evolution, of this family of teleosts and this gene cluster containing cichlid-specific genes could have amplified this capacity (Galis and Metz 1998). As we continue to genomically dissect the developmental mechanisms generating convergent adaptations that characterize the exceptional diversity and species richness of lineages such as cichlid fishes, it will be key to more extensively recognize the genomic and developmental interplay between evolutionary conservation and novelty.

Materials and Methods

All fish were raised in aquaria at the University of Konstanz animal care facility. Prior to tissue extraction, specimens were euthanized with an overdose of tricain (MS-222). For RNA-sequencing and in situ hybridization, we used the African cichlids *Astatoreochromis alluaudi* and *Astatotilapia burtoni*, respectively, and species listed in figure 1B were used for quantitative PCR analysis.

Gene Expression Analyses

For RNA-seq analysis, we dissected three distinct tissues from *Astatoreochromis alluaudi* including the LOJs ($n = 10$), the most posterior GA ($n = 10$) and LPJs ($n = 20$) and stored them in RNAlater (Sigma-Aldrich, St. Louis). These fish ranged in size from 46 to 110 mm. Because cichlids replace their teeth throughout their lifetimes, teeth of fish at these sizes should show high transcriptional activity and allow us to primarily observe expression in tooth genes. Prior to storage of tissue in RNA-later at $-20\,^{\circ}$C, we removed soft connective and muscle tissue from the elements. We then extracted RNA using the ReliaPrep miRNA cell and tissue miniprep system (Promega, Madison, WI) according to Schneider et al. (2014). We then prepared RNA libraries for each sample with the SENSE mRNA library preparation kit (Lexogen, Vienna, Austria) with an input of 100 ng total RNA per sample and unique barcode sequences for each individual sample. After equimolar pooling of RNA libraries, we size-selected all fragments in a range of 300–600 bp and sequenced libraries (paired-end 2×150 bp) on an Illumina Hiseq X-ten platform at the Beijing Genomics Institute (BGI, Hong Kong, China).

For quality control, we removed low-quality bases from each read. Bases at the beginning and end of each read with a quality score <3 were removed. Each read then was scanned with a four-base sliding window and bases were clipped if the average quality score within the sliding window dropped <15. Further, adapter sequences were clipped from each read. Afterward, all reads shorter than 145 bp were removed. Remaining high-quality reads were mapped to the publicly available *O. niloticus* reference transcriptome (Orenil_1.0, GCA_000188253.1) using the program Salmon v0.8.1 (Patro et al. 2017) that allowed us to obtain normalized expression values for each transcript. For differential expression analyses, we used the Bioconductor package limma (Ritchie et al. 2015) in the R statistical environment and discarded loci falling under a minimum threshold of 20 TPM in at least five samples per tissue. The high threshold was used to focus on highly expressed genes and to reduce the analyzed set of genes from 33,220 to 6,407. Differential expression for the remaining genes was analyzed for two contrasts: 1) genes differentially expressed between the LOJ and GA as well as 2) genes differentially expressed between LPJ and GA (fig. 1A). We concentrated on these contrasts involving the GA because this structure represents a nontoothed serially homologous element that shares developmental similarity with both of the toothed jaws. Gene-wise linear models with tissues as explanatory variables and expression of each gene as the response variable, as implemented in limma, were used to determine

Fig. 3. Continued
Orthologous sequences are connected with similarly colored areas indicating inversions when hourglass-shaped. For each gene, exon sequences are shown, except for black sequences which represent orthologous genomic regions without an open reading frame. The gray line connecting the genes is scaled to 1 kb except for the double breaks between gene pairs that indicate different sized gaps within the cluster. (B) For six species pairs, we examined size-standardized tooth area and expression levels of the four genes from the cluster in lower pharyngeal jaws (P values for each species pair is shown in the upper right corner of comparisons). The phylogenetic topology reflects the established evolutionary relationships among the species (Friedman et al. 2013; Ilves et al. 2018). Normalized expression data for each gene are depicted as the fold change (mean ± SD) in the cichlid species with larger teeth (gray) compared with the species with smaller teeth (white).

significance over all three tissues to allow for integrated inferences over the whole experiment (Law et al. 2014). P values were further corrected using the Benjamini–Hochberg method (Benjamini and Hochberg 1995) (FDR corrected P values below $\alpha = 0.05$). An absolute $\log_2$-fold change threshold ($|\log_2 = 1.5|$) was applied on significantly differentially expressed genes for each independent contrast to focus on genes highly differentiated among the toothed and nontoothed structures. Gene ontology enrichment analysis was performed as described in Hart et al. (2018) using g: Profiler (version e98_eg45_p14_bca6d38, Raudvere et al. 2019). We used the g: SCS method, taking into account the structure of functional annotations, to determine significant terms with P value <0.05 (Reimand et al. 2007). Gene functions were visualized with REVIGO (Supek et al. 2011).

