Title
Functional Genetic Analysis of Stickleback Craniofacial Evolution

Author
Erickson, Priscilla Ashley

Publication Date
2016

Peer reviewed|Thesis/dissertation

Functional Genetic Analysis of Stickleback Craniofacial Evolution

By

Priscilla Ashley Erickson

A dissertation submitted in partial satisfaction of the

requirements for the degree of

Doctor of Philosophy

in

Molecular and Cell Biology

in the

Graduate Division

of the

University of California, Berkeley

Committee in charge:

Professor Craig T. Miller, Chair
Professor Erica Bree Rosenblum
Professor Nipam H. Patel
Professor Gian Garriga

Spring 2016

Functional Genetic Analysis of Stickleback Craniofacial Evolution

By
Priscilla Ashley Erickson

Abstract

Functional Genetic Analysis of Stickleback Craniofacial Evolution

by

Priscilla Ashley Erickson

Doctor of Philosophy in Molecular and Cell Biology

University of California, Berkeley

Professor Craig T. Miller, Chair

The biosphere contains an incredible level of natural morphological diversity, and most differences within and between species can be explained by evolved differences in their genetic code. While traditional genetics has made great strides to connect genes to phenotypes in laboratory strains of model organisms, understanding the link between genotype and phenotype in natural populations is one of the greatest challenges of modern biology. Acquiring the genome sequences of organisms has been facilitated by rapidly advancing technologies, but connecting genetic variants to evolved differences remains elusive. What types of genetic changes underlie adaptive differences in morphology? How predictable is the path of evolution? Do individual mutations control multiple adaptive phenotypes? Answering these questions requires harnessing the power of modern genetics in a system with naturally evolved phenotypic variation.

Chapter one outlines key questions in the fields of evolutionary developmental biology and adaptation genetics. It describes the *cis*-regulatory hypothesis for the genetic basis of morphological evolution and occurrence of supergenes that control multiple evolved phenotypes. It describes the natural history of the threespine stickleback fish and explains why the stickleback system is an outstanding model to tackle these questions. *Gasterosteus aculeatus* has adapted to unique habitats across the Northern Hemisphere and is amenable to both forward and reverse genetic studies. Marine and freshwater sticklebacks eat different foods and have different adaptations in the skeletal elements used to process food. Quantitative trait locus (QTL) mapping has demonstrated that a large number of genomic regions control the evolution of these skeletal traits, but a few key regions control a disproportionate number of traits. The following chapters investigate the developmental and genetic bases of two evolved skeletal changes that are controlled by the same genomic regions.

Chapter two explores the genetic and developmental basis of the elongation of the branchial bones of the throat in freshwater sticklebacks. Elongation of these bones expands the buccal cavity, likely enabling freshwater fish to consume larger prey items. This increase in bone length is found in both wild and lab-reared fish from two populations, suggesting heritable convergent evolution in freshwater environments. In one population, an early increase in cartilage size contributes to increased bone length, and in both populations the bones grow faster throughout development. In both freshwater populations, the increase in bone length maps to two chromosomes: 4 and 21, with distinct effects of these two chromosomes on individual bones over the course of development, but similar effects in each cross. Collectively, these results suggest a largely parallel genetic and developmental basis of evolved bone length gain in two populations.

Chapter three describes further mapping and functional testing of the chromosome 21 bone length QTL. While pharyngeal tooth gain maps to a regulatory haplotype of the gene *Bone Morphogenetic Protein 6* (*Bmp6*), bone length gain maps to a nearby region containing the gene *Tfap2a* in two freshwater populations of sticklebacks. Therefore, evolved pharyngeal tooth gain and bone length gain are controlled by separate loci. *Tfap2a* is an important transcriptional regulator of craniofacial development and produces severe craniofacial phenotypes when mutated in vertebrates. In sticklebacks, the freshwater allele of *Tfap2a* is downregulated in the developing branchial skeleton of hybrid animals and deletion of *Tfap2a* causes a nearly complete absence of pharyngeal arch-derived skeletal elements. Heterozygous loss of *Tfap2a* alters branchial bone length, suggesting that dosage of this gene is important to determining bone patterning. Combined with previous findings in the lab, these results suggest that closely linked regulatory changes to two key developmental patterning genes produce skeletal gain phenotypes.

Chapter four investigates the extent of genetic parallelism for repeated phenotypic evolution. In British Columbia, several lakes have independently evolved two freshwater stickleback ecotypes: a bottom-dwelling benthic form and an open-water limnetic form. Using crosses of benthic populations from three lakes, this study tests whether the genetic architecture underlying skeletal differences between benthic and marine individuals is repeatable across lakes. The majority of genomic regions underlying skeletal differences are unique to an individual lake, but there is more parallelism of QTL than expected by chance in simulations. Furthermore, the chromosome 21 QTL controlling bone length and tooth number were identified in multiple lakes, suggesting that these loci may be adaptive in the benthic habitat. These findings suggest that benthic evolution in three lakes has a significantly parallel but largely nonparallel basis.

Chapter five examines the regulation of the gene *Bmp6*, which, like the bone length QTL, is found on chromosome 21 and likely underlies evolved tooth gain in sticklebacks via a *cis*-regulatory down-regulation of the freshwater allele. A short conserved regulatory element upstream of *Bmp6* drives robust reporter gene expression during tooth development in both sticklebacks and distantly related zebrafish. This enhancer responds to TGFß signaling, likely via SMAD3 binding, and the enhancer is required for normal expression of *Bmp6*. Therefore, changes to additional regulatory loci

controlling *Bmp6* and interacting with this enhancer may underlie pharyngeal tooth number evolution.

Finally, the future of stickleback molecular genetics will rely on functional genetic manipulations that will be facilitated by the emerging genome-editing revolution. The Appendix outlines a protocol for generating transgenic sticklebacks (carrying both transgenes and genome-edited alleles) using techniques developed and optimized over the course of the experiments described in chapters three and five. This protocol is intended to serve as a resource for the fish evolution and development community.

Combined, the results described here offer several insights towards the molecular genetic and developmental basis of evolved skeletal change. Two adaptive alleles controlling related traits (tooth number and bone length gain) are found tightly linked in the genome, indicating that linkage of the QTL controlling these phenotypes may be adaptive for rapid colonization of freshwater habitats. Both QTL are associated with *cis*-regulatory down-regulation of candidate genes with highly pleiotropic roles during development. This finding suggests that skeletal gain traits may be readily accomplished by a loss of gene expression. Future studies will attempt to identify the causative mutations responsible for each trait and examine their frequencies and evolutionary histories in natural stickleback populations. Additional studies will attempt to identify the precise developmental effects of the regulatory mutations underlying the evolved differences.

Dedicated in memory of my dad, Dr. Leonard C. Erickson (1946-2013), an accomplished molecular biologist who inspired me with his love for nature, photography, and adventure; helped me get excited about research at a young age; and was a constant source of encouragement and support throughout my academic pursuits. He is greatly loved, appreciated, and missed.

Acknowledgements

My advisor, Dr. Craig Miller has been a tremendous influence on my scientific development, most especially in encouraging clear and concise communication in both writing and speaking. I appreciate his commitment to my success in the lab and his support of my future career goals—not many PIs invest in their students as much as Craig does. I feel fortunate to be among the first members of the Miller lab and am so excited to see what great work comes out of it in the future. I am grateful to my committee members, past and present: Drs. Nipam Patel, Michael Levine, Gian Garriga, and especially Bree Rosenblum, for their scientific input as well as career guidance. Richard "Rudy" Rudersdorf at the University of Wisconsin taught me almost everything I know about molecular biology techniques, which was immensely helpful as I got started on my own projects. I am indebted to my past mentors at Kenyon College: Drs. Robert Mauck, Mark Haussmann, Karen Hicks, and Wade Powell for providing me a phenomenal undergraduate education and the foundation in scientific thinking that was the cornerstone for my work in graduate school. I hope that one day I will be as influential to future students as they were to me.

Nick Ellis and I joined the Miller lab at the same time and have been through it all together. He has been an incredibly kind, understanding, and generous coworker and, most importantly, a great friend. Phillip Cleves and Andrew Glazer were invaluable mentors and friends as I got started in the Miller lab. Phil's enthusiasm for science made every day interesting, and Andrew's generosity in sharing data analysis expertise were a tremendous help in several projects. James Hart has generously shared his computational biology skills, and I look forward to seeing him carry on the Miller lab tradition. I have had the privilege of mentoring a number of talented undergraduates, including Alyson Smith, Joan Baek, Clement Kao, Aloukika Shah, and Kristen Huang, who were a joy to work with. Alyson and Joan each stayed in the lab for multiple years and made substantial contributions to the projects in this dissertation. I truly appreciate their patience with me as I have learned to be a better mentor, and I wish them all the best as their future careers in science unfold. Dozens of other undergraduates in the lab performed dissections, phenotyping, and genotyping that contributed to the data in this thesis, and I appreciate their diligence and hard work to build a tremendously powerful genetic dataset. Their enthusiasm and curiosity also helped make the lab a fun and rewarding place to work.

I couldn't have made it through graduate school without a wonderful group of supportive friends. Ashley Stone has been my family away from home in California. Sarah May has made my last year in Berkeley so much fun since she moved to Oakland. I've enjoyed many movie nights with Aisha Ellahi, Alisha Ellis (and baby Lincoln!), and Debbie Thurtle. I have also been fortunate to explore the natural beauty of California with Hanna Engelke and too many other backpacking, kayaking, and climbing buddies to individually name. Those trips into the wilderness will always be among my favorite memories from graduate school. Finally, the Cal Aquatics Masters swim team has helped me maintain my physical and mental health. I am especially grateful for the friendship of Lisa Kronstad, Scott Adams, Jim Lunt, Peter Hepburn, and many other cheerful lane-

mates and enthusiastic coaches who have made swimming such a fun and important part of my life. I am also grateful to my therapists, Virginia and Carey, who helped me navigate the rough patches and reminded me to prioritize my goals, values, and well-being.

Anne Dodson and I decided to be roommates on a whim, and it is one of the best decisions I have ever made. Through six years and three apartments, we have been through all the ups and downs of graduate school together. We've cooked nearly every recipe on Smitten Kitchen, hosted parties, and been to a slightly embarrassing number of Chris Thile concerts. We have shared many laughs, fun memories, and even occasional tears, and it is hard to imagine a "home" without her there.

I am so grateful to have been raised in a family that valued and encouraged inquiry and education. Although my dad, Len Erickson, is not here today to read this, he truly inspired me to get started in science and provided thoughtful advice at every step of the way. My mom, Pam Erickson, and sister, Laura Erickson have been wonderfully supportive and encouraging throughout my studies. Our family has undergone many changes and challenges in the past three years, and though we are geographically far apart, I am so grateful for their love and support. My mom is always there when I need to talk, and I have so much respect for her open-mindedness and diplomatic approach to life. Traveling to England to be part of Laura's wedding to Adam McGinness was certainly a highlight of the past several years and I am excited to have him as a part of the family. I also thank my extended family, including my grandma, Uncle Cliff, Aunt Charl, and cousins Kari and Kirsten and their families for their support through both happy and sad times.

Lastly, I have to thank (Professor!) Miles Johnson. We met when I was halfway through my PhD, and I can't imagine the past three years without him. Even though we have been apart for most of the last two years, he is always there to listen (usually with a silly joke to make me smile). He has been my best friend and greatest cheerleader, and I am so grateful to have had him and his family as a part of my life. I have many incredible memories of the time we've spent together and the adventures we've had in Hawaii, the high Sierra, the deserts of Southern California, and everywhere in between. I am so grateful for his love and support, and incredibly proud of his persistence and his successes. I can't wait for our upcoming travels and look forward to starting the next stage of our lives together in Virginia.

Table of Contents

3. Genetic dissection of a supergene implicates *Tfap2a* as underlying craniofacial evolution in threespine sticklebacks

5. Partially repeatable genetic basis of benthic adaptation in threespine sticklebacks

A. Microinjection for transgenesis and genome editing in threespine sticklebacks

1. Insights towards the genetic and developmental basis of morphological diversity using threespine sticklebacks

"This is the assembly of life that took a billion years to evolve. It has eaten the storms—folded them into its genes—and created the world that created us."

- E. O. Wilson, *The Diversity of Life* (1999)

The natural world is filled with breathtaking biological diversity across every kingdom of life. Biology has found innumerable adaptations to even the most extreme environments, and within animals, a spectacular array of shapes and forms has evolved to facilitate locomotion, feeding, and reproduction in diverse habitats. Since the Modern Synthesis (Huxley 1942) explained how natural selection could act upon genes and mutations to promote evolution, understanding the genetic changes underlying the evolution of these "*endless forms most beautiful*" (Darwin 1859) has been a key goal in evolutionary biology. While the bizarre and beautiful outcomes of evolution are tantalizing to contemplate, this dissertation focuses on the power of quantitative and genetically tractable natural morphological variation to elucidate the genetic basis of evolution.

A brief history of natural variation and the genetic basis of adaptation

"Nothing at first can appear more difficult to believe than that the more complex organs and instincts should have been perfected … by the accumulation of innumerable slight variations, each good for the individual possessor. Nevertheless, this difficulty … cannot be considered real if we admit the following propositions, namely, -- that gradations in the perfection of any organ or instinct, which we may consider, either do now exist or could have existed, each good of its kind, -- that all organs and instincts are, in ever so slight a degree, variable, -- and, lastly, that there is a struggle for existence leading to the preservation of each profitable deviation of structure or instinct."

- Charles Darwin, *The Origin of Species,* 1859

Darwin (1859) initially put forth a theory of evolution by means of natural selection, suggesting that the fittest individuals in a given environment would survive and produce the most offspring. He examined cases of this process occurring in the wild, most famously on the Galapagos Islands, where each island has birds, reptiles, and plants uniquely adapted to the particular environment. He wrote that these observations "*are explicable on the view of colonisation from the nearest and readiest source, together with*

the subsequent modification and better adaptation of the colonists to their new homes" (Darwin 1859) However, he also observed this process occurring on a recent and dramatic scale in the hands of humans, which have created hundreds of varieties of domesticated animals selected to fit unique aesthetic and functional purposes: "*The key is man's power of accumulative selection: nature gives successive variations; man adds them up in certain directions useful to him*." Ultimately, Darwin concluded that natural variation was required for selection to occur: "*Without variability, nothing can be effected; slight individual differences, however, suffice for the work, and are probably the chief or sole means in the production of new species."* While Darwin had no explanation for the mechanism by which offspring would inherit more fit phenotypes, his observations about the power of the accumulation of slight differences to drive evolutionary change would eventually provide the foundation for elucidating detailed mechanisms of evolution.

Working in the same era as Darwin, the Austrian monk Gregor Mendel made meticulous observations of pea phenotypes in experimental crosses that established the three fundamental laws governing a particulate means of genetic inheritance. Although Mendel may have recognized the significance of his findings for evolution ("*This seems to be the one correct way of finally reaching a solution to a question whose significance for the evolutionary history of organic forms cannot be underestimated*" (Mendel 1865 quoted in Charlesworth and Charlesworth 2009)), it wasn't until decades later that the work of Mendel was combined with the theories of Darwin to form a cohesive theory of evolution by natural selection of genetic alleles that produce fitness advantages. Mutations arising in natural populations can produce variation, which may improve survival or reproduction. Mutations may have different frequencies in different environments, and mutations that prevent interbreeding can lead to speciation (Huxley 1942).

But how do mutations actually change phenotypes and fitness? Fisher (1930) proposed that evolution occurred by countless mutations of tiny effect, the so-called "infinitesimal model" of evolution, as larger changes would be more likely to overshoot adaptive optima. This theory was later revised by Kimura, who suggested that small effect mutations would be subject to random loss by genetic drift, and therefore mutations of intermediate effect would be most likely to fix in populations (Kimura 1983). In the most recent rethinking of this model, Orr's work proposes that the fitness effects of adaptive mutations have an exponential distribution, that large fitness gains often occur early in adaptation, and that adaptation may skip over many beneficial alleles (Orr 1998, 2002, 2005). This informative theoretical work focuses on abstract phenotypes and mutations. Connecting real-world mutations to evolved phenotypes remains a primary goal for modern evolutionary biology.

Individual phenotypes can be controlled by one or many genetic loci. Traits controlled by a single locus (sometimes called "Mendelian") are frequently studied because the phenotype of interest can be readily followed in genetic crosses or pedigrees. Decades of groundbreaking work have tracked the link between mutations and Mendelian phenotypes in laboratory model organisms and human disease. However the

majority of traits seen in nature are complex (or polygenic or quantitative) phenotypes, which are controlled by two or more genomic loci. Complex phenotypes can be mapped to regions of the genome by looking for linkage between molecular markers and phenotypes (Lander and Botstein 1989). This linkage can be established by studying the offspring of crosses of two individuals with differing phenotypes (quantitative trait locus (QTL) mapping, Weller 1986), or by looking for association between genotype and phenotype in large interbreeding populations, such as humans (genome wide association mapping) (Hirschhorn and Daly 2005). Tremendous work is often required to identify the underlying mutations and conclusively link them to a phenotype, though recent decades have seen success towards these goal in both human medical genetics and evolutionary genetic.

The modern era of genetics therefore fulfills Darwin's prediction: *"A grand and almost untrodden field of inquiry will be opened, on the causes and laws of variation…"* (1859).

The *cis*-regulatory hypothesis and the evolution of morphology

"We suggest that evolutionary changes in anatomy and way of life are more often based on changes in the mechanisms controlling the expression of genes than on sequence changes in protein. We therefore propose that regulatory mutations account for the major biological differences between humans and chimpanzees."

--Mary-Claire King and Allan C. Wilson (1975)

In principle, an individual gene can vary in one of two main ways to change a phenotype. The *coding sequence*—the number and order of amino acids in the final protein, can vary by changing, adding, or removing codons in a gene. Changes to the protein coding sequence might affect its activity, interactions with other proteins, and/or localization in order to produce a phenotypic outcome. However, the *expression* of the protein in both time and space can also vary. Mechanisms to modulate gene and protein expression can include changes to the number of mRNA molecules produced, the stability of the mRNA molecules, the post-transcriptional splicing and processing of mRNA, the localization of mRNA at a cellular and organismal level, as well as regulation of translation and post-translational modifications. For a given gene of interest, changes within its own regulatory sequences are considered to be "*cis*" changes, whereas changes to an outside regulatory factor are changes in "*trans*" (Prud'homme et al. 2007; Pai and Gilad 2014).

The differentiation of cells and development of organisms is tightly controlled by the regulation of gene expression to pattern tissues. Many genes promoting developmental processes are expressed in different cell types during different stages of development. Expression of genes in time and space is controlled at a broad scale by chromatin topology, and at a finer scale by specific regulatory DNA sequences that bind combinations of transcriptional activators and repressors that are themselves expressed in

subsets of tissues in time and space. In a simplified model of gene regulation, a given enhancer DNA sequence controls expression of its target gene in one particular location or tissue in response to a unique combination of signaling inputs that bind the enhancer and help recruit RNA polymerase to the promoter. Fine tuning of this regulation through multiple enhancers and repressors allows complex tissue patterns and phenotypes to develop (reviewed in Pennacchio et al. 2013).

Because many genes are widely expressed, changes to them have the potential to be pleiotropic (affect multiple tissues or cell types). The current predominant theory for the evolution of morphological changes is that mutations affecting gene expression (*cis*-regulatory changes), rather than changes to amino acid sequence (coding changes), are common because changes to gene regulation avoid the potentially deleterious pleiotropic consequences of coding changes to widely-used genes (Stern 2000; Wray 2007; Carroll 2008). This idea was first proposed by King and Wilson (1975) upon the demonstration that the protein coding sequences of chimpanzees and humans are highly similar, suggesting that changes to regulation must underlie the vast differences between the great apes. Many genes affecting development of morphological structures are part of canonical signaling pathways (e.g. Wnt, Bmp, Fgf) that have been repeatedly coopted for the development of a wide variety of structures (Pires-daSilva and Sommer 2003). Thus, significant changes to the function of one gene could have substantial pleiotropic effects in all anatomical locations using the same signaling pathway. Nonetheless, coding changes have been found to underlie a number of instances of evolution, especially for physiological processes and coloration (Steiner et al. 2007; Rosenblum et al. 2010; McGlothlin et al. 2014; Vickrey et al. 2015).

Understanding regulatory variation is critical to the dissection of complex traits in humans and other organisms (Ward and Kellis 2012; Albert and Kruglyak 2015). The extent to which coding and regulatory changes affect the evolution of morphology will continue to be tested as the genetic basis of more evolved traits is determined. Furthermore, a better understanding of the rules governing the regulation of gene expression will facilitate interpretation of genome-wide association studies for disease, since the majority of loci associated with human disease are in non-coding regions (reviewed in Mathelier et al. 2015). Enhancers evolve rapidly, even for organs with highly conserved form and function, such as the mammalian liver (Villar et al. 2015), but binding site turnover can maintain similar functions with highly divergent sequences (Fisher et al. 2006). Characterizing the regulation of evolutionarily important genes will inform a broad-scale understanding of logic and structure of gene regulation and its evolution.

Adaptive radiations as windows into the evolution of morphology

"Actually, the entire ascent of life can be presented as an adaptive radiation in the time dimension. From the beginning of replicating molecules … each of these steps permitted the utilization of a different set of environmental resources, that is, the occupation of a different adaptive zone."

- Ernst Mayr, *What Evolution Is* (2001)

Adaptive radiations provide a particularly remarkable system to study the evolution of morphology. In adaptive radiations, a large number of species rapidly evolve from a common ancestor to fill unique environmental niches (Schluter 2000). Classic examples of adaptive radiations include the *Anolis* lizards of the Caribbean, the Galapagos finches, and the cichlids of the African great lakes. Because the species within adaptive radiations are often closely related, identifying the genes and genomic regions responsible for specific adaptive divergences may be facilitated because the members of the radiation will likely be highly genetically similar, except in parts of the genome that are important for controlling adaptive phenotypic differences (Berner and Salzburger 2015). Indeed, recent years have seen an explosion of studies identifying genetic differentiation between uniquely adapted populations and species. However, for many of these studies, connecting phenotype to genotype has proved to be quite challenging. Heroic efforts have identified some fascinating genetic changes associated with adaptive radiations, including changes to the gene *ALX1* associated with beak morphology of Galapagos finches (Lamichhaney et al. 2015), regulatory changes to *optix* associated with the red pattern of *Heliconius* butterflies (Reed et al. 2011), and regulatory changes to the *Pax7* gene associated with the orange blotch phenotype found in some cichlid species (Roberts et al. 2009). Yet a tremendous amount of diversity remains to be explained at a molecular level, both within and beyond the classic adaptive radiations.

Supergenes and the clustering of adaptive loci

"Thus, whenever we can discern a sufficiently strong tie, or cooperation, between structures or entities…we might expect to find all the features of polymorphism…adjusted to work together for their common advantage, the adjustment depending on a common genotype…"

- Kenneth Mather (1955)

It is easy to imagine that in certain environments, it may be beneficial to inherit phenotypic changes in multiple traits as a single genetic package. This can be accomplished in one of two ways: a single genetic change that affects multiple phenotypes (a pleiotropic mutation), or the very close linkage of multiple alleles affecting different phenotypes. Supergenes are defined as tightly linked loci controlling multiple phenotypes in which certain allele combinations are almost always inherited together (Schwander et al. 2014; Thompson and Jiggins 2014). Supergenes are perhaps best known for controlling distinct mimicry wing patterns in *Heliconius* butterflies (Joron et al. 2011) but were first described in grouse locusts (Nabours 1929, 1933) and later in *Primula* flowers (Mather 1950). Supergenes are also involved in shell patterning in *Partula* snails (Murray and Clarke 1976a,b) and pigmentation and behavior in sparrows and ruffs (Thomas et al. 2008; Küpper et al. 2016; Lamichhaney et al. 2016; Tuttle et al. 2016). Most described supergenes are controlled by chromosomal inversions that prevent

recombination between alleles that are inherited together. However, a single gene that controls a cascade leading to multiple phenotypes could also be responsible for a similar phenomenon (e.g. Kunte et al. 2014). For example, a change to a gene affecting skeletal growth could cause increases in size in a number of bones. In this case, the size of each tissue would map to the same genomic region. Relatively few pleiotropic QTL have been closely dissected, but multiple changes in the regulation of a single gene (*Agouti*) affect various aspects of pigmentation in mice (Linnen et al. 2013). Described cases of supergenes usually involve dramatic changes to morphology and/or behavior. Whether supergenes exist on quantitative scales of variation is not known. Are genetically linked phenotypes usually due to pleiotropy or due to different changes in closely linked genes?

The genetic and developmental basis of parallel and convergent evolution

"The operation of selecting agents, gradually and steadily bringing about the deceptive resemblance of a species to some other definite object, produces the impression of there being some innate principle in species which causes an advance of organization in a special direction."

- Henry Walter Bates (1862)

Convergent evolution is defined as the repeated evolution of similar phenotypes to serve similar purposes, and is often used as an argument that evolution might have predictable directions under certain conditions (Stern 2013). One of the most striking example of convergent evolution is the mimetic phenotypes of butterflies, in which non-toxic species mimic unrelated toxic species to gain protection from potential predators (Bates 1862). Other examples of convergent evolution include the evolution of pale colors in light, sandy environments in mice and lizards (Steiner et al. 2009; Rosenblum et al. 2010), the evolution of microbes to nutrient-limiting environments (Herron and Doebeli 2013), the evolution of limb length variation associated with habitat in Caribbean anoles (Losos 2009), and the evolution of distinct feeding morphologies in cichlid fish (Muschick et al. 2012). When such similar changes occur, do they use the same genetic loci? In other words, do parallel genetic changes underlie convergent phenotypic changes? Work in a wide variety of model systems has indicated that a full spectrum of possibilities, from different changes in different genes, to similar changes in the same genetic pathway, to identical yet independent changes in the same gene. When evolved changes require specific biochemical pathways, evolution may be constrained and often uses parallel genetics: evolving populations of bacteria undergoing experimental evolution (Dettman et al. 2012) and adapting to the lungs of patients with cystic fibrosis (Huse et al. 2010) repeatedly use the same genetic changes, and multiple species of herbivorous insect have evolved changes in the same ion transporter to resist milkweed toxin (Zhen et al. 2012). Other times, evolution recurrently uses the same genetic pathway: a wide variety of species use mutations in the same pigmentation gene, *MC1R* and its inhibitor, *Agouti*, to evolve light coloration on light backgrounds (Mundy 2005; Hoekstra 2006). However,

convergent radiations of Nicaraguan cichlids appear to use largely non-parallel genetic routes (Elmer et al. 2014), and further studies may reveal similar trends in other systems.

Parallelism and convergence can also be examined at a developmental level. For evolved morphology, many developmental pathways may feed into creating a single anatomical structure. Therefore, multiple options may be available to produce convergent phenotypes. For example in the aforementioned Caribbean *Anolis* model system, Sanger *et al (*2012*)* tested whether elongation of limbs in multiple species adapted to ground habitats was due to similar or different developmental mechanisms. The authors found that differences in very early patterning of the limb, prior to the formation of the cartilage template, result in longer limbs. In this case, similar developmental differences underlie similar changes, but one could imagine that some populations might modify limb length due to a difference in growth rate. In forest mice that have evolved longer tails than open field mice, both the number and the length of the vertebra is increased during development (Kingsley et al. 2016), suggesting that multiple developmental processes have contributed to the evolved changes. Pinpointing the developmental basis of evolved morphological differences provides information about the pathways available to and constraints that limit evolutionary processes.

The threespine stickleback model system

"Even those who measure the value of a science by its immediate application to human affairs can learn some important lessons from the study of this insignificant little fish."
-Niko Tinbergen, "The Curious Behavior of the Stickleback" (1952)

Tinbergen earned a Nobel prize for his classic studies on the mating behavior of the stickleback (1951, 1965), which he claimed could "*be hauled in numbers out of every ditch*" (Tinbergen 1952). This widespread abundance proved to be a great advantage for evolutionary biologists as well. The adaptive radiation of the threespine stickleback (*Gasterosteus aculeatus*) has produced a remarkable array of phenotypes and is an outstanding model system for studying the genetic basis of morphological evolution. Ancestral marine sticklebacks have colonized countless freshwater lakes and streams across the northern hemisphere, often in coastal areas that were exposed and flooded at the end of the last glacial maximum. Because sticklebacks have colonized so many different freshwater habitats, they have not only radiated into diverse environments, they have also evolved in parallel in similar environments, making them excellent models for the study of both adaptive radiation and parallel evolution (reviewed in Bell and Foster 1994).

In novel freshwater environments, sticklebacks encounter changes to diet and predation regimes. While marine fish frequently experience abundant predators and filter feed on small plankton, freshwater fish experience lower predation and consume macroscopic invertebrates. As a result, freshwater fish have repeatedly evolved losses in defensive armor (lateral plates and dorsal spines) as well as changes to craniofacial bones

associated with feeding (including the jaw and gill rakers, McPhail 1984, 1992; Schluter and McPhail 1992). These morphological changes have profound functional implications for the evolution of feeding and trophic diversity (McGee and Wainwright 2013; McGee et al. 2013).

In addition to the jaw, fish feed using the branchial skeleton, a set of serially homologous skeletal elements found in the throat. The branchial skeleton consists of three bilaterally paired pharyngeal tooth plates (two dorsal pairs and one ventral pair) as well as nine bilaterally paired endochondral bones (five dorsal pairs and four dorsal pairs). The pharyngeal skeleton is highly important for fish feeding. Pharyngeal teeth are a primary point of mastication, and the branchial bones in particular are important levers for the muscles that open and close the pharyngeal jaws (Wainwright 2006). Therefore, the stickleback branchial skeleton can be used the study the genetic and developmental basis of the evolution of multiple skeletal phenotypes.

Besides a fascinating evolutionary history and striking morphological diversity, sticklebacks are an outstanding model for studying the genetic and developmental basis of evolution for several reasons. First, marine and freshwater fish can be intercrossed and produce large numbers of offspring (~100-300 embryos per mating), which permits a forward genetics mapping approach for studying the genetic basis of evolution. Second, sticklebacks have a relatively compact genome (~450 Mb) with a high quality assembly that has been sequenced in 21 populations (Jones et al. 2012b). Recent work provided a high-resolution linkage map that further improved the genome assembly (Glazer et al. 2015). Third, stickleback embryos are large, fertilized externally, and develop relatively quickly (~10 days to hatching), allowing transgenesis via microinjection (Kingsley et al. 2004). Additionally, the embryos are transparent, allowing characterization of fluorescent transgene expression patterns for the analysis of regulatory elements. Combined, these features mean that regions of the genome controlling traits of interest can be genetically mapped and candidate genes can be functionally tested through functional manipuation.

The genetic and developmental basis of stickleback evolution

Extensive genetic mapping work has begun to identify some of the genetic changes involved in these evolved differences between sticklebacks, and a few classic and recent studies have successfully found *cis*-regulatory differences in candidate genes (Shapiro et al. 2004; Colosimo et al. 2005; Miller et al. 2007; Chan et al. 2010; O'Brown et al. 2015; Indjeian et al. 2016). However, many of the traits mapped to genes to date have been large effect, nearly Mendelian phenotypes in which a single locus controls most of the variance. The genetic basis of loci with smaller but tractable phenotypic effects has been less investigated due to the challenges of mapping more polygenic traits.

To begin to resolve the genetic basis of quantitative traits in sticklebacks, Miller et al. (2014) mapped quantitative trait loci (QTL) controlling dozens of skeletal traits,

including the number of pharyngeal teeth and length of branchial bones in a genetic mapping cross using a benthic male from Paxton Lake, British Columbia, and a female from a Japanese marine population. This study identified 8-16 QTL controlling each phenotype, and the majority of the identified QTL had freshwater alleles that increased branchial bone length and tooth number. However, at the time of the study, little was known about the phenotypes of the grandparental populations. Cleves et al. (2014) went on to show that tooth number is nearly two-fold higher in the Paxton Benthic (PAXB) population relative to marine fish, and that the difference is heritable in laboratory environments. Ellis et al (2015) found that two populations of stickleback increase the number of pharyngeal teeth due to quite different developmental mechanisms: PAXB adds more new teeth to the edge of the tooth plate whereas another population (Cerrito Creek from the San Francisco Bay Area) increases the rate of tooth replacement within the tooth plate. Furthermore, tooth number is controlled by completely different genetic loci in the two populations. Do other traits besides teeth follow this pattern of non-convergence? Chapter 2 of this thesis addresses whether other skeletal traits differ between marine and freshwater fish. It demonstrates that branchial bone length in six branchial long bones differs heritably between populations. Evolved gains in bone length in two stickleback populations are caused by similar increases in bone growth rate, and two shared regions of the genome are used by both populations to control these differences. Both populations use increased bone growth rate to evolve longer bones, and one population uses an additional increase in cartilage template size. Therefore, in the case of branchial bone length evolution, the developmental and genetic basis is remarkably, but not completely, parallel.

One outstanding question about parallel genetic evolution is whether the same changes occur repeatedly via independent mutations, or whether a single mutation has been selected repeatedly from standing variants (Barrett and Schluter 2008). Because marine and freshwater sticklebacks frequently come into contact when marine fish breed in freshwater streams, standing variants have been proposed to be a main driver of parallel evolution in sticklebacks (Schluter and Conte 2009; Bell and Aguirre 2013). In the case of lateral armor reduction, a standing variant allele of the gene *Eda* causes plate reduction in the vast majority of cases (Colosimo et al. 2005). Recent work in other systems has revealed a high degree of introgression as a major source of genetic variation underlying adaptation of wide variety of species, including humans, finches, and monkeyflowers (e.g. Huerta-Sánchez et al. 2014; Lamichhaney et al. 2015; Stankowski et al. 2015). In contrast, stickleback pelvic loss has been caused by recurrent mutations in the same enhancer of the hindlimb developmental regulatory *Pitx1* (Chan et al. 2010). The work in chapter three of this thesis shows that the same small region of stickleback chromosome 21 controls the evolution of branchial bone length in two independently-derived freshwater populations. Future work to identify the haplotypes associated with the pharyngeal skeletal traits studied here will help determine whether these phenotypes evolved independently or via selection on rare genotypes found within the marine population.

In a select few lakes in British Columbia, an extreme case of convergent evolution has occurred. In each lake, two stickleback ecomorphs are present: a deep-bodied form adaptive to live in the benthic (lake-bottom) habitat, as well as a long, slender morph adapted to the limnetic (open-water) environment (McPhail 1984, 1992, 1993). These pairs represent a particularly spectacular occurrence of convergent evolution, as each lake was independently colonized by two invasions of marine fish (Taylor and McPhail 1999, 2000), yet evolved highly similar species pairs. The descendants of the first invasion were displaced to the benthic environment by the second wave of colonization, which went on to adapt to the limnetic environment. Genomic analysis indicated that a few genomic regions show signs of natural selection across multiple lakes, but the connection between these regions and specific benthic-limnetic phenotypic differences is unknown (Jones et al. 2012a) Chapter four of this thesis tests for parallel genetic evolution in crosses of sticklebacks from three benthic populations. This work shows that QTL controlling similar traits in the three crosses overlap more often than expected by chance but are mostly unique across lakes. Several QTL controlling skeletal phenotypes are found in all three lakes.

One striking finding from the Miller et al. 2014 study of the genetics of stickleback skeletal differences was that the QTL controlling skeletal traits were significantly clustered on three chromosomes: 4, 20, and 21. Chapter four additionally shows that two clusters of skeletal QTL are found in multiple benthic populations of sticklebacks, suggesting that the clustering of QTL in these regions may be adaptive in the benthic environment. A key question arising from this work was whether each these QTL clusters represented single loci with pleiotropic effects, very tightly linked loci that affect multiple phenotypes ("supergenes"), or many spread out QTL whose intervals happen to overlap. Both cases have been observed in stickleback: *Eda* controls both lateral plates and neuromast number and positioning (Mills et al. 2014) but genetically separate regulatory regions of *Gdf6* are associated with lateral plate height and width (Indjeian et al. 2016). Chapter three addresses the clustering of two QTL on stickleback chromosome 21. Branchial bone length and pharyngeal tooth number both map to a small region of the chromosome near *Bmp6*. Cleves et al (2014) showed that the pharyngeal tooth QTL maps to a region including a regulatory allele of *Bmp6*, but does this region also control bone length? Using recombinant chromosomes, the bone length QTL is genetically separable from the tooth number QTL (*Bmp6*) in two populations of stickleback, but the regions controlling these two traits are genomically quite closely linked (likely within 300 kb). The region controlling bone length contains an excellent candidate gene, *Tfap2a*. Functional tests of the role of *Tfap2a* expression levels during stickleback craniofacial development show that a reduction of *Tfap2a* expression in freshwater fish may underlie evolved changes to bone length. These results support the *cis*-Regulatory Hypothesis (Carroll 2008) and demonstrate genetic dissection of a supergene cluster.

Finally, understanding the molecular basis of loci controlling evolved traits requires gaining a fine scale examination of the function and regulation of genes of interest. The largest effect QTL controlling stickleback tooth gain (part of the

chromosome 21 QTL cluster and found in all three benthic populations) maps to a regulatory allele of *Bmp6* in the PAXB population (Cleves et al. 2014). BMPs are known to have both activating and inhibiting roles in mammalian tooth development (Chen et al. 1996; Aberg et al. 1997; Bei and Maas 1998; Zhao et al. 2000). A *cis*-regulatory loss of *Bmp6* expression is associated with increased tooth number in Paxton Benthic sticklebacks, suggesting that *Bmp6* inhibits tooth development or replacement (Cleves et al. 2014). However, *Bmp6* has a highly complex spatiotemporal expression pattern during tooth development. In order to understand the evolved change to *Bmp6* regulation, it is critical to understand the genetic regulation of *Bmp6* expression. Chapter five identifies a key regulatory element of *Bmp6* that responds to TGFß signaling and is required for normal *Bmp6* expression. Although a second *Bmp6* enhancer has since been identified and likely contains mutations that have affected tooth evolution (Cleves et al., in prep) understanding the complex regulatory landscape of *Bmp6* will be necessary to fully understand its evolution.

Summary

The findings outlined in this dissertation suggest preliminary answers to a number of key questions about the genetic basis of evolution. Combined with previous studies, these results suggest that loci controlling different co-adaptive traits can be tightly clustered in the genome, perhaps to facilitate rapid adaptation to novel environments. However, in the case of tooth and bone gain, separate but closely linked loci underlie two cases of evolved skeletal gain. Further studies will test whether the haplotypes underlying the bone length and tooth number QTL are in linkage disequilibrium in marine fish; it is conceivable that inheriting these alleles together would be advantageous upon colonization of freshwater. Furthermore, regulatory changes to highly pleiotropic developmental signaling molecules (such as *Bmp6* and *Tfap2a*) seem to cause a substantial portion of skeletal evolution in the stickleback system. Interestingly, the *gain* of these two phenotypes is associated with a *loss* of gene expression. Evolutionarily, losing or reducing the function of an activator/enhancer of gene expression may be more likely than regulatory changes that gain new expression.

Stickleback evolution is often described as "microevolution" due to the relatively quick rate of evolution and the relatively subtle nature of the phenotypic differences. However, like most aspects of biology, evolution occurs on a spectrum. It is not difficult to imagine how, just as Darwin envisioned, the relatively small genetic differences controlling skeletal elements described here, when multiplied over millions of generations and perhaps more dramatic selective regimes, could produce dramatic phenotypic differences. Therefore, these studies into quantitative natural variation within a single species complex may offer valuable insight towards the genetic mechanisms that have produced the "*Tree of Life, which fills with its dead and broken branches the crust of the earth, and covers the surface with its ever branching and beautiful ramification*" (Darwin 1859).

References

Aberg, T., J. Wozney, and I. Thesleff. 1997. Expression patterns of bone morphogenetic proteins (Bmps) in the developing mouse tooth suggest roles in morphogenesis and cell differentiation. Dev. Dyn. Off. Publ. Am. Assoc. Anat. 210:383–396.

Albert, F. W., and L. Kruglyak. 2015. The role of regulatory variation in complex traits and disease. Nat. Rev. Genet. advance online publication.

Barrett, R. D. H., and D. Schluter. 2008. Adaptation from standing genetic variation. Trends Ecol. Evol. 23:38–44.

Bates, H. W. 1862. Contributions to an Insect Fauna of the Amazon Valley. Lepidoptera: Heliconidæ. Trans. Linn. Soc. Lond. 23:495–566.

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

Bell, M. A., and W. E. Aguirre. 2013. Contemporary evolution, allelic recycling, and adaptive radiation of the threespine stickleback. Evol. Ecol. Res. 15:377–411.

Bell, M. A., and S. A. Foster. 1994. The Evolutionary Biology of the Threespine Stickleback. Oxford University Press.

Berner, D., and W. Salzburger. 2015. The genomics of organismal diversification illuminated by adaptive radiations. Trends Genet. 31:491–499.

Carroll, S. B. 2008. Evo-Devo and an Expanding Evolutionary Synthesis: A Genetic Theory of Morphological Evolution. Cell 134:25–36.

Chan, Y. F., M. E. Marks, F. C. Jones, G. Villarreal, M. D. Shapiro, S. D. Brady, A. M. Southwick, D. M. Absher, J. Grimwood, J. Schmutz, R. M. Myers, D. Petrov, B. Jónsson, D. Schluter, M. A. Bell, and D. M. Kingsley. 2010. Adaptive evolution of pelvic reduction in sticklebacks by recurrent deletion of a Pitx1 enhancer. Science 327:302–305.

Charlesworth, B., and D. Charlesworth. 2009. Darwin and Genetics. Genetics 183:757–766.

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

Cleves, P. A., N. A. Ellis, M. T. Jimenez, S. M. Nunez, D. Schluter, D. M. Kingsley, and C. T. Miller. 2014. Evolved tooth gain in sticklebacks is associated with a cis-regulatory allele of Bmp6. Proc. Natl. Acad. Sci. 111:13912–13917.

Colosimo, P. F., K. E. Hosemann, S. Balabhadra, G. Villarreal, M. Dickson, J. Grimwood, J. Schmutz, R. M. Myers, D. Schluter, and D. M. Kingsley. 2005. Widespread Parallel

Evolution in Sticklebacks by Repeated Fixation of Ectodysplasin Alleles. Science 307:1928–1933.

Darwin, C. R. 1859. The Origin of Species. J. Murray, London.

Dettman, J. R., N. Rodrigue, A. H. Melnyk, A. Wong, S. F. Bailey, and R. Kassen. 2012. Evolutionary insight from whole-genome sequencing of experimentally evolved microbes. Mol. Ecol. 21:2058–2077.

Ellis, N. A., A. M. Glazer, N. N. Donde, P. A. Cleves, R. M. Agoglia, and C. T. Miller. 2015. Distinct developmental and genetic mechanisms underlie convergently evolved tooth gain in sticklebacks. Development 142:2442–2451.

Elmer, K. R., S. Fan, H. Kusche, M. Luise Spreitzer, A. F. Kautt, P. Franchini, and A. Meyer. 2014. Parallel evolution of Nicaraguan crater lake cichlid fishes via non-parallel routes. Nat. Commun. 5:5168.

Fisher, R. A. 1930. The Genetical Theory of Natural Selection. Oxford University Press, Oxford.

Fisher, S., E. A. Grice, R. M. Vinton, S. L. Bessling, and A. S. McCallion. 2006. Conservation of RET regulatory function from human to zebrafish without sequence similarity. Science 312:276–279.

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 g3.115.017905.

Herron, M. D., and M. Doebeli. 2013. Parallel Evolutionary Dynamics of Adaptive Diversification in Escherichia coli. PLOS Biol 11:e1001490.

Hirschhorn, J. N., and M. J. Daly. 2005. Genome-wide association studies for common diseases and complex traits. Nat. Rev. Genet. 6:95–108.

Hoekstra, H. E. 2006. Genetics, development and evolution of adaptive pigmentation in vertebrates. Heredity 97:222–234.

Huerta-Sánchez, E., X. Jin, Asan, Z. Bianba, B. M. Peter, N. Vinckenbosch, Y. Liang, X. Yi, M. He, M. Somel, P. Ni, B. Wang, X. Ou, Huasang, J. Luosang, Z. X. P. Cuo, K. Li, G. Gao, Y. Yin, W. Wang, X. Zhang, X. Xu, H. Yang, Y. Li, J. Wang, J. Wang, and R. Nielsen. 2014. Altitude adaptation in Tibetans caused by introgression of Denisovan-like DNA. Nature 512:194–197.

Huse, H. K., T. Kwon, J. E. A. Zlosnik, D. P. Speert, E. M. Marcotte, and M. Whiteley. 2010. Parallel Evolution in Pseudomonas aeruginosa over 39,000 Generations In Vivo. mBio 1:e00199-10.

Huxley, J. 1942. Evolution, the Modern Synthesis. George Allen Unwin, London.

Indjeian, V. B., G. A. Kingman, F. C. Jones, C. A. Guenther, J. Grimwood, J. Schmutz, R. M. Myers, and D. M. Kingsley. 2016. Evolving New Skeletal Traits by cis-Regulatory Changes in Bone Morphogenetic Proteins. Cell 164:45–56.

Jones, F. C., Y. F. Chan, J. Schmutz, J. Grimwood, S. D. Brady, A. M. Southwick, D. M. Absher, R. M. Myers, T. E. Reimchen, B. E. Deagle, D. Schluter, and D. M. Kingsley. 2012a. A Genome-wide SNP Genotyping Array Reveals Patterns of Global and Repeated Species-Pair Divergence in Sticklebacks. Curr. Biol. 22:83–90.

Jones, F. C., M. G. Grabherr, Y. F. Chan, P. Russell, E. Mauceli, J. Johnson, R. Swofford, M. Pirun, M. C. Zody, S. White, E. Birney, S. Searle, J. Schmutz, J. Grimwood, M. C. Dickson, R. M. Myers, C. T. Miller, B. R. Summers, A. K. Knecht, S. D. Brady, H. Zhang, A. A. Pollen, T. Howes, C. Amemiya, Broad Institute Genome Sequencing Platform & Whole Genome Assembly Team, E. S. Lander, F. Di Palma, K. Lindblad-Toh, and D. M. Kingsley. 2012b. The genomic basis of adaptive evolution in threespine sticklebacks. Nature 484:55–61.

Joron, M., L. Frezal, R. T. Jones, N. L. Chamberlain, S. F. Lee, C. R. Haag, A. Whibley, M. Becuwe, S. W. Baxter, L. Ferguson, P. A. Wilkinson, C. Salazar, C. Davidson, R. Clark, M. A. Quail, H. Beasley, R. Glithero, C. Lloyd, S. Sims, M. C. Jones, J. Rogers, C. D. Jiggins, and R. H. ffrench-Constant. 2011. Chromosomal rearrangements maintain a polymorphic supergene controlling butterfly mimicry. Nature 477:203–206.

Kimura, M. 1983. The Neutral Theory of Molecular Evolution. Cambridge University Press, Cambridge.

King, M. C., and A. C. Wilson. 1975. Evolution at two levels in humans and chimpanzees. Science 188:107–116.

Kingsley, D. M., B. Zhu, K. Osoegawa, P. J. D. Jong, J. Schein, M. Marra, C. Peichel, C. Amemiya, D. Schluter, S. Balabhadra, B. Friedlander, Y. M. Cha, M. Dickson, J. Grimwood, J. Schmutz, W. S. Talbot, and R. Myers. 2004. New Genomic Tools for Molecular Studies of Evolutionary Change in Threespine Sticklebacks. Behaviour 141:1331–1344.

Kingsley, E. P., K. M. Kozak, S. Pfeifer, D.-S. Yang, and H. E. Hoekstra. 2016. Multiple genetic changes underlie the evolution of long-tailed forest deer mice. bioRxiv 41699.

Kunte, K., W. Zhang, A. Tenger-Trolander, D. H. Palmer, A. Martin, R. D. Reed, S. P. Mullen, and M. R. Kronforst. 2014. doublesex is a mimicry supergene. Nature 507:229–232.

Küpper, C., M. Stocks, J. E. Risse, N. dos Remedios, L. L. Farrell, S. B. McRae, T. C. Morgan, N. Karlionova, P. Pinchuk, Y. I. Verkuil, A. S. Kitaysky, J. C. Wingfield, T. Piersma, K. Zeng, J. Slate, M. Blaxter, D. B. Lank, and T. Burke. 2016. A supergene determines highly divergent male reproductive morphs in the ruff. Nat. Genet. 48:79–83.

Lamichhaney, S., J. Berglund, M. S. Almén, K. Maqbool, M. Grabherr, A. Martinez-Barrio, M. Promerová, C.-J. Rubin, C. Wang, N. Zamani, B. R. Grant, P. R. Grant, M. T. Webster, and L. Andersson. 2015. Evolution of Darwin's finches and their beaks revealed by genome sequencing. Nature 518:371–375.

Lamichhaney, S., G. Fan, F. Widemo, U. Gunnarsson, D. S. Thalmann, M. P. Hoeppner, S. Kerje, U. Gustafson, C. Shi, H. Zhang, W. Chen, X. Liang, L. Huang, J. Wang, E. Liang, Q. Wu, S. M.-Y. Lee, X. Xu, J. Höglund, X. Liu, and L. Andersson. 2016. Structural genomic changes underlie alternative reproductive strategies in the ruff (Philomachus pugnax). Nat. Genet. 48:84–88.

Lander, E. S., and D. Botstein. 1989. Mapping mendelian factors underlying quantitative traits using RFLP linkage maps. Genetics 121:185–199.

Linnen, C. R., Y.-P. Poh, B. K. Peterson, R. D. H. Barrett, J. G. Larson, J. D. Jensen, and H. E. Hoekstra. 2013. Adaptive Evolution of Multiple Traits Through Multiple Mutations at a Single Gene. Science 339:1312–1316.

Losos, J. B. 2009. Lizards in an Evolutionary Tree: Ecology and Adaptive Radiation of Anoles. University of California Press.

Mathelier, A., W. Shi, and W. W. Wasserman. 2015. Identification of altered cis-regulatory elements in human disease. Trends Genet. 31:67–76.

Mather, K. 1955. Polymorphism as an Outcome of Disruptive Selection. Evolution 9:52–61.

Mather, K. 1950. The Genetical Architecture of Heterostyly in Primula sinensis. Evolution 4:340–352.

Mayr, E. 2001. What Evolution is. Basic Books.

McGee, M. D., D. Schluter, and P. C. Wainwright. 2013. Functional basis of ecological divergence in sympatric stickleback. BMC Evol. Biol. 13:277.

McGee, M. D., and P. C. Wainwright. 2013. Convergent Evolution as a Generator of Phenotypic Diversity in Threespine Stickleback. Evolution 67:1204–1208.

McGlothlin, J. W., J. P. Chuckalovcak, D. E. Janes, S. V. Edwards, C. R. Feldman, E. D. Brodie, M. E. Pfrender, and E. D. Brodie. 2014. Parallel Evolution of Tetrodotoxin Resistance in Three Voltage-Gated Sodium Channel Genes in the Garter Snake Thamnophis sirtalis. Mol. Biol. Evol. msu237.

McPhail, J. D. 1992. Ecology and evolution of sympatric sticklebacks (Gasterosteus): evidence for a species-pair in Paxton Lake, Texada Island, British Columbia. Can. J. Zool. 70:361–369.

McPhail, J. D. 1984. Ecology and evolution of sympatric sticklebacks (Gasterosteus): morphological and genetic evidence for a species pair in Enos Lake, British Columbia. Can. J. Zool. 62:1402–1408.

McPhail, J. D. 1993. Ecology and evolution of sympatric sticklebacks (Gasterosteus): origin of the species pairs. Can. J. Zool. 71:515–523.

Mendel, G. J. 1865. Versuche über Pflanzen-Hybriden. Verhandlungen Naturforschenden Vereines Brünn. IV:3–47.

Miller, C. T., S. Beleza, A. A. Pollen, D. Schluter, R. A. Kittles, M. D. Shriver, and D. M. Kingsley. 2007. cis-Regulatory Changes in Kit Ligand Expression and Parallel Evolution of Pigmentation in Sticklebacks and Humans. Cell 131:1179–1189.

Mills, M. G., A. K. Greenwood, and C. L. Peichel. 2014. Pleiotropic effects of a single gene on skeletal development and sensory system patterning in sticklebacks. EvoDevo 5:5.

Mundy, N. I. 2005. A window on the genetics of evolution: MC1R and plumage colouration in birds. Proc. R. Soc. B Biol. Sci. 272:1633–1640.

Murray, J., and B. Clarke. 1976a. Supergenes in polymorphic land snails. I. Partula taeniata. Heredity 37:253–269.

Murray, J., and B. Clarke. 1976b. Supergenes in polymorphic land snails. II. Partula suturalis. Heredity 37:271–282.

Muschick, M., A. Indermaur, and W. Salzburger. 2012. Convergent evolution within an adaptive radiation of cichlid fishes. Curr. Biol. CB 22:2362–2368.

Nabours, R. K. 1933. Inheritance of Color Patterns in the Grouse Locust ACRYDIUM ARENOSUM Burmeister (Tettigidae). Genetics 18:159–171.

Nabours, R. K. 1929. The Genetics of the Tettigidae (Grouse Locusts). Springer.

O'Brown, N. M., B. R. Summers, F. C. Jones, S. D. Brady, and D. M. Kingsley. 2015. A recurrent regulatory change underlying altered expression and Wnt response of the stickleback armor plates gene EDA. eLife e05290.

Orr, H. A. 2005. The genetic theory of adaptation: a brief history. Nat. Rev. Genet. 6:119–127.

Orr, H. A. 2002. The Population Genetics of Adaptation: The Adaptation of Dna Sequences. Evolution 56:1317–1330.

Orr, H. A. 1998. The Population Genetics of Adaptation: The Distribution of Factors Fixed during Adaptive Evolution. Evolution 52:935–949.

Pai, A. A., and Y. Gilad. 2014. Comparative studies of gene regulatory mechanisms. Curr. Opin. Genet. Dev. 29:68–74.

Pennacchio, L. A., W. Bickmore, A. Dean, M. A. Nobrega, and G. Bejerano. 2013. Enhancers: five essential questions. Nat. Rev. Genet. 14:288–295.

Pires-daSilva, A., and R. J. Sommer. 2003. The evolution of signalling pathways in animal development. Nat. Rev. Genet. 4:39–49.

Prud'homme, B., N. Gompel, and S. B. Carroll. 2007. Emerging principles of regulatory evolution. Proc. Natl. Acad. Sci. 104:8605–8612.

Reed, R. D., R. Papa, A. Martin, H. M. Hines, B. A. Counterman, C. Pardo-Diaz, C. D. Jiggins, N. L. Chamberlain, M. R. Kronforst, R. Chen, G. Halder, H. F. Nijhout, and W. O. McMillan. 2011. optix drives the repeated convergent evolution of butterfly wing pattern mimicry. Science 333:1137–1141.

Roberts, R. B., J. R. Ser, and T. D. Kocher. 2009. Sexual Conflict Resolved by Invasion of a Novel Sex Determiner in Lake Malawi Cichlid Fishes. Science 326:998–1001.

Rosenblum, E. B., H. Römpler, T. Schöneberg, and H. E. Hoekstra. 2010. Molecular and functional basis of phenotypic convergence in white lizards at White Sands. Proc. Natl. Acad. Sci. 107:2113–2117.

Sanger, T. J., L. J. Revell, J. J. Gibson-Brown, and J. B. Losos. 2012. Repeated modification of early limb morphogenesis programmes underlies the convergence of relative limb length in Anolis lizards. Proc. R. Soc. B Biol. Sci. 279:739–748.

Schluter, D. 2000. The Ecology of Adaptive Radiation. Oxford University Press.

Schluter, D., and G. L. Conte. 2009. Genetics and ecological speciation. Proc. Natl. Acad. Sci. 106:9955–9962.

Schluter, D., and J. D. McPhail. 1992. Ecological Character Displacement and Speciation in Sticklebacks. Am. Nat. 140:85–108.

Schwander, T., R. Libbrecht, and L. Keller. 2014. Supergenes and Complex Phenotypes. Curr. Biol. 24:R288–R294.

Shapiro, M. D., M. E. Marks, C. L. Peichel, B. K. Blackman, K. S. Nereng, B. Jónsson, D. Schluter, and D. M. Kingsley. 2004. Genetic and developmental basis of evolutionary pelvic reduction in threespine sticklebacks. Nature 428:717–723.

Stankowski, S., J. M. Sobel, and M. A. Streisfeld. 2015. The geography of divergence with gene flow facilitates multitrait adaptation and the evolution of pollinator isolation in Mimulus aurantiacus. Evolution 69:3054–3068.

Steiner, C. C., H. Römpler, L. M. Boettger, T. Schöneberg, and H. E. Hoekstra. 2009. The Genetic Basis of Phenotypic Convergence in Beach Mice: Similar Pigment Patterns but Different Genes. Mol. Biol. Evol. 26:35–45.

Steiner, C. C., J. N. Weber, and H. E. Hoekstra. 2007. Adaptive variation in beach mice produced by two interacting pigmentation genes. PLoS Biol. 5:e219.

Stern, D. L. 2000. Perspective: Evolutionary Developmental Biology and the Problem of Variation. Evolution 54:1079–1091.

Stern, D. L. 2013. The genetic causes of convergent evolution. Nat. Rev. Genet. 14:751–764.

Taylor, E. B., and J. D. McPhail. 1999. Evolutionary history of an adaptive radiation in species pairs of threespine sticklebacks (Gasterosteus): insights from mitochondrial DNA. Biol. J. Linn. Soc. 66:271–291.

Taylor, E. B., and J. D. McPhail. 2000. Historical contingency and ecological determinism interact to prime speciation in sticklebacks, Gasterosteus. Proc. R. Soc. Lond. B Biol. Sci. 267:2375–2384.

Thomas, J. W., M. Cáceres, J. J. Lowman, C. B. Morehouse, M. E. Short, E. L. Baldwin, D. L. Maney, and C. L. Martin. 2008. The Chromosomal Polymorphism Linked to Variation in Social Behavior in the White-Throated Sparrow (Zonotrichia albicollis) Is a Complex Rearrangement and Suppressor of Recombination. Genetics 179:1455–1468.

Thompson, M. J., and C. D. Jiggins. 2014. Supergenes and their role in evolution. Heredity 113:1–8.

Tinbergen, N. 1965. Social behaviour in animals: with special reference to vertebrates. Chapman and Hall.

Tinbergen, N. 1952. The curious behavior of the stickleback. Sci. Am. 187:22–26.

Tinbergen, N. 1951. The Study of Instinct. Clarendon Press.

Tuttle, E. M., A. O. Bergland, M. L. Korody, M. S. Brewer, D. J. Newhouse, P. Minx, M. Stager, A. Betuel, Z. A. Cheviron, W. C. Warren, R. A. Gonser, and C. N. Balakrishnan. 2016. Divergence and Functional Degradation of a Sex Chromosome-like Supergene. Curr. Biol. 0.

Vickrey, A. I., E. T. Domyan, M. P. Horvath, and M. D. Shapiro. 2015. Convergent evolution of head crests in two domesticated columbids is associated with different missense mutations in EphB2. Mol. Biol. Evol. msv140.

Villar, D., C. Berthelot, S. Aldridge, T. F. Rayner, M. Lukk, M. Pignatelli, T. J. Park, R. Deaville, J. T. Erichsen, A. J. Jasinska, J. M. A. Turner, M. F. Bertelsen, E. P. Murchison, P.

Flicek, and D. T. Odom. 2015. Enhancer Evolution across 20 Mammalian Species. Cell 160:554–566.

Wainwright, P. C. 2006. Functional Morphology of the Pharyngeal Jaw Apparatus. P. *in* Fish Physiology: Fish Biomechanics. Academic Press.

Ward, L. D., and M. Kellis. 2012. Interpreting noncoding genetic variation in complex traits and human disease. Nat. Biotechnol. 30:1095–1106.

Weller, J. I. 1986. Maximum Likelihood Techniques for the Mapping and Analysis of Quantitative Trait Loci with the Aid of Genetic Markers. Biometrics 42:627–640.

Wilson, E. O. 1999. The Diversity of Life. W. W. Norton & Company.

Wray, G. A. 2007. The evolutionary significance of cis-regulatory mutations. Nat. Rev. Genet. 8:206–216.

Zhao, X., Z. Zhang, Y. Song, X. Zhang, Y. Zhang, Y. Hu, S. H. Fromm, and Y. Chen. 2000. Transgenically ectopic expression of Bmp4 to the Msx1 mutant dental mesenchyme restores downstream gene expression but represses Shh and Bmp2 in the enamel knot of wild type tooth germ. Mech. Dev. 99:29–38.

Zhen, Y., M. L. Aardema, E. M. Medina, M. Schumer, and P. Andolfatto. 2012. Parallel Molecular Evolution in an Herbivore Community. Science 337:1634–1637.

This chapter appeared in *Proceedings of the Royal Society of London B: Biological Sciences* 281 (1788), 2014. All authors have agreed to its use in this dissertation.

2. Two developmentally temporal QTL underlie convergent evolution of increased branchial bone length in sticklebacks

Priscilla A. Erickson[1], Andrew M. Glazer[1], Phillip A. Cleves[1], Alyson S. Smith[1] and Craig T. Miller[1*]

[1] Department of Molecular and Cell Biology, University of California, Berkeley, CA, 94720

Abstract

In convergent evolution, similar phenotypes evolve repeatedly in independent populations, often reflecting adaptation to similar environments. Understanding whether convergent evolution proceeds via similar or different genetic and developmental mechanisms offers insight towards the repeatability and predictability of evolution. Oceanic populations of threespine stickleback fish, *Gasterosteus aculeatus*, have repeatedly colonized countless freshwater lakes and streams, where new diets lead to morphological adaptations related to feeding. Here, we show that heritable increases in branchial bone length have convergently evolved in two independently derived freshwater stickleback populations. In both populations, an increased bone growth rate in juveniles underlies the convergent adult phenotype, and one population also has a longer cartilage template. Using F2 crosses from these two freshwater populations, we show that two quantitative trait loci (QTL) control branchial bone length at distinct points in development. In both populations, a QTL on chromosome 21 controls bone length throughout juvenile development, and a QTL on chromosome 4 controls bone length only in adults. In addition to these similar developmental profiles, these QTL show similar chromosomal locations in both populations. Our results suggest that sticklebacks have convergently evolved longer branchial bones using similar genetic and developmental programs in two independently derived populations.

Introduction

Independent populations that converge on similar evolved phenotypes may do so by using similar genetic and developmental mechanisms, suggesting that evolution is, at times, constrained and predictable [1,2]. When convergent phenotypic evolution is caused by parallel genetic mechanisms, the parallelism may occur on a number of different hierarchical levels, including changes in the same nucleotide, gene, genetic pathway, or genomic region (reviewed in [2-4]; for examples see [5-8]). While numerous cases of convergent evolution have been documented across natural and experimental populations of animals, plants and microbes, fewer studies have investigated whether these convergently evolved phenotypes arise in the same way during development [9]. Furthermore, most studies of convergent evolution have focused on traits with a simple genetic architecture, and less is known about whether more complex traits, which are more common in nature, convergently evolve via parallel developmental genetic features.

In vertebrates, the skeleton contributes to organismal form and function, and evolved changes in skeletal elements occur repeatedly as populations adapt to new environments. The skeleton forms largely from two types of bone: endochondral, which develops from a cartilage template, and dermal, which ossifies directly without a cartilage intermediate [10]. Atchley and Hall [11] proposed that skeletal evolution may proceed through a number of cellular mechanisms (e.g. the size of the cartilage template or the rate of bone growth). In support of this proposal, a dramatic difference in jaw size between quails and ducks results from several differences in the specification and morphogenesis of the midbrain and midbrain neural crest cells from which the jaw is derived [12]. In *Anolis* lizards, however, a smaller number of cellular mechanisms appear to underlie convergent skeletal evolution. On at least four islands of the Caribbean, *Anolis* limb morphologies have repeatedly evolved in different ecological habitats [13]. While multiple pathways in pre- and post-embryonic development could contribute to differing limb length, increased adult limb size in four different long-limbed species arises from an increase in the size of the embryonic limb template, followed by growth rates equal to those in shorter limbed species [14]. Whether this mechanism of evolved bone length differences is seen in other convergently evolved skeletal changes is largely unknown.

The threespine stickleback (*Gasterosteus aculeatus*) provides an excellent model system for studying both the developmental and genetic basis of convergent skeletal evolution. Ancestral marine sticklebacks have colonized thousands of freshwater environments throughout the Northern Hemisphere and have evolved numerous adaptations to these new freshwater environments [15]. For example, freshwater sticklebacks have repeatedly evolved changes to their head skeletons to improve feeding efficiency on new foods in freshwater environments, including convergent decreases in gill raker number [16-18], as well as increases in jaw width [17,19] and suction feeding index [20].

Here we hypothesized that other trophic skeletal elements may also differ between marine and freshwater sticklebacks. The branchial skeleton (see Figure 2.1, adapted from [21]) is primarily made up of bilateral, segmentally-reiterated bones: 5 ventral pairs (ceratobranchials, CB1-CB5) and 4 dorsal pairs (epibranchials EB1-EB4). In fish, these

bones arise from neural crest cells in the pharyngeal arches during development, and the dorsal and ventral bones are segmental homologs of the upper and lower jaw, respectively [22,23]. These long dorsal and ventral bones of the branchial skeleton are endochondral and resemble mammalian long bones (e.g. the femur) in appearance [24]. In fish, the branchial cartilages start to form late in embryogenesis, just before hatching [23,25]. As development continues, this cartilage is mostly replaced with bone deposited by osteoblasts that originate both outside and within the cartilage template [26]. The bones then elongate as the fish grows larger [24]. Thus, two key developmental processes contribute to the length of the bone: the establishment of the cartilage template early in embryonic development and the rate of subsequent bone growth.

A previous genome-wide linkage mapping study of the genetic basis of skeletal variation in sticklebacks identified 14 quantitative trait loci (QTL) with significant effects on the length of branchial bones in a marine-by-freshwater F2 cross, including two QTL on chromosomes 4 and 21 with large effects [27]. Combined, these two QTL explain ~27% of the variance in length of the dorsal EB1 and ~25% of variance in length of the ventral bones. Most (11) of these QTL, including both of the large effect QTL, had effects in the same direction, with freshwater alleles conferring longer bones [27]. However, this study did not measure the bone length phenotypes of the parental populations. Orr [28] proposed that a concerted sign of QTL effect indicates a trait is under natural selection, as similar directions of effect would unlikely be observed by chance. Here we test the hypothesis that the two previously identified large effect QTL on chromosomes 4 and 21 are used in a second independently derived freshwater population. By studying the developmental trajectories of evolved increases in bone length, as well as the developmental timing of two bone length QTL, we also test whether similar developmental and genetic effects contribute to these evolved increases in bone length.

Methods

Wild collections

Wild anadromous marine fish were collected from the Little Campbell River (LITC) in British Columbia under a fish collection permit from the British Columbia Ministry of Environment (permit #SU08-44549). Wild freshwater fish were collected from Fishtrap Creek (FTC) in Washington under fish scientific collection from the Washington Department of Fish and Wildlife (permit #08-284). All animal work was approved by the Institutional Animal Care and Use Committees of the University of California-Berkeley or Stanford University (protocol numbers R330 and 13834). Wild sticklebacks were collected in the summer of 2008. All wild and lab-reared fish were euthanized with 0.08% Tricaine and stored in 100% ethanol until staining and dissection.

Fish husbandry and crosses:

For the FTC x LITC cross, a wild male FTC fish was crossed to a wild female LITC. For the Paxton Benthic (PAXB, British Columbia, Canada) x LITC cross, a lab-reared male offspring of wild PAXB fish was crossed to two wild LITC females. Adult F1 fish were then intercrossed to their siblings to create F2 families, which were grown to ages of 20, 40,

and 80 days post fertilization (dpf), or adults (see Table 2.1). All fish were raised in 3ppt salinity (~10% seawater) at 18 degrees C in 110 L (29 gallon) tanks. Fish were fed a diet of live *Artemia* as young fry, live *Artemia* and frozen *Daphnia* as juveniles, and frozen bloodworms and *Mysis* shrimp as adults.

	Time-point	total fish	standard length (mm)	number of F2 families	individuals per family
FTC x LITC	20 dpf	94	8.61 ± 0.57	3	43, 39, 12
	40 dpf	96	15.02 ± 1.24	2	54, 42
	80 dpf	96	24.32 ± 1.73	3	49, 27, 20
	adult	279	37.33 ± 2.95	4	96, 78, 54, 51
PAXB x LITC	20 dpf	96	7.96 ± 0.64	1	96
	40 dpf	141	15.08 ± 2.22	2	88, 53
	80 dpf	108	23.57 ± 2.61	3	42, 38, 28
	adult	270	40.99 ± 5.40	8	59, 46, 37, 34, 30, 25, 23, 16

Table 2.1: Summary of fish included in QTL analysis.

F2 families were raised to indicated ages from each cross. Mean ± standard deviation of standard length is given for each time point and the size of each F2 family listed.

Branchial bone phenotyping in juveniles (>20dpf) and adults:

Standard length (tip of snout to caudal peduncle) and/or total length (tip of snout to most posterior tip of tail) were measured with digital calipers following ethanol fixation. Pectoral and caudal fins were stored in ethanol for DNA extraction and genotyping. All fish were fixed in 4% paraformaldehyde in 1X PBS or 10% neutral buffered formalin for 1-6 days. For staining, fish were rinsed with water, stained in 0.008% Alizarin red in 1% KOH for 24-72 hours, rinsed again with water, washed in 1% KOH for 24 hours, and finally stored in 50% glycerol, 0.5% KOH until dissection. Branchial dissections were performed as previously described [1]. Bones were imaged with a DFC420 camera on a Leica M165FC dissecting microscope. Left side bone length measurements were taken by drawing a straight line in ImageJ software using previously defined landmarks: the anterior medial tip to the anterior distal tip of the Alizarin-positive bone [1]. Due to their complex three-dimensional shapes, EB2-4 were not measured. Bone lengths of wild fish were regressed to standard length within each population and residuals back-transformed to a 60 mm fish. Wild LITC bone lengths were also corrected for the sex of the fish (see table 2.4; FTC bones were not sexually dimorphic in the wild). Residual bone lengths of lab-reared fish were back-transformed to a fish of length 45 mm and corrected for sex when appropriate (see Table 2.4). All lab-reared fish were 5-7 months of age for the adult time points.

Cartilage and bone phenotyping in fry (≤ 20 dpf):

For the 20 dpf QTL time points, the posterior half of the fish was removed for DNA isolation, and fish heads were stained with both Alizarin red and Alcian blue as previously described [2]. For the cartilage early time point comparisons (10 dpf and 14 dpf), cartilage was stained using 0.1% Alcian blue in 70% ethanol and 0.37% hydrochloric acid. CB4 cartilage measurements were imaged and measured as described above. For measurement of EB1 at the 14 dpf timepoint, the first branchial arch (CB1+EB1) was removed and mounted separately from the rest of the branchial skeleton. Flattened cartilages were imaged on a Leica DM2500 compound microscope under differential interference contrast optics and photographed with a Leica DFC500 camera. Left and right cartilage measurements were averaged. Cartilage lengths in Figure 2.6(*a*) were regressed to standard length and residuals back-transformed to a 5 mm fry.

Marker genotyping:

Genomic DNA was isolated from stickleback fins using phenol-chloroform extraction and ethanol precipitation. Primers for genetic mapping were previously described [1,3] or were designed flanking microsatellites or indels in the stickleback genome using Primer3 (http://biotools.umassmed.edu/bioapps/primer3_www.cgi) and the Gramene SSR finder (http://www.gramene.org/db/markers/ssrtool). Some primers were directly labeled with a fluorophore; others were labeled with an M13 sequence tag, and a three-primer PCR [4] was performed using a fluorescently labeled M13 primer (see table 2.2 for all primers used). Other markers were analyzed on agarose gels (unlabeled indel primers) or using fragment analysis (fluorescently labeled and M13-labeled primers) on an ABI3730xl capillary sequencer. All PCRs were performed in 10 uL reactions using 10 ng of genomic DNA template and the following reaction conditions: 10mM Tris (pH 8.5), 50mM KCl, 1.5mM $MgCl_2$, 0.1% Triton-X100, and 200 uM of each dNTP. Cycling conditions were 1 cycle of 3 minutes at 94°; 35 cycles of 15s at 94°, 15s at 56°, and 15s at 72°; and a final incubation of 5m at 72°. For M13-labeled, 3-primer reactions, the conditions were: 1 cycle of 30s at 94°; 30 cycles of 30s at 94°, 15s at 56°, 15s at 72°; 7 cycles of 30s at 94°, 15s at 53°, 15s at 72°; final incubation of 10 minutes at 72°. Sex was determined as previously described [5] or by using primers 5'CATATTGCTGCTTGTGTGGAAG and 5'GATCCTCCTCGTTCCTACAG which are closely linked to the stickleback sex determining region [5]. These primers amplify a 186 bp band from the Y chromosome and a 229 bp band from the X chromosome.

