

# Whole-genome Comparisons Identify Repeated Regulatory Changes Underlying Convergent Appendage Evolution in Diverse Fish Lineages

Heidi I. Chen <sup>1</sup>, Yatish Turakhia <sup>†,2</sup>, Gill Bejerano <sup>\*,1,3,4,5</sup> and David M. Kingsley <sup>\*,1,6</sup>

<sup>1</sup>Department of Developmental Biology, Stanford University School of Medicine, Stanford, CA, USA

<sup>2</sup>Department of Electrical and Computer Engineering, University of California, San Diego, CA, USA

<sup>3</sup>Department of Biomedical Data Science, Stanford University School of Medicine, Stanford, CA, USA

<sup>4</sup>Department of Computer Science, Stanford University School of Engineering, Stanford, CA, USA

<sup>5</sup>Department of Pediatrics, Stanford University School of Medicine, Stanford, CA, USA

<sup>6</sup>Howard Hughes Medical Institute, Stanford University, Stanford, CA, USA

<sup>†</sup>Work performed at Department of Electrical Engineering, Stanford University School of Engineering, Stanford, CA, USA

\*Corresponding author: E-mails: bejerano@stanford.edu; kingsley@stanford.edu.

Associate editor: Dr. Emma Teeling

## Abstract

Fins are major functional appendages of fish that have been repeatedly modified in different lineages. To search for genomic changes underlying natural fin diversity, we compared the genomes of 36 percomorph fish species that span over 100 million years of evolution and either have complete or reduced pelvic and caudal fins. We identify 1,614 genomic regions that are well-conserved in fin-complete species but missing from multiple fin-reduced lineages. Recurrent deletions of conserved sequences in wild fin-reduced species are enriched for functions related to appendage development, suggesting that convergent fin reduction at the organismal level is associated with repeated genomic deletions near fin-appendage development genes. We used sequencing and functional enhancer assays to confirm that *PelA*, a *Pitx1* enhancer previously linked to recurrent pelvic loss in sticklebacks, has also been independently deleted and may have contributed to the fin morphology in distantly related pelvic-reduced species. We also identify a novel enhancer that is conserved in the majority of percomorphs, drives caudal fin expression in transgenic stickleback, is missing in tetraodontiform, syngnathid, and synbranchid species with caudal fin reduction, and alters caudal fin development when targeted by genome editing. Our study illustrates a broadly applicable strategy for mapping phenotypes to genotypes across a tree of vertebrate species and highlights notable new examples of regulatory genomic hotspots that have been used to evolve recurrent phenotypes across 100 million years of fish evolution.

**Key words:** repeated evolution, pelvic reduction, genome-wide species comparisons, phenotype-genotype mapping, fish genomics, regulatory deletions.

## Introduction

Extensive efforts are underway to sequence the genomes of all eukaryotic species (Lewin et al. 2018, 2022)—including roughly 74,000 extant vertebrates (IUCN 2022). Despite dramatic progress in the cost and throughput of whole-genome sequencing, it remains challenging to identify the genomic basis of traits that have evolved repeatedly in wild species. The typical vertebrate genome spans several hundred megabases to several gigabases in length, of which only ~2% encode proteins (Biscotti et al. 2019; Gregory 2022). Innovative new methods will be needed to compare and interpret genomes in order to identify both the protein-coding and the noncoding regulatory differences that underlie “endless forms most beautiful” that have evolved across the tree of life (Darwin 1859).

Fish species constitute nearly half of all vertebrates, and this enormous radiation exhibits especially remarkable phenotypic diversity (Norman 1949; Nelson, Grande and Wilson 2016; IUCN 2022). A small handful of fishes (e.g., zebrafish, medaka, killifish, stickleback, cichlid) have been developed as useful model organisms for studying vertebrate genetics and development, environmental monitoring, ecological interactions, and evolutionary change (Bell and Foster 1994; Schier and Talbot 2005; Ostlund-Nilsson, Mayer and Huntingford 2006; Katsiadaki et al. 2007; Leras-Forero, Winkler and Schulte-Merker 2020; Patton, Zon and Langenau 2021; Reid, Bell and Veeramah 2021). Focused studies on these few models have been highly successful, but many more insights will likely come from comparative studies on thousands of additional fish species that have evolved in a diversity of environments.

© The Author(s) 2023. Published by Oxford University Press on behalf of Society for Molecular Biology and Evolution.

This is an Open Access article distributed under the terms of the Creative Commons Attribution License (<https://creativecommons.org/licenses/by/4.0/>), which permits unrestricted reuse, distribution, and reproduction in any medium, provided the original work is properly cited.

Open Access

A useful phenomenon within the teleost radiation includes the abundance of traits that have recurrently evolved across different clades. For example, changes in vertebral number, skeletal armor, teeth, scales, sensory modalities, locomotion (i.e., swimming mechanics, fin “walking,” etc.), osmoregulation, temperature tolerance, and duration of lifespan have evolved multiple times in independent lineages of fish (Norman 1949; Nelson, Grande and Wilson 2016; Kolora et al. 2021). Fin modifications—including extensive loss, dramatic expansion, and/or structural ornamentation—are particularly interesting given the outsized effects of fins on mobility, defense/predation, and reproductive success, and how easily fins can be scored visually or by nondestructive methods (Norman 1949; Davenport 1994; Westneat et al. 2004; Yamanoue, Setiamarga and Matsuura 2010; Price, Friedman and Wainwright 2015; Nelson, Grande and Wilson 2016; Goldberg et al. 2019; Giammona 2021; Sowersby et al. 2022). Because fins are homologous to tetrapod limbs, modification of these major body appendages in fish may also inform a variety of traits and diseases in other animals, including humans (Clack 2009; Tanaka 2016; Larouche et al. 2017, 2019; Letelier et al. 2021; Tzung et al. 2023).

Recent studies suggest that recurrent evolution of phenotypic traits may often take place through reuse of particular genes (Conte et al. 2012; Martin and Orgogozo 2013; Courtier-Orgogozo et al. 2020). As more species’ genomes are sequenced, it may therefore become possible to identify loci controlling certain traits by comparing phenotypes and genotypes over large phylogenetic trees where a trait of interest has evolved multiple times (Smith et al. 2020). This principle has previously been used to identify genomic loci involved in the recurrent evolution of traits as diverse as vitamin C dependence, echolocation, antler loss, flightlessness in birds, and hairlessness in mammals (Hiller et al. 2012a, 2012b; Marcovitz et al. 2019; Sackton et al. 2019; Wang et al. 2019; Kowalczyk et al. 2022). Here, we use a growing number of sequenced fish species to identify genomic regions associated with recurrent fin modifications.

## Results

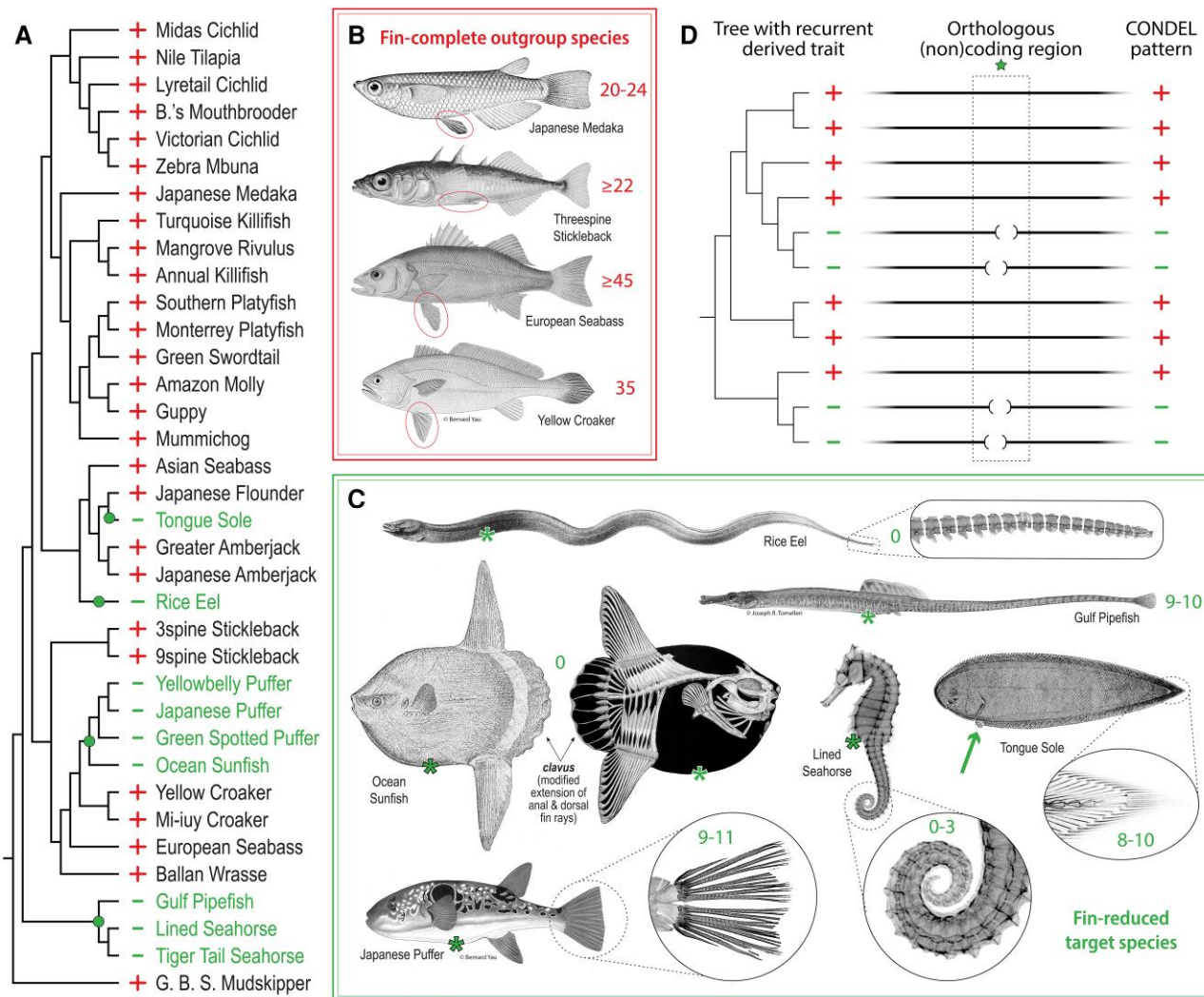
### A Computational Screen to Identify Genomic Deletions Recurrently Associated With Pelvic Reduction

We selected 36 publicly available, full nuclear genome assemblies that pass a scaffold contiguity criterion of  $L50 \leq 300$  (see Materials and Methods) and that represent percomorph fish species informative for convergent pelvic fin evolution (fig. 1A, see [supplementary spreadsheet 1, Supplementary Material](#) online for assembly accession identifiers and sources). Most of the included species have complete bilateral pelvic fins and represent the ancestral, *outgroup* trait status (fig. 1A and B) (Nelson 1989). In addition, we included representative species from four independent *target* lineages that have

evolved dramatic loss of pelvic appendages (fig. 1A and C; [supplementary table S1, Supplementary Material](#) online). The two pelvic-loss clades that were best represented among the assemblies include three members of the family Syngnathidae (pipefish and seahorses) and four members of the order Tetraodontiformes (pufferfishes and Ocean Sunfish). Also included among our target species were one member of the family Synbranchidae (Rice Eel) and one member of the family Cynoglossidae (Tongue Sole, which has one pelvic fin remaining on the blind side of the fish; see [supplementary table S1, Supplementary Material](#) online).

Based on overall genome assembly quality, extent of functional annotation, and outgroup trait status, we chose the Japanese Medaka genome assembly (abbreviated oryLat04) to be the reference against which all 35 other query genome sequences were aligned (Ichikawa et al. 2017). Using only orthology-confident alignment chains, we then searched for genomic sequences that were highly conserved in most outgroup species that had intact pelvic fins but were missing (deleted or extensively diverged) in multiple syngnathid and tetraodontiform target species that showed pelvic loss (see fig. 1D and Materials and Methods). We termed these regions *percomorph conserved sequence deletions* or *pCONDELs* and reasoned that some of these candidate intervals are likely involved in the control of pelvic fin development. Because multiple species in Synbranchidae and Cynoglossidae were not able to be included, a genotype–phenotype match in these single-representative target clades was not strictly required in the computational screen but was considered in selecting candidates for subsequent functional experiments.

With these criteria, we scanned intervals spanning 200 kb upstream to 200 kb downstream of the transcription start site for all successfully mapped reference gene orthologs. In total, we identified 1,614 predicted conserved sequences that were missing in multiple individuals of at least two independent target clades with pelvic reduction (see Materials and Methods). Of these pCONDELs, 9.5% intersected protein-coding exons, 49.6% intersected non-coding regions within genes, and 40.9% were located in intergenic regions of the Japanese Medaka reference assembly (see [supplementary spreadsheet 2, Supplementary Material](#) online). Notably, the 3,489 gene orthologs that were linked to these candidate regions were most enriched for functions related to *medial fin development* (GO:0033338, 4.45-fold enrichment,  $q$ -value = 0.022) and *embryonic appendage morphogenesis* (GO:0035113, 3.63-fold enrichment,  $q$ -value = 0.026). These two most-enriched ontology terms suggest that convergent fin loss at the phenotypic level is associated with repeated genomic deletions that remove conserved sequences near fin and appendage development genes, including *Acvr11*, *Bmp4*, *Dlx6a*, *Ext2*, *Extl3*, *Fgf10a*, *Fndc3a*, *Hmcn2*, *Lef1*, *Rspo2*, *Sall4*, *Shha*, *Smo*, *Sox9*, *Sp9*, *Tbx5a*, and *Wnt2ba* (see [supplementary spreadsheets 2 and 3, Supplementary Material](#) online for the full list of genes linked to pCONDELs as well as all enriched ontology terms).



