Title
Genetic and Developmental Analysis of Evolved Stickleback Gill Raker Reduction

Author
Glazer, Andrew Michael

Publication Date
2015

Peer reviewed|Thesis/dissertation

Genetic and Developmental Analysis of Evolved Stickleback Gill Raker Reduction

by

Andrew Michael Glazer

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 Miller, Chair

Professor Daniel Rokhsar

Professor Richard Harland

Professor Marvalee Wake

Spring 2015

Genetic and Developmental Analysis of Evolved
Stickleback Gill Raker Reduction

Abstract

Genetic and Developmental Analysis of Evolved

Stickleback Gill Raker Reduction

by

Andrew Michael Glazer

Doctor of Philosophy in Molecular and Cellular Biology

University of California, Berkeley

Professor Craig Miller, Chair

Much is known about the genetic and developmental basis of mutant phenotypes in the laboratory. However, much less is known about the mechanisms that control naturally evolved phenotypes in the wild. Are specific types of genes or mutations preferred? How are developmental programs modified to generate altered phenotypes? Finally, in cases where the same phenotype has evolved in independent lineages (convergent evolution), are the same genes or mutations used?

The threespine stickleback fish (*Gasterosteus aculeatus)* is an emerging model organism in which the approaches and tools of developmental biology, genetics, and genomics can be applied to understand evolutionary changes. Marine sticklebacks have repeatedly colonized and adapted to new freshwater environments throughout the Northern Hemisphere. Freshwater sticklebacks have evolved many changes in their skeleton relative to marine fish. For example, freshwater sticklebacks have evolved changes to their defensive armor in response to decreased predation and changes in bones used for feeding in response to differences in diet. Despite their phenotypic differences, marine and freshwater sticklebacks can be bred in the lab. Crosses between marine and freshwater sticklebacks can be used to decipher the genetic basis of their evolved phenotypic differences.

In chapter 2 of this thesis, I present the results of a large quantitative genetic study of stickleback skeletal evolution. I mapped regions of the genome, known as quantitative trait loci (QTL), affecting a wide range of skeletal traits. Over 100 QTL were mapped, and a meta-analysis of their properties was performed. Using these QTL, several general questions in evolutionary genetics were addressed. First, I explored the dominance of the QTL. The dominance of a QTL is a genetic property that might affect likelihood of being used during adaptation. Surprisingly, a large proportion of skeletal QTL had an additive (intermediate) dominance. Second, I investigated the modularity of QTL. To do this, I examined serially repeating skeletal elements such as vertebrae. Most (76%) QTL that influenced such traits had specific, modular effects on just a subset of the possible

domains of the phenotype. Finally, I examined whether the QTL clustered in the genome. Many large-effect skeletal QTL were clustered on chromosomes 4, 20, and 21. Inheritance of these clusters of linked adaptive alleles might allow multiple aspects of skeletal morphology to evolve rapidly and simultaneously. In summary, the large set of skeletal QTL was used to explore the genetic properties of evolved mutations. These QTL are a starting point for the identification of the molecular basis of evolved changes in the vertebrate skeleton.

In chapter 3 of this thesis, I focus on one skeletal trait, gill raker number. Gill rakers are periodically patterned bones in a fish's throat important for feeding. In response to a change in diet in freshwater populations, hundreds of freshwater stickleback populations have independently evolved a reduction in gill raker number. This phenomenon is known as gill raker reduction. I found that heritable marine/freshwater differences in gill raker number were present early in development. The gill raker number differences were due to a difference in the spacing between adjacent gill rakers. The marine/freshwater number and spacing differences were present at the earliest detectable point of gill raker specification. This result suggests that gill raker reduction is caused by an early-acting change to a lateral inhibition process controlling raker bud spacing. Next, I performed linkage mapping in F2 fish from crosses with three independently derived freshwater populations. In all three crosses, gill raker QTL mapped to chromosomes 4 and 20, suggesting a similar genetic basis. Finally, the chromosome 4 and 20 QTL affected the early spacing of gill raker buds. Collectively, these data demonstrated that parallel developmental genetic features underlie the convergent evolution of gill raker reduction in freshwater sticklebacks. These results suggest that even highly polygenic adaptive traits can have a predictable developmental genetic basis.

In chapter 4 of this thesis, I present an economical method that uses next-generation sequencing to perform genome-wide genotyping of hundreds of multiplexed samples. This method was used to genotype two large marine by freshwater F2 crosses of over 350 fish each. The resulting maps significantly improved the stickleback genome assembly by making over 100 changes to the order and orientation of the scaffolds that compose the genome assembly. In the revised genome assembly, 95% of the assembly was anchored to a chromosome, compared to 87% in the original assembly. To assess linkage map quality, I mapped quantitative trait loci (QTL) controlling lateral plate number. Plate number mapped as expected to a 200 kilobase genomic region containing *Ectodysplasin*. In addition, plate number mapped to a chromosome 7 QTL overlapping a previously identified modifier QTL. Finally, I examined gill raker length. I found that the two freshwater populations convergently evolved shorter gill rakers relative to marine fish. I mapped eight QTL controlling gill raker length in the two crosses. However, none of the QTL were overlapping between the two crosses. Thus, the convergent evolution gill raker length reduction has a surprisingly non-parallel genetic basis.

In chapter 5 of this thesis, I present additional experiments to determine the genetic basis of evolved reduction in gill raker number. Using the genotyping method developed in chapter 4, I mapped QTL controlling raker number and spacing in two large

marine by freshwater F2 crosses of over 350 fish each. Overlapping QTL on chromosomes 4, 16, and 20 were found in both crosses. In addition, 14 unique modifier QTL were mapped. Together these results indicate that a combination of similar and different genetic basis underlies the convergent evolution of gill raker number. Using crosses with recombinant chromosomes, I fine-mapped the chromosome 4 and 20 QTL, substantially narrowing the interval sizes. Surprisingly, the chromosome 4 QTL in the two crosses mapped to non-overlapping regions. Therefore, even though the chromosome 4 QTL appeared to map to similar genomic regions in the two crosses, a different genetic basis was actually used. An excellent candidate gene, *Fibroblast Growth Factor 20 (Fgf20)* was located within one of the fine-mapped chromosome 4 QTL. I mutated the coding region of *Fgf20* with a genome-editing technology called Transcription Activator-Like Effector Nucleases (TALENs). Induced loss of function mutations in *Fgf20* resulted in a decrease in gill raker number. The *Fgf20* mutants had a concomitant increase in gill raker spacing, similar to the evolved freshwater phenotype. Induced mutations in a second gene, *Smad5*, also resulted in a raker phenotype. However, *Smad5* was excluded from the fine-mapped QTL intervals, so it likely does not underlie gill raker evolution. No coding changes in *Fgf20* likely cause evolved gill raker reduction. These results suggest a model in which *cis*-regulatory mutations that affect *Fgf20* expression levels or spatial patterns contribute to evolved gill raker reduction.

Together, this work makes significant progress towards understanding the genetic and developmental basis of skeletal evolution in stickleback fish. The results have broad implications for understanding the process of how genetic variation contributes to adaptive evolution.

Table of Contents

Chapter 4:
Genome assembly improvement and mapping convergently evolved skeletal traits in sticklebacks with Genotyping-by-Sequencing

Chapter 5:
Induced mutations in *Fgf20* phenocopy evolved changes in gill raker spacing

List of Figures

Chapter 2:

Chapter 3:

Chapter 4:

Chapter 5:

List of Tables

Chapter 2:

Chapter 3:

Chapter 4:

Chapter 5:

Acknowledgements

I thank Craig Miller, for being an excellent mentor throughout my graduate project. Throughout my time at Berkeley, I could always look to Craig to provide great scientific, career, and life advice. I also thank Dan Rokhsar, Richard Harland, Marvalee Wake, Rachel Brem, and Craig Moritz for guidance and support as members of my qualifying exam and thesis committees. Thanks to David Kingsley and Dolph Schluter, who provided excellent guidance for the skeletal QTL mapping project, and to Dan Rokhsar for his guidance on the GBS project. Thanks to the many coauthors and others who contributed so much to the success of this project. Further scientific acknowledgements are located at the end of each chapter.

I thank all the members of the Miller Lab, who provided a great atmosphere to work in. Phil Cleves was never afraid to think big and push me to do the same, whether in science or on Youtube. Priscilla Erickson provided great scientific input and shared my passion for figuring out the best way to present data. Nick Ellis never let me forget that I was ultimately working on a developmental biology project. James Hart gave me great genomic and programming advice. Chris Martin provided a valuable outsider's viewpoint on how to interpret and package my thesis work for publication. I also thank the many excellent undergraduates I worked with. I would especially like to thank Angela Lam, Emily Killingbeck, and Kari Miller, who each took on big chunks of this project and made them their own. It was really gratifying to see each of them grow as a scientist during their time in the lab.

Thanks to my family and friends for their support and camaraderie through this process. Thanks to my dad, for his encouragement and instilling in me a love of learning. Thanks to my mom, for teaching me that being organized, responsible, and making a strong effort are important to accomplishing anything of value. Thanks to my sister Rachel for her friendship. And many thanks to my wife Jenn, for her support and partnership during my time in grad school. Most of the time we've been together has spanned my time in grad school, so I look back on my time in Berkeley as intertwined with many wonderful, formative years of our relationship.

Chapter 1:
An introduction to evolutionary genetics, stickleback adaptive radiation, and gill raker evolution

Major questions in evolutionary genetics

In order to adapt to changing environments, populations of organisms have evolved a dramatic variety of phenotypes. Understanding the process of how genetic variation contributes to adaptive evolution is a longstanding question in biology. In the 1930s and 1940s, the Modern Synthesis merged Mendelian genetics with population genetics to create the modern theoretical understanding of how traits evolve (Huxley 1942). This work, expanded upon by generations of population geneticists, has provided theoretical models of how genetic alleles arise and change in frequency due to forces such as drift and selection. More recently, evolutionary biologists have harnessed the tools of molecular biology, genetics, and genomics to empirically study specific evolved phenotypes and understand their genetic basis. Using a variety of model and non-model organisms from microbes to plants to animals, ongoing work is attempting to address several major questions in evolutionary genetics, including the properties of adaptive alleles, the extent of parallel evolution, how evolved changes affect developmental mechanisms, and what genes are used.

Properties of adaptive alleles

Some phenotypes are completely controlled by a single "Mendelian" locus. For example, whether a person has cystic fibrosis is completely determined by the gene CFTR. However, most traits in nature do not segregate as a simple Mendelian manner. These traits are "complex" or "quantitative", controlled by the combined action of multiple genes. Modern linkage mapping techniques, first developed for human medical genetics (Lander and Botstein 1989), can be used to map specific genomic regions contributing to evolved quantitative traits. The mapping of specific genomic regions that control a trait, called Quantitative Trait Loci (QTL), allows the investigation of the general properties of evolved alleles.

Some of the major properties of adaptive alleles include (1) whether they have a large or small effect on phenotype, (2) whether they are dominant, additive, or recessive, (3) whether the causative mutations are located within or outside or genes, and (4) whether alleles have a global or a modular effect on phenotypes. Understanding these properties of adaptive alleles can yield a broader understanding of the process of adaptive evolution.

Large or small effect

Mutations can have either a large or a small effect on phenotype. Does adaptive evolution proceed by a small number of mutations with a large effect on phenotype, or a large number of mutations, each with a small effect? This question is important, because large and small effect mutations might have different evolutionary properties (Rockman 2012). Fisher proposed a geometric model in which, due to negative pleiotropic effects of large effect mutations, small effect mutations predominate (Fisher 1930). More recent theoretical models have suggested that medium or large effect mutations could play a role in adaptive evolution. This result was because when smaller effect mutations first arise, they are more likely to be lost due to genetic drift (Kimura 1983; Orr 1998). This work predicts an exponential distribution of effect sizes, with a few large-effect loci and many small-effect loci underlying an evolved phenotype (Orr 1998). However, this work is a theoretical model that was based on a relatively idealized situation, where in response to a new phenotypic optimum, adaptive alleles arise due to new mutations, then sweep to fixation. In nature, adaptive alleles are likely a mixture of new mutations that

undergo “hard sweeps” to fixation and preexisting variants, already present at low frequency, that undergo “soft sweeps.” In addition, adaptive alleles might be supplied through migration or gene flow. Theoretical models predict that increased migration might promote selection for larger effect mutations (Yeaman and Whitlock 2011; Griswold 2006). The empirical distribution of effect sizes in various organisms with different evolutionary dynamics has not been thoroughly explored, and therefore remains unclear. In my thesis, I explore the distribution of effect sizes of a large set of over 100 QTL.

Dominance

An important genetic property of a QTL that may affect its likelihood of fixation during selection is its dominance. In diploids, the dominance of an allele is a measure of whether one copy of that allele is sufficient to confer a phenotype. Alleles can be dominant (heterozygotes have the same phenotype as individuals homozygous for the variant allele), recessive (heterozygotes do not have a phenotype) or additive (heterozygotes have an intermediate phenotype). According to evolutionary theory, in the case of a new mutation undergoing positive selection, dominant alleles should be more likely to sweep to fixation than recessive alleles, because rare heterozygotes will have the adaptive phenotype and therefore be “seen” by natural selection (Haldane 1927). However, adaptive alleles may also arise from “standing variants,” preexisting mutations that increase in frequency after a shift in environment. For standing variants, there is little difference in the likelihood of alleles of different dominances to sweep to fixation, because of counteracting biases in initial allele frequency and fixation probability (Orr and Betancourt 2001). Therefore if selection from standing variation predominates in an evolutionary system, the underlying distribution of dominances of advantageous mutations should translate into a similar distribution of observed dominances of successful alleles that sweep to fixation.

Most new mutations are thought to be recessive (Fisher 1928; Orr 1991). However, advantageous mutations may have a different distribution of dominances than all mutations. Few studies have described empirically the dominance distributions of large numbers of naturally or artificially evolved alleles. Three exceptions are large QTL studies in tomatoes, mice, and sunflowers, which describe a surprising bias towards additive QTL (Burke et al. 2002; deVicente and Tanksley 1993; Kenney-Hunt et al. 2008). The dominance of an allele might at least partially reflect the molecular nature of the allele (e.g. regulatory mutations, which contribute to subtle variations in gene expression, might tend to be more additive). Thus, knowing general trends in dominances of alleles using during adaptation might help predict the typical underlying molecular nature of these alleles. In my thesis, I investigate the distribution of dominances of a large set of fish skeletal QTL.

Coding vs. regulatory

Mutations inside the protein-coding region of a gene can directly affect protein function by altering the amino acid composition of a protein. However, mutations outside the coding region can also affect function. These mutations, called “*cis*-regulatory” mutations, can affect DNA elements, such as enhancers, that affect the expression level or pattern of a gene. One major question is the extent to which adaptive phenotypic changes arise through coding vs. *cis*-regulatory mutations. In multicellular organisms, protein coding changes in a gene are likely to alter that gene's function at all locations and times where the gene is normally expressed. If a gene has multiple, pleiotropic functions during development, coding mutations will affect all of

those functions. In contrast, *cis*-regulatory mutations can alter a gene's expression at a highly specific subset of times and/or locations. This specific is possible because enhancers that control the spatial and temporal expression of a gene are modular in the genome, limiting phenotypic effects to subdomains of a gene's normal functions. The *cis*-Regulatory Hypothesis (Carroll 2008; Stern and Orgogozo 2008) states that mutations used during adaptive evolution are more likely to be *cis*-regulatory than coding mutations, because *cis*-regulatory mutations can have specific, non-pleiotropic effects.

What is the extent of coding vs. cis-regulatory mutations underlying adaptive evolution in nature? Empirical evidence so far has been mixed: some studies have identified *cis*-regulatory mutations underlying evolutionary change while other studies have found clear cases of adaptive coding mutations (Hoekstra and Coyne 2007; Martin and Orgogozo 2013; Stern and Orgogozo 2008). Famous examples of coding variation include *Alx1* controlling beak shape in in Darwin's finches (Lamichhaney et al. 2015), *Edar* in humans controlling hair thickness (Fujimoto et al. 2008), and *Mc1r* controlling pigmentation changes in birds and other vertebrates (Mundy 2005). Famous examples of noncoding variation include *shavenbaby* controlling trichome patterning in flies (Frankel et al. 2011), lactase persistence in humans (Tishkoff et al. 2007), and *Pitx1* and *Eda* in sticklebacks controlling armor changes (Colosimo et al. 2005; Shapiro et al. 2004). It is generally easier to identify coding mutations than *cis*-regulatory mutations, and therefore a discovery bias might exist. More case studies are needed to understand the prevalence of *cis*-regulatory vs. coding mutations and understand the conditions in which each is favored. In my thesis, I attempt to discover the genetic basis of an important evolved trait, gill raker reduction, and discover whether the underlying QTL are due to coding or regulatory mutations.

Extent of modularity

Many regions of multicellular organisms involve segmented, repeated, or periodic structures. For example, in vertebrates, the vertebrae, digits, branchial arches, teeth, and hair are all repeated or periodic structures. Despite this repetition, individual regions or segments can have distinguishing features, a phenomenon is known as modularity (Wagner et al. 2007). For example, different vertebrae or teeth can have modular specializations and features. In cases where a repeated structure exists, evolution often acts in a modular way on one aspect of that structure. For example, even though fish have multiple dorsal spines with clear homology, angler fish have evolved a specific modular extension of the first dorsal spine to use as a lure for prey. How is this type of evolution accomplished? Are there evolved alleles that have specific modular effects on a subset of possible domains?

One reason to suspect that modular QTL might be prevalent is that many genes have specific roles in regional patterning of segmented structures. For example, Hox genes have modular effects on specifying body regions along the anterior-posterior axis, and the Dlx genes have modular effects on dorsal/ventral patterning in the branchial arches (reviewed in Minoux and Rijli 2010). However, few studies have examined large numbers of evolved traits to examine the modularity of evolved alleles. In my thesis, I analyze a set of over 100 QTL to detect the extent of modular QTL in fish skeletal evolution.

Convergent evolution and parallel genetic mechanisms

One dramatic pattern of evolutionary change is convergent evolution, the repeated evolution of the same phenotypes in independent lineages. Convergent evolution has been observed for a

wide variety of traits, including morphological traits such as the body shape of sharks and tunas (Donley et al. 2004) and trunk-ground lizard ecomorphs (Kolbe et al. 2011), physiological traits such as human lactase persistence (Tishkoff et al. 2007) and hemoglobin changes in waterfowl to adapt to high altitudes (McCracken et al. 2009), and behavior traits such as web-building in spiders (Blackledge and Gillespie 2004) and feeding posture in cavefish (Kowalko et al. 2013). One striking conclusion from these studies is that convergent evolution often occurs via parallel genetic mechanisms, with the same genomic regions, genes, and sometimes even alleles used for evolutionary change. For example, genetic mapping in flies revealed that similarly mapping QTL underlie the convergent evolution of body size (Calboli et al. 2003), and an allele of *Optix* is repeatedly used in the convergent evolution of butterfly wing pattern (Reed et al. 2011). These observations of convergent evolution, among many others, suggest that some evolutionary trajectories are constrained and perhaps even predictable.

Why some evolved phenotypes appear to have a predictable genetic basis remains a major unanswered question in biology, but could result from topology of genetic networks, constraints to developmental programs, constraints to available genetic variation, correlated response to selection on another trait, or even coincidence (Jeffery 2009; Kopp 2009; Losos 2011; Stern 2000, 2013). Within an evolving species, one factor that might promote parallel evolution through the same genetic basis is the presence of standing variation, where mutations are present in the population at low frequency prior to adaptation. These mutations are then available for selection upon repeated environmental shifts. This pattern, recently termed "collateral evolution" (Stern 2013), has been documented in several cases such as stickleback lateral plate reduction (Colosimo et al. 2005) or wing coloration changes in *Heliconius* butterflies (Heliconius Genome 2012). So far, the documented cases have all had fairly simple genetic architectures with single alleles of large effect. The extent to which collateral evolution has occurred for more complex, highly polygenic traits is poorly understood, and is a question that I will address in my thesis.

Developmental mechanisms?

How do evolved changes affect developmental mechanisms? Several recent studies have begun to examine the developmental timing and mechanisms of evolved traits. For example, mice with evolved pigmentation changes have late-acting changes in *Agouti* expression that delayed differentiation of melanocytes (Manceau et al. 2011) and lizards with evolved changes in limb lengths have early modifications to limb morphogenesis programs (Sanger et al. 2012). However, most studies of evolved alleles focus on their role on adult phenotypes, and few studies have directly interrogated the developmental effects of evolved alleles. A full mechanistic understanding of evolutionary change will include an understanding of how developmental mechanisms are altered to control adult phenotypes. One open question is whether the same types of changes to developmental processes underlie convergently evolved phenotypes. Distinct (non-parallel) developmental processes can generate convergent phenotypes, as was seen in the evolution of fly sex combs through two different developmental mechanisms to produce nearly identical adult phenotypes (Tanaka et al. 2009). In my thesis, I investigate the developmental basis of one important adaptive trait, stickleback gill raker reduction.

What genes are used?

In the past two decades, a growing number of researchers have attempted to discover the gene(s) and/or mutation(s) underlying specific evolved phenotypes. Are there preferred types of genes or mutations that underlie evolved changes? There might be key nodes in a genetic network that are common targets of evolution, because they give rise to a desired phenotype (Stern and Orgogozo 2008). Developmental biologists have identified a small number of signaling pathways that are repeatedly used during the development of many different body parts. Are genes in these pathways common targets of evolution?

Identifying the genes underlying an evolved QTL can help answer specific questions about the underlying genetic properties of that locus. For example, identifying the specific gene(s) underlying a QTL would reveal whether that QTL is due to protein coding or *cis*-regulatory mutations. By identifying the genes underlying many individual cases of evolution across different organisms, the evolutionary genetics field may learn general principles about the genetic basis of evolution. In my thesis, I attempt to discover the genes underlying one important evolved trait, gill raker reduction in stickleback fish.

Threespine sticklebacks: a model for evolutionary genetics

Stickleback adaptive radiation

The threespine stickleback fish (*Gasterosteus aculeatus*) provides an excellent opportunity to study mechanisms of adaptive evolution. Sticklebacks have undergone a dramatic adaptive radiation, in which ancestral marine fish have independently colonized thousands of freshwater environments (Bell and Foster 1994). To adapt to new freshwater environmental conditions, freshwater sticklebacks have evolved a characteristic suite of phenotypic changes. Two of the major changes include reductions in skeletal defensive armor in response to decreased predation (e.g. Gross 1978) and a set of changes to trophic (feeding) morphology in response to a change in prey from plankton to larger prey (e.g. Hagen 1967). Many freshwater populations were formed from colonization of new freshwater lakes and streams generated by widespread melting of glaciers at the end of the last Ice Age approximately 15,000 years ago. Therefore, evolution of new freshwater adaptations in these populations can be dated to within 15,000 years. Strikingly, rapid evolution of several traits over the timespan of a few decades has been documented (Aguirre and Bell 2012; Bell et al. 2004; Kitano et al. 2008). This widespread and rapid morphological evolution makes sticklebacks an excellent organism to dissect the processes and patterns of evolutionary change.

In addition to the marine/freshwater sticklebacks, two other stickleback variants have been used to study the processes of evolutionary change, the benthic/limnetic and lake/stream contrasts. The benthic/limnetic species pairs are pairs of different morphs of sticklebacks that co-inhabit the same environment. The pairs have been found in five lakes in British Columbia, Canada (Schluter and McPhail 1992). The benthic morph is bottom-dwelling and feeds on large prey, and has a suite of behavioral and morphological differences relative to the limnetic morph, which dwells in the open water and eats smaller prey. The lake/stream contrast consists of lake-dwelling sticklebacks and sticklebacks inhabiting the lake's inlet and outlet streams (Berner et al. 2010). These fish display evolved differences, including differences in body shape and feeding morphology.

Genetic mapping of evolved stickleback changes

Forward genetics is the unbiased search for the genetic basis of a phenotype, without prior knowledge of the gene(s) involved. The forward genetic mapping approach has traditionally been used in a small number of model organisms. Traditionally, mutagens are applied to induce random genetic mutations, some of which cause phenotypes of interest. When a phenotype of interest arises, the genetic crosses with mutant organisms can be used to map the underlying gene(s) that were mutated. In addition phenotypes in the laboratory using induced mutations, forward genetic mapping can also be used to investigate the mutations underlying evolved phenotypes. Instead of using mutagens to produce a phenotype, this approach uses "natural mutants" from the wild that have already evolved a phenotype of interest. Two different wild forms can be interbred to map the genetic basis of the evolved phenotype of interest.

This forward genetic approach relies on the ability to generate genetic crosses between wild organisms with different phenotypes. Despite their phenotypic differences, marine and freshwater sticklebacks can be crossed in the laboratory to generate fertile hybrids, and these hybrids can be intercrossed to create F2s. These F2s can be genotyped for molecular markers genome-wide and phenotyped for traits of interest. If a statistical association is found between genotype at a region of the genome with phenotype, that region is termed a quantitative trait locus, or QTL. This approach has been used fruitfully in sticklebacks to study the genetic basis of a wide variety of evolved traits, including skeletal armor (Colosimo et al. 2004; Peichel et al. 2001; Shapiro et al. 2004), body shape (Albert et al. 2008), pigmentation (Greenwood et al. 2011; Miller et al. 2007), tooth number (Cleves et al. 2014), branchial bone length (Erickson et al. 2014), and schooling behavior (Greenwood et al. 2015; Greenwood et al. 2013). From the initial mapping using hundreds of F2 progeny, QTL are typically resolved to dozens to hundreds of genes. Previously mapped stickleback QTL have ranged in effect size from controlling less than 5% of the variance in the F2 phenotype to nearly 100%. A few of the largest-effect stickleback QTL have been fine-mapped to further narrow the underlying genetic interval. A variety of fine-mapping approaches have been used, including analyzing the phenotypes of fish with recombination events inside the QTL interval (Cleves et al. 2014; Miller et al. 2007), association mapping using the genotypes and phenotypes of large numbers of wild fish (Colosimo et al. 2005), and scans of genomes of wild fish for molecular signatures of a haplotypes that have undergone a selective sweep (Chan et al. 2010; Jones et al. 2012b). Due to the fully sequenced stickleback genome (Jones et al. 2012b), the genes within a QTL interval can easily be identified. Candidate genes for underlying the QTL can be tested using a variety of expression and functional studies. Further functional and sequencing experiments can identify the Quantitative Trait Nucleotide(s)—a causative mutation or mutations underlying the QTL (Chan et al. 2010; O'Brown et al. 2015).

In three cases in sticklebacks, linkage mapping and further experiments have been used to identify specific genes underlying evolved phenotypes. These three cases are pelvic spine reduction (mapped to *Pitx1*), lateral plate loss (*Eda*), and pigmentation changes (*Kit ligand*) (Chan et al. 2010; Colosimo et al. 2005; Colosimo et al. 2004; Miller et al. 2007; Shapiro et al. 2004). All three genes are major developmental regulators, and in each case *cis*-regulatory variation of the genes was linked to changes in the evolved phenotype. For these three cases, the QTL identified in the initial linkage mapping experiment had a very large effect, controlling over 50% of the variance of the phenotype in the F2s. In contrast to these cases with relatively simple genetic architectures, the genes underlying traits with a more complicated quantitative genetic

basis remain unclear. In my thesis, I map QTL underlying a large set of skeletal traits, including many with a complicated polygenic basis.

Parallel evolution in sticklebacks

Genetic studies in sticklebacks have revealed that the same genomic regions (Colosimo et al. 2004; Coyle et al. 2007; Cresko et al. 2004; Shapiro et al. 2004), genes (Chan et al. 2010), and even alleles (Colosimo et al. 2005; Miller et al. 2007) can be reused in separate cases of freshwater adaptation. Intriguingly, recent genome-wide genotyping and genome resequencing studies in sticklebacks have identified striking re-use of many genomic variants during freshwater adaptation (Hohenlohe et al. 2010; Jones et al. 2012a; Jones et al. 2012b). These results suggest that parallel genetic evolution is common in sticklebacks. One factor that might promote parallel evolution in sticklebacks is the fact that many freshwater stickleback populations maintain a low level of contact and gene flow with marine populations. Mutations that are adaptive for freshwater environments might sweep to a high frequency in freshwater, then be exported back to marine environments. Under the "transporter" model of stickleback evolution, adaptive freshwater variants are maintained at low frequency in marine environments, and can make their way back into new freshwater environments, where they are re-assembled into the freshwater genotype (Bell and Aguirre 2013; Schluter and Conte 2009). Sticklebacks might therefore be more likely to use standing variation to rapidly evolve freshwater phenotypes than other organisms where gene flow between different forms does not occur.

The best studied examples of stickleback parallel evolution are the genes *Ectodysplasin* and *Kit Ligand.* Alleles of these two genes control adaptive differences in lateral plate number and pigmentation. These alleles are present at low frequency in marine environments and have been repeatedly selected for in freshwater environments. However, these previously studied QTL each explain over half of the variance in the trait. Large effect QTL appear to be rare, but are easier to study than small effect mutations (Rockman 2012). One outstanding question in evolutionary biology is whether principles discovered for traits with a relatively simple genetic basis also apply to traits with a more polygenic basis. The degree to which highly polygenic traits evolve using a parallel genetic basis is largely unknown. In this thesis, I study the extent to which the evolution of more complex, polygenic stickleback traits occurs through parallel evolution.

Stickleback genomics

The recent development of next-generation high-throughput DNA sequencing technologies has resulted in an explosion of genomic tools and methods for analyzing evolved differences. The relatively small (~460 Megabase) stickleback genome has been sequenced and a high-quality assembly has been constructed (Jones et al. 2012b). Scaffolds were anchored in the stickleback reference assembly using a linkage map made from an F2 cross of 92 fish. The assembly consists of 113 large anchored scaffolds on 21 chromosomes (87% of the assembly), as well as 1,822 unanchored scaffolds. Further understanding the structure of the stickleback genome should further aid efforts to understand the evolutionary dynamics of the stickleback genome and discover the genes underlying stickleback adaptations. In my thesis I use high-quality linkage maps to improve the stickleback genome assembly.

Twenty additional stickleback genomes have been resequenced, revealing widespread use of standing variants present at low frequency in marine fish that repeatedly sweep to high

frequency in freshwater environments to accomplish freshwater adaptation (Jones et al. 2012b). Additional genomic studies using high-throughput genotyping or genome sequencing of wild sticklebacks has identified specific genomic regions that show signals of selection between or within populations (Chain et al. 2014; Chan et al. 2010; Deagle et al. 2013; Feulner et al. 2015; Hohenlohe et al. 2010; Jones et al. 2012a; Roesti et al. 2012). With the low cost of high-throughput sequencing, genomic analysis using next-generation genome sequencing promises to be a powerful tool towards understanding the genomic basis of stickleback evolution.

Forward and reverse genetics in sticklebacks

In addition to the excellent genomic tools for sticklebacks, a combination of forward and reverse genetic approaches is a promising avenue for determining the genetic basis of evolved traits. In sticklebacks, forward genetic approaches using recombinant fine-mapping has helped determine the genes underlying the QTL controlling lateral plate number and pigmentation (Colosimo et al. 2005; Miller et al. 2007). Recombinant mapping approaches have also been used to fine-map smaller-effect QTL in wasps and flies (Loehlin and Werren 2012; McGregor et al. 2007). These approaches should be feasible in sticklebacks because of the generation time (~5 months) and ability to raise large genetic crosses.

Reverse genetics is process where the phenotype of a specific gene is determined through targeted manipulations of that gene's function. Genome editing technologies such as Transcription Activator-Like Effector Nucleases (TALENs) can be used for reverse genetic testing of a gene's function by inducing targeted mutations (Gaj et al. 2013). TALENs induce targeted double strand breaks, which can be repaired by the Non-Homologous End Joining pathway. This error-prone process results in small insertions or deletions, which can result in frameshifted, nonsense mutations. This method works especially efficiently in fish, where external fertilization allows collection of one-celled zygotes that can be injected with these genome editing reagents (e.g. Bedell et al. 2012; Erickson et al. 2014). Thus, reverse genetic approaches that directly test the function of candidate genes should be feasible in sticklebacks. In my thesis, I use a combination of genomics and forward and reverse genetics to attempt to discover genomic regions and genes underlying evolved stickleback phenotypes.

Evolved stickleback gill raker reduction

A classic set of phenotypes studied by evolutionary biologists are trophic or feeding traits, as many radiations (e.g. the Galapagos finches, African cichlids, threespine sticklebacks) display striking correlations between a population's craniofacial pattern and diet (Albertson et al. 2003; Albertson et al. 2005; Bernatchez and Dodson 1990; Grant and Grant 2006; Hulsey et al. 2008; Mallarino et al. 2011; McGee and Wainwright 2013; Muschick et al. 2012; Ruber et al. 1999; Schluter and McPhail 1992; Schluter et al. 1985). A classic example of trophic adaptation is evolved changes in gill raker number and length in fishes (Schluter 2000). Gill rakers are periodically spaced bones in the throat of the fish that are involved in filter feeding tiny prey. Typically, fish that feed on small prey items such as zooplankton have longer and more numerous, densely-spaced gill rakers than fish that feed on larger prey such as invertebrates. Gill rakers have variable morphologies within the same species (Kahilainen et al. 2011; Schluter 2000) as well as between species, including broad macroevolutionary differences (Magnuson and Heitz 1971; Noramly and Morgan 1998). Imaging of gill rakers during feeding in multiple

species suggested that instead of acting as a dead-end sieve, gill rakers typically act as a crossflow filter (Sanderson et al. 2001). In addition, gill rakers produce mucus, which may affect the dynamics of fluid flow or prey retention (Smith and Sanderson 2007). Therefore, evolved variations in gill raker morphology likely affect prey retention by altering the fluid dynamics of filtration through the branchial cavity (Cheer et al. 2012).

Stickleback gill raker evolution

In threespine sticklebacks, gill raker reduction has been documented in over a hundred independently derived stickleback populations (and has likely evolved thousands of times) from three main ecological contrasts: marine vs. freshwater, limnetic vs. benthic zones within a lake, and lake vs. adjoining inlet or outlet streams (e.g. Bell 1982; Berner et al. 2008; Berner et al. 2009; Gross and Anderson 1984; Hagen 1967, 1973; Hagen and Gilbertson 1972; Hendry et al. 2002; Lavin and Mcphail 1985, 1986, 1993; Mcphail 1984; Miller and Hubbs 1969; Moodie and Reimchen 1976; Raeymaekers et al. 2007; Reimchen et al. 1985; Schluter 1993; Schluter and McPhail 1992). Marine sticklebacks feed primarily on tiny zooplankton in the water column, whereas freshwater sticklebacks that have adapted to feeding in the benthic zone have shifted to a diet of larger invertebrates living in sediments or attached to vegetation (Gross and Anderson 1984; Kislalioglu and Gibson 1977). As a consequence, ancestral oceanic fish have many, long gill rakers compared to typical freshwater adapted creek populations. Marine sticklebacks typically have around 22 row 1 gill rakers, whereas highly reduced freshwater populations have about 16 row 1 gill rakers (Gross and Anderson 1984; Hagen and Gilbertson 1972). The repeated evolution of gill raker reduction throughout the Northern Hemisphere suggests that gill raker number is under strong natural selection. Gill raker number predicts feeding efficiency, with high gill raker counts correlating with better foraging on zooplankton (Roesch et al. 2013; Robinson 2000), and low gill raker counts correlating with better foraging on benthos (Lundsgaard-Hansen et al. 2013).

The ecological importance of gill raker reduction has been best studied in five benthic-limnetic species pairs in British Columbia (Schluter and McPhail 1992). The limnetic morphs feed in open water on smaller planktonic prey, and have many long gill rakers (even more than marine populations). In contrast, the benthic morphs feed on larger prey in the substrate and have fewer, shorter gill rakers (Schluter and McPhail 1992). In addition to gill raker differences, several other trophic differences are present including differences in gape width and other traits important for suction feeding, epaxial musclulature, and skull morphology (McGee et al. 2013; Willacker et al. 2010). Transplantation experiments where benthic and limnetic fish were raised in opposite habitats show that these trophic differences predict foraging success and growth (Bentzen and Mcphail 1984; Hatfield and Schluter 1999; Larson 1976; Lavin and Mcphail 1986; Schluter 1993, 1995). In the benthic and limnetic species' preferred habitats, they can grow up to twice as fast (Schluter 1995). Although many differences in trophic morphology likely contribute to this difference in fitness, gill raker differences likely contribute substantially to the different abilities of limnetics and benthics to feed on zooplankton and larger prey. In an F2 cross in an artificial pond, gill raker number was a major component of niche score, a measure of benthic vs. limnetic-like specialization, which in turn predicted diet specialization and fitness (Arnegard et al. 2014).

Genetic control of gill raker variation

Genetic studies in sticklebacks have revealed a strong genetic component to divergent gill raker numbers. Heritability studies on gill raker number in sticklebacks have shown that gill raker number has a high heritability (0.51, Aguirre et al. 2004; 0.58, Hagen 1973; 0.34-0.47 for ventral raker number, Hermida et al. 2002; 0.36, Schluter 1996) and a strong genetic component has been supported by several other studies (Day et al. 1994; Hagen 1967; Hatfield 1997; Lavin and Mcphail 1986, 1987). An experiment in which the diets of benthic and limnetic sticklebacks were switched identified almost no contribution of phenotypic plasticity to gill raker number differences (a not statistically significant 1% shift in phenotype, Day et al. 1994). Gill raker lengths were partially altered by diet switching (38% reduction in phenotypic differences), but showed a large heritable component as well (Day et al. 1994). A study (Hatfield 1997) estimated the number of genes underlying gill raker differences by analyzing phenotype distributions of parental, F1, and F2 fish. Using two variants of Wright's estimator (Wright 1968), this study estimated that approximately 17 or 50 loci controlled gill raker number, indicating a highly polygenic genetic basis. Interestingly, an additive model had the best fit, with no significant dominance or epistasis between alleles. Consistent with a large genetic contribution, specific QTL underlying gill raker number have been mapped in both sticklebacks (Arnegard et al. 2014; Peichel et al. 2001) and whitefish (Gagnaire et al. 2013; Rogers and Bernatchez 2007). QTL mapping of gill raker differences should allow for identification of additional regions underlying evolved gill raker reduction.

Gill rakers as a model for epithelial appendage evolution

Gill rakers appear to have genetic and developmental similarities to other vertebrate epithelial appendages, a broad class of periodically patterned organs that include hair, teeth, feathers, sweat glands, and scales (reviewed in Chuong 1998; Chuong et al. 2013). These structures form embryonically from placodes—transient, regularly spaced, epithelial thickenings that signal to underlying mesenchyme to make an epithelial organ (reviewed in Chuong 1998). Evolved changes in patterning of epithelial appendages have occurred repeatedly during vertebrate evolution. For example, in Asian human populations, a derived allele of the *EDAR* gene affecting hair, sweat gland, and mammary gland morphology underwent one of the strongest selective sweeps in the genome (Kamberov et al. 2013). Epithelial appendages might represent a "hot spot" target for evolution, as they form and function at the interface between an organism and its environment (reviewed in Sadier et al. 2014).

Ectodysplasin (Eda) and the gene encoding the EDA receptor, *Ectodysplasin receptor (Edar)* play highly conserved roles in the development of placodes. Mice and humans with strong loss-of-function mutations in either gene have ectodermal dysplasia, with defects in teeth, hair, and sweat glands (reviewed in Sadier et al. 2014). During development of epithelial appendages, *Edar* is typically expressed in the placodes, flanked by a complementary expression pattern of *Eda* around the non-placode forming part of the field (Drew et al. 2007; Houghton et al. 2005; Laurikkala et al. 2002; Pispa et al. 2003; Tucker et al. 2004). Interestingly, in zebrafish, *Eda* and *Edar* are required for proper formation of gill rakers, as well as teeth and scales (Harris et al. 2008). In cichlid larvae, *Edar* is expressed within developing gill rakers and *Eda* is expressed between gill rakers (Fraser et al. 2008). This shared genetic requirement and complementary expression pattern of *Eda* and *Edar* suggests that gill rakers and other epithelial appendages develop by similar co-opted developmental genetic regulatory networks.

Two other signaling pathways, the Fibroblast Growth Factor (FGF) and Bone

Morphogenetic Protein (BMP) pathways, also play major roles in patterning epithelial appendages. Typically FGF promotes an increased number of denser placodes, whereas BMP plays an inhibitory role, promoting fewer, less dense placodes (Botchkarev et al. 1999; Jung et al. 1998; Noramly and Morgan 1998; Pispa and Thesleff 2003; Wells et al. 2012). In chickens, selection for fitness in hot climates resulted in the evolution of breeds with featherless necks, caused by the upregulation of an inhibitory gene, *Bmp12*, during feather placode development (Mou et al. 2011). A nonsense mutation in *Fgf20* results in featherless chickens (Wells et al. 2012). Intriguingly, in carp, mutants for an FGF receptor *Fgfr1* have fewer scales and quantitative reductions in gill raker number (Golovinskaya 1940; Rohner et al. 2009). Since both activating and inhibitory genes control the spacing of other epithelial appendages, mutations that contribute to an increase in the spacing of gill raker primordia could increase the strength of inhibitory genes, decrease the strength of activating genes, or both. The extent to which EDA, FGF and BMP pathway genes affect gill raker patterning and evolution in sticklebacks is unknown. Understanding the developmental and genetic mechanisms underlying stickleback gill raker evolution might further shed light on general principles of epithelial appendage evolution.

Summary

In this thesis I address several questions about the genetic basis of evolved phenotypes. In chapter 2, I present the results of a large quantitative genetic study of over 100 stickleback skeletal traits. Over 100 QTL were identified controlling bone numbers, sizes, and patterns. Using these QTL several general questions in evolutionary genetics are addressed, including: the extent to which loci are clustered in the genome, the dominance distribution of evolutionary alleles, and the proportion of loci that have modular effects. These QTL are a starting point for the identification of the molecular basis of many evolved changes in the vertebrate skeleton.

In chapter 3 of this thesis, I focus on the genetic and developmental basis of one convergently evolved skeletal trait, gill raker reduction. Dense developmental time courses are analyzed to identify the timing of divergence between marine and freshwater morphology and likely developmental processes that are altered. To ask whether the convergent evolution of highly polygenic traits uses similar genetic mechanisms, I map gill raker number in F2 crosses with three independently derived freshwater populations. Specifically, I test whether the two largest effect QTL on chromosomes 4 and 20 are used in each cross. Collectively, the study tests the extent to which parallel developmental genetic features underlie the convergent evolution of gill raker reduction in freshwater sticklebacks.

In chapter 4 of this thesis, I present a new method that uses next-generation sequencing to economically generate genome-wide genotypes of hundreds of multiplexed samples. I use this method to genotype two large marine by freshwater F2 crosses of over 350 fish each. The resulting linkage maps significantly improve the genome assembly by making over 100 changes to scaffold order and orientation. These crosses are used to determine the genetic basis of gill raker length reduction. I also test whether the convergent evolution of this quantitative trait is due to modular and/or parallel QTL.

In chapter 5 of this thesis, I present additional experiments to determine the genetic basis of evolved reduction in gill raker number. Using the genotyping method developed in chapter 4, I map QTL controlling gill raker number and spacing in two large marine by freshwater F2 crosses of over 350 fish each. Using small genetic crosses with recombinant chromosomes, the

chromosome 4 and 20 QTL from these two freshwater populations are fine-mapped, substantially narrowing the interval sizes. Finally, I induce loss of function mutations in two candidate genes, *Fibroblast Growth Factor 20 (Fgf20)* and *Smad5* to determine whether these genes affect gill raker patterning and evolution in sticklebacks.

Together this work makes significant progress towards understanding the genetic and developmental basis of skeletal evolution in stickleback fish. The results have broad implications for understanding the process of how genetic variation contributes to adaptive evolution.

REFERENCES

Aguirre, W.E., and M.A. Bell, 2012 Twenty years of body shape evolution in a threespine stickleback population adapting to a lake environment. Biological Journal of the Linnean Society 105: 817-831.

Aguirre, W.E., P.K. Doherty, and M.A. Bell, 2004 Genetics of lateral plate and gillraker phenotypes in a rapidly evolving population of threespine stickleback. Behaviour 141: 1465-1483.

Albert, A.Y., S. Sawaya, T.H. Vines, A.K. Knecht, C.T. Miller *et al.*, 2008 The genetics of adaptive shape shift in stickleback: pleiotropy and effect size. Evolution 62: 76-85.

Albertson, R.C., J.T. Streelman, and T.D. Kocher, 2003 Directional selection has shaped the oral jaws of Lake Malawi cichlid fishes. Proc. Natl. Acad. Sci. U. S. A 100: 5252-5257.

Albertson, R.C., J.T. Streelman, T.D. Kocher, and P.C. Yelick, 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.

Arnegard, M.E., M.D. McGee, B. Matthews, K.B. Marchinko, G.L. Conte *et al.*, 2014 Genetics of ecological divergence during speciation. Nature 511: 307-311.

Bedell, V.M., Y. Wang, J.M. Campbell, T.L. Poshusta, C.G. Starker *et al.*, 2012 *In vivo* genome editing using a high-efficiency TALEN system. Nature 491: 114-118.

Bell, M.A., 1982 Differentiation of Adjacent Stream Populations of Threespine Sticklebacks. Evolution 36: 189-199.

Bell, M.A., and W.E. Aguirre, 2013 Contemporary evolution, allelic recycling, and adaptive radiation of the threespine stickleback. Evolutionary Ecology Research 15: 377-411.

Bell, M.A., W.E. Aguirre, and N.J. Buck, 2004 Twelve years of contemporary armor evolution in a threespine stickleback population. Evolution 58: 814-824.

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

Bernatchez, L., and J.J. Dodson, 1990 Allopatric origin of sympatric populations of lake whitefish (*Coregonus clupeaformis*) as revealed by mitochondrial DNA restriction analysis. Evolution 44: 1263-1271.

Berner, D., D.C. Adams, A.C. Grandchamp, and A.P. Hendry, 2008 Natural selection drives patterns of lake-stream divergence in stickleback foraging morphology. Journal of Evolutionary Biology 21: 1653-1665.

Berner, D., A.C. Grandchamp, and A.P. Hendry, 2009 Variable Progress toward Ecological Speciation in Parapatry: Stickleback across Eight Lake-Stream Transitions. Evolution 63: 1740-1753.

Berner, D., M. Roesti, A.P. Hendry, and W. Salzburger, 2010 Constraints on speciation suggested by comparing lake-stream stickleback divergence across two continents. Mol Ecol 19: 4963-4978.

Blackledge, T.A., and R.G. Gillespie, 2004 Convergent evolution of behavior in an adaptive radiation of Hawaiian web-building spiders. Proc. Natl. Acad. Sci. U. S. A 101: 16228-16233.

Botchkarev, V.A., N.V. Botchkareva, W. Roth, M. Nakamura, L.H. Chen *et al.*, 1999 Noggin is a mesenchymally derived stimulator of hair-follicle induction. Nature Cell Biology 1: 158-164.

Burke, J.M., S. Tang, S.J. Knapp, and L.H. Rieseberg, 2002 Genetic analysis of sunflower domestication. Genetics 161: 1257-1267.

Calboli, F.C., 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.

Carroll, S.B., 2008 Evo-devo and an expanding evolutionary synthesis: a genetic theory of morphological evolution. Cell 134: 25-36.

Chain, F.J., P.G. Feulner, M. Panchal, C. Eizaguirre, I.E. Samonte *et al.*, 2014 Extensive copy-number variation of young genes across stickleback populations. Plos Genetics 10: e1004830.

Chan, Y.F., M.E. Marks, F.C. Jones, G. Villarreal, Jr., M.D. Shapiro *et al.*, 2010 Adaptive evolution of pelvic reduction in sticklebacks by recurrent deletion of a *Pitx1* enhancer. Science 327: 302-305.

Cheer, A., S. Cheung, T.C. Hung, R.H. Piedrahita, and S.L. Sanderson, 2012 Computational fluid dynamics of fish gill rakers during crossflow filtration. Bull Math Biol 74: 981-1000.

Chuong, C.-M., 1998 *Molecular Basis of Epithelial Appendage Morphogenesis*. Austin, TX: R.G. Landes.

Chuong, C.M., C.Y. Yeh, T.X. Jiang, and R. Widelitz, 2013 Module based complexity formation: periodic patterning in feathers and hairs. WIREs Dev. Biol. 2: 97-112.

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

Colosimo, P.F., K.E. Hosemann, S. Balabhadra, G. Villarreal, Jr., M. Dickson *et al.*, 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 *et al.*, 2004 The genetic architecture of parallel armor plate reduction in threespine sticklebacks. PLoS Biol. 2: 635-641.

Coyle, S.M., F.A. Huntingford, and C.L. Peichel, 2007 Parallel evolution of *Pitx1* underlies pelvic reduction in Scottish threespine stickleback (*Gasterosteus aculeatus*). J. Hered. 98: 581-586.

Cresko, W.A., A. Amores, C. Wilson, J. Murphy, M. Currey *et al.*, 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.

Day, T., J. Pritchard, and D. Schluter, 1994 A Comparison of 2 Sticklebacks. Evolution 48: 1723-1734.