QTL Analysis:

All F2 bone lengths within an age group (20 dpf, 40 dpf, 80 dpf, and adult) were tested for association with standard length (SL), sex and F2 family using a linear model in R. Standard length was always significant; see Table 2.3 for significance of sex and family in each group. Bone lengths were then corrected by taking residuals from a linear model containing SL and sex and/or family when appropriate ($p<0.05$ in initial model). For QTL mapping, these residuals were z-scored within age group to permit the combination of residuals from different age groups.

Linkage maps for chromosomes 4 and 21 were generated using JoinMap 4 software [6]. Chromosome 4 and 21 QTL were initially mapped with the *scanone* function in R/qtl using Haley-Knott regression. Subsequently, the *fitqtl* and *refineqtl* functions were used to account for the effect of one chromosome while calculating the LOD scores and percent variance explained (PVE) for the other. Dominance (d/a) was calculated using the formula d = difference between the heterozygous phenotype and the midpoint between the homozygous phenotypes; a = half the phenotypic difference between the homozygous genotypes [7]. Significance thresholds of $\propto$ = 0.05, 0.01, and 0.001 were calculated with 10,000 permutations of each phenotype on the two linkage groups tested. The markers Stn489 (chromosome 21 peak marker in the FTC x LITC cross) and Stn382 (chromosome 4 peak marker in the PAXB x LITC cross) were used to test for the effect of each QTL at each developmental timepoint by ANOVA analysis in Figures 4 and 5 as they were the peak markers of a previous genetic study of bone length [1].

Statistical analysis:

All statistical analyses were performed using the R statistical software package (http://www.r-project.org/). QTL analysis was performed using R/qtl (www.rqtl.org).

primer name	forward sequence	reverse sequence	chr.	physical position (Mb)	FTC x LITC position (cM)	PAXB x LITC position (cM)	label type
Stn38	GCAGGTGACATCTTCAGGG	TTTCATTAGGACCCAGGACG	4	3	0	0	M13
Stn40	CCAGAATCTCTCTCTCATTTGG	TGCCCTTATCCAGTGAAACG	4	4.46	NA	12.59	M13
Stn42	ACACGCAGCTTGACTGTTCC	GCGTATACGTTACACGCCG	4	6.11	29.19	NA	M13
Gac4147	CCGCGATGATGAGAGTG	GTGAAATGCGACAGATGATG	4	11.58	NA	26.83	M13
Stn382	CCCTTAGAGAATTTCCTAGCA	CTTGTCCCGGATCATACGC	4	12.8	32.55	28.97	(indel)
Stn46	GCTCGGAGCTCATCTACACC	CAGCCGTTTCAAGACAACC	4	15.23	34.93	NA	M13
Stn266	CTATGCGTCTAGCTTTCGGG	CGAGAAGATTCTCCACACGG	4	15.62	35.83	NA	M13
Stn253	AACCACCCAGACCACTAAGC	ATGTCACGTATAGGTCGGCG	4	21.35	40.49	40.1	M13
chr4_280	CGTCCAGTACGTCCTAATCCC	AGGTCCGTGGTGAGCTAATG	4	30.5	52.09	54.41	M13; DL
chr21_0049	CTGGACGGCTCTTTTTGGTA	CTGGACGGCTCTTTTTGGTA	21	0.49	0	0	(indel)
chr21_0163	AGACCAGCATCCAAAGAGGA	CTTGTAAGCCAAGTGGGAGC	21	1.63	NA	1.46	M13
Stn487	CACGGCAAACAGGTGAGAC	TCGATGGGCTGTAAATCCTC	21	2.56	1.6	2.17	M13; DL
Stn423	CCTCCAGGACGAATTCAAAG	CTGCATCTCGGCTGTGTGG	21	3.77	2.19	2.52	M13
Stn489	AGTGACGAATCCCTCTTCTGC	CACACCTTGTTGTGTTTGTAGC	21	4.23	2.7	2.73	DL
Stn491	AACGTTAACCAGTTGCAGTCC	GATGTCGACACAGAATCTCTTAGC	21	5.12	3.5	3.69	M13
Stn424	AGGTCTCGGTTCGATTACCA	TTGTGCGCTTGCATATGCAT	21	NA	7.55	6.94	M13; DL
Stn425	AGTCGATTCAACCGACCCAACAC	ACACAGTCCAAACCGTCTCT	21	9.04	11.75	10.42	M13
chr21_0991	ATCGTGCAGAGGTCAGTGC	TGTGGTGACACAACCTCCAT	21	9.91	19.37	NA	M13
chr21_0997	CCGGGTTGTTATTGACTGCT	AAACTCTGCCTCACAGCGAt	21	9.97	NA	19.57	M13
chr21_1136	ACAGTGGTCGGTGAAGGTC	GAGCGCATTTGATTTGTGAA	21	11.36	NA	44.34	M13
chr21_1170	AGTCAAGCGTTGCCATTTCT	CAAAGTCAAGTCCTCTCGGC	21	11.7	39.63	NA	M13

Table 2.2: Markers and genetic map used in QTL analysis.

Markers were genotyped using standard methods and the linkage map was calculated using JoinMap. Label types: M13=5' M13 sequence tag for 3-primer amplification; DL=directly labeled with fluorophore (NED, PET, VIC, or FAM); indel=large insertion/deletion scored on agarose gel. Physical position refers to the stickleback genome assembly gas/Acu1 (Jones *et al* 2012).

	CB1	CB2	CB3	CB4	CB5	EB1
FTC x LITC						
20 dpf	fam	NA	NA	fam	fam	sex
40 dpf	fam	fam	fam	fam	sex, fam	fam
80 dpf	NA	NA	NA	NA	NA	NA
adult	sex, fam	sex, fam	sex, fam	sex, fam	fam	sex, fam
PAXB x LITC						
20 dpf	sex	NA	NA	NA	NA	NA
40 dpf	fam	fam	fam	fam	fam	fam
80 dpf	fam	fam	fam	fam	fam	sex, fam
adult	sex, fam	fam	sex, fam	sex, fam	fam	sex, fam

Table 2.3: Corrections applied in QTL mapping.

Effects of sex and F2 family (fam) were tested in a linear model with standard length as a covariate for each timepoint in each cross. Factors with significant effects in the combined model ($p<0.05$) were corrected in the final model used to produce residuals for mapping. All bone lengths were significantly associated with standard length. Phenotypes used in final QTL analysis were corrected for listed variables by taking residuals from a linear regression that included all significant variables.

Results

Population differences in bone length

To test the hypothesis that wild marine and freshwater fish differ in branchial bone length, we analyzed wild-caught marine (Little Campbell, British Columbia, Canada [LITC]) and freshwater (Fish Trap Creek, Washington [FTC]) sticklebacks for differences in length of the dorsal (epibranchial, EB1) and ventral (ceratobranchial, CB1-5) branchial long bones (Figure 2.1). All six branchial long bones differed significantly in length (Figure 2.3A), with freshwater fish having longer bones relative to standard length than marine fish (ventral bones were 8.8-17.1% longer; dorsal bone was 23.8% longer in 60 mm fish). Because a strong genetic component of bone length was previously observed in a large F2 cross [27], we next hypothesized that these differences in bone length were heritable in multiple freshwater populations. We tested these hypotheses by raising adult marine and freshwater fish under common lab-reared conditions. Supporting our hypotheses, fish from both FTC and a second freshwater population, Paxton Benthic (PAXB, British Columbia, Canada) raised in the laboratory had significantly longer branchial bones than marine LITC fish (Figure 2.3B, Figure 2.2). FTC bones were consistently slightly longer than PAXB bones, and the increase in FTC over LITC bone length ranged from 7.6% to 19.7% for ventral bones and was 33.6% for EB1.

In wild fish, LITC bones were sexually dimorphic, with males having longer bones than females, but FTC bone lengths did not differ significantly between sexes (Table 2.4). In lab-reared fish, LITC branchial bones were all sexually dimorphic; some ventral FTC bones (CB1, 2, and 4) were sexually dimorphic, while no PAXB bones were significantly sexually dimorphic. This observation matches previous findings for sticklebacks: marine fish are sexually dimorphic for body shape and feeding kinematic phenotypes, while freshwater fish have lost this sexual dimorphism, with both sexes having an overall phenotype more similar to marine males [29,30].

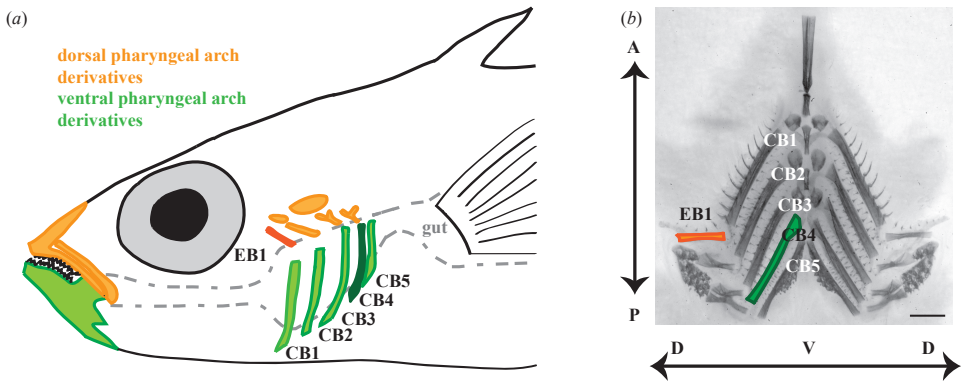


Figure 2.1. Anatomy of stickleback branchial bones.

(*a*) Fish ingest food into the buccal cavity, which is flanked bilaterally by dorsal (epibranchial [EB], orange) and ventral (ceratobranchial [CB], green) pharyngeal arch bones between the mouth and the gut. Also shown are the upper and lower oral jaw, segmental homologs of the branchial bones in the first pharyngeal arch. In some experiments we focused on the highlighted bones: EB1 and CB4, the dorsal and ventral bone with the strongest detected genetic effects. Some bones have been omitted for clarity. *(b)* Dissected and flat-mounted Alizarin red stained branchial skeleton with dorsal (EB1) and ventral (CB1-5) bones indicated. Dorsal-ventral (DV) and anterior-posterior (AP) axes are labeled with arrows. Scale bar = 1 mm.

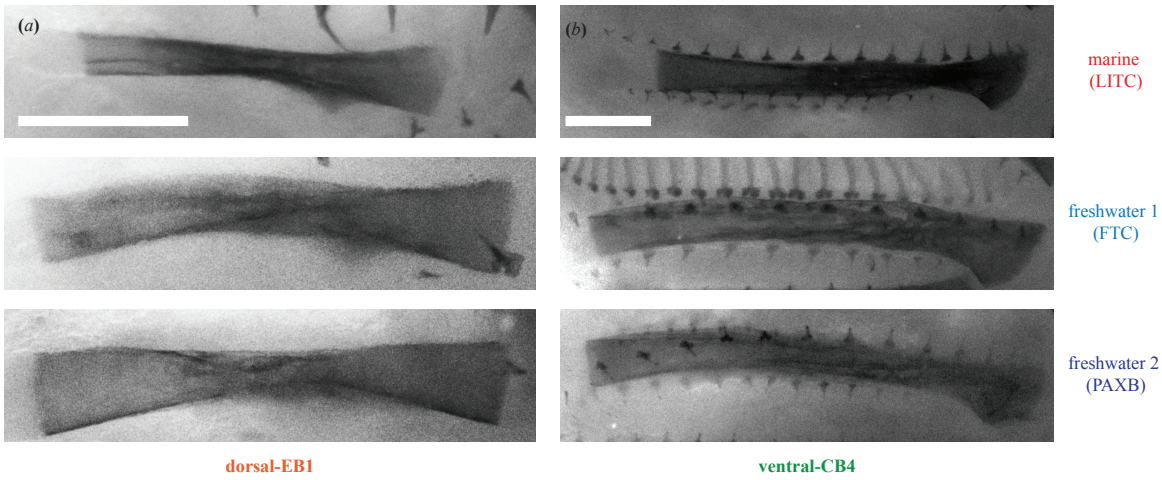


Figure 2.2. Comparison of marine and freshwater branchial bones.

Shown are dissected and flat-mounted Alizarin red stained (*a*) epibranchial 1 (EB1) and (*b*) ceratobranchial 4 (CB4) bones of fish from a marine population (LITC) and two freshwater populations (FTC and PAXB). All fish are lab-reared, 5.3-6.3 months old, and 41.7 to 41.8 mm standard length. Scale bar = 500 µm.

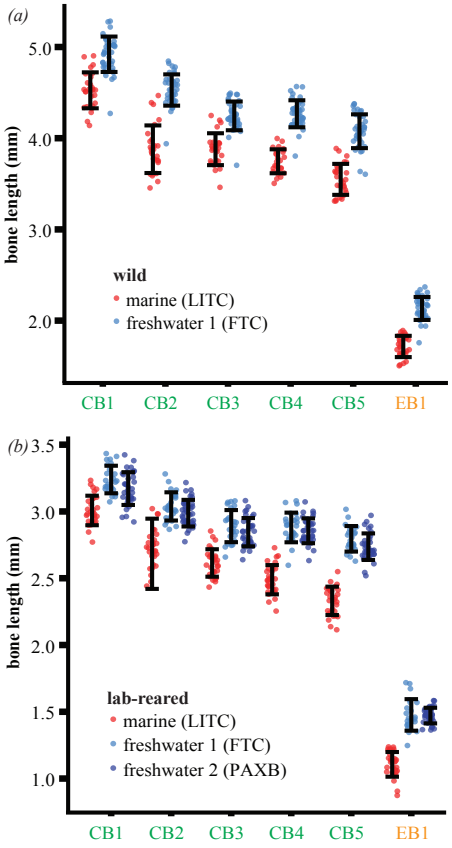


Figure 2.3 Heritable increases in branchial bone lengths in freshwater sticklebacks.

(a) Wild Fishtrap Creek (FTC) fish have significantly longer dorsal (EB1) and ventral (CB1-5) branchial bones relative to size-matched wild marine Little Campbell (LITC) fish. All six bones are significantly longer in FTC fish than LITC fish ($p < 10^{-10}$ for all bones, n_{LITC}=27, n_{FTC}=40, Welch's T-test). *(b)* Increased bone lengths are heritable in adult lab-reared fish, and longer bones are also found in a second lab-reared freshwater population, Paxton benthic (PAXB). All bones are significantly longer in each freshwater population relative to LITC marine (Tukey HSD: $p < 10^{-5}$, n_{LITC}=32, n_{FTC}=25, n_{PAXB}=36). PAXB and FTC bone lengths do not significantly differ ($p > 0.05$). Error bars = standard deviation of the mean. Red=LITC, light blue=FTC, dark blue=PAXB.

Developmental basis of bone length differences

We hypothesized that stickleback bone length differences, like evolved *Anolis* limb length [14], would manifest during development as transposition of the y-intercept, but not slope, of a regression of bone length against standard length. We collected lab-reared fish from each population at regular developmental time points, resulting in fish varying in total length from 10 to 40 mm. We looked for differences in bone growth rate and initial bone size using an ANCOVA with standard length as the covariate and population

as an interacting factor. Contrary to our prediction, we observed significant differences in the slopes (population*standard length interaction term) of dorsal and posterior ventral bone lengths relative to standard length between marine and freshwater fish, suggesting that freshwater bones grow more rapidly relative to body size (Figure 2.4 (*a*) and (*b*), Figure 2.5, Table 2.5). Thus, unlike in *Anolis* lizards, the convergent evolution of increased bone length in two derived freshwater stickleback populations appears to utilize a similar faster bone growth rate in both populations.

The significant differences in y-intercepts in the bone development time courses (Table 2.5) led us to hypothesize that the cartilage templates that prefigure branchial bones may be larger in freshwater fish relative to marine fish. For ventral cartilages, we focused on CB4 because it had a large marine-freshwater difference and had strong genetic effects in a previous cross [27]. We raised FTC and LITC fry to stage 26 (~10 dpf) [31] to measure the CB4 cartilage and stage 28 (~13-14 dpf) to measure EB1 cartilage. We found that both cartilage templates were longer in FTC relative to both LITC and PAXB (Figure 2.4 (*c*) and (*d*)). Thus, despite the convergent increased bone growth rates, one unique developmental difference contributes to the convergent evolution, with one freshwater population (FTC) but not a second (PAXB) evolving a longer cartilage template early in development.

	female bone length (mm)	male bone length (mm)	d.f	t	*p*
Wild LITC	*n=16*	*n=11*			
CB1	4.47 ± 0.29	4.90 ± 0.15	23.50	-4.88	**<0.0001**
CB2	3.85 ± 0.28	4.12 ± 0.29	21.28	-2.37	**0.0273**
CB3	3.84 ± 0.19	4.16 ± 0.22	19.92	-3.84	**0.0010**
CB4	3.71 ± 0.19	4.03 ± 0.15	24.54	-4.96	**<0.0001**
CB5	3.51 ± 0.22	3.81 ± 0.17	24.42	-3.90	**0.0007**
EB1	1.70 ± 0.13	1.85 ± 0.13	21.90	-3.03	**0.0061**
Wild FTC	*n=28*	*n=12*			
CB1	4.91 ± 0.20	4.93 ± 0.18	22.99	-0.31	0.7579
CB2	4.52 ± 0.18	4.54± 0.15	24.78	-0.43	0.6690
CB3	4.24 ± 0.16	4.26 ± 0.15	21.94	-0.41	0.6882
CB4	4.25 ± 0.15	4.29 ± 0.15	20.48	-0.61	0.5500
CB5	4.07 ± 0.20	4.08 ± 0.16	22.61	-0.16	0.8747
EB1	2.13 ± 0.14	2.15 ± 0.08	34.14	-0.67	0.5062
Lab-reared LITC	*n=18*	*n=12*			
CB1	2.99 ± 0.25	3.19 ± 0.31	26.96	-4.29	**0.0002**
CB2	2.68 ± 0.43	2.93 ± 0.27	20.65	-2.96	**0.0075**
CB3	2.61 ± 0.21	2.74 ± 0.27	24.89	-3.20	**0.0037**
CB4	2.49 ± 0.22	2.65 ± 0.22	24.00	-3.94	**0.0006**
CB5	2.32 ± 0.19	2.49 ± 0.23	24.57	-4.00	**0.0005**
EB1	1.10 ± 0.13	1.19 ± 0.14	26.18	-2.59	**0.0156**
Lab-reared FTC	*n=17*	*n=8*			
CB1	3.25 ± 0.37	3.36 ± 0.28	11.68	-2.23	**0.0457**
CB2	3.04 ± 0.31	3.14 ± 0.25	15.40	-2.20	**0.0433**
CB3	2.87 ± 0.30	2.94 ± 0.23	10.64	-1.28	0.2259
CB4	2.88 ± 0.24	2.99 ± 0.23	13.04	-2.30	**0.0388**
CB5	2.77 ± 0.30	2.85 ± 0.29	8.93	-1.84	0.0989
EB1	1.44 ± 0.17	1.52 ± 0.20	10.10	-1.39	0.1921
Lab Reared PAXB	*n=21*	*n=14*			
CB1	3.14 ± 0.36	3.21 ± 0.17	32.87	-1.93	0.0621
CB2	2.96 ± 0.31	3.02 ± 0.18	27.79	-1.60	0.1196
CB3	2.82 ± 0.31	2.88 ± 0.17	28.31	-1.61	0.1186
CB4	2.84 ± 0.30	2.87 ± 0.16	29.81	-1.23	0.2277
CB5	2.72 ± 0.32	2.76 ± 0.17	22.55	-1.31	0.2024
EB1	1.47 ± 0.18	1.47 ± 0.10	28.21	-0.21	0.8363

Table 2.4. Sexual dimorphism of branchial bone lengths.

Bone lengths of lab-reared and wild fish were regressed to standard length and back-transformed to a 45 or 60 mm fish, respectively. Mean length ± standard deviation in mm is shown for males and females. Statistics shown are for a Welch's two sample T-test. Significant differences (p<0.05) are highlighted in bold; these bones were corrected for sex in Figure 2.3.

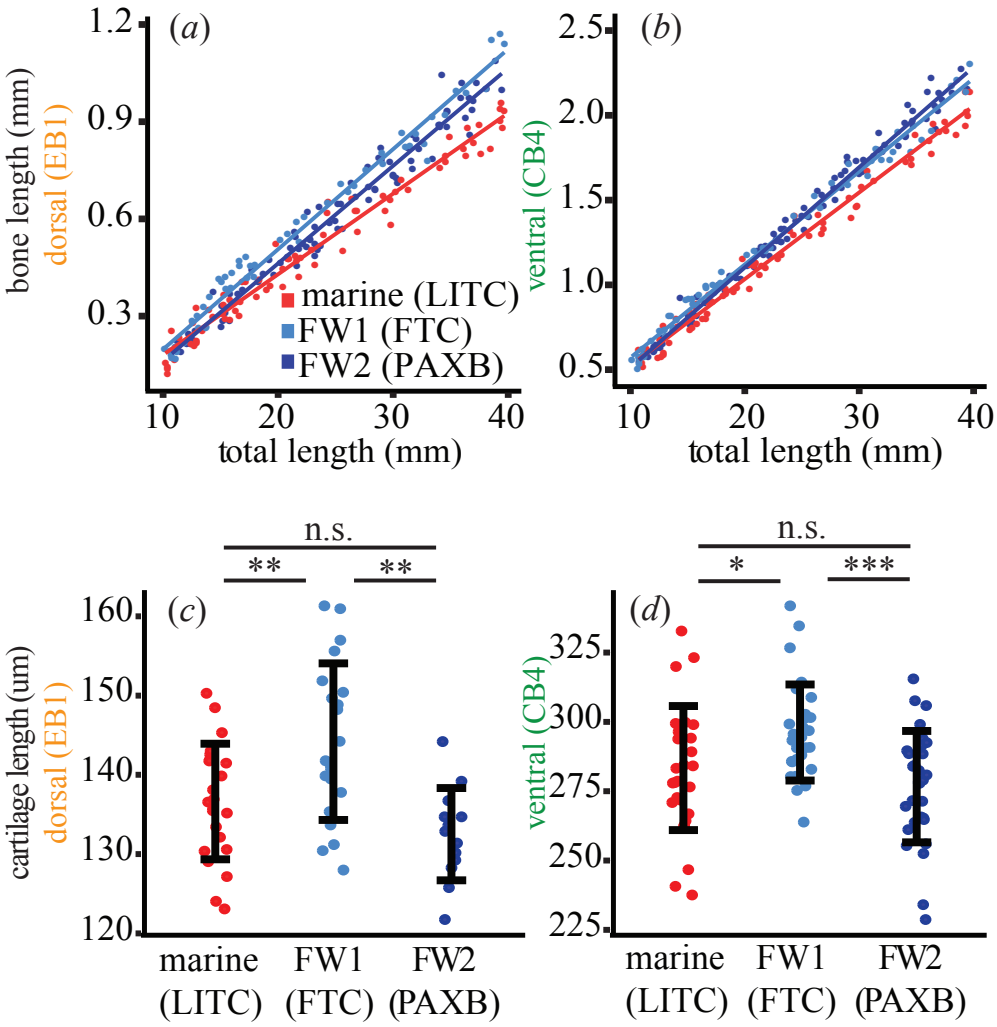


Figure 2.4. Developmental basis of dorsal and ventral bone length differences.

Developmental time course of dorsal (EB1, panel A) and ventral (CB4, panel B) bone lengths plotted against total length of lab-reared fish under 5 months of age. Both bones show statistically significant differences in slope (bone growth rate) as well as y-intercept between marine and freshwater. ANCOVA statistics are shown in Table 2.5; additional bones are shown in Figure 2.5. Red=LITC, light blue=FTC, dark blue=PAXB. EB1 (*c*) and CB4 (*d*) cartilages are longer in FTC relative to LITC and PAXB fry (Tukey HSD $p < 0.05$ for LITC-FTC and PAXB-FTC comparisons of both cartilages). In (*c*) the FTC fish were slightly shorter in total length than the LITC and PAXB fish (Tukey HSD test $p < 0.05$), which makes the cartilage size increase even greater relative to body size. Error bars=standard deviation. Asterisks indicate Tukey HSD *p* values: n.s=not significant; * = $p < 0.05$, ** $p < 0.01$, ***$p < 0.001$.

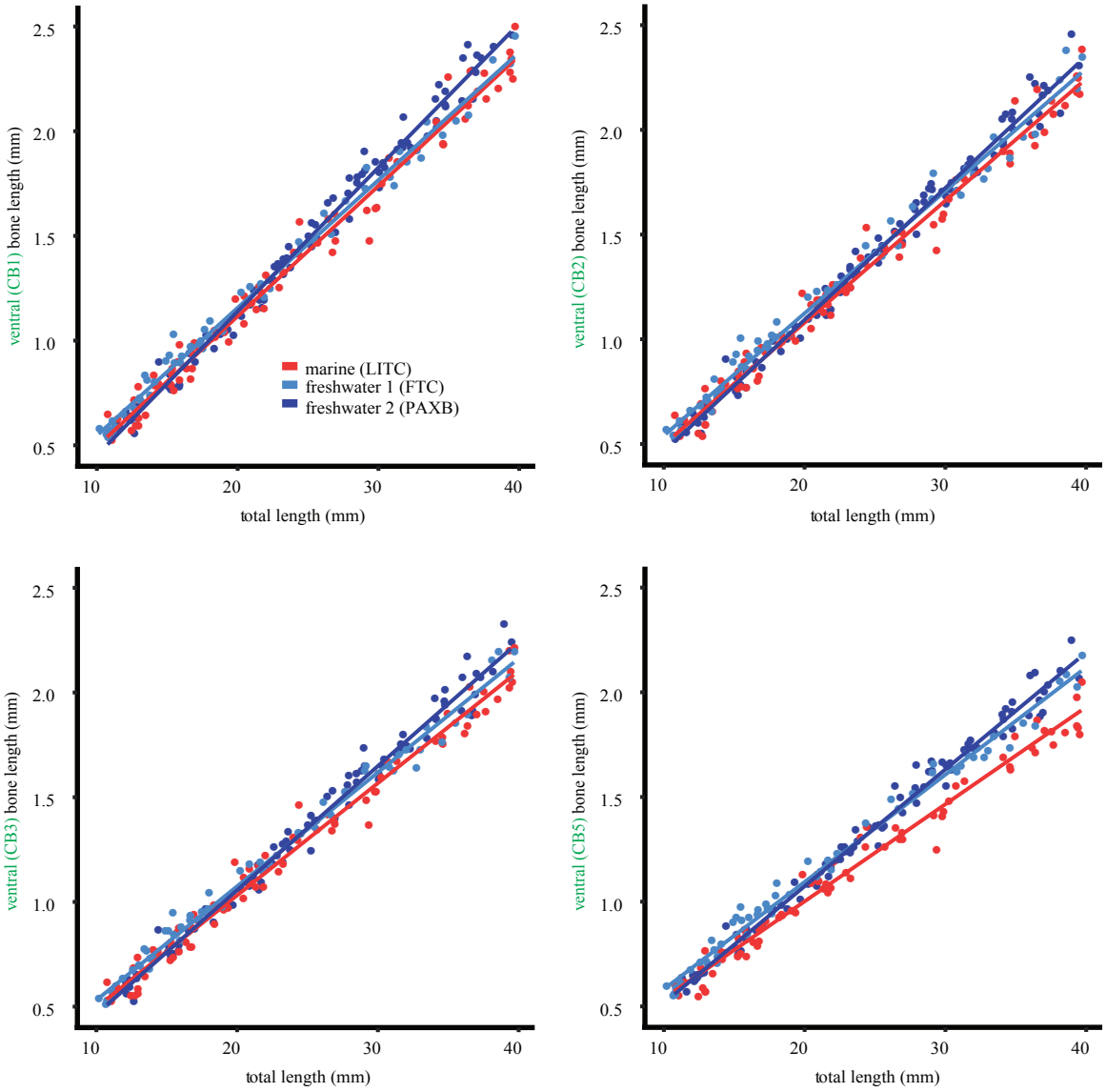


Figure 2.5. Developmental time courses of other ventral branchial bones.

Ventral ceratobranchial (CB, see Fig. 1) bone lengths of a developmental time course of fish plotted against fish total length. ANCOVA statistics are shown in Table 2.5.

	slopes			intercepts		
	d.f.	*F*	*p*	d.f.	*F*	*p*
LITC vs FTC						
CB1	1, 134	0.25	0.617	1, 135	14.11	**0.0003**
CB2	1, 132	0.01	0.918	1, 133	15.76	**0.0001**
CB3	1, 133	0.09	0.765	1, 134	21.39	**< 0.0001**
CB4	1, 132	8.38	**0.004**	1, 133	86.58	**< 0.0001**
CB5	1, 132	12.29	**0.001**	1, 133	106.4	**< 0.0001**
EB1	1, 127	54.2	**< 0.0001**	1, 128	142.6	**< 0.0001**
LITC vs PAXB						
CB1	1, 167	28.41	**< 0.0001**	1, 168	22.65	**< 0.0001**
CB2	1, 165	14.07	**0.0002**	1, 166	10.4	**0.0011**
CB3	1, 163	22.73	**< 0.0001**	1, 164	25.04	**< 0.0001**
CB4	1, 166	52.98	**< 0.0001**	1, 167	127	**< 0.0001**
CB5	1, 167	69.86	**< 0.0001**	1, 168	156.1	**< 0.0001**
EB1	1, 163	48.19	**< 0.0001**	1, 164	64.6	**< 0.0001**

Table 2.5: Developmental differences in marine and freshwater bone growth rates.
Results of ANCOVA for bone length with standard length as covariate. Bones grew faster in freshwater fish (significant difference in slope, $p < 0.05$, for the ANCOVA standard length*population interaction term) except Fishtrap Creek (FTC) anterior ventral bones (CB1-3).

Genetic basis of bone length differences

QTL mapping provides a powerful first test of possible parallel genetic mechanisms underlying convergent evolution. We hypothesized that previously identified bone length QTL might be reused in multiple freshwater stickleback populations due to extensive sharing of the genetic basis of evolved traits in stickleback populations [32,33] and the similar increased bone growth rates in FTC and PAXB. Because there are likely multiple developmental mechanisms that can be altered to change bone length [11], we further predicted that these QTL might exert different effects at specific points in development.

We focused on the two largest effect QTL controlling adult bone length in a previous cross (chromosomes 4 and 21, see [27]) and observed strikingly similar developmental profiles of these QTL in our two crosses. We raised F2 fish to four time points (20, 40, and 80 days post fertilization (dpf), and adults, see Table 2.1) and tested for the effect of chromosomes 4 and 21 at each time point using the peak markers from the previous cross [27]. We found that freshwater alleles of chromosome 21 increased dorsal bone length at all stages in the FTC x LITC cross and at all stages except 20 dpf in the

PAXB x LITC cross (Figure 2.6; Table 2.6; ANOVA $p < 0.05$ for marker Stn489 on chromosome 21), suggesting that the freshwater allele of this QTL acts relatively early in development to increase bone size. The 20 dpf time point in the PAXB x LITC cross strongly trends in the same direction ($p = 0.06$). The dorsal effect of chromosome 21 was strongest at 80 dpf in both crosses. Chromosome 21 controlled ventral (CB4) bone length at the 80 dpf timepoint in the FTC x LITC cross and was nearly significant ($p = 0.056$) in the PAXB x LITC cross at 80 dpf. These results suggest that the freshwater allele may specifically increase both dorsal and ventral bone growth rates during juvenile stages. In contrast, chromosome 4 exhibited effects on dorsal and ventral bones only in adults of the two crosses (Figure 2.7, Table 2.6, ANOVA $p < 0.05$), suggesting that this QTL acts later in bone development. In both crosses, the effect of the chromosome 4 QTL was greatest on CB4 (Figure 2.8).

Finally, to test whether the QTL overlap in the two crosses, we genotyped markers across each chromosome and tested for association with bone length, focusing on the ventral effect of chromosome 4 and dorsal effect of chromosome 21. In support of a parallel genetic basis of these QTL, we saw similar localization of the two QTL in the two crosses (Figure 2.9). The chromosome 4 peak marker from the PAXB x LITC cross (Stn382) was identical to the previously identified peak marker [27] and was within 3 cM (6 Mb) of the peak marker of the FTC x LITC cross (Stn42). The peak chromosome 21 marker from the PAXB x LITC cross (Stn491) was only 0.9 cM (0.9 Mb) away from the peak marker of the FTC x LITC cross (Stn489), which was the peak marker in the previous study. Furthermore, the dorsal chromosome 21 and ventral chromosome 4 QTL are additive in both cross (dominances between -0.15 and 0.21, Table 2.7). Combined with the QTL developmental profiles, these localization and dominance data suggest that FTC and PAXB share several parallel genetic features for evolved bone length gain, including overlapping QTL on chromosomes 4 and 21.

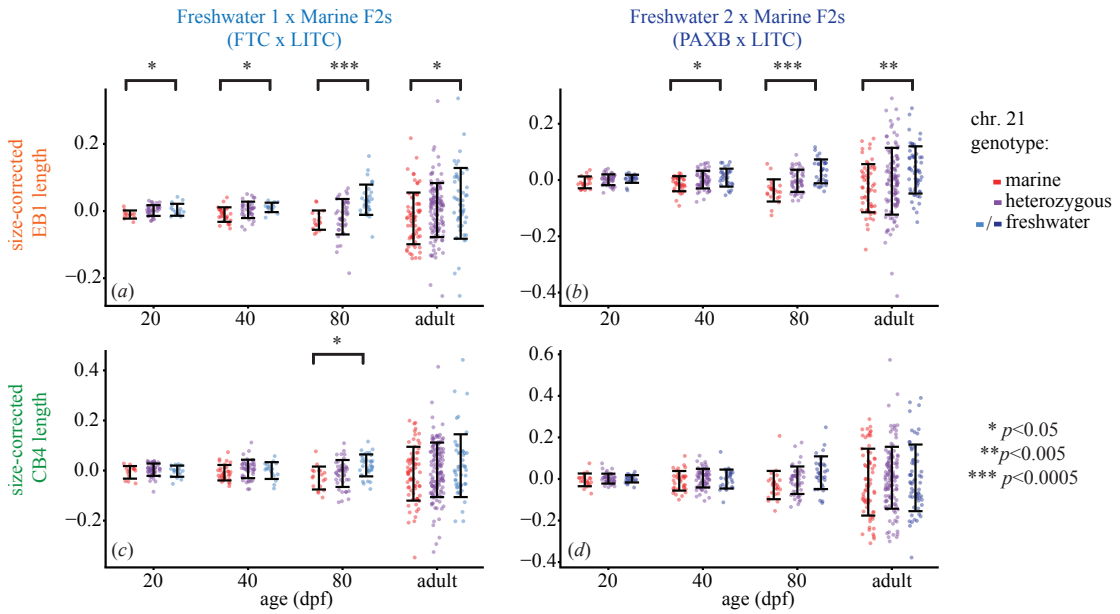


Figure 2.6. Similar developmental effects of chromosome 21 QTL in two independently derived freshwater populations.

F2 fish from two marine x freshwater F2 crosses [(*a*) and (*c*), FTC x LITC; (*b*) and (*d*), PAXB x LITC] were raised to four time points (given in days post-fertilization, dpf; or adults >150dpf) and tested for effects of chromosome 21 genotype on size-corrected bone length residuals. See Table 2.1 for a summary of the fish included in each time point. Chromosome 21 controlled dorsal [EB1, (*a*) and (*b*)] bones at all time points in the FTC cross and at all time points except 20 dpf in the PAXB cross, and the effect was strongest at 80 dpf in both crosses. The effect of chromosome 21 on ventral bone length (CB4, (*c*) and (*d*)) was significant at 80dpf in the FTC cross and nearly significant at 80 dpf in the PAXB cross. ANOVA *p*-values for the marker Stn489 are indicated with asterisks when significant ($p < 0.05$, see Table 2.6 for complete listing of ANOVA results for all branchial bones). Error bars = standard deviation.

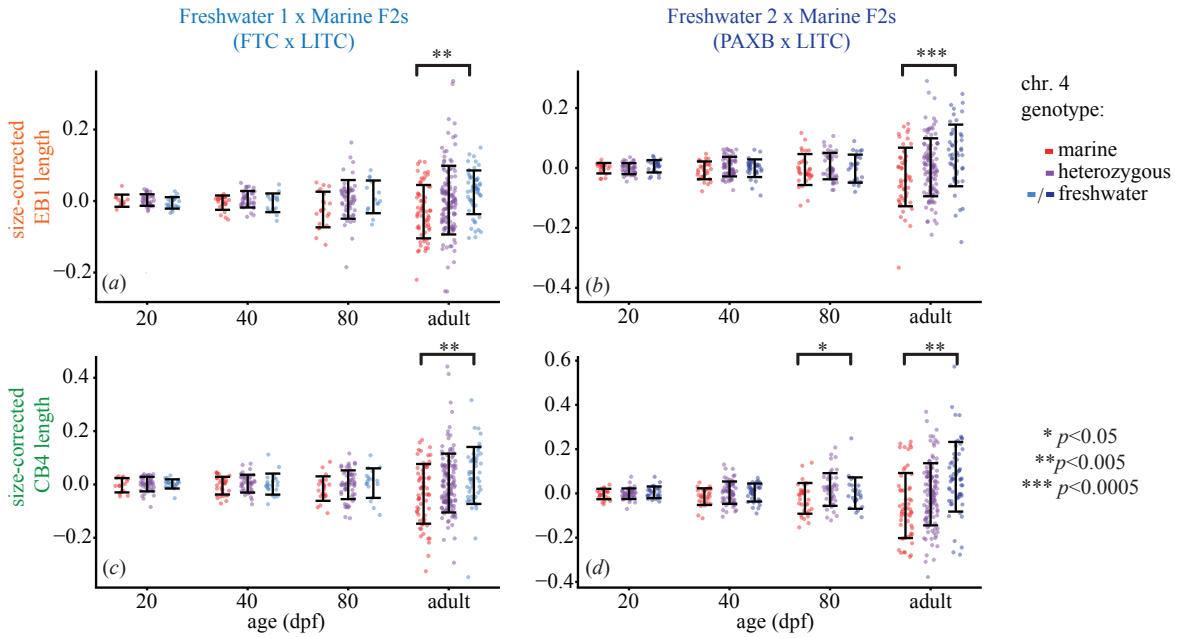


Figure 2.7. Similar developmental effects of chromosome 4 QTL in two independently derived freshwater populations.

F2 fish from two crosses [(*a*) and (*c*), FTC x LITC; (*b*) and (*d*), PAXB x LITC] were raised to four time points (given in dpf for sub-adults) and tested for effects of chromosome 4 genotype on size-corrected bone length residuals. In both crosses, a chromosome 4 QTL was only significant in adults. ANOVA *p*-values for the marker Stn382 are indicated with asterisks when significant ($p < 0.05$, see Table 2.6 for complete listing of ANOVA results). Error bars = standard deviation.

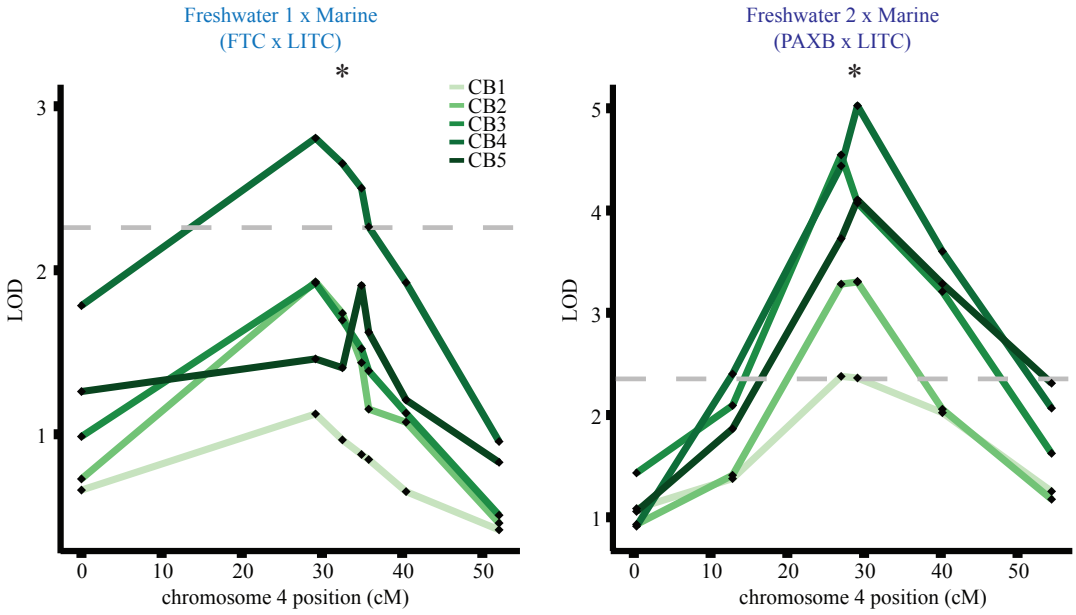


Figure 2.8. Effects of chromosome 4 on ventral bones in two crosses.

The chromosome 4 QTL was mapped using size-corrected bone lengths from adults in each cross (adjusted for sex and F2 family when appropriate). The significance threshold of $\propto$ = 0.05 (calculated based on 1000 phenotype permutations) is shown as a dashed grey line. Asterisks indicate the location of marker Stn382. Markers (black diamonds) from left to right: (*a*) Stn38, Stn42, Stn382, Stn46, Stn266, Stn253, chr4_280; (*b*) Stn38, Stn40, Gac4147, Stn382, Stn253, chr4_280

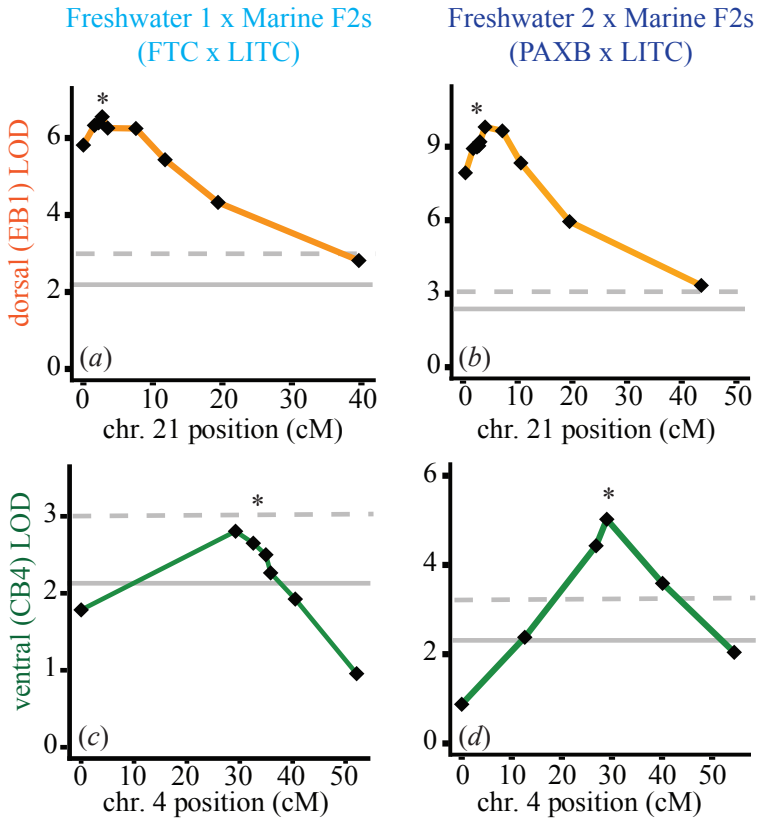


Figure 2.9: Similar localization of chromosome 21 dorsal bone QTL and chromosome 4 ventral bone QTL in two crosses.

The chromosome 21 QTL was mapped in fish from the 40 dpf, 80 dpf, and adult time points in each cross, while the chromosome 4 QTL was mapped in adults only. The significance threshold of $\propto = 0.05$, calculated based on 10,000 permutations of the phenotype data, is shown as a solid gray line. The 0.01 significance threshold is shown as a dashed line. The 0.001 significance thresholds were 4.20, 3.92, 4.10, and 4.21, for panels (*a*)-(*d*), respectively. Genetic markers (black diamonds) from left to right: (*a*) chr21_0049, Stn487, Stn423, Stn489, Stn491, Stn424, Stn425, chr21_0991, chr21_1170; (*b*) chr21_0049, chr21_0163, Stn487, Stn423, Stn489, Stn491, Stn424, Stn425, chr21_0997, chr21_1136; (*c*) Stn38, Stn42, Stn382, Stn46, Stn266, Stn253, chr4_280; (*d*) Stn38, Stn40, Gac4147, Stn382, Stn253, chr4_280. Marker Stn489 is starred for chromosome 21 and marker Stn382 is starred for chromosome 4 (see table 2.2 for a linkage map of each cross).

			chromosome 4				chromosome 21			
			20 dpf	*40 dpf*	*80 dpf*	*adult*	*20 dpf*	*40 dpf*	*80 dpf*	*adult*
FTC x LITC	CB1	aa	-2.3±7.5	-0.7±7.8	-8.3±13.7	-28.4±14.8	-10.7±8.7	-9.8±7.1	-19.9±15.2	-20±14.9
		ab	0.3±4.8	2.7±5.9	-4.7±10.1	5.1±10.8	2.7±3.9	3.4±5.8	-23.6±12	0.1±10.3
		bb	3.6±5	-5.1±11.1	4.4±19.9	22.3±17.2	3.2±6.4	10.4±12.6	28.8±10.1	26.8±19.5
		P	*0.817*	*0.796*	*0.856*	*0.066*	*0.288*	*0.214*	***0.004***	*0.129*
	CB2	aa	-10.1±8.2	-1.7±6.3	-12.1±13.2	-33.5±14.3	-14.4±8.1	-9±5.9	-19.7±13.7	-7.1±12.6
		ab	3.6±5.2	3.9±5.3	-2.4±9.6	12.8±10.9	5.5±4.1	3.4±5.3	-20.7±10.6	5.5±11
		bb	3.7±3.5	-4.7±8.3	1.4±16.8	21.5±15.5	0.3±5.9	8.3±9.1	26.7±10.6	8.5±18.8
		P	*0.273*	*625*	*0.801*	***0.018***	*0.07*	*0.18*	***0.004***	*0.73*
	CB3	aa	-5.3±9.5	-0.8±6.2	-11±11.4	-28.3±14.4	-16.4±9.6	-9.2±5.7	-11±14	-11±12.1
		ab	1±5.2	2±5.6	-4.4±8.7	6.4±10.1	5.1±4	6.7±4.7	-15±9	10.6±10.3
		bb	3.6±3.2	-2.6±6.9	7.9±15.4	27.7±14	0.1±5.9	0.1±10.4	14.7±10.7	0.6±17
		P	*0.666*	*0.872*	*0.623*	***0.024***	*0.067*	*0.127*	*0.089*	*0.437*
	CB4	aa	-3.1±6.9	-4.7±5.9	-15.6±9.8	-35.4±14	-7.2±6.7	-8.6±5.4	-29.9±10.8	-12.7±12.8
		ab	1.1±4.1	3±5.2	-1.2±7.3	5.2±9.7	3.4±3.4	6.3±5.5	-11.8±8.2	2.9±9.5
		bb	1.9±3.6	1.3±8.4	4.9±13.1	33.7±14.5	-2.5±5.3	0±8.1	20.6±7.5	19.5±17.9
		P	*0.81*	*0.639*	*0.424*	***0.003***	*0.297*	*0.17*	***0.001***	*0.303*
	CB5	aa	-1.7±6.5	-3.4±5.6	-6.5±10	-30.9±14.3	-7.2±6.5	-8.2±6.1	-21.4±10.8	-5.8±14.4
		ab	0.9±4.4	3.2±5.9	1.1±8.2	6.6±11.6	2.5±3.6	6.1±4.9	-14.4±9.2	0.9±10.6
		bb	3.4±3.9	-1±6.9	-7.8±19.6	22.3±14.8	1.8±6.1	-0.6±8.6	21.1±11.2	14.6±19.4
		P	*0.835*	*0.718*	*0.826*	***0.046***	*0.443*	*0.194*	***0.015***	*0.677*
	EB	aa	1±5.1	-4.5±4.2	23.6±10.9	-29.7±9.6	-10.3±3.5	-10.4±4.1	-27±7	-21.8±9.3
		ab	2.8±2.7	4.9±3.7	4.6±7.4	2.6±8.6	1.6±2.6	3.7±4	-16.8±8.2	3±7.2
		bb	-5.2±3.5	-4.9±6.2	11.7±11.4	24.4±8.4	3.9±4.5	10.9±3.8	33.8±7.7	23±15.2
		P	*0.207*	*0.179*	*0.066*	***0.003***	***0.048***	***0.006***	***<0.001***	***0.018***
PAXB x LITC	CB1	aa	-2.3±6.3	-17.8±11.2	-16.8±13.0	-38.5±22.5	-7.1±6.4	-6.5±8.8	-23.4±18.9	-11.1±20
		ab	-0.8±3.8	10.8±6.8	15.7±10.8	1±14.5	-0.3±3.5	4.5±6.7	-4.8±9.3	5.3±16.1
		bb	2.7±5.1	-6±6.9	0.8±14.9	60.9±22.9	4.1±5	-4.4±9.9	24.6±12.8	0.2±21.5
		P	*0.796*	***0.042***	*0.166*	***<0.001***	*0.324*	*0.557*	*0.051*	*0.826*
	CB2	aa	-0.8±5	-15.7±8.7	24.2±12.6	-41.7±20.1	-5.4±5.6	-2.8±8.3	-17.7±16.5	-3.7±19.7
		ab	-0.2±3.5	7.1±6.8	17.4±10.4	-2.3±13.4	2.5±3.6	4.3±6.5	-7.6±9.9	8.6±14.3
		bb	-0.5±4.4	-4.2±7	4.1±13.7	62.8±20.9	-0.2±3.1	-8±8.5	24.1±11.8	-12.5±19.7
		P	*0.994*	*0.113*	***0.039***	***<0.001***	*0.407*	*0.508*	*0.061*	*0.672*
	CB3	aa	-5.8±5.5	-13.9±7	27.6±12.8	-34.6±19.9	-2.2±5.7	-4±6.9	-13.8±16.0	1.0±1.9
		ab	-0.6±3.6	5.5±5.8	17.8±9.7	-8.0±11.7	-0.1±4	4.1±5.4	-8.2±9.8	3.1±13.6
		bb	4.7±5.2	-1.9±6.7	9.7±11.2	70.8±19.9	1.2±4.3	-6.1±8.1	22.3±10.9	-6.3±17.7
		P	*0.341*	*0.124*	***0.011***	***<0.001***	*0.895*	*0.474*	*0.083*	*0.918*
	CB4	aa	-2.7±5	-14.5±6.8	-24.9±11.2	-54.8±19.8	-4.3±6.5	-8.8±8	-19.1±14.3	-15.2±19.8
		ab	-1.5±3.6	3.4±6.3	18.7±10.2	-4±13.1	1.7±3.5	4.1±5.6	-6.5±9.1	5.9±13.8
		bb	4.9±5.2	3.9±7.1	3.3±14.8	75.2±20.7	0.8±3.6	-0.3±8.3	23.4±13.6	5.4±20.5
		P	*0.483*	*0.157*	*0.023*	***<0.001***	*0.632*	*0.41*	***0.056***	*0.647*
	CB5	aa	-5±5.3	-5.5±8.3	-18±10.2	-46.5±22.4	-4.2±5.7	-8.4±7.8	**-12.4±14.3**	1.6±20
		ab	-2.2±3	3.7±5.9	14±10.7	-2.2±12.7	-0.2±3.4	3.9±5.5	-10.9±9.6	-4±14.1
		bb	8.6±5.4	-5.7±6.9	0.5±14.9	71.5±19.7	1.8±4.9	-2.4±8.4	24.3±12.1	5.6±19.8
		P	0.1	0.512	0.131	***<0.001***	0.693	0.437	***0.05***	0.922
	EB	aa	-0.7±3.7	-7.9±5.4	-8.0±15.3	-30±13.7	-8.3±4.6	-13.1±4.5	-30.2±7.6	-29.1±11.1
		ab	-2.2±2.8	4.6±4.2	8.6±6.2	2.3±9.4	1±2.9	1.5±4	-2.7±5.8	-4.2±11.6
		bb	5.7±4.1	-0.8±5.1	-2.4±9.4	41.9±14.9	4.4±2.9	8.5±5.7	25.7±6.7	35.9±11.1
		P	*0.237*	*0.196*	*0.234*	***0.002***	*0.062*	***0.012***	***0.002***	***0.003***

Table 2.6 (previous page): Effects of chromosome 4 and 21 QTL on branchial bones.
Bone lengths were corrected for size, sex, and F2 family when appropriate (see Table 2.3) using linear regression analysis. Mean of the residual bone lengths (in microns) ± standard error is presented for each genotypic class for chromosome 21 (marker Stn489) and chromosome 4 (marker Stn382). MM=homozygous marine, MF=heterozygous, FF=homozygous freshwater. The *p*-value for an ANOVA of each marker is shown in italics below, and significant results are highlighted in bold.

	chr.	phenotype	age	peak marker	LOD	PVE	*d*/*a*
FTC x LITC	21	EB1	40, 80, adult	Stn489	6.64	6.69	0.15
	4	CB4	adult	Stn42	2.88	5.08	-0.09
PAXB x LITC	21	EB1	40, 80, adult	Stn491	9.84	9.30	-0.15
	4	CB4	adult	Stn382	5.08	9.03	0.21

Table 2.7: Genetic properties of chromosome 4 and 21 QTL.
QTL were mapped in R/qtl using the *fitqtl* and *refineqtl* functions to simultaneously map the effects of both chromosomes. Dominance (*d*/*a*) was calculated using the formula *d* = difference between the heterozygous phenotype and the midpoint between the homozygous phenotypes; *a* = half the phenotypic difference between the homozygous genotypes. PVE = percent variance explained.

Discussion

A heritable increase in branchial bone length in two freshwater stickleback populations is likely a trophic adaptation

A previous QTL mapping study found that most (11/14) freshwater alleles controlling stickleback branchial bone length produced longer bones [27], suggesting increased branchial bone length is under natural selection in freshwater environments. Supporting this prediction, we show that marine and freshwater bone lengths differ in the wild and that two populations of freshwater stickleback show strongly heritable increases in branchial bone length. This elongation of branchial bones may facilitate the processing of larger prey items in freshwater by providing a larger buccal cavity for food to pass through, generating greater crushing force, and/or offering increased muscle attachment area for the crushing of freshwater prey. While many studies have focused on evolutionary loss, these evolved increases in bone length demonstrate that despite the predictable loss of several skeletal elements including gill rakers, dorsal spines, and armor plates in freshwater environments [18,32,34], other parts of the skeleton (i.e. the branchial bones) increase in size despite the much lower environmental calcium concentration in

freshwater. In both freshwater populations studied here, the increased bone length differences are most pronounced in the dorsal (EB1) and posterior ventral bones (CB4 and CB5 demonstrate larger marine-freshwater differences than the more anterior 3 CBs). These findings suggest that the entire branchial skeleton is not uniformly enlarged relative to standard length in freshwater fish, but rather that independent genetic and developmental mechanisms have led to modular changes in the relative sizes of bones in the branchial skeleton. Heritable and similarly modular increases in bone length in two independent freshwater populations suggest that this trait may be adaptive in these environments [35].

A convergent increase in bone growth rate underlies bone elongation in freshwater sticklebacks

Here, we find two developmental mechanisms of evolved bone elongation (increased cartilage template size and bone growth rate) are at work in freshwater stickleback populations. Relative to marine fish, both freshwater populations have evolved an increased bone growth rate. All PAXB branchial bones and dorsal and posterior ventral FTC branchial bones have an increased growth rate relative to marine bones. Early cartilage template size is also increased in FTC freshwater fish. Therefore, the convergent evolution of these independent stickleback populations uses one shared developmental feature (increased bone growth rate) as well as at least one unique feature (increased cartilage template size in only one freshwater population). Differences in juvenile bone growth rates have been observed in the limbs of large and small mouse strains [36] and the elongated craniofacial bones of needlefish [37]. Multiple aspects of chondrocyte hypertrophy (the enlargement of chondrocytes that promotes bone growth) are altered to produce elongated digits in bats [38] and elongated limbs in jerboas [39]. Thus, developmental modulation of bone growth rates seems to be a shared mechanism of altering skeletal proportions in multiple taxa, including sticklebacks.

Shared QTL on chromosomes 4 and 21 suggest a parallel developmental genetic basis for freshwater bone length increase

Consistent with the convergent increased bone growth rate in two freshwater populations, these populations also share two overlapping QTL with strong effects on bone length at various stages of development. These two QTL, initially identified in the PAXB freshwater population [27] were successfully replicated here by crossing a different PAXB fish to a different marine background and also were observed in a second freshwater population, Fishtrap Creek. The developmental profiles of the QTL are remarkably similar between the two crosses. The effect of chromosome 4 is only seen in adult bones in both crosses. This QTL may only act late in development, or its earlier effects may only be apparent when fish reach a larger size. In contrast, chromosome 21 seems to exert its effects earlier than chromosome 4 in both crosses. Thus, similar developmental genetic features underlie the convergent evolution of longer branchial

bones, suggesting that even complex traits can have at least partially predictable genetic bases.

Although QTL in vertebrates are typically mapped in adults, a handful of studies have linked adult phenotypes to changes in juveniles. QTL for stickleback juvenile pigmentation and standard length have been identified [40]. In a finding similar to ours, a QTL controls shank growth rate in chickens during a specific time period of juvenile development [41]. Additionally, a QTL controlling adult hair color in beach mice can be traced to differential expression of the *Agouti* gene early in development [42]. These studies demonstrate that genetic changes that manifest at specific developmental stages can contribute to differences in final adult phenotype.

Future fine-mapping work will also determine whether FTC and PAXB share the alleles on chromosomes 4 and 21 that control bone size. Since freshwater stickleback populations are derived from a large oceanic ancestral population and often share alleles controlling evolved morphological changes [32,33,43,44], we parsimoniously hypothesize that the same alleles are at work in the FTC and PAXB populations. This recycling of ancestral alleles to produce a convergent phenotype has been called "collateral" evolution [3]. This hypothesis is supported by the similar developmental profiles of the QTL and the similarities in chromosomal location and dominances of the QTL in each cross. However, sticklebacks have also been shown to independently evolve alleles in multiple populations in the case of pelvic reduction [45], so the possibility remains that unique alleles have evolved in each population.

In conclusion, we find evidence of similar genetic and developmental properties underlying evolved increases in bone length in two independently derived freshwater stickleback populations. Both derived freshwater populations share an increased rate of growth of some bones relative to the bones of their marine counterparts, and the two QTL on chromosomes 4 and 21 demonstrate strikingly similar effects throughout development in crosses of each population. Our developmental genetic evidence supports a model that the same chromosome 4 and 21 genomic regions were selected independently in two freshwater populations to produce quantitative changes in a convergently evolved trophic phenotype. Future studies of other freshwater populations and crosses will test whether this evolved gain trait and the use of bone length QTL on chromosomes 4 and 21 are predictable features of freshwater adaptation.

Acknowledgements:

Andrew Glazer and Phillip Cleves performed most of the crosses and fish rearing for the F2s used in this study, as well as supervised the genotyping and phenotyping of many of the animals. Alyson Smith was involved in the studies of cartilage length as well as some of the F2 genotyping and phenotyping.

We thank Monica Jimenez, Nihar Patel, Brittany Bartolome, Jessica Grindheim, and Jiyeon Baek for assistance in genotyping and phenotyping, Anthony Lee and Patrick Lee for expertise in fish husbandry, and David Kingsley, Dolph Schluter, and Alex Pollen for help and support in wild fish collection. This work was supported in part by the NIH (R01-DE021475), a March of Dimes Basil O'Connor Starter Scholar Award, a Hellman

Family Faculty award (CTM), an NIH Predoctoral Training Grant (5T32GM007127) to PAE, AMG, and PAC, and National Science Foundation Graduate Research Fellowships (AMG and PAC).

References

[1] Losos, J. B. 2011 Convergence, adaptation, and constraint. *Evolution.* **65**, 1827-1840.

[2] Stern, D. L. & Orgogozo, V. 2008 The loci of evolution: how predictable is genetic evolution? *Evolution.* **62**, 2155-2177.

[3] Stern, D. L. 2013 The genetic causes of convergent evolution. *Nature Reviews Genetics.* **14**, 751-764.

[4] Martin, A. & Orgogozo, V. 2013 The loci of repeated evolution: a catalog of genetic hotspots of phenotypic variation. *Evolution.* **67**, 1235-1250.

[5] Jones, F. C., Grabherr, M. G., Chan, Y. F., Russell, P., Mauceli, E., Johnson, J., Swofford, R., Pirun, M., Zody, M. C., White, S. et al. 2012 The genomic basis of adaptive evolution in threespine sticklebacks. *Nature.* **484**, 55-61.

[6] Steiner, C. C., Weber, J. N. & Hoekstra, H. E. 2007 Adaptive variation in beach mice produced by two interacting pigmentation genes. *PLoS Biology.* **5**, e219.

[7] Feldman, C. R., Brodie, E. D., Brodie, E. D., & Pfrender, M. E. 2012 Constraint shapes convergence in tetrodotoxin-resistant sodium channels of snakes. *Proc. Natl. Acad. Sci. U. S. A.* **109**, 4556-4561. (DOI 10.1073/pnas.1113468109; 10.1073/pnas.1113468109).

[8] Tishkoff, S. A., Reed, F. A., Ranciaro, A., Voight, B. F., Babbitt, C. C., Silverman, J. S., Powell, K., Mortensen, H. M., Hirbo, J. B. & Osman, M. 2007 Convergent adaptation of human lactase persistence in Africa and Europe. *Nat. Genet.* **39**, 31-40.

[9] Hall, B. K. 2012 Parallelism, deep homology, and evo-devo. *Evol. Dev.* **14**, 29-33.

[10] Hall, B. K. 2005 *Bones and cartilage: developmental and evolutionary skeletal biology*. San Diego, CA: Academic Press.

[11] Atchley, W. R. & Hall, B. K. 1991 A model for development and evolution of complex morphological structures. *Biological Reviews.* **66**, 101-157.

[12] Fish, J. L., Sklar, R. S., Woronowicz, K. C. & Schneider, R. A. 2014 Multiple developmental mechanisms regulate species-specific jaw size. *Development.* **141**, 674-684.

[13] Losos, J. 2011 *Lizards in an Evolutionary Tree: Ecology and Adaptive Radiation of Anoles*, pp. 428: University of California Press.

[14] Sanger, T. J., Revell, L. J., Gibson-Brown, J. J. & Losos, J. B. 2012 Repeated modification of early limb morphogenesis programmes underlies the convergence of relative limb length in *Anolis* lizards. *Proceedings of the Royal Society B: Biological Sciences.* **279**, 739-748.

[15] Foster, S. A. & Bell, M. A. 1994 *The Evolutionary Biology of the Threespine Stickleback*. New York: Oxford University Press.

[16] Hagen, D. & Gilbertson, L. 1972 Geographic variation and environmental selection in *Gasterosteus aculeatus L.* in the Pacific Northwest, America. *Evolution.* **26**, 32-51.

[17] Schluter, D. & McPhail, J. D. 1992 Ecological character displacement and speciation in sticklebacks. *Am. Nat.* **140**, 85-108.

[18] Gross, H. P. & Anderson, J. 1984 Geographic variation in the gillrakers and diet of European threespine sticklebacks, *Gasterosteus aculeatus. Copeia.* **1**, 87-97.

[19] Lavin, P. & McPhail, J. 1986 Adaptive divergence of trophic phenotype among freshwater populations of the threespine stickleback (*Gasterosteus aculeatus*). *Can. J. Fish. Aquat. Sci.* **43**, 2455-2463.

[20] McGee, M. D. & Wainwright, P. C. 2013 Convergent evolution as a generator of phenotypic diversity in threespine stickleback. *Evolution.* **67**, 1204-1208.

[21] Anker, G. 1974 Morphology and kinetics of the head of the stickleback, *Gasterosteus aculeatus. The Transactions of the Zoological Society of London.* **32**, 311-416.

[22] Schilling, T. F. & Kimmel, C. B. 1994 Segment and cell type lineage restrictions during pharyngeal arch development in the zebrafish embryo. *Development.* **120**, 483-494.

[23] Schilling, T. F. & Kimmel, C. B. 1997 Musculoskeletal patterning in the pharyngeal segments of the zebrafish embryo. *Development.* **124**, 2945-2960.

[24] Haines, R. 1934 Epiphysial growth in the branchial skeleton of fishes. *QJ Microsc Sci.* **77**, 77-97.

[25] Kimmel, C. B., Ballard, W. W., Kimmel, S. R., Ullmann, B. & Schilling, T. F. 1995 Stages of embryonic development of the zebrafish. *Developmental Dynamics.* **203**, 253-310.

[26] Hammond, C. L. & Schulte-Merker, S. 2009 Two populations of endochondral osteoblasts with differential sensitivity to Hedgehog signalling. *Development.* **136**, 3991-4000.

[27] Miller, C. T., Glazer, A. M., Summers, B. R., Blackman, B. K., Norman, A. R., Shapiro, M. D., Cole, B. L., Peichel, C. L., Schluter, D. & Kingsley, D. M. 2014 Modular skeletal evolution in sticklebacks is controlled by additive and clustered quantitative trait loci. *Genetics.* **197:** 405-420. (DOI 10.1534/genetics.114.162420).

[28] Orr, H. A. 1998 Testing natural selection vs. genetic drift in phenotypic evolution using quantitative trait locus data. *Genetics.* **149**, 2099-2104.

[29] Albert, A. Y., Sawaya, S., Vines, T. H., Knecht, A. K., Miller, C. T., Summers, B. R., Balabhadra, S., Kingsley, D. M. & Schluter, D. 2008 The genetics of adaptive shape shift in stickleback: pleiotropy and effect size. *Evolution.* **62**, 76-85.

[30] McGee, M. D. & Wainwright, P. C. 2013 Sexual dimorphism in the feeding mechanism of threespine stickleback. *J. Exp. Biol.* **216**, 835-840.

[31] Swarup, H. 1958 Stages in the development of the stickleback *Gasterosteus aculeatus (L.). J. Embryol. Exp. Morphol.* **6**, 373-383.

[32] Colosimo, P. F., Hosemann, K. E., Balabhadra, S., Villarreal Jr, G., Dickson, M., Grimwood, J., Schmutz, J., Myers, R. M., Schluter, D. & Kingsley, D. M. 2005 Widespread parallel evolution in sticklebacks by repeated fixation of ectodysplasin alleles. *Science.* **307**, 1928-1933.

[33] Bell, M. A. & Aguirre, W. E. 2013 Contemporary evolution, allelic recycling, and adaptive radiation of the threespine stickleback. *Evol. Ecol. Res.* **15**, 377-411.

[34] Gross, H. P. 1978 Natural selection by predators on the defensive apparatus of the three-spined stickleback, *Gasterosteus aculeatus L. Can. J. Zool.* **56**, 398-413.

[35] Schluter, D. 2000 *The ecology of adaptive radiation*. New York: Oxford University Press.

[36] Sanger, T. J., Norgard, E. A., Pletscher, L. S., Bevilacqua, M., Brooks, V. R., Sandell, L. J. & Cheverud, J. M. 2011 Developmental and genetic origins of murine long bone length variation. *Journal of Experimental Zoology Part B: Molecular and Developmental Evolution.* **316**, 146-161.

[37] Gunter, H. M., Koppermann, C. & Meyer, A. 2014 Revisiting de Beer's textbook example of heterochrony and jaw elongation in fish: calmodulin expression reflects

heterochronic growth, and underlies morphological innovation in the jaws of belonoid fishes. *EvoDevo.* **5**, 8. (DOI 10.1186/2041-9139-5-8; 10.1186/2041-9139-5-8).

[38] Farnum, C., Tinsley, M. & Hermanson, J. 2008 Forelimb versus hindlimb skeletal development in the big brown bat, *Eptesicus fuscus*: functional divergence is reflected in chondrocytic performance in autopodial growth plates. *Cells Tissues Organs (Print).* **187**, 35-47.

[39] Cooper, K. L., Oh, S., Sung, Y., Dasari, R. R., Kirschner, M. W. & Tabin, C. J. 2013 Multiple phases of chondrocyte enlargement underlie differences in skeletal proportions. *Nature.* **495**, 375-378.

[40] Greenwood, A. K., Jones, F. C., Chan, Y. F., Brady, S. D., Absher, D. M., Grimwood, J., Schmutz, J., Myers, R. M., Kingsley, D. M. & Peichel, C. L. 2011 The genetic basis of divergent pigment patterns in juvenile threespine sticklebacks. *Heredity.* **107**, 155-166.

[41] Gao, Y., Du, Z., Feng, C., Deng, X., Li, N., Da, Y. & Hu, X. 2010 Identification of quantitative trait loci for shank length and growth at different development stages in chicken. *Anim. Genet.* **41**, 101-104.

[42] Manceau, M., Domingues, V. S., Mallarino, R. & Hoekstra, H. E. 2011 The developmental role of Agouti in color pattern evolution. *Science.* **331**, 1062.

[43] Schluter, D. & Conte, G. L. 2009 Genetics and ecological speciation. *Proc. Natl. Acad. Sci. U. S. A.* **106**, 9955-9962. (DOI 10.1073/pnas.0901264106; 10.1073/pnas.0901264106).

[44] Miller, C. T., Beleza, S., Pollen, A. A., Schluter, D., Kittles, R. A., Shriver, M. D. & Kingsley, D. M. 2007 *cis*-Regulatory changes in *Kit ligand* expression and parallel evolution of pigmentation in sticklebacks and humans. *Cell.* **131**, 1179-1189.

[45] Chan, Y. F., Marks, M. E., Jones, F. C., Villarreal Jr, G., Shapiro, M. D., Brady, S. D., Southwick, A. M., Absher, D. M., Grimwood, J. & Schmutz, J. 2010 Adaptive evolution of pelvic reduction in sticklebacks by recurrent deletion of a *Pitx1* enhancer. *Science.* **327**, 302-305.

3. Genetic dissection of a supergene implicates *Tfap2a* as underlying craniofacial evolution in threespine sticklebacks

Abstract

In nature, functionally related phenotypes often coevolve and can be controlled by two or more tightly linked genetic loci (supergenes). Dissecting the genetic basis of these linked phenotypes is a major challenge in evolutionary genetics. Multiple populations of threespine stickleback fish (*Gasterosteus aculeatus*) have convergently evolved longer branchial bones and increased pharyngeal tooth number in freshwater habitats, likely adaptations to dietary differences between environments. Prior quantitative trait locus (QTL) mapping showed that both traits are partially controlled by overlapping genomic regions on chromosome 21, and that a regulatory change to *Bmp6* likely underlies one tooth number QTL. Here we show that branchial bone length maps to a 155 kb, 8-gene interval tightly linked to but excluding the coding regions of *Bmp6* and containing the candidate gene *Tfap2a*. Further recombinant mapping revealed this QTL is separable into at least two fractions. During embryonic and larval development, *Tfap2a* is expressed in the branchial bone primordia, in which the freshwater allele of *Tfap2a* is expressed at lower levels relative to the marine allele in hybrid fish. Induced loss-of-function mutations in *Tfap2a* reveal an essential role in stickleback craniofacial development and show that bone length is sensitive to *Tfap2a* dosage in heterozygotes. Combined, these results suggest that closely linked but genetically separable *cis*-regulatory changes in *Bmp6* and *Tfap2a* contribute to a genomic region underlying evolved skeletal gain in multiple freshwater populations. Since both evolved changes are downregulation, perhaps evolved gains of morphology commonly arise from *cis*-regulatory reduction of genes inhibiting tissue growth.

Introduction

Intraspecific morphological variation offers a unique opportunity to dissect the genetic changes that underlie adaptation to novel environments and to answer questions about the molecular basis of evolution. What types of genetic changes underlie morphological evolution? When adaptive changes for multiple phenotypes cluster into genomic "hotspots" of evolution, are adaptive alleles composed of multiple genetic changes? Do similar changes in phenotype use similar genetic changes across lineages?