**FIG. 1.** Genome-wide computational screen to identify conserved sequence deletions (CONDELs) associated with fin reduction. (A) Pelvic and caudal fin reduction appears to have evolved at least four independent times among the 36 percomorph species examined in this study, as indicated by dots annotated on the phylogenetic tree. Fin-complete “outgroup” species and fin-reduced “target” lineages are denoted with plus (+) and minus (–) signs, respectively. (B) Key outgroup species of the study include Japanese Medaka (the reference genome species and origin of the cell line used for in vitro functional experiments), Threespine Stickleback (the species in which all in vivo functional experiments were performed), and European Seabass and Yellow Croaker (additional outgroups to which target lineages were compared in functional studies). All outgroup species exhibit the ancestral state of complete, bilateral pelvic fins (circled) and of complete caudal fins supported by 20 or more bony rays (as indicated by the numerical annotations adjacent to each tail fin). (C) Representative species of target lineages exhibiting pelvic and caudal fin reduction include Tongue Sole (of Cynoglossidae); Rice Eel (of Synbranchidae); four members of the order Tetraodontiformes, including Japanese Puffer and Ocean Sunfish; and three members of the family Syngnathidae, including Gulf Pipefish and Lined Seahorse. With the exception of Tongue Sole (which has one pelvic fin on the fish’s blind side, thick arrow), all target lineages are pelvic-absent, as indicated by asterisks (\*) positioned where pelvic appendages might *once* have existed in the ancestors of these lineages. Rice Eel, Ocean Sunfish, and the two seahorse species exhibit complete or near-complete caudal reduction, as indicated by three or fewer bony rays remaining at their caudal extremes. The tail fins of other target lineages exhibit no more than 11 caudal rays (indicated by the numerical annotations adjacent to the caudal extreme of each species). See [supplementary tables S1 and S6, Supplementary Material](#) online, respectively, for references documenting key species phenotypic status and for the sources of illustrations shown in (B) and (C). (D) The genome-wide association strategy implemented in this study requires a tree of broadly related species in which two or more independent sub-lineages have evolved a recurrent trait. Orthologous whole-genome alignments are then used to focus subsequent experimental interrogation on regions (like the one indicated by a star) with tight correlation between genotype and phenotype—such as the regions of *conserved sequence deletions* (CONDELs) associated with fin reduction identified in the present study.

For an indication of pCONDEL age, we used BLAST and chained LASTZ whole-genome alignments to identify sequences that were likely already present in the common ancestor of Japanese Medaka and the nonpercomorph Zebrafish (see Materials and Methods). We manually inspected each homology candidate using UCSC Genome

Browser visualizations for oryLat04 and danRer11 (Kent et al. 2002; Raney et al. 2014) and determined that at least ~18% ( $n = 286/1,614$ , see column “conserved to zebrafish” in [supplementary spreadsheet 2, Supplementary Material](#) online) of the pCONDEL regions have sequence homologs that were present in the most recent common

actinopterygian ancestor of these taxa, which existed ~240 million years ago (Kumar et al. 2022). Of the 286 actinopterygian-conserved pCONDELS, 140 intersected protein-coding exons, 69 intersected noncoding regions within genes, and 77 were located in intergenic regions of the Japanese Medaka reference assembly (see [supplementary spreadsheet 2, Supplementary Material](#) online).

### Recurrent Deletions at the *PelA* Pelvic Enhancer of *Pitx1*

One of the regions identified by the computational screen for recurrent deletions in pelvic-less percomorph clades (pCONDEL.329) is a 50 bp interval that exhibits BLAST sequence similarity to the *PelA* pelvic enhancer of the limb and pituitary development gene *Pitx1*. In particular, pCONDEL.329 appears to be orthologous specifically to part of the minimal 500 bp core functional element of the *PelA* enhancer sequence (fig. 2A) (Chan et al. 2010). Previous studies found that recurrent pelvic reduction in multiple freshwater populations of Threespine Stickleback is largely attributable to repeated, de novo deletion of this noncoding pelvic control region (Shapiro et al. 2004; Chan et al. 2010; Xie et al. 2019). As in stickleback, the Japanese Medaka ortholog of the *PelA* enhancer is also upstream of the *Pitx1* gene. To confirm the presence of *PelA* deletions in other pelvic-reduced lineages, we used PCR to amplify across the orthologous region from relevant target species and their nearest outgroup (see Materials and Methods, [supplementary fig. S1, text 1, and file 4, Supplementary Material](#) online). Sanger sequencing confirmed that most pelvic-reduced species show independent deletions of different sizes and breakpoints as well as varied nucleotide substitutions (see Materials and Methods, GenBank accession numbers in [supplementary table S2 and fig. S2, Supplementary Material](#) online), resulting in the loss of at least 78 bp of the 500 bp core *PelA* enhancer (corresponding to gasAcu1-4.chrP:129,576–129,653 in the Threespine Stickleback reference genome).

To test whether these natural deletions also altered *PelA* function, we amplified the orthologous sequences denoted by the orange bar in figure 2A from several key species, cloned them upstream of a minimal promoter and GFP reporter gene, and injected the constructs into fertilized eggs from pelvic-complete Threespine Stickleback. The *PelA* sequence from outgroup Yellow Croaker drove bright GFP expression in developing pelvic structures of transgenic larvae (fig. 2B,  $n = 11/16$  fish from two clutches), in patterns similar to that of the previously characterized marine stickleback ortholog ( $n = 11/18$  fish from 1 clutch, data not shown) (Chan et al. 2010). Only weak and diffuse pelvic expression was seen from the cloned *PelA* region of the pelvic-reduced Green Spotted Puffer (fig. 2C,  $n = 7/10$  fish from 3 clutches), and this expression appeared substantially reduced compared to control eye expression known to occur from the *hsp70* minimal

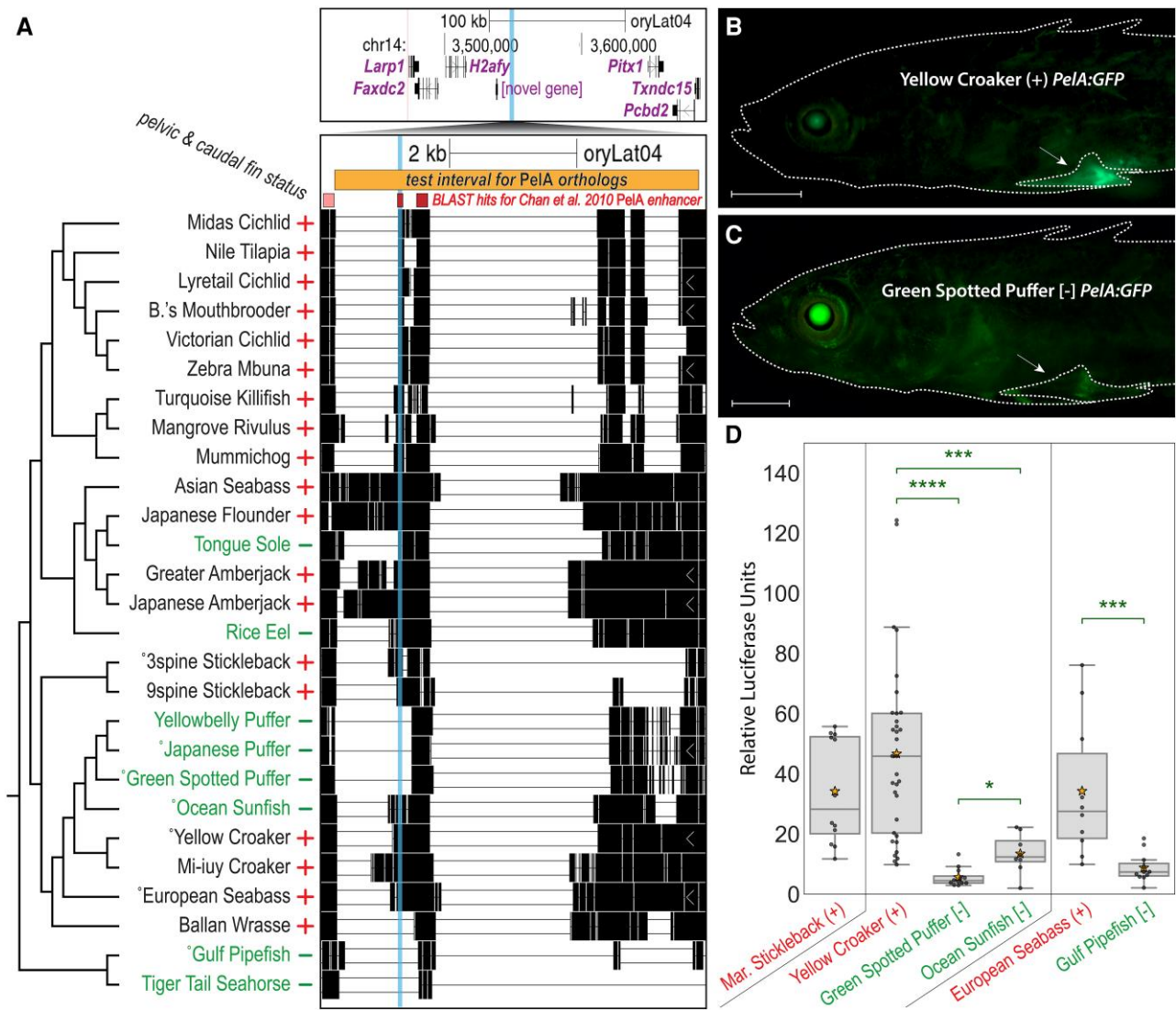
promoter of the transgenic construct (fig. 2B and C) (Nagayoshi et al. 2008).

For a more quantitative assessment of the activity of various *PelA* orthologs, we cloned these regions upstream of a luciferase reporter gene and compared levels of luciferase activity following transfection into OLHNI-2 cultured fin cells derived from Japanese Medaka (Hirayama, Mitani and Watabe 2006). Sequence from the previously characterized marine stickleback *PelA* region drove substantial expression in cultured fin cells, as did sequences from pelvic-complete outgroup species such as Yellow Croaker and European Sea Bass. In contrast, orthologs from two different pelvic-absent tetraodontiform species (Ocean Sunfish and Green Spotted Puffer), and from pelvic-absent Gulf Pipefish, drove significantly weaker luciferase activity compared to their nearest pelvic-complete outgroup (fig. 2D). Notably, the relative strengths of the *PelA* alleles from Ocean Sunfish and Green Spotted Puffer (both pelvic-absent) were inversely correlated with the amount of sequence (containing putative transcription factor binding sites) deleted in this region ([supplementary fig. S2, Supplementary Material](#) online). Previous studies have associated repeated deletions of *PelA* with recurrent pelvic reduction occurring over the last 10,000–20,000 years of stickleback evolution (Shapiro et al. 2004; Chan et al. 2010; Xie et al. 2019). Our results identify similar deletions also operating over a significantly wider, 100+ million-year span of fish evolution (Kumar et al. 2022), and confirm that our whole-genome comparative screen can identify functionally important lesions in known pelvic development loci.

### A Novel, Recurrently Deleted Regulatory Region Near *Faf1*

To study a completely novel region identified by the screen, we chose to explore pCONDEL.1189, whose presence or absence was even more tightly correlated with pelvic fin status than *PelA*. Located in an intron of *Elavl4*, this 135 bp region appeared to be either missing or disrupted in all scorable syngnathids and tetraodontiforms as well as in Rice Eel (fig. 3A and [supplementary fig. S3, Supplementary Material](#) online). We again confirmed the computational predictions in this region by amplifying and sequencing across the genomic interval pictured in figure 3A from relevant target species and their nearest outgroups (see Materials and Methods, GenBank accession numbers in [supplementary table S2, Supplementary Material](#) online).