Deagle, B.E., F.C. Jones, D.M. Absher, D.M. Kingsley, and T.E. Reimchen, 2013 Phylogeography and adaptation genetics of stickleback from the Haida Gwaii archipelago revealed using genome-wide single nucleotide polymorphism genotyping. Mol Ecol 22: 1917-1932.

deVicente, M.C., and S.D. Tanksley, 1993 QTL analysis of transgressive segregation in an interspecific tomato cross. Genetics 134: 585-596.
Donley, J.M., C.A. Sepulveda, P. Konstantinidis, S. Gemballa, and R.E. Shadwick, 2004 Convergent evolution in mechanical design of lamnid sharks and tunas. Nature 429: 61-65.
Drew, C.F., C.M. Lin, T.X. Jiang, G. Blunt, C. Mou *et al.*, 2007 The Edar subfamily in feather placode formation. Dev. Biol. 305: 232-245.
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 Biol Sci 281: 20140822.
Feulner, P.G., F.J. Chain, M. Panchal, Y. Huang, C. Eizaguirre *et al.*, 2015 Genomics of divergence along a continuum of parapatric population differentiation. Plos Genetics 11: e1004966.
Fisher, R.A., 1928 The possible modification of the response of the wild type to recurrent mutations. American Naturalist 62: 115-126.
Fisher, R.A., 1930 *The genetical theory of natural selection*. Oxford,: The Clarendon press.
Frankel, N., D.F. Erezyilmaz, A.P. McGregor, S. Wang, F. Payre *et al.*, 2011 Morphological evolution caused by many subtle-effect substitutions in regulatory DNA. Nature 474: 598-603.
Fraser, G.J., R.F. Bloomquist, and J.T. Streelman, 2008 A periodic pattern generator for dental diversity. BMC Biol. 6.
Fujimoto, A., R. Kimura, J. Ohashi, K. Omi, R. Yuliwulandari *et al.*, 2008 A scan for genetic determinants of human hair morphology: EDAR is associated with Asian hair thickness. Hum Mol Genet 17: 835-843.
Gagnaire, P.A., E. Normandeau, S.A. Pavey, and L. Bernatchez, 2013 Mapping phenotypic, expression and transmission ratio distortion QTL using RAD markers in the Lake Whitefish (Coregonus clupeaformis). Mol Ecol 22: 3036-3048.
Gaj, T., C.A. Gersbach, and C.F. Barbas, 3rd, 2013 ZFN, TALEN, and CRISPR/Cas-based methods for genome engineering. Trends Biotechnol 31: 397-405.
Golovinskaya, K., 1940 Pleiotropic effect of scale genes in carp. Comptes Rendus De L Academie Des Sciences De L Urss 28: 533-536.
Grant, P.R., and B.R. Grant, 2006 Evolution of character displacement in Darwin's finches. Science 313: 224-226.
Greenwood, A.K., R. Ardekani, S.R. McCann, M.E. Dubin, A. Sullivan *et al.*, 2015 Genetic Mapping of Natural Variation in Schooling Tendency in the Threespine Stickleback. G3 (Bethesda).
Greenwood, A.K., F.C. Jones, Y.F. Chan, S.D. Brady, D.M. Absher *et al.*, 2011 The genetic basis of divergent pigment patterns in juvenile threespine sticklebacks. Heredity (Edinb) 107: 155-166.
Greenwood, A.K., A.R. Wark, K. Yoshida, and C.L. Peichel, 2013 Genetic and neural modularity underlie the evolution of schooling behavior in threespine sticklebacks. Current Biology 23: 1884-1888.
Griswold, C.K., 2006 Gene flow's effect on the genetic architecture of a local adaptation and its consequences for QTL analyses. Heredity (Edinb) 96: 445-453.

Gross, H.P., 1978 Natural-Selection by Predators on Defensive Apparatus of 3-Spined Stickleback, Gasterosteus-Aculeatus L. Canadian Journal of Zoology-Revue Canadienne De Zoologie 56: 398-413.
Gross, H.P., and J.M. Anderson, 1984 Geographic variation in the gillrakers and diet of European threespine sticklebacks, *Gasterosteus aculeatus*. Copeia: 87-97.
Hagen, D.W., 1967 Isolating Mechanism in Threespine Sticklebacks (Gasterosteus). Journal of the Fisheries Research Board of Canada 24: 1637-1692.
Hagen, D.W., 1973 Inheritance of Numbers of Lateral Plates and Gill Rakers in Gasterosteus-Aculeatus. Heredity (Edinb) 30: 303-312.
Hagen, D.W., and L.G. Gilbertson, 1972 Geographic variation and environmental selection in *Gasterosteus aculeatus* L. in the Pacific Northwest, America. Evolution 26: 32-51.
Haldane, J.B.S., 1927 A mathematical theory of natural and artificial selection, Part V: Selection and mutation. Proceedings of the Cambridge Philosophical Society 23: 838-844.
Harris, M.P., N. Rohner, H. Schwarz, S. Perathoner, P. Konstantinidis *et al.*, 2008 Zebrafish *eda* and *edar* mutants reveal conserved and ancestral roles of ectodysplasin signaling in vertebrates. PLoS Genet. 4: e1000206.
Hatfield, T., 1997 Genetic divergence in adaptive characters between sympatric species of stickleback. Am. Nat. 149: 1009-1029.
Hatfield, T., and D. Schluter, 1999 Ecological speciation in sticklebacks: Environment-dependent hybrid fitness. Evolution 53: 866-873.
Heliconius Genome, C., 2012 Butterfly genome reveals promiscuous exchange of mimicry adaptations among species. Nature 487: 94-98.
Hendry, A.P., E.B. Taylor, and J.D. McPhail, 2002 Adaptive divergence and the balance between selection and gene flow: Lake and stream stickleback in the misty system. Evolution 56: 1199-1216.
Hermida, M., C. Fernandez, R. Amaro, and E. San Miguel, 2002 Heritability and "evolvability" of meristic characters in a natural population of *Gasterosteus aculeatus*. Can. J. Zool. 80: 532-541.
Hoekstra, H.E., and J.A. Coyne, 2007 The locus of evolution: Evo devo and the genetics of adaptation. Evolution 61: 995-1016.
Hohenlohe, P.A., S. Bassham, P.D. Etter, N. Stiffler, E.A. Johnson *et al.*, 2010 Population genomics of parallel adaptation in threespine stickleback using sequenced RAD tags. PLoS Genet. 6: e1000862.
Houghton, L., C. Lindon, and B.A. Morgan, 2005 The ectodysplasin pathway in feather tract development. Development 132: 863-872.
Hulsey, C.D., R.J. Roberts, A.S.P. Lin, R. Guldberg, and J.T. Streelman, 2008 Convergence in a mechanically complex phenotype: Detecting structural adaptations for crushing in cichlid fish. Evolution 62: 1587-1599.
Huxley, J., 1942 *Evolution, the modern synthesis*. London,: G. Allen & Unwin ltd.
Jeffery, W.R., 2009 Regressive evolution in *Astyanax* cavefish. Annu. Rev. Genet. 43: 25-47.
Jones, F.C., Y.F. Chan, J. Schmutz, J. Grimwood, S.D. Brady *et al.*, 2012a A genome-wide SNP genotyping array reveals patterns of global and repeated species-pair divergence in sticklebacks. Current Biology 22: 83-90.
Jones, F.C., M.G. Grabherr, Y.F. Chan, P. Russell, E. Mauceli *et al.*, 2012b The genomic basis of adaptive evolution in threespine sticklebacks. Nature 484: 55-61.

Jung, H.S., P.H. Francis-West, R.B. Widelitz, T.X. Jiang, S. Ting-Berreth *et al.*, 1998 Local inhibitory action of BMPs and their relationships with activators in feather formation: implications for periodic patterning. Dev Biol 196: 11-23.

Kahilainen, K.K., A. Siwertsson, K.O. Gjelland, R. Knudsen, T. Bohn *et al.*, 2011 The role of gill raker number variability in adaptive radiation of coregonid fish. Evolutionary Ecology 25: 573-588.

Kamberov, Y.G., S.J. Wang, J.Z. Tan, P. Gerbault, A. Wark *et al.*, 2013 Modeling recent human evolution in mice by expression of a selected EDAR variant. Cell 152: 691-702.

Kenney-Hunt, J.P., B. Wang, E.A. Norgard, G. Fawcett, D. Falk *et al.*, 2008 Pleiotropic patterns of quantitative trait loci for 70 murine skeletal traits. Genetics 178: 2275-2288.

Kimura, M., 1983 *The neutral theory of molecular evolution*. Cambridge Cambridgeshire ; New York: Cambridge University Press.

Kislalioglu, M., and R.N. Gibson, 1977 Feeding relationship of shallow water fishes in a Scottish sea loch. Journal of Fish Biology 11: 257-266.

Kitano, J., D.I. Bolnick, D.A. Beauchamp, M.M. Mazur, S. Mori *et al.*, 2008 Reverse evolution of armor plates in the threespine stickleback. Current Biology 18: 769-774.

Kolbe, J.J., L.J. Revell, B. Szekely, E.D. Brodie, 3rd, and J.B. Losos, 2011 Convergent evolution of phenotypic integration and its alignment with morphological diversification in Caribbean *Anolis* ecomorphs. Evolution 65: 3608-3624.

Kopp, A., 2009 Metamodels and phylogenetic replication: a systematic approach to the evolution of developmental pathways. Evolution 63: 2771-2789.

Kowalko, J.E., N. Rohner, T.A. Linden, S.B. Rompani, W.C. Warren *et al.*, 2013 Convergence in feeding posture occurs through different genetic loci in independently evolved cave populations of *Astyanax mexicanus*. Proc. Natl. Acad. Sci. U. S. A 110: 16933-16938.

Lamichhaney, S., J. Berglund, M.S. Almen, K. Maqbool, M. Grabherr *et al.*, 2015 Evolution of Darwin's finches and their beaks revealed by genome sequencing. Nature 518: 371-375.

Lander, E.S., and D. Botstein, 1989 Mapping Mendelian Factors Underlying Quantitative Traits Using Rflp Linkage Maps. Genetics 121: 185-199.

Larson, G.L., 1976 Social-Behavior and Feeding Ability of 2 Phenotypes of Gasterosteus-Aculeatus in Relation to Their Spatial and Trophic Segregation in a Temperate Lake. Canadian Journal of Zoology-Revue Canadienne De Zoologie 54: 107-121.

Laurikkala, J., J. Pispa, H.S. Jung, P. Nieminen, M. Mikkola *et al.*, 2002 Regulation of hair follicle development by the TNF signal ectodysplasin and its receptor Edar. Development 129: 2541-2553.

Lavin, P.A., and J.D. Mcphail, 1985 The Evolution of Fresh-Water Diversity in the Threespine Stickleback (Gasterosteus-Aculeatus) - Site-Specific Differentiation of Trophic Morphology. Canadian Journal of Zoology-Revue Canadienne De Zoologie 63: 2632-2638.

Lavin, P.A., and J.D. Mcphail, 1986 Adaptive divergence of trophic phenotype among freshwater populations of the threespine stickleback (*Gasterosteus aculeatus*). Canadian Journal of Fisheries and Aquatic Sciences 43: 2455-2463.

Lavin, P.A., and J.D. Mcphail, 1987 Morphological Divergence and the Organization of Trophic Characters among Lacustrine Populations of the Threespine Stickleback (Gasterosteus-Aculeatus). Canadian Journal of Fisheries and Aquatic Sciences 44: 1820-1829.

Lavin, P.A., and J.D. Mcphail, 1993 Parapatric Lake and Stream Sticklebacks on Northern Vancouver Island - Disjunct Distribution or Parallel Evolution. Canadian Journal of Zoology-Revue Canadienne De Zoologie 71: 11-17.
Loehlin, D.W., and J.H. Werren, 2012 Evolution of shape by multiple regulatory changes to a growth gene. Science 335: 943-947.
Losos, J.B., 2011 Convergence, adaptation, and constraint. Evolution 65: 1827-1840.
Lundsgaard-Hansen, B., B. Matthews, P. Vonlanthen, A. Taverna, and O. Seehausen, 2013 Adaptive plasticity and genetic divergence in feeding efficiency during parallel adaptive radiation of whitefish (*Coregonus* spp.). J. Evol. Biol. 26: 483-498.
Magnuson, J.J., and J.G. Heitz, 1971 Gill Raker Apparatus and Food Selectivity among Mackerels, Tunas, and Dolphins. Fishery Bulletin of the National Oceanic and Atmospheric Administration 69: 361-&.
Mallarino, R., P.R. Grant, B.R. Grant, A. Herrel, W.P. Kuo *et al.*, 2011 Two developmental modules establish 3D beak-shape variation in Darwin's finches. Proc. Natl. Acad. Sci. U. S. A 108: 4057-4062.
Manceau, M., V.S. Domingues, R. Mallarino, and H.E. Hoekstra, 2011 The developmental role of Agouti in color pattern evolution. Science 331: 1062-1065.
Martin, A., and V. Orgogozo, 2013 The Loci of Repeated Evolution: A Catalog of Genetic Hotspots of Phenotypic Variation. Evolution 67: 1235-1250.
McCracken, K.G., C.P. Barger, M. Bulgarella, K.P. Johnson, S.A. Sonsthagen *et al.*, 2009 Parallel evolution in the major haemoglobin genes of eight species of Andean waterfowl. Mol. Ecol. 18: 3992-4005.
McGee, M.D., D. Schluter, and P.C. Wainwright, 2013 Functional basis of ecological divergence in sympatric stickleback. Bmc Evolutionary Biology 13.
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 *et al.*, 2007 Morphological evolution through multiple *cis*-regulatory mutations at a single gene. Nature 448: 587-590.
Mcphail, J.D., 1984 Ecology and Evolution of Sympatric Sticklebacks (Gasterosteus) - Morphological and Genetic-Evidence for a Species Pair in Enos Lake, British-Columbia. Canadian Journal of Zoology-Revue Canadienne De Zoologie 62: 1402-1408.
Miller, C.T., S. Beleza, A.A. Pollen, D. Schluter, R.A. Kittles *et al.*, 2007 *Cis*-regulatory changes in *Kit ligand* expression and parallel evolution of pigmentation in sticklebacks and humans. Cell 131: 1179-1189.
Miller, R.R., and C.L. Hubbs, 1969 Systematics of Gasterosteus Aculeatus with Particular Reference to Intergradation and Introgression Along Pacific Coast of North America - a Commentary on a Recent Contribution. Copeia: 52-&.
Minoux, M., and F.M. Rijli, 2010 Molecular mechanisms of cranial neural crest cell migration and patterning in craniofacial development. Development 137: 2605-2621.
Moodie, G.E.E., and T.E. Reimchen, 1976 Phenetic Variation and Habitat Differences in Gasterosteus Populations of Queen Charlotte Islands. Systematic Zoology 25: 49-61.
Mou, C., F. Pitel, D. Gourichon, F. Vignoles, A. Tzika *et al.*, 2011 Cryptic patterning of avian skin confers a developmental facility for loss of neck feathering. PLoS Biol. 9: e1001028.

Mundy, N.I., 2005 A window on the genetics of evolution: MC1R and plumage colouration in birds. Proc Biol Sci 272: 1633-1640.
Muschick, M., A. Indermaur, and W. Salzburger, 2012 Convergent evolution within an adaptive radiation of cichlid fishes. Current Biology 22: 2362-2368.
Noramly, S., and B.A. Morgan, 1998 BMPs mediate lateral inhibition at successive stages in feather tract development. Development 125: 3775-3787.
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 4: e05290.
Orr, H.A., 1991 A test of Fisher's theory of dominance. Proc Natl Acad Sci U S A 88: 11413-11415.
Orr, H.A., 1998 The population genetics of adaptation: The distribution of factors fixed during adaptive evolution. Evolution 52: 935-949.
Orr, H.A., and A.J. Betancourt, 2001 Haldane's sieve and adaptation from the standing genetic variation. Genetics 157: 875-884.
Peichel, C.L., K.S. Nereng, K.A. Ohgi, B.L. Cole, P.F. Colosimo *et al.*, 2001 The genetic architecture of divergence between threespine stickleback species. Nature 414: 901-905.
Pispa, J., M.L. Mikkola, T. Mustonen, and I. Thesleff, 2003 *Ectodysplasin*, *Edar* and *TNFRSF19* are expressed in complementary and overlapping patterns during mouse embryogenesis. Gene Expr. Patterns 3: 675-679.
Pispa, J., and I. Thesleff, 2003 Mechanisms of ectodermal organogenesis. Dev Biol 262: 195-205.
Raeymaekers, J.A.M., J.K.J. Van Houdt, M.H.D. Larmuseau, S. Geldof, and F.A.M. Volckaert, 2007 Divergent selection as revealed by P-ST and QTL-based F-ST in three-spined stickleback (Gasterosteus aculeatus) populations along a coastal-inland gradient. Mol. Ecol. 16: 891-905.
Reed, R.D., R. Papa, A. Martin, H.M. Hines, B.A. Counterman *et al.*, 2011 *Optix* drives the repeated convergent evolution of butterfly wing pattern mimicry. Science 333: 1137-1141.
Reimchen, T.E., E.M. Stinson, and J.S. Nelson, 1985 Multivariate Differentiation of Parapatric and Allopatric Populations of Threespine Stickleback in the Sangan River Watershed, Queen-Charlotte-Islands. Canadian Journal of Zoology-Revue Canadienne De Zoologie 63: 2944-2951.
Robinson, B.W., 2000 Trade offs in habitat-specific foraging efficiency and the nascent adaptive divergence of sticklebacks in lakes. Behaviour 137: 865-888.
Rockman, M.V., 2012 The QTN program and the alleles that matter for evolution: all that's gold does not glitter. Evolution 66: 1-17.
Roesch, C., B. Lundsgaard-Hansen, P. Vonlanthen, A. Taverna, and O. Seehausen, 2013 Experimental evidence for trait utility of gill raker number in adaptive radiation of a north temperate fish. J. Evol. Biol. 26: 1578-1587.
Roesti, M., A.P. Hendry, W. Salzburger, and D. Berner, 2012 Genome divergence during evolutionary diversification as revealed in replicate lake-stream stickleback population pairs. Mol. Ecol. 21: 2852-2862.

Rogers, S.M., and L. Bernatchez, 2007 The genetic architecture of ecological speciation and the association with signatures of selection in natural lake whitefish (Coregonus sp. Salmonidae) species pairs. Mol Biol Evol 24: 1423-1438.
Rohner, N., M. Bercsenyi, L. Orban, M.E. Kolanczyk, D. Linke *et al.*, 2009 Duplication of fgfr1 permits Fgf signaling to serve as a target for selection during domestication. Current Biology 19: 1642-1647.
Ruber, L., E. Verheyen, and A. Meyer, 1999 Replicated evolution of trophic specializations in an endemic cichlid fish lineage from Lake Tanganyika. Proc. Natl. Acad. Sci. U. S. A 96: 10230-10235.
Sadier, A., L. Viriot, S. Pantalacci, and V. Laudet, 2014 The ectodysplasin pathway: from diseases to adaptations. Trends Genet. 30: 24-31.
Sanderson, S.L., A.Y. Cheer, J.S. Goodrich, J.D. Graziano, and W.T. Callan, 2001 Crossflow filtration in suspension-feeding fishes. Nature 412: 439-441.
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. Biol. Sci. 279: 739-748.
Schluter, D., 1993 Adaptive Radiation in Sticklebacks - Size, Shape, and Habitat Use Efficiency. Ecology 74: 699-709.
Schluter, D., 1995 Adaptive Radiation in Sticklebacks - Trade-Offs in Feeding Performance and Growth. Ecology 76: 82-90.
Schluter, D., 1996 Adaptive radiation along genetic lines of least resistance. Evolution 50: 1766-1774.
Schluter, D., 2000 *The ecology of adaptive radiation*. Oxford: Oxford University Press.
Schluter, D., and G.L. Conte, 2009 Genetics and ecological speciation. Proc. Natl. Acad. Sci. U. S. A 106: 9955-9962.
Schluter, D., and J.D. McPhail, 1992 Ecological character displacement and speciation in sticklebacks. Am. Nat. 140: 85-108.
Schluter, D., T.D. Price, and P.R. Grant, 1985 Ecological character displacement in Darwin's finches. Science 227: 1056-1059.
Shapiro, M.D., M.E. Marks, C.L. Peichel, B.K. Blackman, K.S. Nereng *et al.*, 2004 Genetic and developmental basis of evolutionary pelvic reduction in threespine sticklebacks. Nature 428: 717-723.
Smith, J.C., and S.L. Sanderson, 2007 Mucus function and crossflow filtration in a fish with gill rakers removed versus intact. J Exp Biol 210: 2706-2713.
Stern, D.L., 2000 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.
Tanaka, K., O. Barmina, and A. Kopp, 2009 Distinct developmental mechanisms underlie the evolutionary diversification of *Drosophila* sex combs. Proc. Natl. Acad. Sci. U. S. A 106: 4764-4769.
Tishkoff, S.A., F.A. Reed, A. Ranciaro, B.F. Voight, C.C. Babbitt *et al.*, 2007 Convergent adaptation of human lactase persistence in Africa and Europe. Nat. Genet. 39: 31-40.

Tucker, A.S., D.J. Headon, J.M. Courtney, P. Overbeek, and P.T. Sharpe, 2004 The activation level of the TNF family receptor, Edar, determines cusp number and tooth number during tooth development. Dev. Biol. 268: 185-194.

Wagner, G.P., M. Pavlicev, and J.M. Cheverud, 2007 The road to modularity. Nat Rev Genet 8: 921-931.

Wells, K.L., Y. Hadad, D. Ben-Avraham, J. Hillel, A. Cahaner *et al.*, 2012 Genome-wide SNP scan of pooled DNA reveals nonsense mutation in FGF20 in the scaleless line of featherless chickens. BMC Genomics 13: 257.

Willacker, J.J., F.A. Von Hippel, P.R. Wilton, and K.M. Walton, 2010 Classification of threespine stickleback along the benthic-limnetic axis. Biological Journal of the Linnean Society 101: 595-608.

Wright, S., 1968 *Evolution and the genetics of populations; a treatise*. Chicago,: University of Chicago Press.

Yeaman, S., and M.C. Whitlock, 2011 The Genetic Architecture of Adaptation under Migration-Selection Balance. Evolution 65: 1897-1911.

Chapter 2:
Additive and clustered quantitative trait loci control anatomically regional skeletal evolution in sticklebacks

The following chapter was originally published as an article:
Genetics 2014 Vol 197, 405-420

Craig T. Miller[*,§,1,7], Andrew M. Glazer[*,1], Brian R. Summers[§], Benjamin K. Blackman[§,2], Andrew R. Norman[§,3], Michael D. Shapiro[§,4], Bonnie L. Cole[§,5], Catherine L. Peichel[§,6], Dolph Schluter[†], and David M. Kingsley[§,7]

[*] Molecular and Cell Biology Department, University of California-Berkeley, Berkeley, California 94720
[§] Department of Developmental Biology and Howard Hughes Medical Institute, Stanford University, Stanford, California 94305
[†] Department of Zoology, University of British Columbia, Vancouver, British Columbia, Canada V6T 1Z4

[1] these authors contributed equally to this work
[2] present address: Department of Biology, University of Virginia, Charlottesville, Virginia 22904
[3] present address: Department of Biochemistry, University of California, San Francisco, San Francisco, California 94158
[4] present address: Department of Biology, University of Utah, Salt Lake City, Utah 84112
[5] present address: Department of Anatomic Pathology, University of Washington, and Seattle Children's Hospital, Department of Laboratories, Seattle, Washington 98105
[6] present address: Divisions of Human Biology and Basic Sciences, Fred Hutchinson Cancer Research Center, Seattle, Washington 98109
[7] corresponding authors

ABSTRACT

Understanding the genetic architecture of evolutionary change remains a long-standing goal in biology. In vertebrates, skeletal evolution has contributed greatly to adaptation in body form and function in response to changing ecological variables like diet and predation. Here we use genome-wide linkage mapping in threespine stickleback fish to investigate the genetic architecture of evolved changes in many armor and trophic traits. We identify over 100 quantitative trait loci (QTL) controlling the pattern of serially repeating skeletal elements, including gill rakers, teeth, branchial bones, jaws, median fin spines, and vertebrae. We use this large collection of QTL to address long-standing questions about the anatomical specficity, genetic dominance, and genomic clustering of loci controlling skeletal differences in evolving populations. We find that most QTL (76%) that influence serially repeating skeletal elements have anatomically regional effects. In addition, most QTL (71%) have at least partially additive effects, regardless of whether the QTL controls evolved loss or gain of skeletal elements. Finally, many QTL with high LOD scores cluster on chromosomes 4, 20, and 21. These results identify a modular system that can control highly specific aspects of skeletal form. Because of the general additivity and genomic clustering of major QTL, concerted changes in both protective armor and trophic traits may occur when sticklebacks inherit either marine or freshwater alleles at linked or possible "supergene" regions of the stickleback genome. Further study of these regions will help identify the molecular basis of both modular and coordinated changes in the vertebrate skeleton.

INTRODUCTION

Understanding the quantitative genetic architecture underlying evolutionary change in nature remains a major goal in genetics. The past two decades have seen a rapid increase in experimental data from various model systems, generating vigorous debate over the relative importance of coding vs. regulatory alleles, the prevalence of pleiotropy, and the role of large-effect mutations during adaptation to new environments (Rockman 2012; Stern and Orgogozo 2008; Streisfeld and Rausher 2011).

One particularly interesting genetic architecture found in several natural systems is close linkage of loci controlling multiple, often co-adaptive, phenotypes. Such trait clusters, sometimes called "supergenes", have been observed in primroses (Darwin 1877; Li et al. 2011; Mather 1950), butterflies (Clarke et al. 1968; Joron et al. 2006; Mallet 1989), snails (Murray and Clarke 1976), and fish (Protas et al. 2008; Roberts et al. 2009; Tripathi et al. 2009; Winge 1927). Trait clusters could result from recombination suppression (Noor et al. 2001), for example through chromosomal inversions (Fishman 2013; Joron et al. 2011; Lowry and Willis 2010). Alternatively, trait clusters could result from tightly linked loci or pleiotropic effects of individual genes (Mallet 1989; Studer and Doebley 2011). Having multiple different phenotypes controlled by the same genomic region could greatly facilitate rapid adaptive evolution (Feder et al. 2011; Kirkpatrick and Barton 2006; Yeaman and Whitlock 2011).

Adaptive mutations may arise *de novo*, or be selected from preexisting standing variants that become favorable following environmental change. When selection acts on newly arising mutations, dominant alleles should have a higher probability of fixation than recessive alleles (Haldane 1927). However, if previously unfavorable standing variant alleles become advantageous following environmental change, there is little bias in the likelihood of alleles of different dominances to sweep to fixation (Orr and Betancourt 2001). Therefore in systems where selection from standing variation predominates, the observed distribution of dominances should largely reflect the underlying distribution of dominances of advantageous mutations. Although most new mutations are recessive (Fisher 1928; Orr 1991), advantageous mutations may have a different distribution of dominances than all mutations. Dominance distributions of adaptive mutations are still poorly characterized, particularly for alleles underlying morphological traits in natural vertebrate populations.

cis-regulatory changes may predominate during morphological evolution because of the highly pleiotropic effects of developmental regulatory genes (Carroll 2008; Stern 2000). Protein coding changes in such genes will alter the gene's function at all sites of expression. In contrast, *cis*-regulatory changes can alter expression at highly specific times or locations, limiting phenotypic effects to subdomains of a gene's function. This idea predicts that QTL controlling adaptive morphological changes may typically act in subsets of anatomical regions. Although this idea can be tested by looking for regional vs. global effects among evolutionary QTL that influence serially repeating morphology, few studies have examined large numbers of traits to test the prevalence of modular genetic effects in naturally evolved species (Wagner et al. 2007).

The threespine stickleback (*Gasterosteus aculeatus*) species complex provides a powerful system for forward genetic dissection of repeated evolution in nature. Migratory marine sticklebacks colonized thousands of new freshwater lakes and streams following the last Ice Age. Newly established freshwater populations evolved similar phenotypes in response to similar ecological conditions, providing strong evidence that the corresponding traits evolve by natural

selection (Schluter 2000). Despite dramatic morphological and physiological differences among sticklebacks, intercrosses between populations produce viable and fertile offspring, making it possible to study the genetic and genomic mechanisms that underlie adaptive evolution in new environments (reviewed in Kingsley 2007; Schluter et al. 2010). The remarkably compact genome size (~460 Mb) has facilitated a high-quality genome assembly and resequencing of fish from twenty different populations, revealing abundant reuse of standing variants as one of several mechanisms underlying evolutionary differences in this system (Jones et al. 2012b).

Previous studies have identified many trophic and defensive armor traits that evolve repeatedly in freshwater (Bell and Foster 1994). A classic case of ecology-driven natural selection is the reduction in number of gill raker bones (Schluter 2000) in countless freshwater stickleback populations throughout the northern hemisphere (Gross and Anderson 1984; Hagen and Gilbertson 1972). Oceanic fish primarily feed on tiny zooplankton in the water column, while freshwater fish adapted to the benthic zone (bottom of lake) have shifted to a diet of larger invertebrates living in sediments or attached to vegetation (Gross and Anderson 1984; Kislalioglu and Gibson 1977). Both reduced gill raker number and larger jaw gape are found in benthic-adapted species (Schluter and Mcphail 1992). While large jaws and low gill raker counts correlate with more successful benthos foraging (Lavin and Mcphail 1986), small jaws and high gill raker counts correlate with more successful foraging of small prey from the water column (Bentzen and Mcphail 1984). Benthic-adapted stickleback forms also display changes in skull morphology that distinguish them from forms adapted to eat smaller prey items (McGee et al. 2013; Willacker et al. 2010). Collectively, these studies suggest that a concerted set of craniofacial changes allow freshwater populations to forage more efficiently on new diets in freshwater habitats. In addition to head skeletal traits, aspects of the median fin and vertebral skeleton are known to vary and be under selection in stickleback populations. These include dorsal spine lengths (Bell 2006; Gross 1978; Hunt et al. 2008), the number and position of dorsal and anal fin rays and their supporting pterygiophores, and vertebral number and positioning (Ahn and Gibson 1999; Swain 1992b; Swain 1992a).

Here we apply genome-wide linkage mapping to investigate the genetic architecture of over a hundred trophic, armor, and serially repeating skeletal traits in sticklebacks. Using a large set of newly identified quantitative trait loci (QTL) we address several general questions in evolutionary genetics, including: the extent to which loci are clustered in the genome, the dominance distribution of evolutionary alleles, and the proportion of loci that have anatomically regional effects. Our results show that loci controlling both regressive (loss) and constructive (gain) traits are clustered in the stickleback genome, making it possible to shape multiple aspects of both trophic and defensive morphology by co-inheritance of marine or freshwater alleles at linked loci.

METHODS

Ethics statement

All animal work was approved by University of British Columbia and Stanford Institutional Animal Care and Use Committees (protocol A97-0298 and 13834).

Grandparental population phenotyping

Skeletal morphologies of Japanese marine and Paxton benthic adult wild fish were compared by micro-computerized tomography using a Scanco uCT 40 scanned at 55kvp, 145 uA, at high resolution averaging four frames.

Phenotyping skeletal traits in F2 fish

We phenotyped 110 skeletal traits using a variety of methods described below. All traits were quantified on the left side, except for (1) premaxilla height and length which were quantified on the right side, and (2) premaxilla width, frontal width, supraoccipital traits, median fin and vertebral position traits, which are bilateral or midline measurements (see **Figure 2.2**). All linear measurements were quantified using an eye reticule on a Nikon SMZ1500 dissecting microscope unless noted otherwise.

Dissection method for branchial trait phenotyping

We developed a method to dissect out the entire branchial skeleton and mount it flat on a coverslip (**Figure 2.2A**). Briefly, under a Zeiss STEMI 2000 dissecting microscope with watchmaker's forceps, eyes were removed, and four cuts were made with iris or vannas scissors: two bilateral cuts dorsal to the opercle and hyomandibula through to the eye sockets, a cut across the frontal bone through the eye sockets, and a cut through the midline parasphenoid bone. Ventrally, the ceratohyals were disarticulated from the basihyal, and the urohyal removed. Next, the entire facial skeleton was removed, exposing the branchial skeleton. The epibranchials were detached from the neurocranium and the branchial skeleton removed by pulling the gut tube away from the rest of the fish. Soft tissue including the gut was removed and a single midline incision was made between the dorsal tooth plates to allow mounting the branchial skeletons flat on bridged coverslips as in **Figure 2.2A**. This method enables visualization of the entire branchial skeletal pattern from a dorsal view, as well as previously described variation in the pigmentation of the gill filaments from a ventral view (Miller et al. 2007).

Gill raker phenotyping

Along the anterior/posterior axis, gill rakers are distributed across nine rows projecting from both the anterior and posterior faces of all five branchial segments, except for the fifth branchial segment, which has only an anterior row (**Figure 2.2A**). Using the edge of Alizarin-positive branchial bone staining, we defined four dorsal-ventral raker domains as follows: (1) hypo (all rakers medial to the ceratobranchial), (2) cerato (bounded by the edges of the ceratobranchial bones), (3) joint (between epibranchial and ceratobranchial), and (4) epi (dorsal to the epibranchial) (**Figure 2.2B-E**). If a raker spanned these bone landmarks, the center of the raker base was used to assign each raker to a domain. We recorded raker number in each of these 25 anterior-posterior and dorsal-ventral domains using a Zeiss STEMI 2000 dissecting microscope. We also combined the individual domain phenotypes into 19 composite phenotypes in the following possible developmental modules: rows, segments (branchial arches), odd rows, even rows, all, and dorsal/ventral domains (hypo, cerato, joint, and epi).

In addition to gill raker number, we directly measured the inter-raker spacing distance at three positions (lateral, middle, and medial, **Figure 2.2F**) along row 2 rakers. Lateral spacing was measured between the second and third raker from the ceratobranchial-epibranchial joint, middle spacing was measured between two rakers in the middle of the ceratobranchial, and medial spacing was measured between the second and third raker from the hypobranchial-

ceratobranchial joint. All three spacing measurements were made between the center of the base of the two rakers being measured. For all three spacing measurements, if an atypical raker spacing was present following the above landmarks, an adjacent raker space was recorded if it appeared more typical of the spacing within the row.
We phenotyped gill raker length (from raker tip to ceratobranchial bone of the third raker from the ceratobranchial/epibranchial joint in rows 1 and 2), but after no significant genetic effect was detected after scoring 92 F2 males, we did not pursue this trait further.

Pharyngeal dentition phenotyping
We quantified pharyngeal tooth number on all three pharyngeal toothplates: the two dorsal toothplates (DTP1 and DTP2) attached to the pharyngobranchials (Anker 1974), and the one ventral pharyngeal toothplate (VTP) attached to the fifth ceratobranchial (**Figure 2.2G**). Teeth were counted using a Zeiss Axiophot compound microscope with DIC optics. Baby teeth that were visible under DIC but did not stain with Alizarin red were not counted. In addition, we measured the lengths and widths of all three toothplates (**Figure 2.2H**) by recording the longest and widest point-to-point measurements between Alizarin-positive toothplate bone.

Branchial bone phenotyping
Along the dorsal-ventral axis, the branchial skeleton consists of: four epibranchials (EBs, dorsal bones in the roof of the buccal cavity); five ceratobranchials (CBs, long ventral bones in the floor of the buccal cavity), and three hypobranchials (HBs, short ventral bones in between the ceratobranchials and the midline). We measured the lengths of all five ceratobranchials and the first epibranchial using the two anterior corners of Alizarin-positive bone as landmarks (see **Figure 2.2I**). The lengths of the highly three-dimensional epibranchials 2-4 and the widths of all ceratobranchials and epibranchials were not measured due to marked variation in mounting angles.

Phenotyping jaw traits
Premaxillas were manually removed then soaked for several minutes in a dilute 2.5% bleach solution to remove soft tissue before measuring height, width, and length as in **Figure 2.2J**. Lower jaw measurements were quantified by dissecting out and separating the left dentary and articular as in **Figure 2.2K**, acquiring digitized images with an Evolution MP camera using ImageProPlus on a Leica MZFLIII microscope, then using ImageProPlus software to make linear measurements as in **Figure 2.2K,L**.

Phenotyping skull and opercle traits
We quantified four skull traits: the linear measurement of frontal width or interorbital distance (**Figure 2.2M**), and three measurements of the supraoccipital crest (**Figure 2.2N**). Supraoccipital traits were quantified from digital images of the dorsal view of the skull taken with an Evolution MP camera on a Leica MZFLIII microscope and analyzed with ImageProPlus software. Three measurements of opercle size and shape were made: the length and width of the opercle, and a measurement of the width of the neck of the opercle (**Figure 2.2O**).

Phenotyping median fin and vertebral traits

Spine serrations were scored from digital images of the second dorsal spine acquired with a Nikon D1X camera fixed to a Nikon SMZ-U microscope. Area of the anterior surface of the spine was calculated by counting the number of pixels in Photoshop (Adobe) and converting to square millimeters. The serration area (SRA) was calculated by subtracting a digitally smoothened dorsal spine area (i.e., a spine without serrations, SDSA) from the total spine area (SPA, **Figure 2.2P**). Pterygiophore and fin ray number, and anal spine lengths were quantified under a Leica S8APO microscope with an eye reticule. For all vertebral traits, animals were first X-rayed (**Figure 2.2Q**) at 5x magnification for 15-20 seconds at 20 kV in a Micro–50 cabinet specimen radiography machine (Faxitron). Positions of bones in the median skeleton were assigned a numerical value corresponding to the closest vertebra as described(Ahn and Gibson 1999). The position of the last dorsal and anal fin ray was determined based on the position of the pterygiophore that supported the fin ray. On occasion, the element was judged to be equidistant from two vertebrae and was assigned a value that was an average of the two vertebrae.

Genome-wide linkage map construction

A set of 275 microsatellites was genotyped in a single full-sibling family ("Family 4") of 370 fish from a Japanese marine (JAMA) by Paxton benthic (PAXB) freshwater F2 cross (Colosimo et al. 2004). These markers consisted of previously described sets of genome-wide microsatellites (Colosimo et al. 2004; Albert et al. 2008; Peichel et al. 2001) and markers near previously mapped genes (Colosimo et al. 2005; Knecht et al. 2007; Miller et al. 2007; Shapiro et al. 2004). In addition, we added 16 new markers to the genetic map by genotyping new microsatellites near candidate genes with important roles in pharyngeal arch patterning in other vertebrates (*Dlx1/2*, *Dlx5/6*, *Dlx3*, *Msx1*, *Edn1*) as well as new positional markers. New markers were identified using a variety of methods including degenerate PCR, bacterial artificial chromosome (BAC) screening by radioactively labeled overgo hybridization, BAC end sequencing, physical map information, and publicly available previously sequenced BAC ends, as described below. A linkage map was constructed with JoinMap 3.0 (Kyazma), using previously described settings (Peichel et al. 2001) but by accepting more conservative LOD 6 groupings. The total map length is 1287.8 cM over 21 linkage groups, resulting in an average marker spacing of 5.1 cM. For each linkage group, proper phase was determined from the grandparental genotypes.

Cloning Dlx and Msx genes

Intergenic (*Dlx5/*6) or genic (*Msx1*, *Dlx3a*, and *Dlx3b*) regions of new genes added to map were amplified by PCR using the following primers (all sequences 5' to 3'). For *Dlx5/6,* PCR primers GGTGGGAAAGTGTTTGCACACC and CTGAGACAATCCGCATTCCTGTGG were designed to conserved intergenic sequences (Zerucha et al. 2000) which were found to flank Stn339 in intervening genomic sequence. For *Dlx3a* and *Dlx3b*, portions of two stickleback *Dlx3* genes were amplified and sequenced using a common forward degenerate primer (GGGTGAAGATHTGGTTYCARAA) and a reverse degenerate primer for either *Dlx3a* (CGGGCTGRTACCARTTYTGRTG) or *Dlx3b* (CGCCCTGYTGRTACCARTGRTT). The resulting *Dlx3* sequences were used to design two gene specific overgoes (see below) for BAC screening. Both *Dlx5/6* and *Dlx3* PCRs used Little Campbell marine genomic DNA as a template. For *Msx1,* degenerate RT-PCR primers CCGTTCAGCGTCGARGCNCTNATGGC and GGGGTGRTACATRCTRTANCC were used with oligo-dT reverse primed cDNA harvested

from a 1 cM long Little Campbell marine fry. The resulting RT-PCR amplicon sequence was used to design overgoes (see below) for BAC screening.

Overgo screening, BAC end sequencing, and genotyping

Overgo screening was performed as described at www.chori.org/bacpac/overgohyb.htm. *Dlx1/2* overgoes were directly designed to conserved intergenic sequence (Ghanem et al. 2003). Forward and reverse overgo sequences (5' to 3') for each marker or gene were: *Msx1*: CGGTAGTCTGGATACTTCAGTTCC and GCCCATCGATAAAGCAGGAACTGA; Stn207: TTTCAGCAGGTGCAACGTTTCCAC and AACTAAGAAGGCGAGCGTGGAAAC; *Dlx1/2:* ACCAAGATCTCGAGTGCACAATGT and CCTCATTACGCTGATGACATTGTG; *Dlx3a*: GGCGGCAGTATTAAGAGTAATGCG and CGGTGGGATCCACAAGCGCATTAC; *Dlx3b*: CCGACGCACAGCTCGTCGCCGCCA and TATAATCCTCCAGGTATGGCGGCG; Stn48: GTGCCAGAAAACTTGCATTCCAGG and ATCCCCTCACGTCACACCTGGAAT; EaccMgtg: GCAGGGTGATTGAATGTCTTCACT and GTCCTTAGGAAGATGCAGTGAAGA; 48B15.t7: AACAGTGTTGAGCGCTGAAATGCC and ACCTGTATGCACACACGGCATTTC; Stn292: AAGATACGGGCTGATGAGCAGTGA and TTCTTACTACGCCTCCTCACTGCT; Stn222: TCGCACTTCAGACACTAAGCCTTG and TGAAGGGTGTCCAAACCAAGGCTT; Bmp6: TGTGACGTTGACCTCAGCTAGACT and GAGGATTTAAACCGGGAGTCTAGC.
For overgo screening, three pairs of labeled overgoes were combined in one hybridization bottle containing four filters, and positive BACs subsequently identified using a combination of the physical map (Kingsley et al. 2004) and PCR screening. BAC ends were sequenced using ABI3730xl manufacturer's suggestions, using 8 uL of ABI BigDye per 20uL sequencing reaction. Genotypes were generated essentially as previously described (Miller et al. 2007; Peichel et al. 2001).

QTL mapping and analyses

A family of 370 full-sibling F2 fish derived from a Japanese Pacific marine grandmother and a Paxton Lake (British Columbia, Canada) benthic freshwater grandfather (Colosimo et al. 2004) was genotyped with 275 microsatellite markers and phenotyped for 110 skeletal traits across 8 trait classes. QTL within the same trait class with overlapping 1.5 LOD intervals were filtered, keeping the QTL with the highest LOD and removing lesser effect QTL to avoid redundant QTL sampling. This filtered QTL set was used for all dominance and clustering analyses. All raw phenotype, adjusted phenotype (see **Table 2.1**), and genotype data used for QTL mapping are presented in **File 2.1**. Details on the genetic positions, effect sizes, and dominances of all QTL are presented in **File 2.2**.

Trait processing and analysis was performed in R (http://www.R-project.org/). A custom pipeline was made to correct each trait for sex and/or size-dependence, log-transform if appropriate, and to remove phenotypic outliers as follows. First each trait was tested for size dependence by linear regression vs. standard length (SL), for sex dependence by a one-way ANOVA using sex as a factor, and for sex and size dependence using SL as a covariate and sex as a main effect in a General Linear Model (GLM) ANOVA. If the trait was neither sex nor size dependent, raw trait values were used for QTL mapping. If there was SL-dependence but no sex-dependence, traits were regressed against SL to obtain residuals. If there was sex-dependence but not size dependence, sex was corrected for using the residuals of a one-way ANOVA with sex as

a factor. If traits were significantly dependent upon both sex and size, the residuals of the GLM ANOVA were used for QTL mapping. Outliers (defined as fish that had trait values greater than four standard deviations from the mean trait value) were removed and ANOVAs, regressions, and GLM ANOVAs were redone without outliers. Outliers were rare and consisted of only 35 values for 17 total traits (AH, DH, DL, DS1L, DS2L, FDP, IL1, IL2, OPL, OPN, PD4, PML, PMW, SDS2A, SOL, SPA, and SRA). Traits were log-transformed when the transformation equalized variances (in sextiles ranked by standard length) by Levene's test for equality of variances, and/or normalized the residuals by an Anderson-Darling test of normality.

QTL mapping was performed in R/qtl (Broman and Sen 2009). Initial QTL mapping was performed with *scanone* with Haley-Knott regressions (hk). For each phenotype, ten thousand permutations were performed to determine a LOD threshold at which alpha equals 0.05. The average of these trait-specific thresholds was 4.1; thus this value was used as the QTL significance threshold for all traits. All significant QTL by *scanone* were also identified by *stepwise* mapping, so the larger *stepwise* set of QTL are presented here. The *stepwise* algorithm was performed by an automated forward-backward stepwise search for QTL using *stepwiseqtl* with a main penalty of 4.1, which was the average penalty from 100 *scantwo* permutations for each trait. QTL peak markers and LOD scores were calculated using *refineqtl* and percent variances explained were calculated with *fitqtl*. For a small number of traits (n=11), the *stepwiseqtl* output included markers with a LOD less than 4.1. These markers were conservatively removed. In 10 cases, the *stepwiseqtl* output included two markers on the same chromosome. Only cases where both peak markers had LODs greater than 4.1 and also had non-overlapping 1.5 LOD intervals were considered as two QTL. In cases where the two linked markers had overlapping 1.5 LOD intervals, only the peak marker with the highest LOD was considered a QTL. LOD scores for QTL on chromosomes that did not have significant effects were determined with *addqtl*, adjusting for QTL that were identified from the stepwise search. Additional QTL were included from *addqtl* if they surpassed a 4.1 LOD score threshold and LODs were recalculated as above. This *addqtl/refineqtl/fitqtl* process was iteratively repeated for three rounds and all QTL that had a LOD score above 4.1 in the final *fitqtl* model were included in the final QTL set. LOD scores for phenotypes with no significant QTL were determined by *scanone* with Haley-Knott regressions. Heat maps in **Figures 2.3-2.4** use color schemes from http://colorbrewer2.org/.

Anatomical specificity of QTL

For investigating the anatomical specificity of QTL, the subset of QTL with clearly or likely serially homologous domains (QTL controlling gill raker number, pharyngeal tooth number, branchial bone length, upper and lower jaw size, and dorsal spine lengths) were considered. QTL were considered regional if they affected a subset of domains, and global if they affected all domains. QTL controlling raker spacing were excluded because this phenotype was only quantified on one segment, and QTL controlling toothplate size and tooth number were analyzed separately.

Dominance analyses

To calculate the dominance of each QTL, Z-scored residual phenotypes were first calculated from a linear regression of the phenotype against all other peak marker genotypes affecting that phenotype. For calculating dominance, the equation *d/a* (Falconer 1989) was used, with *a*

representing the additive effect of one additional benthic allele (i.e. half the phenotypic difference between the homozygous benthic and homozygous marine genotypic classes). *d* represents the dominance effect: the difference between the heterozygous phenotype and the midpoint between homozygous parental phenotypes. Similar to a sunflower domestication QTL study (Burke et al. 2002), we used the following *d/a* ranges to classify the dominance effect of benthic alleles: <-1.25 for underdominant, -1.25 to -0.75 for recessive, -0.75 to -0.25 for partially recessive, -0.25 to 0.25 for additive, 0.25 to 0.75 for partially dominant, 0.75 to 1.25 for dominant and >1.25 for overdominant. For each QTL, one value of *a* and two values of *d* were calculated by Haley-Knott regression, as two classes of heterozygous F2 animals were present. For two out of 342 QTL, the chromosome had only one heterozygous genotypic class (because F1 parents had same heterozygous genotypes across this chromosome). For these two QTL, the single value of *d* was counted twice. For the other 340 QTL, *d* for both the M_1B_1 and M_2B_2 (M = Marine, B = Benthic) heterozygous genotypic classes was calculated.

Tests of trait clustering

To determine if QTL were significantly clustered in the genome, we took the observed number of QTL per trait class for (1) all QTL or (2) large-effect QTL, defined as the top quartile of QTL by LOD score (LOD >8.95). We then simulated 1000 random placements of peak markers in the genome, allowing only one QTL per trait class per chromosome (similar to the QTL filtering method described above). For each simulation, we determined the number of QTL with peak markers less than 5 cM away. We calculated *P* values by comparing the observed number of QTL having peak markers within 5 cM to this null distribution. To determine if the number of QTL on a single chromosome was significantly enriched relative to a null hypothesis of independent and evenly distributed QTL, simulations were also performed with all of the QTL or the top quartile of QTL by LOD score. For each simulation, the observed total number of trait classes with QTL on each chromosome was determined. This observed set of QTL for each trait class was distributed randomly to chromosomes without replacement with probability in proportion to 1) the genetic length of the chromosome, 2) the physical length of the chromosome (Jones et al. 2012b), or 3) the number of Ensembl-predicted genes (Jones et al. 2012b) within the chromosome. For each case, 10,000 simulations were performed to calculate a null distribution of QTL per chromosome as well as a mean number of "expected" QTL. For every chromosome, the true number of QTL was compared to the null distribution to calculate a *P* value. Since sexually dimorphic traits represent a genetic effect of the sex chromosome (chromosome 19) and this effect was largely statistically removed prior to QTL mapping, chromosome 19 was excluded from the clustering simulations and analysis.