The repeated evolution of similar phenotypes (convergent evolution) can be used to test each of these questions in replicate. Testing whether the same or different genetic differences underlie similar phenotypes may shed light on the predictability and repeatability of evolution (Stern and Orgogozo 2008; Stern 2013; Rosenblum et al. 2014). Convergent phenotypic evolution may be caused by independent mutations, similar

mutations in shared pathways, or by introgression of shared alleles. Convergent evolution via different genetic mechanisms illustrates the ability of biological systems to achieve similar outcomes via different pathways, whereas convergent evolution via similar or shared mutations is indicative of constraints on the types of mutation or developmental processes that can achieve a particular phenotype. Both theoretical and empirical studies indicate that parallel evolution can occur quite often (Orr 2005; Conte et al. 2012). However, given the relatively small number of cases where a convergently evolved phenotype has been mapped to the level of a gene, the extent of genetic parallelism remains unknown, especially for quantitative traits.

The loci underlying evolved traits can be mapped to genomic regions using quantitative trait locus (QTL) mapping. QTL for multiple phenotypes that map to the same genomic region raise the question of whether a single pleiotropic locus affects multiple traits, or whether separate closely linked alleles affect each trait. Clustering of two or more adaptive loci into "supergenes" is a commonly described feature of evolution (Schwander et al. 2014; Thompson and Jiggins 2014), likely because close linkage of loci allows advantageous combinations of traits to be inherited together. Supergenes have been shown to control pigmentation patterns in locusts (Nabours 1933), mimetic patterns in butterflies (Joron et al. 2011; Kunte et al. 2014), color and patterning in snails (Murray and Clarke 1976a,b), breeding behavior and morphology in birds (Thomas et al. 2008; Küpper et al. 2016; Lamichhaney et al. 2016; Tuttle et al. 2016), and life history, morphology, and pollination syndromes in plants (Mather 1950; Lowry and Willis 2010; Hermann et al. 2013). The supergenes studied molecularly to date are often controlled by chromosomal inversions with large phenotypic effects (e.g. Thomas et al. 2008; Lowry and Willis 2010; Küpper et al. 2016; Lamichhaney et al. 2016). In *Heliconius* butterflies, a supergene in one species was shown to be an inversion containing tightly linked loci that are uninverted and can recombine in other species (Joron et al. 2006, 2011). Whether non-inversion regions controlling quantitative variation in phenotypes can also act as supergenes is an outstanding question. Determining whether clustering of QTL controlling different traits is caused by close linkage or pleiotropy requires careful genetic dissection of the regions of interest.

While QTL for different traits can cluster in the genome, separate loci affecting a single phenotype can also be clustered within a QTL. Whether a given QTL for a single trait typically represents an individual locus, or alternatively, multiple tightly linked loci, remains an outstanding question in quantitative genetics. The regions identified by QTL mapping are often broad and contain many genes, so this question often cannot be answered within a F2 cross with a single generation of recombination between parental genomes. Follow-up studies have revealed examples of single and multiple loci responsible for QTL. For example, QTL mapping in corn x teosinte crosses has revealed some large effect QTL fractionate (Studer and Doebley 2011) including one that was fractionated into up to five distinct tightly linked loci (Lemmon and Doebley 2014). In contrast, other corn x teosinte QTL were mapped to reveal simpler, single causative loci (Wang et al. 2005; Hung et al. 2012; Wills et al. 2013). In mice, one QTL controlling pigmentation maps to a single coding mutation in *Mc1r* (Steiner et al. 2007) but a second QTL is controlled by multiple smaller effect mutations affecting different aspects of

Agouti expression (Linnen et al. 2013). Interspecific differences in *Drosophila* pigmentation map to a single regulatory element of *tan* (Jeong et al. 2008), but differences in *Drosophila* trichome patterning are caused by changes to multiple nucleotides in different regulatory elements of the *shaven baby* gene that are spread out over 50 kb (McGregor et al. 2007; Frankel et al. 2011). Understanding how loci control complex traits and clusters of traits requires model systems with the ability to perform high-resolution genetic mapping on naturally varying phenotypes.

The threespine stickleback fish has convergently evolved countless freshwater forms from a common marine ancestor and has emerged as a powerful model system for studying both the genetic basis and repeatability of morphological evolution. In freshwater environments, sticklebacks repeatedly evolve a suite of craniofacial and other morphological adaptations to cope with differences in diet, predation, and other environmental variables (Bell and Foster 1994). Marine and freshwater forms are easily intercrossed to produce large clutches in the laboratory (Peichel et al. 2001), the stickleback genome is well assembled and annotated (Jones et al. 2012; Glazer et al. 2015), and reverse genetic techniques are available (Erickson et al. 2016a) facilitating both genetic and genomic dissection of the molecular basis of evolved traits.

Marine and freshwater sticklebacks occupy different trophic niches: while marine fish feed on small planktonic prey, freshwater fish typically consume diets of larger macroinvertebrates (Kislalioglu and Gibson 1977; Gross and Anderson 1984). Sticklebacks process food primarily with the branchial skeleton and pharyngeal jaw, used to chew and crush food en route to the gut (McGee and Wainwright 2013; McGee et al. 2013). The branchial skeleton is derived from the posterior five pharyngeal arches and consists of segmental homologs of the upper and lower jaw. Composed of 5 bilateral pairs of ventral bones, 4 bilateral pairs of dorsal bones, and 3 bilateral pairs of tooth plates, the branchial skeleton is also lined by gill rakers that help trap food particles (Anker 1974). While differences in jaw morphology and gill raker number have long been described as trophic adaptations to freshwater and benthic (lake bottom) environments (Bentzen and McPhail 1984; Schluter and McPhail 1992), we have recently described increases in both pharyngeal tooth number and branchial bone length as repeatable and heritable features of freshwater adaptation (Cleves et al. 2014; Erickson et al. 2014; Ellis et al. 2015). We hypothesize that these increases permit freshwater fish to process larger prey items via a larger pharyngeal cavity and greater chewing capacity. The first epibranchial (EB1) bone, a serial homolog of the upper jaw, serves as a critical lever for the mastication motion of the pharyngeal jaw (Wainwright 2006). Comparing branchial skeletal morphology between marine and freshwater populations revealed EB1 to be the most elongated branchial bone in freshwater sticklebacks (Erickson et al. 2014).

The branchial bones are endochondral bones that form from cartilage templates during late embryonic development. The cartilage templates ossify in the middle and elongate via chondrocyte maturation in the growth plates at either end of the bone (Haines 1934), much like mammalian long bones (Haines 1942). Changes to both the early patterning of cartilage and the relative growth of bones are both predicted to be important determinants of skeletal evolution and have been shown to affect skeletal evolution in other systems (Farnum et al. 2008a,b, Sanger et al. 2011, 2012).

Marine and freshwater sticklebacks can be intercrossed, permitting forward genetic analysis of quantitative trait loci (QTL) controlling evolved phenotypes. Previous QTL mapping of stickleback skeletal evolution has identified over a hundred QTL controlling a variety of traits, including pharyngeal tooth number and branchial bone length (Cleves et al. 2014; Erickson et al. 2014, 2016b; Miller et al. 2014). These trophic traits are highly polygenic, with up to 14 different genomic regions affecting branchial bone length and eight regions affecting tooth number. Strikingly, despite the genetic complexity of each trait, the QTL controlling skeletal adaptation are significantly clustered on three chromosomes, including chromosome 21 (Miller et al. 2014), and QTL clustering was replicated in additional crosses (Erickson et al. 2016b). Clustering of adaptive loci is predicted by population genetic theory when migration occurs between differentially adapted populations (Yeaman 2013), such as occurs between marine and freshwater adapted sticklebacks (Schluter and Conte 2009; Bell and Aguirre 2013). Further genetic analysis is required to determine whether the clustering of QTL is due to mutations in pleiotropic loci controlling multiple skeletal traits, or multiple tightly linked loci.

Increases in both branchial bone length and pharyngeal tooth number map to overlapping regions of chromosome 21 in the Paxton Benthic population from British Columbia (Miller et al. 2014). Although chromosome 21 contains an inversion that typically differs between marine and freshwater populations (Jones et al. 2012), the pharyngeal tooth number QTL was fine-mapped to a genomic region over a megabase outside this inversion which contains a *cis*-regulatory allele of the gene *Bone morphogenetic protein 6* (*Bmp6*) (Cleves et al. 2014). Additional work mapped branchial bone length QTL to peaks near *Bmp6* on chromosome 21 in Paxton Benthic and a second freshwater population, Fishtrap Creek, Washington (Erickson et al. 2014). BMP ligands are critical to both tooth and bone development (Balic and Thesleff 2015; Salazar et al. 2016), so *Bmp6* is an excellent candidate gene for both QTL. We sought to determine the molecular basis of the bone gain QTL in both of these crosses to answer three questions: 1) are the bone length and tooth number QTL genetically separable? 2) what is the developmental genetic basis of the bone length QTL? and 3) does convergent evolution of branchial bone gain in two freshwater populations have a parallel genetic basis?

Methods

Animal statement

All animal work was approved by UCB animal protocol #R330. Fish were reared as previously described (Erickson et al. 2014).

Recombinant mapping and statistical analysis

Fish segregating the chromosome 21 epibranchial length QTL from the PAXB x LITC (Little Campbell Marine) and FTC x LITC crosses (Erickson et al. 2014) were propagated and genotyped with markers Stn487 (Cleves et al. 2014), PAE309, PAE323, and/or PAE349 (Table 3.1) to ensure that marine and freshwater alleles of the QTL interval were passed on to each generation and to look for recombination events within the QTL. To test the

phenotypic effects of recombinant chromosomes, recombinant individuals were identified and crossed to related fish that were either heterozygous for the QTL or homozygous for the marine chromosome within the QTL interval. Offspring were grown to 25-28 days post fertilization (dpf) (10-12mm, FTC cross) or to roughly 80 dpf (~20 mm, PAXB cross). These fish were fixed, stained with Alizarin red, cleared, dissected, and photographed, and epibranchial 1 (EB1) bone length was measured as previously described (Erickson et al. 2014; Miller et al. 2014). Pharyngeal tooth number was counted as described (Cleves et al. 2014). EB1 length was measured on both the left and right side and averaged. Bone length was corrected for fish standard length in each clutch separately and residuals of bone length or back-transformed residuals were used for all analyses.

Name	Chr. 21 position (Mb)	Forward Primer	Reverse Primer	purpose
PAE309/310	3.9	CAGGGTGATTGTCAGCCATT	CCCGATTTATTAAAATGCCAAA	indel marker
PAE311/312	4.15	AACGTCCAAAGCAATGTTGG	CGGACTAGCAGAGCAATGTG	indel marker
PAE317/318	4.37	GGCCGTACTTCCACATTATGA	ACTCCATGGCAACCAGATCC	indel marker
PAE323/324	4.42	AGAGGGAGCCTTGTTGCAT	GCTCTGGTCTGTCAGCGTAA	indel marker
PAE325/326	4.44	ATGCTTCCCGATAAATCGATAC	TGCCCAAAATAGGAAATGACA	indel marker
PAE349/350	5.96	CCTGCAAGATGTCCCCTAAA	CCTCCCCTAGTTGCATTGTG	indel marker
PAE428/429	4.2	AAAAACAGATGTGCCCCTTTT	TACAGGCCTCTCAGGCAAAT	indel marker
PAE438/439	4.35	ACGTGTTTTGGTCCGATACC	CCTCCAATTGGCCCTTAATA	indel marker
PAE462/463	4.98	TGCAGGTACCAAAATAGCAAGA	GCTCCGGGACTAGTTGCAC	indel marker
PAE464/465	5.03	GATTTGGCTCTTGGGGTGTA	AGGCAAATGAAGCACAGAGG	indel marker
PE444/445	2.75	AAAGGCAGTGTCGATGTTGG	TTGGGAAGATGTGTGGACAG	indel marker
PE363/364	5.6	TGATCACATGGTTCTTGTCCA	AAACAGGGTTTGGGACTGG	indel marker
PAE511/512	4.59	GTGCCAGAAATAGCGTCGGC	GGGTATTACGTACCGTGATGATGATT	indel marker
PAE513/514	4.72	CATAGCTGCCGATTAATCCCCCT	TGAGCGTGCACTACTGGGGA	indel marker
PAE436/437	4.32	TGATGCAAGAGAGCTGGAAA	GCTGGGTTGGAAACGATAAA	indel marker
PAE438/439	4.36	ACGTGTTTTGGTCCGATACC	CCTCCAATTGGCCCTTAATA	indel marker
PAE540/541	4.21	AAAACTGGTGGCGCGTGTTT	CAACTGCCTTTCGTCAGTCTGC	indel marker
PAE542/543	4.231	ACCGACGTGCACAACACAA	ACCTGCCACCAGCAGAGTCT	indel marker
PAE496/497	3.963	ACAATATTTGACAGGTTATCTAGTGTCC	GTGAGAGTTGTGAGGACACAAAACAA	indel marker
PAE414/416	4.26	gccggcagcAACAACAACATCTCCTCCATCCCC	gccggcagcCCCAGCAGGGAGGCGTTGAG	AP2 ASE
PAE499/519	3.85	gccggcagcACGAAGCTCATGACCATGTCTGC	gccggcagcATCGTGGACCGGTACGAACC	Bmp6 ASE
PAE379/381	4.26	GGGTCGTTGACGTGCGAGTAA	AGCGGGACAACGTCATCACTTA	talen genotyping

Table 3.1: Primers used in this study.

For genotyping, primers were designed to amplify insertion/deletions (indels) that differed between the grandparents of the PAXB x LITC and FTC x LITC crosses.

For recombinant fish crossed to heterozygotes, the R package *lmtest* was used to perform a likelihood ratio test (*lrtest*) for whether the recombinant chromosome behaved as marine or freshwater (Cleves et al. 2016 in prep). Genotypes were coded in two possible ways: the recombinant chromosome (R) was treated as (1) either a marine (M) or (2) a freshwater (F) chromosome. Depending on the setup of the cross, one of these genotype codes contained three possible genotypes (MM, MF, and FF) and one contained two possible genotypes (MM & MF or MF & FF). The *lrtest* function was used to compare nested models of bone length: a model with both possible genotype variables (1+2) was compared to a model with only one of the possible genotypes (1 or 2). A significant difference between the model containing one variable and a model containing both variables indicated that the addition of the second genotype significantly improved the model, and therefore that the second genotype best represented the behavior of the recombinant chromosome. For recombinant chromosomes crossed to homozygous marine fish (MM), a T-test was used to determine whether the recombinant behaved like a freshwater chromosome (i.e. there was a significant increase in bone length between the MM and MR genotypic classes).

Genome resequencing and analysis of Tfap2a coding sequence

A coding sequence variant in *Tfap2a* was initially identified by examining the genome sequences of the grandparents of a previously studied LITC x FTC cross (Glazer et al. 2015) and the PAXB x LITC cross (Cleves et al., in prep). This coding change disrupts an AvaI cut site. Additional cross grandparents were screened for this mutation by amplifying with PAE414/416 and digesting the PCR product with AvaI; presence of an uncut band indicated that the cross parent carried the coding change.

The TFAP2A predicted protein coding sequence for sticklebacks was acquired from Ensembl (ENSGACT00000003048.1) (Yates et al. 2016). Additional vertebrate TFAP2A sequences (human, mouse, frog, zebrafish, and salmon) were gathered from RefSeq (Pruitt et al. 2014). Sequences were aligned with Clustal Omega (Sievers et al. 2011) and shared residues visualized with Boxshade (http://www.ch.embnet.org/software/BOX_form.html).

Analysis of chromosome 21 QTL in previously published crosses

Genotype data from Glazer et al. (2015) and Erickson et al. (2016b) were used to test for the presence of a bone length QTL at the marker containing *Tfap2a* (binned marker 16_9) in the LITC x FTC and LITC x ENOB crosses. Bone lengths were measured for 210 fish from the Glazer study as described above and processed to correct for size and sex as described in Erickson et al. 2016b. Processed bone lengths for the ENOB x LITC cross were used as described in Erickson et al. 2016b.

Allele specific expression (ASE) assay for Tfap2a and Bmp6

We modified a genotype by sequencing (GBS) protocol (Elshire et al. 2011; Glazer et al. 2015) to assay allele-specific expression (ASE) in barcoded PCR products amplified from cDNA. PAXB x LITC fish segregating the QTL interval (F9 generation) were intercrossed and their F10 offspring were sacrificed with MS-222. The branchial skeleton or EB1 was immediately dissected on ice (within 5 minutes of sacrifice) and stored in 500 uL TRI reagent (Ambion) at -80 °C. Tissue was collected from five time points: 9 dpf, 13 dpf, 17 dpf, 22 dpf, and 35 dpf. At 9 and 13 dpf, the entire branchial skeleton was collected. At 17 and 22 dpf, the dorsal portion of the branchial skeleton was collected, and at 35 dpf individual EB1 bones were isolated. The remaining tissue was stored in ethanol. DNA was isolated from the excess tissue by digesting overnight with proteinase K in lysis buffer at 55 °C and then performing a phenol-chloroform extraction followed by ethanol precipitation (Green and Sambrook 2012). DNA was genotyped with indel markers PAE309/310 and PAE311/312 (flanking *Tfap2a*) to identify heterozygous fish. RNA was extracted from heterozygous fish following the TRI reagent manufacturer's protocol, with the final RNA pellet resuspended in 20 uL RNAse free water. For each time point, 24 fish were collected, yielding approximately 12 heterozygotes.

To prepare cDNA, 4.5 uL RNA was treated with 0.5 uL amplification grade DNAse 1 (Invitrogen) according to manufacturer's protocol. cDNA was synthesized with Superscript III (Invitrogen) using random hexamers according to the manufacturer's protocol but halving all volumes to produce a final volume of 10 uL cDNA. One uL of cDNA or 50 ng of genomic DNA from each fish were used as templates in separate Phusion PCR reactions with primers PAE414 and PAE416, designed to amplify the nonsynonymous SNP in *Tfap2a* at position chrXXI:4,265,995 which was shared in both the PAXB and FTC cross parents. For *Bmp6*, the primers PAE499 and PAE519 were used to amplify a synonymous SNP at chrXXI: 3,831,812, which was also shared in the grandparents of each cross. The 5' end of each primer contained an ApeKI cut site so that the product could be ligated into the GBS adapters. These PCR products were purified in a 96-well format by the UC Berkeley DNA Sequencing Center and quantified using a PicoGreen assay (ThermoFisher) on a BioTek Flx800 plate reader.

Up to 25 ng of each purified PCR product was used in the GBS library preparation as previously described (Elshire et al. 2011; Glazer et al. 2015). Briefly, the PCR products were combined with 1.8 ng of barcoded adaptors and digested with ApeKI (NEB) for 2 hours at 75 °C. The adapters were ligated to the PCR products with T4 DNA Ligase (NEB) for 1 h at 22 °C. Then, 5 uL of each ligation product was pooled for a PCR purification (Qiagen). Sequencing adapters were added with 10 cycles of PCR using Taq 2X master mix (NEB). The PCR product was cleaned up and size-selected twice with 1.5 volumes of Sera-Mag beads prepared as previously described (Rohland and Reich 2012). Six libraries were pooled equally into one Illumina HiSeq 2000 lane and sequenced on 50 bp single-end rapid run mode at the UC Berkeley DNA Vincent Coates sequencing center using a custom indexing primer. Indexed, barcoded sequences were de-multiplexed using custom scripts and aligned to the expected PCR product sequence, using a custom Python script to count marine and freshwater alleles at the polymorphic site. Samples receiving fewer than 300 aligned reads were removed from the analysis (n = 4 removed for *Tfap2a*, n = 8

removed for *Bmp6*), as our preliminary experiments suggested that approximately 300 reads were required to obtain reliable 1:1 allelic ratios for genomic DNA PCR products. Two samples were removed that had gDNA ratios less than 0.5 for *Bmp6*, and 0Outliers greater than 3 standard deviations from the mean were also removed (n = 2 for *Tfap2a*). The gDNA and cDNA marine:freshwater ratios were compared at each time point to test for ASE using the *wilcox.test* function in R.

Genome editing of Tfap2a

TALENs were generated to target the second exon of *Tfap2a* (see Table 3.2 for RVD design) following established protocols (Cermak et al. 2011; Doyle et al. 2012) and injected into one-cell stickleback embryos as previously described (Erickson et al. 2015, 2016a). A subset of injected embryos (n = 10 to 12) were screened for mutations by amplifying with primers PAE379 and PAE381 in a standard Phusion (NEB) reaction, digesting the PCR product with PvuII, and running the digested product on a 1% agarose gel. Undigested product of ~297 base pairs indicated molecular lesions that disrupted a PvuII sequence at the expected DNA cleavage site. Clutches carrying lesions were raised to adulthood and outcrossed to wild-type fish. Offspring were screened for molecular lesions by fin biopsy before crossing. Lesions were sequenced by extracting the undigested PCR band from an agarose gel followed by Sanger sequencing (see Table 3.3 for lesions studied). F1 individuals carrying lesions were propagated for further analysis.

	TALEN 1			TALEN 2		
position	**pFUS_A 5'**	**pFUS_B 5'**	**pLR 5'**	**pFUS_A 3'**	**pFUS_B 3'**	**pLR 3'**
1	HD	HD	HD	NN	NN	HD
2	NN	NG		NN	NN	
3	NG	HD		NN	HD	
4	NN	NG		NI	NG	
5	NG	NN		HD	NI	
6	NI	NN		NI	HD	
7	NN			NN	HD	
8	NN			HD	NG	
9	NN			HD		
10	NI			NI		

Table 3.2: TALEN RVDs.

Description	Sequence	Product
wild-type	ACAGCCAGGCTACCTCAGCTGGGCGGCGTGGGCC	427 AA
6 bp deletion/22 bp insertion (*)	ACAGCCAGGCTACCTCAGC******GCGTGGGCC	early stop after 158 AA
13 bp deletion	ACAGCC---C-A-CTCA--------GCGTGGGCC	early stop after 142 AA
22 bp deletion	ACAGC----------GTGGGCC	early stop after 139 AA
8 bp deletion	ACAGCCAGGCTACCTCAGC--------GTGGGCC	early stop after 150 AA
10 bp deletion	ACAGCCAGGCTACCTCAG----------TGGGCC	early stop after 143 AA

Table 3.3: Tfap2a mutations studied.

Sequenced mutations induced by TALENs. ** indicates 22 bp inserted sequence: ACCTCACCTCAGCTCCCTACAC. Product indicates predicted protein product length, which is frame-shifted and results in an early stop for all mutations (AA = amino acids).

TALEN phenotyping

Tfap2a +/- heterozygotes were intercrossed to produce clutches containing homozygous mutants. These larval fish were sacrificed immediately after hatching (9-10dpf), fixed overnight in 4% paraformaldehyde in 1X phosphate buffered saline, stained with Alcian blue, digested with trypsin, and mounted in glycerol as previously described (Kimmel et al. 1998), and imaged on a DM2500 compound microscope. For bone length phenotyping, heterozygous fish were outcrossed to wild-type lab-reared individuals. In the FTC background, a total of 144 fish were studied from three families carrying a 16 bp insertion, an 8 bp deletion, and a 10 bp deletion. In LITC, two families totaling 96 fish carrying a 13 bp deletion were studied. Fish were raised to 25-28 dpf and genotyped and phenotyped as described above.

In situ hybridization

In situ hybridization for *Tfap2a* was performed on whole embryos as previously described (Cleves et al. 2014).

Results

Fine mapping of the bone length QTL

We previously identified three stickleback chromosomes enriched for quantitative trait loci (QTL) controlling evolved skeletal changes (Miller et al. 2014). One of these "supergene" regions maps to a 13 cM region on chromosome 21, where seven different skeletal traits map, including the constructive gain of both branchial bone length and pharyngeal tooth number. We tested two hypotheses about these overlapping skeletal QTL. First, we hypothesized that like tooth number, evolved bone length is also mediated by an intronic haplotype of *Bmp6* (Cleves et al. 2014; Cleves et al., in prep.), as they are both skeletal gain traits and BMPs are important regulators of skeletal development

(Salazar et al. 2016). Second, we hypothesized that the same genomic region underlies the bone length QTL in both freshwater populations. We tested these hypotheses in two freshwater x marine crosses known to carry the bone length QTL (Erickson et al. 2014): Paxton Benthic (PAXB, British Columbia) and Fishtrap Creek (FTC, Washington), both crossed to Little Campbell River marine (LITC, British Columbia).

We analyzed a series of recombinant chromosomes derived from later generations of the original crosses used to map bone length. For each recombinant chromosome, we bred dozens to hundreds of offspring of varying genotypes and used statistical tests to ask whether each recombinant chromosome produced a phenotypic effect significantly more similar to a marine or freshwater chromosome. If the chromosome behaved like a freshwater chromosome (significantly increased branchial bone length relative to a marine chromosome or did not differ from a freshwater chromosome), we concluded that the QTL was contained within the freshwater portion of the recombinant (blue in Figure 3.1). If the chromosome behaved like a marine chromosome (did not increase branchial bone length relative to a marine chromosome or caused significantly shorter bone length when compared to a freshwater chromosome), we concluded that the QTL was contained within the marine portion (red in Figure 3.1).

Using this approach, we identified recombinant chromosomes in each cross with recombination events located within ~2Mb of *Bmp6.* In the PAXB cross, we identified one recombinant chromosome that separated the *Bmp6* locus from the region controlling bone length (Figure 3.1, chromosome R_{P_2}). This particular recombinant chromosome was previously shown to increase pharyngeal tooth number (Cleves et al, in prep.), but did not affect EB1 length in our analysis. Therefore, the chromosome 21 bone length QTL and tooth number QTL are controlled by different genetic loci in the PAXB x LITC cross. In the Fishtrap Creek cross, we also found a chromosome that carried a freshwater allele of *Bmp6* but did not affect bone length (Figure 3.1, chromosome R_{F_2}, Table 3.4). We found that this recombinant chromosome increased pharyngeal tooth number (Figure 3.2), suggesting that the FTC population also harbors a tooth QTL in a portion of the chromosome containing *Bmp6*, and like in the PAXB population, the tooth number and bone length QTL are genetically separable. Thus, the coding sequence of *Bmp6* is excluded from the bone length QTL interval in both crosses.

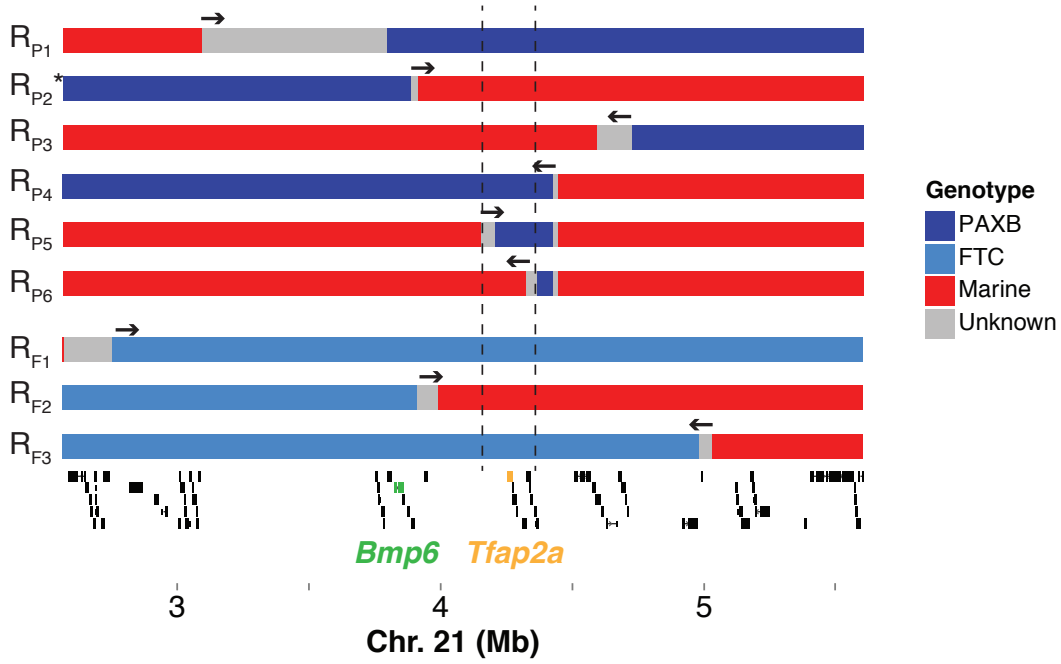


Figure 3.1: Recombinant mapping of bone length QTL in two crosses.

Recombinant chromosomes were identified in later generations (>F4) of the original F2 crosses described in Erickson et al. 2014 by screening for recombination events between indel markers that distinguish marine and freshwater chromosomes. Recombinant fish were crossed to known-genotype relatives descended from the same F2 cross. Colors indicate the marine/freshwater identity along each recombinant chromosome; regions between confirmed breakpoint genotypes are colored grey (see key). Arrows above each chromosome indicate the direction of the bone length QTL relative to the recombination break point; see Table 3.4 for full statistical analysis of each cross. Note that recombinant chromosomes R_{P5} and R_{P6} in the PAXB cross are both double recombinants derived from chromosome R_{P4}. The asterisk on R_{P2} indicates that chromosome that was previously shown to carry a tooth number QTL (Cleves et al. 2016, in prep). Statistical analysis of each chromosome is reported in Table 3.4. The vertical dashed lines indicate the boundaries of the fine-mapped region in the PAXB x LITC cross. An Ensembl-based gene prediction track from the stickleback genome (chrXXI: 2.56-5.60 Mb) is shown below. Thick lines indicated predicted exons and thin lines indicate introns. *Bmp6* and *Tfap2a* are highlighted in green and yellow, respectively.

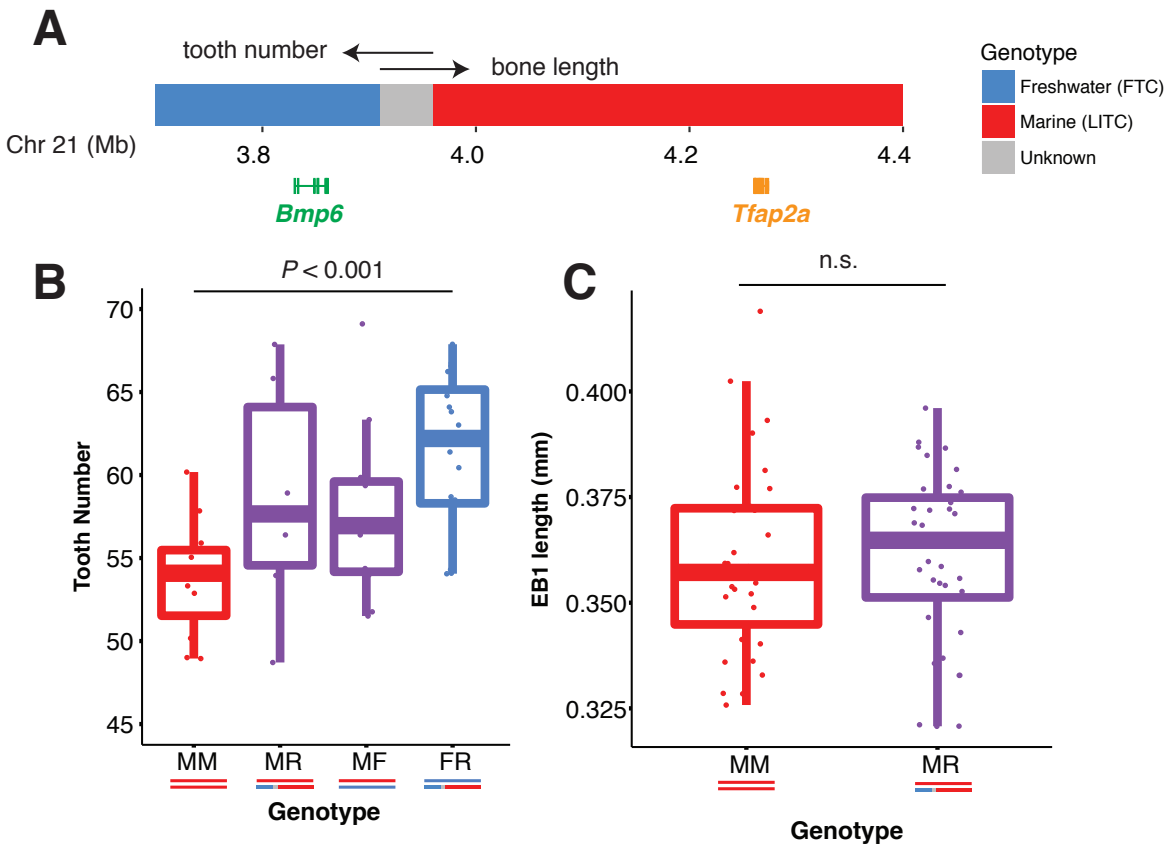


Figure 3.2: A FTC x LITC recombinant chromosome controls tooth number but not bone length.

(A) Depiction of a region of chromosome 21 (3.7-4.4 Mb) indicating recombinant breakpoints, *Bmp6*, and *Tfap2a* locations. (B) This chromosome affects pharyngeal tooth number in a small cross of ~120 dpf juveniles (ANOVA, df = 43, F = 5.785, *P* = 0.002), consistent with previous mapping of a pharyngeal tooth QTL to *Bmp6* (Cleves et al. 2014, 2016). (C) The recombinant chromosome does not control EB1 length at 28 dpf (T test, df = 64.289, t = -0.2287, *P* = 0.82). Bone lengths from Table 3.4 were back-transformed to a 10.5 mm individual. See Table 3.4 for additional data on this chromosome.

name	cross	design	generation	n	clutches	Genotype 1	Genotype 2	Genotype 3	Genotype 4	LR test (acts like M)	LR test (acts like F)	T test (R different from M)	Conclusion
P1	PAXB x LITC	MR x MM	F7	125	3	MM (-4.7 ± 21.2)	MR (5.6 ± 21.1)	-	-	-	-	**0.008**	right of 3.09 Mb
P2	PAXB x LITC	MR x MM	F7	96	2	MM (1.5 ± 15.8)	MR (-1.2 ± 17.7)	-	-	-	-	0.43	right of 3.88 Mb
P3	PAXB x LITC	MR x MF	F6	192	2	MM (-3.4 ± 27.8)	MR (-10.4 ± 25)	MF (2.6 ± 22.2)	FR (9.3 ± 29.0)	**0.04**	0.99	-	left of 4.89 Mb
P4	PAXB x LITC	MR x MF	F6	171	2	MM (-10.4 ± 20.1)	MR (3.1 ± 30.9)	MF (0.4 ± 23.4)	FR (8.6 ± 25.2)	0.59	**0.005**	-	left of 4.4 Mb
P5	PAXB x LITC	FR x MM	F8	62	1	MR (6.9 ± 35.7)	MF (-6.6 ± 48.7)	-	-	-	-	0.21	between 4.15 and 4.44 Mb
P6	PAXB x LITC	MR x MM	F9	78	1	MM (6.3 ± 42.3)	MR (2.8 ± 41.9)	-	-	-	-	0.72	excludes 4.32-4.44 Mb
F1	FTC x LITC	FR x MF	F8	96	2	MR (-11.1 ± 14.6)	MF (-8.4 ± 17.1)	FR (8.4 ± 14.8)	FF (14.0 ± 15.5)	0.22	**0.0002**	-	right of 2.57 Mb
F2	FTC x LITC	MR x MM	F8	76	2	MM (-2.9 ± 17.6)	MR (1.5 ±15.0)	-	-	-		0.25	right of 3.91 Mb
F3	FTC x LITC	FR x MF	F8	63	1	MM (-17.7 ± 15.6)	MR (2.6 ± 20.1)	MF (2.7 ± 19.2)	FR (16.9 ± 20.3)	0.93	**0.0004**	-	left of 5.03 Mb

Table 3.4: Statistical analysis of recombinant chromosomes.

Each row of the table corresponds to one chromosome in Figure 3.1 (in order from top to bottom). All bone length measurements were size-corrected within individual clutches and then pooled. Reported values are mean residual bone lengths (in μm) per genotypic class +/- standard deviation. *P* values for linear model tests or T tests are reported when appropriate; see methods for details on the statistical tests. Arrows in Figure 3.1 indicate the conclusion for each chromosome.

With subsequent fine mapping, we narrowed the QTL in the PAXB cross (Figure 3.1, Table 3.4). Notably, a double recombinant chromosome in the PAXB cross (chromosome R_{P5}) shows no phenotypic difference from the original recombinant chromosome (R_{P4}), which had a strong effect on bone length (Table 3.4), suggesting that the freshwater alleles within the double recombinant region are sufficient to increase bone length. A second double recombinant sharing the same 3' breakpoint (recombinant R_{P6}) did not differ from a marine chromosome, suggesting that this smaller freshwater portion of the chromosome does not carry the QTL. Combined these two results map the bone length QTL to a 155 kb region (chrXXI:4,200,364-4,355,895) in the PAXB cross, containing eight Ensembl-predicted genes in the UCSC stickleback genome browser (*Tfap2a*, *Tmem14b*, *Mak*, *Plcxd2*, *Phldb2*, *Tmem56, ENSGACG00000002373,* and *Bc01*). Seven of these eight genes have no known roles in mouse skeletal development (Smith et al. 2014). However, one gene, *Transcription Factor Activating Protein 2 alpha (Tfap2a)*, is an outstanding candidate for craniofacial evolution because it has roles in both patterning the craniofacial skeleton humans, mice, and zebrafish (Schorle et al. 1996; Zhang et al. 1996; Knight et al. 2005; Milunsky et al. 2008) as well as the growth and maturation of mammalian chondrocytes (reviewed in Wenke and Bosserhoff 2010).

Fine mapping in the FTC cross supported a somewhat larger 1.13 Mb genomic interval (chrXXI: 3,906,104-5,003,790 bp, Table 3.4), also containing *Tfap2a*. The entire PAXB fine-mapped interval is contained within the FTC fine-mapped interval. Therefore, together these data support a model of a shared genomic basis for branchial bone length gain in the two independently derived freshwater populations.

Fractionation of the QTL in the FTC cross

For one FTC recombinant chromosome, we found evidence of fractionation of the bone length QTL. The breakpoint of this recombinant was within a large (~200kb) gene desert between *Bmp6* and *Tfap2a* (Figure 3.3A). We used crosses to compare the chromosome to both marine and freshwater chromosomes that were nonrecombinant within the QTL region (MR x MM and FR x MM). We found that this chromosome produced significantly longer bones than a marine chromosome (Figure 3.3B), however it also produced significantly shorter bones than a freshwater chromosome (Figure 3.3C). Overall, the chromosome conferred bone length phenotypes more similar to a marine chromosome (compare Figure 3.3B to 3.3 C). These data suggest that the effects of the QTL interval were fractionated in this chromosome, with the larger effect portion of the QTL effect mapping to the region containing the *Tfap2a* locus, and the smaller effect portion mapping to the non-coding gene desert downstream of *Tfap2a*.

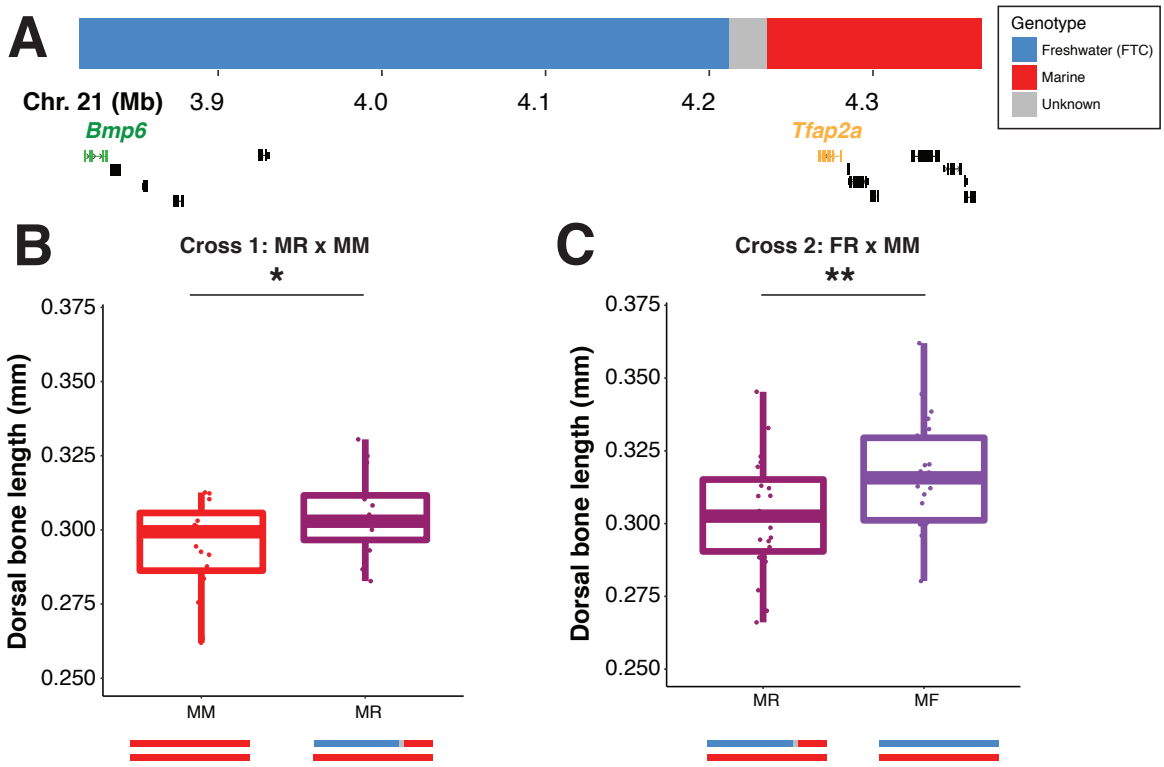


Figure 3.3: Fractionation of the QTL in the FTC x LITC cross.

A single recombinant chromosome with breakpoint between 4,210,854 and 4,231,226 bp (A) was identified and then bred for a generation to produce fish with genotypes MR and FR. The Ensembl gene predictions for this region (chrXXI:3.85-4.35 Mb) are shown below with *Bmp6* and *Tfap2a* highlighted. Each genotype was crossed to an MM fish and bone length (size-corrected and back transformed to a 9 mm fish) was compared between genotypes in each cross. The recombinant chromosome diagrammed in (A) results in bones significantly longer than those from an M chromosome (B, T-test: df = 38.98, t = -2.51, P = 0.017) but significantly shorter than those from an F chromosome (C, T-test: df = 49.13, t = 2.75, P = 0.008). Schematics depicting the genotypes of the offspring are illustrated below the x-axis in B and C. * $P < 0.05$, ** $P < 0.001$.

Chromosome 21 controls bone length immediately after ossification

Our previous findings that cartilage template size was larger and bone growth rate was accelerated in the FTC population (Erickson et al. 2014) raised the question of whether the chromosome 21 QTL controlled cartilage development, bone development, or both. To address this question, we raised closely related F5 fish to 13 days post fertilization (dpf) (EB1 cartilage template formation) and 20 dpf (immediately following EB1 ossification) to test for an effect of the chromosome 21 QTL on cartilage and bone length. The chromosome 21 QTL did not have a significant effect on EB1 cartilage length in an MF x MM backcross at 13 dpf, but the chromosome 21 QTL had a strong effect on bone length by 20 dpf (Figure 3.4). Therefore, the chromosome 21 QTL likely has an effect on either initial bone ossification or bone elongation, but does not appear to have a substantial effect on cartilage length. This result is consistent with the presence of the chromosome 21 QTL in both the FTC and PAXB populations, even though PAXB does not have a significantly larger cartilage template size (Erickson et al. 2014).

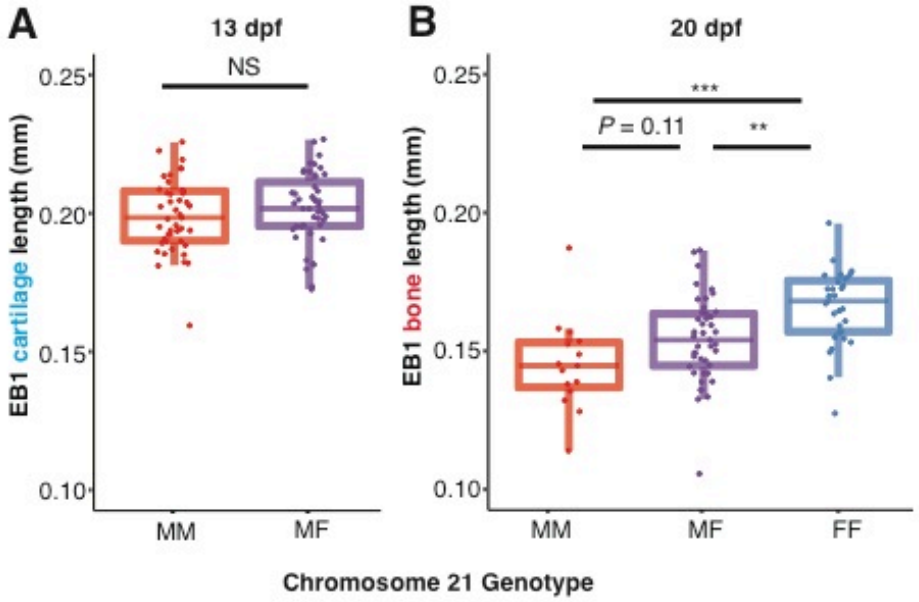


Figure 3.4: The chromosome 21 QTL has effects immediately after bone formation. (A) At 13dpf, chromosome 21 does not control the size of the EB1 template in F5 fish from the FTC x LITC cross (T test, df = 89.7, P = 0.28). (B) By 20 dpf (immediately following initial ossification of EB1), chromosome 21 controls bone length in related F5 fish (ANOVA, df = 86, F = 10.73 P = 6.9 x 10^{-5}). Fish were genotyped for Stn423 and EB measurements were back transformed to fish of total length = 8.5 mm for (a) and standard length = 9 mm for (b).

A coding change in Tfap2a

Fine mapping the QTL in the PAXB cross identified a genomic interval containing coding sequence for eight Ensembl-predicted genes. Using genome resequencing data from a previous study (Glazer et al. 2015), we discovered that the FTC grandparent of that cross was homozygous for a coding change in *Tfap2a*: E237D. Although this change substitutes one acidic amino acid for another, this amino acid position and the surrounding residues are highly conserved across vertebrates (Figure 3.5). We then tested the grandparents of both the FTC x LITC and PAXB x LITC crosses studied here, as well as the grandparents of three benthic x marine crosses (Erickson et al. 2016b), for the coding mutation using a RFLP assay. All five freshwater grandparents were homozygous for the E237D mutation and no marine grandparents carried the mutation. We previously reported a relatively weak EB1 bone length QTL at the far end of chromosome 21 (41.7 cM away from *Tfap2a*) in a LITC x ENOB cross (Erickson et al. 2016b), but in this cross there is no association between *Tfap2a* genotype and EB1 length (Figures 3.6A). These finding suggests that the SNP is not sufficient for the QTL because two of these crosses do not show evidence of a bone length QTL overlapping the *Tfap2a* region despite the presence of the coding change. Additionally, the cross in which the mutation was originally identified, a LITC x FTC cross (Glazer et al. 2014, 2015), also shows no evidence for a bone length QTL near *Tfap2a* (Figure 3.6B). Finally, the recombinant chromosome that fractionated the QTL was marine at this SNP (Figure 3.3A), yet still produced a significant increase in bone length (Figure 3.3B). Therefore, this SNP may contribute to the bone length QTL in these crosses but is not sufficient to produce the entire effect of the QTL.

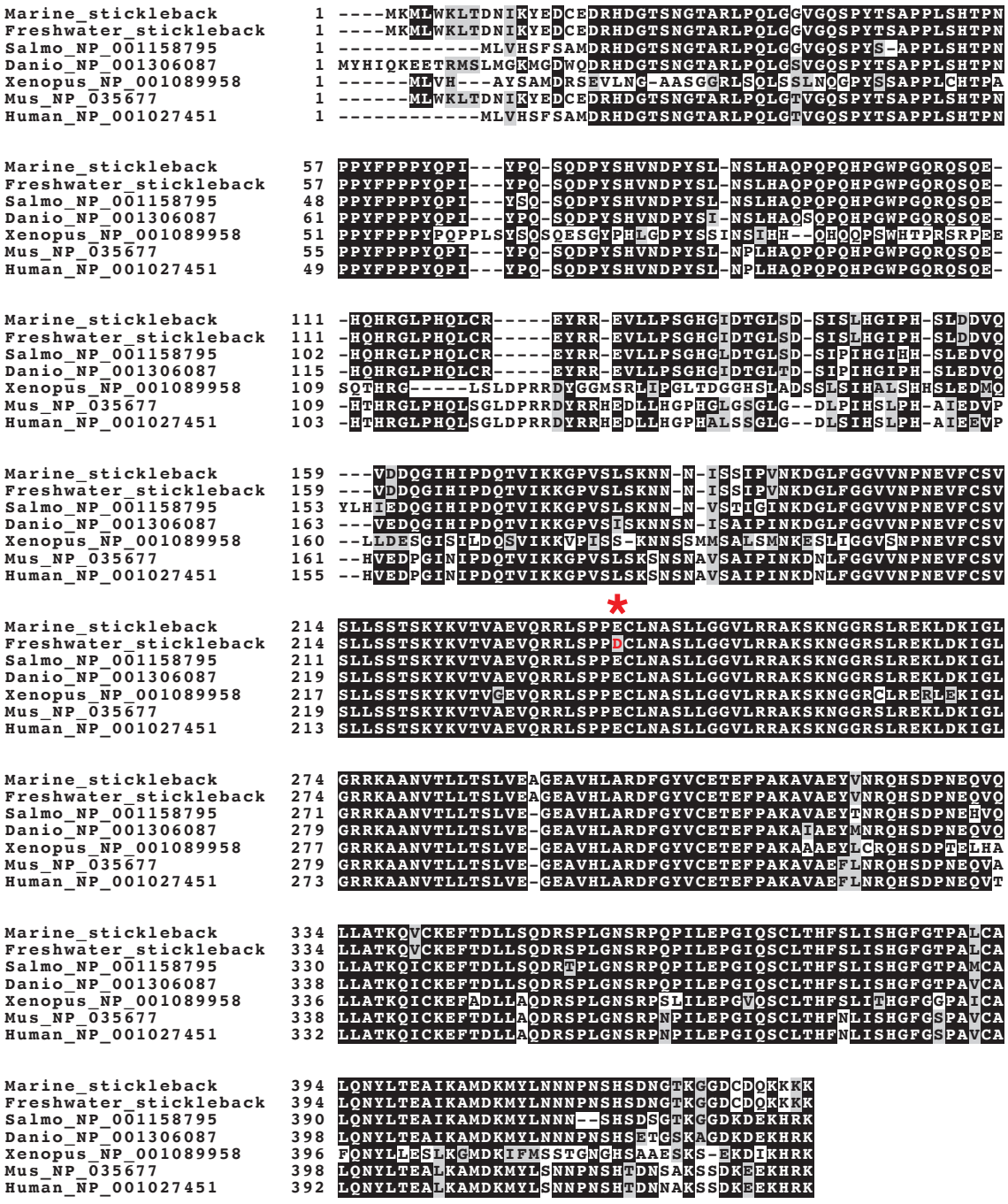


Figure 3.5: Amino acid alignment of marine and freshwater stickleback TFAP2A with other vertebrate sequences.

Stickleback predicted sequence is based on the predicted transcript Ensembl ENSGACT00000003048.1). The RefSeq accession number is given next to each non-stickleback species name. The red asterisk asterisk indicates the location of the stickleback E-to-D polymorphism

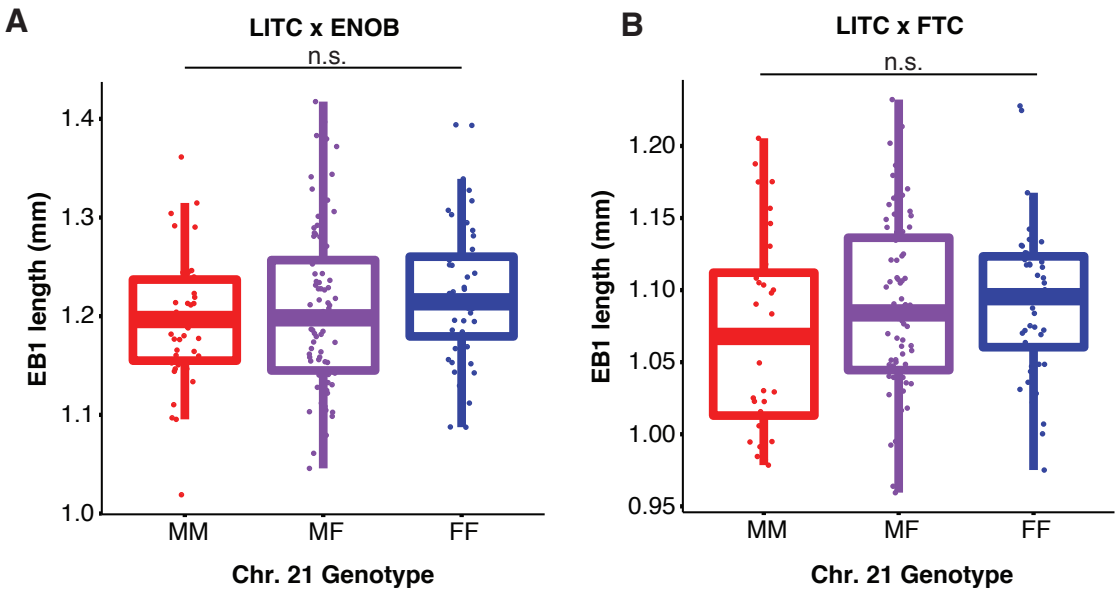


Figure 3.6: Two crosses with a coding mutation in *Tfap2a* lack the chromosome 21 bone length QTL.

(A) Epibranchial 1 bone length was measured in 210 fish from A LITC x FTC cross (Glazer et al 2015). The GBS binned marker containing *Tfap2a* (marker 16_9) was used to test for the presence of a QTL at *Tfap2a*. No significant relationship between genotype and phenotype was observed (ANOVA, df = 153, F = 1.135, *P* = 0.3). (B) An LITC x Enos Benthic cross was reported to have a suggestive bone length QTL near the end of chromosome 21 (Erickson et al 2016). However, the marker containing *Tfap2a* does not have a significant effect on EB1 bone length in this cross (ANOVA; df = 172, F = 1.597, *P* = 0.2).

Expression of Tfap2a in developing branchial skeletons

We previously showed that *Tfap2a* is expressed in undifferentiated mesenchymal cells in the dorsal branchial skeletal primordia prior to epibranchial chondrification but not in developing teeth (Cleves et al. 2014). To further characterize the temporal emergence of this dorsal branchial expression domain, we assayed *Tfap2a* expression at earlier stages in development. At 4 dpf, *Tfap2a* was broadly expressed in the posterior pharyngeal arches (Figure 3.7A, B). By 5 dpf, expression was detected in the dorsal pharyngeal arches (Figure 3.7C). At 7 dpf, as previously reported, expression was detected in undifferentiated mesenchymal cells in the dorsal branchial arches (Figure 3.7D-F). We detected no obvious qualitative differences in expression patterns or levels between marine and freshwater embryos.

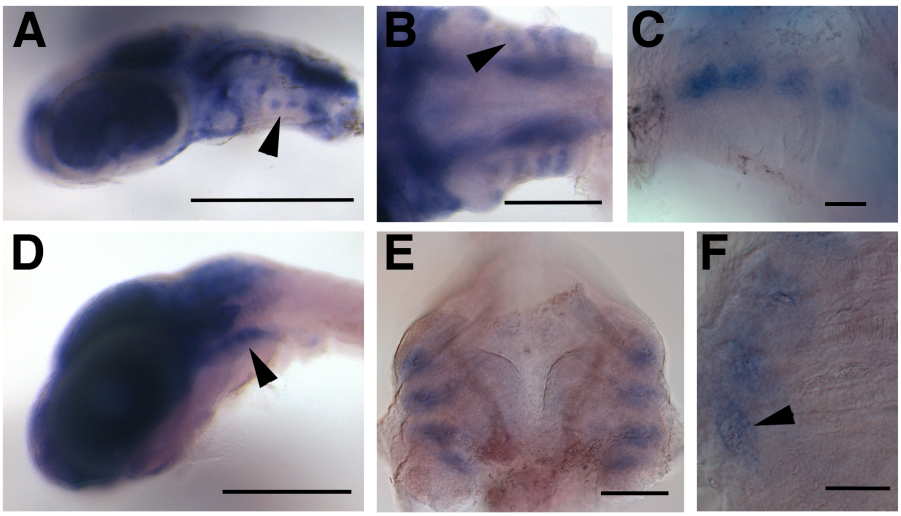


Figure 3.7: In situ hybridization for *Tfap2a*.
(A) Lateral view of a 4 dpf embryonic head. Expression in dorsal pharyngeal arch postmigratory cranial neural crest is indicated with an arrowhead. (B) Ventral view of a 5dpf embryo; arrowheads indicated pharyngeal arch expression. (C) Ventrolateral view of a 5dpf embryo; *Tfap2a* expression is visible at the dorsal ends of the pharyngeal pouches. (D) Lateral view of 7 dpf embryo. *Tfap2a* expression remains in the dorsal pharyngeal arches. (E) Dorsal view of a dissected 7dpf branchial skeleton. (F) Higher magnification view of a 7 dpf branchial skeleton. Arrowhead indicates expression in condensing epibranchial mesenchymal cells. Scale bars: A, D = 500 μm; B = 200 μm; C, E, F = 100 μm.

Allele specific expression of Tfap2a

Since the coding change in *Tfap2a* replaces one acidic amino acid with another, and in two crosses is not sufficient for the presence of a bone length QTL, we also tested for *cis*-regulatory differences in *Tfap2a*. We compared allelic ratios of marine (PAXB) and freshwater (LITC) *Tfap2a* cDNA in developing branchial skeletal tissue to look for allele specific expression (Cowles et al. 2002; Yan et al. 2002; Wittkopp et al. 2004) by adapting a barcoded next-generation sequencing assay originally developed for genotyping by sequencing (Elshire et al. 2011; Glazer et al. 2015). We tested cDNA from a series of developmental stages from the PAXB x LITC cross for ASE of *Tfap2a*. Because *Tfap2a* was broadly expressed in the head during embryonic development (Figure 3.7B) and we could not reliably dissect the branchial skeleton prior to hatching, the earliest stage collected for ASE was hatching (9 dpf) when the branchial skeleton could be accurately dissected. We collected additional time points at 13 dpf, 17 dpf (immediately before ossification of EB1), 22 dpf (immediate following ossification of EB1) and 35 dpf (during growth of EB1). We compared the freshwater:marine allelic ratio of cDNA to that of gDNA controls (which should have a ratio of 1:1 in heterozygous fish). While only one developmental stage showed a significant difference between the gDNA ratio and the cDNA ratio, at

every stage the median expression of the freshwater allele was lower than the marine allele (Figure 3.8), and this ratio was lower in the cDNA samples than in the gDNA controls. When the data from all five stages were pooled, the freshwater: marine allelic ratio was significantly reduced in cDNA samples relative to gDNA controls (P = 0.001, Figure 3.8). These findings suggest that *Tfap2a* might have an inhibitory role for bone development, as reduction of expression of the freshwater allele of *Tfap2a* is associated with longer bones. We also tested for ASE of *Bmp6* in the first three time points. Consistent with previous results that ASE of *Bmp6* arises late in development, (Cleves et al. 2014) we found no evidence for early ASE of *Bmp6* in 9, 13 or 17 dpf branchial skeletons (Figure 3.9, P = 0.19).

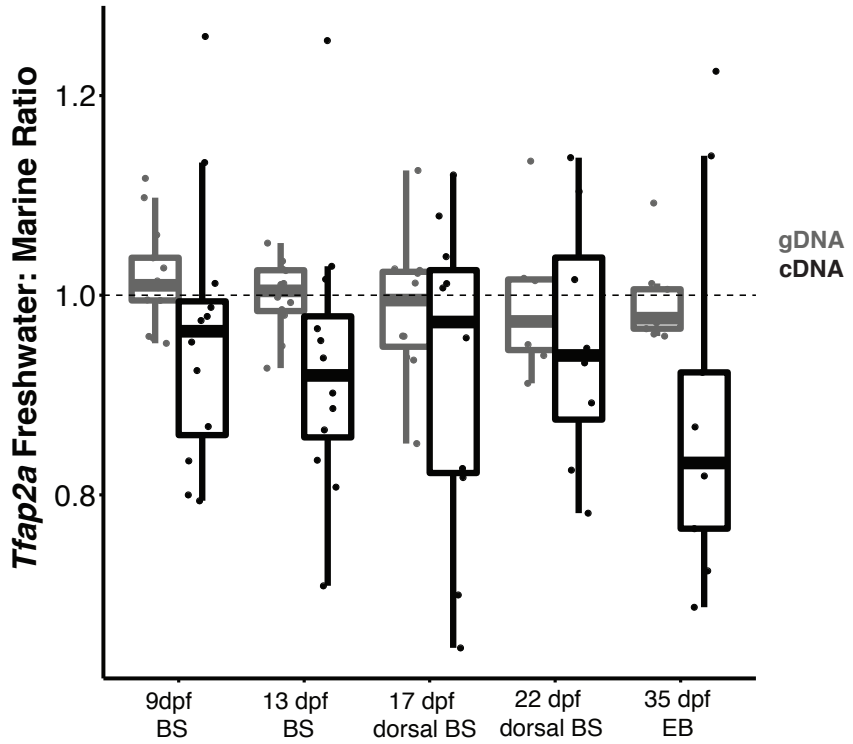


Figure 3.8: Allele specific expression of *Tfap2a*.

A SNP within exon 2 of *Tfap2a* was amplified by PCR from genomic DNA and cDNA collected from approximately 12 individuals at each of 5 stages: 9 and 13 dpf branchial skeleton (BS), 17 and 22 dpf dorsal branchial skeleton, and 35 dpf EB1 tissue. The freshwater: marine allele ratio was calculated for each individual, and the genomic ratio was compared to the cDNA ratio for each time point using a Mann-Whitney U test. While only one developmental stage was significant (9dpf: W = 43, P = 0.06; 13 dpf: W = 35, ***P* = 0.03**; 17 dpf: W = 50, P = 0.51; 22 dpf W = 21, P = 0.46; 35 dpf: W = 18 P = 0.05), collectively the reduced expression of the freshwater allele across the entire data set was highly significant (W = 851, P = 0.001).

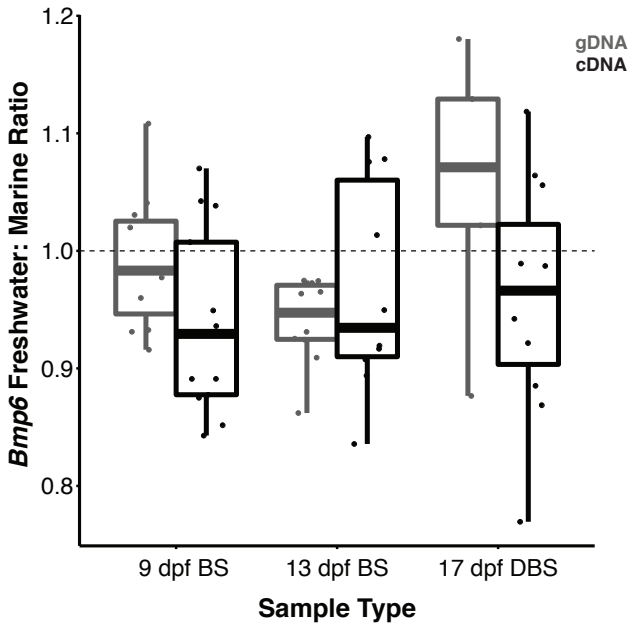


Figure 3.9: No allele specific expression of *Bmp6* during early branchial skeleton development.

A synonymous SNP in *Bmp6* was amplified by PCR from genomic DNA and cDNA collected from approximately 12 individuals at each of 3 stages: 9 and 13 dpf branchial skeleton (BS), and 17 dorsal branchial skeleton. The freshwater: marine allele ratio was calculated for each individual, and the genomic ratio was compared to the cDNA ratio for each time point using a Mann-Whitney U test. No individual time points were significantly different (9 dpf: W = 43, P = 0.11; 13 dpf: W = 51, P = 0.97; 17 dpf: W = 13, P = 0.12), and collectively there was no expression difference across the samples (W = 530, P = 0.19).

Tfap2a dosage affects branchial bone length and craniofacial development

Based on the results of the fine mapping and allele-specific expression studies, we hypothesized that manipulation of *Tfap2a* levels could affect branchial bone length. We used TALENs to induce predicted loss-of-function mutations (see Table 3.3) in *Tfap2a* in the FTC genetic background. We intercrossed stable heterozygous mutants to produce *trans*-heterozygotes. As expected based upon the zebrafish *Tfap2a* mutant phenotype (Holzschuh et al. 2003; Knight et al. 2003), we found that homozygous mutants with a severe lethal craniofacial phenotype (Figure 3.10A-B) were obtained at the expected 25% proportion in each of nine crosses carrying various combinations of alleles in Table 3.3. Homozygous mutant fish were almost entirely lacking a pharyngeal skeleton, while in contrast, the mesodermally derived posterior neurocranium was present and well formed (Figure 3.10E-F). The remnants of the pharyngeal skeletal elements were severely malformed, often asymmetric, and not identifiable (Figure 3.10C-D). The branchial skeleton was particularly hypoplastic, and no identifiable ceratobranchial or epibranchial

cartilages were present. Presence of this phenotype was perfectly concordant with homozygous lesions indicated by loss of a PvuII cut site (see Methods). The craniofacial defects were seen in homozygous mutants in both FTC and LITC genetic backgrounds. Consistent with the known roles of *Tfap2a* in controlling neural crest development and migration (Knight et al. 2003, 2004), we also observed pigmentation defects in homozygous mutants, including reduced melanophore numbers (Figure 5G-H). Like the zebrafish *lockjaw* mutant (Knight et al. 2004), xanthophores appeared unaffected and iridiphores were reduced in the mutant (Figure S6).

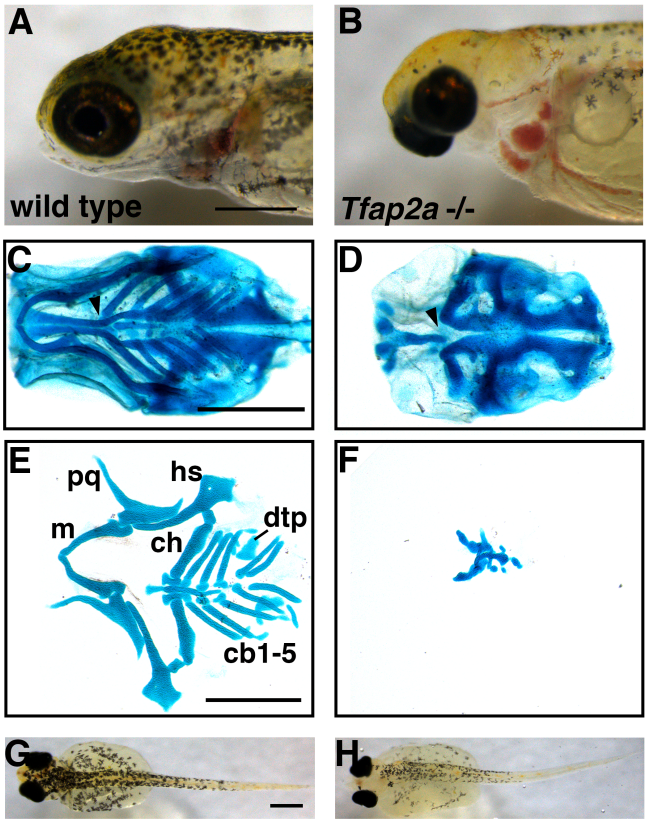


Figure 3.10: Induced homozygous mutations in *Tfap2a* result in severe craniofacial defects and a reduction in trunk pigmentation.

(A, B) Relative to wild-type (A) fish, mutants (B) have severely hypoplastic ventral craniofacial tissue after hatching that results in an inability to feed. (C, D) Ventral view of the heads of Alcian blue stained wild-type (C) and mutant (D) fish indicate defects to the anterior neurocranium in mutant fish (arrowheads). (E,F) Flat-mounted, Alcian blue-stained pharyngeal arch cartilage elements from wild-type (E) and a representative mutant (F) fish. The remnant cartilage elements in the mutants are unidentifiable. (G,H) Dorsal views of 9 dpf wild-type (G) and mutant (H) larvae showing reduced melanophore pigmentation in *Tfap2a* mutants. Anterior is to the left for all images; scale bars = 500 µm. Abbreviations: m = Meckel's cartilage (lower jaw); pq = palatoquadrate; hs = hyosymplectic; ch = ceratohyal; dtp = dorsal tooth plate; cb = ceratobranchial.

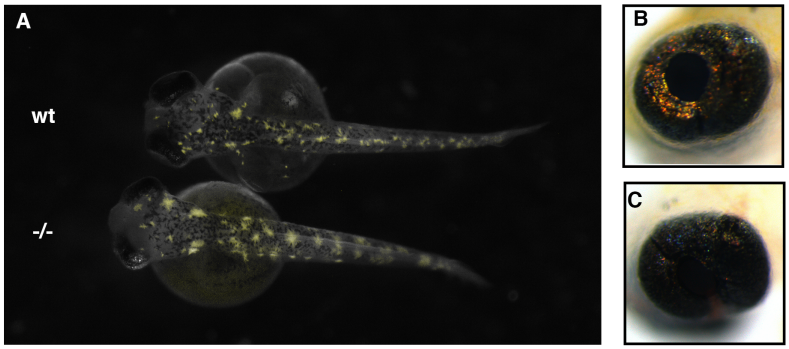


Figure 3.11. Iridiphore but not xanthophore defects in stickleback *Tfap2a* mutants. (A) Xanthophores appear unaffected in *Tfap2a* homozygous mutants. (B, C) Iridiphores were reduced but present in the eyes of mutants (C) relative to wild type (B). Xanthophores were photographed using a GFP filter and iridiphores were imaged under brightfield light. All fish were imaged at 9 dpf.

Because we could not study branchial bone length in homozygous mutants, we outcrossed heterozygous fish to wild-type fish to compare bone length in wild-type and heterozygous siblings. We grew fish to ~4 weeks post fertilization (~10 mm SL, a stage just after the branchial bone length QTL first appears). For the *Tfap2a* deletion on the FTC background, we found that EB1 length was slightly but significantly decreased in heterozygous fish (Figure 3.12A), while surprisingly, the fourth and fifth ceratobranchials (CB4 and CB5) were slightly but significantly increased (Figure 3.13). However, on the LITC background, none of the branchial bones measured (CB1-5 and EB) were significantly different between wild-type and heterozygous fish (Figure 3.12B, Figure 3.13). Although our allele specific expression data predict that partial loss of *Tfap2a* should lead to an increase in branchial bone length, our functional data suggest that halving the dosage of *Tfap2a* on a freshwater genetic background can influence branchial bone length: negatively for dorsal EB1 and positively for CB5.

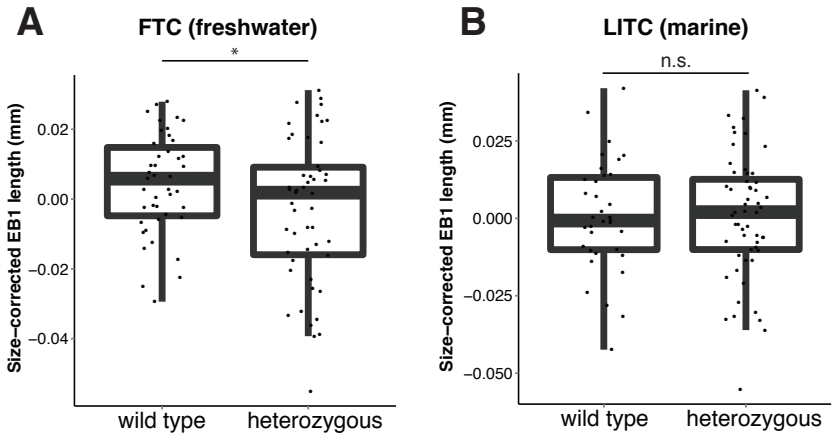


Figure 3.12: Heterozygous loss of *Tfap2a* produces subtle EB1 bone length phenotypes.

FTC fish (25-28 dpf) heterozygous for a *Tfap2a* loss of function mutations have significantly shorter EB1 length relative to wild-type (T-test: t = 1.99, df = 87.1, *P* = 0.049, panel A). LITC heterozygous fish have no significant difference in EB1. Additional bone length phenotypes are shown in Figure 3.13.