Comparative Genomics

The comparative analysis of the investigated genomic cluster containing the four genes unk, odam, rpfA, and scpp5 was performed using BLAST (2.9.0) for similar (megablast) as well as more dissimilar sequences (discontinuous megablast). We compared the genomes of 23 different fish species including six cichlid species (supplementary fig. 6, Supplementary Material online) against which we blasted the whole genomic region, coding sequences, individual exons as well as amino acid sequences from O. niloticus, respectively. We only considered hits significant if they were >50 bp with an E-value $<1.0^{-10}$. Genomic locations for each gene were extracted from publicly available annotations in the ensemble database (Zerbino et al. 2018; ensembl 97).

Fluorescent In Situ Hybridization

We performed in situ hybridization of putative tooth genes on cryo-sectioned LOJ and LPJ of adult Astatotilapia burtoni ($n = 6$). Jaws were fixed in 4% paraformaldehyde in phosphate-buffered saline (PBS) overnight at $+4\,^{\circ}\mathrm{C}$. Following fixation, jaws were demineralized in 10% acetic acid in 4% PFA in PBS for 7 days at $4\,^{\circ}\mathrm{C}$. Jaws were then incubated in 18% sucrose in PBS overnight at $4\,^{\circ}\mathrm{C}$ to increase tissue stability during freezing and sectioning. Afterward, we incubated jaws in 50% Tissue-Tek O.C.T. Compound (Sakura Finetek, The Netherlands) for several hours prior to embedding the jaws in pure Tissue-Tek O.C.T. Compound at $-80\,^{\circ}\mathrm{C}$. We prepared $25\,\mu\mathrm{m}$ cryo-sections of the jaw tissues at $-20\,^{\circ}\mathrm{C}$ and stored the sections until further use at $-80\,^{\circ}\mathrm{C}$.

Jaw sections were stained for either one of the four closely examined genes (unk, odam, rpfA, scpp5) with a fluorescently labeled probe according to the protocol in Woltering et al. (2009). Probes were labeled with DIG-labeled UTP during reverse transcription of template DNA. Hybridization was performed at $65\,^{\circ}\mathrm{C}$ overnight, and subsequent antibody incubation was performed overnight at room temperature with an Anti-Digoxigenin-Peroxidase (Merck, Germany) for detection of digoxigenin-labeled probes. Fluorescent detection was performed with a thyramid amplification system (TSA Amplification System, Perkin Elmer) for 20 min at room temperature. Both bright-field and fluorescent photographs of in situ gene expression were obtained with a Leica DM6B

microscope and original colors replaced in ImageJ (Schindelin et al. 2012) for presentation in figure 2.

qPCR Assay for Validation of Differential Gene Expression

To determine if large-toothed pharyngeal jaws convergently differed in gene expression levels, we examined 12 different cichlid species that form six pairs of closely related species. These pairs were chosen to be phylogenetically distributed across cichlid diversity (Friedman et al. 2013). The pairs of Amphilophus citrinellus and Archocentrus centrarchus, Amatitlania siquia and Hypsophrys nematopus, and Thorichthys meeki and Trichromis salvini are native to the Neotropics. Neolamprologus christy and N. brichardi, Astatoreochromis calliptera and Melanochromis auratus, and Mylochromis mola and Copadichromis borleyi are all native to Africa. Six individuals per species were examined except for the species Thorichthys meeki, Trichromis salvini, N. brichardi, and M. mola in which five individuals were examined. Pairs were chosen specifically because previous observations indicated that one species likely exhibited enlarged molar-like teeth and the other displayed only smaller, more pencil-like teeth (see full list of species in fig. 3B) on the pharyngeal jaw. All individuals used were adult (>50 mm) lab-reared fish. We photographed each fish as well as their respective LPJ from a dorsal view using a Leica MZ10F stereomicroscope. These digital photographs were imported into Image J (Schindelin et al. 2012) and the standard length as well as dorsal tooth areas was measured. The area of the six most posterior teeth along the midline of the jaw that tend to be the largest teeth and are likely to be homologous across all species were averaged for each individual. These average tooth areas were then square root transformed and adjusted for size by taking residuals of the regression on individual standard length for each species pair. Subsequently, we extracted RNA from LPJs as described above (see Gene Expression Analyses).