Principal Components Analysis

To determine the major axes of skeletal variation in the dataset, we performed Principal Components Analysis (PCA) using the *FactoMineR* package in R. Phenotypes were size/sex/log-adjusted as necessary (see **Table 2.1**) and Z-scored. Missing data were imputed using the *imputePCA* command, then weighted PCA was performed using phenotype weights such that the total weight for each phenotype class was equal. We performed PCA on all phenotypes, excluding composite phenotypes where the non-composite phenotypes comprising the composite phenotype were also present. The first five principal components explained 18.4, 9.4, 4.9, 4.6,

and 4.4 percent of the phenotypic variance, respectively. The coordinates for each fish for the five largest principal components were extracted and QTL were mapped as described above.

Investigating biases in dominance

For simulations investigating QTL detection biases for dominance, two cases of QTL were compared: dominant QTL (d/a=1; heterozygotes have the same mean phenotype as the benthic B_1B_2 genotype) and additive QTL (d/a=0; heterozygous mean phenotype equals the mean of the M_1M_2 and B_1B_2 homozygous phenotypes). 400 samples for each value of dominance were used. Effect sizes span the boundaries of detection. Quantities in the simulation were based on the results of the analysis of the trait "DTP2," using Haley-Knott regression and a step size of 5 and a LOD threshold of 4.5. The simulations model a normally distributed trait, with QTL effect sizes ranging from 0 to 5, and a constant residual variance of 30 within each genotypic class, and assume no genotyping errors and no missing values. After detecting four QTL, all QTL were entered into a linear model in R/qtl. Obtained effect sizes for the four QTL ranged from LODs of 5 to 8, with a residual variance of about 28. Exploration with these numbers showed that effect sizes between 0 and 5 led to probabilities of detection ranging from about 0 to about 1.

Overlap with marine-freshwater divergent regions

The number of marine-freshwater divergent genomic regions that show evidence of repeated selection [the HMM or CSS signals of selection from (Jones et al. 2012b)] within the 1.5 LOD interval of each QTL was determined. To test for enrichment of signals of selection within various groups of QTL, the mean number of overlaps of the QTL group was divided by the mean number of overlaps from 1000 simulations of random placement of signals of selection across the genome. P values were calculated by comparing the mean number of signal of selection-QTL overlaps to a null distribution of simulated placements of signals of selection. For determining the number of signals of selection overlapping the three trait clusters, the following coordinates were used: 2.34-28.56 Mb, 1.71-14.68 Mb, and 0-8.94 Mb for chromosomes 4, 20, and 21, respectively. These physical coordinates correspond to the genetic range on each chromosome that spans all of the clustered QTL shown in **Figure 2.12**, based upon markers flanking the 1.5-LOD interval listed in **File 2.2**. These coordinates were also used to identify putative developmental regulatory genes within the trait clusters with Gene Ontology (GO) terms of "multicellular organismal development," "growth factor activity," or "regulation of transcription, DNA-dependent."

RESULTS

Major skeletal differences between marine and freshwater fish

The skeletons of Paxton benthic freshwater (PAXB) sticklebacks show multiple obvious reductions in external bones compared to Japanese Pacific marine (JAMA) animals, including reduced size and number of armor plates, loss of pelvic fins, and reduced length of dorsal spines (**Figure 2.1A,B**). In addition, computerized tomography (**Figure 2.1C-H**) revealed a mixture of both regressive ("loss") and constructive ("gain") traits in the skull and internal branchial skeleton of PAXB sticklebacks. The derived freshwater fish show dramatic reductions in the number and length of gill rakers (**Figure 2.1C,D**), as expected based on previous studies (Kitano 2007; McPhail 1992). However, PAXB branchial bones, especially the first epibranchial, have

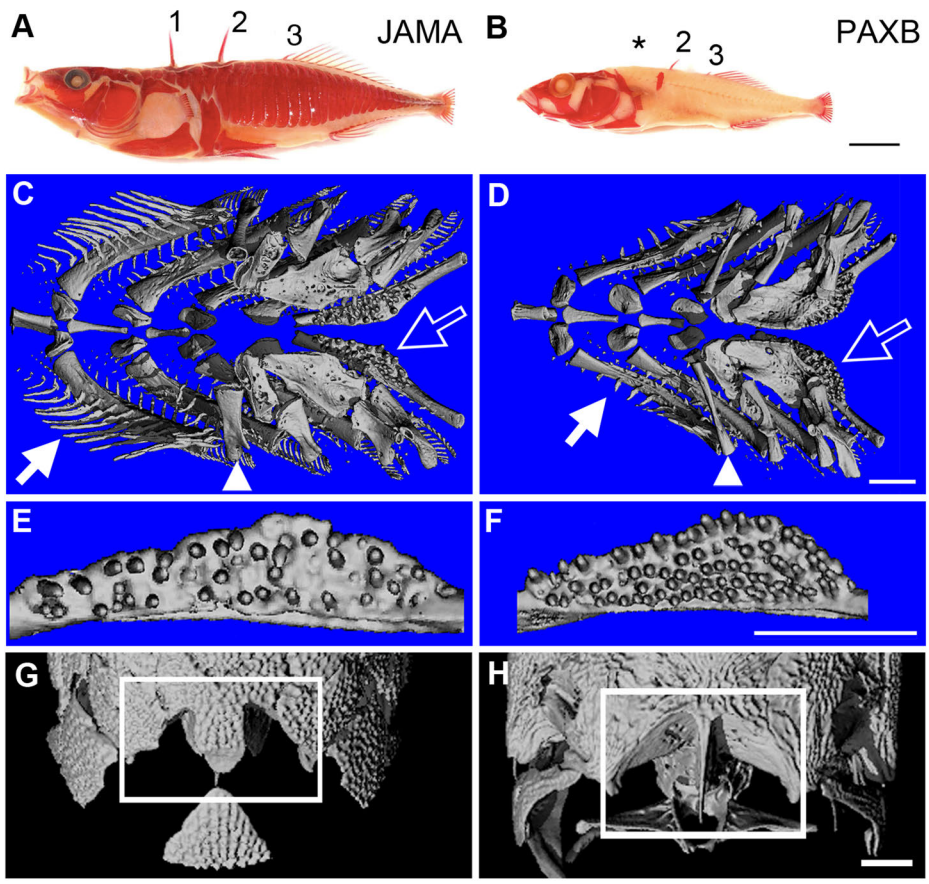


Figure 2.1: Evolved skeletal differences between marine and benthic sticklebacks
Skeletal morphology revealed by Alizarin red staining (A-B) or micro-computerized tomography (C-H) of adult Japanese Pacific marine (JAMA; A, C, E, G) and Paxton benthic freshwater (PAXB; B, D, F, H) fish. (A,B) Lateral views of bone-stained adults reveal differences in dorsal spine lengths. Three dorsal spines are numbered in A. The first dorsal spine is missing (asterisk) in this PAXB fish. (C-D) Dorsal views of branchial skeletons reveal fewer and less densely spaced gill rakers (white arrows) and longer branchial bones (white arrowheads) in PAXB. (E,F) Ventral pharyngeal toothplates (labeled with open white arrow in C,D) reveal higher tooth number in PAXB. (G,H) Dorsal views of skulls reveal differences in the size and shape of the supraoccipital crest (white boxes). The JAMA supraoccipital crest is shorter and wider but larger in area, while the PAXB supraoccipital crest is longer and narrower, smaller in area, and flanked by more robust insertion points for the epaxial muscles. Scale bars = 1 mm.

increased in length compared to their marine counterparts (arrowheads **Figure 2.1C,D**). In addition, PAXB fish have roughly twice the number of ventral pharyngeal teeth seen in JAMA marine animals (**Figure 2.1E,F**). Compared to marine animals, PAXB fish also show a longer and thinner supraoccipital crest, a posterior process on the supraoccipital bone at the back of the skull that serves as the insertion point for muscles involved in buccal cavity opening (**Figure 2.1G-H**). Increased size of these muscles is a characteristic feature of PAXB fish, and is thought to increase force generation and suction pressure for feeding on attached littoral prey items (McGee et al. 2013).

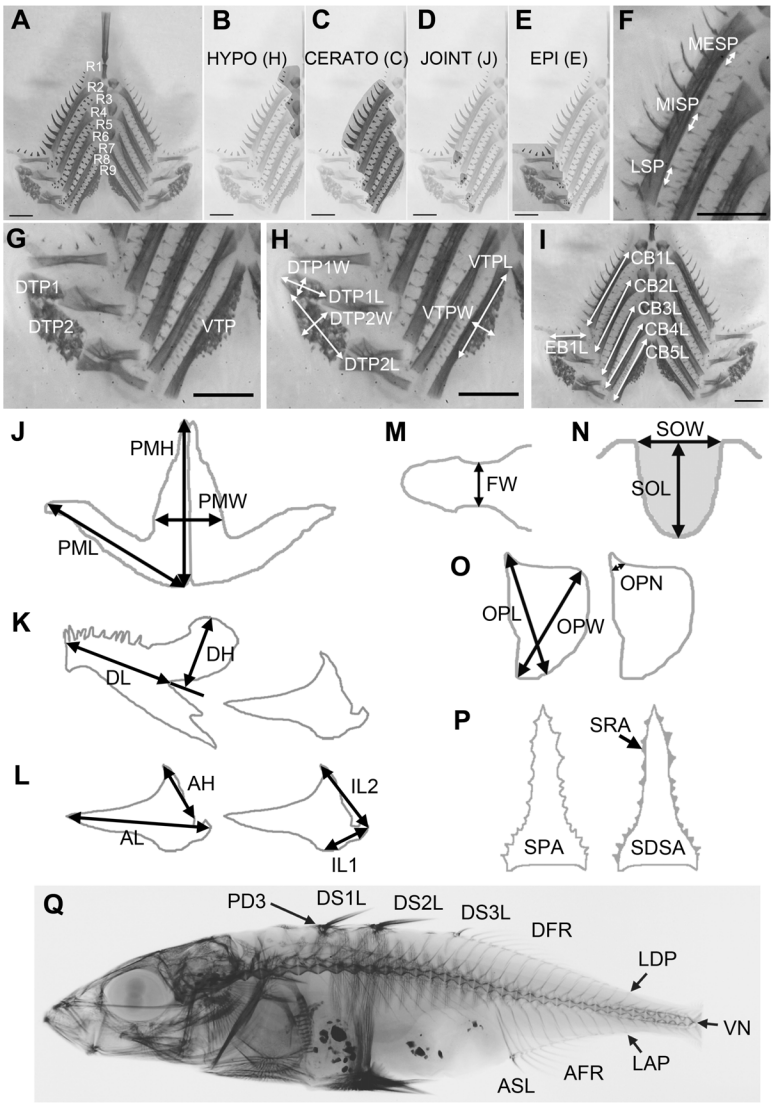


Figure 2.2: Additional skeletal phenotypes
(A-I) Branchial skeleton from Alizarin red-stained F2 fish, dissected and flattened into two-dimensional prep by single incision along the dorsal midline. All branchial bones, including pharyngeal toothplates and gill rakers are readily visible. Scale bars = 1 mm. (A) Nine rows of gill rakers (R1-9), pseudocolored black on left half, line the anterior and posterior faces of each segment, except for the last segment which lacks gill rakers on its posterior side. (B-E) Dorsal-ventral domains of gill rakers, defined by edges of branchial bones from ventromedial to dorsal: hypo (B), cerato (C), joint (D), and epi (E) gill raker domains. (F) Three inter-raker spacing measurements in lateral (LSP), middle (MISP), and medial (MESP) regions of row 2 cerato rakers. (G) Three pharyngeal toothplates (two dorsal and one ventral) are present on each side. (H) Pharyngeal toothplate lengths and widths. (I) Branchial bone lengths. (J) Dorsal view of premaxilla bone (upper jaw) traits. (K) Lateral view of dentary bone (lower jaw, shown with articular posteriorly). (L) Lateral view of articular bone traits (M) Dorsal view of frontal bone width (or interorbital distance). (N) Dorsal view of caudal end of supraoccipital bone, with supraoccipital crest area shaded grey (also see **Figure 2.1G-H**). (O) Lateral view of opercle bone traits. (P) Second dorsal spine area and serration traits. (Q) X-ray showing spine, median fin ray, and vertebral position landmarks. Abbreviations (defined in **Table 2.1**): AFR, number of anal fin rays; AH, articular height; AL, articular length; ASL, anal spine length; CB1L, ceratobranchial 1 length; CB2L, ceratobranchial 2 length; CB3L, ceratobranchial 3 length; CB4L, ceratobranchial 4 length; CB5L, ceratobranchial 5 length; DFR, dorsal fin ray number; DH, dentary height; DL, dentary length; DS1L, dorsal spine 1 length; DS2L, dorsal spine 2 length; DS3L, dorsal spine 3 length; DTP1, dorsal toothplate 1; DTP1L, dorsal toothplate 1 length; DTP1W, dorsal toothplate 1 width; DTP2, dorsal toothplate 2; DTP2L, dorsal toothplate 2 length; DTP2W, dorsal toothplate 2 width; EB1L, epibranchial 1 length; FW, frontal width; IL1, in-lever 1 of the articular; IL2, in-lever 2 of the articular; LAP, vertebrae number of last anal pterygiophore; LDP, vertebrae number of last dorsal pterygiophore; LSP, lateral row 2 raker spacing; MESP, medial row 2 raker spacing; MISP, middle row 2 raker spacing; OPL, opercle length; OPN, opercle neck width; OPW, opercle width; PD3, vertebrae number of third predorsal pterygiophore; PMH, premaxilla height; PML, premaxilla length; PMW, premaxilla width; R1-9, rows 1-9 of gill rakers; SDSA, smoothened dorsal spine 2 area; SOL, supraoccipital crest length; SOW, supraoccipital crest width; SPA, dorsal spine 2 area; SRA, spine 2 serration area; VN, total vertebrae number; VTP, ventral toothplate; VTPL, ventral toothplate length; VTPW, ventral toothplate width.

Trait	Description	Class	Model	Number of QTL
R1E	Row 1 epi raker number	raker	Raw	1
R2E	Row 2 epi raker number	raker	SL lm	1
R3E	Row 3 epi raker number	raker	Raw	1
R4E	Row 4 epi raker number	raker	Glm	0
R5E	Row 5 epi raker number	raker	Raw	0
R6E	Row 6 epi raker number	raker	Sex lm	1
R7E	Row 7 epi raker number	raker	Raw	0
R2J	Row 2 joint raker number	raker	SL lm	3
R4J	Row 4 joint raker number	raker	SL lm	1
R6J	Row 6 joint raker number	raker	Sex lm	2
R7J	Row 7 joint raker number	raker	Sex lm	0
R8J	Row 8 joint raker number	raker	Sex lm	0
R1C	Row 1 cerato raker number	raker	Glm	3
R2C	Row 2 cerato raker number	raker	Sex lm	2
R3C	Row 3 cerato raker number	raker	Glm	4
R4C	Row 4 cerato raker number	raker	Glm	4
R5C	Row 5 cerato raker number	raker	SL lm	2
R6C	Row 6 cerato raker number	raker	Glm	3
R7C	Row 7 cerato raker number	raker	Glm	1
R8C	Row 8 cerato raker number	raker	Sex lm	4
R9C	Row 9 cerato raker number	raker	SL lm	2
R1H	Row 1 hypo raker number	raker	Glm	0
R3H	Row 3 hypo raker number	raker	Raw	1
R5H	Row 5 hypo raker number	raker	SL lm	1
R7H	Row 7 hypo raker number	raker	Sex lm	1
E	Epi raker number	raker	Glm	1
J	Joint raker number	raker	Glm	4
C	Cerato raker number	raker	Glm	5
H	Hypo raker number	raker	Glm	6
R1	Row 1 raker number	raker	SL lm	5
R2	Row 2 raker number	raker	Sex lm	2
R3	Row 3 raker number	raker	Raw	5
R4	Row 4 raker number	raker	Sex lm	2
R5	Row 5 raker number	raker	Raw	2
R6	Row 6 raker number	raker	SL lm	2
R7	Row 7 raker number	raker	Glm	0
R8	Row 8 raker number	raker	Raw	1
BA1	Branchial arch 1 raker number	raker	Sex lm	6
BA2	Branchial arch 2 raker number	raker	Sex lm	6
BA3	Branchial arch 3 raker number	raker	Glm	4
BA4	Branchial arch 4 raker number	raker	SL lm	4
ODD	Odd row raker number	raker	Glm	5
EVEN	Even row raker number	raker	Sex lm	5
ALL	All raker number	raker	Glm	8
LSP	Lateral raker spacing	raker	SL lm	5
MISP	Middle raker spacing	raker	Glm	5
MESP	Medial raker spacing	raker	SL lm	1
DTP1	Dorsal toothplate 1 tooth number	teeth	Raw	6
DTP2	Dorsal toothplate 2 tooth number	teeth	Raw	7
VTP	Ventral toothplate tooth number	teeth	SL lm	5
PMT	Premaxilla tooth number	teeth	Glm	0
PMTR	Number of tooth rows on premaxilla	teeth	Glm	0
DTP1L	Dorsal toothplate 1 length	teeth	Glm	5
DTP1W	Dorsal toothplate 1 width	teeth	Glm	3
DTP2L	Dorsal toothplate 2 length	teeth	SL lm	6

DTP2W	Dorsal toothplate 2 width	teeth	Glm	5
VTPL	Ventral toothplate length	teeth	SL lm	3
VTPW	Ventral toothplate width	teeth	Glm	5
EB1L	Epibranchial 1 length	branchial	Glm	8
CB1L	Ceratobranchial 1 length	branchial	Glm	9
CB2L	Ceratobranchial 2 length	branchial	Glm	8
CB3L	Ceratobranchial 3 length	branchial	Glm	6
CB4L	Ceratobranchial 4 length	branchial	Glm	10
CB5L	Ceratobranchial 5 length	branchial	Glm	9
PML	Premaxilla length	jaw	Glm	9
PMW	Premaxilla width	jaw	Glm	2
PMH	Premaxilla height	jaw	Glm	5
DL	Dentary length	jaw	Glm	4
DH	Dentary height	jaw	Glm	4
AL	Articular length	jaw	Glm	7
AH	Articular height	jaw	Glm	3
IL1	In-lever 1 of articular length	jaw	Glm	3
IL2	In-lever 2 of articular length	jaw	Glm	3
FW	Frontal width	skull	Glm	7
SOL	Supraoccipital crest length	skull	SL lm	2
SOW	Supraoccipital crest width	skull	SL lm	4
SOA	Supraoccipital crest area	skull	SL lm	6
OPL	Opercle length	opercle	Glm	3
OPW	Opercle width	opercle	Glm	4
OPN	Opercle neck width	opercle	Sex lm	2
DS1L	Dorsal spine 1 length	median fin	SL lm	8
DS2L	Dorsal spine 2 length	median fin	Glm	9
DS3L	Dorsal spine 3 length	median fin	Glm	5
ASL	Anal spine length	median fin	Glm	3
SR	Serration number on dorsal spine 2	median fin	Glm	2
SRA	Serration area on dorsal spine 2	median fin	Glm	2
SPA	Dorsal spine 2 area	median fin	SL lm	2
SDS2A	Smoothened dorsal spine 2 area (SPA-SRA)	median fin	SL lm	1
DFR	Dorsal fin ray number	median fin	Sex lm	1
AFR	Anal fin ray number	median fin	Glm	1
PD	Predorsal pterygiophore number	median fin	Sex lm	0
PE	Non-ray-bearing postdorsal pterygiophore number	median fin	Raw	1
PN	Non-ray-bearing postanal pterygiophore number	median fin	Raw	1
TPDP	Total postdorsal pterygiophore number	median fin	Sex lm	1
TAP	Total postanal pterygiophore number	median fin	Sex lm	2
TDP	Total dorsal pterygiophore number	median fin	Sex lm	1
PD3	Third predorsal pterygiophore position	vertebrae	Raw	2
PD4	Fourth predorsal pterygiophore position	vertebrae	Raw	0
PD5	Fifth predorsal pterygiophore position	vertebrae	Sex lm	0
FDP	First postdorsal pterygiophore position	vertebrae	Sex lm	0
LDP	Last postdorsal pterygiophore position	vertebrae	Raw	4
FAP	First postanal pterygiophore position	vertebrae	Sex lm	0
LAP	Last postanal pterygiophore position	vertebrae	Raw	5
FHS	First hemal spine position	vertebrae	Sex lm	0
LDFR	Last dorsal fin ray position	vertebrae	SL lm	0
LAFR	Last anal fin ray position	vertebrae	Raw	3
VN	Vertebrae number	vertebrae	Sex lm	2
AVN	Abdominal vertebrae number	vertebrae	Sex lm	0
CVN	Caudal vertebrae number	vertebrae	Sex lm	1
VR	Vertebrae ratio (AVN/CVN)	vertebrae	Sex lm	0

Table 2.1 (previous page): Trait descriptions and transformations
Depending on the relationship of phenotype with standard length (SL) and sex, the phenotype used for QTL mapping was unadjusted (raw), residuals from a SL regression (SL lm), residuals from one-way ANOVA using sex as a factor (Sex lm), or residuals from a SL+Sex General Linear Model ANOVA (Glm). Units are meristic counts for number traits, millimeters for linear measurements, millimeters^2 for area traits, vertebrae number for vertebral position traits or unitless for VR. For 92 of 110 traits, at least one QTL was detected.

To investigate the genetic architecture underlying the evolution of these trophic, armor, and other skeletal traits, we used genome-wide linkage mapping in a large marine x benthic F2 genetic cross previously studied for lateral plate, pelvic spine, and pigmentation patterning (Colosimo et al. 2005; Colosimo et al. 2004; Miller et al. 2007; Shapiro et al. 2004). We phenotyped 370 full-sibling F2 fish from this cross, then mapped quantitative trait loci (QTL) influencing 110 different skeletal phenotypes (**Table 2.1**), including a large number of traits in the branchial skeleton (**Figure 2.2A-I**); multiple aspects of jaw, skull, and opercle morphology (**Figure 2.2J-O**); dorsal and anal spine lengths, and degree of spine serrations (**Figure 2.2P-Q**); and several median fin and vertebral traits (Ahn and Gibson 1999) proposed to be important for freshwater adaptation (**Figure 2.2Q**).

Most traits are sexually dimorphic
As most traits (93 of 110) are size and/or sex dependent, we systematically corrected for size and/or sex (**Table 2.1**). In total, 72 traits showed significant differences between the sexes. The traits that show dimorphism, and the direction of dimorphism, were largely consistent with previous reports. For example, males had bigger jaws, more oral teeth, more vertebrae, fewer abdominal vertebrae, more dorsal and anal fin rays, more pterygiophores, and dorsal spines with more serrations in both the current study and in previous studies (Caldecutt et al. 2001; Kitano 2007; Kitano et al. 2009; Lindsey 1962; McPhail 1992; Reimchen and Nelson 1987).

Most QTL are anatomically specific
Raw or corrected trait values (**Table 2.2**) and a genome-wide linkage map were used to map QTL with a multiple-QTL mapping approach in R/qtl (Broman and Sen 2009). This analysis identified 342 total QTL for 92 of the 110 traits (**File 2.2**). Based on segmental homology and likely embryonic origin, we divided the 110 skeletal traits into eight trait classes: rakers, teeth, branchial bones, jaw, skull, opercle, median fin, and vertebrae (**Figure 2.3**). In cases where a particular chromosome region had an effect on multiple phenotypes within the same trait class, it is parsimonious to assume that a single underlying QTL affected the trait across several domains. Thus, for our analysis of the properties of these QTL, we conservatively considered such QTL only once. To define a minimally non-redundant set of QTL by trait class, we included only QTL whose 1.5 LOD intervals (an approximate 95% confidence interval, Dupuis and Siegmund 1999) did not overlap. This filtering defined a set of 118 QTL across the eight trait classes (**Table 2.2, Figure 2.3**). The overall distribution of percent variance explained (PVE) of these QTL consisted of many small-effect QTL and few large-effect QTL (**Figure 2.4**), similar to a previous stickleback shape QTL study (Albert et al. 2008). For trait classes that had multiple serially homologous anatomical domains or elements (raker number, teeth, branchial, jaw, and spine

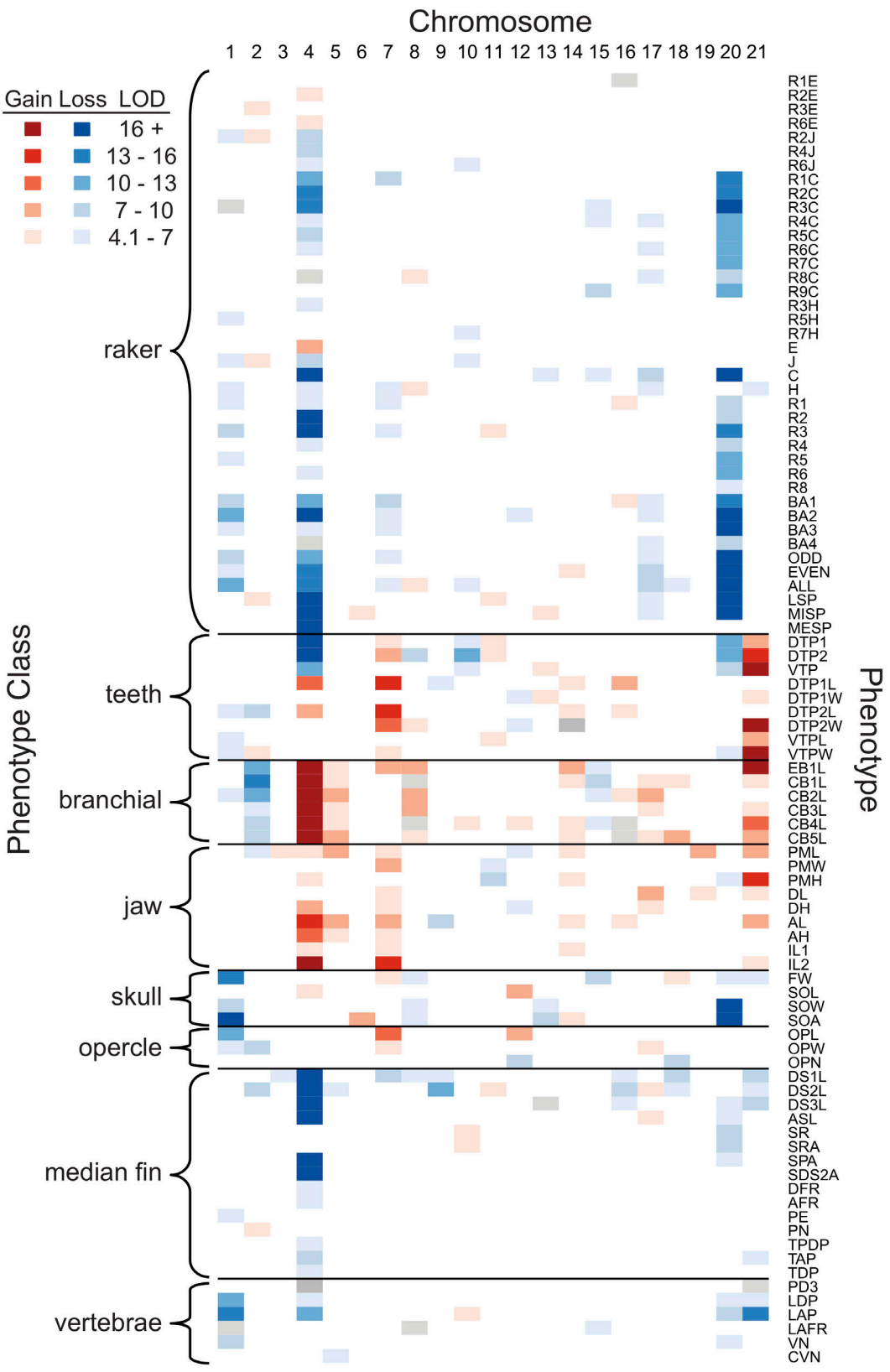

Chromosome
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21
Gain Loss LOD
16 +
13 - 16
10 - 13
7 - 10
4.1 - 7
Phenotype Class
raker
teeth
branchial
jaw
skull
opercle
median fin
vertebrae
Phenotype
R1E
R2E
R3E
R6E
R2J
R4J
R6J
R1C
R2C
R3C
R4C
R5C
R6C
R7C
R8C
R9C
R3H
R5H
R7H
E
J
C
H
R1
R2
R3
R4
R5
R6
R8
BA1
BA2
BA3
BA4
ODD
EVEN
ALL
LSP
MISP
MESP
DTP1
DTP2
VTP
DTP1L
DTP1W
DTP2L
DTP2W
VTPL
VTPW
EB1L
CB1L
CB2L
CB3L
CB4L
CB5L
PML
PMW
PMH
DL
DH
AL
AH
IL1
IL2
FW
SOL
SOW
SOA
OPL
OPW
OPN
DS1L
DS2L
DS3L
ASL
SR
SRA
SPA
SDS2A
DFR
AFR
PE
PN
TPDP
TAP
TDP
PD3
LDP
LAP
LAFR
VN
CVN

Figure 2.3 (previous page): Genome-wide overview of detected skeletal QTL
Classes of traits are grouped on the left, and individual traits listed on the right. Abbreviations are defined in Figure 2.2 legend and Table 2.1. For each trait, the LOD score for each QTL on each chromosome is indicated by the heat map shown in the upper left, with "gain" traits (benthic allele confers more or bigger bones) colored red and "loss" traits (benthic allele confers fewer or smaller bones) colored blue. Heterotic QTL (homozygous marine and benthic F2 fish do not differ significantly in phenotype by two-tailed t-test) and vertebral position QTL (trait is neither loss nor gain) are shaded gray.

classes), we asked whether QTL controlling these traits had anatomically regional or global effects. Of this set of QTL, a large majority (76%) affected only a subset of the possible domains while 24% affected all domains (**Figure 2.5, Table 2.3**).

Gill rakers

Gill rakers are present in nine rows from anterior to posterior and in four regions from dorsal to ventral (**Figure 2.2A-E**). Given well-established regional developmental genetic control of pharyngeal segments along the anterior/posterior and dorsal/ventral axes (e.g. Hox gene control of segmental identity and Dlx gene control of D/V patterning, reviewed in Minoux and Rijli 2010), we scored and mapped gill raker number separately for each individual A/P and D/V domain, as well as composite phenotypes that represented putative developmental domains [e.g. all ventral (cerato) rakers]. Overall, we found 23 QTL controlling gill raker number or spacing. The raker QTL displayed a high degree of regional specificity, with no QTL having significant effects in all possible domains (**Table 2.3, Figure 2.5**). For example, the largest effect gill raker QTL mapped to chromosome 20, and had strictly ventral (cerato)-specific effects on raker number, while chromosomes 10 and 7 had regional effects on joint and hypo rakers, respectively (**Figure 2.5**). Additionally, on chromosome 4, the second largest effect raker QTL mapped to one end of the chromosome and was largely ventral (cerato)-specific, whereas a non-overlapping region more centrally located on chromosome 4 had a dorsal (epi)-specific effect on raker number (see below). Both the chromosome 4 and 20 large-effect gill raker number QTL controlled gill raker number at least in part through controlling gill raker spacing, as row 2 inter-raker spacing mapped strongly to an overlapping region of chromosomes 4 and 20. Four raker QTL (right chromosome 4, chromosomes 6, 11, and 13) had a raker spacing measurement map more strongly than any raker number trait.

Teeth

We identified 20 QTL on 14 chromosomes controlling tooth number or toothplate size (**Figure 2.2G-H**), including two QTL with large effects on tooth number (**Table 2.2**). A QTL on chromosome 21 explained nearly a third of the variance in ventral pharyngeal tooth number, while a QTL on chromosome 4 had large effects on dorsal pharyngeal tooth number. All tooth number QTL had stronger effects (>2 LOD units difference) on either dorsal or ventral pharyngeal tooth number. Genetic control of oral and pharyngeal tooth variation appeared largely independent. Although oral tooth number in the upper jaw was sexually dimorphic as previously reported for several wild stickleback populations (Caldecutt et al. 2001), no autosomal QTL for oral tooth number were detected (**Table 2.1**).

Trait	Class	LG	Marker	LOD	PVE	Mean +/- Standard Error (n)			
						M_1M_2	M_1B_1	M_2B_2	B_1B_2
R3C	raker	1	Stn2	4.7	3.8	10.8 +/- 0.085	10.85 +/- 0.086	10.85 +/- 0.084	10.59 +/- 0.1
BA2	raker	1	Stn5	10.7	7.5	30.4 +/- 0.15	29.9 +/- 0.13	29.8 +/- 0.13	29.2 +/- 0.19
R3E	raker	2	Stn17	5.1	6.4	1.387 +/- 0.057	1.63 +/- 0.061	1.521 +/- 0.056	1.765 +/- 0.05
E	raker	4	Stn38	8.6	10.6	11.83 +/- 0.2	12.78 +/- 0.17	13.06 +/- 0.17	13.55 +/- 0.17
LSP	raker	4	Stn47	27.7	22.6	0.254 +/- 0.0019	0.264 +/- 0.0025	0.2721 +/- 0.0026	0.2885 +/- 0.0024
MISP	raker	6	Stn62	4.9	3.4	0.2921 +/- 0.0025	0.282 +/- 0.0025	0.2841 +/- 0.0021	0.2851 +/- 0.0019
H	raker	7	Stn258	6.1	5.7	9.113 +/- 0.097	8.975 +/- 0.11	8.682 +/- 0.11	8.555 +/- 0.11
R1C	raker	7	Stn75	7.8	7.0	11.39 +/- 0.082	11.07 +/- 0.078	11.23 +/- 0.083	10.8 +/- 0.085
R8C	raker	8	Stn300	5.3	5.5	10.71 +/- 0.081	10.68 +/- 0.081	10.82 +/- 0.074	10.98 +/- 0.087
H	raker	8	Stn95	5.0	4.6	8.679 +/- 0.11	9.03 +/- 0.11	8.684 +/- 0.11	8.965 +/- 0.11
R6J	raker	10	Stn124	5.6	6.4	2.968 +/- 0.059	2.704 +/- 0.058	2.879 +/- 0.062	2.632 +/- 0.056
LSP	raker	11	Stn244	6.0	4.2	0.278 +/- 0.003	0.2694 +/- 0.003	0.2685 +/- 0.002	0.2656 +/- 0.0031
BA2	raker	12	Nemo	4.2	2.8	30.06 +/- 0.16	29.91 +/- 0.17	29.77 +/- 0.16	29.67 +/- 0.14
MISP	raker	13	Stn155	5.2	3.6	0.2873 +/- 0.0018	0.2837 +/- 0.0022	0.2887 +/- 0.0024	0.2828 +/- 0.0022
EVEN	raker	14	Stn166	4.5	3.4	57.11 +/- 0.28	57.6 +/- 0.28	57.66 +/- 0.28	58.28 +/- 0.27
R9C	raker	15	Stn230	7.3	7.8	12.98 +/- 0.081	12.82 +/- 0.083	13.05 +/- 0.089	12.42 +/- 0.079
BA1	raker	16	Stn176	6.6	5.0	33.75 +/- 0.21	34.6 +/- 0.19	33.55 +/- 0.18	34.3 +/- 0.19
H	raker	17	Stn323	5.5	5.1	9.067 +/- 0.11	9.017 +/- 0.1	8.905 +/- 0.11	8.419 +/- 0.11
ALL	raker	17	Stn205	8.4	5.1	137.7 +/- 0.49	137.5 +/- 0.54	136 +/- 0.69	134.4 +/- 0.61
ALL	raker	18	Stn305	4.6	2.8	137.6 +/- 0.9	135.1 +/- 0.63	136.6 +/- 0.49	136.7 +/- 0.54
C	raker	20	Stn215	32.9	25.3	105.6 +/- 0.37	103.4 +/- 0.44	101.5 +/- 0.39	98.88 +/- 0.52
R8C	raker	20	Gac1125	9.0	9.5	11.03 +/- 0.065	10.9 +/- 0.083	10.58 +/- 0.083	10.56 +/- 0.087
H	raker	21	Stn223	4.6	4.2	9.128 +/- 0.1	8.858 +/- 0.11	8.676 +/- 0.11	8.682 +/- 0.11
VTPW	teeth	1	Stn272	6.3	4.8	0.4276 +/- 0.0059	0.4197 +/- 0.0038	0.4191 +/- 0.0044	0.398 +/- 0.0045
DTP2L	teeth	1	Stn13	4.5	3.7	1.493 +/- 0.0077	1.48 +/- 0.0077	1.469 +/- 0.0074	1.471 +/- 0.0074
DTP2L	teeth	2	Stn259	7.8	6.7	1.507 +/- 0.0079	1.478 +/- 0.008	1.47 +/- 0.0068	1.463 +/- 0.0071
VTPW	teeth	2	Stn24	6.5	5.0	0.4169 +/- 0.0046	0.4202 +/- 0.0053	0.4082 +/- 0.004	0.4245 +/- 0.0053
DTP2	teeth	4	Eda	34.9	24.2	53.96 +/- 0.68	49.06 +/- 0.66	47.3 +/- 0.66	44.68 +/- 0.61
DTP1	teeth	7	Stn72	6.9	5.1	15.63 +/- 0.29	16.59 +/- 0.34	16.27 +/- 0.29	17.53 +/- 0.35
DTP2L	teeth	7	Stn75	15.2	13.7	1.444 +/- 0.0071	1.47 +/- 0.0062	1.488 +/- 0.0079	1.515 +/- 0.0076
DTP2	teeth	8	Stn300	7.1	4.1	50.06 +/- 0.67	49.11 +/- 0.71	47.8 +/- 0.69	47.67 +/- 0.89
DTP2W	teeth	8	Stn95	5.1	4.2	0.662 +/- 0.0054	0.6832 +/- 0.0037	0.6658 +/- 0.004	0.6852 +/- 0.0048
DTP1L	teeth	9	Stn102	5.6	4.7	0.8897 +/- 0.0056	0.8756 +/- 0.0055	0.866 +/- 0.0057	0.8527 +/- 0.0058
DTP2	teeth	10	Stn211	12.7	7.6	49.97 +/- 0.61	49.28 +/- 0.84	49.41 +/- 0.88	46.03 +/- 0.58
DTP1	teeth	11	Dlx3a	6.1	4.5	15.82 +/- 0.32	15.67 +/- 0.29	16.63 +/- 0.29	17.43 +/- 0.35
DTP2W	teeth	12	Stn232	5.4	4.4	0.6858 +/- 0.0042	0.6727 +/- 0.0042	0.6802 +/- 0.0053	0.6631 +/- 0.0042
DTP1W	teeth	12	Stn142	4.7	5.2	0.4122 +/- 0.0043	0.3963 +/- 0.0033	0.3958 +/- 0.0037	0.3915 +/- 0.0039
VTP	teeth	13	Stn153	5.5	3.9	45.24 +/- 0.7	47.98 +/- 0.81	46.41 +/- 0.8	48.34 +/- 0.7
DTP2W	teeth	14	Stn198	7.0	5.8	0.6828 +/- 0.0047	0.6637 +/- 0.0038	0.6706 +/- 0.0047	0.685 +/- 0.0047
DTP1L	teeth	14	Stn237	6.6	5.6	0.8505 +/- 0.0051	0.8798 +/- 0.0059	0.8716 +/- 0.0062	0.882 +/- 0.0055
DTP1L	teeth	16	Stn178	7.4	6.4	0.8617 +/- 0.0065	0.8639 +/- 0.0059	0.873 +/- 0.0051	0.8844 +/- 0.0058
DTP2	teeth	20	Stn215	12.8	7.6	50.99 +/- 0.61	49.35 +/- 0.74	48.14 +/- 0.7	44.82 +/- 0.8
VTP	teeth	21	Stn422	31.0	26.2	42.28 +/- 0.56	46.48 +/- 0.69	46.93 +/- 0.58	52.48 +/- 0.76
CB2L	branchial	1	Stn248	6.2	4.2	2.821 +/- 0.012	2.804 +/- 0.011	2.789 +/- 0.011	2.786 +/- 0.011
EB1L	branchial	2	Stn21	11.6	6.7	1.308 +/- 0.01	1.288 +/- 0.0091	1.269 +/- 0.0085	1.257 +/- 0.0083
CB1L	branchial	2	Stn268	14.9	10.4	3.056 +/- 0.011	3.019 +/- 0.013	3.007 +/- 0.01	2.973 +/- 0.012
CB4L	branchial	4	Stn45	29.3	19.1	2.514 +/- 0.0084	2.555 +/- 0.01	2.574 +/- 0.01	2.64 +/- 0.0087

CB5L	branchial	5	Stn52	8.6	5.9	2.46 +/- 0.01	2.49 +/- 0.01	2.5 +/- 0.01	2.533 +/- 0.0087
EB1L	branchial	5	Stn312	6.6	3.7	1.254 +/- 0.011	1.269 +/- 0.0087	1.281 +/- 0.0077	1.303 +/- 0.0089
EB1L	branchial	7	Stn71	7.4	4.2	1.262 +/- 0.0085	1.265 +/- 0.0089	1.276 +/- 0.0091	1.312 +/- 0.0092
EB1L	branchial	8	Stn239	9.6	5.5	1.244 +/- 0.0083	1.3 +/- 0.0092	1.267 +/- 0.0092	1.296 +/- 0.0088
CB4L	branchial	10	Edn1	6.0	3.3	2.551 +/- 0.0088	2.547 +/- 0.012	2.583 +/- 0.0096	2.607 +/- 0.011
CB4L	branchial	12	Stn287	6.2	3.5	2.54 +/- 0.01	2.574 +/- 0.0095	2.572 +/- 0.011	2.6 +/- 0.01
EB1L	branchial	14	Stn163	7.1	4.0	1.259 +/- 0.0075	1.283 +/- 0.0093	1.266 +/- 0.0096	1.303 +/- 0.0096
CB1L	branchial	15	Stn173	7.4	4.9	3.04 +/- 0.014	3.027 +/- 0.0094	2.998 +/- 0.011	2.973 +/- 0.013
CB4L	branchial	16	Edar	6.9	3.9	2.558 +/- 0.011	2.558 +/- 0.011	2.597 +/- 0.009	2.574 +/- 0.011
CB2L	branchial	17	Stn325	7.1	4.9	2.773 +/- 0.012	2.811 +/- 0.0097	2.783 +/- 0.013	2.829 +/- 0.011
CB5L	branchial	18	Stn280	9.8	6.7	2.454 +/- 0.011	2.508 +/- 0.011	2.501 +/- 0.0083	2.517 +/- 0.0097
EB1L	branchial	21	Stn219	24.4	15.5	1.231 +/- 0.0079	1.273 +/- 0.009	1.289 +/- 0.0081	1.322 +/- 0.0088
PML	jaw	2	Stn24	5.3	3.7	2.39 +/- 0.013	2.405 +/- 0.014	2.361 +/- 0.012	2.37 +/- 0.015
PML	jaw	3	Stn31	5.2	3.7	2.344 +/- 0.012	2.405 +/- 0.015	2.363 +/- 0.012	2.415 +/- 0.014
IL2	jaw	4	Stn45	17.9	16.9	1.264 +/- 0.0082	1.297 +/- 0.008	1.302 +/- 0.009	1.366 +/- 0.0091
PML	jaw	5	Stn52	8.3	6.0	2.333 +/- 0.013	2.403 +/- 0.014	2.389 +/- 0.013	2.396 +/- 0.013
IL2	jaw	7	Stn75	13.1	11.9	1.267 +/- 0.0084	1.318 +/- 0.0085	1.305 +/- 0.0091	1.351 +/- 0.0099
PMW	jaw	7	Stn257	7.4	8.3	0.2965 +/- 0.0039	0.3029 +/- 0.004	0.3133 +/- 0.0042	0.3222 +/- 0.0035
AL	jaw	9	Stn102	8.0	5.6	2.329 +/- 0.013	2.304 +/- 0.014	2.289 +/- 0.014	2.22 +/- 0.013
PMH	jaw	11	Stn224	7.8	6.9	2.427 +/- 0.012	2.387 +/- 0.012	2.37 +/- 0.011	2.355 +/- 0.011
DH	jaw	12	Stn318	6.8	6.9	1.313 +/- 0.0085	1.279 +/- 0.0078	1.283 +/- 0.0096	1.249 +/- 0.0083
PMH	jaw	14	Stn331	5.9	5.2	2.345 +/- 0.013	2.399 +/- 0.011	2.372 +/- 0.012	2.419 +/- 0.01
AL	jaw	16	Dlx1/2	6.1	4.3	2.267 +/- 0.014	2.26 +/- 0.015	2.299 +/- 0.012	2.312 +/- 0.015
DL	jaw	17	Stn325	7.2	7.5	1.729 +/- 0.014	1.743 +/- 0.011	1.771 +/- 0.012	1.802 +/- 0.012
PML	jaw	19	Stn194	7.6	5.4	2.358 +/- 0.012	2.398 +/- 0.013	2.338 +/- 0.012	2.417 +/- 0.014
PMH	jaw	20	Stn216	5.2	4.5	2.408 +/- 0.011	2.396 +/- 0.012	2.373 +/- 0.012	2.34 +/- 0.012
PMH	jaw	21	Stn422	13.4	12.3	2.322 +/- 0.011	2.386 +/- 0.011	2.406 +/- 0.012	2.424 +/- 0.011
SOA	skull	1	Stn272	17.7	10.5	313.6 +/- 13	278.2 +/- 9.5	250.9 +/- 11	215.9 +/- 11
SOL	skull	4	Stn253	4.9	5.6	1.272 +/- 0.012	1.304 +/- 0.013	1.297 +/- 0.012	1.354 +/- 0.012
SOA	skull	6	Stn279	8.4	4.7	242.8 +/- 12	290.1 +/- 13	239.8 +/- 9.9	293.2 +/- 11
FW	skull	7	Stn72	6.8	5.4	2.4 +/- 0.014	2.387 +/- 0.015	2.453 +/- 0.012	2.474 +/- 0.016
SOA	skull	8	Stn88	6.3	3.5	289.3 +/- 9.9	269.6 +/- 10	267.5 +/- 12	234.2 +/- 13
SOL	skull	12	Stn318	7.8	9.0	1.253 +/- 0.013	1.298 +/- 0.012	1.339 +/- 0.013	1.34 +/- 0.011
SOA	skull	13	Stn156	8.3	4.7	298.9 +/- 12	268.9 +/- 11	264.3 +/- 12	236.7 +/- 10
SOA	skull	14	Stn167	6.2	3.4	265.8 +/- 11	258.9 +/- 10	254.5 +/- 13	282.9 +/- 11
FW	skull	15	Stn230	7.0	5.6	2.465 +/- 0.015	2.408 +/- 0.014	2.436 +/- 0.015	2.41 +/- 0.012
FW	skull	18	Stn280	4.8	3.8	2.4 +/- 0.014	2.414 +/- 0.015	2.435 +/- 0.014	2.457 +/- 0.014
SOA	skull	20	Stn216	43.0	30.4	340.2 +/- 11	257.6 +/- 9.4	251.1 +/- 9.6	183.2 +/- 7.2
FW	skull	21	Stn223	6.2	4.9	2.457 +/- 0.014	2.4 +/- 0.014	2.457 +/- 0.016	2.4 +/- 0.013
OPL	opercle	1	Stn272	11.4	10.9	3.928 +/- 0.019	3.919 +/- 0.015	3.858 +/- 0.017	3.794 +/- 0.017
OPW	opercle	2	Stn297	7.3	7.5	3.879 +/- 0.02	3.869 +/- 0.018	3.829 +/- 0.017	3.772 +/- 0.017
OPL	opercle	7	Stn79	11.6	11.1	3.797 +/- 0.016	3.893 +/- 0.017	3.897 +/- 0.019	3.942 +/- 0.016
OPW	opercle	7	Stn81	5.3	5.4	3.81 +/- 0.018	3.842 +/- 0.019	3.79 +/- 0.017	3.888 +/- 0.018
OPL	opercle	12	Stn275	9.6	9.1	3.818 +/- 0.017	3.839 +/- 0.016	3.915 +/- 0.018	3.929 +/- 0.017
OPW	opercle	17	Stn231	6.8	7.0	3.8 +/- 0.018	3.808 +/- 0.016	3.84 +/- 0.021	3.887 +/- 0.018
OPN	opercle	18	Stn196	8.5	9.2	0.2704 +/- 0.008	0.2063 +/- 0.0073	0.2229 +/- 0.0063	0.2364 +/- 0.0067
PE	medianfin	1	Stn248	4.5	5.5	3.023 +/- 0.062	2.943 +/- 0.05	2.911 +/- 0.07	2.615 +/- 0.064
DS2L	medianfin	2	Stn21	7.0	3.9	3.982 +/- 0.05	3.892 +/- 0.046	4.001 +/- 0.045	3.841 +/- 0.035
DS1L	medianfin	3	Stn33	6.1	3.8	3.182 +/- 0.054	3.122 +/- 0.047	3.182 +/- 0.039	3.013 +/- 0.052
DS2L	medianfin	4	Eda	51.1	38.2	4.302 +/- 0.042	3.96 +/- 0.033	3.751 +/- 0.038	3.685 +/- 0.032

TAP	medianfin	4	Stn420	9.7	11.0	10.79 +/- 0.059	10.47 +/- 0.061	10.51 +/- 0.063	10.2 +/- 0.064
DS2L	medianfin	5	Stn326	4.8	2.6	3.959 +/- 0.049	3.886 +/- 0.043	3.96 +/- 0.038	3.869 +/- 0.045
DS1L	medianfin	7	Stn76	7.1	4.5	3.195 +/- 0.046	3.143 +/- 0.051	3.125 +/- 0.045	3.015 +/- 0.052
DS1L	medianfin	8	Stn92	6.3	4.0	3.226 +/- 0.052	3.164 +/- 0.05	3.077 +/- 0.038	3.034 +/- 0.048
DS2L	medianfin	9	Stn102	12.2	7.0	4.067 +/- 0.047	3.89 +/- 0.046	3.922 +/- 0.038	3.8 +/- 0.038
SRA	medianfin	10	Stn121	6.0	6.7	0.01785 +/- 0.0035	0.02191 +/- 0.0037	0.03213 +/- 0.0049	0.04105 +/- 0.0038
DS2L	medianfin	11	Stn224	4.5	2.5	3.845 +/- 0.04	3.887 +/- 0.045	3.915 +/- 0.042	4.018 +/- 0.044
DS3L	medianfin	13	Stn155	5.5	4.6	1.127 +/- 0.02	1.15 +/- 0.016	1.09 +/- 0.018	1.121 +/- 0.02
DS2L	medianfin	16	Stn299	8.0	4.5	4.059 +/- 0.051	3.906 +/- 0.045	3.902 +/- 0.038	3.815 +/- 0.036
ASL	medianfin	17	Stn286	5.7	5.5	1.158 +/- 0.018	1.251 +/- 0.019	1.237 +/- 0.022	1.262 +/- 0.017
DS1L	medianfin	18	Stn196	7.5	4.8	3.247 +/- 0.049	3.085 +/- 0.047	3.134 +/- 0.042	3.028 +/- 0.05
SRA	medianfin	20	Stn212	7.9	9.0	0.04322 +/- 0.004	0.03082 +/- 0.0036	0.02228 +/- 0.0041	0.01077 +/- 0.0037
DS3L	medianfin	21	Stn421	7.9	6.7	1.175 +/- 0.019	1.13 +/- 0.018	1.131 +/- 0.02	1.054 +/- 0.016
LAP	vertebrae	1	Stn240	13.1	10.9	25.68 +/- 0.05	25.49 +/- 0.049	25.33 +/- 0.056	25.21 +/- 0.051
LAP	vertebrae	4	Stn419	10.4	8.5	25.61 +/- 0.055	25.48 +/- 0.053	25.46 +/- 0.057	25.21 +/- 0.044
CVN	vertebrae	5	Stn58	5.0	6.4	16.32 +/- 0.062	16.43 +/- 0.062	16.12 +/- 0.059	16.13 +/- 0.042
LAFR	vertebrae	8	Stn87	5.9	14.9	22.55 +/- 0.15	22.33 +/- 0.11	22 +/- 0.14	22.52 +/- 0.085
LAP	vertebrae	10	Stn120	5.5	4.4	25.4 +/- 0.056	25.38 +/- 0.055	25.34 +/- 0.055	25.58 +/- 0.049
LAFR	vertebrae	15	Stn170	4.3	10.5	22.61 +/- 0.12	22.26 +/- 0.11	22.57 +/- 0.12	22.11 +/- 0.12
LAP	vertebrae	20	Stn340	8.3	6.7	25.55 +/- 0.05	25.37 +/- 0.052	25.52 +/- 0.055	25.25 +/- 0.057
LAP	vertebrae	21	Stn223	14.6	12.3	25.62 +/- 0.048	25.33 +/- 0.055	25.53 +/- 0.056	25.24 +/- 0.05

Table 2.2: Summary of filtered QTL

Only the non-overlapping QTL with highest LOD score within each trait class are shown. Phenotypes are presented as phenotype of a fish of 40 mm SL, with residuals added to each genotypic class (M_1M_2 is homozygous marine, B_1B_2 homozygous benthic, and MB classes the heterozygotes). For traits that were only size corrected, the linear regression equation was used, for traits both size and sex corrected, the GLM equation was used.

