Title
Genetic and Genomic Bases of Evolved Increases in Stickleback Dentition

Author
Hart, James Clinton

Publication Date
2018

Peer reviewed|Thesis/dissertation

Genetic and Genomic Bases of Evolved Increases in Stickleback Dentition

by

James Clinton Hart

A dissertation submitted in partial satisfaction of the

requirements for the degree of

Doctor of Philosophy

in

Molecular and Cell Biology

and the Designated Emphasis

in

Computational Biology

in the

Graduate Division

of the

University of California, Berkeley

Committee in charge:

Professor Craig T. Miller, Chair
Professor Michael B. Eisen
Professor Daniel S. Rokhsar
Professor Michael W. Nachman

Spring 2018

Genetic and Genomic Bases of Evolved Increases in Stickleback Dentition

Abstract

Genetic and Genomic Bases of Evolved Increases in Stickleback Dentition

by

James Clinton Hart

Doctor of Philosophy in Molecular and Cell Biology
and the Designated Emphasis in
Computational Biology

University of California, Berkeley

Professor Craig T. Miller, Chair

Evolution - the great tinkerer - has produced the astounding diversity of form within and between existing species. It is a fundamental goal of evolutionary biology to understand the origin of such diversity. What types of genes underlie evolved changes in morphology? Are certain types of mutations (notably changes within regulatory regions) more likely to be used to produce adaptive changes in form? When distinct populations evolve similar morphological changes, are the underlying genetic bases changes to the same genes, the same genetic pathways, or largely independent? Are changes in form modular, or are their concerted changes to multiple developmentally similar organs? The ever cheapening cost of sequencing, coupled the availability of high-quality reference genomes, allows high-throughput approaches to identifying the loci of evolution. The emergence of a robust genome engineering system, CRISPR/Cas9, allows for efficient and direct testing of a gene's phenotype. Combining both of these techniques with a model system with naturally evolved phenotypic variation, the threespine stickleback, allows for systems-level answers to the many evolutionary questions.

Chapter one outlines the field of evolutionary developmental biology. It proposes two alternative viewpoints for thinking about the evolution of form. The first is the view of the 'Modern Synthesis', linking Mendelian inheritance with Darwinian natural selection, which explains evolution as the change in allele frequencies over time. The second views evolution through the lens of deep homology, focusing on changes to developmental programs over time, even across related organs within the same animal. It then introduces key concepts within evolutionary and developmental biology, including *cis*-regulation of gene expression, and gene regulatory networks. It then provides examples of evolution reusing similar gene regulatory networks, including *Hox* genes, *Pax6* dependent eye initiation, and ectodermal placode development. Teeth use highly conserved signaling pathways, during both their initiation and replacement. Threespine sticklebacks *Gasterosteus aculeatus* have repeatedly adapted following a shift from marine to freshwater environments, with many independently

derived populations sharing common morphological traits, including a gain in tooth number. The following chapters investigate this gain in tooth number in multiple distinct populations of sticklebacks.

Chapter two describes the discovery and mapping of a spontaneous stickleback albino mutation, named *casper*. *casper* is a sex-linked recessive mutation that results in oculocutaneous albinism, defective swim bladders, and blood clotting defects. Bulked segregant mapping of *casper* mutants revealed a strong genetic signal on chromosome 19, the stickleback X chromosome, proximal to the gene *Hps5*. *casper* mutants had a unique insertion of a G in the 6$^{\text{th}}$ exon on *Hps5*. As mutants in the human orthologue of *Hps5* resulted in similar albino and blood clotting phenotypes, *Hps5* is a strong candidate underlying the *casper* phenotype. Further supporting this model, genome editing of *Hps5* phenocopied *casper*. Lastly, we show that *casper* is an excellent tool for visualizing the activity of fluorescent transgenes at late developmental stages due to the near-translucent nature of the mutant animals.

Chapter three details the fine mapping of a quantitative trail locus (QTL) on chromosome 21 controlling increases in tooth number in a Canadian freshwater stickleback population. Recombinant mapping reduced the QTL-containing region to an 884kb window. Repeated QTL mapping experiments showed the presence of this QTL on multiple, but not all, wild-derived chromosomes from the Canadian population. Comparative genome sequencing revealed the perfect correlation with genetic data of ten variants, spanning 4.4kb, all within the 4$^{\text{th}}$ intron of the gene *Bmp6*. Transgenic analysis of this intronic region uncovered its role as a robust tooth enhancer. TALEN induced mutations in *Bmp6* revealed required roles for the gene in stickleback tooth development. Finally, comparative RNA-seq between *Bmp6* wild-type and mutant dental tissue showed a loss of mouse hair stem cell genes in *Bmp6* mutant fish teeth, suggesting deep homology of the regeneration of these two organs.

Chapter four investigates the evolved changes in gene expression that accompany evolved increases in tooth number in two distinct freshwater populations. Independently derived stickleback populations from California and Canada have both evolved increases in tooth number, and previous work suggested that these populations used distinct genetic changes during their shared morphological changes. RNA-seq analysis of dental tissue from both freshwater populations compared to marine revealed a gain in critical regulators of tooth development in both freshwater populations. These evolved changes in gene expression can be partitioned in *cis* changes (mutations within regulatory elements of a gene) and *trans* changes (changes to the overall regulatory environment) using phased RNA-seq data from marine-freshwater F_1 hybrids. Many genes show evidence for stabilizing selection of expression levels, with *cis* and *trans* changes in opposing directions. Most evolved changes in gene expression are due to changes in the *trans* environment, and these *trans* changes are more likely to be shared among the high-toothed freshwater populations. Thus, Californian and Canadian sticklebacks have convergently evolved similar *trans* regulatory environments through distinct *cis* regulatory changes.

Chapter five identifies candidate genes underlying evolved tooth gain in multiple geographically distinct freshwater populations. Many populations of freshwater sticklebacks have evolved increases in both oral and pharyngeal tooth number. QTL mapping of this

evolved gain in pharyngeal tooth number revealed that a 438bp regulatory haplotype of *Bmp6* is associated with increased tooth number in five distinct Pacific Northwest populations, though not in the high-toothed California population. QTL mapping of evolved oral tooth gain in California fish reveals the surprisingly modular nature of evolved changes in dentition. Correlation analysis of gene expression data from 33 separate samples across multiple populations and genotypes revealed *Plod2* and *Pitx2* as dentally expressed candidate genes underlying evolved tooth gain. CRISPR/Cas9 genome editing of *Plod2* resulted in mutants displaying increases in pharyngeal but decreases in oral tooth number. Mutations in *Pitx2* are homozygous lethal and show a recessive near-complete loss of dentition across all tooth fields. The pleitropic effects of the coding mutations and the lack of evolved coding changes suggest that modular regulatory changes to *Plod2* and *Pitx2* underlie increases in tooth number.

Combined, these results make significant contributions to our understanding of the evolutionary genetics underlying an adaptive change in morphology. Modular *cis*-regulatory alleles appear to play critical roles during the evolution of increased tooth number. Some alleles, such as the regulatory haplotype of *Bmp6*, are repeatedly used by multiple independently derived freshwater populations, suggesting both that the haplotype is adaptive and that evolution is partially repeatable. The Californian specific use of *Plod2* and *Pitx2* shows that evolution is not entirely predictable, and that there are many ways to modify teeth. Additionally, the use of high-throughput expression assays and genome sequencing, combined with genome editing with CRISPR/Cas9, allowed for rapid identification and testing of candidate genes underlying evolved changes in morphology. Additional studies could use these approaches to further identify the loci of evolved changes in morphology.

To my parents, Jay Hart and Lauren Baker-Hart

Contents

List of Figures

List of Tables

Acknowledgments

First and foremost, I would like to thank my advisor, Craig Miller, for all of his support during my time in his lab. Craig has had a tremendous influence on my development as a scientist, not only in the minutia of conceiving, planning, and interpreting scientific experiments, but especially in clearly and concisely communicating my work. Craig was always able to find time to talk about my or others science throughout my time at Berkeley. I also thank Michael Eisen, my unofficial co-advisor, for scientific mentorship, and for the ability to present at his group meetings. I was constantly forced to think about my science from an outsider's perspective, and received countless pieces of useful advice from his lab. I also extend my appreciation to my other committee members Daniel Rokhsar and Michael Nachman for their helpful suggestions and questions raised during my committee meetings.

I would also like to extend my gratitude to Sorin Istrail, who first sparked my interest in computational biology and provided my first research opportunity. I would also like to thank Istrail lab members Ryan Tarpine, Derek Aguiar, and Timothy Johnstone for their mentoring, advice, support, and generally camaraderie. Thanks also to Richard Freiman, for teaching me the fundamentals of developmental and molecular biology.

I thank members of the Miller lab, both past and present, for creating and maintaining a collaborative, productive, and lively environment to do science in. Thanks to Andrew Glazer for many late nights spent discussing the finer points of figure making and data analysis. To Phillip Cleves for his constant mentorship and guidance, for his countless experimental ideas and feedback, and for spreading his boundless enthusiasm for his projects. To Nicholas Ellis for his developmental biology expertise, his patient and clear advice, and for our relaxed but highly productive collaborations. To Priscilla Erickson for teaching me everything I know about molecular biology, and for instilling in me that one can be a space cadet and still be a successful grad student, so long as one forces themselves to be highly organized. I am thankful to the current members of the Miller lab, and know they will keep the scientific torch moving forward. Thanks to Mark Stepaniak for tolerating me at my most stressed-out moments, and for many conversations about films of all genres. Thanks to Alyssa Borman, her scientific and organizational skills will continue to add to the lab. Thanks also to Tyler Square, for his excellent developmental images and high levels of scientific enthusiasm.

I have had the privilege of working with a large number of talented undergraduate researchers. Thanks to former undergraduates Siegen McKeller and Aluokika Shah - your hard work and dedication made you a pleasure to work with. Thanks to Mary Chen, your high levels of organization helped all the projects you worked on. Thanks especially to Nathan Cramer - I have enjoyed watching you grow into a careful, introspective, skeptical, smart, and independent scientist.

Throughout graduate school I have nearly continuously explored the beauty of the wilds of California with a number of fantastic people, notably Akshay Tambe, Maxwell Goodman, Amy Strom, Shion and John Lim, Laura Wetzel, Gabrielle Boisrame, Adam Coogan, and many others. Our constant adventures to the backcountry kept the majority of my sanity intact, and helped develop what I expect to be a life-long love of the outdoors.

I am extremely fortunate to have an incredibly supporting family through my whole life, and especially during graduate school. My dad, Jay Hart, has instilled in me a curiosity about the world and the forces that shape it. My mom, Lauren Baker-Hart, has given me all her encouragement and support, life advice and guidance, and taught me how to write. My brother, Russell Hart, has always been there to talk, and to help blow off steam over an online game. My yearly pilgrimages back to Boston to see the family have been a constant and reliable break from the daily stresses of Berkeley.

Lastly, I have to thank Diane Taylor. We first met the summer of my 4^{th} year of my PhD, and I can't imagine life without her. She has been my best friend, my favorite backpacking buddy, my main climbing partner, and my best travel companion. We've been on many adventures, from Maine to Chapel Hill, from the low deserts of Death Valley to the alpine lakes of the high Sierra, from Montreal to Iceland, and many places in between. I am so thankful for her constant love and support throughout graduate school, and am excited to be able to start the next stage of my life with her.

Chapter 1

Introduction

1.1 Diversity of animal form

"From so simple a beginning endless forms most beautiful and most wonderful have been, and are being, evolved" (Charles Darwin) [1]

Animals display an almost endless diversity of form. Forelimbs range from webbed swimming fins to clawed grasping hands to delicate feathered wings. Eyes vary from the compound eyes of insects to the telescopic eyes of birds to the camera eyes of cephalopods. Coloration spans the ghost white of cave dwelling albinos to the incredible hues of birds of paradise and tropical fish to the ever-changing pigmentations of chameleons and cuttlefish. As all of these organisms share a common ancestor, the diversity observed today must have an ancient evolutionary origin. This diversity raises the key question in evolutionary developmental biology - how do differences in animal form arise?

There are two different schools of thought that attempt to answer the question - "How does evolution produce the diversity of life we see today?" The first finds its origins in the Modern Synthesis [2], combining Darwinian natural selection [1] with Mendelian inheritance [3] and population genetics [4]. Mutations are inherited following Mendel's laws, and the resulting genetic diversity within a population can give rise to a corresponding phenotypic diversity. If a population is subject to natural selection, the more fit phenotypes survive, also ensuring survival of the underlying genotypes. These genotypes are again transmitted through Mendelian processes, and the cycle continues in subsequent generations. In this view, evolution is the change in allele frequencies over time.

An alternate, and not necessarily mutually exclusive, view addresses the origin of animal form through homology [5]. An organisms development gives rise to a wide variety of morphological traits. Many existing traits share a common evolutionary origin, whether looking at the same trait class in different animals [6,7], or even different traits within the same animal [8]. These similarities go beyond just surface-level morphology, with many of the same genes and genetic circuits active during development of these homologous structures.

These homologous structures continue to look similar in part because small changes to development are more tolerated, and more likely to be beneficial. Organisms are not clay balls, to be molded as natural selection sees fit; rather organisms are contingent collections of their evolutionary and developmental history, with some transitions of form no longer possible. In this view, the origin of animal form is highly dependent on modifications to the developmental program of the ancestral, homologous structure. Attempts to answer questions about the evolution of morphology must address both viewpoints, not only finding genetic changes that underlie changes in morphology, but also understanding the conserved developmental program that forms the structure in the first place.

1.2 Regulation of gene expression during development

"Spooky action at a distance" (Albert Einstein)

Precise control of the spatiotemporal expression patterns of thousands of genes is necessary for an organism's development. Gene expression can be regulated at many different stages. mRNA translation can be modulated by the presence of upstream open reading frames (uORFs) [9], the blocking of ribosome binding due to mRNA secondary structure [10], or the action of microRNAs (miRNAs) [11]. mRNA stability can be affected by *trans* acting RNA binding proteins or miRNAs triggering deadenylation or exonuclease cleavage, eventually resulting in transcript destruction [12]. mRNAs can be regulated by direct post-transcriptional modifications, such as the formation of N6-methyladenosine [13]. Co-transcriptional RNA splicing can result in different isoforms of mRNA[14]. The rate and cell type specificity of transcription itself can also be tightly controlled.

Transcriptional regulation is a critical part of the control of gene expression during development. Eukaryotic transcription is regulated by a combination of *trans*-acting transcription factors and the *cis*-acting sequences they bind to. Transcription of a gene is initiated at proximal promoter sequences, with several alternate promoters occasionally available [15,16]. Specific *cis* sequences in the promoter, such as the TATA-box or the B recognition element, are bound by general transcription factors such as the TATA-binding protein (TBP) and general transcription factor II B (TFIIB). General transcription factor binding, together with tissue-specific TBP-associated factors (TAFs), recruits RNA polymerase II, assembles the transcription pre-initiation complex, and eventually initiates transcription [17-19].

Distal *cis*-acting regulatory sequences, often termed enhancers, bind tissue-specific transcription factors. The DNA-bound transcription factors will then loop out the intervening sequence of DNA and interact with promoter-bound proteins, allowing enhancers to act distally from the gene or genes they regulate [20]. A single gene can be regulated by multiple enhancers, with one estimate that, in a given cell type, >50% of transcription start sites are regulated by more than one distal sequence, with up to as many as 20 interactions [21]. Enhancers are often associated with specific epigenetic modifications, especially the presence of acetylated lysine 27 of histone 3 (H3K27Ac) [22] or general open chromatin as assayed by

DNase I hypersensitivity (DNase-seq) [23] or transposase-accessible chromatin (ATAC-seq) [24]. Many enhancers are transcribed [25], either as longer, polyadenylated, unidirectional transcripts or short, bidirectional RNAs lacking polyadenylation. It is still unclear the exact role of this transcription, though there is increasing evidence that it plays a functional role in the regulation of nearby genes [26].

1.3 Regulatory developmental networks

"Cell differentiation is based almost certainly on the regulation of gene expression" (Eric Davidson) [27]

Gene expression is controlled by the given state the gene regulatory network (GRN) of the cell is in. A gene regulatory network includes all regulatory events (the edges), as well as the molecules (whether DNA, RNA, or protein) responsible for such events (the nodes). Gene regulatory networks seem to exhibit scale-free topology, where most nodes participate in only one or two interactions, and a small number of nodes (hubs) participate in many, with node connectivity degree following a power law [28]. As regulators are capable of regulating other regulators, including themselves, GRNs are full of cyclic feedback loops. Development and cellular differentiation can be thought of as changing the GRN of a given cell from one metastable state into another through the activation or repression of key regulators [29].

One of the best studied examples of a gene regulatory network is used during the formation of segments during the early development of Drosophila melanogaster. Initially, maternally deposited transcription factors set up the anterior-posterior axis of the embryo Bicoid and Hunchback at the anterior, and Caudal at the posterior. These maternal effect transcription factors freely diffuse in the syncytium of the early embryo, creating morphogen gradients. These gradients of transcription factors activate spatially restricted expression of gap genes, which encode other transcription factors, including *giant*, *huckebein*, *knirps*, *Krppel*, *tailless*, and the zygotically transcribed *hunchback*. These gap genes activate another class of genes encoding transcription factors the pair rule genes *even-skipped*, *odd-skipped*, *hairy*, *paired*, and *runt*. These pair rule genes are expressed in alternating parasegments, which later guide the formation of proper segments. Lastly, the orientation of each parasegment is specified by the segment polarity genes *engrailed*, *hedgehog*, *wingless*, *frizzled*, and *naked cuticle*. Each layer of regulation in this gene regulatory network involves the spatial control of transcription of transcription factors. These transcription factors are able to regulate the transcription of additional targets, often including themselves or their regulators. This creates a stable, robust network capable of repeatedly developing segments in a precise and predictable manner [30,31].

1.4 Evolutionary reuse of developmental networks

"Anything found to be true of E. coli must also be true of elephants, only more so" (Jacques Monod)

Evolution is a tinkerer [32] in that it is evolutionarily easier to modify an existing structure than it is to create a novel structure from scratch. The genetic code is highly conserved across all domains of life yet studied. Central metabolism is shared between eukaryotes and bacteria. Vertebrate wings have repeatedly evolved from arms, rather than alongside arms. Indeed, there has been little variation of phylum level body plan of animals since the early Cambrian period [33].

Many animals share *Hox* genes, first discovered in *Drosophila* as a cluster of genes which caused homeotic transformations when mutated [34]. Following their molecular characterization, it was found that all of these *Drosophila* homeotic genes possessed a similar DNA motif, termed the homeobox [35], which encodes the DNA-binding homeodomain of this class of proteins. Surprisingly, this homeobox appeared to be highly conserved to vertebrates as well, suggesting a conserved role for *Hox* genes during development [36]. Overexpressing mouse *Hox* genes in *Drosophila* resulted in a similar homeotic transformation as overexpressing *Drosophila* orthologs. Loss-of-function mutations in the mouse *HoxC8* gene resulted in homeotic phenotypes [37], and ectopic expression of mouse *HoxB6* in *Drosophila* results in homeotic phenotypes similar to ectopic expression of the fly ortholog [38]. Other key developmental regulators, such as the morphogen Hedgehog, originally discovered as a key regulator of *Drosophila* development [31], are conserved across many animal phyla [39]. Certain *cis*-regulatory sequences are deeply conserved, such as a *SoxB2* class enhancer that drives expression in the central nervous system in vertebrates and cnidarians [40].

Networks of key developmental regulators are often conserved across many animal phyla. The regulatory factors *eyeless* (*Pax6*), *atonal* (*Math5*), *eyes absent* (*Eya*), and *optix* (*Six3/6*) have been shown to be critical for eye development from organisms ranging from flies to mice [7,41,42]. Flies and mice lacking wild-type copies of any of these genes lack fully developed eyes. Ectopic expression of *eyeless* or *eyes absent* results in ectopic eye formation in *Drosophila* [43]. Surprisingly, expression of the eyeless mouse homologue, *Pax6*, was also able to induce ectopic compound eyes in flies, demonstrating the cross-phyla conservation of eye induction circuitry. Even in animals as distant as the jellyfish *Tripedalia cystophora*, *Pax* gene family members are involved in eye development, with the *PaxB* gene of *Tripedalia* also able to induce ectopic eyes in *Drosophila* [44]. Even though it has been estimated that eyes have independently evolved dozens of times [7], the core developmental regulators are conserved across vast evolutionary distances.

Eyes start development as part of a broader class of developmental structures called placodes. Epidermal appendages, such as hair, scales, sweat glands, and teeth, also all begin development as placodes, with the thickening of an epithelial layer overlying mesenchymal cells. These placodes are typically thought to be specified by a Turing-style reaction-diffusion mechanism [45], though there is little evidence of long-range diffusion of morphogens within

vertebrate tissues [46,47]. An alternative mechanism for placode formation is the mechanical forces stretching the epithelial layer of cells [48,49], which appears to be a conserved process [50]. Epithelial placodes of diverse developmental fates are marked by the expression of highly conserved regulatory molecules including *Bmp2/4*, *Fgf10*, *Eda*, and members of the Wnt family [51].

Following initial placode formation, the epithelial layer of cells begins proliferating, and either buds downward into the mesenchyme (in the case of hair, teeth, and mammary glands), or outwards (in the case of feathers) [51]. The identity of this epithelial growth is determined by signals coming from the underlying layer of mesenchymal cells. Transplantations of corneal epithelium covering dermal mesenchyme result in the ectopic formation of hair follicles, while corneal epithelium covering plantar dermis results in sweat glands [52]. This process of mesoderm-driven specification appears to be conserved to all ectodermal placodes [51].

1.5 Tooth Development

"If I only knew how many teeth and of what kind every animal had I should perhaps be able to work out a perfectly natural system for the arrangement of all quadrupeds" (Carl Linnaeus)

Teeth belong to a class of structures which start development as placodes, and have long been a model system for vertebrate organogenesis [51]. The epithelial layer then invaginates into the underlying mesenchyme during the bud stage, and eventually forms the start of a crown during the cap stage. The mesenchymal cells continue to condense during the bell stage, where the tooth begins to grow to its full size. The tooth becomes ossified due to the action of mesenchymal dentin producing osteoblasts, and epithelial enamel (enameloid in teleosts) producing ameloblasts [53]. This process involves a series of highly conserved signaling pathways notably the transforming growth factor β (TGF-β), bone morphogenetic protein (BMP), fibroblast growth factor (FGF), sonic hedgehog (SHH), and Wnt signaling pathways [53]. The most upstream signals in mouse tooth development seem to be the Wnt and BMP pathways, which form a feedback loop [54].

Though the mouse oral teeth are the most studied dental developmental system, much of tooth development is conserved across species and tooth types. In snakes, early tooth germs show expression of Wnt signaling components *Lef1*, *Axin2*, *Wnt10b*, and *Wnt6*, and inhibition of Wnt or SHH signaling results in developmental arrest [55]. Catshark tooth development shows similar conserved patterns of expression of Wnt components *β-catenin*, *Lef1*, as well as *Pitx2* and *Sox2* [56]. These conserved patterns of gene expression extend to pharyngeal teeth. In cichlids, oral and pharyngeal tooth number is correlated, and appear to have a similar genetic architecture[57]. Both cichlid pharyngeal and oral teeth share conserved expression patterns of *Bmp2*, *Bmp4*, *Dlx2*, *Pitx2*, *Runx2*, and *Shh* [58]. Zebrafish pharyngeal tooth development also shows conserved early expression patterns of *Dlx2a*, *Dlx2b*, and *Pitx2*, and later expression of *Fgf3* and *Fgf4*. Inhibition of FGF signaling arrests zebrafish pharyngeal

tooth development, similar to mouse oral teeth [59]. Thus, though tooth morphology may differ between organisms, the core developmental genetic circuitry appears deeply conserved.

Continuous tooth replacement (polyphyodonty) appears to be the ancestral jawed vertebrate condition [60]. However, mice are monophyodonts and do not replace their teeth, necessitating the need to study other polyphyodont systems. Elasmobranchs (sharks) replace their teeth using a conveyor-belt system where new teeth are fully formed by an odontogenic band before being functionally used [56,61]. Teleosts often undergo a one-for-one tooth replacement scheme, where a functional tooth is able to form a single replacement tooth, which eventually replaces the functional tooth in the tooth field [62]. In cichlids and zebrafish, the dental lamina proliferates and invaginates, forming a successional lamina capable of generating the next replacement tooth [63]. Unlike many tetrapods, geckos are polyphyodonts that exhibit a many-for-one mode of tooth replacement. Similar to the replacement of mammalian hair, gecko teeth harbor a niche of slow-cycling epithelial stem-like cells, marked by the gene expression of *Lgr5*, *Dkk3*, and *Igfbp5*, and these putative stem cells are potentially able to give rise to the new replacement tooth [64].

1.6 Radial evolution of sticklebacks

"Evolution is a process of constant branching and expansion" (Stephen Jay Gould)

The threespine stickleback (*Gasterosteus aculeatus*) is an emergent supermodel [65] for the study of the evolution and development of morphology. Ancestral marine populations of sticklebacks have colonized freshwater lakes and streams throughout the northern hemisphere, leading to a series of isolated populations independently adapting to a similar shift in environment. These evolutionary replicates, as well as a compact 460 Mb genome with a high-quality reference [66,67], large clutch size, and ease of transgenesis and genome editing [68] make for an excellent model system for evolutionary and developmental genetics.

Stickleback freshwater colonization has been thought to follow a source-sink model, where marine populations colonize isolated freshwater environments [69]. Consistent with this, marine stickleback genomes display little population structure, and even geographically diverse marine populations appear to share more genomic similarity than freshwater populations [70,71]. When colonizing a new freshwater environment, certain rare alleles will switch from neutral or deleterious to advantageous and will increase in frequency following natural selection. However, due to founder effects, each newly derived freshwater population will potentially carry a different set of adaptive alleles. Further complicating matters, alleles might be only advantageous in specific types of freshwater environments (e.g. lakes but not streams [72]).

Freshwater populations of sticklebacks nevertheless share many derived phenotypes and genotypes. The best studied example is the loss of armored lateral plates in most freshwater stickleback populations [69], due to a reuse of an ancient regulatory allele of the *Eda* gene [73,74], though this allele may be pleiotropic [75,76]. Genome-wide scans for signs of parallel

freshwater adaptation have revealed multiple regions that appear to be reused in diverse populations [66,77]. This collateral evolution [78] suggests a model where freshwater populations could act as reservoirs of freshwater adaptive alleles, which would percolate into the marine gene pool by occasional introgression. These alleles could then be reused during colonization of new freshwater environments, further increasing their migration rate into the marine gene pool [79]. Supporting this, the low-plated *Eda* allele was found at 0.2% frequency in marine populations [73], far above the frequency expected by mutation-selection balance, and more in line with migration-selection balance.

Not all shared evolved morphological changes in freshwater stickleback populations have a shared genetic basis. A recent study found a mix of potentially shared and population specific QTL controlling the evolved change in a wide array of phenotypes in three geographically proximal freshwater populations [80]. Two additional studies, and the fourth chapter of this thesis, revealed distinct genetic bases of evolved tooth gain in benthic freshwater fish from Paxton Lake, Canada and creek fish from Cerrito Creek, California [81,82]. These cases of convergent evolution reveal that evolution is able to access a wide variety of genetic changes to evolve an adaptive change in form.

1.7 References

1. Darwin C. The Origin of Species. P. F. Collier and Son; 1859.
2. Huxley J. Evolution, the Modern Synthesis. G. Allen and Unwin Limited; 1942.
3. Mendel G. Versuche ber Pflanzen-Hybriden. Verh Naturforsch Ver Brnn. 1865;4: 347.
4. Fisher RA. XV.The Correlation between Relatives on the Supposition of Mendelian Inheritance. Earth Environ Sci Trans R Soc Edinb. 1919;52: 399433.
doi:10.1017/S0080456800012163
5. Wagner GP. Homology, Genes, and Evolutionary Innovation. Princeton University Press; 2014.
6. Shubin N, Tabin C, Carroll S. Fossils, genes and the evolution of animal limbs. Nature. 1997;388: 639648. doi:10.1038/41710
7. Shubin N, Tabin C, Carroll S. Deep homology and the origins of evolutionary novelty. Nature. 2009;457: 818823. doi:10.1038/nature07891
8. Li A, Lai Y-C, Figueroa S, Yang T, Widelitz RB, Kobielak K, et al. Deciphering principles of morphogenesis from temporal and spatial patterns on the integument. Dev Dyn Off Publ Am Assoc Anat. 2015;244: 905920. doi:10.1002/dvdy.24281
9. Barbosa C, Peixeiro I, Romo L. Gene Expression Regulation by Upstream Open Reading Frames and Human Disease. PLOS Genet. 2013;9: e1003529.
doi:10.1371/journal.pgen.1003529
10. Jacob F, Perrin D, Sanchez C, Monod J. [Operon: a group of genes with the expression coordinated by an operator]. Comptes Rendus Hebd Seances Acad Sci. 1960;250: 17271729.
11. Sonenberg N, Hinnebusch AG. Regulation of Translation Initiation in Eukaryotes: Mechanisms and Biological Targets. Cell. 2009;136: 731745. doi:10.1016/j.cell.2009.01.042

12. Guhaniyogi J, Brewer G. Regulation of mRNA stability in mammalian cells. Gene. 2001;265: 1123. doi:10.1016/S0378-1119(01)00350-X
13. Fu Y, Dominissini D, Rechavi G, He C. Gene expression regulation mediated through reversible m6A RNA methylation. Nat Rev Genet. 2014;15: 293306. doi:10.1038/nrg3724
14. Baralle FE, Giudice J. Alternative splicing as a regulator of development and tissue identity. Nat Rev Mol Cell Biol. 2017;18: 437451. doi:10.1038/nrm.2017.27
15. Batut P, Dobin A, Plessy C, Carninci P, Gingeras TR. High-fidelity promoter profiling reveals widespread alternative promoter usage and transposon-driven developmental gene expression. Genome Res. 2013;23: 169180. doi:10.1101/gr.139618.112
16. Ayoubi TA, Ven WJVD. Regulation of gene expression by alternative promoters. FASEB J. 1996;10: 453460.
17. Louder RK, He Y, Lpez-Blanco JR, Fang J, Chacn P, Nogales E. Structure of promoter-bound TFIID and model of human pre-initiation complex assembly. Nature. 2016;531: 604609. doi:10.1038/nature17394
18. Dynlacht BD, Hoey T, Tjian R. Isolation of coactivators associated with the TATA-binding protein that mediate transcriptional activation. Cell. 1991;66: 563576.
19. Hiller MA, Lin T-Y, Wood C, Fuller MT. Developmental regulation of transcription by a tissue-specific TAF homolog. Genes Dev. 2001;15: 10211030. doi:10.1101/gad.869101
20. Schoenfelder S, Clay I, Fraser P. The transcriptional interactome: gene expression in 3D. Curr Opin Genet Dev. 2010;20: 127133. doi:10.1016/j.gde.2010.02.002
21. Sanyal A, Lajoie BR, Jain G, Dekker J. The long-range interaction landscape of gene promoters. Nature. 2012;489: 109113. doi:10.1038/nature11279
22. Creyghton MP, Cheng AW, Welstead GG, Kooistra T, Carey BW, Steine EJ, et al. Histone H3K27ac separates active from poised enhancers and predicts developmental state. Proc Natl Acad Sci. 2010; 201016071. doi:10.1073/pnas.1016071107
23. Thurman RE, Rynes E, Humbert R, Vierstra J, Maurano MT, Haugen E, et al. The accessible chromatin landscape of the human genome. Nature. 2012;489: 7582. doi:10.1038/nature11232
24. Buenrostro JD, Giresi PG, Zaba LC, Chang HY, Greenleaf WJ. Transposition of native chromatin for fast and sensitive epigenomic profiling of open chromatin, DNA-binding proteins and nucleosome position. Nat Methods. 2013;advance online publication. doi:10.1038/nmeth.2688
25. Kim T-K, Hemberg M, Gray JM, Costa AM, Bear DM, Wu J, et al. Widespread transcription at neuronal activity-regulated enhancers. Nature. 2010;465: 182187. doi:10.1038/nature09033
26. Lam MTY, Cho H, Lesch HP, Gosselin D, Heinz S, Tanaka-Oishi Y, et al. Rev-Erbs repress macrophage gene expression by inhibiting enhancer-directed transcription. Nature. 2013;498: 511515. doi:10.1038/nature12209
27. Davidson EH, Britten RJ. Note on the control of gene expression during development. J Theor Biol. 1971;32: 123130. doi:10.1016/0022-5193(71)90140-8
28. Barabsi A-L, Oltvai ZN. Network biology: understanding the cells functional organization. Nat Rev Genet. 2004;5: 101113. doi:10.1038/nrg1272

29. Britten RJ, Davidson EH. Gene regulation for higher cells: a theory. Science. 1969;165: 349357.
30. Rivera-Pomar R, Jckle H. From gradients to stripes in Drosophila embryogenesis: filling in the gaps. Trends Genet TIG. 1996;12: 478483.
31. Nsslein-Volhard C, Wieschaus E. Mutations affecting segment number and polarity in Drosophila. Nature. 1980;287: 795801.
32. Jacob F. Evolution and tinkering. Science. 1977;196: 11611166.
doi:10.1126/science.860134
33. Gould SJ. Wonderful Life: The Burgess Shale and the Nature of History. W. W. Norton and Company; 1989.
34. Lewis EB. A gene complex controlling segmentation in Drosophila. Nature. 1978;276: 565570. doi:10.1038/276565a0 35. McGinnis W, Levine MS, Hafen E, Kuroiwa A, Gehring WJ. A conserved DNA sequence in homoeotic genes of the Drosophila Antennapedia and bithorax complexes. Nature. 1984;308: 428433.
36. McGinnis W, Garber RL, Wirz J, Kuroiwa A, Gehring WJ. A homologous protein-coding sequence in Drosophila homeotic genes and its conservation in other metazoans. Cell. 1984;37: 403408.
37. Le Mouellic H, Lallemand Y, Brlet P. Homeosis in the mouse induced by a null mutation in the Hox-3.1 gene. Cell. 1992;69: 251264.
38. Malicki J, Schughart K, McGinnis W. Mouse Hox-2.2 specifies thoracic segmental identity in Drosophila embryos and larvae. Cell. 1990;63: 961967. doi:10.1016/0092-8674(90)90499-5
39. Riddle RD, Johnson RL, Laufer E, Tabin C. Sonic hedgehog mediates the polarizing activity of the ZPA. Cell. 1993;75: 14011416.
40. Royo JL, Maeso I, Irimia M, Gao F, Peter IS, Lopes CS, et al. Transphyletic conservation of developmental regulatory state in animal evolution. Proc Natl Acad Sci U S A. 2011;108: 1418614191. doi:10.1073/pnas.1109037108
41. Gehring WJ, Ikeo K. Pax 6: mastering eye morphogenesis and eye evolution. Trends Genet. 1999;15: 371377. doi:10.1016/S0168-9525(99)01776-X
42. Gehring WJ. New perspectives on eye development and the evolution of eyes and photoreceptors. J Hered. 2005;96: 171184. doi:10.1093/jhered/esi027
43. Halder G, Callaerts P, Gehring WJ. Induction of ectopic eyes by targeted expression of the eyeless gene in Drosophila. Science. 1995;267: 17881792.
44. Kozmik Z, Daube M, Frei E, Norman B, Kos L, Dishaw LJ, et al. Role of Pax Genes in Eye Evolution. Dev Cell. 2003;5: 773785. doi:10.1016/S1534-5807(03)00325-3
45. Walton KD, Whidden M, Kolterud , Shoffner SK, Czerwinski MJ, Kushwaha J, et al. Villification in the mouse: Bmp signals control intestinal villus patterning. Development. 2016;143: 427436. doi:10.1242/dev.130112
46. Alexandre C, Baena-Lopez A, Vincent J-P. Patterning and growth control by membrane-tethered Wingless. Nature. 2014;505: 180185. doi:10.1038/nature12879
47. van Boxtel AL, Chesebro JE, Heliot C, Ramel M-C, Stone RK, Hill CS. A Temporal Window for Signal Activation Dictates the Dimensions of a Nodal Signaling Domain. Dev Cell. 2015;35: 175185. doi:10.1016/j.devcel.2015.09.014

48. Shyer AE, Tallinen T, Nerurkar NL, Wei Z, Gil ES, Kaplan DL, et al. Villification: How the Gut Gets Its Villi. Science. 2013;342: 212218. doi:10.1126/science.1238842
49. Shyer AE, Huycke TR, Lee C, Mahadevan L, Tabin CJ. Bending gradients: how the intestinal stem cell gets its home. Cell. 2015;161: 569580. doi:10.1016/j.cell.2015.03.041
50. Shyer AE, Rodrigues AR, Schroeder GG, Kassianidou E, Kumar S, Harland RM. Emergent cellular self-organization and mechanosensation initiate follicle pattern in the avian skin. Science. 2017; eaai7868. doi:10.1126/science.aai7868
51. Pispa J, Thesleff I. Mechanisms of ectodermal organogenesis. Dev Biol. 2003;262: 195205.
52. Ferraris C, Chevalier G, Favier B, Jahoda CA, Dhouailly D. Adult corneal epithelium basal cells possess the capacity to activate epidermal, pilosebaceous and sweat gland genetic programs in response to embryonic dermal stimuli. Dev Camb Engl. 2000;127: 54875495.
53. Jernvall J, Thesleff I. Tooth shape formation and tooth renewal: evolving with the same signals. Development. 2012;139: 34873497. doi:10.1242/dev.085084
54. OConnell DJ, Ho JWK, Mammoto T, Turbe-Doan A, OConnell JT, Haseley PS, et al. A WNT-BMP feedback circuit controls intertissue signaling dynamics in tooth organogenesis. Sci Signal. 2012;5: ra4. doi:10.1126/scisignal.2002414
55. Handrigan GR, Richman JM. A network of Wnt, hedgehog and BMP signaling pathways regulates tooth replacement in snakes. Dev Biol. 2010;348: 130141. doi:10.1016/j.ydbio.2010.09.003
56. Rasch LJ, Martin KJ, Cooper RL, Metscher BD, Underwood CJ, Fraser GJ. An ancient dental gene set governs development and continuous regeneration of teeth in sharks. Dev Biol. 2016;415: 347370. doi:10.1016/j.ydbio.2016.01.038
57. Hulsey CD, Machado-Schiaffino G, Keicher L, Ellis-Soto D, Henning F, Meyer A. The Integrated Genomic Architecture and Evolution of Dental Divergence in East African Cichlid Fishes (Haplochromis chilotes x H. nyererei). G3 GenesGenomesGenetics. 2017;7: 31953202. doi:10.1534/g3.117.300083
58. Fraser GJ, Hulsey CD, Bloomquist RF, Uyesugi K, Manley NR, Streelman JT. An ancient gene network is co-opted for teeth on old and new jaws. PLoS Biol. 2009;7: e1000031. doi:10.1371/journal.pbio.1000031
59. Jackman WR, Draper BW, Stock DW. Fgf signaling is required for zebrafish tooth development. Dev Biol. 2004;274: 139157. doi:10.1016/j.ydbio.2004.07.003
60. Donoghue PCJ, Rcklin M. The ins and outs of the evolutionary origin of teeth. Evol Dev. 2016;18: 1930. doi:10.1111/ede.12099
61. Martin KJ, Rasch LJ, Cooper RL, Metscher BD, Johanson Z, Fraser GJ. Sox2+ progenitors in sharks link taste development with the evolution of regenerative teeth from denticles. Proc Natl Acad Sci U S A. 2016;113: 1476914774. doi:10.1073/pnas.1612354113
62. Tucker AS, Fraser GJ. Evolution and developmental diversity of tooth regeneration. Semin Cell Dev Biol. doi:10.1016/j.semcdb.2013.12.013
63. Fraser GJ, Bloomquist RF, Streelman JT. Common developmental pathways link tooth shape to regeneration. Dev Biol. 2013;377: 399414. doi:10.1016/j.ydbio.2013.02.007
64. Handrigan GR, Leung KJ, Richman JM. Identification of putative dental epithelial

stem cells in a lizard with life-long tooth replacement. Development. 2010;137: 35453549. doi:10.1242/dev.052415
65. Gibson G. The Synthesis and Evolution of a Supermodel. Science. 2005;307: 18901891. doi:10.1126/science.1109835
66. Jones FC, Grabherr MG, Chan YF, Russell P, Mauceli E, Johnson J, et al. The genomic basis of adaptive evolution in threespine sticklebacks. Nature. 2012;484: 5561. doi:10.1038/nature10944
67. Glazer AM, Killingbeck EE, Mitros T, Rokhsar DS, Miller CT. Genome Assembly Improvement and Mapping Convergently Evolved Skeletal Traits in Sticklebacks with Genotyping by Sequencing. G3 Genes Genomes Genet. 2015; g3.115.017905. doi:10.1534/g3.115.017905
68. Erickson PA, Ellis NA, Miller CT. Microinjection for Transgenesis and Genome Editing in Threespine Sticklebacks. JoVE J Vis Exp. 2016; e54055e54055. doi:10.3791/54055
69. Bell MA, Foster SA. The evolutionary biology of the threespine stickleback. Oxford University Press; 1994.
70. Catchen J, Bassham S, Wilson T, Currey M, OBrien C, Yeates Q, et al. The population structure and recent colonization history of Oregon threespine stickleback determined using RAD-seq. Mol Ecol. 2013;22: 28642883. doi:10.1111/mec.12330
71. Hohenlohe PA, Bassham S, Currey M, Cresko WA. Extensive linkage disequilibrium and parallel adaptive divergence across threespine stickleback genomes. Philos Trans R Soc B Biol Sci. 2012;367: 395408. doi:10.1098/rstb.2011.0245
72. Kueng B, Berner D, Moser D, Roesti M. The genomics of ecological vicariance in threespine stickleback fish. Nat Commun. 2015;6: 8767. doi:10.1038/ncomms9767
73. Colosimo PF, Hosemann KE, Balabhadra S, Villarreal G, Dickson M, Grimwood J, et al. Widespread parallel evolution in sticklebacks by repeated fixation of ectodysplasin alleles. Science. 2005;307: 19281933. doi:10.1126/science.1107239
74. OBrown NM, Summers BR, Jones FC, Brady SD, Kingsley DM. A recurrent regulatory change underlying altered expression and Wnt response of the stickleback armor plates gene EDA. eLife. 2015;4: e05290. doi:10.7554/eLife.05290
75. MacColl ADC, Bradley JE, Robertson S. Eda haplotypes in three-spined stickleback are associated with variation in immune gene expression. Sci Rep. 2017;7: 42677. doi:10.1038/srep42677
76. Greenwood AK, Mills MG, Wark AR, Archambeault SL, Peichel CL. Evolution of Schooling Behavior in Threespine Sticklebacks Is Shaped by the Eda Gene. Genetics. 2016;203: 677681. doi:10.1534/genetics.116.188342
77. Terekhanova NV, Logacheva MD, Penin AA, Neretina TV, Barmintseva AE, Bazykin GA, et al. Fast Evolution from Precast Bricks: Genomics of Young Freshwater Populations of Threespine Stickleback Gasterosteus aculeatus. PLOS Genet. 2014;10: e1004696. doi:10.1371/journal.pgen.1004696
78. Stern DL. The genetic causes of convergent evolution. Nat Rev Genet. 2013;14: 751764. doi:10.1038/nrg3483
79. Bassham S, Catchen J, Lescak E, Hippel FA von, Cresko WA. Sweeping genomic re-

modeling through repeated selection of alternatively adapted haplotypes occurs in the first decades after marine stickleback colonize new freshwater ponds. bioRxiv. 2017; 191627. doi:10.1101/191627

80. Erickson PA, Glazer AM, Killingbeck EE, Agoglia RM, Baek J, Carsanaro SM, et al. Partially repeatable genetic basis of benthic adaptation in threespine sticklebacks: repeatable evolution in benthic sticklebacks. Evolution. 2016;70: 887902. doi:10.1111/evo.12897

81. Cleves PA, Ellis NA, Jimenez MT, Nunez SM, Schluter D, Kingsley DM, et al. Evolved tooth gain in sticklebacks is associated with a cis-regulatory allele of Bmp6. Proc Natl Acad Sci. 2014;111: 1391213917. doi:10.1073/pnas.1407567111

82. Ellis NA, Glazer AM, Donde NN, Cleves PA, Agoglia RM, Miller CT. Distinct developmental genetic mechanisms underlie convergently evolved tooth gain in sticklebacks. Development. 2015;142: 24422451. doi:10.1242/dev.124248

Chapter 2

Sequence-based mapping and genome editing reveal mutations in stickleback *Hps5* cause albinism and the *casper* phenotype

The following chapter was originally published as an article:
G3 (Bethesda). 2017 Sep 7;7(9):3123-3131

James C. Hart and Craig T. Miller
Department of Molecular and Cell Biology, University of California-Berkeley, Berkeley CA, 94720, USA

2.1 Abstract

Here we present and characterize the spontaneous X-linked recessive mutation *casper*, which causes oculocutaneous albinism in threespine sticklebacks (*Gasterosteus aculeatus*). In humans, Hermansky-Pudlak syndrome results in pigmentation defects due to disrupted formation of the melanin-containing lysosomal-related organelle (LRO), the melanosome. *casper* mutants display not only reduced pigmentation of melanosomes in melanophores, but also reductions in the iridescent silver color from iridophores, while the yellow pigmentation from xanthophores appears unaffected. We mapped *casper* using high-throughput sequencing of genomic DNA from bulked *casper* mutants to a region of the stickleback X chromosome (chromosome 19) near the stickleback ortholog of *Hermansky-Pudlak syndrome 5* (*Hps5*). *casper* mutants have an insertion of a single nucleotide in the 6th exon of *Hsp5*, predicted to generate an early frameshift. Genome editing using CRISPR/Cas9 induced lesions in *Hsp5* and phenocopied the *casper* mutation. Injecting single or paired *Hps5* guide RNAs revealed higher incidences of genomic deletions from paired guide RNAs compared to single gRNAs. Stickleback *Hps5* provides a genetic system where a hemizygous locus in XY males and a diploid locus in XX females can be used to generate an easily scored visible phenotype, facilitating quantitative studies of different genome editing approaches. Lastly, we show the ability to better visualize patterns of fluorescent transgenic reporters in *Hps5* mutant fish. Thus, *Hps5* mutations present an opportunity to study pigmented LROs in the emerging stickleback model system, as well as a tool to aid in assaying genome editing and visualizing enhancer activity in transgenic fish.

2.2 Introduction

The combination of new genome editing methods and next-generation sequencing technologies has ushered in a new era in genetics. High throughput DNA sequencing enables rapid forward genetic mapping of Mendelian (Schneeberger et al. 2009; Cuperus et al. 2010; Zuryn et al. 2010; Doitsidou et al. 2010; Bowen et al. 2012; Obholzer et al. 2012) and quantitative (Glazer et al. 2014, 2015; Jamann et al. 2015) loci. The remarkably high efficiency of the CRISPR/Cas9 system across diverse species (Jinek et al. 2013; Hwang et al. 2013; Friedland et al. 2013; Guo et al. 2014; Square et al. 2015; Martin et al. 2016) now allows for functional reverse genetic analysis in any species competent for delivery of genome editing reagents. However, the optimization of genome editing protocols is still in its infancy, with the efficiency of DNA double-strand break formation and repair still requiring characterization in many species. Particularly powerful loci for this characterization are those that mutate to cause an obvious viable visible phenotype, such as loci that affect pigmentation (Dahlem et al. 2012; Guo et al. 2014; Irion et al. 2014; Square et al. 2015; Hoshijima et al. 2016; Burger et al. 2016).

The diversity of vertebrate skin pigmentation is due to an interplay between four main groups of pigment-containing cells known as chromatophores (Fujii 2000; Kelsh 2004; Mills

and Patterson 2009; Irion et al. 2016). Chromatophores originate from the neural crest, which migrate and then differentiate into pigment cell types during embryonic development (Fujii 2000; Kelsh et al. 2009). Melanophores possess black or dark brown melanin-containing organelles, melanosomes, which are also present in the retinal pigment epithelium (RPE) (Marks and Seabra 2001; Wasmeier et al. 2008). Iridophores appear iridescent and typically silver due to the presence of light-reflective guanine-containing platelets, the size and orientation of which determine the reflective color (Cooper et al. 1990; Oshima 2001). Xanthophores appear yellow-red due to the presence of pteridine within pterinosomes (Ziegler 2003). Erythrophores contain vesicles with red carotenoids obtained through the diet, and are a critical mating cue in sticklebacks (Milinski and Bakker 1990; Wedekind et al. 1998; Fujii 2000). Pigment cells in insects also possess melanin, ommochrome, or drosopterin containing organelles (Shoup 1966).

Pigment containing organelles belong to a larger class of cellular organelles, the lysosome-related organelles (LROs) (DellAngelica et al. 2000; DellAngelica 2004). LROs encompass a wide variety of organelles in different cellular contexts, including platelet granules, basophil granules, neutrophil azurophil granules, major histocompatibility complex class II compartments, lamellar bodies, osteoclast granules, and lytic granules (DellAngelica et al. 2000; Marks and Seabra 2001; DellAngelica 2004). The best studied LRO, the melanosome, has a well characterized biogenesis. Starting at stage I, premelanosomes already have internal vesicles and interluminal fibers that become parallel and organized during stage II, darkening during stage III until they are obscured by melanin in fully formed stage IV melanosomes (Marks and Seabra 2001). Patients with melanosome maturation defects often exhibit defects in other LROs, revealing that different LROs share similar biogenesis pathways (Marks and Seabra 2001).

In a wide range of vertebrates and invertebrates, a highly conserved set of genes including members of biogenesis of lysosome-related organelles complexes (BLOC) 1,2, and 3 (Helip-Wooley et al. 2007; Wei et al. 2013) regulate the formation and maturation of LROs, including pigment producing LROs, and are required for wild-type pigmentation (Helip-Wooley et al. 2007; Wei et al. 2013). In humans, mutations in the BLOC-2 complex member *Hps5* result in Hermansky-Pudlak Syndrome type 5, which is characterized by oculocutaneous albinism, and bleeding diathesis (Huizing et al. 2004). Mutations in orthologs of *Hps5* result in oculocutaneous albinism in a wide variety of model organisms, including the *ruby-eyed 2* phenotype in mice (Zhang et al. 2003), the *snow white* phenotype in zebrafish (Daly et al. 2013), the *pink* phenotype in Drosophila (Falcn-Prez et al. 2007; Syrzycka et al. 2007), and the *translucent* phenotype in silkworms (Fujii et al. 2012). In vertebrates, *Hps5* is required for the maturation of type I to type II melanosomes (Nguyen et al. 2002), and also binds to and stabilizes other HPS proteins including HPS3 and HPS6 (Daly et al. 2013). In addition to conserved roles in melanosome maturation, *Hps5* and *Hps6* are required for iridophore development in zebrafish and Xenopus, respectively (Daly et al. 2013; Nakayama et al. 2016).

In zebrafish, the *snow white* mutant phenotype, oculocutaneous albinism, was shown to result from an I76N point mutation in the WD40 domain of *Hps5* (Daly et al. 2013). This

mutation results in relatively mild reductions in RPE and melanophore melanization due to fewer and smaller melanosomes, and loss of iridophores. *snow white* mutants display early larval lethality, representing the only lethal *Hps5* allele reported, as well as the only mutation within the N-terminal WD40 domain (Daly et al. 2013).

Here we report the discovery and characterization of a spontaneous threespine stickleback X-linked recessive albino mutation *casper*. A mapping-by-sequencing approach revealed *casper* to be tightly linked to the stickleback ortholog of *Hps5*. *casper* mutants have a one base pair insertion resulting in a predicted frame-shift mutation in *Hps5* that results in an early truncation of the highly conserved protein product. Inducing mutations in *Hps5* using the CRISPR/Cas9 system phenocopied the *casper* mutation. Lastly, we show the usefulness of *casper* embryos as a tool to visualize fluorescent reporters in adult fish. Together these data provide a new locus in an emergent genetic supermodel (Gibson 2005) to facilitate studies of genome editing, transgene expression, and pigmentation biology.

2.3 Materials and Methods

Animal Husbandry and Imaging

Fish were raised in brackish water (3.5g/L Instant Ocean salt, 0.217mL/L 10% sodium bicarbonate) at 18C in eight hours of light per day in 110L aquaria. Fry with standard length (SL) less than 10 mm were fed a diet of live *Artemia*, with frozen *Daphnia* added as fish reached 10 mm SL. Adults with SL greater than 20 mm were fed a combination of frozen bloodworms and *Mysis* shrimp. To map the *casper* mutation, six crosses were generated by crossing four different marine males to heterozygous *casper* mutant females (Table 2.1). Experiments were approved by the Institutional Animal Care and Use Committees of the University of California-Berkeley (protocol AUP-2015-01-7117). Embryos were visualized using Montage z-stacks on a Leica M165 FC dissecting microscope, using a GFP2 filter to visualize xanthophores, or on a Keyence VHX-5000 microscope. Adult fish were imaged using a Cannon Powershot S95 digital camera.

Male Population	Male Number	casper Embryos	Wild-Type Embryos
RABS	1	19 (23%)	63 (77%)
RABS	2	16 (25%)	47 (75%)
RABS	3	15 (38%)	24 (62%)
RABS	3	22 (23%)	72 (77%)
LITC	4	16 (21%)	61 (79%)
LITC	4	15 (23%)	63 (77%)

Table 2.1: **Outcrossing female *casper* carriers yields *casper* animals.** Each row lists a single cross between a male and a *casper* carrier female. "Male population" describes the stickleback population of the male parent of the cross (RABS = Rabbit Slough, Alaska, LITC = Little Campbell River, British Columbia). "Male number is a unique identifier of a single male. Male 3 and 4 sired two different crosses. "*casper* embryos lists the number of embryos with a textitcasper phenotype in the clutch, and "wild-type embryos lists the number of wild-type embryos in the clutch.

DNA Purification and Sequencing

For sequencing and genotyping, DNA was extracted from caudal fin tissue from the original *casper* male, the F_0 female he was crossed to, as well as 47 individual *casper* mutant F_2 embryos. Caudal fin tissue or embryos were digested for 12 hours at 55C in 600 µl of tail digestion buffer (10mM Tris pH 8.0, 100mM NaCl, 10mM EDTA, 0.05% SDS, 2.5 µl Proteinase K (Ambion AM2546)).

DNA from whole *casper* mutant F_2 embryos was diluted to 10 ng/µl for each fish and pooled. Barcoded Nextera libraries (Illumina FC-121-1031) were created from 50 ng of genomic DNA from the pooled embryos, as well as 50 ng of genomic DNA from the original *casper* male and F_0 female mate following the manufacturers instructions. Quality was assayed on an Agilent bioanalyzer, and the resulting libraries were sequenced on a single lane of an Illumina HiSeq4000 to generate 100 bp single-end reads.

Bulked Segregant Analysis

Reads were mapped to a revised assembly of the stickleback genome (Jones et al. 2012; Glazer et al. 2015) using bowtie2 (Langmead and Salzberg 2012) with parameters '-q –sensitive' (Table 2.2). Resulting SAM files were converted to BAM files and sorted using Samtools version 0.1.18 (Li et al. 2009). Read groups were added, CIGAR strings fixed, mate pair information was fixed and PCR duplicates were removed using picard tools v 1.51 (http://broadinstitute.github.io/picard). The Genome Analysis Tool Kit (McKenna et al. 2010; DePristo et al. 2011; Van der Auwera et al. 2013) (GATK)'s (v3.2-2) IndelRealigner (parameter: '-LOD 0.4'), BaseRecalibrator, and PrintReads were used to finalize preprocessing of BAM files. Finally, Unified Genotyper was used to call variants, with parameters

'–genotype_likelihoods_model BOTH -stand_call_conf 50.' The resulting VCF file was filtered for variants which had qual score >40, had more than five reads covering the variant, and were not found in the F_0 female mated to the original *casper* male. Analysis was performed using a custom ipython notebook. Briefly, we computed the proportion of each variant that matched the *casper* allele, and results were smoothed by plotting the proportion of *casper* reads within a 50 variant sliding window, advancing five variants at a time. As we found *casper* to be X-linked, we reasoned that the hemizygous XY *casper* animals should all be identical in sequence around the causative locus. We computed, as an additional measure, the proportion of variants within the window with reads matching more than one allele, which is expected to be 0 at the causative locus.

Sample	Total Reads	Mapped Reads	Final Reads	Expected Coverage
Original *casper*	122762025	103307721	67090400	14.5x
Wild-type RABS	117041920	96661604	63849076	13.8x
Bulked Segregants	78373157	67075791	48780347	10.5x

Table 2.2: **Mapping *casper* by bulked segregant analysis.** "Total reads lists the reads matching the sample barcode. "Mapped reads indicates the number of reads successfully mapped to a revised assembly of the stickleback reference genome (Jones et al 2012, Glazer et al. 2015). "Final reads indicates the number of reads that passed quality filters including removal of PCR duplicates. "Expected Coverage indicates the expected fold-coverage of each base in the 463 Mb assembled genome (Jones et al 2012) given the number of Final Reads.

PCR, Cloning, and Sanger Sequencing Validation

PCR primers (Table 2.3) were designed using Primer3 (Rozen and Skaletsky 2000) and ordered from IDT. PCR was performed using Phusion DNA polymerase (Macro lab, UC-Berkeley) and Phusion Buffer (NEB B0518S). To obtain sequence of single clones, PCR products were purified using a Qiagen PCR purification kit, and digested with *Xho*I (NEB R0146L) and *Xba*I (NEB R0145L) in cutsmart buffer for 1 hour at 37C. Digested products were ligated into a pBluescript II SK+ vector cut with *Xba*I and *Xho*I, transformed and plated onto LB agar plates with ampicillin, IPTG, and X-gal. White colonies were picked and used as input to PCR as described above, adding a 5 minute incubation at 95C before thermocycling. Resulting reactions were purified using a Qiagen PCR purification kit. Purified PCR product (20 ng) was Sanger sequenced by the UC Berkeley DNA sequencing facility and results visualized using abiview.

Name	Gene	Forward sequence	Reverse sequence	Restriction Cut Sites
casper genotyping	*Hsp5*	CCAGTGGCTGTTT GAACTGA	GGCCACGGTGTT TTTCATTA	NA
gRNA validation	*Hsp5*	CTTTGGCACTGG GTCGTAAT	GCGAGTACGGAG CATGTTTT	NA
gRNA validation cloning	*Hsp5*	CGGCctcgagCTTTG GCACTGGGTCGT AAT	CGGCtctagaGCGAG TACGGAGCATGT TTT	*Xho*I, *XbaI*
Hps5 guide1 template	NA	GCGTAATACGACT CACTATAGGGAG AGGAGGACAGGC CAGGTTTTAGAG CTAGAAATAGC	AAAGCACCGACTC GGTGCCACTTTTT CAAGTTGATAAC GGACTAGCCTTAT TTTAACTTGCTAT TTCTAGCTCTAAA AC	NA
Hps5 guide2 template	*Hsp5*	GCGTAATACGACT CACTATAGGTCG GGCGGAGTACAA CAGGTTTTAGAG CTAGAAATAGC	AAAGCACCGACTC GGTGCCACTTTTT CAAGTTGATAAC GGACTAGCCTTAT TTTAACTTGCTAT TTCTAGCTCTAAA AC	NA
Template amplification	NA	GCGTAATACGACT CACTATAG	AAAGCACCGACTC GGTGCCAC	NA
Sex determination	NA	CATATTGCTGCTT GTGTGGAAG	GATCCTCCTCGTT CCTACAG	NA

Table 2.3: **Sequences of primers used for PCR and creation of sgRNA templates.** "Gene indicates the corresponding gene in the stickleback genome the primer sequences are derived from, "Forward and reverse sequence give the sequence of the primers, and "Restriction cut sites indicates the presence of cut sites for the given enzyme on the 5′ end of the primers. The "Sex determination primers amplify fragment sizes of 186 bp and 229 bp from the X and Y chromosomes, respectively (Glazer et al., 2014).

Genome Editing of *Hps5*

Genome editing reagents were designed as previously described (Talbot and Amacher 2014). Briefly, pCS2-nCas9n (Addgene 7929) was linearized following digestion with *Not*I. Lin-

earized plasmid (600 ng) was used as input to the mMessage SP6 kit, following the manufacturers instructions. mRNA quality was verified by running 0.5 µL of the reaction in 0.1%SDS on a Tris acetic acid EDTA (TAE) gel.

Guide RNAs (gRNAs) were designed using ZiFiT (Sander et al. 2010). DNA oligos were ordered from IDT, and gRNA templates were created with T7 promoters by PCR using Phusion polymerase. Resulting PCR products were gel purified, and 100 ng of the resulting elution was used as input to the MAXIscript T7 kit (Ambion), and guide RNA quality verified by running 0.5 µl of the reaction in a 50% formamide buffer on a TAE gel. Resulting gRNAs were precipitated using lithium chloride, incubated at -80C following addition of 75 µl 100% ethanol, and centrifuged at 15000 rpm for 1 hour at 4C. Following a wash with 200 µL of 75% ethanol and an additional centrifugation at 15000 rpm for 10 minutes at 4C, RNAs were resuspended in 20 µl of DEPC treated water.