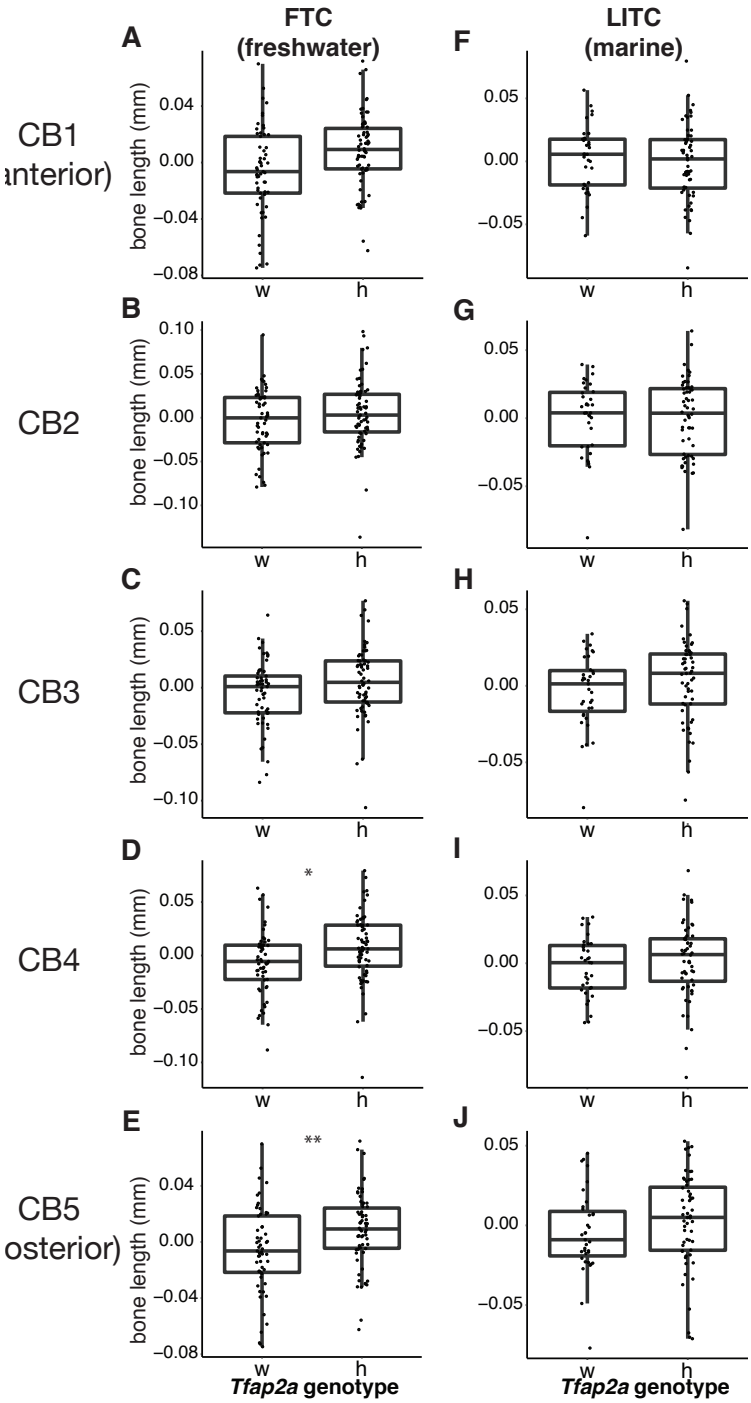


Figure 3.13. Heterozygous *Tfap2a* mutation affects posterior ventral bone length in the FTC background.

(A-E) FTC fish with heterozygous loss of *Tfap2a* have significantly longer ventral posterior bones (CB4 and CB5, T-test $P = 0.014$ (*) and $P = 0.006$ (**), respectively). (F-J) No significant effect of the TALEN on bone length was seen in any ventral bone in the LITC background. w = wild-type; h = heterozygous.

Discussion

Dissection of a supergene

The increasing degree of pleiotropy with increasing organismal complexity has been considered a "cost of complexity" that might slow adaptation (Orr 2000). Putatively pleiotropic QTL affecting traits of evolutionary, economic, and biomedical interest have been discovered (Feitosa et al. 2006; Hall et al. 2006; Ookawa et al. 2010; Stearns 2010; Saatchi et al. 2014), but determining whether the causative loci are truly the same remains a challenge (Wagner and Zhang 2011). Using recombinant mapping in two genetic crosses, we show that two separate but physically close genes, *Bmp6* and *Tfap2a*, form a supergene that increases both the length of the bones and the number of teeth in the stickleback branchial skeleton. In two previously studied large crosses with dense genotyping, these genes were separated by 0.9 and 1.4 cM (Glazer et al. 2015), suggesting that the loci are in tight linkage. The underlying mutations on chromosome 21 causing both evolved bone length gain and tooth gain may have additional pleiotropic effects on other phenotypes, including those previously mapped to chromosome 21 (Miller et al. 2014). However, a simple model in which these similar skeletal gain phenotypes are caused by a single pleiotropic mutation can be ruled out by the data presented here.

In addition to separating the loci controlling bone length and tooth number, we also separate the bone length QTL into at least two fractions in the FTC population. The majority of the phenotypic effect maps nearer to *Tfap2a*, however, a portion of the phenotypic effect is attributable to a large non-coding region between *Tfap2a* and *Bmp6*. QTL fractionation into multiple loci has been observed for a variety of *Drosophila* life history traits (reviewed in Mackay 2004), for several complex phenotypes in tomatoes (Johnson et al. 2012) and for emotional behavior traits in mice (Yalcin et al. 2004; Willis-Owen and Flint 2006). To our knowledge, this is the first demonstration of QTL fractionating for a vertebrate morphological trait.

Tfap2a as a candidate gene for craniofacial evolution

Our mapping results support *Tfap2a* as a candidate for craniofacial evolution in stickleback. *Tfap2a* was initially identified as an activating protein that bound the SV40 viral enhancer to promote transcription (Williams et al. 1988) and has roles in both activating and inhibiting expression of a variety of human genes (Mitchell et al. 1987; Williams and Tjian 1991; Gaubatz et al. 1995; Pfisterer et al. 2002; Eckert et al. 2005). *Tfap2a* is expressed in a variety of tissues, including the neural crest, during development in both mice and zebrafish (Mitchell et al. 1991; Thisse et al. 2001). Further studies linked *Tfap2a* to a critical role as a master regulator of development of the neural crest, (de Crozé et al. 2011; Rada-Iglesias et al. 2012), an embryonic migratory cell population that gives rise to a variety of tissues, including the craniofacial skeleton and pigment cells (Bronner and LeDouarin 2012; Mayor and Theveneau 2013). Mice and zebrafish with loss of function mutations have severe craniofacial defects (Schorle et al. 1996; Zhang et al. 1996; Holzschuh et al. 2003; Knight et al. 2003, 2004) likely due to *Tfap2a*'s role in the pharyngeal ectoderm (Knight et al. 2005). *Tfap2a* is expressed in craniofacial sutures postnatally in mice, and conditional knockout of *Tfap2a* in the frontonasal prominence

results in a failure of craniofacial outgrowth during postnatal development (Nelson and Williams 2004). These findings suggest a role for *Tfap2a* in later stages of craniofacial development. *Tfap2a* is also required for human craniofacial development: familial mutations in *Tfap2a* cause branchio-oculo-facial syndrome (Milunsky et al. 2008; Stoetzel et al. 2009; Tekin et al. 2009).

Evolved differences in neural crest cell regulation and patterning are hypothesized to underlie craniofacial evolution in primates, birds, and cichlids, as well as the domestication of a variety of mammals (Fish et al. 2014; Powder et al. 2014; Wilkins et al. 2014; Prescott et al. 2015). Therefore, one mechanism by which *Tfap2a* might affect stickleback branchial bone length is by affecting the early patterning of the neural crest cells that eventually form the pharyngeal skeleton. Consistent with this possibility, we observed tissue-restricted expression of *Tfap2a* in the dorsal region of developing pharyngeal arches as well as primordia of stickleback branchial bones. While conserved craniofacial regulatory elements of *Tfap2a* have been described (Zhang and Williams 2003; Donner and Williams 2006; Feng et al. 2008), a neural crest-specific enhancer has not been reported. In sticklebacks, *Tfap2a* expression appears restricted to the dorsal branchial skeleton by 5 days post-fertilization. Modification of a regulatory element that affects the dorsal-ventral partitioning of cells within the pharyngeal arches is one mechanism by which *Tfap2a* might regulate dorsal branchial skeleton morphology.

In addition to its roles in craniofacial development, *Tfap2a* also regulates the differentiation and maintenance of cell types important to skeletal development. *Tfap2a* mutant mice and *Tfap2a* mutant chimeric mice have partially penetrant absence of zeugopods and polydactyly (Schorle et al. 1996; Zhang et al. 1996; Nottoli et al. 1998). During mammalian long bone development [analogous to fish branchial bone development (Haines 1942)], an initial cartilage template for the bone is formed. Osteoblasts then secrete bone matrix surrounding this template, but a region of cartilage cells (chondrocytes) is left behind as a growth plate at either end to proliferate and differentiate into additional cartilage template for further bone elongation (Kronenberg 2003; Hall 2005; Karsenty et al. 2009). Furthermore, although the precise mechanism has not been elucidated, *Tfap2a* appears to play an important role(s) in the differentiation and maintenance of chondrocytes in the growth plate. TFAP2A protein is expressed in the growth plate of mouse long bones (Davies et al. 2002), is thought to regulate a variety of genes related to chondrocyte development (Xie et al. 1998; Tuli et al. 2002), and is a negative regulator of chondrocyte differentiation *in vitro* (Huang et al. 2004). Therefore, a downregulation or reduced function of *Tfap2a* in freshwater fish might cause an increase in chondrocyte differentiation, resulting in greater long bone growth. The allele specific expression of *Tfap2a* seen during early bone development, combined with the manifestation of the bone length QTL observed in juveniles, suggests that a combination of its roles in neural crest development as well as chondrocyte differentiation might underlie the evolved bone gain. Furthermore, the fractionation of the QTL in the FTC cross indeed suggests that at least two regions surrounding *Tfap2a* contribute to the final phenotype.

Surprisingly, the effect of *Tfap2a* mutation was more severe in sticklebacks than in zebrafish. While zebrafish *lockjaw* mutants still have identifiable pharyngeal cartilage

elements in the jaw and branchial skeleton (Knight et al. 2003), stickleback trans-heterozygous mutants have only a few severely hypoplastic, unidentifiable pharyngeal skeletal elements, as well as defects in the anterior neurocranium. The existing chondrocytes appear to be well developed in mutant fish, so the main craniofacial defect is in tissue patterning, not cartilage differentiation. Therefore, compared to zebrafish, sticklebacks may have fewer redundant functions of other genes with *Tfap2a* during craniofacial development.

We found that while genome-edited mutations in *Tfap2a* caused lethal craniofacial defects, heterozygous loss of *Tfap2a* was sufficient to alter branchial bone length in the FTC background. Based on our ASE data, we predict that freshwater fish likely have lower levels of *Tfap2a*, so they may be more sensitive to its loss than marine fish. Loss of *Tfap2a* led to a slightly shorter dorsal (EB1) bone, but a longer ventral (CB5) bone. These opposite results suggest that *Tfap2a* could play some role in allocating cells to dorsal and ventral pharyngeal elements, but also suggests that branchial skeleton morphology can be sensitive to *Tfap2a* dosage, similar to the quantitative craniofacial phenotypes reported in heterozygous *Tfap2a* mutant mice (Green et al. 2015). The complete loss of function mutation generated by TALENs likely does not accurately recapitulate the likely *cis*-regulatory mutations in *Tfap2a* that occur in nature, which may affect the timing, location, and/or level of gene expression.

Parallel genetic basis of convergent evolved bone gain

Our data support a model in which *cis* regulatory changes, and perhaps a coding change, to both the FTC and PAXB alleles of *Tfap2a* underlie evolved bone length gain. However, we note that the mapping is less resolved in the FTC population. Our data suggest that the bone length QTL is segregating in the wild FTC population as we have identified one FTC cross with the QTL and one without. Parallel genetic evolution could occur through two mechanisms: repeated selection of a shared ancestral haplotype found in the marine population, or independent mutations that both affect *Tfap2a* regulation in the two freshwater populations. The former model has been shown in a classic example of stickleback evolution, the recurrent fixation of a regulatory allele of *Eda* to reduce armor plating in freshwater (Colosimo et al. 2005; O'Brown et al. 2015). In contrast, the latter model of recurrent regulatory mutations underlies the repeated loss of pelvic spines in multiple freshwater populations (Shapiro et al. 2004; Chan et al. 2010). Identifying the same causative mutations in both freshwater populations would suggest recurrent selection of an ancestral allele. Identifying different mutations would suggest that evolution can achieve similar phenotypic outcomes using distinct but physically close genetic changes.

Although our data suggest that separate mutations underlie branchial bone gain and pharyngeal tooth gain, the mutations causing these two phenotypes could form an adaptive supergene cluster. Close genetic linkage of two alleles affecting trophic morphology could facilitate rapid adaptation to freshwater diets upon colonization of new habitats. After the causative mutation(s) for both QTL have been determined, testing for linkage disequilibrium of these alleles in wild marine populations would test this hypothesis. Additionally, branchial bones are serial homologs of the jaw, and stickleback

jaw phenotypes also map to chromosome 21 (Miller et al. 2014). Thus, testing whether alleles of *Bmp6* and/or *Tfap2a* additionally affect jaw morphology would help to further define the potentially pleiotropic roles of this putative supergene.

Cis-regulatory changes and morphological evolution

Evolution of cis-regulatory modules is thought to be an important general feature of morphological evolution (Stern 2000; Carroll 2008). Because important developmental regulatory genes (such as *Bmp6* and *Tfap2a*) are often expressed in a wide variety of tissues at time points throughout development and are required for normal development, changes to their regulation, rather than changes to their coding sequences, may prevent deleterious pleiotropic effects. Both *Tfap2a* and *Bmp6* are expressed dynamically in multiple tissues during development (Erickson et al. 2015 and this study) and have deleterious and lethal loss-of-function phenotypes (Cleves et al. in prep, and this study). Here, we provide evidence that a subtle regulatory difference in *Tfap2a* is associated with evolved skeletal changes related to dietary adaptation. Although fine-mapping supported an interval containing eight genes, we cannot exclude the possibility that a long distance regulatory element of another gene outside the interval (including *Bmp6*) could underlie at least part of the QTL effect (Montavon et al. 2011; Marinić et al. 2013; Smallwood and Ren 2013; Anderson and Hill 2014). However, the paralogous tightly linked *Tfap2c* and *Bmp7* genes in mice have been found to be in distinct topologically-associating domains (TADs) (Tsujimura et al. 2015).

In addition to evidence for a regulatory change, we also identified a mutation that is predicted to change a highly conserved amino acid position in the freshwater grandparents of both crosses. This mutation is in a predicted DNA binding domain of the protein (Williams and Tjian 1991) and is not found in other vertebrates with sequenced genomes. However, the amino acid change is from glutamate to aspartate, which are both acidic side chains with a similar chemical structure that may not affect function of the protein. As the crystal structure of TFAP2A is not reported, we cannot make further predictions about the effect of this mutation. However, we have two lines of evidence that this coding change does not solely underlie the QTL. First, our QTL fractionation result shows that part of the bone length QTL maps away from this coding change. Second, two additional F2 crosses segregating this coding change do not show any association between *Tfap2a* genotype and bone length phenotype. Although it is formally possible that in these crosses, epistatic interactions mask the effect of the coding change, given the significant *cis*-regulatory change, we favor a model that the coding change has little or no functional effect on the protein that is relevant to bone length.

According to the stickleback genome assembly (Jones et al. 2012), *Tfap2a* and *Bmp6* are separated by a large gene desert that is replete with conserved non-coding sequence. These putative regulatory elements might provide fodder for *cis*-regulatory changes controlling evolved morphological differences. Regulatory changes to other developmental signaling molecules have also been implicated in stickleback evolution (Shapiro et al. 2004; Colosimo et al. 2005; Miller et al. 2007; Chan et al. 2010), suggesting that changes to gene regulation may be used to fine-tune developmental processes to produce novel phenotypes. Interestingly, our studies indicate similar mechanisms for

branchial bone length gain and pharyngeal tooth gain in freshwater stickleback. Both evolved changes are associated with *cis*-regulatory reductions in expression. Combined, these findings suggest that evolved traits may make "more from less": gain traits may commonly use loss of repressors to evolve morphological gain phenotypes.

Acknowledgements

A talented undergraduate, Joan Baek, analyzed many of the mapping crosses shown here, as well as assisted with the TALEN experiments and in situ hybridization. James Hart helped develop the ASE assay and wrote the Python scripts used to analyze the data. Phillip Cleves was involved in generation and analysis of some of the PAXB x LITC recombinants. Craig Miller supervised the study and was involved in writing the manuscript.

We thank Monica Jimenez for assistance in designing the *Tfap2a* TALENs, Alyson Cook for assistance in cartilage phenotyping, and Andrew Glazer for helpful discussions on the ASE assay development. This work was funded in part by NIH R01 #DE021475 (C.M.), NIH Predoctoral Training Grants #5T32GM007127 (P.E.) and # 5T32HG000047-15 (J.H.), and an NSF Graduate Research Fellowships (P.C.). Experiments in this study used the Vincent J. Coates Genomics Sequencing Laboratory at UC Berkeley, supported by NIH S10 Instrumentation Grants

References

Anderson, E., and R. E. Hill. 2014. Long range regulation of the sonic hedgehog gene. Curr. Opin. Genet. Dev. 27:54–59.

Anker, G. C. 1974. Morphology and kinetics of the head of the stickleback, Gasterosteus aculeatus. J. Zool. 32:311–416.

Balic, A., and I. Thesleff. 2015. Chapter Seven - Tissue Interactions Regulating Tooth Development and Renewal. Pp. 157–186 *in* Y. Chai, ed. Current Topics in Developmental Biology. Academic Press.

Bell, M. A., and W. E. Aguirre. 2013. Contemporary evolution, allelic recycling, and adaptive radiation of the threespine stickleback. Evol. Ecol. Res. 15:377–411.

Bell, M. A., and S. A. Foster. 1994. The Evolutionary Biology of the Threespine Stickleback. Oxford University Press.

Bentzen, P., and J. D. McPhail. 1984. Ecology and evolution of sympatric sticklebacks (Gasterosteus): specialization for alternative trophic niches in the Enos Lake species pair. Can. J. Zool. 62:2280–2286.

Bronner, M. E., and N. M. LeDouarin. 2012. Development and evolution of the neural crest: An overview. Dev. Biol. 366:2–9.

Carroll, S. B. 2008. Evo-Devo and an Expanding Evolutionary Synthesis: A Genetic Theory of Morphological Evolution. Cell 134:25–36.

Cermak, T., E. L. Doyle, M. Christian, L. Wang, Y. Zhang, C. Schmidt, J. A. Baller, N. V. Somia, A. J. Bogdanove, and D. F. Voytas. 2011. Efficient design and assembly of custom TALEN and other TAL effector-based constructs for DNA targeting. Nucleic Acids Res. gkr218.

Chan, Y. F., M. E. Marks, F. C. Jones, G. Villarreal, M. D. Shapiro, S. D. Brady, A. M. Southwick, D. M. Absher, J. Grimwood, J. Schmutz, R. M. Myers, D. Petrov, B. Jónsson, D. Schluter, M. A. Bell, and D. M. Kingsley. 2010. Adaptive evolution of pelvic reduction in sticklebacks by recurrent deletion of a Pitx1 enhancer. Science 327:302–305.

Cleves, P. A., N. A. Ellis, M. T. Jimenez, S. M. Nunez, D. Schluter, D. M. Kingsley, and C. T. Miller. 2014. Evolved tooth gain in sticklebacks is associated with a cis-regulatory allele of Bmp6. Proc. Natl. Acad. Sci. 111:13912–13917.

Colosimo, P. F., K. E. Hosemann, S. Balabhadra, G. Villarreal, M. Dickson, J. Grimwood, J. Schmutz, R. M. Myers, D. Schluter, and D. M. Kingsley. 2005. Widespread Parallel Evolution in Sticklebacks by Repeated Fixation of Ectodysplasin Alleles. Science 307:1928–1933.

Conte, G. L., M. E. Arnegard, C. L. Peichel, and D. Schluter. 2012. The probability of genetic parallelism and convergence in natural populations. Proc. R. Soc. Lond. B Biol. Sci. 279:5039–5047.

Cowles, C. R., J. N. Hirschhorn, D. Altshuler, and E. S. Lander. 2002. Detection of regulatory variation in mouse genes. Nat. Genet. 32:432–437.

Davies, S. R., S. Sakano, Y. Zhu, and L. J. Sandell. 2002. Distribution of the transcription factors Sox9, AP-2, and [delta]EF1 in adult murine articular and meniscal cartilage and growth plate. J. Histochem. Cytochem. Off. J. Histochem. Soc. 50:1059–1065.

de Crozé, N., F. Maczkowiak, and A. H. Monsoro-Burq. 2011. Reiterative AP2a activity controls sequential steps in the neural crest gene regulatory network. Proc. Natl. Acad. Sci. U. S. A. 108:155–160.

Donner, A. L., and T. Williams. 2006. Frontal nasal prominence expression driven by Tcfap2a relies on a conserved binding site for STAT proteins. Dev. Dyn. 235:1358–1370.

Doyle, E. L., N. J. Booher, D. S. Standage, D. F. Voytas, V. P. Brendel, J. K. VanDyk, and A. J. Bogdanove. 2012. TAL Effector-Nucleotide Targeter (TALE-NT) 2.0: tools for TAL effector design and target prediction. Nucleic Acids Res. 40:W117–W122.

Eckert, D., S. Buhl, S. Weber, R. Jäger, and H. Schorle. 2005. The AP-2 family of transcription factors. Genome Biol. 6:246.

Ellis, N. A., A. M. Glazer, N. N. Donde, P. A. Cleves, R. M. Agoglia, and C. T. Miller. 2015. Distinct developmental and genetic mechanisms underlie convergently evolved tooth gain in sticklebacks. Development 142:2442–2451.

Elshire, R. J., J. C. Glaubitz, Q. Sun, J. A. Poland, K. Kawamoto, E. S. Buckler, and S. E. Mitchell. 2011. A Robust, Simple Genotyping-by-Sequencing (GBS) Approach for High Diversity Species. PLoS ONE 6:e19379.

Erickson, P. A., P. A. Cleves, N. A. Ellis, K. T. Schwalbach, J. C. Hart, and C. T. Miller. 2015. A 190 base pair, TGF-β responsive tooth and fin enhancer is required for stickleback Bmp6 expression. Dev. Biol. 401:310–323.

Erickson, P. A., N. A. Ellis, and C. T. Miller. 2016a. Microinjection for Transgenesis and Genome Editing in Threespine Sticklebacks. J. Vis. Exp. e54055.

Erickson, P. A., A. M. Glazer, P. A. Cleves, A. S. Smith, and C. T. Miller. 2014. Two developmentally temporal quantitative trait loci underlie convergent evolution of increased branchial bone length in sticklebacks. Proc. R. Soc. Lond. B Biol. Sci. 281:20140822.

Erickson, P. A., A. M. Glazer, E. E. Killingbeck, R. M. Agoglia, J. Baek, S. M. Carsanaro, A. M. Lee, P. A. Cleves, D. Schluter, and C. T. Miller. 2016b. Partially repeatable genetic basis of benthic adaptation in threespine sticklebacks. Evolution 70:887–902.

Farnum, C. E., M. Tinsley, and J. W. Hermanson. 2008a. Forelimb versus hindlimb skeletal development in the big brown bat, Eptesicus fuscus: functional divergence is reflected in chondrocytic performance in Autopodial growth plates. Cells Tissues Organs 187:35–47.

Farnum, C. E., M. Tinsley, and J. W. Hermanson. 2008b. Postnatal bone elongation of the manus versus pes: analysis of the chondrocytic differentiation cascade in Mus musculus and Eptesicus fuscus. Cells Tissues Organs 187:48–58.

Feitosa, M. F., T. Rice, K. E. North, A. Kraja, T. Rankinen, A. S. Leon, J. S. Skinner, J. Blangero, C. Bouchard, and D. C. Rao. 2006. Pleiotropic QTL on chromosome 19q13 for triglycerides and adiposity: The HERITAGE family study. Atherosclerosis 185:426–432.

Feng, W., J. Huang, J. Zhang, and T. Williams. 2008. Identification and Analysis of a Conserved Tcfap2a Intronic Enhancer Element Required for Expression in Facial and Limb Bud Mesenchyme. Mol. Cell. Biol. 28:315–325.

Fish, J. L., R. S. Sklar, K. C. Woronowicz, and R. A. Schneider. 2014. Multiple developmental mechanisms regulate species-specific jaw size. Dev. Camb. Engl. 141:674–684.

Frankel, N., D. F. Erezyilmaz, A. P. McGregor, S. Wang, F. Payre, and D. L. Stern. 2011. Morphological evolution caused by many subtle-effect substitutions in regulatory DNA. Nature 474:598–603.

Gaubatz, S., A. Imhof, R. Dosch, O. Werner, P. Mitchell, R. Buettner, and M. Eilers. 1995. Transcriptional activation by Myc is under negative control by the transcription factor AP-2. EMBO J. 14:1508–1519.

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 g3.115.017905.

Green, M. R., and J. Sambrook. 2012. Molecular cloning : a laboratory manual. Cold Spring Harbor, N.Y. : Cold Spring Harbor Laboratory Press, c2012.

Green, R. M., W. Feng, T. Phang, J. L. Fish, H. Li, R. A. Spritz, R. S. Marcucio, J. Hooper, H. Jamniczky, B. Hallgrímsson, and T. Williams. 2015. Tfap2a-dependent changes in mouse facial morphology result in clefting that can be ameliorated by a reduction in Fgf8 gene dosage. Dis. Model. Mech. 8:31–43.

Gross, H. P., and J. M. Anderson. 1984. Geographic Variation in the Gillrakers and Diet of European Threespine Sticklebacks, Gasterosteus aculeatus. Copeia 1984:87–97.

Haines, R. W. 1934. Epiphyseal growth in the branchial skeleton of fishes. Q. J. Microsc. Sci. 77:77–97.

Haines, R. W. 1942. The Evolution of Epiphyses and of Endochondral Bone. Biol. Rev. 17:267–292.

Hall, B. K. 2005. Bones and Cartilage: Developmental and Evolutionary Skeletal Biology. Academic Press.

Hall, M. C., C. J. Basten, and J. H. Willis. 2006. Pleiotropic Quantitative Trait Loci Contribute to Population Divergence in Traits Associated With Life-History Variation in Mimulus guttatus. Genetics 172:1829–1844.

Hermann, K., U. Klahre, M. Moser, H. Sheehan, T. Mandel, and C. Kuhlemeier. 2013. Tight Genetic Linkage of Prezygotic Barrier Loci Creates a Multifunctional Speciation Island in Petunia. Curr. Biol. 23:873–877.

Holzschuh, J., A. Barrallo-Gimeno, A.-K. Ettl, K. Durr, E. W. Knapik, and W. Driever. 2003. Noradrenergic neurons in the zebrafish hindbrain are induced by retinoic acid and

require tfap2a for expression of the neurotransmitter phenotype. Dev. Camb. Engl. 130:5741–5754.

Huang, Z., H. Xu, and L. Sandell. 2004. Negative regulation of chondrocyte differentiation by transcription factor AP-2alpha. J. Bone Miner. Res. Off. J. Am. Soc. Bone Miner. Res. 19:245–255.

Hung, H.-Y., L. M. Shannon, F. Tian, P. J. Bradbury, C. Chen, S. A. Flint-Garcia, M. D. McMullen, D. Ware, E. S. Buckler, J. F. Doebley, and J. B. Holland. 2012. ZmCCT and the genetic basis of day-length adaptation underlying the postdomestication spread of maize. Proc. Natl. Acad. Sci. U. S. A. 109:E1913-1921.

Jeong, S., M. Rebeiz, P. Andolfatto, T. Werner, J. True, and S. B. Carroll. 2008. The Evolution of Gene Regulation Underlies a Morphological Difference between Two Drosophila Sister Species. Cell 132:783–793.

Johnson, E. B., J. E. Haggard, and D. A. St.Clair. 2012. Fractionation, Stability, and Isolate-Specificity of QTL for Resistance to Phytophthora infestans in Cultivated Tomato (Solanum lycopersicum). G3 2:1145–1159.

Jones, F. C., M. G. Grabherr, Y. F. Chan, P. Russell, E. Mauceli, J. Johnson, R. Swofford, M. Pirun, M. C. Zody, S. White, E. Birney, S. Searle, J. Schmutz, J. Grimwood, M. C. Dickson, R. M. Myers, C. T. Miller, B. R. Summers, A. K. Knecht, S. D. Brady, H. Zhang, A. A. Pollen, T. Howes, C. Amemiya, Broad Institute Genome Sequencing Platform & Whole Genome Assembly Team, E. S. Lander, F. Di Palma, K. Lindblad-Toh, and D. M. Kingsley. 2012. The genomic basis of adaptive evolution in threespine sticklebacks. Nature 484:55–61.

Joron, M., L. Frezal, R. T. Jones, N. L. Chamberlain, S. F. Lee, C. R. Haag, A. Whibley, M. Becuwe, S. W. Baxter, L. Ferguson, P. A. Wilkinson, C. Salazar, C. Davidson, R. Clark, M. A. Quail, H. Beasley, R. Glithero, C. Lloyd, S. Sims, M. C. Jones, J. Rogers, C. D. Jiggins, and R. H. ffrench-Constant. 2011. Chromosomal rearrangements maintain a polymorphic supergene controlling butterfly mimicry. Nature 477:203–206.

Joron, M., R. Papa, M. Beltrán, N. Chamberlain, J. Mavárez, S. Baxter, M. Abanto, E. Bermingham, S. J. Humphray, J. Rogers, H. Beasley, K. Barlow, R. H. ffrench-Constant, J. Mallet, W. O. McMillan, and C. D. Jiggins. 2006. A Conserved Supergene Locus Controls Colour Pattern Diversity in Heliconius Butterflies. PLOS Biol 4:e303.

Karsenty, G., H. M. Kronenberg, and C. Settembre. 2009. Genetic Control of Bone Formation. Annu. Rev. Cell Dev. Biol. 25:629–648.

Kimmel, C. B., C. T. Miller, G. Kruze, B. Ullmann, R. A. BreMiller, K. D. Larison, and H. C. Snyder. 1998. The shaping of pharyngeal cartilages during early development of the zebrafish. Dev. Biol. 203:245–263.

Kislalioglu, M., and R. N. Gibson. 1977. The feeding relationship of shallow water fishes in a Scottish sea loch. J. Fish Biol. 11:257–266.

Knight, R. D., Y. Javidan, S. Nelson, T. Zhang, and T. Schilling. 2004. Skeletal and pigment cell defects in the lockjaw mutant reveal multiple roles for zebrafish tfap2a in neural crest development. Dev. Dyn. 229:87–98.

Knight, R. D., Y. Javidan, T. Zhang, S. Nelson, and T. F. Schilling. 2005. AP2-dependent signals from the ectoderm regulate craniofacial development in the zebrafish embryo. Development 132:3127–3138.

Knight, R. D., S. Nair, S. S. Nelson, A. Afshar, Y. Javidan, R. Geisler, G.-J. Rauch, and T. F. Schilling. 2003. lockjaw encodes a zebrafish tfap2a required for early neural crest development. Dev. Camb. Engl. 130:5755–5768.

Kronenberg, H. M. 2003. Developmental regulation of the growth plate. Nature 423:332–336.

Kunte, K., W. Zhang, A. Tenger-Trolander, D. H. Palmer, A. Martin, R. D. Reed, S. P. Mullen, and M. R. Kronforst. 2014. doublesex is a mimicry supergene. Nature 507:229–232.

Küpper, C., M. Stocks, J. E. Risse, N. dos Remedios, L. L. Farrell, S. B. McRae, T. C. Morgan, N. Karlionova, P. Pinchuk, Y. I. Verkuil, A. S. Kitaysky, J. C. Wingfield, T. Piersma, K. Zeng, J. Slate, M. Blaxter, D. B. Lank, and T. Burke. 2016. A supergene determines highly divergent male reproductive morphs in the ruff. Nat. Genet. 48:79–83.

Lamichhaney, S., G. Fan, F. Widemo, U. Gunnarsson, D. S. Thalmann, M. P. Hoeppner, S. Kerje, U. Gustafson, C. Shi, H. Zhang, W. Chen, X. Liang, L. Huang, J. Wang, E. Liang, Q. Wu, S. M.-Y. Lee, X. Xu, J. Höglund, X. Liu, and L. Andersson. 2016. Structural genomic changes underlie alternative reproductive strategies in the ruff (Philomachus pugnax). Nat. Genet. 48:84–88.

Lemmon, Z. H., and J. F. Doebley. 2014. Genetic dissection of a genomic region with pleiotropic effects on domestication traits in maize reveals multiple linked QTL. Genetics 198:345–353.

Linnen, C. R., Y.-P. Poh, B. K. Peterson, R. D. H. Barrett, J. G. Larson, J. D. Jensen, and H. E. Hoekstra. 2013. Adaptive Evolution of Multiple Traits Through Multiple Mutations at a Single Gene. Science 339:1312–1316.

Lowry, D. B., and J. H. Willis. 2010. A Widespread Chromosomal Inversion Polymorphism Contributes to a Major Life-History Transition, Local Adaptation, and Reproductive Isolation. PLoS Biol 8:e1000500.

Mackay, T. F. 2004. The genetic architecture of quantitative traits: lessons from Drosophila. Curr. Opin. Genet. Dev. 14:253–257.

Marinić, M., T. Aktas, S. Ruf, and F. Spitz. 2013. An Integrated Holo-Enhancer Unit Defines Tissue and Gene Specificity of the Fgf8 Regulatory Landscape. Dev. Cell 24:530–542.

Mather, K. 1950. The Genetical Architecture of Heterostyly in Primula sinensis. Evolution 4:340–352.

Mayor, R., and E. Theveneau. 2013. The neural crest. Development 140:2247–2251.

McGee, M. D., D. Schluter, and P. C. Wainwright. 2013. Functional basis of ecological divergence in sympatric stickleback. BMC Evol. Biol. 13:277.

McGee, M. D., and P. C. Wainwright. 2013. Convergent Evolution as a Generator of Phenotypic Diversity in Threespine Stickleback. Evolution 67:1204–1208.

McGregor, A. P., V. Orgogozo, I. Delon, J. Zanet, D. G. Srinivasan, F. Payre, and D. L. Stern. 2007. Morphological evolution through multiple cis-regulatory mutations at a single gene. Nature 448:587–590.

Miller, C. T., S. Beleza, A. A. Pollen, D. Schluter, R. A. Kittles, M. D. Shriver, and D. M. Kingsley. 2007. cis-Regulatory Changes in Kit Ligand Expression and Parallel Evolution of Pigmentation in Sticklebacks and Humans. Cell 131:1179–1189.

Miller, C. T., A. M. Glazer, B. R. Summers, B. K. Blackman, A. R. Norman, M. D. Shapiro, B. L. Cole, C. L. Peichel, D. Schluter, and D. M. Kingsley. 2014. Modular skeletal evolution in sticklebacks is controlled by additive and clustered quantitative trait Loci. Genetics 197:405–420.

Milunsky, J. M., T. A. Maher, G. Zhao, A. E. Roberts, H. J. Stalker, R. T. Zori, M. N. Burch, M. Clemens, J. B. Mulliken, R. Smith, and A. E. Lin. 2008. TFAP2A Mutations Result in Branchio-Oculo-Facial Syndrome. Am. J. Hum. Genet. 82:1171–1177.

Mitchell, P. J., P. M. Timmons, J. M. Hébert, P. W. Rigby, and R. Tjian. 1991. Transcription factor AP-2 is expressed in neural crest cell lineages during mouse embryogenesis. Genes Dev. 5:105–119.

Mitchell, P. J., C. Wang, and R. Tjian. 1987. Positive and negative regulation of transcription in vitro: enhancer-binding protein AP-2 is inhibited by SV40 T antigen. Cell 50:847–861.

Montavon, T., N. Soshnikova, B. Mascrez, E. Joye, L. Thevenet, E. Splinter, W. de Laat, F. Spitz, and D. Duboule. 2011. A regulatory archipelago controls Hox genes transcription in digits. Cell 147:1132–1145.

Murray, J., and B. Clarke. 1976a. Supergenes in polymorphic land snails. I. Partula taeniata. Heredity 37:253–269.

Murray, J., and B. Clarke. 1976b. Supergenes in polymorphic land snails. II. Partula suturalis. Heredity 37:271–282.

Nabours, R. K. 1933. Inheritance of Color Patterns in the Grouse Locust ACRYDIUM ARENOSUM Burmeister (Tettigidae). Genetics 18:159–171.

Nelson, D. K., and T. Williams. 2004. Frontonasal process-specific disruption of AP-2alpha results in postnatal midfacial hypoplasia, vascular anomalies, and nasal cavity defects. Dev. Biol. 267:72–92.

Nottoli, T., S. Hagopian-Donaldson, J. Zhang, A. Perkins, and T. Williams. 1998. AP-2-null cells disrupt morphogenesis of the eye, face, and limbs in chimeric mice. Proc. Natl. Acad. Sci. 95:13714–13719.

O'Brown, N. M., B. R. Summers, F. C. Jones, S. D. Brady, and D. M. Kingsley. 2015. A recurrent regulatory change underlying altered expression and Wnt response of the stickleback armor plates gene EDA. eLife e05290.

Ookawa, T., T. Hobo, M. Yano, K. Murata, T. Ando, H. Miura, K. Asano, Y. Ochiai, M. Ikeda, R. Nishitani, T. Ebitani, H. Ozaki, E. R. Angeles, T. Hirasawa, and M. Matsuoka. 2010. New approach for rice improvement using a pleiotropic QTL gene for lodging resistance and yield. Nat. Commun. 1:132.

Orr, H. A. 2000. Adaptation and the cost of complexity. Evolution 54:13–20.

Orr, H. A. 2005. The probability of parallel evolution. Evolution 59:216–220.

Peichel, C. L., K. S. Nereng, K. A. Ohgi, B. L. E. Cole, P. F. Colosimo, C. A. Buerkle, D. Schluter, and D. M. Kingsley. 2001. The genetic architecture of divergence between threespine stickleback species. Nature 414:901–905.

Pfisterer, P., J. Ehlermann, M. Hegen, and H. Schorle. 2002. A Subtractive Gene Expression Screen Suggests a Role of Transcription Factor AP-2α in Control of Proliferation and Differentiation. J. Biol. Chem. 277:6637–6644.

Powder, K. E., H. Cousin, G. P. McLinden, and R. C. Albertson. 2014. A Nonsynonymous Mutation in the Transcriptional Regulator lbh Is Associated with Cichlid Craniofacial Adaptation and Neural Crest Cell Development. Mol. Biol. Evol. 31:3113–3124.

Prescott, S. L., R. Srinivasan, M. C. Marchetto, I. Grishina, I. Narvaiza, L. Selleri, F. H. Gage, T. Swigut, and J. Wysocka. 2015. Enhancer Divergence and cis-Regulatory Evolution in the Human and Chimp Neural Crest. Cell 163:68–83.

Pruitt, K. D., G. R. Brown, S. M. Hiatt, F. Thibaud-Nissen, A. Astashyn, O. Ermolaeva, C. M. Farrell, J. Hart, M. J. Landrum, K. M. McGarvey, M. R. Murphy, N. A. O'Leary, S. Pujar, B. Rajput, S. H. Rangwala, L. D. Riddick, A. Shkeda, H. Sun, P. Tamez, R. E. Tully, C.

Wallin, D. Webb, J. Weber, W. Wu, M. DiCuccio, P. Kitts, D. R. Maglott, T. D. Murphy, and J. M. Ostell. 2014. RefSeq: an update on mammalian reference sequences. Nucleic Acids Res. 42:D756-763.

Rada-Iglesias, A., R. Bajpai, S. Prescott, S. A. Brugmann, T. Swigut, and J. Wysocka. 2012. Epigenomic Annotation of Enhancers Predicts Transcriptional Regulators of Human Neural Crest. Cell Stem Cell 11:633–648.

Rohland, N., and D. Reich. 2012. Cost-effective, high-throughput DNA sequencing libraries for multiplexed target capture. Genome Res. gr.128124.111.

Rosenblum, E. B., C. E. Parent, and E. E. Brandt. 2014. The Molecular Basis of Phenotypic Convergence. Annu. Rev. Ecol. Evol. Syst. 45:203–226.

Saatchi, M., R. D. Schnabel, J. F. Taylor, and D. J. Garrick. 2014. Large-effect pleiotropic or closely linked QTL segregate within and across ten US cattle breeds. BMC Genomics 15:442.

Salazar, V. S., L. W. Gamer, and V. Rosen. 2016. BMP signalling in skeletal development, disease and repair. Nat. Rev. Endocrinol. 12:203–221.

Sanger, T. J., E. A. Norgard, L. S. Pletscher, M. Bevilacqua, V. R. Brooks, L. J. Sandell, and J. M. Cheverud. 2011. Developmental and genetic origins of murine long bone length variation. J. Exp. Zoolog. B Mol. Dev. Evol. 316B:146–161.

Sanger, T. J., L. J. Revell, J. J. Gibson-Brown, and J. B. Losos. 2012. Repeated modification of early limb morphogenesis programmes underlies the convergence of relative limb length in Anolis lizards. Proc. R. Soc. B Biol. Sci. 279:739–748.

Schluter, D., and G. L. Conte. 2009. Genetics and ecological speciation. Proc. Natl. Acad. Sci. 106:9955–9962.

Schluter, D., and J. D. McPhail. 1992. Ecological Character Displacement and Speciation in Sticklebacks. Am. Nat. 140:85–108.

Schorle, H., P. Meier, M. Buchert, R. Jaenisch, and P. J. Mitchell. 1996. Transcription factor AP-2 essential for cranial closure and craniofacial development. Nature 381:235–238.

Schwander, T., R. Libbrecht, and L. Keller. 2014. Supergenes and Complex Phenotypes. Curr. Biol. 24:R288–R294.

Shapiro, M. D., M. E. Marks, C. L. Peichel, B. K. Blackman, K. S. Nereng, B. Jónsson, D. Schluter, and D. M. Kingsley. 2004. Genetic and developmental basis of evolutionary pelvic reduction in threespine sticklebacks. Nature 428:717–723.

Sievers, F., A. Wilm, D. Dineen, T. J. Gibson, K. Karplus, W. Li, R. Lopez, H. McWilliam, M. Remmert, J. Söding, J. D. Thompson, and D. G. Higgins. 2011. Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. Mol. Syst. Biol. 7:n/a-n/a.

Smallwood, A., and B. Ren. 2013. Genome organization and long-range regulation of gene expression by enhancers. Curr. Opin. Cell Biol. 25:387–394.

Smith, C. M., J. H. Finger, T. F. Hayamizu, I. J. McCright, J. Xu, J. Berghout, J. Campbell, L. E. Corbani, K. L. Forthofer, P. J. Frost, D. Miers, D. R. Shaw, K. R. Stone, J. T. Eppig, J. A. Kadin, J. E. Richardson, and M. Ringwald. 2014. The mouse Gene Expression Database (GXD): 2014 update. Nucleic Acids Res. 42:D818–D824.

Stearns, F. W. 2010. One Hundred Years of Pleiotropy: A Retrospective. Genetics 186:767–773.

Steiner, C. C., J. N. Weber, and H. E. Hoekstra. 2007. Adaptive variation in beach mice produced by two interacting pigmentation genes. PLoS Biol. 5:e219.

Stern, D. L. 2000. Perspective: Evolutionary Developmental Biology and the Problem of Variation. Evolution 54:1079–1091.

Stern, D. L. 2013. The genetic causes of convergent evolution. Nat. Rev. Genet. 14:751–764.

Stern, D. L., and V. Orgogozo. 2008. The Loci of Evolution: How Predictable Is Genetic Evolution? Evolution 62:2155–2177.

Stoetzel, C., S. Riehm, V. Bennouna Greene, V. Pelletier, J. Vigneron, B. Leheup, V. Marion, S. Hellé, J. m. Danse, C. Thibault, L. Moulinier, F. Veillon, and H. Dollfus. 2009. Confirmation of TFAP2A gene involvement in branchio-oculo-facial syndrome (BOFS) and report of temporal bone anomalies. Am. J. Med. Genet. A. 149A:2141–2146.

Studer, A. J., and J. F. Doebley. 2011. Do large effect QTL fractionate? A case study at the maize domestication QTL teosinte branched1. Genetics 188:673–681.

Tekin, M., A. Sırmacı, B. Yüksel-Konuk, S. Fitoz, and L. Sennaroğlu. 2009. A complex TFAP2A allele is associated with branchio-oculo-facial syndrome and inner ear malformation in a deaf child. Am. J. Med. Genet. A. 149A:427–430.

Thisse, B., S. Pfumio, M. Furthauer, B. Loppin, V. Heyer, A. Degrave, R. Woehl, A. Lux, T. Steffan, X. Charbonnier, and C. Thisse. 2001. Expression of the zebrafish genome during embryogenesis. ZFIN -Line Publ.

Thomas, J. W., M. Cáceres, J. J. Lowman, C. B. Morehouse, M. E. Short, E. L. Baldwin, D. L. Maney, and C. L. Martin. 2008. The Chromosomal Polymorphism Linked to Variation

in Social Behavior in the White-Throated Sparrow (Zonotrichia albicollis) Is a Complex Rearrangement and Suppressor of Recombination. Genetics 179:1455–1468.

Thompson, M. J., and C. D. Jiggins. 2014. Supergenes and their role in evolution. Heredity 113:1–8.

Tsujimura, T., F. A. Klein, K. Langenfeld, J. Glaser, W. Huber, and F. Spitz. 2015. A Discrete Transition Zone Organizes the Topological and Regulatory Autonomy of the Adjacent Tfap2c and Bmp7 Genes. PLOS Genet 11:e1004897.

Tuli, R., M. R. Seghatoleslami, S. Tuli, M. S. Howard, K. G. Danielson, and R. S. Tuan. 2002. p38 MAP Kinase Regulation of AP-2 Binding in TGF-β1-Stimulated Chondrogenesis of Human Trabecular Bone-Derived Cells. Ann. N. Y. Acad. Sci. 961:172–177.

Tuttle, E. M., A. O. Bergland, M. L. Korody, M. S. Brewer, D. J. Newhouse, P. Minx, M. Stager, A. Betuel, Z. A. Cheviron, W. C. Warren, R. A. Gonser, and C. N. Balakrishnan. 2016. Divergence and Functional Degradation of a Sex Chromosome-like Supergene. Curr. Biol. 0.

Wagner, G. P., and J. Zhang. 2011. The pleiotropic structure of the genotype–phenotype map: the evolvability of complex organisms. Nat. Rev. Genet. 12:204–213.

Wainwright, P. C. 2006. Functional Morphology of the Pharyngeal Jaw Apparatus. P. *in* Fish Physiology: Fish Biomechanics. Academic Press.

Wang, H., T. Nussbaum-Wagler, B. Li, Q. Zhao, Y. Vigouroux, M. Faller, K. Bomblies, L. Lukens, and J. F. Doebley. 2005. The origin of the naked grains of maize. Nature 436:714–719.

Wenke, A.-K., and A. K. Bosserhoff. 2010. Roles of AP-2 transcription factors in the regulation of cartilage and skeletal development. FEBS J. 277:894–902.

Wilkins, A. S., R. W. Wrangham, and W. T. Fitch. 2014. The “Domestication Syndrome” in Mammals: A Unified Explanation Based on Neural Crest Cell Behavior and Genetics. Genetics 197:795–808.

Williams, T., A. Admon, B. Lüscher, and R. Tjian. 1988. Cloning and expression of AP-2, a cell-type-specific transcription factor that activates inducible enhancer elements. Genes Dev. 2:1557–1569.

Williams, T., and R. Tjian. 1991. Analysis of the DNA-binding and activation properties of the human transcription factor AP-2. Genes Dev. 5:670–682.

Willis-Owen, S. A. G., and J. Flint. 2006. The genetic basis of emotional behaviour in mice. Eur. J. Hum. Genet. 14:721–728.

Wills, D. M., C. J. Whipple, S. Takuno, L. E. Kursel, L. M. Shannon, J. Ross-Ibarra, and J. F. Doebley. 2013. From many, one: genetic control of prolificacy during maize domestication. PLoS Genet. 9:e1003604.

Wittkopp, P. J., B. K. Haerum, and A. G. Clark. 2004. Evolutionary changes in cis and trans gene regulation. Nature 430:85–88.

Xie, W. F., S. Kondo, and L. J. Sandell. 1998. Regulation of the mouse cartilage-derived retinoic acid-sensitive protein gene by the transcription factor AP-2. J. Biol. Chem. 273:5026–5032.

Yalcin, B., S. A. G. Willis-Owen, J. Fullerton, A. Meesaq, R. M. Deacon, J. N. P. Rawlins, R. R. Copley, A. P. Morris, J. Flint, and R. Mott. 2004. Genetic dissection of a behavioral quantitative trait locus shows that Rgs2 modulates anxiety in mice. Nat. Genet. 36:1197–1202.

Yan, H., W. Yuan, V. E. Velculescu, B. Vogelstein, and K. W. Kinzler. 2002. Allelic Variation in Human Gene Expression. Science 297:1143–1143.

Yates, A., W. Akanni, M. R. Amode, D. Barrell, K. Billis, D. Carvalho-Silva, C. Cummins, P. Clapham, S. Fitzgerald, L. Gil, C. G. Girón, L. Gordon, T. Hourlier, S. E. Hunt, S. H. Janacek, N. Johnson, T. Juettemann, S. Keenan, I. Lavidas, F. J. Martin, T. Maurel, W. McLaren, D. N. Murphy, R. Nag, M. Nuhn, A. Parker, M. Patricio, M. Pignatelli, M. Rahtz, H. S. Riat, D. Sheppard, K. Taylor, A. Thormann, A. Vullo, S. P. Wilder, A. Zadissa, E. Birney, J. Harrow, M. Muffato, E. Perry, M. Ruffier, G. Spudich, S. J. Trevanion, F. Cunningham, B. L. Aken, D. R. Zerbino, and P. Flicek. 2016. Ensembl 2016. Nucleic Acids Res. 44:D710-716.

Yeaman, S. 2013. Genomic rearrangements and the evolution of clusters of locally adaptive loci. Proc. Natl. Acad. Sci. 110:E1743–E1751.

Zhang, J., S. Hagopian-Donaldson, G. Serbedzija, J. Elsemore, D. Plehn-Dujowich, A. P. McMahon, R. A. Flavell, and T. Williams. 1996. Neural tube, skeletal and body wall defects in mice lacking transcription factor AP-2. Nature 381:238–241.

Zhang, J., and T. Williams. 2003. Identification and regulation of tissue-specific cis-acting elements associated with the human AP-2α gene. Dev. Dyn. 228:194–207.

This chapter appeared in *Developmental Biology* 401 (2), 2015. All authors have agreed to its use in this dissertation.

4. A 190 base pair, TGF-β responsive tooth and fin enhancer is required for stickleback *Bmp6* expression

Priscilla A. Erickson, Phillip A. Cleves, Nicholas A. Ellis, Kevin T. Schwalbach, James C. Hart, and Craig T. Miller*

Department of Molecular and Cell Biology, University of California, Berkeley, CA, 94720

Abstract

The ligands of the Bone Morphogenetic Protein (BMP) family of developmental signaling molecules are often under the control of complex *cis*-regulatory modules and play diverse roles in vertebrate development and evolution. Here, we investigated the *cis*-regulatory control of stickleback *Bmp6*. We identified a 190 bp enhancer ~2.5 kilobases 5' of the *Bmp6* gene that recapitulates expression in developing teeth and fins, with a core 72 bp sequence that is sufficient for both domains. By testing orthologous enhancers with varying degrees of sequence conservation from outgroup teleosts in transgenic reporter gene assays in sticklebacks and zebrafish, we found that the function of this regulatory element appears to have been conserved for over 250 million years of teleost evolution. We show that a predicted binding site for the TGFß effector Smad3 in this enhancer is required for enhancer function and that pharmacological inhibition of TGFß signaling abolishes enhancer activity and severely reduces endogenous *Bmp6* expression. Finally, we used TALENs to disrupt the enhancer *in vivo* and find that *Bmp6* expression is dramatically reduced in teeth and fins, suggesting this enhancer is necessary for expression of the *Bmp6* locus. This work identifies a relatively short regulatory sequence that is required for expression in multiple tissues and, combined with previous work, suggests that shared regulatory networks control limb and tooth development.

Introduction

Bone Morphogenetic Protein (BMP) ligands, the largest subfamily of TGFβ ligands, play multiple essential roles during vertebrate development (Hogan, 1996; Kingsley, 1994; Massagué, 2012), including during craniofacial and tooth development (Nie et al., 2006). Many vertebrate organs develop through reciprocal permissive and instructive signaling between adjacent epithelial and mesenchymal tissues, often involving multiple BMP ligands (Bellusci et al., 1996; Dassule and McMahon, 1998; Dudley et al., 1999; Jung et al., 1998). These pleiotropic functions of BMP ligands are orchestrated by typically large, modular, regulatory regions, which work together to drive complex spatiotemporally restricted expression patterns (Pregizer and Mortlock, 2009).

In humans, regulatory variation in *Bmp* genes has been associated with developmental disorders including brachydactyly and other birth defects (Dathe et al., 2009; Justice et al., 2012), as well as colorectal cancer (Houlston et al., 2008; Lubbe et al., 2012). In other animals, variation in the expression of *Bmp* genes has also been associated with major evolved changes in morphology, including beak shape in Darwin's finches (Abzhanov et al., 2004), jaw size and shape in cichlid fish (Albertson et al., 2005), and tooth number in stickleback fish (Cleves et al 2014).

While the *cis*-regulatory architecture of *Bmp2*, *Bmp4*, *Bmp5*, and *Bmp7* has been studied in mice (Adams et al., 2007; Chandler et al., 2007; Guenther et al., 2008; Jumlongras et al., 2012), less is known about *Bmp6* and *Bmp* gene regulation in other vertebrates. Although not required for viability in the mouse, *Bmp6* is required for axial skeletal patterning (Solloway et al., 1998), kidney function (Dendooven et al., 2011), and physiological iron regulation (Andriopoulos et al., 2009). Non-coding variants in human *Bmp6* have been associated with human height variation (Gudbjartsson et al., 2008; Wood et al., 2014), as well as orofacial clefting birth defects (Shi et al., 2012). A *cis*-regulatory allele of stickleback *Bmp6* with reduced *Bmp6* expression in developing tooth tissue has recently been shown to be associated with evolved increases in tooth number in derived freshwater sticklebacks, likely adaptive for the shift in diet in freshwater sticklebacks relative to their marine ancestors (Cleves et al., 2014).

BMP signaling plays complex and, in general, poorly understood roles during the development of placodes. During tooth development, multiple *Bmp* genes are expressed dynamically in developing odontogenic epithelia and mesenchyme (Aberg et al., 1997; Vainio et al., 1993). Several lines of evidence reveal BMP signaling plays activating roles during odontogenesis. First, epithelial BMP4 activates *Msx* expression in the mesenchyme, and exogenous BMP from a bead (Bei and Maas, 1998; Chen et al., 1996) or transgene (Zhao et al., 2000) can partially rescue tooth development in *Msx1* mutant mice. Second, in mice, teeth arrest at the bud-to-cap transition in *Bmpr1a* mutants (Andl et al., 2004; Liu et al., 2005). Third, exogenous BMP4 beads can induce molar development in mice (Kavanagh et al., 2007). Fourth, in fish, pharmacological inhibition of BMP signaling can inhibit tooth formation in cichlids (Fraser et al., 2013). In contrast, other evidence supports BMP signaling playing inhibitory effects during the development of teeth and other placodes. In mice, *Pax9* expression marks early dental mesenchyme, and BMP2 and BMP4 inhibit *Pax9* expression (Neubüser et al., 1997). In zebrafish,

inhibition of BMP signaling produces supernumerary teeth with altered morphology (Jackman et al., 2013). During development of both feather and hair placodes, BMPs play inhibitory roles (Botchkarev et al., 1999; Jung et al., 1998; Mou et al., 2006, 2011), and suppression of epithelial BMP signaling is required for hair placode induction (reviewed in Biggs and Mikkola, 2014). Together these results suggest that complex positive and negative interactions between epithelial and mesenchymal BMPs are critical for placode development, yet the regulation of these interactions remains less well understood.

Despite the major role BMP signaling plays during tooth development, little is known about the *cis*-regulatory sequences that drive dynamic *Bmp* expression in early developing odontogenic epithelia and mesenchyme. In mice, a late-stage ameloblast enhancer has been identified for the *Bmp4* gene (Feng et al., 2002); however this enhancer is not reported to be active during embryogenesis, or in dental mesenchyme. A second enhancer of mouse *Bmp4* has been described that is active during embryogenesis and drives expression in dental epithelium but not mesenchyme (Jumlongras et al., 2012). Tooth epithelial and mesenchymal enhancers of the mouse *Bmp2* gene have been localized to a ~150 kb region 3' of *Bmp2* (Chandler et al., 2007), however these enhancers have not yet been further mapped, and in general, *cis*-regulation of BMPs in dental mesenchyme is poorly understood. Furthermore, since mice are monophyodonts that form one wave of primary teeth and no replacements, less is known about *cis*-regulatory elements that drive expression in developing and replacement teeth in polyphyodont vertebrates (such as fish) that replace their teeth continuously. Because of the recently identified *cis*-regulatory allele of *Bmp6* associated with evolved changes in stickleback tooth number (Cleves et al., 2014) and to dissect epithelial and mesenchymal *cis*-regulation of vertebrate BMP signaling, we sought to begin to identify the *cis*-regulatory architecture of the stickleback *Bmp6* gene.

Methods:

Animal statement and fish husbandry:

All animal work was approved by the Institutional Animal Care and Use Committee of the University of California-Berkeley (protocol number R330). Sticklebacks (*Gasterosteus aculeatus*) were raised in ~10% seawater (3.5 g/l Instant Ocean salt, 0.217 ml/l 10% sodium bicarbonate) at 18° C, and crosses were generated by *in vitro* fertilization. Zebrafish (*Danio rerio*) were raised in a recirculating system under standard conditions, and embryos were collected either from natural spawning or *in vitro* fertilization and raised at 28.5 degrees (Westerfield, 2007).

BAC Isolation and Recombineering:

Bacterial Artificial Chromosomes (BACs) from the CHORI-213 and CHORI-215 (Salmon River marine and Paxton benthic freshwater stickleback, respectively) BAC libraries were identified by overgo screening (Ross et al., 1999) using the following overgoes: 5'- TGTGACGTTGACCTCAGCTAGACT-3' and 5'-GAGGATTTAAACCGGGAGTCTAGC -3'.

BAC ends were sequenced using Sp6 and T7 primers and mapped to the stickleback genome using the UCSC browser. BAC CHORI-215-29E12 was chosen for reporter analysis because *Bmp6* was relatively centrally located in the BAC. Inverted Tol2 sites were recombineered into the Lox511 site of the pTarbac2.1 backbone according to Suster et al. (2011) using primers PTARBAC_tol2FWD and PTARBAC_tol2REV, and ampicillin resistance was used to select successfully recombineered BAC clones. To place GFP into exon 1 of *Bmp6* as a reporter, a GFP/kanamycin resistant cassette was amplified from pGFP-FRT-Kan-FRT (Suster et al., 2011) using primers GFP_Bmp6_for and GFP_Bmp6_rev (Table 4.1), which contained 50 bp homology to the beginning and end of the first exon of stickleback *Bmp6,* respectively. This construct was then recombineered into the BAC containing iTol2 sites to produce the final reporter BAC (see Figure 4.14A-C).

Enhancer Constructs:

The vector for the stickleback 2.8 kb enhancer/promoter construct was generated using pENTRbasGFP and pTolDest (Villefranc et al., 2007) using Gateway cloning to produce a construct with the carp *ß-actin* basal promoter (Scheer and Campos-Ortega, 1999) upstream of EGFP, flanked by Tol2 sites (Urasaki et al., 2006). Next, a 2,810 bp sequence upstream of the predicted *Bmp6* transcriptional start site was PCR amplified from BAC CHORI-213-256N24 using primers Gac_3kb_for and Gac_3kb_rev and cloned upstream of the carp *ß-actin* promoter using a ClaI restriction site. Blocks of conserved sequences within the 2.8 kb construct were identified as CS1, CS2, and CS3 from the UCSC 8 species Multiz conservation track (see Figure 4.4A). These sequences were cloned into ClaI site of the carp *ß-actin* reporter construct using primers shown in Table 4.1. CS1 was cloned with Gac_3kb_for and Gac_CS1_rev. CS2 was cloned with Gac_CS2_for and Gac_CS2_rev. CS3 was cloned with Gac_CS3_for and Gac_3kb_rev. CS2+3 was cloned with Gac_CS2_for and Gac_3kb_rev. Because the CS1 fragment drove weak expression with the *ß-actin* promoter, we switched to using a well-characterized zebrafish *hsp70* promoter construct, which we found to drive much brighter expression in transgenic stickleback embryos. CS1 and CS2+3 were also cloned into the *hsp70* promoter construct for additional testing using the same genomic primer sequences but with Nhe and BamHI restriction sites in place of ClaI. The 190 bp and 72 bp enhancer sequences were amplified from the 2.8 kb construct with primers indicated in Table 4.1 and cloned into the *hsp70* construct.

The orthologous enhancer sequences were identified in other teleost genomes using the UCSC genome browser (genome.ucsc.edu) to identify sequence conservation. Zebrafish and medaka (*Oryzias latipes*) wild-type genomic DNA was isolated by standard phenol-chlorofom extraction and enhancers were amplified using primers (Table 4.1) designed from the respective genome assemblies (zv9/danRer7 and oryLat2) and cloned into the *hsp70* promoter construct. The Atlantic cod (*Gadus morhua*) enhancer DNA sequence was identified by sequence conservation on contig CAEA01327401 of the Atlantic cod genome assembly (UCSC, gadMor1). This short, unassembled contig is flanked by repetitive sequence, but the intervening sequence contains a 94 bp stretch that has 92.4% sequence identity to the stickleback enhancer and is likely the orthologous

sequence. We synthesized a 130 bp construct of Atlantic cod sequence by using two primers for amplification (Gmo_for and Gmo_rev, see Table 4.1) and two additional overlapping oligonucleotides as template (Gmo_temp1 and Gmo_temp2). The template oligonucleotides were added to standard Phusion (NEB) PCR reaction at a concentration of 0.05 µM to amplify the full 130bp sequence, which was then cloned into the Tol2 construct as described above.

Sequence Analysis:

Sequence alignments were generated using ClustalW2 (http://www.ebi.ac.uk/Tools/msa/clustalw2/) (Larkin et al., 2007) and Boxshade (http://www.ch.embnet.org/software/BOX_form.html). Binding sites were predicted with the UniProbe database (http://the_brain.bwh.harvard.edu/uniprobe/) (Newburger and Bulyk, 2009) and PROMO (http://alggen.lsi.upc.es/cgi-bin/promo_v3/promo/promoinit.cgi?dirDB=TF_8.3) (Farre et al., 2003; Messeguer et al., 2002).

Imaging and Microscopy:

Transgenic lines were imaged using a Leica DM2500 compound microscope equipped with a Leica DFC500 camera, a Leica M165FC dissecting microscope equipped with a DFC340 FX camera, or a Zeiss 700 confocal microscope. Transgenic fish were fixed for 4 hours at 4°C in either 4% paraformaldehyde in 1X PBS or 10% neutral buffered formalin. For Alizarin red fluorescent counterstaining of GFP lines, 0.01% Alizarin red was added to the fixative. Tooth number was counted on the DM2500 with TX2 filter to visualize Alizarin-stained teeth. Tooth germs with GFP+ epithelia were counted on photographs of GFP fluorescence.

Enhancer constructs

Primer Name	Sequence	Purpose
Gac_3kb_for	GCCGATCGATATAGGAAGGCTGGACAACGA	stickleback 3kb forward
Gac_3kb_rev	GCCGATCGATAGAACACAGCGGGGAAACACC	stickleback 3kb reverse
Gac_CS1_rev	GCCGATCGATAGTATGGTGCGTGTGTGCAT	stickleback CS1 reverse
Gac_CS2_for	GCCGATCGATATGCACACACGCACCATACT	stickleback CS2 forward
Gac_CS2_rev	GCCGATCGATGAAACAGCAAGCAATGACGA	stickleback CS2 reverse
Gac_CS3_for	GCCGATCGATTCGTCATTGCTTGCTGTTTC	stickleback CS3 forward
Gac_190_for	GCCGGCTAGCGCGCTCGCTTGAAAAGAGAGC	stickleback 190bp forward
Gac_190_rev	GCCGGGATCCTGTGTGTTTGCGCACATCCC	stickleback 190bp reverse
Gac_72_for	GCCGGCTAGCAGGAGGTGTCCTGTCTAGACA	stickleback 72bp forward
Gac_72_rev	GCCGGGATCCGAGGGAGGAGGAGGAAACATTAAA	stickleback 72bp rev
Dre_for	GCCGGCTAGCCCCTGAAGTTCTGTGCTTTGATCA	zebrafish forward
Dre_rev	GCCGGGATCCAAGCTGGACATTCCTCTGCAAATG	zebrafish reverse
Gmo_for	GCCGGCTAGCTGTGTACTAGGTGGAGGAGGGAGGGACCCAGGGAGGGGGGGAGGACATT	cod forward
Gmo_temp1	GACGGCCTGATGACAGGACACAGAGCTTCTGTTTAATGTCCTCCCCCCC	cod template 1
Gmo_temp2	CTGTCATCAGGCCGTCTAGACAGGACACCTCCTAGACCTAATGAGGTC	cod template 2
Gmo_rev	GCCGGGATCCGTGTGGGAGACAGAGAAAGACCTCATTAGGTCTAGGAGG	cod reverse
Ola_for	AGTCGCTAGCAATGGAAGCAGTGTGGAGGAGG	medaka forward
Ola_rev	AGCTGGATCCGGCCCTAATCAGTTGTGTTCTGCA	medaka reverse

Mutagenesis constructs

Primer Name	Sequence	Purpose
Smad3_mut1_for	ATTAGGTCTAGGAGGTGTCCTAAATAGACAGTGTGATGACAGGAC	SMAD3 mut. first round forward
Smad3_mut1_rev	GTCCTGTCATCACACTGTCTATTTAGGACACCTCCTAGACCTAAT	SMAD3 mut. first round reverse
Smad3_mut2_for	GTCCTGTCATCACACTAAATATTTAGGACACCTCCTAGACCTAATGAGGT	SMAD3 mut. second round forward
Smad3_mut2_rev	ACCTCATTAGGTCTAGGAGGTGTCCTAAATATTTAGTGTGATGACAGGAC	SMAD3 mut. second round reverse
Pea3_mut1_for	CTCCTCCTCCCTCTACTTTTAATTCACCCGCCGAACAC	PEA3 mut. first round forward
Pea3_mut1_rev	GTGTTCGGCGGGTGAATTAAAAGTAGAGGGAGGAGGAG	PEA3 mut. first round reverse
Pea3_mut2_for	AGGACACAGAACCTCTGTTTAATGTTTGGCCTCCTCCCTCTAC	PEA3 mut. second round forward
Pea3_mut2_rev	GTAGAGGGAGGAGGCCAAACATTAAACAGAGGTTCTGTGTCCT	PEA3 mut. second round reverse
RAR_mut1_for	CTCCTCCTCCCTCTACTTCCAATGGGCCCGCCGAACAC	RAR mut. first round forward
RAR_mut1_rev	GTGTTCGGCGGGCCCATTGGAAGTAGAGGGAGGAGGAG	RAR mut. first round reverse

RAR_mut2_for	TTCACCCGCCGAACACACACAGGGCCTGCTCTGCC	RAR mut. second round forward
RAR_mut2_rev	GGCAGAGCAGGCCCTGTGTGTGTTCGGCGGGTGAA	RAR mut. second round reverse
TCF_mut1_for	GCGCTCGCTTGAAAAGAGTCCGATTCAAGCAGACAAAG	TCF mut. 1st round forward
TCF_mut1_rev	CTTTGTCTGCTTGAATCGGACTCTTTTCAAGCGAGCGC	TCF mut. 1st round reverse
TCF_mut2_for	GTGATGACAGGACACAGAACGACTGTTTAATGTTTCCTCCTC	TCF mut. 2nd round forward
TCF_mut2_rev	GAGGAGGAAACATTAAACAGTCGTTCTGTGTCCTGTCATCAC	TCF mut. 2nd round reverse
TCF_mut3_for	AATGTTTCCTCCTCCTCCGACTACTTCCAATTCACCCG	TCF mut. 3rd round forward
TCF_mut3_rev	CGGGTGAATTGGAAGTAGTCGGAGGAGGAGGAAACATT	TCF mut. 3rd round reverse
TCF_mut4_for	GAAAAGAGTCCGATTCAAGCAGATGCGGACCTCATTAGGTCTAGGAGGTG	TCF mut. 4th round forward
TCF_mut4_rev	CACCTCCTAGACCTAATGAGGTCCGCATCTGCTTGAATCGGACTCTTTTC	TCF mut. 4th round reverse
Homeo_mut1_for	GCGATTCAAGCAGACAAAGACCGGGGTAGGTCTAGGAGGTGTCCTGTC	Homeodomain mut. forward
Homeo_mut1_rev	GACAGGACACCTCCTAGACCTACCCCGGTCTTTGTCTGCTTGAATCGC	Homeodomain mut. reverse

BAC recombineering

Primer Name	**Sequence**	**Purpose**
GFP_Bmp6_for	CTGCAGCTCCAAGAGAGACCCACTTGGGACAGCGGAGAACACAGCGGGGAGCCACCATGG TGAGCAAGGGCGAGGAGCTGTTC	GFP>Bmp6 recombineering
GFP_Bmp6_rev	CCAAGGTTAACGAAGCTCATGACCATGTCTGCGTCATTTAGAAAGGCACTCCGCGTGTAG GCTGGAGCTGCTTC	GFP>Bmp6 recombineering
PTARBAC_tol2FWD	GCGTAAGCGGGGCACATTTCATTACCTCTTTCTCCGCACCCGACATAGATCCCTGCTCGAG CCGGGCCCAAGTG	iTol2 recombineering
PTARBAC_tol2REV	CGCGGGGCATGACTATTGGCGCGCCGGATCGATCCTTAATTAAGTCTACTAATTATGATCC TCTAGATCAGATCT	iTol2 recombineering

Table 4.1. Primers used in this study.

All primers were designed from genomic sequences obtained from UCSC. Gac=*Gasterosteus aculeatus* (stickleback), Dre=*Danio rerio* (zebrafish), Gmo=*Gadus morhua* (Atlantic cod), Ola=*Oryzias latipes* (medaka). For constructs with multiple mutations, the order in which the mutations were introduced is indicated.

Fish injections and line generation:

Transposase mRNA was produced from the pCS2-TP plasmid (Kawakami et al., 2004) with the mMessage mMachine SP6 *in vitro* transcription kit (Ambion) according to manufacturer's instructions and purified with a Qiagen RNeasy column. Zebrafish injections were performed with 25 ng/µL plasmid DNA and 37.5 ng/µL transposase and 0.05% phenol red as previously described (Fisher et al., 2006). Because stickleback embryos are much larger than zebrafish embryos, the DNA and RNA concentrations were increased to 37.5 and 75 ng/µL respectively. Stable transgenic lines were generated by outcrossing injected fish to non-transgenic fish and visually screening for fluorescent transgenic offspring. At least two stable lines were observed for each construct to ensure fluorescent patterns were due to the transgene and not artifacts of the transgene integration sites.

Site directed mutagenesis:

Mutagenesis primers were designed using the online Quickchange tool (http://www.genomics.agilent.com/primerDesignProgram). For constructs containing multiple mutations, the mutagenesis was performed in multiple rounds. Mutagenesis reactions were performed with 125 ng of each primer, 50 ng plasmid template, 200 µM dNTPs, and Pfu Turbo polymerase and buffer. Cycling conditions were 95°C for 30 seconds, followed by 16 cycles of 95°C / 30 s, 55°C / 60 s, and 68°C /780 s. Primer sequences can be found in supplementary Table 4.1; the mutated sequences are shown in Figure 4.10A. DpnI was added immediately after cycling, and the reaction was incubated for 1 hr at 37°C, then immediately transformed into Top10 chemically competent *E. coli* cells.