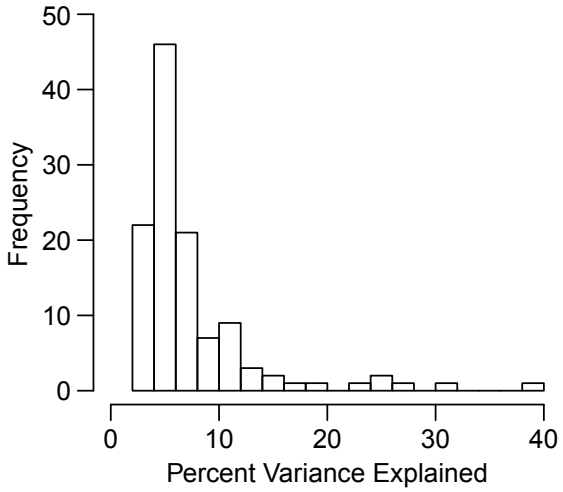


Figure 2.4: Distribution of percent variance explained (PVE). Histogram of PVE for the 118 filtered QTL. There are many QTL of small effect and few QTL of large effect.

Branchial bones

We measured the lengths of all five ceratobranchials (large ventral bones in the branchial skeleton), as well as the first epibranchial, a dorsal branchial bone (**Figure 2.2I**). We identified 16 QTL on 14 chromosomes controlling branchial bone length, including two QTL on chromosomes 4 and 21 that had large effects on ventral and dorsal bones, respectively (**Table 2.2**). QTL with anatomically regional or broad effects were both detected in the cross: four QTL (chromosomes 1, 7, 10, and 12) had significant effects on only one of six bones analyzed, while four other QTL (chromosomes 2, 4, 5, and 8) had significant effects on all six branchial bones analyzed.

Jaw

Unlike the branchial bones, which ossify endochondrally, the premaxilla and dentary, the major bones of the fish upper and lower jaw (**Figure 2.2J-L**), respectively, are dermal bones that form without a cartilage intermediate (Anker 1974; Cubbage and Mabee 1996). We detected 15 QTL on 14 chromosomes controlling jaw morphology (**Table 2.2**). Eleven QTL influenced the size of the premaxilla, with the largest effect QTL on chromosome 21. Ten QTL influenced the size of the dentary and the associated articular bone, with the largest effect QTL mapping to chromosome 4. Three QTL (chromosomes 2, 3, and 20) controlled upper but not lower jaw size and three other QTL (chromosomes 9, 16, and 17) controlled lower but not upper jaw size. In contrast, seven QTL (chromosomes 4, 5, 7, 12, 14, 19, and 21) had significant effects on both the upper and lower jaw.

Skull

In the posterior skull, the supraoccipital crest is longer and thinner in benthic fish than marine fish (**Figures 2.1G-H, 2.2M-N**). The supraoccipital crest serves as the attachment points of the

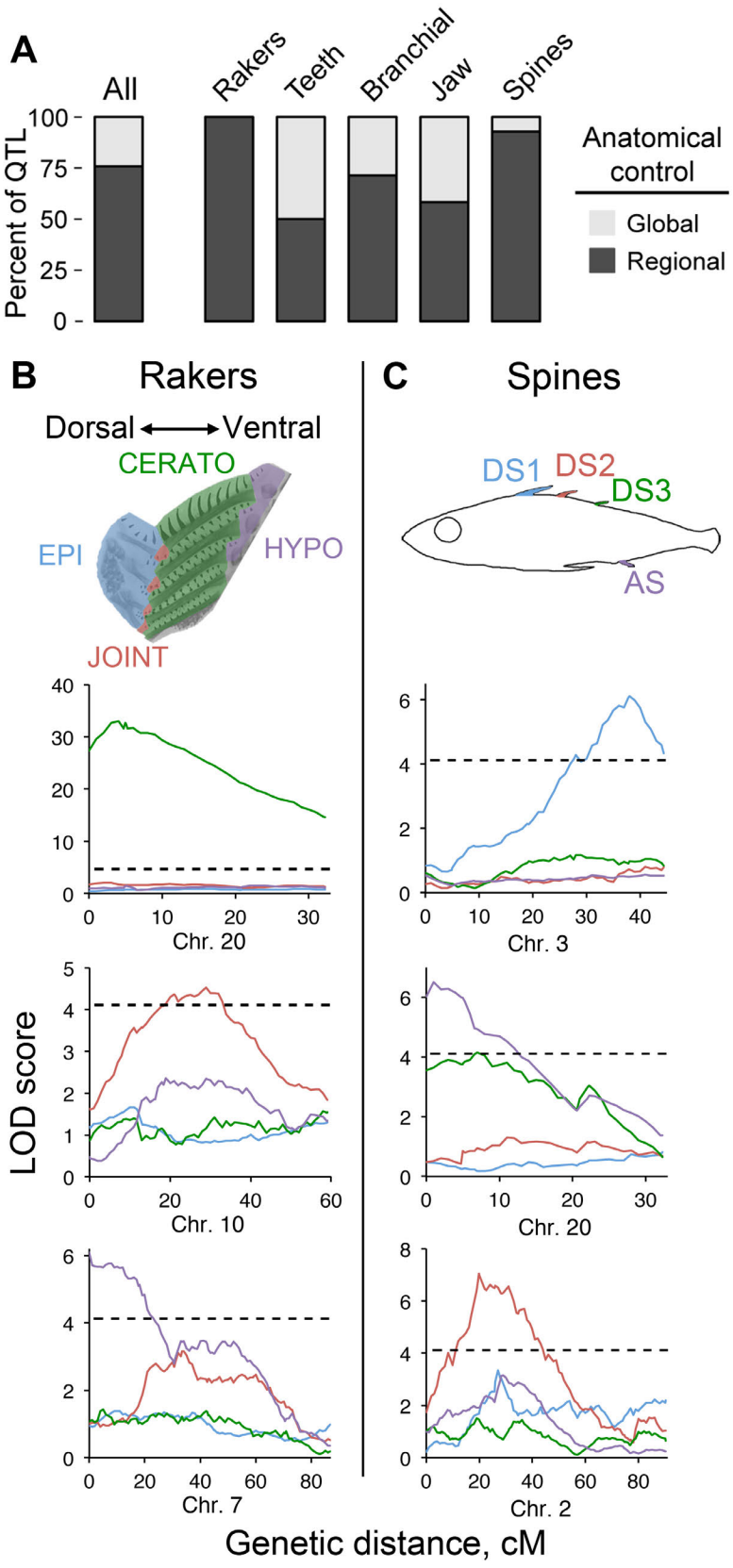

A
All
Rakers
Teeth
Branchial
Jaw
Spines
Percent of QTL
100
75
50
25
0
Anatomical control
Global
Regional
B
Rakers
Dorsal
Ventral
CERATO
EPI
HYPO
JOINT
C
Spines
DS1
DS2
DS3
AS
LOD score
Chr. 20
Chr. 3
Chr. 10
Chr. 20
Chr. 7
Chr. 2
Genetic distance, cM

Figure 2.5 (previous page): Skeletal QTL with anatomically regional effects
(A) Proportion of QTL with regional vs. global effects. For trait classes containing multiple serial or developmental homologous elements or domains, percentages of QTL controlling some (black) or all (white) elements or domains within the trait class are shown. (B) Gill raker and (C) median fin spine morphology are controlled mainly by QTL with highly regional effects. Top left: color-coded dorsal-ventral domains of gill rakers. From dorsal to the ventral midline, rakers are present in EPI (dorsal, blue), JOINT (intermediate, red), CERATO (ventral, green), and HYPO (ventromedial, purple) domains. Examples of three gill raker QTL with regional effects in these dorsal-ventral domains, top to bottom: chromosome 20, 10, and 7 have regional cerato (ventral), joint, and hypo specific domains, respectively. Raker totals in each domain are mapped separately and results color-coded as in the raker schematic. Top right: three dorsal spines (DS1-3) and one anal spine are color coded in the spine schematic. Below are examples of three spine QTL on chromosomes 3, 2, and 20 having regional effects on DS1, DS3+AS, and DS2 respectively. In each QTL plot, genetic distance in cM is on the x-axis and LOD score on the y-axis. Dashed lines are significance thresholds.

epaxial muscles that generate force during suction feeding. These muscles are larger in Paxton benthic fish, contributing to a derived increase in suction index (McGee and Wainwright 2013). We hypothesized that these differences in supraoccipital crest morphology are adaptive for benthic feeding and have a heritable genetic component. We measured supraoccipital crest length, width, and area, and a fourth skull trait, frontal width (interorbital distance). We identified 12 QTL that affect these skull traits (**Table 2.2**), including a large-effect QTL on chromosome 20 that significantly influenced supraoccipital crest width and area.

Opercle

The shape of the opercle bone (**Figure 2.2O**) also varies between marine and freshwater populations, which might reflect differences in feeding and/or respiration (Kimmel et al. 2012; Kimmel et al. 2005). Opercle size in this cross was strongly sexually dimorphic, and was also controlled by seven autosomal QTL controlling opercle length, width, or neck width (**Table 2.2**).

Spines and median fins

We mapped QTL on 16 chromosomes controlling dorsal or anal spine length (**Table 2.2**). A QTL on chromosome 4 affecting the length of the second dorsal spine was the most significant QTL in our entire data set, with a LOD of 51. Chromosome 4 also had large effects on the lengths of the other two dorsal spines and the anal spine. In contrast, QTL on seven chromosomes (2, 3, 5, 7, 8, 11, and 13) had regional effects with significant effects restricted to only one of the four spines measured. Chromosomes 3 and 2 had specific effects on dorsal spine 1 and 2, respectively, while chromosome 20 had regional effects on dorsal spine 3 and the anal spine (**Figure 2.5**). QTL controlling the number and area of serrations on the second dorsal spine (**Figure 2.2P**) mapped independently of the QTL controlling the length of the second spine, consistent with previous studies showing that presence or absence of serrations varies substantially among sticklebacks populations, even among populations with prominent second dorsal spines (Gross 1978).

Category of homologous skeletal elements	Number and description of anatomical domains	Regional QTL (controlling subset of domains)	Global QTL (controlling all domains)
Gill raker number	25 domains in 9 anterior/posterior and 4 dorsal/ventral regions	10	0
Pharyngeal tooth number	2 dorsal, 1 ventral toothplate	4	4
Branchial bone length	1 dorsal, 5 ventral bones	10	4
Jaw size	1 dorsal, 1 ventral bone	7	5
Spine length	3 dorsal, 1 anal spine	13	1
Total	--	44	14

Table 2.3: Regional or global QTL controlling serially homologous skeletal elements
For each set of likely serially homologous skeletal elements, QTL controlling these elements were classified as controlling only a subset of the domains (anatomically regional) or controlling all the domains (global). Most QTL are anatomically regional, and this trend holds broadly across different trait classes.

Vertebrae traits

We found vertebral number to be sexually dimorphic, consistent with previous studies (Reimchen and Nelson 1987). Vertebrae number and position of axial landmarks were also under autosomal control, mapping to eight QTL. Although caudal vertebrae number mapped to one significant QTL, the caudal to abdominal vertebrae ratio, proposed to be important for larval fitness (Swain 1992b), had no detected QTL.

Covariance of traits and multivariate analysis of covariance

We analyzed patterns of trait covariance across all trait classes, and found that in general, traits within a trait class tended to covary (**Figure 2.6**). As expected, traits that mapped strongly to the same chromosome (e.g. gill raker and dorsal spine reduction, which both map strongly to chromosome 4) also tended to covary (**Figure 2.6**). The mean absolute correlation was 0.30 and 0.12 for traits within and between trait classes, respectively. We performed principal components analysis with all traits quantified in this study, and mapped the first five principal components (**Figures 2.7-2.8**). The first principal component maps strongly to chromosomes 4 and 21 (**Figure 2.7**), and as expected, traits that map strongly to these two chromosomes (e.g. branchial bone length, jaw size, ventral pharyngeal tooth gain) load heavily onto this component (**Figure 2.8**). The second principal component maps strongly to chromosome 20 (**Figure 2.7**), and traits that map strongly to this chromosome (e.g. gill raker reduction, supraoccipital crest shape) load heavily onto this component (**Figure 2.8**). Four of the top five principal components map significantly to chromosome 4 (**Figure 2.7**), suggesting that the patterns of trait covariance and integration are complex, but frequently involve particular stickleback chromosomes.
The genomic regions underlying trait clusters on chromosomes 4, 20, and 21 contain 48, 35, and 11 genes respectively that have GO annotations suggesting key roles in early developmental patterning and signaling. Changes in such genes may contribute to the multiple traits and covariances that map to particular stickleback chromosomes.

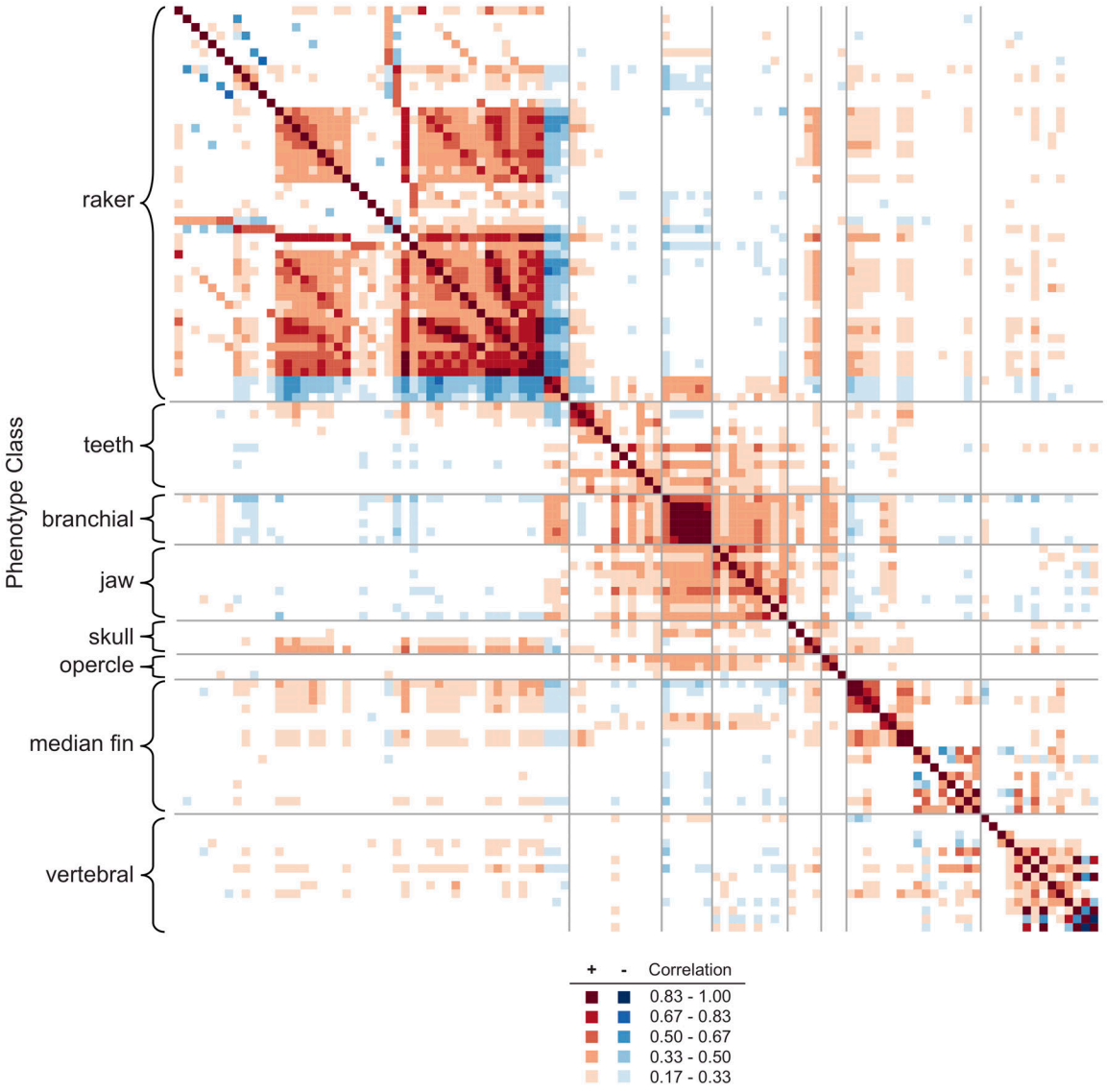


Figure 2.6: Trait covariance heat map
Classes of traits are grouped on the left, and abbreviations (defined in Table 2.1) of individual traits are listed on the right. For each pair of traits, the covariance of Z-scored phenotypes (correlation) is indicated by the heat map shown at bottom, with positive covariances colored red and negative covariances colored blue.

Most QTL are additive or partially additive

We estimated the dominance of each QTL using the formula *d/a* (Falconer 1989) where benthic alleles with strictly recessive, additive, and dominant effects have *d/a* values of -1, 0, and 1, respectively. Across all trait classes, there was a tendency for QTL to act additively, with the distribution of dominance values centered around 0 (**Figure 2.9A**). We defined dominance classes using *d/a* ranges as in a large sunflower domestication QTL study (Burke et al. 2002).

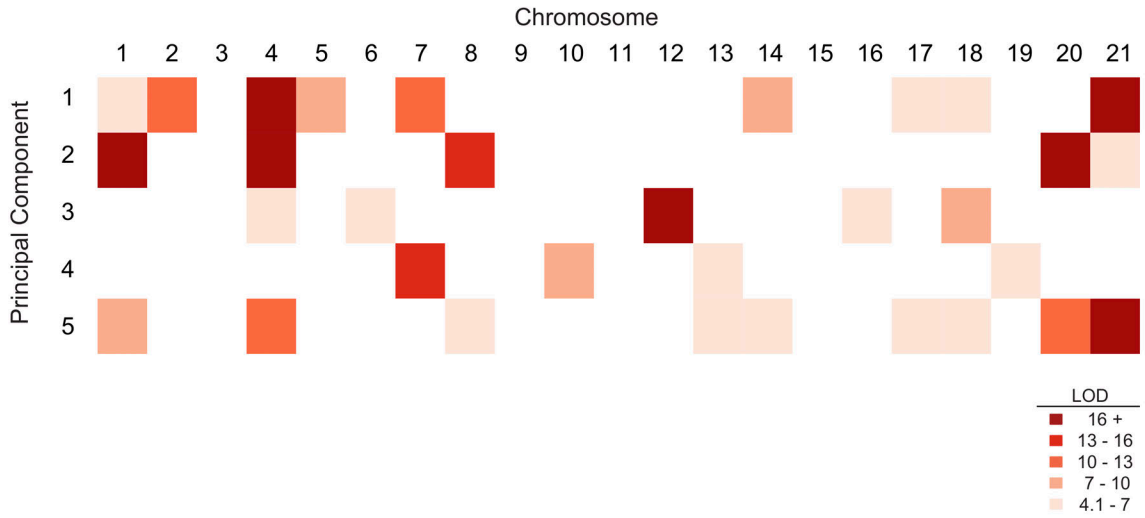


Figure 2.7: Genome-wide overview of principal component QTL. For each of the top five principal components, the LOD score for each significant QTL is indicated by the heat map shown in the bottom right. Further details of the QTL are presented in File 2.3.

Using these ranges, 28% of QTL are additive, 22% partially recessive, 21% partially dominant, 11% recessive, 5% dominant, 6% underdominant, and 6% overdominant (**Table 2.4**). QTL in different trait classes had similar patterns of dominance, with a consistent bias towards additive and partially recessive/dominant QTL (**Table 2.4**). We also observed an apparent relationship between effect size and dominance, with larger effect QTL having more of an additive effect (**Figure 2.9B**), which simulations show is at least in part driven by a lower precision of dominance estimates of small-effect QTL (data not shown). To investigate possible bias in the detection of QTL differing in dominance, we carried out a simulation to determine the detection probability of additive and dominant QTL. Since more variance is present among the mean phenotypes of genotypic classes of a completely dominant or recessive QTL than an additive QTL, we predicted that additive QTL would be harder to detect than dominant QTL by interval mapping. As expected, the probability of detection was slightly higher for a dominant QTL than for an additive QTL (**Figure 2.10**). Thus, the detection of more QTL with additive effects does not result from a detection bias.

Although many different types of mutations can lead to loss or gain of structures during development, we tested the hypothesis that regressive or "loss" QTL (where the freshwater benthic allele contributes to smaller or fewer bones) might more often be recessive, while constructive or "gain" QTL (where the benthic allele contributes to larger or additional bones) might show more dominance. However, the sets of loss and gain QTL contained similar proportions of dominant QTL (5 percent), and the set of gain QTL actually showed a higher percentage of recessive QTL (16 percent versus 7 percent, **Table 2.4**). For both sets of QTL, most loci were at least partially additive (69 percent for gain, 74 percent for loss, **Table 2.4**), and there was no significant difference ($P = 0.46$, Mann-Whitney U test) between the distribution of dominances of loss and gain QTL.

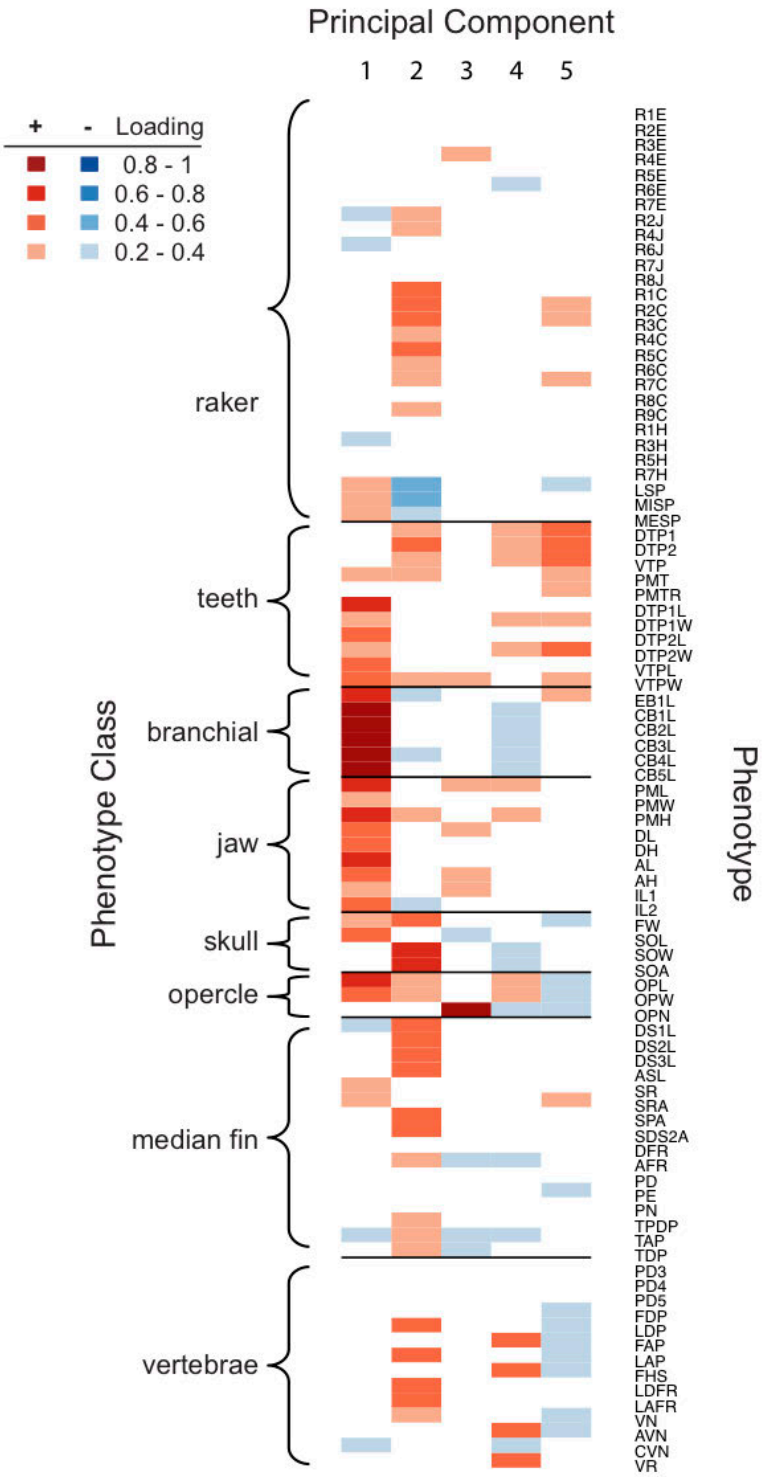


Figure 2.8: Trait loadings for top five principal components
Classes of traits are grouped on the left, and abbreviations (defined in Table 2.1) of individual traits are listed on the right. For each principal component, the loading for each trait is indicated by the heat map shown in the upper left, with positive loadings colored red and negative loadings colored blue.

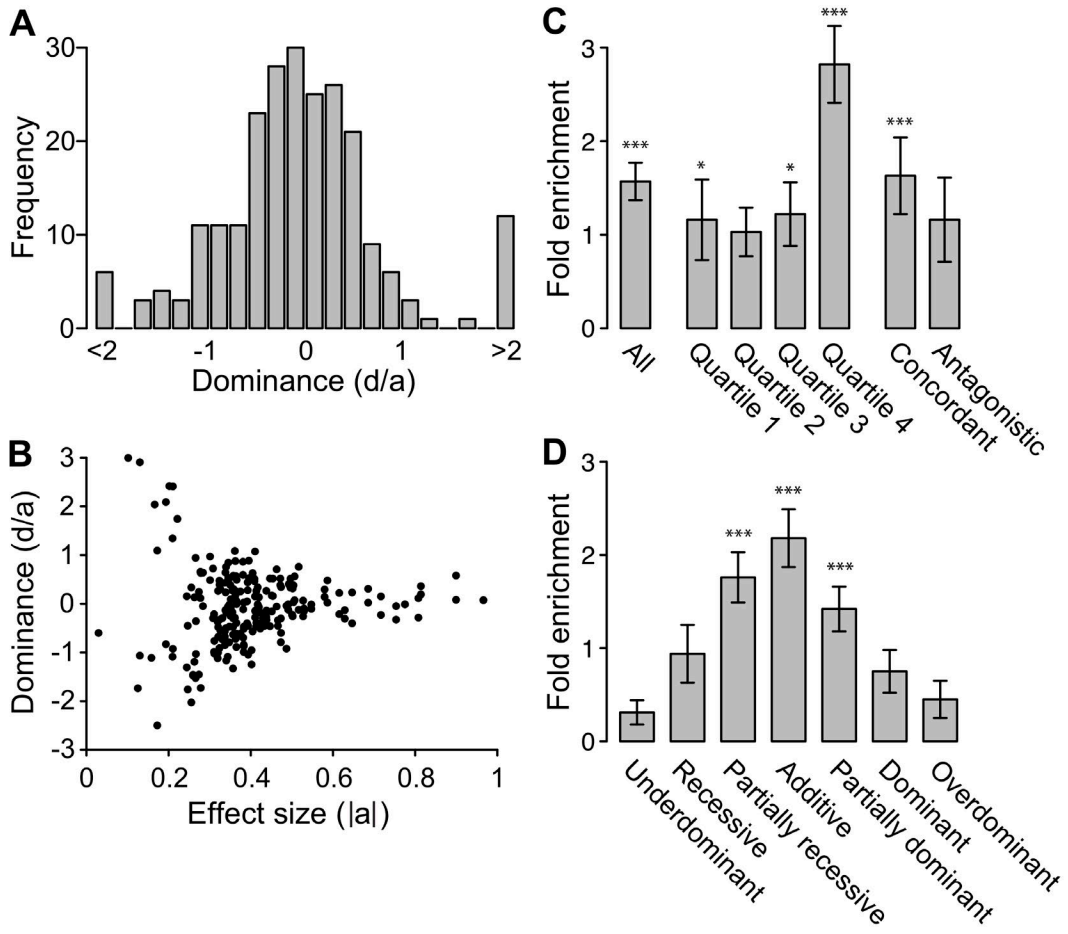


Figure 2.9: QTL dominance patterns and overlap with genomic intervals repeatedly selected during global marine-freshwater stickleback divergence.
(A) Histogram of dominance values (*d/a*) of QTL reveal a tendency towards additive QTL (*d/a* = 0). (B) Dominance values for each QTL plotted against effect size (absolute value of *a*). Ten outlier dominance values that were either over 3 or less than -3 (File 2.2) are not shown. As effect size increases, QTL tend to be more additive. (C) Fold enrichment of QTL 1.5 LOD interval overlaps with genomic regions showing parallel marine-freshwater divergence in whole genome sequence comparisons (Jones *et al.* 2012b) for all QTL, quartiles of QTL by LOD score, concordant or antagonistic QTL. Highly significant enrichment is seen for all QTL, highest LOD QTL (Quartile 4), and concordant QTL. (D) Fold enrichment of QTL 1.5 LOD interval overlap with signals of selection (Jones *et al.* 2012b) by dominance class. Highly significant enrichment is seen for partially recessive, additive, and partially dominant QTL, but not for other dominance classes. * $p<0.05$, *** $p<0.001$.

QTL can also be classified into sets whose effects are either concordant or antagonist to the overall direction of evolutionary change, (i.e. where substitution of a benthic allele confers a more benthic-like or a more marine-like phenotype, respectively). Most (66%) QTL with a predicted evolutionary direction (based on known phenotypes from the grandparental populations) were in the concordant direction. Although we hypothesized that concordant and

Class	Under-dominant	Recessive	Partially recessive	Additive	Partially dominant	Dominant	Over-dominant
raker	0.09	0.15	0.26	0.17	0.15	0.04	0.13
teeth	0.08	0.05	0.20	0.35	0.30	0.00	0.02
branchial	0.03	0.16	0.09	0.41	0.25	0.03	0.03
jaw	0.03	0.10	0.27	0.17	0.33	0.07	0.03
skull	0.04	0.00	0.25	0.50	0.12	0.08	0.00
opercle	0.07	0.21	0.21	0.21	0.14	0.00	0.14
median fin	0.09	0.12	0.24	0.29	0.15	0.06	0.06
vertebrae	0.06	0.12	0.31	0.06	0.19	0.12	0.12
gain	0.07	0.16	0.17	0.29	0.23	0.05	0.04
loss	0.06	0.07	0.27	0.27	0.20	0.05	0.08
all	0.06	0.11	0.22	0.28	0.21	0.05	0.06

Table 2.4 Dominance of QTL by phenotypic class
Dominance (*d/a*) ranges for different dominance classes follow cut-offs of Burke *et al.*: < -1.25 for underdominant, -1.25 to -0.75 for recessive, -0.75 to -0.25 for partially recessive, -0.25 to 0.25 for additive, 0.25-0.75 for partially dominant, 0.75-1.25 for dominant and >1.25 for overdominant. For each dominance class, the proportion of filtered QTL that have dominance values within that range are listed. A tendency towards additive and partially additive QTL is seen broadly across trait classes. Constructive (gain) QTL and regressive (loss) QTL also show a trend towards additivity.

antagonistic QTL might show different dominance distributions, we observed no significant differences for the QTL identified in this study ($P = 0.61$, Mann-Whitney U test).

Recent whole genome resequencing studies in sticklebacks have identified a genome-wide set of regions that are consistently differentiated between marine and freshwater fish populations around the world, and have likely been selected repeatedly to produce marine-freshwater differences (Jones et al. 2012b). The 1.5 LOD genetic intervals controlling skeletal traits in this QTL study were significantly enriched for the genomic regions that show consistent marine-freshwater sequence differences ($P < 0.001$, **Figure 2.9C**). The biggest enrichment was found for the genetic intervals that had the most significant effects on morphology, with a 2.8-fold enrichment ($P < 0.001$) observed for QTL in the top quartile of LOD score (Quartile 4, **Figure 2.9C**). Interestingly, we also observed significant enrichment for the set of concordant QTL that act in the same direction as overall evolutionary change (1.6 fold enriched, $P < 0.001$), but no significant enrichment for the set of antagonistic QTL (**Figure 2.9C**), as expected if marine-freshwater differentiated regions represent genomic intervals that are repeatedly selected to produce the consistent morphological differences observed in marine and freshwater environments. However, there was a trend towards enrichment for discordant QTL, and the difference in fold enrichment between concordant and antagonistic QTL was not significant ($P = 0.44$, two-tailed Student's t-test).

Some variants controlling freshwater stickleback phenotypes are carried at low frequency in marine populations (Colosimo et al. 2005; Miller et al. 2007), and the dominance of such variants may affect their carrier frequency in marine populations, or the rate at which they increase in frequency following colonization of new freshwater environments. We therefore

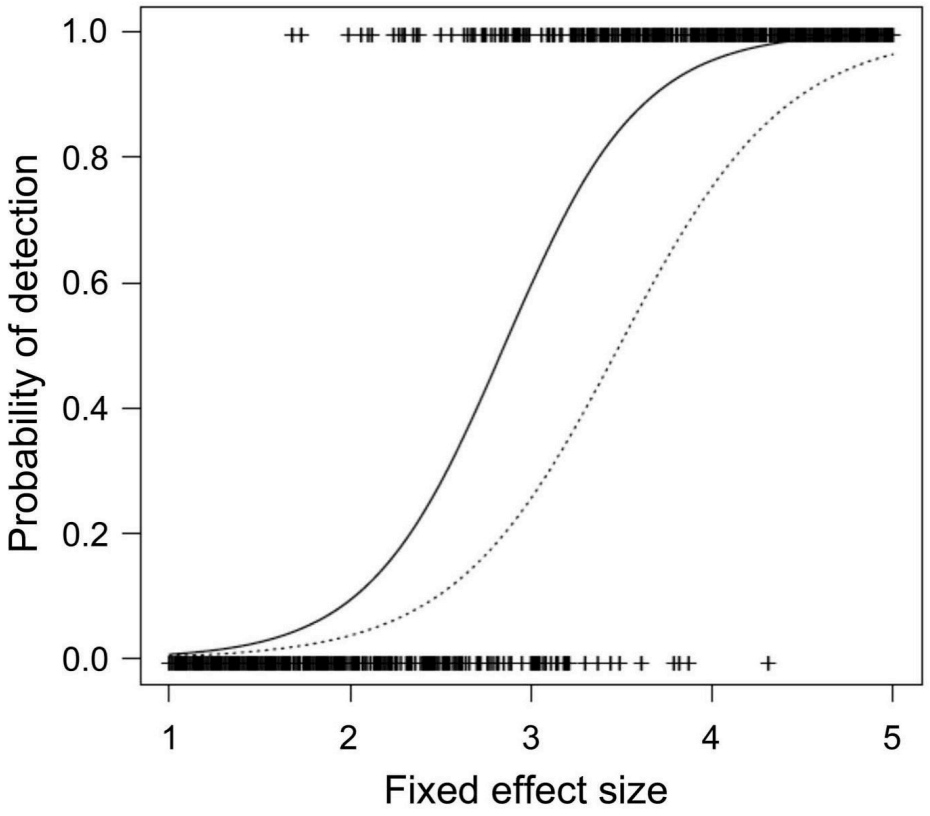


Figure 2.10: Probability of detection of QTL for simulated varying effect sizes
Logistic regressions fitted to the 400 data points on detection (0 = no, 1 = yes), showing probability of detection for QTL with dominance (*d/a*) of 1 (dominant, solid curve) or 0 (additive; dashed curve). Additive QTL are less likely to be detected.

tested whether QTL in different dominance classes were differentially enriched for the marine-freshwater genomic regions that show evidence of repeated selection. Neither recessive nor dominant QTL were enriched for overlap with the genomic regions identified in the Jones et al. 2012 study (**Figure 2.9D**). In contrast, partially recessive, additive, and partially dominant QTL were all strongly enriched for overlap with signals of repeated genomic selection ($P < 0.001$, **Figure 2.9D**).

Trait clusters on chromosomes 4, 20, and 21

As detailed above, inspection of QTL results revealed many complex and non-overlapping patterns of genetic control within and among trait classes (**Figure 2.3**). However, certain chromosomes appeared enriched for QTL, especially QTL with high LOD scores, spanning multiple trait classes (**Figure 2.3, Table 2.2**). To examine possible clustering in greater detail, we first tested whether detected QTL were more likely to have peak markers within 5 cM of each other, compared to randomly distributed QTL (Protas et al. 2008). For both (1) all QTL and (2) large-effect QTL (defined as being in the top quartile by LOD), peak markers were significantly

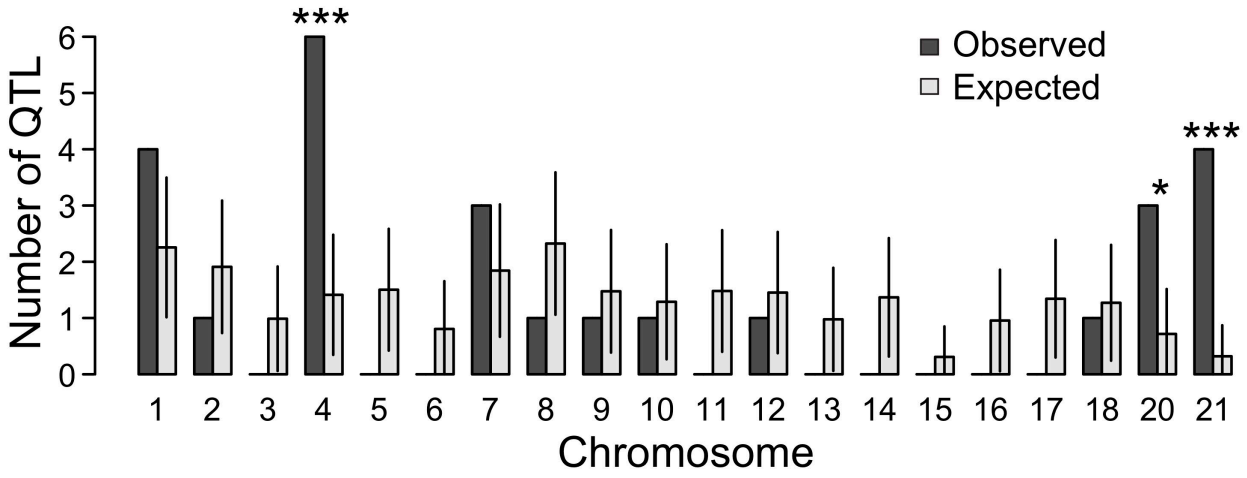


Figure 2.11: Large-effect QTL are enriched on chromosomes 4, 20, and 21
Observed (black) and expected (white) number of large-effect QTL (top quartile by LOD) per chromosome. Data for the expected values represents the mean and standard deviation of values generated from 10,000 simulations. This analysis excluded the sex chromosome (chromosome 19). * $p<0.05$, *** $p<0.001$.

clustered ($P = 0.01$ and $P < 0.001$ for all QTL and large-effect QTL, respectively). In the entire set of filtered QTL, as expected, LOD score and percent variance explained (PVE) were highly correlated (Pearson's correlation coefficient = 0.95).

To ask whether specific chromosomes were significantly enriched for QTL, we used simulations to ask whether (1) all QTL or (2) large-effect QTL were over-represented on any autosome. Relative to a null prediction where QTL are distributed in the genome in proportion to the genetic lengths of autosomes, large-effect QTL were significantly enriched on chromosomes 4, 20, and 21 (**Figure 2.11**). In contrast, only chromosome 21 was statistically enriched when we analyzed all QTL, not just those of large effect (**Table 2.5**). Since we statistically corrected for the effect of sexual dimorphism for each trait, we did not include the sex chromosome (chromosome 19) in these calculations.

Given that QTL are more likely to be detected in regions of low recombination ("the Noor effect," Noor et al. 2001), a low recombination rate on a particular chromosome could contribute to an enrichment of detected QTL on that chromosome. We therefore asked whether trait clustering was significant even when considering physical distance or gene number of each chromosome in the recently published stickleback genome assembly (Jones et al. 2012b). The enrichment of large-effect QTL on chromosomes 4 and 21 remains significant when compared to a null distribution of QTL generated in proportion to either chromosome length or gene number (**Table 2.5**), while the enrichment of QTL on chromosome 20 is suggestive, but not significant ($P = 0.14$ after correcting for either chromosome length or gene number). For chromosomes 4 and 21, clustered traits included both loss and gain QTL, with benthic alleles in the same trait cluster contributing to bone loss for some traits and bone gain for others (**Figure 2.12**). For example, chromosome 4 had large effects on gill raker and dorsal spine loss, but also on jaw size gain. Chromosome 21 had large effects on tooth and branchial bone gain, but also on dorsal

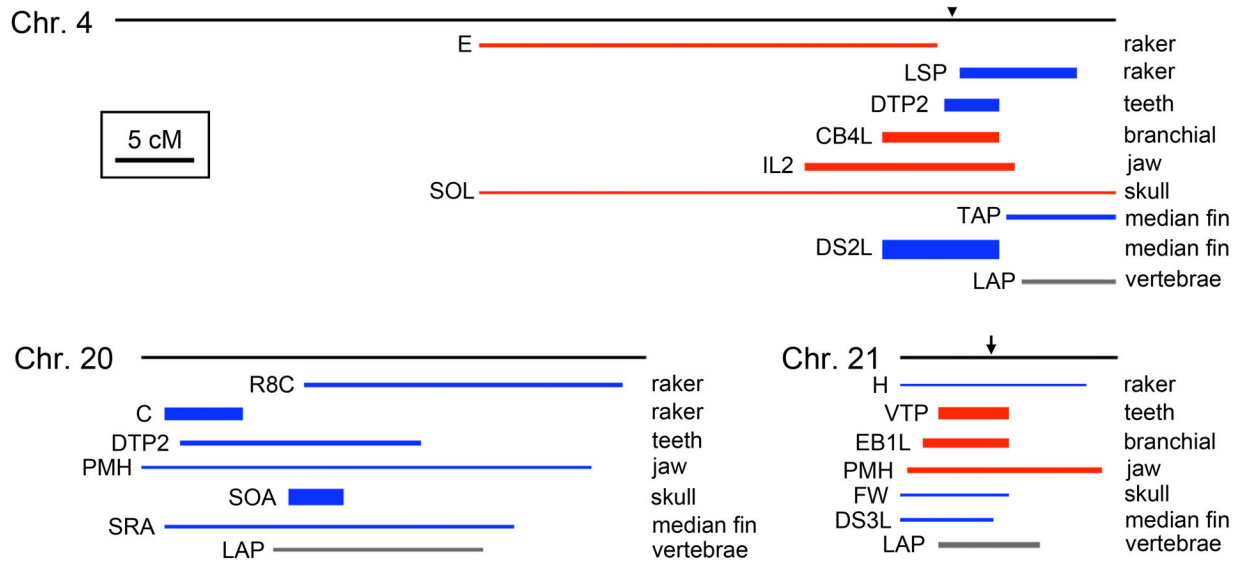


Figure 2.12: Trait clusters on chromosomes 4, 20, and 21
For each chromosome, the genetic length of the chromosome is shown as the black line at the top of each panel (scale bar = 5 centiMorgans). Each blue, red, or gray bar represents a QTL, with the length of the bar denoting the 1.5 LOD interval. Trait abbreviations (Table 2.1) are listed to the right of each bar. The height of each bar is proportional to the LOD score of the QTL. For each QTL, direction of effect is indicated by red (benthic allele confers more or bigger bones), blue (benthic allele confers fewer or smaller bones), or gray (for neither bone loss nor gain). The position of *Ectodysplasin* on chromosome 4 is marked with an arrowhead. A 1.7 megabase inversion on chromosome 21 (arrow), with different orientations typical in marine and freshwater fish (Jones *et al.* 2012b), mapped within the 1.5 LOD interval of all seven QTL on this chromosome, although the peak markers for each QTL (Table 2.2) mapped left of the inversion. See Figure 2.3 and File 2.2 for more details on traits controlled by these three chromosomes. Abbreviations (see Figure 2.2 and Table 2.1): C, total cerato (ventral) gill raker number; CB4L, ceratobranchial 4 length; DS2L, dorsal spine 2 length; DS3L, dorsal spine 3 length; DTP2, dorsal toothplate 2 tooth number; E, epi (dorsal) gill raker number; EB1L, epibranchial 1 length; FW, frontal width; H, total hypo gill raker number; IL2, in-lever 2 of the articular; LAP, vertebrae number of last anal pterygiophore; LSP, lateral row 2 raker spacing; PMH, premaxilla height; R8C, row 8 cerato (ventral) gill raker number; SOL, supraoccipital crest length; SOA, supraoccipital crest area; SRA, spine 2 serration area; TAP, total postanal pterygiophore number; VTP, ventral toothplate tooth number.

spine loss (**Figure 2.12**). In contrast, QTL mapping to chromosome 20 were mostly in the direction where the benthic allele conferred loss or reduction of bone size across multiple trait classes.