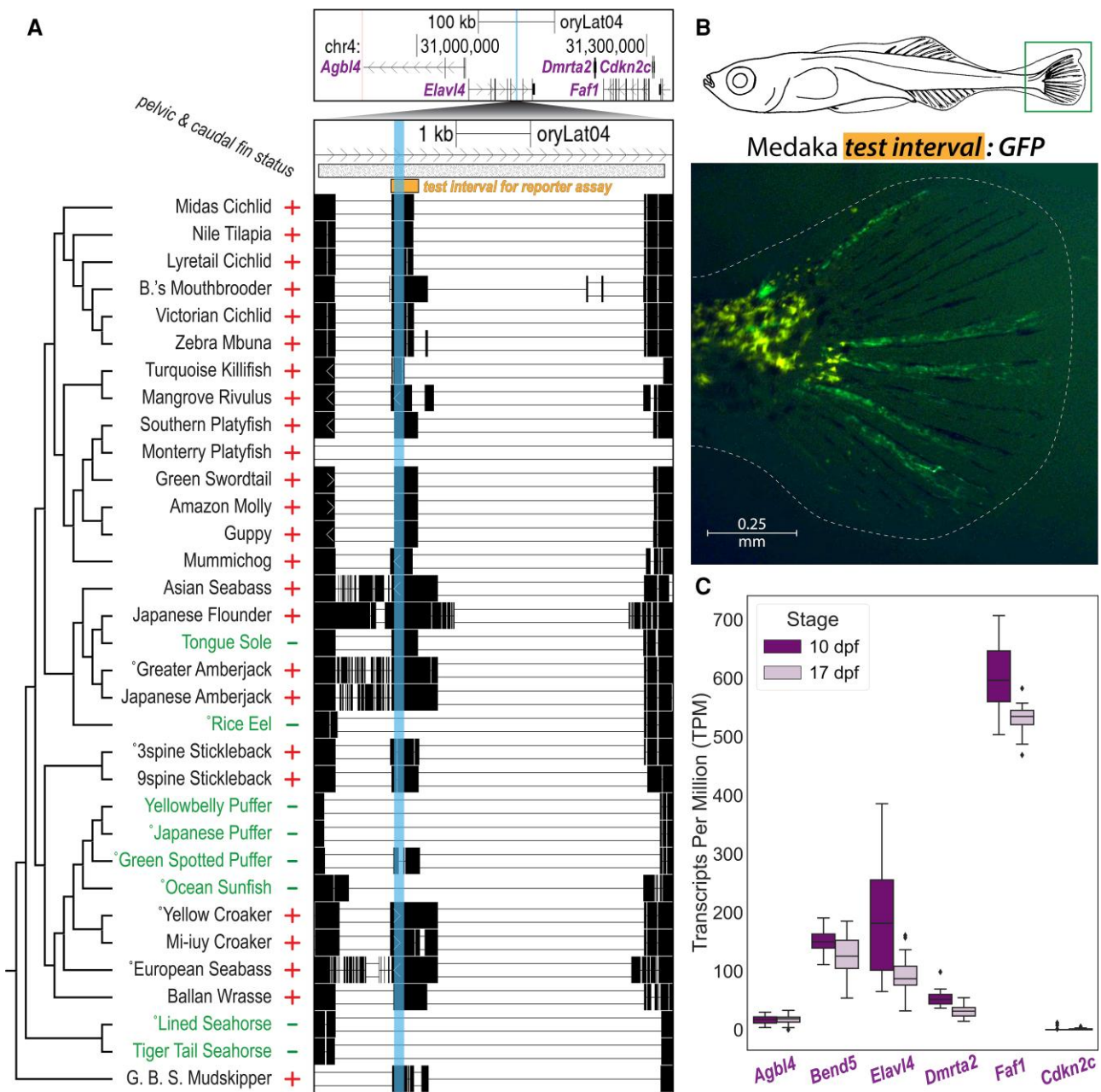
Given the absence of previous functional information about this candidate interval, we tested whether the conserved sequence element could drive informative GFP reporter expression in transgenic assays of enhancer activity in vivo. To do this, we cloned a ~500 bp test interval (as indicated by the orange bar in fig. 3A) from Japanese Medaka genomic DNA upstream of a transposable GFP reporter and injected the construct into fertilized eggs of



**Fig. 2.** Recurrent deletions in multiple fish species reduce functional activity of the *PelA* enhancer. (A) The 50 bp candidate interval pCONDEL.329 (vertical cyan highlight) is located ~100 kb upstream of the hind appendage control gene *Pitx1* (top genome browser view), a region orthologous to the *PelA* pelvic enhancer of the Threespine Stickleback *Pitx1* gene (Chan et al. 2010; Xie et al. 2019). Regions of BLAST similarity to the 2.5 kb Threespine Stickleback *PelA* sequence are shown in an annotation track as shaded (pink and red) bars. The darker (red) bars overlap pCONDEL.329 and denote BLAST hits specifically to the 500 bp core functional element of *PelA*. All existing pairwise orthologous alignment chains mapped to *Pitx1* are shown. The bar labeled “test interval for *PelA* orthologs” (oryLat04.chr14:3,547,836–3,553,572) marks the interval of orthologous sequences tested in the subsequent functional experiments depicted in (B–D) and/or verified by Sanger sequencing from the indicated (\*) target and outgroup species. Plus (+) and minus (–) signs denote phenotypic trait status (presence or absence of pelvic and caudal fin reduction). See [supplementary figure S2, Supplementary Material](#) online for the nucleotide-level multiple sequence alignment of pCONDEL.329. (B, C) GFP reporter expression in transgenic Threespine Stickleback driven by *PelA* orthologs from pelvic-complete (+) Yellow Croaker (B) and pelvic-absent [–] Green Spotted Puffer (C). Scale bars are 1 mm. White arrows point to the developing left pelvic girdle and spine. (D) Relative luciferase expression in cultured fin cells driven by *PelA* orthologs from: pelvic-complete Marine Threespine Stickleback; pelvic-reduced Green Spotted Puffer and Ocean Sunfish and their nearest pelvic-complete outgroup Yellow Croaker; and pelvic-reduced Gulf Pipefish and its nearest pelvic-complete outgroup European Seabass. Boxes show quartiles; stars denote the mean of all points in the category. Two-tailed Mann–Whitney *U* test *P*-values: \*, < 0.02; \*\*\*, < 4e–3; \*\*\*\*, < 1e–6.

pelvic-complete Threespine Stickleback. Although we hypothesized that we would see enhancer activity in the developing pelvis of transgenic larvae, we instead observed consistent reporter expression in the developing caudal fin (fig. 3B,  $n = 15/21$  transgenic fish from six clutches). These results suggested the candidate region might actually be a tail fin enhancer that was lost in many of the target lineages. A review of published literature and publicly

accessible radiographs showed that the pelvic-reduced species in our study also exhibited varying and significant degrees of reduction in the caudal fin (Larouche et al. 2017). Pelvic-complete outgroup species showed totals of 20 or more segmented plus unsegmented bony fin rays (lepidotrichia) present in their caudal fins (see numerical annotations in fig. 1B and references in [supplementary table S1, Supplementary Material](#) online). In contrast,



**Fig. 3.** A novel, recurrently deleted regulatory region near *Faf1*. (A) The 135 bp candidate pCONDEL.1189 (vertical cyan highlight) is located within an intron of *Elavl4*. The top genome browser view shows a region spanning 200 kb on both sides of pCONDEL.1189 (Japanese Medaka sequence space). Pairwise alignment chains for all species with orthologous mappings to *Elavl4* and *Faf1* are shown in the bottom genome browser view. The bar labeled “test interval for reporter assay” (oryLat04.chr4:31,129,554–31,130,065) marks the interval of orthologous sequence from Japanese Medaka tested in (B). The hatched bar (oryLat04.chr4:31,128,601–31,133,355) marks the interval verified by Sanger sequencing from the indicated (\*) target and outgroup species orthologs. Plus (+) and minus (–) signs denote phenotypic trait status (presence or absence of pelvic and caudal fin reduction). See [supplementary figure S3, Supplementary Material](#) online for the nucleotide-level multiple sequence alignment of pCONDEL.1189. (B) GFP reporter expression in the caudal fin (demarcated with a box) of a transgenic Threespine Stickleback larva at Swarup stage 30 (schematic modified from [Swarup 1958](#)) driven by the test interval in (A) cloned from Japanese Medaka genomic DNA. (C) Normalized gene expression data in the developing caudal fins of Threespine Stickleback ( $n = 12$  at each stage, before and after fin ray development) for all orthologous genes with transcription start sites predicted to lie within 200 kb of pCONDEL.1189 as shown in (A). Data for *Bend5*—a gene between *Agbl4* and *Elavl4* found in Threespine Stickleback but not in Japanese Medaka—are also included. *dpf*, days post fertilization.

pelvic-reduced species in our screen showed totals of 11 or fewer caudal fin rays (see numerical annotations in [fig. 1C](#); significance of change in number of caudal fin rays determined by two-tailed Welch’s test:  $t = 5.64$ ,  $df = 6.42$ ,  $P = 1.16e-3$ ). Note that all ecotypes of adult Threespine

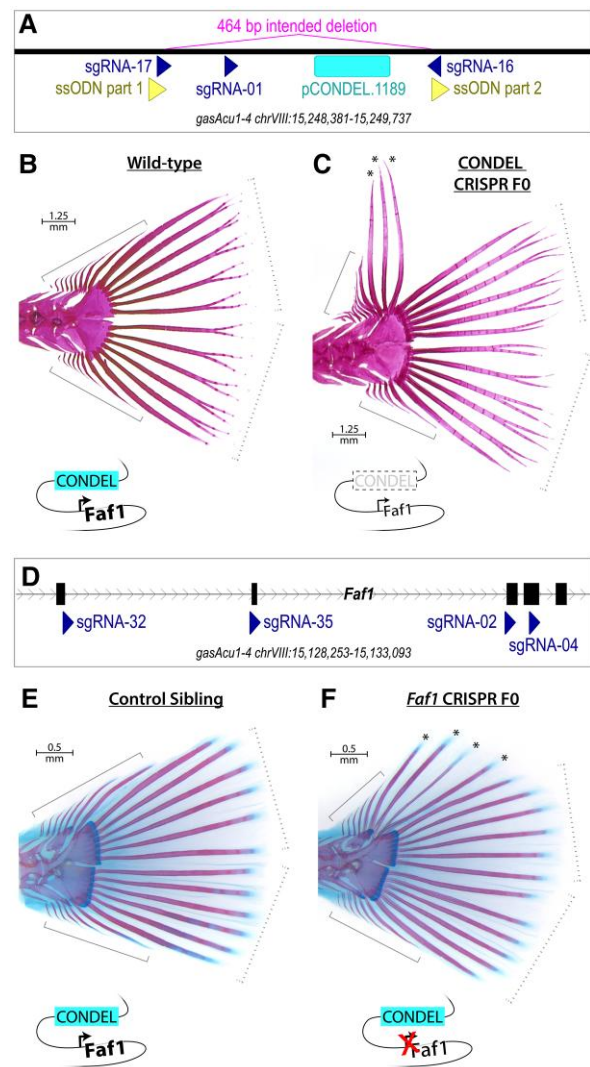
Stickleback possess 22 or more lepidotrichia ([supplementary fig. S4, Supplementary Material](#) online and see Plate III of the study published by [Huxley 1859](#); [Lindsey 1962](#)).

Given the caudal fin enhancer activity of the pCONDEL.1189 region, we performed RNAseq to determine whether any genes in the surrounding genomic interval were also expressed in the developing caudal fins of Threespine Stickleback fry before and during lepidotrichia formation (10 and 17 days post fertilization, respectively) (Swarup 1958). Of the five genes with transcription start sites within 200 kb of pCONDEL.1189 in the Japanese Medaka genome assembly, *FAS-associated factor 1* (*Faf1*) was most highly expressed (fig. 3A and C). *Faf1* encodes a protein that associates with the cell death receptor FAS (TNFRSF6) and has been shown to modulate apoptosis and cell proliferation, including in skeletal tissues (see Discussion).

### Morphological Effects of Engineered Mutations in pCONDEL.1189 and *Faf1*

To test whether pCONDEL.1189 was required for normal caudal fin development, we used CRISPR-Cas9 editing in Threespine Stickleback to remove the conserved sequence from an outgroup species that normally exhibits complete caudal and pelvic fin development (fig. 4A). Following injections of Cas9 and single-guide RNAs (sgRNAs) into fertilized stickleback eggs, genotyping and sequencing confirmed high efficiency removal of the conserved element in resulting fish (supplementary fig. S5A–C, Supplementary Material online). Across 17 genome-edited clutches, 8% of the pCONDEL.1189-targeted fish ( $n = 14/175$ ) exhibited ectopic overgrowth of the typically unsegmented procurrent rays on the dorsal edge of the tail (fig. 4B and C; supplementary figs. S5D and S6, Supplementary Material online). To control for potentially confounding effects of Cas9 nuclease injection at the zygote stage, we simultaneously injected clutch siblings with equal concentrations of editing reagents (see supplementary table S3, Supplementary Material online) that instead targeted the previously characterized *Slc24a5* gene of the *golden* pigment-control locus (Lamason et al. 2005; Wucherpfennig, Miller and Kingsley 2019). Only a single *golden* knockout control sibling also showed tail abnormalities (0.43%,  $n = 1/236$ , supplementary fig. S6, Supplementary Material online), confirming a significant effect of pCONDEL.1189 targeting on caudal fin development (two-tailed Boschloo's Exact Test: statistic =  $4.40e-5$ ,  $P = 3.55e-5$ ). In three clutches generated by intercrossing pCONDEL.1189 mosaic founder fish, no mutant tail fin phenotypes were seen, even in offspring missing both copies of the conserved sequence at the candidate region ( $n = 85$ ; see supplementary fig. S7, Supplementary Material online and additional discussion below).

To test whether the hypothesized target gene, *Faf1*, was similarly required for normal caudal fin development, we also injected Threespine Stickleback zygotes with a pool of 4 sgRNAs targeting early exons of the gene (fig. 4D). Sequencing confirmed the creation of indels disrupting *Faf1* coding exons in resulting fish (supplementary fig. S8A–C, Supplementary Material online). Across 10 genome-edited clutches, we observed that 6% of fish



**Fig. 4.** Morphological effects of engineered mutations in pCONDEL.1189 and *Faf1*. (A) The conserved element at pCONDEL.1189 (labeled bar) was deleted from Threespine Stickleback by injecting zygotes with Cas9 protein, three sgRNAs (darkly shaded triangles), and a 60 bp single-stranded DNA oligonucleotide (ssODN, a concatemer of the two lightly shaded triangles) to promote homology-directed repair and creation of the indicated 464 bp deletion. (B) Alizarin red-stained wild-type Threespine Stickleback exhibiting the typical six segmented *principal* rays on each of the dorsal and ventral halves of the tail fin (dashed brackets) as well as a variable number of smaller, unsegmented *simple* rays more anteriorly (solid brackets). (C) Alizarin red-stained mosaic F0 crisperant targeted for deletion at the pCONDEL.1189 region showing additional long segmented rays (\*) where typically only unsegmented simple rays exist. (D) *Faf1*, putative target gene of the conserved enhancer at pCONDEL.1189, was disrupted by injecting Threespine Stickleback zygotes with Cas9 protein and four sgRNAs (darkly shaded triangles) targeting exons 2–5 of the gene (furthest right exon 6 is also pictured). (E) Alizarin red and alcian blue-stained unmodified control sibling exhibiting the typical six segmented *principal* rays on each of the dorsal and ventral halves of the tail fin (dashed brackets) as well as a variable number of smaller, unsegmented *simple* rays more anteriorly (solid brackets). (F) Alizarin red and alcian blue-stained fish targeted for inactivation of *Faf1* showing ectopic segmented rays (\*) where typically only unsegmented simple rays exist. Schematics summarizing the genome modifications performed near pCONDEL.1189 and *Faf1*, as well as their hypothesized mechanisms of action, are included below (B, C) and (E, F).