Stickleback embryos at the one cell stage were microinjected as described (Erickson et al. 2016) with some modifications. First, the concentration of gRNA in the 0.2M KCl injection mixture was increased to 50 ng/µl. The Cas9 mRNA concentration was also doubled to 160ng/µl, with 0.025% phenol red used as a tracking dye. Embryos were scored for casper-like phenotypes at four days post fertilization (dpf). Embryos were scored by eye, with embryos with embryos with any sign of mosaic albinism in their retinal pigment epithelium (RPE) classified as mosaic, with embryos with 75% albino RPE classified as severe. DNA from uninjected, wild-type injected, mosaic injected, and *casper*-like injected embryos was purified at 4-6dpf, and deletions in *Hps5* were validated using Sanger sequencing of PCR products as described above.

2.4 Results

casper mutants display severely reduced pigmentation early during embryogenesis

We discovered a single spontaneous mutant male stickleback displaying severe pigmentation defects. We named this mutation *casper*, and recovered the mutation in subsequent generations (see below). *casper* mutants display oculocutaneous hypopigmentation in unhatched embryos when pigment first becomes apparent, becoming readily apparent by seven days post fertilization (dpf) (Figure 2.1). Mutants appear fully viable and fertile (see below). *casper* mutants display severely reduced melanization of their retinal pigment epithelium (RPE), the most obvious visible phenotype (Figure 2.1A-D). Additionally, sexually mature *casper* males displayed severely reduced pigmentation in their testes relative to their wild-type siblings (Figure 2.2). Chromatophores are differentially affected in *casper* mutants. Beginning at the time of their first appearance (four dpf), melanophores in *casper* mutants are present, but display severe reductions in melanization relative to their wild-type siblings (Figure 2.1E-H). The silver pigmentation from iridophores appears absent from older *casper* fish (Figure 2.1I,J). The red erythrophores, which contain diet-supplied carotenoids

(Wedekind et al. 1998), were never observed in the throats of sexually mature casper mutant males. However, the yellow xanthophores (autofluorescent in sticklebacks) appear unaffected by the *casper* mutation in 10 dpf *casper* mutants (Figure 2.3).

casper mutants appear to initially inflate their swim bladder (Figure 2.1E-H). Older mutants display variably penetrant minor difficulties in maintaining their position in the water column, suggesting possible swim bladder defects (14/15 one month old juvenile *casper* mutants were found within 1 cm of the bottom of their tank compared to 2/21 wild-type siblings, $P < .01$, Fishers exact test). Lastly, *casper* mutants display a bleeding phenotype, possibly due to a decreased clotting rate, following euthanization in 0.04% tricaine relative to their wild-type siblings (Figure 2.1I-J).

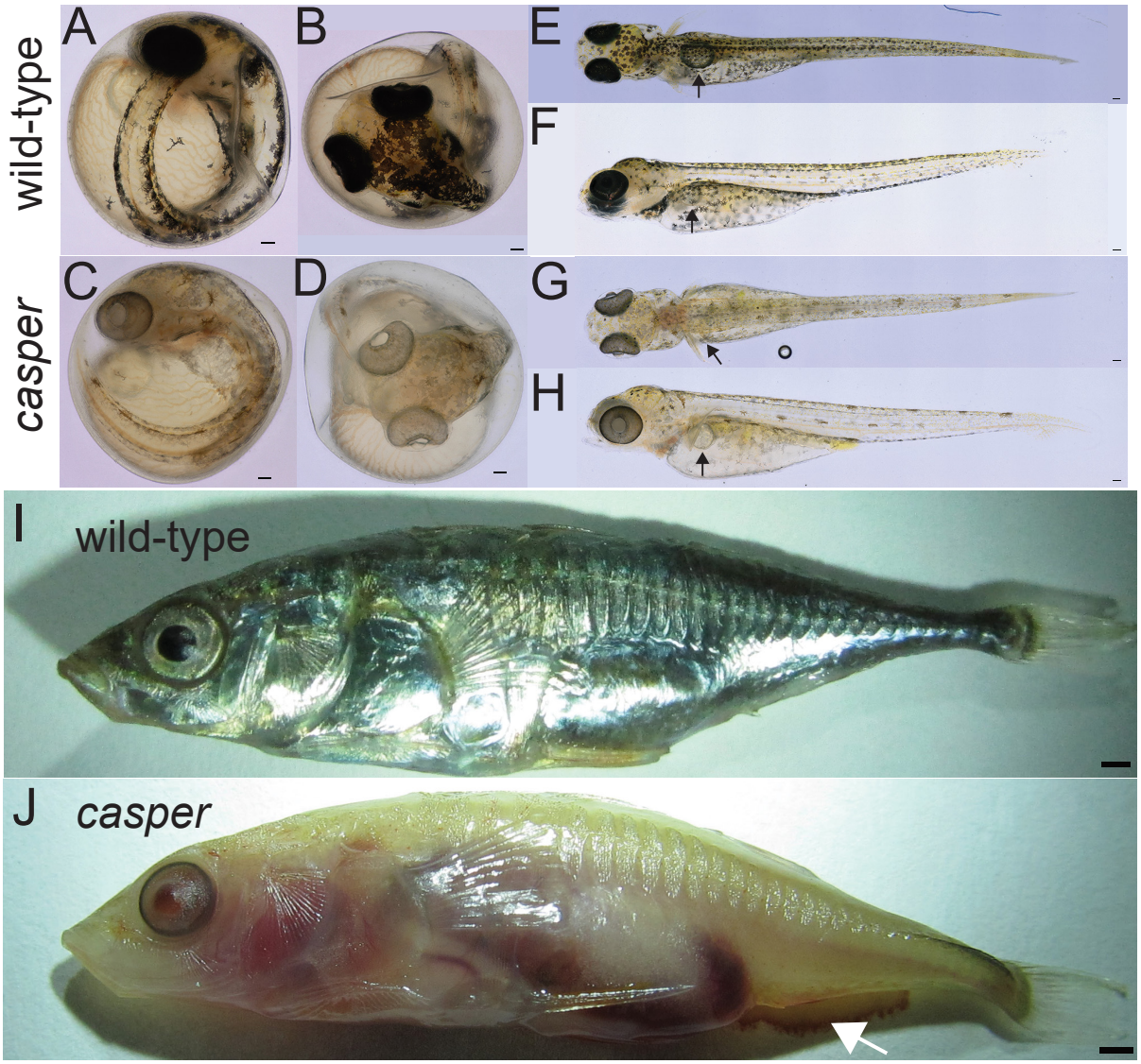

Figure 2.1: ***casper* mutants display severely reduced eye and body pigmentation.** (A-D) *casper* mutants are hypopigmented at 7 days post fertilization (dpf) and have severely reduced melanization of retinal pigmented epithelium (RPE): lateral (A,C) and dorsal (B,D) views. (E-H) Hypopigmentation of *casper* mutants persists at 14 dpf, with severely reduced melanophores but slightly melanized RPE: dorsal (E,G) and lateral (F,H) views. Black arrows indicate inflated swim bladder. (I, J) *casper* males are translucent at five months, with severely reduced iridophores and melanophore pigmentation, and highly reduced melanization of RPE. Mutants also bleed after euthanization (arrow in J). scale bars = 100 μm (A-H), 1 mm (I).

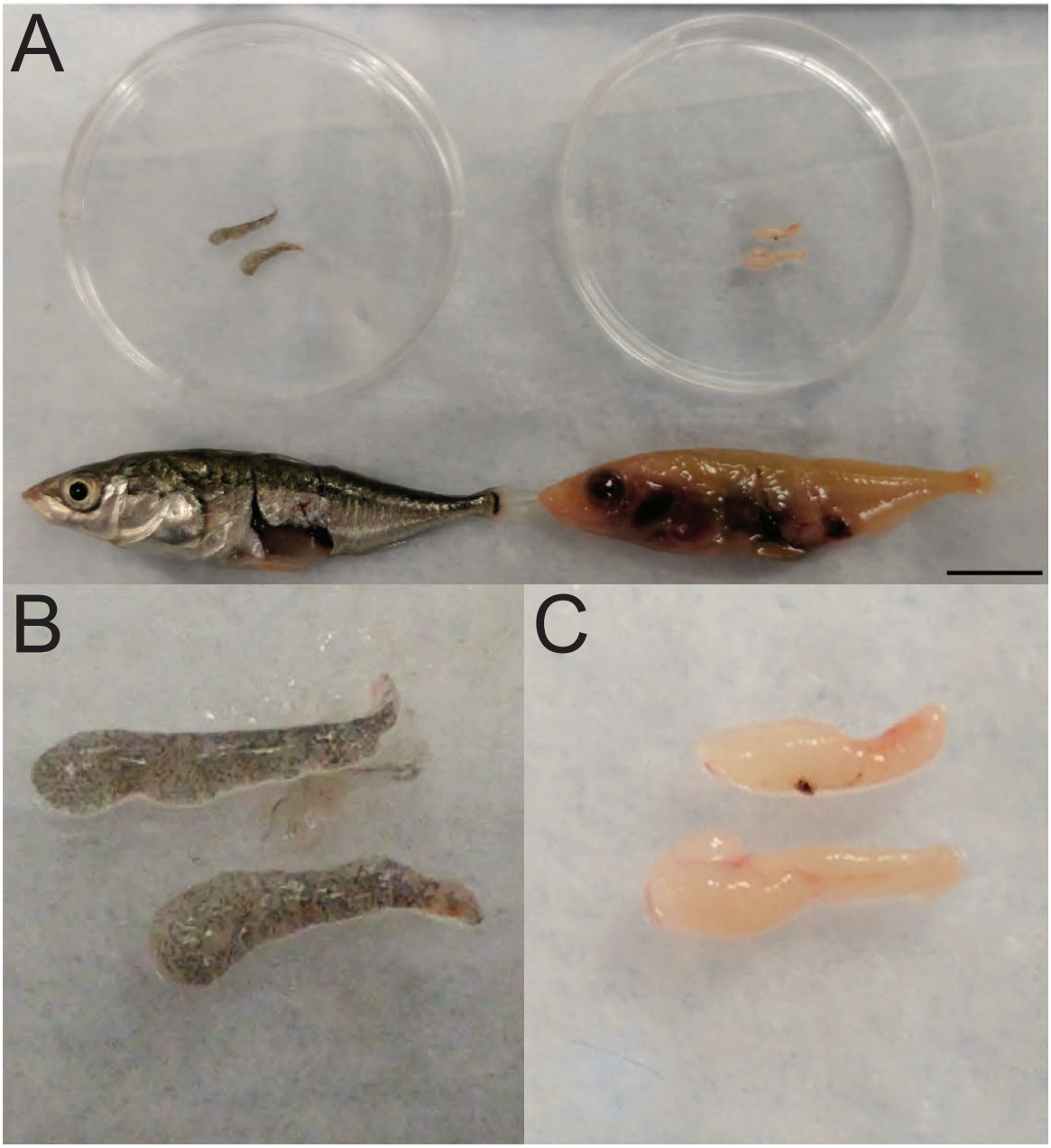

Figure 2.2: ***casper* mutant testes have reduced pigmentation.** (A) *casper* male (right) displays reduced pigmentation in its dissected testes relative to its wild-type sibling (left). (B) Wild-type dissected testes, displaying normal pigmentation patterns. (C) *casper* dissected testes are severely lacking in pigmentation.

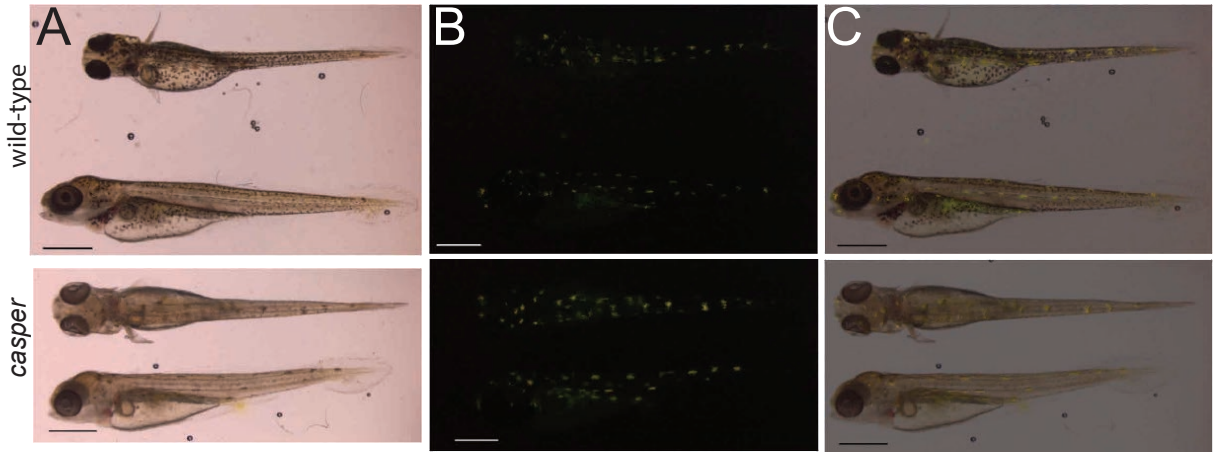

Figure 2.3: **Xanthophores appear unaffected in *casper* mutants.** Wild-type (top) and *casper* mutant (bottom) 15 dpf larvae both possess pigmented xanthophores, which fluoresce yellow under GFP fluorescence filters (B). *casper* xanthophores are the most visible pigmented cell type even under bright-field conditions (A), scale bars = 1mm.

casper is a spontaneous, X-linked recessive albino mutation

The original *casper* mutant fish was a single male, and first appeared in a clutch of 80 fish from a cross between a marine male from the Rabbit Slough (RABS), Alaska population, and a freshwater creek female from the Cerrito Creek (CERC), California population (Figure 2.4A). To map the locus responsible for the *casper* phenotype, the original *casper* male was outcrossed to a female fish from a different marine population (Table 2.1). All the resulting F_1 progeny were phenotypically wild-type, suggesting the *casper* mutation was either recessive or mosaic, with the germline of the original mutant fish not containing the *casper* mutation. As sticklebacks have a simple XY sex determination mechanism (Peichel et al. 2004), a spontaneous X-linked recessive mutation would be displayed in the original hemizygous male, but not in F_1 progeny. We thus hypothesized that *casper* was X-linked. To test this hypothesis, we generated six outcrosses of F_1 females to four males from two other populations and observed the F_2s. Consistent with an X-linked mutation, 103/419 (24.6%) of the F_2 offspring were *casper* mutants (Table 2.1), and molecular genotyping (Table 2.3) of 47 of the *casper* mutants showed that they were all male, confirming the sex-linked nature of *casper* (P <1.4e-14, two-tailed binomial test).

We next mapped the *casper* locus using a bulk segregant approach (Schneeberger et al. 2009; Cuperus et al. 2010; Zuryn et al. 2010; Doitsidou et al. 2010; Bowen et al. 2012; Obholzer et al. 2012). A barcoded Illumina sequencing library was created using genomic DNA pooled from 47 F_2 *casper* males from six different F_1 crosses. Additional barcoded libraries were created using DNA from the original *casper* fish as well as the F_0 female crossed to *casper*, and all libraries were sequenced to moderate (10-14x) coverage (Table

2.2). We mapped *casper* by examining the proportion of reads at each variant position that matched the *casper* male's allele. As an additional measure, we also looked for a loss of variants with mixed mapped reads (reflecting positions where most or all mutant male fish have the same X-chromosome genotype), as measured by the proportion of variants which only have reads supporting a single allele within a 50 variant genomic window (Figure 2.4B). Both measures had similar peaks along stickleback chromosome 19, the stickleback X chromosome (Peichel et al. 2004), near the stickleback ortholog of a human oculocutaneous albinism gene, *Hermansky-Pudlak syndrome 5* (*Hps5*) (Figure 2.4B).

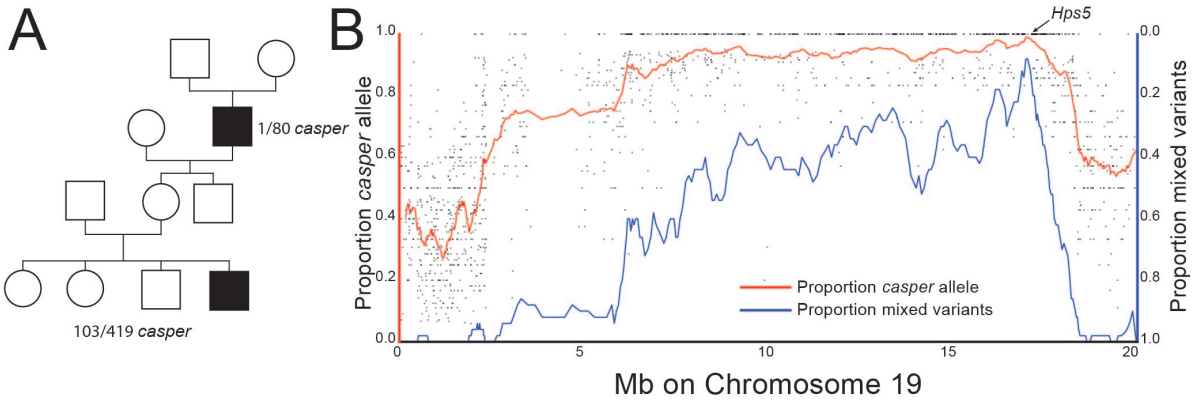

Figure 2.4: ***casper* is X-linked and maps to a region of chromosome 19 near *Hps5*.** Pedigree of the spontaneous appearance of the original *casper* mutant and X-linked transmission in subsequent generations. Of the 103 *casper* mutants, 47 were genotyped and all 47 were male by molecular genotyping with sex-specific primers. (B) Bulk segregant analysis of 47 *casper* mutants by high-throughput sequencing. Each point is the proportion of the variant allele with the *casper* genotype, with the red line showing a sliding window average across 50 variants. The blue line shows the proportion of variants called as heterozygous across a 50 variant sliding window. The peak of both red and blue lines is near the stickleback *Hps5* gene. X-axis shows the chromosome 19 revised genome assembly (Glazer et al. 2015). The top of the Y-axis for the proportion of mixed variants (blue line) is zero.

casper is the result of the insertion of a single base-pair into the coding sequencing of *Hps5*

We next sought to determine the mutation responsible for the *casper* phenotype. Genome-wide variant discovery using high-throughput sequencing data from the original *casper* male and the F_0 wild-type female he was crossed to revealed 14 total variants within the predicted *Hps5* coding frame. Of these, ten were synonymous point mutations, and two were substitutions found in both the affected *casper* male and unaffected female. The unaffected female had a mutation that results in a substitution from alanine to valine, which appears neutral (score of 0) in the BLOSUM62 matrix (Henikoff and Henikoff 1992), and which is also found in the orthologous mouse HPS5 sequence (Figure 2.5A). The only remaining and highest impact variant was the insertion of a G in the sixth exon of *Hps5*, resulting in a frameshift and predicted early stop appearing seven codons following the novel insertion (Figure 2.5A, Figure 2.6). This variant is only present within the *casper* male and not in his unaffected female mate. Sanger sequencing in both the *casper* male (Figure 2.5B) and wild-type female (Figure 2.5C) validated this insertion.

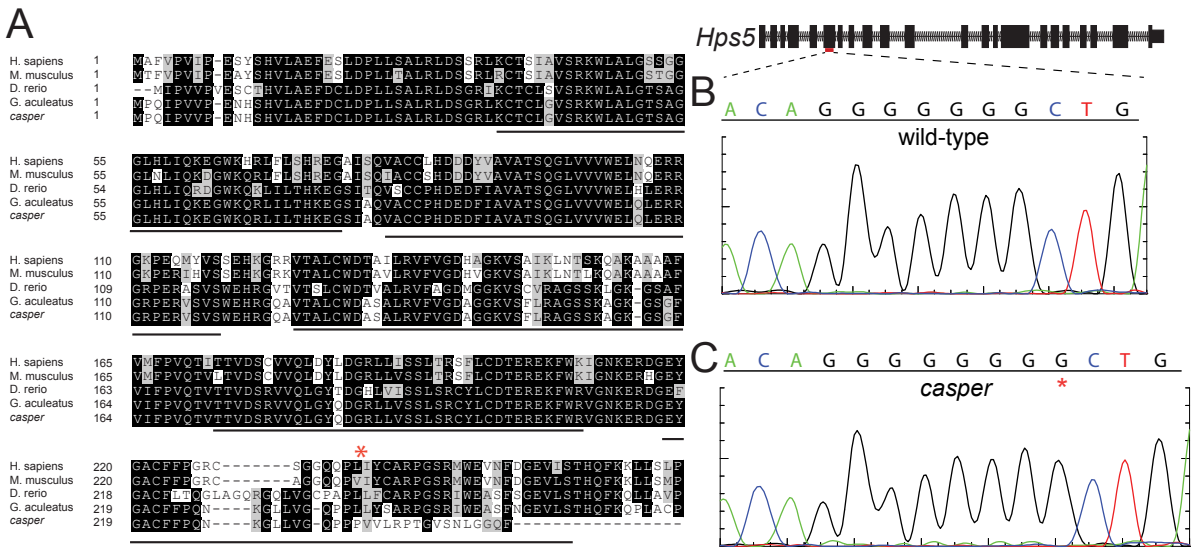

Figure 2.5: ***casper* mutants contain a 1 base-pair insertion in exon 6 of *Hps5*.** (A) A Clustal Omega (Sievers et al. 2011) multiple alignment of predicted amino-terminal HPS5 sequences from vertebrates with a known loss of function phenotype, as well as the predicted sequence from the original *casper* mutant (*casper*), and his wild-type F_0 mate (G. acu). Black lines indicate WD40 repeats predicted in human HPS5, and the red asterisk indicates the position of the *casper* insertion. See Figure 2.6 for full amino acid alignment (B) Sanger sequencing of the red indicated region of *Hps5* in wild-type fish. (C) Sanger sequencing of the red indicated region of *Hps5* in the original *casper* mutant. The red * indicates the inserted G, which results in a predicted frame-shift and early truncation of the HPS5 protein.

Figure 2.6: **Full alignment of predicted HPS5 orthologues.** The Clustal Omega (Sievers et al. 2011) alignment of the full predicted HPS5 protein sequences in a variety of animals shows the highly conserved nature of the protein.

Genome editing of *Hps5* phenocopies *casper*

We next tested whether other predicted loss-of-function mutations in stickleback *Hps5* could phenocopy *casper* mutants. The CRISPR/Cas9 system has been shown to be effective in genome editing in a wide range of model organisms, including another teleost, zebrafish (Hwang et al. 2013; Talbot and Amacher 2014). Two guide RNAs (gRNAs) were designed to target the sixth exon of stickleback *Hps5* (Figure 2.7A). Injection of either of the two gRNAs alone as well as Cas9 mRNA at the one cell stage resulted in a wide range of insertions/deletions (indels) within the *Hps5* coding region of representative injected embryos (Figure 2.7B). Co-injection of the two gRNAs with Cas9 mRNA resulted in local indels around each gRNA target, but also larger deletions between the two, with an overall increase in indel size over single guides (P <0.05, 1-tailed Mann-Whitney U, Figure 2.7B).

Hps5 gRNA injected embryos phenocopied *casper* mutants, with severely reduced melanization in both the RPE and melanophores (Figure 2.7C). Coinjection of these two gRNAs resulted in highly efficient induction of pigmentation phenotypes. Only 3/32 (9%) of the surviving embryos had fully wild-type pigmentation at 6dpf, while 15/32 (46.9%) displayed a severe loss of pigment, and 14/32 (44%) appeared mosaic (Figure 2.7C). Embryos injected with only a single gRNA displayed a similar loss of pigmentation, though with decreased efficiency (P <0.01, binomial test, Table 2.4). Overall, we observed severe or mosaic casper-like pigmentation phenotypes in 29/32 (91%) of injected embryos, and 8/9 (90%) of sequenced target regions contained indels near a protospacer adjacent motif (PAM) (Figure 2.7B). *Hps5* injected embryos were viable and displayed mosaic reduced RPE and melanophore melanization into adulthood, as well as a mosaic loss of iridophore pigmentation (Figure 2.7D).

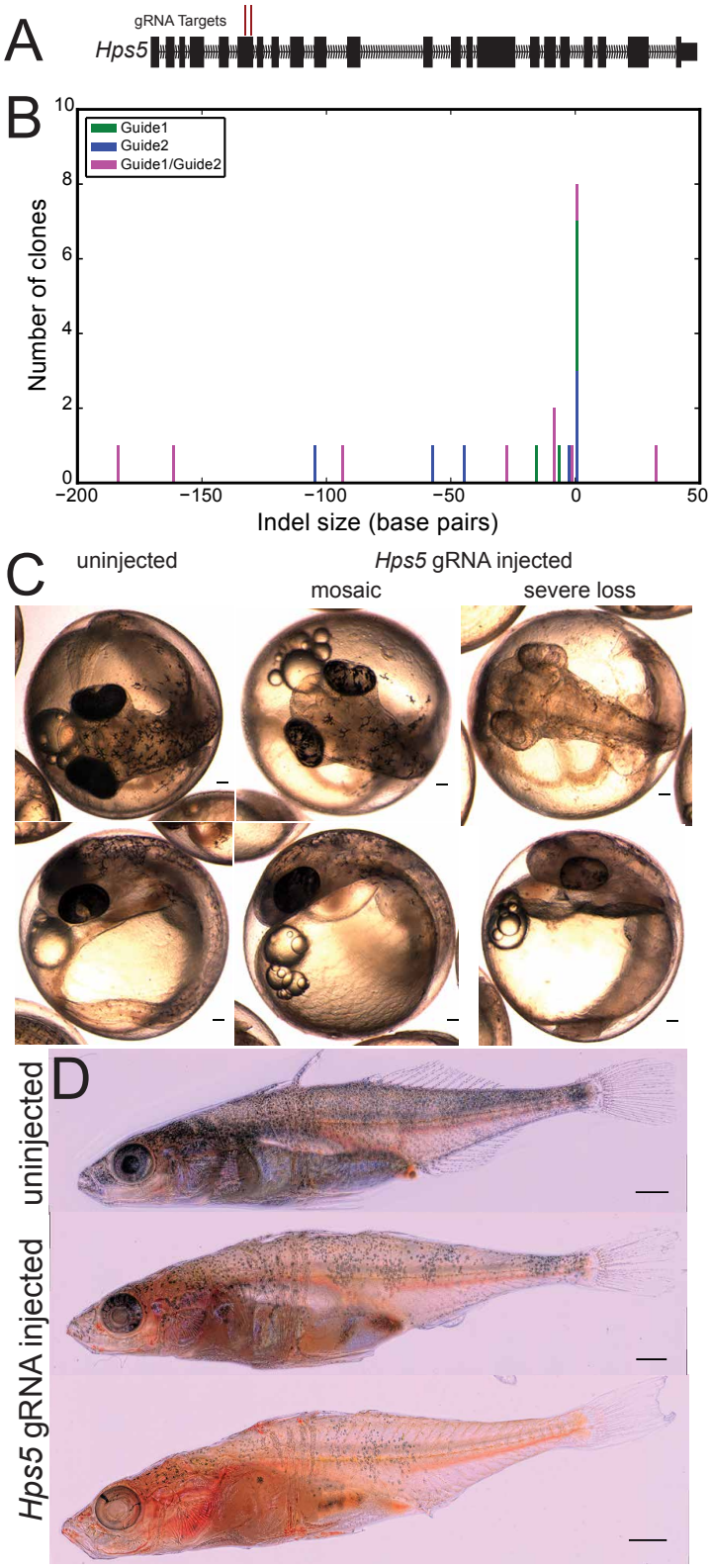

Figure 2.7: **CRISPR/Cas9 induced mutations in *Hps5* phenocopy the *casper* mutation.** (A) Two guide RNA sequences were targeted to the 6th exon of stickleback *Hps5* and were singly or co-injected along with Cas9 mRNA at the 1 cell stage. (B) Sanger sequencing of clones derived from single representative *Hps5* injected fish reveals the highly efficient and mosaic nature of Cas9 mediated indel formation. (C) Most [29/32 (91%)] embryos injected with *Hps5* gRNAs displayed pigmentation reductions, with 14/32 (44%) displaying a mosaic loss of pigment in melanophores and RPE, and 15/32 (47% of F_0 injected fish) displayed >75% loss of RPE (severe loss). (D) Adult *Hps5* gRNA injected sticklebacks show a mosaic loss of pigmented melanophores and iridophores and are partially translucent. scale bars = 100 μm (C), 1mm (D).

Guide RNA	Wild-Type	Mosaic	*casper*-like
Hps5 guide 1	2 (8%)	19 (73%)	5 (19%)
Hps5 guide 2	6 (15%)	25 (64%)	8 (21%)
Hps5 guide 1 + *Hps5* guide 2	3 (9%)	14 (44%)	15 (47%)

Table 2.4: **Highly efficient generation of mosaically albino embryos with CRISPR/Cas9.** Each row lists all viable embryos of a single injected clutch. Guide RNA indicates whether a single guide or two guides was co-injected along with Cas9 mRNA. Wild-type is the number of embryos displaying no visible albino clones. Mosaic indicates the number of embryos with albino clones (1% to 75% albino). *casper*-like indicates the number of embryos that were severely affected (>75% albino).

Visualizing fluorescent transgenic reporters in *casper* mutants

Fish embryos are highly transparent and develop externally, allowing for easy visualization of early embryogenesis. Combined with efficient transgene incorporation using Tol2 transgenesis (Kawakami 2005; Erickson et al. 2016), sticklebacks represent a powerful system for assaying the activity of developmental enhancers using fluorescent reporter constructs (Erickson et al. 2015). However, as fish develop and become more pigmented, imaging becomes extremely difficult, and investigations into late-acting enhancer elements require microdissection. We next sought to test whether the reduced pigmentation of *casper* embryos would allow for easier imaging of enhancer patterns in juvenile fish.

We crossed a heterozygous *casper* female to a male carrying a single copy of a fluorescent reporter of a previously described 190 base pair *Bmp6* enhancer (Erickson et al. 2015), previously described to be active in the fins and teeth. Two months post fertilization, wild-type males showed reported green fluorescent protein (GFP) expression in the lens of the eye, a known internal positive control domain of expression of the zebrafish *heat-shock 70-like* (*Hsp70l*) promoter used (Erickson et al. 2015), with other expression domains obscured by pigmented cells (Figure 2.8A). *casper* mutants carrying the transgene showed similar robust lens expression, but also better revealed other visible juvenile expression domains in oral and pharyngeal teeth, as well as revealed a previously unreported major expression pattern in the liver (Figure 2.8B-D).

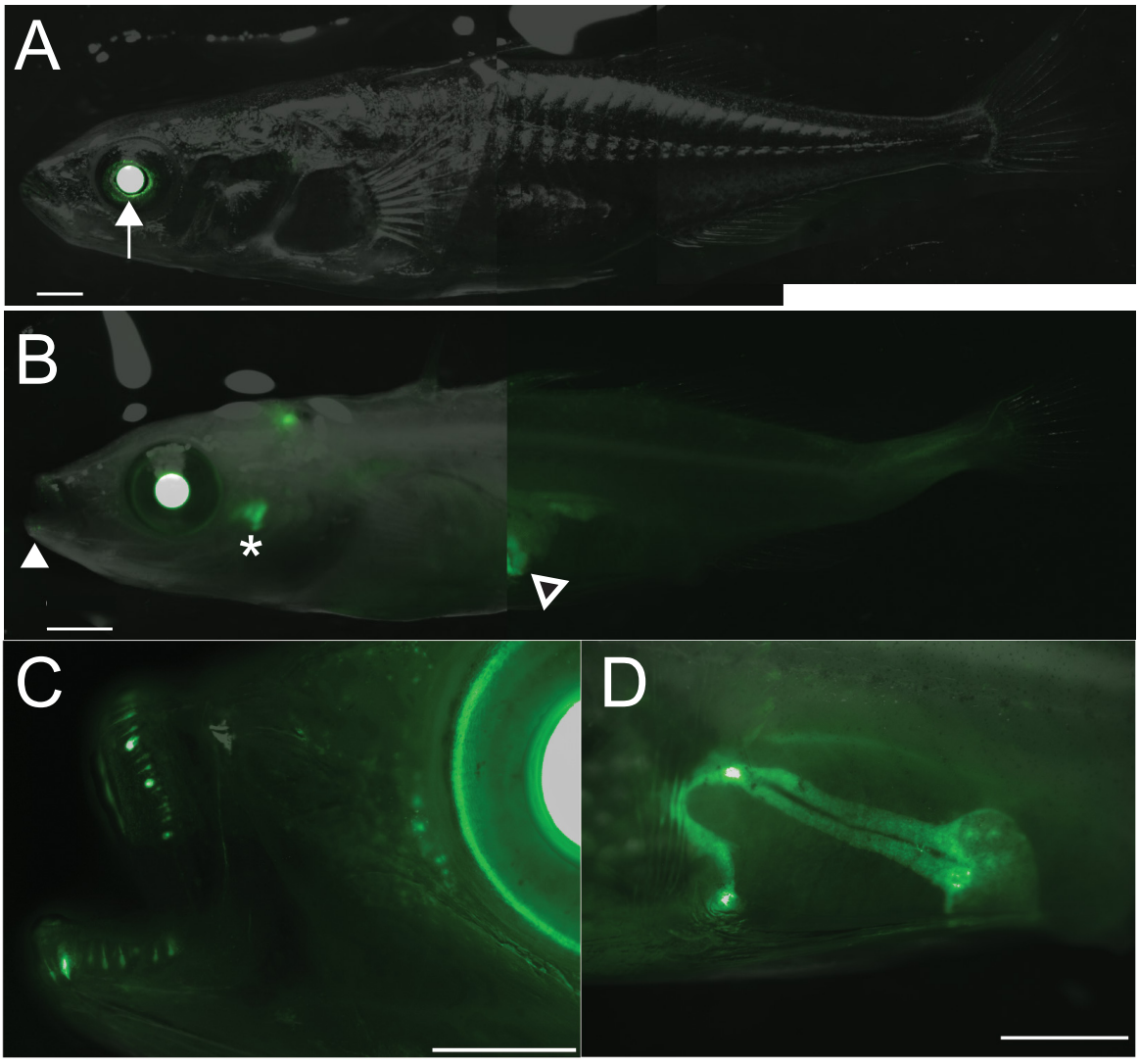

Figure 2.8: ***casper* mutants allow improved visualization of fluorescent reporter genes** (A) A fluorescent GFP reporter of a previously described tooth enhancer (Erickson et al. 2015) is only clearly visible in the lenses (white arrow) of wild-type fish. (B) A *casper* sibling with the same stable integration of the fluorescent reporter reveals more readily apparent oral (white filled arrowhead) and pharyngeal teeth (asterisk) domains, and clearly reveals a major previously unreported liver expression domain (black filled arrowhead). (C) GFP reporter activity during oral tooth development and replacement. (D) The left side of the liver expression domain in *casper* fish (right visible in B). Scale bars = 1 mm.

2.5 Discussion

Here we present the first stickleback model of Hermansky-Pudlak syndrome, the spontaneous X-linked recessive *casper* phenotype, resulting from a frame-shifting insertion in *Hps5*. *Hps5* mutants display oculocutaneous albinism, with severely decreased pigment in melanophores, iridophores, and erythrophores, though interestingly not xanthophores. These pigment phenotypes suggest that pterinosomes in xanthophores develop in a *Hps5* independent manner, while the melanins, carotenoids, or guanine crystal containing organelles in melanophores, erythrophores, and iridophores develop in a *Hps5* dependent manner. Pigment was also drastically reduced in the retinal pigmented epithelium (RPE) of *Hps5* mutants, suggesting defects in LRO biogenesis are not restricted to chromatophores, as the RPE does not contain chromatophores (Schraermeyer and Heimann 1999). Additionally, we observed a bleeding phenotype in casper mutants, potentially similar to the bleeding diathesis phenotype seen in human *HPS5* mutants (Huizing et al., 2004). Overall, these phenotypes agree with the reported phenotypes of mutations in BLOC-2 complex members in other species (Zhang et al. 2003; Daly et al. 2013; Nakayama et al. 2016).

It is unclear why the stickleback *casper* mutation in exon 6 of *Hps5* causes more severe pigmentation phenotypes than the zebrafish *snow white* mutation (Daly et al. 2013), which is lethal, unlike the stickleback *Hps5* mutations reported here. Whether more N-terminal mutations within or prior to the WD40 domains also cause lethality, as previously proposed (Daly et al. 2013), in sticklebacks could be tested by inducing mutations more N-terminal in *Hps5*.

Hps5 underlies the casper phenotype

Our mapping-by-sequencing approach using bulked segregant analysis of *casper* mutants revealed a peak genetic signal near *Hps5* on chromosome 19, the stickleback X chromosome. By Sanger sequencing we found the insertion of a single guanine to a heptaguanine run in the coding frame of *Hps5* in *casper* mutants. This spontaneous insertion might be due to the presence of this homopolymer repeat, as the rate of indel formation is elevated at long homopolymer runs (Montgomery et al. 2013), potentially due to polymerase slippage (Levinson and Gutman 1987). Lastly, we showed injection of Cas9 mRNA and *Hps5* guide RNAs resulted in induced mutations in *Hps5* and embryos displaying oculocutaneous albinism phenotypes similar to *casper* mutants, demonstrating *Hps5* disruption underlies the *casper* phenotype. Although our approach using 47 pooled mutants identified a strong candidate gene, future mapping-by-sequencing of other mutations could generate improved genetic resolution by pooling even more mutant DNAs.

Live imaging of fluorescent reporters in *casper* embryos

casper and *Hps5* mutant embryos are both semitransparent even as adults, unlike their wild-type siblings. This transparency allows for better live imaging of fluorescent transgenic

reporters in *Hps5* mutant fish. Imaging a previously characterized enhancer of *Bmp6* (Erickson et al. 2015) in *Hps5* mutants, we discovered a major unreported expression domain in the liver. We also found that imaging the previously reported dynamic tooth expression domains (Erickson et al. 2015) to be greatly facilitated by the depigmented phenotype of *Hps5* mutants. The viable nature of the *casper* mutation allows for the creation of stable transgenic lines to more easily visualize reporter gene expression, especially at older postembryonic stages. The X-linked nature of the *casper* mutation allows for the recovery of *casper* males from outcrosses to different stable lines within a single generation.

Genome editing with CRISPR/Cas9

We report the first successful generation of loss-of-function mutations using the CRISPR/Cas9 system in sticklebacks. Coinjection of Cas9 mRNA along with either of two guide RNAs (gRNAs) targeted to exon 6 of *Hps5* resulted in a high frequency (>90%) of embryos with severe or mosaic pigmentation phenotypes. Furthermore, strong pigment phenotypes were seen in both XY males and XX females, implying that Cas9 is able to induce bi-allelic hits in stickleback embryos. We also see evidence for high efficiency in our single clone Sanger sequencing, as most (8/9) of our sequenced clones contained induced mutations. This high rate suggests that *Hps5* guide RNAs could be used as a marker for other guide RNAs in a co-CRISPR approach (Kim et al. 2014; Kane et al. 2017), with more albino embryos representing embryos with high levels of nuclear Cas9 activity. As stickleback testes are pigmented, screening for albino testes might further enrich for germline mutations from other co-injected gRNAs.

Injection of even a single gRNA is sufficient to induce large (>25 bp) deletions around the genomic target, similar in size to zebrafish reports but larger than in human cells (Hwang et al. 2013; Paquet et al. 2016). Co-injection of two gRNAs resulted in an increase in efficiently edited embryos, suggesting co-injection results in a high F_0 mutation induction efficiency and allows phenotypic analysis of F_0 injected embryos, as we have done here. Furthermore, co-injection of two gRNAs significantly increased the induction of deletions of the intervening sequence between our two *Hps5* guide RNAs, showing that Cas9 can efficiently induce genomic deletions in stickleback embryos. These deletions could have a stronger effect on gene function, and allow for easy and inexpensive genotyping of stable mutants. Additionally, these deletions could be targeted to non-coding DNA such as enhancers, which might not be as sensitive to small deletions as a coding frame. Inducing deletions of regulatory elements will allow functional genetic tests of candidate regulatory elements that underlie evolved changes.

2.6 Acknowledgements

We thank Anthony Lee for first noticing the original *casper* male. This work used the Vincent J. Coates Genomics Sequencing Laboratory at UC Berkeley, supported by NIH S10

Instrumentation Grants S10RR029668 and S10RR027303. This work was funded by NIH R01DE021475 (to C.T.M.), and NIH genomics training grant 5T32HG000047-15 (to J.C.H.)

2.7 Literature Cited

Bowen, M. E., K. Henke, K. R. Siegfried, M. L. Warman, and M. P. Harris, 2012 Efficient mapping and cloning of mutations in zebrafish by low-coverage whole-genome sequencing. Genetics 190: 10171024.

Burger, A., H. Lindsay, A. Felker, C. Hess, C. Anders et al., 2016 Maximizing mutagenesis with solubilized CRISPR-Cas9 ribonucleoprotein complexes. Development 143: 20252037.

Cooper, K. M., R. T. Hanlon, and B. U. Budelmann, 1990 Physiological color change in squid iridophores. II. Ultrastructural mechanisms in Lolliguncula brevis. Cell Tissue Res. 259: 1524.

Cuperus, J. T., T. A. Montgomery, N. Fahlgren, R. T. Burke, T. Townsend et al., 2010 Identification of MIR390a precursor processing-defective mutants in Arabidopsis by direct genome sequencing. Proc. Natl. Acad. Sci. U. S. A. 107: 466471.

Dahlem, T. J., K. Hoshijima, M. J. Jurynec, D. Gunther, C. G. Starker et al., 2012 Simple methods for generating and detecting locus-specific mutations induced with TALENs in the zebrafish genome. PLoS Genet. 8: e1002861.

Daly, C. M. S., J. Willer, R. Gregg, and J. M. Gross, 2013 snow white, a zebrafish model of Hermansky-Pudlak Syndrome type 5. Genetics 195: 481494.

DellAngelica, E. C., 2004 The building BLOC(k)s of lysosomes and related organelles. Curr. Opin. Cell Biol. 16: 458464.

DellAngelica, E. C., C. Mullins, S. Caplan, and J. S. Bonifacino, 2000 Lysosome-related organelles. FASEB J. 14: 12651278.

DePristo, M. A., E. Banks, R. Poplin, K. V. Garimella, J. R. Maguire et al., 2011 A framework for variation discovery and genotyping using next-generation DNA sequencing data. Nat. Genet. 43: 491498.

Doitsidou, M., R. J. Poole, S. Sarin, H. Bigelow, and O. Hobert, 2010 C. elegans mutant identification with a one-step whole-genome-sequencing and SNP mapping strategy. PloS One 5: e15435.

Erickson, P. A., P. A. Cleves, N. A. Ellis, K. T. Schwalbach, J. C. Hart et al., 2015 A 190 base pair, TGF- responsive tooth and fin enhancer is required for stickleback Bmp6 expression. Dev. Biol. 401: 310323.

Erickson, P. A., N. A. Ellis, and C. T. Miller, 2016 Microinjection for transgenesis and genome editing in threespine sticklebacks. J. Vis. Exp. e54055e54055.

Falcn-Prez, J. M., R. Romero-Caldern, E. S. Brooks, D. E. Krantz, and E. C. DellAngelica, 2007 The Drosophila pigmentation gene pink (p) encodes a homologue of human HermanskyPudlak Syndrome 5 (HPS5). Traffic 8: 154168.

Friedland, A. E., Y. B. Tzur, K. M. Esvelt, M. P. Colaicovo, G. M. Church et al., 2013 Heritable genome editing in C. elegans via a CRISPR-Cas9 system. Nat. Methods 10: 741743.

Fujii, R., 2000 The regulation of motile activity in fish chromatophores. Pigment Cell Res. 13: 300319.

Fujii, T., Y. Banno, H. Abe, S. Katsuma, and T. Shimada, 2012 A homolog of the human Hermansky-Pudluck syndrome-5 (HPS5) gene is responsible for the oa larval translucent mutants in the silkworm, Bombyx mori. Genetica 140: 463468.

Gibson, G., 2005 The Synthesis and Evolution of a Supermodel. Science 307: 18901891.

Glazer, A. M., P. A. Cleves, P. A. Erickson, A. Y. Lam, and C. T. Miller, 2014 Parallel developmental genetic features underlie stickleback gill raker evolution. EvoDevo 5: 19.

Glazer, A. M., E. E. Killingbeck, T. Mitros, D. S. Rokhsar, and C. T. Miller, 2015 Genome assembly improvement and mapping convergently evolved skeletal traits in sticklebacks with genotyping-by-sequencing. G3 (Bethesda) 5: 14631472.

Guo, X., T. Zhang, Z. Hu, Y. Zhang, Z. Shi et al., 2014 Efficient RNA/Cas9-mediated genome editing in Xenopus tropicalis. Development 141: 707714.

Helip-Wooley, A., W. Westbroek, H. M. Dorward, A. Koshoffer, M. Huizing et al., 2007 Improper trafficking of melanocyte-specific proteins in Hermansky-Pudlak syndrome type-5. J. Invest. Dermatol. 127: 14711478.

Henikoff, S., and J. G. Henikoff, 1992 Amino acid substitution matrices from protein blocks. Proc. Natl. Acad. Sci. U. S. A. 89: 1091510919.

Hoshijima, K., M. J. Jurynec, and D. J. Grunwald, 2016 Precise editing of the zebrafish genome made simple and efficient. Dev. Cell 36: 654667.

Huizing, M., R. Hess, H. Dorward, D. A. Claassen, A. Helip-Wooley et al., 2004 Cellular, molecular and clinical characterization of patients with HermanskyPudlak Syndrome Type 5. Traffic 5: 711722.

Hwang, W. Y., Y. Fu, D. Reyon, M. L. Maeder, S. Q. Tsai et al., 2013 Efficient genome editing in zebrafish using a CRISPR-Cas system. Nat. Biotechnol. 31: 227229.

Irion, U., J. Krauss, and C. Nsslein-Volhard, 2014 Precise and efficient genome editing in zebrafish using the CRISPR/Cas9 system. Development 141: 48274830.

Irion, U., A. P. Singh, and C. Nsslein-Volhard, 2016 The Developmental genetics of vertebrate color pattern formation: lessons from zebrafish. Curr. Top. Dev. Biol. 117: 141169.

Jamann, T. M., P. J. Balint-Kurti, and J. B. Holland, 2015 QTL mapping using high-throughput sequencing. Methods Mol. Biol. Clifton NJ 1284: 257285.

Jinek, M., A. East, A. Cheng, S. Lin, E. Ma et al., 2013 RNA-programmed genome editing in human cells. eLife 2:e00471

Jones, F. C., M. G. Grabherr, Y. F. Chan, P. Russell, E. Mauceli et al., 2012 The genomic basis of adaptive evolution in threespine sticklebacks. Nature 484: 5561.

Kane, N. S., M. Vora, K. J. Varre, and R. W. Padgett, 2017 Efficient screening of CRISPR/Cas9-induced events in Drosophila using a co-CRISPR strategy. G3 (Bethesda) 7: 8793.

Kawakami, K., 2005 Transposon tools and methods in zebrafish. Dev. Dyn. 234: 244254.

Kelsh, R. N., 2004 Genetics and evolution of pigment patterns in fish. Pigment Cell Res. 17: 326336.

Kelsh, R. N., M. L. Harris, S. Colanesi, and C. A. Erickson, 2009 Stripes and belly-spotsA review of pigment cell morphogenesis in vertebrates. Semin. Cell Dev. Biol. 20: 90104.

Kim, H., T. Ishidate, K. S. Ghanta, M. Seth, D. Conte et al., 2014 A co-CRISPR strategy for efficient genome editing in Caenorhabditis elegans. Genetics 197: 10691080.
Langmead, B., and S. L. Salzberg, 2012 Fast gapped-read alignment with Bowtie 2. Nat. Methods 9: 357359.
Levinson, G., and G. A. Gutman, 1987 Slipped-strand mispairing: a major mechanism for DNA sequence evolution. Mol. Biol. Evol. 4: 203221.
Li, H., B. Handsaker, A. Wysoker, T. Fennell, J. Ruan et al., 2009 The Sequence Alignment/Map format and SAMtools. Bioinformatics 25: 20782079.
Marks, M. S., and M. C. Seabra, 2001 The melanosome: membrane dynamics in black and white. Nat. Rev. Mol. Cell Biol. 2: 738748.
Martin, A., J. M. Serano, E. Jarvis, H. S. Bruce, J. Wang et al., 2016 CRISPR/Cas9 mutagenesis reveals versatile roles of Hox genes in crustacean limb specification and evolution. Curr. Biol. 26: 1426.
McKenna, A., M. Hanna, E. Banks, A. Sivachenko, K. Cibulskis et al., 2010 The Genome Analysis Toolkit: A MapReduce framework for analyzing next-generation DNA sequencing data. Genome Res. 20: 12971303.
Milinski, M., and T. C. M. Bakker, 1990 Female sticklebacks use male coloration in mate choice and hence avoid parasitized males. Nature 344: 330333.
Mills, M. G., and L. B. Patterson, 2009 Not just black and white: pigment pattern development and evolution in vertebrates. Semin. Cell Dev. Biol. 20: 7281.
Montgomery, S. B., D. L. Goode, E. Kvikstad, C. A. Albers, Z. D. Zhang et al., 2013 The origin, evolution, and functional impact of short insertiondeletion variants identified in 179 human genomes. Genome Res. 23: 749761.
Nakayama, T., K. Nakajima, A. Cox, M. Fisher, M. Howell et al., 2016 no privacy, a Xenopus tropicalis mutant, is a model of human Hermansky-Pudlak Syndrome and allows visualization of internal organogenesis during tadpole development. Dev. Biol. 2. pii: S0012-1606(16)30371-2.
Nguyen, T., E. K. Novak, M. Kermani, J. Fluhr, L. L. Peters et al., 2002 Melanosome morphologies in murine models of Hermansky-Pudlak Syndrome reflect blocks in organelle development. J. Invest. Dermatol. 119: 11561164.
Obholzer, N., I. A. Swinburne, E. Schwab, A. V. Nechiporuk, T. Nicolson et al., 2012 Rapid positional cloning of zebrafish mutations by linkage and homozygosity mapping using whole-genome sequencing. Development 139: 42804290.
Oshima, N., 2001 Direct reception of light by chromatophores of lower vertebrates. Pigment Cell Res. 14: 312319.
Paquet, D., D. Kwart, A. Chen, A. Sproul, S. Jacob et al., 2016 Efficient introduction of specific homozygous and heterozygous mutations using CRISPR/Cas9. Nature 533: 125129.
Peichel, C. L., J. A. Ross, C. K. Matson, M. Dickson, J. Grimwood et al., 2004 The master sex-determination locus in threespine sticklebacks is on a nascent Y chromosome. Curr. Biol. 14: 14161424.
Rozen, S., and H. J. Skaletsky, 2000 Primer3. Methods Mol Biol. 2000:132:365-86
Sander, J. D., M. L. Maeder, D. Reyon, D. F. Voytas, J. K. Joung et al., 2010 ZiFiT

(Zinc Finger Targeter): an updated zinc finger engineering tool. Nucleic Acids Res. 38: W462W468.
Schneeberger, K., S. Ossowski, C. Lanz, T. Juul, A. H. Petersen et al., 2009 SHOREmap: simultaneous mapping and mutation identification by deep sequencing. Nat. Methods 6: 550551.
Schraermeyer, U., and K. Heimann, 1999 Current understanding on the role of retinal pigment epithelium and its pigmentation. Pigment Cell Res. 12: 219236.
Shoup, J. R., 1966 The development of pigment granules in the eyes of wild type and mutant Drosophila melanogaster. J. Cell Biol. 29: 223249.
Sievers, F., A. Wilm, D. Dineen, T. J. Gibson, K. Karplus et al., 2011 Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. Mol. Syst. Biol. 7: 539.
Square, T., M. Romek, D. Jandzik, M. V. Cattell, M. Klymkowsky et al., 2015 CRISPR/Cas9-mediated mutagenesis in the sea lamprey Petromyzon marinus: a powerful tool for understanding ancestral gene functions in vertebrates. Development 142: 41804187.
Syrzycka, M., L. A. McEachern, J. Kinneard, K. Prabhu, K. Fitzpatrick et al., 2007 The pink gene encodes the Drosophila orthologue of the human Hermansky-Pudlak syndrome 5 (HPS5) gene. Genome 50: 548556.
Talbot, J. C., and S. L. Amacher, 2014 A streamlined CRISPR pipeline to reliably generate zebrafish frameshifting alleles. Zebrafish 11: 583585.
Van der Auwera, G. A., M. O. Carneiro, C. Hartl, R. Poplin, G. Del Angel et al., 2013 From FastQ data to high confidence variant calls: the Genome Analysis Toolkit best practices pipeline. Curr. Protoc. Bioinformatics 43: 11.10.1-33.
Wasmeier, C., A. N. Hume, G. Bolasco, and M. C. Seabra, 2008 Melanosomes at a glance. J Cell Sci 121: 39953999.
Wedekind, C., P. Meyer, M. Frischknecht, U. A. Niggli, and H. Pfander, 1998 Different carotenoids and potential information content of red coloration of male three-spined stickleback. J. Chem. Ecol. 24: 787801.
Wei, A.-H., X. He, and W. Li, 2013 Hypopigmentation in HermanskyPudlak syndrome. J. Dermatol. 40: 325329.
Zhang, Q., B. Zhao, W. Li, N. Oiso, E. K. Novak et al., 2003 Ru2 and Ru encode mouse orthologs of the genes mutated in human Hermansky-Pudlak syndrome types 5 and 6. Nat. Genet. 33: 145153.
Ziegler, I., 2003 The pteridine pathway in zebrafish: regulation and specification during the determination of neural crest cell-fate. Pigment Cell Res. 16: 172182.
Zuryn, S., S. Le Gras, K. Jamet, and S. Jarriault, 2010 A strategy for direct mapping and identification of mutations by whole-genome sequencing. Genetics 186: 427430.

Chapter 3

An intronic enhancer of *Bmp6* underlies evolved tooth gain in sticklebacks

A version of this article has been submitted to PLoS Genetics.

Phillip A. Cleves, James C. Hart, Rachel M. Agoglia, Monica T. Jimenez, Priscilla A. Erickson, Linda Gai, and Craig T. Miller
Department of Molecular and Cell Biology, University of California-Berkeley, Berkeley CA, 94720, USA

3.1 Abstract

Threespine stickleback fish offer a powerful system to dissect the genetic basis of morphological evolution in nature. Marine sticklebacks have repeatedly invaded and adapted to numerous freshwater environments throughout the Northern hemisphere. In response to new diets in freshwater habitats, changes in craniofacial morphology, including heritable increases in tooth number, have evolved in derived freshwater populations. Using a combination of quantitative genetics and genome resequencing, here we fine-mapped a quantitative trait locus (QTL) regulating evolved tooth gain to a cluster of ten QTL-associated single nucleotide variants, all within intron four of *Bone Morphogenetic Protein 6* (*Bmp6).* Transgenic reporter assays revealed this intronic region contains a tooth enhancer. We induced mutations in *Bmp6*, revealing required roles for survival, growth, and tooth patterning. Transcriptional profiling of *Bmp6* mutant dental tissues identified significant downregulation of a set of genes whose orthologs were previously shown to be expressed in quiescent mouse hair stem cells. Collectively these data support a model where mutations within a *Bmp6* intronic tooth enhancer contribute to evolved tooth gain, and suggest that ancient shared genetic circuitry regulates the regeneration of diverse vertebrate epithelial appendages including mammalian hair and fish teeth.

3.2 Introduction

Finding the genes and ultimately the mutations that drive the evolution of animal form remains an important goal in biology [1]. The *cis*-regulatory hypothesis proposes that *cis*-regulatory changes are the preferred substrate for morphological evolution because these mutations are more likely to bypass the negative pleiotropy typically generated by coding mutations in developmental regulatory genes [2]. Although many studies in a variety of organisms have found *cis*-regulatory alleles underlying morphological evolution, less is known about why or how *cis*-regulatory alleles are used [3,4]. For example, for genes found to have *cis*-regulatory alleles associated with evolved differences, whether coding mutations generate negative pleiotropy and/or reduced fitness remains largely untested in many natural populations.

Teeth are a classic model system for studying organ development and evolution in vertebrates [5,6]. During tooth development, epithelial and mesenchymal cells reciprocally signal to each other, integrating dynamic BMP, TGF-β, FGF, SHH, Notch, Activin, EDA, and Wnt signals to orchestrate the formation of a mature tooth [7,8]. Bone Morphogenetic Protein (BMP) signaling plays multiple critical roles during tooth development. During tooth initiation, epithelial *Bmp4* inhibits expression of *Pax9* and *Pitx2*, developmental markers of the forming tooth placode [9,10]. These results suggest an inhibitory role of BMP signaling on tooth development. However, several lines of evidence support an activating role of BMPs on tooth development. For example, exogenous BMP4 can rescue tooth development in *Msx1* mutant mice and accelerate tooth development in cultured tooth mandibles, suggesting an

activating role of BMP signaling [11,12]. Furthermore, mice with dental epithelial ablation of the BMP receptor, *Bmpr1a*, or transgenic for a construct overexpressing a BMP antagonist, *Noggin*, in dental epithelium have tooth arrest at the bud and placode stage, respectively [13,14]. Together, these results suggest that there are both activating and inhibitory roles of BMP signaling during tooth development. However, the roles of many BMP signaling components are not fully understood. Furthermore, the genetic pathways of early tooth pattern and initiation have been extensively studied and well characterized in mice. Because mice are monophyodont rodents that do not replace their teeth, considerably less is known about the developmental genetic basis of tooth replacement. Polyphyodont vertebrates (e.g. sharks, teleosts, and reptiles) that continuously replace their teeth offer an opportunity to study the genetic and developmental basis of tooth regeneration [6].

Threespine stickleback fish (*Gasterosteus aculeatus*) are an excellent model for understanding the molecular genetic basis of natural variation, including evolved differences in tooth number [15,16]. Sticklebacks have undergone a dramatic adaptive radiation in which ancestral marine sticklebacks have colonized freshwater lakes and streams throughout the Northern hemisphere [17]. Recent genetic studies have implicated *cis*-regulatory changes of developmental signaling molecules as underlying several aspects of stickleback morphological evolution [18-23]. Genome-wide searches for regions under selection during freshwater adaptation have found an enrichment in non-coding elements of the genome, further implicating *cis*-regulatory changes in underlying stickleback evolution [24].

Freshwater sticklebacks have evolved several morphological adaptations in their head skeleton, some likely due to the shift to feeding on larger prey in freshwater niches [25]. While many freshwater adaptations in sticklebacks involve skeletal loss, a constructive gain of pharyngeal tooth number is seen in freshwater benthic (adapted to lake bottom) and creek populations [19,26]. Pharyngeal teeth lie in the pharynx of fish and are serial and phylogenetic homologs of mammalian oral teeth [27]. Pharyngeal jaw patterning is an adaptive trait in fish that covaries with diet and ecological niche [28]. Many aspects of the developmental genetic circuitry regulating tooth development are conserved from mice to fish [29-31]. Thus, evolved tooth gain in sticklebacks provides a powerful opportunity to understand the evolutionary genetics of tooth development and replacement.

Evolved tooth gain in benthic freshwater fish from Paxton Lake in British Columbia is accompanied by an increase in the size of the tooth field, a decrease in tooth spacing, and an increase in tooth replacement rate late in development [19,26] (Table 3.11, columns 1-3). Previously we showed that this derived tooth pattern is partially explained by a large effect quantitative trait locus (QTL) on chromosome 21 that is associated with a late-acting *cis*-regulatory downregulation of *Bmp6* expression from benthic alleles in dental tissue [19] (Table 3.11, column 4). These results make *Bmp6* an excellent candidate gene for underlying evolved tooth gain by regulating tooth patterning and replacement. As no coding changes were found between marine and benthic freshwater alleles of *Bmp6* [19], we sought to map candidate regulatory regions of *Bmp6* associated with evolved tooth gain. Here, we use a combination of recombinant mapping, comparative genomics, genome editing, and transcriptional profiling to further dissect the molecular genetic basis of evolved tooth gain

and the role of *Bmp6* during tooth development in threespine sticklebacks.