Finally, we asked whether the three trait clusters on chromosomes 4, 20, and 21 are enriched for the genome-wide set of regions that are consistently differentiated between marine and freshwater stickleback populations (Jones et al. 2012b). These trait clusters on chromosomes 4, 20, and 21 overlapped 70, 26, and 6 marine-freshwater divergent regions from this set, respectively. Compared to the genome-wide average for an equivalently sized chromosome segment, the chromosome 4 and 20 trait clusters are significantly enriched ($P < 0.001$ for both), with a 4.5-fold and 3.4-fold enrichment, respectively. The chromosome 21 trait cluster, in contrast, has only a 1.1-fold enrichment for these marine-freshwater divergent regions, which is

Chr	All QTL			Large effect QTL		
	Genetic	Physical	Gene number	Genetic	Physical	Gene number
1	0.59	0.39	0.33	0.16	0.09	0.07
2	0.78	0.58	0.36	0.90	0.83	0.75
3	0.99	0.99	1.00	1.00	1.00	1.00
4	0.22	0.50	0.37	**0.00**	**0.01**	**0.00**
5	0.97	0.64	0.80	1.00	1.00	1.00
6	0.98	1.00	0.99	1.00	1.00	1.00
7	0.41	0.37	0.36	0.27	0.27	0.27
8	0.90	0.41	0.38	0.94	0.77	0.76
9	1.00	0.99	0.99	0.81	0.78	0.80
10	0.76	0.56	0.62	0.76	0.70	0.73
11	0.97	0.85	0.95	1.00	1.00	1.00
12	0.57	0.38	0.49	0.80	0.75	0.81
13	0.81	0.93	0.93	1.00	1.00	1.00
14	0.80	0.53	0.55	1.00	1.00	1.00
15	0.08	0.83	0.83	1.00	1.00	1.00
16	0.51	0.68	0.60	1.00	1.00	1.00
17	0.78	0.49	0.50	1.00	1.00	1.00
18	0.75	0.59	0.56	0.75	0.71	0.70
20	0.09	0.43	0.42	**0.03**	0.14	0.14
21	**0.00**	**0.02**	**0.01**	**0.00**	**0.01**	**0.00**

Table 2.5: Trait clustering *P* values
P values for enrichment of QTL on each chromosome for all QTL or large effect QTL were calculated by comparing the actual number of QTL per chromosome to the distribution from 1000 simulated placements of QTL. Simulations were ran assuming QTL were randomly placed in proportion to the 1) the genetic length of the chromosome, 2) the physical length of the chromosome, or 3) the number of Ensembl-predicted genes on the chromosome. *P* values less than 0.05 are in bold. Relative to simulated QTL placed in proportion to the genetic length of each chromosome, large effect QTL (the top quartile of QTL by LOD) are significantly enriched on chromosomes 4, 20, and 21.

not significant ($P = 0.28$). Thus, two of the three trait clusters are significantly associated with characteristic marine-freshwater divergent genomic regions.

DISCUSSION

Regional control of skeletal anatomy

A main finding of this study is that the genetic control of evolved skeletal morphology in sticklebacks involves both differential genetic control between trait classes, as well as highly specific control of individual skeletal elements within a trait class. It is not surprising to find differential genetic control between trait classes, given the likely different embryonic origins of skeletal elements in different trait classes. Perhaps more surprising is the extent of highly specific anatomical control among skeletal elements thought to be serially homologous.

Consider, for example, the genetic mapping data for gill raker number. Gill rakers form throughout the pharynx, projecting from dorsal, joint, and ventral regions of branchial arches. The reduction in gill raker number in derived freshwater fish occurs in each branchial segment (Gross and Anderson 1984) and is typically described by summing all anterior-facing rakers on the first branchial segment (row 1; e.g. Hagen and Gilbertson 1972). Our mapping data revealed that in the Paxton benthic population, gill raker reduction was accomplished genetically in a piecemeal fashion by at least 23 QTL with specific effects in particular dorsal/ventral domains. Both of the large-effect gill raker QTL had regionally specific effects, the QTL on right chromosome 4 controlling anterior ventral gill raker number and spacing, and the chromosome 20 raker QTL controlling strictly ventral gill raker number (**Figures 2.4-2.5**). This decoupling of the genetic control of dorsal and ventral gill rakers is also consistent with a previous ecological study that found ventral, but not dorsal, gill raker number to have predictive value in discriminating different wild freshwater populations, perhaps reflecting population-specific diets (Reimchen et al. 1985). The anatomical specificity of QTL might be even more complex; even within the ventral domain, the chromosome 20 raker QTL had regional effects, controlling only lateral and middle, but not medial spacing (**Figure 2.4**).

Another example of highly specific anatomical effects within a trait class is the genetic control of dorsal spine development. Nine of 14 QTL controlling dorsal spine lengths were specific to one of the three dorsal spines. Additional QTL controlled the number and area of barb-like serrations along the surface of the second dorsal spine, and these QTL mapped to different genomic regions than QTL that control length of the second dorsal spine. Previous studies have shown that natural populations of sticklebacks differ in the number of spines, the length of particular spines, and the degree of barb development along spines, likely reflecting the key roles of dorsal spine morphology in defense against different types of predators, as well as possible functions in display and dorsal pricking interactions during stickleback courtship (Gross 1978; Hoogland 1957; Kitano et al. 2009; Reimchen 1980). Across larger phylogenetic distances, many other fish groups show striking changes in the length or morphology of individual spines, for example, the specific elongation of the first dorsal spine in trigger fish and angler fish. These dramatic species-specific modifications also likely depend on precise anatomical control of spine growth by genetic mechanisms that do not cause comparable changes in all members of a developmentally related series.

The high degree of regional control for skeletal QTL in sticklebacks is consistent with the idea that anatomically specific changes may avoid negative pleiotropy during development, and will therefore predominate during morphological evolution in natural populations (Carroll 2008; Stern 2000). Highly specific skeletal effects may be controlled by genes whose expression patterns are themselves highly restricted along developmental axes (e.g. *Hox* and *Dlx* genes), or by *cis*-regulatory changes that alter a particular subset of the expression domains of more broadly expressed genes. Further molecular dissection of the QTL mapped in this study, using genetic fine mapping and transgenic methods similar to those that have been successfully applied to other stickleback traits (Chan et al. 2010; Colosimo et al. 2005; Miller et al. 2007), should help illuminate the detailed mechanisms that vertebrates use to shape the size and number of individual skeletal elements as they evolve in different environments.

Most QTL had additive or partially additive effects

A second major conclusion from this study is that the majority of the detected evolutionary QTL had additive or partially additive effects, regardless of skeletal trait class, and regardless of the overall direction of their effects on the skeleton (either gain or loss of bone in the derived freshwater form; **Table 2.4**). Our simulations indicated that detection bias does not explain the enrichment in additive QTL; instead it might contribute to an underdetection of additive QTL (**Figure 2.10**). The strong tendency toward additivity across trait classes suggests that this trend may be a general feature of evolved stickleback traits. Previous studies have shown that repressive genetic interactions tend to be as common as activating genetic mechanisms during development (Davidson and Levine 2008). Since constructive traits could be due to either loss of repressors or gain of activators, and regressive traits could arise by either loss of activators or gain of repressors, it is perhaps not surprising that a range of skeletal traits, including both increases and decreases of bony tissue, tend to show similar genetic architectures.

Two main models have been proposed for the effect of the dominance of a mutation on its likelihood of fixation during adaptation. "Haldane's sieve" predicts that new advantageous mutations are more likely to increase in frequency if they are not recessive (Haldane 1927). In contrast, Orr and Betancourt showed that if standing variants preexist in populations at mutation-selection balance and are disadvantageous prior to, but favored after, an environmental change, then probability of fixation in the new selective regime is largely independent of dominance (Orr and Betancourt 2001). In the stickleback system, detailed case histories of the specific variants underlying armor and pigment traits (Colosimo et al. 2005; Miller et al. 2007), and more recent genome-wide surveys of parallel evolving freshwater populations (Hohenlohe et al. 2010; Jones et al. 2012b), show that repeated selection of ancient standing variants plays a substantial, but not exclusive (Chan et al. 2010), role in repeated marine-freshwater divergence. The overall distribution of dominance we observe for skeletal QTL in sticklebacks is thus likely based on a mixture of *de novo* mutations which have arisen during the divergence of the particular populations studied, and older standing variants that likely exist at selection-migration balance in ocean populations, and become favorable when introduced into new freshwater environments. We observe a strong tendency towards additivity for QTL, which can not be simply explained by either Haldane's sieve or the Orr/Betancourt model. However, the Haldane and Orr/Betancourt predictions are for fitness, and it is possible that the dominance for the skeletal traits studied here do not reflect the dominances for fitness, as previously seen in the dominances of *Eda* and chromosome 4 genotype on lateral plate morphology and fitness (Barrett et al. 2008).

Similar trends towards additivity of QTL have been observed in genetic studies of traits under artificial selection in mice and outbreeding plants (Burke et al. 2002; deVicente and Tanksley 1993; Kenney-Hunt et al. 2008; Ronfort and Glemin 2013), as well as naturally evolved differences between surface and cave-adapted fish (Protas et al. 2008). This trend towards additivity for evolutionary QTL could result at least in part from a bias in the dominance distribution of the types of mutations favored by selection. For example, segregating *cis*-regulatory alleles have been found to be additive more often than *trans*-regulatory alleles (Gruber et al. 2012; Lemos et al. 2008; McManus et al. 2010). If selection favors *cis*-regulatory mutations, then additive QTL are expected to be common. Furthermore, the Orr/Betancourt model assumes that standing variation is at mutation-selection balance, whereas much of the standing variation reused by stickleback populations may be at migration-selection balance (Barrett and Schluter 2008), maintained in the ancestral marine population by introgression from freshwater populations. Such variation has already been filtered by selection: to be present in the

sea it likely had increased in frequency already in freshwater populations. As a result, standing variation should be biased towards the kinds of mutations that selection favors in freshwater, which we hypothesize to be *cis*-regulatory mutations due to their low pleiotropy and a tendency to be additive.

Although we observed a strong tendency towards additivity of QTL, the 12% of overdominant and underdominant QTL observed likely indicates that some of the genetic effects observed in this cross result from complex interactions between the divergent grandparental genomes used for the cross, some of which may not be typical for very recent divergence between more closely related populations. Furthermore, although the extant Japanese Pacific marine population was used in this study as a living proxy for the marine ancestor of Pacific basin derived freshwater fish including Paxton benthics, modern day marine fish can not be equated with the ancestor of Paxton Lake fish. Given that genetic effects including dominance of QTL are likely context (e.g. genetic background, environment) dependent, additional crosses will be needed to test general patterns of evolved genetic effects in this system.

Overall, the QTL identified in this study show significant enrichment for overlap with the previously identified haplotypes that are consistently differentiated between marine and freshwater fish populations around the world (Jones et al. 2012b). This enrichment suggests that a subset of the genomic regions repeatedly used for freshwater adaptation is selected for their effects on skeletal morphology. Large-effect QTL and additive QTL display the strongest enrichment, while small-effect, recessive, or dominant QTL showed no or less enrichment for overlap with these haplotypes. These enrichment differences could at least partly result from a higher proportion of false positives in the set of small-effect QTL, which are more likely to be recessive or dominant. In addition, small-effect, recessive, and dominant classes might be enriched for new mutations (rather than standing genetic variation), which have a lower probability of detection by the method used in the Jones et al. study. Further analysis of the QTL intervals identified in this study will test the hypothesis that the enrichment of these signals of selection in the QTL intervals is driven by particular genomic regions that act to control specific skeletal traits mapped in this study. Future population genetic studies in marine and Paxton benthic populations can also test whether haplotypes inside the QTL intervals identified here are outliers for metrics such as Fst.

Clustering of QTL on chromosomes 4, 20, and 21

A third major finding of this study is that multiple trophic and armor traits map strongly to chromosomes 4, 20, and 21. We found that QTL from six (chromosome 20) or seven (chromosomes 4 and 21) of the eight trait classes mapped to each trait cluster (**Figure 2.3**). Although all QTL were only enriched on chromosome 21, large-effect (top quartile of QTL by LOD) QTL were enriched on chromosomes 4, 20, and 21 (**Figure 2.11**). All three trait clusters controlled specific subsets of skeletal traits and are thus unlikely to represent loci generally involved in bone formation. For example, some skeletal traits, such as opercle size, mapped strongly to multiple genomic locations but were not significantly controlled by any of the three large-effect trait clusters.

The trait clusters could result from single genes with pleiotropic effects, or from the combined effects of multiple linked genes. Several QTL studies have identified loci that are thought to have pleiotropic consequences (Albert et al. 2008; Kimura et al. 2007; Studer and Doebley 2011), including a large-scale study of QTL controlling skeletal differences between

mice artificially selected for large or small body size (Kenney-Hunt et al. 2008). In contrast, genetic studies in butterflies, pinthrum, and *Petunia* have reported some trait clusters that are due to closely linked but separable loci, rather than pleiotropic effects of a single gene (Ferguson et al. 2010; Hermann et al. 2013; Joron et al. 2006; Kurian and Richards 1997). As the degree of pleiotropy of a QTL increases, the relative frequency of antagonistic effects (effect in the opposite direction of the direction of evolutionary change) is predicted to increase during selection (Griswold and Whitlock 2003), which perhaps at least partially explains our observation that a significant fraction (34%) of QTL are antagonistic. However, antagonistic effects could also result from stabilizing selection, from genetic drift (Griswold and Whitlock 2003; Rieseberg et al. 2002), or from pleiotropic mutations that overshoot the optimum phenotype.

For two of the stickleback trait clusters presented here, genetic resolution of linked traits argues against pleiotropy. For example, two of the linked raker QTL on chromosome 4, as well as two linked raker and supraoccipital crest QTL on chromosome 20, appear spatially distinct from each other, with non-overlapping 1.5 LOD intervals (**Figure 2.12**). In addition, for the trait clusters on chromosomes 4 and 21, benthic alleles do not act in a consistent phenotypic direction (**Figure 2.12**). For example, chromosome 4 benthic alleles reduce gill raker number, pharyngeal tooth number, and dorsal and anal spine lengths, but also increase upper and lower jaw sizes, branchial bone sizes, and the length of the supraoccipital crest. Given the opposite directions of phenotypic effects, and the genetic resolution separating some of the linked QTL, we favor a model where several individual, linked QTL exist, possibly including a "supergene" complex with multiple effects on both armor and trophic phenotypes. Increased genetic resolution of these overlapping QTL is needed to test whether the QTL are separable, and whether some of the overlapping traits might resolve to a supergene complex. In cases where loss and gain QTL overlap, it is possible these traits share developmental interactions (e.g. the genetically encoded loss of a trait might result indirectly in the gain of another).

In cases of multiple linked QTL, trait clustering may be due to genomic intervals of decreased recombination. For example, inversions suppress recombination, and in *Mimulus* and *Heliconius* appear to lock in a suite of co-adaptive polymorphisms (Fishman 2013; Joron et al. 2011; Lowry and Willis 2010). Recent stickleback genome sequencing revealed a 1.7 megabase inversion on chromosome 21 that displays strong signals of selection, whereby marine and freshwater populations have high and low allele frequencies, respectively, of the inversion (Jones et al. 2012b). Jones et al. proposed that this inversion may hold several distinct adaptive loci together, and both the current study of skeletal QTL, as well as another recent study of lateral line QTL (Wark et al. 2012), confirm that many QTL map to chromosome 21, with confidence intervals that overlap the position of the inversion. Although this study identifies a large number of new traits that may be controlled by an inversion/supergene complex in sticklebacks, we note that the peak markers for each of the chromosome 21 QTL map left of the inversion. Ongoing fine mapping studies using crosses that generate recombination events in and around the inversion will provide useful information on both the position and identity of the genes and mutations that underlie one of the most distinctive trait clusters in the stickleback genome.

Previous studies have identified multiple QTL mapping to chromosome 4 in sticklebacks, including QTL for lateral plate number and lateral plate size (Colosimo et al. 2004; Cresko et al. 2004), pelvic spine length (Shapiro et al. 2004), and multiple aspects of body shape (Albert et al. 2008; Rogers et al. 2012). The data presented here reveal that a surprisingly large number of

additional traits also map to chromosome 4, including gill raker number, pharyngeal tooth number, branchial bone size, premaxilla size, dentary and articular size, supraoccipital crest length, dorsal and anal spine length, and aspects of vertebral positioning. Many of these traits, including larger jaws and fewer gill rakers, shorter dorsal and pelvic spine lengths, reductions in lateral plate number, and changes in overall body shape, appear to have adaptive significance in benthic environments, as multiple benthic species independently evolve these morphological changes in recurrent stickleback species pairs (Schluter and Mcphail 1992). Linkage of large-effect QTL controlling multiple aspects of both trophic morphology and anti-predator defense may preserve combinations of traits that function together in different ecological environments. For example, fish foraging in open water environments not only specialize on different food sources, but also tend to encounter different predators. Thus, tight linkage of genes controlling feeding and armor traits may provide a fitness advantage to offspring of contrasting ecotypes, and theory predicts that such linked assemblages will evolve under conditions where strongly contrasting forms sometimes meet and hybridize (Kirkpatrick and Barton 2006), as frequently occurs in marine-stream and benthic-limnetic stickleback species pairs.

Parallel evolution of polygenic traits

Large-effect QTL for armor plate, pigment, and pelvic development that were previously mapped in this cross do not appear to be specific to this cross. Instead the same major loci (Colosimo et al. 2004; Coyle et al. 2007; Cresko et al. 2004; Shapiro et al. 2004), the same underlying genes (Chan et al. 2010), and sometimes even the same freshwater alleles (Colosimo et al. 2005; Miller et al. 2007) are used repeatedly in other populations that have evolved similar phenotypes (Jones et al. 2012a). All of these well studied examples involve QTL that control half or more of the variance in the corresponding trait, and it remains unclear whether QTL with smaller effects, like many of those identified here, will also be used in parallel in other populations. Previous studies have mapped two gill raker number and four dorsal spine length QTL in a Priest lake benthic x limnetic cross (Peichel et al. 2001). Both of the Priest lake raker QTL overlap raker QTL found in this study, although anatomical domains affected by these QTL differ. In contrast, only one of the four Priest lake dorsal spine QTL overlaps any of the spine QTL presented here. Although the Priest cross also used benthic forms, it was a backcross to freshwater limnetic fish, and trophic and armor selective pressures likely differ on limnetic versus marine fish. It is also likely that some genetic variation is not fixed within a population, and that the spectrum of QTL observed in a genetic cross could be different if different individuals from the same population were used. Additional crosses are needed to test whether similar genetic loci underlie repeated evolution of similar trophic and armor phenotypes in many benthic lake and stream forms that have evolved from marine ancestors.

ACKNOWLEDGEMENTS

The work in this chapter was the result of a highly collaborative effort by many people over many years. All of the generation of genotypes and phenotypes and several initial QTL analyses were performed in the Kingsley lab before I began working on this project. I re-mapped the QTL using uniform methods, and performed much of the meta-analysis of QTL properties, with guidance from Craig Miller, David Kingsley, and Dolph Schluter. I thank the other coauthors for their many contributions, listed below:

Conceived and designed the experiments: DMK, CLP, DS, CTM, AMG, BRS, BKB, ARN, MDS, BLC. Performed the experiments: CTM, AMG, BRS, BKB, ARN, MDS, BLC, CLP. Analyzed the data: AMG, CTM, DMK, DS, BRS, BKB, ARN, MDS, CLP. Contributed reagents/materials/analysis tools: DS. Wrote the paper: CTM, AMG, DMK, DS.

In addition to the coauthors listed above, I thank Steve Arnott, Anne Knecht, and Joy Yang for helpful discussions. This work was supported in part by a Jane Coffin Childs Fund Fellowship, a March of Dimes Basil O'Connor award, a Pew Scholar in Biomedical Sciences award supported by the Pew Charitable Trusts, and National Institutes of Health (NIH) grant DE021475 (to C.T.M.); National Science Foundation Graduate Research Fellowships (to A.M.G. and B.R.S.); a Helen Hay Whitney Foundation fellowship (to M.D.S.); Burroughs Wellcome Fund Career Awards in the Biomedical Sciences (to M.D.S. and C.L.P.); a Canada Research Chair and grants from the National Sciences and Engineering Research Council and the Canada Foundation for Innovation (to D.S.); and an NIH Center for Excellence in Genomic Studies grant (5P50HG002568) and investigator position at the Howard Hughes Medical Institute (to D.M.K.).

REFERENCES

Ahn, D.G., and G. Gibson, 1999 Axial variation in the three-spine stickleback: genetic and environmental factors. Evol Dev 1: 100-112.

Albert, A.Y., S. Sawaya, T.H. Vines, A.K. Knecht, C.T. Miller *et al.*, 2008 The genetics of adaptive shape shift in stickleback: pleiotropy and effect size. Evolution 62: 76-85.

Anker, G.C., 1974 Morphology and kinetics of the head of the stickleback, *Gasterosteus aculeatus*. Trans. Zool. Soc. Lond. 32: 311-416.

Barrett, R.D.H., S.M. Rogers, and D. Schluter, 2008 Natural selection on a major armor gene in threespine stickleback. Science 322: 255-257.

Barrett, R.D.H., and D. Schluter, 2008 Adaptation from standing genetic variation. Trends in Ecology & Evolution 23: 38-44.

Bell, M.A., and S.A. Foster, 1994 *The evolutionary biology of the threespine stickleback*. New York: Oxford University Press.

Bell, M.A., Travis, M. P. and Blouw, D. M., 2006 Inferring natural selection in a fossil threespine stickleback. Paleobiology 32: 562-577.

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.

Broman, K.W., and S. Sen, 2009 *A guide to QTL mapping with R/qtl*. Dordrecht: Springer.

Burke, J.M., S. Tang, S.J. Knapp, and L.H. Rieseberg, 2002 Genetic analysis of sunflower domestication. Genetics 161: 1257-1267.

Caldecutt, W.J., M.A. Bell, and J.A. Buckland-Nicks, 2001 Sexual dimorphism and geographic variation in dentition of threespine stickleback, *Gasterosteus aculeatus*. Copeia: 936-944.

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, Jr., M.D. Shapiro *et al.*, 2010 Adaptive evolution of pelvic reduction in sticklebacks by recurrent deletion of a Pitx1 enhancer. Science 327: 302-305.

Clarke, C.A., P.M. Sheppard, and I.W. Thornton, 1968 Genetics of mimetic butterfly *Papilio Memnon* L. Philosophical Transactions of the Royal Society of London Series B-Biological Sciences 254: 37-89.

Colosimo, P.F., K.E. Hosemann, S. Balabhadra, G. Villarreal, Jr., M. Dickson *et al.*, 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 *et al.*, 2004 The genetic architecture of parallel armor plate reduction in threespine sticklebacks. PLoS Biol 2: e109.

Coyle, S.M., F.A. Huntingford, and C.L. Peichel, 2007 Parallel evolution of *Pitx1* underlies pelvic reduction in Scottish threespine stickleback (*Gasterosteus aculeatus*). J Hered 98: 581-586.

Cresko, W.A., A. Amores, C. Wilson, J. Murphy, M. Currey *et al.*, 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.

Cubbage, C.C., and P.M. Mabee, 1996 Development of the cranium and paired fins in the zebrafish *Danio rerio* (*Ostariophysi, cyprinidae*). Journal of Morphology 229: 121-160.
Darwin, C., 1877 *The different forms of flowers on plants of the same species*. London: J. Murray.
Davidson, E.H., and M.S. Levine, 2008 Properties of developmental gene regulatory networks. Proc Natl Acad Sci U S A 105: 20063-20066.
deVicente, M.C., and S.D. Tanksley, 1993 QTL analysis of transgressive segregation in an interspecific tomato cross. Genetics 134: 585-596.
Dupuis, J., and D. Siegmund, 1999 Statistical methods for mapping quantitative trait loci from a dense set of markers. Genetics 151: 373-386.
Falconer, D.S., 1989 *Introduction to quantitative genetics*. New York: Pearson.
Feder, J.L., R. Gejji, T.H. Powell, and P. Nosil, 2011 Adaptive chromosomal divergence driven by mixed geographic mode of evolution. Evolution 65: 2157-2170.
Ferguson, L., S.F. Lee, N. Chamberlain, N. Nadeau, M. Joron *et al.*, 2010 Characterization of a hotspot for mimicry: assembly of a butterfly wing transcriptome to genomic sequence at the *HmYb/Sb* locus. Mol Ecol 19 Suppl 1: 240-254.
Fisher, R.A., 1928 The possible modification of the response of the wild type to recurrent mutations. American Naturalist 62: 115-126.
Fishman, L., Stathos, A., Beardsley, P.M., Williams, C.F., and Hill, J.P., 2013 Chromosomal rearrangements and the genetics of reproductive barriers in *Mimulus* (monkeyflowers). Evolution 67: 2547-2560.
Ghanem, N., O. Jarinova, A. Amores, Q. Long, G. Hatch *et al.*, 2003 Regulatory roles of conserved intergenic domains in vertebrate *Dlx* bigene clusters. Genome Res 13: 533-543.
Griswold, C.K., and M.C. Whitlock, 2003 The genetics of adaptation: the roles of pleiotropy, stabilizing selection and drift in shaping the distribution of bidirectional fixed mutational effects. Genetics 165: 2181-2192.
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.
Gross, H.P., and J.M. Anderson, 1984 Geographic variation in the gillrakers and diet of European threespine sticklebacks, *Gasterosteus aculeatus*. Copeia: 87-97.
Gruber, J.D., K. Vogel, G. Kalay, and P.J. Wittkopp, 2012 Contrasting properties of gene-specific regulatory, coding, and copy number mutations in *Saccharomyces cerevisiae*: frequency, effects, and dominance. PLoS Genet 8: e1002497.
Hagen, D.W., and L.G. Gilbertson, 1972 Geographic variation and environmental selection in *Gasterosteus aculeatus* L. in the Pacific Northwest, America. Evolution 26: 32-51.
Haldane, J.B.S., 1927 A mathematical theory of natural and artificial selection, Part V: Selection and mutation. Proceedings of the Cambridge Philosophical Society 23: 838-844.
Hermann, K., U. Klahre, M. Moser, H. Sheehan, T. Mandel *et al.*, 2013 Tight genetic linkage of prezygotic barrier loci creates a multifunctional speciation island in petunia. Curr Biol 23: 873-877.
Hohenlohe, P.A., S. Bassham, P.D. Etter, N. Stiffler, E.A. Johnson *et al.*, 2010 Population genomics of parallel adaptation in threespine stickleback using sequenced RAD tags. PLoS Genet 6: e1000862.
Hoogland, R.D., Morris, D. and Tinbergen, N. , 1957 The spines of sticklebacks (*Gasterosteus* and *Pygosteus*) as means of defense against predators (*Perca* and *Esox*). Behaviour 10: 205-230.

Hunt, G., M.A. Bell, and M.P. Travis, 2008 Evolution toward a new adaptive optimum: phenotypic evolution in a fossil stickleback lineage. Evolution 62: 700-710.
Jones, F.C., Y.F. Chan, J. Schmutz, J. Grimwood, S.D. Brady *et al.*, 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 *et al.*, 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 *et al.*, 2011 Chromosomal rearrangements maintain a polymorphic supergene controlling butterfly mimicry. Nature 477: 203-206.
Joron, M., R. Papa, M. Beltran, N. Chamberlain, J. Mavarez *et al.*, 2006 A conserved supergene locus controls colour pattern diversity in *Heliconius* butterflies. Plos Biology 4: 1831-1840.
Kenney-Hunt, J.P., B. Wang, E.A. Norgard, G. Fawcett, D. Falk *et al.*, 2008 Pleiotropic patterns of quantitative trait loci for 70 murine skeletal traits. Genetics 178: 2275-2288.
Kimmel, C.B., W.A. Cresko, P.C. Phillips, B. Ullmann, M. Currey *et al.*, 2012 Independent axes of genetic variation and parallel evolutionary divergence of opercle bone shape in threespine stickleback. Evolution 66: 419-434.
Kimmel, C.B., B. Ullmann, C. Walker, C. Wilson, M. Currey *et al.*, 2005 Evolution and development of facial bone morphology in threespine sticklebacks. Proc Natl Acad Sci U S A 102: 5791-5796.
Kimura, T., A. Shimada, N. Sakai, H. Mitani, K. Naruse *et al.*, 2007 Genetic analysis of craniofacial traits in the medaka. Genetics 177: 2379-2388.
Kingsley, D.M., B.L. Zhu, K. Osoegawa, P.J. De Jong, J. Schein *et al.*, 2004 New genomic tools for molecular studies of evolutionary change in threespine sticklebacks. Behaviour 141: 1331-1344.
Kingsley, D.M.a.C.L.P., 2007 The molecular genetics of evolutionary change in sticklebacks, pp. 41-81 in *Biology of the three-spine stickleback*, edited by I.M.a.F.A.H. S. Ostlund-Nilsson. CRC Press
Kirkpatrick, M., and N. Barton, 2006 Chromosome inversions, local adaptation and speciation. Genetics 173: 419-434.
Kislalioglu, M., and R.N. Gibson, 1977 Feeding relationship of shallow water fishes in a Scottish sea loch. Journal of Fish Biology 11: 257-266.
Kitano, J., Mori, S., and Peichel, C.L., 2007 Sexual dimorphism in the external morphology of the threespine stickleback (*Gasterosteus aculeatus*). Copeia: 336-349.
Kitano, J., J.A. Ross, S. Mori, M. Kume, F.C. Jones *et al.*, 2009 A role for a neo-sex chromosome in stickleback speciation. Nature 461: 1079-1083.
Knecht, A.K., K.E. Hosemann, and D.M. Kingsley, 2007 Constraints on utilization of the EDA-signaling pathway in threespine stickleback evolution. Evol Dev 9: 141-154.
Kurian, V., and A.J. Richards, 1997 A new recombinant in the heteromorphy '*S*' supergene in *Primula*. Heredity 78: 383-390.
Lavin, P.A., and J.D. Mcphail, 1986 Adaptive divergence of trophic phenotype among freshwater populations of the threespine stickleback (*Gasterosteus aculeatus*). Canadian Journal of Fisheries and Aquatic Sciences 43: 2455-2463.

Lemos, B., L.O. Araripe, P. Fontanillas, and D.L. Hartl, 2008 Dominance and the evolutionary accumulation of *cis*- and *trans*-effects on gene expression. Proc Natl Acad Sci U S A 105: 14471-14476.

Li, J., M.A. Webster, M.C. Smith, and P.M. Gilmartin, 2011 Floral heteromorphy in *Primula vulgaris*: progress towards isolation and characterization of the *S* locus. Ann Bot 108: 715-726.

Lindsey, C.C., 1962 Experimental study of meristic variation in a population of threespine stickleback, *Gasterosteus aculeatus*. Can. J. Zool. 40: 271-312.

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.

Mallet, J., 1989 The genetics of warning colour in Peruvian hybrid zones of *Heliconius Erato* and *H. Melpomene*. Proceedings of the Royal Society of London Series B-Biological Sciences 236: 163-185.

Mather, K., 1950 The genetical architecture of heterostyly in *Primula Sinensis*. Evolution 4: 340-352.

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.

McManus, C.J., J.D. Coolon, M.O. Duff, J. Eipper-Mains, B.R. Graveley *et al.*, 2010 Regulatory divergence in Drosophila revealed by mRNA-seq. Genome Res 20: 816-825.

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.

Miller, C.T., S. Beleza, A.A. Pollen, D. Schluter, R.A. Kittles *et al.*, 2007 *cis*-Regulatory changes in *Kit ligand* expression and parallel evolution of pigmentation in sticklebacks and humans. Cell 131: 1179-1189.

Minoux, M., and F.M. Rijli, 2010 Molecular mechanisms of cranial neural crest cell migration and patterning in craniofacial development. Development 137: 2605-2621.

Murray, J., and B. Clarke, 1976 Supergenes in polymorphic land snails I. *Partula Taeniata*. Heredity 37: 253-269.

Noor, M.A., A.L. Cunningham, and J.C. Larkin, 2001 Consequences of recombination rate variation on quantitative trait locus mapping studies: Simulations based on the *Drosophila melanogaster* genome. Genetics 159: 581-588.

Orr, H.A., 1991 A test of Fisher's theory of dominance. Proc Natl Acad Sci U S A 88: 11413-11415.

Orr, H.A., and A.J. Betancourt, 2001 Haldane's sieve and adaptation from the standing genetic variation. Genetics 157: 875-884.

Peichel, C.L., K.S. Nereng, K.A. Ohgi, B.L. Cole, P.F. Colosimo *et al.*, 2001 The genetic architecture of divergence between threespine stickleback species. Nature 414: 901-905.

Protas, M., I. Tabansky, M. Conrad, J.B. Gross, O. Vidal *et al.*, 2008 Multi-trait evolution in a cave fish, *Astyanax mexicanus*. Evol Dev 10: 196-209.

Reimchen, T.E., 1980 Spine deficiency and polymorphism in a population of *Gasterosteus aculeatus* - an adaptation to predators. Can. J. Zool. 58: 1232-1244.

Reimchen, T.E., and J.S. Nelson, 1987 Habitat and morphological correlates to vertebral number as shown in a teleost, *Gasterosteus aculeatus*. Copeia: 868-874.
Reimchen, T.E., E.M. Stinson, and J.S. Nelson, 1985 Multivariate differentiation of parapatric and allopatric populations of threespine stickleback in the Sangan River Watershed, Queen Charlotte Islands. Can. J. Zool. 63: 2944-2951.
Rieseberg, L.H., A. Widmer, A.M. Arntz, and J.M. Burke, 2002 Directional selection is the primary cause of phenotypic diversification. Proc Natl Acad Sci U S A 99: 12242-12245.
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.
Rockman, M.V., 2012 The QTN program and the alleles that matter for evolution: all that's gold does not glitter. Evolution 66: 1-17.
Rogers, S.M., P. Tamkee, B. Summers, S. Balabahadra, M. Marks *et al.*, 2012 Genetic signature of adaptive peak shift in threespine stickleback. Evolution 66: 2439-2450.
Ronfort, J., and S. Glemin, 2013 Mating system, Haldane's sieve, and the domestication process. Evolution 67: 1518-1526.
Schluter, D., 2000 *The ecology of adaptive radiation*. Oxford: Oxford University Press.
Schluter, D., K.B. Marchinko, R.D.H. Barrett, and S.M. Rogers, 2010 Natural selection and the genetics of adaptation in threespine stickleback. Philosophical Transactions of the Royal Society B-Biological Sciences 365: 2479-2486.
Schluter, D., and J.D. Mcphail, 1992 Ecological character displacement and speciation in sticklebacks. American Naturalist 140: 85-108.
Shapiro, M.D., M.E. Marks, C.L. Peichel, B.K. Blackman, K.S. Nereng *et al.*, 2004 Genetic and developmental basis of evolutionary pelvic reduction in threespine sticklebacks. Nature 428: 717-723.
Stern, D.L., 2000 Evolutionary developmental biology and the problem of variation. Evolution 54: 1079-1091.
Stern, D.L., and V. Orgogozo, 2008 The loci of evolution: how predictable is genetic evolution? Evolution 62: 2155-2177.
Streisfeld, M.A., and M.D. Rausher, 2011 Population genetics, pleiotropy, and the preferential fixation of mutations during adaptive evolution. Evolution 65: 629-642.
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.
Swain, D.P., 1992a The functional basis of natural selection for vertebral traits of larvae in the stickleback *Gasterosteus aculeatus*. Evolution 46: 987-997.
Swain, D.P., 1992b Selective predation for vertebral phenotype in *Gasterosteus aculeatus*: reversal in the direction of selection at different larval sizes. Evolution 46: 998-1013.
Tripathi, N., M. Hoffmann, E.M. Willing, C. Lanz, D. Weigel *et al.*, 2009 Genetic linkage map of the guppy, *Poecilia reticulata*, and quantitative trait loci analysis of male size and colour variation. Proceedings of the Royal Society B-Biological Sciences 276: 2195-2208.
Wagner, G.P., M. Pavlicev, and J.M. Cheverud, 2007 The road to modularity. Nat Rev Genet 8: 921-931.
Wark, A.R., M.G. Mills, L.H. Dang, Y.F. Chan, F.C. Jones *et al.*, 2012 Genetic architecture of variation in the lateral line sensory system of threespine sticklebacks. G3 2: 1047-1056.

Willacker, J.J., F.A. von Hippel, P.R. Wilton, and K.M. Walton, 2010 Classification of threespine stickleback along the benthic-limnetic axis. Biol J Linn Soc Lond 101: 595-608.

Winge, O., 1927 The location of eighteen genes in *Lebistes reticulatus*. Journal of Genetics 18: 1-43.

Yeaman, S., and M.C. Whitlock, 2011 The genetic architecture of adaptation under migration-selection balance. Evolution 65: 1897-1911.

Zerucha, T., T. Stuhmer, G. Hatch, B.K. Park, Q. Long *et al.*, 2000 A highly conserved enhancer in the *Dlx5/Dlx6* intergenic region is the site of cross-regulatory interactions between *Dlx* genes in the embryonic forebrain. J Neurosci 20: 709-721.

Chapter 3:
Parallel developmental genetic features underlie stickleback gill raker evolution

The following chapter was originally published as an article:
EvoDevo 2014 5:19.

Andrew M. Glazer, Phillip A. Cleves, Priscilla A. Erickson, Angela Y. Lam and Craig T. Miller

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

ABSTRACT

Convergent evolution, the repeated evolution of similar phenotypes in independent lineages, provides natural replicates to study mechanisms of evolution. Cases of convergent evolution might have the same underlying developmental and genetic bases, implying that some evolutionary trajectories might be predictable. In a classic example of convergent evolution, most freshwater populations of threespine stickleback fish have independently evolved a reduction of gill raker number to adapt to novel diets. Gill rakers are a segmentally reiterated set of dermal bones important for fish feeding. A previous large quantitative trait locus (QTL) mapping study using a marine x freshwater F2 cross identified QTL on chromosomes 4 and 20 with large effects on evolved gill raker reduction. By examining skeletal morphology in adult and developing sticklebacks, we find heritable marine/freshwater differences in gill raker number and spacing that are specified early in development. Using the expression of the *Ectodysplasin receptor (Edar)* gene as a marker of raker primordia, we find that the differences are present before the budding of gill rakers occurs, suggesting an early change to a lateral inhibition process controlling raker primordia spacing. Through linkage mapping in F2 fish from crosses with three independently derived freshwater populations, we find in all three crosses QTL overlapping both previously identified QTL on chromosomes 4 and 20 that control raker number. These two QTL affect the early spacing of gill raker buds. Collectively, these data demonstrate that parallel developmental genetic features underlie the convergent evolution of gill raker reduction in freshwater sticklebacks, suggesting that even highly polygenic adaptive traits can have a predictable developmental genetic basis.

INTRODUCTION

Convergent evolution, the repeated evolution of similar phenotypes in different lineages, provides evolutionary replicates to test for possible constraints on evolutionary trajectories. This repeated evolution of similar traits has been observed for a wide variety of morphological (e.g. Albert et al. 1992; Bossuyt and Milinkovitch 2000; Donley et al. 2004; Kolbe et al. 2011), physiological (Chen et al. 1997; Christin et al. 2008; Dobler et al. 2012; McCracken et al. 2009; Tishkoff et al. 2007), and behavioral (Blackledge and Gillespie 2004; Hodgkin and Doniach 1997; Kowalko et al. 2013) traits. Numerous recent studies in a variety of microbes, plants, and animals have begun to address the extent to which convergent phenotypic evolution occurs via parallel genetic bases (reviewed in Conte et al. 2012; Martin and Orgogozo 2013; Nadeau and Jiggins 2010; Stern 2013; but see Arendt and Reznick 2008). One striking conclusion from these studies is that convergent evolution often occurs via parallel genetic mechanisms, with the same genomic regions, genes, and sometimes even alleles used for evolutionary change. This genetic parallelism of convergent evolution has been observed in naturally (Calboli et al. 2003; Gross et al. 2009; Perry et al. 2007; Protas et al. 2006; Reed et al. 2011; Sucena et al. 2003) and artificially (Andersson et al. 2012; Parker et al. 2009; Sutter et al. 2007) selected populations of animals, as well as in plants (Cooley et al. 2011; Johanson et al. 2000; Streisfeld and Rausher 2009; Yoon and Baum 2004), and experimentally evolved microbes (Herring et al. 2006; Tenaillon et al. 2012; van Ditmarsch et al. 2013). These common phenomena of convergent and parallel evolution suggest that some evolutionary trajectories are constrained and perhaps even predictable.

Why some evolved phenotypes appear to have a predictable genetic basis remains a major unanswered question in biology, but could result from topology of genetic networks, constraints to developmental programs, constraints to available genetic variation, correlated response to selection on another trait, or even coincidence (Jeffery 2009; Kopp 2009; Losos 2011; Stern 2000, 2013). One test for parallelism underlying convergent phenotypes is to compare how convergent traits arise during development, as distinct (non-parallel) developmental processes can generate convergent phenotypes (Tanaka et al. 2009). The developmental processes affected by most evolved morphological traits are in general poorly understood, although several recent studies have begun to examine the developmental trajectories of evolved traits (Cooley et al. 2012; Manceau et al. 2011; Sanger et al. 2012; Yamamoto et al. 2009).

The adaptive radiation of the threespine stickleback (*Gasterosteus aculeatus)* provides a powerful system to study convergent evolution (reviewed in Bell and Foster 1994). Ancestral marine populations of sticklebacks have repeatedly colonized and adapted to countless freshwater lakes and streams throughout the Northern Hemisphere. Despite their evolved differences, ancestral marine and derived freshwater forms can be crossed, generating viable and fertile hybrid offspring, allowing for forward genetic crosses to map genomic regions controlling evolved change. Genetic studies from this system have revealed that the same genomic regions (Colosimo et al. 2004; Coyle et al. 2007; Cresko et al. 2004; Shapiro et al. 2004), genes (Chan et al. 2010), and even alleles (Colosimo et al. 2005; Miller et al. 2007) can be reused in freshwater adaptation. However, these previously studied traits (pelvic skeleton, lateral plates, and pigmentation) are each primarily controlled by a single large-effect locus that explains over half of the variance in the trait. One outstanding question in evolutionary biology is whether rules of

traits with a relatively simple genetic basis apply to traits with a more polygenic basis, which are much more common in nature (reviewed in Rockman 2012). In particular, the degree to which highly polygenic traits evolve using a parallel genetic basis is largely unknown. Intriguingly, recent genome-wide genotyping and genome resequencing studies in sticklebacks have identified striking re-use of many genomic variants during freshwater adaptation (Hohenlohe et al. 2010; Jones et al. 2012a; Jones et al. 2012b). These results suggest that parallel genetic evolution is common in sticklebacks, perhaps through reuse of adaptive variants of both large and small effect (Bell and Aguirre 2013; Schluter and Conte 2009).

A classic set of phenotypes studied by evolutionary biologists are trophic traits, as many radiations (e.g. the Galapagos finches, African cichlids, threespine sticklebacks) display striking correlations between a population's craniofacial pattern and the diet they eat (Albertson et al. 2003; Albertson et al. 2005; Bernatchez and Dodson 1990; Grant and Grant 2006; Hulsey et al. 2008; Mallarino et al. 2011; McGee and Wainwright 2013; Muschick et al. 2012; Ruber et al. 1999; Schluter and McPhail 1992; Schluter et al. 1985). In fish, the patterning of gill rakers, a segmentally reiterated set of dermal bones important for feeding, often correlates with a population's diet composition and prey size (Magnuson and Heitz 1971; reviewed in Schluter 2000). Gill raker number predicts feeding efficiency, with high gill raker counts correlating with better foraging on zooplankton (Roesch et al. 2013; Robinson 2000), and low gill raker counts correlating with better foraging on benthos (Lundsgaard-Hansen et al. 2013). While marine sticklebacks primarily eat small zooplankton suspended in the water column, freshwater sticklebacks typically eat larger prey items (Gross and Anderson 1984; Kislalioglu and Gibson 1977). Correlated with this dietary shift, many freshwater stickleback populations have evolved a reduction in gill raker number. Gill raker reduction has been documented in over a hundred independently derived stickleback populations (and likely has evolved thousands of times) from three main ecological contrasts: marine vs. freshwater, limnetic vs. benthic zones within a lake, and lake vs. adjoining inlet or outlet streams(Berner et al. 2009; Gross and Anderson 1984; Hagen and Gilbertson 1972; McPhail 1993; Moodie and Reimchen 1976; Raeymaekers et al. 2007; Reimchen et al. 1985). This repeated evolution of gill raker reduction throughout the Northern Hemisphere suggests that gill raker number is under strong natural selection.

Gill raker number in both marine and freshwater populations is highly heritable (Aguirre et al. 2004; Day et al. 1994; Hagen 1967, 1973; Hermida et al. 2002). Tests of phenotypic plasticity have revealed that gill raker number, unlike gill raker length, has no significant plastic response to a shifted diet (Day et al. 1994). Genetic studies in both sticklebacks (Peichel et al. 2001) and whitefish (Gagnaire et al. 2013; Rogers and Bernatchez 2007) have revealed that gill raker number is a polygenic trait, controlled by multiple quantitative trait loci (QTL). In a large F2 cross between marine fish from Japan and benthic freshwater fish from Paxton Lake, British Columbia, we previously mapped gill raker number and spacing to QTL on 17 chromosomes (Miller et al. 2014). Two large-effect QTL on chromosomes 4 and 20 QTL explained 23% and 25% of the variance of ventral gill raker patterning, respectively. Each of the additional modifier QTL had much weaker effects, explaining 3-8% of the variance of gill raker number or spacing.

Despite the well-established gill raker patterning differences in wild adult populations, little is known about the developmental basis of these patterning changes. Gill rakers appear to have genetic and developmental similarities to other vertebrate epithelial appendages, a broad class of periodically patterned organs that include hair, teeth, feathers, sweat glands, and scales (reviewed in Chuong 1998; Chuong et al. 2013). These structures form embryonically from

placodes—transient, regularly arrayed, epithelial thickenings that signal to underlying mesenchyme to make an epithelial organ (reviewed in Chuong 1998). *Ectodysplasin (Eda)* and the gene encoding the EDA receptor, *Ectodysplasin receptor (Edar)* play highly conserved roles in the development of placodes. Mice and humans with strong loss-of-function mutations in either gene have ectodermal dysplasia, with defects in teeth, hair, and sweat glands (reviewed in Sadier et al. 2014). During development of epithelial appendages, *Edar* is typically expressed in the placodes, flanked by a complementary expression pattern of *Eda* around the non-placode forming part of the field (Drew et al. 2007; Houghton et al. 2005; Laurikkala et al. 2002; Pispa et al. 2003; Tucker et al. 2004). Interestingly, in zebrafish, *Eda* and *Edar* are required for proper formation of gill rakers, as well as teeth and scales (Harris et al. 2008). In cichlid larvae, *Edar* is expressed within developing gill rakers and *Eda* is expressed between gill rakers (Fraser et al. 2008). This shared genetic requirement and complementary expression pattern of *Eda* and *Edar* suggests that gill rakers and other epithelial appendages develop by similar co-opted developmental genetic regulatory networks.

Here we examine a time course of gill raker number and spacing in developing stickleback fry from multiple populations to test whether the convergent evolution of gill raker reduction has evolved by parallel developmental mechanisms. We also test the hypothesis that convergent reduction of gill raker number has a parallel genetic basis involving QTL on chromosomes 4 and 20 using genetic crosses between fish from a marine population and three independently derived freshwater populations.

RESULTS

Heritable evolution of differences in gill raker pattern in three freshwater populations

To test whether multiple freshwater populations have evolved a heritable change in gill raker number and spacing, we compared skeletal morphology in marine and freshwater wild and lab-reared fish. Stickleback gill rakers were present in nine rows along the anterior-posterior axis, protruding anteriorly (odd rows) and posteriorly (even rows) from the five branchial arches (**Figure 3.1A**). They were also present in both ventral and dorsal domains (overlaying the ceratobranchial and epibranchial bones, respectively, **Figure 3.1B**). We first compared gill rakers from adult wild and lab-reared fish from the anadromous marine population from the Little Campbell River (LITC) in British Columbia to the Fishtrap Creek (FTC) freshwater population from Washington state (**Figure 3.2A-B**). These populations were previously described as having high and low gill raker counts, respectively, in the wild (Hagen 1967; Hagen and Gilbertson 1972). We observed highly significant differences in ventral gill raker number and spacing between marine LITC and freshwater FTC fish for both wild and lab-reared fish ($P < 10^{-10}$) for each comparison by Tukey's HSD test; **Figures 3.2C-D, 3.3**). In lab-reared fish, mean LITC raker number was 41% higher than FTC, with a concomitant 40% increase in mean FTC raker spacing compared to LITC (measured from center-to-center, **Figure 3.1B**). Next we examined lab-reared fish from two additional freshwater populations: Bear Paw Lake in Alaska and benthic fish from Paxton Lake, British Columbia (PAXB). PAXB wild and lab-reared fish have been previously characterized as low-rakered (Day et al. 1994; McPhail 1992). Lab-reared BEPA fish have also been described as low-rakered (Bell and Aguirre 2013). As with FTC, we also observed highly significant differences between marine LITC and freshwater PAXB and BEPA lab-reared fish for both ventral raker number and spacing ($P < 10^{-10}$ for each comparison

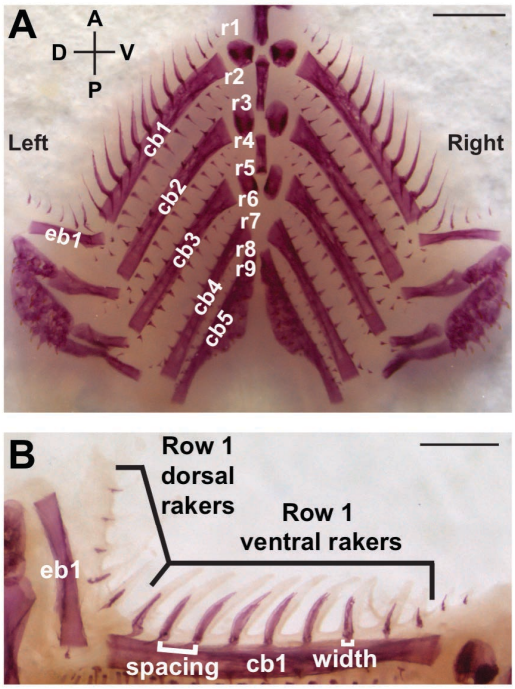


Figure 3.1 Diagram of gill raker domains in the stickleback branchial skeleton
A) Adult Alizarin red-stained stickleback branchial skeleton. Gill rakers are present in nine anterior-posterior rows (r1-r9). They protrude anteriorly and posteriorly from ventral ceratobranchials 1-4 (cb1-4), epibranchials 1-4 (eb1-4), and anteriorly from ceratobranchial 5 (cb5). A = anterior, P = posterior, D = dorsal, V = ventral. Scale bar = 1 mm.
B) Adult Alizarin red-stained stickleback branchial skeleton, zoomed in on left side row 1 gill rakers. Ventral and dorsal gill rakers protrude anteriorly from ceratobranchial 1 (cb1) and epibranchial 1 (eb1), respectively. Raker spacing measurements were obtained by measuring the mean center-to-center distance of all ventral rakers. Raker width measurements were obtained by measuring the width of the Alizarin-positive region of the raker base. Scale bar = 500 um.

by Tukey's HSD test; **Figure 3.2C-D**). Across the lab-reared and wild datasets ventral row 1 raker number and spacing were generally moderately anti-correlated (**Figure 3.4**). These data show that relative to ancestral marine fish, fish from these three derived freshwater populations have convergently evolved a heritable decrease in adult gill raker number and increase in gill raker spacing.