( $n = 16/265$ ) exhibited a phenotype very similar to that of the pCONDEL.1189 enhancer deletion fish, with unusually long and segmented rays on the dorsal edge of the tail (fig. 4E and F; supplementary figs. S8D and S9, Supplementary Material online). No mutant tail fin phenotypes were seen in 277 unmodified control siblings from the same clutches, confirming a significant effect of *Faf1* targeting on caudal fin development (two-tailed Boschloo's Exact Test: statistic =  $8.40e-6$ ,  $P = 6.66e-6$ ).

## Discussion

Previous studies have identified “hotspots of evolution”—particular genomic regions that are used repeatedly when similar traits evolved in multiple lineages across the tree of life (Martin and Orgogozo 2013). Over 115 hotspot loci have been cataloged that show both computational and functional experimental evidence of having been used repeatedly during independent evolution of traits across closely related populations within a species (intraspecific hotspots). In contrast, about 4-fold fewer loci are currently known that underlie repeated evolution across higher taxonomic levels (i.e., hotspots reused between individuals of different genera and even more distant phylogenetic taxa) (Courtier-Orgogozo et al. 2020).

Extensive linkage mapping, mutation, transgenic, and genome-editing studies have previously shown that the *PelA* enhancer of the *Pitx1* gene corresponds to an intraspecific genomic hotspot for repeated pelvic loss within Threespine Stickleback (*Gasterosteus aculeatus*) (Shapiro et al. 2004; Chan et al. 2010; Xie et al. 2019). Multiple *Gasterosteus* populations have independently evolved total loss of pelvic fins in new freshwater lakes generated by widespread melting of glaciers in the last 20,000 years. The mutation spectrum in these pelvic-reduced *Gasterosteus* populations consists of deletions of a few hundred to a few thousand bases that completely eliminate a shared 369 bp region of the *PelA* pelvic enhancer control region (Chan et al. 2010; Xie et al. 2019). Pelvic reduction can be phenocopied by targeted deletions of the *PelA* region (Xie et al. 2019), or rescued by reintroducing *PelA:Pitx1* constructs into pelvic-less fish (Chan et al. 2010), confirming that *PelA* is a key causal locus for loss of pelvic structures within the *Gasterosteus aculeatus* species complex.

Genetic studies also suggest that changes in the *Pitx1* locus underlie some examples of pelvic reduction in the more distantly related Ninespine Stickleback lineage (*Pungitius pungitius*) (Shapiro et al. 2009; Shikano et al. 2013; Kemppainen et al. 2021). However, the genetic basis of pelvic reduction appears to be heterogeneous across different *Pungitius* populations, and the causative molecular lesions in most populations are still unknown (Shapiro et al. 2009; Shikano et al. 2013; Kemppainen et al. 2021).

Our current study has identified recurrent deletions of the *PelA* enhancer occurring in distantly related wild fish species that have independently evolved pelvic reduction since their divergence over 100 million years ago. These

results show that the *PelA* locus is not only an intraspecific evolutionary hotspot but also an interordinal hotspot that is reused when pelvic reduction evolves in different orders of fishes. Most previous examples of hotspot loci reused between distant phylogenetic groups and cataloged in “GePheBase: The Database of Genotype-Phenotype Relationships” have involved repeated amino acid changes in particular proteins (Courtier-Orgogozo et al. 2020). *PelA* is one of a very few experimentally validated *cis*-regulatory regions now known to underlie naturally evolving traits that have converged in species belonging to separate genera, families, orders, or phyla (Sagai et al. 2004; Miller et al. 2007; Guerreiro et al. 2013; Kvon et al. 2016; Sackton et al. 2019; Courtier-Orgogozo et al. 2020; Wucherpfennig et al. 2022).

Although the *PelA* region had previously been implicated in pelvic reduction, most of the other genomic regions recovered in our genome-wide comparative analysis do not have previously known functions. Over 90% map outside of protein-coding exons of genes, so, like *PelA*, may also correspond to regulatory elements in the genome. The genomic regions near these recurrently lost sequences are enriched for genes involved in fin and appendage development, and we hypothesize that regulatory deletions occurring near such genes are an important contributor to recurrent fin evolution in fishes.

Our studies clearly illustrate the value of experimental validation of novel loci identified by comparative genomic scans. Although we selected target and outgroup species to screen for regions associated with pelvic reduction, when we tested the *in vivo* activity of the conserved element at pCONDEL.1189, we found that this noncoding regulatory region drove expression in caudal fins rather than pelvic fins of transgenic larvae. Retrospective trait inspection showed that pelvic and caudal fin reduction are correlated with each other in the original species tree. Thus, the genomic regions recovered in our whole-genome screen might be associated with either of these fin phenotypes or with any other trait (such as branchial arch modification) that co-occurs with pelvic and caudal fin reduction in the same target lineages (Larouche et al. 2017). We have not yet experimentally tested other regions from the computational screen for *in vivo* expression patterns and phenotypes, so do not know what fraction of pCONDELs may regulate expression in pelvic fins, median fins, or other tissues. Future surveys characterizing tissue-specific patterns of chromatin accessibility might aid in discerning which tissue(s) and trait(s) are most likely associated with particular pCONDELs identified in the computational screen (e.g., Sackton et al. 2019).

Our genome-editing experiments further confirm that the pCONDEL.1189 region functions during caudal fin development. Deletion of the region results in the formation of additional long segmented fin rays in the caudal fins of Threespine Stickleback. Targeting of the nearby *Faf1* gene produces a very similar phenotype, suggesting the pCONDEL.1189 region likely acts by regulating *Faf1* function. *Faf1* is named for its association with the *Fas1*



receptor, which activates cell death pathways in response to intercellular signaling (Chu, Niu and Williams 1995; Ryu et al. 2003). The extra fin rays seen in our targeting experiments may thus be the result of decreasing the activity of a known apoptotic pathway during caudal fin development. Since recurrent loss of the region is correlated with recurrent reduction of fin rays in natural fish species, we speculate that evolution of caudal fin morphology may involve a mixture of both reductive and compensatory mutations.

It is interesting that tail phenotypes were only observed in a fraction of CRISPR/Cas9-edited founder fish: approximately 8% of pCONDEL.1189-targeted fish and 6% of *Faf1*-targeted fish (supplementary figs. S5D and S8D, Supplementary Material online), versus  $\geq 50\%$  of founder fish exhibiting obvious phenotypes when targeted at previously reported pigmentation loci (Wucherpfennig, Miller and Kingsley 2019). In addition, we did not observe similar tail phenotypes when we intercrossed mosaic founders to produce F1 fish that were homozygous for mutant pCONDEL.1189 alleles in every cell during normal development (supplementary fig. S7, Supplementary Material online). These results might be attributable to the polar coordinate model of limb development first described by French, Bryant and Bryant (1976). Based on observations of supernumerary limb and digit formation in insects and amphibians after mirror-image appendage grafting, this model posits that every cell has a particular positional identity relative to all other cells in a given tissue, as if arranged in a polar coordinate system. When normally non-adjacent identities come into contact, either by injury, tissue grafting, or—we suggest—mosaic genome editing, cells proliferate and/or intercalate to restore the original cell–cell adjacencies and thereby generate supernumerary structures. We hypothesize that the ectopic fin ray phenotype in our mosaic founders may occur when *Faf1* activity is disrupted in subsets of developing fin cells, and perhaps only when those cells occur in juxtaposition with other cells that have different relative levels of *Faf1* activity. Future studies could modulate *Fas1/Faf1* activity in subgroups of cells during caudal fin development to test whether manipulation of this pathway in specific anatomical regions leads to predictable morphology that resembles fin ray changes in wild fishes.

Our current work on *PelA* and pCONDEL.1189 illustrates how genome-wide approaches can identify both known and novel loci contributing to repeated evolution of fish fin modifications. In our genomic comparisons, we focused on a specific type of convergent DNA change (deletions), because such structural changes seemed likely to yield phenotypic consequences and were already associated with repeated evolution of pelvic reduction in multiple stickleback populations (Chan et al. 2010; Xie et al. 2019). However, we note that similar genome-wide association approaches can be designed to identify many other types of DNA alterations, which likely differ in degree of pleiotropy and selective constraints (e.g., regions of accelerated base pair substitutions, convergent amino acid changes, transcription factor binding site modifications,

splicing changes, insertions, copy number differences, or multiple types of lesions co-occurring in the same gene or gene pathway in different lineages (Hiller et al. 2012a, 2012b; Chikina, Robinson and Clark 2016; Marcovitz, Jia and Bejerano 2016; Prudent et al. 2016; Lowe et al. 2017; Partha et al. 2017; Berger et al. 2018; Marcovitz et al. 2019; Sackton et al. 2019; He et al. 2020; Kowalczyk et al. 2020; Turakhia et al. 2020; Kaplow et al. 2022; Kowalczyk, Chikina and Clark 2022; Roscito et al. 2022)). We also note that our screen used a comparative genomic analysis that was anchored to a particular reference species assembly. This strategy will miss potential candidate regions which are not present in the reference assembly itself (due to either biological or artifactual reasons), and we encourage future studies to extend our efforts by using reference-free multiple sequence alignments, such as those generated by Progressive Cactus (Armstrong et al. 2020).

It is still not clear how often repeated phenotypic evolution is due to convergent mutational changes in the same genomic regions (Gompel and Prud'homme 2009; Conte et al. 2012; Martin and Orgogozo 2013). Recent reviews of known cases suggest the answer may vary depending on the phylogenetic distance between lineages in which convergent phenotypes arise, with more closely related species being most likely to evolve through similar genetic pathways (Conte et al. 2012; Ord and Summers 2015). On the other hand, reuse of certain genomic loci still appears to be appreciable even among distantly related lineages (Conte et al. 2012; Martin and Orgogozo 2013; Courtier-Orgogozo et al. 2020), and our results provide striking new examples of genomic convergence occurring across fish species that last shared a common ancestor over 100 million years ago. Although we focused on fin modifications in this study, many other types of traits also evolve repeatedly when different lineages adapt to similar ecological pressures (McGhee 2011). As fish make up nearly half of vertebrate species and also account for a large fraction of known examples of repeated evolution (Ord and Summers 2015), they may provide an especially powerful system for linking phenotypes and genotypes using tree-wide association and functional genomic studies.

## Materials and Methods

### Computational Screen to Identify Conserved Sequence Deletions

#### Genome Assemblies

See supplementary spreadsheet 1, Supplementary Material online for accession identifiers and sources. We required assemblies used in the screen to have a scaffold L50  $\leq 300$ . The scaffold L50 statistic is defined as the smallest number of scaffolds whose lengths sum to half of the total genome size. Unlike the N50 statistic, it is normalized to genome size and is thus more informative when assessing the relative quality of different species' (variably sized) genome assemblies.

### Repeat Masking

For assemblies hipEre01 and synSco01, RepeatMasker v4.1.0 (with NCBI/RMBLAST [2.10.0+] and Complete Master RepeatMasker DatabaseCONS-Dfam\_3.1-rb20181026) was used to soft-mask repetitive sequences prior to whole-genome alignment (Smit, Hubley and Green 2015). Specifically, the following commands were run:

```
RepeatMasker -engine rmbblast -species "hippocampus erectus" -s -no_is -cutoff 255 -frag 20000 hipEre01.fa
```

```
RepeatMasker -engine rmbblast -species "syngnathus scovelli" -s -no_is -cutoff 255 -frag 20000 synSco01.fa
```

### Whole-genome Alignment and Orthology Mapping

Based on its high level of contiguity and functional annotation, the Japanese Medaka genome assembly ASM223467v1 (GenBank accession # GCA\_002234675.1, assembly abbreviation oryLat04) was used as the reference to which all 35 other percomorph query genome sequences were aligned (Ichikawa et al. 2017). Ensembl release 98 gene models for ASM223467v1 served as the genomic landmarks for identifying orthologous alignment chains (Cunningham et al. 2022). LASTZ-based pairwise whole-genome alignment chains (Kent et al. 2003) were generated using the doBlastzChainNet.pl tool (<https://github.com/ENCODE-DCC/kentUtils/>, last accessed 12 Sep 2022) and the parameters listed in [supplementary spreadsheet 1, Supplementary Material](#) online. The chainLinearGap parameter was set to loose for alignment to the nonpercomorph Zebrafish assembly (danRer11), and to medium for all other percomorph genome alignments. Only alignment chains containing confident gene ortholog mappings, identified as previously described (Turakhia et al. 2020) with gene-in-syteny and second-best-chain-ratio thresholds of 10, were used in subsequent computational steps to identify conserved sequence deletions (CONDELs). All orthologous pairwise alignment chains can be viewed at the following UCSC Genome Browser assembly hub (Kent et al. 2002; Raney et al. 2014): [https://genome.ucsc.edu/s/hchen17/oryLat04\\_public](https://genome.ucsc.edu/s/hchen17/oryLat04_public).