Drug treatments:

SB431542 and XAV939 (Sigma) were dissolved in DMSO to concentrations of 100 µM and 10 µM, respectively. The drug was then diluted into stickleback water or zebrafish system water to working concentrations (25-100 µM for SB431542 and 5-10 µM for XAV939). A DMSO vehicle control was done in parallel with all drug treatments. Drug treatment was performed in 6- or 24-well cell culture dishes. Sticklebacks were treated from 2 dpf to 5 dpf for observation of pectoral and median fin expression, and for 5-7 days post-hatching for observation of tooth GFP. Zebrafish were treated beginning at 10 hpf for observation of median fin and beginning at 24 hpf for pectoral fin and tooth expression. For multiday treatments, fresh solution was applied every 48 hours until the end of the experiment.

In situ hybridization (ISH):

Bmp6 in situ hybridization was performed on embryos and newly-hatched juveniles as previously described (Cleves et al., 2014). For pharyngeal tooth and gill *in situs*, the branchial skeleton was dissected out of the embryo and cut along the dorsal midline prior to the hybridization step.

Mutagenesis using TALENs:

TAL Effector Nucleases (TALENs) were targeted to the predicted Smad3 binding site within the 190 bp enhancer using TAL Effector Nuclear Targeter 2.0 (https://tale-nt.cac.cornell.edu/) using the Cermak architecture (Cermak et al., 2011; Doyle et al., 2012). TALEN plasmids were generated using the RVDs shown in Table 4.2. TALEN mRNAs were produced with the Mmessage Mmachine kit (Ambion), purified with Qiagen RNeasy columns, and injected into one-cell stickleback embryos at a concentration of 40 ng/μL for each mRNA plus 0.05% phenol red. Embryos and juvenile fish were screened for lesions in the Smad3 site by screening for loss of an XbaI cut site in a 144 bp PCR product amplified with primers Gac_190_for and Gac_72_rev (see Figure 4.11G). F1 animals with deletions visible on a 2% agarose gel (~15 bp or larger) were crossed to generate animals used in *in situ* hybridization. Because the F1 parents carried different TALEN-induced lesions, the F2 animals used for ISH were transheterozygotes for two slightly different alleles of the enhancer deletion (see Figure 4.14E).

	PfusA RVDs	PfusB RVDs	pLR RVD
5':	NG1	NI1	HD
	NI2	NN2	
	NN3	NN3	
	NN4	NG4	
	NG5	NN5	
	HD6	NG6	
	NG7		
	NI8		
	NN9		
	NN10		
3':	NG1	NN1	NG
	HD2	NG2	
	NG3	HD3	
	NN4	NI4	
	NG5		
	NN6		
	NG7		
	HD8		
	HD9		
	NG10		

Table 4.2. RVDs used for TALEN construction.

Individual RVD monomers were cloned into pFUS_A and the appropriate pFUS_B plasmid. The completed pFUS_A and pFUS_B plasmids were then combined into pTal3-DD (5') and pTal3-RR (3') with the appropriate pLR and sequence-verified by Sanger sequencing (Cermak et al., 2011).

Results:

A Bmp6 reporter BAC recapitulates endogenous Bmp6 expression

To begin to identify the *cis*-regulatory architecture of the stickleback *Bmp6* gene, we generated a *Bmp6* reporter line by identifying a bacterial artificial chromosome (CHORI BAC215-29E12) containing 180 kb of sequence starting ~52 kb upstream of *Bmp6*. Inverted Tol2 sequences were recombineered into the backbone of this BAC, and the first exon of *Bmp6* was replaced with GFP coding sequence. This transgenic construct drove GFP reporter expression in a variety of tissues throughout development (Figure 4.1), including the embryonic tailbud following somitogenesis (3.5 dpf), the embryonic heart and ventrolateral cells in the pharyngeal region (4 dpf), the distal edge of the developing pectoral fin, and the distal edge of the median fin (5 dpf). After hatching (10-15 dpf), expression was seen in oral and pharyngeal teeth, the pericardium, cells surrounding the opercle and branchiostegal rays, gill buds, and gill rakers.

We compared this transgene expression pattern to the expression pattern of endogenous *Bmp6* mRNA via *in situ* hybridization. We observed *Bmp6* expression in nearly all of the same domains as the reporter BAC (Figure 4.2), including the tailbud (at 3.5 dpf), heart, the distal edges of the median and pectoral fins (at 5 dpf), gills, gill rakers, and in the previously described (Cleves et al., 2014) epithelium and mesenchyme of developing teeth (assayed at ~12 dpf). However, several domains observed by *in situ* hybridization were not observed in the BAC transgenic line, including the notochord, the dorsal medial diencephalon, the eyes, and the ears (Figure 4.2), suggesting that regulatory elements lying outside of the 180 kb of genomic sequence contained within the BAC control these *Bmp6* expression domain.

A conserved 190 bp enhancer drives tooth, median fin, and pectoral fin expression in both stickleback and zebrafish

To begin to identify regulatory elements contained within this 180 kb genomic interval containing *Bmp6*, we first cloned a construct containing ~2.8 kb of sequence immediately upstream of stickleback *Bmp6* containing regions of sequence conserved among other teleosts (Figure 4.4A). This construct drove GFP expression in a number of tissues that were similar to expression patterns driven by the BAC (Figure 4.3, compare to Figure 4.1), including the tailbud, the heart, pectoral and median fins, oral and pharyngeal teeth, gills, and the pericardium. Other domains driven by the BAC were not observed in the 5' construct, including gill rakers, opercle, and branchiostegal rays; these domains are likely driven by more distal regulatory elements contained within the BAC but excluded from the 2.8 kb sequence. Combined, these results suggest that much of the regulatory information for *Bmp6* is contained within the 2.8 kb upstream sequence, but that other regulatory elements drive additional expression domains.

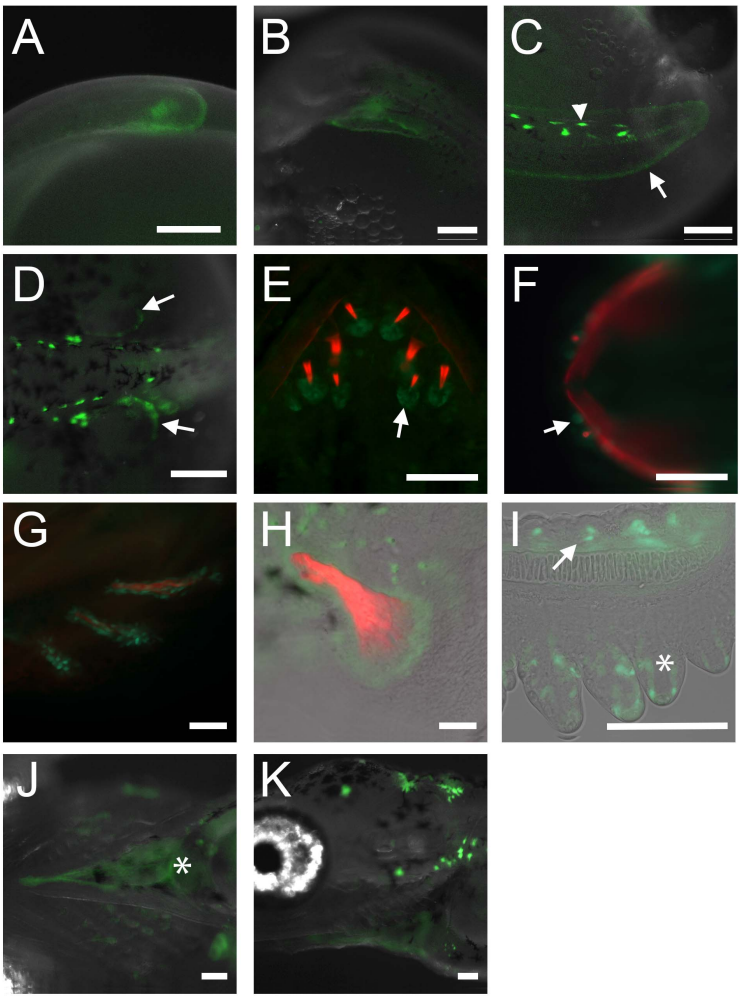


Figure 4.1. Domains of GFP expression in a stickleback *Bmp6* BAC reporter.
The first exon of *Bmp6* was replaced with GFP in a 180 kb stickleback BAC (see Figure 4.14). Stable lines carrying this reporter construct displayed GFP expression in a variety of tissues. Expression was detected in the distal edge of the forming median caudal fin (A) and ventrolateral cells surrounding the heart and pharyngeal region (B) at 3 dpf when viewed laterally. At 5 dpf, expression was observed in cells in the distal edge of the median fin (arrow in C, arrowhead points to autofluorescent pigment cell) and the distal edge of the developing pectoral fins (arrows in D). Soon after hatching (at 11-12 dpf), expression was observed in pharyngeal (E) and oral (F, ventral view) teeth. Additionally, GFP^+ cells were observed surrounding the branchiostegal rays (G), opercle (H), and gill rakers (arrow I). Cells in the soft tissue of the gill buds were also seen (asterisk in I). GFP was also observed in cells surrounding the heart (asterisk in J, ventral view and K, lateral view). Bone is fluorescently counterstained with Alizarin red in E-G. Scale bars = 200 µm (A-D); 100 µm (E-K).

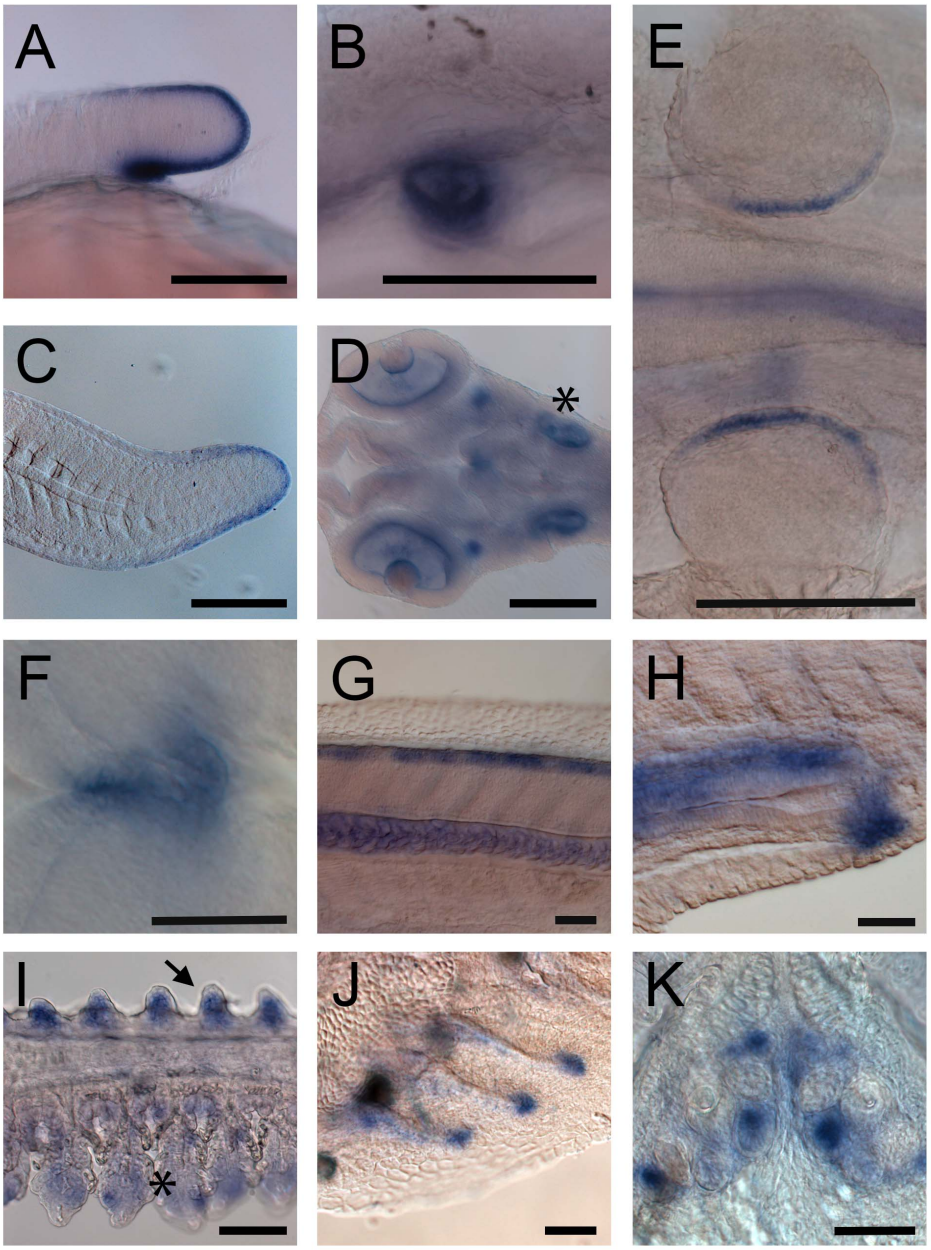


Figure 4.2. Expression domains of stickleback Bmp6.

Bmp6 expression was assayed by whole mount *in situ* hybridization at 3 dpf (A, B), 5 dpf (C-H), and 12 dpf (I-K). Expression was observed in the forming median fin in the tailbud (A, C), heart (lateral view in B), eyes and ears (asterisk) (D), distal edge of the developing pectoral fins (E), dorsal medial diencephalon (F), notochord and dorsal neural tube (G), hindgut and cloaca (H), gill rakers and gill buds (arrow and asterisk in I), branchiostegal rays (J), and pharyngeal teeth (K). Scale bars = 200 μm (A-E); 50 μm (F-K).

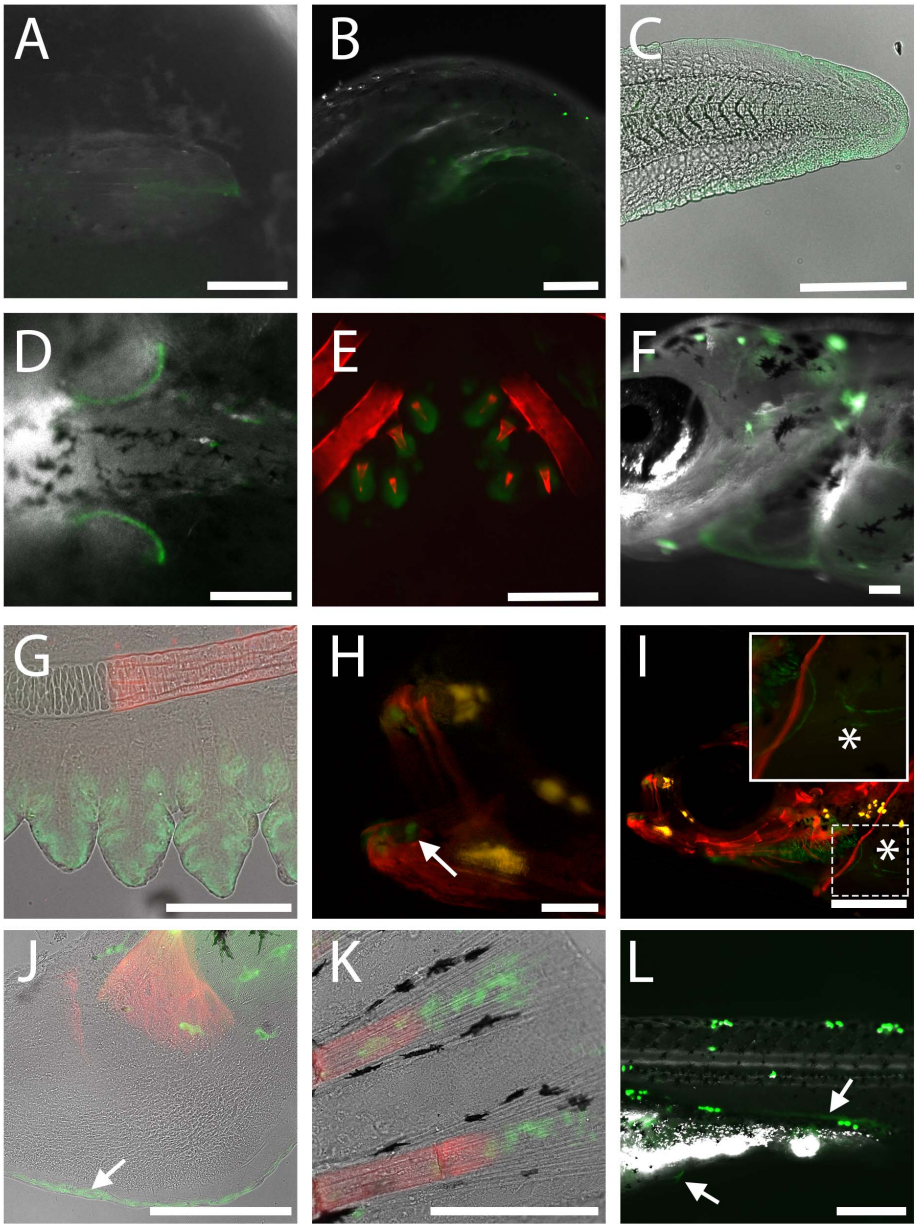


Figure 4.3. Expression driven by 2.8 kb of genomic sequence upstream of Bmp6. During early development, the 2.8 kb reporter construct drove expression in the forming median fin in the tailbud at 3 dpf (A), cells in the developing heart and pharyngeal pouches at 4 dpf (B), the distal edge of the median (C) and pectoral (D) fins at 5 dpf. After hatching (11-14 dpf), additional expression was observed in pharyngeal teeth (E), pericardial cells (F), the developing gills (G), oral teeth (H), the scapulocoracoid cartilage (I), and the distal edge of the opercle (J). In fry (22-30 dpf), expression was observed in the distal tips of fin rays (K) and the developing pelvic spine (arrow) and kidney (asterisk) (L). Red in E, G-K is Alizarin red counterstaining of bone, and yellow spots in H-J are autofluorescent pigment cells. Scale bars = 200 μm (A-D); 100 μm (E-H, J, K); 500 μm (I, L).

We hypothesized that the different anatomical sites of expression driven by the 2.8 kb fragment result from multiple anatomically specific enhancers within this sequence. We first tested three non-overlapping subclones, each containing a block of evolutionarily conserved sequence (Figure 4.4A). While the most 5' subclone (CS1) drove robust reporter gene expression in most domains of the 2.8 kb fragment, neither the middle (CS2) nor 3' subclone (CS3) drove detectable GFP expression in fins, teeth, or other domains driven by the 2.8 kb fragment at the 3-5 dpf or post-hatching (10-13 dpf) stages. Furthermore, a construct containing CS2 + CS3 also drove no detectable pattern of GFP with either the *ß-actin* or *hsp70* promoter. Next, we focused on the 5'-most region (CS1), and tested a 190 bp fragment highly conserved within teleosts (Figure 4.4B). This 190 bp fragment drove robust GFP expression in the distal edges of the pectoral and median fins, and oral and pharyngeal teeth (Figure 4.4C-E). Within developing teeth, GFP expression was observed in the inner dental epithelium (IDE) for all constructs (Figure 4.5) as well as the interior mesenchyme of mature functional teeth (Figure 4.4D), similar to endogenous *Bmp6* expression during tooth development (Cleves et al., 2014). Robust tooth GFP expression was seen in all teeth at all stages examined including in juveniles and adults, suggesting tooth enhancer activity is present in both primary and replacement teeth (Figure 4.4D-E, data not shown). Some domains, including the gills, were lost when CS1 was reduced to the 190 bp fragment, suggesting that flanking sequence is required for these domains. When the orientation of the enhancer was flipped with respect to the *hsp70* promoter, 77% (38/49) of injected fish had pectoral and/or median fin expression at 5 dpf, and 69% (27/39) had oral and/or pharyngeal tooth expression at 13 dpf. This result suggests that this enhancer functions regardless of orientation to the promoter. Combined, our results suggest that most domains driven by the 2.8 kb enhancer are driven by the short 190 bp conserved sequence. This 190 bp minimal sequence does not differ between marine and freshwater sticklebacks, though several marine-freshwater sequence differences exist in the surrounding sequences of CS1.

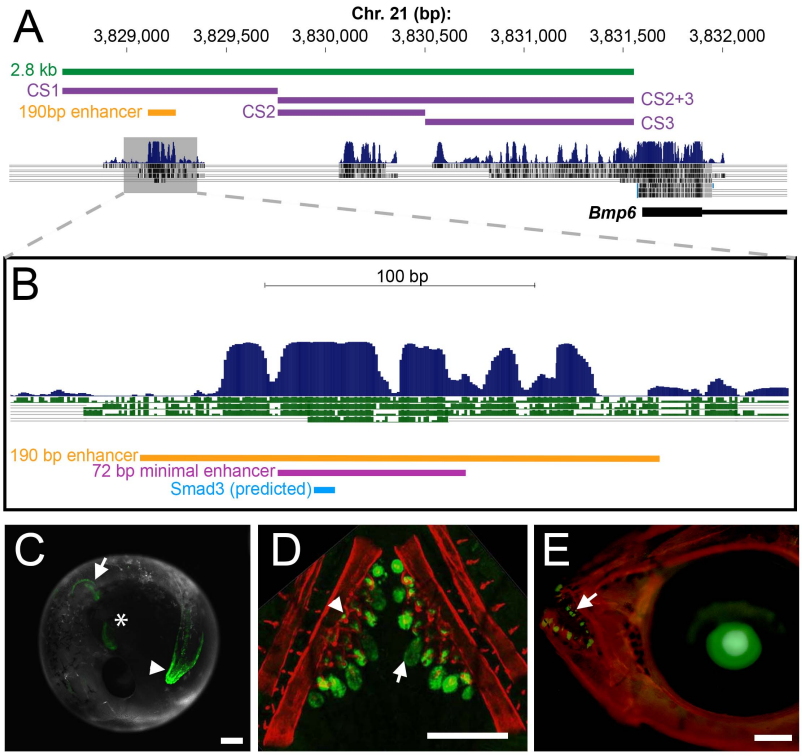


Figure 4.4. A conserved 190 bp enhancer upstream of Bmp6 drives gene expression in several domains.

(A) The 5' region of stickleback *Bmp6* from the UCSC genome browser (http://genome.ucsc.edu/). The region of genomic DNA used in the 2.8 kb enhancer construct (see Figure 4.3) is shown in green, conserved sequences (CS) 1-3 are shown in purple, and the subcloned 190 bp enhancer is shown in yellow. The first exon and part of the first intron of *Bmp6* are shown in thick and thin black lines, respectively (bottom). Conservation peaks and alignments (dark blue and grey) are shown from the 8-Species MultiZ track. (B) Zoom in on the middle of CS1, approximately 2.5 kb upstream of the predicted *Bmp6* transcription start site. The 190 bp enhancer, the 72 bp minimal enhancer (see Figure 4.8), and a predicted Smad3 binding site (see Figure 4.10-11) are shown in yellow, pink, and blue, respectively. The conservation track is shown as dark blue peaks, above green alignments showing conservation to medaka, tetraodon, fugu, and zebrafish, from top to bottom. (C) GFP reporter expression pattern driven by the 190 bp enhancer in a 5 dpf (stage 22, (Swarup, 1958)) stickleback embryo. Strong expression was seen in the distal edge of the developing pectoral fin (arrow), the heart (asterisk), and the distal edge of the median fin (arrowhead). (D) Confocal projection of GFP reporter expression in the ventral pharyngeal tooth plate in a ~10 mm stickleback fry. Expression was observed in the epithelium of developing tooth germs (arrow) and the odontogenic mesenchyme (arrowhead) in the cores of ossified teeth. Bones are fluorescently stained with Alizarin red. (E) GFP reporter expression in the oral teeth (arrow) of a 30 dpf stickleback fry. GFP in the lens is an internal control for the zebrafish *hsp70* promoter used in the transgenic construct. Scale bars = 200 μm.

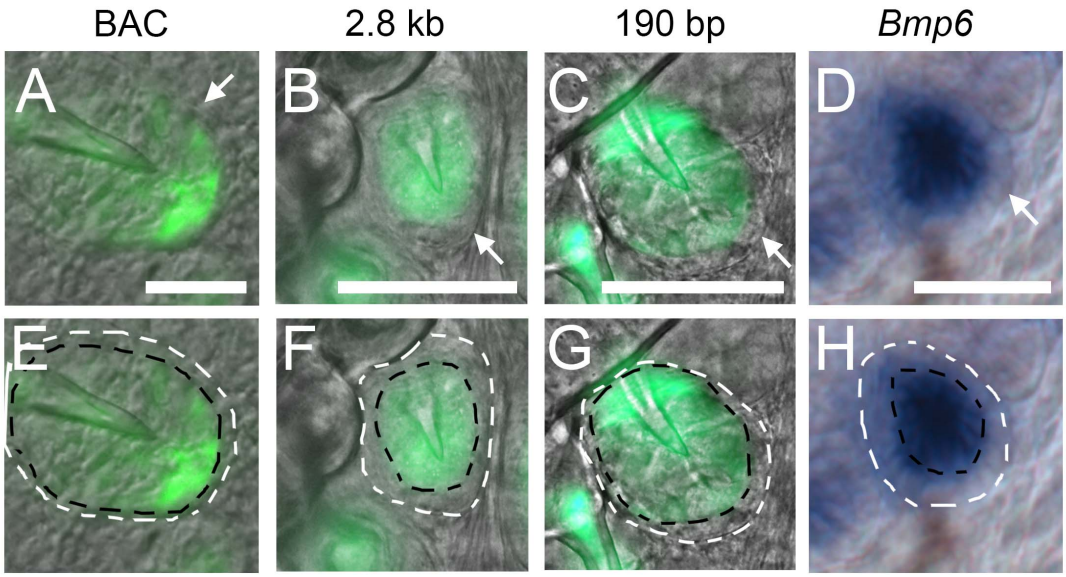


Figure 4.5. Enhancer GFP and Bmp6 expression are detected in the inner but not outer dental epithelium.

(A-C) GFP expression driven by the reporter BAC (A), 2.8 kb reporter construct (B), and 190 bp reporter construct (C) was limited to the inner dental epithelium (IDE) as visualized under differential interference contrast optics. (D) *Bmp6* mRNA expression was also restricted to the IDE as previously reported (Cleves et al 2014). The outer dental epithelium (ODE) is indicated with white arrows in A-D. (E-H) Images from A-D with the outer edge of the ODE traced with white dashed lines and the outer edge of the IDE traced with black dashed lines. Scale bars = 100 μm.

Conservation of cis regulatory elements and trans machinery in teleosts

Because we used evolutionary sequence conservation to identify the 190 bp minimal enhancer and the sequence was partially conserved to zebrafish, we hypothesized that this 190 bp stickleback enhancer would show similar activity in transgenic zebrafish. Stickleback and zebrafish are ~250 million years divergent (Near et al., 2012) and share only 3 short blocks (totaling 28 bp, Figure 4.6A) of perfectly conserved nucleotides in the middle of the enhancer. However, the stickleback enhancer robustly drove a highly similar expression pattern in zebrafish, with expression in the distal edges of the median and pectoral fins, and pharyngeal tooth epithelium and mesenchyme (Figure 4.6B-D), suggesting that the *trans* factors activating the enhancer are conserved in distantly related teleosts. We next asked whether the orthologous sequence from the zebrafish genome had similar enhancer activity in both zebrafish and sticklebacks. A construct containing 477 bp of sequence from the orthologous region of the zebrafish genome drove weak expression in these expression domains (distal edges of median and pectoral fins, and teeth) in a subset of transgenic zebrafish offspring obtained (Figure 4.6E-G and Table 4.3). In sticklebacks, seven stable transgenic lines with the zebrafish sequence driving GFP had no fin expression, although one transgenic line displayed very faint expression in the distal edges of the median and pectoral fins (Figure 4.6H-I). None of the eight lines had GFP expression in teeth (Figure 4.6J). Therefore, sticklebacks and zebrafish likely share the *trans* machinery sufficient to drive expression from the stickleback sequence, but the *cis* regulatory information present in the zebrafish orthologous sequence is not sufficient to drive tooth expression in the stickleback *trans* environment.

Because the zebrafish enhancer shows much less sequence conservation to sticklebacks relative to other teleosts (Figure 4.6A), we hypothesized that the loss of robustness and loss of tooth expression may be unique to the zebrafish *cis*-regulatory element. We generated constructs containing the orthologous enhancer sequences of a beloniform (medaka) and a gadiform (Atlantic cod), which fall between zebrafish and sticklebacks in the teleost phylogeny (Near et al., 2012). We found that sequences from both additional species drove expression in fins and teeth in both stickleback and zebrafish embryos (Figure 4.7, Table 4.3), although the cod enhancer appeared to be slightly less robust (Table 4.3).

Based on the apparent partial conservation of enhancer function in zebrafish and the conserved activities of the medaka and cod enhancers, we further shortened the stickleback enhancer to contain the sequence most highly conserved among teleosts, a 72 bp sequence near the center of the 190 bp construct, and hypothesized that it would drive the tooth, median fin, and pectoral fin expression domains. In support of this hypothesis, this construct in a stable line of zebrafish was sufficient to drive strong GFP expression in teeth and median and pectoral fins (Figure 4.8). Notably, the heart domain driven by this construct was considerably brighter relative to the 190 bp enhancer, suggesting that this short sequence may have lost additional repressor elements that limit expression in the heart. A similar pattern of brighter heart expression was observed in stickleback injected with this construct compared to the 190 bp larger construct (data not shown). These results suggest that the flanking conserved sequences are not required for the basic

enhancer pattern in fins and teeth, but may be important for fine-tuning the transcriptional output.

	***cis*-regulatory element**	**# lines with median fin expression**	**# lines with pectoral fin expression**	**# lines with tooth expression**
Stickleback *trans*	zebrafish	1/8	1/8	0/8
	cod	5/7	5/7	4/7
	medaka	4/5	4/5	4/5
	stickleback	6/6	6/6	6/6
Zebrafish *trans*	zebrafish	4/7	5/7	3/7
	cod	2/2	2/2	2/2
	medaka	5/5	5/5	5/5
	stickleback	2/2	2/2	2/2

Table 4.3. Enhancer activity of cis-regulatory sequences from four species in stickleback and zebrafish trans environments.

Fish injected with each construct were outcrossed to wild-type fish, and offspring were scored for GFP fluorescence in the distal edge of the median fin, distal edge of the pectoral fin, and the pharyngeal teeth for each independent line. For stickleback, median and pectoral fins were scored at 5 dpf and teeth were scored post-hatching (12-20 dpf). For zebrafish, median fins were scored at 24 hpf, pectoral fins were scored at 48 hpf, and teeth were scored at 5 dpf.

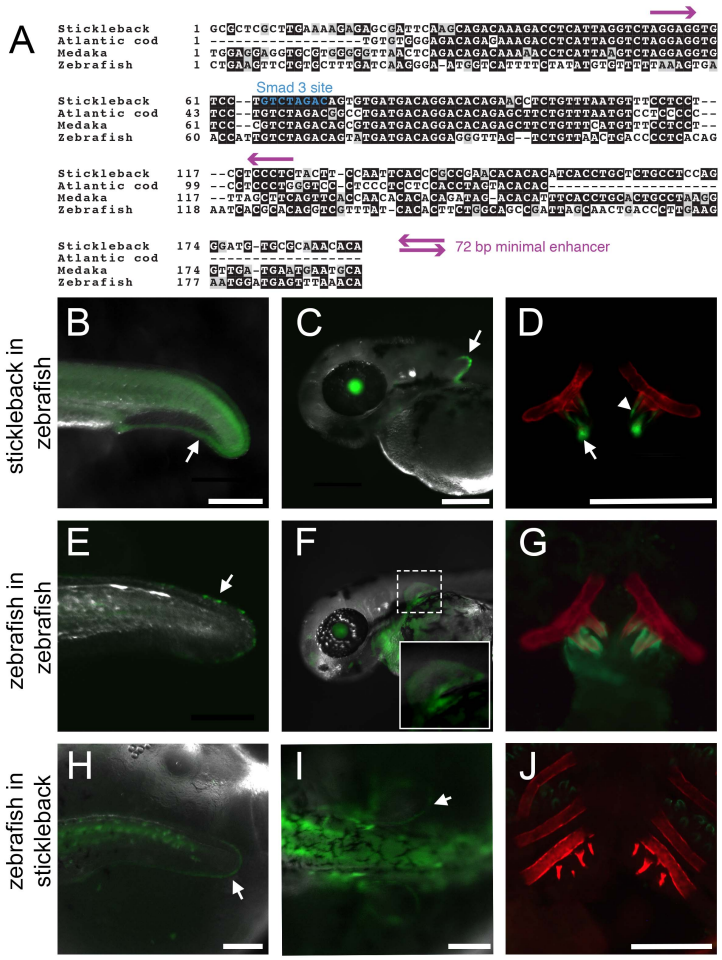


Figure 4.6. Evolutionary functional conservation of the Bmp6 enhancer in teleosts. (A) Sequence alignments of four teleost sequences relative to the 190 bp stickleback enhancer. The perfectly conserved Smad3 dimer binding site is marked in blue, and purple arrows mark the boundaries of the 72 bp minimal enhancer (see Figure 4.8). (B-D) The stickleback sequence reporter construct stably integrated into the zebrafish genome drove expression in the distal edge of the median fin at 24 hpf (arrow in B), the distal edge of the pectoral fin at 48 hpf (arrow in C), and tooth epithelium (arrow) and mesenchyme (arrowhead) at 5 dpf (D). (E-G) A 477 bp construct of zebrafish genomic sequence centered around the conserved sequence of the enhancer drove similar, but weaker expression in the median fin of a 33 hpf zebrafish (arrow in E), pectoral fins of a 48 hpf zebrafish (inset of F), and teeth of a 5 dpf zebrafish (G). (H-I) Although not detected in seven of eight stable lines, in one of eight stable lines, the zebrafish sequence drove faint expression in the distal edges of the median fin (arrow in H) and pectoral fins (arrow in I) of 5 dpf stickleback. However, no expression was detected in tooth germs in newly hatched fry in any line (J). See Table 4.3 for quantification of expression domains of transgenic lines. Bone is fluorescently stained with Alizarin red in (D, G, J). Scale bars = 200 μm.

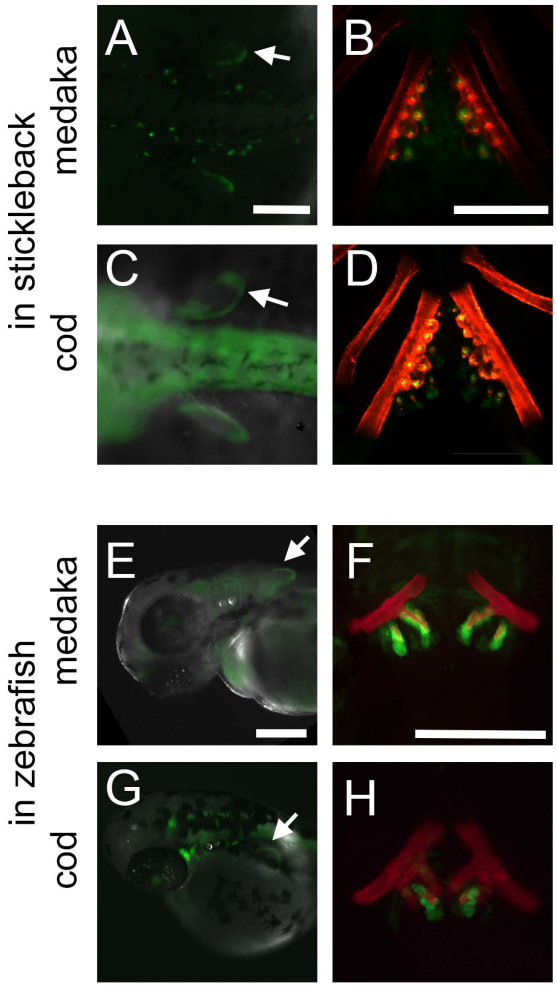


Figure 4.7. Atlantic cod and medaka enhancers drive fin and tooth expression in both stickleback and zebrafish.

Orthologous *Bmp6* enhancer sequences from two species from clades that evolved between zebrafish and sticklebacks, medaka and Atlantic cod, drove similar expression patterns in stickleback (A-D) and zebrafish (E-H). Expression was observed in the distal edges of the pectoral fins (arrows) at 5 dpf in stickleback (A, C) or 48-56 hpf zebrafish (E, G). Later in development, pharyngeal tooth expression was observed at 20 dpf in stickleback (B, D) or 5 dpf zebrafish (F, H). Bright neural expression in (C) was not seen in other lines and was likely an artifact of the transgene integration site. Scale bars = 200 μm.

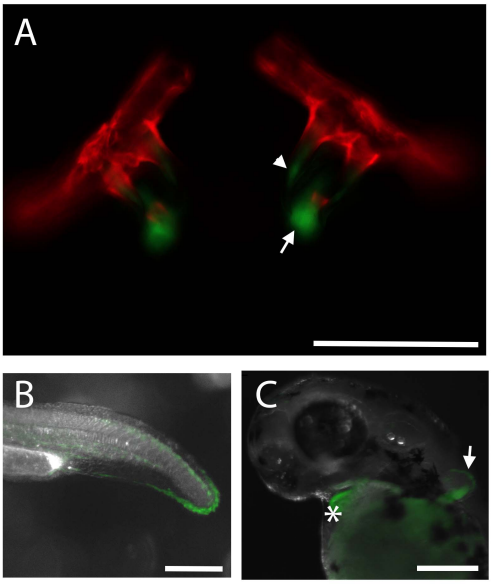


Figure 4.8. 72bp of conserved stickleback genomic sequence is sufficient for enhancer domains but increases heart expression.

The minimally sufficient 72 bp construct drove expression in (A) mesenchyme (arrowhead) and epithelium (arrow) of a 5 dpf zebrafish ventral tooth plate, (B) the distal edge of the median fin in a 24 hpf zebrafish and (C) the distal edge of the pectoral fin (arrow) in a 48 hpf zebrafish. The intensity of heart expression was noticeably increased (asterisk, compare to Figure 4.6C), suggesting that the shortened sequence had lost some repressor activity. Scale bars = 100 μm (A); 200 μm (B-C).

A predicted Smad3 binding site is required for enhancer function.

To identify candidate transcription factor binding sites within the 190 bp enhancer, we used UniProbe and PROMO (Newburger and Bulyk, 2009; Farre et al., 2003; Messeguer et al., 2002) and found predicted binding sites of transcription factors in several signaling pathways involved in developmental regulation: FGF (PEA3), retinoic acid (RAR-γ), Wnt (TCF/Lef), and TGFβ (Smad3), as well as a predicted homeodomain binding site (Figure 4.10A). We were particularly interested in the homeodomain binding site given the known crosstalk between the *Msx1* and *Bmp4* genes during mouse tooth development (Bei and Maas, 1998; Chen et al., 1996; Jumlongras et al., 2012), and the predicted TCF/Lef sites, given the previously described roles of Wnt signaling regulating *Bmp4* dental mesenchyme expression in mice (Fujimori et al., 2010; O'Connell et al., 2012). We quantified the number of stickleback embryos showing pectoral and/or median fin, as well as pharyngeal and/or oral tooth expression, when injected with constructs containing mutated binding sites. The mutation of TCF/Lef and Smad3 binding sites significantly decreased the percentage of fish with median and/or pectoral fin expression domains, whereas the predicted PEA3, RAR-γ, and homeodomain mutations did not (Figure 4.10B). Likewise, only the mutations in predicted TCF/Lef and Smad3 sites affected tooth expression, with especially reduced expression when the predicted Smad3 binding site was mutated (Figure 4.10C). We made stable zebrafish lines for each of the Smad3 and TCF/Lef mutated enhancers and found that the Smad3-mutated reporter construct did not drive robust expression in zebrafish fins or teeth, while the TCF/Lef mutated construct did drive these domains, albeit at apparently reduced levels (Figure 4.9). Since the Smad3-mutated construct did not drive fin or tooth expression in zebrafish, we generated a stable line in sticklebacks and found that this line similarly did not drive detectable median fin, pectoral fin, or tooth expression (Figure 4.11J). Therefore, the predicted Smad3 site is required for normal enhancer output, while TCF/Lef sites may be responsible for expression level but not tissue specificity.

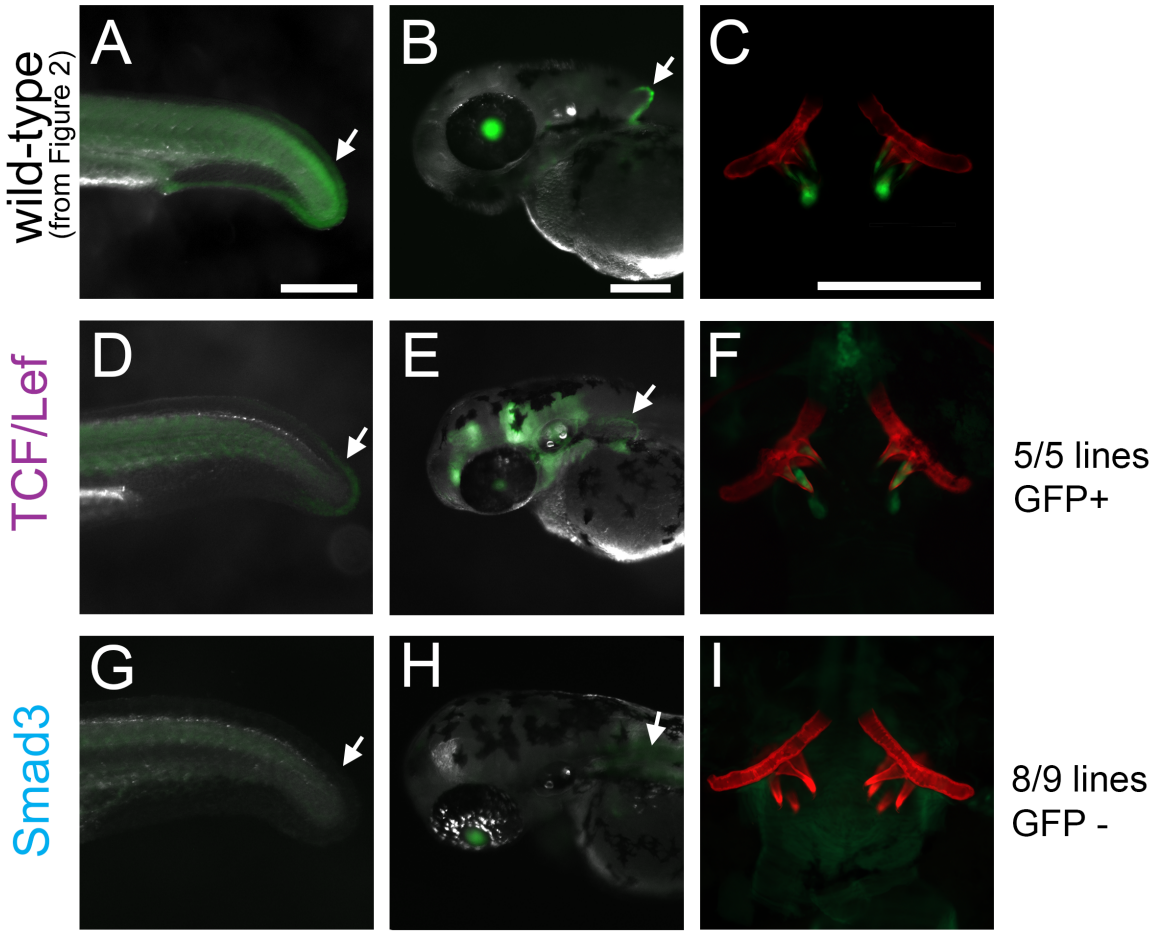


Figure 4.9. Mutation of Smad3 but not TCF/Lef predicted binding sites affects reporter expression in zebrafish.

Zebrafish stable lines were obtained for two constructs that appeared to show reduced activity in sticklebacks. (A-C) The wild-type 190 bp stickleback enhancer drove expression in the distal edge of the median fin (A), distal edge of the pectoral fin (B) and pharyngeal teeth (C) of zebrafish. Images in A-C are the same as in Figure 4.6 for comparison with D-J. (D-F) The TCF/Lef mutated construct showed expression in the median fin at 24 hpf (arrow in D), pectoral fin at 48 hpf (arrow in E), and pharyngeal teeth at 5 dpf (F) in all lines observed. Brain expression in E was not typical and is likely an artifact of the transgene integration site. (G-I) In nearly all (8/9) lines observed, the Smad3 mutated construct lacked expression in the median fin (arrow in G), pectoral fin (arrow in H), and teeth (I). One of 9 lines had very faint expression in these domains. Scale bars = 200 μm.

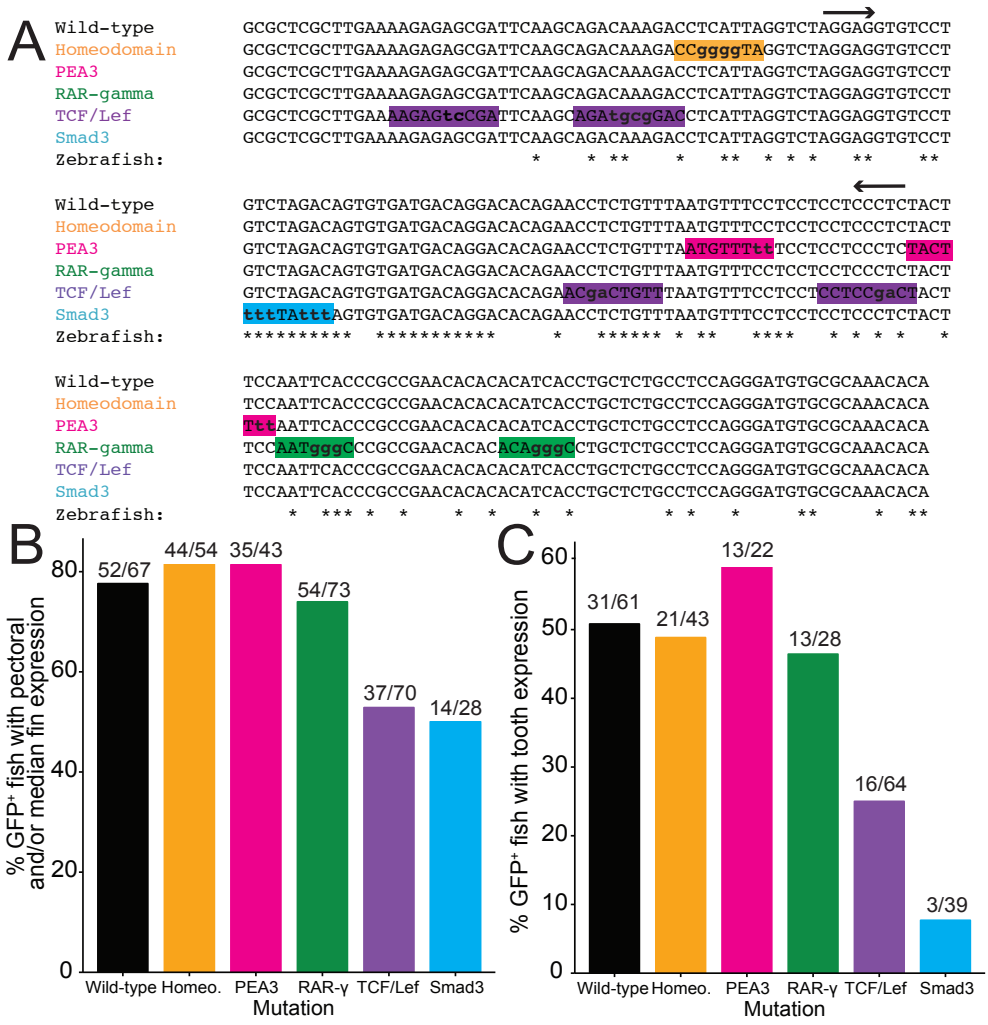


Figure 4.10. Mutations in a predicted Smad3 binding site severely reduce enhancer function.

(A) Binding sites predicted by UniProbe and PROMO are highlighted with a unique color for each signaling pathway. Highlighted sequences represent the "predicted sequence" from PROMO or the "K-mer" from UniProbe. Mutated base pairs are shown with lowercase letters. Nucleotide positions conserved to zebrafish are indicated with an asterisk, and arrows indicate the 72 bp minimal enhancer sequence. (B-C) Sticklebacks were injected with each mutated construct and scored for pectoral fin and/or median fin expression at 5 dpf (B) and oral and/or pharyngeal tooth expression at 12-13 dpf (C). Frequency of expression in these domains is shown as a percentage of the total number of GFP-positive fish (scored as GFP expression driven by the *hsp70* promoter anywhere at 5 dpf or in the lens at 12-13 dpf) on the y-axis.

A small molecule inhibitor of TGFβ signaling, but not a small molecule inhibitor of Wnt signaling, abolishes enhancer function

Since the predicted Smad3 binding site was necessary for enhancer function, we hypothesized that reducing TGFβ signaling (mediated by Smad3) would result in a loss of expression driven by the enhancer. To pharmacologically inhibit TGFβ signaling, we treated transgenic sticklebacks and zebrafish embryos with SB431542, a specific inhibitor of ALK4/5 phosphatase activity that abrogates TGF-β signaling in zebrafish (Inman et al., 2002; Sun et al., 2006). After 6 days of treatment in sticklebacks, GFP expression driven by the 190 bp enhancer was reduced in a dose-dependent manner in the epithelium, but not mesenchyme, of developing pharyngeal teeth, with tooth epithelial expression abolished at 50 μM and reduced at 25 μM (Figure 4.11A-C). Tooth mesenchymal expression was slightly diminished at 50 μM and apparently unaffected at 25 μM. Similarly, GFP reporter expression was lost in the pharyngeal teeth of newly hatched zebrafish upon treatment with SB431542 from 24 hpf until 5 dpf (Figure 4.11D-F). In sticklebacks, we also saw a reduction, but not complete loss, of pectoral and median fin expression driven by the transgene upon treatment with SB431542 (Figure 4.12), while the reduction was more severe in the fins of zebrafish. Combined with our site-directed mutagenesis of the Smad3 binding site result, these pharmacological data suggest that TGFβ signaling mediated by ALK4/5 (likely signaling via Smad3 binding) is necessary for tooth epithelium enhancer activity. However other signals likely contribute to the expression in the pectoral and median fins and tooth mesenchyme, as drug treatment did not completely abolish these expression domains in sticklebacks.

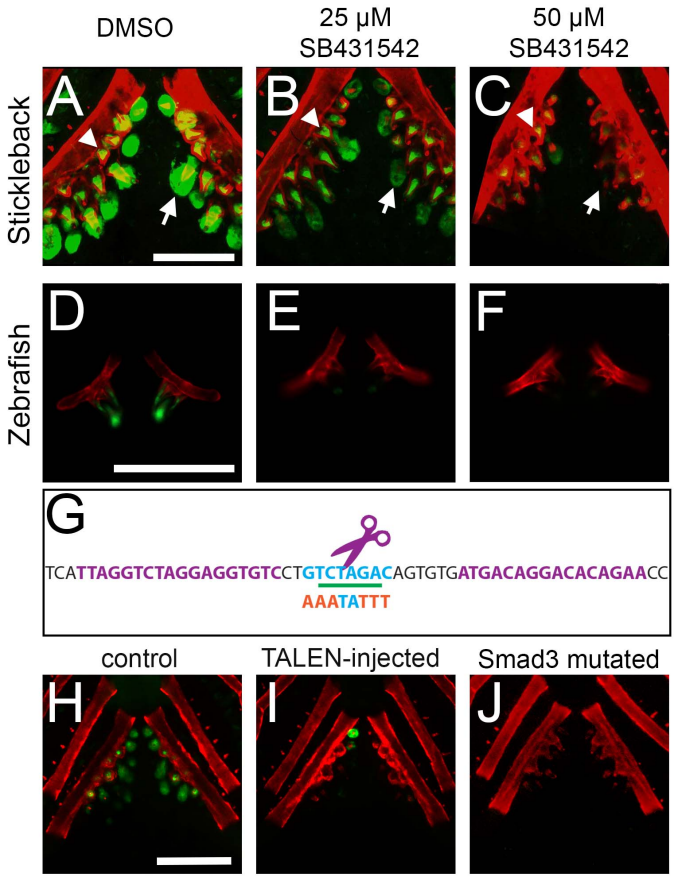


Figure 4.11. Pharmacological disruption of TGFβ signaling or TALEN-induced mutations of the predicted Smad3 binding site reduce enhancer activity.

(A-C) Treatment of stickleback fry for 7 days in SB431542 (an ALK5 inhibitor) severely reduced GFP expression driven by the 190 bp enhancer in a dose-dependent manner. Expression was severely reduced in the epithelia (arrows), but not mesenchyme (asterisks), of pharyngeal teeth at both low (25 μM, B) and high (50 μM, C) doses relative to controls (A). (D-F) SB431542 also eliminated GFP driven by the stickleback enhancer in a zebrafish *trans* environment. (G) The sequence targeted by TALENs contains a predicted Smad3 homodimer binding site (blue). The TALEN binding sites are indicated in purple text and the purple scissors indicate the approximate site of endonuclease activity. The XbaI site used for molecular screening is underlined in green, and the mutagenized sequence of the Smad3 binding site, indicated by orange letters, is shown below. (H-I) Injection of the TALENs into stable transgenic fish carrying the 190 bp reporter construct resulted in near complete loss of GFP expression in 95% of injected animals (I) relative to controls (H). Residual GFP seen in (I) is likely the result of the mosaicism of TALEN-induced lesions. (J). Mutating the predicted Smad3 binding site resulted in a loss of GFP expression in both epithelium and mesenchyme of pharyngeal teeth in 3/3 stickleback lines observed. Bone is fluorescently counterstained with Alizarin red. Scale bars = 200 μm.

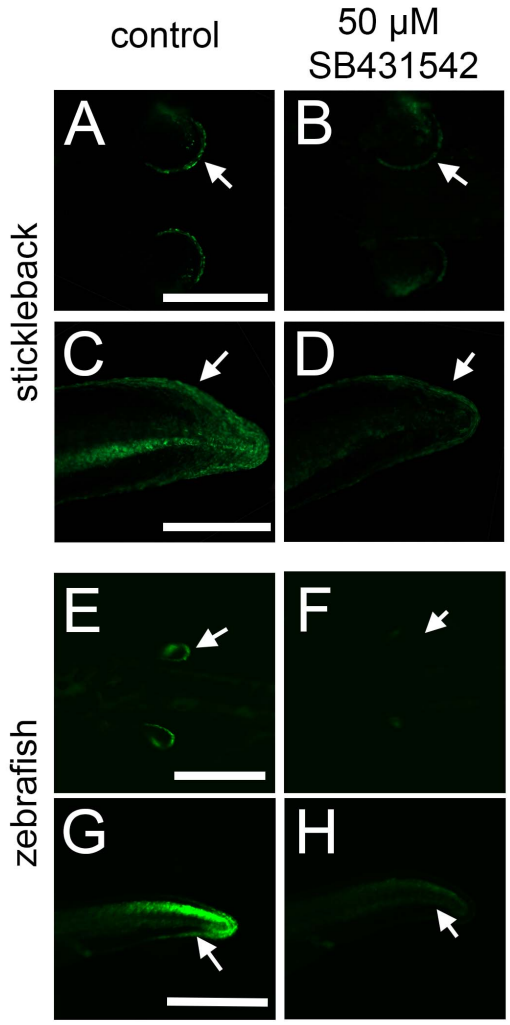


Figure 4.12. SB431542 reduces reporter GFP expression in the median and pectoral fins in both sticklebacks and zebrafish.

Treatment with 50 μM SB431542 reduced, but did not completely eliminate, GFP reporter expression driven by the 190 bp enhancer relative to vehicle (DMSO) controls in the pectoral fins (A, B, E, F) and median fins (C, D, H, G) of both stickleback (A-D) and zebrafish (E-H) embryos. Scale bars = 400 μm.

Since the mutation of TCF/Lef binding sites appeared to decrease enhancer activity in sticklebacks and zebrafish (Figure 4.9-10), we hypothesized that Wnt signaling might be an additional input into the 190 bp *Bmp6* enhancer. To test this hypothesis, we treated transgenic fish with SB431542, XAV939 (a specific inhibitor of the Wnt signaling pathway that is active in zebrafish (Huang et al., 2009)), or both drugs in combination at low and high doses. Treatment with a high-dose combination of XAV939 and SB431542 decreased the standard length of fish (data not shown), possibly indicating a slight developmental delay. With XAV939 or SB431542 treatment alone, there was no effect of the drug on tooth number, suggesting that neither drug alone arrests tooth development. However, the two drugs in combination significantly reduced ventral pharyngeal tooth number (Figure 4.13H), including at the low dose that did not affect fish standard length, suggesting that XAV939 is bioactive in sticklebacks and that reducing Wnt and TGFβ signaling together disrupts tooth development.

There was no obvious qualitatively detectable effect of XAV939 treatment on the intensity of enhancer expression in the teeth, either alone or in combination with SB431542 (Figure 4.13; compare D and E to A, and compare F and G to B and C). However, tooth mesenchymal GFP in the combined drug treatment appeared slightly lower than with SB431542 treatment alone (insets of Figure 4.13). Importantly, we never saw a complete loss of mesenchymal GFP with any drug treatment, but frequently saw complete loss of epithelial GFP with SB431542 treatment. To quantify the effect of drug treatment on epithelial GFP expression, we counted the number of GFP^+ tooth epithelia (regardless of fluorescent intensity) in each treatment and expressed it as a ratio to the total number of Alizarin red-stained teeth. XAV939 had no effect on the relative number of GFP^+ epithelia, while SB431542 had a strong, dose-dependent effect (Figure 4.13I). In combination with SB431542, there was no additional effect of XAV939 on reporter expression (GFP^+ epithelia in the combination treatments did not differ from treatment with SB431542 alone). Combined, our results suggest that SB431542, but not XAV939, affects enhancer activity and that simultaneous inhibition of Wnt and TGFβ signaling affects tooth development.

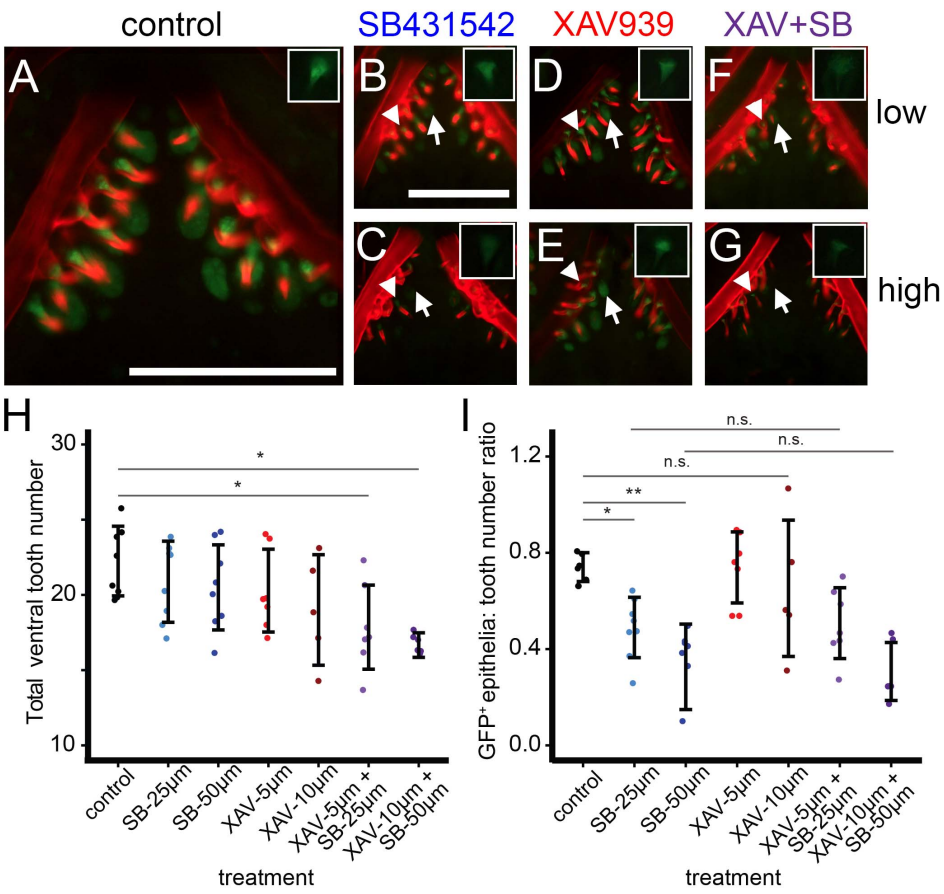


Figure 4.13. Wnt signaling is not required for enhancer function, but Wnt and TGFβ are required for tooth development.

Newly hatched stickleback fry were treated with DMSO (control, A), SB431542 (B-C), XAV939 (D-E), or a combination of the two drugs at low (25µM for SB431542 and 5 µM for XAV939, F) or high (50 µM for SB431542 or 10 µM XAV939, G) doses for 5 days. Main panels show Alizarin red and GFP for the ventral tooth plate; insets show GFP only for mesenchyme of a single tooth from the dorsal tooth plate. (B, C) SB431542 reduced GFP in tooth epithelia (arrows) relative to control (A, and see Figure 4.10). However, mesenchymal GFP (arrowhead, inset) was less severely reduced. (D, E) XAV939 alone did not affect GFP expression in epithelia (arrows) or mesenchyme (arrowheads) at either dose. (F, G) No strong additional effect on GFP expression was seen when XAV939 and SB431542 were combined, though mesenchymal GFP appeared slightly lower in the combined dose. (H) A combination of SB431542 and XAV939 significantly reduced ventral pharyngeal tooth number. (I) Treatment with SB431542, but not XAV939, decreased the number of green tooth epithelia relative to total ventral teeth (ratio is expressed as a decimal). XAV939 had no additional effect on green epithelia in combination with SB431542. Tukey HSD P-values of relevant comparisons are shown above with asterisks (*=P<0.05, ** =P<0.0005, n.s.=P>0.05). Scale bars = 200 µm.

The 190 bp enhancer is necessary for Bmp6 expression

As an additional test of the importance of the predicted Smad3 binding site, we generated a pair of TALENs designed to induce mutations in this region of the enhancer (see Figure 4.11G). This pair of TALENs was highly efficient at producing lesions, detected molecularly by loss of an XbaI restriction site, and confirmed by Sanger sequencing in a subset of individuals (Table 4.4; example deletions shown in Figure 4.14E). Upon injection of these TALENs into a stable transgenic line of the 190 bp enhancer driving GFP, 95% of animals (40 of 42) showed partial or full loss of GFP fluorescence in the pectoral fins and median fin expression at 5 dpf. In those same animals, 95% of animals (39 of 41) also showed partial or complete loss of oral and/or pharyngeal tooth expression at 12-13 dpf (see example in Figure 4.11I). Thus, the lesions generated by these TALENs are highly effective at disrupting activity driven by this 190 bp enhancer.

We next tested whether the sequence targeted by the TALENs was necessary for *Bmp6* expression by injecting the TALENs into a stable transgenic line of the *Bmp6:GFP* BAC reporter. 91% (61/67) of animals had a reduction or complete loss of pectoral and median fin expression, and 89% (8/9) of dissected tooth plates showed severe reductions of GFP expression in the pharyngeal teeth (representative animals shown in Figure 4.14 F-K). Notably, GFP expression in the embryonic and juvenile heart was detectable at seemingly unaffected levels in all animals, suggesting that the enhancer is not necessary for this expression domain. Additionally, gill expression appeared to be reduced but not completely eliminated in all animals observed (n=6), and gill raker expression was only slightly reduced. These data suggest the enhancer is required for some (e.g. pectoral fin, median fin, tooth epithelium), but not all domains of *Bmp6* expression.

Next, we tested the role of the enhancer in driving endogenous *Bmp6* expression by performing *in situ* hybridization for *Bmp6* in fish *trans*-heterozygous for different TALEN-induced mutations in the predicted Smad3 binding site (Figure 4.14E). In these *trans*-heterozygous fish, expression of *Bmp6* was dramatically reduced in fins, tooth epithelia and gills, but gill raker expression appeared similar to wild-type controls (Figure 4.14L-Q). Despite the severe loss of *Bmp6* expression in tooth epithelia in mutant fish, expression in the mesenchyme of developing teeth was still detectable, although at apparently reduced levels (Figure 4.14N-O). Thus, this enhancer is required to maintain normal levels of *Bmp6* expression in developing fins and tooth epithelia.

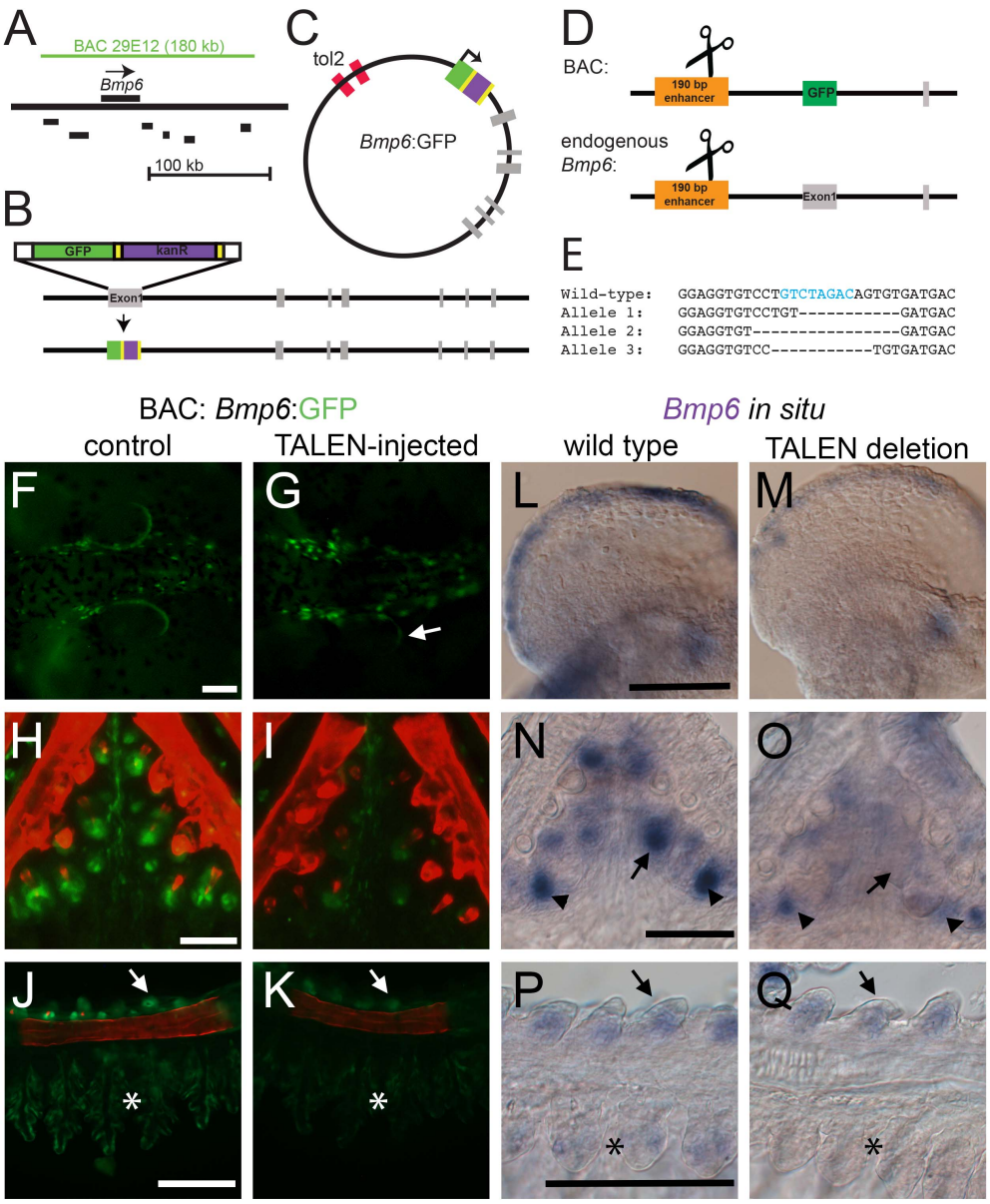


Figure 4.14. The 5' 190 bp enhancer is necessary for Bmp6 expression.

(A) Schematic of the genomic location of the 180 kb CHORI BAC29E12 with respect to *Bmp6* and nearby genes (coding regions shown in black are *Ipo4*, *Pdcd6*, *Txndc5*, *Muted*, *Eefie1*, and *Slc35b3* from left to right). (B) Recombineering strategy for introducing GFP into the first exon *of Bmp6;* grey bars indicate exons. (C) Final circular BAC with inverted Tol2 sites for transposition and GFP reporter (not to scale). (D) Strategy for introducing TALEN lesions into the 190 bp 5' enhancer. The same TALENs were used to target the enhancer in stable transgenic BAC fish and at the endogenous *Bmp6* locus (diagram not to scale). (E) Sequences of stable mutant enhancer alleles. For the endogenous locus targeting, F2 fish trans-heterozygous for two different enhancer mutations were generated. Fish in (M) carried alleles 1 and 2; fish in (O) and (Q) carried alleles 1 and 3.

The predicted Smad3 binding site is indicated with blue text in the wild type sequence. (F, G) In the reporter BAC, TALEN injection frequently severely reduced GFP expression from the pectoral fin relative to controls at 5 dpf. A small patch of mosaic, unaffected GFP is indicated with the arrow in (G). (H, I) TALEN injection also eliminated much of the *Bmp6* tooth expression (I). (J, K) GFP expression was also reduced in gills (asterisk) and slightly reduced in the gill rakers (arrowhead). (L-M). Mutations in the enhancer caused a reduction in pectoral fin expression relative to wild-type siblings. (N, O) *Bmp6* expression was also lost in tooth epithelia (arrows), but was not entirely lost in mesenchyme (arrowheads). (P, Q) Expression was also noticeably reduced in gills (asterisk), though gill raker expression (arrows) appears similar to wild-type sibling controls. Scale bars = 100 μm.

clutch number	**generation**	**% molecular lesions**
1	F0 injected	17/17 (100%)
2	F0 injected	19/19 (100%)
3	F0 injected	9/10 (90%)
Average	*F0 injected*	*98%*
4	F1 outcross	2/10 (20%)
5	F1 outcross	5/10 (50%)
6	F1 outcross	7/10 (70%)
7	F1 outcross	6/9 (67%)
8	F1 outcross	9/10 (90%)
Average	*F1 outcross*	*59%*

Table 4.4. Efficiency of molecular lesions produced by TALENs.

A subset of each TALEN clutch was screened at 2 dpf for TALEN-induced lesions. Molecular lesions were identified by PCR amplification with Gac_190_for and Gac_72_rev and digestion with XbaI (see Figure 4.11G for illustration). An undigested band indicated the presence of a TALEN-induced lesion. Lesions were confirmed by Sanger sequencing for a subset of F1 animals, including parents of animals used for *in situ* hybridization (see Figure 6E).

TGFβ signaling is necessary for normal Bmp6 expression levels

Since enhancer activity was lost upon treatment with a TGFβ inhibitor, and the enhancer is required for normal *Bmp6* expression, we predicted that endogenous *Bmp6* expression would likewise be reduced upon inhibition of TGFβ signaling. By *in situ* hybridization, pectoral fin and tooth epithelium expression of *Bmp6* were both reduced upon 100 µM SB431542 treatment (Figure 4.15A-D). SB431542 treatment also reduced GFP expression in reporter BAC animals in fins and teeth (Figure 4.15E-H). The effect of the drug on BAC-driven GFP was not robustly observed with a 50 µM treatment (data not shown), despite the strong effect that this dose had on enhancer expression (Figure 4.11). Together these data support a model in which TGFβ signaling is required for *Bmp6* expression in teeth and fins and exerts its effect through the putative Smad3 binding site that is necessary for enhancer function.