3.3 Results

Recombinant mapping of chromosome 21 tooth number QTL identifies an 884 kb interval containing *Bmp6*

We previously identified and fine-mapped a large effect tooth number QTL to a 2.56 Mb 1.5-LOD interval on stickleback chromosome 21 containing an excellent candidate gene, *Bone Morphogenetic Protein 6* (*Bmp6*), along with 58 other predicted genes [19,32]. To further fine-map this QTL, we identified three chromosomes with marine-benthic recombination events within the 2.56 Mb fine-mapped interval (Fig 3.1). Fish with each of these recombinant chromosomes were crossed to fish heterozygous for marine and benthic alleles of chromosome 21 to generate large (>100 fish each) crosses to test these recombinant chromosomes for effects on tooth number (Fig 3.1A, Table 3.1). Recombinant chromosomes that increase tooth number compared to marine chromosomes suggest that the tooth controlling region of chromosome 21 lies within the benthic portion of the recombinant chromosome. We used a likelihood ratio test to determine whether each recombinant chromosome behaved more like a marine or benthic chromosome. Recombinant chromosomes one and three increased tooth number, each behaving like a benthic allele of chromosome 21 (P value from likelihood ratio test = 3.0 x 10^{-4} for both) (Fig 3.1B). Recombinant chromosome two did not increase tooth number, behaving like a marine allele of chromosome 21 ($P = 1$ x 10^{-3} from likelihood ratio test) (Fig 3.1B). Together, these recombinant crosses support a new smaller genetic interval, 884 kb in the stickleback reference genome assembly [24], that contains 21 predicted genes including *Bmp6* (Fig 3.1C), reducing the physical size of the interval and number of genes by 65% and 64%, respectively.

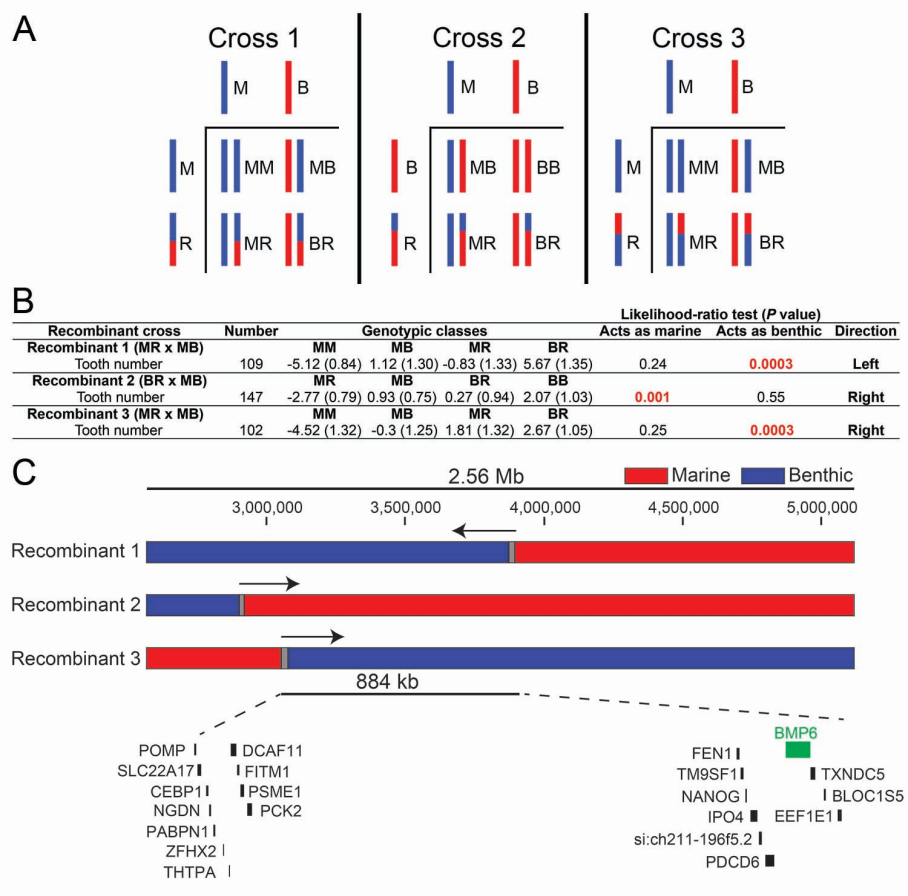

Figure 3.1: **Recombinant mapping of chromosome 21 tooth QTL supports an 884 kb interval containing *Bmp6*.** A). Schematic of three recombinant crosses. In each cross, fish heterozygous for recombinant and marine (Cross 1 and 3) or recombinant and benthic (Cross 2) alleles of chromosome 21 were crossed to fish heterozygous for non-recombinant marine and benthic alleles of chromosome 21. For each cross, cartoons of Punnett squares are shown, with haploid genotypes to the left and top and four classes of resulting diploid genotypes shown in the lower right. (B) Size-corrected total ventral pharyngeal tooth number and standard error are listed for each genotypic class within each of the recombinant crosses. For each cross, parental genotypes of the tooth QTL are listed and coded: marine (M), benthic (B), or recombinant (R). Likelihood ratio tests were used to test whether recombinant chromosome effects on tooth number behaved like a marine or benthic chromosome (see Methods). *P*-values from each likelihood ratio test are listed with the supported direction column in B. (C) The chromosome 21 tooth QTL was previously fine mapped to a 2.56 Mb region containing *Bmp6* along with 58 other Ensembl predicted genes [19]. The three recombinant chromosome 21s tested are shown. Genotypes are colored red for marine, blue for benthic, and grey for unresolved. Arrows denote position of tooth QTL supported by each recombinant chromosome. The final recombinant mapped interval is 884 kb in the reference genome assembly and contains 21 predicted genes, including *Bmp6*.

Cross	N=	Correction	Primer sequences left recombinant position	Primer sequences right recombinant position	Marker Type
1	109	Sex	CACTGAAGCCGG AGGAGAGG, ATC AGAGAGGGTCCA GAACG	AGTCCGCCACTTG TCTTTCC, GTCAT GCAGACCATGATT CC	Size Polymorphism
2	147	SL and Sex	TGAACCAATTGTT TGGAACATC, AAT CGCCATGTCAAAT TCCT	CCCGCAAGAAAG CAATTTAT, TTTG TTTCCTGCCTTCG AGT	Size Polymorphism
3	102	SL and Sex	ATCCAGCCCAGAG TGAAATG, GGCCT ACCAACTTGACCG TA	TGTGTGCAAACAC ACAGCAT,TCTGC TCTGCTTTGCTTC TTC	RFLP (*Ava*II left, *Sal*I right)

Table 3.1: **Summary of recombinant crosses.** Sample sizes of the Paxton benthic x Little Campbell marine recombinant crosses are shown along with the primer sequences for left and right genotyping markers used as the boundaries for the recombination breakpoint. The markers for recombinant 1 and 2 are size polymorphisms and the markers for recombinant 3 are restriction fragment length polymorphisms (RFLPs) using the restriction nuclease shown. Standard fish length (SL) and sex were corrected for when appropriate and corrections performed for each cross are listed. For each left and right marker, the left and right positions in base pairs, respectively, on chromosome 21 in the stickleback genome assembly [24] are listed.

Seven out of eight derived benthic chromosomes have a large effect tooth QTL

To estimate the frequency of the chromosome 21 high tooth number allele within the wild Paxton benthic population, we generated six marine by benthic F_2 crosses testing eight wild-derived benthic chromosomes (named B_1-B_8, Fig 3.2, Table 3.2). These chromosomes had different genotypes at three microsatellite loci located 5′, within, and 3′ of the chromosome 21 tooth QTL, suggesting they are molecularly distinct wild chromosomes (Table 3.2, see Methods). We found that seven of these eight benthic chromosomes had significant effects on tooth number with the same direction and similar magnitude of effect (Fig 3.2, Table 3.2). The benthic chromosome tested in cross 6 (B_8) had no effects on tooth number (Fig 3.2, Table 3.2). These results together suggest that the high tooth number allele on chromosome 21 is at high frequency in the Paxton benthic population, but at least one lower-frequency benthic allele is not associated with an increase in tooth number.

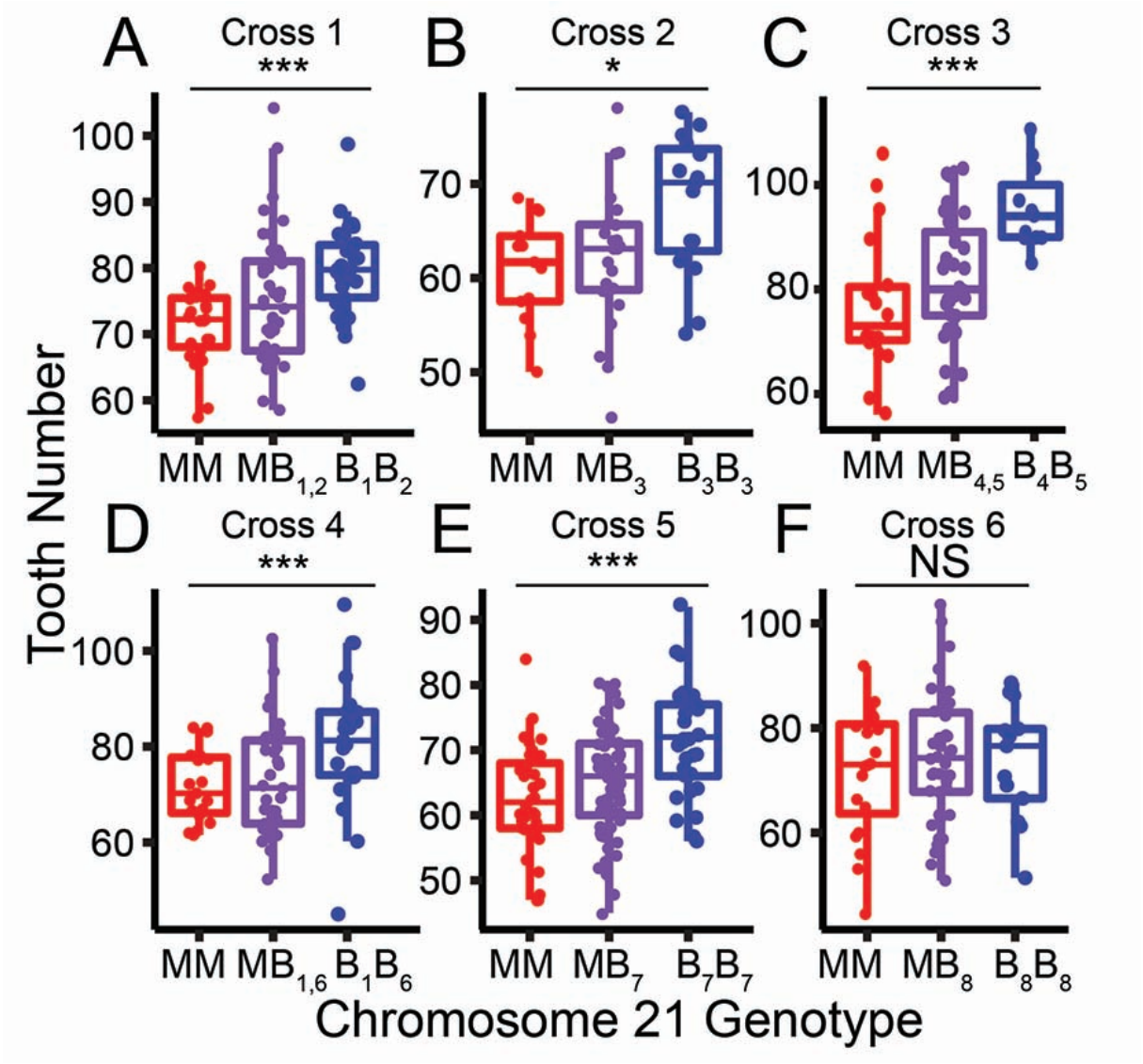

Figure 3.2: **Seven of eight benthic chromosome 21s have a tooth QTL.** (A-F) Results from six benthic by marine F_2 crosses testing eight molecularly distinct (see Methods) benthic chromosome 21s (B1-8) are shown. (A-E) Benthic chromosomes 1-7 had strong effects on tooth number; however (F) the benthic chromosome 8 had no detectable effects on tooth number. Back-transformed total tooth numbers from marine homozygous (red), heterozygous (purple), and benthic homozygous (blue) fish for chromosome 21 are shown (see Methods). *P* values from an ANOVA for cross 1-6 are 0.002, 0.024, 0.0005, 0.004, 2.11×10^{-5}, 0.69, respectively (* = P <0.05, *** = P <0.01). F_2 crosses 1, 3, and 4 are testing two benthic chromosomes each and crosses 2, 5, and 6 each are testing one. Crosses 1 and 4 share a benthic chromosome. See Table 3.2 for more details.

Cross	Populations crossed (male x female)	n F_2s	Marker	MM	MB	BB	Chromosomes tested	ANOVA *P*-values	LRT *P*-values
1	PAXB x JAMA	92	CM1440	71 (1.2)	75 (1.6)	80 (1.3)	B_1 and B_2	**0.002**	**0.01, 0.03**
2	PAXB x JAMA	51	CM1440	61 (1.6)	63 (1.6)	68 (1.9)	B_3	**0.024**	NA
3	PAXB x RABS	62	Stn489	77 (3.1)	82 (2.1)	96 (2.4)	B_4 and B_5	**0.0005**	**0.0005 0.02**
4	JAMA x PAXB	77	Stn489	72 (1.7)	73 (2)	81 (2.9)	B_1 and B_6	**0.004**	**0.04, 0.04**
5	PAXB x LITC	138	Stn487	62 (1.4)	65 (0.9)	72 (1.5)	B_7	**$2.1 * 10^{-5}$**	NA
6	PAXB x LITC	75	Stn487	72 (3.1)	74 (1.9)	74 (2.7)	B_8	.69	NA

Table 3.2: **Benthic x marine F_2 cross summary.** Results from marine by benthic F_2 crosses testing eight benthic chromosomes. Populations, number of F_2 fish, and chromosome 21 marker genotyped for each cross are listed. PAXB = Paxton benthic, JAMA = Japanese marine, RABS = Rabbit Slough marine, LITC = Little Campbell river marine. Sex of each grandparent is indicated in Populations crossed (male x female) column. For each cross, the most informative and completely genotyped marker nearest to the previously reported QTL peak [19] is listed. Standard length effects on total ventral pharyngeal tooth number were corrected for when appropriate and residuals were back-transformed to the mean standard fish length within each cross. Mean and standard error of corrected tooth number are shown for marine homozygotes (MM), heterozygotes (MB), and benthic homozygotes (BB). All PAXB grandparents were different fish, except the grandparent of crosses 5 and 6, which was the same PAXB male fish. The eight different molecularly distinct benthic chromosomes (see Figure 2) are listed in Benthic chromosomes tested column. Crosses 1, 3, and 4 tested two distinct benthic chromosomes and crosses 2, 5, and 6 tested a single benthic chromosome. Crosses 1 and 4 share a benthic chromosome with the same microsatellite genotypes (see Methods). P values from ANOVAs for testing whether genotype significantly effects tooth number phenotype are listed (see Fig 3.2). The last column shows P values from two likelihood ratio (LR) tests comparing the additive model to no effect benthic 1 model and no effect benthic 2 model are shown. The four allele marker used for crosses 1, 3, and 4 were CM1440, Stn223, and CM1440, respectively. The LR tests show that both benthic chromosomes have significant effects on tooth number in crosses 1,3, and 4. F_2 crosses 2, 5, and 6 contain the same benthic chromosome and thus, in these crosses the benthic chromosomes can not be tested individually (since they can not be molecularly distinguished).

Whole genome resequencing reveals a cluster of QTL-associated variants in intron 4 of *Bmp6*

We hypothesized that the Paxton benthic chromosome 21 alleles that increase tooth number ($B_{1\text{-}7}$, Fig 3.2) share sequence variants that underlie evolved tooth gain that are not present on marine alleles or the benthic chromosome 21 allele without the tooth QTL (B_8, Fig 3.2). To test for QTL-associated variants, we resequenced the genomes of the four benthic grandparents from crosses 1-4, two F_2 fish homozygous for chromosomes B_7 and B_8, and the three marine grandparents from crosses 2, 5, and 6 tested in Fig 3.2 (Table 3.3). We identified 372 sequence variants (consisting of 323 SNPs, and 49 indels) within the 884 kb fine-mapped genetic interval that were present on all the benthic chromosomes with a large effect QTL, but not present on marine chromosomes (Fig 3.3A). We gave variants a QTL concordance score: the absolute value of the proportion of times a variant allele is found in the benthic fish with a chromosome 21 tooth QTL minus the proportion of times the same allele was found in fish without a tooth QTL. Only ten of these variants (all SNPs) were perfectly associated with the presence of the tooth QTL (Fig 3.3). Strikingly, all of these variants lie within a 4.4 kb region of *Bmp6* intron 4 (Fig 3.3B, Table 3.4).

Fish	Library Prep Kit	Total Reads	Final Mapped Reads	Est. Coverage
Cross 1 - PAXB Grandparent	NEXTERA XT DNA sample preparation kit	45084360	39022926	8
Cross 2 - PAXB Grandparent	NEXTERA XT DNA sample preparation kit	39144478	33907774	7
Cross 2 - JAMA Grandparent	NEXTERA DNA sample preparation kit	59900872	43098831	9
Cross 3 - PAXB Grandparent	NEXTERA XT DNA sample preparation kit	42228302	26568000	6
Cross 4 - PAXB Grandparent	NEXTERA XT DNA sample preparation kit	38992384	32095135	7
Cross 5 - PAXB F_2	Epicenter NEXTERA	482809124	323700364	70
Cross 5 - LITC Grandparent	NEXTERA DNA sample preparation kit	70365294	58516503	13
Cross 6 - LITC Grandparent	NEXTERA DNA sample preparation kit	69017068	58385657	13
Cross 6 - PAXB F_2	Epicenter NEXTERA	450163966	326212116	71

Table 3.3: **Summary of genome resequencing.** For each fish used for genome resequencing, library preparation kit, total reads, final mapped reads, and estimated coverage are listed. LITC, JAMA, and PAXB refer to the Little Campbell Marine, Japanese Marine, and Paxton Benthic populations, respectively. Estimated coverage was calculated by dividing the final mapped reads by the stickleback genome size for each sample. The two high coverage genomes (>70x) were each sequenced in a full lane and the lower coverage genomes were barcoded and multiplexed with five other fish per sequencing lane. All sequencing was 100 bp paired-end on Illumina HiSeq2000.

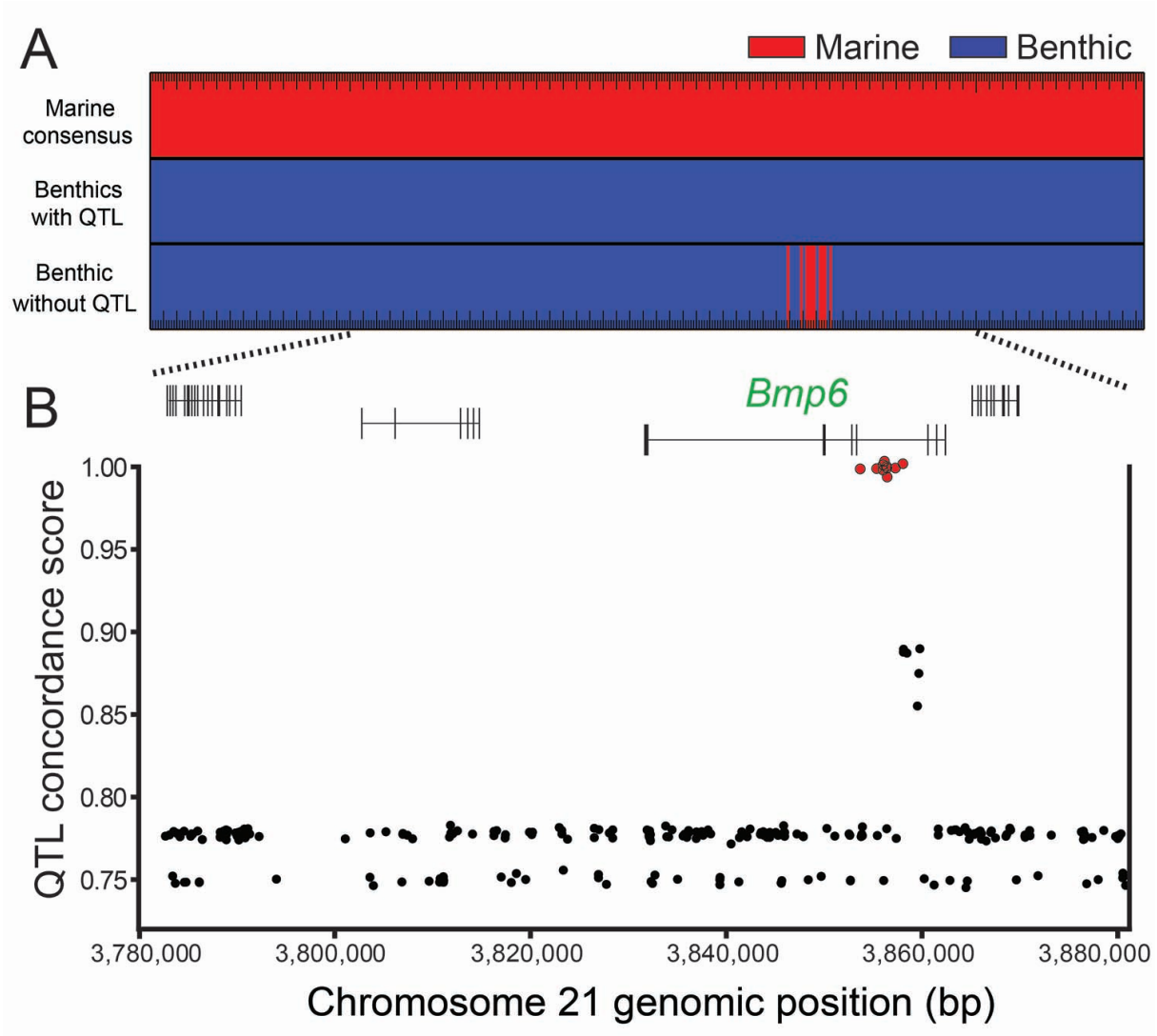

Figure 3.3: **Comparative genomics reveal QTL-associated variants in intron 4 of *Bmp6*.** (A) Comparing genomic sequences of the fine-mapped tooth QTL (from Fig 1C) between marine (n =3, from crosses 2, 5, and 6) and benthic chromosomes with the tooth QTL (n = 7, from crosses 1-5) identified a set of variants with opposite homozygous genotypes, colored red for marine (top) and blue for benthic (middle). Note that only positions with opposite homozygous genotypes within this 884 kb are shown. The benthic chromosome without the QTL (chromosome B_8 from cross 6) had a cluster of variants sharing the consensus marine genotype (bottom). (B) The ten variants with perfect QTL association (red points) all lie within intron 4 of *Bmp6*. The y-axis shows QTL concordance score (see Methods), a metric of concordance between genotype and presence or absence of tooth QTL. Gene model of *Bmp6* and surrounding genes are based on Ensembl predictions [24].

Chr. 21 position	Reference	QTL-associated variant
3853687	G	C
3855372	A	C
3856007	T	A
3856021	C	T
3856164	G	A
3856390	T	A
3856434	G	T
3856444	A	G
3857276	C	T
3858044	C	T

Table 3.4: **QTL-associated variants.** Chr. 21 position indicates position on chromosome 21 in stickleback reference genome assembly. Reference lists genotype at that position in reference genome assembly [24], while QTL-associated variant indicates genotype at that position of variants concordant with presence or absence of tooth QTL (see Fig 3.3).

QTL-associated variants surround a tooth and fin enhancer in intron 4 of *Bmp6* that drives overlapping and distinct expression patterns as the *Bmp6* 5′ enhancer

We previously showed that a *cis*-regulatory decrease in expression of *Bmp6* is associated with the chromosome 21 tooth QTL in Paxton benthic fish, suggesting that changes to *Bmp6* regulatory elements underlie the tooth QTL [19]. We hypothesized that the region of intron 4 containing tooth QTL specific variants is a tooth enhancer of *Bmp6* (Fig 3.3B). To test for enhancer function, we cloned a 2 kb intron 4 genomic fragment from marine fish into a reporter construct (Table 3.5). Transgenic fish for this construct expressed GFP in the distal tips of developing pectoral and median fins at eight days post fertilization (dpf), and pharyngeal and oral teeth at 10 dpf (Fig 3.4). These domains have been previously shown to be endogenous sites of *Bmp6* expression in developing sticklebacks [19,33]. These results demonstrate that the fourth intron of *Bmp6* contains an enhancer active in developing teeth and fins.

To define the minimally sufficient enhancer, we subcloned the 2 kb fragment into two smaller fragments of 1.3 kb and 511 bp based on patterns of sequence conservation (Fig 3.5A, Table 3.5), and tested for enhancer function in marine stickleback fish. The 511 bp construct is highly conserved in fish and contains no QTL-specific variants. The 1.3 kb construct includes the 511 bp region and a less conserved region that contains six of the 10 QTL-specific variants. The 800 bp included in the 1.3 kb construct but not the 511 bp construct drove no consistent expression, and no convincing differences were observed either

between the 1.3 kb construct and the 511 bp construct, or marine and benthic versions of the 1.3 kb construct at early embryonic and larval stages [n >3 injection rounds each, n >20 GFP+ lenses (the internal control domain driven by the *Hsp70l* promoter) for both early embryonic and early larval comparisons]. Both the larger 1.3 kb construct and the 511 bp construct drove expression in the distal edges of the median and pectoral fins at eight dpf (Fig 3.5B). By 13 dpf, the 511 bp enhancer drove expression in mesenchymal cells in developing pharyngeal teeth, as well as expression in the tooth epithelium (Fig 3.5C). In developing teeth, the GFP-positive mesenchymal domain extended from each tooth germ deep into the tooth plate (Fig 3.5C). This tooth expression continued into late juvenile stages when the pharyngeal tooth number differences arise between marine and freshwater populations (Fig 3.5D) [19]. GFP expression was also detected in late juvenile oral teeth (Fig 3.5E). These results demonstrate that the intron 4 tooth QTL-associated variants surround an enhancer sufficient to drive expression in developing fins and teeth.

We previously identified a TGFβ-responsive 5′ enhancer of *Bmp6* that also drives expression in developing teeth and fins in sticklebacks [33]. Because stickleback *Bmp6* expression is spatially and temporally complex in developing teeth [19], we hypothesized that the two regulatory elements may control distinct aspects of *Bmp6* expression in teeth. To test this hypothesis, we compared GFP expression patterns in fish stably transgenic for reporter genes for the 190 bp 5′ tooth enhancer or the 511 bp intron 4 tooth enhancer (Fig 3.5C,F-G). As previously described [33], we found that the 5′ enhancer drives robust expression in developing tooth epithelium and adjacent tooth mesenchyme (Fig 3.5G). We found that the intron 4 enhancer drove expression that appeared distinct from the 5′ enhancer at some stages of tooth development (Fig 3.5C,G). The intronic enhancer drove expression in the mesenchymal cores of mature teeth similar to the expression driven by the 5′ enhancer. However, the intronic enhancer drove deeper mesenchymal expression around the base of the developing tooth compared to the 5′ enhancer (Fig 3.5C,G).

To directly compare the tooth expression domains driven by the two enhancers, we generated fish transgenic for both a 511 bp intron 4 enhancer mCherry reporter construct as well as a 190 bp 5′ tooth enhancer [33] GFP reporter construct (Fig 3.5H). The tooth expression domains were partially overlapping between the two enhancers in developing teeth (Fig 3.5I-J). As was seen comparing the stable lines, both enhancers drive similar mesenchymal expression at early stages of tooth development, but the 5′ enhancer and not the intron 4 enhancer drove strong epithelial expression at these stages (Fig 3.5I, I′). As tooth development progresses, the intron 4 enhancer drove expression at the base of the mineralized tooth, in mesenchymal cells that did not express the 5′ enhancer (Fig 3.5J, J′). These results suggest that *Bmp6* expression in tooth epithelial and mesenchymal cells is driven by at least two enhancers that drive partially overlapping yet distinct expression patterns.

Construct	Fw primer	Rv Primer	Enzyme
2 kb enhancer GFP	GCCGGCTAGCAC CGACACAGCTGTA CTTGG	GCCGGCTAGCAG AGTCCTGATGGC CTCTCC	*Nhe*I
1.3 kb enhancer GFP	GCCGGCTAGCGA GAGCATCCGTCTT GTGGG	GCCGGCTAGCAG AGTCCTGATGGC CTCTCC	*Nhe*I
511 bp enhancer GFP	GCCGGCTAGCGT GTGTGCGCGGTG GAAAATG	GCCGGCTAGCAG AGTCCTGATGGC CTCTCC	*Nhe*I
511 bp enhancer mCh	GCCGGCTAGCGT GTGTGCGCGGTG GAAAATG	GCCGGGATCCAG AGTCCTGATGGC CTCTCC	*Nhe*I and *BamH*I

Table 3.5: **Reporter construct cloning primers.** Sequences of forward and reverse primers used to clone each construct are listed 5′ to 3′ along with the restriction enzyme used to digest the PCR amplicon. The orientation of the inserts in the GFP constructs were tested in the minus direction relative to the promoter since the endogenous enhancer is 3′ to the *Bmp6* promoter in the stickleback genome. The mCherry construct was cloned in the plus orientation to mirror the orientation for the 5′ tooth enhancer transgenic line used in the co-labeling experiment (see Fig 3.5).

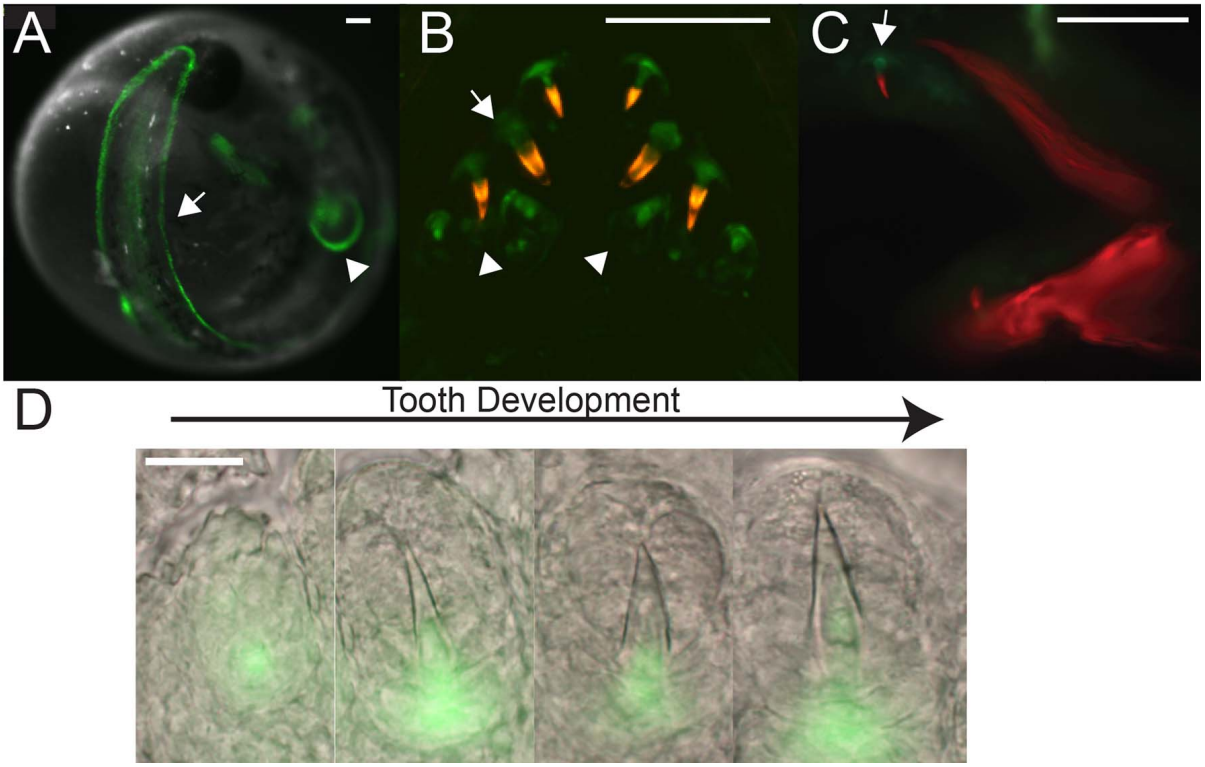

Figure 3.4: **2 kb intron 4 region is an enhancer active in developing fins and teeth.** (A) The marine 2 kb intronic enhancer drove expression at 8 dpf in the distal edges of the developing median fin (arrow) and pectoral fin (arrowhead). (B-C) By 10 dpf, the enhancer drove GFP expression in tooth mesenchyme (arrow) and diffusely in the tooth epithelium (arrowheads) in pharyngeal (B) jaws. GFP expression was also detected in developing tooth germs (arrow) in the oral (C) jaws. In B-C, bone is counterstained with red fluorescence by Alizarin Red. B is a dorsal view of the dissected ventral pharyngeal jaw, while C is a lateral view with anterior to the left of the upper jaw (premaxilla, top) and lower jaw (dentary, bottom). (D) This 2 kb enhancer controlled dynamic expression throughout development, becoming more restricted to the mesenchyme as the tooth matures. Scale bars are 100 μm (A-C) and 50 μm (D).

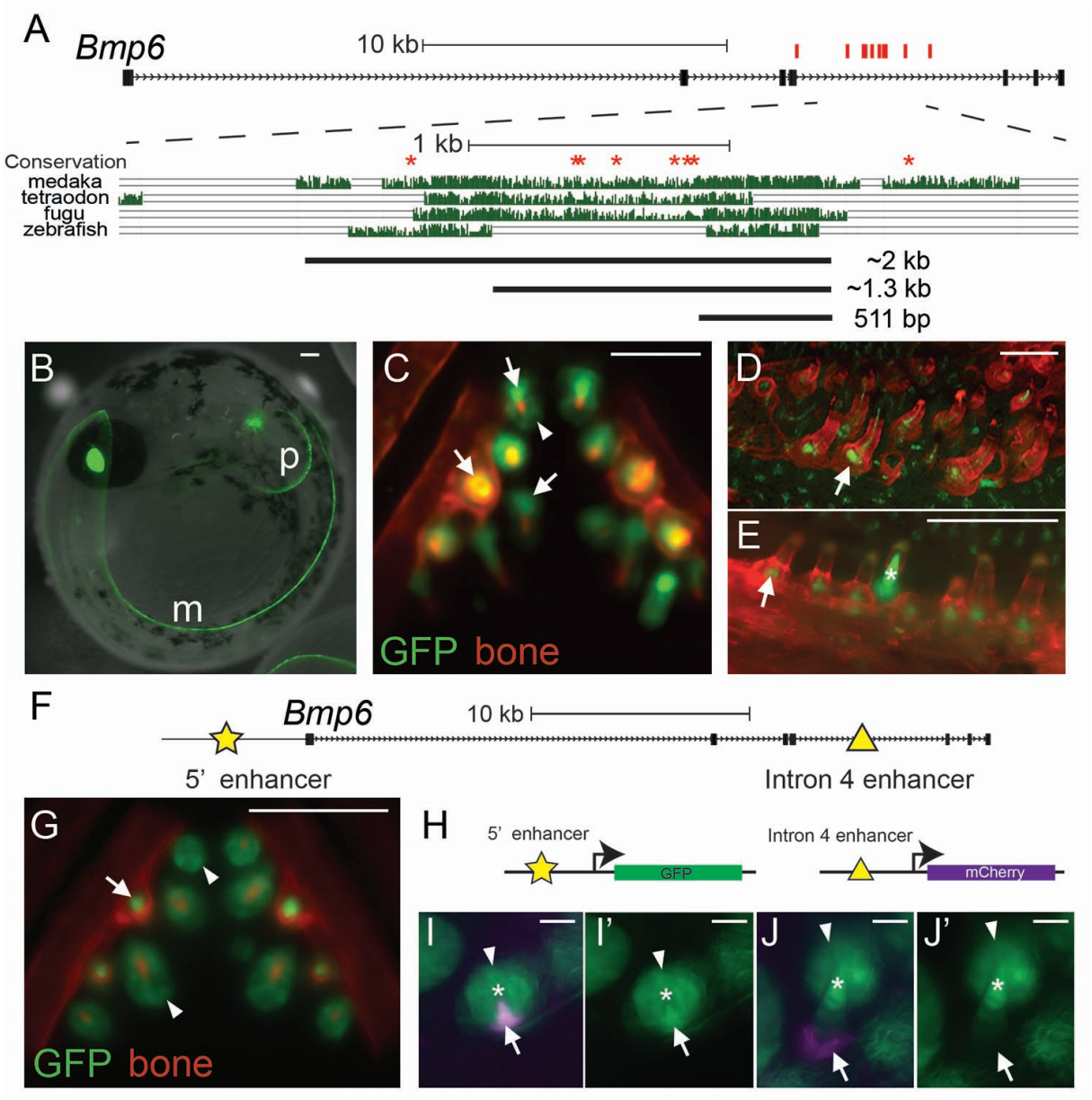

Figure 3.5: **Intron 4 region with QTL-associated variants contains a tooth and fin enhancer.** Caption on following page

Figure 3.5: **Intron 4 region with QTL-associated variants contains a tooth and fin enhancer.** (A) Schematic of *Bmp6* locus. All ten QTL-associated variants (red ticks) are located within intron 4. Eight of these variants (red asterisks) are in conserved sequence, expanded below. Conservation in teleosts is shown from the UCSC genome browser (http://genome.ucsc.edu/). Black bars show the 2 kb, 1.3 kb, and the 511 bp enhancer subclones tested. (B-E) GFP reporter expression from the 511 bp enhancer in stable transgenic fish. (B) At eight days post fertilization (dpf), expression was detected in the developing distal edge of the pectoral ("p) and median fin ("m). (C) By 13 dpf, relatively faint GFP expression was present in developing tooth epithelia (arrowhead) and stronger GFP expression was present in mesenchyme (arrows) of early stage and fully-formed teeth (see Figure S4 in [19] for time course of tooth epithelia and mesenchyme morphology in whole mounts). By late juvenile stages, mesenchymal expression was detected in developing pharyngeal (D) and oral (E) teeth. (F) The 5 (star) and intron 4 (triangle) tooth enhancers of *Bmp6* are shown. (G) The previously described 190 bp 5′ *Bmp6* tooth enhancer [33] drove overlapping but distinct expression than the intronic enhancer. Compared to the intronic enhancer (C), the 5 enhancer drove more persistent expression in tooth epithelial cells (arrowheads), and expression in tooth mesenchyme (arrow). (H) Fish doubly transgenic for the 190 bp 5′ enhancer driving GFP (green) and the 511 bp intron 4 enhancer driving mCherry (magenta) allow enhancer patterns to be directly compared. (I-I′) At early stages of tooth development, both enhancers drove mesenchymal expression (arrows), while the 5′ enhancer, but not the intron 4 enhancer, drove strong epithelial expression (arrowheads). (J, J′) As tooth development progresses, the intron 4 mesenchymal expression (arrow) extended to the base of the developing tooth, in cells not expressing the 5′ enhancer, while the 5′ enhancer continued to drive expression in both epithelial and mesenchymal cells. I and J show both GFP and mCherry channels overlaid, I′ and J′ show the GFP channels only. White asterisks in I, I′, J, and J′ mark mineralized teeth. Bone is counterstained with Alizarin Red in (C-E, G). Scale bars are 100 μm (B-G) and 10 μm (I-J).

Induced mutations in stickleback *Bmp6*

To test whether *Bmp6* is required for tooth patterning in sticklebacks, we used transcription activator-like effector nucleases (TALENs) to generate two predicted loss-of-function mutations in stickleback Bmp6 (Fig 3.6A, Table 3.6). We designed a TALEN pair to target the highly conserved second exon of *Bmp6*, which is 5′ to the exons encoding the predicted secreted ligand. Thus early stop codons would be predicted to generate strong loss-of-function alleles. Injection of these TALEN RNAs into stickleback embryos efficiently induced mutations in the *Bmp6* target sequence. To identify mutations in F_0 injected fish and mutations transmitted through the germline in F_1 fish, we PCR amplified the surrounding sequence around the target site, digested this amplicon with *EcoR*I, then gel extracted and sequenced the uncut band (Fig 3.7). We found that 24-57% of injected F_0 stickleback embryos had detectable deletions, with up to 12% of these embryos appearing to have biallelic mutations

(Fig 3.7). Consistent with previous studies using TALENs in fish, we identified a spectrum of insertions and deletions at the target site (Fig 3.6B) [34]. We generated two mutant alleles that we bred to test for phenotypes: (1) a 13 bp deletion, and (2) a 3 bp deletion plus 4 bp insertion (Fig 3.6B bold). Both of these mutations are predicted to produce frameshifts and an early stop codon 5′ to the secreted BMP ligand and thus are both likely strong loss-of-function alleles (Fig 3.8).

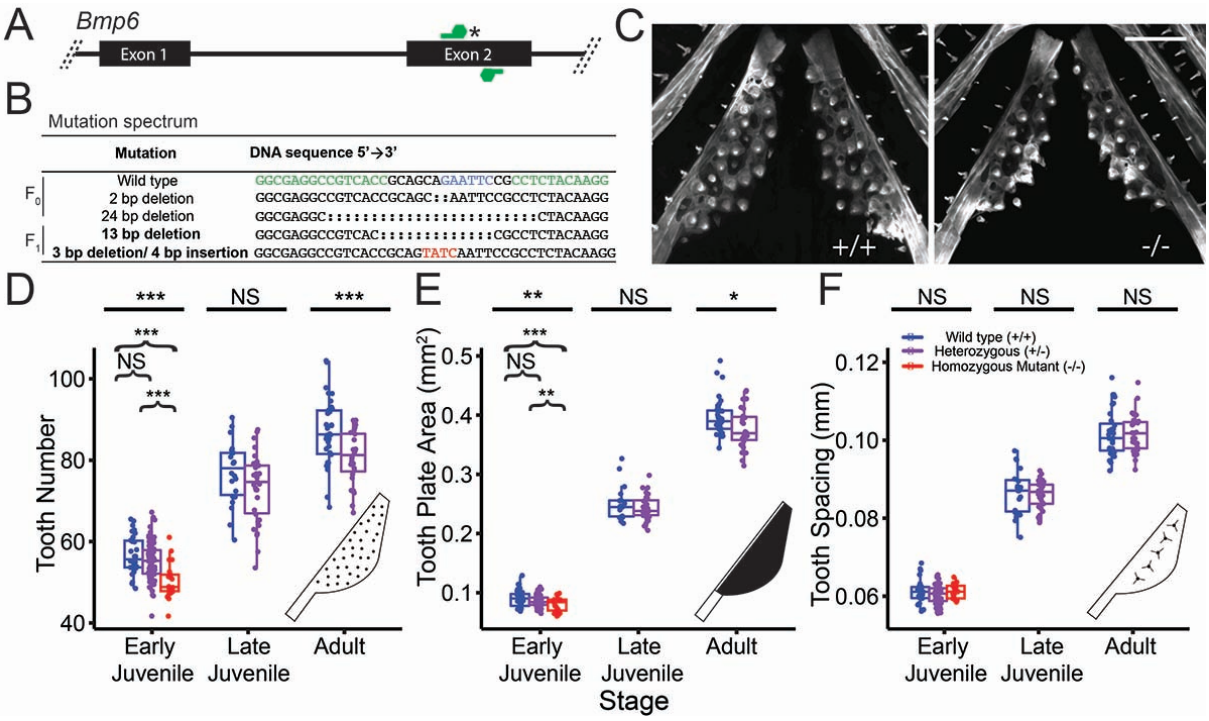

Figure 3.6: ***Bmp6* is required for viability, growth, and tooth patterning.** (A) Schematic of TALEN pair (green) targeting an *EcoR*I site (asterisk) in the second exon of *Bmp6*. (B) Sanger sequencing of F_0 or F_1 fish revealed a spectrum of genomic deletions (colons) and insertions (red) in *Bmp6*. The two mutations used in this study are in bold. In the wild-type sequence, the *EcoR*I site is shown in blue and the edges of the TALEN targeting sequences shown in green. (C) Confocal images of early juvenile (16-17 mm total length) wild-type (left) and homozygous mutant (right) ventral pharyngeal tooth plates showing fewer teeth in mutant. Mutant shown is transheterozygote for 13 bp deletion and 3 bp deletion+4 bp insertion. Scale bar is 200 μm. (D-F) Developmental time course of tooth number (D), tooth plate area (E), and tooth spacing (F) in wild-type (blue), heterozygous (purple), and homozygous mutant (red) fish. (D-E) Homozygous fish have recessive reduction of tooth number and tooth plate area at the early juvenile stage (Tukey post-hoc *P* values comparing wild-type to homozygous mutant are 9.3 $x10^{-6}$ and 0.004, respectively and comparing heterozygous to homozygous mutant are $1.3x10^{-4}$ and 0.08, respectively). Tooth number and area diverges late in development between wild-type and heterozygous fish. (F) Tooth spacing is not significantly different in the mutant at any stage. The late juvenile and adult crosses were heterozygous mutant backcrossed to wild-type fish. For D-F, homozygous mutants include both fish homozygous for the 13 bp deletion mutation and fish transheterozygous for the 13 bp deletion and the 3 bp deletion + 4 bp insertion (see Table 3.7).

Target	TAL1 (left), TAL2 (right) RVDs	TAL1,TAL2 Lengths	Spacer	Target Stickleback Sequence
Bmp6 exon 2	TAL1: HD HD HD NI NN NI NN NN NN HD NN NI NN NN HD HD NN NG HD TAL2: HD NI HD NI HD NI HD NG HD HD HG HG NN NG NI NN NI NN NN	19bp	17 bp	CCCAGAGGGCGA-GGCCGTC accgcagca**gaattc**cg CCTCTACAAGG-AGTGTGTG

Table 3.6: **Custom TALEN design and targets.** Repeat Variable Diresidues (RVDs) used to generate left (TAL1) and right (TAL2) nuclease pairs targeting the second exon of *Bmp6*, and stickleback target sequence is listed. Underlined nucleotides correspond to the 19bp TAL1 and TAL2 targets flanked by the 17bp spacer containing an *EcoR*I restriction site (bold).

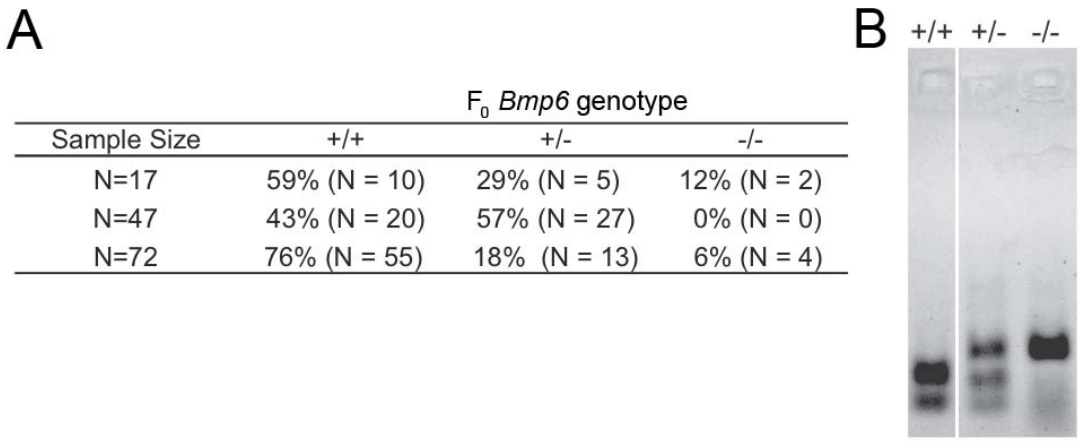

Figure 3.7: **Efficacy of *Bmp6* TALENs in stickleback embryos.** (A) Frequencies of wild-type (+/+), heterozygous (+/-), and homozygous (-/-) mutant F_0-injected three days post fertilization (dpf) embryos are shown for three independent injection rounds. (B) An *EcoR*I site was destroyed by induced mutations. Representative *EcoR*I digest assays on PCR amplicon from genomic DNA from a homozygous wild-type (left, +/+), heterozygous (middle, +/-), and homozygous mutant (right, -/-) injected embryo are shown.

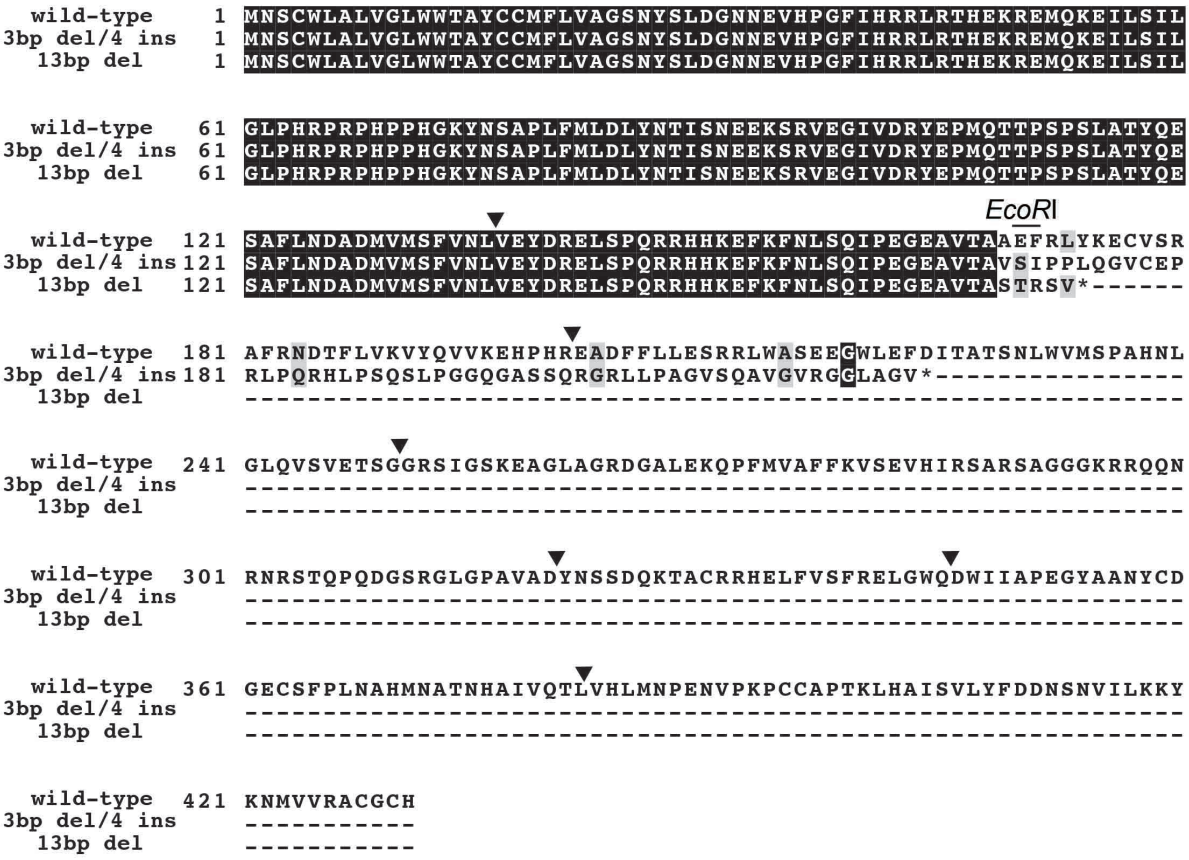

Figure 3.8: **Predicted amino acid alignments of the wild-type, 13bp deletion, and the 3 bp deletion/4 bp insertion alleles of BMP6.** Predicted mutant BMP6 sequences, 3bp deletion/4bp insertion (middle) and 13bp deletion (bottom), aligned to wild-type (top) BMP6 sequence. The 13bp deletion and the 3bp deletion + 4bp insertion generate frameshifts that result in premature stop codons (marked by asterisk) in the 2nd and 3rd exons, respectively, predicted to truncate the protein. Wild-type BMP6 sequences and intron/exon boundaries (marked with arrowheads) were previously described [19]. The position of the *EcoR*I site used as the genotyping assay is noted.

Bmp6 is required for viability, growth and tooth patterning

To test for tooth patterning phenotypes in *Bmp6* mutants, we intercrossed fish that were heterozygous for the 13 bp deletion or the 3 bp deletion plus 4 bp insertion and raised developmental time courses. Homozygous mutants were underrepresented from expected ratios at later developmental stages, suggesting early juvenile lethality (Table 3.7). The surviving homozygous mutants tended to be slightly smaller (Table 3.7). Because of the late stage lethality, we continued the *Bmp6* mutant time course with heterozygous backcrosses for late juvenile and adult stages. To test for required roles of *Bmp6* in tooth patterning,

we quantified ventral pharyngeal tooth number, tooth plate area (size of tooth field), and inter-tooth spacing, three phenotypes controlled by the chromosome 21 tooth QTL [19] in the *Bmp6* mutant time course (Fig 3.6C-F). At the early juvenile stage, homozygous mutants had a reduction of both tooth number and tooth plate area compared to wild-type or heterozygous fish (Fig 3.6D-E; Table 3.11, column 5). Beginning in early juveniles, heterozygous fish had fewer ventral teeth and smaller tooth plate area, which were both significantly more reduced at later time points including adults (Fig 3.6D-E; Table 3.11, column 6). There were no significant differences in inter-tooth spacing at any stage (Fig 5F). These results show that *Bmp6* is required for specifying tooth number and the size of the tooth field.

In addition to the bilateral ventral tooth plates, stickleback pharyngeal teeth are also present on two bilateral dorsal tooth plates, dorsal tooth plate 1 (DTP1) and dorsal tooth plate 2 (DTP2) [35]. We next asked whether *Bmp6* also regulates dorsal pharyngeal tooth number. We found no significant differences in tooth number of either dorsal tooth plate at early developmental stages (Fig 3.9). In adults, DTP2 tooth number was significantly lower in heterozygous mutants, but to a lesser degree than the ventral tooth number differences at the same stage (Fig 3.9). For both dorsal tooth plates, tooth numbers trended in the same direction as seen for the ventral tooth plates, with fewer teeth in mutants than wild types. These results demonstrate that, like the chromosome 21 tooth QTL [32], *Bmp6* dosage has stronger effects on ventral pharyngeal tooth number than dorsal pharyngeal tooth number. To test whether fish transheterozygous for both the 13 bp deletion and the 3 bp deletion/4 bp insertion have tooth patterning phenotypes, we generated a transheterozygote cross using the two different *Bmp6* mutant alleles. We found that fish transheterozygous for the two different mutations had similar tooth patterning phenotypes as fish homozygous for the 13 bp deletion (Table 3.8).

Cross	Genotype	Sample Size	Mortality P value	Mean SL	Stdev	Length P value
A	+/+	6	0.75	14.53	0.80	**0.04**
	+/-	16	-	14.59	0.55	-
	-/-	6	-	13.95	0.72	-
B	+/+	9	0.31	20.56	1.31	**0.04**
	+/-	27	-	20.26	1.44	-
	-/-	8	-	19.21	1.20	-
C	+/+	25	**.005**	25.47	4.63	**0.03**
	+/-	55	-	25.85	4.71	-
	-/-	9	-	22.15	2.60	-
D	+/+	32	0.35	37.61	3.07	0.52
	+/-	25	-	38.21	3.77	-
E	+/+	18	0.87	41.34	2.25	**0.05**
	+/-	17	-	39.38	3.15	-

Table 3.7: ***Bmp6* mutant class survival and fish length.** Sample sizes, mean total fish lengths, and standard deviations for crosses generating wild-type, heterozygous, and homozygous mutant fish (intercrosses, top) or wild-type and heterozygotes (backcrosses, bottom) are shown. Mortality P values from a Chi-square test expecting a 1:2:1 ratio for the intercrosses and a 1:1 ratio for the backcrosses are shown. There was significant deviation from expected 1:2:1 ratios (likely due to mortality) in intercross clutch C, where the fish were the largest. Length P values from an ANOVA are shown for a recessive model (Wild-type and heterozygous classes are merged and compared to the homozygous mutants). In all three intercrosses, homozygous mutant fish were smaller than their heterozygous and wild-type siblings. One of the backcross clutches had a significant size defect, which was not seen in the other clutch. Crosses A, C, D, and E contain the 13 bp deletion allele. Cross B is a transheterozygous cross between a fish heterozygous for the 13 bp deletion and a fish heterozygous for the 3bp deletion+4bp insertion (see Table 3.8).

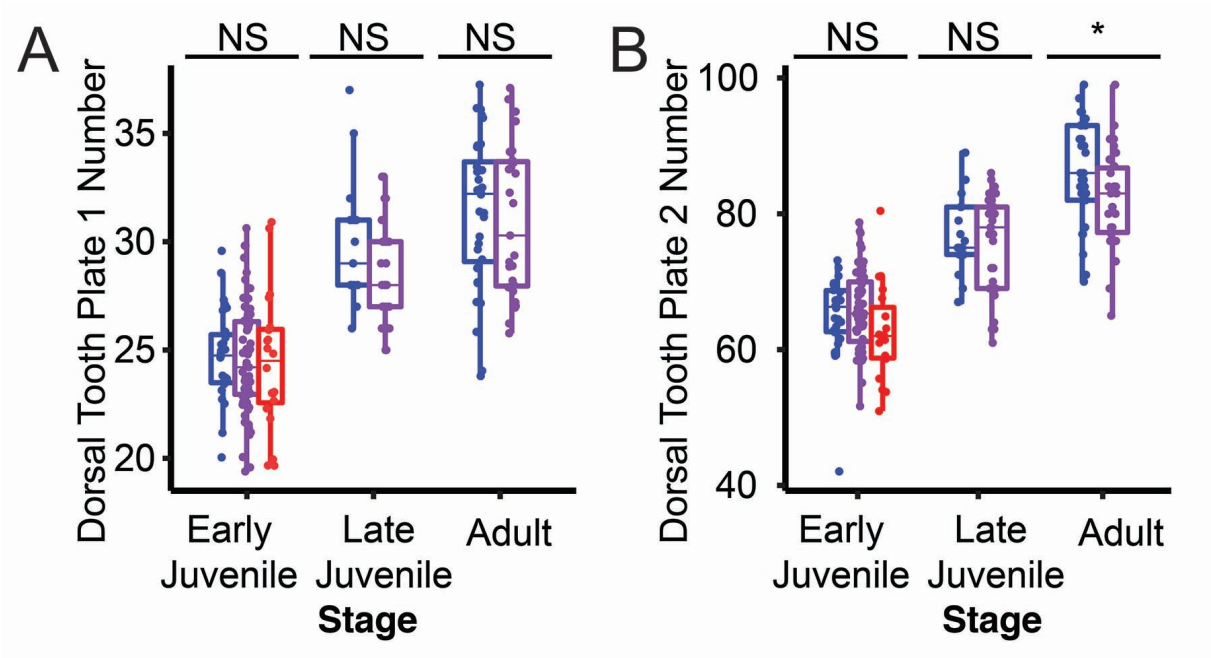

Figure 3.9: ***Bmp6* mutation effects on dorsal pharyngeal teeth.** (A) Size-corrected pharyngeal tooth number on dorsal tooth plate 1 (DTP1) were not significantly different between homozygous mutant (red), heterozygous (purple), and homozygous wild-type (blue) fish at any stage. (B) The dorsal tooth plate 2 (DTP2) tooth numbers were only significant at the adult stage (ANOVA $P = 0.028$) in contrast to the ventral pharyngeal teeth (VTP) results (see Fig 3.6)

Trait	Wild-type (WW, n = 9)	Het (WM, n = 27)	Mutant (MM, n = 8)	WW-WM	WM-MM	WW-MM
Ventral Tooth Number	54.83 (3.34)	54.46 (2.72)	49.73 (4.67)	0.954	**0.002**	**0.007**
Dorsal Tooth Plate 1 Number	24.35 (2.64)	25.4 (1.99)	24.14 (2.6)	0.451	0.355	0.980
Dorsal Tooth Plate 2 Number	66.57 (5.2)	67.5 (5.45)	64.47 (6.07)	0.902	0.378	0.714
Ventral Tooth Plate Area	0.08 (0.01)	0.08 (0.01)	0.07 (0.01)	0.680	**0.026**	**0.016**
Average Tooth Spacing	0.06 (0)	0.06 (0)	0.06 (0)	0.897	0.683	0.938

Table 3.8: **Transheterozygous effects on tooth patterning.** Analysis of a transheterozygous cross, 13bp deletion by the 3bp deletion/4bp insertion, for tooth patterning phenotypes. The effect of fish standard length was removed using a linear regression for each meristic or continuous trait. There were significant recessive differences in the homozygous mutant class for ventral tooth number and ventral tooth plate area, consistent with the results of the mutant time course. Continuous trait means and standard deviations (shown in brackets) for each genotypic class along with P values from a Tukey post-hoc test are shown.

Bmp6 regulates orthologs of BMP target genes, genes in the TGF-β signaling pathway, and genes upregulated in mouse hair follicle stem cells

To begin to identify the genetic networks downstream of *Bmp6*, we performed RNA-seq of early juvenile wild-type and 13 bp deletion homozygous mutant bilateral pharyngeal tooth plates (n = 3 of each genotype, Table 3.9). Following read mapping and gene expression quantification, we performed principal component analysis of normalized read count of the entire dataset (Fig 3.10A). PC1 explains a large fraction of the total variance (31.15%), and discriminates between the *Bmp6* homozygous wild-type and mutant samples (Fig 6A). Furthermore, genes whose expression correlated with the first principal component were

highly enriched for gene ontology terms related to development and cell signaling (Table 3.10).