Early developmental difference in marine/freshwater gill raker spacing
Although gill raker development has not been well studied, development of many other epithelial appendages involves a reaction-diffusion system of activators and inhibitors that control the regular size and spacing of placodes (Chuong et al. 2013; Jung et al. 1998; reviewed in Kondo and Miura 2010; Sick et al. 2006). Therefore, we hypothesized that during gill raker development, freshwater fish have evolved differences in lateral inhibition, a developmental process where cells inhibit other nearby cells from adopting their same fate. The altered lateral inhibition hypothesis predicts that the raker primordia are spaced differently at the time of their

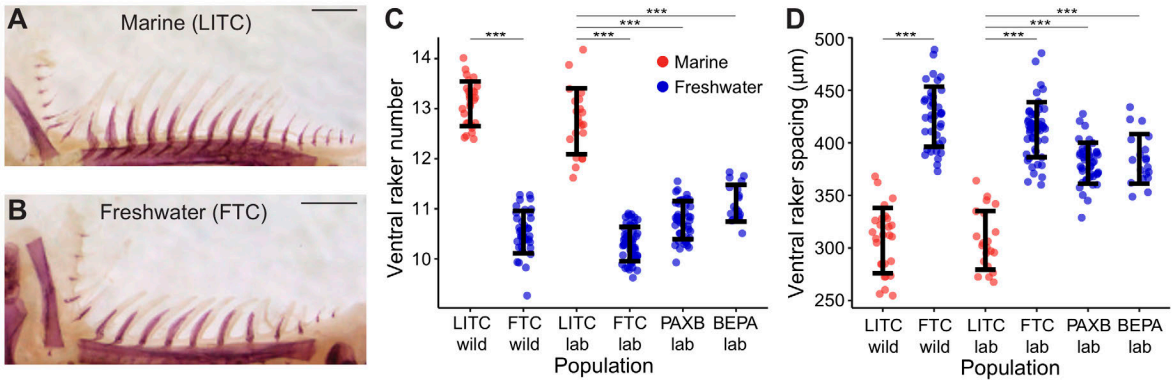


Figure 3.2: Heritable evolution of gill raker number and spacing in three freshwater populations
(A and B) Alizarin red-stained left anterior (row 1) gill rakers from an adult lab-reared Little Campbell (LITC) marine (A) and Fishtrap Creek (FTC) freshwater (B) fish. Scale bar = 500 um. Additional images labeling anatomical features of the branchial skeleton are presented in **Figure 3.1**. (C and D) Mean row 1-9 ventral gill raker number (C) and left row 1 ventral gill raker spacing (D) for wild-caught and/or lab-raised fish from Little Campbell marine (LITC), Fishtrap Creek freshwater (FTC), Paxton Benthic freshwater (PAXB), or Bear Paw freshwater (BEPA). LITC and FTC wild raker number and spacing differences are maintained in lab-reared fish, and fish from the three freshwater populations (blue) have fewer gill rakers that are more widely spaced than fish from the marine population (red). Data in (D) are back-transformed residuals from a standard length regression for a mean length of 40 mm. Error bars depict mean +/- SD. *n* >= 19 per condition. *** $P < 0.001$, Tukey's HSD test.

first appearance, and that these spacing differences are maintained to adulthood. To test this hypothesis, we examined lab-reared fish from the LITC marine and FTC and PAXB freshwater populations. Fish from all three populations were raised to various stages of development and stained for cartilage and bone. Gill rakers were first apparent in ~6 mm total length (TL) fry as non-ossified buds of soft tissue that protruded from the ventral gill-bearing (branchial) arches (**Figure 3.5A**). As development proceeded, these buds grew outwards and dermal bone ossified inside the buds (**Figure 3.5B-C**). In all three populations, the number of ventral rakers in the anterior-most row (row 1) was largely fixed by the 20 mm total length stage. From the earliest point of raker ossification until adulthood, we observed consistent marine-freshwater differences in gill raker number (**Figure 3.5D**). Row 1 inter-raker spacing increased approximately linearly as the fish grew and was also consistently different between marine and freshwater fish throughout development, with freshwater fish having a larger distance between their rakers (**Figure 3.5E**). Throughout development, FTC fish had fewer gill rakers and larger inter-raker spacing than PAXB (**Figures 3.2C-D, 3.5D-E**), but both freshwater populations had fewer, more widely-spaced rakers than marine LITC fish from the earliest stage of the time courses ($P <$ 0.001, Tukey's HSD test of pre-15 mm TL fish). These results establish that evolved reductions in gill raker number in two independently derived freshwater populations arise mainly through a parallel early developmental increase in freshwater inter-raker spacing.

If stickleback freshwater raker reduction were due to an altered lateral inhibition process, the spacing of presumptive gill rakers would differ between marine and freshwater stickleback from the first point of specification, even before the morphological process of budding actually occurs. Therefore, we attempted to detect pre-budded gill rakers by *in situ* hybridization of

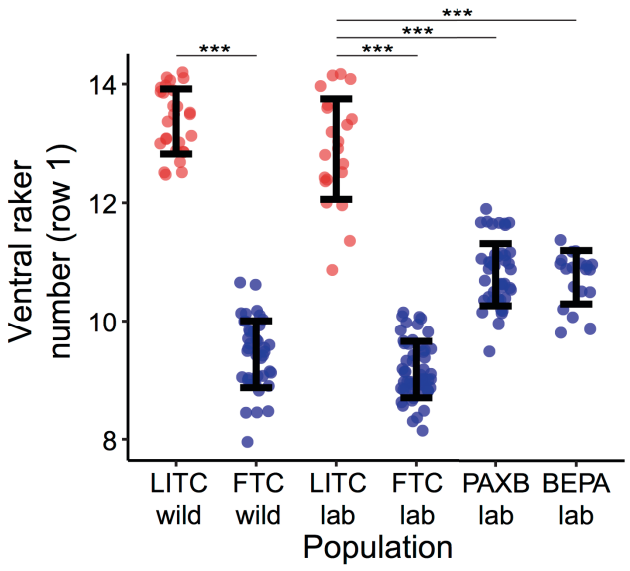


Figure 3.3: Heritable row 1 ventral gill raker reduction in three freshwater populations
Mean row 1 ventral gill raker number for wild-caught and/or lab-raised fish from Little Campbell marine (LITC), Fishtrap Creek freshwater (FTC), Paxton Benthic freshwater (PAXB), or Bear Paw freshwater (BEPA). LITC and FTC wild raker number differences are maintained in lab-reared fish, and fish from the three freshwater populations (blue) have fewer gill rakers than fish from the marine population (red). Compared to an average of all ventral rows (**Figure 3.2**), FTC is especially low-rakered in row 1. Error bars depict mean +/- SD. n >= 19 per condition. *** $P < 0.001$, Tukey's HSD test.

Ectodysplasin receptor (*Edar*), a gene required for gill raker formation in zebrafish (Harris et al. 2008) and a marker of developing gill rakers in cichlids (Fraser et al. 2008). In early-stage (~5.5 mm TL) branchial arches before rakers were visibly budded, *Edar* was detected broadly throughout pharyngeal endodermal epithelia, but appeared to have increased expression in periodic clusters of cells, which we interpreted as specified, pre-budded raker primordia (**Figure 3.6A-B**). We did not detect any specific staining pattern using a control *Edar* sense probe (data not shown). As gill rakers began to bud, *Edar* expression in the buds remained strong, in contrast to the inter-raker expression domains, which lost *Edar* expression (**Figure 3.6C-F**). From the earliest stage that we could detect *Edar*-positive gill raker primordia, we saw a significant difference in both the number of primordia and the spacing between primordia in LITC marine and FTC freshwater fish ($P < 0.001$, two-tailed t-test, **Figure 3.6G-H**). After adjusting for fish size, marine fish had a 45% increase ($P < 0.001$) in mean *Edar*+ foci number compared to freshwater fish (**Figure 3.7B**). There was a concomitant 32% increase ($P < 0.001$) in mean foci spacing in freshwater fish, strongly supporting altered lateral inhibition as a major factor contributing to primordia number differences (**Figure 3.7C**). However, marine fry also had a 15% increase ($P = 0.007$) in field size (the total length of the field containing *Edar*+ primordia) compared to freshwater fish, suggesting that raker primordia field size differences also exist between marine and freshwater fish (**Figure 3.7D**). Freshwater fry also had slightly wider *Edar*+ primordia (14% increase, $P = 0.03$, **Figure 3.7E**); however there was no significant difference between FTC and LITC row 1 raker width in adults ($P = 0.37$, **Figure 3.8**).

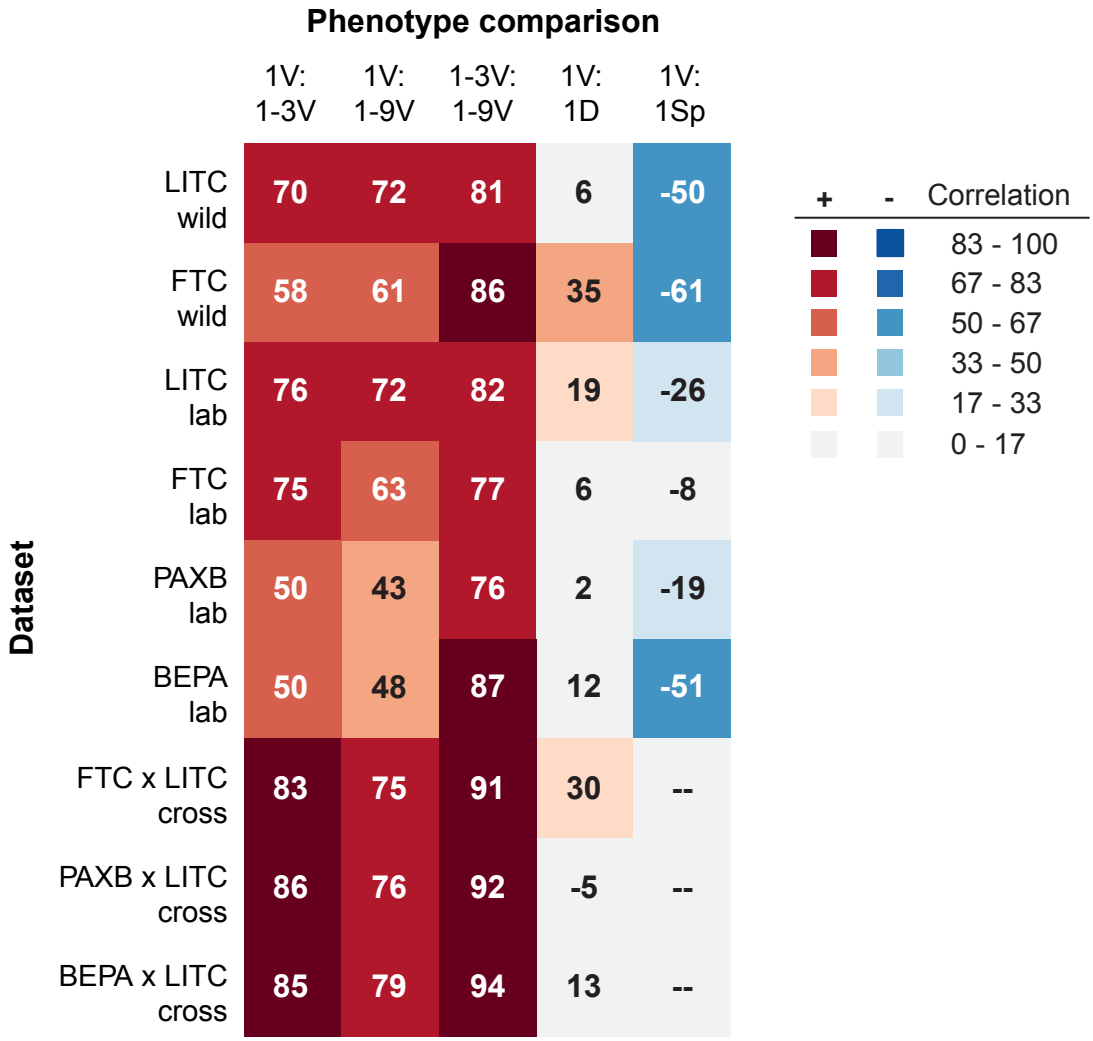


Figure 3.4: Correlations of raker number and spacing measurements
Pearson's correlation coefficients are presented for five comparisons between size and sex-adjusted (as appropriate) raker number and spacing phenotypes for all measured fish (wild and lab-reared datasets) or a sample of 100 fish (crosses). Correlations are presented as values multiplied by 100, i.e. 76 corresponds to a correlation of 0.76. Positive correlations are colored red and negative correlations are colored blue. Phenotypes are abbreviated: 1V = mean row 1 ventral number, 1-3V = mean row 1-3 ventral number, 1-9V = mean row 1-9 ventral number, 1D = mean row 1 dorsal number, 1Sp = left side row 1 spacing.

Parallel genetic and developmental effects of QTL on chromosomes 4 and 20

In a large F2 cross between Japanese marine and Paxton benthic freshwater fish, we previously mapped gill raker number to QTL on 17 chromosomes including the two largest-effect QTL on chromosomes 4 and 20 (Miller et al. 2014). To test the hypothesis that parallel reduction of gill raker number in multiple independently derived freshwater populations involved QTL on chromosomes 4 and 20, we raised three large F2 crosses (n=273, 384, and 418 fish) between PAXB, FTC, or BEPA grandparental freshwater fish each crossed to grandparental marine LITC

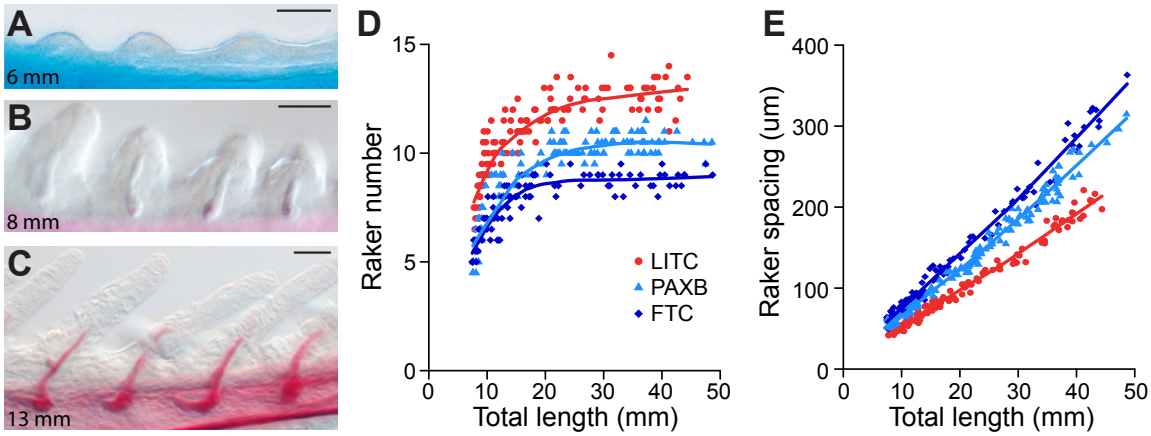


Figure 3.5: Marine/freshwater differences in gill raker number and spacing are specified early in development (A-C) Ventral row 1 developing gill raker buds, stained for cartilage (Alcian blue) and bone (Alizarin red). Scale bar = 25 um. Fish total length is indicated in bottom left. (D and E) Time course of lab-reared mean row 1 ventral raker number (D) or left side row 1 ventral raker spacing (E) vs. total length of fish. Red = Little Campbell marine (LITC), light blue = Paxton benthic freshwater (PAXB), dark blue = Fishtrap Creek freshwater (FTC).

fish. In each cross, we phenotyped gill raker number in each F2 fish. We also identified and genotyped a set of markers that were 1) polymorphic in at least two of three crosses and 2) spanned the previously identified QTL intervals and surrounding regions on chromosomes 4 and 20 (5-8 markers per chromosome per cross, **Tables 3.1, 3.2**). We found high correlations between the number of ventral gill rakers in different rows (**Figure 3.4**); therefore we phenotyped rows 1-3 in the entire set of F2s. Strikingly, we detected QTL with strong effects on gill raker number on chromosomes 4 and 20 in all three crosses (percent variance explained of 10-21% and 10-22% for chromosomes 4 and 20, respectively; **Figure 3.9, Table 3.3**). Furthermore, the localization of the two QTL overlapped in all three crosses (**Table 3.4, Figure 3.10**), and highly colocalized with the originally reported chromosome 4 and 20 gill raker QTL in each cross (Miller et al. 2014). The peak marker of the chromosome 4 QTL in the PAXB cross was Chr4_152, which was also the peak marker in the FTC cross (**Figure 3.9A**). In the BEPA cross, the peak marker of the chromosome 4 QTL was Chr4_131, a marker tightly linked (only 3.3 cM away, **Table 3.2**) to Chr4_152. Although the BEPA peak marker was different, there was a high degree of overlap between the 1.5 LOD intervals (an approximate 95% confidence interval, Dupuis and Siegmund 1999) of the chromosome 4 QTL in the BEPA cross and the PAXB and FTC crosses (**Table 3.4**). The peak marker of the chromosome 20 QTL in the PAXB cross was Stn216, which was also the peak marker in the BEPA cross (**Figure 3.9C**). In the FTC cross, the peak marker was Stn212, 0.4 cM away (**Table 3.2**) from Stn216 in this cross, having a peak LOD only 0.4 LOD units higher than Stn216 (**Figure 3.9C**). Thus, raker number mapped to largely overlapping genomic regions within chromosomes 4 and 20 in all three crosses.

To further test whether the two raker number QTL have parallel genetic features, we asked whether the QTL had similar properties of additivity and epistasis in each cross. In all three crosses, the chromosome 4 and 20 gill raker QTL had additive genetic effects with

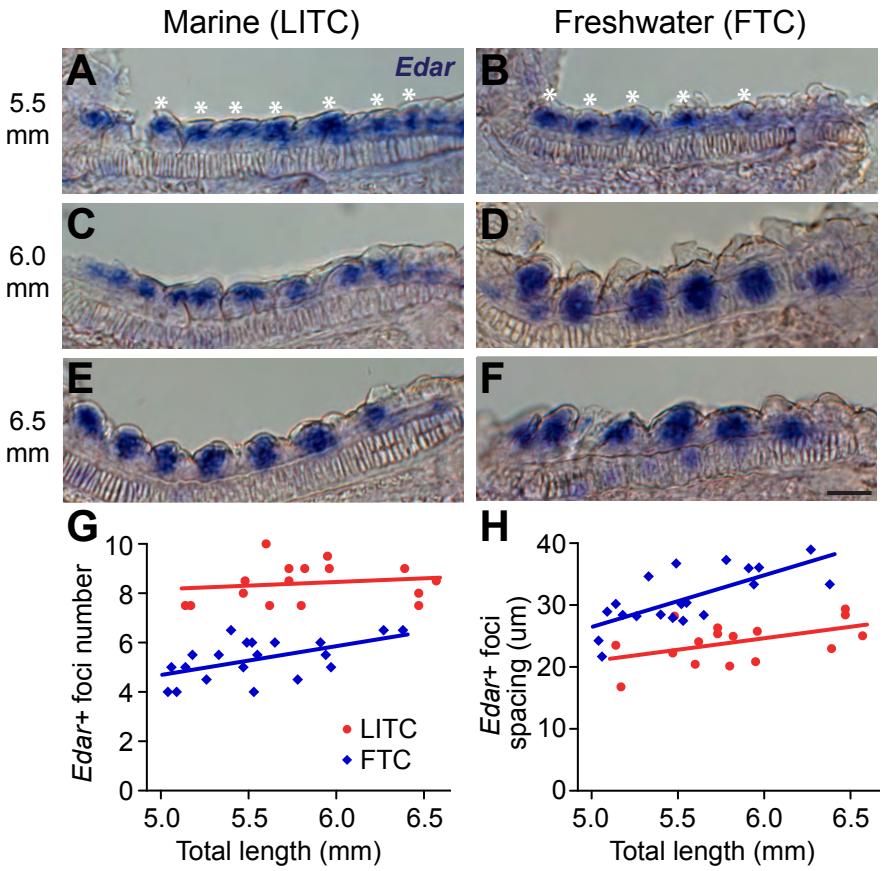


Figure 3.6: Marine/freshwater gill raker spacing differences are specified before gill raker budding (A-F) *Edar* expression in developing ventral row 1 raker primordia in pre-bud (A-D) and early bud (E-F) stages (fish total length in millimeters shown at left) in Little Campbell marine (LITC) (A,C,E) and Fishtrap Creek freshwater (FTC). Scale bar = 25 um.
(G and H) Significant differences in early bud stage ventral row 1 raker number (G) and spacing (H) between LITC (red) and FTC (blue) fish, detected by *Edar in situ* hybridization.

dominance values between -0.30 and 0.23 (dominances of -1, 0, or 1 represent a perfectly recessive, additive, or dominant effect, respectively, of the freshwater allele; **Table 3.4, Figure 3.9B,D**). Furthermore, in each cross, there were no significant epistatic interactions between the chromosome 4 and 20 QTL ($P = 0.18$, 0.37, and 0.10 for the PAXB, FTC, and BEPA crosses, respectively, for a Chromosome 4 peak genotype x Chromosome 20 peak genotype interaction term in an ANOVA).

Next, we asked whether the two raker number QTL have parallel developmental features. Gill rakers are present in both ventral and dorsal domains (**Figures 3.1, 3.2A-B**), and both ventral and dorsal gill raker numbers significantly differ between marine (LITC) and freshwater (FTC/PAXB/BEPA) lab-reared fish ($P < 0.001$ by Tukey's HSD, **Figures 3.2C, 3.11**). Despite the differences in lab-reared phenotypes, in all three crosses the effect of the chromosome 4 and 20 QTL was modular, with a much stronger effect on ventral raker number than dorsal raker number (**Table 3.5**). Consistent with this finding, ventral and dorsal raker numbers had low or no correlation in the three crosses and the lab-reared and wild datasets (**Figure 3.4**). Thus, both

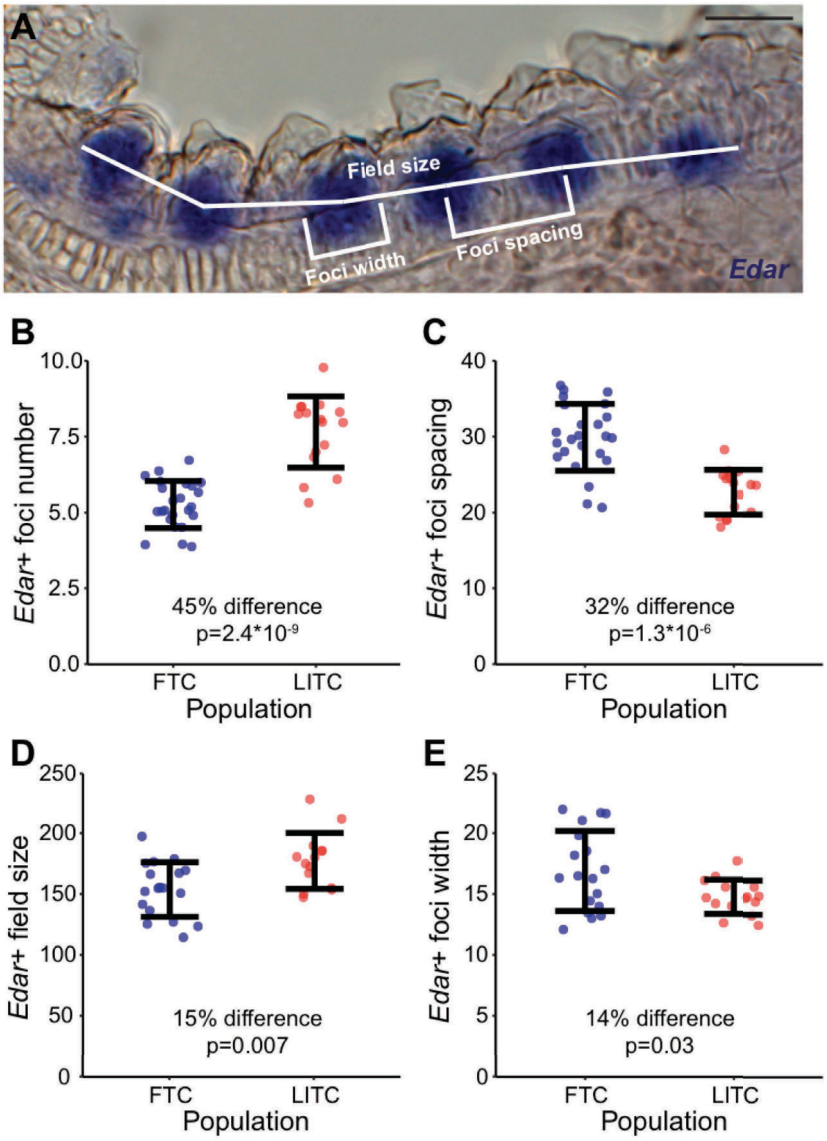


Figure 3.7: Pre-budding marine/freshwater differences in bud number, bud spacing, bud width, and field size
(A) *Edar* expression in developing ventral row 1 raker primordia in early bud stage (6.0 mm total length) fry. Landmarks used for f oci width, foci spacing, and field size are indicated. Scale bar = 25 um. (B-E) Significant differences in early bud stage ventral row 1 *Edar*+ foci number (B), foci spacing (C), field size (D), and foci width (E) between LITC (red) and FTC (blue) fish, detected by *Edar in situ* hybridization. Phenotypes are back transformed residuals for a regression to total length for a mean length of 5.5 mm. Error bars depict mean +/- SD. Displayed *P* values are from a two tailed t-test. Percent difference is from the ratio of mean marine and freshwater values.

raker number QTL display multiple genetic and developmental parallelisms in three independently derived freshwater populations.

Finally, we asked whether the chromosome 4 and 20 QTL affected gill rakers through a

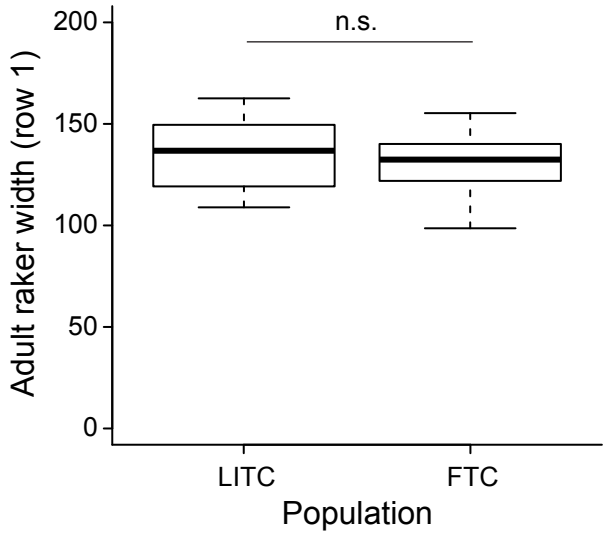


Figure 3.8: Adult marine and freshwater fish do not have significantly different gill raker widths Boxplot of row 1 ventral gill raker width for Little Campbell (LITC) marine and Fish Trap Creek (FTC) freshwater adult lab-reared fish. Values are represented as median +/- interquartile range. $n >= 12$ per population. n.s. = not significant ($P = 0.37$, two-tailed t-test). Refer to **Figure 3.1B** for a diagram of the landmarks used for raker width measurements.

similar developmental mechanism in different freshwater populations. We hypothesized that early in development, the chromosome 4 and 20 QTL were largely responsible for the altered relative strength of a lateral inhibition process controlling raker bud spacing. To test whether the chromosome 4 and 20 QTL directly controlled the early spacing of raker primordia, we raised 96 F2 fish from each of the PAXB x LITC and FTC x LITC crosses to an early stage of 19-20 days post fertilization (~8.5 mm total length), when early gill raker buds were still being specified. In both crosses, gill raker number and spacing were each controlled by both the chromosome 4 and 20 QTL at this early time point (**Figure 3.12** and **Table 3.6**, $P < 0.05$, two-tailed t-test between marine and freshwater homozygous classes). Thus, the differences in early raker patterning between marine and freshwater fish are due in large part to the early action of the chromosome 4 and 20 gill raker QTL, which control the early spacing of gill raker primordia in independently derived freshwater populations.

DISCUSSION

For polygenic quantitative traits that have evolved convergently, QTL mapping is a powerful first test of a parallel genetic architecture. While it is difficult to find the mutation(s) underlying these QTL, one prediction of parallel use of the same genes or genomic regions is that overlapping QTL would be found in multiple crosses from independently derived populations. A main finding of this study is that three independently derived freshwater populations have evolved a reduction in gill raker number involving QTL on chromosomes 4 and 20. Overlapping QTL on these chromosomes from three independently derived freshwater populations could be due to either the same genes underlying the QTL in each freshwater population, or different tightly linked genes in different populations. We parsimoniously hypothesize that the same genes

Marker	Forward sequence (5' to 3')	Reverse sequence (5' to 3')	Type	Ref.	Accession
Stn38	GCAGGTGACATCTTCAGGG	TTTCATTAGGACCCAGGACG	A	1	G72145.1
Gac4174	CCGCGATGATGAGAGTG	GTGAAATGCGACAGATGATG	A	2	NA
Stn45	ACGAGGGTTTGAGTCTCTCC	GTTGTTCAATCCATCCGTCC	B	1	G72247.1
Stn382	CCCTTAGAGAATTTCCTAGCA	CTTGTCCCGGATCATACGC	C	3	NA
Chr4_131	CAGAATGAGTCTCGATCCGC	GATCTCGGGTGTTTCTTTGC	A	This study	Pr032066746
Chr4_152	GGGATTGGAGATGAGTGGAA	TCTGGATGAAGTGTTGGTGG	A	This study	Pr032066747
Chr4_221	GTCCTGCTGGCACAGATCAT	TCATCAGCAGAGGCAGTGTT	B	This study	Pr032066748
Stn253	AACCACCCAGACCACTAAGC	ATGTCACGTATAGGTCGGCG	A	NA	BV678078.1
Chr4_280	CGTCCAGTACGTCCTAATCCC	AGGTCCGTGGTGAGCTAATG	B	This study	Pr032066749
Chr20_204	TGCAACAATATCAGGGACGA	TTATTCCTCTGCGTGGTGTG	A	This study	Pr032066744
Chr20_55	CGCTGTATCAACCAATGTGC	TTGATTCCGCTACACTTCCC	B	This study	Pr032066745
Stn212	TCATGGCATTCATAACCACGC	ACTCAGCTCGACTGTGTTGC	A	4	BV102490.1
Stn216	TGTGCAGTAGAGCAACAGCC	TGTTTCTGGCAGTAGGGTCC	A	4	BV102494.1
Chr20_155	CCGGTGAATAATGTCGAAGC	GGGCCACTCAATCAGTTCAT	B	This study	Pr032066742
Chr20_174	TTTCAATGGCTGTGCAGAAG	TGTTTACCAGCAGCGAGTTG	A	This study	Pr032066743

Table 3.1: Markers used in this study
Three methods of PCR were used in this study to genotype markers. Type A: 3 primer PCR. Method of [105] with M13F (TGTAAAACGACGGCCAGT) added to the 5' of the forward primer. Type B: Direct PCR. Forward primer directly labeled with a fluorophore (FAM/VIC/PET/NED). Type C: Unlabeled PCR. Primers not fluorescently labeled; analyzed by gel electrophoresis. References: (1) Peichel et al., 2001; (2) Largiader et al., 1999; (3) Colosimo et al., 2005; (4) Schuelke 2000.

in each population underlie the two QTL primarily because the QTL map to overlapping regions of chromosomes 4 and 20. Additional support of this hypothesis comes from several shared features of these QTL: the QTL (1) have similar genetic properties of additivity, (2) lack epistatic interactions with the other QTL, (3) have a modular effect with stronger effects on ventral than dorsal gill rakers, and (4) affect the same developmental process of early raker primordia spacing. Although multiple genetic changes underlie raker reduction in stickleback and whitefish genetic crosses (Gagnaire et al. 2013; Peichel et al. 2001; Rogers and Bernatchez 2007), the parallel involvement of QTL on chromosomes 4 and 20 in three independently derived freshwater stickleback populations suggests that the evolution of gill raker reduction is genetically constrained, and that properties of these two QTL bias them towards being selected to result in evolved changes in gill raker pattern. These properties could include the large additive phenotypic effects and specific developmental effects on the spacing of early gill raker primordia that we show here, but also could include pleiotropic effects (or the lack thereof) and/or standing allele frequencies in the oceanic population. The strongly additive effects of raker QTL in this study are consistent with previous findings of additivity for chromosome 4 and 20 gill raker QTL in a large marine x freshwater F2 cross (Miller et al.), and for row 1 total gill raker number in a marine x BEPA F1 cross (Bell and Aguirre 2013).

Repeated use of the same genes during stickleback adaptive radiation has been observed in previous genetic studies of traits with a simple genetic architecture that evolve repeatedly from either standing variation (Colosimo et al. 2005; Miller et al. 2007) or repeated mutation (Chan et al. 2010). However, to our knowledge, this study is the first to demonstrate multiple overlapping QTL controlling a convergently evolved trait in multiple independently derived freshwater stickleback populations. Recent genomic studies in sticklebacks suggest that adaptation through

Chr.	Marker	Physical Position (Mb)	Genetic Position (cM)		
			PAXBxLITC	FTCxLITC	BEPAxLITC
4	Stn38	3	0.0	0.0	0.0
4	Gac4174	11.6	27.2	25.9	NA
4	Stn45	11.7	28.0	25.9	NA
4	Stn382	12.8	28.4	26.8	31.4
4	Chr4_131	15.4	30.3	28.7	35.9
4	Chr4_152	16.9	32.9	29.8	39.2
4	Stn253	21.4*	33.8	30.4	44.6
4	Chr4_280	30.5	46.1	41.1	57.5
20	Chr20_204	18.6^	0.0	0.0	0.0
20	Chr20_55	4.5	12.4	14.6	23.5
20	Stn212	7.3	16.2	19.6	NA
20	Stn216	8.3	17.7	20.0	30.4
20	Chr20_155	13.5	21.3	22.5	33.4
20	Chr20_174	15.7	26.6	28.7	40.4

Table 3.2: Genetic maps of chromosome 4 and 20 used for adult QTL mapping
*The genomic region containing scaffolds 24 and 28 on chromosome 4 (containing Stn253) is inverted in the genome assembly (Roesti et al. 2013). ^Scaffold 46 containing marker Chr20_204 maps to the "left" end of chromosome 20 in all three crosses despite being on the right end of the genome assembly (higher coordinate in the genome assembly). Cytogenetic data are consistent with Scaffold 46 mapping to the left end of the chromosome (lower coordinate in the genome assembly) (Urton et al., 2001). Chr4_221, which was not used for adult QTL mapping, is located at 25.32 Mb in the genome assembly.

the reuse of identical genetic variants is strikingly widespread, although the phenotypes controlled by these reused variants are mostly unknown (Hohenlohe et al. 2010; Jones et al. 2012b). Intriguingly, in the Jones et al. set of standing variant regions under parallel selection in freshwater, chromosomes 4 and 20 are the two chromosomes with the most re-used standing variant regions, including several regions that overlap the raker QTL identified here (**Table 3.4**) (Jones et al. 2012b). Given the widespread use of standing variants in stickleback freshwater adaptation, we hypothesize that the two raker QTL are standing variants, present at low frequency in the oceanic population, that increase in frequency predictably upon freshwater colonization. This pattern, recently termed "collateral evolution," has largely been documented on traits with fairly simple genetic architectures (reviewed in Stern 2013). The extent to which collateral evolution is used for more complex, highly polygenic traits is poorly understood, but evolved gill raker reduction in sticklebacks provides a powerful system to address this question, especially since constant low levels of gene flow between oceanic and freshwater populations provide ample opportunities for adaptive alleles to be recycled and reused again during future freshwater colonizations (Bell and Aguirre 2013; Schluter and Conte 2009). Future work will test the hypothesis of collateral evolution of gill raker QTL by using next-generation sequencing approaches to look for genomic signatures of shared haplotypes that are under strong selection in multiple raker-reduced freshwater populations and present at low frequency in anadromous marine populations (Hohenlohe et al. 2010; Jones et al. 2012b). It will be especially interesting to compare the genetic and developmental mechanisms of evolved changes in gill raker number in

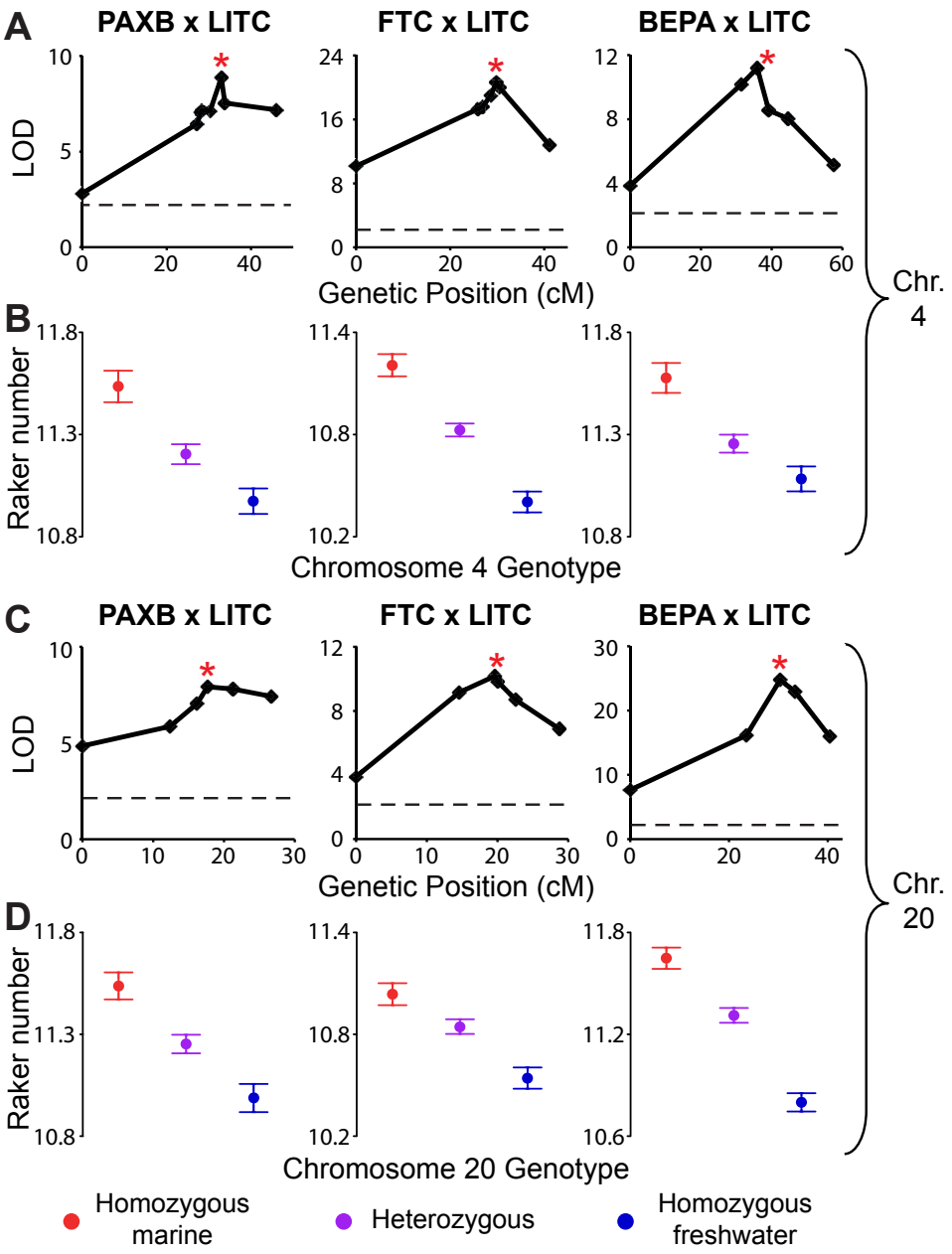


Figure 3.9: Two additive QTL control gill raker number in three crosses with independent freshwater populations

(A and C) Association of mean ventral row 1-3 gill raker number with chromosome 4 (A) or chromosome 20 (C) genotype. The peak marker in the Paxton benthic (PAXB) x Little Campbell marine (LITC) cross is indicated with red asterisks: Chr4_152 for (A) and Stn216 for (C). These two markers are also starred in the Fishtrap Creek (FTC) and Bear Paw Lake (BEPA) crosses. See **Table 3.2** for a list of which markers are present in each plot. (B and D) Mean ventral row 1-3 gill raker number by Chr4_152 (B) or Stn216 (D) genotype of F2s (homozygous marine, red; heterozygous, purple; and homozygous freshwater, blue). Phenotypes are back transformed residuals for a regression to standard length for a mean standard length of 40 mm. Values are presented as mean +/- SEM.

Cross	Chr.	LOD	PVE	Peak Marker	Phenotype Mean +/- Standard Error		
					MM	MF	FF
PAXB x LITC	4	8.87	12.5	Chr4_152	11.54 +/- 0.08	11.2 +/- 0.05	10.97 +/- 0.06
FTC x LITC	4	20.65	20.7	Chr4_152	11.21 +/- 0.07	10.83 +/- 0.04	10.4 +/- 0.06
BEPA x LITC	4	11.19	9.5	Chr4_131	11.58 +/- 0.07	11.26 +/- 0.04	11.08 +/- 0.06
PAXB x LITC	20	7.92	11.5	Stn216	11.54 +/- 0.07	11.25 +/- 0.05	10.99 +/- 0.07
FTC x LITC	20	10.18	9.7	Stn212	11.04 +/- 0.06	10.85 +/- 0.04	10.54 +/- 0.06
BEPA x LITC	20	24.77	22.4	Stn216	11.65 +/- 0.06	11.31 +/- 0.04	10.8 +/- 0.05

Table 3.3: Chromosome 4 and 20 QTL controlling mean ventral row 1 to 3 gill raker number in three marine × freshwater F2 crosses.
Chr.: chromosome, LOD: logarithm of the odds, PVE: percentage of phenotypic variance explained. Phenotype means and standard errors are given by genotypic class of F2 (MM is homozygous marine, MF is heterozygous, and FF is homozygous freshwater). Additional information on the properties of these QTL is presented in **Table 3.4**.

additional stickleback populations, as well as in other fish species that also evolve dramatic changes in gill raker counts in populations adapted to eat different diets (Schluter 2000; Schluter and McPhail 1993).

Although the two chromosome 4 and 20 QTL have many parallel features in multiple freshwater populations, there are still several unexplained aspects of the genetic and developmental basis of convergent gill raker evolution observed in this study. First, the effect sizes of the chromosome 4 and 20 QTL varied in the three crosses. Although some variation in effect size is to be expected by chance, this phenomenon could also suggest that different genetic backgrounds might modulate the effects of the QTL (for example due to varying numbers of other modifier QTL and/or epistatic interactions with other QTL). Alternatively, the variation in effect size could reflect different underlying genetic bases in the different freshwater populations.

A second unexplained feature of gill raker reduction observed in this study is the modularity of dorsal and ventral gill rakers: both the chromosome 4 and 20 QTL have much stronger effects on ventral gill raker number than dorsal gill raker number in all three crosses. In another cross, we previously also found extensive modularity along the dorsal-ventral axis in the genetic control of stickleback gill raker reduction (Miller et al. 2014). In the time course studies presented here, ventral gill rakers form much earlier during development than dorsal gill rakers, so this modularity might reflect the temporal window of developmental effect for the QTL, and/or regionally restricted (e.g. in ventral not dorsal primordia) expression of genes underlying the QTL. Regardless of the developmental genetic mechanism, this strong modularity of gill rakers even within row 1 suggests that separately phenotyping dorsal and ventral row 1 gill rakers might yield different results than summing the total of all row 1 rakers, as is commonly done in ecological and evolutionary studies.

A third unexplained genetic feature of gill raker reduction observed in this study is that one freshwater population (FTC) is more raker reduced than the others (BEPA and PAXB). Interestingly, much of this difference is due to a much stronger reduction of FTC in row 1 gill raker relative to other rows (compare **Figure 3.3** to **Figure 3.2C**), which might reflect

QTL	Adjustments	Threshold	LOD	PVE	1.5 LOD Boundaries			Effect size (a)	Hetero-zygous effect (d)	Dominance (d/a)	Mean +/- Standard Error			Overlapping Jones et al. Regions (kb)
					Left	Peak	Right				MM	MF	FF	
PAXB x LITC 4	SL+Sex	2.21	8.87	12.5	Chr4_131	Chr4_152	Chr4_280	0.28	-0.05	-0.18	11.54 +/- 0.08	11.20 +/- 0.05	10.97 +/- 0.06	15.43, 15.48, 19.77, 19.82, 19.85, 19.88, 20.26, 21.61,
FTC x LITC 4	SL+Sex	2.23	20.65	20.7	Chr4_131	Chr4_152	Chr4_280	0.40	0.02	0.05	11.21 +/- 0.07	10.83 +/- 0.04	10.40 +/- 0.06	21.80, 22.13, 22.30, 23.37, 23.93, 23.94, 23.95, 23.97,
BEPA x LITC 4	SL+Sex	2.17	11.19	9.5	Stn 38	Chr4_131	Chr4_152	0.25	-0.07	-0.30	11.58 +/- 0.07	11.26 +/- 0.04	11.08 +/- 0.06	24.92, 25.22, 25.88, 26.02, 26.08, 26.16, 26.56, 28.35, 28.45
PAXB x LITC 20	SL+Sex	2.21	7.92	11.5	Chr20_55	Stn 216	Chr20_174	0.28	-0.01	-0.04	11.54 +/- 0.07	11.25 +/- 0.05	10.99 +/- 0.07	5.05, 5.19, 5.21, 5.93, 6.30, 7.07, 7.80, 8.35, 8.45, 8.46, 8.47, 8.79, 8.86, 8.87, 8.89, 8.95, 9.30, 11.58, 11.58, 11.61, 13.48
FTC x LITC 20	SL+Sex	2.23	10.18	9.7	Chr20_204	Stn 212	Chr20_174	0.25	0.06	0.23	11.04 +/- 0.06	10.85 +/- 0.04	10.54 +/- 0.06	
BEPA x LITC 20	SL+Sex	2.17	24.77	22.4	Chr20_55	Stn 216	Chr20_155	0.42	0.09	0.21	11.65 +/- 0.06	11.31 +/- 0.04	10.80 +/- 0.05	

Table 3.4: Summary of adult QTL

Statistics for QTL for average ventral rows 1-3 are shown. Genotypic classes of F2 fish are abbreviated: MM = homozygous marine, MF = heterozygous, FF = homozygous freshwater. LOD is the logarithm of the odds and PVE is the percentage of phenotypic variance explained. Genomic coordinates of regions of marine-freshwater divergence (Jones et al. 2012) that overlap with consensus QTL positions from this study are shown (Mb=megabases).