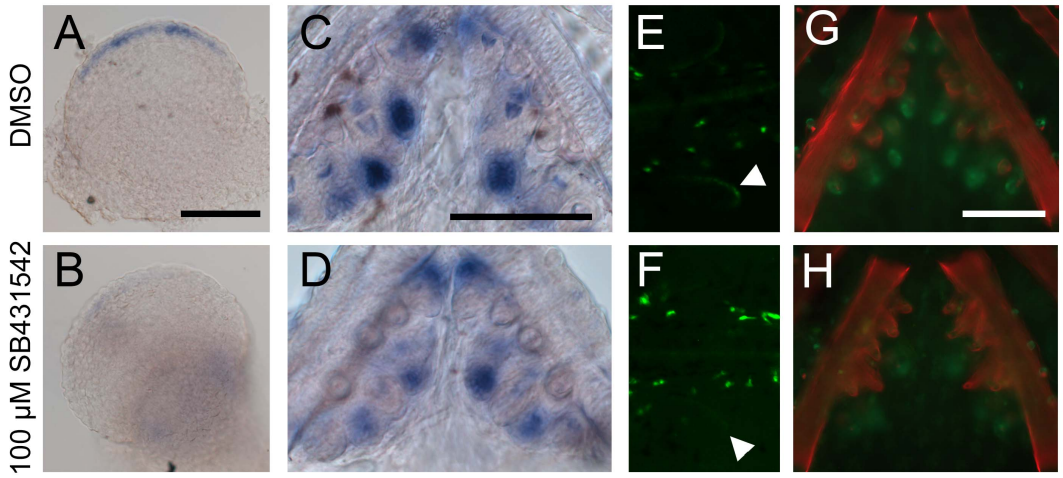


Figure 4.15. Treatment with TGFβ inhibitor SB431542 reduces Bmp6 expression.
(A-D) Sticklebacks were treated with 100 µM SB431542 or vehicle control from 2 to 5 dpf or for 5 days post-hatching, and *Bmp6* expression was assayed by *in situ* hybridization. Drug treatment severely reduced *Bmp6* expression in fins (A, B) and also reduced *Bmp6* expression in tooth epithelia (C, D). Likewise, GFP driven by the *Bmp6* locus in the reporter BAC was also reduced in fins (arrowheads in E, F) and teeth (G, H) after SB431542 treatment. Scale bars = 100 µm.

Discussion

A short, conserved enhancer with pleiotropic expression domains required for Bmp6 tooth and fin expression

Here we have identified a 190 base pair enhancer that is highly conserved in teleosts and is both necessary and sufficient for tooth and fin expression of stickleback *Bmp6*. Site-directed mutagenesis of a predicted Smad3 binding site and pharmacological experiments suggest this enhancer is TGFβ-responsive. Though this enhancer drives expression in several of *Bmp6*'s endogenous domains, our results suggest that like other *Bmp* genes, stickleback *Bmp6* contains a complex *cis*-regulatory architecture composed of multiple modules driving expression in different domains. We detected embryonic expression domains of *Bmp6* by *in situ* hybridization, such as the eye, ear, diencephalon, and notochord, that were not observed in the BAC reporter line, suggesting that the regulatory elements controlling these domains lie outside of the 180 kb of stickleback DNA included in the BAC. Furthermore, while TALEN mutations severely reduced expression in the fins and teeth, every BAC reporter fish injected with TALENs had GFP expression in the heart, suggesting that the enhancer is not required for heart expression. Thus, the short enhancer presented here contributes to a subset of the endogenous *Bmp6* expression domains, with other domains likely driven by other enhancers greater than ~100 kb away. Evidence for long range distant enhancers of stickleback *Bmp6* is expected, given the frequent finding of long distance regulatory elements for developmental regulatory genes, including other vertebrate *Bmp* genes (reviewed in Preziger and Mortlock, 2009). Interestingly, despite the presence of redundant "shadow" enhancers found in many genes (Calle-Mustienes et al., 2005; Marinić et al., 2013; Perry et al., 2010), this enhancer appears to be required for several *Bmp6* expression domains; additional enhancers did not appear to sufficiently compensate in driving *Bmp6* expression when the 5' enhancer was targeted with TALENs.

Another teleost tooth/fin enhancer has been described with overall similar expression patterns observed in this *Bmp6* enhancer. In zebrafish, an FGF-responsive enhancer mediates *Dlx2* expression in teeth and median and pectoral fins (Jackman and Stock, 2006). Additionally, in mice, a *Bmp4* enhancer drives tooth epithelium and limb bud expression by responding to Pitx and Msx homeodomains (Jumlongras et al., 2012). The shared fin/limb and tooth expression domains of these *cis*-regulatory elements and the one described here suggest that fin and tooth development share multiple *cis*-regulatory networks, with at least three signaling pathways (FGF, Pitx/Msx, and TGFß) involved in generating similar gene expression readouts in teeth and fins/limbs. Gene expression patterns of paired fins are thought to be co-opted from median fin expression domains in agnathans (Freitas et al., 2006). The *Bmp6* enhancer presented here appears to be teleost-specific, as we did not find evidence of this conserved enhancer sequence in the genomes of lamprey, elephant shark, or spotted gar. Thus, our results suggest that teleosts may have secondarily coopted components of a gene regulatory network in developing median and pectoral fins and teeth.

Elucidating the *cis*-regulatory architecture of stickleback *Bmp6* and evolved changes in *Bmp6*'s *cis*-regulatory architecture will help test the hypothesis that evolved

changes in *Bmp6 cis*-regulation underlie the evolved increases in freshwater stickleback tooth number we previously described (Cleves et al., 2014). Although the 190 bp core *Bmp6* enhancer presented here contains no nucleotide differences between low-toothed marine and high-toothed freshwater sticklebacks, several nucleotide differences exist in the sequence flanking the enhancer, which might contribute to the *cis*-regulatory differences observed between marine and freshwater alleles of *Bmp6.* Future studies will focus on whether these differences result in differential *cis*-regulatory activity between the marine and freshwater alleles of *Bmp6*.

Conservation and turnover of cis- and trans-regulatory information

It has been proposed that the *cis*-regulatory architecture of developmental control genes often consist of multiple independent modules, each of which drives expression in a particular tissue or cell type (Carroll, 2008; Stern, 2000). Because the *Bmp6* enhancer drives multiple anatomical expression domains and is only partially conserved to zebrafish, we hypothesized that domains may have been sequentially added to the enhancer during teleost evolution, and that the different anatomical domains would be separable. Contrary to these predictions, our site directed mutagenesis and subcloning experiments of the stickleback *Bmp6* enhancer appeared to affect all or none of the different expression domains, suggesting the different anatomical domains might not be separable and instead reflect ability to respond to a signal or signals present in multiple tissues.

Furthermore, enhancers from all four teleost species tested were sufficient to drive fin and tooth expression in zebrafish. However, the zebrafish enhancer, the most evolutionary divergent enhancer tested in this study, did not function robustly in sticklebacks, suggesting that the *trans* factors driving expression might have changed during the divergence of the two species. Similarly, testing a zebrafish *Dlx2* tooth and fin enhancer in both zebrafish and Mexican tetra revealed that loss of oral *Dlx2* expression in zebrafish is caused by changes in *trans* factors, as the *Dlx2* zebrafish tooth enhancer is active in tetra oral teeth (Jackman and Stock, 2006). In both *C. elegans* and *Drosophila*, transgenic testing of *cis*-regulatory elements from two fly or worm species in both fly or worm species revealed that the greater the evolutionary distance separating two regulatory elements, the more likely upstream *trans* differences are to have evolved (Gordon and Ruvinsky, 2012). But, subtle changes in *trans*-acting factors can maintain similar expression patterns despite *cis* changes in divergent lineages (Barrière et al., 2012). Our results suggest a combination of conservation and divergence of *trans* factors, as stickleback sequence worked robustly in zebrafish, but zebrafish sequence was not functional in stickleback. Additionally, SB431542 treatment affected the stickleback enhancer in zebrafish more severely than in stickleback. Even at a low dose of SB431542 (25 µM), the enhancer was completely shut off in both epithelia and mesenchyme of zebrafish teeth (see Figure 4.11E-F). This result supports potential *trans* regulatory divergence between stickleback and zebrafish, because it suggests that the enhancer's expression may be more sensitive to TGFß signaling in zebrafish than in stickleback.

A role for TGFß in the regulation of BMPs

To our knowledge, this study is the first to support a role for TGFβ signaling in controlling Bmp signaling via a *cis*-regulatory input. Conditional deletion of *Tgfbr1* (*Alk5*) in mouse neural crest lineages results in reduced expression of *Bmp4* and delayed tooth initiation (Zhao et al., 2008); however, the mechanism of this interaction has not been described. Other studies have shown both positive and negative correlations between *Bmp6* expression and TGFβ levels: *Smad3* -/- chondrocytes have reduced *Bmp6* expression (Li et al., 2006), whereas *Bmp6* expression is increased in *Smad3* -/- tendons undergoing tissue repair (Katzel et al., 2011). Our data suggest that in sticklebacks, TGFß signaling activates *Bmp6* expression in multiple tissues via a predicted Smad3 binding site. In teeth, blocking TGFß signaling using the inhibitor SB431542 caused loss of epithelial reporter expression, but the effect on the mesenchymal expression was less severe (Figure 4.11C, Figure 4.13). The same pattern was observed in endogenous *Bmp6* expression (Figure 4.14O). This result suggests that epithelial and mesenchymal *Bmp6* expression domains respond to partially different signaling pathways, with epithelial expression much more sensitive to TGFß disruption.

We observed that a higher dose of TGFβ inhibitor SB431542 was required to shut off endogenous *Bmp6* expression relative to expression driven solely by the 190 bp enhancer. While a 50 μM treatment almost completely eliminated enhancer expression (Figure 4.11), at this dose we did not observe a strong difference in GFP expression driven by the reporter BAC. Only at the higher dose of 100 μM did we observe a change in BAC reporter expression and endogenous *Bmp6* expression (Figure 4.15). This finding suggests that in its native genomic context, the enhancer may be less sensitive to TGFβ signaling perturbations than when it is isolated in a reporter construct. There may be additional non-TGFβ regulatory elements that drive *Bmp6* expression in the same tooth and fin domains such that a decrease in TGFβ signaling has a less obvious effect at lower doses. Furthermore, the effect of SB431542 treatment on endogenous *Bmp6* expression and BAC reporter expression was not as dramatic as deletion of the Smad3 binding site with TALENs (compare Figure 4.14 to Figure 4.15). This finding suggests that other non-TGFß factors may bind sequences immediately surrounding the Smad3 binding site to drive enhancer expression. However, the predicted Smad3 site is absolutely required, as loss of this site completely eliminates enhancer activity (Figure 4.11J).

Combined effects of Wnt and TGFß on tooth development

Although our site-directed mutagenesis experiment indicated that TCF/Lef predicted binding sites might be important for enhancer function (Figure 4.10), pharmacological testing with XAV939 did not support the hypothesis that the enhancer requires Wnt signaling inputs for enhancer function. A stable line of zebrafish containing the TCF/Lef mutated reporter also drove robust reporter expression in fins and teeth, providing a second piece of evidence that the enhancer does not require Wnt input. This result was somewhat surprising, as the expression domains driven by the *Bmp6* enhancer are similar to a TCF reporter zebrafish line (Shimizu et al., 2012). The reduction in activity seen from mutating the TCF/Lef sites may have been caused by other unknown binding sites overlapping the mutated base pairs, by inadvertently creating repressive motifs, or

by somehow altering the binding of the Smad3 complex. The mutations may have affected the level, but not pattern, of GFP expression, making the construct appear less robust in our transient transgenic assay. We did note that combined treatment with XAV939 and SB431542 caused a slight decrease in mesenchymal tooth GFP expression (see insets of Figure 4.13), however, this effect was less reproducible than the complete loss of epithelial expression seen upon SB431542 treatment alone.

The combination treatment with SB431542 and XAV939 did reduce tooth number in sticklebacks, suggesting that Wnt and TGFβ signaling pathways together are required for maintaining normal tooth development and patterning. In mice, as well as in diphyodont humans and polyphyodonts including snakes and alligators, Wnt signaling is required for tooth formation and replacement (Adaimy et al., 2007; Bohring et al., 2009; Gaete and Tucker, 2013; Genderen et al., 1994; Liu et al., 2008; Wu et al., 2013). In mice, TGFß signaling is also required for tooth development (Ferguson et al., 1998, 2001; Oka et al., 2007). Antisense abrogation of both *TGFB2* and *TGFBRII* in cultured mandibles resulted in accelerated tooth formation (Chai et al., 1994, 1999), however the *TGFB2* knockout mouse has no reported tooth phenotype (Sanford et al., 1997). While the *TGFBRII* knockout dies prior to tooth formation (Oshima et al., 1996), conditional ablation in neural crest cells prevents terminal differentiation of odontoblasts (Oka et al., 2007), while conditional ablation in *Osx*-expressing odontoblasts revealed a necessary role for *TGFBRII* in molar root development (Wang et al., 2013). Furthermore, Wnt and TGFß signaling are required to activate *Eda* and *Edar* in appropriate patterns in the developing tooth germs (Laurikkala et al., 2001). However, to our knowledge, this study is the first to show a partially redundant requirement for TGFß and Wnt during tooth development, as only XAV939 and SB431542 doubly treated fish had reduced tooth numbers. Future studies of this enhancer will further test the hypothesis that this enhancer responds to TGFß signaling to control *Bmp6* expression during tooth and fin development.

Conclusions

We have identified a 190 base pair conserved enhancer required for tooth, fin, and other expression domains of stickleback *Bmp6*. Site directed mutagenesis and pharmacology experiments support the hypothesis that this enhancer responds to TGFß signaling via a Smad3 binding site. Expression driven by this enhancer in tooth epithelial cells appears more sensitive to TGFß levels than expression in tooth mesenchymal cells. To our knowledge, this is the first demonstration of a likely *cis*-regulatory link between TGFß signaling and *Bmp* expression in teeth. *In vivo* deletion of this enhancer using TALENs caused severe disruption of *Bmp6* expression in fins and tooth epithelia, suggesting this enhancer is required for normal expression patterns in a subset of *Bmp6*'s endogenous domains. Finally, we demonstrate that a combination of TGFß signaling and Wnt signaling is required for normal tooth development in sticklebacks.

Acknowledgements

Phillip Cleves was instrumental in getting this project started by cloning the first 3 kb transgenic construct. Kevin Schwalbach recombineered BACs and he and Nick Ellis helped generate BAC stable lines. James Hart assisted with the site directed mutagenesis of TCF/Lef binding sites. Craig Miller made substantial contributions to the writing of this manuscript.

This work was supported by NIH R01 #DE021475. We thank David Kingsley for support and advice on BAC isolation, Tim Howes and David Kingsley for the generous gift of the Tol2/*hsp70* backbone, Daniel Schlenk and Anita Kuepper for providing medaka specimens, Amy Strom and Katie Sieverman for assistance in cloning the TALENs, Natasha Naidoo for assistance in cloning the medaka reporter construct, and Lisa Kronstad for providing the site-directed mutagenesis protocol.

References:

Aberg, T., Wozney, J., and Thesleff, I. (1997). Expression patterns of bone morphogenetic proteins (Bmps) in the developing mouse tooth suggest roles in morphogenesis and cell differentiation. Dev. Dyn. *210*, 383–396.

Abzhanov, A., Protas, M., Grant, B.R., Grant, P.R., and Tabin, C.J. (2004). Bmp4 and morphological variation of beaks in Darwin's finches. Science *305*, 1462–1465.

Adaimy, L., Chouery, E., Mégarbané, H., Mroueh, S., Delague, V., Nicolas, E., Belguith, H., de Mazancourt, P., and Mégarbané, A. (2007). Mutation in WNT10A Is Associated with an Autosomal Recessive Ectodermal Dysplasia: The Odonto-onycho-dermal Dysplasia. Am. J. Hum. Genet. *81*, 821–828.

Adams, D., Karolak, M., Robertson, E., and Oxburgh, L. (2007). Control of kidney, eye and limb expression of Bmp7 by an enhancer element highly conserved between species. Dev. Biol. *311*, 679–690.

Albertson, R.C., Streelman, J.T., Kocher, T.D., and Yelick, P.C. (2005). Integration and evolution of the cichlid mandible: The molecular basis of alternate feeding strategies. Proc. Natl. Acad. Sci. U. S. A. *102*, 16287–16292.

Andl, T., Ahn, K., Kairo, A., Chu, E.Y., Wine-Lee, L., Reddy, S.T., Croft, N.J., Cebra-Thomas, J.A., Metzger, D., Chambon, P., et al. (2004). Epithelial Bmpr1a regulates differentiation and proliferation in postnatal hair follicles and is essential for tooth development. Development. *131*, 2257–2268.

Andriopoulos, B., Corradini, E., Xia, Y., Faasse, S.A., Chen, S., Grgurevic, L., Knutson, M.D., Pietrangelo, A., Vukicevic, S., Lin, H.Y., et al. (2009). BMP6 is a key endogenous regulator of hepcidin expression and iron metabolism. Nat. Genet. *41*, 482–487.

Barrière, A., Gordon, K.L., and Ruvinsky, I. (2012). Coevolution within and between Regulatory Loci Can Preserve Promoter Function Despite Evolutionary Rate Acceleration. PLoS Genet *8*, e1002961.

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

Bellusci, S., Henderson, R., Winnier, G., Oikawa, T., and Hogan, B.L. (1996). Evidence from normal expression and targeted misexpression that bone morphogenetic protein (Bmp-4) plays a role in mouse embryonic lung morphogenesis. Development *122*, 1693–1702.

Biggs, L.C., and Mikkola, M.L. (2014). Early inductive events in ectodermal appendage morphogenesis. Semin. Cell Dev. Biol. *25–26*, 11–21.

Bohring, A., Stamm, T., Spaich, C., Haase, C., Spree, K., Hehr, U., Hoffmann, M., Ledig, S., Sel, S., Wieacker, P., et al. (2009). WNT10A Mutations Are a Frequent Cause of a Broad Spectrum of Ectodermal Dysplasias with Sex-Biased Manifestation Pattern in Heterozygotes. Am. J. Hum. Genet. *85*, 97–105.

Botchkarev, V.A., Botchkareva, N.V., Roth, W., Nakamura, M., Chen, L.H., Herzog, W., Lindner, G., McMahon, J.A., Peters, C., Lauster, R., et al. (1999). Noggin is a mesenchymally derived stimulator of hair-follicle induction. Nat. Cell Biol. *1*, 158–164.

Calle-Mustienes, E. de la, Feijóo, C.G., Manzanares, M., Tena, J.J., Rodríguez-Seguel, E., Letizia, A., Allende, M.L., and Gómez-Skarmeta, J.L. (2005). A functional survey of the enhancer activity of conserved non-coding sequences from vertebrate Iroquois cluster gene deserts. Genome Res. *15*, 1061–1072.

Carroll, S.B. (2008). Evo-Devo and an Expanding Evolutionary Synthesis: A Genetic Theory of Morphological Evolution. Cell *134*, 25–36.

Cermak, T., Doyle, E.L., Christian, M., Wang, L., Zhang, Y., Schmidt, C., Baller, J.A., Somia, N.V., Bogdanove, A.J., and Voytas, D.F. (2011). Efficient design and assembly of custom TALEN and other TAL effector-based constructs for DNA targeting. Nucleic Acids Res. gkr218.

Chai, Y., Mah, A., Crohin, C., Groff, S., Bringas, P., Le, T., Santos, V., and Slavkin, H.C. (1994). Specific transforming growth factor-beta subtypes regulate embryonic mouse Meckel's cartilage and tooth development. Dev. Biol. *162*, 85–103.

Chai, Y., Zhao, J., Mogharei, A., Xu, B., Bringas Jr., P., Shuler, C., and Warburton, D. (1999). Inhibition of transforming growth factor-β type II receptor signaling accelerates tooth formation in mouse first branchial arch explants. Mech. Dev. *86*, 63–74.

Chandler, R.L., Chandler, K.J., McFarland, K.A., and Mortlock, D.P. (2007). Bmp2 Transcription in Osteoblast Progenitors Is Regulated by a Distant 3′ Enhancer Located 156.3 Kilobases from the Promoter. Mol. Cell. Biol. *27*, 2934–2951.

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

Cleves, P.A., Ellis, N.A., Jimenez, M.T., Nunez, S.M., Schluter, D., Kingsley, D.M., and Miller, C.T. (2014). Evolved tooth gain in sticklebacks is associated with a cis-regulatory allele of Bmp6. Proc. Natl. Acad. Sci. *111*, 13912–13917.

Dassule, H.R., and McMahon, A.P. (1998). Analysis of Epithelial–Mesenchymal Interactions in the Initial Morphogenesis of the Mammalian Tooth. Dev. Biol. *202*, 215–227.

Dathe, K., Kjaer, K.W., Brehm, A., Meinecke, P., Nürnberg, P., Neto, J.C., Brunoni, D., Tommerup, N., Ott, C.E., Klopocki, E., et al. (2009). Duplications Involving a Conserved Regulatory Element Downstream of BMP2 Are Associated with Brachydactyly Type A2. Am. J. Hum. Genet. *84*, 483–492.

Dendooven, A., van Oostrom, O., van der Giezen, D.M., Willem Leeuwis, J., Snijckers, C., Joles, J.A., Robertson, E.J., Verhaar, M.C., Nguyen, T.Q., and Goldschmeding, R. (2011). Loss of Endogenous Bone Morphogenetic Protein-6 Aggravates Renal Fibrosis. Am. J. Pathol. *178*, 1069–1079.

Doyle, E.L., Booher, N.J., Standage, D.S., Voytas, D.F., Brendel, V.P., VanDyk, J.K., and Bogdanove, A.J. (2012). TAL Effector-Nucleotide Targeter (TALE-NT) 2.0: tools for TAL effector design and target prediction. Nucleic Acids Res. *40*, W117–W122.

Dudley, A.T., Godin, R.E., and Robertson, E.J. (1999). Interaction between FGF and BMP signaling pathways regulates development of metanephric mesenchyme. Genes Dev. *13*, 1601–1613.

Farre, D., Roset, R., Huerta, M., Adsuara, J.E., Rosello, L., Alba, M.M., and Messeguer, X. (2003). Identification of patterns in biological sequences at the ALGGEN server: PROMO and MALGEN. Nucleic Acids Res. *31*, 3651–3653.

Feng, J.Q., Zhang, J., Tan, X., Lu, Y., Guo, D., and Harris, S.E. (2002). Identification of Cis-DNA Regions Controlling Bmp4 Expression during Tooth Morphogenesis in vivo. J. Dent. Res. *81*, 6–10.

Ferguson, C.A., Tucker, A.S., Christensen, L., Lau, A.L., Matzuk, M.M., and Sharpe, P.T. (1998). Activin is an essential early mesenchymal signal in tooth development that is required for patterning of the murine dentition. Genes Dev. *12*, 2636–2649.

Ferguson, C.A., Tucker, A.S., Heikinheimo, K., Nomura, M., Oh, P., Li, E., and Sharpe, P.T. (2001). The role of effectors of the activin signalling pathway, activin receptors IIA and IIB, and Smad2, in patterning of tooth. Development *128*, 4605–4613.

Fisher, S., Grice, E.A., Vinton, R.M., Bessling, S.L., Urasaki, A., Kawakami, K., and McCallion, A.S. (2006). Evaluating the biological relevance of putative enhancers using Tol2 transposon-mediated transgenesis in zebrafish. Nat. Protoc. *1*, 1297–1305.

Fraser, G.J., Bloomquist, R.F., and Streelman, J.T. (2013). Common developmental pathways link tooth shape to regeneration. Dev. Biol. *377*, 399–414.

Freitas, R., Zhang, G., and Cohn, M.J. (2006). Evidence that mechanisms of fin development evolved in the midline of early vertebrates. Nature *442*, 1033–1037.

Fujimori, S., Novak, H., Weissenböck, M., Jussila, M., Gonçalves, A., Zeller, R., Galloway, J., Thesleff, I., and Hartmann, C. (2010). Wnt/β-catenin signaling in the dental mesenchyme regulates incisor development by regulating Bmp4. Dev. Biol. *348*, 97–106.

Gaete, M., and Tucker, A.S. (2013). Organized Emergence of Multiple-Generations of Teeth in Snakes Is Dysregulated by Activation of Wnt/Beta-Catenin Signalling. PLoS ONE *8*, e74484.

Genderen, C. van, Okamura, R.M., Fariñas, I., Quo, R.G., Parslow, T.G., Bruhn, L., and Grosschedl, R. (1994). Development of several organs that require inductive epithelial-mesenchymal interactions is impaired in LEF-1-deficient mice. Genes Dev. *8*, 2691–2703.

Gordon, K.L., and Ruvinsky, I. (2012). Tempo and Mode in Evolution of Transcriptional Regulation. PLoS Genet *8*, e1002432.

Gudbjartsson, D.F., Walters, G.B., Thorleifsson, G., Stefansson, H., Halldorsson, B.V., Zusmanovich, P., Sulem, P., Thorlacius, S., Gylfason, A., Steinberg, S., et al. (2008). Many sequence variants affecting diversity of adult human height. Nat. Genet. *40*, 609–615.

Guenther, C., Pantalena-Filho, L., and Kingsley, D.M. (2008). Shaping Skeletal Growth by Modular Regulatory Elements in the Bmp5 Gene. PLoS Genet. *4*.

Hogan, B.L. (1996). Bone morphogenetic proteins: multifunctional regulators of vertebrate development. Genes Dev. *10*, 1580–1594.

Houlston, R.S., Webb, E., Broderick, P., Pittman, A.M., Di Bernardo, M.C., Lubbe, S., Chandler, I., Vijayakrishnan, J., Sullivan, K., Penegar, S., et al. (2008). Meta-analysis of genome-wide association data identifies four new susceptibility loci for colorectal cancer. Nat. Genet. *40*, 1426–1435.

Huang, S.-M.A., Mishina, Y.M., Liu, S., Cheung, A., Stegmeier, F., Michaud, G.A., Charlat, O., Wiellette, E., Zhang, Y., Wiessner, S., et al. (2009). Tankyrase inhibition stabilizes axin and antagonizes Wnt signalling. Nature *461*, 614–620.

Inman, G.J., Nicolás, F.J., Callahan, J.F., Harling, J.D., Gaster, L.M., Reith, A.D., Laping, N.J., and Hill, C.S. (2002). SB-431542 Is a Potent and Specific Inhibitor of Transforming Growth Factor-β Superfamily Type I Activin Receptor-Like Kinase (ALK) Receptors ALK4, ALK5, and ALK7. Mol. Pharmacol. *62*, 65–74.

Jackman, W.R., and Stock, D.W. (2006). Transgenic analysis of Dlx regulation in fish tooth development reveals evolutionary retention of enhancer function despite organ loss. Proc. Natl. Acad. Sci. *103*, 19390–19395.

Jackman, W.R., Davies, S.H., Lyons, D.B., Stauder, C.K., Denton-Schneider, B.R., Jowdry, A., Aigler, S.R., Vogel, S.A., and Stock, D.W. (2013). Manipulation of Fgf and Bmp signaling in teleost fishes suggests potential pathways for the evolutionary origin of multicuspid teeth. Evol. Dev. *15*, 107–118.

Jumlongras, D., Lachke, S.A., O'Connell, D.J., Aboukhalil, A., Li, X., Choe, S.E., Ho, J.W.K., Turbe-Doan, A., Robertson, E.A., Olsen, B.R., et al. (2012). An Evolutionarily Conserved Enhancer Regulates Bmp4 Expression in Developing Incisor and Limb Bud. PLoS ONE *7*, e38568.

Jung, H.-S., Francis-West, P.H., Widelitz, R.B., Jiang, T.-X., Ting-Berreth, S., Tickle, C., Wolpert, L., and Chuong, C.-M. (1998). Local Inhibitory Action of BMPs and Their Relationships with Activators in Feather Formation: Implications for Periodic Patterning. Dev. Biol. *196*, 11–23.

Justice, C.M., Yagnik, G., Kim, Y., Peter, I., Jabs, E.W., Erazo, M., Ye, X., Ainehsazan, E., Shi, L., Cunningham, M.L., et al. (2012). A genome-wide association study identifies susceptibility loci for nonsyndromic sagittal craniosynostosis near BMP2 and within BBS9. Nat. Genet. *44*, 1360–1364.

Katzel, E.B., Wolenski, M., Loiselle, A.E., Basile, P., Flick, L.M., Langstein, H.N., Hilton, M.J., Awad, H.A., Hammert, W.C., and O'Keefe, R.J. (2011). Impact of Smad3 loss of function on scarring and adhesion formation during tendon healing. J. Orthop. Res. Off. Publ. Orthop. Res. Soc. *29*, 684–693.

Kavanagh, K.D., Evans, A.R., and Jernvall, J. (2007). Predicting evolutionary patterns of mammalian teeth from development. Nature *449*, 427–432.

Kawakami, K., Takeda, H., Kawakami, N., Kobayashi, M., Matsuda, N., and Mishina, M. (2004). A Transposon-Mediated Gene Trap Approach Identifies Developmentally Regulated Genes in Zebrafish. Dev. Cell *7*, 133–144.

Kingsley, D.M. (1994). What do BMPs do in mammals? Clues from the mouse short-ear mutation. Trends Genet. *10*, 16–21.

Larkin, M.A., Blackshields, G., Brown, N.P., Chenna, R., McGettigan, P.A., McWilliam, H., Valentin, F., Wallace, I.M., Wilm, A., Lopez, R., et al. (2007). Clustal W and Clustal X version 2.0. Bioinformatics *23*, 2947–2948.

Laurikkala, J., Mikkola, M., Mustonen, T., Åberg, T., Koppinen, P., Pispa, J., Nieminen, P., Galceran, J., Grosschedl, R., and Thesleff, I. (2001). TNF Signaling via the Ligand–Receptor Pair Ectodysplasin and Edar Controls the Function of Epithelial Signaling Centers and Is Regulated by Wnt and Activin during Tooth Organogenesis. Dev. Biol. *229*, 443–455.

Li, T.-F., Darowish, M., Zuscik, M.J., Chen, D., Schwarz, E.M., Rosier, R.N., Drissi, H., and O'Keefe, R.J. (2006). Smad3-deficient chondrocytes have enhanced BMP signaling and accelerated differentiation. J. Bone Miner. Res. *21*, 4–16.

Liu, F., Chu, E.Y., Watt, B., Zhang, Y., Gallant, N.M., Andl, T., Yang, S.H., Lu, M.-M., Piccolo, S., Schmidt-Ullrich, R., et al. (2008). Wnt/β-catenin signaling directs multiple stages of tooth morphogenesis. Dev. Biol. *313*, 210–224.

Liu, W., Sun, X., Braut, A., Mishina, Y., Behringer, R.R., Mina, M., and Martin, J.F. (2005). Distinct functions for Bmp signaling in lip and palate fusion in mice. Development *132*, 1453–1461.

Lubbe, S.J., Pittman, A.M., Olver, B., Lloyd, A., Vijayakrishnan, J., Naranjo, S., Dobbins, S., Broderick, P., Gómez-Skarmeta, J.L., and Houlston, R.S. (2012). The 14q22.2 colorectal cancer variant rs4444235 shows cis-acting regulation of BMP4. Oncogene *31*, 3777–3784.

Marinić, M., Aktas, T., Ruf, S., and Spitz, F. (2013). An Integrated Holo-Enhancer Unit Defines Tissue and Gene Specificity of the Fgf8 Regulatory Landscape. Dev. Cell *24*, 530–542.

Massagué, J. (2012). TGFβ signalling in context. Nat. Rev. Mol. Cell Biol. *13*, 616–630.

Messeguer, X., Escudero, R., Farré, D., Núñez, O., Martínez, J., and Albà, M.M. (2002). PROMO: detection of known transcription regulatory elements using species-tailored searches. Bioinformatics *18*, 333–334.

Mou, C., Jackson, B., Schneider, P., Overbeek, P.A., and Headon, D.J. (2006). Generation of the primary hair follicle pattern. Proc. Natl. Acad. Sci. *103*, 9075–9080.

Mou, C., Pitel, F., Gourichon, D., Vignoles, F., Tzika, A., Tato, P., Yu, L., Burt, D.W., Bed'hom, B., Tixier-Boichard, M., et al. (2011). Cryptic Patterning of Avian Skin Confers a Developmental Facility for Loss of Neck Feathering. PLoS Biol *9*, e1001028.

Near, T.J., Eytan, R.I., Dornburg, A., Kuhn, K.L., Moore, J.A., Davis, M.P., Wainwright, P.C., Friedman, M., and Smith, W.L. (2012). Resolution of ray-finned fish phylogeny and timing of diversification. Proc. Natl. Acad. Sci. *109*, 13698–13703.

Neubüser, A., Peters, H., Balling, R., and Martin, G.R. (1997). Antagonistic Interactions between FGF and BMP Signaling Pathways: A Mechanism for Positioning the Sites of Tooth Formation. Cell *90*, 247–255.

Newburger, D.E., and Bulyk, M.L. (2009). UniPROBE: an online database of protein binding microarray data on protein–DNA interactions. Nucleic Acids Res. *37*, D77–D82.

Nie, X., Luukko, K., and Kettunen, P. (2006). BMP signalling in craniofacial development. Int. J. Dev. Biol. *50*.

O'Connell, D.J., Ho, J.W.K., Mammoto, T., Turbe-Doan, A., O'Connell, J.T., Haseley, P.S., Koo, S., Kamiya, N., Ingber, D.E., Park, P.J., et al. (2012). A Wnt-bmp feedback circuit controls intertissue signaling dynamics in tooth organogenesis. Sci. Signal. *5*, ra4.

Oka, S., Oka, K., Xu, X., Sasaki, T., Bringas Jr., P., and Chai, Y. (2007). Cell autonomous requirement for TGF-β signaling during odontoblast differentiation and dentin matrix formation. Mech. Dev. *124*, 409–415.

Oshima, M., Oshima, H., and Taketo, M.M. (1996). TGF-beta receptor type II deficiency results in defects of yolk sac hematopoiesis and vasculogenesis. Dev. Biol. *179*, 297–302.

Perry, M.W., Boettiger, A.N., Bothma, J.P., and Levine, M. (2010). Shadow enhancers foster robustness of Drosophila gastrulation. Curr. Biol. *20*, 1562–1567.

Pregizer, S., and Mortlock, D.P. (2009). Control of BMP gene expression by long-range regulatory elements. Cytokine Growth Factor Rev. *20*, 509–515.

Ross, M.T., LaBrie, S., McPherson, J., and Stanton, V.P. (1999). Screening large-insert libraries by hybridization. In Current Protocols in Human Genetics, N.C. Dracopoli, J.L. Haines, and B.. Korf, eds. (New York: John Wiley and Sons), pp. 5.6.1–5.6.52.

Sanford, L.P., Ormsby, I., Groot, A.C.G., Sariola, H., Friedman, R., Boivin, G.P., Cardell, E.L., and Doetschman, T. (1997). TGFbeta2 knockout mice have multiple developmental defects that are non-overlapping with other TGFbeta knockout phenotypes. Development *124*, 2659–2670.

Scheer, N., and Campos-Ortega, J.A. (1999). Use of the Gal4-UAS technique for targeted gene expression in the zebrafish. Mech. Dev. *80*, 153–158.

Shi, M., Murray, J.C., Marazita, M.L., Munger, R.G., Ruczinski, I., Hetmanski, J.B., Wu, T., Murray, T., Redett, R.J., Wilcox, A.J., et al. (2012). Genome wide study of maternal and

parent-of-origin effects on the etiology of orofacial clefts. Am. J. Med. Genet. A. *158A*, 784–794.

Shimizu, N., Kawakami, K., and Ishitani, T. (2012). Visualization and exploration of Tcf/Lef function using a highly responsive Wnt/β-catenin signaling-reporter transgenic zebrafish. Dev. Biol. *370*, 71–85.

Solloway, M.J., Dudley, A.T., Bikoff, E.K., Lyons, K.M., Hogan, B.L., and Robertson, E.J. (1998). Mice lacking Bmp6 function. Dev. Genet. *22*, 321–339.

Stern, D.L. (2000). Perspective: Evolutionary Developmental Biology and the Problem of Variation. Evolution *54*, 1079–1091.

Sun, Z., Jin, P., Tian, T., Gu, Y., Chen, Y.-G., and Meng, A. (2006). Activation and roles of ALK4/ALK7-mediated maternal TGFβ signals in zebrafish embryo. Biochem. Biophys. Res. Commun. *345*, 694–703.

Suster, M.L., Abe, G., Schouw, A., and Kawakami, K. (2011). Transposon-mediated BAC transgenesis in zebrafish. Nat. Protoc. *6*, 1998–2021.

Swarup, H. (1958). Stages in the Development of the Stickleback Gasterosteus aculeatus (L.). J. Embryol. Exp. Morphol. *6*, 373–383.

Urasaki, A., Morvan, G., and Kawakami, K. (2006). Functional Dissection of the Tol2 Transposable Element Identified the Minimal cis-Sequence and a Highly Repetitive Sequence in the Subterminal Region Essential for Transposition. Genetics *174*, 639–649.

Vainio, S., Karavanova, I., Jowett, A., and Thesleff, I. (1993). Identification of BMP-4 as a signal mediating secondary induction between epithelial and mesenchymal tissues during early tooth development. Cell *75*, 45–58.

Villefranc, J.A., Amigo, J., and Lawson, N.D. (2007). Gateway compatible vectors for analysis of gene function in the zebrafish. Dev. Dyn. *236*, 3077–3087.

Wang, Y., Cox, M.K., Coricor, G., MacDougall, M., and Serra, R. (2013). Inactivation of Tgfbr2 in Osterix-Cre expressing dental mesenchyme disrupts molar root formation. Dev. Biol. *382*, 27–37.

Westerfield, M. (2007). The Zebrafish Book: A guide for the Laboratory Use of Zebrafish (Danio rerio), 5th Edition (Eugene, OR: University of Oregon Press).

Wood, A.R., Esko, T., Yang, J., Vedantam, S., Pers, T.H., Gustafsson, S., Chu, A.Y., Estrada, K., Luan, J., Kutalik, Z., et al. (2014). Defining the role of common variation in the genomic and biological architecture of adult human height. Nat. Genet. *46*, 1173–1186.

Wu, P., Wu, X., Jiang, T.-X., Elsey, R.M., Temple, B.L., Divers, S.J., Glenn, T.C., Yuan, K., Chen, M.-H., Widelitz, R.B., et al. (2013). Specialized stem cell niche enables repetitive renewal of alligator teeth. Proc. Natl. Acad. Sci. U. S. A. *110*, E2009–E2018.

Zhao, H., Oka, K., Bringas, P., Kaartinen, V., and Chai, Y. (2008). TGF-β type I receptor Alk5 regulates tooth initiation and mandible patterning in a type II receptor-independent manner. Dev. Biol. *320*, 19–29.

Zhao, X., Zhang, Z., Song, Y., Zhang, X., Zhang, Y., Hu, Y., Fromm, S.H., and Chen, Y. (2000). Transgenically ectopic expression of Bmp4 to the Msx1 mutant dental mesenchyme restores downstream gene expression but represses Shh and Bmp2 in the enamel knot of wild type tooth germ. Mech. Dev. *99*, 29–38.

This chapter appeared in *Evolution* **70**:4, 887-902 (2016). All authors have agreed to its use in this dissertation.

5. Partially repeatable genetic basis of benthic adaptation in threespine sticklebacks

Priscilla A. Erickson[1], Andrew M. Glazer[1], Emily E. Killingbeck[1], Rachel M. Agoglia[1], Jiyeon Baek[1], Sara M. Carsanaro[1], Anthony M. Lee[1], Phillip A. Cleves[1], Dolph Schluter[2], and Craig T. Miller[1*]

[1]Department of Molecular and Cell Biology, University of California-Berkeley, California, USA

[2]Biodiversity Research Centre and Zoology Department, University of British Columbia, Vancouver, BC, Canada

Abstract

The extent to which convergent adaptation to similar ecological niches occurs by a predictable genetic basis remains a fundamental question in biology. Threespine stickleback fish have undergone an adaptive radiation in which ancestral oceanic populations repeatedly colonized and adapted to freshwater habitats. In multiple lakes in British Columbia, two different freshwater species have evolved: a deep-bodied benthic form adapted to forage near the lake substrate, and a narrow-bodied limnetic form adapted to forage in open water. Here we use genome-wide linkage mapping in marine x benthic F2 genetic crosses to test the extent of shared genomic regions underlying benthic adaptation in three benthic populations. We identify at least 100 Quantitative Trait Loci (QTL) harboring genes influencing skeletal morphology. The majority of QTL (57%) are unique to one cross. However, four genomic regions affecting eight craniofacial and armor phenotypes are found in all three benthic populations. We find that QTL are clustered in the genome and overlapping QTL regions are enriched for genomic signatures of natural selection. These findings suggest that benthic adaptation has occurred via both parallel and non-parallel genetic changes.

Introduction

Convergent evolution, the evolution of similar phenotypes in independent lineages adapting to similar ecological niches, provides compelling evidence for ecological adaptation (Schluter 2000), as well as natural replicates to study the genetic basis of evolutionary change. Convergent phenotypic evolution sometimes occurs through changes in the same genes in multiple lineages, called parallel evolution (Stern 2013), suggesting some evolutionary trajectories are constrained and partially predictable. How often convergent evolution occurs through parallel and thus predictable genetic changes remains an outstanding and important question (Stern and Orgogozo 2008, 2009; Rosenblum et al. 2014). While many studies have identified similar genetic bases for convergent phenotypes (reviewed in Conte et al. 2012; Stern 2013), few studies have simultaneously tested for genetic parallelism underlying many phenotypic traits in multiple convergently evolved populations.

One critical first step in testing for a parallel genetic basis of convergence is mapping the genomic regions involved in convergent phenotypic evolution. Although quantitative trait locus (QTL) mapping can robustly identify genomic regions underlying complex traits, the genomic intervals are typically large and contain multiple genes, when there are few rounds of recombination in the genetic cross. However, comparing the genomic intervals of QTL between crosses from different populations will either support or refute the hypothesis of genetic parallelism, since QTL can either map to overlapping regions (potentially parallel) or map to non-overlapping regions (non-parallel).

Threespine stickleback fish (*Gasterosteus aculeatus*) provide a powerful vertebrate model system to study convergent phenotypic evolution, as ancestral oceanic forms have repeatedly colonized and rapidly adapted to countless freshwater environments throughout the Northern hemisphere (Bell and Foster 1994). Many morphological phenotypes convergently evolve in freshwater, most of which are likely adaptations to the new ecological pressures in freshwater environments, such as a different diet and predation regime. In five drainages in British Columbia, two freshwater species pairs convergently evolved: a benthic species adapted to feed on the lake bottom, and a limnetic species adapted to forage in the open water (McPhail 1984, 1992; Schluter and McPhail 1992). Across lakes, the different benthic and limnetic forms strikingly resemble each other, yet evolved in isolation (Taylor and McPhail 1999). Once thought to be the result of sympatric speciation, phylogenetic analyses based on allozymes (McPhail 1984, 1992), nuclear microsatellites (Taylor and McPhail 2000), mtDNA haplotypes (Taylor and McPhail 1999), genome-wide SNP genotypes (Jones et al. 2012a), and salinity tolerance experiments (Kassen et al. 1995) instead support a double invasion scenario. In this scenario, the first oceanic colonization event evolved into a freshwater form, and then a second oceanic colonization event displaced the first population to the benthic niche while adapting to the alternative open water limnetic niche (Taylor and McPhail 2000). Thus, the distinct benthic morphs found in species pair lakes are especially divergent from marine ancestors and offer an outstanding system to study the genetic architecture of repeatedly but independently evolved phenotypes.

Differences in a number of skeletal traits have evolved repeatedly in benthic sticklebacks. Benthic fish have fewer gill rakers, a larger jaw, and reduced pelvic and dorsal spines (McPhail 1984, 1992, 1994; Schluter and McPhail 1992). Additionally, benthic fish from at least one of these species pair lakes, Paxton Lake, have evolved more pharyngeal teeth and longer branchial bones relative to marine fish (Cleves et al. 2014; Erickson et al. 2014). Collectively these craniofacial differences have important functional consequences: benthic fish are more efficient at foraging for larger prey items in a benthic substrate, and limnetic fish are more efficient at foraging for plankton from open water (Bentzen and McPhail 1984; Schluter 1993). In turn, the distinct morphologies are important for survival in their respective niches; benthic-limnetic hybrids have reduced survival in nature (Gow et al. 2007). Some of these benthic trophic traits have also been documented in nearby lakes that have a predominantly benthic environment yet only have a single ecomorph (Lavin and Mcphail 1985), suggesting that many of these traits are important for adaptation to a benthic environment, either with or without a limnetic ecomorph.

How often do the same genomic regions underlie convergent phenotypic adaptation in multiple benthic environments? Hybridization occurs between anadromous marine and freshwater sticklebacks (Hagen 1967; Jones et al. 2006, 2008), and likely maintains freshwater-adapted alleles at low frequency in oceanic populations (Barrett and Schluter 2008; Schluter and Conte 2009; Hohenlohe et al. 2012; Jones et al. 2012b; Terekhanova et al. 2014), which might increase the chances of parallel evolution. For example, an ancient allele of *Eda* conferring reduced lateral plate armor is present at low frequency in marine populations and has been reused many times in freshwater adaptation (Colosimo et al. 2004, 2005; O'Brown et al. 2015). Allele sharing by hybridization of differently adapted populations has also been found to be an important contributor to parallel evolution in other species, including *Heliconius* butterflies, Galapagos finches, mice, humans, and *Mimulus* (Song et al. 2011; The Heliconius Genome Consortium 2012; Huerta-Sánchez et al. 2014; Lamichhaney et al. 2015; Stankowski and Streisfeld 2015). However, convergent evolution by novel mutations in the same gene or genetic pathway has also been observed in sticklebacks and other systems (Colosimo et al. 2004; Protas et al. 2006; Kingsley et al. 2009; Chan et al. 2010; Rosenblum et al. 2010; Vickrey et al. 2015). Alternatively, convergent phenotypic evolution may have a mostly different (non-parallel) genetic basis (Wittkopp et al. 2003; Kowalko et al. 2013; Ellis et al. 2015; Glazer et al. 2015). The overall extent to which parallel vs. non-parallel genetic changes and new mutations vs. standing variation are involved in stickleback evolution remains largely unknown (Conte et al. 2015).

Here we use genome-wide linkage mapping to map QTL influencing a variety of trophic and armor traits in three marine x benthic F2 genetic crosses, each using a benthic grandmother from independently derived benthic populations in Paxton, Priest, and Enos Lakes. The shared marine grandfather individual came from an anadromous marine population geographically near the benthic populations (Little Campbell River, BC), and thus serves as an extant proxy for the marine population that likely colonized the lakes. This crossing scheme allows identification of genomic regions responsible for phenotypic differentiation of each benthic population compared to a shared marine

genetic background. In the three crosses, QTL that map to overlapping genomic regions and influence similar phenotypes in multiple crosses are candidates for parallel genetic evolution. Alternatively, and assuming no major chromosomal rearrangements, QTL mapping to unique genomic regions indicate non-parallel evolution.

Methods

Animal Statement and Crosses

A single wild Little Campbell marine (LCM, Little Campbell River, British Columbia) male was crossed to wild benthic freshwater females from Paxton Lake (PAXB, Texada Island, BC), Priest Lake (PRIB, Texada Island, BC), and Enos Lake (ENOB, Vancouver Island, BC) in 2002. See Figure 5.1A for representative fish from each population and maps of population locations (Figure 5.1B and C). Since the Enos species pair collapsed between 1994 and 2002 (Kraak et al. 2001; Taylor et al. 2006), a female from Enos lake that morphologically resembled the benthic ecomorph was used. Fish were raised in 100 l aquaria in 5 ppt saltwater. F1 fish were intercrossed to produce a total of five F2 families (see Table 5.1). One family of 180 F2s was studied for the ENOB cross; 186 F2s (two families of 94 and 92) for the PAXB cross, and 180 F2s (two families of 90) for the PRIB cross. F2s were raised to six months, euthanized, and stored in ethanol. All animal work was approved by the UBC Animal Care committee under protocol A00-191.

Cross	F2 family	# F2s	mean SL (mm)	st. dev. SL (mm)
ENOB	G1-5	180	40.8	3.2
PAXB	G3-2	94	43.5	3.8
PAXB	G3-4	92	42.2	3.6
PRIB	G2-13	90	40.8	2.7
PRIB	G2-18	90	43.3	3

Table 5.1: Description of F2 families included in analysis.

Phenotypes were corrected for size and standard length within families, and then families were combined within each cross for QTL analysis. Ten fish with either more than 50% failed genotypes (8) or duplicate genotypes (2) were excluded from the phenotypic analysis.

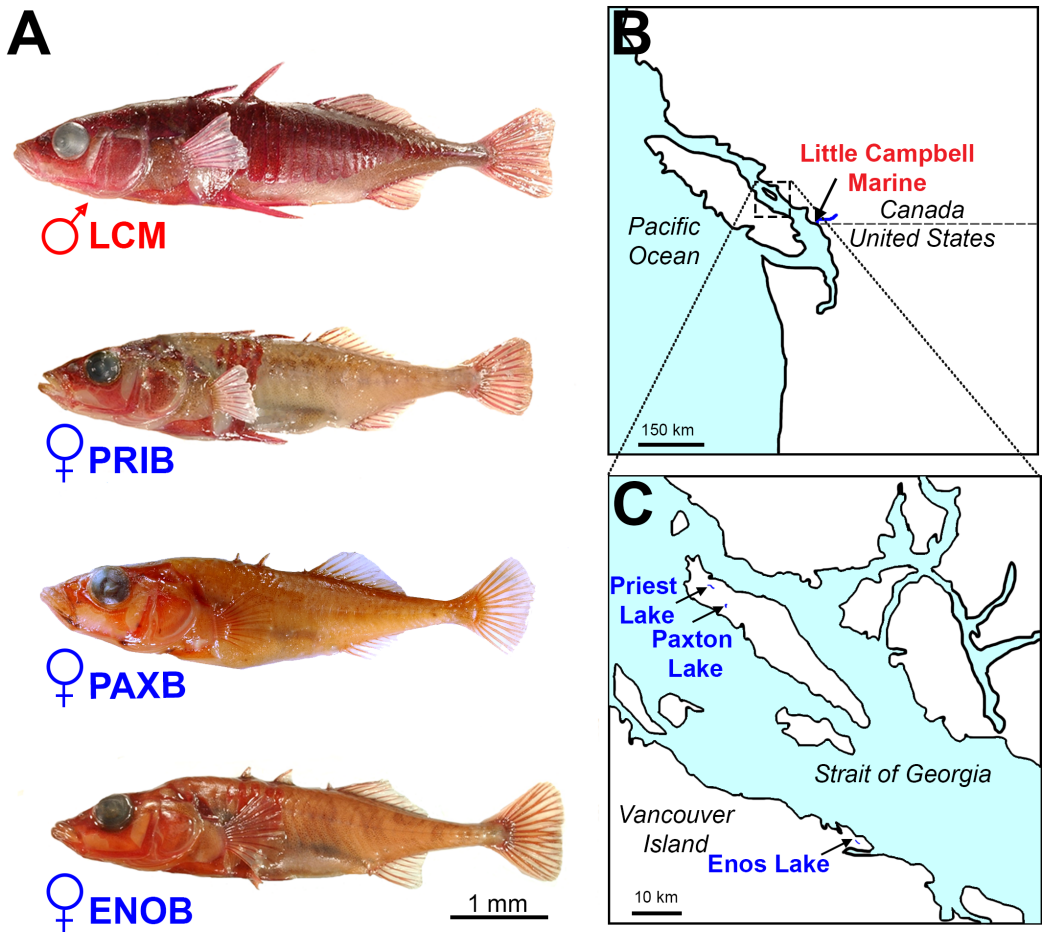


Figure 5.1: Benthic populations studied.

(A) Representative wild-caught adult fish from each population, with bone stained with Alizarin red. LCM = Little Campbell Marine, PAXB = Paxton Benthic, PRIB = Priest Benthic, ENOB = Enos Benthic. (B, C) Location of the marine population source and three species-pair lakes in the Pacific Northwest. The Little Campbell River, where the anadromous LCM population breeds, is indicated as a blue line in (B). Red = marine population (LCM), Blue = Benthic populations.

Phenotyping

Standard length was measured with digital calipers twice on all fish and averaged. All fish were fixed in 10% neutral buffered formalin overnight, stained with 0.008% Alizarin red in 1% potassium hydroxide and cleared in 50% glycerol, 0.5% potassium hydroxide. Most traits phenotyped were previously shown to have a strong genetic basis in a PAXB x marine F2 cross (Miller et al. 2014). Body depth, dorsal spines (DS1-3), left and right pelvic spines (PSL and PSR), and premaxilla height (PMH) were measured using digital calipers, and lateral plates were counted on the left side. Additional external craniofacial phenotypes [frontal width (FW), jaw width (JW), premaxilla length (PML), articular

length (AL), dentary length (DL) opercle length (OPL), and opercle width (OPW)] were measured using a Leica M125 microscope with reticule as described in Miller et al. 2014. Branchial skeletons were dissected and flat-mounted to measure branchial bone length [the basihyal (BH), all five ceratobranchials (CB1-5), and the anterior-most epibranchial (EB1)] and to count ventral and dorsal pharyngeal teeth (VTP, DTP1, and DTP2), ventral gill rakers in rows 1-9 along the ceratobranchials (R1C-R9C), and the total number of row 1-9 ventral rakers (C). See Miller et al. 2014 for more details on branchial skeletal phenotypes. For DTP1 and VTP, the total number of teeth on the right and left tooth plates was counted; for DTP2, the left only was counted. For rakers and branchial bones, the left side was counted or measured. Bone lengths were measured using NIH Image J as described in Erickson et al. 2014.

Phenotypes were tested for significant association with sex and standard length using ANOVA (sex), linear regression (standard length), or general linear model (both sex and standard length) in R (version 3.1.2, http://www.r-project.org/) and corrected for any variable significant at $P \leq 0.05$. Outliers more than 3 standard deviations from the mean were removed and the residuals were recalculated. Each phenotype was tested for increasing variances by a Levene test and deviation from normality by an Anderson-Darling test, and log-transformed if doing so corrected either of these deviations (see Miller et al. 2014 for details). Corrections were performed on phenotypes within each F2 family, and all residuals were back transformed to a 40 mm standard length fish for QTL mapping. See Table 5.2 for corrections applied to each phenotype in each F2 family. Phenotype correlation plots (Figure 5.2) were generated with the *Ellipse* package in R (version 0.3-8, http://CRAN.R-project.org/package=ellipse) using code adapted from https://hlplab.wordpress.com/.

Phenotype	Description	Category	ENOB	PAXB G3-2	PAXB G3-4	PRIB G2-13	PRIB G2-18	penalty
SL	standard length (mm)	SL	NA	NA	NA	NA	NA	yes
depth	body depth (mm)	depth	Size, Log	Size, Sex	Size	Size, Sex	Size	yes
FW	frontal width (mm)	skull	Size, Sex	Size, Sex	Size, Sex	Size, Sex	Size	yes
JW	jaw width (mm)	jaw	Size	Size	Size, Sex	Size, Sex	Size, Sex	yes
PML	premaxillary length (mm)	jaw	Size, Sex, Log	Size, Sex	Size, Sex	Size, Sex	Size, Sex	yes
DL	dentary length (mm)	jaw	Size, Sex	Size, Sex	Size, Sex	Size, Sex	Size, Sex	yes
AL	articular length (mm)	jaw	Size, Sex	Size, Sex	Size, Sex	Size, Sex	Size, Sex	yes
PMH	premaxillary height (mm)	jaw	Size	Size, Sex	Size, Sex	Size, Sex	Size, Sex	yes
OPL	opercle length (mm)	opercle	Size, Sex	Size, Sex	Size, Sex	Size, Sex	Size, Sex	yes
OPW	opercle width (mm)	opercle	Size, Sex, Log	Size, Sex	Size, Sex	Size, Sex	Size, Sex	yes
Plates	left lateral plate count	plates	NA	NA	NA	NA	NA	yes
VTP	total ventral pharyngeal tooth count	teeth	Sex	NA	NA	Size	Size, Sex	yes
DTP1	total dorsal pharyngeal tooth plate 1 count	teeth	Sex	NA	Sex	Size, Sex	Size, Sex	yes
DTP2	left dorsal pharyngeal tooth plate 2 count	teeth	Sex	NA	NA	Size	Size, Sex	yes
BH	basihyal length (mm)	branchial	Size, Sex	Size, Sex	Size, Sex	Size, Sex	Size, Sex	yes
CB1	ceratobranchial 1 length (mm)	branchial	Size, Sex	Size, Sex	Size	Size, Sex	Size, Sex	no
CB2	ceratobranchial 2 length (mm)	branchial	Size, Sex	Size, Sex	Size	Size, Sex	Size, Sex	no
CB3	ceratobranchial 3 length (mm)	branchial	Size, Sex	Size, Sex	Size	Size, Sex	Size, Sex	no
CB4	ceratobranchial 4 length (mm)	branchial	Size, Sex	Size, Sex	Size, Sex	Size, Sex	Size, Sex	no
CB5	ceratobranchial 5 length (mm)	branchial	Size, Sex	Size	Size	Size, Sex	Size, Sex	yes
EB	epibranchial 1 length	branchial	Size, Sex	Size, Sex	Size, Sex	Size, Sex	Size, Sex	yes
R1C	left row 1 cerato gill raker count	rakers	NA	Size	NA	NA	NA	no
R2C	left row 2 cerato gill raker count	rakers	NA	Sex	NA	NA	Sex	no
R3C	left row 3 cerato gill raker count	rakers	NA	NA	NA	NA	Sex	no
R4C	left row 4 cerato gill raker count	rakers	Sex	Sex	Sex	Sex	Sex	no
R5C	left row 5 cerato gill raker count	rakers	NA	Sex	Sex	NA	Sex	no
R6C	left row 6 cerato gill raker count	rakers	NA	Size	Sex	NA	NA	no

Phenotype	Description	Category	ENOB	PAXB G3-2	PAXB G3-4	PRIB G2-13	PRIB G2-18	penalty
R8C	left row 8 cerato gill raker count	rakers	NA	Size	NA	NA	Sex	no
R9C	left row 9 cerato gill raker count	rakers	NA	NA	NA	NA	NA	no
C	total left row 1 through 9 cerato gill raker count	rakers	NA	Size, Sex	Sex	NA	Sex	yes
DS1	dorsal spine 1 length (mm)	median fin	Size, Sex	Size	Size	Size	Size	yes
DS2	dorsal spine 2 length (mm)	median fin	Size, Sex	Size, Sex	Size, Sex	Size, Sex	Size	yes
DS3	dorsal spine 3 length (mm)	median fin	Size, Sex	Size, Sex	Size, Sex	Size, Sex	Size, Sex, Log	yes
PSL	left pelvic spine length (mm)	pelvic spine	Size, Sex	NA	Size, Sex	Size	Size	yes
PSR	right pelvic spine length (mm)	pelvic spine	Size, Sex	NA	Size, Log	Size	Size, Sex	no

Table 5.2. Statistical corrections applied to each phenotype in each F2 family.

Phenotypes were tested for a significant association with sex and standard length (size) in a linear model. All size-corrected residuals were back-transformed to a 40 mm fish. NA indicates that no statistical corrections were applied and raw phenotypic values were used. "Penalty" indicates whether the phenotype was used to calculate *scantwo* penalties via 100 permutations of the phenotype data.

DNA Isolation and Genotyping-By-Sequencing (GBS)

DNA was extracted from pectoral fins using the Qiagen DNAEasy 96 Blood & Tissue kit according to manufacturer's instructions. Concentration was quantified using the Quant-iT PicoGreen kit and 50 ng genomic DNA was used for each sample. Thirteen F2 fish were excluded from GBS due to low DNA concentration. Genomic DNA was digested with ApeKI and barcoded libraries were prepared as described in Glazer et al. 2015 using a set of 96 unique barcodes and four unique adapters. A total of 384 barcoded samples were sequenced in each of two Illumina HiSeq2000 lanes using paired-end 100 bp reads. The grandparents of the crosses were also genotyped with GBS. For the grandparental DNA samples, multiple barcodes were used to generate higher coverage for SNP calling (11 barcodes for the LCM grandfather, 9 barcodes for the PRIB grandmother, and 6 barcodes each for the PAXB and ENOB grandmothers). All GBS reads are deposited in the Sequence Read Archive (accession number SRP 070856.).

Reads were mapped to the stickleback reference genome with Samtools, and SNPs that had opposite homozygous states in the grandparents were called as described in Glazer et al. 2015. Genotypes were assigned using a custom pipeline to combine individual SNP genotypes into bins with a maximum length of 500 kb. Markers that failed in over 25% of fish (n_{PAXB} =98, n_{PRIB} = 105, n_{ENOB} = 83) or were outliers in the expected 1:2:1 distribution by a chi-square test were dropped (n_{PAXB} = 68, n_{PRIB} = 58, n_{ENOB} = 41). F2 fish in which over 50% of markers failed (n_{PAXB} = 1, n_{PRIB} = 3, n_{ENOB} = 4,) were also excluded. Sex was assigned based on coverage of the X chromosome relative to the rest of the genome as in Glazer et al. 2015. See Table 5.3 for a summary of GBS data processing.

Constructing Linkage Maps

A set of 823 binned markers that were informative in all three crosses was initially used to construct linkage maps in Joinmap 4.0 (Kyazma) using the cross-pollination setting. Plots of genetic distance versus physical distance (G x P) were generated using the revised genome assembly described in Glazer et al. 2015, and 18 markers that appeared to be strong outliers in at least one G x P plot were removed from each cross, resulting in a final set of 805 shared markers used for mapping in all three crosses.

Mapping QTL in R/qtl

Initial QTL mapping was performed in R/qtl (version 1.33-7, http://www.rqtl.org/, Broman et al. 2003; Broman and Sen 2009) using the *stepwiseqtl* mapping function. Penalties were calculated for a set of 22 representative phenotypes (see Table 5.2) in each cross using the *scantwo* function with 100 permutations of each phenotype. The penalties of the 22 phenotypes were then averaged across all three crosses to determine a final penalty of 3.7, which was used as a genome-wide significance threshold in a *stepwiseqtl* scan. Following this initial mapping, *addqtl*, *fitqtl* and *refineqtl* were used to search for additional QTL and to calculate the LOD scores and percent variance explained. In cases where two QTL for the same phenotype had overlapping 1.5-LOD intervals, the lower

LOD score QTL was dropped (n = 5). QTL with a LOD score less than 3.7 after the *refineqtl* calculation were dropped (n = 8). The *addqtl*/*refineqtl*/*fitqtl* process was repeated to search for additional QTL. No additional QTL were identified, so the QTL and LOD scores from the second round of analysis were used in the final data set. LOD scores at every marker for every phenotype were calculated with *refineqtl* for chromosomes with a QTL and *scanone* for chromosomes without any QTL.

For analysis of phenotypic variation at specific markers, the *argmax.geno* function of R/qtl was used to calculate the most likely genotype for fish with missing genotypes at that marker. Dominance and additivity of each QTL were calculated based on the phenotypic means for each genotypic class at the peak marker (Falconer and Mackay 1996). In the PAXB cross only, a subset of F2s lacking pelvic spines were excluded from the analysis of pelvic spine length. In this cross, pelvic spine presence/absence was mapped separately as a binary trait with the *scanone* function.

Analysis of potential pleiotropy

Given that overlapping QTL affecting different trait classes in different crosses could be the result of the same underlying parallel genetic change affecting multiple phenotypes, we performed a second overlap analysis, not classifying QTL into any trait classes. Starting with the list of all filtered, genome-wide QTL, we chose the largest effect (highest PVE) QTL for each chromosome in each cross. Then, in order of decreasing PVE, we added any additional non-overlapping QTL. This process generated a single list of QTL that represented all skeletal QTL regions in the cross, with any given genomic position represented at most once per cross. We then performed an overlap analysis and simulations as described above, but did not consider the trait category when counting overlaps.

Identifying Suggestive Parallel QTL

The minimum QTL effect size that can be detected is larger when cross size is smaller (Beavis 1998), so detecting small effect QTL shared in multiple relatively small crosses is even less likely. To minimize the resulting bias against detecting parallelism, especially given the limitations of the cross sizes of ~180 F2s per each of three crosses studied here, we considered QTL identified in the genome-wide search to be "candidate" QTL for each trait class (as in Conte et al. 2015). If no QTL for the same trait class were found on the same chromosome on which the candidate QTL was detected in a second cross, we then looked for suggestive parallel QTL that overlapped the candidate QTL as follows. We tested for QTL having LOD scores above 2.0 on the same chromosome for all phenotypes in the same trait class in the second cross. We then added these QTL to the *refineqtl* model to recalculate the LOD scores and 1.5 LOD intervals for the QTL model. If the peak marker of the suggestive QTL was found within the 1.5 LOD interval of any candidate QTL in the same trait class from either of the other two crosses, we considered the QTL to be a suggestive parallel QTL and included it in the supplemental analysis of suggestive QTL. QTL that did not meet this criterion were removed. All analyses involving

suggestive parallel QTL used the QTL locations and percentage of phenotypic variance explained (PVEs) estimated by the *refineqtl* model that included suggestive parallel QTL.

QTL filtering

To minimize overcounting QTL for multiple similar phenotypes within a trait class, we generated a list of filtered QTL as follows. For each of the 7 trait classes with multiple phenotypes (teeth, gill rakers, branchial bones, median fin, pelvic spines, jaw, and opercle), we identified the largest effect QTL (highest PVE) on each chromosome for that trait category in each cross. Since many of the phenotypes measured are serially repeated traits (such as gill raker count or ceratobranchial length), this filtering method avoids overrepresentation of QTL influencing multiple anatomically similar traits, following a previous analysis of many stickleback skeletal QTL (Miller et al. 2014). The filtering approach was applied to both genome-wide and suggestive parallel QTL. Hereafter, unless otherwise specified, QTL refers to these filtered QTL (Tables S4 and S5).

QTL clustering

We tested for clustering of QTL as described (Arnegard et al. 2014). Given the total number of QTL detected in each cross, we calculated the expected number of QTL per chromosome based on genetic length, physical length and number of Ensembl gene predictions per chromosome. Physical length and gene number were based on the revised assembly of Glazer et al. 2015 and genetic length was based on the linkage map for each cross. We then used the R function *chisq.test* to test whether the observed number of QTL per chromosome differed significantly from the expected number. *P*-values were calculated based on 10,000 permutations of the data. In crosses in which the distribution differed significantly from the expected number ($P < 0.05$), we used the standard residuals from the chi-square test to determine which chromosomes were enriched for QTL, with a standard residual >2 considered enriched.

Overlap analysis

We determined overlap by asking whether the physical 1.5 LOD intervals of QTL for the same trait category overlapped between crosses. Unique QTL were identified if they did not overlap with a QTL found in any other cross. Double overlaps were identified if they overlapped in only two of the three crosses. Triple overlaps were identified if the 1.5 LOD intervals overlapped a region of the genome in all three crosses. We calculated the mean PVE of each double and triple overlap by averaging the individual PVEs of the overlapping QTL. To test whether the QTL datasets were enriched for overlapping QTL, we randomly permuted the physical locations of the QTL in the genome 10,000 times. In each permutation, no two QTL in the same trait category in the same cross could have overlapping 1.5 LOD intervals. We then calculated the number of unique QTL, double overlaps, and triple overlaps for each random permutation. We compared the actual number of overlaps between crosses to the distribution of simulated number of overlaps and calculated a *P*-value for the extent of overlap based on the percentage of permutations in which the number of overlaps was equal to or greater than the observed number.

Analysis of Signals of Selection

Two previous studies identified regions of the stickleback genome that show strong signatures of natural selection in 21 populations (Jones et al. 2012b) or in benthic populations relative to their limnetic counterparts (Jones et al. 2012a). As in the overlap analysis, we tested for an enrichment of these signals of selection in the total set of QTL found here by randomly permuting the locations of the 1.5-LOD intervals of the QTL 10,000 times for each cross. We then compared the number of signals of selection overlapping QTL in the permuted dataset to the number of actual overlaps with the signals of selection. We calculated fold enrichment for each cross based on the ratio of the actual number of overlaps relative to the permuted mean. The *P*-value was calculated as the percent of all permutations in which the number of overlaps was greater than or equal to the observed number of overlaps. For overlapping double or triple QTL, the maximum physical range spanned by the 1.5 LOD intervals of all two or three QTL was used in calculations.

For the genomic regions displaying marine-freshwater signals of selection in 21 stickleback genome sequences (Jones et al. 2012b), we used the union of the HMM and CSS signals of selection (a total of 240 regions) and calculated their locations in the revised genome assembly from Glazer et al. 2015. Each time the 1.5 LOD interval of a QTL overlapped with a signal of selection was counted as an overlap. For the benthic-limnetic signals of selection, we used all 46 F_{ST}-outlier SNPs identified in any one of the three species-pair lakes (Paxton Lake, Priest Lake, or Quarry Lake; Jones et al. 2012a) and converted the SNP location to the revised genome assembly (Glazer et al. 2015). We then counted every overlap between a QTL and an F_{ST}-outlier SNP and compared this number to the simulated number of overlaps.

Results

Overlapping regions of the genome affect armor and craniofacial traits in multiple benthic populations

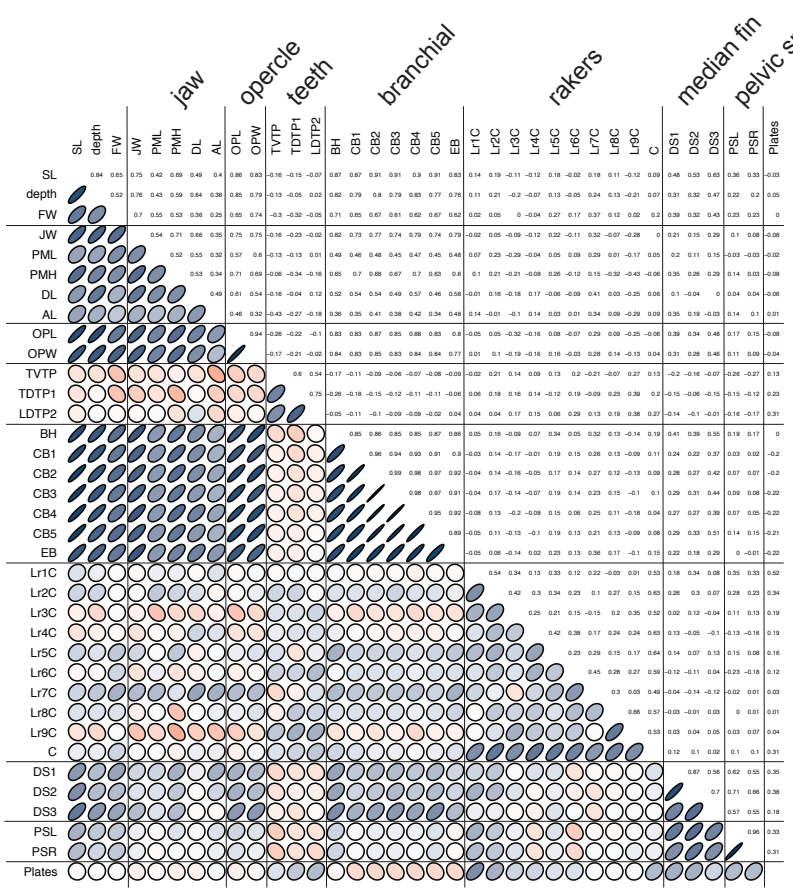
To test whether similar genetic architectures underlie skeletal adaptation in multiple populations of benthic sticklebacks, we phenotyped 36 skeletal traits in ten different trait categories in three marine x benthic crosses (see Table 5.2) and found a strong correlation between phenotypes within each trait category within each cross (see Figure 5.2). We used genotyping-by-sequencing (GBS, Elshire et al. 2011; Glazer et al. 2015) to generate genome-wide genotypes with at least a 50% success rate in 546/554 (98.5%) of F2s sequenced (see Table 5.3). Using random permutations of 22 phenotypes, we calculated a genome-wide significance threshold of LOD 3.7 for QTL mapping. All traits except standard length (SL), basihyal length (BH) and premaxilla length (PML) mapped significantly to at least one chromosome in at least one cross with this cutoff. Our linkage maps were all collinear with the revised genome assembly (Glazer et al. 2015), suggesting no major genome rearrangements occurred in the parents of the crosses.