To test whether stickleback *Bmp6* regulates BMP target genes found in other systems, we compared the genes that were differentially expressed between wild-types and mutants to three different data sets, two from ToothCODE [8] and the third from a microarray study [36]. By combining literature mining of published mouse tooth development studies as well as their own functional analyses, the ToothCODE project collected a list of target genes downstream of BMP signaling in developing tooth epithelium or mesenchyme [8]. We tested whether stickleback orthologs of these epithelial and mesenchymal BMP target gene sets were differentially affected in *Bmp6* mutant tooth plate tissue. Orthologs of mesenchymal BMP target genes as a whole displayed significantly reduced expression in *Bmp6* mutants ($P = 1.25 \times 10^{-2}$), while orthologs of epithelial BMP target genes were not significantly affected (Fig 3.10B). A third set of BMP signaling target genes was identified in a meta-analysis of published microarray studies [36]. We next asked whether stickleback orthologs of this gene set were significantly downregulated in Bmp6 mutant tooth plate tissue. We found this set of orthologs was significantly downregulated in Bmp6 mutants ($P = 3.12 \times 10^{-4}$), with 15/17 displaying a lower mean expression (Fig 3.10B). These results show that stickleback *Bmp6* is required to regulate a conserved battery of BMP-responsive genes.

We hypothesized that the *Bmp6* tooth number phenotype may result from changes in major signaling pathways known to be involved in tooth development [6,7]. The ToothCODE project manually curated a list of genes involved in tooth development in eight major signaling pathways (BMP, FGF, SHH, Wnt, Activin, TGF-β, Notch, and EDA) important for tooth development in mice [8]. We asked whether stickleback genes annotated as being in each of these pathways were concertedly differentially expressed in *Bmp6* mutants compared to wild types. We found the TGF-β signaling pathway to be significantly downregulated ($P = 4.7 \times 10^{-3}$) in *Bmp6* mutant tooth plates (Fig 3.10C). Strikingly, all eight TGF-β components tested had reduced mean expression in *Bmp6* mutant tooth plates (Fig 3.10D). In contrast, none of the other seven signaling pathways had significant expression differences (Fig 3.10C), despite the differences in tooth number in *Bmp6* mutants. Together these data suggest that *Bmp6* positively regulates TGF-β signaling in stickleback tooth plate tissue.

In polyphyodont sharks, fish, reptiles, and mammals, *Sox2* has been implicated in putative epithelial stem cells during tooth replacement [3739]. We found no significant differences in *Sox2* expression between *Bmp6* wild-type and mutant fish [mean FPKMs (Fragments Per Kilobase of transcript per Million mapped reads) of 91 and 97, respectively]. In mice, *Bmp6* inhibits the proliferation of hair follicle stem cells [40,41]. Teeth and hair are epithelial appendages with deep developmental and genetic homology [4245]. Thus, we hypothesized that *Bmp6* may play a conserved role of mediating stem cell quiescence during tooth replacement. A previous study characterized a set of hair follicle stem cell signature genes that are upregulated in the stem cell niche in the mouse hair follicle relative to the proliferating hair germ [46]. *Bmp6* mutants showed a highly significant ($P = 8.5 \times 10^{-12}$) decrease in the expression of stickleback orthologs of these genes (Fig 3.10E-F). The reduced expression of the orthologs of these hair follicle stem cell signature genes supports the hypothesis that *Bmp6* regulates

stem cell quiescence during tooth replacement.

Fish	Total Reads	Final Mapped Reads	SL (mm)
Bmp6 +/+ 1	49144984	30534676	24.70
Bmp6 +/+ 2	53590124	41559304	29.01
Bmp6 +/+ 3	51516122	42258366	26.32
Bmp6 -/- 4	47897146	22799870	25.05
Bmp6 -/- 5	48728442	34295004	22.63
Bmp6 -/- 6	69383016	52728234	22.22

Table 3.9: **RNA sequencing summary statistics.** Total reads, mapped reads and fish standard length are listed for each wild-type (1-3) and mutant fish (4-6) used for sequencing. All libraries were made with the TruSeq Stranded mRNA Library Prep Kit, barcoded, multiplexed and 100 bp paired-end sequenced in a single lane of an Illumina HiSeq2000.

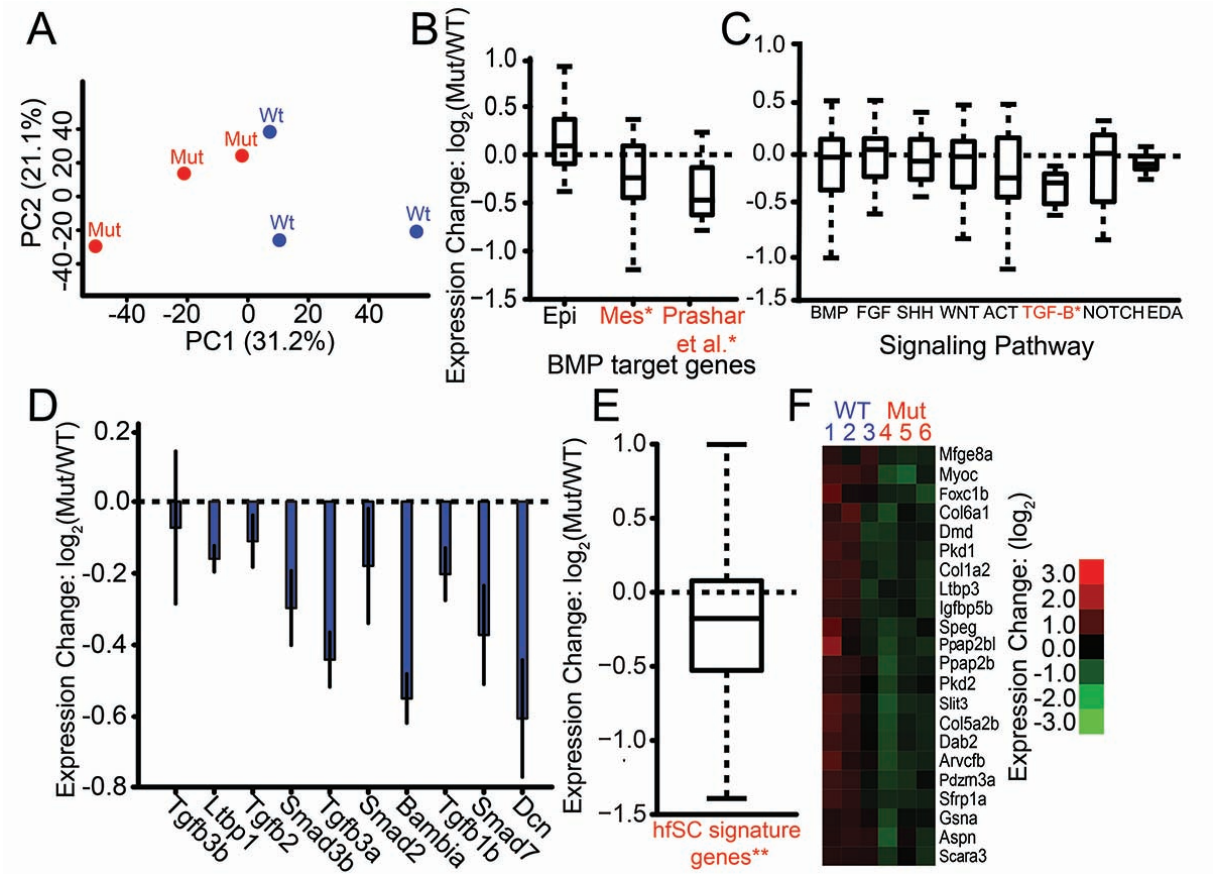

Figure 3.10: **Transcriptional profiling reveals TGF-β signaling components, BMP target genes, and hair follicle stem cell signature genes are downregulated in *Bmp6* mutant tooth plates.** (A) Principal component analysis of genome-wide expression levels in late juvenile ventral pharyngeal tooth plate tissue by RNA-seq separates wild-type (Wt, blue) and *Bmp6* mutants (Mut, red) along PC1. (B) BMP target genes during tooth development [8] were significantly downregulated in the mutant mesenchyme (P = 1.25 x 10^{-2}, middle bar), but not in the epithelium (left bar). A set of BMP target genes [36] was significantly downregulated in mutants (P = 3.12 x 10^{-3}, right bar). For B and C, gene sets with significant expression differences between wild-type and mutant are listed in red and with an asterisk. (C) Expression of ToothCODE signaling pathways. Homozygous mutant fish (Mut) had significantly lower TGF-β pathway expression compared to wild-type fish (WT) (P = 4.7 x 10^{-3}). None of the other pathways showed significant differences. (D) Each of the ToothCODE TGF-β genes was downregulated in the mutant. Error bars are SE of the mean. (E) A previously described set of genes upregulated in the mouse hair follicle stem cell niche [46] was downregulated in *Bmp6* mutants (P = 8.5 x 10^{-12}). hfSC = hair follicle stem cells (F) Hair follicle stem cell signature genes showing significant downregulated expression in *Bmp6* mutants.

GO Term	Description	FDR q-value
GO:0044421	extracellular region part	7.73E-12
GO:0005576	extracellular region	1.44E-11
GO:0005581	collagen trimer	8.85E-11
GO:0005615	extracellular space	1.03E-09
GO:0031012	extracellular matrix	2.58E-07
GO:0032502	developmental process	3.22E-07
GO:0044767	single-organism developmental process	6.04E-07
GO:0048646	anatomical structure formation involved in morphogenesis	6.57E-07
GO:0044707	single-multicellular organism process	8.78E-06
GO:0032501	multicellular organismal process	1.10E-05
GO:0098797	plasma membrane protein complex	1.90E-05
GO:0044459	plasma membrane part	2.25E-05
GO:0044724	single-organism carbohydrate catabolic process	2.48E-05
GO:0007155	cell adhesion	2.60E-05
GO:0016021	integral component of membrane	2.69E-05
GO:0044699	single-organism process	2.87E-05
GO:0022610	biological adhesion	2.97E-05
GO:0007166	cell surface receptor signaling pathway	3.75E-05
GO:0048856	anatomical structure development	4.04E-05
GO:0004872	receptor activity	4.57E-05
GO:0031224	intrinsic component of membrane	4.68E-05
GO:0016052	carbohydrate catabolic process	6.86E-05
GO:0001525	angiogenesis	7.27E-05

Table 3.10: **GO term enrichment of the first principal component of gene expression.** GO term enrichment for a list of genes ranked by expression correlation with the first principal component of the *Bmp6* wild-type and mutant tooth plate expression matrix. Shown are significant GO terms with an FDR q-value less than 1E-04.

Stage	Tooth number in wild fish	Tooth number in lab-reared fish	Bmp6 allele-specific expression in dental tissue	Tooth number in fish homozygous for induced Bmp6 mutations	Tooth number in fish heterozygous for induced Bmp6 mutations	Intron 4 enhancer spatiotemporal and/or quantitative differences
Larval/ early juvenile	?	not diff.	not diff.	-	not diff.	?
Late juvenile	?	+	-	no data (lethal)	not diff.	?
Adult	+	+	-	no data (lethal)	-	?

Table 3.11: **Summary of phenotypes seen at different stages in wild, lab-reared, and *Bmp6* mutant benthic fish.** Each row shows a developmental stage from a previous [19] or this study. Shown are the evolved benthic phenotypes relative to ancestral marine phenotypes (columns 2-4) and the mutant phenotype relative to wild-type phenotype (column 5-6). + = benthic phenotype was significantly higher than marine phenotype, not diff. = no significant differences were observed, - = benthic phenotype was significantly lower than marine phenotype. ? = unknown.

3.4 Discussion

Mapping an evolved tooth gain QTL to a *Bmp6* intronic enhancer

We previously identified a *cis*-regulatory downregulation of *Bmp6* associated with the chromosome 21 tooth QTL [19]. Because there are no reported coding changes between marine and freshwater benthic alleles of *Bmp6* in wild sticklebacks [19], regulatory changes that change the spatiotemporal pattern and/or the quantitative levels of *Bmp6* expression likely modulate natural variation in tooth patterning. Here we combined recombinant mapping and comparative genomics of multiple QTL crosses to fine-map this chromosome 21 tooth QTL to a haplotype within the fourth intron of *Bmp6*. The association of ten variants in intron 4 of *Bmp6* with the chromosome 21 tooth QTL, together with our data showing intron 4 contains a robust tooth enhancer, suggest a model in which these QTL-associated variants at least partially underlie the tooth QTL. Although all of the tooth QTL-associated

mutations are outside of the minimally sufficient 511 bp tooth enhancer, we propose that some or all of these variants underlie the *cis*-regulatory changes in *Bmp6*. One of the most outstanding questions for future research to address is whether these ten variants affect the spatiotemporal patterns and/or quantitative levels of enhancer activity (Table 3.11, column 7). Although comparing the 1.3 kb marine and benthic constructs has revealed no obvious differences at early embryonic and larval stages to date, several technical challenges including mosaicism in F_0s and position effects in stable lines make comparing two enhancers in different fish difficult. Due to the dynamic and complex expression patterns of the intronic enhancer, addressing potential marine/benthic enhancer differences would be facilitated by better tools to precisely compare enhancer activity, either at the same integration site using transgene landing pads, or in the same fish using bicistronic constructs separated by an insulator.

We note that the minimally sufficient 511 tooth enhancer contains a predicted FOXC1 binding site [47]. In mice, *Foxc1* regulates mammalian hair regeneration in part through regulating BMP signaling and appears to directly regulate *Bmp6* [48], so potential FoxC inputs into *Bmp6* expression in replacement teeth are especially intriguing. Of the marine/freshwater differences in the enhancer, one SNP alters a predicted NFATc1 binding site, a critical regulator of stem cell quiescence in the mouse hair follicle stem cell niche [49]. Another SNP affects a predicted GLI binding site, of interest because *Gli* expression is seen in multiple epithelial appendage stem cell niches in mice [50]. Future experiments will dissect what signals regulate this intronic enhancer, as well as what phenotypic consequences, if any, result from mutations in this enhancer.

This intronic enhancer, like the 5′ tooth enhancer [33], also drives embryonic and larval expression in developing pectoral and median fins. One interesting hypothesis these fin domains raise is whether evolved differences in median or pectoral fin morphology are also regulated by this derived intronic haplotype. Perhaps supporting this hypothesis, a QTL regulating median fin morphology (dorsal spine 3 length) was previously mapped to a broad region of chromosome 21 overlapping *Bmp6* [32]. Future experiments will also test whether the marine and freshwater versions of the enhancer have different activity.

Regulation of Bmp6 during tooth development and replacement

Our transgenic assays show that the intronic enhancer of *Bmp6* drives both overlapping and distinct domains of expression as the previously characterized 5′ Bmp6 enhancer [33]. Both enhancers drive overlapping expression in the mesenchymal cores of developing teeth. However, relative to the 5′ enhancer, the intronic enhancer also appears to drive deeper and broader mesenchymal expression and more restricted epithelial expression. These differences in expression patterns from the two enhancers suggest different signaling inputs control the mesenchymal and epithelial expression of *Bmp6* in developing teeth. Our finding that the 5′ and intronic *Bmp6* enhancers drive partially non-overlapping expression patterns is reminiscent of the mouse *Bmp5* gene, which has two rib enhancers that drive expression in largely complementary patterns [51]. A modular *cis*-regulatory architecture is likely a

common feature of *Bmp* genes, and could predispose these genes to frequently be used in morphological evolution [21,52-54].

Required roles for *Bmp6* in survival, growth, and tooth patterning

This QTL confers late-acting (juvenile stage, >5mm in fish length) increases in tooth number and tooth field size, and decreases in tooth spacing [19]. Here we generated fish with induced mutations in *Bmp6* to directly test whether *Bmp6* played any required role in regulating tooth patterning. Strikingly, fish heterozygous for induced mutations in *Bmp6* also had developmentally late differences in tooth number and tooth field size, similar to the tooth QTL (Table 3.11). A second phenotypic similarity between the tooth QTL and induced mutations in *Bmp6* is a stronger effect on ventral tooth number than dorsal tooth number [19,32]. However, the direction of the *cis*-regulatory allele, where the high-toothed allele drives reduced *Bmp6* expression in *cis* relative to a marine allele [19], would predict that a mutation that lowers *Bmp6* mRNA levels would increase tooth number, while the *Bmp6* coding mutants have fewer teeth (Table 3.11). One explanation for this unexpected direction of effect could be a threshold effect: the *Bmp6* mutations were made in a freshwater benthic genetic background with already reduced levels of *Bmp6* expression, and further lowering of *Bmp6* activity could inhibit tooth development. One test of this model could be to analyze the role of *Bmp6* during tooth development in marine sticklebacks, or in other freshwater populations lacking the benthic *Bmp6* intronic haplotype reported here. Alternatively, the induced mutant coding alleles of *Bmp6* might not recapitulate the evolved *cis*-regulatory differences between marine and freshwater fish. The dynamic expression of *Bmp6* in dental epithelium and mesenchyme at different stages of tooth development is controlled by at least two different *cis*-regulatory elements ([33]; this study), which we show here drive expression at some stages in non-overlapping patterns. The evolved *cis*-regulatory allele of *Bmp6* may change the spatiotemporal pattern and/or levels of *Bmp6* mRNA in different tissues, leading to different phenotypes than the coding mutations. Inducing loss-of-function mutations in the two known stickleback *Bmp6* enhancers and assessing potential changes in tooth patterning could test this hypothesis.

The *cis*-regulatory hypothesis proposes that morphological evolution typically proceeds through *cis*-regulatory mutations that avoid the negative pleiotropy typical of coding mutations [1,2,55]. Recent studies have shown that *cis*-regulatory and coding mutations can drive morphological evolution, and that the type of mutation may depend on the degree of pleiotropy of the gene of interest [18,19,56,57]. The lethality and smaller size of fish homozygous for *Bmp6* coding mutations could explain why *cis*-regulatory changes of *Bmp6* have been used to evolve increases in tooth number.

There were no significant differences in tooth pattern at early developmental stages between wild-type and heterozygous *Bmp6* mutant fish. However, as these heterozygous fish continued to develop to adult stages, when newly forming teeth are likely replacement teeth, the reduction of tooth number and tooth plate area became more dramatic, suggesting that tooth development at late stages is more sensitive to the dosage of *Bmp6*. These differences

could be due to different developmental or genetic constraints at the early juvenile and late adult stages of tooth patterning. For example, there could be more functional redundancy of *Bmp6* with other BMP ligands in teeth at early developmental stages that compensate in *Bmp6* heterozygous mutants. Alternatively, these differences may signify differing roles of *Bmp6* in primary and replacement tooth formation: later developing replacement teeth may be more sensitive to *Bmp6* dosage than primary teeth. However, homozygous mutants had significantly fewer teeth at early juvenile stages, suggesting *Bmp6* is also required for formation of primary teeth.

Downstream targets of *Bmp6* signaling in dental tissue

To test which genes and pathways are downstream of *Bmp6* signaling, we used RNA-seq to compare genome-wide transcriptional profiles of wild-type and homozygous mutant *Bmp6* tooth plates. Seven signaling pathways were not significantly different in this contrast, perhaps surprising given the predicted difference in total tooth number in these samples. However, we found that there is a concerted downregulation of the TGF-β signaling pathway components in homozygous mutants. TGF-βsignaling is required for tooth development [58-60]. Furthermore, TGF-β signaling regulates *Bmp6* expression in stickleback teeth through the previously described 5′ tooth enhancer [33]. These results suggest that TGF-β signaling is involved both upstream and downstream of *Bmp6* during tooth development.

During tooth development in mice, reciprocal signaling events involving *Bmp4* and *Msx1* occur between developing tooth epithelium and mesenchyme: *Bmp4* expression is first detected in dental epithelium, is required to induce *Msx1* expression in underlying mesenchyme, which in turn is required to induce *Bmp4* expression in dental mesenchyme [11,61-63]. Thus, *Bmp4* is thought to play critical roles during tooth development in both dental epithelium and mesenchyme. A large mouse gene expression study revealed sets of genes regulated by *Bmp2/4/7* in dental epithelium and mesenchyme [8]. We hypothesized that mouse BMPs and stickleback *Bmp6* regulate a conserved set of downstream genes in developing teeth. We tested this hypothesis by asking whether orthologs of known mouse BMP signaling target genes are differentially regulated in stickleback *Bmp6* mutant tooth plate tissue. Surprisingly, we found significantly reduced expression of the set of genes responsive to BMP signaling in mouse dental mesenchymal cells, while the set of genes responsive to BMP signaling in mouse dental epithelial cells was not significantly altered. Perhaps consistent with a relatively less effect on dental epithelia than mesenchyme in the *Bmp6* mutant, *Sox2*, implicated in epithelial stem cells during tooth replacement in other polyphyodonts [3739], was not significantly affected in *Bmp6* mutants.

Potential parallels between tooth and hair regeneration

In other vertebrates that undergo tooth replacement, dental stem cells have been proposed to mediate tooth replacement [37-39,64-66]. Teeth develop from placodes, transient epithelial thickenings that grow outwards or inwards to form epithelial appendages [42,43]. Teeth are

developmentally deeply homologous to other placode-derived organs, such as mammalian hair [44,45,67]. Mammalian hairs, like fish teeth, are constantly replaced throughout adult life. During mammalian hair regeneration, *Bmp6* regulates stem cell quiescence in the hair follicle stem cell niche [40,46]. Additionally, conditional knockout of the BMP receptor *Bmpr1a* in mouse hair follicles resulted in a loss of both hair regeneration and stem cell signature genes [46]. Thus, we hypothesized that stickleback *Bmp6* might regulate similar genetic pathways during tooth replacement as during hair regeneration. Supporting this hypothesis, in *Bmp6* mutant tooth plate tissue, we found a significant downregulation of mouse hair follicle stem cell signature genes, a set of genes previously described to be upregulated in mouse hair follicle stem cells compared to cells in the forming hair germ [46]. This result supports a model where modulating *Bmp6* expression in derived freshwater sticklebacks alters dental stem cell dynamics to result in the elevated tooth replacement rate seen in high-toothed freshwater sticklebacks [26]. Furthermore, this result suggests that the genetic circuitry regulating stem cell quiescence in continuously regenerating mammalian hair may be shared during constant tooth replacement in fish. This shared gene set might reflect an ancient highly conserved pathway regulating vertebrate epithelial appendage regeneration. If so, further identifying this core conserved gene regulatory network would provide profound insights into vertebrate development, regeneration, and evolution.

3.5 Methods

Ethics statement

All animal experiments (including euthanasia by immersion in a buffered 250 mg/L tricaine methane sulfonate solution) were done with the approval of the Institutional Animal Care and Use Committee from University of California, Berkeley (protocol R330).

Stickleback husbandry

Stickleback fish were raised in 29-gallon tanks in 1/10th ocean water (3.5 g/l Instant Ocean salt, 0.4 mL/l NaHCO3) and fed live brine shrimp as larvae, then frozen daphnia, bloodworms, and Mysis shrimp as juveniles and adults. All fish crosses were conducted using artificial fertilization.

Recombinant mapping

Further F_3-F_5 generations of a Paxton Benthic freshwater by Little Campbell marine F_2 cross [68] were propagated by intercrossing fish heterozygous for marine and benthic alleles of chromosome 21 (identified by heterozygosity at Stn487 and Stn489). Recombinant fish in F_4- F_5 generation were identified using microsatellite markers Stn487 and Stn489 which flank the genetic interval surrounding *Bmp6*. Caudal fin tissue was genotyped by first isolating

DNA by incubating for 20′ at 94C, then digesting with 2.5 μL of 20mg/ml proteinase K in lysis buffer (10mM Tris, pH 8.3 ; 50 mM KCL ; 1.5 mM MgCl2 ; 0.3% Tween-20 0.3% NP-40) for an hour at 55C followed by 20′ at 94C. One μl of undiluted DNA was used directly in the genotyping PCR. Once recombinant fish were identified, recombinant breakpoints were further mapped using a combination of microsatellite markers and restriction fragment length polymorphisms (RFLPs). Primer sequences for the left and right markers used to refine each recombinant chromosome used in this study are shown in Table 3.1. Gene content was determined by hand annotating the Ensembl predicted gene list.

Recombinant fish were crossed to F_4-F_5 fish heterozygous for marine and benthic chromosome 21 that were also derived from the same F_2 grandparents. The recombination events in crosses 1-3 were between markers Stn488 and Stn489 (cross 1), or between markers Stn487 and Stn488 (crosses 2 and 3). Genotypes of chromosome 21 in these three crosses were scored as M (marine), B (benthic), or R (recombinant) based upon the two locus genotypes of Stn488/Stn489 (cross 1) or Stn487/Stn488 (crosses 2 and 3).

Recombinant crosses were raised to 30 mm standard length. Fish were stained for bone with Alizarin Red, cleared, and pharyngeal teeth were quantified as previously described [19]. If tooth number was significantly correlated with standard fish length, sex, or family, we corrected for each using a linear model and used residuals from that regression for statistical analysis (Table 3.1). To test whether each recombinant chromosome contained the tooth number QTL, we performed a likelihood-ratio test comparing two models, one with the recombinant chromosome behaving as a benthic chromosome and one with the recombinant chromosome behaving as a marine chromosome.

Benthic by marine F_2 crosses

Lab-reared stocks of Paxton Benthic fish used for F_2 crosses were generated by incrossing wild-derived fish from Paxton Benthic lake, British Columbia. Five benthic fish were crossed to marine fish and F_1s subsequently incrossed to generate six F_2 crosses. The specifics of marine populations used in each cross are presented in Table 3.2. Three microsatellite markers spanning the chromosome 21 tooth QTL were genotyped: CM1440 (primer sequences 5 to 3: AAATGTGCTCCTGGATGTGC and CTTTCTCCTTCTGCCAAACG), Stn489, and Stn488; this set of genotypes was used to define molecularly distinct chromosome 21s. F_2 crosses 5 and 6 shared a benthic grandparent. This marker analysis suggests that there are eight molecularly distinct chromosome 21s in the five benthic grandparents.

To determine the effect of chromosome 21 on tooth number, the F_2 crosses were genotyped using microsatellites markers on chromosome 21 near the tooth QTL (see Table 3.2 for details). The effects of fish size on tooth number were removed by linear regression and the residuals were back-transformed to the mean standard fish length in each cross. Statistical association between chromosome 21 genotype and back-transformed phenotypes was tested using an ANOVA in R. To determine if both benthic chromosomes had an effect on tooth number in each cross, we performed a likelihood-ratio test for each wild benthic chromosome

comparing a model where that chromosome does not have an effect on tooth number to a model where both benthic chromosomes have an equal effect on tooth number.

Genome sequences of marine and benthic stickleback fish

We resequenced the genomes of the four benthic grandparents from crosses 1-4 and F_2 fish homozygous for chromosome B_7 and B_8. We also sequenced the marine Little Campbell grandparents from crosses 5-6, and the Japanese marine grandparent from cross 3 (Fig 2). Caudal fins were digested overnight at 55C in Tail Digestion Buffer (10 mM Tris, pH 8.0, 100 mM NaCl, 10 mM EDTA, pH 8.0, 0.5% SDS, 10 µl of 20mg/ml proteinase K). Genomic DNA was purified with a phenol:choloroform extraction followed by ethanol precipitation. Genomic libraries were generated using the Nextera DNA Sample Prep Kit (Epicentre Biotechologies), the Nextra DNA Sample Preparation Kit (Illumina), or the Nextera XT DNA Library Preparation Kit (Illumina). Paired-end reads (100 bp) were sequenced using an Illumina HiSeq2000. See Table 3.3 for details of library preparation and sequencing summary for each library.

Variant calling and tooth specific variant identification

Resulting reads were aligned to the repeat masked verision of the reference stickleback genome [24] using the bwa aln and bwa sampe modules of the burrows-wheeler aligner [69]. As the genome assemblies in the minimal 884 kb meiotic interval are identical in the Jones et al. and Glazer et al. assemblies [24,70], the original Jones et al. assembly was used [24]. Samtools (version 0.1.17) [71] was used to create a sorted and indexed BAM file, and Picard tools (version 1.51) (http://broadinstitute.github.io/picard/) was used to fix mate information, add read groups, and remove PCR duplicates. GATK's Unified Genotyper (parameters: 'genotype likelihoods model INDEL', 'stand call conf 25', and 'stand emit conf 25') RealignerTargetCreator, IndelRealigner (parameter: 'LOD 0.4') was used to call potential target indels and perform realignment around indels. Base quality recalibration was accomplished using BaseRecalibrator. HaplotypeCaller (parameters: 'emitRefConfidence GVCF', 'variant index type LINEAR', and 'variant index parameter 128000') was used to generating a genomic VCF (gVCF) file for each library. The resulting gVCFs were merged and variants were called using the GenotypeGVCFs module [7274]. High quality variants were selected using the following criteria: 1) Variants must have a variant quality score greater than 400. 2) Variants must not be called 'missing' or have a quality score of less than 10 in either high-coverage benthic genome. 3) Variants must not be called 'missing' or have a quality score of less than ten in no more than two genomes. To further remove stickleback specific repeats, we removed variants with >99% of the 100bp flanking sequence matching more than six places in the genome using blastn with an e-value of less than 1x10-30 [75]. QTL concordance score is the absolute value of the proportion of times a variant was present in benthic fish with a chromosome 21 tooth QTL minus the proportion of times the same

variant was found in fish without a tooth QTL. QTL Concordance scores were calculated using a custom python script.

Generation of transgenic enhancer stickleback lines

To generate GFP reporter constructs, each of the intron 4 fragments from the Little Campbell marine grandparent from cross 5 was cloned upstream of the *Hsp70l* promoter in a Tol2 expression construct using *Nhe*I [33]. For the mCherry construct, we cloned mCherry into the *Hsp70l* reporter construct using *Sal*I and *Cla*I and the inserts were cloned upstream using *Nhe*I and *BamH*I. Primers for construct generation and sequencing are shown in Table 3.5.

To generate transgenic stickleback, transposase messenger RNA was synthesized from pCS2-TP [76] plasmid linearized with *Not*I and transcribed using the mMessage SP6 *in vitro* transcription kit (Ambion) and purified using the Qiagen RNeasy column. One-cell marine stickleback embryos were injected with a mixture of 37.6 ng/μL plasmid DNA and 75 ng/μL RNA with 0.05% phenol red as previously described [33]. All transgene images presented are from stable lines except for the mCherry expression in Fig. 3.5I-J and the 2kb fragment in Fig. 3.4 (which were mosaic).

Generation of TALEN construct targeting stickleback *Bmp6*

To generate a TALEN pair to target the stickleback *Bmp6* gene, we used the TAL effector Nucleotide Targeter 2.0 (https://tale-nt.cac.cornell.edu/node/add/talen) to scan the second exon sequence of *Bmp6* for potential target sites [77,78]. We chose TALEN parameters as described [34]. We chose a target site that is unique to *Bmp6* in the stickleback genome and contains a common restriction site, *Eco*RI, which can be used to detect molecular deletions. We assembled the two TALEN constructs using Golden Gate cloning into the destination vectors pCS2TALDD and pCS2TALRR and verified correct assembly using Sanger sequencing as described [34]. See Table 3.6 for the specifics of the *Bmp6* TALEN design.

Synthesis and injection of TALEN RNA into stickleback embryos

5-capped mRNA for each TALEN pair was transcribed using the SP6 mMessage Machine (Ambion) after the TALEN plasmid templates had been linearized with *Not*I. Pooled TALEN mRNA was injected into one-cell PAXB freshwater benthic stickleback embryos at a concentration of 40 ng/μL for each mRNA with 0.05% phenol red.

Talen mutation identification

To genotype fish for TALEN induced mutations, DNA was extracted as described above from adult fish caudal fin tissue or homogenized whole 1-3 dpf embryos. Genotyping PCR was performed using forward primer 5- ACAAGCCGCTAAAAAGGACA-3 and reverse primer 5- GCACGTGTGCATGCTTTAGA -3. The reaction profile for the NEB Phusion reaction

was 98C for 30 seconds, 39 cycles of 98C for 10 seconds, 58C for 15 seconds, 72C for 30 seconds, followed by 72C for 10 minutes. The PCR products were cut directly with *EcoR*I. The products from the wild-type and mutant alleles are cut and not cut, respectively, by this assay (See Fig 3.7).

Tooth patterning quantification

Dorsal and ventral pharyngeal tooth number was quantified on a DM2500 Leica microscope using a TX2 filter as previously described [19]. For both ventral and dorsal tooth counts, total tooth number equals the sum of the left and right sides (of ventral and dorsal pharyngeal teeth, respectively). Tooth plate area and spacing of the ventral pharyngeal tooth plate were quantified from a gray scale image taken with a DFC340 FX camera on a Leica M165FC as previously described [19]. Area and spacing of the ventral pharyngeal tooth plates are the averages of the left and right tooth plate. Skeletal traits were binned by total fish length for three stages: early juvenile <27 mm, late juvenile 27-37 mm, and adults >37 mm.

RNA purification, sequencing, and alignment

Ventral tooth plates from three wild-type and homozygous mutant (for the 13 bp deletion allele) *Bmp6* female sticklebacks (standard length 25 mm) were dissected, placed into TRI reagent (Sigma-Aldrich) on ice, ground with a disposable pestle, and frozen overnight at -80C. The next day, RNA was extracted, isopropanol precipitated, and resuspended in DEPC-treated water. 200 ng of purified RNA was used with Illumina's Truseq Stranded mRNA Library Prep Kit to create sequencing libraries. The resulting bar-coded libraries were pooled and 100 bp paired end reads were generated using a single lane of an Illumina HiSeq2000. Reads were mapped to the stickleback reference genome [24] using STAR (parameters: '–alignIntronMax 200000' '–alignMatesGapMax 200000' '–outFilterMultimapNmax 8') [79] . BAM files were created, sorted, and indexed using Samtools (version 0.1.17)[71]. Picard tools (version 1.51) was used to fix mate information, add read groups, and remove PCR duplicates (http://broadinstitute.github.io/picard/). Using the Ensembl reference transcriptome [24], transcripts were quantified using cuffquant version 2.2.1 (parameters: '-u' '–library-type fr-firststrand') and normalized using cuffnorm [80,81]. Principal component analysis of the resulting transcript abundances was done using the PCA package of FactoMineR (http://factominer.free.fr/index.html) in R, and was plotted in R. GO term enrichment for genes ranked by expression correlation with the first principal component of the RNAseq expression matrix was performed using GOrilla [82,83]. Hierarchical clustering was done using Cluster3.0 (parameters: '-l' '-cg a' '-g 2' '-e 0' '-m c') [84], and the results were visualized using JavaTreeView (version 1.1.6r4)[85]. Additional figures and analyses were done using custom python scripts and figures created using matplotlib.

Gene Set Enrichment Analysis

ToothCODE gene sets were downloaded from the ToothCODE database (http://compbio.med.harvard.edu/ToothCODE/). ToothCODE identified downstream targets of Bmp signaling by literature mining manipulations of *Bmp2*, *Bmp4*, and *Bmp7*. Targets that were upregulated when BMP signaling increased or downregulated when BMP signaling was decreased were termed BMP target genes. Stickleback orthologs of mouse hair follicle stem cell signature genes, genes upregulated in the hair follicle bulge relative to the hair germ [46] were identified using Ensembl predictions. Statistical enrichment was done similar to the methods as previously described [86]. Each gene in a set was subject to a t-test, obtaining a list of z-scores. The null hypothesis, that the gene set displays no differential expression enrichment, (i.e. t-test z-scores are drawn from a standard normal distribution) was tested using a 1-sample t-test, with resulting P values subject to a Bonferroni correction. The significance cutoff for the 1-sample t-test was confirmed by creating a simulated null distribution, using 10,000 permutations of an equal number of genes as in each gene set, randomly chosen without replacement. Cutoff test statistic values were chosen by taking the values at the 100-(2.5/N) and 2.5/N percentile in the simulated null distribution, where N was the number of hypotheses being tested. Analysis was done using a set of custom python scripts, available upon request.

3.6 Acknowledgements

We thank Sara Carsanaro, Yijia Hong, and Anthony Lee for help with phenotypic analyses; Andrew Glazer and Michael Bronski for help constructing some of the Nextera genome resequencing libraries, Mario Cleves for statistical consultation, and Deena Emera and Tyler Square for comments on the manuscript.

3.7 References

1. Carroll SB. Evo-devo and an expanding evolutionary synthesis: a genetic theory of morphological evolution. Cell. 2008;134: 2536.
2. Wray GA. The evolutionary significance of cis-regulatory mutations. Nat Rev Genet. 2007;8: 206216. doi:10.1038/nrg2063
3. Martin A, Orgogozo V. The loci of repeated evolution: a catalog of genetic hotspots of phenotypic variation. Evolution. 2013;67: 12351250. doi:10.1111/evo.12081
4. Stern DL, Orgogozo V. The loci of evolution: how predictable is genetic evolution? Evolution. 2008;62: 215577.
5. Jernvall J, Thesleff I. Tooth shape formation and tooth renewal: evolving with the same signals. Development. 2012;139: 34873497. doi:10.1242/dev.085084
6. Tucker AS, Fraser GJ. Evolution and developmental diversity of tooth regeneration. Semin Cell Dev Biol. 2014;2526: 7180. doi:10.1016/j.semcdb.2013.12.013

7. Catn J, Tucker AS. Current knowledge of tooth development: patterning and mineralization of the murine dentition. J Anat. 2009;214: 502515. doi:10.1111/j.1469-7580.2008.01014.x
8. OConnell DJ, Ho JWK, Mammoto T, Turbe-Doan A, OConnell JT, Haseley PS, et al. A Wnt-bmp feedback circuit controls intertissue signaling dynamics in tooth organogenesis. Sci Signal. 2012;5: ra4. doi:10.1126/scisignal.2002414
9. Neubser A, Peters H, Balling R, Martin GR. Antagonistic interactions between FGF and BMP signaling pathways: a mechanism for positioning the sites of tooth formation. Cell. 1997;90: 247255.
10. St Amand TR, Zhang Y, Semina EV, Zhao X, Hu Y, Nguyen L, et al. Antagonistic signals between BMP4 and FGF8 define the expression of Pitx1 and Pitx2 in mouse tooth-forming anlage. Dev Biol. 2000;217: 323332. doi:10.1006/dbio.1999.9547
11. Bei M, Kratochwil K, Maas RL. BMP4 rescues a non-cell-autonomous function of Msx1 in tooth development. Development. 2000;127: 47114718.
12. Kavanagh KD, Evans AR, Jernvall J. Predicting evolutionary patterns of mammalian teeth from development. Nature. 2007;449: 42732.
13. Andl T, Ahn K, Kairo A, Chu EY, Wine-Lee L, Reddy ST, et al. Epithelial Bmpr1a regulates differentiation and proliferation in postnatal hair follicles and is essential for tooth development. Development. 2004;131: 22572268. doi:10.1242/dev.01125
14. Wang Y, Li L, Zheng Y, Yuan G, Yang G, He F, et al. BMP activity is required for tooth development from the lamina to bud stage. J Dent Res. 2012;91: 6905.
15. Kingsley DM, Peichel CL. The molecular genetics of evolutionary change in sticklebacks. In: S. Ostlund-Nilsson IM, Huntingford FA, editors. Biology of the three-spine stickleback. CRC Press; 2007. pp. 4181.
16. Peichel CL, Marques DA. The genetic and molecular architecture of phenotypic diversity in sticklebacks. Philos Trans R Soc Lond B Biol Sci. 2017;372. doi:10.1098/rstb.2015.0486
17. Bell MA, Foster SA. The evolutionary biology of the threespine stickleback. New York: Oxford University Press; 1994.
18. Chan YF, Marks ME, Jones FC, Villarreal G, Shapiro MD, Brady SD, et al. Adaptive evolution of pelvic reduction in sticklebacks by recurrent deletion of a Pitx1 enhancer. Science. 2010;327: 302305. doi:10.1126/science.1182213
19. Cleves PA, Ellis NA, Jimenez MT, Nunez SM, Schluter D, Kingsley DM, et al. Evolved tooth gain in sticklebacks is associated with a cis-regulatory allele of Bmp6. Proc Natl Acad Sci U S A. 2014;111: 1391213917. doi:10.1073/pnas.1407567111
20. Colosimo PF, Hosemann KE, Balabhadra S, Villarreal G, Dickson M, Grimwood J, et al. Widespread parallel evolution in sticklebacks by repeated fixation of Ectodysplasin alleles. Science. 2005;307: 19281933. doi:10.1126/science.1107239
21. Indjeian VB, Kingman GA, Jones FC, Guenther CA, Grimwood J, Schmutz J, et al. Evolving new skeletal traits by cis-regulatory changes in Bone Morphogenetic Proteins. Cell. 2016;164: 4556. doi:10.1016/j.cell.2015.12.007
22. Miller CT, Beleza S, Pollen AA, Schluter D, Kittles RA, Shriver MD, et al. cis-Regulatory changes in Kit ligand expression and parallel evolution of pigmentation in sticklebacks and humans. Cell. 2007;131: 11791189. doi:10.1016/j.cell.2007.10.055

23. OBrown NM, Summers BR, Jones FC, Brady SD, Kingsley DM. A recurrent regulatory change underlying altered expression and Wnt response of the stickleback armor plates gene EDA. eLife. 2015;4: e05290. doi:10.7554/eLife.05290
24. Jones FC, Grabherr MG, Chan YF, Russell P, Mauceli E, Johnson J, et al. The genomic basis of adaptive evolution in threespine sticklebacks. Nature. 2012;484: 5561. doi:10.1038/nature10944
25. Schluter D, Mcphail JD. Ecological character displacement and speciation in sticklebacks. Am Nat. 1992;140: 85108.
26. Ellis NA, Glazer AM, Donde NN, Cleves PA, Agoglia RM, Miller CT. Distinct developmental genetic mechanisms underlie convergently evolved tooth gain in sticklebacks. Development. 2015;142: 24422451. doi:10.1242/dev.124248
27. Wise SB, Stock DW. Conservation and divergence of Bmp2a, Bmp2b, and Bmp4 expression patterns within and between dentitions of teleost fishes. Evol Dev. 2006;8: 511523. doi:10.1111/j.1525-142X.2006.00124.x
28. Muschick M, Indermaur A, Salzburger W. Convergent evolution within an adaptive radiation of cichlid fishes. Curr Biol. 2012;22: 23628.
29. Fraser GJ, Graham A, Smith MM. Conserved deployment of genes during odontogenesis across osteichthyans. Proc Biol Sci. 2004;271: 23112317. doi:10.1098/rspb.2004.2878
30. Fraser GJ, Hulsey CD, Bloomquist RF, Uyesugi K, Manley NR, Streelman JT. An ancient gene network is co-opted for teeth on old and new jaws. PLoS Biol. 2009;7: e31. doi:10.1371/journal.pbio.1000031
31. Jackman WR, Draper BW, Stock DW. Fgf signaling is required for zebrafish tooth development. Dev Biol. 2004;274: 139157. doi:10.1016/j.ydbio.2004.07.003
32. Miller CT, Glazer AM, Summers BR, Blackman BK, Norman AR, Shapiro MD, et al. Modular skeletal evolution in sticklebacks is controlled by additive and clustered quantitative trait loci. Genetics. 2014;197: 405420. doi:10.1534/genetics.114.162420
33. Erickson PA, Cleves PA, Ellis NA, Schwalbach KT, Hart JC, Miller CT. A 190 base pair, TGF- responsive tooth and fin enhancer is required for stickleback Bmp6 expression. Dev Biol. 2015;401: 310323. doi:10.1016/j.ydbio.2015.02.006
34. Dahlem TJ, Hoshijima K, Jurynec MJ, Gunther D, Starker CG, Locke AS, et al. Simple methods for generating and detecting locus-specific mutations induced with TALENs in the zebrafish genome. PLoS Genet. 2012;8: e1002861. doi:10.1371/journal.pgen.1002861
35. Anker GC. Morphology and kinetics of the head of the stickleback, Gasterosteus aculeatus. Trans Zool Soc Lond. 1974;32: 311416.
36. Prashar P, Yadav PS, Samarjeet F, Bandyopadhyay A. Microarray meta-analysis identifies evolutionarily conserved BMP signaling targets in developing long bones. Dev Biol. 2014;389: 192207. doi:10.1016/j.ydbio.2014.02.015
37. Juuri E, Jussila M, Seidel K, Holmes S, Wu P, Richman J, et al. Sox2 marks epithelial competence to generate teeth in mammals and reptiles. Dev Camb Engl. 2013;140: 14241432. doi:10.1242/dev.089599
38. Abduweli D, Baba O, Tabata MJ, Higuchi K, Mitani H, Takano Y. Tooth replacement and putative odontogenic stem cell niches in pharyngeal dentition of medaka (Oryzias

latipes). Microsc Oxf Engl. 2014;63: 141153. doi:10.1093/jmicro/dft085
39. Martin KJ, Rasch LJ, Cooper RL, Metscher BD, Johanson Z, Fraser GJ. Sox2+ progenitors in sharks link taste development with the evolution of regenerative teeth from denticles. Proc Natl Acad Sci U S A. 2016;113: 1476914774. doi:10.1073/pnas.1612354113
40. Blanpain C, Lowry WE, Geoghegan A, Polak L, Fuchs E. Self-renewal, multipotency, and the existence of two cell populations within an epithelial stem cell niche. Cell. 2004;118: 635648. doi:10.1016/j.cell.2004.08.012
41. Genander M, Cook PJ, Ramskld D, Keyes BE, Mertz AF, Sandberg R, et al. BMP signaling and its pSMAD1/5 target genes differentially regulate hair follicle stem cell lineages. Cell Stem Cell. 2014;15: 619633. doi:10.1016/j.stem.2014.09.009
42. Ahn Y. Signaling in tooth, hair, and mammary placodes. Curr Top Dev Biol. 2015;111: 421459. doi:10.1016/bs.ctdb.2014.11.013
43. Biggs LC, Mikkola ML. Early inductive events in ectodermal appendage morphogenesis. Semin Cell Dev Biol. 2014;2526: 1121. doi:10.1016/j.semcdb.2014.01.007
44. Darwin C. The variation of animals and plants under domestication. London: John Murray; 1875.
45. Pispa J, Thesleff I. Mechanisms of ectodermal organogenesis. Dev Biol. 2003;262: 195205.
46. Kandyba E, Leung Y, Chen Y-B, Widelitz R, Chuong C-M, Kobielak K. Competitive balance of intrabulge BMP/Wnt signaling reveals a robust gene network ruling stem cell homeostasis and cyclic activation. Proc Natl Acad Sci U S A. 2013;110: 13511356. doi:10.1073/pnas.1121312110
47. Berry FB, Skarie JM, Mirzayans F, Fortin Y, Hudson TJ, Raymond V, et al. FOXC1 is required for cell viability and resistance to oxidative stress in the eye through the transcriptional regulation of FOXO1A. Hum Mol Genet. 2008;17: 490505. doi:10.1093/hmg/ddm326
48. Wang L, Siegenthaler JA, Dowell RD, Yi R. Foxc1 reinforces quiescence in self-renewing hair follicle stem cells. Science. 2016;351: 613617. doi:10.1126/science.aad5440
49. Horsley V, Aliprantis AO, Polak L, Glimcher LH, Fuchs E. NFATc1 balances quiescence and proliferation of skin stem cells. Cell. 2008;132: 299310. doi:10.1016/j.cell.2007.11.047
50. Naveau A, Seidel K, Klein OD. Tooth, hair and claw: comparing epithelial stem cell niches of ectodermal appendages. Exp Cell Res. 2014;325: 96103.
doi:10.1016/j.yexcr.2014.02.003
51. Guenther C, Pantalena-Filho L, Kingsley DM. Shaping skeletal growth by modular regulatory elements in the Bmp5 gene. PLoS Genet. 2008;4: e1000308.
doi:10.1371/journal.pgen.1000308
52. Abzhanov A, Protas M, Grant BR, Grant PR, Tabin CJ. Bmp4 and morphological variation of beaks in Darwins finches. Science. 2004;305: 14625.
53. Albertson RC, Streelman JT, Kocher TD, Yelick PC. Integration and evolution of the cichlid mandible: the molecular basis of alternate feeding strategies. Proc Natl Acad Sci U A. 2005;102: 1628792.
54. Mou C, Pitel F, Gourichon D, Vignoles F, Tzika A, Tato P, et al. Cryptic patterning of avian skin confers a developmental facility for loss of neck feathering. PLoS Biol. 2011;9:

e1001028.
55. King MC, Wilson AC. Evolution at two levels in humans and chimpanzees. Science. 1975;188: 107116.
56. Hoekstra HE, Hirschmann RJ, Bundey RA, Insel PA, Crossland JP. A single amino acid mutation contributes to adaptive beach mouse color pattern. Science. 2006;313: 101104. doi:10.1126/science.1126121
57. Stern DL, Frankel N. The structure and evolution of cis-regulatory regions: the shavenbaby story. Philos Trans R Soc Lond B Biol Sci. 2013;368: 20130028. doi:10.1098/rstb.2013.0028
58. Ferguson CA, Tucker AS, Christensen L, Lau AL, Matzuk MM, Sharpe PT. Activin is an essential early mesenchymal signal in tooth development that is required for patterning of the murine dentition. Genes Dev. 1998;12: 26362649.
59. Ferguson CA, Tucker AS, Heikinheimo K, Nomura M, Oh P, Li E, et al. The role of effectors of the activin signalling pathway, activin receptors IIA and IIB, and Smad2, in patterning of tooth development. Development. 2001;128: 46054613.
60. Oka S, Oka K, Xu X, Sasaki T, Bringas P, Chai Y. Cell autonomous requirement for TGF-beta signaling during odontoblast differentiation and dentin matrix formation. Mech Dev. 2007;124: 409415. doi:10.1016/j.mod.2007.02.003
61. Bei M, Maas R. FGFs and BMP4 induce both Msx1-independent and Msx1-dependent signaling pathways in early tooth development. Development. 1998;125: 43254333.
62. Chen Y, Bei M, Woo I, Satokata I, Maas R. Msx1 controls inductive signaling in mammalian tooth morphogenesis. Development. 1996;122: 303544.
63. Vainio S, Karavanova I, Jowett A, Thesleff I. Identification of BMP-4 as a signal mediating secondary induction between epithelial and mesenchymal tissues during early tooth development. Cell. 1993;75: 4558.
64. Handrigan GR, Richman JM. A network of Wnt, hedgehog and BMP signaling pathways regulates tooth replacement in snakes. Dev Biol. 2010;348: 13041.
65. Handrigan GR, Leung KJ, Richman JM. Identification of putative dental epithelial stem cells in a lizard with life-long tooth replacement. Development. 2010;137: 35459.
66. Wu P, Wu X, Jiang T-X, Elsey RM, Temple BL, Divers SJ, et al. Specialized stem cell niche enables repetitive renewal of alligator teeth. Proc Natl Acad Sci U S A. 2013;110: E2009-2018. doi:10.1073/pnas.1213202110
67. Cooper RL, Martin KJ, Rasch LJ, Fraser GJ. Developing an ancient epithelial appendage: FGF signalling regulates early tail denticle formation in sharks. EvoDevo. 2017;8: 8. doi:10.1186/s13227-017-0071-0
68. Glazer AM, Cleves PA, Erickson PA, Lam AY, Miller CT. Parallel developmental genetic features underlie stickleback gill raker evolution. EvoDevo. 2014;5: 19. doi:10.1186/2041-9139-5-19
69. Li H, Durbin R. Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinforma Oxf Engl. 2009;25: 17541760. doi:10.1093/bioinformatics/btp324
70. Glazer AM, Killingbeck EE, Mitros T, Rokhsar DS, Miller CT. Genome assembly improvement and mapping convergently evolved skeletal traits in sticklebacks with genotyping-

by-sequencing. G3. 2015;5: 14631472. doi:10.1534/g3.115.017905
71. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, et al. The Sequence Alignment/Map format and SAMtools. Bioinforma Oxf Engl. 2009;25: 20782079. doi:10.1093/bioinformatics/btp352
72. DePristo MA, Banks E, Poplin R, Garimella KV, Maguire JR, Hartl C, et al. A framework for variation discovery and genotyping using next-generation DNA sequencing data. Nat Genet. 2011;43: 491498. doi:10.1038/ng.806
73. McKenna A, Hanna M, Banks E, Sivachenko A, Cibulskis K, Kernytsky A, et al. The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. Genome Res. 2010;20: 12971303. doi:10.1101/gr.107524.110
74. Van der Auwera GA, Carneiro MO, Hartl C, Poplin R, Del Angel G, Levy-Moonshine A, et al. From FastQ data to high confidence variant calls: the Genome Analysis Toolkit best practices pipeline. Curr Protoc Bioinforma. 2013;43: 11.10.1-33. doi:10.1002/0471250953.bi1110s43
75. Camacho C, Coulouris G, Avagyan V, Ma N, Papadopoulos J, Bealer K, et al. BLAST+: architecture and applications. BMC Bioinformatics. 2009;10: 421. doi:10.1186/1471-2105-10-421
76. Kawakami K. Transgenesis and gene trap methods in zebrafish by using the Tol2 transposable element. Methods Cell Biol. 2004;77: 201222.
77. Cermak T, Doyle EL, Christian M, Wang L, Zhang Y, Schmidt C, et al. Efficient design and assembly of custom TALEN and other TAL effector-based constructs for DNA targeting. Nucleic Acids Res. 2011;39: e82. doi:10.1093/nar/gkr218
78. Doyle EL, Booher NJ, Standage DS, Voytas DF, Brendel VP, Vandyk JK, et al. TAL Effector-Nucleotide Targeter (TALE-NT) 2.0: tools for TAL effector design and target prediction. Nucleic Acids Res. 2012;40: W117-122. doi:10.1093/nar/gks608
79. Dobin A, Davis CA, Schlesinger F, Drenkow J, Zaleski C, Jha S, et al. STAR: ultrafast universal RNA-seq aligner. Bioinforma Oxf Engl. 2013;29: 1521. doi:10.1093/bioinformatics/bts635
80. Trapnell C, Williams BA, Pertea G, Mortazavi A, Kwan G, van Baren MJ, et al. Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation. Nat Biotechnol. 2010;28: 511515. doi:10.1038/nbt.1621
81. Trapnell C, Hendrickson DG, Sauvageau M, Goff L, Rinn JL, Pachter L. Differential analysis of gene regulation at transcript resolution with RNA-seq. Nat Biotechnol. 2013;31: 4653. doi:10.1038/nbt.2450
82. Eden E, Lipson D, Yogev S, Yakhini Z. Discovering motifs in ranked lists of DNA sequences. PLoS Comput Biol. 2007;3: e39. doi:10.1371/journal.pcbi.0030039
83. Eden E, Navon R, Steinfeld I, Lipson D, Yakhini Z. GOrilla: a tool for discovery and visualization of enriched GO terms in ranked gene lists. BMC Bioinformatics. 2009;10: 48. doi:10.1186/1471-2105-10-48
84. de Hoon MJL, Imoto S, Nolan J, Miyano S. Open source clustering software. Bioinforma Oxf Engl. 2004;20: 14531454. doi:10.1093/bioinformatics/bth078
85. Saldanha AJ. Java Treeview–extensible visualization of microarray data. Bioinforma

Oxf Engl. 2004;20: 32463248. doi:10.1093/bioinformatics/bth349
86. Irizarry RA, Wang C, Zhou Y, Speed TP. Gene set enrichment analysis made simple. Stat Methods Med Res. 2009;18: 565575. doi:10.1177/0962280209351908

Chapter 4

Convergent evolution of gene expression in two high-toothed stickleback populations

A version of this article has been submitted to PLoS Genetics.

James C. Hart[1], Nicholas A. Ellis[1], Michael B. Eisen[1,2], and Craig T. Milller[1]
[1]Department of Molecular and Cell Biology, University of California-Berkeley, CA, USA
[2]Howard Hughes Medical Institute, University of California, Berkeley, CA, USA

4.1 Abstract

Changes in developmental gene regulatory networks enable evolved changes in morphology. These changes can be in *cis* regulatory elements that act in an allele-specific manner, or changes to the overall *trans* regulatory environment that interacts with *cis* regulatory sequences. Here we address several questions about the evolution of gene expression accompanying a convergently evolved constructive morphological trait, increases in tooth number in two independently derived freshwater populations of threespine stickleback fish (*Gasterosteus aculeatus*). Are convergently evolved *cis* and/or *trans* changes in gene expression associated with convergently evolved morphological evolution? Do *cis* or *trans* regulatory changes contribute more to gene expression changes accompanying an evolutionary gain of a morphological trait? Transcriptome data from dental tissue of ancestral low-toothed and two independently derived high-toothed stickleback populations revealed significantly shared gene expression changes that have convergently evolved in the two high-toothed populations. Comparing *cis* and *trans* regulatory changes using phased gene expression data from F_1 hybrids, we found that *trans* regulatory changes were predominant and more likely to be shared among both high-toothed populations. In contrast, while *cis* regulatory changes have evolved in both high-toothed populations, overall these changes were distinct and not shared among high-toothed populations. Together these data suggest that a convergently evolved trait can occur through genetically distinct regulatory changes that converge on similar *trans* regulatory environments.

4.2 Introduction

Development is controlled by a complex series of interlocking gene regulatory networks. Much of this regulation occurs at the level of transcription initiation, where *trans* acting factors bind to *cis* regulatory elements to control their target genes expression [1,2]. Evolved changes in an organism's morphology are the result of changes in this developmental regulatory landscape. It has been proposed that the genetic bases of many of these evolved changes are mutations within the *cis*-regulatory elements of genes [35]. Indeed, recent work in evolutionary genetics suggests the molecular bases of a diverse array of traits from *Drosophila* wing spots [6] to mouse pigmentation [7] to stickleback armored plate number [8,9] and size [10] are changes in the activity of *cis*-regulatory elements.

Evolved changes in gene expression can be divided into two broad regulatory classes. *Cis* regulatory changes occur within the proximal promoter [11], distal enhancer [12], or the gene body itself [13]. *Trans* regulatory changes modify the overall regulatory environment [14,15], but are genetically unlinked to the expression change. The total evolved gene expression differences can be partitioned into changes in *cis* and *trans* by quantifying expression differences between two populations and also testing for expression differences between alleles in F_1 hybrids between the two populations [16]. As both alleles in F_1 hybrids animals are exposed to the same regulatory environment, any difference in their expression must be

due to a *cis*-regulatory change. The remaining difference in gene expression between the two populations not explained by this *cis* change is due to changes in the *trans* regulatory environment. Several studies have attempted to characterize evolved *cis* and *trans*-regulatory changes at a transcriptome-wide level [17-21]. Though the relative contribution of *cis* and *trans* regulatory changes varies extensively among studies, *cis* changes have been found to dominate [17,18,21] or at least be approximately equivalent [19,20] to *trans* changes [22]. Additionally, compensatory changes (*cis* and *trans* changes in opposing directions) have been found to be enriched over neutral models [17,18], showing evidence for selection for stable gene expression levels. However, none of these studies examined contribution of *cis* and *trans* gene expression changes during convergent morphological evolution.

Populations evolve new traits following a shift to a novel environment, due to a mixture of drift and selection. Truly adaptive traits can often be repeatedly observed in multiple populations following a similar ecological shift. Threespine sticklebacks are an excellent system for the study of evolved changes in phenotypes, including gene expression [23-27]. Marine sticklebacks have repeatedly colonized freshwater lakes and streams along the coasts of the Northern hemisphere [28]. Each of these freshwater populations has independently adapted to its new environment; however, several morphological changes, including a loss in armored plates and a gain in tooth number, are shared among multiple newly derived populations [29,30]. The repeated evolution of lateral plate loss is due to repeated selection of a standing variant regulatory allele of the *Eda* gene within marine populations [8,9] and genome sequencing studies found over a hundred other shared standing variant alleles present in geographically diverse freshwater populations [31]. These studies suggest the genetic basis of freshwater adaptation might typically involve repeated reuse of the same standing variants to evolve the same adaptive freshwater phenotype.

However, more recent evidence has shown that similar traits have also evolved through different genetic means in freshwater stickleback populations. A recent study which mapped the genetic basis of a gain in pharyngeal tooth number in two independently derived freshwater populations showed a largely non-overlapping genetic architecture [30]. Another study using three different independently derived benthic (adapted to the bottom of a lake) populations showed that, even when adapting to geographically and ecologically similar environments, the genetic architecture of evolved traits is a mix of shared and unique changes [32]. Even in cases where the same gene is targeted by evolution in multiple populations (the loss of *Pitx1* expression resulting in a reduction in pelvic spines), the individual mutations are often independently derived [33,34]. All of these genomic scale studies have looked at the genetic control of morphological changes, while the extent and nature of genome-wide gene expression changes has been less studied. It remains an open question as to whether similar gene expression patterns evolve during the convergent evolution of morphology, and if so, to what extent those potential shared gene expression changes are due to shared *cis* or *trans* changes.