Extracted RNA was reverse transcribed using the GoScript Reverse transcription system (Promega, Madison, WI) with an input of 200 ng total RNA according to manufacturer's protocol. Synthesized cDNA was diluted 1:5 in nuclease free water and $2\,\mu\mathrm{l}$ were used for qPCRs according to manufacturer's protocol (GoTaq qPCR MasterMix, Promega, Madison, WI). Expression levels of the four candidate genes (unk, odam, rpfA, scpp5) as well as two housekeeping genes (actinR, twinfilin; Gunter et al. 2013) were analyzed by qPCR with the GoTaq qPCR System (Promega, Madison, WI) as described in Karagic et al. (2018). Significant differences in each genes' expression values were compared for each species pair using individuals as replicates with a t-test ($\alpha = 0.05$). Since analysis of replicate sister taxa provides one of the most robust frameworks for phylogenetic comparative methods (Felsenstein 1985), we also generated cichlid-wide P values over all species pairs globally with an ANOVA to assess if gene expression showed cichlid-wide evidence of convergent associations with tooth size.

Supplementary Material

Supplementary data are available at *Molecular Biology and Evolution* online.

References

Alappat S, Zhang ZY, Chen YP. 2003. *Msx* homeobox gene family and craniofacial development. *Cell Res*. 13(6):429–442.

Assaraf-Weill N, Gasse B, Silvent J, Bardet C, Sire JY, Davit-Béal T. 2014. Ameloblasts express type I collagen during amelogenesis. *J Dent Res*. 93(5):502–507.

Bei M, Maas R. 1998. *FGFs* and *BMP4* induce both *Msx1*-independent and *Msx1*-dependent signaling pathways in early tooth development. *Development* 125(21):4325–4333.

Benjamini Y, Hochberg Y. 1995. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J R Stat Soc B Methodol*. 57(1):289–300.

Burress ED. 2015. Cichlid fishes as models of ecological diversification: patterns, mechanisms, and consequences. *Hydrobiologia* 748(1):7–27.

Chen S, Zhang YE, Long M. 2010. New genes in *Drosophila* quickly become essential. *Science* 330(6011):1682–1685.

Chen Y, Bei M, Woo I, Satokata I, Maas R. 1996. *Msx1* controls inductive signaling in mammalian tooth morphogenesis. *Development* 122(10):3035–3044.

Dassule HR, Lewis P, Bei M, Maas R, McMahon AP. 2000. Sonic hedgehog regulates growth and morphogenesis of the tooth. *Development* 127(22):4775–4785.

Ellis NA, Glazer AM, Donde NN, Cleves PA, Agoglia RM, Miller CT. 2015. Distinct developmental genetic mechanisms underlie convergently evolved tooth gain in sticklebacks. *Development* 142(14):2442–2451.

Felsenstein J. 1985. Phylogenies and the comparative method. *Am Nat*. 125(1):1–15.

Fraser GJ, Bloomquist RF, Streelman JT. 2013. Common developmental pathways link tooth shape to regeneration. *Dev Biol*. 377(2):399–414.

Fraser GJ, Hulsey CD, Bloomquist RF, Uyesugi K, Manley NR, Streelman JT. 2009. An ancient gene network is co-opted for teeth on old and new jaws. *PLoS Biol*. 7(2):e1000031.

Friedman M, Keck BP, Dornburg A, Eytan RI, Martin CH, Hulsey CD, Wainwright PC, Near TJ. 2013. Molecular and fossil evidence place the origin of cichlid fishes long after Gondwanan rifting. *Proc R Soc B*. 280(1770):20131733.

Fryer G. Iles TD. 1972. Cichlid fishes of the great lakes ofafrica. Edinburgh,UK: Oliver and Boyd.

Galis F, Metz JA. 1998. Why are there so many cichlid species? *Trends Ecol Evol*. 13(1):1–2.

Geffeney S, Brodie ED, Ruben PC, Brodie ED. 2002. Mechanisms of adaptation in a predator-prey arms race: TTX-resistant sodium channels. *Science* 297(5585):1336–1339.

Gene expression in tooth. 1996. Developmental Biology Programme of the University of Helsink [Internet]. Available from http://bite-it.he lsinki.fi. Accessed March 12, 2020.