	PAXB	PRIB	ENOB	LCM
Number of barcodes used for grandparents	6	9	6	11
Grandparent mapped reads	12,488,892	15,963,364	20,486,268	21,305,328
SNPs called in grandparents	133,228	140,604	157,821	-
Homozygous different SNPs in grandparents	81,396	81,618	93,722	-
F2s sequenced	188	183	184	-
F2 mapped reads	133,029,878	130,366,646	124,438,779	-
Average reads per F2	707,606	716,300	676,298	-
High quality SNPs in F2s	60,715	59,358	64,061	-
Initial number of sex chromosome bins	36	35	35	-
Initial number of autosomal bins	1,063	1,060	1,049	-
High quality binned markers (including sex chromosomes)	933	932	960	-
Final shared markers in analysis	805	805	805	-
Number F2s dropped (failed or duplicate genotypes)	2	3	4	-

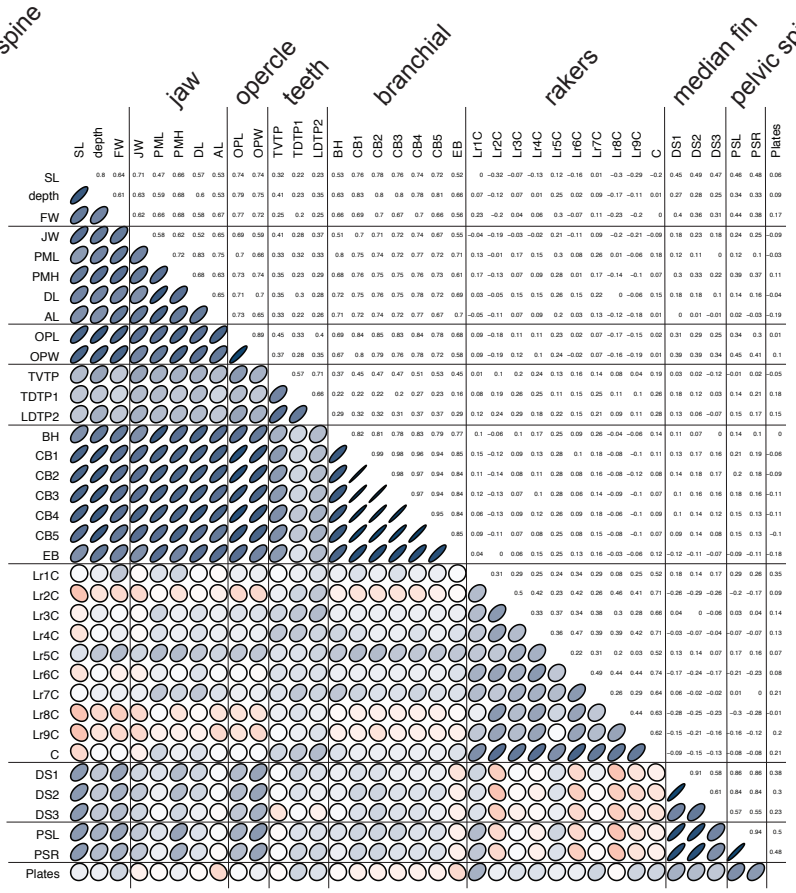
Table 5.3. Data for processing of Illumina reads into genotypes for 3 crosses.

Genomic DNA of grandparents and F2s was digested with ApeKI and barcoded with unique barcodes for genotyping-by-sequencing (GBS). Libraries were sequenced in two Illumina HiSeq2000 lanes with 384 barcodes per lane. Reads were mapped to the stickleback genome and processed using a custom pipeline (Glazer et al., 2015) to create binned markers with a maximum bin length of 500 kb. A shared set of markers informative in all three crosses was used for QTL mapping. LCM = Little Campbell Marine grandfather used to call SNPs in all three crosses.

PAXB phenotype correlations



PRIB phenotype correlations



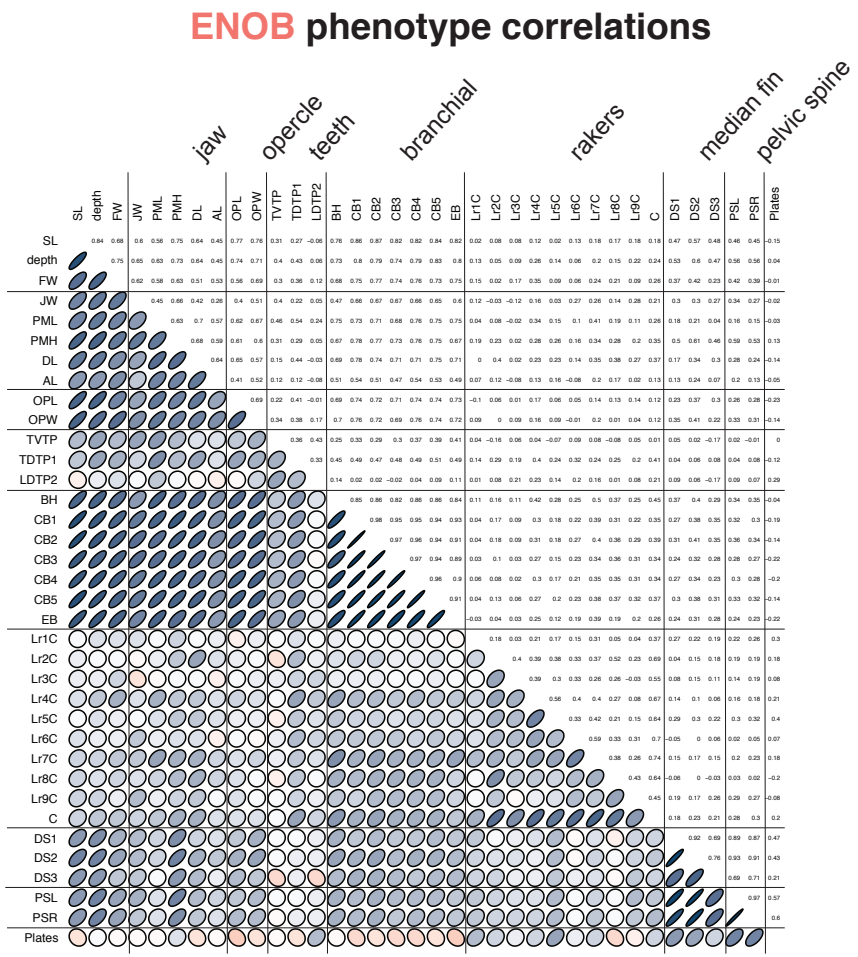


Figure 5.2: Correlation matrix of all raw phenotypes measured.

(starts on previous page) Below the diagonal, the color, shape, and orientation of the ellipse indicate the strength of the correlation (darker blue and upward sloping indicates a more positive correlation, darker red and downward sloping represents a more negative correlation; narrower ellipses indicate stronger correlations, more circular ellipses indicate weak correlations). The correlation coefficient is presented above the diagonal. Horizontal and vertical lines separate trait categories, which are labeled above

We identified a total of 157 QTL (46 PAXB, 64 PRIB, and 47 ENOB) significantly affecting skeletal traits in the three crosses (see Figure 5.4A). We then filtered the QTL from each cross such that only the QTL with the highest percentage of phenotypic variance explained (PVE) in a trait class was kept for each chromosome in order to minimize redundant oversampling of QTL (following Miller et al. 2014). This filtering resulted in a total of 100 QTL (33 PAXB, 40 PRIB, and 27 ENOB). Overall, the effect sizes of filtered and unfiltered QTL were quite similar: most QTL were small effect (PVE < 20), with a few QTL of large effect (see Figure 5.3 for the distribution of PVE of all QTL and all filtered QTL in each cross). QTL overlapped if the physical ranges of the 1.5 LOD intervals overlapped in two or three crosses. We observed that 43% of all filtered QTL overlapped with a QTL influencing a trait in the same category in at least one other cross.

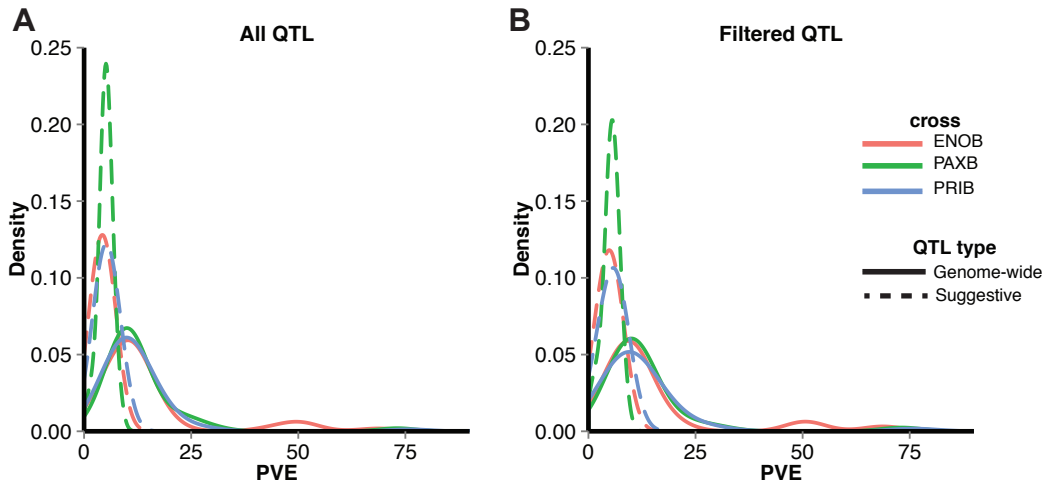


Figure 5.3: Distribution of QTL effect sizes.

Density curves for percentage of phenotypic variance explained (PVE) in each cross. Genome-wide QTL are indicated with solid lines and suggestive parallel QTL are indicated with dashed lines. (A) All QTL for all phenotypes. (B) Filtered QTL. Filtering QTL did not dramatically change the PVE distribution. Red = ENOB, green = PAXB, blue = PRIB.

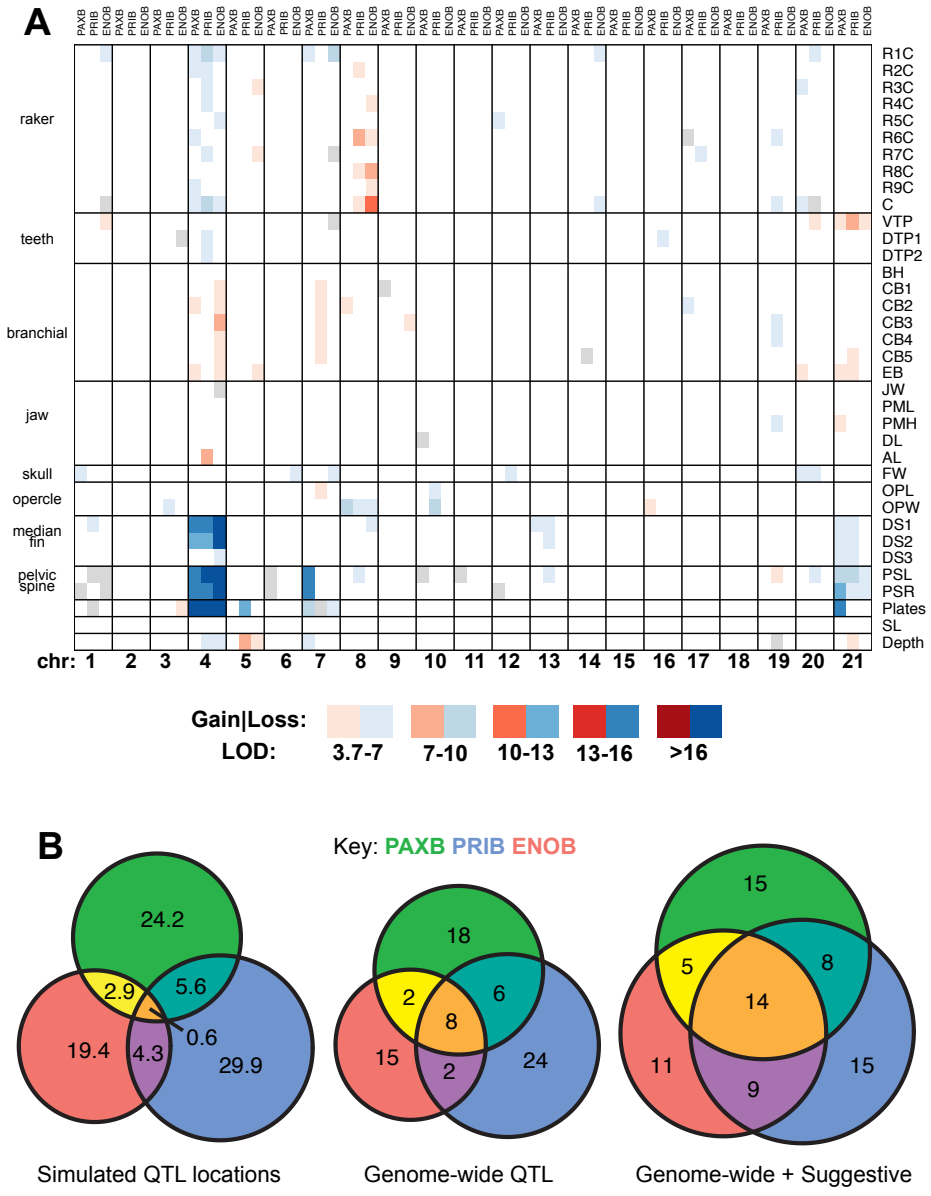


Figure 5.4: Overview of QTL.

(A) Summary of QTL detected at a LOD 3.7 genome-wide significance threshold. Chromosomes 1-21 are separated by vertical lines and numbered below. Within each chromosome, QTL for PAXB, PRIB, and ENOB (labeled on top) are indicated from left to right. Trait categories are labeled at left and separated by horizontal lines; descriptions of phenotype abbreviations at right can be found in Table 52. Color intensity indicates magnitude of LOD score (see key). Red colors indicate skeletal gain QTL (freshwater allele confers more bone); blue colors indicate skeletal loss QTL (freshwater allele confers less bone). QTL in which the phenotypes of homozygous genotypes do not differ by a Student's T-test ($P > 0.05$) are shaded in gray. (B) Venn diagrams of simulated QTL overlap (left), all genome-wide QTL overlap (middle), and all QTL including suggestive parallel QTL (right).

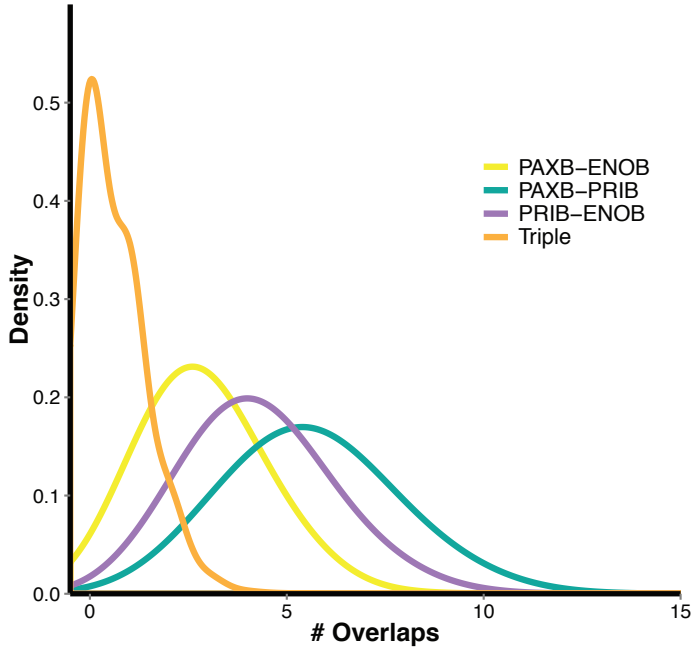


Figure 5.5: Results of QTL overlap simulations.

Density plots of the number of simulated overlaps for all QTL at a genome wide significance threshold (LOD = 3.7) are shown based on 10,000 permutations without replacement for each trait category. These distributions were used to calculate the statistical significance of the detected overlapping QTL.

We found six overlapping QTL between PAXB and PRIB, two between PAXB and ENOB, and two between PRIB and ENOB (Figure 5.4B). Ninety percent (9/10) of all QTL overlapping in two crosses had effects in the same direction. Eight QTL overlapped in all three crosses (Figure 5.4B, Figure 5.6, Figure 5.7). Five of these QTL affected armor traits and were found on chromosome 4 (dorsal spine length, pelvic spine length and lateral plate number, Figure 5.6A-C), chromosome 21 (pelvic spine, Figure 5.6D), and chromosome 7 (lateral plates, Figure 5.6E). Notably, the chromosome 4 QTL affecting dorsal and pelvic spines (Figure 5.6A and 5.6B) had large effects in all three crosses and mapped to a similar region of chromosome 4 in all three crosses. However, the genetic basis of pelvic and dorsal spine length are markedly different: several QTL were identified that affect pelvic but not dorsal spine length (Figure 5.4). For all triple-overlapping armor QTL, the benthic allele conferred a reduction in the number or size of skeletal element measured, except for the chromosome 7 lateral plate QTL in PRIB, for which heterozygotes had the fewest plates.

The genetic and developmental bases of several QTL influencing freshwater trophic adaptation in sticklebacks have been studied extensively (Cleves et al. 2014; Erickson et al. 2014; Glazer et al. 2014; Ellis et al. 2015). We were particularly interested in whether previously identified QTL affecting the branchial skeleton (the major food-processing apparatus in fish) were found in multiple benthic populations, which could suggest that these QTL are under selection in benthic environments. A large-effect QTL increasing ventral pharyngeal tooth gain on chromosome 21 (Miller et al. 2014; Cleves et

al. 2014) was indeed found in all three crosses (Figure 5.7), with the benthic allele conferring more teeth. However, the PRIB cross appeared to have two QTL peaks on chromosome 21 (Figure 5.7A). We found that a gill raker reduction QTL on chromosome 4, previously described to overlap in three independently derived freshwater populations (including PAXB, Glazer et al. 2014), was also found in all three benthic populations, as judged by overlapping 1.5 LOD intervals (Figure 5.7B). However, a second previously identified QTL on chromosome 20 was found only in the PAXB and PRIB crosses. Previously described QTL for increased branchial bone length (Erickson et al. 2014) on chromosomes 4 and 21 were double-overlapping QTL but were found in all three crosses when suggestive parallel QTL were included (Figure 5.7C-D). However, the suggestive ENOB chromosome 21 QTL appeared to map to a different region of the chromosome (although was counted as an overlap because the peak marker was found within the 1.5 LOD interval of the PRIB CB5 QTL, which was in the same branchial bone trait category). We also found a previously unreported QTL on chromosome 8 that influenced opercle width in all three crosses (Figure 5.7E).

Benthic QTL for similar traits overlap more than expected by random chance

We hypothesized that abundant standing genetic variation (Colosimo et al. 2005; Conte et al. 2012; Bell and Aguirre 2013) and similar selective pressures in the benthic freshwater environments would lead to more genetic parallelism in QTL affecting skeletal traits than expected by chance. To test for significant parallelism, we randomly permuted the locations of the QTL in the genome without replacement for each trait category and calculated the number of overlapping QTL in each permutation (Figure 5.4B, see Figure 5.5 for the distributions of simulated overlapping QTL). We found that the observed number of triple-overlapping QTL significantly exceeded the number of triple QTL expected by chance (eight observed triple overlaps versus a maximum of five in 10,000 permutations of the data, Table 5.4). However, double QTL were not significantly enriched in any population pair. All three crosses had significantly fewer unique QTL than expected by chance at a $P < 0.05$ significance level (98-99% of permutations had equal or more unique QTL, respectively, Table 5.4). Combined, these results suggest that strong selection on some skeletal traits may drive genetic parallelism for the QTL that were found to overlap in all three lakes, which then may result in a concomitant dearth of double and unique QTL. Despite this finding, the majority of detected QTL (57/100 QTL) were unique to a single cross, suggesting that benthic adaptation also has a large non-parallel component.

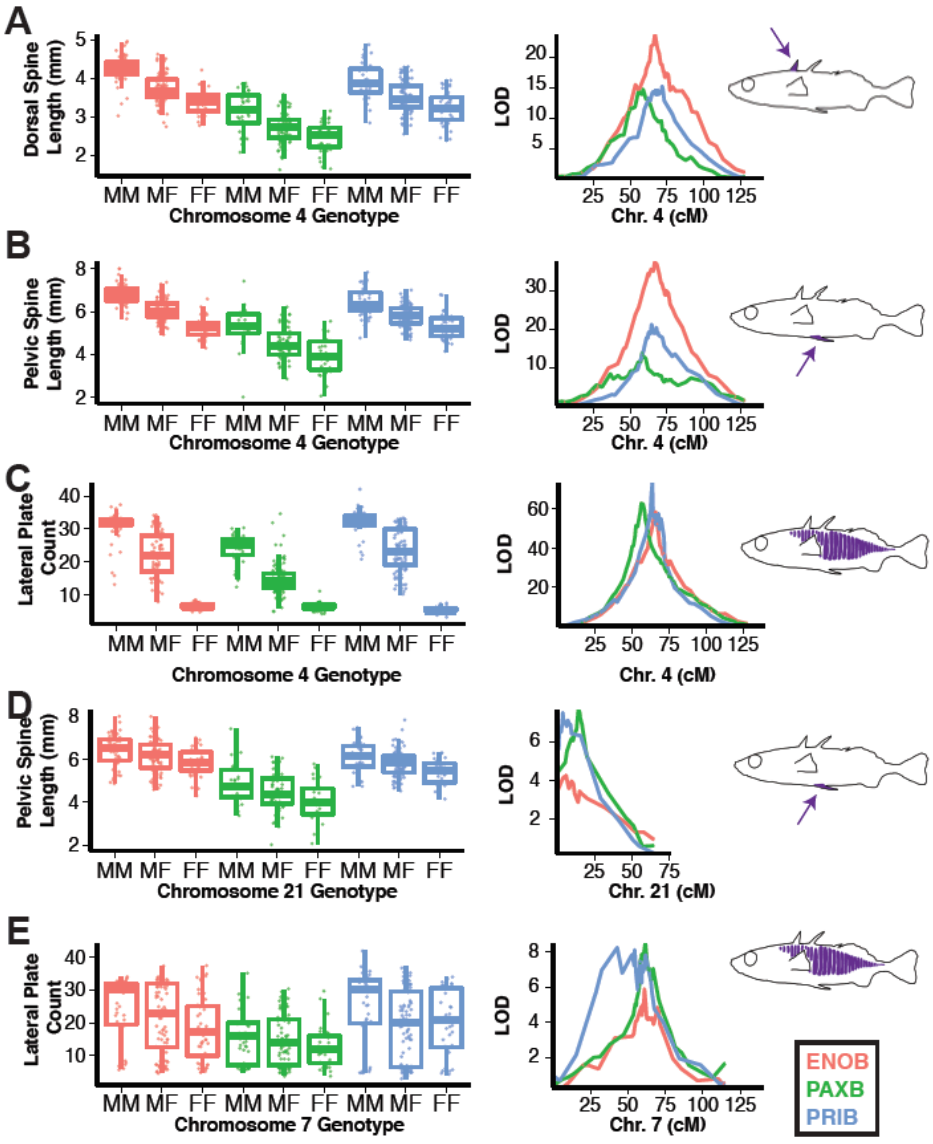


Figure 5.6: Armor QTL identified in all three crosses at the genome-wide (LOD 3.7) significance level.

Phenotypes for each trait are shown for fish with MM, MF, or FF genotypes at the peak marker in each cross on the left, where M = marine allele and F = freshwater allele. For all five QTL, the freshwater allele produces smaller or fewer skeletal elements. The LOD profiles of each QTL plotted relative to genetic distance (cM) are shown in the middle. Since genetic distance varies between crosses, the position of each marker was scaled relative to the total mean genetic length of the chromosome in all three crosses. Cartoon illustrations of armor phenotypes measured or counted are on the right, with skeletal elements highlighted in purple. (A) Dorsal spine 1 length, chr. 4; (B) Left pelvic spine length, chr. 4; (C) Lateral plate count, chr. 4; (D) Left pelvic spine length, chr. 21; (E) Lateral plate count, chr. 7. PAXB = green, PRIB = blue, ENOB = red.

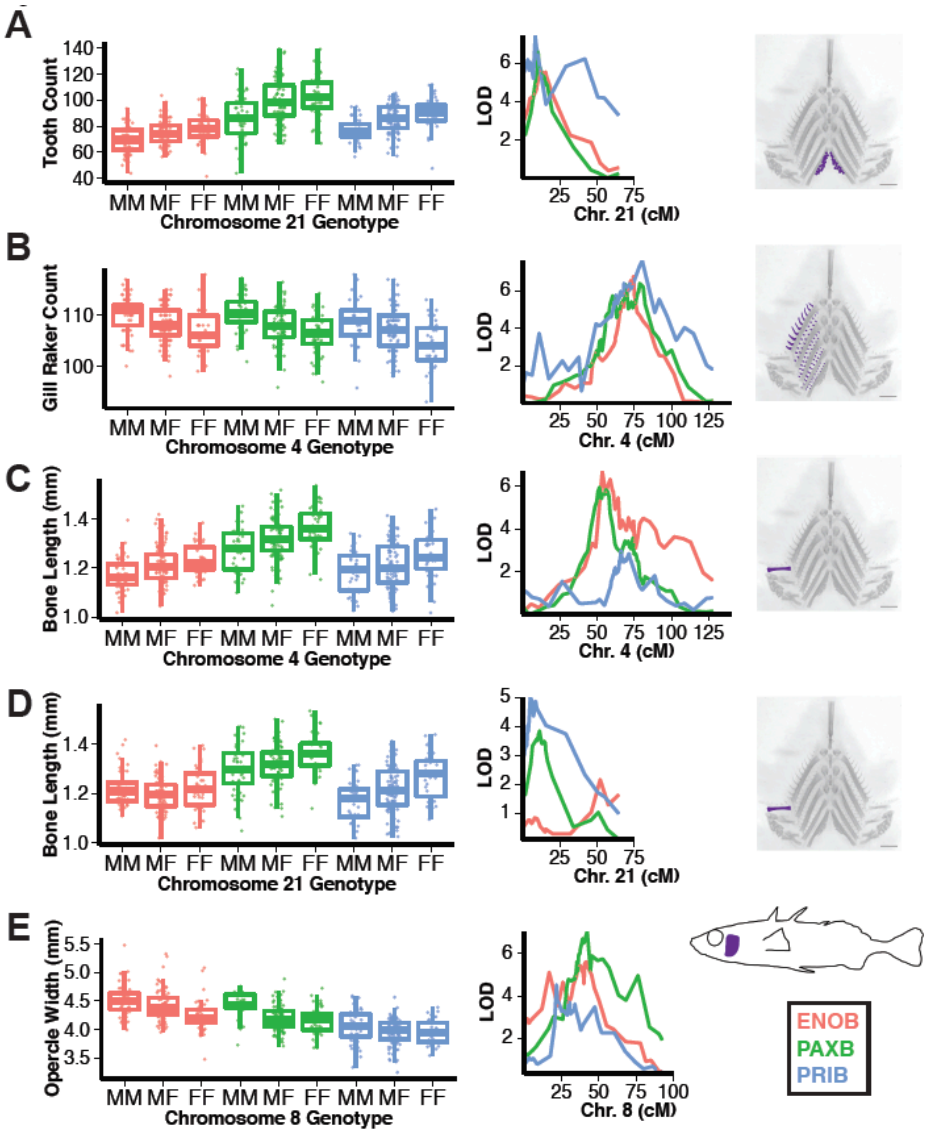


Figure 5.7: Shared craniofacial QTL.

QTL for gill raker number (chr. 4), ventral pharyngeal tooth number (chr. 21) and branchial bone length (chr. 4 and chr. 21) have previously been described in PAXB x marine crosses (Miller and Glazer et al. 2014, Glazer et al. 2014, Erickson et al. 2014, Cleves et al. 2014). Phenotypes for each genotypic class at the peak marker are shown on the left (M = marine, F = freshwater). The LOD profiles of each QTL plotted relative to genetic distance (cM) are shown in the middle, as in Figure 5.6. Craniofacial bones measured or counted are highlighted in purple to the right; images of flat-mounted branchial skeletons are shown with anterior at top. Panels A-C show triple-overlapping QTL at the genome-wide threshold; D and E each include one suggestive parallel QTL. (A): Ventral pharyngeal teeth, chr. 21; (B); All row 1-9 gill rakers, chr. 4; (C) Epibranchial 1 length, chr. 4; (D) Epibranchial 1 length, chr. 21; (E) Opercle width, chr. 8. PAXB = green, PRIB = blue, ENOB = red.

The number of overlapping QTL might be over-counted if mutations are present that affect the development of skeletal elements belonging to more than one trait class that map to the same genomic region. In total, four distinct genomic regions contain triple QTL, which is still significant in a modified permutation test that disallows multiple triple QTL in the same genomic region (P = 0.001). Therefore, we still find significant parallelism when we account for potential pleiotropy on chromosomes 4 and 21. As a second, even more conservative, control for the potential pleiotropy of QTL, we performed an additional filtering in our simulations that allows a genomic position to be covered by at most one skeletal QTL per cross, regardless of trait. Although five genomic regions contained a QTL affecting at least one trait in all three crosses, neither double nor triple QTL were statistically significantly enriched relative to the expectations from these simulations (see Table 5.5).

We might fail to detect some overlapping QTL due to small effect sizes that fail to meet the strict genome-wide LOD cutoff (Beavis 1998). To test this possibility, we performed a second search for QTL by looking for suggestive parallel QTL in the regions of the genome where QTL had been identified in at least one cross for the trait class. This analysis identified a total of 43 new suggestive QTL (13 PAXB, 13 PRIB, and 17 ENOB, Figure 5.8, Figure 5.9), including six new triply overlapping QTL that were not previously identified at the genome-wide significance threshold (Figure 5.4B). We found that 68% and 71% of the suggestive double and triple-overlapping QTL, respectively, had effects in the same direction. When these QTL are included, 67% of all QTL overlap with at least one other QTL, and 33% are unique to a single cross. Thus, the relatively small sizes of our crosses (~180 F2s) may have prevented us from detecting some overlapping QTL, causing us to underestimate parallelism in the main analysis.

QTL type	observed	mean simulated	2.5% simulated	97.5% simulated	*P*(simulated ≥ observed)
PAXB unique	18	24.2	19	29	0.9964
PRIB unique	24	29.9	25	35	0.994
ENOB unique	15	19.4	15	23	0.9826
PAXB-PRIB overlapping	6	5.6	2	10	0.5014
PAXB-ENOB overlapping	2	2.9	0	6	0.8149
PRIB-ENOB overlapping	2	4.3	1	8	0.9477
triple overlapping	8	0.6	0	2	**<0.0001**

Table 5.4: Results of QTL overlap simulation.

The physical locations of the QTL were randomly permuted 10,000 times and tested for overlap between crosses in each permutation. The total number of pairwise overlaps and triple overlaps was counted. The mean, 2.5 and 97.5 percentiles of permuted overlapping QTL are presented. The *P*-value was calculated based on the number of permutations with an equal or greater number of overlaps than the actual observed overlaps. See Figure 5.5 for the distribution of simulated numbers of QTL overlaps.

QTL type	observed	mean simulated	2.5% simulated	97.5% simulated	*P*(simulated ≥ observed)
PAXB unique	5	5.6	2	9	0.7286
PRIB unique	2	4.3	1	8	0.9679
ENOB unique	5	3.8	1	7	0.3241
PAXB-PRIB overlapping	7	6.6	3	11	0.5057
PAXB-ENOB overlapping	2	4.1	1	8	0.9485
PRIB-ENOB overlapping	3	3.8	1	7	0.7818
triple overlapping	5	3.3	1	6	0.1888

Table 5.5: Results of QTL overlap simulations without respect to QTL category.

QTL were filtered so that only the largest effect QTL covering any given chromosome region was counted. The physical locations of the filtered QTL were randomly simulated 10,000 times and tested for overlap between crosses in each simulation, regardless of trait category of the QTL. The total number of pairwise overlaps and triple overlaps was counted. The mean, 2.5 and 97.5 percentiles of permuted overlapping QTL are presented. The *P*-value was calculated based on the number of simulations with an equal or greater number of overlaps than the actual observed overlaps.

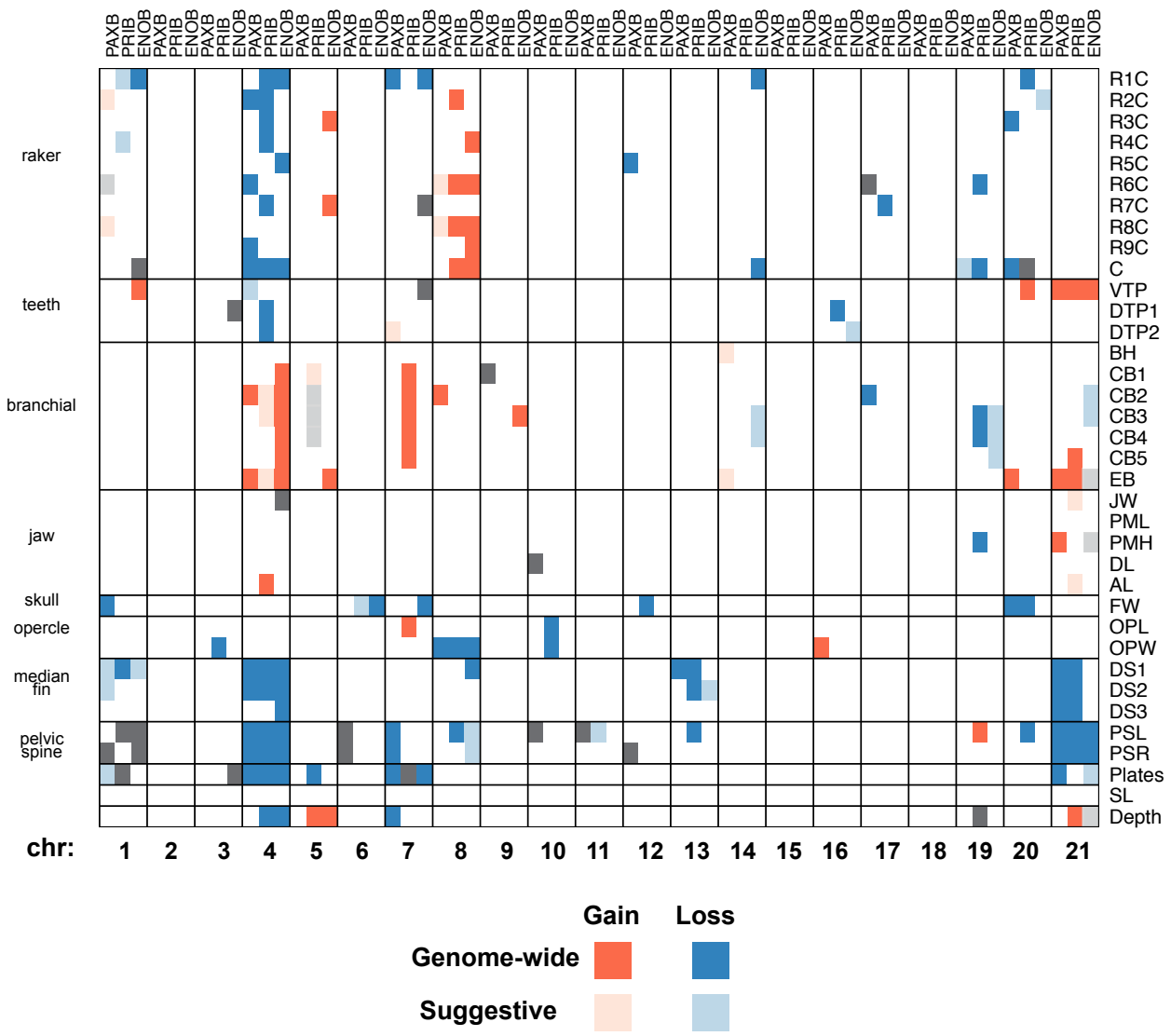


Figure 5.8: Overview of all suggestive parallel QTL relative to genome-wide QTL.

Chromosomes 1-21 are separated by vertical lines and numbered below. On each chromosome, QTL for PAXB, PRIB, and ENOB are indicated from left to right and labeled on top. Horizontal lines separate trait categories. Color intensity indicates the detection threshold: dark colors indicate genome-wide QTL; light colors indicate suggestive parallel QTL, and white indicates no detected QTL (see key). Red indicates skeletal gain QTL (freshwater allele confers more bone), blue indicates skeletal loss QTL (freshwater allele confers less bone), and grey indicates QTL with no difference between homozygous genotypes.

Figure 5.9: Location of all filtered QTL.

The physical position of each GBS marker is indicated as a vertical grey line. The physical size of the 1.5-LOD interval for each QTL is indicated with a horizontal line. The thickness of the line is proportional to the PVE of each QTL and each QTL is labeled with its trait category. Genome-wide QTL are indicated with black text and bold colors and suggestive parallel QTL are indicated with grey text and light colors. Red = ENOB, green = PAXB, blue = PRIB. The positions of three previously identified chromosomal inversions showing marine-freshwater ecotype-specific allele frequencies (Jones et al. 2012b) are marked with purple arrows.

Most QTL are not shared between lakes

Despite the significant enrichment for overlap of genomic regions influencing similar phenotypes in the three crosses, the majority of QTL identified (57%) were not shared. Several striking non-parallel genetic patterns were observed. For example, severe pelvic reduction has been described in Paxton benthics, but not in Priest or Enos benthics. Consistent with the previously described role of *Pitx1* in mediating pelvic reduction in PAXB fish (Shapiro et al. 2004; Chan et al. 2010), we detected a large effect QTL controlling presence or absence of pelvic spines in the PAXB cross that mapped to the end of chromosome 7 containing *Pitx1* (Figure 5.10). Over half of all gill raker QTL (9/16) were unique to a single cross, and of 13 QTL affecting branchial bone length, only one double-overlapping QTL was observed.

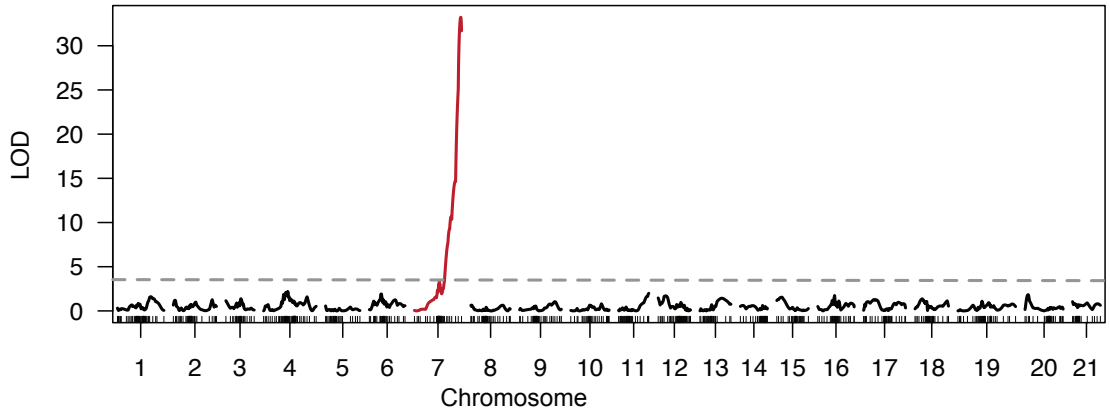


Figure 5.10: Pelvic spine presence/absence maps to chromosome 7 in PAXB.

Spine presence or absence was mapped as a binary trait using the *scanone* function in R/qtl. A single QTL (LOD = 32.4) was detected on the right end of chromosome 7 (red line). Grey dashed line indicates the genome-wide significance threshold of 3.7. *Pitx1,* shown to control pelvic spine presence/absence, is located at this right end of chromosome 7 (Shapiro et al. 2004; Chan et al. 2010).

Weak relationship between QTL effect size and parallelism

We tested the prediction of Conte et al. (2015) that large effect QTL would be more likely to overlap in multiple benthic populations than small effect QTL. Briefly, population genetics theory predicts that evolution via new mutations or standing variation should produce a positive correlation between parallelism and QTL effect size. We tested this prediction by examining the relationship between the average PVE of each QTL within a trait class and the number of overlaps for that QTL. Because traits with many small effect QTL might have QTL that overlap by chance, we restricted our analysis of PVE vs. effect size to only the largest effect QTL affecting each trait class in each cross, to reduce oversampling of trait categories with many QTL (as in Conte et al. 2015). Because small-effect QTL are less likely to be detected in parallel, we included suggestive QTL if they

were the largest effect QTL for the trait. We found a significant relationship between PVE and effect size ($\rho = 0.50$, n = 20, $P = 0.02$, Figure 5.11), though this effect was driven by the large-effect, parallel QTL on chromosome 4. When we performed this correlation analysis on the list of maximally filtered QTL regions, to reduce oversampling of individual chromosomes, the relationship was not significant ($\rho = 0.139$, n = 28, $P = 0.49$). Therefore, our results confirm those of Conte et al. (2015)—the relationship between parallelism and effect size is at most weak and driven by a few QTL on the same chromosome.

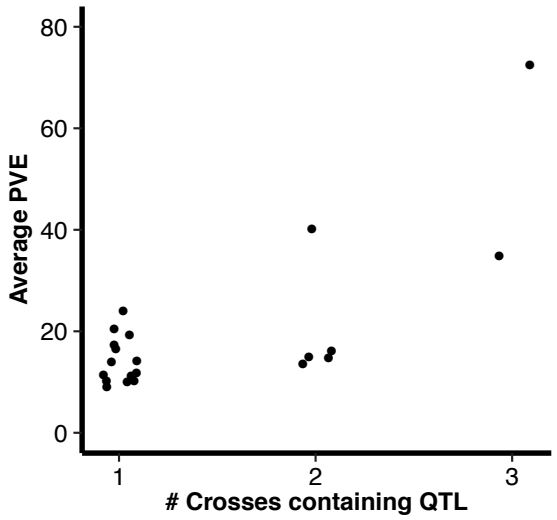


Figure 5.11: Larger effect QTL overlap in multiple benthic crosses.
The highest PVE QTL controlling each trait class, including suggestive QTL, was identified for each cross to reduce oversampling of phenotypes with multiple QTL. QTL were classified as overlapping in 1, 2, or 3 crosses based on overlap of the physical 1.5 LOD intervals for QTL controlling the same trait category. For overlapping QTL, the average PVE was calculated between crosses. Average PVE is plotted against the number of crosses in which overlapping QTL were found, and points are jittered along the x-axis to show all points. A significant relationship between effect size and parallelism was observed (Spearman rank correlation, $\rho = 0.50$, n = 20, $P = 0.02$).

Benthic QTL are clustered in the genome

In a previous PAXB x marine cross, three chromosomes (4, 20, and 21) were enriched for QTL for the studied traits (Miller et al. 2014). We hypothesized that enrichment of QTL on these chromosomes is a general feature of adaptation to freshwater benthic environments. We tested for significant clustering of QTL in all three crosses using a chi-square test and null expectations based on physical length, genetic length, and Ensembl-predicted gene number (using the revised assembly from Glazer et al. 2015). We found that, by all three expectations, the PAXB cross was enriched for QTL on chromosome 21, the PRIB cross was enriched for QTL on chromosomes 4 and 21, and the ENOB cross was

enriched for QTL on chromosome 4 (Table 5.6). Additionally, enrichment for QTL on chromosomes 19 and 5 was seen in the PRIB and ENOB crosses, respectively, for two of these three tests of clustering. Therefore, two of the three previously identified trait clusters were found in benthic populations other than PAXB and had significantly clustered QTL on chromosomes when accounting for chromosome genetic length, physical length, or gene number. Interestingly, using the same chi-square test to compare predicted gene number on each chromosome to its physical length based on the revised genome assembly of Glazer et al. 2015, we found that chromosomes 4 and 21 have significantly fewer genes than expected based on their physical length (standard residuals of -5.70 and -7.61, respectively). Despite this low gene number, they have more QTL for per unit physical length.

	By Genes (Ensembl)		By Physical Length (Mb)		By Genetic Length (cM)	
Cross	***P***	**Enriched Chr.**	***P***	**Enriched Chr.**	***P***	**Enriched Chr.**
PAXB	0.004	21	0.04	21	0.01	21
PRIB	0.008	4, 19, 21	0.034	4, 19, 21	0.019	4, 21
ENOB	0.008	4	0.018	4, 5	0.005	4,5

Table 5.6: Results of QTL clustering analyses.

A chi-square test was used to test whether the distribution of QTL across chromosomes was proportional to gene number (based on Ensembl predictions), physical length (Mb), or genetic length (cM). *P*-values were based on 10,000 permutations of the data, and enriched chromosomes are chromosomes that had standard residuals > 2 (as in Arnegard et al. 2014).

Benthic QTL are enriched for genomic signatures of natural selection

Next, we hypothesized that QTL important for benthic adaptation would be enriched for loci showing signatures of natural selection in two analyses of stickleback divergence: marine-freshwater (Jones et al. 2012b) and benthic-limnetic (Jones et al. 2012a). These studies looked for genetic variants that were shared among freshwater (or benthic) populations and differed from marine (or limnetic) populations. We tested for enrichment by calculating the number of selected loci overlapping with benthic skeletal QTL compared to a randomly permuted set of QTL. We found that shared QTL (double- and triple-overlapping) were enriched for loci showing marine-freshwater signals of selection ($P < 0.05$, Table 5.7). However, unshared QTL, those found in single lakes, were not enriched ($P = 0.18$ to 0.54, Table 5.7). Shared QTL were also enriched for a set of 46 SNPs found to have high F_{ST} in at least one benthic-limnetic species pair (Jones et al. 2012a), whereas the unshared, lake-specific QTL were not enriched (with the exception of Paxton lake, which was slightly enriched, Table 5.7). When we included the suggestive parallel QTL in the analysis, the results were similar, with double- and triple-overlapping

QTL enriched for both marine-freshwater and benthic-limnetic signals of selection (Table 5.8). Thus, the overlapping genomic regions underlying similar skeletal QTL are enriched for loci showing population genetic signals of selection, suggesting these genomic regions are under strong natural selection during benthic adaptation.

	Signals of Selection:	
QTL set:	**Marine-freshwater**	**Benthic-limnetic**
PAXB unique (n=18)	1.36 (0.18)	1.85 (**0.03**)
PRIB unique (n=24)	1.15 (0.28)	1.34 (0.11)
ENOB unique (n=15)	0.88 (0.54)	0.99 (0.44)
all double (n=10)	2.24 (**0.0011**)	1.91 (**0.002**)
all triple (n=8)	3.59 (**<0.0001**)	2.35 (**0.0006**)

Table 5.7: Shared QTL are enriched for marine-freshwater and benthic-limnetic genomic signals of selection.

The number of overlaps between loci with signals of selection and benthic QTL were counted and compared to 10,000 random permutations of the QTL locations. Values given are the fold enrichment followed by the *P*-value in parentheses. Marine-freshwater loci with signals of selection based on Jones et al. 2012b and benthic-limnetic loci with signals of selection based on Jones et al. 2012a

	Signals of Selection	
QTL set:	**Marine-freshwater**	**Benthic-limnetic**
PAXB unique (n=15)	1.24 (0.26)	2.26 (**0.0078**)
PRIB unique (n=15)	1.35 (0.18)	1.16 (0.0555)
ENOB unique (n=11)	0.55 (0.75)	1.03 (0.3676)
all double (n=23)	2.00 (**0.0001**)	1.59 (**0.0053**)
all triple (n=14)	2.68 (**<0.0001**)	2.04 (**0.0003**)

Table 5.8: Shared suggestive QTL are enriched for marine-freshwater and benthic-limnetic signals of selection.

The number of overlaps between signals of selection and benthic QTL (including suggestive QTL) were counted and compared to 10,000 random simulations of the QTL locations. Values given are the fold enrichment followed by the *P*-value in parentheses. Marine-freshwater signals of selection based on Jones et al (2012b) and benthic-limnetic signals of selection based on Jones et al (2012a).

Most QTL do not overlap three previously identified inversions

Chromosomal inversions are theoretically predicted to contribute to adaptation when there is gene flow and multiple loci with selected alleles within the inversion (Kirkpatrick and Barton 2006; Hoffmann and Rieseberg 2008). Supporting these predictions, inversions contribute to evolved differences between populations of butterflies, sparrows, and monkeyflowers (Thomas et al. 2008; Lowry and Willis 2010; Joron et al. 2011; Fishman et al. 2013; Kunte et al. 2014). In sticklebacks, three chromosomal inversions typically have different orientations in marine and freshwater populations (Jones et al. 2012b). A total of 11 detected QTL overlap one of these three inversions, a significant enrichment relative to QTL placed randomly in the genome (P = 0.02, based on 10,000 permutations). This enrichment was driven mainly by the QTL cluster on chromosome 21 (Figure 5.9). We have evidence that at least some of the QTL overlapping this inversion genetically map outside of the inversion (see Discussion).

Discussion

Parallel QTL are enriched, but the majority of QTL are non-parallel

The benthic-limnetic stickleback species pairs provide a powerful system to study ecological adaptation and incipient speciation (Schluter and Rambaut 1996; Schluter 2001). One long-standing question has been the extent of genetic parallelism underlying the benthic and limnetic phenotypic convergence across lakes (Schluter and Conte 2009; Schluter et al. 2010). Prior to this study, the only study addressing the extent of genetic parallelism underlying phenotypic convergence in this system used QTL mapping in benthic-limnetic crosses from two lakes containing species pairs (Conte et al. 2015). Here we significantly extend our understanding of the genetic basis of convergent adaptation by studying the benthic ecotype from three lakes with species pairs and using a common marine genetic background. To our knowledge, this study represents one of the first to use genome-wide linkage mapping in three independently derived, convergently evolved lineages to study the genetic basis of repeated adaptive divergence.

We found that the genomic regions underlying benthic adaptation in three independently derived populations significantly overlap, supporting the hypothesis of a parallel genetic component to convergent skeletal evolution. We found that 47% (16/33) of PAXB QTL, 40% (16/40) of PRIB QTL, and 44% (12/27) of ENOB QTL overlapped a QTL affecting a similar phenotype in at least one other benthic population. Furthermore, eight QTL underlying similar phenotypes, accounting for 20-29% of all QTL found in each cross, overlapped in all three populations, with 88% having effects in the same direction in all three crosses. It is important to note that since all QTL identified here contain multiple genes, finding overlapping QTL does not necessarily mean that the underlying genes or genetic changes are the same. The genetic resolution of these QTL is coarse (due to a small cross size of ~180 F2s per cross), and many QTL regions contain hundreds of genes. These results suggest that some shared large genomic regions repeatedly underlie benthic adaptation. However, only identifying the actual genes underlying these evolved phenotypes can answer the question of whether true genetic parallelism has occurred.

Despite this uncertainty of whether QTL parallelism reflects genetic parallelism, observing non-overlapping QTL is more straightforward to interpret, as non-overlapping QTL strongly support a non-parallel genetic basis. In this study, despite the significant enrichment for overlapping QTL, 57% of all detected QTL were found in only one population. This partially predictable but largely nonparallel basis of convergent evolution is consistent with previous findings that although 35% of divergent genomic regions within a single marine-freshwater contrast were shared with marine-freshwater divergence worldwide, 65% were not (Jones et al. 2012). Furthermore, our results of partially repeatable evolution are consistent with the degree of genetic parallelism found in a previous study of benthic and limnetic sticklebacks (Conte et al. 2015) as well as a meta-analysis of parallelism in a wide variety of organisms (Conte et al. 2012). Our results highlight a remarkable outcome: benthic adaptation has occurred three times via largely different genetic mechanisms. The largely non-parallel genetic basis for dozens of phenotypes suggests that previous findings on sticklebacks that identified a parallel genetic basis for freshwater traits (Colosimo et al. 2004, 2005; Cresko et al. 2004; Miller et al. 2007; Chan et al. 2010) are not representative of all traits.

Our results showing a partially predictable genomic basis of convergent evolution fit within a spectrum of previous work showing both repeated and non-repeated bases of convergent evolution at the QTL level in diverse organisms. Three species of *Mimulus* have all convergently evolved changes in leaf shape. In all three species, leaf shape mapped to the same two genomic regions (Ferris et al. 2015). Likewise, multiple populations of clinally-adapted *Drosophila melanogaster* share overlapping QTL for wing size (Gockel et al. 2002; Calboli et al. 2003). However, in *D. simulans*, similar clinal variation in wing size maps to one genomic region that overlaps a *melanogaster* QTL and one distinct region (Lee et al. 2011). Similar to our finding, two strains of weedy rice have adapted to the agricultural environment through non-overlapping QTL for three traits, but partially overlapping QTL for a fourth trait (Thurber et al. 2013). Thus our work adds to a growing list of partial parallelism at the QTL level, suggesting that the same genomic regions only sometimes underlie convergent adaptation, and enabling future work to test genetic parallelism.

Pleiotropy, QTL clustering, and inversions

Pleiotropic loci influencing multiple adaptive phenotypes have been observed in monkeyflowers, rice, flies, and mice (Hall et al. 2006; Yan et al. 2011; Linnen et al. 2013; Paaby et al. 2014) and could explain the multiple triple-overlapping QTL found on chromosome 4. The peak marker and 1.5 LOD intervals were highly similar for the chromosome 4 pelvic and dorsal spine length QTL within each cross, and both overlapped a gill raker QTL and the previously described *Eda* lateral plates QTL (Colosimo et al. 2004, 2005), which was found in all three crosses. Parsimoniously, a shared locus with pleiotropic effects could underlie all four phenotypes in benthic populations. Although the *Eda* haplotype that controls lateral plates has not been reported to affect gill raker or spine morphology, *Eda* mRNA expression is observed in dorsal and pelvic spine tissue (Colosimo et al. 2004, 2005; O'Brown et al. 2015), *Eda*

receptor expression is detected in forming gill raker buds (Glazer et al. 2014), and zebrafish *Eda* mutants lack dorsal fins, pelvic fins and gill rakers (Harris et al. 2008). However, in PRIB, the dorsal spine QTL does not overlap *Eda*, and in PAXB, the gill raker QTL does not overlap any other QTL, suggesting these traits are affected by separate tightly linked loci. Thus, we hypothesize chromosome 4 has at least two armor reduction loci (*Eda* plus at least one spine length locus), as well as a gill raker reduction locus, that form a skeletal reduction supergene (Schwander et al. 2014). Identifying the genes underlying these QTL and testing whether freshwater alleles are in linkage disequilibrium with the low-armor *Eda* allele in marine fish could test this hypothesis. These results combined with the overall genomic clustering of QTL suggest an important role for pleiotropic QTL and/or supergenes as drivers of parallel adaptation in benthic environments. Approaches to infer pleiotropy in QTL studies have been developed (Jiang and Zeng 1995), which future analyses could apply to correlated traits in sticklebacks.

Like previous studies, (Arnegard et al. 2014; Liu et al. 2014; Miller et al. 2014) we mapped multiple QTL to chromosomes 4 and 21 in some crosses. Unlike Miller et al. (2014), we did not identify a chromosome 20 QTL cluster, perhaps because the PAXB grandparent used in this cross and the grandparent used in the previous cross had different chromosome 20 alleles. Additionally, the relatively smaller sizes of our cross and reduced number of phenotypes scored may have prevented us from detecting some loci that contribute to clustering (e.g. we did not phenotype several bones that were part of the cluster found in Miller et al. 2014). Furthermore, because clustering was seen when we adjusted for genetic length, physical length, or gene number, clustering does not appear to result simply from recombination suppression or differential gene density between chromosomes.

Theoretical work has proposed that inversions could evolve because they cluster adaptive loci and prevent their recombination (Kirkpatrick and Barton 2006; Hoffmann and Rieseberg 2008). In sticklebacks, inversions on chromosomes 1, 11, and 21 are oppositely fixed in most marine and freshwater populations including the PAXB population (Jones et al. 2012). These three inversions were significantly enriched within detected QTL intervals, including two triple QTL and two double QTL on chromosome 21. However, in the PAXB population, tooth number fine-maps to a genomic interval over a megabase outside of the inversion (Cleves et al. 2014), so the triple-overlapping tooth number QTL is likely not in the inversion. Likewise, the PAXB pelvic spine length QTL also maps entirely outside the inversion, so the triple-overlapping pelvic spine length QTL is also unlikely to be in the inversion. Chromosome 4, highly enriched for QTL and triple-overlapping QTL, does not contain one of these inversions. Therefore, although a few QTL could be due to mutations within inversions, most QTL involved in benthic skeletal adaptation (at least 89/100, since only 11/100 QTL overlap one of these three inversions) do not map to these three previously described inversions.

Shared QTL and freshwater adaptation

Many of the triple QTL we identified have been found in previous studies, but little was known about their parallelism. A chromosome 4 gill raker reduction QTL was found in

three freshwater populations (Glazer et al. 2014) and was detected in all three benthic populations here. A chromosome 21 tooth QTL maps to a *cis*-regulatory allele of the *Bmp6* gene in the PAXB population (Cleves et al. 2014), and the presence of a similar tooth gain QTL in two other benthic populations suggests that increased tooth number is adaptive in benthic environments. Chromosomes 4 and 21 also increase branchial bone length in multiple freshwater populations (Erickson et al. 2014), and each QTL was found in a second benthic population, suggesting increased branchial bone length may also be adaptive. Continued mapping of these QTL will help determine whether the parallelism we observe is due to different tightly linked genes, different mutations in the same gene, or repeated selection of standing variants. Given the geographic proximity of the lakes, we hypothesize that common shared ancestral alleles underlie the shared QTL, whereas QTL unique to a single cross are due to rarer ancestral variants or new mutations.

Genomic techniques such as RAD-seq and genome sequencing have enabled the discovery of local and temporal genomic signatures of selection in a variety of organisms including maize (Hufford et al. 2012), monkeyflowers (Stankowski and Streisfeld 2015), flies (Bergland et al. 2014), stick insects (Soria-Carrasco et al. 2014), wolves (Schweizer et al. 2015), cichlids (Ford et al. 2015), whitefish (Laporte et al. 2015), salmon (Seeb et al. 2014), and sheep (Kardos et al. 2015). However, an understanding of the phenotypes underlying these regions is often far more limited. QTL mapping of diverse phenotypes that differ between populations can provide a starting point to connect genomic signatures of selection to loci affecting morphology and physiology. We found that overlapping QTL were strongly enriched for genomic signatures of recurrent natural selection in multiple freshwater populations (Jones et al. 2012b). Importantly, these genomic regions are divergent in multiple freshwater and marine populations, so the QTL enrichment in these regions may be related to skeletal changes involved in general freshwater adaptation, rather than benthic adaptation. However, shared QTL were also enriched for SNPs that are F_{ST} outliers between limnetic and benthic fish (Jones et al. 2012a), suggesting some of the QTL might underlie benthic adaptation. Whether the QTL we identified are due to genetic variants that drive these signals of selection or whether the QTL are hitchhiking along with other loci important for freshwater adaptation remains currently unknown. However, the signals of selection can pinpoint interesting candidate genes for the QTL intervals, which ultimately could link these population genetic signals of selection to adaptive phenotypes.

Acknowledgements

Andrew Glazer wrote many of the computer programs involved in genotype calling and high-throughput QTL mapping to develop an analysis pipeline. Emily Killingbeck performed most of the molecular biology for library construction and was involved in developing the GBS method. Rachel Agoglia, Anthony Lee, and Sara Carsanaro performed extensive phenotyping and DNA extractions, and Jiyeon Baek performed additional phenotyping. Phillip Cleves was instrumental in initiating the collaboration that led to this project and supervised much of the phenotyping. Dolph Schluter generated the crosses and provided significant input on the statistical analysis.

We thank Chris Martin for helpful suggestions on data analysis. This work was supported in part by NIH R01 #DE021475 (C.M.), an NIH Predoctoral Training Grant #5T32GM007127 (P.E.), NSF Graduate Research Fellowships (A.G. and P.C.), and Discovery grants from the Natural Sciences and Engineering Research Council of Canada (D.S.). This work used the Vincent J. Coates Genomics Sequencing Laboratory at UC Berkeley, supported by NIH S10 Instrumentation Grants S10RR029668 and S10RR027303.

References

Arnegard, M. E., M. D. McGee, B. Matthews, K. B. Marchinko, G. L. Conte, S. Kabir, N. Bedford, S. Bergek, Y. F. Chan, F. C. Jones, D. M. Kingsley, C. L. Peichel, and D. Schluter. 2014. Genetics of ecological divergence during speciation. Nature 511:307–311.

Barrett, R. D. H., and D. Schluter. 2008. Adaptation from standing genetic variation. Trends Ecol. Evol. 23:38–44.

Beavis, W. 1998. Limitations of QTL mapping: power, precision, and accuracy. Pp. 145–162 *in* A. H. Paterson, ed. Molecular Dissection of Complex Traits. CRC Press, Boca Raton, FL.

Bell, M. A., and W. E. Aguirre. 2013. Contemporary evolution, allelic recycling, and adaptive radiation of the threespine stickleback. Evol. Ecol. Res. 15:377–411.

Bell, M. A., and S. A. Foster. 1994. The evolutionary biology of the threespine stickleback. Oxford Univ. Press, New York.

Bentzen, P., and J. D. McPhail. 1984. Ecology and evolution of sympatric sticklebacks (*Gasterosteus*): specialization for alternative trophic niches in the Enos Lake species pair. Can. J. Zool. 62:2280–2286.

Bergland, A. O., E. L. Behrman, K. R. O'Brien, P. S. Schmidt, and D. A. Petrov. 2014. Genomic evidence of rapid and stable adaptive oscillations over seasonal time scales in *Drosophila*. PLoS Genet 10:e1004775.

Calboli, F. C. F., W. J. Kennington, and L. Partridge. 2003. QTL mapping reveals a striking coincidence in the positions of genomic regions associated with adaptive variation in body size in parallel clines of *Drosophila melanogaster* on different continents. Evolution 57:2653–2658.

Chan, Y. F., M. E. Marks, F. C. Jones, G. Villarreal, M. D. Shapiro, S. D. Brady, A. M. Southwick, D. M. Absher, J. Grimwood, J. Schmutz, R. M. Myers, D. Petrov, B. Jónsson, D. Schluter, M. A. Bell, and D. M. Kingsley. 2010. Adaptive evolution of pelvic reduction in sticklebacks by recurrent deletion of a *Pitx1* enhancer. Science 327:302–305.

Cleves, P. A., N. A. Ellis, M. T. Jimenez, S. M. Nunez, D. Schluter, D. M. Kingsley, and C. T. Miller. 2014. Evolved tooth gain in sticklebacks is associated with a *cis*-regulatory allele of *Bmp6*. Proc. Natl. Acad. Sci. 111:13912–13917.

Colosimo, P. F., K. E. Hosemann, S. Balabhadra, G. Villarreal, M. Dickson, J. Grimwood, J. Schmutz, R. M. Myers, D. Schluter, and D. M. Kingsley. 2005. Widespread parallel evolution in sticklebacks by repeated fixation of *Ectodysplasin* alleles. Science 307:1928–1933.

Colosimo, P. F., C. L. Peichel, K. Nereng, B. K. Blackman, M. D. Shapiro, D. Schluter, and D. M. Kingsley. 2004. The genetic architecture of parallel armor plate reduction in threespine sticklebacks. PLoS Biol 2:e109.

Conte, G. L., M. E. Arnegard, J. Best, Y. F. Chan, F. C. Jones, D. M. Kingsley, D. Schluter, and C. L. Peichel. 2015. Extent of QTL reuse during repeated phenotypic divergence of sympatric threespine stickleback. Genetics 201:1189-1200.

Conte, G. L., M. E. Arnegard, C. L. Peichel, and D. Schluter. 2012. The probability of genetic parallelism and convergence in natural populations. Proc. R. Soc. Lond. B Biol. Sci. 279:5039–5047.

Cresko, W. A., A. Amores, C. Wilson, J. Murphy, M. Currey, P. Phillips, M. A. Bell, C. B. Kimmel, and J. H. Postlethwait. 2004. Parallel genetic basis for repeated evolution of armor loss in Alaskan threespine stickleback populations. Proc. Natl. Acad. Sci. U. S. A. 101:6050–6055.

Ellis, N. A., A. M. Glazer, N. N. Donde, P. A. Cleves, R. M. Agoglia, and C. T. Miller. 2015. Distinct developmental and genetic mechanisms underlie convergently evolved tooth gain in sticklebacks. Development 142: 2442-2451.

Elshire, R. J., J. C. Glaubitz, Q. Sun, J. A. Poland, K. Kawamoto, E. S. Buckler, and S. E. Mitchell. 2011. A robust, simple genotyping-by-sequencing (GBS) approach for high diversity species. PLoS ONE 6:e19379.

Erickson, P. A., A. M. Glazer, P. A. Cleves, A. S. Smith, and C. T. Miller. 2014. Two developmentally temporal quantitative trait loci underlie convergent evolution of increased branchial bone length in sticklebacks. Proc. R. Soc. Lond. B Biol. Sci. 281:20140822.

Falconer, D. S., and T. F. C. Mackay. 1996. Introduction to quantitative genetics. Longman.

Ferris, K. G., T. Rushton, A. B. Greenlee, K. Toll, B. K. Blackman, and J. H. Willis. 2015. Leaf shape evolution has a similar genetic architecture in three edaphic specialists within the *Mimulus guttatus* species complex. Ann. Bot. 116:213-223.

Fishman, L., A. Stathos, P. M. Beardsley, C. F. Williams, and J. P. Hill. 2013. Chromosomal rearrangements and the genetics of reproductive barriers in *Mimulus* (monkey flowers). Evolution 67:2547–2560.

Ford, A. G. P., K. K. Dasmahapatra, L. Rüber, K. Gharbi, T. Cezard, and J. J. Day. 2015. High levels of interspecific gene flow in an endemic cichlid fish adaptive radiation from an extreme lake environment. Mol. Ecol. 24:3421–3440.

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 5:1463-1472.

Gockel, J., S. J. W. Robinson, W. J. Kennington, D. B. Goldstein, and L. Partridge. 2002. Quantitative genetic analysis of natural variation in body size in *Drosophila melanogaster*. Heredity 89:145–153.

Gow, J. L., C. L. Peichel, and E. B. Taylor. 2007. Ecological selection against hybrids in natural populations of sympatric threespine sticklebacks. J. Evol. Biol. 20:2173–2180.

Hagen, D. W. 1967. Isolating mechanisms in threespine sticklebacks (*Gasterosteus*). J. Fish. Res. Board Can. 24:1637–1692.

Hall, M. C., C. J. Basten, and J. H. Willis. 2006. Pleiotropic quantitative trait loci contribute to population divergence in traits associated with life-history variation in *Mimulus guttatus*. Genetics 172:1829–1844.

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

Hoffmann, A. A., and L. H. Rieseberg. 2008. Revisiting the impact of inversions in evolution: from population genetic markers to drivers of adaptive shifts and speciation? Annu. Rev. Ecol. Evol. Syst. 39:21–42.

Hohenlohe, P. A., S. Bassham, M. Currey, and W. A. Cresko. 2012. Extensive linkage disequilibrium and parallel adaptive divergence across threespine stickleback genomes. Philos. Trans. R. Soc. Lond. B Biol. Sci. 367:395–408.

Huerta-Sánchez, E., X. Jin, Asan, Z. Bianba, B. M. Peter, N. Vinckenbosch, Y. Liang, X. Yi, M. He, M. Somel, P. Ni, B. Wang, X. Ou, Huasang, J. Luosang, Z. X. P. Cuo, K. Li, G. Gao, Y. Yin, W. Wang, X. Zhang, X. Xu, H. Yang, Y. Li, J. Wang, J. Wang, and R. Nielsen. 2014. Altitude adaptation in Tibetans caused by introgression of Denisovan-like DNA. Nature 512:194–197.

Hufford, M. B., X. Xu, J. van Heerwaarden, T. Pyhäjärvi, J.-M. Chia, R. A. Cartwright, R. J. Elshire, J. C. Glaubitz, K. E. Guill, S. M. Kaeppler, J. Lai, P. L. Morrell, L. M. Shannon, C. Song, N. M. Springer, R. A. Swanson-Wagner, P. Tiffin, J. Wang, G. Zhang, J. Doebley, M. D. McMullen, D. Ware, E. S. Buckler, S. Yang, and J. Ross-Ibarra. 2012. Comparative population genomics of maize domestication and improvement. Nat. Genet. 44:808–811.

Jiang, C., and Z. B. Zeng. 1995. Multiple trait analysis of genetic mapping for quantitative trait loci. Genetics 140:1111–1127.

Jones, F. C., C. Brown, and V. A. Braithwaite. 2008. Lack of assortative mating between incipient species of stickleback from a hybrid zone. Behaviour 145:463–484.

Jones, F. C., C. Brown, J. M. Pemberton, and V. A. Braithwaite. 2006. Reproductive isolation in a threespine stickleback hybrid zone. J. Evol. Biol. 19:1531–1544.

Jones, F. C., Y. F. Chan, J. Schmutz, J. Grimwood, S. D. Brady, A. M. Southwick, D. M. Absher, R. M. Myers, T. E. Reimchen, B. E. Deagle, D. Schluter, and D. M. Kingsley. 2012a. A genome-wide SNP genotyping array reveals patterns of global and repeated species-pair divergence in sticklebacks. Curr. Biol. 22:83–90.

Jones, F. C., M. G. Grabherr, Y. F. Chan, P. Russell, E. Mauceli, J. Johnson, R. Swofford, M. Pirun, M. C. Zody, S. White, E. Birney, S. Searle, J. Schmutz, J. Grimwood, M. C. Dickson, R. M. Myers, C. T. Miller, B. R. Summers, A. K. Knecht, S. D. Brady, H. Zhang, A. A. Pollen, T. Howes, C. Amemiya, Broad Institute Genome Sequencing Platform & Whole Genome Assembly Team, E. S. Lander, F. Di Palma, K. Lindblad-Toh, and D. M. Kingsley. 2012b. The genomic basis of adaptive evolution in threespine sticklebacks. Nature 484:55–61.

Joron, M., L. Frezal, R. T. Jones, N. L. Chamberlain, S. F. Lee, C. R. Haag, A. Whibley, M. Becuwe, S. W. Baxter, L. Ferguson, P. A. Wilkinson, C. Salazar, C. Davidson, R. Clark, M. A. Quail, H. Beasley, R. Glithero, C. Lloyd, S. Sims, M. C. Jones, J. Rogers, C. D. Jiggins, and R. H. Ffrench-Constant. 2011. Chromosomal rearrangements maintain a polymorphic supergene controlling butterfly mimicry. Nature 477:203–206.

Kardos, M., G. Luikart, R. Bunch, S. Dewey, W. Edwards, S. McWilliam, J. Stephenson, F. W. Allendorf, J. T. Hogg, and J. Kijas. 2015. Whole genome resequencing uncovers molecular signatures of natural and sexual selection in wild bighorn sheep. Mol. Ecol. 24:5616-5632.

Kassen, R., D. Schluter, and J. D. McPhail. 1995. Evolutionary history of threespine sticklebacks (*Gasterosteus* spp) in British Columbia: insights from a physiological clock. Can. J. Zool. 73:2154–2158.

Kingsley, E. P., M. Manceau, C. D. Wiley, and H. E. Hoekstra. 2009. Melanism in *Peromyscus* is caused by independent mutations in *Agouti*. PLoS ONE 4:e6435.

Kirkpatrick, M., and N. Barton. 2006. Chromosome inversions, local adaptation and speciation. Genetics 173:419–434.

Kowalko, J. E., N. Rohner, T. A. Linden, S. B. Rompani, W. C. Warren, R. Borowsky, C. J. Tabin, W. R. Jeffery, and M. Yoshizawa. 2013. Convergence in feeding posture occurs through different genetic loci in independently evolved cave populations of *Astyanax mexicanus*. Proc. Natl. Acad. Sci. 110:16933–16938.

Kraak, S. B. M., B. Mundwiler, and P. J. B. Hart. 2001. Increased number of hybrids between benthic and limnetic three-spined sticklebacks in Enos Lake, Canada; the collapse of a species pair? J. Fish Biol. 58:1458–1464.

Kunte, K., W. Zhang, A. Tenger-Trolander, D. H. Palmer, A. Martin, R. D. Reed, S. P. Mullen, and M. R. Kronforst. 2014. *doublesex* is a mimicry supergene. Nature 507:229–232.

Lamichhaney, S., J. Berglund, M. S. Almén, K. Maqbool, M. Grabherr, A. Martinez-Barrio, M. Promerová, C.-J. Rubin, C. Wang, N. Zamani, B. R. Grant, P. R. Grant, M. T. Webster, and L. Andersson. 2015. Evolution of Darwin's finches and their beaks revealed by genome sequencing. Nature 518:371–375.

Laporte, M., S. M. Rogers, A.-M. Dion-Côté, E. Normandeau, P.-A. Gagnaire, A. C. Dalziel, J. Chebib, and L. Bernatchez. 2015. RAD-QTL mapping reveals both genome-level parallelism and different genetic architecture underlying the evolution of body shape in lake whitefish (*Coregonus clupeaformis*) species pairs. G3 5:1481-1491.

Lavin, P. A., and J. D. Mcphail. 1985. The evolution of freshwater diversity in the threespine stickleback (*Gasterosteus aculeatus*): site-specific differentiation of trophic morphology. Can. J. Zool. 63:2632–2638.

Lee, S. F., L. Rako, and A. A. Hoffmann. 2011. Genetic mapping of adaptive wing size variation in *Drosophila simulans*. Heredity 107:22–29.

Linnen, C. R., Y.-P. Poh, B. K. Peterson, R. D. H. Barrett, J. G. Larson, J. D. Jensen, and H. E. Hoekstra. 2013. Adaptive evolution of multiple traits through multiple mutations at a single gene. Science 339:1312–1316.