Teeth belong to a class of vertebrate epithelial appendages (including mammalian hair) that develop from placodes, and have long served as a model system for studying organogenesis and epithelial-mesenchymal interactions in vertebrates [35]. Odontogenesis is initiated

and controlled by complex interactions between epithelial and mesenchymal cell layers, and involves several deeply conserved signaling pathways [36-38]. Sticklebacks retain the ancestral jawed vertebrate condition of polyphyodonty, or continuous tooth replacement, and offer an emergent model system for studying tooth replacement. Previous work has supported the hypothesis that two independently derived freshwater stickleback populations have evolved an increase in tooth replacement rate, potentially mediated through differential odontogenic stem cell dynamics [30] (Cleves et al, 2018, under review). Recent studies have found teeth and taste bud development to be linked, with one study supporting a model where teeth and taste buds are copatterned from a shared oral epithelial source [39], and another study supporting a model where teeth and taste buds share a common progenitor stem cell pool [40].

We sought to examine the evolution of the regulatory landscape controlling stickleback tooth development and replacement. Using high-throughput RNA sequencing (RNA-seq), we found that two independently derived high-toothed freshwater populations display highly convergent gene expression changes, especially in orthologs of known tooth-expressed genes in other vertebrates, likely reflecting the convergently evolved tooth gain phenotype and the deep homology of teeth across all jawed vertebrates. We also quantitatively partitioned these evolved gene expression changes into *cis* and *trans* regulatory changes [16,19] in both populations at a transcriptome-wide level using RNA-seq on F_1 marine-freshwater hybrids. We found that *trans* regulatory changes predominate evolved changes in gene expression in dental tissue. Additionally, we found that the *trans* regulatory changes are more likely to be shared between the freshwater populations than *cis* regulatory changes. Thus, similar downstream transcription networks controlling tooth development and replacement have convergently evolved largely through different upstream genetic regulatory changes.

4.3 Results

Convergent evolution of tooth gain in two freshwater populations

To test whether multiple freshwater populations have evolved increases in tooth number compared to multiple ancestral marine populations [30,41], we quantified total ventral pharyngeal tooth number of lab reared sticklebacks from four distinct populations: (1) a marine population from the Little Campbell river ($LITC_M$) in British Columbia, Canada, (2) a second marine population from Rabbit Slough ($RABS_M$) in Alaska, (3) a benthic freshwater population from Paxton Lake ($PAXB_{FW}$) in British Columbia, Canada, USA, and (4) a second freshwater population from Cerrito Creek ($CERC_{FW}$) in California, USA (Fig 4.1A,B). Freshwater fish from both populations had more pharyngeal teeth than marine fish at this 35-50mm standard length (SL) stage, consistent with previous findings [30,41] of increases in tooth number in freshwater sticklebacks (Fig 4.1B,C).

To estimate the genomic relatedness of these populations, we resequenced the genomes of three marine and six freshwater sticklebacks from the four different populations (Table

4.1). We aligned the resulting reads to the stickleback reference genome [31] using Bowtie2 [42], and called variants using the Genome Analysis Toolkit (GATK) [43-45]. As it has been previously shown that Pacific marine stickleback populations are an outgroup to freshwater populations from Canada ($PAXB_{FW}$) and California ($CERC_{FW}$) [31], we hypothesized the two high-toothed populations would be more related to each other genomically than either marine population. A phylogeny constructed using genome-wide variant data cleanly separated freshwater populations from each other and from marine fish (Fig 4.2A). Principal component analysis of the genome-wide variants revealed that the first principle component explains nearly half (44%) of the overall variance and separates $PAXB_{FW}$ sticklebacks from both $CERC_{FW}$ and marine fish (Fig 4.2B), representing the independent evolution of $PAXB_{FW}$ genomes. The second principal component separated both freshwater populations from marine populations, showing partially shared freshwater genome evolution. These results further support the model that populations of freshwater sticklebacks used a combination of shared and independent genetic changes [31,32] when evolving a set of similar morphological changes in response to a new environment.

Convergent evolution of tooth gain in two freshwater populations

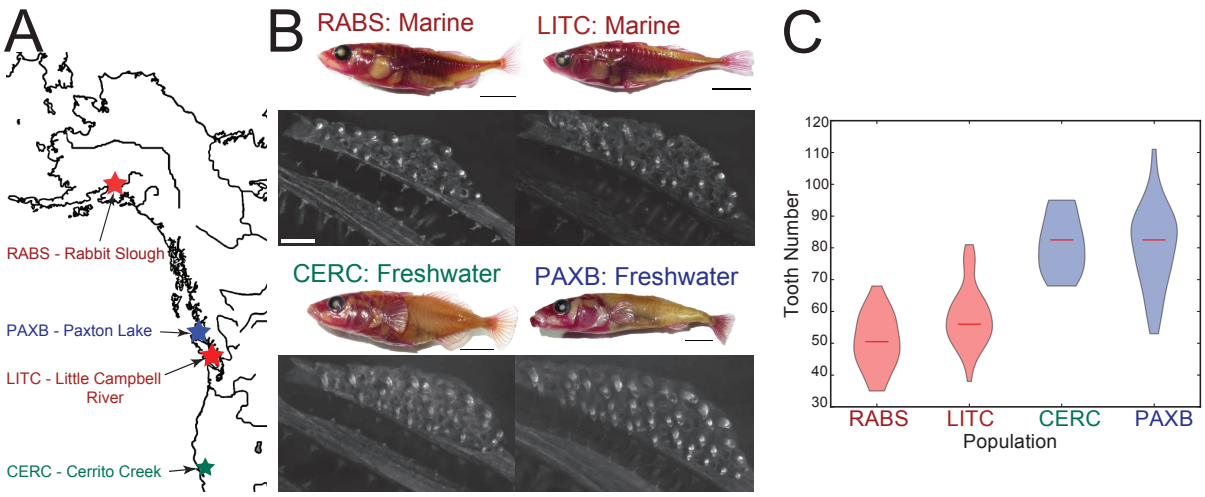

Figure 4.1: **Evolved tooth gain in two freshwater populations.** (A) Stickleback population locations. (B) Representative Alizarin red stained adult lab-reared sticklebacks (top, scale bars = 1 cm) and dissected ventral pharyngeal tooth plates (scale bars = 100μm). (C) Total ventral pharyngeal tooth number of 35-50 millimeter standard length lab-reared adult fish from each population. N = 44, 52, 12, 32 for $RABS_M$, $LITC_M$, $CERC_{FW}$, and $PAXB_{FW}$, respectively.

Fish	Reads	Final Mapped	In Cleves et al, 2018?
PAXB1	54682734	41536045	NO
PAXB2	32925802	27561555	NO
PAXB3	39144478	33907774	YES
CERC1	43204268	36319583	NO
CERC2	51249116	37310922	NO
CERC3	61773485	34238217	NO
LITC1	70365294	58516503	YES
LITC2	69017068	58385657	YES
RABS	39016218	33380182	NO

Table 4.1: **Genomic DNA sequencing reads.** For each fish, population and biological replicate number (Fish), the total number of barcoded reads from each fish (reads), and number of reads that mapped and passed all filters (final mapped) is listed. Some genomes are part of an additional study, indicated by (in Cleves et al, 2018?).

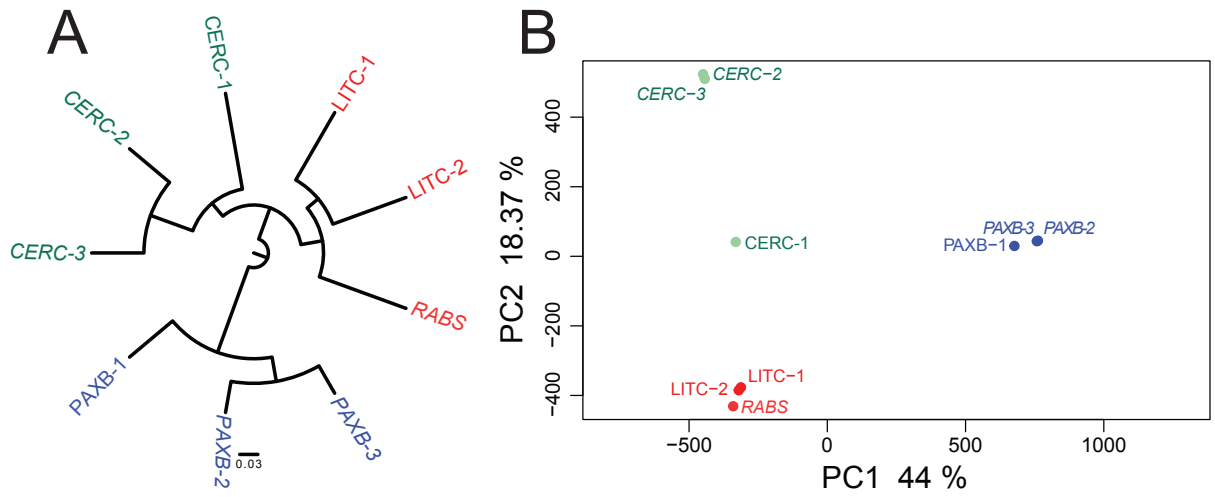

Figure 4.2: **Independent freshwater evolutionary history.** (A) Genome-wide maximum-likelihood phylogeny created from genomic resequencing data. Wild-caught fish are non-italicized. All nodes have 100% posterior probability. Scale bar shows 3% sequence divergence at variant positions. (B) Principal component analysis of genome-wide genotypes separates marine and CERC_{FW} populations from the PAXB_{FW} lake population, with the 2$^{\text{nd}}$ PC separating marine and freshwater populations.

Convergent evolution of gene expression

As morphological changes are often the result of changes in gene expression patterns and levels, we sought to identify evolved changes in gene expression during tooth development at stages soon after the evolved differences emerge [41]. We quantified gene expression in ventral pharyngeal dental tissue for three females each from the two high-toothed freshwater and Alaskan ($RABS_M$) low-toothed marine populations using RNA-seq (Fig 4.3A, Table 4.2). Principal component (PC) analysis of the resulting gene expression matrix showed a clustering of gene expression by population, with the first PC separating $PAXB_{FW}$ samples, and the second PC separating both $PAXB_{FW}$ and $CERC_{FW}$ samples from marine, similar to the PC analysis of the genome-wide variants (Fig 4.3B) [46].

Given the convergently evolved morphological change of increases in tooth number, we hypothesized that convergent evolution has occurred at the gene expression level in freshwater dental tissue. To test this hypothesis, we performed a differential expression analysis, comparing the evolved change in gene expression in $PAXB_{FW}$ dental tissue ($PAXB_{FW}$ expression vs marine) to the evolved change in $CERC_{FW}$ dental tissue ($CERC_{FW}$ expression vs marine). We found 6,693 and 3,501 genes (out of a total of 22,442) with significant (as determined by cuffdiff2 [47], see methods) evolved expression changes in $PAXB_{FW}$ and $CERC_{FW}$ respectively. Of these genes with evolved expression changes, 2,223 were called differentially expressed in both populations, with 1,898 (85%) showing expression changes in the same direction relative to marine.

At a genome-wide level, correlated changes in gene expression levels have evolved in the two high-toothed freshwater populations (Fig 4.3C, Spearman's r = 0.43). We next asked if orthologs of genes implicated in tooth development in other vertebrates showed an increase in correlated evolved expression changes. We compared the gene expression changes of stickleback orthologs of genes in the BiteIt (http://bite-it.helsinki.fi/) [48] or ToothCODE (http://compbio.med.harvard.edu/ToothCODE/) [36] databases (hereafter referred to as the BiteCode gene set), two databases of genes implicated in mammalian tooth development. Consistent with the conserved roles of gene regulatory networks regulating mammalian and fish teeth [49-52] and the major evolved increases in tooth number in both freshwater populations (Fig 4.1C), these predicted dental genes showed an increase in their correlated evolved gene expression change (Fig 4.3C red points, Spearman's r = 0.68), and tended to have an overall increase in gene expression (Fig 4.4, P = 7.36e-6, GSEA, see methods). This correlation coefficient was higher than any we observed over 100,000 bootstrapped gene sets of the same size from the same gene expression matrix. We also examined the expression levels of genes whose orthologs are annotated as being expressed in zebrafish pharyngeal teeth (www.zfin.org). Within this gene set, 27 of 40 genes were significantly more highly expressed in at least one freshwater population, with no genes expressed significantly higher (as determined by cuffdiff2 [47,53-55], see Materials and Methods) in marine samples than either freshwater population (Fig 4.3D).

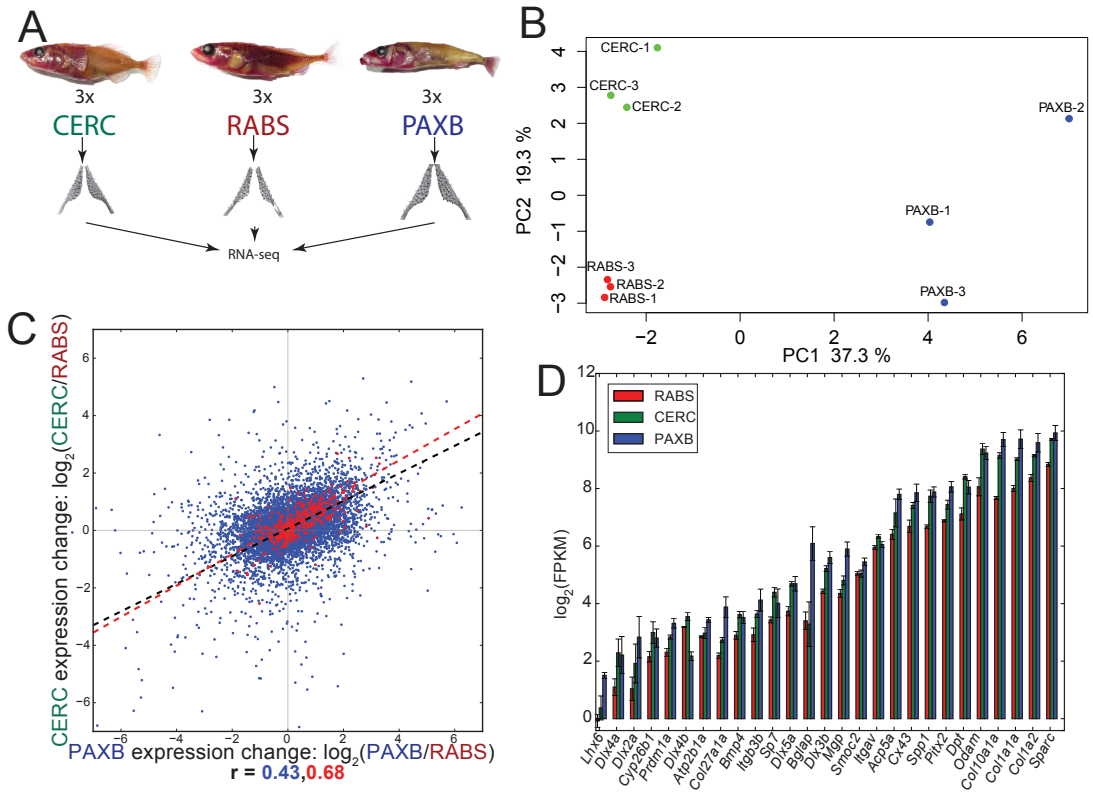

Figure 4.3: **Convergent evolution of gene expression in dental tissue.** (A) Ventral pharyngeal tooth plates from three different populations were dissected and gene expression quantified by RNA-seq. (B) Principal component analysis of dental tissue gene expression shows population specific expression profiles. (C) Freshwater dental tissue exhibited correlated gene expression changes for all genes (blue), with increased correlation observed for orthologs of genes known to be expressed during mammalian tooth development (red). (D) Expression of genes annotated as expressed in zebrafish teeth (zfin.org) which were significantly upregulated in one or both freshwater populations.

Sample	SL	Total Reads	Mapped Reads	Final Reads	Run1	Run2
CERC 1	44.58	84621832	84230042	60347896	43194152	41427680
CERC 2	45.51	82625088	77628572	58426762	42819976	39805112
CERC 3	46.36	78572698	77382826	59167308	38798290	39774408
RABS 1	44.7	87094088	88102796	68419470	46740088	40354000
RABS 2	46.19	82342214	80995352	60918592	43241272	39100942
RABS 3	47.44	86400410	82290040	62717370	43021352	43379058
PAXB 1	42.4	81773488	77302504	56014832	41827240	39946248
PAXB 2	43.73	82346498	81319300	64412654	41285630	41060868
PAXB 3	41.7	91013392	105993542	81377888	48187408	42825984

Table 4.2: **RNA-seq reads.** For each fish, population of parents and biological replicate number (sample), standard length (SL), total reads (generated by HiSeq2000 over two different runs (run1 and run2)), mapped reads (reads that mapped to the genome), and final reads (excludes reads filtered due to low quality or PCR duplication) is listed.

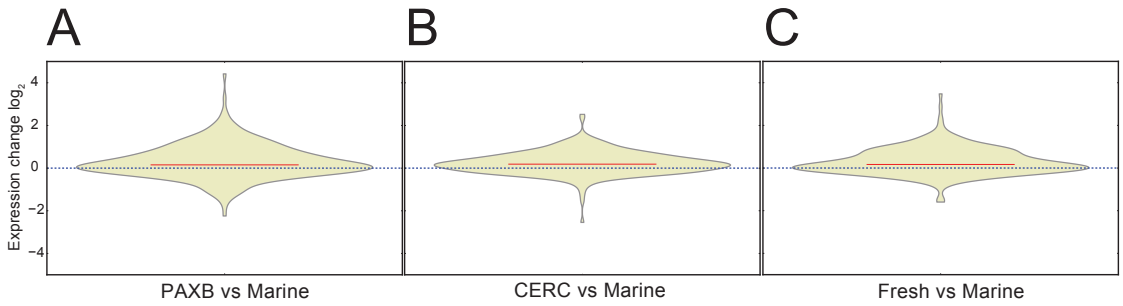

Figure 4.4: **Freshwater upregulation of putative dental genes.** (A) $PAXB_{FW}$ upregulation of BiteCode genes (P = 9.8e-3, GSEA). (B) $CERC_{FW}$ upregulation of BiteCode genes (P = 2.1e-5, GSEA). (C) $PAXB_{FW}$ and $CERC_{FW}$ upregulation of BiteCode genes (P = 5.1e-6, GSEA). 282 BiteCode genes were expressed in stickleback ventral tooth plates.

Increased freshwater expression of stem cell maintenance genes

Tooth development is controlled by several deeply conserved developmental signaling pathways [50,52]. To test whether expression changes in the components of specific developmental signaling pathways have evolved in the two high-toothed freshwater populations, we next analyzed the expression levels of stickleback orthologs of genes implicated in mammalian tooth development and annotated as components of different signaling pathways [36]. When comparing gene expression levels in freshwater dental tissue to marine dental tissue, genes annotated as part of the TGF-β signaling pathway displayed significantly increased expression in freshwater dental tissue (Fig 4.5A-F).

Since these two freshwater populations have a largely different developmental genetic basis for their evolved tooth gain [30], we next asked whether any pathways were upregulated or downregulated specifically in one freshwater population. When comparing the expression of genes in $PAXB_{FW}$ dental tissue to expression in $CERC_{FW}$ or marine dental tissue, genes not only in the TGF-β pathway, but also in the Wnt signaling pathway, displayed significantly increased expression, consistent with the differing genetic basis of tooth gain in these populations (Fig 4.5B). Genes upregulated in freshwater dental tissue were enriched for Gene Ontology (GO) terms involved in anatomical structure development, signaling, and regulation of cell proliferation (Fig 4.6A). Genes upregulated in $PAXB_{FW}$ dental tissue over marine were enriched for GO terms involved in cell proliferation, division and cell cycle regulation, as well as DNA replication (Fig 4.6B), while genes upregulated in $CERC_{FW}$ over marine were enriched for GO terms involved in cell locomotion, movement, and response to lipids (Fig 4.6C).

As teeth are constantly being replaced in polyphyodont adult fish, potentially due to the action of dental stem cells [40], we hypothesized that genes involved in stem cell maintenance have evolved increased expression in freshwater tooth plates, given the higher rate of newly forming teeth previously found in adults [30], and the possibly greater number of stem cell niches in high-toothed fish (Cleves et al, 2018, under review). We further hypothesized that since teeth are developmentally homologous to hair, perhaps an ancient genetic circuit regulating vertebrate placode replacement controls both fish tooth and mammalian hair replacement. For example, the *Bmp6* gene, previously described as expressed in all stickleback teeth [41] was significantly upregulated in $CERC_{FW}$ fish, consistent with the evolved major increases in tooth number in this population. In contrast, no such significant upregulation was observed in the expression of $PAXB_{FW}$ *Bmp6*, consistent with the observed evolved *cis*-regulatory decrease in $PAXB_{FW}$ *Bmp6* expression [41]. Further supporting this hypothesis, the expression of the stickleback orthologs of a previously published set of mouse hair follicle stem cell (HFSC) signature genes [56] were significantly upregulated in freshwater dental tissue (Fig 4.5A). $CERC_{FW}$ dental tissue displayed a small but significant increase in expression of this set of HFSC orthologs relative to both $PAXB_{FW}$ and marine samples (Fig 4.5C).

In cichlid fish, pharmacology experiments revealed that reductions in tooth density can be accompanied by concomitant increases or decreases in taste bud density [39]. To begin

to test whether derived high-toothed stickleback populations have also evolved significantly altered levels of known taste bud marker gene expression, we examined the expression levels of known taste bud markers *Calbindin2* and *Phospholipase Beta 2* [57], as well as taste receptors such as *Taste 1 Receptor Member 1*, *Taste 1 Receptor Member 3*, and *Polycystin 2 Like 1*[58]. Although four of these five genes had detectable significant expression changes between different populations, no consistent freshwater upregulation or downregulation of taste bud marker genes was seen (Fig 4.7).

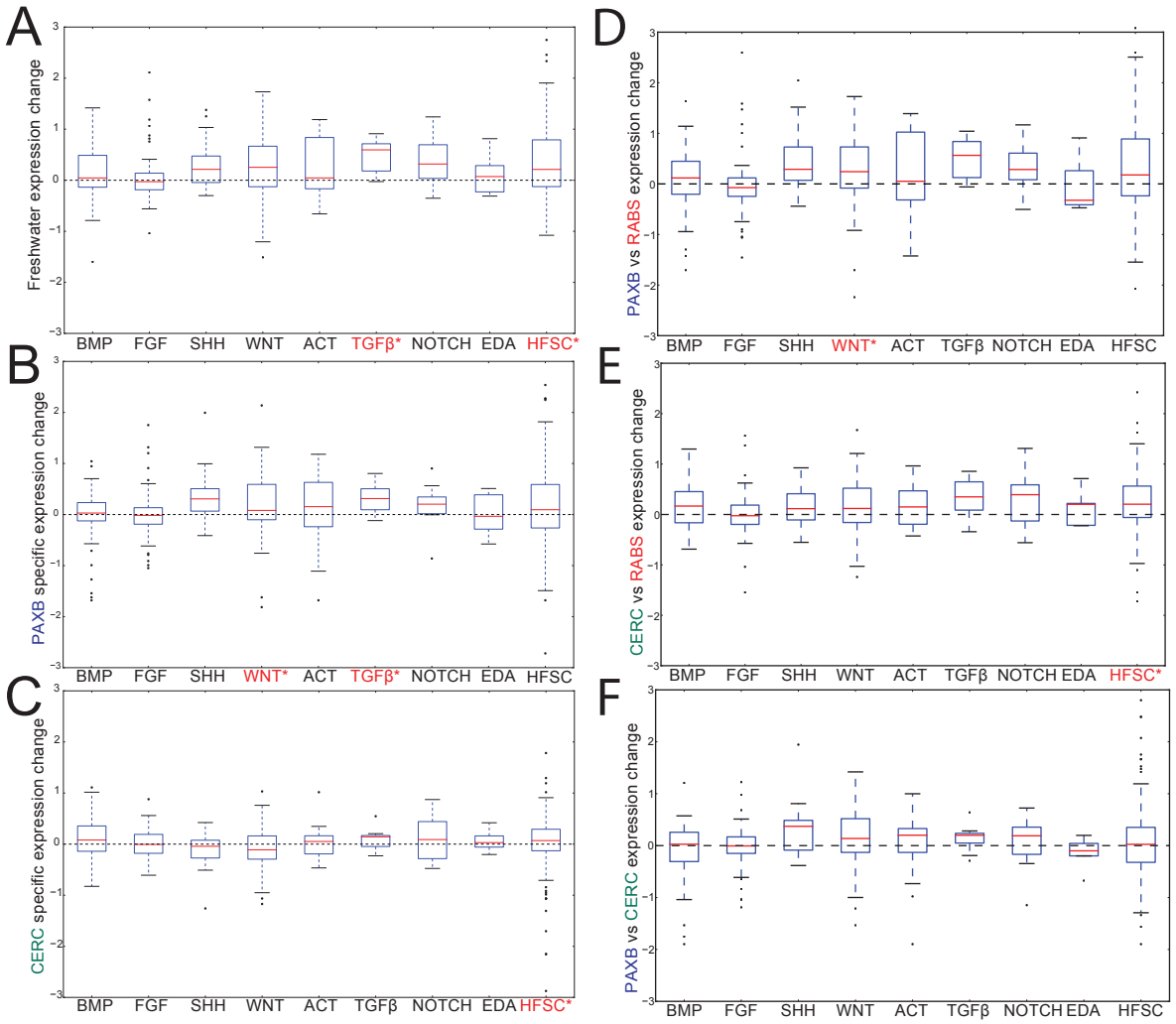

Figure 4.5: **Concerted changes in stem cell markers and signaling pathways.** (A-F) Changes in gene expression changes of genes annotated as components of the indicated signaling pathways (BMP, FGF, SHH, Wnt, ACT, TGFβ, NOTCH, or EDA, containing 59, 60, 28, 75, 19, 11, 12, and 6 expressed orthologs, respectively) [36] or orthologs of a described set of mouse hair follicle stem cell signature genes (HFSC, containing 254 expressed orthologs) [56]. Violin plots show the mean expression change of genes in the pathway. (A) Change in freshwater ($PAXB_{FW}$ + $CERC_{FW}$) relative to marine. (B) $PAXB_{FW}$ specific changes ($PAXB_{FW}$ relative to $CERC_{FW}$ + marine). (C) $CERC_{FW}$ specific changes ($CERC_{FW}$ relative to $PAXB_{FW}$ + marine). (D) $PAXB_{FW}$ evolved changes ($PAXB_{FW}$ relative to marine) (E) $CERC_{FW}$ evolved changes ($CERC_{FW}$ relative to marine) (F) $PAXB_{FW}$ vs $CERC_{FW}$ changes ($PAXB_{FW}$ relative to $CERC_{FW}$).

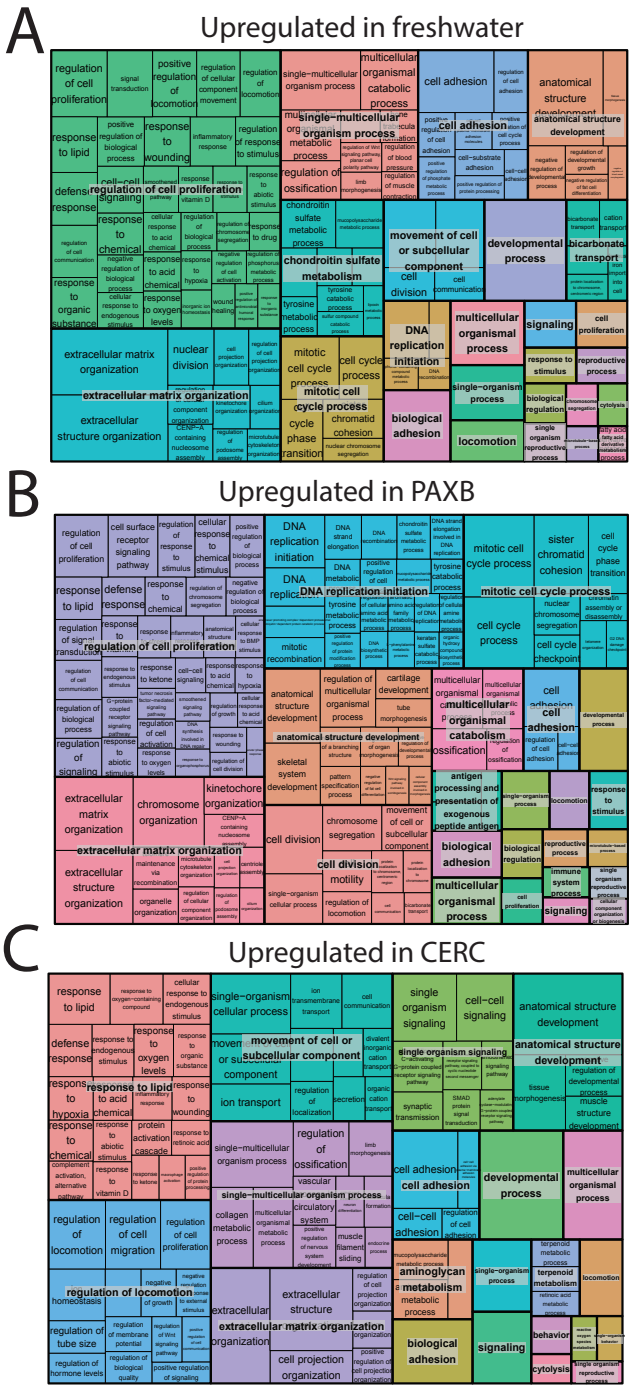

Figure 4.6: **Gene ontology of freshwater upregulated genes.** (A-C) GO enrichment of genes upregulated in PAXB_{FW} (A), CERC_{FW} (B), or both (C). GO analysis was preformed using Gorilla [68], with the results visualized with Revigo [70].

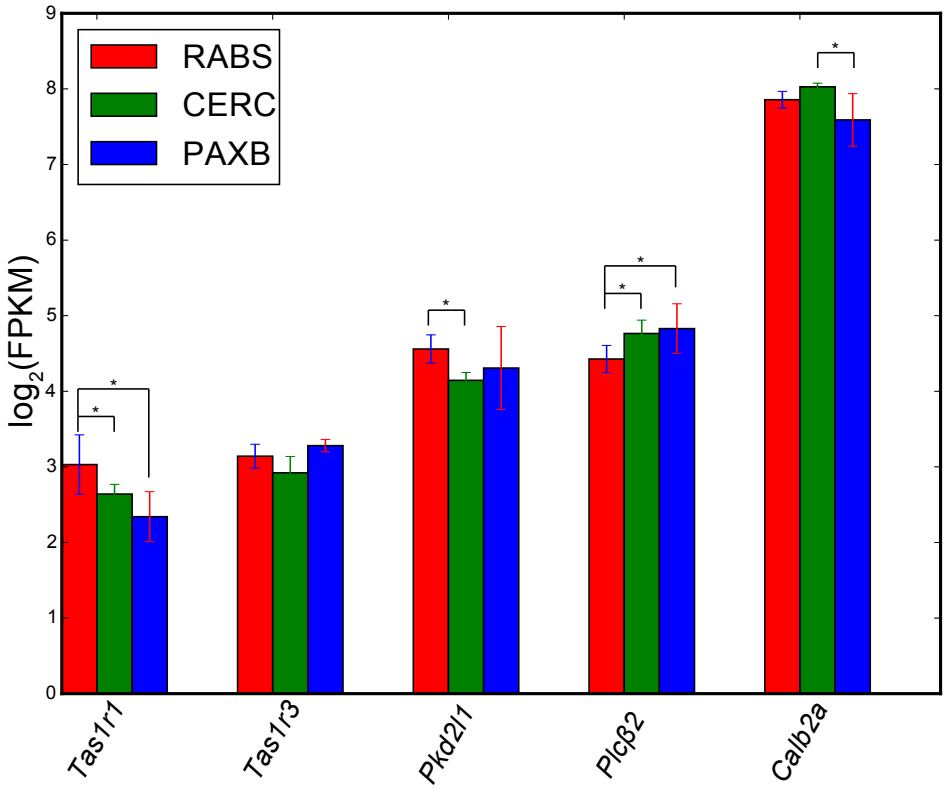

Figure 4.7: **Expression of taste bud marker genes.** Expression levels of known taste bud marker genes in marine, $PAXB_{FW}$ and $CERC_{FW}$ tooth plates as assayed by RNA-seq. * indicates differentially expressed genes. Error bars are standard error of the mean.

Cis and *trans* regulatory changes in gene expression

Evolved changes in gene expression are due to a combination of *cis* acting changes that are linked to the genes they act on, and *trans* acting changes which usually are genetically unlinked to the gene or genes they regulate. Since the genetic basis of freshwater tooth gain mapped to non-overlapping intervals in these two populations [30] (Cleves et al, 2018, under review), we hypothesized that the observed shared freshwater gene expression changes

were the result of a similar trans environment, but a largely different set of *cis* changes. To test this hypothesis, we measured evolved *cis* expression changes in marine-freshwater F_1 hybrids, which have marine and freshwater alleles present in the same *trans* environment. We raised both $CERC_{FW}$-marine and $PAXB_{FW}$-marine F_1 hybrids to the late juvenile stage, dissected their ventral pharyngeal tooth plates, then generated and sequenced five barcoded RNA-seq libraries per population (10 total). We then quantified the *cis* expression change as the ratio of the number of reads mapping uniquely to the freshwater allele of a gene to the number of uniquely mapping marine reads (Fig 4.8, Table 4.3). *Trans* expression changes were calculated by factoring the *cis* change out from the overall parental expression change [19].

We found 11,832 and 8,990 genes in $PAXB_{FW}$ and $CERC_{FW}$ F_1 hybrids, respectively, that had a fixed marine-freshwater sequence difference which had more than 20 total reads mapping to it. We observed no significant bias towards either the marine or freshwater allele in either set of F_1 hybrids (Fig 4.8B). We next classified genes into one of four categories (*cis* change only, *trans* change only, concordant *cis* and *trans* changes, discordant *cis* and *trans* changes). We found 1640 and 1116 $PAXB_{FW}$ (Fig. 4.8C) and $CERC_{FW}$ (Fig. 4.8D) genes, respectively, with only significant *cis* changes, and 1873 and 1048 genes, respectively, with only significant *trans* changes. We also found 478 and 359 genes with significant *cis* and *trans* changes in the same direction, which we term concordant changes in gene expression. Conversely, we found 772 and 607 genes with significant *cis* and *trans* changes in opposing directions, which we termed discordant changes. Discordant *cis* and *trans* changes were more common in both populations, suggesting selection for stable levels of gene expression.

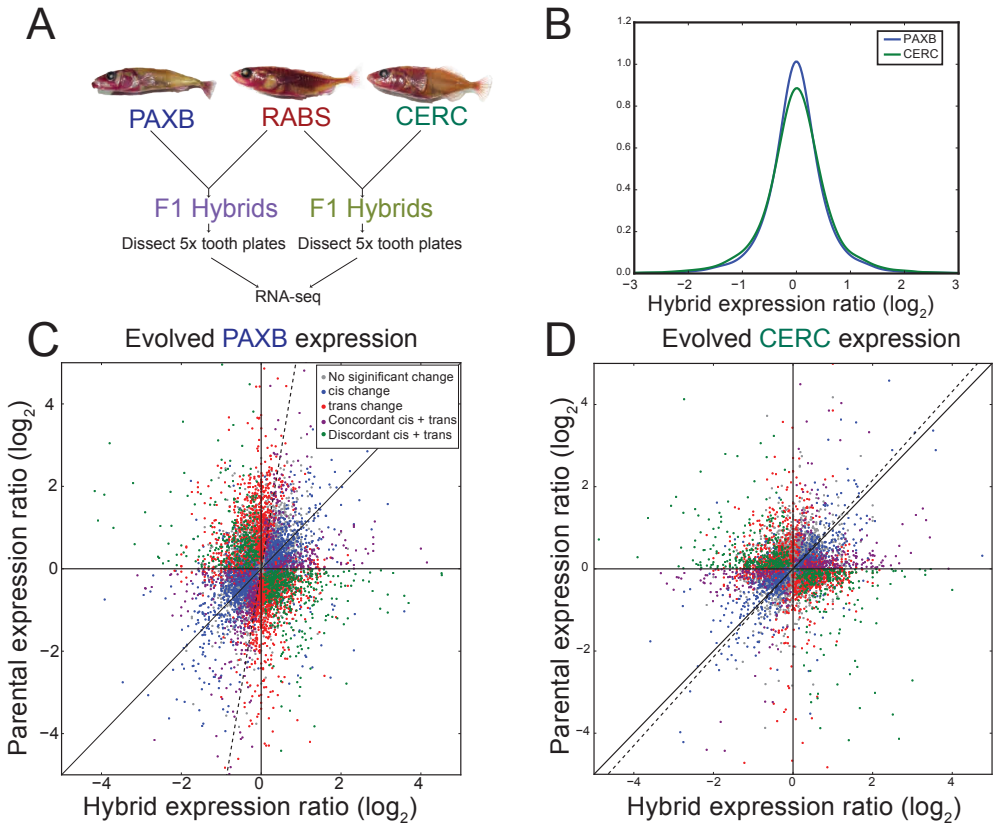

Figure 4.8: **Evolved changes in *cis*-regulation.** (A) Ventral pharyngeal tooth plates from marine- $PAXB_{FW}$ and marine- $CERC_{FW}$ F_1 hybrids were dissected and *cis* regulatory changes assayed using phased RNA-seq reads. (B) Density plot showing the measured *cis*-regulatory changes. Neither population displayed a significant allelic bias, as measured by a Wilcoxon signed-rank test. (C-D) Gene expression changes in both parental and hybrid dental tissue genes are color-coded based on the role of *cis* and/or *trans* change in $PAXB_{FW}$ (C) or $CERC_{FW}$ (D) dental tissue. Dashed line indicates the first principal component axis.

Sample	SL	Total Reads	Mapped Reads	Final Reads	Unique Reads
PAXB x RABS F_1 VTP1	41.08	38355384	29212068	24905322	5782725
PAXB x RABS F_1 VTP2	42.08	33832370	30173302	24512858	4455969
PAXB x RABS F_1 VTP3	42.35	34599652	31312244	25269588	4361345
PAXB x RABS F_1 VTP4	43.86	40963412	33975682	28630960	5260891
PAXB x RABS F_1 VTP5	43.69	32711530	36708220	29882826	4112675
CERC x RABS F_1 VTP1	43.4	32665968	32865322	27760198	3045175
CERC x RABS F_1 VTP2	41.95	40213188	40903260	32618470	3870225
CERC x RABS F_1 VTP3	40.5	34838262	35239962	25191170	3176683
CERC x RABS F_1 VTP4	43.25	33460702	34050608	28898530	3156749
CERC x RABS F_1 VTP5	42.25	33753920	34579258	29857006	3214069

Table 4.3: **F_1 hybrid RNA-seq reads.** For each ventral pharyngeal tooth plate (VTP), population of parents and biological replicate number (sample), standard length (SL), total reads (generated by HiSeq2000), mapped reads (reads that mapped to the genome), final reads (excludes reads filtered due to low quality or PCR duplication), and unique reads (reads that mapped uniquely to one haplotype) is listed.

Trans regulatory changes dominate

We next wanted to determine the relative contribution of *cis* and *trans* gene expression changes to evolved changes in gene expression. We restricted our analysis to differentially expressed genes (as determined by cuffdiff2 [47]) to examine only genes with a significant evolved difference in gene expression and quantifiable (i.e. genes with transcripts containing a polymorphic variant covered by at least 20 reads) *cis* and *trans* expression changes. When evolving a change in gene expression, the *cis* and *trans* regulatory basis for this change can be concordant (*cis* and *trans* effects both increase or decrease expression) or discordant (*cis* effects increase and *trans* decrease or vice versa). We hypothesized that genes would tend to display more discordant expression changes, as stabilizing selection has been found to buffer

gene expression levels [17,22,59]. To test this hypothesis, we binned differentially expressed genes into a 2x2 contingency table, with genes classified as *cis* or *trans* based on which effect controlled the majority of the evolved expression change, and discordant or concordant based on the direction of the *cis* and *trans* changes (Fig 4.9A, B). In the CERC_{FW} population, significantly more discordant changes than expected by a neutral model (P =1.35e-7, binomial test) have evolved. In both populations, we found increased discordant changes when the *trans* effect is larger than the *cis* effect (P =1.29e-7, 1.44e-13, PAXB_{FW} and CERC_{FW} respectively, binomial test). In both populations, we observe the opposite (an enrichment of concordant changes) when the *cis* effect is stronger, relative to the ratio when the *trans* effect is dominant (P =1.34e-36, 8.2e-11 PAXB_{FW} and CERC_{FW} respectively, binomial test). When considering all (not just differentially expressed) genes with quantifiable *cis* and *trans* expression changes, discordant changes dominated regardless of the relative strength of the *cis* effect (Fig 4.10).

If all gene expression changes were due to changes only in *cis*, we would expect to see the measured *cis* ratios in the hybrids match the parental expression ratios. Instead, in both cases of evolved change, we saw parental expression ratios of a greater magnitude than F_1 hybrid ratios, indicating a stronger contribution of *trans* changes to overall gene expression changes (Fig 4.8C-D). Indeed, when we examined the overall percentage of expression changes of differentially expressed genes that were due to changes in *cis*, we observed median per gene values of only 25.2% and 32.5% of PAXB_{FW} and CERC_{FW} gene expression changes, respectively (Fig 4.9C). Comparing the expression levels of orthologs of known dentally expressed genes from the BiteIt [48] and ToothCODE [36] databases revealed a similarly small number of gene expression changes explained by changes in *cis*, relative to the genome-wide average (Fig 4.9D). Evolved changes in CERC_{FW} gene expression were more due to changes in *cis* than PAXB_{FW} genes (Fig 4.9D, P = 1.25e-22, Mann-Whitney U test). Thus, *trans* effects on gene expression dominate the evolved freshwater gene expression changes.

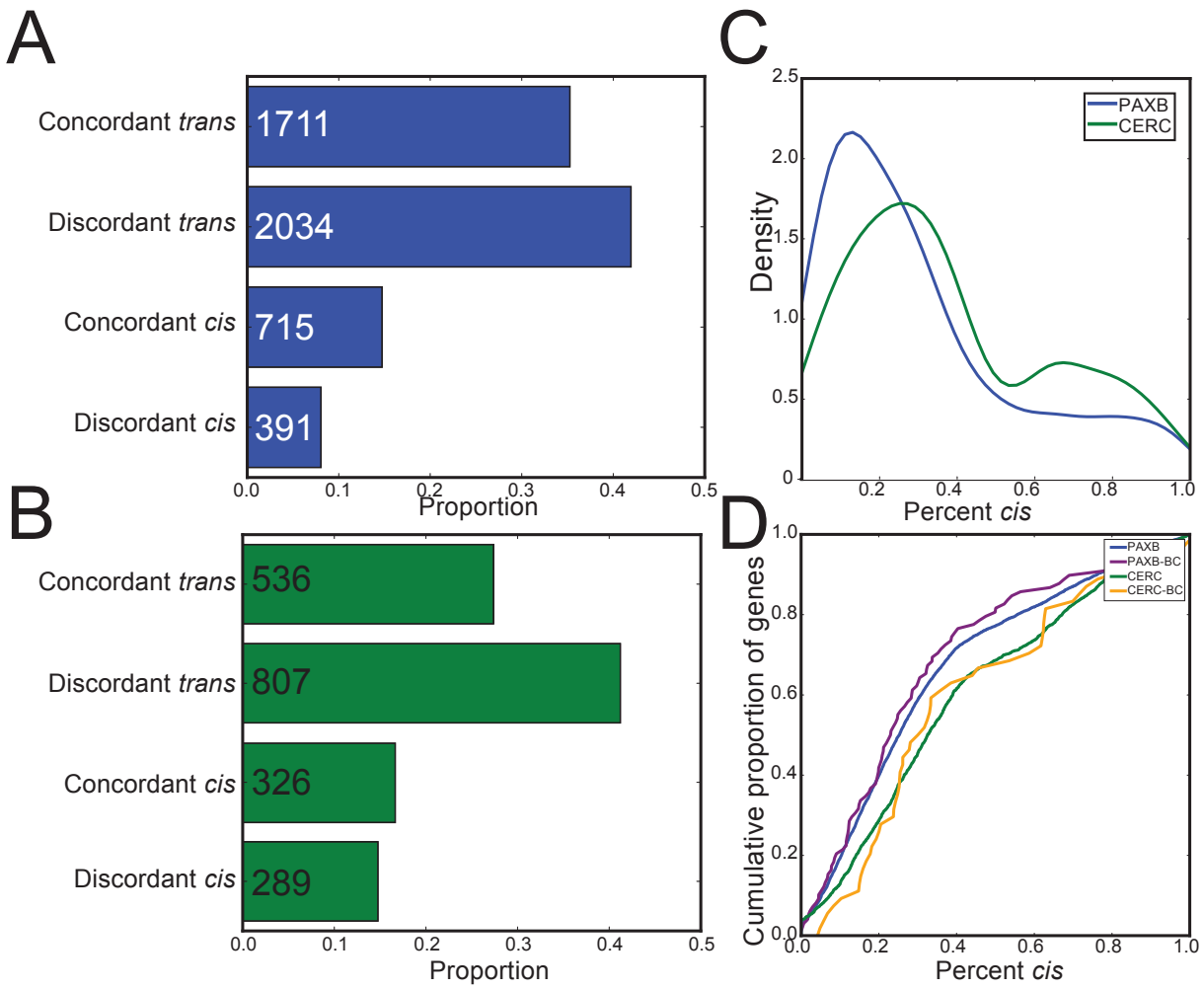

Figure 4.9: ***Trans* changes predominate evolved dental gene expression changes.** (A-B) Proportion of differentially expressed genes displaying opposing and concordant *cis* and *trans* changes in PAXB_{FW} (A) or CERC_{FW} (B) dental tissue. Genes whose expression differences were mostly explained by *cis* changes tended to be more concordant (P =5.0e-17, 0.002 for PAXB_{FW} and CERC_{FW}, respectively) than those mostly explained by *trans* changes. (C) Density of the relative percentage of gene expression differences which are explained by *cis* changes in PAXB_{FW} and CERC_{FW} dental tissue. (D) Cumulative percentage of percentage of gene expression due to *cis* changes. Genes in CERC_{FW} samples display a higher percentage *cis* change than genes in PAXB_{FW} samples (P = 1.25e-22, Mann-Whitney U test).

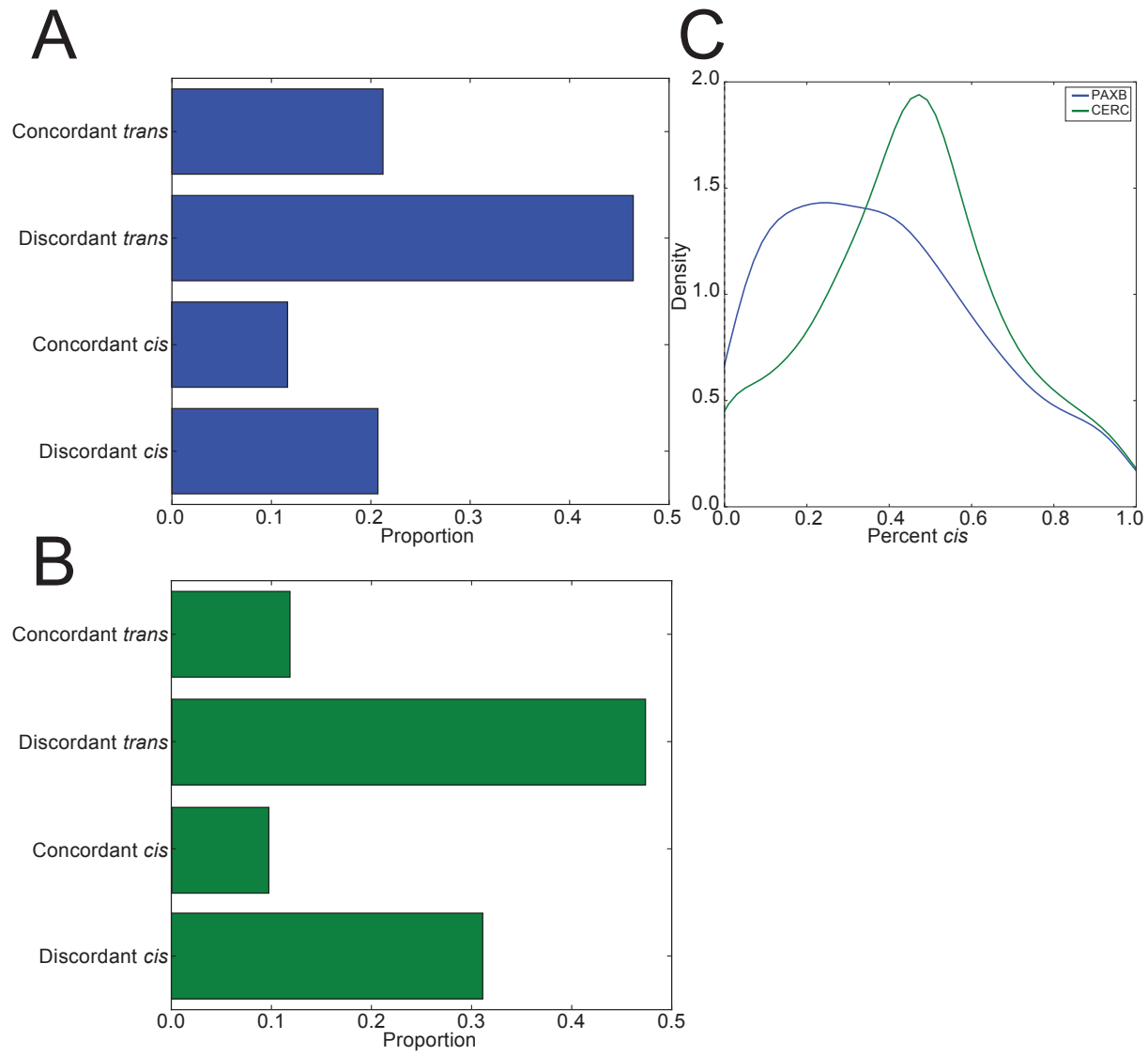

Figure 4.10: **Compensatory changes dominate genes with no significant evolved gene expression difference.** (A-B) Proportion of genes with quantifiable (i.e. genes with transcripts containing a polymorphic SNP covered by at least 20 reads) hybrid expression displaying opposing and concordant cis and trans changes in $PAXB_{FW}$ (A) or $CERC_{FW}$ (B) dental tissue. Similar to Fig 4.9, but here showing all genes, not just genes with significantly different expression levels compared to marine. *Trans* regulatory changes predominate, as do opposing over concordant changes. (C) Density plot of the percentage of gene expression changes explained by *cis*-regulatory changes.

Trans regulatory changes are more likely to be shared between freshwater populations

We next wanted to test the hypothesis that the shared freshwater gene expression changes were primarily due to shared *trans* changes, rather than shared *cis* changes. We first compared the overall expression levels of genes called differentially expressed between $PAXB_{FW}$ and marine as well as $CERC_{FW}$ and marine. We restricted our analysis to differentially expressed genes whose *cis*-regulatory change we were able to measure in our F_1 hybrids, including genes without a significant *cis* change. Similar to the genome-wide comparison, we found a highly significant non-parametric correlation coefficient (Spearman's r = 0.62, P =1.2e-132) for the expression change of these shared differentially expressed genes (Fig 4.11A). When comparing the $PAXB_{FW}$ *cis* changes of these genes to the $CERC_{FW}$ *cis* changes, however, we found a much lower (though still significant) correlation coefficient (Spearman's r = 0.13, P =5.1e-6) (Fig 4.11B). We calculated *trans* changes for each of these differentially expressed genes, defined as the difference between the expression change in the freshwater parent relative to marine and the freshwater allele relative to the marine in the F_1 hybrid [18,19,60]. When comparing the calculated *trans* changes for these shared differentially expressed genes, we observed much higher correlation coefficient (Spearman's r = 0.51, P =1.2e-80) (Fig 4.11C). When comparing all, not just differentially expressed, genes, *trans* changes are still likely to be more shared than *cis* (Fig 4.12). Additionally, 35/38 of the shared differentially expressed putative dental genes have shared regulatory increases or decreases in both freshwater populations relative to marine in overall expression difference, with 32/38 in *trans*, but only 25/38 in *cis* (Fig 4.11G-I). Thus, the *trans* effects on evolved gene expression are more likely to be shared by both freshwater populations than the *cis* changes.

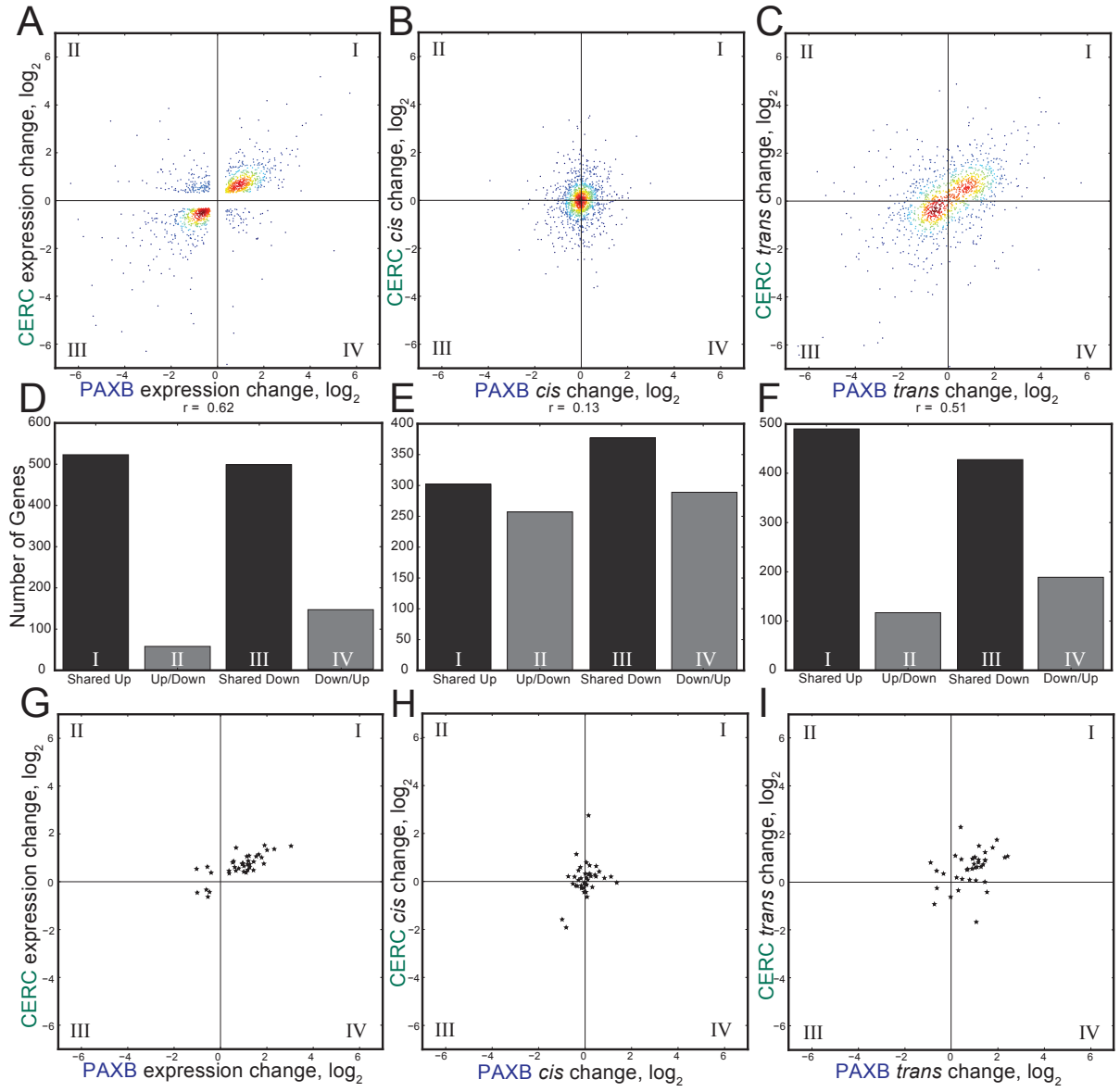

Figure 4.11: ***Trans* changes are more likely to be shared across populations.** (A) Genes with significantly different evolved expression in both freshwater populations relative to marine fish, showing significantly correlated changes in gene expression in PAXB_{FW} and CERC_{FW} dental tissue. (B) Freshwater dental tissue had a significant but small number of shared *cis*-regulatory changes. (C) Freshwater dental tissue showed significantly correlated changes in *trans* expression changes. A-C show genes with significant expression changes between populations and quantifiable (i.e. genes with transcripts containing a polymorphic SNP covered by at least 20 reads) *cis*-regulatory changes in both populations. Density (color) was estimated with a Gaussian kernal density estimator. BiteCode genes (see Methods) are indicated with black stars. D-F Bar graphs show the number of genes with shared or divergent expression patterns from the above panels. G-I are similar to A-C, but show only genes in the BiteCode gene set.

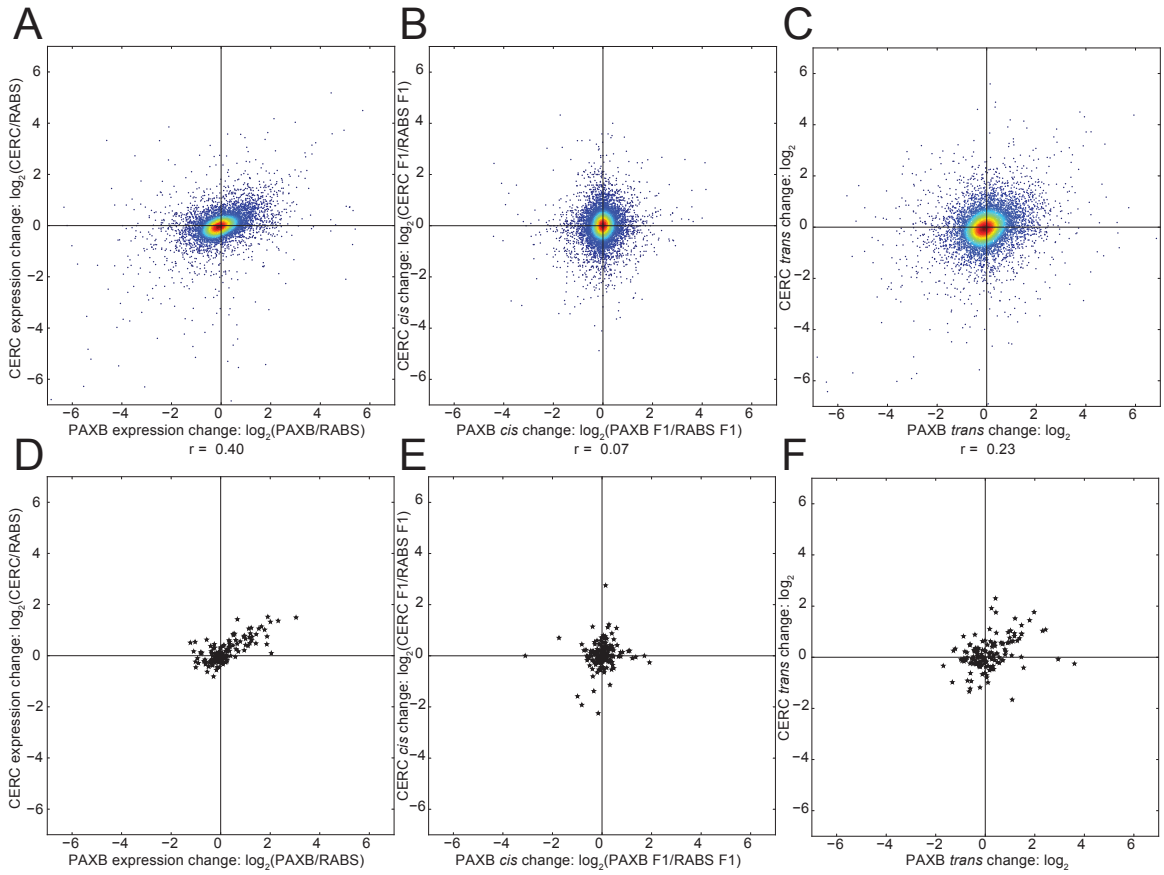

Figure 4.12: **Genome-wide *Trans* changes, not restricted to differentially expressed genes, are more likely to be shared across populations.** (A) Expression changes of genes with quantifiable (i.e. genes with transcripts containing a polymorphic SNP covered by at least 20 reads) hybrid expression in both freshwater populations relative to marine fish, showing significantly correlated changes in gene expression in PAXB_{FW} and CERC_{FW} tooth plates. (B) *cis* regulatory changes of genes with quantifiable hybrid expression in freshwater dental tissue overall do not display correlated evolved changes. (C) *trans* regulatory changes of genes with quantifiable hybrid expression in freshwater dental tissue. Density (color) was estimated with a Gaussian kernel density estimator. (D-F) Similar to A-C, but show only genes in the BiteCode gene set, revealing that these orthologs have evolved highly convergent changes in the two freshwater populations (D), despite non-convergent *cis* regulatory changes (E).