### Identifying Conserved Sequences and Gaps in Alignment Chains

Orthologous pairwise alignment chains allow for convenient programmatic distinction between reference genome regions that are either intact and conserved at the orthologous position of the aligned query genome versus those that are missing or extensively diverged (Kent et al. 2003). We thus used the chains with gene ortholog mappings from the previous step to record single- and double-sided chain gaps (intervals predicted to be missing from the aligned genome relative to the reference) from each target species. To avoid mistaking incomplete genome assembly for genuinely missing sequence, we excluded (subtracted) chain gap intervals that are within 100 bp of an assembly gap longer than  $[N]_5$  in any given target genome. For each alignment chain, we merged any chain gap intervals that are within 20 bp of each other.

Using orthologous chains of outgroup species, we identified conserved genomic intervals by (a) recording regions in each query genome that exhibit an intact alignment block; (b) computing percent sequence identity relative to the oryLat04 reference using 10, 25, 50, and 100 bp sliding windows across the alignment blocks; (c) sorting the windows by percent sequence identity; and (d) keeping the top  $M$  most-conserved windows that, when merged and flattened, cover 5% of oryLat04, and together were deemed per-species conserved elements. For each alignment chain, we merged conserved elements within 20 bp of each other. Note that the conserved elements incorporate all additional windows with sequence identity scores tied with that of the  $M^{\text{th}}$  (lowest-scoring) window included to reach the 5% reference coverage threshold. Summary statistics for the sliding windows used to delineate conserved elements are in [supplementary table S4, Supplementary Material](#) online.

### Screen for Genomic Regions Associated With Fin Reduction

For each Japanese Medaka reference gene that was successfully mapped to at least 17 pelvis-complete outgroups and 5 pelvis-reduced target species, we scanned the interval spanning 200 kb upstream and downstream of the canonical (longest) isoform's transcription start site (400,001 bp total). Within this interval, we identified target deletions by recording regions where sequence is missing (i.e., a valid chain gap is present) in at least  $\frac{2}{3}$  of all species in each of the two officially screened target clades (Tetraodontiformes and Syngnathidae). We then intersected the target deletions with regions that were covered by conserved elements in at least 17 outgroups (not including the reference assembly itself, which is sequence-conserved by definition) and that spanned at least 20 bp to identify raw candidate CONDELs. We merged raw CONDEL candidates within 20 bp of each other and only recorded final candidate intervals that are 50 bp or larger and that do not overlap a chain gap (i.e., have a genotype-phenotype violation) in more than one scorable outgroup chain. In this procedure, a CONDEL can be "called by" (linked to) multiple gene orthologs if they are less than 200 kb from the same candidate region. The set of pCONDEL-linked orthologs was used for the functional enrichment analysis described in the following section. In [supplementary spreadsheet 2, Supplementary Material](#) online, we report unique percomorph CONDEL intervals and note all linked orthologs that call a given candidate. CONDEL candidates (as well as reference gene annotations) can also be visualized at the following UCSC Genome Browser assembly hub (Kent et al. 2002; Raney et al. 2014): [https://genome.ucsc.edu/s/hchen17/oryLat04\\_public](https://genome.ucsc.edu/s/hchen17/oryLat04_public).

### pCONDEL Functional Enrichment Test

Using GO Ontology (Ashburner et al. 2000; The Gene Ontology Consortium 2021) Biological Process Complete annotations (DOI: 10.5281/zenodo.6799722 Released 2022-07-01) and false discovery rate multiple hypothesis

testing correction, a PANTHER binomial overrepresentation test (Released 20220712) was performed on the 3,489 oryLat04 orthologs that yielded one or more pCONDEL candidates in the computational screen (Mi et al. 2019). All oryLat04 protein-coding genes (provided by PANTHER) constituted the background set for the binomial test. See [supplementary spreadsheet 3, Supplementary Material](#) online for a complete list of all ontology terms with  $q$ -value  $<0.05$ , all input genes, and all unmapped input genes (which are almost exclusively noncoding RNA genes).

#### *Identifying CONDELs With Homology in the Zebrafish Genome*

BLASTN (Altschul et al. 1997; version 2.7.1+; task specified to blastn; E-value threshold set to  $1e-4$ ; and word sizes set to 7 and 11) was used to search for homology between the Japanese Medaka sequence of each pCONDEL and the Zebrafish reference assembly danRer11 (GenBank accession # GCA\_000002035.4). BLAST hits to chrUn and “alt” chromosomes were disregarded. In addition, we considered homology suggested by any pCONDEL conserved element that intersected (by at least 10 bp) an alignment block in the liftOver chains between oryLat04 and danRer11 (generated by doBlastzChainNet.pl as described above; Kent et al. 2003). All homology candidates were then manually inspected via UCSC Genome Browser visualizations of oryLat04 and danRer11 to verify orthology or paralogy for intragenic loci as well as preservation of genomic synteny for intergenic loci (Kent et al. 2002; Raney et al. 2014). Candidate regions that were identified by the initial BLAST and LASTZ alignment criteria above but that did not show convincing homology were deemed false positive hits and were noted accordingly in [supplementary spreadsheet 2, Supplementary Material](#) online (see column “conserved to zebrafish”).

#### *Locating PelA in the Japanese Medaka Genome*

BLASTN (version 2.7.1+) (task specified to blastn) was used to locate regions of sequence similarity between the Japanese Medaka reference genome sequence and the Threespine Stickleback 500 bp and 2.5 kb *PelA* enhancer sequences described by Chan et al. (Altschul et al. 1997; Chan et al. 2010).

#### *Prediction of Transcription Factor Binding Sites Within CONDEL Candidates*

The R/Bioconductor packages TFBSTools (version 1.34.0) and JASPAR2022 (version 0.99.7) were used to identify binding site predictions for all latest-version vertebrate transcription factors in the JASPAR 2022 database with a position weight matrix match score  $\geq 925$  (out of 1000) to either the forward or reverse complement of the Japanese Medaka CONDEL sequence (Tan and Lenhard 2016; Baranasic 2022; Castro-Mondragon et al. 2022). See script predictTFBS.R in code repository.

#### *Generation of CONDEL Multiple Sequence Alignment From Orthologous Chains*

The script getGappedMSAfromENSGmappedChains\_v1.sh (see code repository) was used to extract each nonreference species' orthologous sequence at the CONDEL coordinates from the associated pairwise alignment chains. Then, a gapped multiple sequence alignment was manually collated.

#### *Phylogenetic Tree and Branch Length Calculation*

To create a multiple sequence alignment for estimating the phylogenetic distance between species in the screen, we passed the pairwise netted alignments generated by doBlastzChainNet.pl into ROAST v3 (multiz) (Blanchette et al. 2004). We then used msa\_view to identify and extract sufficient statistics for 4-fold degenerate sites in coding regions as defined by Ensembl release 98 ASM223467v1 gene models (Cunningham et al. 2022). We estimated the rate of neutral evolution at 4-fold degenerate sites using phyloFit (Hubisz, Pollard and Siepel 2011) under the REV substitution model, and used PhyloDM (Mussig 2022) to compute pairwise distances between all species in the screen ([supplementary spreadsheet 4, Supplementary Material](#) online). We based the percomorph phylogenetic tree topology on the consensus of several recent studies (see [supplementary fig. S1 and text 1, Supplementary Material](#) online) (Alfaro et al. 2018; Hughes et al. 2018; Mu et al. 2022). MEGA11 (version 11.0.13) was used to format and draw the resulting branch length-calibrated tree (Tamura, Stecher and Kumar 2021).

#### *Nearest Outgroup Definition*

To identify a given target clade's nearest outgroup for use in functional experiments, we considered the following factors in the listed order:

- 1) Proximity based on the most recent common ancestor by phylogenetic tree topology ([fig. 1A](#))
- 2) Pairwise genetic distance based on the neutral model of evolution described above (see [supplementary fig. S1, text 1, and file 4, Supplementary Material](#) online)
- 3) Practical availability of high-quality genomic DNA or adequately preserved tissue samples

#### *Testing for Significant Phenotypic Differences Between Target and Outgroup Species*

“Phylogenetically corrected” means for the numbers of pelvic fins and caudal rays in target and outgroup species were estimated using the phyloMean function of the MOTMOT R package (Thomas and Freckleton 2012; Puttick et al. 2019). Then, two-tailed Welch's  $t$ -tests were performed to determine if the traits differed significantly between target and outgroup species. See comparePhyloMeans.R script in the code repository.

#### *Percomorph Tissue and Genomic DNA Samples*

Percomorph genomic DNA was isolated from ethanol-preserved fin, muscle, or liver tissue by incubation in lysis

buffer (10 mM Tris, pH 8, 100 mM NaCl, 10 mM EDTA, 0.5% SDS, 500 µg/ml Proteinase K) at 55 °C for 4–12 h, followed by extraction with 25:24:1 phenol:chloroform:isoamyl alcohol, 1–2 washes in chloroform, ethanol precipitation, and resuspension of the DNA pellet in water or TE buffer. Sample and tissue sources are listed in [supplementary table S2, Supplementary Material](#) online.

### Species Verification by Cytochrome Oxidase Subunit I Sequencing

A segment of the mitochondrial gene *cytochrome oxidase subunit I* (*COI*) was amplified using DreamTaq Green PCR Master Mix (Thermo Scientific catalog # K1081) and primers percomorph-COI-fwd and percomorph-COI-rev (see [supplementary table S5, Supplementary Material](#) online), according to manufacturer recommendations. The subsequent Sanger sequencing reads were cross-referenced with the Barcode of Life Data System (<http://v4.boldsystems.org/>) to confirm species identity prior to further empirical interrogation of the genetic material (Ratnasingham and Hebert 2007). *COI* amplicon sequences of non-*Gasterosteus* species have been deposited in GenBank (see [supplementary table S2, Supplementary Material](#) online for accession numbers).

### Sequence Verification and Cloning of Candidate CONDEL Regions

We verified the computationally derived pCONDEL candidates near *Pitx1* and *Faf1* by PCR and Sanger sequencing to confirm the DNA sequence represented by the genome assemblies and alignments, according to the following procedure. To compare equivalent genomic sequence between species, we first identified—in the Japanese Medaka genome assembly—the nearest intervals upstream and downstream of the pCONDEL candidate that exhibit alignment blocks in nearly every outgroup and target species with a successful orthologous mapping to *Pitx1* or *Elavl4* and *Faf1*, respectively (see *Whole-genome Alignment and Orthology Mapping* methods above). We then used the coordinates of these flanking conservation anchors to extract comparable sequence from each query genome assembly using the script `getConsAnchoredOrthologousFastas.sh` (see code repository).

We designed amplification primers for the extracted assembly sequences using Geneious Prime (Biomatters Ltd) and generated PCR amplicons using Q5 High-Fidelity 2X Master Mix (New England Biolabs catalog # M0492), Phusion High-Fidelity PCR Master Mix with HF Buffer (New England Biolabs catalog # M0531), and/or LongAmp® Hot Start Taq DNA Polymerase (New England Biolabs catalog # M05345) with Expand HF Buffer (Roche catalog # 05917131103) according to manufacturer recommendations. The PCR amplicons were then integrated into vector pCR4Blunt-TOPO using the Zero Blunt TOPO PCR Cloning Kit (Invitrogen catalog # 450031) for subsequent Sanger sequencing. Primers introducing homology arms for Gibson Assembly were used to

clone candidate sequences into either the GFP reporter vector pT2HE (Howes, Summers and Kingsley 2017) and/or the firefly luciferase reporter vector pGL4.23[luc2/minP] (Promega, Genbank Accession # DQ904455.1) using NEBuilder HiFi DNA Assembly Master Mix (New England Biolabs catalog # E2621). Reporter constructs and candidate region amplicon sequences have been deposited in GenBank (see [supplementary table S2, Supplementary Material](#) online for accession numbers).

### Threespine Stickleback Husbandry

All Threespine Stickleback in this study were raised in 29-gallon tanks under standard aquarium conditions (3.5 g/L Instant Ocean salt, 18 °C) and fed live brine shrimp as larvae, then frozen daphnia, bloodworms, and/or mysis shrimp as juveniles and adults—in accordance with the Guide for the Care and Use of Laboratory Animals of the National Institutes of Health under Protocol #13834 of the Institutional Animal Care and Use Committee of Stanford University. All live fish in this study were lab-raised descendants of wild-caught fish. All quantitative in vivo experiments that tested specific hypotheses adhered to ARRIVE guidelines (Percie du Sert et al. 2020).

### Enhancer Reporter Assays in Transgenic Threespine Stickleback

Microinjection of freshly fertilized Threespine Stickleback eggs with *Tol2* transposase mRNA and GFP reporter constructs was performed as previously described (Chan et al. 2010; Thompson et al. 2018). All resulting larvae were raised under standard aquarium conditions to Swarup stage 30 (Swarup 1958) and then euthanized in 600 mg/L tricaine, pH 7.5 for phenotyping. GFP expression was documented using a Leica MZFLIII fluorescence microscope (Leica Microsystems) fitted with a GFP2 filter and ProgResCF camera (Jenoptik AG). Only fish exhibiting bilateral green eyes (from background expression by the *hsp70* minimal promoter of the expression construct, indicating extent of transgenesis) were phenotyped (Nagayoshi et al. 2008). The percomorph *PelA* constructs were assayed on lab-raised pelvic- and caudal-complete stickleback descended from fish collected at Rabbit Slough, Alaska, USA. The construct containing the Japanese Medaka ortholog of the *Faf1* CONDEL candidate region was assayed on lab-raised pelvic- and caudal-complete Threespine Stickleback descended from fish collected at Matadero Creek, California, USA and at Little Campbell River, British Columbia, Canada. The experiment did not include blinding, but post hoc PCR confirmation of integrated plasmid identity was performed.