Liu, J., T. Shikano, T. Leinonen, J. M. Cano, M.-H. Li, and J. Merilä. 2014. Identification of major and minor QTL for ecologically important morphological traits in three-spined sticklebacks (*Gasterosteus aculeatus*). G3 4:595–604.

Lowry, D. B., and J. H. Willis. 2010. A widespread chromosomal inversion polymorphism contributes to a major life-history transition, local adaptation, and reproductive isolation. PLoS Biol 8:e1000500.

McPhail, J. D. 1992. Ecology and evolution of sympatric sticklebacks (*Gasterosteus*): evidence for a species-pair in Paxton Lake, Texada Island, British Columbia. Can. J. Zool. 70:361–369.

McPhail, J. D. 1984. Ecology and evolution of sympatric sticklebacks (*Gasterosteus*): morphological and genetic evidence for a species pair in Enos Lake, British Columbia. Can. J. Zool. 62:1402–1408.

McPhail, J. D. 1994. Speciation and the evolution of reproductive isolation in the sticklebacks (*Gasterosteus*) of south-western British Columbia. Pp. 399–437 *in* The evolutionary biology of the threespine stickleback. Oxford Univ. Press, New York.

Miller, C. T., S. Beleza, A. A. Pollen, D. Schluter, R. A. Kittles, M. D. Shriver, and D. M. Kingsley. 2007. *cis*-Regulatory changes in *Kit Ligand* expression and parallel evolution of pigmentation in sticklebacks and humans. Cell 131:1179–1189.

Miller, C. T., A. M. Glazer, B. R. Summers, B. K. Blackman, A. R. Norman, M. D. Shapiro, B. L. Cole, C. L. Peichel, D. Schluter, and D. M. Kingsley. 2014. Modular skeletal evolution in sticklebacks is controlled by additive and clustered quantitative trait loci. Genetics 197:405–420.

O'Brown, N. M., B. R. Summers, F. C. Jones, S. D. Brady, and D. M. Kingsley. 2015. A recurrent regulatory change underlying altered expression and Wnt response of the stickleback armor plates gene *EDA*. eLife e05290.

Paaby, A. B., A. O. Bergland, E. L. Behrman, and P. S. Schmidt. 2014. A highly pleiotropic amino acid polymorphism in the *Drosophila* insulin receptor contributes to life-history adaptation. Evolution 68:3395–3409.

Protas, M. E., C. Hersey, D. Kochanek, Y. Zhou, H. Wilkens, W. R. Jeffery, L. I. Zon, R. Borowsky, and C. J. Tabin. 2006. Genetic analysis of cavefish reveals molecular convergence in the evolution of albinism. Nat. Genet. 38:107–111.

Rosenblum, E. B., C. E. Parent, and E. E. Brandt. 2014. The molecular basis of phenotypic convergence. Annu. Rev. Ecol. Evol. Syst. 45:203–226.

Rosenblum, E. B., H. Römpler, T. Schöneberg, and H. E. Hoekstra. 2010. Molecular and functional basis of phenotypic convergence in white lizards at White Sands. Proc. Natl. Acad. Sci. 107:2113–2117.

Schluter, D. 1993. Adaptive radiation in sticklebacks: size, shape, and habitat use efficiency. Ecology 74:699–709.

Schluter, D. 2001. Ecology and the origin of species. Trends Ecol. Evol. 16:372–380.

Schluter, D. 2000. The ecology of adaptive radiation. Oxford University Press.

Schluter, D., and G. L. Conte. 2009. Genetics and ecological speciation. Proc. Natl. Acad. Sci. 106:9955–9962.

Schluter, D., K. B. Marchinko, R. D. H. Barrett, and S. M. Rogers. 2010. Natural selection and the genetics of adaptation in threespine stickleback. Philos. Trans. R. Soc. Lond. B Biol. Sci. 365:2479–2486.

Schluter, D., and J. D. McPhail. 1992. Ecological character displacement and speciation in sticklebacks. Am. Nat. 140:85–108.

Schluter, D. 1996. Ecological speciation in postglacial fishes. Philos. Trans. R. Soc. Lond. B Biol. Sci. 351:807–814.

Schwander, T., R. Libbrecht, and L. Keller. 2014. Supergenes and complex phenotypes. Curr. Biol. 24:R288–R294.

Schweizer, R. M., B. M. vonHoldt, R. Harrigan, J. C. Knowles, M. Musiani, D. Coltman, J. Novembre, and R. K. Wayne. 2015. Genetic subdivision and candidate genes under selection in North American gray wolves. Mol. Ecol. 25:380-402.

Seeb, L. W., R. K. Waples, M. T. Limborg, K. I. Warheit, C. E. Pascal, and J. E. Seeb. 2014. Parallel signatures of selection in temporally isolated lineages of pink salmon. Mol. Ecol. 23:2473–2485.

Shapiro, M. D., M. E. Marks, C. L. Peichel, B. K. Blackman, K. S. Nereng, B. Jónsson, D. Schluter, and D. M. Kingsley. 2004. Genetic and developmental basis of evolutionary pelvic reduction in threespine sticklebacks. Nature 428:717–723.

Song, Y., S. Endepols, N. Klemann, D. Richter, F.-R. Matuschka, C.-H. Shih, M. W. Nachman, and M. H. Kohn. 2011. Adaptive introgression of anticoagulant rodent poison resistance by hybridization between old world mice. Curr. Biol. 21:1296–1301.

Soria-Carrasco, V., Z. Gompert, A. A. Comeault, T. E. Farkas, T. L. Parchman, J. S. Johnston, C. A. Buerkle, J. L. Feder, J. Bast, T. Schwander, S. P. Egan, B. J. Crespi, and P. Nosil. 2014. Stick insect genomes reveal natural selection's role in parallel speciation. Science 344:738–742.

Stankowski, S., and M. A. Streisfeld. 2015. Introgressive hybridization facilitates adaptive divergence in a recent radiation of monkeyflowers. Proc R Soc B 282:20151666.

Stern, D. L. 2013. The genetic causes of convergent evolution. Nat. Rev. Genet. 14:751–764.

Stern, D. L., and V. Orgogozo. 2009. Is genetic evolution predictable? Science 323:746–751.

Stern, D. L., and V. Orgogozo. 2008. The loci of evolution: how predictable is genetic evolution? Evolution 62:2155–2177.

Taylor, E. B., J. W. Boughman, M. Groenenboom, M. Sniatynski, D. Schluter, and J. L. Gow. 2006. Speciation in reverse: morphological and genetic evidence of the collapse of a three-spined stickleback (*Gasterosteus aculeatus*) species pair. Mol. Ecol. 15:343–355.

Taylor, E. B., and J. D. McPhail. 1999. Evolutionary history of an adaptive radiation in species pairs of threespine sticklebacks (*Gasterosteus*): insights from mitochondrial DNA. Biol. J. Linn. Soc. 66:271–291.

Taylor, E. B., and J. D. McPhail. 2000. Historical contingency and ecological determinism interact to prime speciation in sticklebacks, *Gasterosteus*. Proc. R. Soc. Lond. B Biol. Sci. 267:2375–2384.

Terekhanova, N. V., M. D. Logacheva, A. A. Penin, T. V. Neretina, A. E. Barmintseva, G. A. Bazykin, A. S. Kondrashov, and N. S. Mugue. 2014. Fast evolution from precast bricks: genomics of young freshwater populations of threespine stickleback *Gasterosteus aculeatus*. PLoS Genet 10:e1004696.

The Heliconius Genome Consortium. 2012. Butterfly genome reveals promiscuous exchange of mimicry adaptations among species. Nature 487:94–98.

Thomas, J. W., M. Cáceres, J. J. Lowman, C. B. Morehouse, M. E. Short, E. L. Baldwin, D. L. Maney, and C. L. Martin. 2008. The chromosomal polymorphism linked to variation in social behavior in the white-throated sparrow (*Zonotrichia albicollis*) is a complex rearrangement and suppressor of recombination. Genetics 179:1455–1468.

Thurber, C. S., M. H. Jia, Y. Jia, and A. L. Caicedo. 2013. Similar traits, different genes? Examining convergent evolution in related weedy rice populations. Mol. Ecol. 22:685–698.

Vickrey, A. I., E. T. Domyan, M. P. Horvath, and M. D. Shapiro. 2015. Convergent evolution of head crests in two domesticated columbids is associated with different missense mutations in *EphB2*. Mol. Biol. Evol. 32:2657-2664.

Wittkopp, P. J., B. L. Williams, J. E. Selegue, and S. B. Carroll. 2003. *Drosophila* pigmentation evolution: divergent genotypes underlying convergent phenotypes. Proc. Natl. Acad. Sci. 100:1808–1813.

Yan, W.-H., P. Wang, H.-X. Chen, H.-J. Zhou, Q.-P. Li, C.-R. Wang, Z.-H. Ding, Y.-S. Zhang, S.-B. Yu, Y.-Z. Xing, and Q.-F. Zhang. 2011. A major QTL, *Ghd8*, plays pleiotropic roles in regulating grain productivity, plant height, and heading date in rice. Mol. Plant 4:319–330.

This appendix will appear as a video and article in the *Journal of Visualized Experiments*, 2016. All authors have agreed to its use in this dissertation.

A. Microinjection for transgenesis and genome editing in threespine sticklebacks

Priscilla A. Erickson, Nicholas. A. Ellis and Craig T. Miller

Department of Molecular and Cell Biology
University of California, Berkeley

Abstract:

The threespine stickleback fish has emerged as a powerful system to study the genetic basis of a wide variety of morphological, physiological, and behavioral phenotypes. The remarkably diverse phenotypes that have evolved as marine populations adapt to countless freshwater environments, combined with the ability to cross marine and freshwater forms, provide a rare vertebrate system in which genetics can be used to map genomic regions controlling evolved traits. Excellent genomic resources are now available, facilitating molecular genetic dissection of evolved changes. While mapping experiments generate lists of interesting candidate genes, functional genetic manipulations are required to test the roles of these genes. Gene regulation can be studied with transgenic reporter plasmids and BACs integrated into the genome using the Tol2 transposase. Functions of specific candidate genes and *cis*-regulatory elements can be assessed by inducing targeted mutations with TALEN and CRISPR/Cas9 genome editing reagents. All methods require introducing nucleic acids into fertilized one-cell stickleback embryos, a task made challenging by the thick chorion of stickleback embryos and the relatively small and thin blastomere. Here, a detailed protocol for microinjection of nucleic acids into stickleback embryos is described for transgenic and genome editing applications to study gene expression and function, as well as techniques to assess the success of transgenesis and recover stable lines.

Introduction:

One fundamental component of understanding how biodiversity arises is determining the genetic and developmental bases of evolved phenotypic changes in nature. The threespine stickleback fish, *Gasterosteus aculeatus*, has emerged as an excellent model for studying the genetic basis of evolution. Sticklebacks have undergone many adaptive evolutionary changes as marine fish have colonized countless freshwater environments around the northern hemisphere, resulting in dramatic morphological, physiological, and behavioral changes[1]. The genomes of individuals from twenty-one stickleback populations have been sequenced and assembled, and a high density linkage map has been generated to further improve the assembly[2,3]. Genetic mapping experiments have identified genomic regions underlying evolved phenotypes[4–6], and in a few cases, the functional roles of specific candidate genes have been tested[7,8]. A number of genomic regions underlying morphological changes have been identified with promising candidate genes, but these candidates have not yet been functionally tested[9–12]. In addition, sticklebacks are common models for studies of population genetics/genomics[13,14], speciation[15], behavior[1], endocrinology[16], ecotoxicology[17], immunology[18] and parasitology[19]. Future studies in each of these fields will benefit from the ability to perform functional genetic manipulations in sticklebacks. In addition to manipulating their coding sequences, the roles of candidate genes can be assessed by studying their *cis*-regulatory sequences and by functionally increasing, decreasing, or eliminating expression of the candidate gene. Microinjection and transgenesis methods in sticklebacks are well established[7,8,20] and were initially developed using a meganuclease-mediated method[21] first described in medaka[22]. The

modified microinjection method presented here has been optimized for both Tol2-mediated transgenesis and recently developed genome editing reagents including TALENs and CRISPRs.

Changes to *cis*-regulatory elements are thought to be critical to morphological evolution, as *cis*-regulatory changes can avoid the negative pleiotropic consequences of coding mutations[23]. Therefore, testing and comparing putative *cis*-regulatory sequences has become a central goal of an increasing number of evolutionary studies. In addition, most human disease variants are regulatory variants[24,25], and model vertebrate systems are sorely needed to study *cis*-regulatory element function and logic. Fish that fertilize their embryos externally in large numbers offer powerful vertebrate systems to study *cis*-regulation. The Tol2 transposon system, in which foreign DNA to be integrated in the genome is flanked by Tol2 transposase binding sites and co-injected with Tol2 transposase mRNA, works with high efficiency for successfully integrating plasmid constructs into fish genomes[26–28]. Typically, a potential enhancer is cloned upstream of a basal promoter (such as *hsp70l*[29]) and fluorescent reporter molecule such as EGFP (enhanced green fluorescent protein) or mCherry in a Tol2 backbone and injected with transposase mRNA[26]. Observation of expression of the fluorescent reporter, either in injected embryos or offspring with stably integrated transgenes, provides information about the spatiotemporal regulation of gene expression driven by the putative enhancer. In further experiments, validated enhancers can be used to drive tissue-specific overexpression of genes of interest.

For analysis of larger *cis*-regulatory regions, high quality large-insert genomic libraries using bacterial artificial chromosomes (BACs) have been constructed for both marine and freshwater sticklebacks[30]. These BACs can be recombineered to replace a gene with a fluorescent reporter molecule in the context of a large (150-200 kb) genomic region[31]. The fluorescent reporter is then expressed in a spatiotemporal pattern as determined by regulatory sequences within the BAC. For studies in fish, Tol2 sites can be added to the BAC to facilitate genomic integration[32,33]. In later stages of development when *in situ* hybridization is technically challenging, the fluorescent readout of the BAC can be used to study patterns of gene expression, as has been shown for stickleback *Bone morphogenetic protein 6* (*Bmp6*)[20]. Additionally, fluorescent expression patterns in an individual can be tracked over time, which cannot be accomplished with *in situ* hybridization. BACs can also be used to add an additional copy of a genomic region to increase dosage of a gene of interest.

For the study of gene function, genome editing is an explosively expanding field that can be used to produce targeted changes to genomic sequences in a wide variety of organisms[34]. Transcription activator-like effector nucleases (TALENs) are modular, sequence-specific nucleases originally isolated from plant pathogens that can be precisely engineered to bind directly to a genomic sequence of choice and generate a double strand break[35,36]. Clustered regularly interspaced short palindromic repeats (CRISPRs) were originally found in bacteria and use a guide RNA and the Cas9 mRNA or protein to

generate a break in a target DNA sequence complementary to the guide[37]. The subsequent repair of the double strand break created by both TALENs and CRISPRs often leaves behind a small insertion or deletion, which can disrupt the function of the target sequence[35-37]. In sticklebacks, TALENs have been used to disrupt gene expression by targeting an enhancer[20], and both TALENs and CRISPRs have successfully produced mutations in coding sequences (unpublished data). A detailed protocol for the generation of CRISPRs for use in zebrafish can be used as a guideline to develop CRISPRs for sticklebacks[38].

Transgenic and genome editing experiments require introduction of nucleic acids into a newly fertilized one-cell embryo. By introducing the transgene or genome-editing tool early in development, the number of genetically manipulated daughter cells in the embryo is maximized. Injected embryos are then visually screened for fluorescence or molecularly screened for genome modifications. If cells contributing to the germline are successfully targeted, the transgene or mutation can be passed on to a subset of offspring, even when post-injection lethality is high. The mosaic fish can be outcrossed or intercrossed and their offspring screened to recover the mutant alleles or a stably integrated transgene of interest. This protocol describes methods for introducing transgenes and genome editing reagents into one-cell stickleback embryos and monitoring for successful genomic modifications.

Protocol

All fish work was approved by the Institutional Animal Care and Use Committee of the University of California-Berkeley (protocol number R330).

1. Prepare nucleic acids for injection

1.1) Tol2 plasmid transgenesis (adapted from Fisher[26]).

1.1.1) Cut 10 µg transposase plasmid (pCS-Tp)[39] with 10 U NotI in supplied buffer for 1 hr at 37 °C to linearize. Note: Material Transfer Agreements may be required to obtain Tol2 plasmids.

1.1.2) Extract the cut plasmid with a 25:24:1 mixture of phenol:chloroform:isoamyl alcohol and ethanol precipitate with sodium acetate according to standard protocols[40]. Resuspend plasmid in 50 µL RNase-free water. (Note: phenol-chloroform should be used in a hood and the waste must be properly disposed according to institutional guidelines.)

1.1.3) Set up Sp6 transcription reaction according to manufacturer's instructions.

1.1.4) Use RNA isolation kit to clean up transcription reaction according to manufacturer's instructions; resuspend RNA in 50 µL RNase-free water.

1.1.5) Remove a 1 µL aliquot of RNA. Heat to 65 °C for 5 minutes to denature secondary structures then immediately chill on ice. Freeze remaining transcription reaction at -80 °C.

1.1.6) Run the RNA aliquot on a 1% agarose gel in 0.5X TAE (Tris base, acetic acid, Ethylenediaminetetraacetic acid) running buffer with an RNA size standard in one lane. The expected product is 2200 bp; discard if >5% of the total RNA appears in a smear smaller than 2200 bp, which indicates extensive degradation (Figure A.1).

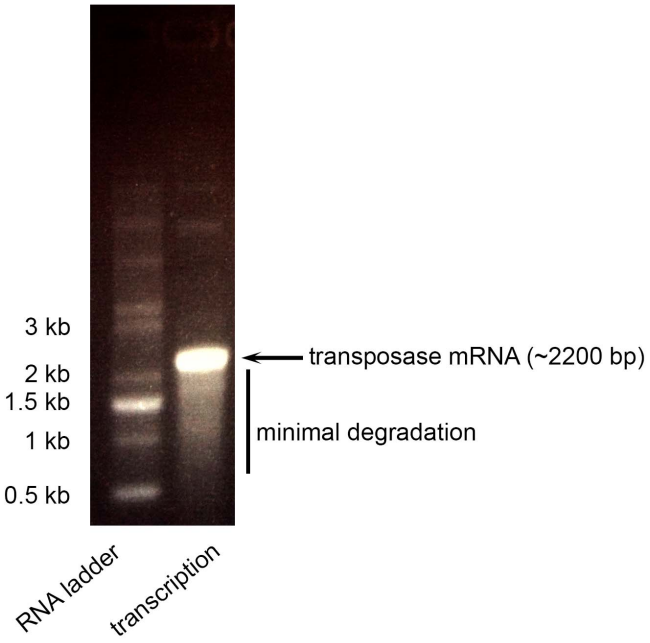


Figure A.1: Transposase mRNA gel.
Purified transcription reaction product (1 µL) was heated to 65 °C, chilled on ice, and run on a 1% agarose gel with 0.5X TAE running buffer at 100V. The sizes of the RNA ladder in kilobases (kb) are indicated to the left. The full length transposase mRNA is a bright band at ~2.2 kb. A small but acceptable amount of degraded or incomplete mRNA is seen below 2.2 kb.

1.1.7) Quantify RNA using a spectrophotometer at 260 nm. Dilute to 350 ng/µL in RNase-free water and store 1 µL aliquots at -80 °C (good for at least two years).

1.1.8) Clone Tol2 reporter plasmid (for example, using pT2HE[8] or plasmids from the Tol2 kit[41]) with *cis*-regulatory element of interest. Briefly, PCR amplify a genomic DNA sequence of interest with primers containing restriction sites found in the plasmid, digest the PCR product and vector with the enzyme(s), ligate the insert into the vector, and

transform resulting plasmid into competent *E. coli* [40]. Isolate plasmid with a kit that includes an endotoxin rinse according to manufacturer's protocol.

1.1.9) Perform a second purification of the Tol2 plasmid with a PCR purification kit according to manufacturer's protocol. Elute in 30 µL RNase-free water.

Note: the yield from this step may be low (sometimes under 50% of the input plasmid).

1.1.10) Dilute plasmid to ~125 ng/µL in RNase-free water.

1.2) BAC transgenesis (see Suster[32,33] for BAC recombineering techniques)

1.2.1) Prepare BAC from *E. coli* using BAC purification kit according to manufacturer's protocol. Use ethanol precipitation to recover DNA with a standard sodium acetate-ethanol extraction[40] and resuspend DNA at ~250 ng/µL in RNase-free water.

1.2.2) Prepare transposase mRNA as in section 1.1.

1.3) Mutation induction with TALENs (see Cermak[35] for design of TALENs)

1.3.1) Use the web-based application to design TALENs for the gene of interest[42]. If possible, design TALENs to disrupt a restriction enzyme cut site to facilitate molecular analysis.

1.3.2) Clone TALENs and prepare plasmids for transcription following published protocol[35].

1.3.3) Transcribe TALEN mRNA with a Sp6 transcription reaction according to manufacturer's instructions and clean up mRNA as described for transposase in section 1.1.4. Quantify with a spectrophotometer and dilute to 200 ng/µL in RNase free water. Run TALEN mRNA on a gel to ensure it is the proper size and not degraded as described in step 1.1.6.

1.3.4) Design a pair of PCR primers to amplify approximately 100-200 bp surrounding the TALEN target sequence using a primer design tool and the target DNA sequence[43]. Order the appropriate restriction enzyme to test for lesions at the target site based on step 1.3.1.

1.4) CRISPR transgenesis (see Talbot and Amacher[38] for design and preparation of CRISPRs):

1.4.1) Design and prepare CRISPRs and Cas9 mRNA according to protocol[38], and order appropriate verification primers and restriction enzymes as described in step 1.3.4.

2. Prepare injection reagents

2.1) Use borosilicate capillaries to prepare needles as described below. Note: these capillaries are not the standard capillaries used for zebrafish microinjection, and are made of a thicker and stronger glass that is critical to puncture the tough stickleback chorion.

2.1.1) Always wear nitrile or latex gloves when pulling needles, and do not allow needles to contact skin or skin oils.

2.1.2) Determine micropipette pulling parameters empirically by ramp tests following the micropipette puller's manufacturer's instructions. For example, with a box filament, the following parameters were determined to be optimal: (Heat=515, Pull=60, Velocity=60, Delay=85, Pressure=500). These settings produce a needle that tapers steeply at approximately 12° for ~2mm and then a long extension that tapers at approximately 2° for ~6mm (Figure A.2).

Note: The proper parameters will vary by puller and filament, and blindly using a program without determining the parameters first through ramp tests can permanently destroy the puller's filament, which is difficult and expensive to replace.

2.1.3) Follow manufacturer's instructions to pull at least 4 micropipette needles from borosilicate glass with the settings determined in 2.1.2.

2.1.4) Store needles vertically in capillary storage jar with the sharp end facing down.

2.1.5) Before injecting, place capillary storage jar on ice to chill needles. Add a piece of moist paper towel to the jar to prevent evaporation once the needles are filled.

2.2) Fertilize eggs (all steps performed at room temperature)

2.2.1) Strip egg clutch from gravid female stickleback by gently squeezing the abdomen and stroking in an anterior to posterior direction to push the eggs out through the cloaca and into a 35 x 10 mm Petri dish. Add a moist piece of paper towel on one side of the Petri dish (not touching the eggs) to create humidity chamber. Place lid on Petri dish so eggs stay moist.

2.2.2) Euthanize male stickleback in 0.025% Tricaine-S buffered with 0.1% sodium bicarbonate.

2.2.3) Cut open the abdomen, remove testes and macerate in 250 µL Hank's solution (see Westerfield[44] for full protocol for Hank's solution preparation).

2.2.4) Fertilize at most 100 eggs with 50 µL sperm solution and gently stir with pipette tip to ensure all eggs are fertilized. If the clutch is >100 eggs, fertilize half of the eggs later to

ensure that all embryos are at a one-cell stage at the time of injection. Eggs can be left unfertilized at room temperature for up to an hour, and sperm generally lasts for 1-7 days at 4 °C in Hank's solution.

2.2.5) Keep embryos covered with Petri dish lid after fertilization to prevent drying. Allow 20-25 minutes for the first cell to emerge and swell up (prepare injection materials during this time).

2.2.6) Fill a 150 mm x 15 mm Petri dish with stickleback water. (To make stickleback water, first prepare 10% sodium bicarbonate dissolved in deionized water. Then add 3.5 g artificial seawater mix and 0.217 mL of 10% sodium bicarbonate per 1L of deionized water, and stir/shake vigorously to dissolve salt.)

2.3 Prepare injection solution (while eggs are fertilizing)

2.3.1) Prepare injection solution according to Table A.1 and store on ice. Note: the concentrations of some nucleic acids have been increased from those published for zebrafish due to the increased volume of the stickleback blastomere.

Reagent	Tol2 injection	BAC injection	TALEN injection	CRISPR injection
Tol2 mRNA	350 ng	350 ng	-	-
DNA	150-200 ng plasmid	200-300 ng BAC	-	-
TALEN mRNA	-	-	200 ng each	-
CRISPR guide RNA	-	-	-	200 ng
Cas9 mRNA	-	-	-	400 ng
0.5% phenol red in Dulbecco's PBS	0.5 uL	0.5 uL	0.5 uL	0.5 uL
RNAse free water	to 5 uL	to 5 uL	to 5 uL	to 5 uL

Table A.1: Injection reagents.
All mixtures should be prepared to a total volume on 5 µL and stored on ice.

2.4) Fill needles (on ice; allow at least 10 minutes for needles to fill).

2.4.1) Backfill at least three needles by pipetting 0.5 µL injection solution onto the blunt top end of the needle while needles are hanging vertically in capillary storage jar. Be careful that the drop stays on top and does not drip down the side and avoid bubbles.

2.4.2) After the red liquid has mostly drained to the pointed tip of the needle, add another 0.5 µL to the blunt end of the needle and allow it to drain.

3. Prepare inject rig and needle for microinjection

Note: These steps can usually be done after fertilizing the eggs.

3.1) Turn on transillumination light for the dissecting microscope and place a ~13 cm x 13 cm glass plate on the microscope light base with 15 cm plaster saw blade on top of the glass plate[7]. Orient the saw perpendicular to the injection apparatus with the indentations facing towards the needle holder.

3.2) Turn on air supply and ensure pressure is set to ~200 kPa from the regulator.

3.3) Turn on the control box and adjust settings. Set pressure to ~150-175 kPa. Set injection duration to 180 ms.

3.4) Loosen the needle holder, insert a filled needle into the holder until resistance of the rubber holder can be felt, and tighten until finger tight.

3.5) Adjust the needle angle to approximately 45°.

3.6) Use micromanipulator controls to adjust the needle so the end is centered in the field of view. Zoom in to ~40x magnification and focus on the tip of the needle, which should not be touching the glass below.

3.7) Gently break the tip of the needle by grasping it with watchmaker's forceps. Ideally, do not break perpendicularly, but rather at a ~60° angle. Break close to the tip (not more than 2-3 forceps widths away from the end—Figure A.2).

3.8) Press the injection foot pedal several times to test whether the needle is broken. After a few taps, tiny red droplets should begin to come out of the end. If not, try breaking the needle slightly higher.

3.9) If the needle has an air bubble, increase the back pressure unit and press the pedal several times quickly to work the bubble out.

3.10) Adjust the back pressure:

3.10.1) Use the disposable transfer pipette to place a few drops of stickleback water on the glass plate.

3.10.2) Gently lower the needle into the water.

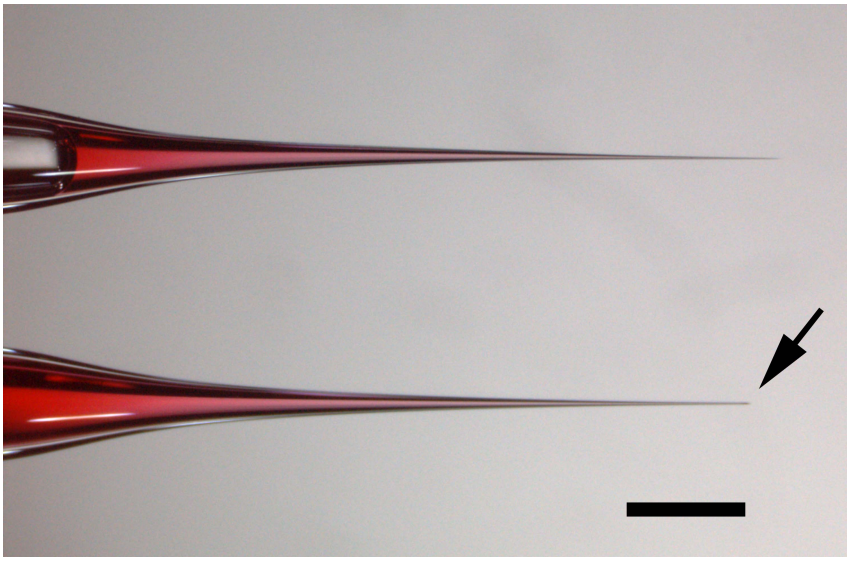


Figure A.2: Unbroken and broken microinjection needles.
The top needle is unbroken and the tip of the bottom needle has been broken with forceps (arrow). Needles are filled with a solution containing 0.05% phenol red. Scale bar = 1 mm.

3.10.3) Increase the back pressure until a faint stream of pink liquid emerges from the needle (indicating positive pressure).

Note: If there is not enough back pressure, the needle will draw up the cytoplasm by capillary action. If there is too much back pressure, the injection volume may be inadvertently too large.

3.10.4) Retract the needle as far as possible so that it will not be damaged while preparing the embryos (see below).

3.11) Alternatively, adjust the back pressure to a higher pressure setting so that a constant strong stream of liquid exits the needle when it is submerged in water. Then, pressing the foot pedal to inject becomes unnecessary; however, the injection must be performed quickly to avoid over-injecting.

Note: Do not attempt this technique when first learning to inject.

4. Microinjection

4.1) About 25 minutes after fertilization, use two 10 µL pipet tips to remove 5-10 embryos from the clutch and transfer to the glass plate.

4.2) Still using the pipet tips, gently separate the embryos into individual indentations of the saw blade. Use caution not to puncture embryos.

4.3) Using a transfer pipet with end cut off so stickleback embryos will fit inside, add enough stickleback water to cover the embryos, leave the water on for 3-5 seconds, then remove the excess water with the pipet, leaving a small volume of water coating each embryo.

Note: Too much water will cause the chorions to harden and break the needle, but a small volume of water is necessary to lift the chorion away from the cell and yolk (Figure A.3A-B).

4.4) Starting with the embryo furthest away, slide the glass plate and zoom in so that the first embryo fills approximately 25% of the field of vision.

4.5) Lower the needle into the field of vision, then use the 10 µL pipet tip to gently rotate the embryo to identify the blastomere, a grainy, slightly yellow raised bump of cytoplasm on top of the yolk (Figure A.3B), and then rotate so that the blastomere is directly perpendicular to the end of the needle (while keeping the embryo in the indentation of the saw blade, Figure A.3C).

Note: the yolk droplets may move as the embryo is rotated, so do not use them as a frame of reference for the location of the blastomere.

4.6) Lower the needle into the cytoplasm but do not push through to the underlying yolk. Apply pressure slowly and evenly when piercing the chorion to avoid breaking the needle. If the needle bends severely, retract and try again in a slightly different location.

4.7) Depress the foot pedal 3-4 times to inject so that a red bolus with slightly diffuse edges fills about ~1/8 the diameter of the cytoplasm (see Figure A.3D).

4.7.1) Avoid obtaining a red bolus with sharp edges that do not begin to diffuse, which indicates injection into the yolk beneath the cytoplasm. (Figure A.3E).

4.7.2) If a bright pinkish-red spot diffuses quickly, insert the needle further to puncture the blastomere.

4.7.3) If the injection bolus turns yellow instantly, rotate the embryo to ensure the blastomere has been targeted and inject again (Figure A.3F).

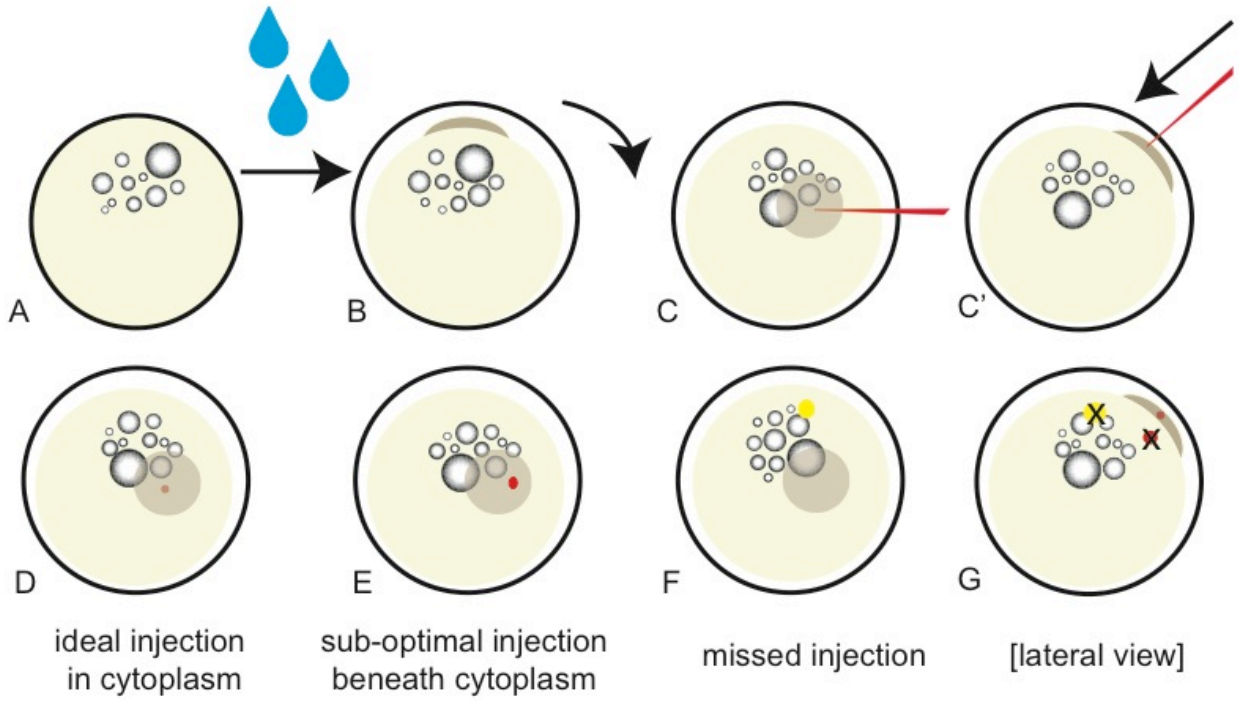


Figure A.3: Appearance of stickleback embryos before and after injection.
All embryos are drawn from the perspective of looking down through the microscope except C' and G. (A) Before adding water, fertilized embryos have a uniform appearance with oil droplets floating near the top of the yolk. (B) After adding water, the chorion swells, revealing a blastomere that protrudes from the yolk and is visible in profile. (C) Rotation of the embryo so that the needle enters perpendicularly to the chorion and blastomere. (C') Lateral view of a needle that has punctured the chorion with the tip in the cytoplasm. (D) Injection into the cytoplasm results in a red spot with diffuse edges that fade over time. (E) Injection into the yolk underlying the cytoplasm results in a red spot with defined edges. (F) Injection into the yolk opposite the cytoplasm results in a pH-induced color shift from red to yellow. (G) Lateral view of injection outcomes, with Xs over sub-ideal injection locations.

4.8) Use the micromanipulator controls to retract the needle, using the 10 µL pipet tip to hold down the embryo if it sticks to the needle.

4.9) After retracting the needle, slide the glass plate to align the next embryo with the needle.

4.10) After injecting all embryos, use the transfer pipet to add water to the embryos, then collect them in the transfer pipette and place in 150 mm Petri dish full of stickleback water.

4.11) Dry off the glass plate with a paper towel to avoid accumulating too much water.

4.12) Distribute fresh embryos onto the saw and repeat steps 4.4 through 4.10.

4.13) Keep ~10% of embryos as uninjected controls to ensure that uninjected embryos develop normally and to use as wild-type controls for the molecular assays described below.

5. Post injection care

5.1) Incubate embryos in Petri dishes at 18 °C after injection. Following hatching, rear larvae in aquaria as described[10].

5.2) Gently pour off the stickleback water one day after injection and replace with fresh stickleback water. Replace with fresh stickleback water at least every other day after that.

5.3) Check for dead or malformed embryos daily. Remove such embryos to prevent decay from contaminating the water.

6. Analysis of injection results

6.1) For injection of fluorescent reporters, monitor embryos daily in a darkened room using a fluorescent dissecting microscope with a GFP or RFP filter (depending on fluorescent transgene). Record anatomical patterns of gene expression with digital photographs and/or written descriptions and tabulation of the number of fish with different expression domains (Figures A.4 and A.5).

Note: sticklebacks have autofluorescent, stellate pigment cells that are visible under GFP filters beginning at 4 days post fertilization (dpf).

6.1.1) To generate stable lines, save embryos with detectable fluorescence and grow to adulthood as described[10].

6.1.2) Outcross injected adult fish to wild-type fish using the *in vitro* fertilization procedure described in section 2.2 and screen offspring for fluorescence as described in step 6.1 to look for fluorescent offspring, indicating successful transgene transmission.

6.1.3) To visualize fluorescence in hatched, free-swimming larvae, add 500 µL 0.8% Tricaine to the 150 mm Petri dish to anaesthetize fish and wait until fish stop moving to image. Immediately rinse several times with fresh stickleback water following observation and imaging.

6.1.4) Optionally, to preserve fluorescence for further imaging, euthanize larvae in 0.025% (250 mg/L) Tricaine buffered with 0.1% sodium bicarbonate and fix for 4 hours in 4% paraformaldehyde in 1X Phosphate Buffered Saline (PBS) at 4 °C. Store in 1X PBS for up to two weeks. Note: background auto-fluorescence increases over time.

6.2) Diagnostic PCR/digestion genotyping for mutation induction with CRISPRs or TALENs --best performed at 2 days post fertilization.

6.2.1) Use a transfer pipet to place 10 injected embryos (2 dpf, still in chorions) into the first 10 wells of a 12-well PCR strip tube. Place uninjected control fish in the last two wells.

6.2.2) Remove excess stickleback water.

6.2.3) Add 50 µL lysis buffer (10 mM Tris pH 8.3, 50 mM KCl, 1.5 mM $MgCl_2$, 0.3% Tween-20, and 0.3% NP-40) to each well.

6.2.4) Place caps on tubes and incubate at 95 °C for 20 minutes in a thermocycler. Note: The yolk will turn white and rubbery after the boiling step.

6.2.5) Remove lids and use a different clean 1000 µL pipet tip to macerate the embryo in each tube.

Note: White debris will collect at the bottom of the tube.

6.2.6) Add 2.5 µL of 10 mg/mL proteinase K to each well.

6.2.7) Replace lids and incubate at 55 °C for 1 hr to digest protein, followed by 95 °C for 20 minutes to inactivate proteinase K. Store at -20 °C to avoid mold growth.

6.2.8) Perform PCR using a high-fidelity polymerase following manufacturer's instructions. Use the embryo lysate as DNA template.

Note: Be careful to remove liquid from top of tube for DNA template, avoiding any visible chorion or yolk debris at the bottom of the tube.

6.2.9) Mix the PCR product in equal volume with a restriction enzyme master mix containing 1X enzyme-specific buffer and 0.25 µL enzyme per sample. Always save half of the uncut PCR product to assay on a gel to confirm PCR products are the predicted size. Incubate the PCR mixed with enzyme at the appropriate conditions for the restriction enzyme.

Note: Some enzymes may require other ratios of enzyme buffer to PCR product; adjust the buffer concentration if the uninjected controls do not show complete digestion.

6.2.10) Following digestion, run products on a 1% agarose gel next to a DNA size ladder to confirm expected product sizes.

Note: Uncut bands in injected embryos indicate the presence of molecular lesions (Figure A.6) and uninjected controls must be fully digested to interpret results.

6.2.11) To confirm lesions, cut out uncut bands from agarose gels and purify DNA with a gel extraction kit. Use Sanger sequencing, ideally using a sequencing primer internal to the PCR primers, to confirm lesions. In F_0 injected embryos, a mix of lesions will likely be present, causing the quality of the Sanger read to degrade near the target site.

6.2.12) To produce a stable line, save all injected embryos from clutches that screen positive for molecular lesions. Grow up fish, outcross, then screen a subset of F_1 embryos for molecular lesions as described in steps 6.2.1 through 6.2.11. If heterozygous carriers are identified, grow the remaining F_1 embryos to adulthood and identify heterozygous individuals using DNA extracted from a caudal fin clip.

Representative Results

For reporter gene transgenes that have enhancer activity, successful injection will result in specific, cellular expression of the transgene (Figure A.4A, 4C). Injected fish can then be outcrossed to produce stable lines (example of a BAC stable line shown in Figure A.4B). Injecting DNA into stickleback embryos typically results in far higher lethality than RNA alone. It is typical to see up to 50% (sometimes even more) lethality or malformation (see Figure A.4D-F, I) after injecting Tol2 reporter constructs (similar to the previously described meganuclease method[21]). However, the results vary widely based on the specific construct, the embryo morphology, and skill level. For an active enhancer, 40-50% of embryos generally will show tissue-specific transgene expression, for example in the median and pectoral fins (Figure A.5). The degree of background and nonspecific fluorescence (Figure A.4G-I) varies widely based on the promoter used; the zebrafish *hsp70l* promoter tends to be leaky, especially in muscle and neural tissue, and BACs tend to have high background expression in the yolk (similar to Figure A.4G). The carp *beta actin* [41] promoter is less leaky but also drives considerably fainter GFP expression. Transmission of Tol2 plasmid transgenes can be high, with up to 100% of GFP+ F_0 fish producing transgenic offspring (Table A.2). However, the percent of offspring carrying the transgene varies widely, from <1% to 72% (Table A.2). Saving only GFP+ injected embryos generally increases transmission efficiency. BACs tend to have far lower transgenesis rates, with only up to 10% of F_0 injected stickleback embryos showing fluorescence in expected tissues. The transmission rate of BACs is lower than that of plasmid constructs, with only up to 14% of screened stickleback transmitting the BAC (Table A.2), which is similar to the reported transmission rate of 15% in zebrafish[32].

In contrast to the relatively low efficiency of BAC transgenesis, typically 70%-100% of screened fish injected with TALENs have mosaic lesions (in n=10 injected clutches that were screened for lesions). This number could be lower with a less efficient TALEN pair, and may vary with injection quality. Figure A.6 shows a PCR/restriction digestion for 10 embryos from a single clutch injected with TALENs targeting a PvuII cut site within *Tfap2a*. An uncut amplicon in each of the injected embryos (lanes 1-10) indicates that a portion of the cells in each embryo carry lesions at the target locus, though each embryo is highly mosaic with a significant wild-type cut band. The amplicon from uninjected embryos in lanes 11-12 are fully digested with PvuII. TALEN-induced mutations are readily transmitted to the next generation. With two different TALEN sets, 50% and 90% of screened F_0s transmitted lesions to offspring, with 20-90% of offspring carrying lesions in positive clutches (Table A.3). While CRISPR/Cas9 efficiency has not been optimized in stickleback, with one CRISPR guide targeting *Pitx2*, one out of three injected embryos had lesions based on Sanger sequencing of a PCR product of the CRISPR target (the uncut band was sequenced following restriction digestion, Figure A.7). The loss of sequence quality at the predicted cut site indicates a mix of molecular lesions are present. Fin clipping adult F_0 fish and screening for lesions using a PCR and restriction digestion assay found 10/22 fish with somatic lesions in the fin. A representative subset of these animals are shown in Figure A.8; individual #3 has a high percentage of DNA with lesions, while individual #2 has a low percentage of DNA with lesions.

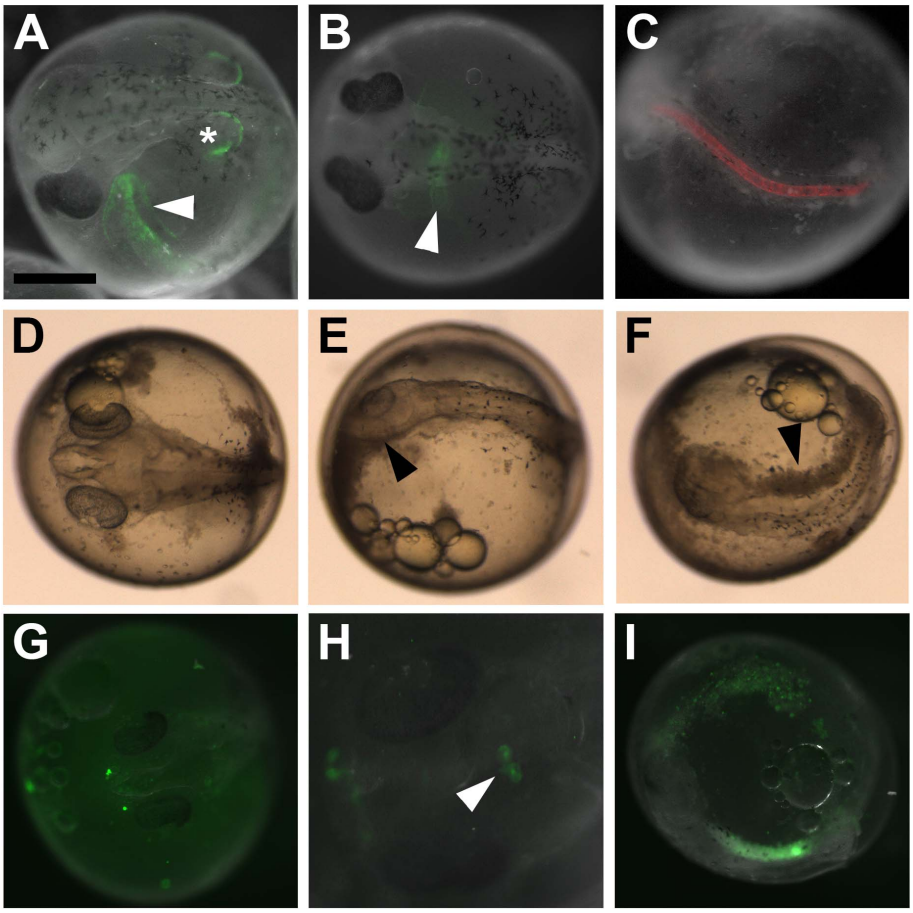


Figure A.4: Examples of injected embryos.

(A) Mosaic embryo injected with an enhancer that drives GFP expression in pectoral (asterisk) and median fins (arrowhead) at 5 days post fertilization (dpf). Scale bar = 500 μm. (B) Stable line of a reporter BAC that drives GFP expression in the embryonic heart at 4 dpf. (C) Mosaic embryo injected with a *Col2* enhancer that drives mCherry expression in the notochord at 4 dpf. The *Col2* enhancer was cloned from stickleback DNA with primers 5'-CGCTCCTTGAGGGTTTGAGCTG-3' and 5'-ATACTGTGCTCATTTCGGCCGT-3' which amplify the conserved orthologous enhancer reported in Dale and Topczewski 2011[45]. (D) Example of a normally developing injected embryo. (E-F) Examples of malformed embryos with injection trauma; E is lacking the left eye and F has necrotic tissue along the right side (arrowheads). (G) Example of diffuse GFP expression in yolk, likely the result of injection into the yolk rather than the blastomere. (H) Example of non-specific GFP expression in epidermis (arrowhead). (I) Bright, non-specific, granular GFP expression.

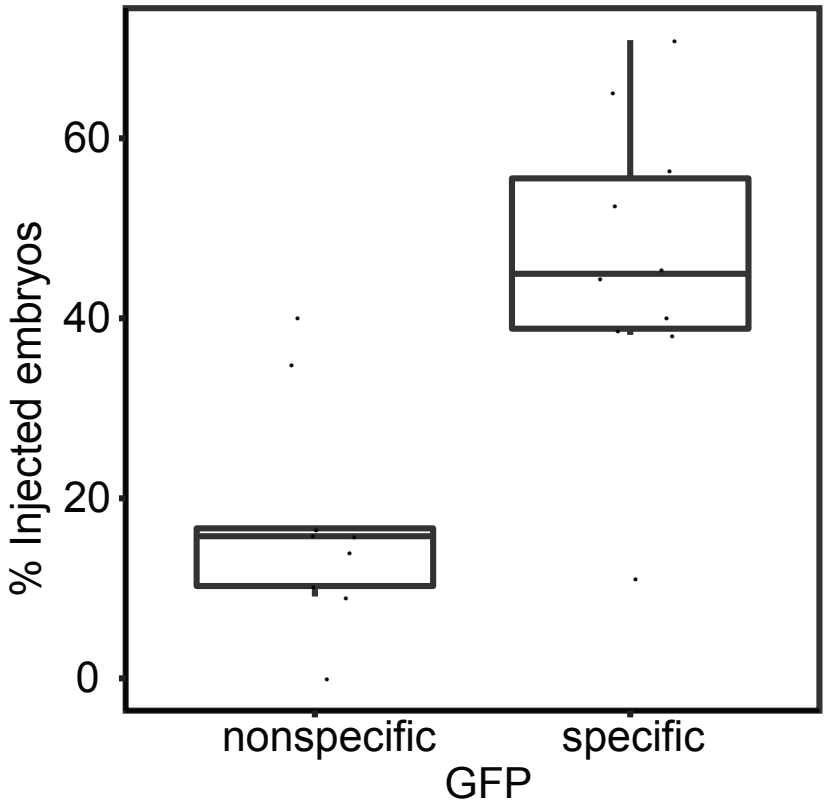


Figure A.5: Efficiency of reporter construct injection.

Ten clutches were injected with a 190 bp stickleback *Bmp6* enhancer that drives pectoral fin and median fin expression at 5 dpf[20]. From each clutch, at least 20 embryos were scored for having tissue-specific (pectoral and/or median fin) and/or nonspecific (all other tissues) GFP expression. The percentage of all surviving injected embryos having non-specific and tissue-specific expression is shown as a boxplot. Horizontal lines indicate the first quartile, median and third quartile; whiskers extend to datum within 1.5 IQR (interquartile range) of the first and third quartile. Data are adapted from Erickson et al. 2015.

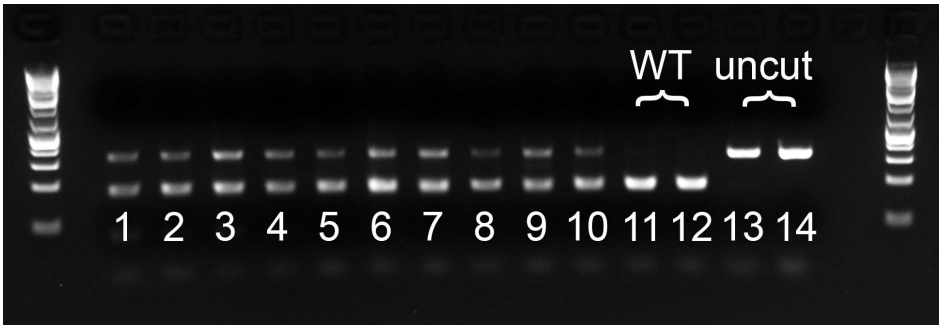


Figure A.6: PCR and restriction digestion to screen for TALEN-induced lesions.
A TALEN pair targeting *Tfap2a* was generated and injected as described. DNA was prepared as described above from 2 dpf injected embryos and a 297 bp fragment surrounding the TALEN target sequence was PCR amplified by a high fidelity DNA polymerase using primers 5'-GGGTCGTTGACGTGCGAGTAA-3' and 5'-AGCGGGACAACGTCATCACTTA-3'. Lanes 1-10 are injected, lanes 11-12 are uninjected, digested controls, and lanes 13-14 are uninjected, undigested controls. PvuII cuts the wild-type sequence into two approximately equal size bands. Uncut bands indicate presence of molecular lesions in the target sequence. All injected embryos in lanes 1-10 show signs of molecular lesions (undigested bands), however all of the embryos either have monoallelic mutations and/or are mosaic as they also have cut (wild-type) bands.

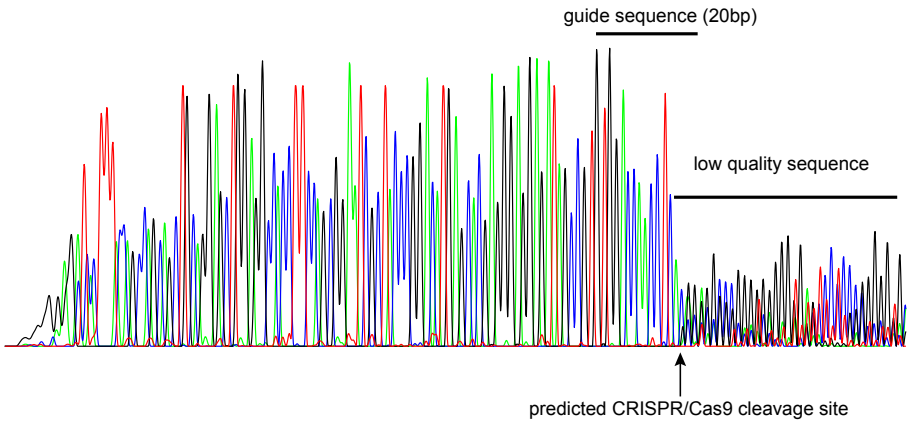


Figure A.7: Sanger sequencing from mosaic F0 CRISPR/Cas9 injected embryo.
A CRISPR guide RNA (5'-GTGGACCAACCTCACGG-3') against *Pitx2* (shown at top) was co-injected with Cas9 mRNA (transcribed from pCS2-nCas9n plasmid as described[38]) and embryos were screened for lesions using a restriction enzyme assay. The uncut band was gel extracted and sequenced by Sanger sequencing. The sequence quality degrades at the predicted cleavage site (arrow below) due to the mosaic mix of lesions present in the injected embryo.

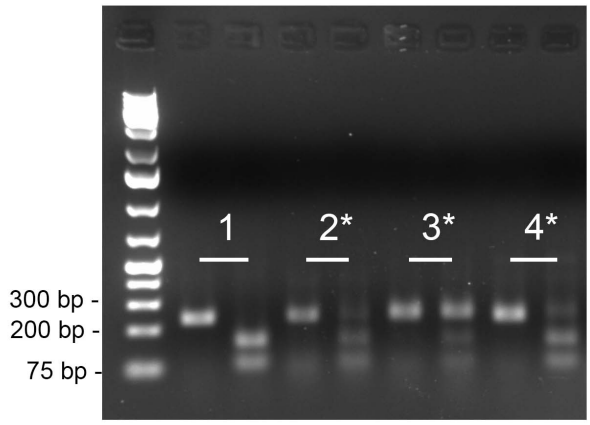


Figure A.8: Analysis of CRISPR Fo caudal fin clips.

DNA was prepared from fin clips from F_0 juveniles raised from embryos injected with CRISPRs against *Pitx2*. The CRISPR/Cas9 target was amplified with primers 5'-CTCGGATGACCCTTCAAAAA-3' and 5'-GGCCCAAATTACCCACATTT-3', and the product was digested with EcoRI. Four individuals are shown; an uncut PCR product is on the left and digested PCR product is on the right for each numbered individual. Product the size of the uncut band (~230bp) in the digested lane indicates the presence of a lesion. Individuals 2, 3 and 4 all show signs of a molecular lesion, indicated with asterisks, with varying mosaicism between individuals. Relevant ladder sizes are indicated on left.

Construct	GFP offspring/total Fo screened fish (%)	% F1 offspring positive	note
BAC A	6/46 (13%)	4%-19%	
BAC B	1/41 (2%)	4%	
BAC C	5/42 (12%)	3%-40%	
BAC D	3/22 (14%)	5%-15%	
plasmid A	2/38 (5%)	2%-5%	all Fo embryos screened, not just GFP+
plasmid B	3/16 (19%)	<1%-8%	all Fo embryos screened, not just GFP+
plasmid C	1/11 (9%)	10%	
plasmid D	2/11 (18%)	1%-45%	plasmid D injected into 2 genetic backgrounds
plasmid D	5/5 (100%)	16%-72%	
plasmid E	2/3 (67%)	2%-22%	
plasmid F	2/6 (33%)	<1%-65%	
plasmid G	3/8 (38%)	2%-56%	
plasmid H	3/18 (17%)	not scored	
plasmid I	5/24 (21%)	not scored	

Table A.2: Transgene transmission efficiencies for BACs and enhancer constructs.

F_0 injected embryos were raised to adulthood and then outcrossed to wild-type fish and the F_1 offspring scored for GFP fluorescence. The number of F_0 individuals that transmitted the transgene is expressed as a percentage of all screened F_0 fish. The range of percentages of F_1 offspring carrying the transgene is also shown for those clutches that had GFP positive fish.

TALEN	% Fo transmitting lesions	% F1 offspring positive
A	9/10 (90%)	20%-90%
B	4/8 (50%)	20%-72%

Table A.3: Transmission efficiencies for two TALEN pairs.

F_0 injected embryos were raised to adulthood and then outcrossed to wild-type fish and the F_1 offspring screened for TALEN lesions. The percentage of injected individuals transmitting lesions is shown, as well as the range of percentages of offspring with lesions in those clutches that transmitted lesions. TALEN A targeted a *Bmp6* enhancer[20], TALEN B targeted *Tfap2a* (unpublished).

Discussion

Injecting one-cell stickleback embryos for transgenesis or genome editing presents three main challenges. First, relative to zebrafish embryos, the stickleback embryonic chorion is tough and will often break needles. This problem can be partially overcome by using thicker and stronger glass micropipettes and injecting perpendicular to the chorion (see Protocol, Figure A.2). Ensuring that as little water as possible is added to the embryos (just enough to cause the chorion to swell and lift away from the cell) helps to reduce chorion hardness. The chorion hardens over time, so working quickly after moistening the embryos is important. Keeping the embryos in a humidity chamber prior to injection so that they do not dry out is also critical. Some clutches and even individual embryos simply have much thicker and tougher chorions; sometimes moving on to the next embryo or trying with a new clutch is the easiest solution. It is much easier to skip one difficult embryo than to replace a damaged needle. Having backup needles ready will allow injections to continue in the case of needle breakage. When injecting BACs, it is common for the needle to clog. The needle can be unclogged by gently scraping or re-breaking the tip with forceps, or by using the constant air pressure switch to purge the clog.

Second, identifying the first cell in the embryo is challenging; it is often quite flattened and difficult for beginners to see, and is especially invisible when looking directly at it. Often the blastomere can only be seen as a slightly raised bump in profile. Therefore, it is best to identify the cell in profile (Figure A.3B) and then gently rotate the embryo forward and to the side so the cell will face the end of the angled needle (though the cell will often be invisible from this angle; the darker color of the blastomere in Figure A.3 is exaggerated for clarity).

Third, targeting the cytoplasm is also difficult, especially if the first cell is especially flat. Aiming for the fattest part of the blastomere (usually the center) improves the chance of injecting into cytoplasm. While injecting into the yolk near the cytoplasm can successfully produce transgenic fish, the efficiency seems to be increased and lethality decreased when the cytoplasm is targeted with a single needle puncture. Sometimes, individual clutches will have particularly thin blastomeres, such that avoiding the yolk is nearly impossible. Waiting longer than 25 minutes to begin injections may help (some clutches do not form a full size blastomere until ~45 minutes post fertilization), but if the blastomeres never increase in size, it is often easier to obtain a new clutch of eggs than to try to inject flattened cells.

Excessive lethality following injections may occur for several reasons. Blunt needles and/or too large a needle bore size may cause too much damage to the cell and/or cause cytoplasm to leak out. Some DNA constructs seem to be especially lethal; lowering the concentration of DNA may improve survival but lower transgenesis rates. Cleaning up plasmids first with a midiprep kit that contains an endotoxin rinse followed by a second PCR cleanup kit reduces construct toxicity. Finally, genome editing may produce a loss of

function mutation that is lethal to developing embryos. Reducing the concentration of the CRISPR or TALEN mRNAs can increase the mosaicism of the embryo to prevent lethality, but may reduce mutant allele recovery efficiency.

A previously published protocol for generating transgenic sticklebacks using a meganuclease method[21] reported a 4-7% transgene germline transmission rate from F_0 founder fish. The Tol2 method reported here resulted in up to a 72% transgene germline transmission rate from F_0 founder fish (indicating multiple genomic integrations). The previous study using a meganuclease method reported 40% of injected embryos showing specific GFP expression, similar to that reported here. Thus while similar rates of transgenic F_0 founders are observed for transgenic fish generated by both the meganuclease and Tol2 methods, the germline transmission rate appears much higher for Tol2 mediated transgenics.

As genome editing technologies continue to advance, even more specific genetic manipulations will become possible in stickleback and other fish species. For example, directed repair[46] and homologous recombination will allow allele swaps between marine and freshwater stickleback genomes, and modified CRISPRs can be used to specifically activate or inhibit gene expression[47]. These exciting technologies for genome editing and analysis will lead to new insights about the genetic basis of many interesting morphological, physiological, and behavioral phenotypes in sticklebacks and other fish species.

Acknowledgments

Nick Ellis was involved in the development of BAC and CRISPR transgenesis techniques.

This work was funded in part by NIH R01 #DE021475 (CTM), an NIH Predoctoral Training Grant 5T32GM007127 (PAE), and an NSF Graduate Research Fellowship (NAE). We thank Kevin Schwalbach for performing BAC recombineering and injections, Nick Donde for generating CRISPR Sanger sequencing data, and Katherine Lipari for helpful feedback on the injection protocol.

References

1. Bell, M. A. & Foster, S. A. *The Evolutionary Biology of the Threespine Stickleback*. Oxford University Press. (1994).
2. Jones, F. C., *et al.* The genomic basis of adaptive evolution in threespine sticklebacks. *Nature* **484** (7392), 55–61, doi:10.1038/nature10944 (2012).
3. Glazer, A. M., Killingbeck, E. E., Mitros, T., Rokhsar, D. S. & Miller, C. T. Genome Assembly Improvement and Mapping Convergently Evolved Skeletal Traits in

Sticklebacks with Genotyping-by-Sequencing. *G3 GenesGenomesGenetics* , g3.115.017905, doi:10.1534/g3.115.017905 (2015).

4. Miller, C. T., *et al.* Modular skeletal evolution in sticklebacks is controlled by additive and clustered quantitative trait Loci. *Genetics* **197** (1), 405–420, doi:10.1534/genetics.114.162420 (2014).
5. Shapiro, M. D., *et al.* Genetic and developmental basis of evolutionary pelvic reduction in threespine sticklebacks. *Nature* **428** (6984), 717–723, doi:10.1038/nature02415 (2004).
6. Colosimo, P. F., *et al.* Widespread Parallel Evolution in Sticklebacks by Repeated Fixation of Ectodysplasin Alleles. *Science* **307** (5717), 1928–1933, doi:10.1126/science.1107239 (2005).
7. Chan, Y. F., *et al.* Adaptive evolution of pelvic reduction in sticklebacks by recurrent deletion of a Pitx1 enhancer. *Science* **327** (5963), 302–305, doi:10.1126/science.1182213 (2010).
8. O'Brown, N. M., Summers, B. R., Jones, F. C., Brady, S. D. & Kingsley, D. M. A recurrent regulatory change underlying altered expression and Wnt response of the stickleback armor plates gene EDA. *eLife* , e05290, doi:10.7554/eLife.05290 (2015).
9. Cleves, P. A., *et al.* Evolved tooth gain in sticklebacks is associated with a cis-regulatory allele of Bmp6. *Proc. Natl. Acad. Sci.* **111** (38), 13912–13917, doi:10.1073/pnas.1407567111 (2014).
10. Erickson, P. A., Glazer, A. M., Cleves, P. A., Smith, A. S. & Miller, C. T. Two developmentally temporal quantitative trait loci underlie convergent evolution of increased branchial bone length in sticklebacks. *Proc. R. Soc. Lond. B Biol. Sci.* **281** (1788), 20140822, doi:10.1098/rspb.2014.0822 (2014).
11. Glazer, A. M., Cleves, P. A., Erickson, P. A., Lam, A. Y. & Miller, C. T. Parallel developmental genetic features underlie stickleback gill raker evolution. *EvoDevo* **5** (1), 19, doi:10.1186/2041-9139-5-19 (2014).
12. Ellis, N. A., *et al.* Distinct developmental and genetic mechanisms underlie convergently evolved tooth gain in sticklebacks. *Development* , dev.124248, doi:10.1242/dev.124248 (2015).
13. Hohenlohe, P. A., Bassham, S., Currey, M. & Cresko, W. A. Extensive linkage disequilibrium and parallel adaptive divergence across threespine stickleback genomes. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* **367** (1587), 395–408, doi:10.1098/rstb.2011.0245 (2012).
14. Hohenlohe, P. A., *et al.* Population Genomics of Parallel Adaptation in Threespine Stickleback using Sequenced RAD Tags. *PLoS Genet* **6** (2), e1000862, doi:10.1371/journal.pgen.1000862 (2010).
15. McKinnon, J. S. & Rundle, H. D. Speciation in nature: the threespine stickleback model systems. *Trends Ecol. Evol.* **17** (10), 480–488, doi:10.1016/S0169-5347(02)02579-X (2002).
16. Shao, Y. T., *et al.* Androgen feedback effects on LH and FSH, and photoperiodic control of reproduction in male three-spined sticklebacks, Gasterosteus aculeatus. *Gen. Comp. Endocrinol.* **182**, 16–23, doi:10.1016/j.ygcen.2012.10.017 (2013).

17. Katsiadaki, I., *et al.* Three-spined stickleback: an emerging model in environmental endocrine disruption. *Environ. Sci. Int. J. Environ. Physiol. Toxicol.* **14** (5), 263–283 (2007).
18. Lenz, T. L., Eizaguirre, C., Rotter, B., Kalbe, M. & Milinski, M. Exploring local immunological adaptation of two stickleback ecotypes by experimental infection and transcriptome-wide digital gene expression analysis. *Mol. Ecol.* **22** (3), 774–786, doi:10.1111/j.1365-294X.2012.05756.x (2013).
19. Barber, I. Sticklebacks as model hosts in ecological and evolutionary parasitology. *Trends Parasitol.* **29** (11), 556–566, doi:10.1016/j.pt.2013.09.004 (2013).
20. Erickson, P. A., *et al.* A 190 base pair, TGF-β responsive tooth and fin enhancer is required for stickleback Bmp6 expression. *Dev. Biol.* **401** (2), 310–323, doi:10.1016/j.ydbio.2015.02.006 (2015).
21. Hosemann, K. E., Colosimo, P. F., Summers, B. R. & Kingsley, D. M. A Simple and Efficient Microinjection Protocol for Making Transgenic Sticklebacks. *Behaviour* **141** (11/12), 1345–1355, doi: 10.1163/1568539042948097 , (2004).
22. Thermes, V., *et al.* I-SceI meganuclease mediates highly efficient transgenesis in fish. *Mech. Dev.* **118** (1–2), 91–98, doi:10.1016/S0925-4773(02)00218-6 (2002).
23. Carroll, S. B. Evo-Devo and an Expanding Evolutionary Synthesis: A Genetic Theory of Morphological Evolution. *Cell* **134** (1), 25–36, doi:10.1016/j.cell.2008.06.030 (2008).
24. Maurano, M. T., *et al.* Systematic localization of common disease-associated variation in regulatory DNA. *Science* **337** (6099), 1190–1195, doi:10.1126/science.1222794 (2012).
25. Hindorff, L. A., *et al.* Potential etiologic and functional implications of genome-wide association loci for human diseases and traits. *Proc. Natl. Acad. Sci. U. S. A.* **106** (23), 9362–9367, doi:10.1073/pnas.0903103106 (2009).
26. Fisher, S., *et al.* Evaluating the biological relevance of putative enhancers using Tol2 transposon-mediated transgenesis in zebrafish. *Nat. Protoc.* **1** (3), 1297–1305, doi:10.1038/nprot.2006.230 (2006).
27. Kawakami, K., Shima, A. & Kawakami, N. Identification of a functional transposase of the Tol2 element, an Ac-like element from the Japanese medaka fish, and its transposition in the zebrafish germ lineage. *Proc. Natl. Acad. Sci. U. S. A.* **97** (21), 11403–11408, doi:10.1073/pnas.97.21.11403 (2000).
28. Kikuta, H. & Kawakami, K. Transient and Stable Transgenesis Using Tol2 Transposon Vectors. *Methods in Moleculario Biology: Zebrafish*, 69–84, doi: 10.1007/978-1-60327-977-2_5 (2009).
29. Halloran, M. C., *et al.* Laser-induced gene expression in specific cells of transgenic zebrafish. *Dev. Camb. Engl.* **127** (9), 1953–1960 (2000).
30. Kingsley, D. M., *et al.* New Genomic Tools for Molecular Studies of Evolutionary Change in Threespine Sticklebacks. *Behaviour* **141** (11/12), 1331–1344 , doi: 10.1163/1568539042948150 (2004).
31. Mortlock, D. P., Guenther, C. & Kingsley, D. M. A General Approach for Identifying Distant Regulatory Elements Applied to the Gdf6 Gene. *Genome Res.* **13** (9), 2069–2081, doi:10.1101/gr.1306003 (2003).

32. Suster, M. L., Abe, G., Schouw, A. & Kawakami, K. Transposon-mediated BAC transgenesis in zebrafish. *Nat. Protoc.* **6** (12), 1998–2021, doi:10.1038/nprot.2011.416 (2011).
33. Suster, M. L., Sumiyama, K. & Kawakami, K. Transposon-mediated BAC transgenesis in zebrafish and mice. *BMC Genomics* **10** (1), 477, doi:10.1186/1471-2164-10-477 (2009).
34. Kim, H. & Kim, J.-S. A guide to genome engineering with programmable nucleases. *Nat. Rev. Genet.* **15** (5), 321–334, doi:10.1038/nrg3686 (2014).
35. Cermak, T., *et al.* Efficient design and assembly of custom TALEN and other TAL effector-based constructs for DNA targeting. *Nucleic Acids Res.* , gkr218, doi:10.1093/nar/gkr218 (2011).
36. Miller, J. C., *et al.* A TALE nuclease architecture for efficient genome editing. *Nat. Biotechnol.* **29** (2), 143–148, doi:10.1038/nbt.1755 (2011).
37. Jinek, M., *et al.* Programmable Dual-RNA–Guided DNA Endonuclease in Adaptive Bacterial Immunity. *Science* **337** (6096), 816–821, doi:10.1126/science.1225829 (2012).
38. Talbot, J. C. & Amacher, S. L. A Streamlined CRISPR Pipeline to Reliably Generate Zebrafish Frameshifting Alleles. *Zebrafish* **11** (6), 583–585, doi:10.1089/zeb.2014.1047 (2014).
39. Kawakami, K., *et al.* A Transposon-Mediated Gene Trap Approach Identifies Developmentally Regulated Genes in Zebrafish. *Dev. Cell* **7** (1), 133–144, doi:10.1016/j.devcel.2004.06.005 (2004).
40. Green, M. R. & Sambrook, J. *Molecular cloning : a laboratory manual*. Cold Spring Harbor, N.Y. : Cold Spring Harbor Laboratory Press, (2012).
41. Villefranc, J. A., Amigo, J. & Lawson, N. D. Gateway compatible vectors for analysis of gene function in the zebrafish. *Dev. Dyn.* **236** (11), 3077–3087, doi:10.1002/dvdy.21354 (2007).
42. Doyle, E. L., *et al.* TAL Effector-Nucleotide Targeter (TALE-NT) 2.0: tools for TAL effector design and target prediction. *Nucleic Acids Res.* **40** (W1), W117–W122, doi:10.1093/nar/gks608 (2012).
43. Untergasser, A., *et al.* Primer3—new capabilities and interfaces. *Nucleic Acids Res.* **40** (15), e115, doi:10.1093/nar/gks596 (2012).
44. Westerfield, M. *The Zebrafish Book: A guide for the Laboratory Use of Zebrafish (Danio rerio), 5th Edition*. University of Oregon Press: Eugene, OR (2007).
45. Dale, R. M. & Topczewski, J. Identification of an evolutionarily conserved regulatory element of the zebrafish col2a1a gene. *Dev. Biol.* **357** (2), 518–531, doi:10.1016/j.ydbio.2011.06.020 (2011).
46. Ran, F. A., *et al.* Genome engineering using the CRISPR-Cas9 system. *Nat. Protoc.* **8** (11), 2281–2308, doi:10.1038/nprot.2013.143 (2013).
47. Qi, L. S., *et al.* Repurposing CRISPR as an RNA-Guided Platform for Sequence-Specific Control of Gene Expression. *Cell* **152** (5), 1173–1183, doi:10.1016/j.cell.2013.02.022 (2013).