4.4 Discussion

We sought to test the relative contribution of *cis* and *trans* gene regulatory changes during convergent evolution of tooth gain, as well as to ask whether the same or different regulatory changes underlie evolved changes in gene expression during this case of convergent evolution. We quantified the overall regulatory divergence, as well as the specific contribution of *cis* and *trans* changes, between ancestral low-toothed marine and two different independently derived populations of high-toothed freshwater sticklebacks. Similar overall changes in gene expression have evolved in both freshwater populations, especially in orthologs of known dental regulators in mammals. In this system, *trans*-regulatory changes play a larger role than *cis* changes in both populations. Furthermore, *trans* acting changes were much more likely to be shared between freshwater populations than *cis* changes, suggesting the two high-toothed populations evolved their similar gene expression patterns through independent genetic changes.

Convergent evolution of dental gene expression

Convergent evolution at the gene expression level occurs when similar gene expression levels evolve in different populations. Both the $PAXB_{FW}$ and $CERC_{FW}$ stickleback populations have adapted from an ancestral marine form to their current freshwater environments. The genomic nature of their derived changes appears largely divergent, with major axis of variation separating $PAXB_{FW}$ genomes from the geographically proximal marine populations ($LITC_{M}$), as well as the more distant marine ($RABS_{M}$) and $CERC_{FW}$ populations. However, when looking at the gene expression basis of their convergently evolved gain in tooth number, orthologs of genes implicated in mammalian dental development showed strong correlated freshwater gains in expression. This correlation suggests both that sticklebacks deploy conserved genetic circuits regulating tooth formation during tooth replacement, but also that both populations have convergently evolved changes to similar downstream transcriptional circuits resulting in a gain of tooth number.

Though both freshwater populations showed strongly correlated changes in evolved gene expression at the *trans* regulatory level, the *cis* changes were largely not shared across populations. This was especially true for putative dentally expressed genes with evolved expression changes the vast majority of the *trans* but not *cis* expression changes were shared between both freshwater populations. This suggests that the similar freshwater gene expression patterns evolved through independent genetic changes. It is possible that the small number of shared *cis* changes are sufficient to drive the observed changes to the overall *trans* regulatory environments. However previous work has shown that the genetic basis of tooth gain in these two populations is distinct [30] (Cleves et al 2018 under review), and it seems parsimonious that the genetic basis of a gain in dental gene expression is also mostly independent. Thus, convergent freshwater gene expression changes appear to be largely due to distinct, independent population-specific regulatory changes. This finding suggests that

there are many regulatory alleles that are accessible during the evolution of an adaptive trait.

Trans effects dominate

Other studies have used RNA-seq to compare the relative contribution of *cis* and *trans*-regulatory changes in the evolution of gene expression. In mice, evolved gene expression changes in the liver [18] and the retina [61] were driven primarily by *cis*-regulatory changes. In *Drosophila*, work on organismal-wide evolved gene expression changes on the genome-wide level has shown the opposite, with *trans*-regulatory effects playing a larger role in the evolution of gene expression [19,22]. Other studies have found *trans* effects contribute more to intraspecific comparisons, while *cis* effects contribute more to interspecific comparisons [17,20,60]. Consistent with this, we observe *trans* effects dominating in both of our intraspecific comparisons.

Another key distinction could be that *cis*-regulatory effects dominate when looking at more cellularly homogenous tissues, while *trans*-regulatory effects dominate when looking at more heterogeneous tissues. Stickleback tooth plates likely fall into an intermediate category, less heterogenous in cell type composition than a full adult fly or fly head, but more heterogeneous than a specialized tissue such as the mouse retina. Overall, freshwater tooth plates are more morphologically similar to each other than marine, with freshwater tooth plates possessing a larger area, increased tooth number, and decreased intertooth spacing [30,41]. Freshwater tooth plates likely have more similar cell type abundances and compositions (e.g. more developing tooth germs with inner and outer dental epithelia, and odontogenic mesenchyme) compared to each other than to marine tooth plates. Similar cell types tend to have similar gene expression patterns, even when compared across different species [62]. Much of the shared freshwater increase in dental gene expression could be due to an increase in dental cell types in both freshwater populations. As other evolved changes to stickleback morphology have been shown to be due to *cis* regulatory changes to key developmental regulatory genes [8,33,41,63], this *trans* regulatory increase in cell type abundance could be due to a small number of *cis* regulatory changes. These initially evolved developmental regulatory changes could result in similar downstream changes in the developmental landscape, resulting in the shared increase in dental cell types. Consistent with this interpretation, stickleback orthologs of genes known to be expressed during mammalian tooth development were found here to have a much greater incidence of convergently evolved increase in *trans* regulatory gene expression.

Compensatory *cis* and *trans* regulatory changes

Previous studies [17,18] have shown compensatory *cis* and *trans* changes are essential for the evolution of gene expression. These findings are consistent with the idea that the main driving force in the evolution of gene expression is stabilizing selection [59] where compensatory changes to regulatory elements are selected for to maintain optimal gene expression levels.

In both $PAXB_{FW}$ and $CERC_{FW}$ dental tissue, when considering all genes with a quantifiable (i.e. polymorphic and covered by 20 reads, see Methods) *cis* effects, discordant compensatory *cis* and *trans* changes were far more common than concordant ones. This trend could be driven by some initial selection on pleiotropic *trans* changes, followed by selection for compensatory *cis* changes to restore optimal gene expression levels [17,18,22]. However, the *trans*, but not the *cis*, evolved changes in gene expression were highly shared among the two freshwater populations. Thus, collectively our data support a model where two independently derived populations have convergently evolved both similar genome-wide expression levels as well as ecologically relevant morphological changes through different genetic means.

Potential parallels between teeth and hair regeneration

$PAXB_{FW}$ and $CERC_{FW}$ sticklebacks have an increased rate of new tooth formation in adults relative to their marine ancestors [30]. In constantly replacing polyphyodonts, it has been proposed that teeth are replaced through a dental stem cell intermediate [37,38]. A strong candidate gene underlying a large effect $PAXB_{FW}$ tooth quantitative trait locus (QTL) is the secreted ligand *Bone Morphogenetic Protein 6* (*Bmp6*) [41] (Cleves et al 2018 under review), which is also a key regulator of stem cells in the mouse hair follicle [56]. Freshwater dental tissue displayed significantly increased expression of known signature genes of mouse hair follicle stem cells, perhaps reflecting more stem cell niches supporting the higher tooth numbers in freshwater fish. Genes upregulated in freshwater dental tissue also were significantly enriched for GO terms involved in the cell cycle and cell proliferation. Together these findings suggest that both freshwater populations have evolved an increased tooth replacement rate through an increased activity or abundance of their dental stem cells. Additionally, these findings suggest the genetic circuitry regulating mammalian hair and fish tooth replacement might share an ancient, underlying core gene regulatory network.

4.5 Materials and Methods

Stickleback husbandry

Fish from all populations were raised in 110L aquaria in brackish water (3.5g/L Instant Ocean salt, 0.217mL/L 10% sodium bicarbonate) at 18C in 8 hours of light per day. Young fry [standard length (SL) <10 millimeters (mm)] were fed a diet of live Artemia, early juveniles (SL 10 - 20 mm) a combination of live Artemia and frozen Daphnia, and older juveniles (SL >20 mm) and adults a combination of frozen bloodworms and Mysis shrimp. Experiments were approved by the Institutional Animal Care and Use Committee of the University of California-Berkeley (protocol R330).

Skeletal staining and imaging

Sticklebacks were fixed in 10% neutral buffered formalin overnight at 4C. Fish were washed once with water and then stained in 1% KOH, 0.008% Alizarin Red for 24 hours. Following a water rinse, fish were cleared in 0.25% KOH, 50% glycerol for 2-3 weeks. Branchial skeletons were dissected as previously described [64]. Pharyngeal teeth were quantified with fluorescent illumination using a TX2 filter on a Leica DM2500 microscope. Representative tooth plates were created using montage z-stacks on a Leica M165 FC using the RhodB filter. Adult fish were imaged using a Canon Powershot S95. Some tooth count data from the $CERC_{FW}$, $RABS_{M}$, and $PAXB_{FW}$ populations; n = 11, 13, 29, respectively, have been previously published [30].

DNA preparation and genome resequencing

Caudal fin tissue was placed into 600μL tail digestion buffer [10mM Tris pH 8.0, 100mM NaCl, 10mM EDTA, 0.05% SDS, 2.5μL ProK (Ambion AM2546)] for 12 hours at 55C. Following addition of 600μL of 1:1 phenol:chloroform solution and an aqueous extraction, DNA was precipitated with the addition of 1ml 100% ethanol, centrifuged, washed with 75% ethanol, and resuspended in water. 50ng of purified genomic DNA was used as input for the Nextera Library prep kit (Illumina FC-121-1031), and barcoded libraries were constructed following the manufacturers instructions. Library quality was verified using an Agilent Bioanalyzer. Libraries were pooled and sequenced on an Illumina HiSeq 2000 (see Table 4.1 for details), resulting in a mean of 52.8 million reads per sample, with a max of 70.3 million reads and a minimum of 39 million reads (Table 4.1).

RNA purification and creation of RNA-seq libraries

Late juvenile stage female sticklebacks (SL 40mm) were euthanized in 0.04% Tricaine. Dissected [64] bilateral ventral pharyngeal tooth plates were placed into 500μL TRI reagent, then incubated at room temperature for 5 minutes. Following addition of 100μL of chloroform, a further 10 minute incubation and centrifugation, the aqueous layer was extracted. Following addition of 250μL isopropyl alcohol and 10 minute incubation, RNA was precipitated by centrifugation, washed with 75% EtOH, and dissolved in 30μL of DEPC-treated water. RNA integrity was assayed by an Agilent Bioanalyzer. 500ng of RNA from each fish was used as input to the Illumina stranded TruSeq polyA RNA kit (Illumina RS-122-2001), and libraries were constructed following the manufacturers instructions. Library quality was analyzed on an Agilent Bioanalyzer, and libraries were pooled and sequenced on an Illumina HiSeq2000 (see Table 4.2). We obtained a mean of 84.1 million reads among the parental samples, with a max of 91.0 million and a minimum of 78.6 million (Table 4.2).

Gene expression quantification and analysis

RNA-seq reads were mapped to the stickleback reference genome [31] using the STAR aligner [65] (version 2.3, parameters = –alignIntronMax 100000 –alignMatesGapMax 200000 –outFilterMultimapNmax 20 –outFilterMismatchNmax 999 –outFilterMismatchNoverLmax 0.04 –outFilterType BySJout), using ENSEMBL genes release 85 as a reference transcriptome. The resulting SAM files were sorted and indexed using Samtools version 0.1.18 [66], PCR duplicates were removed, read groups added and mate pair information fixed using Picard tools (version 1.51) (http://broadinstitute.github.io/picard/) with default settings. Gene expression was quantified with the Cufflinks suite (v 2.2.1) [47,5355] using ENSEMBL genes as a reference transcriptome, with gene expression quantified with cuffquant (-u –library-type fr-firststrand) and normalized with cuffnorm. Differentially expressed genes were found using cuffdiff2, with parameters (-u –FDR .1 –library-type fr-firststrand, using the reference genome for bias correction). Genes with a mean expression less than 0.1 FPKM were filtered from further analysis.

Gene set and gene ontology enrichment

The BiteCode gene set was generated by combining all genes in the BiteIt (http://bite-it.helsinki.fi/) or ToothCODE (http://compbio.med.harvard.edu/ToothCODE/) [36] databases. Stickleback orthologs or co-orthologs were found using the annotated names of ENSEMBL stickleback genes. Gene set expression change statistical enrichment was done as previously described [67]. Briefly, a t-test was performed for each gene to test for a difference in mean expression between the two treatments. The resulting t-values were subject to a 1-sample t-test, with the null model that the mean of the t-values was 0. Cutoffs were validated using 10,000 bootstrapped replicate gene sets drawn from the same gene expression matrix. Stickleback orthologs of mouse or human genes were determined using annotated ENSEMBL orthologs. Sorted lists of genes, ranked by log2 expression change in $PAXB_{FW}$ dental tissue relative to marine, $CERC_{FW}$relative to marine, or the mean of $CERC_{FW}$ and $PAXB_{FW}$ relative to marine, were generated using the measured gene expression data. Gene Ontology enrichment was done using Gorilla [68,69], and results were visualized using REVIGO [70].

Detection of genomic and transcriptomic variants

Genomic resequencing reads were aligned to the stickleback reference genome [31] using the bwa aln and bwa sampe modules of the Burrows-Wheeler Alignment tool (v 0.6.0-r85) [71]. Resulting SAM files were converted to BAM files, sorted and indexed by Samtools version 0.1.18 [66], with PCR duplicates removed by Picard tools. GATK's (v3.2-2) IndelRealigner (parameter: '-LOD 0.4'), BaseRecalibrator, and PrintReads were used on the resulting BAM files. BAM files from the above RNA-seq alignment were readied for genotype calling using GATK's SplitNCigarReads, BaseRecalibrator, and PrintReads. Finally, the UnifiedGenotyper was used to call variants from the RNA-seq and DNA-seq BAM files, with parameters

(-stand call conf 30 -stand emit conf 30 -U ALLOW N CIGAR READS –genotype likelihoods model BOTH) [43,45]. This analysis identified a set of 8,341,326 variants.

Principal components analysis of the genome-wide set of variants was performed by first filtering all multiallelic variants or variants with a missing genotype, resulting in a set of 1,690,729 variants. PCA was performed using FactoMiner [46] and a set of custom R scripts. Phylogenetic trees were constructed using the set of variants, downsampled to 67,507 SNPs (no indels) for use with BEAST and SNAPP [72,73]. We constructed phylogenies using SNAPP, estimating substitution rate and proportion invariant from the data, and ran 1 million generations of MCMC simulations. The best tree was picked with TreeAnnotator and visualized with FigTree.

To accurately phase RNA-seq data from F_1 hybrids, pseudo-transcriptomes were created for each hybrid. The pseudo-transcriptomes consist of the predicted sequence for each allele within an F_1 hybrid, with all predicted splicing variants of a gene collapsed to a single transcript. A variant was added to the pseudo-transcriptome if and only if it was homozygous in the sequenced parents (or parents sibling in the case of the $RABS_M$ parent of the $CERC_{FW}$ x $RABS_M$ F_1 hybrids) and called heterozygous in the F_1 hybrid.

Cis and *trans* regulatory divergence quantification

RNA-seq reads from F_1 hybrid sticklebacks were aligned to the individuals pseudo-transcriptome using STAR (v 2.3) with the parameters: –outFilterMultimapNmax 1 and –outFilterMultimapScoreRange 1. By only looking at uniquely aligning reads, we ensured we only considered reads which overlapped a heterozygous variant site. Counting these unique reads minimizes double counting a single read that supports two different variant positions. Total *cis* divergence in each F_1 hybrid was quantified by comparing the number of reads mapping uniquely to each allele in the pseudo-transcriptome.

Following *cis* divergence quantification in all F_1 hybrids, we considered the overall *cis* change in the different freshwater populations. Genes which only had 20 or fewer uniquely mapping reads across all replicates were filtered from further analysis. We excluded genes with more than a 32-fold change, as a manual inspection revealed these to be either genotyping errors or mitochondrial genes. Reported *cis* ratios were calculated by comparing the ratio of uniquely mapped freshwater reads to uniquely mapped marine reads. Evolved *trans* changes were quantified as the difference between the log of the overall gene expression change between the freshwater and marine parents and the log of measured *cis* freshwater expression change. Percent *cis* change was calculated as the absolute value of the log of the *cis* change divided by the sum of the absolute value of the log of the *cis* change and the absolute value of the log of the trans change. Statistical significance of *cis* changes was determined by a binomial test comparing overall reads mapping to the freshwater allele to a null model of no *cis* divergence, with a false discovery rate of 1% applied using the Benjamini-Hochberg method. Statistical significance of trans changes was determined by a G-test, comparing the expected (based on the measured *cis* change) and observed ratios of marine and freshwater, with a 1% false discovery rate.

Data Availability

All sequencing reads are available on the Sequence Read Archive (SRP142616). All scripts used for analysis are available on GitHub (https://github.com/trahsemaj/Conv_Evo_Gene_Exp).

Acknowledgements

We thank Nikunj Donde for collecting a subset of tooth count data.

Funding statement

This work was supported by NIH R01 DE021475 (to C.T.M.), NIH genomics training grant 5T32HG000047-15 (to J.C.H.), and NSF Graduate Research Fellowship (to N.A.E.). This work used the Vincent J. Coates Genomics Sequencing Laboratory at UC Berkeley, supported by NIH S10 Instrumentation Grants S10RR029668 and S10RR027303. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

4.6 References

1. Banerji J, Rusconi S, Schaffner W. Expression of a beta-globin gene is enhanced by remote SV40 DNA sequences. Cell. 1981;27: 299308.
2. Ong C-T, Corces VG. Enhancer function: new insights into the regulation of tissue-specific gene expression. Nat Rev Genet. 2011;12: 283293. doi:10.1038/nrg2957
3. Stern DL. Evolutionary developmental biology and the problem of variation. Evolution. 2000;54: 10791091.
4. Britten RJ, Davidson EH. Repetitive and non-repetitive DNA sequences and a speculation on the origins of evolutionary novelty. The Quarterly Review of Biology. 1971;46: 111138.
5. King MC, Wilson AC. Evolution at two levels in humans and chimpanzees. Science. 1975;188: 107116.
6. Gompel N, Prudhomme B, Wittkopp PJ, Kassner VA, Carroll SB. Chance caught on the wing: cis-regulatory evolution and the origin of pigment patterns in Drosophila. Nature. 2005;433: 481487. doi:10.1038/nature03235
7. Manceau M, Domingues VS, Mallarino R, Hoekstra HE. The developmental role of agouti in color pattern evolution. Science. 2011;331: 10621065. doi:10.1126/science.1200684
8. OBrown NM, Summers BR, Jones FC, Brady SD, Kingsley DM. A recurrent regulatory change underlying altered expression and Wnt response of the stickleback armor plates gene EDA. eLife. 2015;4: e05290. doi:10.7554/eLife.05290
9. Colosimo PF, Hosemann KE, Balabhadra S, Villarreal G, Dickson M, Grimwood J, et al. Widespread parallel evolution in sticklebacks by repeated fixation of ectodysplasin alleles.

Science. 2005;307: 19281933. doi:10.1126/science.1107239
10. Indjeian VB, Kingman GA, Jones FC, Guenther CA, Grimwood J, Schmutz J, et al. Evolving new skeletal traits by cis-regulatory changes in bone morphogenetic proteins. Cell. 2016;164: 4556. doi:10.1016/j.cell.2015.12.007
11. Schor IE, Degner JF, Harnett D, Cannav E, Casale FP, Shim H, et al. Promoter shape varies across populations and affects promoter evolution and expression noise. Nature Genetics. 2017;49: 550. doi:10.1038/ng.3791
12. McCoy RC, Wakefield J, Akey JM. Impacts of Neanderthal-introgressed sequences on the landscape of human gene expression. Cell. 2017;168: 916-927.e12. doi:10.1016/j.cell.2017.01.038
13. Ritter DI, Dong Z, Guo S, Chuang JH. Transcriptional enhancers in protein-coding rxons of vertebrate developmental genes. PLOS ONE. 2012;7: e35202. doi:10.1371/journal.pone.0035202
14. Gibbons TC, Metzger DCH, Healy TM, Schulte PM. Gene expression plasticity in response to salinity acclimation in threespine stickleback ecotypes from different salinity habitats. Mol Ecol. 2017;26: 27112725. doi:10.1111/mec.14065
15. Yao C, Joehanes R, Johnson AD, Huan T, Liu C, Freedman JE, et al. Dynamic role of trans regulation of gene expression in relation to complex traits. The American Journal of Human Genetics. 2017;100: 571580. doi:10.1016/j.ajhg.2017.02.003
16. Wittkopp PJ, Haerum BK, Clark AG. Evolutionary changes in cis and trans gene regulation. Nature. 2004;430: 8588. doi:10.1038/nature02698
17. Mack KL, Campbell P, Nachman MW. Gene regulation and speciation in house mice. Genome Res. 2016; gr.195743.115. doi:10.1101/gr.195743.115
18. Goncalves A, Leigh-Brown S, Thybert D, Stefflova K, Turro E, Flicek P, et al. Extensive compensatory cis-trans regulation in the evolution of mouse gene expression. Genome Res. 2012;22: 23762384. doi:10.1101/gr.142281.112
19. McManus CJ, Coolon JD, Duff MO, Eipper-Mains J, Graveley BR, Wittkopp PJ. Regulatory divergence in Drosophila revealed by mRNA-seq. Genome Res. 2010;20: 816825. doi:10.1101/gr.102491.109
20. Lemmon ZH, Bukowski R, Sun Q, Doebley JF. The role of cis regulatory evolution in maize domestication. PLOS Genet. 2014;10: e1004745. doi:10.1371/journal.pgen.1004745
21. Osada N, Miyagi R, Takahashi A. cis- and trans-regulatory effects on gene expression in a natural population of Drosophila melanogaster. Genetics. 2017; 117.201459. doi:10.1534/genetics.117.201459
22. Coolon JD, McManus CJ, Stevenson KR, Graveley BR, Wittkopp PJ. Tempo and mode of regulatory evolution in Drosophila. Genome Res. 2014;24: 797808. doi:10.1101/gr.163014.113
23. Peichel CL, Marques DA. The genetic and molecular architecture of phenotypic diversity in sticklebacks. Phil Trans R Soc B. 2017;372: 20150486. doi:10.1098/rstb.2015.0486
24. Pritchard VL, Viitaniemi HM, McCairns RJS, Meril J, Nikinmaa M, Primmer CR, et al. Regulatory architecture of gene expression variation in the threespine stickleback Gasterosteus aculeatus. G3. 2017;7: 165178. doi:10.1534/g3.116.033241

25. Ishikawa A, Kusakabe M, Yoshida K, Ravinet M, Makino T, Toyoda A, et al. Different contributions of local- and distant-regulatory changes to transcriptome divergence between stickleback ecotypes. Evolution. 2017;71: 565581. doi:10.1111/evo.13175
26. Di Poi C, Blanger D, Amyot M, Rogers S, Aubin-Horth N. Receptors rather than signals change in expression in four physiological regulatory networks during evolutionary divergence in threespine stickleback. Mol Ecol. 2016;25: 34163427. doi:10.1111/mec.13690
27. Huang Y, Chain FJJ, Panchal M, Eizaguirre C, Kalbe M, Lenz TL, et al. Transcriptome profiling of immune tissues reveals habitat-specific gene expression between lake and river sticklebacks. Mol Ecol. 2016;25: 943958. doi:10.1111/mec.13520
28. Bell MA, Foster SA. The evolutionary biology of the threespine stickleback. Oxford University Press; 1994.
29. Hagen DW, Gilbertson LG. Geographic variation and environmental selection in Gasterosteus aculeatus L. in the Pacific Northwest, America. Evolution. 1972;26: 3251. doi:10.2307/2406981
30. Ellis NA, Glazer AM, Donde NN, Cleves PA, Agoglia RM, Miller CT. Distinct developmental genetic mechanisms underlie convergently evolved tooth gain in sticklebacks. Development. 2015;142: 24422451. doi:10.1242/dev.124248
31. Jones FC, Grabherr MG, Chan YF, Russell P, Mauceli E, Johnson J, et al. The genomic basis of adaptive evolution in threespine sticklebacks. Nature. 2012;484: 5561. doi:10.1038/nature10944
32. Erickson PA, Glazer AM, Killingbeck EE, Agoglia RM, Baek J, Carsanaro SM, et al. Partially repeatable genetic basis of benthic adaptation in threespine sticklebacks: repeatable evolution in benthic sticklebacks. Evolution. 2016;70: 887902. doi:10.1111/evo.12897
33. Chan YF, Marks ME, Jones FC, Villarreal G, Shapiro MD, Brady SD, et al. Adaptive evolution of pelvic reduction in sticklebacks by recurrent deletion of a Pitx1 enhancer. Science. 2010;327: 302305. doi:10.1126/science.1182213
34. Shapiro MD, Marks ME, Peichel CL, Blackman BK, Nereng KS, Jnsson B, et al. Genetic and developmental basis of evolutionary pelvic reduction in threespine sticklebacks. Nature. 2004;428: 717. doi:10.1038/nature02415
35. Pispa J, Thesleff I. Mechanisms of ectodermal organogenesis. Dev Biol. 2003;262: 195205.
36. OConnell DJ, Ho JWK, Mammoto T, Turbe-Doan A, OConnell JT, Haseley PS, et al. A WNT-BMP feedback circuit controls intertissue signaling dynamics in tooth organogenesis. Sci Signal. 2012;5: ra4. doi:10.1126/scisignal.2002414
37. Jernvall J, Thesleff I. Tooth shape formation and tooth renewal: evolving with the same signals. Development. 2012;139: 34873497. doi:10.1242/dev.085084
38. Tucker AS, Fraser GJ. Evolution and developmental diversity of tooth regeneration. Seminars in Cell & Developmental Biology. doi:10.1016/j.semcdb.2013.12.013
39. Bloomquist RF, Parnell NF, Phillips KA, Fowler TE, Yu TY, Sharpe PT, et al. Coevolutionary patterning of teeth and taste buds. PNAS. 2015;112: E5954E5962. doi:10.1073/pnas.1514298112
40. Martin KJ, Rasch LJ, Cooper RL, Metscher BD, Johanson Z, Fraser GJ. Sox2+ progeni-

tors in sharks link taste development with the evolution of regenerative teeth from denticles. Proc Natl Acad Sci. 2016;113: 1476914774. doi:10.1073/pnas.1612354113
41. Cleves PA, Ellis NA, Jimenez MT, Nunez SM, Schluter D, Kingsley DM, et al. Evolved tooth gain in sticklebacks is associated with a cis-regulatory allele of Bmp6. PNAS. 2014;111: 1391213917. doi:10.1073/pnas.1407567111
42. Langmead B, Salzberg SL. Fast gapped-read alignment with Bowtie 2. Nat Meth. 2012;9: 357359. doi:10.1038/nmeth.1923
43. McKenna A, Hanna M, Banks E, Sivachenko A, Cibulskis K, Kernytsky A, et al. The Genome Analysis Toolkit: A MapReduce framework for analyzing next-generation DNA sequencing data. Genome Res. 2010;20: 12971303. doi:10.1101/gr.107524.110
44. DePristo MA, Banks E, Poplin R, Garimella KV, Maguire JR, Hartl C, et al. A framework for variation discovery and genotyping using next-generation DNA sequencing data. Nat Genet. 2011;43: 491498. doi:10.1038/ng.806
45. Van der Auwera GA, Carneiro MO, Hartl C, Poplin R, Del Angel G, Levy-Moonshine A, et al. From FastQ data to high confidence variant calls: the Genome Analysis Toolkit best practices pipeline. Curr Protoc Bioinformatics. 2013;43: 11.10.1-33. doi:10.1002/0471250953.bi1110s43
46. Sbastien L, Julie Josse, Franois Husson. FactoMineR: an R package for multivariate analysis — L — Journal of Statistical Software. Journal of Statistical Software. 2008;25. Available: https://www.jstatsoft.org/article/view/v025i01
47. Trapnell C, Hendrickson DG, Sauvageau M, Goff L, Rinn JL, Pachter L. Differential analysis of gene regulation at transcript resolution with RNA-seq. Nat Biotech. 2013;31: 4653. doi:10.1038/nbt.2450
48. Gene expression in tooth (WWW database), http://bite-it.helsinki.fi. Developmental Biology Programme of the University of Helsink; 1996.
49. Fraser GJ, Hulsey CD, Bloomquist RF, Uyesugi K, Manley NR, Streelman JT. An ancient gene network is co-opted for teeth on old and new jaws. PLoS Biol. 2009;7: e1000031. doi:10.1371/journal.pbio.1000031
50. Fraser GJ, Bloomquist RF, Streelman JT. Common developmental pathways link tooth shape to regeneration. Dev Biol. 2013;377: 399414. doi:10.1016/j.ydbio.2013.02.007
51. Stock DW. Zebrafish dentition in comparative context. J Exp Zool B Mol Dev Evol. 2007;308: 523549. doi:10.1002/jez.b.21187
52. Jackman WR, Draper BW, Stock DW. Fgf signaling is required for zebrafish tooth development. Dev Biol. 2004;274: 139157. doi:10.1016/j.ydbio.2004.07.003
53. Trapnell C, Williams BA, Pertea G, Mortazavi A, Kwan G, van Baren MJ, et al. Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation. Nat Biotech. 2010;28: 511515. doi:10.1038/nbt.1621
54. Roberts A, Pimentel H, Trapnell C, Pachter L. Identification of novel transcripts in annotated genomes using RNA-Seq. Bioinformatics. 2011; btr355. doi:10.1093/bioinformatics/btr355
55. Roberts A, Trapnell C, Donaghey J, Rinn JL, Pachter L. Improving RNA-Seq expression estimates by correcting for fragment bias. Genome Biology. 2011;12: R22. doi:10.1186/gb-

2011-12-3-r22
56. Kandyba E, Leung Y, Chen Y-B, Widelitz R, Chuong C-M, Kobielak K. Competitive balance of intrabulge BMP/Wnt signaling reveals a robust gene network ruling stem cell homeostasis and cyclic activation. Proc Natl Acad Sci USA. 2013;110: 13511356. doi:10.1073/pnas.1121312110
57. Yasuoka A, Aihara Y, Matsumoto I, Abe K. Phospholipase C-beta 2 as a mammalian taste signaling marker is expressed in the multiple gustatory tissues of medaka fish, Oryzias latipes. Mechanisms of Development. 2004;121: 985989. doi:10.1016/j.mod.2004.03.009
58. Bachmanov AA, Beauchamp GK. Taste Receptor Genes. Annu Rev Nutr. 2007;27: 389414. doi:10.1146/annurev.nutr.26.061505.111329
59. Gilad Y, Oshlack A, Rifkin SA. Natural selection on gene expression. Trends in Genetics. 2006;22: 456461. doi:10.1016/j.tig.2006.06.002
60. Wittkopp PJ, Haerum BK, Clark AG. Regulatory changes underlying expression differences within and between Drosophila species. Nat Genet. 2008;40: 346350. doi:10.1038/ng.77
61. Shen SQ, Turro E, Corbo JC. Hybrid mice reveal parent-of-origin and cis- and trans-regulatory effects in the retina. PLOS ONE. 2014;9: e109382.
doi:10.1371/journal.pone.0109382
62. Brawand D, Soumillon M, Necsulea A, Julien P, Csrdi G, Harrigan P, et al. The evolution of gene expression levels in mammalian organs. Nature. 2011;478: 343348. doi:10.1038/nature10532
63. Chen H, Capellini TD, Schoor M, Mortlock DP, Reddi AH, Kingsley DM. Heads, shoulders, elbows, knees, and toes: modular Gdf5 enhancers control different joints in the vertebrate skeleton. PLOS Genetics. 2016;12: e1006454. doi:10.1371/journal.pgen.1006454
64. Ellis NA, Miller CT. Dissection and flat-mounting of the threespine stickleback branchial skeleton. Journal of Visualized Experiments. 2016; doi:10.3791/54056
65. Dobin A, Davis CA, Schlesinger F, Drenkow J, Zaleski C, Jha S, et al. STAR: ultrafast universal RNA-seq aligner. Bioinformatics. 2012; bts635. doi:10.1093/bioinformatics/bts635
66. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, et al. The sequence alignment/map format and SAMtools. Bioinformatics. 2009;25: 20782079.
doi:10.1093/bioinformatics/btp352
67. Irizarry RA, Wang C, Zhou Y, Speed TP. Gene set enrichment analysis made simple. Stat Methods Med Res. 2009;18: 565575. doi:10.1177/0962280209351908
68. Eden E, Navon R, Steinfeld I, Lipson D, Yakhini Z. GOrilla: a tool for discovery and visualization of enriched GO terms in ranked gene lists. BMC Bioinformatics. 2009;10: 48. doi:10.1186/1471-2105-10-48
69. Eden E, Lipson D, Yogev S, Yakhini Z. Discovering motifs in ranked lists of DNA sequences. PLOS Comput Biol. 2007;3: e39. doi:10.1371/journal.pcbi.0030039
70. Supek F, Bonjak M, kunca N, muc T. REVIGO summarizes and visualizes long lists of gene ontology terms. PLOS ONE. 2011;6: e21800. doi:10.1371/journal.pone.0021800
71. Li H, Durbin R. Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics. 2009;25: 17541760. doi:10.1093/bioinformatics/btp324
72. Bouckaert R, Heled J, Khnert D, Vaughan T, Wu C-H, Xie D, et al. BEAST 2: A Soft-

ware Platform for Bayesian Evolutionary Analysis. PLOS Computational Biology. 2014;10: e1003537. doi:10.1371/journal.pcbi.1003537

73. Bryant D, Bouckaert R, Felsenstein J, Rosenberg NA, RoyChoudhury A. Inferring Species Trees Directly from Biallelic Genetic Markers: Bypassing Gene Trees in a Full Coalescent Analysis. Mol Biol Evol. 2012;29: 19171932. doi:10.1093/molbev/mss086

Chapter 5

Modular and convergent evolution of a gain in stickleback tooth number through distinct sets of loci

James C. Hart[1], Nathan T. Cramer[1], Siegen A. McKellar[1], Tyler A. Square[1], Andrew H. Chang[1], Nicholas A. Ellis[1], Jiyeon Baek[1], Andrew M. Glazer[1], Tiana Kohlsdorf[2], and Craig T. Miller[1].
[1]Department of Molecular and Cell Biology, University of California-Berkeley, CA, USA
[2]Departamento de Biologia, University of São Paulo, São Paulo, Brazil

5.1 Abstract

Developmental regulatory networks are often reused and redeployed during the production of similar organs in different anatomical contexts. Dental development shares similar gene expression patterns conserved not only between teeth in diverse species, but also between regionally distinct oral and pharyngeal teeth. Freshwater stickleback populations (*Gasterosteus aculeatus*) have evolved increased tooth number on both their oral and pharyngeal jaws. Comparative QTL mapping of evolved changes in freshwater oral and pharyngeal dentitions reveal a surprisingly region-specific genetic architecture of tooth gain. Evolved tooth gain is associated with either regulatory alleles of *Bmp6* or *Plod2* and *Pitx2* in multiple high-toothed freshwater populations. Here we show all of these genes are required to regulate tooth development in the oral and pharyngeal jaw. We find that the modular nature of evolved increases in dentition is driven by regionally distinct regulatory changes of critical regulators of tooth development.

5.2 Introduction

The genetic basis of evolved changes in morphology has been proposed to more often result from modifications to tissue-specific *cis*-regulatory elements rather than protein-coding sequences [1-5]. For genes with multiple expression domains regulated by multiple *cis*-regulatory elements, mutations affecting the activity of a *cis*-regulatory element typically perturb expression only within the subset of a genes expression domains regulated by that element. In contrast, mutations affecting coding sequences would affect all tissues the gene is expressed in, barring differences in mRNA splicing. Recent work in evolutionary genetics has supported the hypothesis that evolution favors *cis*-regulatory alleles to enact morphological change [48]. The genetic bases of evolved differences in patterns of trichomes in *Drosophila* embryos are a series of modular changes to distinct enhancers of *svb* [9]. Evolved differences in Drosophila wing spot pattering are due in part to changes within a *cis*-regulatory element of the *yellow* gene [10]. In stickleback fish, differences in the number and size of armored plates [11,12], loss of pelvic spines [13,14], the size of pharyngeal jaw bones [15], and the number of pharyngeal teeth [16] are linked to evolved changes in *cis* regulation.

Many similar yet distinct traits have a shared developmental genetic underpinning. Teeth on the oral and pharyngeal jaw of fish share common developmental genetic bases, including shared gene expression patterns [17-19]. Indeed, genetic mapping studies in cichlid fish found a shared genetic basis of oral and pharyngeal tooth gain [20]. However, evolution is able to act on each of these tooth domains in a modular fashion for instance, stickleback fish show strong sexual dimorphism for tooth number on their oral, but not their pharyngeal, jaw [21,22].

Ancestral marine populations of sticklebacks have colonized freshwater lakes and streams throughout the Northern Hemisphere. These independently derived freshwater populations adapted in isolation to a set of shared ecological conditions, resulting in a series of evolution-

ary replicates. Many freshwater populations evolve similar morphological changes, including a gain in pharyngeal tooth number [16,23]. Due to their presence in multiple independently colonized populations [23], this trait appears to be adaptive in freshwater environments, potentially due to shifts in diet.

In sticklebacks, freshwater adaptive traits appear to have evolved through a mix of shared and independently derived genetic changes. The low-plated phenotype is due to repeated reuse of an *Eda* allele in multiple freshwater populations [11,24]. The loss of pelvic spines is due to repeated, but independently derived, deletions in an enhancer of *Pitx1* [13]. It remains an open question as to how frequently evolution reuses existing genetic variation during convergent morphological adaptation.

Previously we reported that the convergently evolved increase in pharyngeal tooth number in two different freshwater populations appears to be due to distinct genetic changes [23], including a *cis*-regulatory allele of *Bmp6* [16]. Here we tested for evidence of a shared genetic basis of pharyngeal tooth gain in eight freshwater stickleback populations though a combination of genetic analysis and whole genome sequencing. We used genetic mapping to elucidate the shared and modular genetic underpinnings of an evolved gain in both oral and pharyngeal teeth. Additionally, we identified two new candidate genes for underlying evolved gains in pharyngeal teeth, *Pitx2* and *Plod2*, using gene expression data from diverse populations and genetic backgrounds. We then used genome editing to functionally assay these candidates roles during tooth development in both oral and pharyngeal tooth domains.

5.3 Results

Evolved tooth gain in multiple freshwater stickleback populations

We previously described an evolved gain of teeth in freshwater stickleback populations in the pharyngeal jaw, which includes one ventral (VTP) and two dorsal (DTP1 and DTP2) tooth plates [16,19,23]. Less is known about the evolution of teeth in these populations in the oral jaw, which comprise teeth attached to the dentary and premaxilla bones [19]. We aimed to test whether evolved tooth number increases in the pharyngeal jaw would be accompanied by concomitant increases in the oral jaw. To test this, we quantified tooth number of fish derived from many distinct localities raised in the lab, including freshwater populations from Paxton Lake, British Columbia ($PAXB_{FW}$), Fishtrap Creek, Washington (FTC_{FW}), Cerrito Creek, California ($CERC_{FW}$), and marine populations from Japan ($JAMA_{M}$), the Little Campbell River, British Columbia ($LITC_{M}$), and Rabbit Slough, Alaska ($RABS_{M}$) (Fig 5.1A-F).

We observed significant changes in freshwater pharyngeal dentition, with all freshwater populations displaying significant increases in teeth on their ventral tooth plates (VTP) (Fig 5.1E). $PAXB_{FW}$ and $CERC_{FW}$ exhibited evolved tooth gain on both dorsal tooth plates (DTP1 and DTP2), while, surprisingly, FTC_{FW} displayed an evolved decrease in tooth number on DTP1 and no detectable change on DTP2 (Fig 5.1B,C). In line with previously reported results, we found strong sexual dimorphism for tooth number in the oral jaw, with

males showing an increased tooth number on both their dentary and premaxilla relative to females (Fig 5.2A,B) [21]. We found no increase in tooth number on freshwater premaxilla relative to marine fish (Fig 5.1A), though both $CERC_{FW}$ and $PAXB_{FW}$ showed an increased tooth number on their dentary (Fig 5.1D). In line with this modest gain of oral teeth in some freshwater populations, we observed strong correlations for tooth fields within the oral jaw and within the pharyngeal jaw, but lower correlations between tooth fields on the oral jaw compared to the pharyngeal jaw (Fig 5.3). Together, this demonstrates that all tooth fields are able to evolve tooth number modifications while tooth number in other fields can remain static or even change in the other direction, i.e. tooth number changes can evolve in a modular fashion across these five distinct tooth fields.

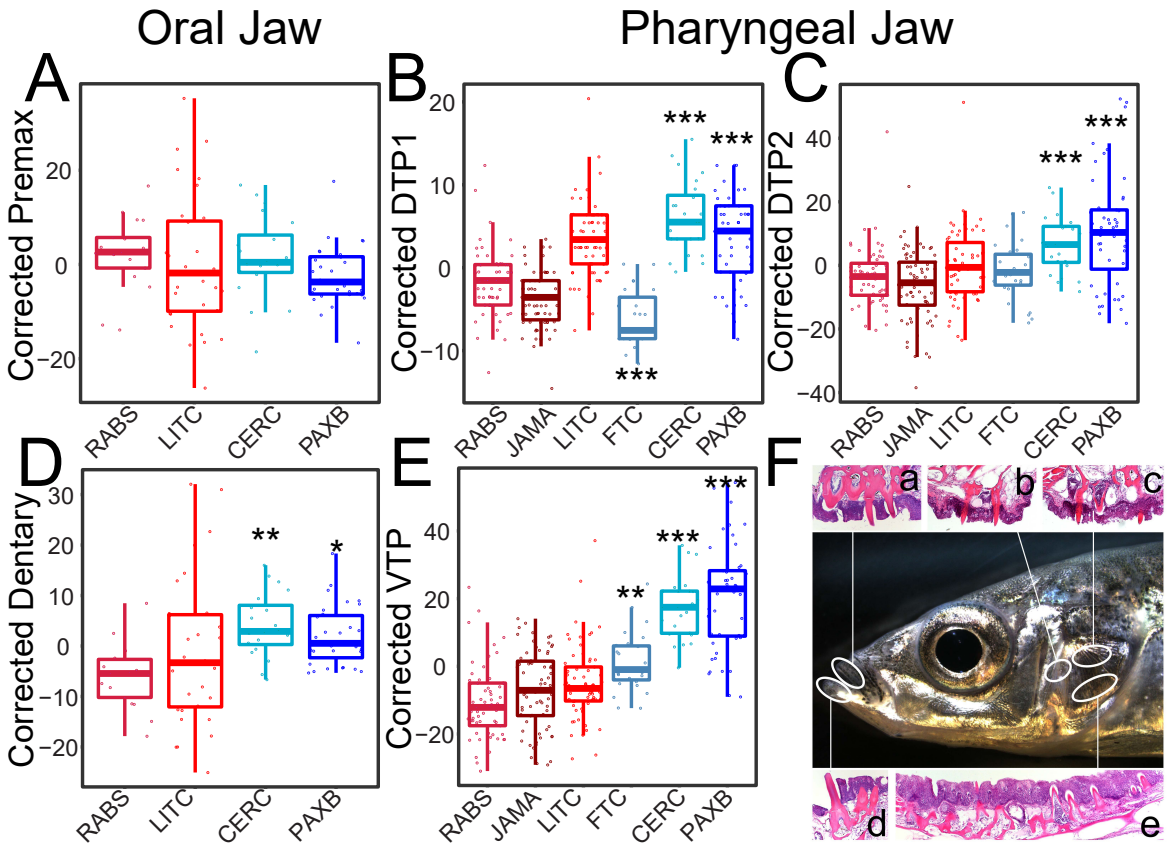

Figure 5.1: **Evolved tooth gain in multiple independently derived freshwater stickleback populations.** (A-E) Dorsal (A-C) and ventral (D-E) oral (A, D) and pharyngeal (B,C,E) tooth counts from lab raised adult sticklebacks, corrected for the effects of size and, for oral teeth, sex. Marine populations (shown in red colors) are $RABS_M$ (Rabbit Slough, Alaska), $JAMA_M$ (Japanese Marine), and $LITC_M$ (Little Campbell River, British Columbia, Canada); freshwater populations (shown in blue colors) are FTC_{FW} (Fishtrap Creek, Washington State), $CERC_{FW}$ (Cerrito Creek, California), and $PAXB_{FW}$ (Paxton Lake, Canada) (subscripts denote marine (M) or freshwater (FW) populations. (F) Stickleback head showing the locations of tooth fields, with lower case letters (a-e) corresponding to the tooth fields presented in (A-E), and showing a histological section of teeth from that field. Evolved freshwater tooth gain or loss relative to marine populations was tested in a linear model, $P < 0.05$: *, $P < 0.01$: **, $P < 0.001$: ***.

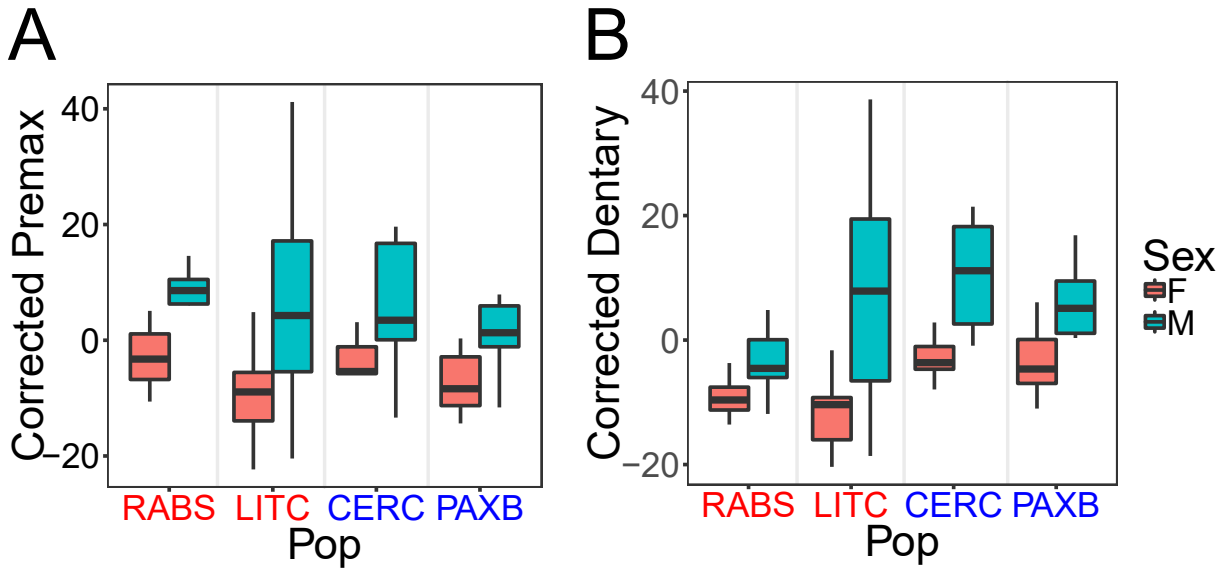

Figure 5.2: **Sexual dimorphism of tooth number on the oral jaw of marine and freshwater stickleback populations.** Size-corrected tooth number of the premaxilla (A) and dentary (B) of adult lab-raised sticklebacks from marine (red, $RABS_M$ and $LITC_M$) and freshwater (blue, $CERC_{FW}$and $PAXB_{FW}$) populations. Sticklebacks display strong sexual dimorphism for oral tooth number, with males displaying greater tooth numbers on their premaxilla (+10.9 teeth, P=1.55e-6, similar to result found in [21]) and dentary (+13.2 teeth, P=2.23e-9).

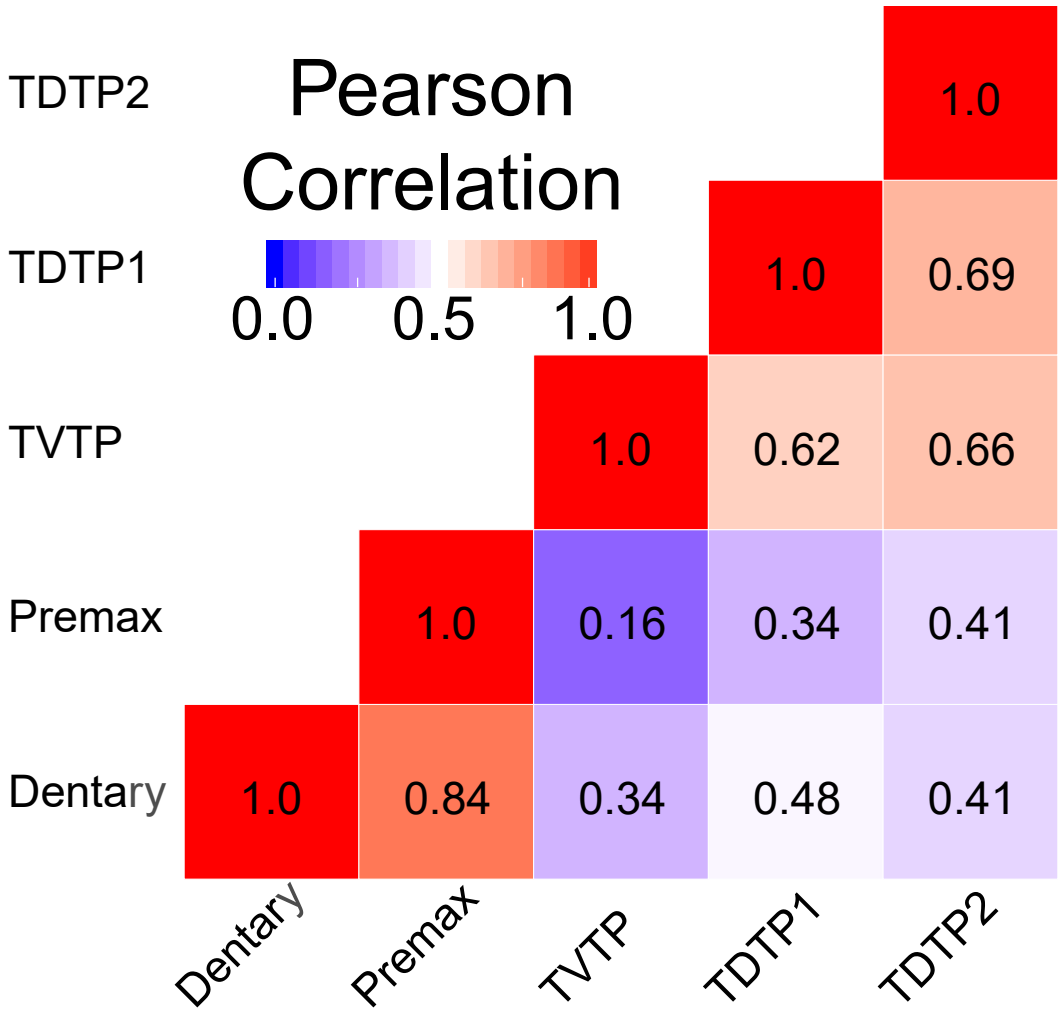

Figure 5.3: **Distinct evolved changes in dentition of the oral and pharyngeal jaw.** Correlations of size and sex corrected tooth counts from two marine and two freshwater populations (same populations as in Fig 5.2). Tooth number is highly correlated for comparisons within the same jaw (Pearsons r $>$ 0.62), and less so between oral and pharyngeal tooth domains (r $<$ 0.48).

Genomic evolution of freshwater sticklebacks

We sought to understand the genetic and genomic basis of stickleback freshwater adaptation. We aimed to determine the degree of genomic diversity among pan-Pacific geographically diverse marine populations as compared to the freshwater populations, which were restricted to the west coast of North America (Fig 5.4A). We resequenced the genomes of 24 fish

from 3 marine and 8 freshwater populations, each of which was the grandparent of an F_2 cross (Table 5.1). After genome-wide variant discovery, we built a phylogenetic tree using a subsample of these identified variants. Surprisingly, we found that the vast majority of genomic variation distinguished freshwater populations, with geographically diverse marine populations comprising a small monophyletic clade (Fig 5.4B). Principal component analysis of a genome-wide set of variants confirmed this result, with PC1 and PC2 separating geographically distinct freshwater populations and explaining nearly half the variance (29.07% and 14.8%, respectively). Additional PC axes explain lower amounts of overall variance but continue to separate freshwater populations from other freshwater and marine populations, rather than marine populations from other marine (Fig 5.5). This wide diversity among freshwater genomes could reflect adaptations to a diverse array of ecological environments, whereas marine environments are generally similar across geographic distances [25].

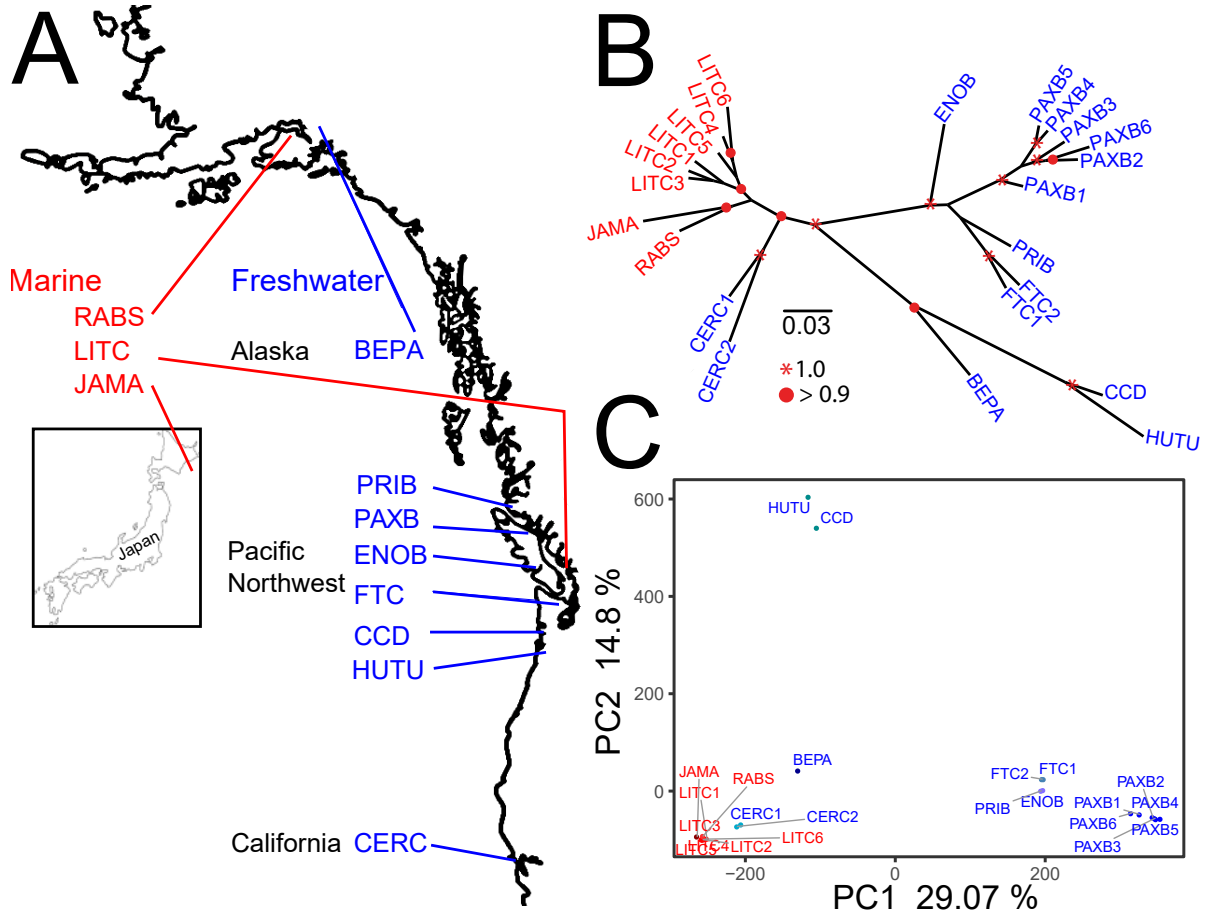

Figure 5.4: **Diverse genomic evolution during adaptation to independent freshwater environments.** (A) Map indicating the collection location for the given stickleback marine (red) and freshwater (blue) populations. (B) Markov chain Monte Carlo (MCMC) phylogenetic tree created using a set of genome-wide set of variants. Nodes with 1.0 and higher than 0.9 posterior probability are indicated by red asterisks and circles, respectively. Scale bar indicates 3% differences at variant positions (C) Principal component analysis of a genome-wide set of variants reveals that the two major axis of variation (PC1 and PC2) separate freshwater genomes from other freshwater genomes, not simply freshwater from marine.

Cross Name	Test Fish	*Bmp6* Marker	*Bmp6* Marker Position	Dad seq?	Mom seq?
BEPAxLITC	374	stn422	3386061	BEPA	LITC6
CCDxLITC	93	JCH115	3856115	CCD	-
CERCxLITC	155	16-6	2718243-3202277	CERC1	-
CERCxRABS	83	JCH137	5187435	CERC2	RABS
FTCxLITC	173	cm1396	3849882	FTC1	LITC4
HUTUxLITC	88	stn490	4598372	HUTU	-
LITCxENOS	172	16-6	2718243-3202277	LITC3	ENOB
LITCxFTC	107	21Sc2-6	2718243-3202277	FTC2	LITC5
LITCxPAXB	183	16-6	2718243-3202277	LITC3	PAXB6
LITCxPRIB	169	16-6	2718243-3202277	LITC3	PRIB
PAXBxJAMA-A	52	cm1288	3794581	PAXB3	-
PAXBxJAMA-L	95	cm1288	3794581	PAXB4	JAMA
PAXBxJAMA-WV	53	cm1288	3794581	PAXB5	-
PAXBxLITC-28	151	cm1430	4231433	PAXB1	LITC1
PAXBxLITC-29	75	cm1284	2565243	PAXB1	LITC2
PAXBxRABS	64	cm1430	4231433	PAXB2	-

Table 5.1: **Multiple crosses show an association between *Bmp6* and increased tooth number.** 'Cross Name' indicates the name of the cross, and 'Tested Fish' indicates the number of fish tested in the cross. '*Bmp6* marker' indicates the name of the marker proximal to *Bmp6* in the cross, with '*Bmp6* marker position' indicating the position in bp on chromosome 21 [68]. Marker sequences are given in Table 5.3. Markers 16-6 and 21Sc2-6 were determined by a genotyping by sequencing method [27], and represent 500kb bins on the chromosome. 'Dad Seq?' and 'Mom Seq?' indicate the name of the parent of the cross if the parent was sequenced.

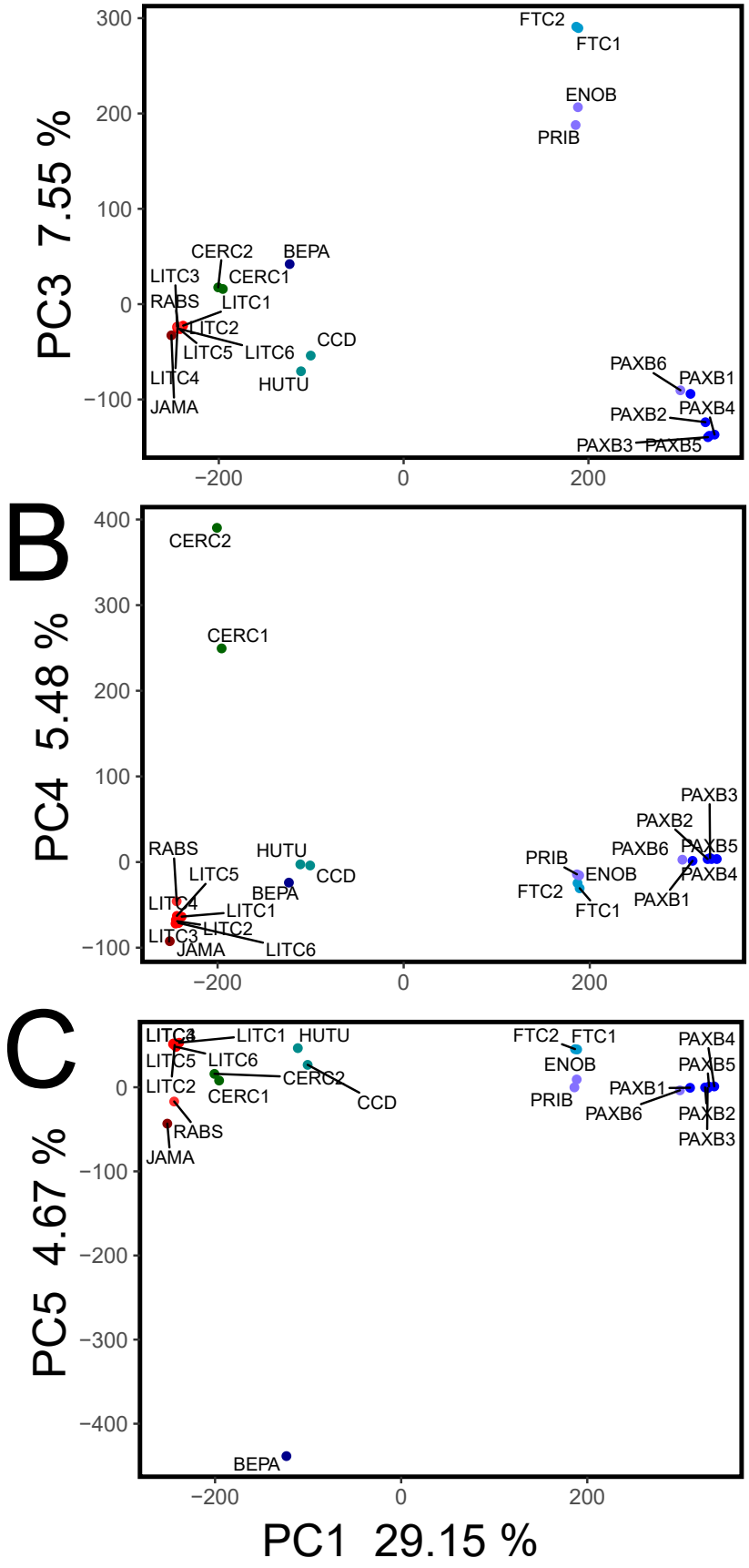

Figure 5.5: **Principal component axes separate freshwater genomes by population.** Principal components analysis of a genome-wide set of variants, with the 1st principal component plotted against the 3rd (A), 4th (B), and 5th (C). The first five principal components separate freshwater genomes by population but consistently cluster all marine genomes together.