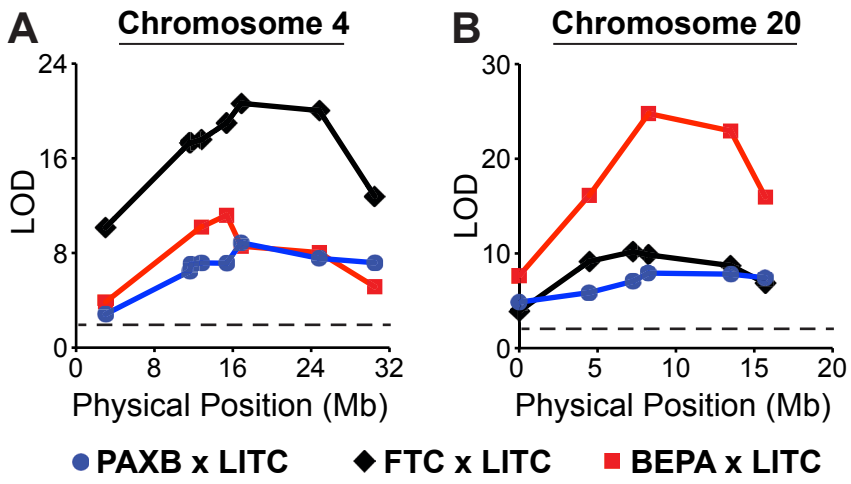


Figure 3.10: Physical positions of chromosome 4 and 20 QTL
(A and B) Association of mean ventral row 1-3 gill raker number with chromosome 4 (A) or chromosome 20 (B) genotype, plotted against adjusted physical position (genome assembly coordinates adjusted as previously described; see Methods). Refer to **Table 3.2** for a list of which markers are present in each plot.

differences in freshwater diets and/or available genetic variation. Future work using genome-wide linkage mapping in multiple F2 crosses will address the extent of genetic constraint, and whether smaller-effect modifier QTL also are repeatedly used to accomplish repeated gill raker reduction. We hypothesize that there are additional and/or stronger effect QTL controlling FTC gill rakers than in PAXB or BEPA, possibly including modular row 1-specific gill raker QTL present in FTC. In addition, further genotyping of the chromosome 4 and 20 QTL in the three crosses, while unlikely to change the main result presented here of overlapping QTL, may improve the resolution of these QTL.

Parallel developmental features underlie stickleback gill raker reduction

Another main finding of this study is that parallel developmental changes underlie convergent evolution of gill raker reduction. Despite the established adaptive significance of evolved changes in gill raker number and the recurrent phenomenon of this trait evolving across many fish clades (reviewed in Lindsey 1981; Schluter 2000), little was previously known about the developmental processes altered by these evolved genetic changes. Here, we find that gill raker spacing is increased in all three freshwater populations in adults, and dense developmental time courses in two of these populations reveal an early developmental increase in the spacing of gill raker primordia that is controlled by the chromosome 4 and 20 QTL. The adult difference in pattern is specified as an increased distance between the budding gill raker primordia at a surprisingly early stage – before hatching, raker ossification, and feeding. A previous study found no significant plastic response in gill raker number to shifts in diet (Day et al. 1994). This result, together with our findings that marine/freshwater gill raker number differences are fixed before the onset of feeding, suggests that gill raker number is largely genetically hard-wired at an early stage in development. This genetically programmed difference in spacing arises somewhere

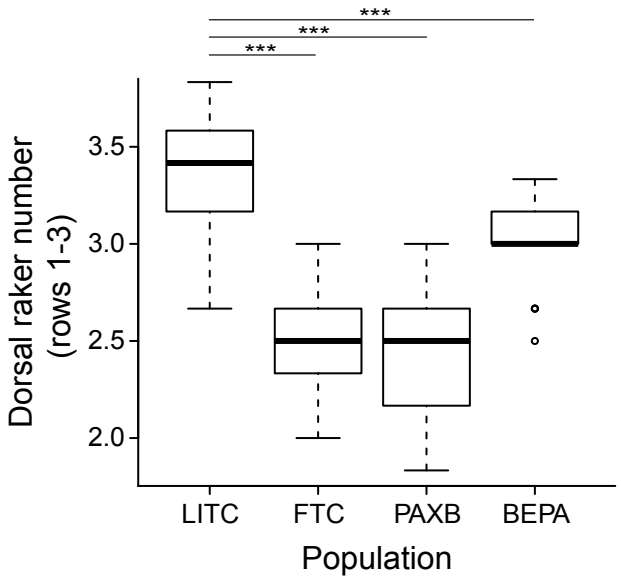


Figure 3.11: Lab-reared freshwater fish have fewer dorsal gill rakers than marine fish
Boxplot of mean row 1-3 dorsal gill raker number for Little Campbell (LITC) marine and Fish Trap Creek (FTC) freshwater, Paxton Benthic freshwater (PAXB), or Bear Paw freshwater (BEPA) adult lab-reared fish. Values are represented as median +/- interquartile range. $n >= 19$ per condition. *** $P < 0.001$, Tukey's HSD test.

upstream of the genetic regulatory networks controlling early spacing of *Ectodysplasin receptor* (*Edar*)-expressing raker primordia, although the precise location of the evolved changes in this pathway in different freshwater populations remains to be determined. A complementary expression pattern of *Edar* and *Eda* in gill raker buds and inter-raker domains, respectively (Fraser et al. 2008, this work) resembles the complementary expression patterns of *Edar* and *Eda* in other epithelial appendage bud and inter-bud domains (Drew et al. 2007; Houghton et al. 2005; Laurikkala et al. 2002; Pispa et al. 2003; Tucker et al. 2004), suggesting a shared genetic program for gill rakers and other epithelial appendages.

Evolved changes in patterning of epithelial appendages have occurred repeatedly during vertebrate evolution. For example, in human populations, a derived allele of the *EDAR* gene affecting hair, sweat gland, and mammary gland morphology underwent one of the strongest selective sweeps in the genome (Kamberov et al. 2013). Ectodysplasin signaling is perhaps used repeatedly during stickleback and human evolution because epithelial appendages are a "hot spot" for evolution, as they form and function at the interface between an organism and its environment (reviewed in Sadier et al. 2014).

The early difference in number and spacing of marine and freshwater *Edar*-positive raker primordia suggests that there is an evolved early-acting difference in a lateral inhibitory process. Early freshwater decreases in *Edar*-positive raker primordia number could be explained largely by increases in freshwater primordia spacing, but also to a smaller extent by a decrease in freshwater field size. Future work will attempt to discover which genes underlie stickleback gill raker reduction, and whether those genes affect lateral inhibition. In chickens, selection for fitness in hot climates resulted in the evolution of breeds with featherless necks, caused by the

Cross	Chromosome	Ventral LOD	Dorsal LOD
PAXB x LITC	4	8.9	1.3
FTC x LITC	4	20.7	0.6
BEPA x LITC	4	11.2	3.9
PAXB x LITC	20	7.9	0.8
FTC x LITC	20	10.2	0.9
BEPA x LITC	20	24.8	1.1

Table 3.5: Ventral modularity of raker number QTL
LOD scores (logarithm of the odds) for ventral (average rows 1-3) and dorsal (average rows 1-3) raker domains in three adult marine x freshwater F2 crosses.

upregulation of an inhibitory gene, *Bmp12*, during feather placode development (Mou et al. 2011). Since both activating and inhibitory genes (e.g. *Edar* and *Bmp12*) control the spacing of other epithelial appendages, mutations that contribute to an increase in the spacing of gill raker primordia could increase the strength of inhibitory genes, decrease the strength of activating genes, or both. Understanding the developmental and genetic mechanisms underlying stickleback gill raker evolution might further shed light on general principles of epithelial appendage evolution.

Pleiotropy and candidate genes

Parallel evolution of gill raker reduction might also be promoted by selection on another trait that is genetically controlled in a linked or pleiotropic manner to gill rakers. Interestingly, we have previously mapped trait clusters of several large-effect QTL controlling various skeletal phenotypes to chromosomes 4 and 20 (Miller et al. 2014). It is possible that the genes underlying the chromosome 4 and 20 gill raker QTL have a pleiotropic effect on multiple adaptive skeletal traits, or that these genes are tightly linked to genes that also confer adaptive phenotypes in freshwater environments, promoting the parallel use of gill raker QTL on chromosome 4 and 20.

One gene in particular, *Eda,* stands out as a candidate for playing a pleiotropic adaptive role in freshwater adaptation. *Eda* is located on chromosome 4, and has been identified as the principal gene underlying freshwater lateral plate reduction and marine/freshwater neuromast differences (Colosimo et al. 2005; Colosimo et al. 2004; Mills et al. 2014). In sticklebacks, the *Eda* genomic region has also been linked to multiple other phenotypes: behavioral preference for alternative salinities (Barrett et al. 2009), aspects of body shape (Albert et al. 2008), and schooling behavior, perhaps through effects on the lateral line (Greenwood et al. 2013; Wark et al. 2012). *Eda* plays pleiotropic roles during fish development, as zebrafish homozygous for strong loss-of-function alleles of *Eda* lack scales (homologous to lateral plates), as well as gill rakers, teeth, and fin rays (Harris et al. 2008). However, genetic resolution of the QTL argues strongly against the previously identified *Eda* haplotype controlling plate number (Colosimo et al. 2005) underlying the chromosome 4 gill raker QTL. In all three crosses, the peak marker of the chromosome 4 gill raker QTL is to the "right" (higher coordinate in the genome assembly) of *Eda* (which is located at Stn382) and in the FTC and PAXB crosses, the coding region of Eda lies well outside the 1.5 LOD interval. This mapping better supports candidate genes to the right of *Eda*, although it is possible that there is a long-range regulatory element of *Eda* that lies

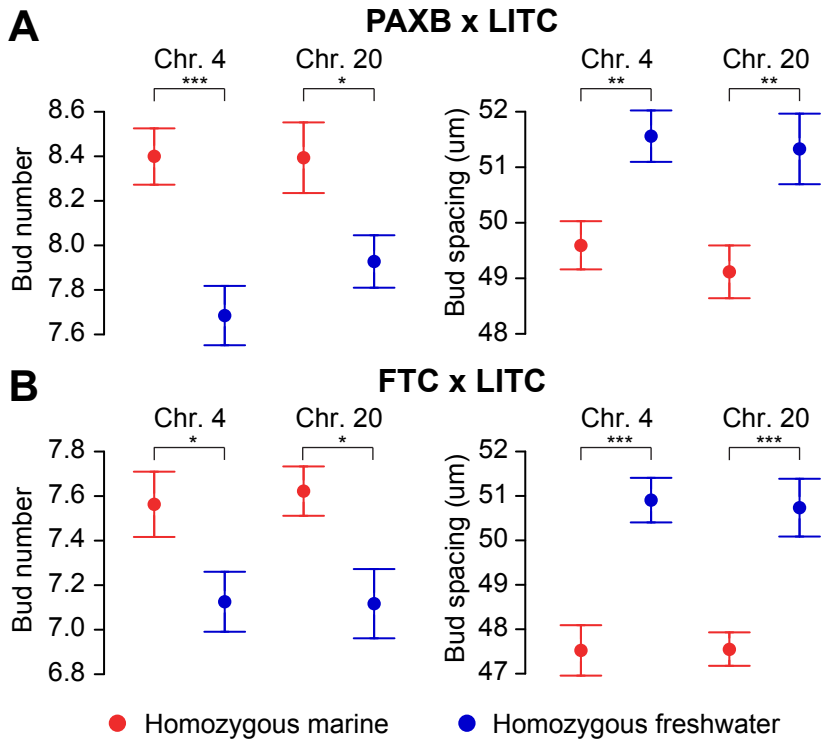


Figure 3.12: QTL on chromosomes 4 and 20 control the number and spacing of early raker buds
(A, B) Left ventral row 1 gill raker bud number (left) and left ventral row 1 to 7 raker bud spacing (right) in 20 dpf fry (approximately 8 mm total length). Mean phenotypes are displayed for F2s homozygous for the marine allele (red) or freshwater allele (blue) of a chromosome 4 marker (Chr4_221) or chromosome 20 marker (Stn212) very tightly linked to the peak marker in both crosses (**Table 3.2**). Phenotypes are back transformed residuals for a regression to total length for a mean length of 8 mm. N = 96 F2s from each of the Paxton benthic freshwater (PAXB) × Little Campbell marine (LITC) (A) and the Fishtrap Creek freshwater (FTC) × Little Campbell marine (LITC) (B) crosses were analyzed. * P <0.05, ** P <0.01, *** P <0.001 by a two-tailed t-test. Values are presented as mean +/- SEM. Additional information on these early time point QTL is presented in **Table 3.6**.

within the consensus QTL interval. Although the chromosome 4 and 20 QTL intervals are broad, several interesting candidate genes lie within the intervals that are members of important developmental signaling pathways known to play a role in epithelial appendage patterning. *Fgf20, Hes7*, *Fgf4,* and *Smad5* on chromosome 4 and *Hey1* and *Gsk3a* on chromosome 20 stand out as intriguing candidates given their roles in FGF, Notch, BMP, or WNT signaling.

METHODS

Stickleback crosses and care

Three marine x freshwater F1 crosses were generated: (1) a wild-caught anadromous marine male from the Little Campbell River (British Columbia, "LITC") fish was crossed to a wild-caught female from Fishtrap Creek (Washington state; "FTC"); (2) a male fish from Bear Paw

QTL	Phenotype	P value	PVE	Effect size (a)	Mean +/- Standard Error	
					MM	FF
PAXB x LITC 4	Number	$2.9x10^{-4}$	19.3	0.36	8.40 +/- 0.13	7.68 +/- 0.13
PAXB x LITC 4	Spacing	$3.2x10^{-3}$	15.4	-0.99	49.59 +/- 0.43	51.56 +/- 0.46
PAXB x LITC 20	Number	0.023	8.0	0.23	8.39 +/- 0.16	7.93 +/- 0.12
PAXB x LITC 20	Spacing	0.008	13.1	-1.11	49.12 +/- 0.48	51.33 +/- 0.64
FTC x LITC 4	Number	0.033	3.5	0.22	7.56 +/- 0.15	7.12 +/- 0.13
FTC x LITC 4	Spacing	$5.4x10^{-5}$	11.4	-1.70	47.52 +/- 0.57	50.91 +/- 0.50
FTC x LITC 20	Number	0.012	9.4	0.25	7.62+/- 0.11	7.12 +/- 0.16
FTC x LITC 20	Spacing	$1.4x10^{-4}$	15.8	-1.60	47.55 +/- 0.38	50.74 +/- 0.65

Table 3.6: Summary of early (20 days post fertilization) QTL
Statistics for QTL for left side ventral row 1 number and left side row 1-7 spacing are shown. Effect size for the spacing phenotypes is in units of microns. Genotypic classes of F2 fish are abbreviated: MM = homozygous marine, FF = homozygous freshwater. PVE is the percentage of phenotypic variance explained.

Lake (Alaska, "BEPA", lab-reared offspring of wild-caught parents) was crossed to a wild-caught female LITC fish; and (3) a male benthic fish from Paxton Lake (British Columbia; "PAXB", lab-reared offspring of wild-caught parents) was crossed to a wild-caught female LITC fish. Fish from each F1 cross were intercrossed to create F2 families. Adult F2 fish (n=273, 384, and 418) were analyzed from 7, 5, and 11 F2 families in the PAXB, FTC, and BEPA crosses respectively. All lab-reared fish were raised at 18^{o}C in 110 liter (29 gallon) aquaria in a common brackish salinity (3.5 g/l Instant Ocean salt, 0.217 ml/l 10% sodium bicarbonate). Lab-reared fish were fed a common diet of live *Artemia nauplii* and frozen *Daphnia* as fry and juveniles, and frozen bloodworms and mysis shrimp as adults. "Adult" F2s were raised to a minimum standard length of at least 20 mm (mean +/- standard deviation of 31.1 +/- 7.3, 38.1 +/- 5.6, and 39.8 +/- 9.0 mm in the PAXB/FTC/BEPA crosses respectively). For the PAXB and FTC crosses, an early time point of F2s was taken at 19-20 days post fertilization ("dpf", n=96 per cross); these datasets are referred to as "20 dpf" or "early" F2 time points. These fish had a total length (TL) average and standard deviation of 8.9 +/- 0.8 and 8.4 +/- 0.6 mm in the PAXB and FTC crosses, respectively. To generate fish for the time course analyses, lab-reared fish from LITC, FTC, and PAXB incrosses were raised as described above to various stages of development from 8 to 50 mm TL.

Bone and cartilage staining
For bone staining, fish were fixed for 1-2 days in 10% neutral buffered formalin or 3-5 days in 4% paraformaldehyde in 1xPBS, washed with water overnight, stained overnight with 0.008% Alizarin Red S in 1% potassium hydroxide, destained in water overnight, then lightly cleared in a 0.25% potassium hydroxide, 50% glycerol solution. For bone and cartilage staining of time course fish and 20 dpf F2s, fish were stained with an acid-free two-color Alizarin/Alcian protocol as described (Walker and Kimmel 2007).

Gill raker phenotyping
Branchial skeletons were dissected out of fish and flat-mounted on a bridged coverslip. For all

adults, time course, and 20 dpf F2s, each branchial skeleton was phenotyped for row 1 or multiple rows of ventral and dorsal gill raker number, counting only Alizarin-positive rakers whose center lay between the Alizarin-positive boundaries of the ceratobranchial (for ventral rakers) or epibranchial (for dorsal rakers) gill arch bones (**Figure 3.1**). When indicated, composite phenotypes such as the average of ventral rows 1-3 or rows 1-7 were determined and averages of left and right side rakers were taken. Genetic mapping in adults was performed with the average of rows 1-3 ventral or dorsal raker number, averaging the left and right side counts. For the early F2 time point, ventral row 1 raker counts and ventral row 1-7 spacing measurements were analyzed; rows 8 and 9 were not scored because these posterior rakers are last to develop and were not consistently present at this time point. Raker primordia phenotypes were measured by mounting the most anterior branchial arch on a bridged cover slip post-*in situ*, then quantifying the number, spacing, and width of distinct *Edar*-positive puncta in row 1 buds (**Figure 3.7A**). All gill raker spacing measurements were obtained by acquiring digital images of rakers on a Leica DM2500 or Leica M165 microscope, determining the x and y coordinates of the center of the base of each raker in imageJ (Schneider et al. 2012), then calculating the average center-to-center spacing between each pair of adjacent rakers with a custom Python script (www.python.org). Raker width measurements were similarly calculated from digital images in imageJ, using the coordinates of the lateral and medial-most extent of Edar-positive cells within raker buds (for early Edar expressing foci) or Alizarin-positive edges of rakers (for adult rakers). Raker field size was calculated by measuring in imageJ the lateral-medial extent of *Edar*+ primordia with a segmented line that followed the path of raker primordia.

Genotyping

DNA was isolated by phenol-chloroform extraction or by a DNeasy 96 Blood and Tissue Kit (Qiagen). Polymerase chain reactions were 10 ul reactions with 10mM Tris (pH 8.5), 50mM KCl, 1.5mM $MgCl_2$, 0.1% Triton-X100, and 200 uM of each dNTP. Molecular markers spanned polymorphic microsatellites or indels on chromosomes 4 and 20. Markers were previously described (Colosimo et al. 2005; Colosimo et al. 2004; Largiader et al. 1999; Peichel et al. 2001) or were designed with Primer3 (Rozen and Skaletsky 2000) around $(AC)_n$ microsatellites found in the stickleback genome assembly (Jones et al. 2012b) with the Gramene SSR finder (Temnykh et al. 2001). All primer sequences and the method used to genotype each marker are listed in **Table 3.1**. Two primer polymerase chain reactions (PCR) with directly labeled fluorescent primers or non-fluorescent primers were performed using cycling conditions of 1 cycle of 94^o for 5 minutes; 35 cycles of 15s at 94^o, 15s at 56^o, and 15s at 72^o; and a final incubation of 5m at 72^o. Alternatively, a three primer PCR was performed as previously described by adding the M13F sequence (TGTAAAACGACGGCCAGT, all sequences listed are 5' to 3') to the 5' of the forward primer and including a fluorescently labeled M13 primer in the reaction (Schuelke 2000). PCR product sizes were determined by agarose gel electrophoresis (non-fluorescent PCR products) or by fragment analysis (fluorescent PCR products) with a 3730xl DNA Analyzer and GeneMapper (Applied Biosystems). Fish sex was determined by PCR amplification with primers CATATTGCTGCTTGTGTGGAAG and GATCCTCCTCGTTCCTACAG and gel electrophoresis. These two primers amplify fragment sizes of 186 bp and 229 bp from the X and Y chromosomes, respectively, from a region tightly linked to the sex-determining region (Peichel et al. 2004). Linkage maps were calculated using Joinmap 4 (Van Ooijen 2006) with regression mapping and default settings.

QTL mapping

For QTL mapping, raker number or spacing was tested for an association with standard length (adult) or total length (early F2 time point) and sex by linear regression in R (www.r-project.org) and corrected for size and/or sex, when appropriate. When association with length was significant ($p<0.05$), residuals were taken from a linear model with fish length and/or sex, then back-transformed to their original units. For adults, phenotypes were back-transformed to values expected for a 40 mm standard length fish. For early F2s, phenotypes were back-transformed to values expected for an 8 mm total length fish. For the early raker primordia (*Edar in situ* hybridization) data set, phenotypes were back-transformed to values expected for a 5.5 mm total length fish. Outliers greater than four standard deviations from the mean (<0.01% of all values) were removed.

Adult QTL mapping was performed in R/qtl (Broman and Sen 2009; Broman et al. 2003). LOD plots and percentage of variance explained were calculated with *fitqtl* and *refineqtl*, adjusting for the effect of another QTL controlling the phenotype when appropriate (e.g. adjusting for chromosome 20 genotype while mapping chromosome 4 QTL). For adult QTL mapping, significance thresholds ($P < 0.05$) were calculated by performing 1000 permutations of the genotypes on the two linkage groups being tested in each cross.

To generate plots of LOD score vs. physical (genome assembly) position, the genomic coordinates of each marker were used, with two exceptions. First, the region on chromosome 4 from 17.82 Megabases (Mb) to 28.36 Mb was inverted to correct for the true orientation and positions of scaffolds 24 and 28 as previously described (Roesti et al. 2013). Second, since Scaffold 46 containing marker Chr20_204 maps to the "left" end of chromosome 20 in all three crosses despite being on the "right" end of the genome assembly (higher coordinate in the genome assembly), this marker was assigned an adjusted physical position of 0 Mb. Cytogenetic data are consistent with Scaffold 46 mapping to the left end of the chromosome (lower coordinate in the genome assembly) (Urton et al. 2011).

Other statistical analyses

For comparisons between lab-reared and wild fish, two-tailed t-tests were performed on raw or back-transformed phenotypes, when appropriate (see above). Best-fit curves for the raker number and spacing time course plots were calculated with the *loess.smooth* function in R with a span of 0.4. Dominance was calculated using the equation d/a (Falconer and Mackay 1996), where a equals the additive effect of one additional freshwater allele (i.e. half the phenotypic difference between the homozygous freshwater and homozygous marine genotypic classes). d equals the dominance effect: the difference between the heterozygous phenotype and the midpoint between homozygous parental phenotypes. Pearson's correlation coefficients were calculated from size and sex-adjusted (as appropriate, see above) raker number and spacing measurements in R.

In situ hybridization

Lab-reared FTC and LITC embryos and fry were fixed with 4% paraformaldehyde in 1xPBS with 1% DMSO overnight at 4^{o}C. Whole mount *in situ* hybridization was performed essentially as described (Thisse and Thisse 2008), with 5-10 minutes of bleaching in a 3% hydrogen peroxide, 0.5% potassium hydroxide solution and 10 minutes of 20 ug/ml Proteinase K treatment

in PBSTween with 1% DMSO. Embryos were hybridized for >36 hours with an *Edar* antisense probe or sense probe as a negative control. *Edar* probes were generated by amplifying a fragment of the stickleback *Edar* gene using primers GCCGCTCGAGTGCCAGTGCAGAGTATTCCA and GCCGTCTAGACAGCTGCTCGTTCTCTGATG from LITC whole fry cDNA, directionally cloning this fragment into pBluescript II SK+ with XhoI and XbaI, linearizing this construct with XhoI, and transcribing the antisense probe with T3 polymerase or linearizing with XbaI and transcribing the sense probe with T7 polymerase. After wholemount *in situ*, first branchial arches were dissected out, transferred to 33%, 66%, and 100% glycerol, mounted flat on a bridged coverslip, and imaged with a Leica DM2500 compound microscope.

Animal statement

Wild anadromous marine fish were collected from the Little Campbell River 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 in Washington under a fish scientific collection permit 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 number R330 and 13834).

CONCLUSIONS

In summary, this work establishes that convergent evolution of gill raker reduction evolves via parallel embryonic shifts in the spacing of gill raker primordia, accomplished at least in part via the parallel use of QTL on chromosomes 4 and 20 in derived freshwater populations from Alaska, British Columbia, and Washington. During embryonic development, gill raker reduction is accomplished largely by an increased spacing between gill raker primordia, which the chromosome 4 and 20 QTL both control. Collectively our data support a model where this classic ecology-driven naturally selected trait evolves repeatedly via parallel developmental genetic mechanisms. Future forward (Loehlin and Werren 2012; McGregor et al. 2007) and reverse (Bedell et al. 2012; Dahlem et al. 2012) genetic approaches will further test how parallel the underlying molecular genetic changes are in this system of parallel adaptive evolution, and how these changes affect evolved differences in the developmental processes controlling epithelial appendage patterning.

ACKNOWLEDGEMENTS

Many of the fish analyzed in this chapter were bred, dissected, or genotyped by Phillip Cleves or Priscilla Erickson. This project benefitted greatly from our combined efforts to study each of our favorite phenotypes in shared sets of fish. Angela Lam spearheaded the QTL mapping efforts in the FTCxLITC cross. Craig Miller provided guidance in designing experiments, analyzing data, and writing the manuscript.

In addition to the coauthors listed above, I thank David Kingsley for his input and support during the beginning of this study, Dolph Schluter and Mike Bell for their generous gifts of Paxton benthic and Bear Paw lake fish, Alex Pollen for assistance collecting Fishtrap Creek and Little Campbell River fish, the Semiahmoo First Nation for permission to collect sticklebacks from the Little Campbell River, Patrick Lee and Anthony Lee for expert assistance in crossing and raising fish, and Jessica Grindheim, Emily Killingbeck, Monica Jimenez, and Nihar Patel for genotyping and phenotyping assistance. This work was supported in part by the NIH (R01- DE021475 to CTM and a Predoctoral Training Grant 5T32GM007127 to AMG, PAC, PAE), the NSF-Graduate Research Fellowship Program (AMG, PAC), a March of Dimes Basil O'Connor Starter Scholar Award (CTM), and a Pew Scholar in Biomedical Sciences award supported by the Pew Charitable Trusts (CTM).

REFERENCES

Aguirre, W.E., P.K. Doherty, and M.A. Bell, 2004 Genetics of lateral plate and gillraker phenotypes in a rapidly evolving population of threespine stickleback. Behaviour 141: 1465-1483.

Albert, A.Y.K., S. Sawaya, T.H. Vines, A.K. Knecht, C.T. Miller *et al.*, 2008 The genetics of adaptive shape shift in stickleback: pleiotropy and effect size. Evolution 62: 76-85.

Albert, V.A., S.E. Williams, and M.W. Chase, 1992 Carnivorous plants: phylogeny and structural evolution. Science 257: 1491-1495.

Albertson, R.C., J.T. Streelman, and T.D. Kocher, 2003 Directional selection has shaped the oral jaws of Lake Malawi cichlid fishes. Proc. Natl. Acad. Sci. U. S. A 100: 5252-5257.

Albertson, R.C., J.T. Streelman, T.D. Kocher, and P.C. Yelick, 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.

Andersson, L.S., M. Larhammar, F. Memic, H. Wootz, D. Schwochow *et al.*, 2012 Mutations in *DMRT3* affect locomotion in horses and spinal circuit function in mice. Nature 488: 642-646.

Arendt, J., and D. Reznick, 2008 Convergence and parallelism reconsidered: what have we learned about the genetics of adaptation? Trends Ecol. Evol. 23: 26-32.

Barrett, R.D.H., T.H. Vines, J.S. Bystriansky, and P.M. Schulte, 2009 Should I stay or should I go? The *Ectodysplasin* locus is associated with behavioural differences in threespine stickleback. Biol. Lett 5: 788-791.

Bedell, V.M., Y. Wang, J.M. Campbell, T.L. Poshusta, C.G. Starker *et al.*, 2012 *In vivo* genome editing using a high-efficiency TALEN system. Nature 491: 114-118.

Bell, M.A., and W.E. Aguirre, 2013 Contemporary evolution, allelic recycling, and adaptive radiation of the threespine stickleback. Evolutionary Ecology Research 15: 377-411.

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

Bernatchez, L., and J.J. Dodson, 1990 Allopatric origin of sympatric populations of lake whitefish (*Coregonus clupeaformis*) as revealed by mitochondrial DNA restriction analysis. Evolution 44: 1263-1271.

Berner, D., A.C. Grandchamp, and A.P. Hendry, 2009 Variable progress toward ecological speciation in parapatry: stickleback across eight lake-stream transitions. Evolution 63: 1740-1753.

Blackledge, T.A., and R.G. Gillespie, 2004 Convergent evolution of behavior in an adaptive radiation of Hawaiian web-building spiders. Proc. Natl. Acad. Sci. U. S. A 101: 16228-16233.

Bossuyt, F., and M.C. Milinkovitch, 2000 Convergent adaptive radiations in Madagascan and Asian ranid frogs reveal covariation between larval and adult traits. Proc. Natl. Acad. Sci. U. S. A 97: 6585-6590.

Broman, K.W., and S. Sen, 2009 *A Guide to QTL Mapping with R/qtl*. Dordrecht: Springer.

Broman, K.W., H. Wu, S. Sen, and G.A. Churchill, 2003 R/qtl: QTL mapping in experimental crosses. Bioinformatics 19: 889-890.

Calboli, F.C., 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, Jr., M.D. Shapiro *et al.*, 2010 Adaptive evolution of pelvic reduction in sticklebacks by recurrent deletion of a *Pitx1* enhancer. Science 327: 302-305.
Chen, L., A.L. DeVries, and C.H. Cheng, 1997 Evolution of antifreeze glycoprotein gene from a trypsinogen gene in Antarctic notothenioid fish. Proc. Natl. Acad. Sci. U. S. A 94: 3811-3816.
Christin, P.A., G. Besnard, E. Samaritani, M.R. Duvall, T.R. Hodkinson *et al.*, 2008 Oligocene CO2 decline promoted C4 photosynthesis in grasses. Current Biology 18: 37-43.
Chuong, C.-M., 1998 *Molecular Basis of Epithelial Appendage Morphogenesis*. Austin, TX: R.G. Landes.
Chuong, C.M., C.Y. Yeh, T.X. Jiang, and R. Widelitz, 2013 Module based complexity formation: periodic patterning in feathers and hairs. WIREs Dev. Biol. 2: 97-112.
Colosimo, P.F., K.E. Hosemann, S. Balabhadra, G. Villarreal, Jr., M. Dickson *et al.*, 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 *et al.*, 2004 The genetic architecture of parallel armor plate reduction in threespine sticklebacks. PLoS Biol. 2: 635-641.
Conte, G.L., M.E. Arnegard, C.L. Peichel, and D. Schluter, 2012 The probability of genetic parallelism and convergence in natural populations. Proc. Biol. Sci. 279: 5039-5047.
Cooley, A.M., J.L. Modliszewski, M.L. Rommel, and J.H. Willis, 2011 Gene duplication in *Mimulus* underlies parallel floral evolution via independent *trans*-regulatory changes. Current Biology 21: 700-704.
Cooley, A.M., L. Shefner, W.N. McLaughlin, E.E. Stewart, and P.J. Wittkopp, 2012 The ontogeny of color: developmental origins of divergent pigmentation in *Drosophila americana* and *D. novamexicana*. Evol. Dev. 14: 317-325.
Coyle, S.M., F.A. Huntingford, and C.L. Peichel, 2007 Parallel evolution of *Pitx1* underlies pelvic reduction in Scottish threespine stickleback (*Gasterosteus aculeatus*). J. Hered. 98: 581-586.
Cresko, W.A., A. Amores, C. Wilson, J. Murphy, M. Currey *et al.*, 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.
Dahlem, T.J., K. Hoshijima, M.J. Jurynec, D. Gunther, C.G. Starker *et al.*, 2012 Simple methods for generating and detecting locus-specific mutations Induced with TALENs in the zebrafish genome. PLoS Genet. 8: e1002861.
Day, T., J. Pritchard, and D. Schluter, 1994 A comparison of two sticklebacks. Evolution 48: 1723-1734.
Dobler, S., S. Dalla, V. Wagschal, and A.A. Agrawal, 2012 Community-wide convergent evolution in insect adaptation to toxic cardenolides by substitutions in the Na,K-ATPase. Proc. Natl. Acad. Sci. U. S. A 109: 13040-13045.
Donley, J.M., C.A. Sepulveda, P. Konstantinidis, S. Gemballa, and R.E. Shadwick, 2004 Convergent evolution in mechanical design of lamnid sharks and tunas. Nature 429: 61-65.

Drew, C.F., C.M. Lin, T.X. Jiang, G. Blunt, C. Mou *et al.*, 2007 The Edar subfamily in feather placode formation. Dev. Biol. 305: 232-245.
Dupuis, J., and D. Siegmund, 1999 Statistical methods for mapping quantitative trait loci from a dense set of markers. Genetics 151: 373-386.
Falconer, D.S., and T.F.C. Mackay, 1996 *Introduction to Quantitative Genetics*. Essex: Pearson.
Fraser, G.J., R.F. Bloomquist, and J.T. Streelman, 2008 A periodic pattern generator for dental diversity. BMC Biol. 6.
Gagnaire, P.A., S.A. Pavey, E. Normandeau, and L. Bernatchez, 2013 The genetic architecture of reproductive isolation during speciation-with-gene-flow in lake whitefish species pairs assessed by RAD sequencing. Evolution 67: 2483-2497.
Grant, P.R., and B.R. Grant, 2006 Evolution of character displacement in Darwin's finches. Science 313: 224-226.
Greenwood, A.K., A.R. Wark, K. Yoshida, and C.L. Peichel, 2013 Genetic and neural modularity underlie the evolution of schooling behavior in threespine sticklebacks. Current Biology 23: 1884-1888.
Gross, H.P., and J.M. Anderson, 1984 Geographic-variation in the gillrakers and diet of European threespine sticklebacks, *Gasterosteus aculeatus*. Copeia 1: 87-97.
Gross, J.B., R. Borowsky, and C.J. Tabin, 2009 A novel role for *Mc1r* in the parallel evolution of depigmentation in independent populations of the cavefish *Astyanax mexicanus*. PLoS Genet. 5: e1000326.
Hagen, D.W., 1967 Isolating mechanisms in threespine sticklebacks (*Gasterosteus*). J. Fish. Res. Bd. Canada 24: 1637-1692.
Hagen, D.W., 1973 Inheritance of numbers of lateral plates and gill rakers in *Gasterosteus aculeatus*. Heredity 30: 303-312.
Hagen, D.W., and L.G. Gilbertson, 1972 Geographic variation and environmental selection in *Gasterosteus aculeatus* L in the Pacific Northwest, America. Evolution 26: 32-51.
Harris, M.P., N. Rohner, H. Schwarz, S. Perathoner, P. Konstantinidis *et al.*, 2008 Zebrafish *eda* and *edar* mutants reveal conserved and ancestral roles of ectodysplasin signaling in vertebrates. PLoS Genet. 4: e1000206.
Hermida, M., C. Fernandez, R. Amaro, and E. San Miguel, 2002 Heritability and "evolvability" of meristic characters in a natural population of *Gasterosteus aculeatus*. Can. J. Zool. 80: 532-541.
Herring, C.D., A. Raghunathan, C. Honisch, T. Patel, M.K. Applebee *et al.*, 2006 Comparative genome sequencing of *Escherichia coli* allows observation of bacterial evolution on a laboratory timescale. Nat. Genet. 38: 1406-1412.
Hodgkin, J., and T. Doniach, 1997 Natural variation and copulatory plug formation in *Caenorhabditis elegans*. Genetics 146: 149-164.
Hohenlohe, P.A., S. Bassham, P.D. Etter, N. Stiffler, E.A. Johnson *et al.*, 2010 Population genomics of parallel adaptation in threespine stickleback using sequenced RAD tags. PLoS Genet. 6: e1000862.
Houghton, L., C. Lindon, and B.A. Morgan, 2005 The ectodysplasin pathway in feather tract development. Development 132: 863-872.
Hulsey, C.D., R.J. Roberts, A.S.P. Lin, R. Guldberg, and J.T. Streelman, 2008 Convergence in a mechanically complex phenotype: Detecting structural adaptations for crushing in cichlid fish. Evolution 62: 1587-1599.

Jeffery, W.R., 2009 Regressive evolution in *Astyanax* cavefish. Annu. Rev. Genet. 43: 25-47.
Johanson, U., J. West, C. Lister, S. Michaels, R. Amasino *et al.*, 2000 Molecular analysis of *FRIGIDA*, a major determinant of natural variation in *Arabidopsis* flowering time. Science 290: 344-347.
Jones, F.C., Y.F. Chan, J. Schmutz, J. Grimwood, S.D. Brady *et al.*, 2012a A genome-wide SNP genotyping array reveals patterns of global and repeated species-pair divergence in sticklebacks. Current Biology 22: 83-90.
Jones, F.C., M.G. Grabherr, Y.F. Chan, P. Russell, E. Mauceli *et al.*, 2012b The genomic basis of adaptive evolution in threespine sticklebacks. Nature 484: 55-61.
Jung, H.S., P.H. Francis-West, R.B. Widelitz, T.X. Jiang, S. Ting-Berreth *et al.*, 1998 Local inhibitory action of BMPs and their relationships with activators in feather formation: Implications for periodic patterning. Dev. Biol. 196: 11-23.
Kamberov, Y.G., S.J. Wang, J.Z. Tan, P. Gerbault, A. Wark *et al.*, 2013 Modeling recent human evolution in mice by expression of a selected EDAR variant. Cell 152: 691-702.
Kislalioglu, M., and R.N. Gibson, 1977 Feeding relationship of shallow water fishes in a Scottish sea loch. J. Fish Biol. 11: 257-266.
Kolbe, J.J., L.J. Revell, B. Szekely, E.D. Brodie, 3rd, and J.B. Losos, 2011 Convergent evolution of phenotypic integration and its alignment with morphological diversification in Caribbean *Anolis* ecomorphs. Evolution 65: 3608-3624.
Kondo, S., and T. Miura, 2010 Reaction-diffusion model as a framework for understanding biological pattern formation. Science 329: 1616-1620.
Kopp, A., 2009 Metamodels and phylogenetic replication: a systematic approach to the evolution of developmental pathways. Evolution 63: 2771-2789.
Kowalko, J.E., N. Rohner, T.A. Linden, S.B. Rompani, W.C. Warren *et al.*, 2013 Convergence in feeding posture occurs through different genetic loci in independently evolved cave populations of *Astyanax mexicanus*. Proc. Natl. Acad. Sci. U. S. A 110: 16933-16938.
Largiader, C.R., V. Fries, B. Kobler, and T.C.M. Bakker, 1999 Isolation and characterization of microsatellite loci from the three-spined stickleback (*Gasterosteus aculeatus* L.). Mol. Ecol. 8: 342-344.
Laurikkala, J., J. Pispa, H.S. Jung, P. Nieminen, M. Mikkola *et al.*, 2002 Regulation of hair follicle development by the TNF signal ectodysplasin and its receptor Edar. Development 129: 2541-2553.
Lindsey, C.C., 1981 Stocks are chameleons - plasticity in gill rakers of coregonid fishes. Can. J. Fish. Aquat. Sci. 38: 1497-1506.
Loehlin, D.W., and J.H. Werren, 2012 Evolution of shape by multiple regulatory changes to a growth gene. Science 335: 943-947.
Losos, J.B., 2011 Convergence, adaptation, and constraint. Evolution 65: 1827-1840.
Lundsgaard-Hansen, B., B. Matthews, P. Vonlanthen, A. Taverna, and O. Seehausen, 2013 Adaptive plasticity and genetic divergence in feeding efficiency during parallel adaptive radiation of whitefish (*Coregonus* spp.). J. Evol. Biol. 26: 483-498.
Magnuson, J.J., and J.G. Heitz, 1971 Gill raker apparatus and food selectivity among mackerels, tunas, and dolphins. Fishery Bulletin of the National Oceanic and Atmospheric Administration 69: 361-370.

Mallarino, R., P.R. Grant, B.R. Grant, A. Herrel, W.P. Kuo *et al.*, 2011 Two developmental modules establish 3D beak-shape variation in Darwin's finches. Proc. Natl. Acad. Sci. U. S. A 108: 4057-4062.
Manceau, M., V.S. Domingues, R. Mallarino, and H.E. Hoekstra, 2011 The developmental role of Agouti in color pattern evolution. Science 331: 1062-1065.
Martin, A., and V. Orgogozo, 2013 The loci of repeated evolution: a catalog of genetic hotspots of phenotypic variation. Evolution 67: 1235-1250.
McCracken, K.G., C.P. Barger, M. Bulgarella, K.P. Johnson, S.A. Sonsthagen *et al.*, 2009 Parallel evolution in the major haemoglobin genes of eight species of Andean waterfowl. Mol. Ecol. 18: 3992-4005.
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 *et al.*, 2007 Morphological evolution through multiple *cis*-regulatory mutations at a single gene. Nature 448: 587-590.
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., 1993 Ecology and evolution of sympatric sticklebacks (*Gasterosteus*) - origin of the species pairs. Can. J. Zool. 71: 515-523.
Miller, C.T., S. Beleza, A.A. Pollen, D. Schluter, R.A. Kittles *et al.*, 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 *et al.*, 2014 Additive and clustered quantitative trait loci control anatomically regional skeletal evolution in sticklebacks. Genetics: in press.
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.
Moodie, G.E.E., and T.E. Reimchen, 1976 Phenetic variation and habitat differences in *Gasterosteus* populations of the Queen Charlotte Islands. Syst. Zool. 25: 49-61.
Mou, C., F. Pitel, D. Gourichon, F. Vignoles, A. Tzika *et al.*, 2011 Cryptic patterning of avian skin confers a developmental facility for loss of neck feathering. PLoS Biol. 9: e1001028.
Muschick, M., A. Indermaur, and W. Salzburger, 2012 Convergent evolution within an adaptive radiation of cichlid fishes. Current Biology 22: 2362-2368.
Nadeau, N.J., and C.D. Jiggins, 2010 A golden age for evolutionary genetics? Genomic studies of adaptation in natural populations. Trends Genet. 26: 484-492.
Parker, H.G., B.M. VonHoldt, P. Quignon, E.H. Margulies, S. Shao *et al.*, 2009 An expressed *fgf4* retrogene is associated with breed-defining chondrodysplasia in domestic dogs. Science 325: 995-998.
Peichel, C.L., K.S. Nereng, K.A. Ohgi, B.L.E. Cole, P.F. Colosimo *et al.*, 2001 The genetic architecture of divergence between threespine stickleback species. Nature 414: 901-905.
Peichel, C.L., J.A. Ross, C.K. Matson, M. Dickson, J. Grimwood *et al.*, 2004 The master sex-determination locus in threespine sticklebacks is on a nascent Y chromosome. Current Biology 14: 1416-1424.

Perry, G.H., N.J. Dominy, K.G. Claw, A.S. Lee, H. Fiegler *et al.*, 2007 Diet and the evolution of human amylase gene copy number variation. Nat. Genet. 39: 1256-1260.
Pispa, J., M.L. Mikkola, T. Mustonen, and I. Thesleff, 2003 *Ectodysplasin*, *Edar* and *TNFRSF19* are expressed in complementary and overlapping patterns during mouse embryogenesis. Gene Expr. Patterns 3: 675-679.
Protas, M.E., C. Hersey, D. Kochanek, Y. Zhou, H. Wilkens *et al.*, 2006 Genetic analysis of cavefish reveals molecular convergence in the evolution of albinism. Nat. Genet. 38: 107-111.
Raeymaekers, J.A.M., J.K.J. Van Houdt, M.H.D. Larmuseau, S. Geldof, and F.A.M. Volckaert, 2007 Divergent selection as revealed by P-ST and QTL-based F-ST in three-spined stickleback (*Gasterosteus aculeatus*) populations along a coastal-inland gradient. Mol. Ecol. 16: 891-905.
Reed, R.D., R. Papa, A. Martin, H.M. Hines, B.A. Counterman *et al.*, 2011 *Optix* drives the repeated convergent evolution of butterfly wing pattern mimicry. Science 333: 1137-1141.
Reimchen, T.E., E.M. Stinson, and J.S. Nelson, 1985 Multivariate differentiation of parapatric and allopatric populations of threespine stickleback in the Sangan River watershed, Queen Charlotte Islands. Can. J. Zool. 63: 2944-2951.
Robinson, B.W., 2000 Trade offs in habitat-specific foraging efficiency and the nascent adaptive divergence of sticklebacks in lakes. Behaviour 137: 865-888.
Rockman, M.V., 2012 The QTN program and the alleles that matter for evolution: All that's gold does not glitter. Evolution 66: 1-17.
Roesch, C., B. Lundsgaard-Hansen, P. Vonlanthen, A. Taverna, and O. Seehausen, 2013 Experimental evidence for trait utility of gill raker number in adaptive radiation of a north temperate fish. J. Evol. Biol. 26: 1578-1587.
Roesti, M., D. Moser, and D. Berner, 2013 Recombination in the threespine stickleback genome--patterns and consequences. Mol. Ecol. 22: 3014-3027.
Rogers, S.M., and L. Bernatchez, 2007 The genetic architecture of ecological speciation and the association with signatures of selection in natural lake whitefish (*Coregonus* sp. Salmonidae) species pairs. Mol. Biol. Evol. 24: 1423-1438.
Rozen, S., and H. Skaletsky, 2000 Primer3 on the WWW for general users and for biologist programmers. Methods Mol. Biol. 132: 365-386.
Ruber, L., E. Verheyen, and A. Meyer, 1999 Replicated evolution of trophic specializations in an endemic cichlid fish lineage from Lake Tanganyika. Proc. Natl. Acad. Sci. U. S. A 96: 10230-10235.
Sadier, A., L. Viriot, S. Pantalacci, and V. Laudet, 2014 The ectodysplasin pathway: from diseases to adaptations. Trends Genet. 30: 24-31.
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. Biol. Sci. 279: 739-748.
Schluter, D., 2000 *The Ecology of Adaptive Radiation*. Oxford: Oxford University Press.
Schluter, D., and G.L. Conte, 2009 Genetics and ecological speciation. Proc. Natl. Acad. Sci. U. S. A 106: 9955-9962.
Schluter, D., and J.D. McPhail, 1992 Ecological character displacement and speciation in sticklebacks. Am. Nat. 140: 85-108.

Schluter, D., and J.D. McPhail, 1993 Character displacement and replicate adaptive radiation. Trends Ecol. Evol. 8: 197-200.
Schluter, D., T.D. Price, and P.R. Grant, 1985 Ecological character displacement in Darwin's finches. Science 227: 1056-1059.
Schneider, C.A., W.S. Rasband, and K.W. Eliceiri, 2012 NIH Image to ImageJ: 25 years of image analysis. Nat. Methods 9: 671-675.
Schuelke, M., 2000 An economic method for the fluorescent labeling of PCR fragments. Nat. Biotechnol. 18: 233-234.
Shapiro, M.D., M.E. Marks, C.L. Peichel, B.K. Blackman, K.S. Nereng *et al.*, 2004 Genetic and developmental basis of evolutionary pelvic reduction in threespine sticklebacks. Nature 428: 717-723.
Sick, S., S. Reinker, J. Timmer, and T. Schlake, 2006 WNT and DKK determine hair follicle spacing through a reaction-diffusion mechanism. Science 314: 1447-1450.
Stern, D.L., 2000 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.
Streisfeld, M.A., and M.D. Rausher, 2009 Genetic changes contributing to the parallel evolution of red floral pigmentation among *Ipomoea* species. New Phytol. 183: 751-763.
Sucena, E., I. Delon, I. Jones, F. Payre, and D.L. Stern, 2003 Regulatory evolution of *shavenbaby/ovo* underlies multiple cases of morphological parallelism. Nature 424: 935-938.
Sutter, N.B., C.D. Bustamante, K. Chase, M.M. Gray, K.Y. Zhao *et al.*, 2007 A single *IGF1* allele is a major determinant of small size in dogs. Science 316: 112-115.
Tanaka, K., O. Barmina, and A. Kopp, 2009 Distinct developmental mechanisms underlie the evolutionary diversification of *Drosophila* sex combs. Proc. Natl. Acad. Sci. U. S. A 106: 4764-4769.
Temnykh, S., G. DeClerck, A. Lukashova, L. Lipovich, S. Cartinhour *et al.*, 2001 Computational and experimental analysis of microsatellites in rice (*Oryza sativa* L.): frequency, length variation, transposon associations, and genetic marker potential. Genome Res. 11: 1441-1452.
Tenaillon, O., A. Rodriguez-Verdugo, R.L. Gaut, P. McDonald, A.F. Bennett *et al.*, 2012 The molecular diversity of adaptive convergence. Science 335: 457-461.
Thisse, C., and B. Thisse, 2008 High-resolution *in situ* hybridization to whole-mount zebrafish embryos. Nat. Protoc. 3: 59-69.
Tishkoff, S.A., F.A. Reed, A. Ranciaro, B.F. Voight, C.C. Babbitt *et al.*, 2007 Convergent adaptation of human lactase persistence in Africa and Europe. Nat. Genet. 39: 31-40.
Tucker, A.S., D.J. Headon, J.M. Courtney, P. Overbeek, and P.T. Sharpe, 2004 The activation level of the TNF family receptor, Edar, determines cusp number and tooth number during tooth development. Dev. Biol. 268: 185-194.
Urton, J.R., S.R. McCann, and C.L. Peichel, 2011 Karyotype differentiation between two stickleback species (Gasterosteidae). Cytogenet. Genome Res. 135: 150-159.
van Ditmarsch, D., K.E. Boyle, H. Sakhtah, J.E. Oyler, C.D. Nadell *et al.*, 2013 Convergent evolution of hyperswarming leads to impaired biofilm formation in pathogenic bacteria. Cell Rep. 4: 697-708.