### Luciferase Enhancer Reporter Assays

OLHNI-2, a Japanese Medaka fibroblast-like fin-derived cell line, was acquired from RIKEN BioResearch Resource Center (Cell No. RCB2942) and verified by sequencing a segment of the mitochondrial gene *COI* (see *Species Verification by COI Sequencing* above) (Hirayama, Mitani

and Watabe 2006). The cells were cultured using Leibovitz's L-15 Medium (Gibco Catalog # 11415064) supplemented to 20% fetal bovine serum (ATCC Catalog # 30-2020) and  $1 \times$  penicillin–streptomycin (Gibco Catalog # 15140122), at 30 °C to 33 °C and ambient CO<sub>2</sub> concentration. Firefly and pRL-SV40 renilla luciferase constructs (Promega, Genbank Accession # AF025845) were transfected using Solution SF (Lonza Catalog # V4SC-2096) and program DN-100 of the Amaxa Nucleofector 96-well Shuttle System according to manufacturer recommendations. Each firefly luciferase plasmid was independently amplified, minipreped (ZymoResearch Catalog # D4210), and tested the following number of times: Marine Stickleback, 12; Yellow Croaker, 33; Green Spotted Puffer, 12; Ocean Sunfish, 8; European Seabass, 10; Gulf Pipefish, 11. Each plasmid preparation replicate was assayed in technical quadruplicate and reported as a single averaged point in figure 2D. For interexperiment normalization, aliquots of a single preparation of the basal firefly vector pGL4.23[luc2/minP] (Promega, Genbank Accession # DQ904455.1) were assayed on each experimental day. Dual-Luciferase Reporter Assay System reagents (Promega Catalog # E1910) and a Dual Injector System for GloMax-Multi Detection System (Promega) luminometer were used to measure luciferase activity. The plasmid locations on each plate were randomized to control for potential positional biases from the multichannel pipettes, nucleofector, and luminometer that were used. Two-tailed Mann–Whitney *U* tests were performed (using python package *scipy* v1.7.3) to compare the distribution of biological replicates (each the average of technical quadruplicates) between constructs, as this statistical test does not require the data to be normally distributed (Virtanen et al. 2020).

### RNAseq

A lab-raised Matadero Creek male Threespine Stickleback and lab-raised Little Campbell River female Threespine Stickleback were crossed to generate one clutch of wild-type F1 hybrid fish. Caudal fin buds were collected into 1.5 ml centrifuge tubes from 12 siblings of the clutch on 10 dpf (after hatching, before lepidotrichia development, Swarup stage 27) and 12 additional siblings on 17 dpf (during lepidotrichia development, Swarup stage 29) (Swarup 1958). The dissected tissues were immediately placed on dry ice and stored at  $-80$  °C until RNA extraction was performed using the RNeasy Plus Micro Kit (Qiagen Catalog # 74034). Each fin bud was first homogenized in 50  $\mu$ l of Buffer RLT (supplemented with  $\beta$ -mercaptoethanol per kit recommendation) for 1 min at RT in a 1.5 ml centrifuge tube using a cordless motor (VWR Catalog # 47747-370) and pestle (USA Scientific Catalog # 1415-5390). The pestle was rinsed with an additional 300  $\mu$ l of Buffer RLT into the 1.5 ml tube; then the resulting 350  $\mu$ l volume was drawn twice into a 1 ml luer slip tip tuberculin syringe (BD Catalog # 309659) fitted with a 27G- $\frac{1}{2}$ " needle. Each RNA sample was eluted in 16  $\mu$ l of RNase-free water and

quantified using the Qubit RNA High Sensitivity Kit (Invitrogen Catalog # Q32852). A subset of the samples from each developmental stage was quality-checked by Bioanalyzer using the RNA 6000 Pico Kit (Agilent Technologies Catalog # 5067-1513). The resulting RNA integrity numbers were between 8.4 and 10, with most values 9.4 or higher.

Sequencing libraries were generated with the Stranded mRNA Prep kit (Illumina Catalog # 20040532) using 68–80 ng of RNA for 10 dpf samples and 150 ng of RNA for 17 dpf samples. The libraries were sequenced ( $2 \times 150$  bp) as a single multiplexed pool on two partial lanes of an Illumina NovaSeq 6000 by Novogene Corporation Inc. The resulting sequencing reads were trimmed with *fastp* v0.21.0 (Chen et al. 2018) and mapped to Threespine Stickleback genome assembly *gasAcu1-4* via two-pass mapping with STAR v2.7.7a (Dobin et al. 2013) according to GATK best practices (Van der Auwera et al. 2013). The program *featureCounts* v2.0.1 (Liao, Smyth and Shi 2014) was used to quantify reads mapped per gene based on *gasAcu1* gene annotations (Ensembl release 99) that were lifted to *gasAcu1-4* as previously described (Roberts Kingman et al. 2021; Cunningham et al. 2022). Each sample's gene-level read counts were normalized to transcripts per million to assess expression levels of genes near *Faf1* at the two developmental stages (Li et al. 2010).

### CRISPR-Cas9 Genome Editing

Single-guide RNAs (sgRNAs) for directing Cas9 nuclease activity in Threespine Stickleback were designed using CHOPCHOP (Labun et al. 2016, 2019) and the genome assembly *gasAcu1*. Templates for the sgRNAs were synthesized via 2-oligo PCR and used for in vitro transcription by T7 RNA polymerase, as previously described (Wucherpennig, Miller and Kingsley 2019). In addition to sgRNAs targeting candidate regions, an sgRNA for inactivating the *golden* pigment gene *Slc24a5* (Lamason et al. 2005) was also included as a visual control for editing status. Supplementary table S5, Supplementary Material online lists all oligonucleotides used to generate sgRNAs.

For recreating the pCONDEL.1189 candidate near *Faf1* in Threespine Stickleback, a 60 bp single-stranded DNA donor template was injected along with the sgRNAs and Cas9 protein. Intended for promoting homology-directed repair, this oligonucleotide was a concatemer of the two 30-mers on either side of the desired 464 bp deletion (see fig. 4A and supplementary table S5, Supplementary Material online), and the two most 5' and two most 3' bases were synthesized to have phosphorothioate (not phosphodiester) bonds to enhance stability in vivo (Integrated DNA Technologies, Inc).

Cas9-2NLS protein (QB3 MacroLab, University of California—Berkeley), sgRNAs, and single-stranded DNA oligonucleotides were injected at the indicated concentrations (supplementary table S3, Supplementary Material online) into freshly fertilized Threespine Stickleback eggs. For experiments in which half of each clutch was injected

with reagents to target the pCONDEL.1189 region and the other half was injected with sgRNAs targeting only the *golden* locus, the order of injection was alternated between each of the 17 genome-edited clutches to control for any potentially confounding effects related to the time of injection after fertilization. To verify editing status in the resulting fish, dorsal spine and dorsal fin biopsies and/or skin mucus swabs were collected from fish after reaching  $\geq 20$  mm standard length and used for subsequent DNA extraction (Breacker et al. 2017). All phenotypic scoring was performed at or after fish reached 15 mm standard length, before any fin tissues were biopsied for genotyping; blinding was not performed. All CRISPR-Cas9 editing experiments were performed on either the Matadero Creek and/or Little Campbell River population background. Two-tailed, Boschloo's exact tests were performed (using python package *scipy* v1.7.3) to ascertain whether the genome editing was significantly associated with phenotypic outcome, as this statistical test is conditioned only on one marginal sum (the number of fish in each treatment or control group) (Boschloo 1970; Mehrotra, Chan and Berger 2003; Ludbrook 2013; Virtanen et al. 2020).

**Sample size calculations:** A pilot experiment to target the pCONDEL.1189 candidate region in Threespine Stickleback from Matadero Creek suggested the ectopic caudal ray phenotype occurred in  $\sim 5\%$  of injected individuals—and the controls for this pilot experiment consisted of unmodified clutch siblings, which all developed wild-type caudal fins. After completing the pilot experiment, we incidentally observed that a spontaneous caudal fin morphology reminiscent of the pCONDEL.1189-edited phenotype appears at very low frequencies ( $\sim 0.5\%$ ) in wild-type, unmodified Threespine Stickleback from Matadero Creek. Given this finding, we performed power calculations ( $\alpha = 0.05$ ,  $\beta = 0.2$ ,  $\text{power} = 0.8$ ) to determine the appropriate sample size for a replication experiment in which control siblings would also be injected to edit the unrelated *golden* pigment locus to account for possible confounding effects from introducing Cas9 nuclease into the zygote. We report full results of the replication experiment here (involving 17 genome-edited clutches total; see [supplementary fig. S5, Supplementary Material](#) online). Because the *golden*-edited controls from the replication experiment above did not suggest confounding effects from the presence of Cas9 nuclease, the subsequent experiment targeting coding exons of *Faf1* used unmodified clutch siblings as controls. A pilot experiment to knockout *Faf1* yielded a similar ( $\sim 5\%$ ) frequency of ectopic caudal ray phenotypes, so we aimed to achieve a similar sample size as above to test the effect of *Faf1* inactivation in a subsequent replication experiment (involving 10 genome-edited clutches total; see [supplementary fig. S8, Supplementary Material](#) online).

### Primers and DNA Oligonucleotides

Sequences for all DNA oligonucleotides and genotyping primers are listed in [supplementary table S5, Supplementary Material](#) online.

### Skeletal Preparations

Adult Threespine Stickleback ( $\geq 30$  mm standard length), a Green Spotted Pufferfish, and a juvenile Rice Eel were fixed in 10% neutral buffered formalin for at least 5 days, rinsed twice in distilled water and twice in 30% saturated sodium borate, and then cleared in a solution of 1% porcine trypsin dissolved in 30% saturated sodium borate for 24–48 h at RT. After equilibrating the cleared specimens to 2% potassium hydroxide (KOH), staining of calcified skeletal tissue was performed by incubation in 0.005% Alizarin Red S dissolved in 2% KOH for 12–24 h; and then pigment was bleached by incubating the fish in 0.375% KOH, 25% glycerol, and 0.0003% hydrogen peroxide. Juvenile Threespine Stickleback ( $\leq 20$  mm standard length) that required genotyping after skeletal characterization were stained using Walker and Kimmel's two-color acid-free method (at 100 mM magnesium chloride concentration) (Walker and Kimmel 2007). Finally, all specimens were passed through a 0.375% KOH:glycerol gradient before imaging and storage in 100% glycerol. An Epson Perfection V800 Photo scanner and a Stemi SV11 stereomicroscope fitted with an AxioCam HRc camera (Carl Zeiss AG), respectively, were used to record brightfield images of adult and juvenile skeletal preparations.

### Code, Data, and Material Availability

A code repository with instructions for replicating the computational screen starting with flat-text genome assembly fasta files is available at <https://github.com/bejerano-lab/percomorphCONDELS>. To aid users who may not have access to a compute cluster, parts of the screen were implemented to optionally use GNU Parallel (Tange 2011). [Supplementary spreadsheet 1, Supplementary Material](#) online lists genome assembly accession identifiers and sources. A copy of precomputed, genome assembly-derived intermediate inputs (including whole-genome alignment chains) for generating the final candidate list has been deposited at <https://doi.org/10.5281/zenodo.7140838>. Gene annotations and pCONDEL candidate regions in the Japanese Medaka reference genome can be viewed at the following UCSC Genome Browser assembly hub (Kent et al. 2002; Raney et al. 2014): [https://genome.ucsc.edu/s/hchen17/oryLat04\\_public](https://genome.ucsc.edu/s/hchen17/oryLat04_public). A copy of the files underlying the Japanese Medaka assembly hub is available at <https://doi.org/10.5281/zenodo.7909719>. [Supplementary table S2, Supplementary Material](#) online lists the source of percomorph tissue and genomic DNA used in this study as well as the GenBank accession numbers for the derived DNA sequences and constructs. Raw RNA-seq data are available through National Center for Biotechnology Information (NCBI) BioProject number PRJNA908888. Other materials will be made available upon reasonable request.

### Supplementary Material

[Supplementary data](#) are available at *Molecular Biology and Evolution* online.

## Acknowledgements

We acknowledge the Ohlone, Dena'ina, and Semiahmoo Nations, from whose traditional territories the Threespine Stickleback used in this study were derived. We are grateful to the following people for providing nonstickleback tissue and/or DNA samples used in this study: Leo Smith (KUBI Ichthyology), Andrew Bentley (KUBI Ichthyology), Pierre-Alexandre Gagnaire, François Bonhomme, Jessica Muniz, Angel Seery, Leo Nico, Chris Smith, Scott Collenberg, Carol Cozzi-Schmarr, Xuegeng Wang, Ramji Bhandari, Susan L. Bassham, William A. Cresko, Molly Schumer, Shreya Banerjee, Anne Brunet, Rahul Nagvekar, Leonor Cancela, and Vincent Laizé. We thank the National Center for Biotechnology Information and the numerous researchers who generated and made freely available the genome assemblies used in this study; the UCSC Genome Browser Team and Hiram Clawson for the generously shared wisdom and tools they have developed for visualizing and analyzing genomic data; Dave Catania (California Academy of Sciences) for assistance in obtaining radiographs of various fish species; Julia Wucherpfennig, Abbey Thompson, Ken Thompson, and Dolph Schluter for help collecting sticklebacks from Little Campbell River; Sarah Frail for contributions to initial CRISPR/Cas9 experiments; William Talbot for generously lending imaging equipment; Shannon Brady and the team of fish feeders for critical husbandry support; and all members of the Bejerano and Kingsley Labs (especially Erich Weiler, Mark Berger, Amir Marcovitz, Johannes Birgmeier, Karthik Jagadeesh, Whitney Heavner, Boyoung Yoo, Yosuke Tanigawa, Janet Song, Amy Herbert, Julia Wucherpfennig, and Hiren Banerjee) as well as Michael A. Bell, William Talbot, and Margaret Fuller for their helpful suggestions. This work was supported by a National Science Foundation Graduate Research Fellowship (1656518), NIH training grant (5T32GM007790), and Stanford CEHG Fellowship to H.I.C.; D.M.K. is an investigator of the Howard Hughes Medical Institute.