Repeated reuse of a regulatory region of *Bmp6* during evolved tooth gain

We previously described a large effect quantitative trail locus (QTL) controlling an increase in tooth number in freshwater $PAXB_{FW}$ fish. This tooth gain QTL was associated with mutations within a tooth enhancer of *Bmp6* and a *cis*-regulatory change in *Bmp6* expression (Cleves et al, 2018, under review) [16]. We hypothesized that this regulatory haplotype was reused by multiple freshwater populations during the evolution of tooth gain, in a manner similar to the low-plated allele of the *Eda* gene [24]. We generated three and reanalyzed 13 [16,23,26-28] marine x freshwater F_2 crosses, totaling 2,246 animals in 16 F_2 crosses (Table 5.1). We found that markers near *Bmp6* were associated with an increased VTP tooth number in nine out of 16 crosses at a false discovery rate (FDR) of 5%, and 11 out of 16 at an FDR of 10%, suggesting a shared genetic basis for evolved tooth gain (Fig 5.6A).

We next asked whether the high-toothed freshwater alleles of *Bmp6* shared any of the 10 previously described mutations within a tooth enhancer of *Bmp6* associated with tooth gain in $PAXB_{FW}$ (Cleves et al, 2018, under review). None of our sequenced marine parents, nor freshwater parents lacking a high-toothed *Bmp6* allele, shared any of these high-tooth associated mutations (Fig 5.6B). In contrast, 10 out of the 11 freshwater parents with a high-toothed allele of *Bmp6* shared a substantial subset of these 10 enhancer mutations. A core set of six mutations, spanning 438 base pairs, was found in common among all 10 of these genomes, suggesting this haplotype as the new minimal set of high-toothed mutations. To further confirm the role of these mutations in the evolution of tooth gain, we generated three outcrosses between FTC_{FW} fish heterozygous for these high-toothed associated mutations and low-toothed marine fish. F_1 fish heterozygous for the high-toothed freshwater allele (HM) displayed more teeth on their VTPs than low-toothed freshwater allele heterozygotes (LM, Fig 5.6C, $P = 0.029$ one-tailed t-test).

We previously reported that the evolution of an increase in tooth number in $PAXB_{FW}$ was associated with a *cis*-regulatory decrease in *Bmp6* expression [16]. As multiple freshwater populations shared a tooth enhancer haplotype associated with an evolved gain in tooth number, we hypothesized that these populations would share a similar *cis*-regulatory decrease in *Bmp6* expression. We outcrossed high-toothed fish from multiple freshwater populations ($PAXB_{FW}$, FTC_{FW}, and $CERC_{FW}$) to marine fish to create F_1 hybrids, dissected VTPs and created cDNA libraries from these hybrids. We developed a Taqman based assay to measure *Bmp6* allele specific expression [29](Fig 5.7). *Bmp6* showed a *cis*-regulatory decrease in activity in the high-toothed $PAXB_{FW}$ and FTC_{FW} populations relative to marine ($P =$ 1.2e-4, 4.3e-15, respectively, nested linear model). $PAXB_{FW}$ and FTC_{FW}, both populations with the high-tooth associated haplotype of *Bmp6*, additionally displayed a *cis*-regulatory decrease in expression relative to $CERC_{FW}$ ($P =$ 9.4e-4 and 3.0e-2, respectively), which lacked the high-toothed *Bmp6* haplotype. This decreased expression further supports the hypothesis that one or more of these mutations decrease *Bmp6* expression in *cis* to reduce tooth number in diverse freshwater populations.

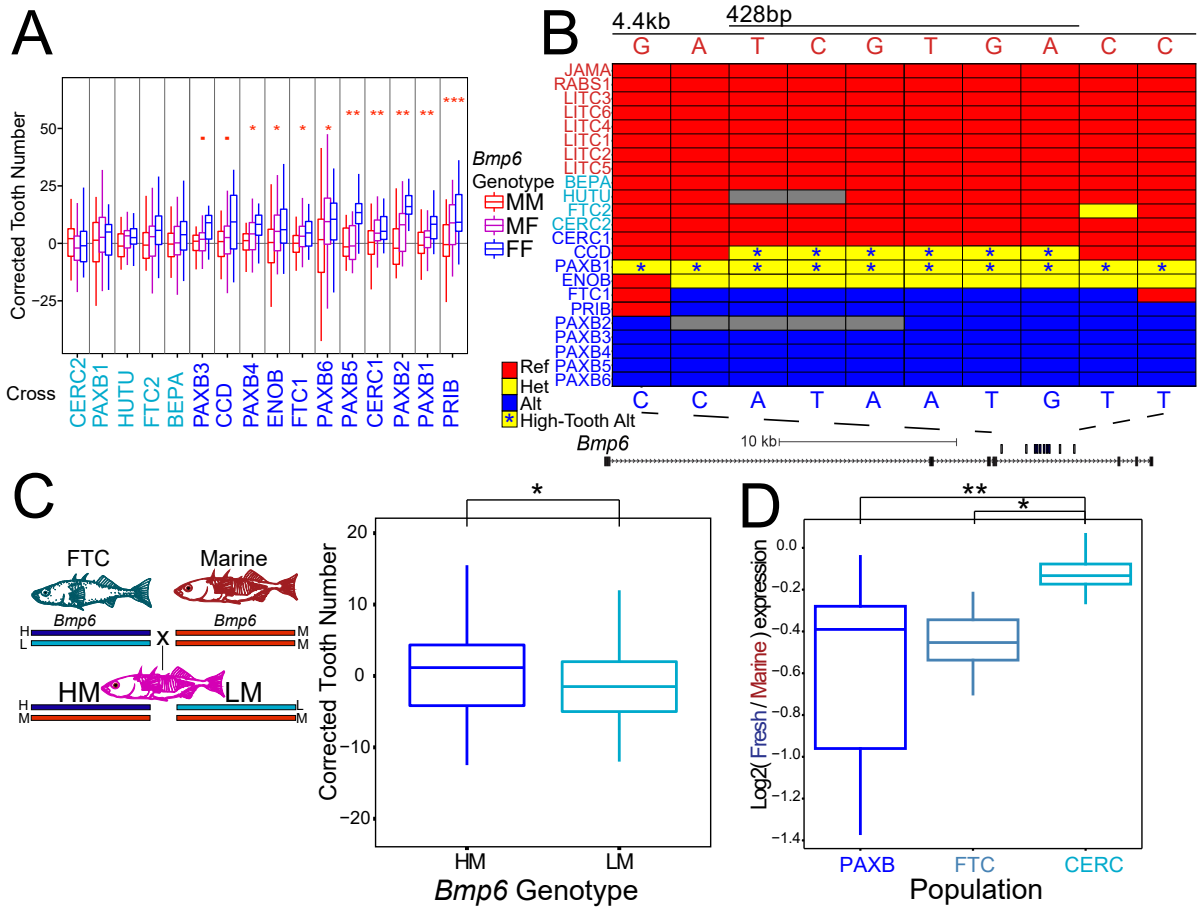

Figure 5.6: **A regulatory haplotype of *Bmp6* is associated with the evolution of tooth gain in multiple populations.**(A) Boxplots showing the effect of genetic markers near *Bmp6* on corrected tooth counts of F_2 animals of a marine x freshwater cross, with the freshwater parent indicated below (see Table 5.1 for details). · : FDR < 0.1 , * : FDR < 0.05, ** : FDR < 0.01, *** : FDR < 0.01. (B) The 10 high-toothed mutations associated with PAXB_{FW} evolved tooth gain (Cleves et al, 2018, under review) are found in other populations with a tooth gain QTL near *Bmp6*. (C) In the FTC_{FW} population, this haplotype is segregating, and genetic cross between heterozygous FTC_{FW} fish and marine fish shows association of the high-toothed *Bmp6* allele with increased tooth number, $P < 0.05$, one-tailed t-test. (D) Allele specific expression assays from *Bmp6* marine-freshwater F_1 hybrids show a *cis*-regulatory decrease of *Bmp6* in both the PAXB_{FW} and FTC_{FW} populations compared to CERC_{FW}. * : $P < 0.05$, ** : $P < 0.01$, nested linear model.

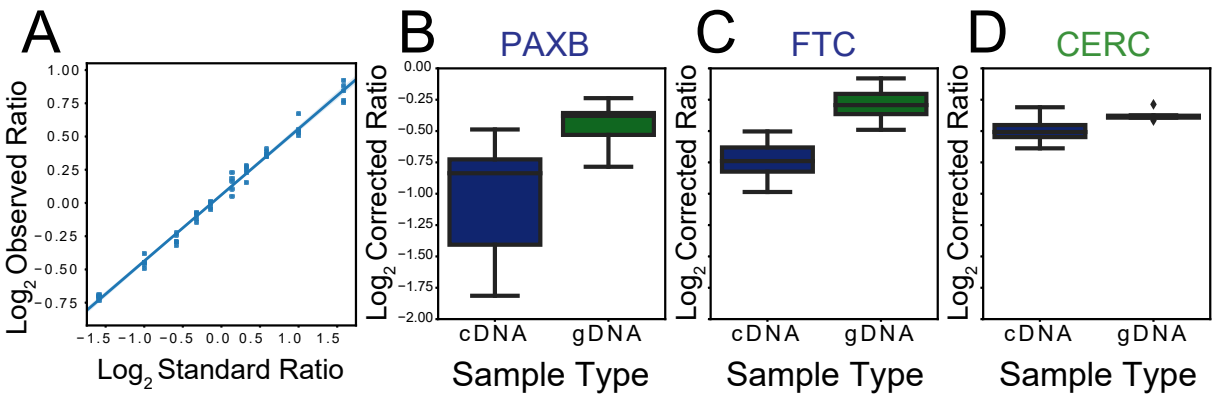

Figure 5.7: **Taqman based assay to measure allele specific expression.** (A) Taqman-based quantification in triplicate of a standard curve constructed from differing ratios of freshwater and marine genomic DNA. Regressing the observed ratios against the expected yields highly precise (Pearsons r = .995) but biased (slope = 0.5) estimates. This bias was corrected for in all future experiments. B-C) Boxplots showing corrected ratios from marine/freshwater F_1 hybrids, from VTP-derived cDNA and genomic DNA. $PAXB_{FW}$ (B) and FTC_{FW} (C) alleles of *Bmp6* showed large significant decreases in *cis*-regulatory activity (P = 1.2e-4, 4.3e-15, respectively, nested linear model). $CERC_{FW}$ alleles (D) showed a slight but significant decrease in cis-regulatory activity (P = 0.014), but this effect was smaller than $PAXB_{FW}$ or FTC_{FW} (P = 9.4e-4 and 3.0e-2, respectively).

As *Bmp6* is expressed in both oral and pharyngeal teeth [16,19], and is required for some aspects of pharyngeal tooth development (Cleves et al, 2018, under review), we hypothesized that *Bmp6* would be required for oral tooth development as well. We dissected oral jaws from a previously described $PAXB_{FW}$ F_2 cross containing a 13bp deletion predicted loss of function allele of *Bmp6*, and quantified oral tooth number. In line with our hypothesis, we found that *Bmp6* heterozygous mutants had fewer teeth on their premaxilla than their wild-type siblings (P = 0.038, 1 tailed t-test), while the dentary tooth domain appeared unaffected (Fig 5.8). Therefore, *Bmp6* has a required role during normal development of teeth on both the oral and pharyngeal jaws.

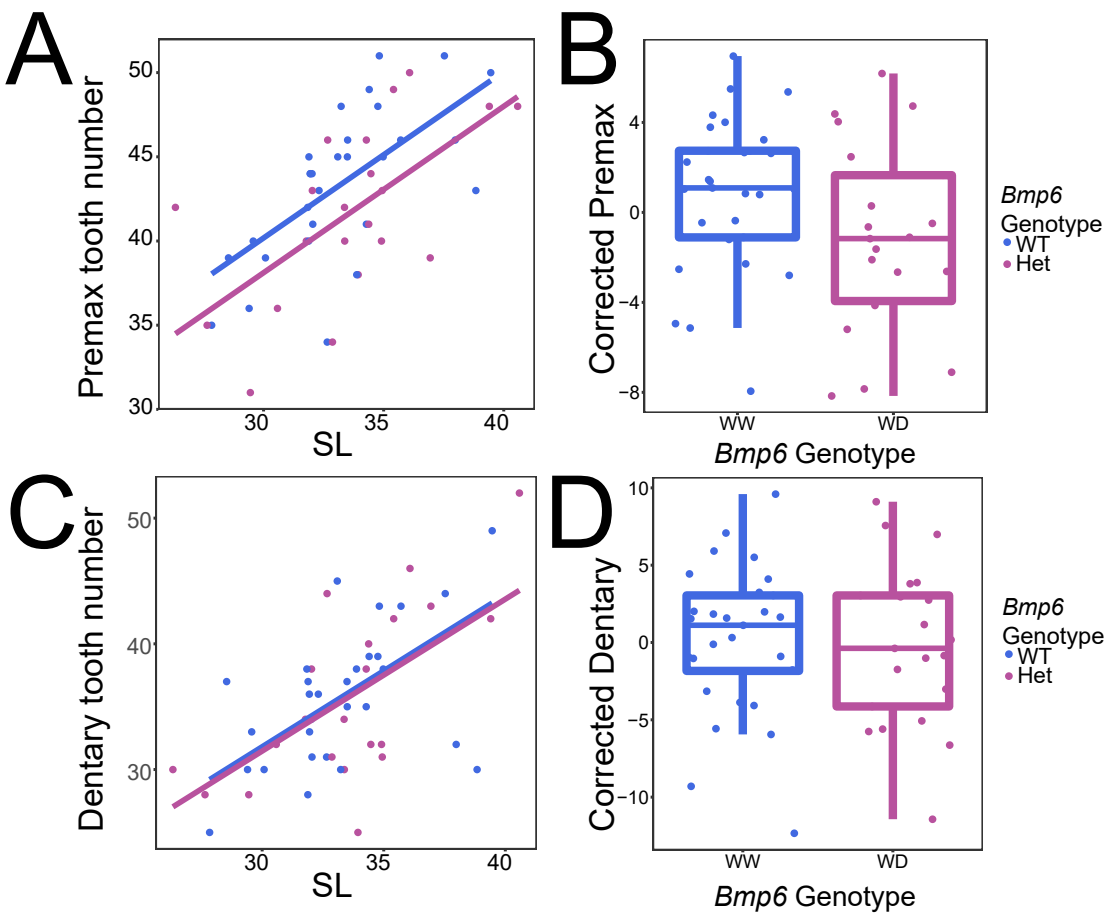

Figure 5.8: **Mutations in *Bmp6* disrupt oral tooth development.** A $PAXB_{FW}$ fish heterozygous for a 13bp deletion in exon 2 of *Bmp6* (Cleves et al, 2018, under review) was outcrossed, and offspring grown to the late juvenile stage. (A) *Bmp6* heterozygotes show a reduced tooth count on their premaxilla relative to their wild-type siblings ($P = 0.031$, 1-tailed t-test). (B) Boxplot showing size-corrected premaxilla tooth number of *Bmp6* wild-types and heterozygotes. (C) *Bmp6* heterozygotes show no detectable decrease in tooth count on their dentary. (D) Boxplot showing size corrected dentary tooth number of *Bmp6* wild-types and heterozygotes.

Modular evolution of tooth gain

Oral teeth and pharyngeal teeth show a modest evolutionary correlation, with $PAXB_{FW}$ and $CERC_{FW}$ populations showing an increased tooth number on components of their oral and pharyngeal jaws. We hypothesized that the genetic bases of these evolved gains in tooth number would be similar when comparing tooth fields within the oral or pharyngeal jaw, but less so when comparing oral to pharyngeal. To test this, we quantified tooth number on both the dentary and premaxilla of a previously published $CERC_{FW}$ x $LITC_{M}$ cross [23].

Similar to the evolved tooth gain correlations (Fig 5.3), sex and size corrected tooth number was more highly correlated when comparing within the oral or pharyngeal jaw (Pearsons r $>$ 0.64), but less correlated when comparing across jaws (Pearsons r $<$ 0.45, Fig 5.9A).

We next mapped QTL for both of these traits, and compared the genetic architecture controlling evolved changes in oral dentition to evolved changes in pharyngeal dentition (Fig 5.9B). We found a single QTL on chromosome 17 controlling an evolved freshwater gain in dentary tooth number (Fig 5.10, Table 5.2). Additionally, we found two QTL, on chromosomes 16 and 17, where the freshwater allele contributed to a gain in premaxilla tooth number, though this trait had no significant differences between these populations (Fig 5.1A). Genome-wide scans for epistasis revealed a significant interaction (LOD = 6.6) between the premaxilla QTL on chromosomes 16 and 17 (Fig 5.11). Additionally, we found an interaction between VTP QTL on chromosomes 4 and 21 (Fig 5.11), though not significant at the genome-wide level (LOD = 3.1). Of the five QTL controlling VTP tooth number, four did not overlap QTL controlling other tooth domains, and appeared to be specific for this domain (Fig 5.9B). In contrast, tooth number on premaxilla, dentary, and VTP was partially controlled by overlapping QTL on chromosome 17, and tooth number on premaxilla and DTP2 was partially controlled by overlapping QTL on chromosome 16. Thus, though we observe several cases of potentially shared genetic architecture underlying tooth gain in different tooth domains, evolved changes in tooth number appear largely modular.

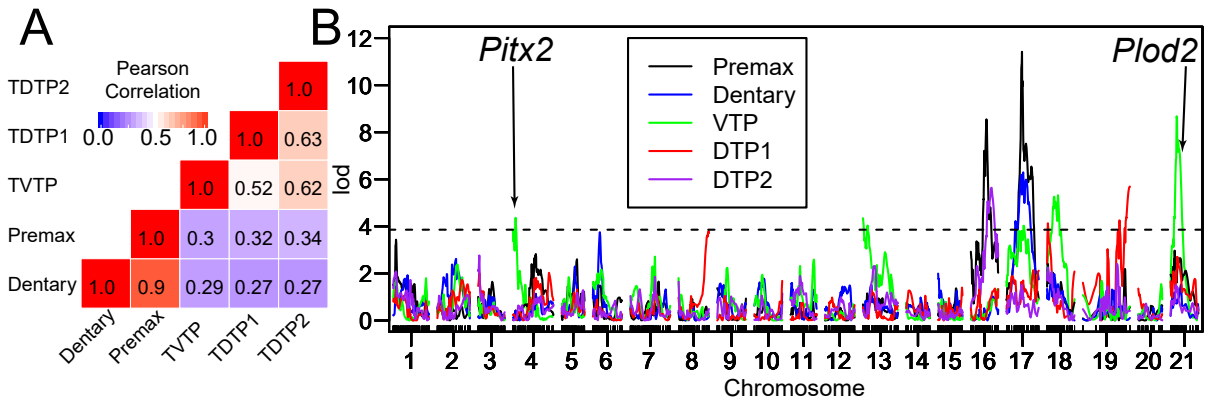

Figure 5.9: **Anatomically modular genetic basis of evolved tooth gain.** (A) Heatmap of correlation coefficients of corrected tooth number from anatomically distinct tooth domains (see Fig 5.1F) in $CERC_{FW}$ x $LITC_{M}$ F_2 fish, showing oral and pharyngeal teeth have a partially shared genetic basis. (B) Regionally distinct and shared genetic basis of evolved tooth gain in F_2 fish. LOD profiles for individual tooth domains (colored as in key) are shown. The genome-wide significance ($P = 0.05$) threshold for the VTP trait is indicated by the dashed line.

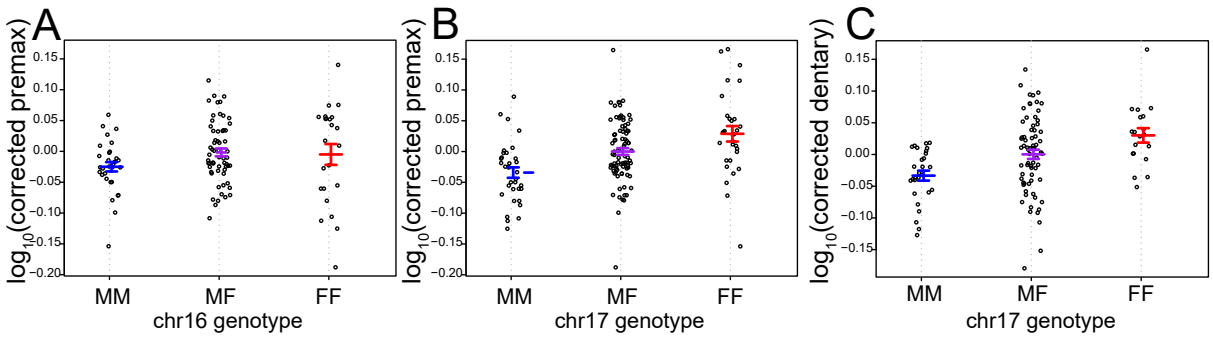

Figure 5.10: **Effect plots of QTL freshwater tooth gain on dentary or premaxilla.** (A-C) Effect plots showing log transformed corrected tooth counts for QTL controlling oral tooth number on premaxilla (A + B) and dentary (C). Bars show standard error and mean.

Trait	Chromosome	Peak Position (cM)	1.5 LOD interval	LOD score	PVE
DTP2 Tooth Number	16	62	47.1-72.0	5.6	14.1
Premaxilla Tooth Number	16	46.6	39.9-49.5	8.6	18.3
Premaxilla Tooth Number	17	46	42.0-48.7	11.4	25.4
Dentary Tooth Number	17	49.6	36.6-71.7	6.3	15.8
VTP Tooth Number	17	51	32.8-71.7	4.5	7.3

Table 5.2: **Overlapping oral and pharyngeal QTL controlling evolved tooth gain.** 'Trait' gives the trait that was mapped, 'chromosome' the chromosome the QTL was found on, with 'Peak Position' the position in cM, and a '1.5 LOD interval' (also in cM) for the QTL. 'LOD score' gives the LOD of the peak marker and 'PVE' gives the percentage of the variance explained by the QTL.

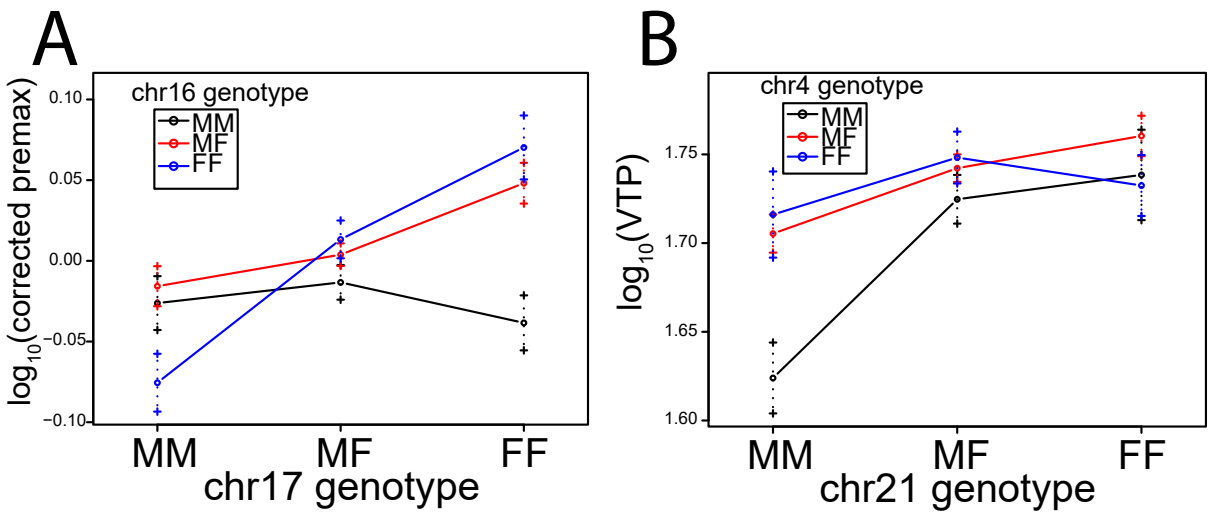

Figure 5.11: **Epistatic interaction between QTL controlling evolved tooth gain.**(A) Effect plots of a LOD 6.6 epistatic interaction between QTL controlling premaxilla tooth number. (B) Effect plots of LOD 3.1 epistatic interaction between QTL controlling VTP tooth number. Dots show standard error.

Pitx2 and *Plod2* are dentally expressed candidate genes underlying evolved tooth gain

We sought to identify the genes underlying the VTP-specific QTL on chromosomes 4 and 21 controlling the convergent evolution of tooth gain in $CERC_{FW}$. Each of the 1.5 LOD intervals of these QTL contains a large number of genes. As genes controlling tooth number are likely expressed in teeth, we first tried to identify all dentally expressed genes within the QTL-containing intervals. We leveraged new and published RNA-seq data sets (33 in total) using VTP-derived RNA from diverse stickleback populations and mutants. As the majority of gene expression changes between tooth plates from different populations are changes in trans, likely reflecting differences in cell-type composition within the VTPs (Hart et al, 2018, under review), we reasoned that much of the variance in this combined RNA-seq data set would be due to differences in cell-type composition between samples. Consistent with this prediction, genes known to mark specific cell types (*Bmp6*: teeth [16,19], *Calb2a*: taste buds [30,31], *Sncb*: neurons [3234], *Ttn*: muscle [35,36]) show highly correlated expression with other putative marker genes of the same tissue (Fig 5.12). To identify putative dentally expressed genes, we first performed a genome-wide correlation analysis to set a baseline correlation profile for each gene in the genome. We then correlated the expression profile of each gene to a set of genes, BiteCode, known to be expressed in developing teeth in other systems [37,38] (Hart et al, 2018, under review). We then performed a Mann-Whitney U test for each gene asking if the set of correlation coefficients from the BiteCode genes showed enrichment relative to the genome-wide set of correlation coefficients. As we observed an enrichment of

low p-values (Fig 5.13), we determined a p-value threshold of 1.58e-11, corresponding to the top 10% most enriched putative dental genes. Cross-validation with a hand-annotated set of genes known to be expressed in developing teleost teeth (TeToG) confirmed the ability of this method to find known tooth genes (Fig 5.13), including *Bmp6* and *Pitx2* (Fig 5.14A).

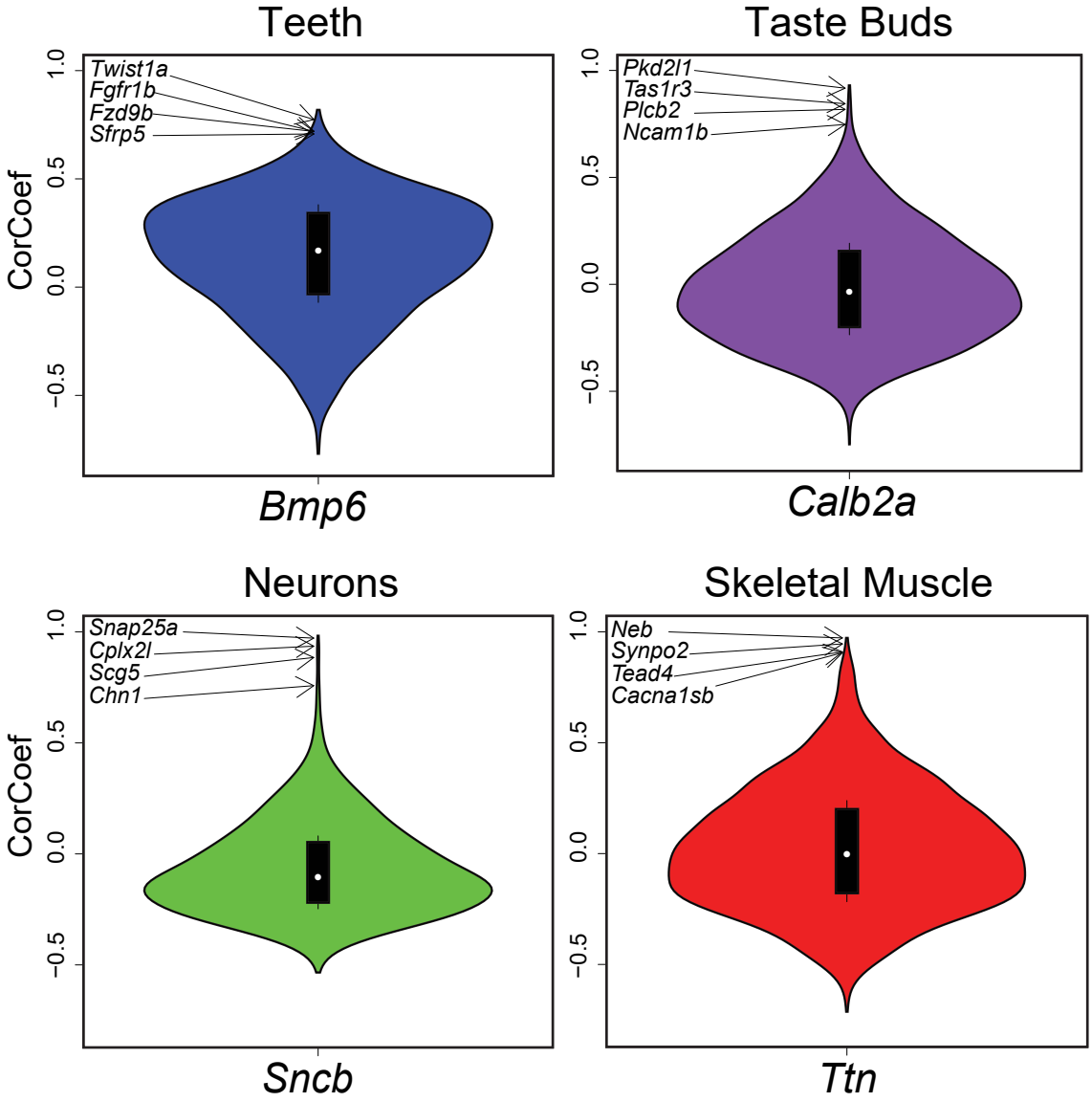

Figure 5.12: **Correlated expression of known tissue marker genes.** Genes reported to mark tissue types correlate highly with other known marker genes. (A-D) are violin plots of genome-wide correlation coefficients for all genes in the genome, compared with the given gene indicated below. A GO analysis [98] of the top 200 most correlated genes with *Calb2a* and *Ttn* showed 'sensory perception of taste (GO:0050909, $P = 2.1e\text{-}10$) and 'muscle structure development (GO:0061061, $P = 2.65e\text{-}9$) as the most enriched GO terms, respectively. A similar GO analysis of the top 200 genes most highly correlated with *Sncb* revealed an enrichment of the GO term 'nervous system process (GO:0050877, $P = 1.0e\text{-}3$). Genes within the top 200 most correlated genes genome-wide with Entrez Gene (https://www.ncbi.nlm.nih.gov/gene) descriptions including the given cell type as indicated by arrows.

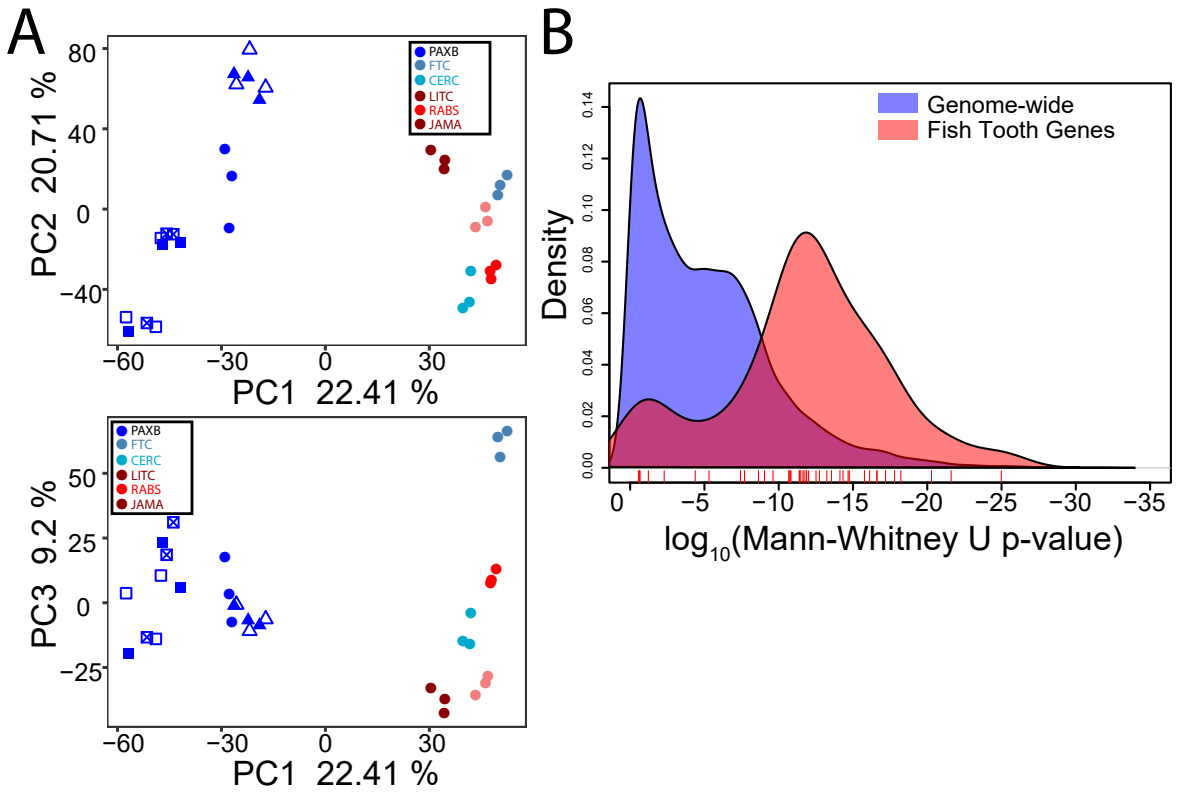

Figure 5.13: **Genes expressed in teleost teeth show highly correlated expression with rodent tooth genes.** (A-B) Principal component analysis of a gene expression matrix shows clustering of samples by population and genotype. Squares indicate a family of 13bp exon2 *Bmp6* mutants, and triangles a family of mutants in a 5' tooth enhancer [80] of *Bmp6*, with wild-types filled, heterozygotes crossed, and homozygous mutants unfilled. (C) A set of known teleost tooth genes (TeToG) display highly increased correlation with a set of mouse tooth genes (BiteCode). A rugplot shows $\log_{10}$ transformed p-value from a Mann-Whitney U test, comparing the genome-wide correlation coefficients to correlation coefficients of BiteCode genes for each gene in the TeToG set.

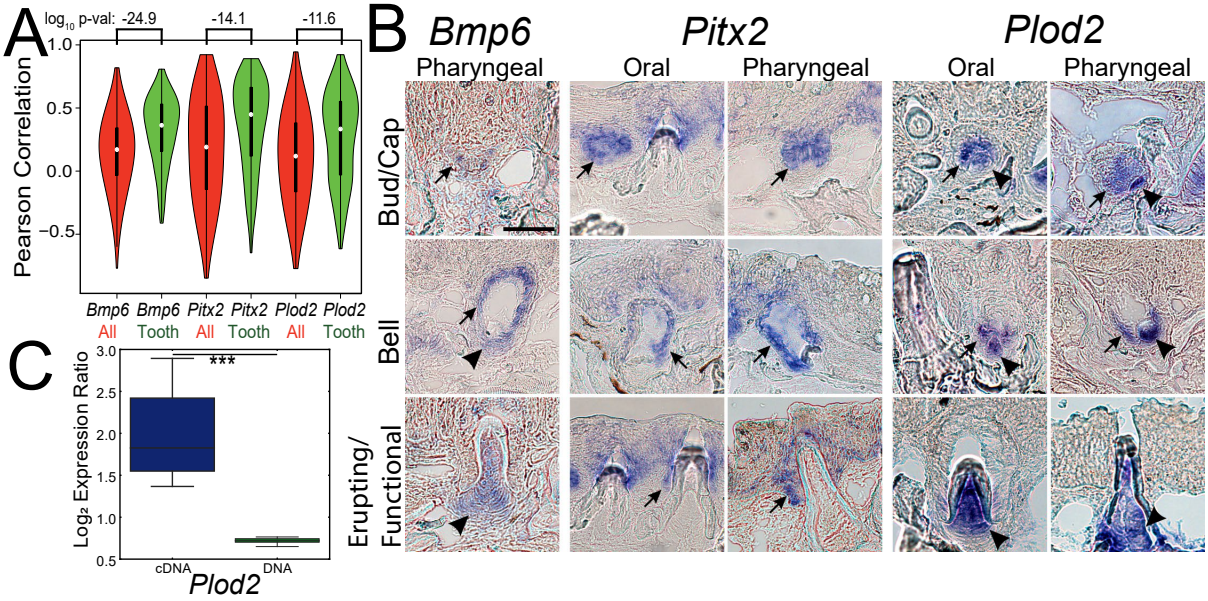

Figure 5.14: **Two new dentally-expressed candidate genes, *Pitx2* and *Plod2*, for underlying evolved increase in stickleback dentition.** (A) Correlation analysis using RNA-seq data derived from ventral pharyngeal tooth plates from diverse populations reveals that *Bmp6*, *Pitx2*, and *Plod2* display increased correlation with known dental genes, with $\log_{10}$ transformed p-values from a Mann-Whitney U test reported above. (B) in *Bmp6*, *Pitx2*, and *Plod2* show similar patterns of gene expression in developing teeth in both the oral and pharyngeal jaws as assayed by *in situ* hybridization (C) *Plod2* displays a *cis*-regulatory increase in expression in CERC_{FW} fish, with DNA controls shown on the right ($P < .001$, nested linear model).

The CERC_{FW} VTP chromosome 4 and chromosome 21 QTL 1.5 LOD intervals are fairly broad, encompassing 3.14Mb and 5.32Mb of sequence and 99 and 228 ENSEMBL predicted genes [39], respectively, of which nine and 19 are putative dentally expressed genes. As many of the genes underlying morphological QTL in sticklebacks have been found to have *cis*-regulatory changes [12-16], we leveraged our previous genome-wide scan for changes in *cis*-regulatory activity in CERC_{FW} VTP (Hart et al, 2018). We found five and 30 genes within the chromosome 4 and 21 interval, respectively, with a significant evolved *cis*-regulatory change, three and one of which were putative dental genes. As these regions were not previously found within a PAXB_{FW} QTL interval, we looked for genes which had a CERC_{FW} specific regulatory change, resulting in a total of one and two QTL candidates. The chromosome 4 QTL interval contained only one gene (*Rab28*) which passed all filters, and it seemed a poor candidate based on its described roles only in photoreceptors [40]. We instead suggest *Pitx2* as a candidate gene, due to its strong expression in teeth and its evolved expression increase in CERC_{FW} teeth relative to marine [fragments per kilobase per million reads (FPKM) 176 in CERC_{FW}, 116 in RABS_{M}, FDR = .038, Hart et al, 2018]. Unfortunately, the lack of

consistent marine-freshwater polymorphisms prevented testing for an evolved *cis*-regulatory change. The chromosome 21 QTL contains 2 genes which passed all our filters, *Limd2* and *Plod2*. We suggest *Plod2* is a stronger candidate underlying the evolution of tooth gain, as the only GO term in zfin (www.zfin.org) associated with *Limd2* is 'metal ion binding, while *Plod2* mutations disrupt skeletal development in animals from zebrafish to humans [41-44].

We found that *Bmp6*, *Pitx2*, and *Plod2* expression domains overlap in tooth epithelium during cap and early bell stages of tooth development (arrows in Fig 5.14B). Additionally, *Pitx2* and *Bmp6* are coexpressed in tooth epithelium from the onset of differentiation, though this expression domain is only occupied by *Pitx2* from mid-bell until eruption. *Plod2* and *Bmp6* are coexpressed in tooth mesenchyme at mid-cap stages and throughout tooth eruption. Importantly, these expression domains appeared identical between all oral and pharyngeal teeth we observed, with no apparent unique or missing domains occurring in any tooth field. This similar expression in all tooth fields suggests that modular changes to tooth development in different tooth fields is unlikely to be driven by coding mutations in these genes, which would presumably affect all teeth. Together with the QTL data, this dental expression supports *Bmp6*, *Pitx2*, and *Plod2* as plausible target loci for evolved changes in tooth development, specifically via quantitative and/or subtle temporal changes in the expression of these genes rather than large-scale gain or loss of expression domains.

$CERC_{FW}$ *Plod2* has a single predicted coding (E409Q) change relative to marine fish, but this change is found at a non-conserved residue with no predicted strong effect on protein function (BLOSSUM80 [45] score of +2) . Additionally, we observe this coding change in other low-toothed marine genomes. Therefore, we hypothesized the *cis*-regulatory change in *Plod2* was responsible for the evolved change in dentition and sought to replicate this result. Using our Taqman allele specific expression assay, we replicated our previous finding that the $CERC_{FW}$ allele of *Plod2* displays a significantly increased cis-regulatory activity relative to $RABS_M$ (P = 4.5e-7, nested linear model, Fig 5.14C). This *cis*-regulatory upregulation of *Plod2* could be specific to $CERC_{FW}$, as *Plod2* was not upregulated in $PAXB_{FW}$ (a non-significant 2% decrease in $PAXB_{FW}$ cis-regulatory activity relative to marine) in a previous study (Hart et al, 2018, under review).

Plod2 mutations disrupt normal oral and pharyngeal tooth development

Plod2 is a strong candidate gene underlying the evolution of tooth gain in $CERC_{FW}$, due to its dental expression (Fig. 5.14B) and evolved increase in *cis*-regulatory activity (Fig 5.14C). However, no functional data exists on the role of *Plod2* during tooth replacement. To functionally characterize the role of *Plod2* during tooth replacement in sticklebacks, we induced a series of predicted strong loss-of-function frame-shifting mutations in the 5^{th} exon of *Plod2* in both $CERC_{FW}$ and $RABS_M$ using the CRISPR/Cas9 system (Fig 5.15) [46-49]. We crossed stable $CERC_{FW}$ F_1s heterozygous for a 4bp deletion to $RABS_M$ F_1s heterozygous for either a 5bp deletion (cross 1) or a 1bp deletion (cross 2), grew up the resulting F_2s and

quantified tooth number).

Plod2 mutants display a highly significant gain in VTP tooth number, with homozygous mutants showing 4.2 more teeth than their heterozygous siblings (Fig 5.16A-C, $P =$ 1.91e-3). This tooth gain effect was found at both early (clutch 1, SL 15-20mm) and later juvenile (clutch 2, SL 25-35mm) stages (Fig 5.16A). In contrast, *Plod2* mutants additionally displayed a significant reduction of tooth number on their premaxilla, with mutants showing 3.3 fewer teeth than heterozygous siblings (Fig 5.16D-F, $P = 1.75$e-3). However, *Plod2* mutations did not appear to affect tooth number on the dentary, DTP1, or DTP2 (Fig 5.17). *Plod2* mutants did appear to have a modest growth defect (Fig 5.18, $P = 1.04$e-5), in line with previously published roles of *Plod2* during teleost development [50]. Thus, *Plod2* has required but opposing roles during tooth development in both the pharyngeal and oral jaw and is an excellent candidate gene underlying evolved tooth gain in CERC_{FW}.

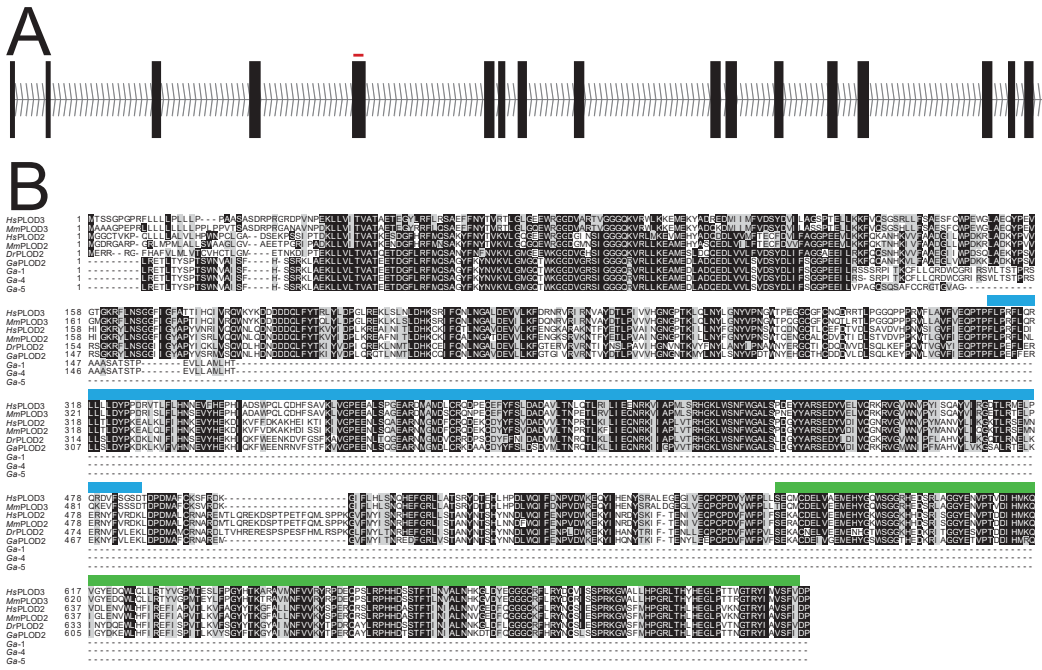

Figure 5.15: **Creation of *Plod2* mutant alleles using CRIPSR/Cas9.** (A) ENSEMBL predicted intron/exon structure of the stickleback *Plod2* gene, with the guide RNA targeting site shown in red. (B) Multiple alignment of human (Hs) and mouse (Mm) PLOD2 and PLOD3 with zebrafish PLOD2 (Dr) and stickleback (Ga) wild-type and predicted mutant PLOD2 alleles. Blue indicates a predicted Glycosyltransferase like family 2 domain, and green indicates a predicted Prolyl 4-hydroxylase domain [99]. Interestingly, human PLOD2 only has a predicted Prolyl 4-hydroxylase domain, while human PLOD3 has both Glycosyltransferase and Prolyl 4-hydroxylase domains, at E-value cutoff of 0.1 [99].

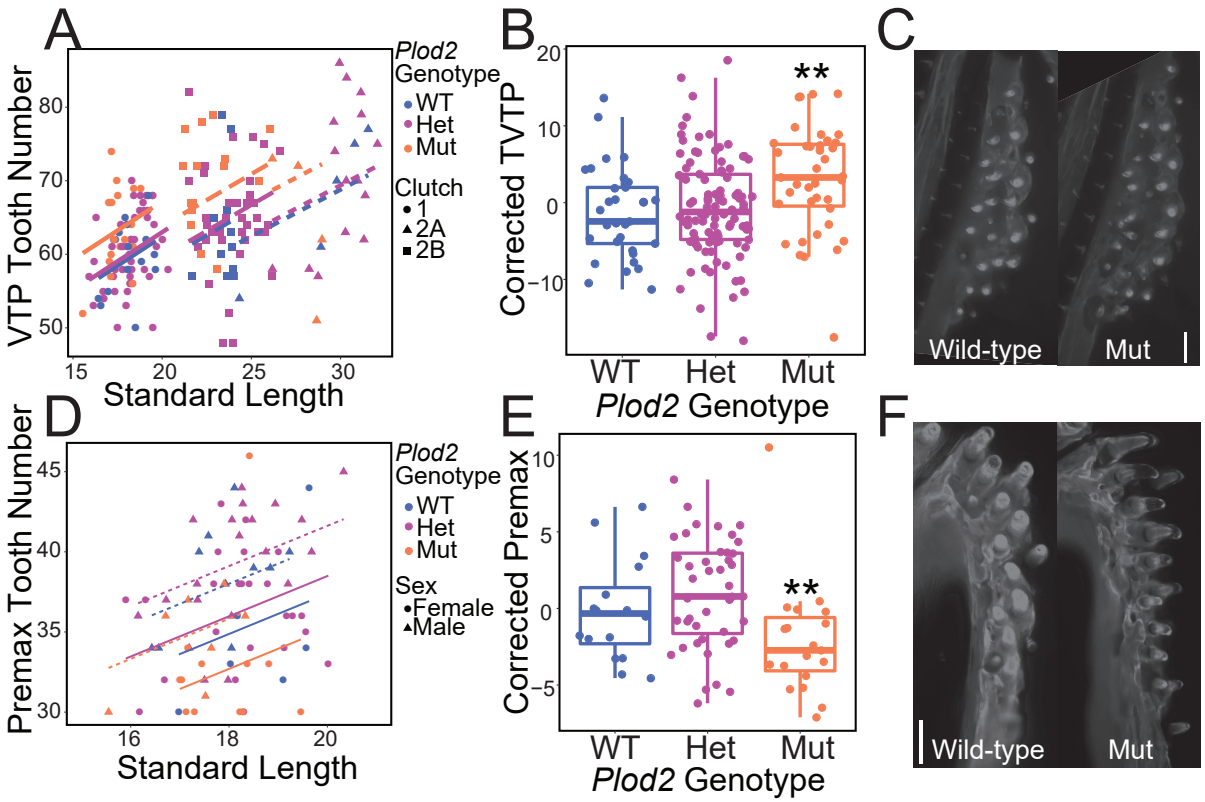

Figure 5.16: **Mutations in *Plod2* increase pharyngeal tooth number and reduce oral tooth development.**(A) *Plod2* mutations increase ventral pharyngeal tooth number at early and late developmental stages. (B) Boxplot showing ventral pharyngeal (VTP) tooth counts for each *Plod2* genotype, correcting for the length of the fish in each clutch separately. *Plod2* homozygous mutants have more teeth than heterozygotes (P = 1.91e-3). (C) Representative images of ventral tooth plates from *Plod2* wild-type and mutant fish from clutch 1. (D) *Plod2* mutants display a decreased premaxilla tooth count in clutch 1. (E) Boxplot showing premaxilla tooth counts, correcting for the effects of size and sex. *Plod2* homozygous mutants have fewer teeth than the heterozygotes (P = 1.75e-3). (F) Representative images of the premaxilla tooth field of wild-type (left) and *Plod2* homozygous mutant (right) fish. Scale bars = 10μM.

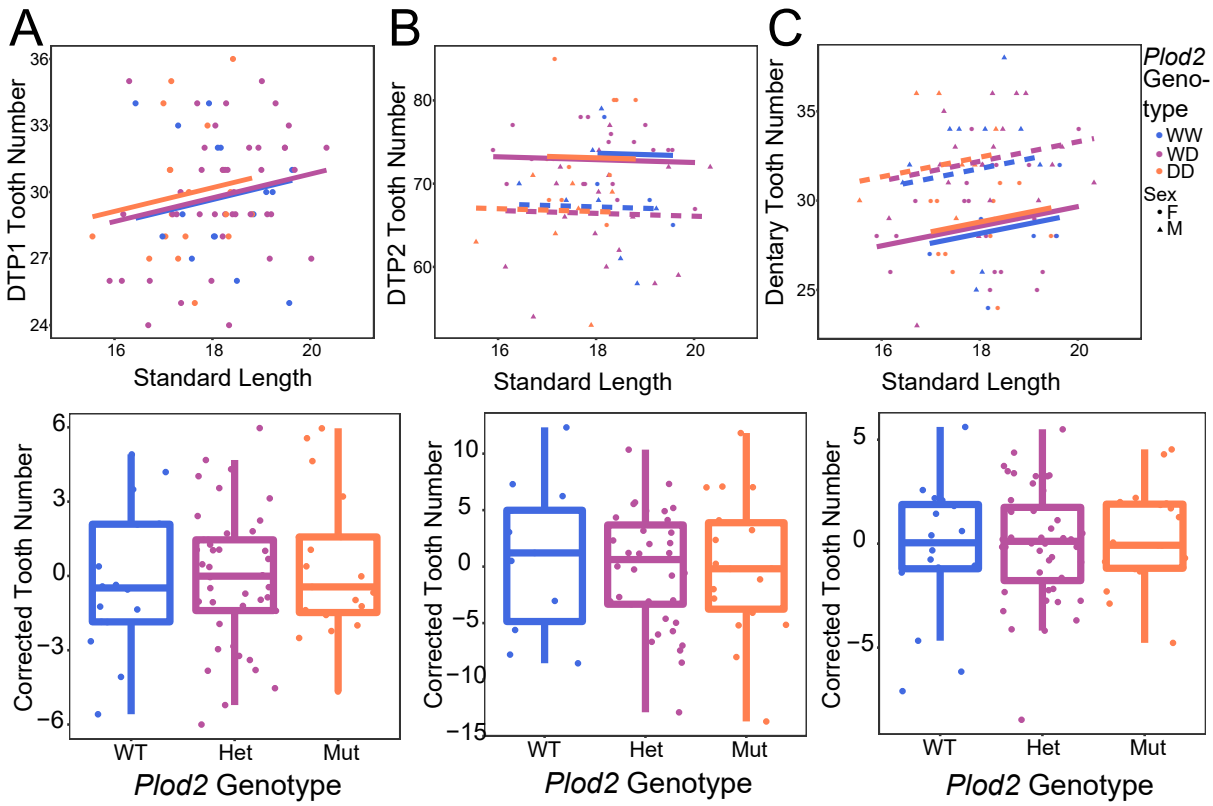

Figure 5.17: **Mutations in Plod2 do not affect dorsal pharyngeal or ventral oral tooth number.**(A-C) *Plod2* mutations in clutch 1 have no detectable effect on tooth number in DTP1 (A), DTP2 (B), or the dentary (C). Boxplots below show tooth number after correcting for size (DTP1), or size and sex (DTP2, dentary).

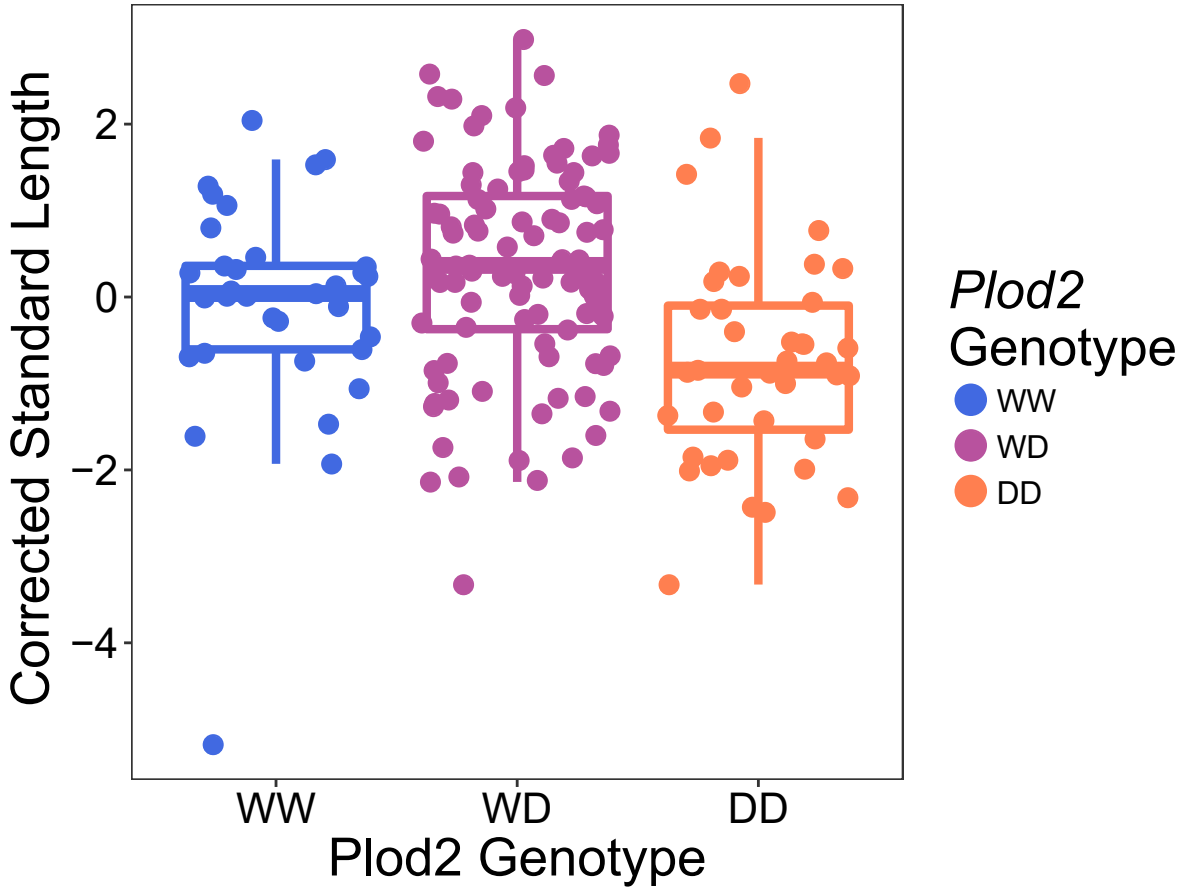

Figure 5.18: ***Plod2* mutants show a recessive growth defect.** *Plod2* mutants display a recessive growth defect in clutch 1 + 2 relative to their heterozygous siblings (P = 1.04e-5). Boxplot shows clutch-corrected standard length.

Pitx2 mutations are lethal and inhibit tooth development in all tooth domains

Mutations in *Pitx2* are associated with oral tooth development defects in mice and humans [5153]. In zebrafish, *Pitx2* mutations lead to defects in the formation of teeth on the pharyngeal jaw [54,55]. However, zebrafish lack teeth on their oral jaw, thus the role of *Pitx2* during oral tooth development had not been tested [56]. To elucidate the role of *Pitx2* during stickleback tooth development, we generated a series of predicted loss-of-function alleles of *Pitx2* using the CRISPR/Cas9 system (Fig 5.19). We targeted the homeodomain with CRISPR/Cas9 [4649] to ensure all potential DNA binding isoforms of *Pitx2* had disrupted function. We incrossed *Pitx2* heterozygous F_1 animals carrying either a one base-pair deletion or a four base-pair deletion, then quantified tooth number in these F_2 crosses.

Pitx2 mutants displayed a severe recessive reduction in tooth number on VTP, with homozygous mutants displaying less than half the teeth of their wild-type siblings (P = 1.62e-7, Fig 5.20A-C). *Pitx2* mutants additionally displayed a recessive loss of tooth number on their dentary (P = 1.28e-12, Fig 5.20D), and a complete loss of teeth on their premaxilla (P = 6.80e-11, Fig 5.20E,F). Both deletion alleles of *Pitx2* were homozygous lethal, with no homozygous mutants detected past 19 dpf (Table 5.3). *Pitx2* heterozygotes did not show any tooth number or growth defect, even at later stages (Fig 5.21).

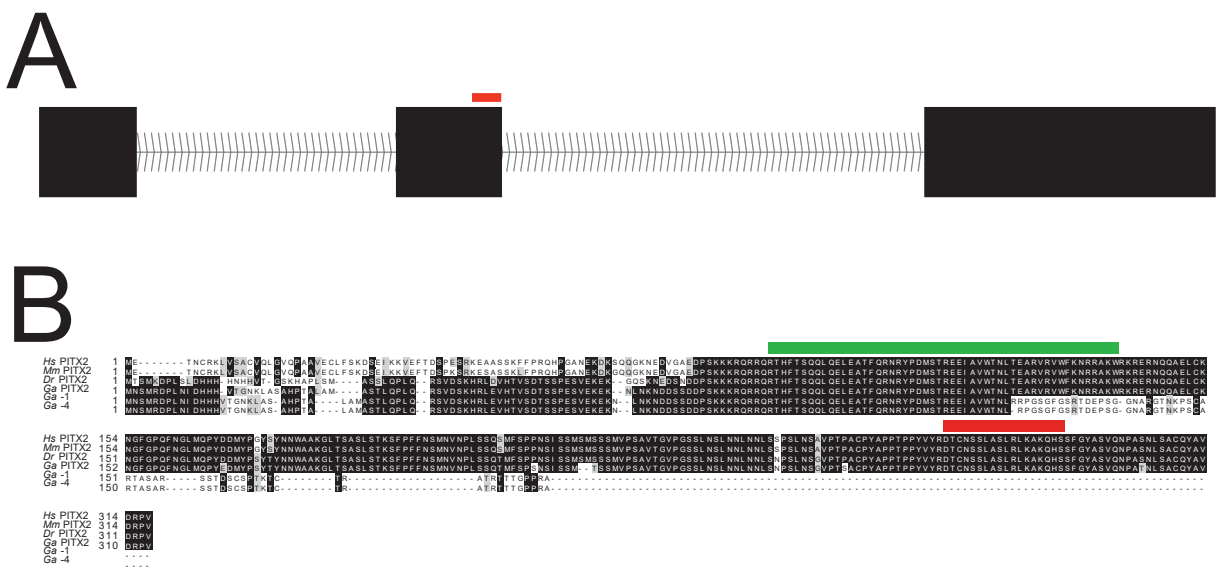

Figure 5.19: **Creation of *Pitx2* mutant alleles using CRIPSR/Cas9.** ENSEMBL predicted *Pitx2* transcript with intron/exon structure. gRNA target is shown in red. B) Multiple alignment with human (Hs), mouse (Mm), zebrafish (Dr), and stickleback (Ga) PITX2 sequences, with the predicted sequence of the PITX2 mutant alleles shown. Green indicates a predicted homeodomain, and red indicates a predicted OAR domain [99].