Gunter HM, Fan S, Xiong F, Franchini P, Fruciano C, Meyer A. 2013. Shaping development through mechanical strain: the transcriptional basis of diet-induced phenotypic plasticity in a cichlid fish. *Mol Ecol*. 22(17):4516–4531.

Härer A, Meyer A, Torres-Dowdall J. 2018. Convergent phenotypic evolution of the visual system via different molecular routes: how Neotropical cichlid fishes adapt to novel light environments. *Evol Lett*. 2(4):341–354.

Harris MP, Rohner N, Schwarz H, Perathoner S, Konstantinidis P, Nüsslein-Volhard C. 2008. Zebrafish *eda* and *edar* mutants reveal conserved and ancestral roles of ectodysplasin signaling in vertebrates. *PLoS Genet*. 4(10):e1000206.

Hart JC, Ellis NA, Eisen MB, Miller CT. 2018. Convergent evolution of gene expression in two high-toothed stickleback populations. *PLoS Genet*. 14(6):e1007443.

Hulsey CD. 2006. Function of a key morphological innovation:Fusion of the cichlid pharyngeal jaw. *Proc Roy Soc B-Biol Sci*. 273(1587):669–675.

Hulsey CD, Fraser GJ, Meyer A. 2016. Biting into the genome to phenome map: developmental genetic modularity of cichlid fish dentitions. *Integr Comp Biol*. 56(3):373–388.

Hulsey CD, Roberts RJ, Lin AS, Guldberg R, Streelman JT. 2008. Convergence in a mechanically complex phenotype: detecting structural adaptations for crushing in cichlid fish. *Evolution* 62(7):1587–1599.

Hulsey CD, Zheng J, Holzman R, Alfaro ME, Olave M, Meyer A. 2018. Phylogenomics of a putatively convergent novelty: did hypertrophied lips evolve once or repeatedly in Lake Malawi cichlid fishes? *BMC Evol Biol*. 18(1):179.

Ilves KL, Torti D, López-Fernández H. 2018. Exon-based phylogenomics strengthens the phylogeny of Neotropical cichlids and identifies remaining conflicting clades (Cichliformes: cichlidae: cichlinae). *Mol Phylogenet Evol*. 118:232–243.

Jacob F. 1977. Evolution and tinkering. *Science* 196(4295):1161–1166.

Jernvall J, Keränen SV, Thesleff I. 2000. Evolutionary modification of development in mammalian teeth: quantifying gene expression patterns and topography. *Proc Natl Acad Sci U S A*. 97(26):14444–14448.

Joron M, Papa R, Beltrán M, Chamberlain N, Mavárez J, Baxter S, Abanto M, Bermingham E, Humphray SJ, Rogers J, et al. 2006. A conserved supergene locus controls colour pattern diversity in *Heliconius* butterflies. *PLoS Biol*. 4(10):e303.

Karagic N, Härer A, Meyer A, Torres-Dowdall J. 2018. Heterochronic opsin expression due to early light deprivation results in drastically shifted visual sensitivity in a cichlid fish: Possible role of thyroid hormone signaling. *J Exp Zool (Mol Dev Evol)*. 330(4):202–214.

Kawasaki K, Suzuki T, Weiss KM. 2004. Genetic basis for the evolution of vertebrate mineralized tissue. *Proc Natl Acad Sci U S A*. 101(31):11356–11361.

Kawasaki K, Suzuki T, Weiss KM. 2005. Phenogenetic drift in evolution: the changing genetic basis of vertebrate teeth. *Proc Natl Acad Sci U S A*. 102(50):18063–18068.

Kawasaki K, Weiss KM. 2003. Mineralized tissue and vertebrate evolution: The secretory calcium-binding phosphoprotein gene cluster. *Proc Natl Acad Sci U S A*. 100(7):4060–4065.

Kocher TD. 2004. Adaptive evolution and explosive speciation: the cichlid fish model. *Nat Rev Genet*. 5(4):288–298.

Kratochwil CF, Liang Y, Gerwin J, Woltering JM, Urban S, Henning F, Machado-Schiaffino G, Hulsey CD, Meyer A. 2018. Agouti-related peptide 2 facilitates convergent evolution of stripe patterns across cichlid fish radiations. *Science* 362(6413):457–460.

Law CW, Chen Y, Shi W, Smyth GK. 2014. voom: precision weights unlock linear model analysis tools for RNA-seq read counts. *Genome Biol*. 15(2):R29.