## Author Contributions

H.I.C., G.B., and D.M.K. conceptualized the computational screen. Y.T. provided code for identifying orthologous alignment chains. H.I.C. curated and aligned the publicly available genome assemblies and wrote the code for performing the computational screen. H.I.C. and D.M.K. conceptualized the functional experiments for testing candidate regions. H.I.C. carried out the investigations and performed all data analysis and visualization. H.I.C. and D.M.K. wrote the manuscript with input from all authors. G.B. and D.M.K. supervised the study, managed resources, and acquired funding.

**Conflict of interest statement.** H.I.C., Y.T., G.B., and D.M.K. have no conflicts of interest to declare.

## References

Alfaro ME, Faircloth BC, Harrington RC, Sorenson L, Friedman M, Thacker CE, Oliveros CH, Černý D, Near TJ. 2018. Explosive

- diversification of marine fishes at the Cretaceous–Palaeogene boundary. *Nat Ecol Evol.* **2**(4):688–696.
- Altschul SF, Madden TL, Schäffer AA, Zhang J, Zhang Z, Miller W, Lipman DJ. 1997. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res.* **25**(17):3389–3402.
- Armstrong J, Hickey G, Diekhans M, Fiddes IT, Novak AM, Deran A, Fang Q, Xie D, Feng S, Stiller J, *et al.* 2020. Progressive cactus is a multiple-genome aligner for the thousand-genome era. *Nature* **587**(7833):246–251.
- Ashburner M, Ball CA, Blake JA, Botstein D, Butler H, Cherry JM, Davis AP, Dolinski K, Dwight SS, Eppig JT, *et al.* 2000. Gene ontology: tool for the unification of biology. *Nat Genet.* **25**(1):25–29.
- Baranasik D. 2022. JASPAR2022: Data package for JASPAR database (version 2022). R package version 0.99.7.
- Bell MA, Foster SA. 1994. *The evolutionary biology of the threespine stickleback*. New York: Oxford University Press. p. 1–468.
- Berger MJ, Wenger AM, Guturu H, Bejerano G. 2018. Independent erosion of conserved transcription factor binding sites points to shared hindlimb, vision and external testes loss in different mammals. *Nucleic Acids Res.* **46**(18):9299–9308.
- Biscotti MA, Carducci F, Olmo E, Canapa A. 2019. Vertebrate genome size and the impact of transposable elements in genome evolution. In: *Evolution, origin of life, concepts and methods*. Cham: Springer International Publishing: 233–251.
- Blanchette M, Kent WJ, Riemer C, Elnitski L, Smit AFA, Roskin KM, Baertsch R, Rosenbloom K, Clawson H, Green ED, *et al.* 2004. Aligning multiple genomic sequences with the threaded blockset aligner. *Genome Res.* **14**(4):708–715.
- Boschloo RD. 1970. Raised conditional level of significance for the 2-table when testing the equality of two probabilities. *Stat Neerl.* **24**(1):1–9.
- Breacker C, Barber I, Norton WHJ, McDermid JR, Tilley CA. 2017. A low-cost method of skin swabbing for the collection of DNA samples from small laboratory fish. *Zebrafish.* **14**(1):35–41.
- Castro-Mondragon JA, Riudavets-Puig R, Rauluseviciute I, Berhanu Lemma R, Turchi L, Blanc-Mathieu R, Lucas J, Boddie P, Khan A, Manosalva Pérez N, *et al.* 2022. JASPAR 2022: the 9th release of the open-access database of transcription factor binding profiles. *Nucleic Acids Res.* **50**(D1):D165–D173.
- Chan YF, Marks ME, Jones FC, Villarreal G, Shapiro MD, Brady SD, Southwick AM, Absher DM, Grimwood J, Schmutz J, *et al.* 2010. Adaptive evolution of pelvic reduction in sticklebacks by recurrent deletion of a Pitx1 enhancer. *Science* **327**(5963):302–305.
- Chen S, Zhou Y, Chen Y, Gu J. 2018. Fastp: an ultra-fast all-in-one FASTQ preprocessor. *Bioinformatics* **34**(17):i884–i890.
- Chikina M, Robinson JD, Clark NL. 2016. Hundreds of genes experienced convergent shifts in selective pressure in marine mammals. *Mol Biol Evol.* **33**(9):2182–2192.
- Chu K, Niu X, Williams LT. 1995. A Fas-associated protein factor, FAF1, potentiates Fas-mediated apoptosis. *Proc Natl Acad Sci U S A.* **92**(25):11894–11898.
- Clack JA. 2009. The fin to limb transition: new data, interpretations, and hypotheses from paleontology and developmental biology. *Annu Rev Earth Planet Sci.* **37**(1):163–179.
- Conte GL, Arnegard ME, Peichel CL and Schluter D. 2012. The probability of genetic parallelism and convergence in natural populations. *Proc R Soc B Biol Sci.* **279**(1749):5039–5047.
- Courtier-Orgogozo V, Arnould L, Prigent SR, Wiltgen S, Martin A. 2020. Gephebase, a database of genotype–phenotype relationships for natural and domesticated variation in eukaryotes. *Nucleic Acids Res.* **48**(D1):D696–D703.
- Cunningham F, Allen JE, Allen J, Alvarez-Jarreta J, Amode MR, Armean IM, Austine-Orimoloye O, Azov AG, Barnes I, Bennett R, *et al.* 2022. Ensembl 2022. *Nucleic Acids Res.* **50**(D1):D988–D995.
- Darwin C. 1859. *On the origin of species by means of natural selection, or preservation of favoured races in the struggle for life*. London: John Murray.

- Davenport J. 1994. How and why do flying fish fly?. *Rev Fish Biol Fisheries*. **4**(2):184–214.
- Dobin A, Davis CA, Schlesinger F, Drenkow J, Zaleski C, Jha S, Batut P, Chaisson M, Gingeras TR. 2013. STAR: ultrafast universal RNA-Seq aligner. *Bioinformatics* **29**(1):15–21.
- French V, Bryant PJ, Bryant SV. 1976. Pattern regulation in epimorphic fields: cells may make use of a polar coordinate system for assessing their positions in developing organs. *Science*. **193**(4257):969–981.
- Giammona FF. 2021. Form and function of the caudal fin throughout the phylogeny of fishes. *Integr Comp Biol*. **61**(2):550–572.
- Goldberg DL, Landy JA, Travis J, Springer MS and Reznick DN. 2019. In love and war: the morphometric and phylogenetic basis of ornamentation, and the evolution of male display behavior, in the livebearer genus *Poecilia*. *Evolution* **73**(2):360–377.
- Gompel N, Prud'homme B. 2009. The causes of repeated genetic evolution. *Dev Biol*. **332**(1):36–47.
- Gregory TR. 2022. *Animal Genome Size Database*. <https://www.genomesize.com>
- Guerreiro I, Nunes A, Woltering JM, Casaca A, Nóvoa A, Vinagre T, Hunter ME, Duboule D, Mallo M. 2013. Role of a polymorphism in a HOX/PAX-responsive enhancer in the evolution of the vertebrate spine. *Proc Natl Acad Sci U S A*. **110**(26):10682–10686.
- He Z, Xu S, Zhang Z, Guo W, Lyu H, Zhong C, Boufford DE, Duke NC, The International Mangrove Consortium, Shi S. 2020. Convergent adaptation of the genomes of woody plants at the land–sea interface. *Natal Sci Rev*. **7**(6):978–993.
- Hiller M, Schaar BT, Bejerano G. 2012a. Hundreds of conserved non-coding genomic regions are independently lost in mammals. *Nucleic Acids Res*. **40**(22):11463–11476.
- Hiller M, Schaar BT, Indjeian VB, Kingsley DM, Hagey LR, Bejerano G. 2012b. A “forward genomics” approach links genotype to phenotype using independent phenotypic losses among related species. *Cell Rep*. **2**(4):817–823.
- Hirayama M, Mitani H, Watabe S. 2006. Temperature-dependent growth rates and gene expression patterns of various medaka *Oryzias latipes* cell lines derived from different populations. *J Comp Physiol B*. **176**(4):311–320.
- Howes TR, Summers BR, Kingsley DM. 2017. Dorsal spine evolution in threespine sticklebacks via a splicing change in *MSX2A*. *BMC Biol*. **15**:115.
- Hubisz MJ, Pollard KS, Siepel A. 2011. PHAST and RPHAST: phylogenetic analysis with space/time models. *Brief Bioinform*. **12**(1):41–51.
- Hughes LC, Orti G, Huang Y, Sun Y, Baldwin CC, Thompson AW, Arcila D, Betancur-R R, Li C, Becker L, et al. 2018. Comprehensive phylogeny of ray-finned fishes (Actinopterygii) based on transcriptomic and genomic data. *Proc Natl Acad Sci U S A*. **115**(24):6249–6254.
- Huxley TH. 1859. Observations on the development of some parts of the skeleton of fishes. *Transac Microsc Soc J*. **7**(1):33–46.
- Ichikawa K, Tomioka S, Suzuki Y, Nakamura R, Doi K, Yoshimura J, Kumagai M, Inoue Y, Uchida Y, Irie N, et al. 2017. Centromere evolution and CpG methylation during vertebrate speciation. *Nat Commun*. **8**:1833.
- IUCN. 2022. The IUCN red list of threatened species. Version 2022–1: Table 1a.
- Kaplow IM, Schäffer DE, Wirthlin ME, Lawler AJ, Brown AR, Kleyman M, Pfening AR. 2022. Inferring mammalian tissue-specific regulatory conservation by predicting tissue-specific differences in open chromatin. *BMC Genomics*. **23**:291.
- Katsiadaki I, Sanders M, Sebire M, Nagae M, Soyano K, Scott AP. 2007. Three-spined stickleback: an emerging model in environmental endocrine disruption. *Environ Sci*. **14**(5):263–283.
- Kempainen P, Li Z, Rastas P, Löytynoja A, Fang B, Yang J, Guo B, Shikano T, Merilä J. 2021. Genetic population structure constrains local adaptation in sticklebacks. *Mol Ecol*. **30**(9):1946–1961.
- Kent WJ, Baertsch R, Hinrichs A, Miller W, Haussler D. 2003. Evolution's cauldron: duplication, deletion, and rearrangement in the mouse and human genomes. *Proc Natl Acad Sci U S A*. **100**(20):11484–11489.
- Kent WJ, Sugnet CW, Furey TS, Roskin KM, Pringle TH, Zahler AM, Haussler D. 2002. The human genome browser at UCSC. *Genome Res*. **12**(6):996–1006.
- Kolora SRR, Owens GL, Vazquez JM, Stubbs A, Chatla K, Jainese C, Seeto K, McCrea M, Sandel MW, Vianna JA, et al. 2021. Origins and evolution of extreme life span in Pacific Ocean rockfishes. *Science* **374**(6569):842–847.
- Kowalczyk A, Chikina M, Clark NL. 2022. Complementary evolution of coding and noncoding sequence underlies mammalian hairlessness. *eLife*. **11**:e76911.
- Kowalczyk A, Partha R, Clark NL, Chikina M. 2020. Pan-mammalian analysis of molecular constraints underlying extended lifespan. *eLife*. **9**:e51089.
- Kumar S, Suleski M, Craig JM, Kasprowicz AE, Sanderford M, Li M, Stecher G, Hedges SB. 2022. Timetree 5: an expanded resource for species divergence times. *Mol Biol Evol*. **39**(8):msac174.
- Kvon EZ, Kamneva OK, Melo US, Barozzi I, Osterwalder M, Mannion BJ, Tissières V, Pickle CS, Plajzer-Frick I, Lee EA, et al. 2016. Progressive loss of function in a limb enhancer during snake evolution. *Cell* **167**(3):633–642.
- Labun K, Montague TG, Gagnon JA, Thyme SB, Valen E. 2016. CHOPCHOP V2: a web tool for the next generation of CRISPR genome engineering. *Nucleic Acids Res*. **44**(W1):W272–W276.
- Labun K, Montague TG, Krause M, Torres Cleuren YN, Tjeldnes H, Valen E. 2019. CHOPCHOP V3: expanding the CRISPR web toolbox beyond genome editing. *Nucleic Acids Res*. **47**(W1):W171–W174.
- Lamason RL, Mohideen M-APK, Mest JR, Wong AC, Norton HL, Aros MC, Juryneć MJ, Mao X, Humphreville VR, Humbert JE, et al. 2005. SLC24A5, A putative cation exchanger, affects pigmentation in zebrafish and humans. *Science*. **310**(5755):1782–1786.
- Larouche O, Zelditch ML, Cloutier R. 2017. Fin modules: an evolutionary perspective on appendage disparity in basal vertebrates. *BMC Biol*. **15**:32.
- Larouche O, Zelditch ML, Cloutier R. 2019. A critical appraisal of appendage disparity and homology in fishes. *Fish Fisheries*. **20**(6):1138–1175.
- Letelier J, Naranjo S, Sospedra-Arrufat I, Martínez-Morales JR, López-Ríos J, Shubin N, Gómez-Skarmeta JL. 2021. The *Shh/Gli3* gene regulatory network precedes the origin of paired fins and reveals the deep homology between distal fins and digits. *Proc Natl Acad Sci U S A*. **118**(46):e2100575118.
- Lewin HA, Richards S, Lieberman Aiden E, Allende ML, Archibald JM, Bálint M, Barker KB, Baumgartner B, Belov K, Bertorelle G, et al. 2022. The Earth BioGenome Project 2020: starting the clock. *Proc Natl Acad Sci U S A*. **119**(4):e2115635118.
- Lewin HA, Robinson GE, Kress WJ, Baker WJ, Coddington J, Crandall KA, Durbin R, Edwards SV, Forest F, Gilbert MTP, et al. 2018. Earth BioGenome Project: sequencing life for the future of life. *Proc Natl Acad Sci U S A*. **115**(17):4325–4333.
- Li B, Ruotti V, Stewart RM, Thomson JA, Dewey CN. 2010. RNA-Seq gene expression estimation with read mapping uncertainty. *Bioinformatics* **26**(4):493–500.
- Liao Y, Smyth GK, Shi W. 2014. featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics* **30**(7):923–930.
- Lindsey CC. 1962. Experimental study of meristic variation in a population of threespine sticklebacks, *Gasterosteus aculeatus*. *Can J Zool*. **40**(2):271–312.
- Lleras-Forero L, Winkler C, Schulte-Merker S. 2020. Zebrafish and medaka as models for biomedical research of bone diseases. *Dev Biol*. **457**(2):191–205.
- Lowe CB, Sanchez-Luege N, Howes TR, Brady SD, Daugherty RR, Jones FC, Bell MA, Kingsley DM. 2017. Detecting differential