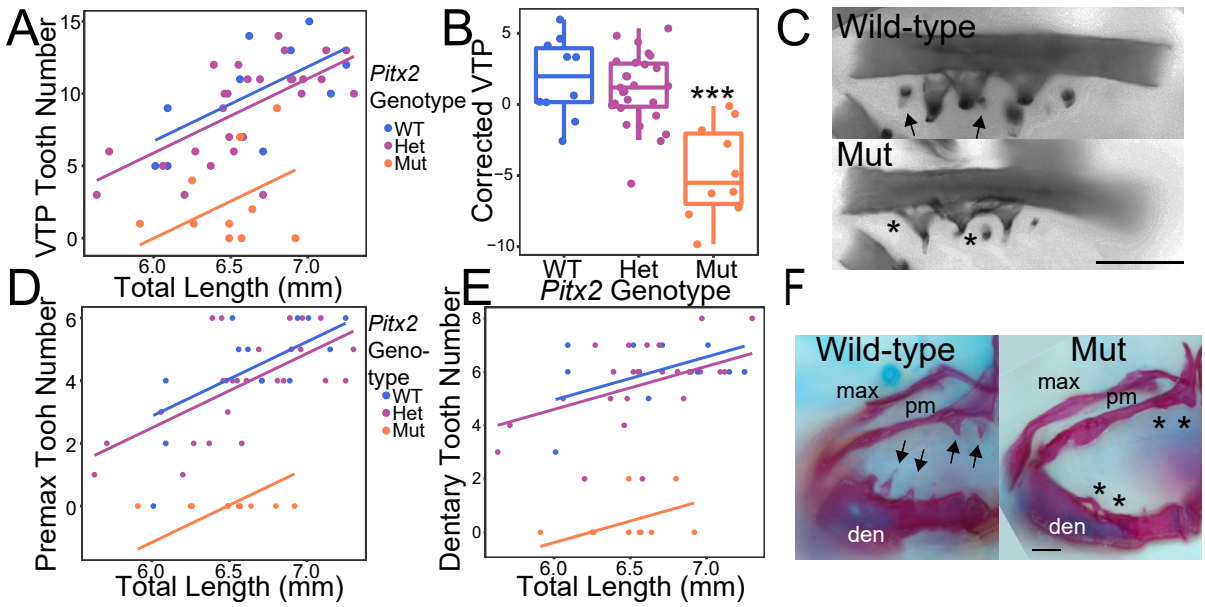

Figure 5.20: **Mutations in *Pitx2* disrupt oral and pharyngeal tooth development.**(A) *Pitx2* homozygous mutants display a marked defect in ventral pharyngeal (VTP) tooth development. Genotypes are color-coded as in key: wild-type (WT) in blue, *Pitx2* heterozygous mutant (Het) in purple, and *Pitx2* homozygous mutant (Mut) in orange (B) Boxplot showing tooth number of wild-type (WT), *Pitx2* heterozygous mutant (Het), and *Pitx2* homozygous mutant (Mut) siblings corrected for the effect of fish total length. *Pitx2* mutants have reduced tooth number relative to heterozygotes (P = 9.03e-8). (C) Representative images of unilateral 19dpf wild-type (top) and *Pitx2* homozygous mutant (bot) ventral pharyngeal tooth plates. Arrows show teeth present in wild-type, and asterisks show corresponding missing tooth positions in *Pitx2* mutants. (D-E) *Pitx2* mutants display a severe reduction in premaxilla (premax, D) and dentary (E) tooth number in the oral jaw. Genotypes are color coded as in (A-B). (F) Oral jaws of 19dpf wild-type (left) and *Pitx2* homozygous mutant (right) oral jaws, showing teeth in wild-type (arrows) that are missing in the mutant (asterisks). max = maxilla, pm = premaxilla, den = dentary. Scale bars = 10μM.

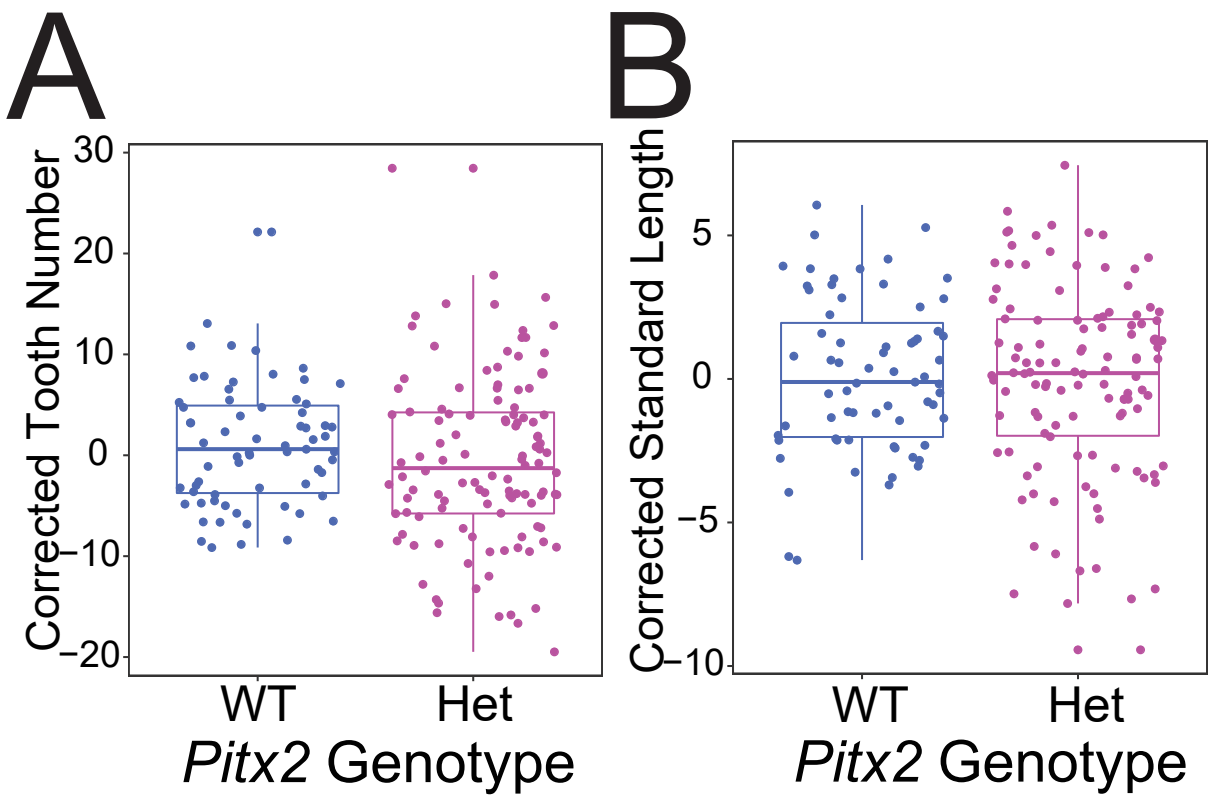

Figure 5.21: ***Pitx2* heterozygous mutations have no detectable effect on growth or tooth number.**(A) Boxplot showing size and clutch corrected tooth counts from *Pitx2* heterozyogus incrosses and outcrosses. *Pitx2* heterozygous mutations have no detectable effect on VTP tooth number. (B) Boxplot showing clutch corrected standard length from the same *Pitx2* crosses. There was no detectable effect of heterozygous *Pitx2* mutations on growth, though *Pitx2* homozygous mutants were never observed after 19dpf.

Dpf	Wild-type	Het	Mut	Exp(Mut)
1	4	3	1	2
3	2	0	1	0.75
5	2	2	0	1
6	1	3	0	1
7	3	3	1	1.75
8	1	2	0	0.75
10	1	5	2	2
11	1	3	1	1.25
12	0	5	2	1.75
13	1	3	2	1.5
14	2	1	2	1.25
15	4	3	0	1.75
16	3	10	7	5
17	5	6	5	4
18	0	6	0	1.5
19	7	14	2	5.75

Table 5.3: **Timecourse of viability of *Pitx2* mutants.** Two *Pitx2* het incrosses were raised, and embryos sacrificed and genotyped. 'Dpf' indicates the day post fertilization the embryos were sacrificed on, with the observed counts of each genotype given in the 'Wild-type', 'Het' and 'Mut' columns. 'Exp(mut)' gives the number of expected homozygous mutants.

5.4 Discussion

We sought to identify the genetic bases of an evolved gain in tooth number in multiple independently colonized freshwater stickleback populations, as well as to test the modular nature of the evolution of tooth gain on both the oral and pharyngeal jaw. We found a set of six core mutations within an intronic enhancer of *Bmp6* were associated with pharyngeal tooth gain in five geographically distinct freshwater populations. These mutations were associated with a *cis*-regulatory decrease in expression of the critical dental regulator *Bmp6* (Cleves et al, 2018, under review) in multiple populations. We identify two new candidate genes, *Plod2* and *Pitx2*, to underlie convergently evolved tooth gain. Both genes have evolved regulatory changes in a high-toothed freshwater population, and both we found both genes play required roles during oral and pharyngeal tooth development. QTL mapping revealed the largely modular nature of the evolution of increased tooth number. In contrast, genome editing revealed that mutations in candidate genes contributing to evolved increases in pharyngeal tooth number affected tooth development in both oral and pharyngeal domains. Thus, modular changes to the regulation of key dental developmental genes appear

to underlie evolved adaptive changes in morphology.

Modular morphological evolution

Tooth development is a highly conserved process, with similar developmental gene expression patterns observed in phylogenetically diverse species [57,58]. The paired-like homeodomain transcription factor *Pitx2* is an early marker of the dental epithelium in many species [19,59-64], and mutations in *Pitx2* result in dental defects in diverse systems [51,54,55,65,66]. Tooth development is also well conserved across different dental fields within the same organism - gene expression patterns are largely shared across oral and pharyngeal tooth development [17,19]. However, evolution has been able to act on specific tooth domains - zebrafish have retained only their ventral pharyngeal dentition [56,67,68], and sticklebacks have evolved strong sexual dimorphism in their oral but not pharyngeal jaws [21,22].

We found that multiple freshwater stickleback populations have evolved changes in tooth number on both their oral and pharyngeal jaw (mostly increases), though the pharyngeal tooth number changes were more pronounced. However, within an F_2 cross we found strong correlations between tooth domains within the oral or pharyngeal jaw, but weaker correlations when comparing tooth domains on different jaws. In line with evolved tooth number changes being modular, of the five previously described QTL controlling pharyngeal tooth gain in $CERC_{FW}$, only one overlapped with a QTL controlling oral tooth gain. However, the evolved tooth number changes are quantitative and relatively subtle compared to more severe tooth number changes seen in mutants with coding changes, so its possible that larger cross sizes would reveal more pharyngeal tooth QTL also regulate oral tooth number. Mutations in all candidate genes we tested (*Bmp6*, *Plod2*, and *Pitx2*) affected both oral and pharyngeal tooth development. Strikingly, mutations in *Plod2* resulted in increased tooth number in the pharyngeal jaw but reduced tooth number on the oral jaw, providing evidence for differing regulatory environments between the two jaws. Additionally, we observe stronger effects for tooth number on the premaxilla than the dentary for all three mutants analyzed here, indicating distinct regulatory environments even within the oral jaw. Given the overall more severe phenotypes in mutants with coding changes than the natural QTL alleles, we hypothesize that changes to the *cis* regulation of key developmental regulators underlies the modular evolution of tooth number change in sticklebacks.

Evolutionary reuse of a *cis*-regulatory haplotype of *Bmp6*

We have previously described a large effect QTL on chromosome 21, associated with a *cis*-regulatory decrease in *Bmp6* expression, controlling an evolved increase of pharyngeal tooth number in $PAXB_{FW}$ [16]. Repeated QTL mapping and genome sequencing revealed a shared set of 10 high-tooth associated mutations within a *Bmp6* intronic haplotype that contains a robust tooth enhancer of *Bmp6* (Cleves et al, 2018, under review). Here we showed that a core set of six of these variants spanning 438 base pairs is associated with tooth gain in four additional freshwater stickleback populations, though not in the high-toothed $CERC_{FW}$

population. The *cis*-regulatory decrease in *Bmp6* expression tracks with these mutations in FTC_{FW} fish, but not in high-toothed $CERC_{FW}$ fish lacking these variants. This correlation further supports the role of these mutations, and their associated *cis*-regulatory allele of *Bmp6*, in the evolution of tooth gain. Additionally, this supports the convergent nature of evolved tooth gain in $CERC_{FW}$, as the partially overlapping chromosome 21 QTL appear to have distinct underlying genetic bases.

These high-toothed mutations are not found in the genomes of nearby marine fish, nor in any marine genomes we have sequenced, though they appear to be present at high frequency in many freshwater populations. This phylogenetic pattern is reminiscent of the low-plated allele of the *Eda* gene, which is present at high frequencies in diverse freshwater populations but at very low (0.2%) allele frequency in marine fish [24]. Though we were unable to detect these mutations in our eight sampled marine genomes, we lacked sampling power to detect low-frequency variants within these populations. Therefore, it seems probable these high-toothed variants were present at low frequency in the founders of new freshwater environments or arrived through migration. In these new freshwater environments, these alleles seem to be adaptive, as they are associated with an increased tooth number, and gradually increased in allele frequency to the levels observed today. Future population genetic tests for natural selection could test whether this intronic haplotype displays molecular signals of selection such as reduced heterozygosity as was found for the *Pitx1* pelvic reduction allele [13].

Genome-wide scans to identify candidate genes controlling dental evolution

QTL mapping is a powerful method to identify the regions of the genome controlling evolved changes in morphology. However, the QTL-containing intervals are often broad (> 1Mb), and contain potentially hundreds of genes [22]. Narrowing these hundreds of genes down to a single causative locus requires either laborious recombination mapping [15](Cleves et al, 2018, under review), and/or a candidate gene approach, selecting genes with known roles in development. While this approach has proven successful in the identification of the genetic bases of a number of traits [11-13,69], it is highly biased against the discovery of genes with undescribed functions.

Here, we used an unbiased approach to identify dentally expressed candidate genes with evolved changes in cis-regulatory activity. We first leveraged our existing genome-wide phased RNA-seq data from marine-freshwater F_1 hybrids (Hart et al, 2018, under review), allowing us to identify genes with evolved changes in *cis*-regulation. We then used a novel approach to identify genes expressed in teeth, looking for increased correlations in 33 RNA-seq datasets among known dentally expressed genes compared to a genome-wide baseline. This allowed us to identify *Bmp6* and *Pitx2*, two known dentally expressed genes, but also *Plod2*, a gene with no previously described role in tooth development. We believe that further unbiased scans using similar gene expression datasets can reveal more genes with novel

roles or confirm that genes with known roles are indeed more likely to be substrates for morphological evolution. However, further fine-mapping of these QTL intervals is required to confirm the role of these candidate genes during evolved changes in tooth number.

Derived directed deletions differentially disrupt distinct dental developmental domains

Plod2 mutations in humans result in Bruck syndrome [41,42,70,71], and *Plod2* mutations result in skeletal defects in zebrafish [50]. Human *PLOD2* mutants have no reported dental phenotype [72,73]. However, human *PLOD3* mutants are reported to have a hair phenotype [74], and mouse *Plod3* mutants fail to make proper basement membranes [75]. It seems possible that stickleback *Plod2* is more multifunctional than its direct human ortholog, with stickleback *Plod2* mutations affecting teeth, an integumental organ with deep homology to hair (Cleves et al, 2018, under review)[76,77].

Plod2 mutants display an increased pharyngeal tooth number than their wild-type siblings, suggesting that *Plod2* mutants may have an accelerated tooth regeneration rate. To our knowledge, this represents the first induced mutation in a polyphyodont (continuous tooth replacement) system that accelerates rather than inhibits tooth development. It remains to be seen if this is truly due to an increased replacement rate, rather than a slower tooth shedding rate or increased size of the tooth field. *Plod2* mutants also show a reduction of tooth number on their premaxilla, in their oral jaw. This gain of teeth in one dental domain, with an accompanying loss in another, suggests that different tooth development and replacement dynamics are active in each domain.

Surprisingly, a *cis*-regulatory gain in *Plod2* expression is associated with an increase in pharyngeal tooth number, while *Plod2* loss-of-function mutants drive increases in tooth number. We suggest that, as *Plod2* has pleiotropic affects in other systems [41,50,70], that *Plod2* has activating and inhibiting roles during tooth development. The *cis*-regulatory change could affect a subset of those domains, while a coding mutation affects all, resulting in the observed similar phenotype.

Conserved role of *Pitx2* during tooth development

Pitx2 has been previously shown to be critical for the development of oral teeth in mice [66] and humans [53,65], as well as pharyngeal teeth in zebrafish [54,55]. To our knowledge, this study represents the first test of the role of *Pitx2* during tooth development in the oral and pharyngeal jaws simultaneously. Consistent with other loss-of-function studies, predicted loss-of-function mutations in *Pitx2* had strong effects on tooth number in both jaws, with a near ablation of teeth in the oral jaw. Unlike the zebrafish *Pitx2* loss-of-function mutations [55], stickleback alleles were homozygous lethal, and homozygotes were never observed past 19dpf. These phenotypes confirm the known role of *Pitx2* as a critical regulator of dental development and suggest *Pitx2* likely plays additional critical roles during stickleback development.

It has previously been reported that *Pitx2* directly binds to and regulates *Plod2* in mice [78,79]. We found that *Pitx2* and *Plod2* were partially overlapping in their gene expression patterns as assayed by *in situ* hybridization, suggesting they are present in similar cell types at similar times. Possibly supporting this regulatory interaction between *Plod2* and *Pitx2* in sticklebacks is the epistatic interaction between QTL on chromosome 4 and 21. If this direct interaction is conserved from sticklebacks to mice, it could imply that certain developmental pathways may prove to be hotspots of evolutionary change. Future work further identifying the genetic bases of evolved changes in multiple independently derived populations will allow tests for concerted changes in similar pathways, potentially uncovering critical pathways regulating tooth replacement.

5.5 Methods

Animal husbandry

Fish of all populations and genotypes were raised in 110L aquaria with eight hours of light per day, at 18C in brackish water (3.5g/L Instant Ocean salt, 0.217mL/L 10% sodium bicarbonate). Fry were fed live Artemia brine shrimp until reaching a standard length (SL) of 10 mm, when frozen Daphnia was added to the diet. Upon reaching 20 mm SL, fish were fed a combination of frozen Mysis shrimp and bloodworms. Experiments were approved by the Institutional Animal Care and Use Committees of the University of California-Berkeley (protocol AUP-2015-01-7117).

Marine-freshwater F_2 crosses were generated by crossing a marine and freshwater grandparent to create F_1 hybrids. F_1s were then blindly incrossed, generating F_2 families (Table 5.1). *Bmp6* mutant crosses were previously described (Cleves et al, 2018. under review). *Bmp6* 5‘ tooth enhancer mutants were generated as described [80] in a $PAXB_{FW}$ background, outcrossed to make stable F_1s, and incrossed to make homozygous mutant and wild-type sibling F_2s. Three different FTC_{FW} high-tooth haplotype heterozygotes were outcrossed to three different marine animals. FTC_{FW} cross 1 and cross 2 were FTC_{FW} males crossed to $RABS_M$ females, and cross 3 was a $LITC_M$ male crossed to a FTC_{FW} female. *Plod2* crosses were generated by crossing stable F_1 $CERC_{FW}$ four bp deletion to a stable F_1 $RABS_M$ five bp deletion heterozygote (cross 1) or to a one bp deletion heterozygote (cross 2). *Pitx2* cross 1 and cross 4 were outcrosses between a one base pair deletion and a wild-type sibling. *Pitx2* cross 2 was an incross between one and four base pair deletion heterozygotes, and *Pitx2* cross 3 and 5 were incrosses between one base pair deletion heterozygotes. All *Pitx2* alleles were in a $CERC_{FW}$ background.

Skeletal staining and imaging

Stickleback corpses were fixed overnight in 10% neutral buffered formalin, washed and stained with 0.008% alizarin red in 1% KOH for 24 hours. Fish were cleared in 2% KOH for 1 week,

and transferred to 0.25 % KOH 50% glycerol for storage. Branchial skeletons were dissected and mounted following [81]. Pharyngeal tooth number was quantified under fluorescent microscopy using a TX2 filter on a Leica DM2500 microscope. Tooth plate images were created using the z-stage module of the Leica Application Suite (LAS). Some tooth count data in Fig 5.1 and some of the data QTL data and genome sequences in Fig 5.4 have been previously published.

Genomic DNA purification and resequencing

Genomic DNA was purified from caudal fins following (Hart et al, 2018, under review). Briefly, caudal fin tissue was digested in tail digestion buffer (10mM Tris pH 8.0, 100mM NaCl, 10mM EDTA, 0.05% SDS, 2.5μL ProK (Ambion AM2546)) overnight at 55C. Following a phenol:chloroform purification and resuspension in water, 50ng of genomic DNA was used as input to the Nextera Library prep kit (Illumina FC-121-1031). Resulting barcoded libraries were sequenced on an Illumina HiSeq 2000 or 4000 at the Vincent J. Coates Genomics Sequencing Laboratory.

Genomic and phylogenetic analysis

Resequencing reads were mapped to the stickleback reference genome [39] using bowtie2 [82]. Resulting SAM files were turned into BAM files, sorted and indexed using Samtools [83], and PCR duplicates removed using Picard Tools (version 1.51) (http://broadinstitute.github.io/picard/). The Genome Analysis Toolkit (GATK)s BaseRecalibrator, AnalyzeCovariates, and PrintReads were used on the resulting BAM files to recalibrate base quality scores [8486]. Variants were called by first creating gVCF files using GATKs HaplotypeCaller (–emitRefConfidence GVCF) and combined using GenotypeGVCFs. The resulting VCF file was used to create a Nexus file using a set of custom python scripts. MrBayes [87] was used to create genome-wide phylogenies, using a General Time Reversible nucleotide substitution model with a gamma distributed rate variation. We compared two separate Markov chain Monte Carlo (MCMC) runs over 200,000 generations, with the average standard deviation of split frequencies equal to .007, below the generally accepted cutoff of .01. The VCF was further filtered to remove rows with any sites either missing or with a GQ quality score less than five, and used for principal components analysis using the FactoMineR R package [88].

Quantitative genetic analysis of tooth patterning

As tooth number can be correlated with standard length, growth conditions, and genetic background, we built and tested several models to correct for these confounding variables. We tested a combination of variables in linear models when tooth number was modeled as a function of genotype, as well as standard length, the tank the fish were raised in, an interaction between tank and length, and, when available, the sex of the fish. All possible

combinations were tested, and the model with the lowest AIC was used for further analysis. If AIC differences between potential best models were less than 2, we defaulted to the one that included an interaction between length and tank as well as an additive term for genotype, as this was usually the best model in other cases. Tooth number was corrected for the effects of confounding variables by taking the residual error values of the best fit model but excluding the genotype term from the model. Statistical effect of the genotype in the repeated QTL crosses was determined by performing an ANOVA in all crosses and correcting the resulting p-value for multiple hypothesis testing using the method of Holm-Bonferroni [89].

Genome-Wide QTL Mapping

Genome-wide genetic markers from a set of F_2s from a $CERC_{FW}$ x $LITC_M$ cross [23] were combined with previously published pharyngeal tooth counts (VTP, DTP1, DTP2), as well as new oral tooth data (premaxilla, dentary). Phenotypes were log transformed if they were not normally distributed (as determined by a Shapiro test p-value ¡ 0.05) and corrected using a linear model for the effects of sex and standard length if such terms were significant in the model. QTL mapping of all phenotypes was preformed using R/qtl [90], following a similar approach as [23]. We calculated marginal genotype probabilities using calc.genoprob (step = 2, err.prob=.001), and set phenotype-specific cutoffs using scan2s permutation testing, calculating an alpha of .05 using CalcErrorProb [90]. Multiple QTL scans were performed using stepwise [90], and LOD profiles were visualized using PlotLodProfile.

PCR and Cloning

PCR primers (Table 5.4) were designed using Primer3 [91] and synthesized by IDT. PCR was performed using Phusion (NEB M0530L) following the manufacturers instructions. To digest the PCR products, restriction enzyme was added directly to the PCR buffer (for *EcoN*I, NEB 0521L), or added to a mix of PCR product and 0.5X of the recommended NEB buffer (NEB3.1 for *BspE*I, NEB R0540L). Resulting genotyping products were run on a 2% Lithium Acetate Boric acid (LAB) agarose gel [92] and visualized using SYBR safe.

For cloning, the PCR product was purified using the QIAquick PCR purification kit (Qiagen 28104) and digested with *Xho*I and *Xba*I (NEB R0146L and R0145L, respectively) in cutsmart buffer. Following an additional PCR purification, digested products were ligated in pBluescript II SK+, transformed into competent cells, and grown on ampicillin plates. Plasmid was cut with *Xba*I and transcribed using T7 polymerase (Promega P2075) with a DIG RNA labeling mix (Roche 11175025910) mix to create probes for *in situ* hybridization.

Primer Pair Name	Forward Primer Sequence	Reverse Primer Sequence	Purpose
CERC 21 marker (JCH137)	CCGGTTTTTAATC CCATTTG	GTCGGTTTCTTG GAGCATTC	chr21 marker

Bmp6 RFLP marker (JCH115)	GCCGGCAGCCAAGCGTGAGTTACTGTGTGC	GGAGCAGCCAAATGTAGGAA	*Bmp6 Ava*II RFLP
*Plod2 BspE*I	CCGGAGGGGAAATAGATGAT	GACCGAGGGGTACTTGTCAG	*Plod2* genotyping/*BspE*I
*Pitx2 EcoN*I	CGGGACTCTTAACCCAATCA	GGCCCAAATTACCCACATTT	*Pitx2* genotyping/*EcoN*I
gRNA Template Amplification	GCGTAATACGACTCACTATAG	AAAGCACCGACTCGGTGCCAC	gRNA template amplification
Guide-Constant	AAAGCACCGACTCGGTGCCACTTTTTCAAGTTGATAACGGACTAGCCTTATTTTAACTTGCTATTTCTAGCTCTAAAAC	NA	gRNA constant
Pitx2 guide template	GCGTAATACGACTCACTATAGGTGGACCAACCTCACGGGTTTTAGAGCTAGAAATAGC	NA	*Pitx2* gRNA
Plod2 guide template	GCGTAATACGACTCACTATAGGCCTGCTGGAACTTCCGGGTTTTAGAGCTAGAAATAGC	NA	*Plod2* gRNA
Plod2 UTR	CGGCtctagaCAGTCTTTGGACTCCGCTCT	CGGCctcgagCAATTTCCATTAAAGAAGGCAGA	*Plod2* probe cloning
Plod2 cDNA	TGGGACCAGAAGAAAACCTG	CGATTTGCCTCATGTGAATG	*Plod2* cDNA cloning
Sex Marker	CATATTGCTGCTTGTGTGGAAG	GATCCTCCTCGTTCCTACAG	Sex Marker
cm1396	CTGTACCATGCTTGTCTCTCC	CGTACTCCACTGTGGCATGG	chr21 marker
cm1284	CACGGCAAACAGGTGAGAC	TCGATGGGCTGTAAATCCTC	chr21 marker
cm1288	AATTACACTGCCTGCACTTGG	GTCAGATGGACGGACAGACG	chr21 marker
cm1430	AGTGACGAATCCCTCTTCTGC	CACACCTTGTTGTGTTTGTAGC	chr21 marker
stn422	CTGCCTCATATGGCATGAAG	CCCAGTTGTTGAGTTGGTTG	chr21 marker

stn490	ATGAGGTCACCCT GCCTAAC	CGCCTGTCATATA CACATTGC	chr21 marker

Table 5.4: **Primer sequences.** 'Primer pair name' gives the name of the primer pair, with the forward sequence given in 'Forward Primer Sequence' and reverse in 'Reverse Primer Sequence', and the purpose of the primers given in 'Purpose'

RNA isolation and sequencing

Late juvenile stage (SL 40mm) females were euthanized in 0.04% tricaine. Ventral pharyngeal tooth plates were dissected on ice, placed into TRIzol (Invitrogen 15596026), and ground with a plastic pestle. RNA was purified following the manufacturers instructions. RNA quantity, quality, and integrity were verified on a Bioanalyzer, and samples with RIN ¿8.0 used to create sequencing libraries. 500ng of total RNA from each sample was used as input to the Illumina stranded TruSeq polyA RNA kit. Following an additional quality check on the bioanalyzer, libraries were sequenced on either the Illumina HiSeq2000 or HiSeq4000.

Allele specific expression assay

RNA was isolated from ventral pharyngeal tooth plates and DNA from caudal fins of each fish as described above. Purified RNA was subjected to DNaseI digestion for 15 min at room temperature, with the reaction halted addition of EDTA to a concentration of 2.5mM and heat inactivated at 65C for 10 min. cDNA libraries were created using the Superscipt III kit (Invitrogen 18080093) and reverse transcribing with random hexamers. Taqman probe and primers were designed using ThermoFishers custom Custom TaqMan Assay Design Tool (https://www.thermofisher.com/order/custom-genomic-products/tools/genotyping/), creating a custom *Bmp6* and *Plod2* primer and probe set (ID: ANAACGU and ANEPVXN, respectively). Reactions were run using Taqman genotyping master mix (ThermoFisher 4371355) kit on a Bio-Rad CFX96 qPCR machine. First, a standard curve was generated by mixing differing amounts of equivalently concentrated genomic DNA from marine and freshwater fish. For *Plod2*, as our assay spanned an exon-intron junction, we created cDNA clones for each allele, and used plasmid DNA instead of genomic DNA. We found a tight-fitting curve (Pearsons r = 0.995 for both genes) and calculated a correction factor (the slope, 0.5 for *Bmp6*, 0.54 for *Plod2*) to go from observed differences to true differences. cDNA Taqman reactions were run in triplicate, with one genomic DNA replicate per sample on the same plate for *Bmp6*. For *Plod2*, two replicates each of four independently created 1:1 mixtures of plasmid clones of the two alleles were used as controls. Statistical analysis was performed in a custom jupyter notebook, creating a linear model nesting each replicate within its sample, and each sample within a given sample type (gDNA or cDNA).

Gene expression correlation analysis

The predicted stickleback transcriptome was created using bedtools getfasta [93], with all isoforms collapsed into a single reference transcript. Transcript abundance was calculated using Kallisto [94]. As we noticed biases in quantification based on read length or paired vs single end, we trimmed all reads to the shortest length, 50bp, and kept only the first read of each mate-pair. High ($>$ 2000 transcripts per million, tmp) and low abundance ($<$ 3 tpm) transcripts were excluded from further analysis, and data was normalized to the mean expression among replicates using a custom jupyter notebook. Correlation analysis of the resulting gene expression matrix was performed in R.

In situ hybridizations on sections

Euthanized 20 mm $CERC_{FW}$ juveniles were decapitated and fixed overnight in 4% formaldehyde in Phosphate-buffered saline (PBS) at 4C with heavy agitation, washed 3x 20 min with phosphate-buffered saline with tween (PBST) on a nutator, then decalcified for 5-7 days in 20% Ethylenediaminetetraacetic acid (EDTA) at room temperature on a nutator. Once decalcified, specimens were again washed 3x 20 min in PBST, then stepped into 100% EtOH via 15-60 min washes in 30, 50, 70, and 95% EtOH in H2O. Samples were sometimes stored at this stage for up to 3 weeks. Samples were then washed for 1 hr+ in 50/50 EtOH/Hemo DE at room temperature, then 1 hr+ in 100% Hemo De at room temperature, then 1 hr+ in 50/50 Hemo De/paraffin (Paraplast x-tra, Fisher) at 65C, then rinsed and washed overnight at 65 C in 100% paraffin. Stickleback heads were embedded in plastic molds with 100% paraffin (heated to 65C), mounted, then sectioned sagittally with a Microm HM 340 E microtome. Sections were captured on Superfrost Plus slides, sometimes stored for up to 3 weeks prior to ISH. To prepare slides for ISH, slides were de-parafinnized (5 min incubation at 65C, let cool, submerge for 5 then 10 min in 100% Hemo De, 5 min 100% etoh, 10 min 80% EtOH/H2O, 10 min 100% H2O). Slides were sometimes stored at this stage for up to 3 weeks. From this point, all steps are performed in LockMailer microscope slide jars (Sigma-Aldrich) in a volume of 9-11 mLs. To begin the in situ process, slides were washed for 5 min in PBST, 10 min in proteinase K solution (15 µg/mL), rinsed quickly with PBST, then re-fixed for 30 min at room temperature in 4% formaldehyde in PBS. Fixative was then washed out with one rinse and 2x 20 min PBST washes before pre-hybridization. Slides were washed for 5 min at room temperature followed by a long incubation in hybridization buffer (no probe, pre-hyb step) for 1-4 hours at 67C in a rotating hybridization oven (hybridization buffer is 50% formamide, 5x SSC, 0.1% Tween, 5 mg/mL CHAPS, 1 mg/mL yeast RNA, 0.1 mg/mL heparin, pH 6.0 with citric acid). Riboprobes for *Bmp6*, *Pitx2*, and *Plod2* were generated as previously described [16]. Riboprobes were added at a concentration of 100 ng/mL in 10 mL of hybridization buffer and agitated overnight in a rotating hybridization oven at 67C. The following day, six pre-heated hybridization washes at 67C in a rotating hybridization oven were performed for 20-90 min each, totaling 5-6 hours of total hyb wash time (generally, shorter washes in the beginning, longer washes towards the end; hyb wash is

the same recipe as hyb buffer, but excluding CHAPS, RNA, and heparin). Slides were then rinsed and washed in pre-heated maleic acid buffer with Tween (MABT) at 67C for 20 min, then washed in pre-heated MABT for 20 minutes at room temperature (to allow for slow cooling). Slides were then removed from the slide jars, placed in a humidor and blocked with 50-100 µL of 2% Boheringer blocking reagent (BBR), covered with parafilm, for one to three hours at room temperature. Following the block step, block was poured off each slide, and anti-Digoxygenin Alkaline Phosphatase conjugated antibody was added at a concentration of 1:2000 in 2% BBR (50-100 µL) and incubated at 4C overnight. The following day, one MABT rinse and 5x 20-50 min MABT washes over the course of 3-4 hours, agitated at room temperature were done to wash out residual antibody. To begin the coloration process, slides were changed into NTMT (0.1M Tris pH 9.5, 0.05M MgCl2, 0.1M NaCl, 0.1% Tween) via 3x 5-10 min washes before removing the final NTMT wash and replacing it with 10 mL of coloration solution (NTMT with 25 µg/mL Nitro blue tetrazolium chloride [NBT] and 175 µg/mL 5-bromo 4-chloro 3-indolyl phosphate [BCIP]). Signal development was carried out for 12-30 hours to visualize mRNA localization. Once adequately developed, slides were rinsed then washed for 10 min in PBST, fixed in 4 formaldehyde in PBS for 1-5 days at 4C, then washed into deionized water for imaging. Slides were coverslipped with H20 and imaged on a Leica DM2500 microscope. Hematoxylin and eosin staining was performed following [95].

CRISPR/Cas9 Genome Editing

Genome editing was performed as described in [48,49]. Briefly, linearized pCS2-nCas9n (Addgene 7929) was used as input for the mMessage (Ambion AM1340) SP6 transcription kit, with mRNA integrity verified on a Tris-EDTA (TAE) agarose gel. gRNA targets were identified using ZiFiT [96], and DNA oligos designed to create T7 gRNA templates were designed following [97] and ordered from IDT. gRNAs were transcribed using the MAXIscript T7 kit (Ambion AM1312), with integrity validated on a TAE gel and purified. A mixture of 150 ng/l cas9 mRNA, 50 ng/l gRNA, 0.025% phenol red in 0.2M KCl was injected into stickleback embryos at the one cell stage. F_0 injected mosaics were used to create stable F_1 mutants, which were then incrossed to produce F_2s. Mutants were identified using a restriction protection assay following PCR, with mutations disrupting a restriction site and protecting the fragment from digestion (Table 5.4).

Data Availability

All sequencing reads are available on the Sequence Read Archive (SRP142636). All scripts used for analysis are available on GitHub (https://github.com/trahsemaj/Modular_Tooth).

Acknowledgements

We thank Michael Eisen for scientific advice and feedback on multiple aspects of this project. We thank Phillip Cleves for help with initial screens for tooth associated variants in multiple populations, and Priscilla Erickson for work on the *Bmp6* mutant oral tooth analysis. We also thank Nikunj Donde and Mary Chen for their work on the *Pitx2* mutants.

Funding statement

This work was supported by NIH R01 DE021475 (to C.T.M.), NIH genomics training grant 5T32HG000047-15 (to J.C.H.), and NSF Graduate Research Fellowship (to N.A.E. and A.M.G.). This work used the Vincent J. Coates Genomics Sequencing Laboratory at UC Berkeley, supported by NIH S10 Instrumentation Grants S10RR029668 and S10RR027303. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

5.6 References

1. Davidson EH, Britten RJ. Note on the control of gene expression during development. Journal of Theoretical Biology. 1971;32: 123130. doi:10.1016/0022-5193(71)90140-8
2. Davidson EH, Erwin DH. Gene regulatory networks and the evolution of animal body plans. Science. 2006;311: 796800. doi:10.1126/science.1113832
3. Stern DL. Evolutionary developmental biology and the problem of variation. Evolution. 2000;54: 10791091.
4. Carroll SB. Evolution at Two Levels: On Genes and Form. PLoS Biol. 2005;3: e245. doi:10.1371/journal.pbio.0030245
5. Carroll SB. Evo-Devo and an Expanding Evolutionary Synthesis: A Genetic Theory of Morphological Evolution. Cell. 2008;134: 2536. doi:10.1016/j.cell.2008.06.030
6. Prudhomme B, Gompel N, Carroll SB. Emerging principles of regulatory evolution. PNAS. 2007;104: 86058612. doi:10.1073/pnas.0700488104
7. Wray GA. The evolutionary significance of cis-regulatory mutations. Nat Rev Genet. 2007;8: 206216. doi:10.1038/nrg2063
8. Stern DL, Orgogozo V. The Loci of Evolution: How Predictable is Genetic Evolution? Evolution. 2008;62: 21552177. doi:10.1111/j.1558-5646.2008.00450.x
9. Stern DL, Frankel N. The structure and evolution of cis-regulatory regions: the shaven-baby story. Philos Trans R Soc Lond, B, Biol Sci. 2013;368: 20130028. doi:10.1098/rstb.2013.0028
10. Gompel N, Prudhomme B, Wittkopp PJ, Kassner VA, Carroll SB. Chance caught on the wing: cis-regulatory evolution and the origin of pigment patterns in Drosophila. Nature. 2005;433: 481487. doi:10.1038/nature03235
11. OBrown NM, Summers BR, Jones FC, Brady SD, Kingsley DM. A recurrent regulatory change underlying altered expression and Wnt response of the stickleback armor plates gene

EDA. eLife. 2015;4: e05290. doi:10.7554/eLife.05290
12. Chen H, Capellini TD, Schoor M, Mortlock DP, Reddi AH, Kingsley DM. Heads, shoulders, elbows, knees, and toes: modular Gdf5 enhancers control different joints in the vertebrate skeleton. PLOS Genetics. 2016;12: e1006454. doi:10.1371/journal.pgen.1006454
13. Chan YF, Marks ME, Jones FC, Villarreal G, Shapiro MD, Brady SD, et al. Adaptive evolution of pelvic reduction in sticklebacks by recurrent deletion of a Pitx1 enhancer. Science. 2010;327: 302305. doi:10.1126/science.1182213
14. Shapiro MD, Marks ME, Peichel CL, Blackman BK, Nereng KS, Jnsson B, et al. Genetic and developmental basis of evolutionary pelvic reduction in threespine sticklebacks. Nature. 2004;428: 717. doi:10.1038/nature02415
15. Erickson PA, Baek J, Hart JC, Cleves PA, Miller CT. Genetic Dissection of a Supergene Implicates Tfap2a in Craniofacial Evolution of Threespine Sticklebacks. Genetics. 2018; genetics.300760.2018. doi:10.1534/genetics.118.300760
16. Cleves PA, Ellis NA, Jimenez MT, Nunez SM, Schluter D, Kingsley DM, et al. Evolved tooth gain in sticklebacks is associated with a cis-regulatory allele of Bmp6. PNAS. 2014;111: 1391213917. doi:10.1073/pnas.1407567111
17. Fraser GJ, Hulsey CD, Bloomquist RF, Uyesugi K, Manley NR, Streelman JT. An ancient gene network is co-opted for teeth on old and new jaws. PLoS Biol. 2009;7: e1000031. doi:10.1371/journal.pbio.1000031
18. Fraser GJ, Bloomquist RF, Streelman JT. Common developmental pathways link tooth shape to regeneration. Dev Biol. 2013;377: 399414. doi:10.1016/j.ydbio.2013.02.007
19. Ellis NA, Donde NN, Miller CT. Early development and replacement of the stickleback dentition. Journal of Morphology. 2016;277: 10721083. doi:10.1002/jmor.20557
20. Hulsey CD, Machado-Schiaffino G, Keicher L, Ellis-Soto D, Henning F, Meyer A. The Integrated Genomic Architecture and Evolution of Dental Divergence in East African Cichlid Fishes (Haplochromis chilotes x H. nyererei). G3 (Bethesda). 2017;7: 31953202. doi:10.1534/g3.117.300083
21. Caldecutt WJ, Bell MA, Buckland-Nicks JA. Sexual Dimorphism and Geographic Variation in Dentition of Threespine Stickleback, Gasterosteus aculeatus. Copeia. 2001;2001: 936944.
22. Miller CT, Glazer AM, Summers BR, Blackman BK, Norman AR, Shapiro MD, et al. Modular Skeletal Evolution in Sticklebacks Is Controlled by Additive and Clustered Quantitative Trait Loci. Genetics. 2014;197: 405420. doi:10.1534/genetics.114.162420
23. Ellis NA, Glazer AM, Donde NN, Cleves PA, Agoglia RM, Miller CT. Distinct developmental genetic mechanisms underlie convergently evolved tooth gain in sticklebacks. Development. 2015;142: 24422451. doi:10.1242/dev.124248
24. Colosimo PF, Hosemann KE, Balabhadra S, Villarreal G, Dickson M, Grimwood J, et al. Widespread parallel evolution in sticklebacks by repeated fixation of ectodysplasin alleles. Science. 2005;307: 19281933. doi:10.1126/science.1107239
25. Lescak EA, Bassham SL, Catchen J, Gelmond O, Sherbick ML, Hippel FA von, et al. Evolution of stickleback in 50 years on earthquake-uplifted islands. PNAS. 2015;112: E7204E7212. doi:10.1073/pnas.1512020112

26. Glazer AM, Cleves PA, Erickson PA, Lam AY, Miller CT. Parallel developmental genetic features underlie stickleback gill raker evolution. EvoDevo. 2014;5: 19. doi:10.1186/2041-9139-5-19
27. Glazer AM, Killingbeck EE, Mitros T, Rokhsar DS, Miller CT. Genome Assembly Improvement and Mapping Convergently Evolved Skeletal Traits in Sticklebacks with Genotyping-by-Sequencing. G3: Genes, Genomes, Genetics. 2015; g3.115.017905. doi:10.1534/g3.115.017905
28. Erickson PA, Glazer AM, Killingbeck EE, Agoglia RM, Baek J, Carsanaro SM, et al. Partially repeatable genetic basis of benthic adaptation in threespine sticklebacks: repeatable evolution in benthic sticklebacks. Evolution. 2016;70: 887902. doi:10.1111/evo.12897
29. Sun C, Southard C, Witonsky DB, Olopade OI, Di Rienzo A. Allelic imbalance identifies novel tissue specific cis-regulatory variation for human UGT2B15. Hum Mutat. 2010;31: 99107. doi:10.1002/humu.21145
30. Bloomquist RF, Parnell NF, Phillips KA, Fowler TE, Yu TY, Sharpe PT, et al. Coevolutionary patterning of teeth and taste buds. PNAS. 2015;112: E5954E5962. doi:10.1073/pnas.1514298112
31. Rebello MR, Aktas A, Medler KF. Expression of Calcium Binding Proteins in Mouse Type II Taste Cells. J Histochem Cytochem. 2011;59: 530539. doi:10.1369/0022155411402352
32. Jakes R, Spillantini MG, Goedert M. Identification of two distinct synucleins from human brain. FEBS Lett. 1994;345: 2732.
33. Tanji K, Mori F, Nakajo S, Imaizumi T, Yoshida H, Hirabayashi T, et al. Expression of beta-synuclein in normal human astrocytes. Neuroreport. 12: 28452848.
34. Sarath Babu N, Murthy CLN, Kakara S, Sharma R, Brahmendra Swamy CV, Idris MM. 1-Methyl-4-phenyl-1,2,3,6-tetrahydropyridine induced Parkinsons disease in zebrafish. Proteomics. 2016;16: 14071420. doi:10.1002/pmic.201500291
35. Labeit S, Kolmerer B. Titins: Giant Proteins in Charge of Muscle Ultrastructure and Elasticity. Science. 1995;270: 293296. doi:10.1126/science.270.5234.293
36. Myhre JL, Hills JA, Prill K, Wohlgemuth SL, Pilgrim DB. The titin A-band rod domain is dispensable for initial thick filament assembly in zebrafish. Developmental Biology. 2014;387: 93108. doi:10.1016/j.ydbio.2013.12.020
37. Gene expression in tooth (WWW database), http://bite-it.helsinki.fi. Developmental Biology Programme of the University of Helsink; 1996.
38. OConnell DJ, Ho JWK, Mammoto T, Turbe-Doan A, OConnell JT, Haseley PS, et al. A WNT-BMP feedback circuit controls intertissue signaling dynamics in tooth organogenesis. Sci Signal. 2012;5: ra4. doi:10.1126/scisignal.2002414
39. Jones FC, Grabherr MG, Chan YF, Russell P, Mauceli E, Johnson J, et al. The genomic basis of adaptive evolution in threespine sticklebacks. Nature. 2012;484: 5561. doi:10.1038/nature10944
40. Roosing S, Rohrschneider K, Beryozkin A, Sharon D, Weisschuh N, Staller J, et al. Mutations in RAB28, Encoding a Farnesylated Small GTPase, Are Associated with Autosomal-Recessive Cone-Rod Dystrophy. Am J Hum Genet. 2013;93: 110117. doi:10.1016/j.ajhg.2013.05.005

41. Puig-Hervs MT, Temtamy S, Aglan M, Valencia M, Martnez-Glez V, Ballesta-Martnez MJ, et al. Mutations in PLOD2 cause autosomal-recessive connective tissue disorders within the Bruck syndrome–osteogenesis imperfecta phenotypic spectrum. Hum Mutat. 2012;33: 14441449. doi:10.1002/humu.22133
42. Ha-Vinh R, Alanay Y, Bank RA, Campos-Xavier AB, Zankl A, Superti-Furga A, et al. Phenotypic and molecular characterization of Bruck syndrome (osteogenesis imperfecta with contractures of the large joints) caused by a recessive mutation in PLOD2. Am J Med Genet A. 2004;131: 115120. doi:10.1002/ajmg.a.30231
43. Heikkinen J, Risteli M, Wang C, Latvala J, Rossi M, Valtavaara M, et al. Lysyl hydroxylase 3 is a multifunctional protein possessing collagen glucosyltransferase activity. J Biol Chem. 2000;275: 3615836163. doi:10.1074/jbc.M006203200
44. Du H, Pang M, Hou X, Yuan S, Sun L. PLOD2 in cancer research. Biomed Pharmacother. 2017;90: 670676. doi:10.1016/j.biopha.2017.04.023
45. Henikoff S, Henikoff JG. Amino acid substitution matrices from protein blocks. Proc Natl Acad Sci USA. 1992;89: 1091510919.
46. Jinek M, East A, Cheng A, Lin S, Ma E, Doudna J. RNA-programmed genome editing in human cells. eLife. 2013;2. doi:10.7554/eLife.00471
47. Jinek M, Chylinski K, Fonfara I, Hauer M, Doudna JA, Charpentier E. A Programmable Dual-RNAGuided DNA Endonuclease in Adaptive Bacterial Immunity. Science. 2012;337: 816821. doi:10.1126/science.1225829
48. Erickson PA, Ellis NA, Miller CT. Microinjection for Transgenesis and Genome Editing in Threespine Sticklebacks. JoVE (Journal of Visualized Experiments). 2016; e54055e54055. doi:10.3791/54055
49. Hart JC, Miller CT. Sequence-Based Mapping and Genome Editing Reveal Mutations in Stickleback Hps5 Cause Oculocutaneous Albinism and the casper Phenotype. G3 (Bethesda). 2017;7: 31233131. doi:10.1534/g3.117.1125
50. Gistelinck C, Witten PE, Huysseune A, Symoens S, Malfait F, Larionova D, et al. Loss of Type I Collagen Telopeptide Lysyl Hydroxylation Causes Musculoskeletal Abnormalities in a Zebrafish Model of Bruck Syndrome. J Bone Miner Res. 2016;31: 19301942. doi:10.1002/jbmr.2977
51. Lin CR, Kioussi C, OConnell S, Briata P, Szeto D, Liu F, et al. Pitx2 regulates lung asymmetry, cardiac positioning and pituitary and tooth morphogenesis. Nature. 1999;401: 279282. doi:10.1038/45803
52. Lu MF, Pressman C, Dyer R, Johnson RL, Martin JF. Function of Rieger syndrome gene in left-right asymmetry and craniofacial development. Nature. 1999;401: 276278. doi:10.1038/45797
53. Semina EV, Reiter R, Leysens NJ, Alward WL, Small KW, Datson NA, et al. Cloning and characterization of a novel bicoid-related homeobox transcription factor gene, RIEG, involved in Rieger syndrome. Nat Genet. 1996;14: 392399. doi:10.1038/ng1296-392
54. Liu Y, Semina EV. pitx2 Deficiency Results in Abnormal Ocular and Craniofacial Development in Zebrafish. PLOS ONE. 2012;7: e30896. doi:10.1371/journal.pone.0030896
55. Ji Y, Buel SM, Amack JD. Mutations in zebrafish pitx2 model congenital malformations

in Axenfeld-Rieger syndrome but do not disrupt left-right placement of visceral organs. Dev Biol. 2016;416: 6981. doi:10.1016/j.ydbio.2016.06.010
56. Stock DW, Jackman WR, Trapani J. Developmental genetic mechanisms of evolutionary tooth loss in cypriniform fishes. Development. 2006;133: 31273137. doi:10.1242/dev.02459
57. Jernvall J, Thesleff I. Tooth shape formation and tooth renewal: evolving with the same signals. Development. 2012;139: 34873497. doi:10.1242/dev.085084
58. Tucker AS, Fraser GJ. Evolution and developmental diversity of tooth regeneration. Seminars in Cell & Developmental Biology. doi:10.1016/j.semcdb.2013.12.013
59. Fraser GJ, Bloomquist RF, Streelman JT. A periodic pattern generator for dental diversity. BMC Biology. 2008;6: 32. doi:10.1186/1741-7007-6-32
60. Thesleff I, Tummers M. Tooth organogenesis and regeneration. StemBook. Cambridge (MA): Harvard Stem Cell Institute; 2008. Available:
http://www.ncbi.nlm.nih.gov/books/NBK27071/
61. Hjalt TA, Semina EV, Amendt BA, Murray JC. The Pitx2 protein in mouse development. Dev Dyn. 2000;218: 195200. doi:10.1002/(SICI)1097-0177(200005)218:1¡195::AID-DVDY17¿3.0.CO;2-C
62. Fraser GJ, Graham A, Smith MM. Conserved deployment of genes during odontogenesis across osteichthyans. Proc Biol Sci. 2004;271: 23112317. doi:10.1098/rspb.2004.2878
63. Jackman WR, Draper BW, Stock DW. Fgf signaling is required for zebrafish tooth development. Dev Biol. 2004;274: 139157. doi:10.1016/j.ydbio.2004.07.003
64. Fraser GJ, Berkovitz BK, Graham A, Smith MM. Gene deployment for tooth replacement in the rainbow trout (Oncorhynchus mykiss): a developmental model for evolution of the osteichthyan dentition. Evol Dev. 2006;8: 446457. doi:10.1111/j.1525-142X.2006.00118.x
65. Amendt BA. The Role of PITX2 in Tooth Development. The Molecular Mechanisms of Axenfeld-Rieger Syndrome. Springer, Boston, MA; 2005. pp. 81-92. doi:10.1007/0-387-28672-1_8
66. Idrees F, Bloch-Zupan A, Free SL, Vaideanu D, Thompson PJ, Ashley P, et al. A novel homeobox mutation in the PITX2 gene in a family with Axenfeld-Rieger syndrome associated with brain, ocular, and dental phenotypes. Am J Med Genet B Neuropsychiatr Genet. 2006;141B: 184191. doi:10.1002/ajmg.b.30237
67. Wautier K, Van der heyden C, Huysseune A. A quantitative analysis of pharyngeal tooth shape in the zebrafish (Danio rerio, Teleostei, Cyprinidae). Arch Oral Biol. 2001;46: 67-75.
68. Gibert Y, Bernard L, Debiais-Thibaud M, Bourrat F, Joly J-S, Pottin K, et al. Formation of oral and pharyngeal dentition in teleosts depends on differential recruitment of retinoic acid signaling. FASEB J. 2010;24: 32983309. doi:10.1096/fj.09-147488
69. Miller CT, Beleza S, Pollen AA, Schluter D, Kittles RA, Shriver MD, et al. cis-Regulatory Changes in Kit Ligand Expression and Parallel Evolution of Pigmentation in Sticklebacks and Humans. Cell. 2007;131: 11791189. doi:10.1016/j.cell.2007.10.055
70. Zhou P, Liu Y, Lv F, Nie M, Jiang Y, Wang O, et al. Novel Mutations in FKBP10 and PLOD2 Cause Rare Bruck Syndrome in Chinese Patients. PLOS ONE. 2014;9: e107594. doi:10.1371/journal.pone.0107594
71. Yapiciolu H, Ozcan K, Arikan O, Satar M, Narli N, Ozbek MH. Bruck syndrome: os-

teogenesis imperfecta and arthrogryposis multiplex congenita. Ann Trop Paediatr. 2009;29: 159162. doi:10.1179/146532809X440798
72. Forlino A, Cabral WA, Barnes AM, Marini JC. New perspectives on osteogenesis imperfecta. Nature Reviews Endocrinology. 2011;7: 540557. doi:10.1038/nrendo.2011.81
73. Forlino A, Marini JC. Osteogenesis imperfecta. The Lancet. 2016;387: 16571671. doi:10.1016/S0140-6736(15)00728-X
74. Salo AM, Cox H, Farndon P, Moss C, Grindulis H, Risteli M, et al. A Connective Tissue Disorder Caused by Mutations of the Lysyl Hydroxylase 3 Gene. The American Journal of Human Genetics. 2008;83: 495503. doi:10.1016/j.ajhg.2008.09.004
75. Rautavuoma K, Takaluoma K, Sormunen R, Myllyharju J, Kivirikko KI, Soininen R. Premature aggregation of type IV collagen and early lethality in lysyl hydroxylase 3 null mice. PNAS. 2004;101: 1412014125. doi:10.1073/pnas.0404966101
76. Li A, Lai Y-C, Figueroa S, Yang T, Widelitz RB, Kobielak K, et al. Deciphering principles of morphogenesis from temporal and spatial patterns on the integument. Dev Dyn. 2015;244: 905920. doi:10.1002/dvdy.24281
77. Pispa J, Thesleff I. Mechanisms of ectodermal organogenesis. Dev Biol. 2003;262: 195205.
78. Hjalt TA, Amendt BA, Murray JC. Pitx2 Regulates Procollagen Lysyl Hydroxylase (Plod) Gene Expression. The Journal of Cell Biology. 2001;152: 545552. doi:10.1083/jcb.152.3.545
79. Ghosh P, Saha SK, Nandi SS, Bhattacharya S, Roy SS. Involvement of Pitx2, a homeodomain transcription factor, in hypothyroidism associated reproductive disorders. Cell Physiol Biochem. 2007;20: 887898. doi:10.1159/000110449
80. Erickson PA, Cleves PA, Ellis NA, Schwalbach KT, Hart JC, Miller CT. A 190 base pair, TGF- responsive tooth and fin enhancer is required for stickleback Bmp6 expression. Developmental Biology. 2015;401: 310323. doi:10.1016/j.ydbio.2015.02.006
81. Ellis NA, Miller CT. Dissection and flat-mounting of the threespine stickleback branchial skeleton. Journal of Visualized Experiments. 2016; doi:10.3791/54056
82. Langmead B, Salzberg SL. Fast gapped-read alignment with Bowtie 2. Nat Meth. 2012;9: 357359. doi:10.1038/nmeth.1923
83. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, et al. The sequence alignment/map format and SAMtools. Bioinformatics. 2009;25: 20782079. doi:10.1093/bioinformatics/btp352
84. McKenna A, Hanna M, Banks E, Sivachenko A, Cibulskis K, Kernytsky A, et al. The Genome Analysis Toolkit: A MapReduce framework for analyzing next-generation DNA sequencing data. Genome Res. 2010;20: 12971303. doi:10.1101/gr.107524.110
85. DePristo MA, Banks E, Poplin R, Garimella KV, Maguire JR, Hartl C, et al. A framework for variation discovery and genotyping using next-generation DNA sequencing data. Nat Genet. 2011;43: 491498. doi:10.1038/ng.806
86. Van der Auwera GA, Carneiro MO, Hartl C, Poplin R, Del Angel G, Levy-Moonshine A, et al. From FastQ data to high confidence variant calls: the Genome Analysis Toolkit best practices pipeline. Curr Protoc Bioinformatics. 2013;43: 11.10.1-33.

doi:10.1002/0471250953.bi1110s43
87. Ronquist F, Teslenko M, van der Mark P, Ayres DL, Darling A, Hhna S, et al. MrBayes 3.2: Efficient Bayesian Phylogenetic Inference and Model Choice Across a Large Model Space. Syst Biol. 2012;61: 539542. doi:10.1093/sysbio/sys029
88. Sbastien L, Julie Josse, Franois Husson. FactoMineR: an R package for multivariate analysis — L — Journal of Statistical Software. Journal of Statistical Software. 2008;25. Available: https://www.jstatsoft.org/article/view/v025i01
89. Holm S. A Simple Sequentially Rejective Multiple Test Procedure. Scandinavian Journal of Statistics. 1979;6: 6570.
90. Broman KW, Wu H, Sen S, Churchill GA. R/qtl: QTL mapping in experimental crosses. Bioinformatics. 2003;19: 889890.
91. Rozen S, Skaletsky HJ. Primer3. 1998.
92. Singhal H, Ren YR, Kern SE. Improved DNA Electrophoresis in Conditions Favoring Polyborates and Lewis Acid Complexation. PLOS ONE. 2010;5: e11318. doi:10.1371/journal.pone.0011318
93. Quinlan AR, Hall IM. BEDTools: a flexible suite of utilities for comparing genomic features. Bioinformatics. 2010;26: 841842. doi:10.1093/bioinformatics/btq033
94. Bray NL, Pimentel H, Melsted P, Pachter L. Near-optimal probabilistic RNA-seq quantification. Nature Biotechnology. 2016;34: 525527. doi:10.1038/nbt.3519
95. Askary A, Smeeton J, Paul S, Schindler S, Braasch I, Ellis NA, et al. Ancient origin of lubricated joints in bony vertebrates. Elife. 2016;5. doi:10.7554/eLife.16415
96. Sander JD, Maeder ML, Reyon D, Voytas DF, Joung JK, Dobbs D. ZiFiT (Zinc Finger Targeter): an updated zinc finger engineering tool. Nucleic Acids Res. 2010;38: W462W468. doi:10.1093/nar/gkq319
97. Talbot JC, Amacher SL. A Streamlined CRISPR Pipeline to Reliably Generate Zebrafish Frameshifting Alleles. Zebrafish. 2014;11: 583585. doi:10.1089/zeb.2014.1047
98. Eden E, Navon R, Steinfeld I, Lipson D, Yakhini Z. GOrilla: a tool for discovery and visualization of enriched GO terms in ranked gene lists. BMC Bioinformatics. 2009;10: 48. doi:10.1186/1471-2105-10-48
99. Marchler-Bauer A, Bo Y, Han L, He J, Lanczycki CJ, Lu S, et al. CDD/SPARCLE: functional classification of proteins via subfamily domain architectures. Nucleic Acids Res. 2017;45: D200D203. doi:10.1093/nar/gkw1129