Lee HK, Lee DS, Ryoo HM, Park JT, Park SJ, Bae HS, Cho MI, Park JC. 2010. The odontogenic ameloblast-associated protein (*ODAM*) cooperates with *RUNX2* and modulates enamel mineralization via regulation of *MMP*. *J Cell Biochem*. 111(3):755–767. -

Lee H-K, Park S-J, Oh H-J, Kim J-W, Bae H-S, Park J-C. 2012. Expression pattern, subcellular localization, and functional implications of *ODAM* in ameloblasts, odontoblasts, osteoblasts, and various cancer cells. *Gene Expr Patterns*. 12(3–4):102–108.

Liem KF. 1973. Evolutionary strategies and morphological innovations: cichlid pharyngeal jaws. *Syst Zool.* 22(4):425–441.

Losos JB. 2011. Convergence, adaptation, and constraint. *Evolution* 65(7):1827–1840.

Lotterhos KE, Yeaman S, Degner J, Aitken S, Hodgins KA. 2018. Modularity of genes involved in local adaptation to climate despite physical linkage. *Genome Biol.* 19(1):157.

Luo Z-X. 2007. Transformation and diversification in early mammal evolution. *Nature* 450(7172):1011–1019.

McLysaght A, Guerzoni D. 2015. New genes from non-coding sequence: the role of de novo protein-coding genes in eukaryotic evolutionary innovation. *Philos Trans R Soc B.* 370(1678):20140332.

Neme R, Tautz D. 2013. Phylogenetic patterns of emergence of new genes support a model of frequent de novo evolution. *BMC Genomics* 14(1):117.

Ohno S. 1970. Evolution by gene duplication. New York: Springer Science & Business Media.

Olson EN. 2006. Gene regulatory networks in the evolution and development of the heart. *Science* 313(5795):1922–1927.

Papakostas S, Vøllestad LA, Bruneaux M, Aykanat T, Vanoverbeke J, Ning M, Primmer CR, Leder EH. 2014. Gene pleiotropy constrains gene expression changes in fish adapted to different thermal conditions. *Nat Commun.* 5(1):4071.

Patro R, Duggal G, Love MI, Irizarry RA, Kingsford C. 2017. Salmon provides fast and bias-aware quantification of transcript expression. *Nat Methods.* 14(4):417–419.

Popa EM, Buchtova M, Tucker AS. 2019. Revitalising the rudimentary replacement dentition in the mouse. *Development* 146(3):dev171363.

Rasch LJ, Martin KJ, Cooper RL, Metscher BD, Underwood CJ, Fraser GJ. 2016. An ancient dental gene set governs development and continuous regeneration of teeth in sharks. *Dev Biol.* 415(2):347–370.

Raudvere U, Kolberg L, Kuzmin I, Arak T, Adler P, Peterson H, Vilo J. 2019. g: profiler: a web server for functional enrichment analysis and conversions of gene lists (2019 update). *Nucleic Acids Res.* 47(W1):W191–W198.

Reimand J, Kull M, Peterson H, Hansen J, Vilo J. 2007. g: profiler—a web-based toolset for functional profiling of gene lists from large-scale experiments. *Nucleic Acids Res.* 35(Suppl 2):W193–W200.

Richman JM, Handrigan GR. 2011. Reptilian tooth development. *Genesis* 49(4):247–260.

Ritchie ME, Phipson B, Wu D, Hu Y, Law CW, Shi W, Smyth GK. 2015. limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Res.* 43(7):e47.

Schaeffer B, Rosen DE. 1961. Major adaptive levels in the evolution of the actinopterygian feeding mechanism. *Am Zool.* 1(2):187–204.

Schindelin J, Arganda-Carreras I, Frise E, Kaynig V, Longair M, Pietzsch T, Preibisch S, Rueden C, Saalfeld S, Schmid B, et al. 2012. Fiji: an open-source platform for biological-image analysis. *Nat Methods.* 9(7):676–682.

Schneider RF, Li Y, Meyer A, Gunter HM. 2014. Regulatory gene networks that shape the development of adaptive phenotypic plasticity in a cichlid fish. *Mol Ecol.* 23(18):4511–4526.

Shapiro MD, Bell MA, Kingsley DM. 2006. Parallel genetic origins of pelvic reduction in vertebrates. *Proc Natl Acad Sci U S A.* 103(37):13753–13758.

Shapiro MD, Marks ME, Peichel CL, Blackman BK, Nereng KS, Jónsson B, Schluter D, Kingsley DM. 2004. Genetic and developmental basis of evolutionary pelvic reduction in threespine sticklebacks. *Nature* 428(6984):717–723.