- copy number variation between groups of samples. *Genome Res.* **28**:256–265.
- Ludbrook J. 2013. Analysing 2x2 contingency tables: which test is best? *Clin Exp Pharmacol Physiol.* **40**(3):177–180.
- Marcovitz A, Jia R, Bejerano G. 2016. “Reverse genomics” predicts function of human conserved noncoding elements. *Mol Biol Evol.* **33**(5):1358–1369.
- Marcovitz A, Turakhia Y, Chen HI, Gloude-mans M, Braun BA, Wang H, Bejerano G. 2019. A functional enrichment test for molecular convergent evolution finds a clear protein-coding signal in echolocating bats and whales. *Proc Natl Acad Sci U S A.* **116**(42):21094–21103.
- Martin A, Orgogozo V. 2013. The loci of repeated evolution: a catalog of genetic hotspots of phenotypic variation. *Evolution* **67**(5):1235–1250.
- McGhee GR. 2011. *Convergent evolution: limited forms most beautiful.* Cambridge, MA: MIT Press. p. 1–286.
- Mehrotra DV, Chan ISF, Berger RL. 2003. A cautionary note on exact unconditional inference for a difference between two independent binomial proportions. *Biometrics* **59**(2):441–450.
- Mi H, Muruganujan A, Ebert D, Huang X, Thomas PD. 2019. PANTHER Version 14: more genomes, a new PANTHER GO-slim and improvements in enrichment analysis tools. *Nucleic Acids Res.* **47**(D1):D419–D426.
- Miller CT, Beleza S, Pollen AA, Schluter D, Kittles RA, Shriver MD, Kingsley DM. 2007. *Cis*-regulatory changes in *Kit ligand* expression and parallel evolution of pigmentation in sticklebacks and humans. *Cell.* **131**(6):1179–1189.
- Mu X, Yang Y, Sun J, Liu Y, Xu M, Shao C, Chu KH, Li W, Liu C, Gu D, et al. 2022. FishPIE: a universal phylogenetically informative exon markers set for ray-finned fishes. *iScience* **25**:105025.
- Mussig AJ. 2022. PhyloDM (2.1.0). *Zenodo*. <https://doi.org/10.5281/zenodo.6910552>
- Nagayoshi S, Hayashi E, Abe G, Osato N, Asakawa K, Urasaki A, Horikawa K, Ikeo K, Takeda H, Kawakami K. 2008. Insertional mutagenesis by the *Tol2* transposon-mediated enhancer trap approach generated mutations in two developmental genes: *tc7* and *synembryn-like*. *Development* **135**(1):159–169.
- Nelson JS. 1989. Analysis of the multiple occurrence of pelvic fin absence in extant fishes. *Matsya.* **15**:21–38.
- Nelson JS, Grande T, Wilson MVH. 2016. *Fishes of the world.* 5th ed. Hoboken, New Jersey: John Wiley & Sons. p. 1–650.
- Norman JR. 1949. *A history of fishes.* New York: A. A. Wyn. p. 1–436.
- Ord TJ, Summers TC. 2015. Repeated evolution and the impact of evolutionary history on adaptation. *BMC Evol Biol.* **15**:137.
- Ostlund-Nilsson S, Mayer I, Huntingford FA. 2006. *Biology of the three-spined stickleback.* 1st ed. Boca Raton: CRC Press. p. 1–372.
- Partha R, Chauhan BK, Ferreira Z, Robinson JD, Lathrop K, Nischal KK, Chikina M, Clark NL. 2017. Subterranean mammals show convergent regression in ocular genes and enhancers, along with adaptation to tunneling. *eLife.* **6**:e25884.
- Patton EE, Zon LI, Langenau DM. 2021. Zebrafish disease models in drug discovery: from preclinical modelling to clinical trials. *Nat Rev Drug Discov.* **20**(8):611–628.
- Percie du Sert N, Hurst V, Ahluwalia A, Alam S, Avey MT, Baker M, Browne WJ, Clark A, Cuthill IC, Dirnagl U, et al. 2020. The ARRIVE guidelines 2.0: updated guidelines for reporting animal research. *PLoS Biol.* **18**(7):e3000410.
- Price SA, Friedman ST, Wainwright PC. 2015. How predation shaped fish: the impact of fin spines on body form evolution across teleosts. *Proc Royal Soc B Biol Sci.* **282**(1819):20151428.
- Prudent X, Parra G, Schwede P, Roscito JG, Hiller M. 2016. Controlling for phylogenetic relatedness and evolutionary rates improves the discovery of associations between species’ phenotypic and genomic differences. *Mol Biol Evol.* **33**(8):2135–2150.
- Puttick M, Thomas G, Freckleton R, Clarke M, Ingram T, Orme D, Paradis E. 2019. MOTMOT: models of trait macroevolution on trees. R package version 2.1.3.
- Raney BJ, Dreszer TR, Barber GP, Clawson H, Fujita PA, Wang T, Nguyen N, Paten B, Zweig AS, Karolchik D, et al. 2014. Track data hubs enable visualization of user-defined genome-wide annotations on the UCSC genome browser. *Bioinformatics* **30**(7):1003–1005.
- Ratnasingham S, Hebert PDN. 2007. BOLD: the barcode of life data system (<http://www.barcodinglife.org>). *Mol Ecol Notes.* **7**(3):355–364.
- Reid K, Bell MA, Veeramah KR. 2021. Threespine stickleback: a model system for evolutionary genomics. *Annu Rev Genomics Hum Genet.* **22**(1):357–383.
- Roberts Kingman GA, Vyas DN, Jones FC, Brady SD, Chen HI, Reid K, Milhaven M, Bertino TS, Aguirre WE, Heins DC, et al. 2021. Predicting future from past: the genomic basis of recurrent and rapid stickleback evolution. *Sci Adv.* **7**(25):eabg5285.
- Roscito JG, Sameith K, Kirilenko BM, Hecker N, Winkler S, Dahl A, Rodrigues MT, Hiller M. 2022. Convergent and lineage-specific genomic differences in limb regulatory elements in limbless reptile lineages. *Cell Rep.* **38**(3):110280.
- Ryu S-W, Lee S-J, Park M-Y, Jun J, Jung Y-K, Kim E. 2003. Fas-associated factor 1, FAF1, is a member of FAS death-inducing signaling complex. *J Biol Chem.* **278**(26):24003–24010.
- Sackton TB, Grayson P, Cloutier A, Hu Z, Liu JS, Wheeler NE, Gardner PP, Clarke JA, Baker AJ, Clamp M, et al. 2019. Convergent regulatory evolution and loss of flight in paleognathous birds. *Science* **364**(6435):74–78.
- Sagai T, Masuya H, Tamura M, Shimizu K, Yada Y, Wakana S, Gondo Y, Noda T, Shiroishi T. 2004. Phylogenetic conservation of a limb-specific, *cis*-acting regulator of *Sonic hedgehog* (*Shh*). *Mamm Genome.* **15**(1):23–34.
- Schier AF, Talbot WS. 2005. Molecular genetics of axis formation in zebrafish. *Annu Rev Genet.* **39**(1):561–613.
- 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.
- Shapiro MD, Summers BR, Balabhadra S, Aldenhoven JT, Miller AL, Cunningham CB, Bell MA, Kingsley DM. 2009. The genetic architecture of skeletal convergence and sex determination in nine-spine sticklebacks. *Curr Biol.* **19**(13):1140–1145.
- Shikano T, Laine VN, Herczeg G, Vilkki J, Merilä J. 2013. Genetic architecture of parallel pelvic reduction in ninespine sticklebacks. *G3* **3**(10):1833–1842.
- Smit A, Hubley R, Green P. 2015. RepeatMasker Open-4.0.
- Smith SD, Pennell MW, Dunn CW, Edwards SV. 2020. Phylogenetics is the new genetics (for most of biodiversity). *Trends Ecol Evol.* **35**(5):415–425.
- Sowersby W, Eckerström-Liedholm S, Rowiński PK, Balogh J, Eiler S, Upstone JD, Gonzalez-Voyer A, Rogell B. 2022. The relative effects of pace of life-history and habitat characteristics on the evolution of sexual ornaments: a comparative assessment. *Evolution* **76**(1):114–127.
- Swarup H. 1958. Stages in the development of the stickleback *Gasterosteus aculeatus*. *Development* **6**(3):373–383.
- Tamura K, Stecher G, Kumar S. 2021. MEGA11: molecular evolutionary genetics analysis version 11. *Mol Biol Evol.* **38**(7):3022–3027.
- Tan G, Lenhard B. 2016. TFBSTools: an R/Bioconductor package for transcription factor binding site analysis. *Bioinformatics* **32**(10):1555–1556.
- Tanaka M. 2016. Fins into limbs: autopod acquisition and anterior elements reduction by modifying gene networks involving 5’ *Hox*, *Gli3*, and *Shh*. *Dev Biol.* **413**(1):1–7.
- Tange O. 2011. GNU parallel—the command-line power tool. *login: The USENIX Magazine.* **36**(1):42–47.
- The Gene Ontology Consortium. 2021. The Gene Ontology resource: enriching a Gold mine. *Nucleic Acids Res.* **49**(D1):D325–D334.
- Thomas GH, Freckleton RP. 2012. MOTMOT: models of trait macroevolution on trees. *Methods Ecol Evol.* **3**(1):145–151.

- Thompson AC, Capellini TD, Guenther CA, Chan YF, Infante CR, Menke DB, Kingsley DM. 2018. A novel enhancer near the *Pitx1* gene influences development and evolution of pelvic appendages in vertebrates. *eLife*. **7**:e38555.
- Turakhia Y, Chen HI, Marcovitz A, Bejerano G. 2020. A fully-automated method discovers loss of mouse-lethal and human-monogenic disease genes in 58 mammals. *Nucleic Acids Res*. **48**(16):e91.
- Tzung KW, Lalonde RL, Prummel KD, Mahabaleshwar H, Moran HR, Stundl J, Cass AN, Le Y, Lea R, Dorey K, et al. 2023. A median fin derived from the lateral plate mesoderm and the origin of paired fins. *Nature*. **618**:543–549.
- Van der Auwera GA, Carneiro MO, Hartl C, Poplin R, del Angel G, Levy-Moonshine A, Jordan T, Shakir K, Roazen D, Thibault J, et al. 2013. From FastQ data to high confidence variant calls: the Genome Analysis Toolkit best practices pipeline. *Curr Protoc Bioinform*. **43**:11.10.1–11.10.33.
- Virtanen P, Gommers R, Oliphant TE, Haberland M, Reddy T, Cournapeau D, Burovski E, Peterson P, Weckesser W, Bright J, et al. 2020. Scipy 1.0: fundamental algorithms for scientific computing in Python. *Nat Methods*. **17**(3):261–272.
- Walker M, Kimmel C. 2007. A two-color acid-free cartilage and bone stain for zebrafish larvae. *Biotechnic Histochem*. **82**(1):23–28.
- Wang Y, Zhang C, Wang N, Li Z, Heller R, Liu R, Zhao Y, Han J, Pan X, Zheng Z, et al. 2019. Genetic basis of ruminant headgear and rapid antler regeneration. *Science* **364**(6446):eaav6335.
- Westneat MW, Thorsen DH, Walker JA, Hale ME. 2004. Structure, function, and neural control of pectoral fins in fishes. *IEEE J Oceanic Eng*. **29**(3):674–683.
- Wucherpennig JI, Howes TR, Au JN, Au EH, Roberts Kingman GA, Brady SD, Herbert AL, Reimchen TE, Bell MA, Lowe CB, et al. 2022. Evolution of stickleback spines through independent cis-regulatory changes at *HOXD*. *Nat Ecol Evol*. **6**(10):1537–1552.
- Wucherpennig JI, Miller CT, Kingsley DM. 2019. Efficient CRISPR-Cas9 editing of major evolutionary loci in sticklebacks. *Evol Ecol Res*. **20**(1):107–132.
- Xie KT, Wang G, Thompson AC, Wucherpennig JI, Reimchen TE, MacColl ADC, Schluter D, Bell MA, Vasquez KM, Kingsley DM. 2019. DNA fragility in the parallel evolution of pelvic reduction in stickleback fish. *Science* **363**(6422):81–84.
- Yamanoue Y, Setiamarga DHE, Matsuura K. 2010. Pelvic fins in teleosts: structure, function and evolution. *J Fish Biol*. **77**(6):1173–1208.