Singer GA, Lloyd AT, Huminiecki LB, Wolfe KH. 2005. Clusters of co-expressed genes in mammalian genomes are conserved by natural selection. *Mol Biol Evol.* 22(3):767–775.

Smith MM, Coates MI. 1998. Evolutionary origins of the vertebrate dentition: phylogenetic patterns and developmental evolution. *Eur J Oral Sci.* 106(S1):482–500.

Stiassny ML, Meyer A. 1999. Cichlids of the Rift Lakes: the extraordinary diversity of cichlid fishes challenges entrenched ideas of how quickly new species can arise. *Sci Am.* 280(2):64–69.

Stopper GF, Wagner GP. 2005. Of chicken wings and frog legs: a smorgasbord of evolutionary variation in mechanisms of tetrapod limb development. *Dev Biol.* 288(1):21–39.

Streelman JT, Webb JF, Albertson RC, Kocher TD. 2003. The cusp of evolution and development: a model of cichlid tooth shape diversity. *Evol Dev.* 5(6):600–608.

Supek F, Bošnjak M, Škunca N, Šmuc T. 2011. REVIGO summarizes and visualizes long lists of gene ontology terms. *PLoS One* 6(7):e21800.

Thesleff I. 2006. The genetic basis of tooth development and dental defects. *Am J Med Genet.* 140A(23):2530–2535.

Thesleff I, Hurmerinta K. 1981. Tissue interactions in tooth development. *Differentiation* 18(1–3):75–88.

Thesleff I, Keranen S, Jernvall J. 2001. Enamel knots as signaling centers linking tooth morphogenesis and odontoblast differentiation. *Adv Dent Res.* 15(1):14–18.

Torres-Dowdall J, Pierotti ME, Härer A, Karagic N, Woltering JM, Henning F, Elmer KR, Meyer A. 2017. Rapid and parallel adaptive evolution of the visual system of neotropical Midas cichlid fishes. *Mol Biol Evol.* 34(10):2469–2485.

Tucker A, Sharpe P. 2004. The cutting-edge of mammalian development; how the embryo makes teeth. *Nat Rev Genet.* 5(7):499–508.

Tuisku F, Hildebrand C. 1994. Evidence for a neural influence on tooth germ generation in a polyphyodont species. *Dev Biol.* 165(1):1–9.

van Nievelt AF, Smith KK. 2005. To replace or not to replace: the significance of reduced functional tooth replacement in marsupial and placental mammals. *Paleobiology* 31(2):324–346.

Vonk FJ, Admiraal JF, Jackson K, Reshef R, de Bakker MAG, Vanderschoot K, van den Berge I, van Atten M, Burgerhout E, Beck A, et al. 2008. Evolutionary origin and development of snake fangs. *Nature* 454(7204):630–633.

Wang X-P, Suomalainen M, Felszeghy S, Zelarayan LC, Alonso MT, Plikus MV, Maas RL, Chuong C-M, Schimmang T, Thesleff I. 2007. An integrated gene regulatory network controls stem cell proliferation in teeth. *PLoS Biol.* 5(6):e159.

Wise SB, Stock DW. 2006. Conservation and divergence of *Bmp2a*, *Bmp2b*, and *Bmp4* expression patterns within and between dentitions of teleost fishes. *Evol Dev.* 8(6):511–523.

Woltering JM, Vonk FJ, Müller H, Bardine N, Tuduce IL, de Bakker MA, Knöchel W, Sirbu IO, Durston AJ, Richardson MK. 2009. Axial patterning in snakes and caecilians: evidence for an alternative interpretation of the *Hox* code. *Dev Biol.* 332(1):82–89.

Wu D-D, Irwin DM, Zhang Y-P. 2011. De novo origin of human protein-coding genes. *PLoS Genet.* 7(11):e1002379.

Wu D-D, Zhang Y-P. 2013. Evolution and function of de novo originated genes. *Mol Phylogenet Evol.* 67(2):541–545.

Zeng G, Zhang P, Zhang Q, Zhao H, Li Z, Zhang X, Wang C, Yin W-B, Fang W. 2018. Duplication of a Pks gene cluster and subsequent functional diversification facilitate environmental adaptation in Metarhizium species. *PLoS Genet.* 14(6):e1007472.

Zerbino DR, Achuthan P, Akanni W, Amode MR, Barrell D, Bhai J, Billis K, Cummins C, Gall A, Girón CG, et al. 2018. Ensembl. *Nucleic Acids Res.* 46(D1):D754–D761.