Van Ooijen, J.W., 2006 *JoinMap 4, Software for the Calculation of Genetic Linkage Maps in Experimental Populations.* Wageningen, Netherlands: Kyazma B. V.
Walker, M.B., and C.B. Kimmel, 2007 A two-color acid-free cartilage and bone stain for zebrafish larvae. Biotech. Histochem. 82: 23-28.
Wark, A.R., M.G. Mills, L.H. Dang, Y.F. Chan, F.C. Jones *et al.*, 2012 Genetic architecture of variation in the lateral line sensory system of threespine sticklebacks. G3 2: 1047-1056.
Yamamoto, Y., M.S. Byerly, W.R. Jackman, and W.R. Jeffery, 2009 Pleiotropic functions of embryonic *sonic hedgehog* expression link jaw and taste bud amplification with eye loss during cavefish evolution. Dev. Biol. 330: 200-211.
Yoon, H.S., and D.A. Baum, 2004 Transgenic study of parallelism in plant morphological evolution. Proc. Natl. Acad. Sci. U. S. A 101: 6524-6529.

Chapter 4:
Genome assembly improvement and mapping convergently evolved skeletal traits in sticklebacks with Genotyping-by-Sequencing

Andrew M. Glazer, Emily E. Killingbeck, Therese Mitros, Daniel S. Rokhsar, and Craig T. Miller

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

ABSTRACT

Marine populations of the threespine stickleback (*Gasterosteus aculeatus*) have repeatedly colonized and rapidly adapted to freshwater habitats, providing a powerful system to map the genetic architecture of evolved traits. Here we developed and applied a binned Genotyping-by-Sequencing (GBS) method to build dense genome-wide linkage maps of sticklebacks, using two large marine by freshwater F2 crosses of over 350 fish each. The resulting linkage maps significantly improve the genome assembly by anchoring 78 new scaffolds to chromosomes, reorienting 40 scaffolds, and rearranging scaffolds in 4 locations. In the revised genome assembly, 94.6% of the assembly was anchored to a chromosome. To assess linkage map quality, we mapped quantitative trait loci (QTL) controlling lateral plate number, which mapped as expected to a 200 kilobase genomic region containing *Ectodysplasin*, as well as a chromosome 7 QTL overlapping a previously identified modifier QTL. Finally, we mapped eight QTL controlling convergently evolved reductions in gill raker length in the two crosses, which revealed that this classic adaptive trait has a surprisingly modular and non-parallel genetic basis.

INTRODUCTION

Understanding the genetic basis of adaptation remains a major unsolved goal in biology. For example, when the same phenotype evolves in independent lineages (convergent evolution), is the genetic basis predictable (Stern and Orgogozo 2009)? Do evolved loci typically affect phenotypes in a global or in a modular, anatomically specific manner (Wagner et al. 2007)? In systems where differently adapted natural populations are interfertile, QTL mapping provides an entry point to study the genetic architecture of evolved traits.

The threespine stickleback (*Gasterosteus aculeatus*) has undergone a widespread adaptive radiation in which marine fish independently colonized and adapted to countless freshwater habitats (Bell and Foster 1994). Marine and freshwater populations typically differ in many skeletal phenotypes, including a reduction of the number of lateral plates, used for defense against predation (Reimchen 1992; Hagen 1973), and a reduction in the length of gill rakers, a set of bones used for prey retention during feeding (Schluter 2000). Lateral plate reduction in many freshwater populations is largely controlled by a large-effect QTL on chromosome 4 (Berner et al. 2014; Colosimo et al. 2004; Cresko et al. 2004; Liu et al. 2014; Wark et al. 2012), which has been shown to be a regulatory allele of *Ectodysplasin* (*Eda*) (Colosimo et al. 2005; O'Brown et al. 2015). Several smaller-effect modifier QTL also contribute to plate number reduction (Colosimo et al. 2004; Wark et al. 2012). Reductions in gill raker length and number are important trophic adaptations in freshwater sticklebacks and other postglacial fish, and have convergently evolved multiple times (Schluter 2000). Typically, fish that eat small plankton evolve more, longer gill rakers and fish that eat larger prey evolve fewer, shorter gill rakers (Arnegard et al. 2014; Berner et al. 2010a; Berner et al. 2010b; Schluter and McPhail 1992; Theis et al. 2014). Our previous genetic studies found gill raker number to be a highly polygenic trait, controlled by over 15 QTL (Glazer et al. 2014; Miller et al. 2014). While stickleback gill raker length differences can arise due to phenotypic plasticity, there is a large heritable component (Day et al. 1994; Hatfield 1997). The genetic basis of evolved freshwater reductions in stickleback gill raker length is poorly understood, but two QTL were identified in a cross between European lake and stream populations (Berner et al. 2014).

The stickleback genome has been sequenced and scaffolds were anchored in the stickleback reference assembly (Jones et al. 2012) using a linkage map made from an F2 cross of 92 fish. The assembly consists of 113 anchored scaffolds on 21 chromosomes, as well as 1,822 unanchored scaffolds (13.2% of the assembly). Subsequent work inverted the orientations of 13 anchored scaffolds and anchored 18 additional scaffolds (Roesti et al. 2013). Three large chromosomal inversions are typically present between marine and freshwater sticklebacks (Jones 2012), but the extent of other differences between stickleback populations in genomic structure and genome-wide recombination patterns is largely unknown.

Here we used Genotyping-by-Sequencing (GBS) to sample about 100,000 SNPs to low coverage (about 1.5X per sample) for over 350 sticklebacks in each of two marine x freshwater F2 crosses. We binned these low-coverage SNPs into over 1,000 high-coverage (about 150X) markers. Using these markers, we constructed high-density

genome-wide linkage maps, which we used to anchor, reorient, and rearrange scaffolds, and examine genome-wide recombination patterns. We also used these maps to map the genetic basis of two ecologically important phenotypes. First, as a positive control, we mapped QTL controlling lateral plate number. Second, we mapped gill raker length QTL to test two hypotheses: that convergent evolution in two independently derived freshwater populations involves similar genetic architectures, and that gill raker length is genetically controlled by modular QTL affecting the lengths of subsets of gill rakers.

METHODS

Stickleback crosses

Two marine x freshwater F1 crosses were previously described (Glazer et al. 2014). A wild-caught male marine fish from the Little Campbell River (British Columbia, Canada,"LITC") was crossed to a wild-caught female freshwater fish from Fishtrap Creek (Washington state; "FTC") to generate the FTC cross. A male freshwater fish from Bear Paw Lake (Alaska, "BEPA", lab-reared offspring of wild-caught parents) was crossed to a wild-caught marine female LITC fish to produce the BEPA cross. F1s were intercrossed to generate 360 and 363 F2 fish from the FTC and BEPA crosses, respectively. Fish with low genotype coverage (n = 2 from each cross) were removed from the analysis.

Lab-reared fish were raised at 18°C in 110 liter (29 gallon) aquaria in a common brackish salinity (3 ppt, 10% ocean water) and fed a common diet of live *Artemia* nauplii and frozen *Daphnia* as fry and juveniles, and bloodworms and *Mysis* shrimp as adults. Adult F2s were raised to a mean size (standard length) of 38.0 mm and 42.1 mm in the FTC and BEPA crosses, respectively. Fish from F1 crosses were intercrossed to generate F2 families ranging in size from 26-108 fish. Adult F2 fish (n= 360 / 363) were sequenced with GBS from 5 and 9 F2 families in the FTC and BEPA crosses, respectively.

Animal statement

Wild anadromous marine fish were collected from the Little Campbell River 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 in Washington under a fish scientific collection permit 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 number R330 and 13834).

Preparation of GBS Libraries

DNA was isolated by phenol-chloroform extraction or with a DNeasy 96 Blood and Tissue Kit (Qiagen). Genomic DNA concentration was assessed with a NanoDrop 1000 spectrophotometer (Thermo Scientific) and by Quant-iT PicoGreen Assay (Invitrogen). Unless otherwise noted, Genotyping-by-Sequencing (GBS) Illumina sequencing libraries were constructed as previously described (Elshire et al. 2011). 50 ng/sample of genomic DNA was used. Individuals were sequenced in seven libraries (Table 4.2). For libraries

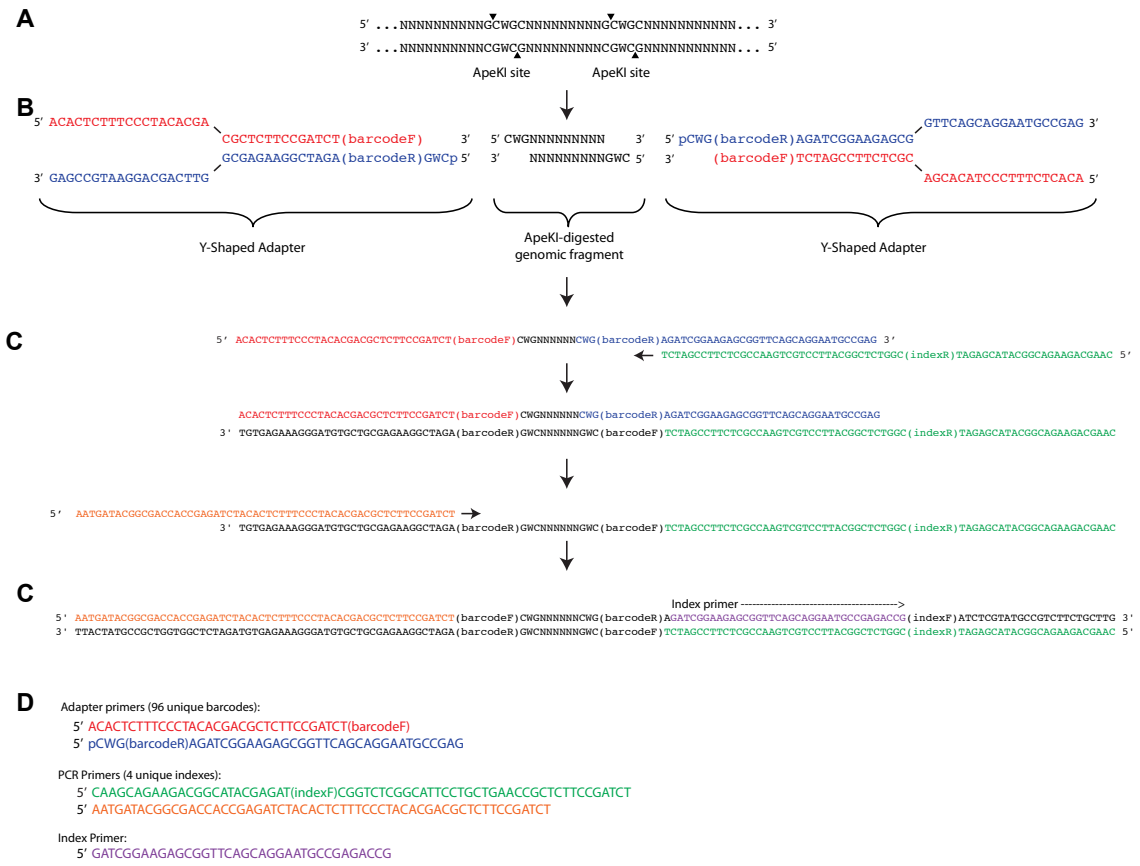


Figure 4.1: Diagram of library making method.
(A) Genomic DNA was digested with ApeKI, creating CWG overhangs. (B) Pairs of adapter primers were annealed together to create a "Y-shaped" primer, with partial base-pairing at one end. These adapters were ligated to both sides of the digested genomic DNA. 96 unique adapters were used, containing different internal barcode sequences ("barcodeF" and "barcodeR" together constitute the double-stranded barcode). (C) PCR primers with one of four index barcodes ("index") are used to add sequence that will anneal to the Illumina flowcell. A custom index primer is used during sequencing to determine the index of each fragment, and the first bases of the R1 and R2 reads will contain the barcode. (D) Primers used in this study. See **Table 4.1** for primer sequences.

1-3, 48 barcode + common adapters were used (Elshire et al. 2011). Libraries 4-6 used 96 ApeKI Y-shaped adapters with internal barcodes, and library 7 used these 96 Y-shaped adapters and 4 different PCR primers with different index barcodes (384 total samples, adapted from Peterson et al. 2012, **Table 4.1** and **Figure 4.1**). For library 7, all volumes in the digestion and ligation reactions were successfully halved relative to the Elshire et al. protocol to reduce reagent costs. ApeKI digestion, ligation, and sample cleanup were performed as described (Elshire et al. 2011). PCR amplification of sequencing libraries was performed in 50 ul reactions with 25 ul Taq 2X Master Mix (NEB), 50-450 ng of primer, and 2 ul of each library at 98°C for 30 seconds; 10-22 cycles at 98°C for 10 seconds, 65°C for 30 seconds, 72°C for 30 seconds; 5 minutes at 72°C; and held at 4°C.

Name	Primer	Type	Barcode
PCR1	CAAGCAGAAGACGGCATACGAGATCGGTCTCGGCATTCCTGCT GAACCGCTCTTCCGATCT	PCR (no indexing)	NA
PCR1_index1	CAAGCAGAAGACGGCATACGAGATcgtgatCGGTCTCGGCATTCCT GCTGAACCGCTCTTCCGATCT	PCR (indexing)	NA
PCR1_index2	CAAGCAGAAGACGGCATACGAGATacatcgCGGTCTCGGCATTCC TGCTGAACCGCTCTTCCGATCT	PCR (indexing)	NA
PCR1_index3	CAAGCAGAAGACGGCATACGAGATgcctaaCGGTCTCGGCATTCC TGCTGAACCGCTCTTCCGATCT	PCR (indexing)	NA
PCR1_index4	CAAGCAGAAGACGGCATACGAGATtggtcaCGGTCTCGGCATTCCT GCTGAACCGCTCTTCCGATCT	PCR (indexing)	NA
PCR2	AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGA CGCTCTTCCGATCT	PCR (indexing/no indexing)	NA
Seq_index	GATCGGAAGAGCGGTTCAGCAGGAATGCCGAGACCG	Index	NA
ApeK1Y_A1F	ACACTCTTTCCCTACACGACGCTCTTCCGATCTctcc	Adapter	ctcc
ApeK1Y_A2F	ACACTCTTTCCCTACACGACGCTCTTCCGATCTtgca	Adapter	tgca
ApeK1Y_A3F	ACACTCTTTCCCTACACGACGCTCTTCCGATCTacta	Adapter	acta
ApeK1Y_A4F	ACACTCTTTCCCTACACGACGCTCTTCCGATCTcaga	Adapter	caga
ApeK1Y_A5F	ACACTCTTTCCCTACACGACGCTCTTCCGATCTaact	Adapter	aact
ApeK1Y_A6F	ACACTCTTTCCCTACACGACGCTCTTCCGATCTgcgt	Adapter	gcgt
ApeK1Y_A7F	ACACTCTTTCCCTACACGACGCTCTTCCGATCTcgat	Adapter	cgat
ApeK1Y_A8F	ACACTCTTTCCCTACACGACGCTCTTCCGATCTgtaa	Adapter	gtaa
ApeK1Y_A9F	ACACTCTTTCCCTACACGACGCTCTTCCGATCTtgcga	Adapter	tgcga
ApeK1Y_A10F	ACACTCTTTCCCTACACGACGCTCTTCCGATCTcgctt	Adapter	cgctt
ApeK1Y_A11F	ACACTCTTTCCCTACACGACGCTCTTCCGATCTtcacc	Adapter	tcacc
ApeK1Y_A12F	ACACTCTTTCCCTACACGACGCTCTTCCGATCTctagc	Adapter	ctagc
ApeK1Y_B1F	ACACTCTTTCCCTACACGACGCTCTTCCGATCTacaaa	Adapter	acaaa
ApeK1Y_B2F	ACACTCTTTCCCTACACGACGCTCTTCCGATCTttctc	Adapter	ttctc
ApeK1Y_B3F	ACACTCTTTCCCTACACGACGCTCTTCCGATCTagccc	Adapter	agccc
ApeK1Y_B4F	ACACTCTTTCCCTACACGACGCTCTTCCGATCTgtatt	Adapter	gtatt
ApeK1Y_B5F	ACACTCTTTCCCTACACGACGCTCTTCCGATCTctgta	Adapter	ctgta
ApeK1Y_B6F	ACACTCTTTCCCTACACGACGCTCTTCCGATCTaccgt	Adapter	accgt
ApeK1Y_B7F	ACACTCTTTCCCTACACGACGCTCTTCCGATCTgctta	Adapter	gctta
ApeK1Y_B8F	ACACTCTTTCCCTACACGACGCTCTTCCGATCTggtgt	Adapter	ggtgt
ApeK1Y_B9F	ACACTCTTTCCCTACACGACGCTCTTCCGATCTggttgt	Adapter	ggttgt
ApeK1Y_B10F	ACACTCTTTCCCTACACGACGCTCTTCCGATCTccagct	Adapter	ccagct
ApeK1Y_B11F	ACACTCTTTCCCTACACGACGCTCTTCCGATCTttcaga	Adapter	ttcaga
ApeK1Y_B12F	ACACTCTTTCCCTACACGACGCTCTTCCGATCTtaggaa	Adapter	taggaa
ApeK1Y_C1F	ACACTCTTTCCCTACACGACGCTCTTCCGATCTgctcta	Adapter	gctcta
ApeK1Y_C2F	ACACTCTTTCCCTACACGACGCTCTTCCGATCTccacaa	Adapter	ccacaa
ApeK1Y_C3F	ACACTCTTTCCCTACACGACGCTCTTCCGATCTcttcca	Adapter	cttcca
ApeK1Y_C4F	ACACTCTTTCCCTACACGACGCTCTTCCGATCTgagata	Adapter	gagata
ApeK1Y_C5F	ACACTCTTTCCCTACACGACGCTCTTCCGATCTatgcct	Adapter	atgcct
ApeK1Y_C6F	ACACTCTTTCCCTACACGACGCTCTTCCGATCTagtgga	Adapter	agtgga
ApeK1Y_C7F	ACACTCTTTCCCTACACGACGCTCTTCCGATCTtattttt	Adapter	tattttt
ApeK1Y_C8F	ACACTCTTTCCCTACACGACGCTCTTCCGATCTcttgctt	Adapter	cttgctt
ApeK1Y_C9F	ACACTCTTTCCCTACACGACGCTCTTCCGATCTatgaaac	Adapter	atgaaac
ApeK1Y_C10F	ACACTCTTTCCCTACACGACGCTCTTCCGATCTaaaagtt	Adapter	aaaagtt
ApeK1Y_C11F	ACACTCTTTCCCTACACGACGCTCTTCCGATCTgaattca	Adapter	gaattca
ApeK1Y_C12F	ACACTCTTTCCCTACACGACGCTCTTCCGATCTgaacttc	Adapter	gaacttc
ApeK1Y_D1F	ACACTCTTTCCCTACACGACGCTCTTCCGATCTggaccta	Adapter	ggaccta
ApeK1Y_D2F	ACACTCTTTCCCTACACGACGCTCTTCCGATCTgtcgatt	Adapter	gtcgatt
ApeK1Y_D3F	ACACTCTTTCCCTACACGACGCTCTTCCGATCTaacgcct	Adapter	aacgcct
ApeK1Y_D4F	ACACTCTTTCCCTACACGACGCTCTTCCGATCTaatatgc	Adapter	aatatgc
ApeK1Y_D5F	ACACTCTTTCCCTACACGACGCTCTTCCGATCTacgactac	Adapter	acgactac
ApeK1Y_D6F	ACACTCTTTCCCTACACGACGCTCTTCCGATCTtagcatgc	Adapter	tagcatgc
ApeK1Y_D7F	ACACTCTTTCCCTACACGACGCTCTTCCGATCTtaggccat	Adapter	taggccat
ApeK1Y_D8F	ACACTCTTTCCCTACACGACGCTCTTCCGATCTtgcaagga	Adapter	tgcaagga
ApeK1Y_D9F	ACACTCTTTCCCTACACGACGCTCTTCCGATCTtggtacgt	Adapter	tggtacgt

ApeK1Y_D10F	ACACTCTTTCCCTACACGACGCTCTTCCGATCTtctcagtc	Adapter	tctcagtc
ApeK1Y_D11F	ACACTCTTTCCCTACACGACGCTCTTCCGATCTccggatat	Adapter	ccggatat
ApeK1Y_D12F	ACACTCTTTCCCTACACGACGCTCTTCCGATCTcgccttat	Adapter	cgccttat
ApeK1Y_E1F	ACACTCTTTCCCTACACGACGCTCTTCCGATCTaggc	Adapter	aggc
ApeK1Y_E2F	ACACTCTTTCCCTACACGACGCTCTTCCGATCTgatc	Adapter	gatc
ApeK1Y_E3F	ACACTCTTTCCCTACACGACGCTCTTCCGATCTtcac	Adapter	tcac
ApeK1Y_E4F	ACACTCTTTCCCTACACGACGCTCTTCCGATCTaggat	Adapter	aggat
ApeK1Y_E5F	ACACTCTTTCCCTACACGACGCTCTTCCGATCTattga	Adapter	attga
ApeK1Y_E6F	ACACTCTTTCCCTACACGACGCTCTTCCGATCTcatct	Adapter	catct
ApeK1Y_E7F	ACACTCTTTCCCTACACGACGCTCTTCCGATCTcctac	Adapter	cctac
ApeK1Y_E8F	ACACTCTTTCCCTACACGACGCTCTTCCGATCTgagga	Adapter	gagga
ApeK1Y_E9F	ACACTCTTTCCCTACACGACGCTCTTCCGATCTggaac	Adapter	ggaac
ApeK1Y_E10F	ACACTCTTTCCCTACACGACGCTCTTCCGATCTgtcaa	Adapter	gtcaa
ApeK1Y_E11F	ACACTCTTTCCCTACACGACGCTCTTCCGATCTtaata	Adapter	taata
ApeK1Y_E12F	ACACTCTTTCCCTACACGACGCTCTTCCGATCTtacat	Adapter	tacat
ApeK1Y_F1F	ACACTCTTTCCCTACACGACGCTCTTCCGATCTtcgtt	Adapter	tcgtt
ApeK1Y_F2F	ACACTCTTTCCCTACACGACGCTCTTCCGATCTacctaa	Adapter	acctaa
ApeK1Y_F3F	ACACTCTTTCCCTACACGACGCTCTTCCGATCTatatgt	Adapter	atatgt
ApeK1Y_F4F	ACACTCTTTCCCTACACGACGCTCTTCCGATCTatcgta	Adapter	atcgta
ApeK1Y_F5F	ACACTCTTTCCCTACACGACGCTCTTCCGATCTcatcgt	Adapter	catcgt
ApeK1Y_F6F	ACACTCTTTCCCTACACGACGCTCTTCCGATCTcgcggt	Adapter	cgcggt
ApeK1Y_F7F	ACACTCTTTCCCTACACGACGCTCTTCCGATCTctatta	Adapter	ctatta
ApeK1Y_F8F	ACACTCTTTCCCTACACGACGCTCTTCCGATCTgccagt	Adapter	gccagt
ApeK1Y_F9F	ACACTCTTTCCCTACACGACGCTCTTCCGATCTggaaga	Adapter	ggaaga
ApeK1Y_F10F	ACACTCTTTCCCTACACGACGCTCTTCCGATCTgtactt	Adapter	gtactt
ApeK1Y_F11F	ACACTCTTTCCCTACACGACGCTCTTCCGATCTgttgaa	Adapter	gttgaa
ApeK1Y_F12F	ACACTCTTTCCCTACACGACGCTCTTCCGATCTtaacga	Adapter	taacga
ApeK1Y_G1F	ACACTCTTTCCCTACACGACGCTCTTCCGATCTtggcta	Adapter	tggcta
ApeK1Y_G2F	ACACTCTTTCCCTACACGACGCTCTTCCGATCTacgtgtt	Adapter	acgtgtt
ApeK1Y_G3F	ACACTCTTTCCCTACACGACGCTCTTCCGATCTattaatt	Adapter	attaatt
ApeK1Y_G4F	ACACTCTTTCCCTACACGACGCTCTTCCGATCTattggat	Adapter	attggat
ApeK1Y_G5F	ACACTCTTTCCCTACACGACGCTCTTCCGATCTcataagt	Adapter	cataagt
ApeK1Y_G6F	ACACTCTTTCCCTACACGACGCTCTTCCGATCTcgctgat	Adapter	cgctgat
ApeK1Y_G7F	ACACTCTTTCCCTACACGACGCTCTTCCGATCTcggtaga	Adapter	cggtaga
ApeK1Y_G8F	ACACTCTTTCCCTACACGACGCTCTTCCGATCTctacgga	Adapter	ctacgga
ApeK1Y_G9F	ACACTCTTTCCCTACACGACGCTCTTCCGATCTgcggaat	Adapter	gcggaat
ApeK1Y_G10F	ACACTCTTTCCCTACACGACGCTCTTCCGATCTtagcgga	Adapter	tagcgga
ApeK1Y_G11F	ACACTCTTTCCCTACACGACGCTCTTCCGATCTtcgaaga	Adapter	tcgaaga
ApeK1Y_G12F	ACACTCTTTCCCTACACGACGCTCTTCCGATCTtctgtga	Adapter	tctgtga
ApeK1Y_H1F	ACACTCTTTCCCTACACGACGCTCTTCCGATCTtgctgga	Adapter	tgctgga
ApeK1Y_H2F	ACACTCTTTCCCTACACGACGCTCTTCCGATCTaaccgaga	Adapter	aaccgaga
ApeK1Y_H3F	ACACTCTTTCCCTACACGACGCTCTTCCGATCTacagggaa	Adapter	acagggaa
ApeK1Y_H4F	ACACTCTTTCCCTACACGACGCTCTTCCGATCTacgtggta	Adapter	acgtggta
ApeK1Y_H5F	ACACTCTTTCCCTACACGACGCTCTTCCGATCTccatgggt	Adapter	ccatgggt
ApeK1Y_H6F	ACACTCTTTCCCTACACGACGCTCTTCCGATCTcgcggaga	Adapter	cgcggaga
ApeK1Y_H7F	ACACTCTTTCCCTACACGACGCTCTTCCGATCTcgtgtggt	Adapter	cgtgtggt
ApeK1Y_H8F	ACACTCTTTCCCTACACGACGCTCTTCCGATCTgctgtgga	Adapter	gctgtgga
ApeK1Y_H9F	ACACTCTTTCCCTACACGACGCTCTTCCGATCTggattggt	Adapter	ggattggt
ApeK1Y_H10F	ACACTCTTTCCCTACACGACGCTCTTCCGATCTgtgagggt	Adapter	gtgagggt
ApeK1Y_H11F	ACACTCTTTCCCTACACGACGCTCTTCCGATCTtatcggga	Adapter	tatcggga
ApeK1Y_H12F	ACACTCTTTCCCTACACGACGCTCTTCCGATCTttcctgga	Adapter	ttcctgga
ApeK1Y_A1R	pCWGggagAGATCGGAAGAGCGGTTCAGCAGGAATGCCGAG	Adapter	NA
ApeK1Y_A2R	pCWGtgcaAGATCGGAAGAGCGGTTCAGCAGGAATGCCGAG	Adapter	NA
ApeK1Y_A3R	pCWGtagtAGATCGGAAGAGCGGTTCAGCAGGAATGCCGAG	Adapter	NA
ApeK1Y_A4R	pCWGtctgAGATCGGAAGAGCGGTTCAGCAGGAATGCCGAG	Adapter	NA
ApeK1Y_A5R	pCWGagttAGATCGGAAGAGCGGTTCAGCAGGAATGCCGAG	Adapter	NA
ApeK1Y_A6R	pCWGacgcAGATCGGAAGAGCGGTTCAGCAGGAATGCCGAG	Adapter	NA

ApeK1Y_A7R	pCWGatcgAGATCGGAAGAGCGGTTCAGCAGGAATGCCGAG	Adapter	NA
ApeK1Y_A8R	pCWGttacAGATCGGAAGAGCGGTTCAGCAGGAATGCCGAG	Adapter	NA
ApeK1Y_A9R	pCWGtcgcaAGATCGGAAGAGCGGTTCAGCAGGAATGCCGAG	Adapter	NA
ApeK1Y_A10R	pCWGaagcgAGATCGGAAGAGCGGTTCAGCAGGAATGCCGAG	Adapter	NA
ApeK1Y_A11R	pCWGggtgaAGATCGGAAGAGCGGTTCAGCAGGAATGCCGAG	Adapter	NA
ApeK1Y_A12R	pCWGgctagAGATCGGAAGAGCGGTTCAGCAGGAATGCCGAG	Adapter	NA
ApeK1Y_B1R	pCWGtttgtAGATCGGAAGAGCGGTTCAGCAGGAATGCCGAG	Adapter	NA
ApeK1Y_B2R	pCWGgagaaAGATCGGAAGAGCGGTTCAGCAGGAATGCCGAG	Adapter	NA
ApeK1Y_B3R	pCWGgggctAGATCGGAAGAGCGGTTCAGCAGGAATGCCGAG	Adapter	NA
ApeK1Y_B4R	pCWGaatacAGATCGGAAGAGCGGTTCAGCAGGAATGCCGAG	Adapter	NA
ApeK1Y_B5R	pCWGtacagAGATCGGAAGAGCGGTTCAGCAGGAATGCCGAG	Adapter	NA
ApeK1Y_B6R	pCWGacggtAGATCGGAAGAGCGGTTCAGCAGGAATGCCGAG	Adapter	NA
ApeK1Y_B7R	pCWGtaagcAGATCGGAAGAGCGGTTCAGCAGGAATGCCGAG	Adapter	NA
ApeK1Y_B8R	pCWGacaccAGATCGGAAGAGCGGTTCAGCAGGAATGCCGAG	Adapter	NA
ApeK1Y_B9R	pCWGacaaccAGATCGGAAGAGCGGTTCAGCAGGAATGCCGAG	Adapter	NA
ApeK1Y_B10R	pCWGagctggAGATCGGAAGAGCGGTTCAGCAGGAATGCCGAG	Adapter	NA
ApeK1Y_B11R	pCWGtctgaaAGATCGGAAGAGCGGTTCAGCAGGAATGCCGAG	Adapter	NA
ApeK1Y_B12R	pCWGttcctaAGATCGGAAGAGCGGTTCAGCAGGAATGCCGAG	Adapter	NA
ApeK1Y_C1R	pCWGtagagcAGATCGGAAGAGCGGTTCAGCAGGAATGCCGAG	Adapter	NA
ApeK1Y_C2R	pCWGttgtggAGATCGGAAGAGCGGTTCAGCAGGAATGCCGAG	Adapter	NA
ApeK1Y_C3R	pCWGtggaagAGATCGGAAGAGCGGTTCAGCAGGAATGCCGAG	Adapter	NA
ApeK1Y_C4R	pCWGtatctcAGATCGGAAGAGCGGTTCAGCAGGAATGCCGAG	Adapter	NA
ApeK1Y_C5R	pCWGaggcatAGATCGGAAGAGCGGTTCAGCAGGAATGCCGAG	Adapter	NA
ApeK1Y_C6R	pCWGtccactAGATCGGAAGAGCGGTTCAGCAGGAATGCCGAG	Adapter	NA
ApeK1Y_C7R	pCWGaaaaataAGATCGGAAGAGCGGTTCAGCAGGAATGCCGAG	Adapter	NA
ApeK1Y_C8R	pCWGaagcaagAGATCGGAAGAGCGGTTCAGCAGGAATGCCGAG	Adapter	NA
ApeK1Y_C9R	pCWGgtttcatAGATCGGAAGAGCGGTTCAGCAGGAATGCCGAG	Adapter	NA
ApeK1Y_C10R	pCWGaacttttAGATCGGAAGAGCGGTTCAGCAGGAATGCCGAG	Adapter	NA
ApeK1Y_C11R	pCWGtgaattcAGATCGGAAGAGCGGTTCAGCAGGAATGCCGAG	Adapter	NA
ApeK1Y_C12R	pCWGgaagttcAGATCGGAAGAGCGGTTCAGCAGGAATGCCGAG	Adapter	NA
ApeK1Y_D1R	pCWGtaggtccAGATCGGAAGAGCGGTTCAGCAGGAATGCCGAG	Adapter	NA
ApeK1Y_D2R	pCWGaatcgacAGATCGGAAGAGCGGTTCAGCAGGAATGCCGAG	Adapter	NA
ApeK1Y_D3R	pCWGaggcgttAGATCGGAAGAGCGGTTCAGCAGGAATGCCGAG	Adapter	NA
ApeK1Y_D4R	pCWGgcatattAGATCGGAAGAGCGGTTCAGCAGGAATGCCGAG	Adapter	NA
ApeK1Y_D5R	pCWGgtagtcgtAGATCGGAAGAGCGGTTCAGCAGGAATGCCGAG	Adapter	NA
ApeK1Y_D6R	pCWGgcatgctaAGATCGGAAGAGCGGTTCAGCAGGAATGCCGAG	Adapter	NA
ApeK1Y_D7R	pCWGatggcctaAGATCGGAAGAGCGGTTCAGCAGGAATGCCGAG	Adapter	NA
ApeK1Y_D8R	pCWGtccttgcaAGATCGGAAGAGCGGTTCAGCAGGAATGCCGAG	Adapter	NA
ApeK1Y_D9R	pCWGacgtaccaAGATCGGAAGAGCGGTTCAGCAGGAATGCCGAG	Adapter	NA
ApeK1Y_D10R	pCWGgactgagaAGATCGGAAGAGCGGTTCAGCAGGAATGCCGAG	Adapter	NA
ApeK1Y_D11R	pCWGatatccggAGATCGGAAGAGCGGTTCAGCAGGAATGCCGAG	Adapter	NA
ApeK1Y_D12R	pCWGataaggcgAGATCGGAAGAGCGGTTCAGCAGGAATGCCGAG	Adapter	NA
ApeK1Y_E1R	pCWGgcctAGATCGGAAGAGCGGTTCAGCAGGAATGCCGAG	Adapter	NA
ApeK1Y_E2R	pCWGgatcAGATCGGAAGAGCGGTTCAGCAGGAATGCCGAG	Adapter	NA
ApeK1Y_E3R	pCWGgtgaAGATCGGAAGAGCGGTTCAGCAGGAATGCCGAG	Adapter	NA
ApeK1Y_E4R	pCWGatcctAGATCGGAAGAGCGGTTCAGCAGGAATGCCGAG	Adapter	NA
ApeK1Y_E5R	pCWGtcaatAGATCGGAAGAGCGGTTCAGCAGGAATGCCGAG	Adapter	NA
ApeK1Y_E6R	pCWGagatgAGATCGGAAGAGCGGTTCAGCAGGAATGCCGAG	Adapter	NA
ApeK1Y_E7R	pCWGgtaggAGATCGGAAGAGCGGTTCAGCAGGAATGCCGAG	Adapter	NA
ApeK1Y_E8R	pCWGtcctcAGATCGGAAGAGCGGTTCAGCAGGAATGCCGAG	Adapter	NA
ApeK1Y_E9R	pCWGgttccAGATCGGAAGAGCGGTTCAGCAGGAATGCCGAG	Adapter	NA
ApeK1Y_E10R	pCWGttgacAGATCGGAAGAGCGGTTCAGCAGGAATGCCGAG	Adapter	NA
ApeK1Y_E11R	pCWGtattaAGATCGGAAGAGCGGTTCAGCAGGAATGCCGAG	Adapter	NA
ApeK1Y_E12R	pCWGatgtaAGATCGGAAGAGCGGTTCAGCAGGAATGCCGAG	Adapter	NA
ApeK1Y_F1R	pCWGaacgaAGATCGGAAGAGCGGTTCAGCAGGAATGCCGAG	Adapter	NA
ApeK1Y_F2R	pCWGttaggtAGATCGGAAGAGCGGTTCAGCAGGAATGCCGAG	Adapter	NA
ApeK1Y_F3R	pCWGacatatAGATCGGAAGAGCGGTTCAGCAGGAATGCCGAG	Adapter	NA

ApeK1Y_F4R	pCWGtacgatAGATCGGAAGAGCGGTTCAGCAGGAATGCCGAG	Adapter	NA
ApeK1Y_F5R	pCWGacgatgAGATCGGAAGAGCGGTTCAGCAGGAATGCCGAG	Adapter	NA
ApeK1Y_F6R	pCWGaccgcgAGATCGGAAGAGCGGTTCAGCAGGAATGCCGAG	Adapter	NA
ApeK1Y_F7R	pCWGtaatagAGATCGGAAGAGCGGTTCAGCAGGAATGCCGAG	Adapter	NA
ApeK1Y_F8R	pCWGactggcAGATCGGAAGAGCGGTTCAGCAGGAATGCCGAG	Adapter	NA
ApeK1Y_F9R	pCWGtcttccAGATCGGAAGAGCGGTTCAGCAGGAATGCCGAG	Adapter	NA
ApeK1Y_F10R	pCWGaagtacAGATCGGAAGAGCGGTTCAGCAGGAATGCCGAG	Adapter	NA
ApeK1Y_F11R	pCWGttcaacAGATCGGAAGAGCGGTTCAGCAGGAATGCCGAG	Adapter	NA
ApeK1Y_F12R	pCWGtcgttaAGATCGGAAGAGCGGTTCAGCAGGAATGCCGAG	Adapter	NA
ApeK1Y_G1R	pCWGtagccaAGATCGGAAGAGCGGTTCAGCAGGAATGCCGAG	Adapter	NA
ApeK1Y_G2R	pCWGaacacgtAGATCGGAAGAGCGGTTCAGCAGGAATGCCGAG	Adapter	NA
ApeK1Y_G3R	pCWGaattaatAGATCGGAAGAGCGGTTCAGCAGGAATGCCGAG	Adapter	NA
ApeK1Y_G4R	pCWGatccaatAGATCGGAAGAGCGGTTCAGCAGGAATGCCGAG	Adapter	NA
ApeK1Y_G5R	pCWGacttatgAGATCGGAAGAGCGGTTCAGCAGGAATGCCGAG	Adapter	NA
ApeK1Y_G6R	pCWGatcagcgAGATCGGAAGAGCGGTTCAGCAGGAATGCCGAG	Adapter	NA
ApeK1Y_G7R	pCWGtctaccgAGATCGGAAGAGCGGTTCAGCAGGAATGCCGAG	Adapter	NA
ApeK1Y_G8R	pCWGtccgtagAGATCGGAAGAGCGGTTCAGCAGGAATGCCGAG	Adapter	NA
ApeK1Y_G9R	pCWGattccgcAGATCGGAAGAGCGGTTCAGCAGGAATGCCGAG	Adapter	NA
ApeK1Y_G10R	pCWGtccgctaAGATCGGAAGAGCGGTTCAGCAGGAATGCCGAG	Adapter	NA
ApeK1Y_G11R	pCWGtcttcgaAGATCGGAAGAGCGGTTCAGCAGGAATGCCGAG	Adapter	NA
ApeK1Y_G12R	pCWGtcacagaAGATCGGAAGAGCGGTTCAGCAGGAATGCCGAG	Adapter	NA
ApeK1Y_H1R	pCWGtccagcaAGATCGGAAGAGCGGTTCAGCAGGAATGCCGAG	Adapter	NA
ApeK1Y_H2R	pCWGtctcggttAGATCGGAAGAGCGGTTCAGCAGGAATGCCGAG	Adapter	NA
ApeK1Y_H3R	pCWGttccctgtAGATCGGAAGAGCGGTTCAGCAGGAATGCCGAG	Adapter	NA
ApeK1Y_H4R	pCWGtaccacgtAGATCGGAAGAGCGGTTCAGCAGGAATGCCGAG	Adapter	NA
ApeK1Y_H5R	pCWGacccatggAGATCGGAAGAGCGGTTCAGCAGGAATGCCGAG	Adapter	NA
ApeK1Y_H6R	pCWGtctccgcgAGATCGGAAGAGCGGTTCAGCAGGAATGCCGAG	Adapter	NA
ApeK1Y_H7R	pCWGaccacacgAGATCGGAAGAGCGGTTCAGCAGGAATGCCGAG	Adapter	NA
ApeK1Y_H8R	pCWGtccacagcAGATCGGAAGAGCGGTTCAGCAGGAATGCCGAG	Adapter	NA
ApeK1Y_H9R	pCWGaccaatccAGATCGGAAGAGCGGTTCAGCAGGAATGCCGAG	Adapter	NA
ApeK1Y_H10R	pCWGaccctcacAGATCGGAAGAGCGGTTCAGCAGGAATGCCGAG	Adapter	NA
ApeK1Y_H11R	pCWGtcccgataAGATCGGAAGAGCGGTTCAGCAGGAATGCCGAG	Adapter	NA
ApeK1Y_H12R	pCWGtccaggaaAGATCGGAAGAGCGGTTCAGCAGGAATGCCGAG	Adapter	NA

Table 4.1: Primers used in this study. Primers were used for PCR, index sequencing ("Index"), or making adapters.

Primer concentration and cycle number were varied to amplify enough product for sequencing. PCR products were purified and size-selected with AMPure XP beads (Agencourt) with a bead:sample ratio of 0.7. Libraries were analyzed on an Agilent 2100 Bioanalyzer High-Sensitivity Chip for quality control and sequenced with 100 base, paired-end sequencing on an Illumina HiSeq 2000 sequencer.

Sequencing of cross grandparents and identification of homozygous SNPs

The two grandparents of the FTC cross and the two grandparents of the BEPA cross were resequenced with a Nextera DNA Sample Preparation kit (Illumina) to approximately 60X and 6X coverage, respectively, with 100 base paired-end sequencing on an Illumina HiSeq 2000 sequencer. Reads were mapped to the reference genome with BWA

Library	Barcodes	F2s	Adapters	Total Reads	Barcoded F2 Reads	Barcoded Reads/F2	Mapped Reads/F2	Mapped Read COV	SNP Coverage	Marker Coverage	Genotype Fail %	Note
1	48	12	Elshire	216	58.8	4.9	3.1	0.5	3.1	371.7	0.04	a
2	48	48	Elshire	212.4	168.3	3.5	2.4	1.03	2.4	288.2	0.5	a
3	48	48	Elshire	418.6	397.7	8.3	5.6	1.18	0.7	85.1	5.4	
4	96	96	Y-shaped	559.7	487.6	5.1	3.9	0.61	3.5	424.7	0.3	
5	96	96	Y-shaped	428.2	363.1	3.8	3	0.59	2.8	228.5	1.1	
6	96	96	Y-shaped	440.5	371.9	3.9	3	0.61	2.8	230.5	0.9	
7	384	332	Y-shaped	498	259.1	0.8	0.6	0.81	0.5	49.1	2.6	b, c
All	816	719	*	2773.3	2101.1	2.9	2.2	1.26	1.7	168.6	1.9	

Table 4.2: Summary of GBS libraries
For each library, the number of barcodes and F2 fish included is listed. Total reads, barcoded F2 reads, barcoded reads/F2 and mapped reads/F2 are reported in units of millions of reads. COV = coefficient of variance (mean/standard deviation). Genotype Fail % indicates percentage of final genotypes that were missing. Note (a): For Libraries 1 and 2, the R2 read failed, resulting in half the expected number of reads. Note (b): ApeK1 digestion and adapter ligation reactions were done at half volume to conserve costs. Note (c): Library 7 included nine F2 samples that were sequenced in Libraries 1-6 but had very low sequencing coverage.

(www.bio-bwa.sourceforge.net). For each grandparent, SNPs relative to the reference genome were called with SAMtools (www.samtools.sourceforge.net). The SNPs in the resulting VCF files were then filtered with a custom Python script (www.python.org) to identify sites where one grandparent was homozygous for the alternate allele (a homozygous alternate SNP was present in the VCF file) and the other grandparent was homozygous for the reference allele (no SNP was present in the VCF file but there was sufficient sequencing coverage to support that the sample was homozygous reference). This analysis identified 1,116,087 and 715,279 sites which were homozygous different between the FTC and BEPA grandparents, respectively ("homozygous SNP positions").

Processing reads from GBS libraries

The two grandparents of the FTC cross and the two grandparents of the BEPA cross were resequenced to approximately 60X and 6X coverage, respectively, and in each cross, sites where one grandparent was homozygous for one allele and the other grandparent was homozygous for a second allele were identified ("homozygous SNP positions"; see Supplemental Methods). GBS reads from F2s were sorted by barcode with a custom Perl script. Reads were mapped to the stickleback reference genome with BWA using default settings (www.bio-bwa.sourceforge.net), allowing up to a 4% difference between reads and the reference genome. We devised a method to identify high quality, segregating SNPs. For each homozygous SNP position, F2 GBS reads overlapping the SNP were considered, and the number of reads supporting marine and freshwater alleles for each homozygous SNP position was determined with SAMtools (www.samtools.sourceforge.net). Genomic positions not identified as homozyogous different in the grandparent resequencing were not examined in the F2s. For each homozygous SNP position, a weighted average of these values was calculated across all F2s, normalized by the total number of mapped reads for each F2. We multiplied the marine and freshwater weighted averages by 10^6 to calculate "Reads Per Million Mapped" (RPMM). Properly segregating SNPs should have an approximately 1:1 ratio of marine:freshwater alleles as was observed for most SNPs (**Figure 4.3C**). However, we observed some genomic regions that had a skewed allele ratio in the F2s, possibly due to meiotic drive and/or the lethality of particular genotypic classes. For example, a region of chromosome 2 centered at marker 0_32 in the FTC cross had a freshwater allele frequency of 0.61. Therefore, a wider range of allele ratios was allowed for individual SNPs (a marine/freshwater RPMM ratio between 4:1 and 1:4). To include SNPs with true segregation bias, skewed markers adjacent to other similarly skewed markers were included, but skewed markers surrounded by non-skewed markers were removed (see below). Additionally, SNPs were filtered for those with an average marine + freshwater RPMM between 0.2 and 3.0 in order to have a set of SNPs with similar coverage levels. A separate set of sieving parameters was used to determine sex chromosome genotypes (see below).

These filtered SNPs were further grouped into bins of at most 500 kb. Bin size was scaled to divide each scaffold into evenly sized bins. Scaffolds smaller than 100 kb were binned into one bin, scaffolds between 100 kb and 1 Mb were binned into 2 equal size bins, scaffolds between 1 Mb and 1.5 Mb were divided into three bins, scaffolds

			Samples			Markers						
Cross	GP SNPs	F2 SNPs used	Initial	Low coverage	Final	Initial	Low coverage	Skewed allele ratio	Final	SNP Coverage	Marker Coverage	Genotype Fail %
FTC	1,116,087	131,091	358	2	356	1,085	59	25	1,001	1.63X	197X	1.9
BEPA	715,279	87,419	361	2	359	1,077	39	60	978	1.73X	140X	1.8

Table 4.3: Cross summary statistics. Numbers of initial samples, dropped samples due to low coverage, and final samples are presented. Numbers of initial markers, dropped markers due to low coverage and skewed allele ratios, and final markers are presented. GP=grandparent. Genotype fail % = percentage of missing genotypes.

between 1.5 and 2 Mb were divided into four bins, *etc*. For each SNP within a bin, marine and freshwater read counts were summed and genotypes were called.

To call genotypes from read counts, a likelihood method was adapted from (Hohenlohe et al. 2010), but with only reads supporting the phased marine or freshwater allele considered. Three possible diploid genotypes were considered: G_i=G_{MM}, G_{MF}, or G_{FF} (M=marine allele, F=freshwater allele). Reads were considered to be independent, and have an equal chance of emanating from the two alleles. An error rate of 1% was assumed for each read, which increased the conservativeness of the genotyping calls. Bayes' theorem was used to calculate the probability of each genotype (G_i) given the observed pattern of reads. For example, for the MM genotype: $p(G_{MM}|reads)=p(reads|G_{MM})/[p(reads|G_{MM})+p(reads|G_{MF})+p(reads|G_{FF})]$. If one genotype G_i was most likely with $p(G_i|reads) > 0.95$, the genotype was assigned as G_i. Otherwise, the genotype was assigned as "MF/FF" or "MM/MF" if $p(G_{MM})<0.05$ or $p(G_{FF})<0.05$ (**Figure 4.2**). These genotypes were frequently called for bins spanning recombinant breakpoints. If no reads were present, a missing genotype ("NA") was assigned.

Fish that had missing genotypes for over 50% of markers were removed from the analysis (n=2 in each cross). Nine additional samples with high rates of missing genotypes in libraries 1-6 were resequenced successfully in library 7. Markers that had missing data for at least 20% of fish were removed from the analysis (n=59 and 39 in the FTC and BEPA crosses, respectively). Markers were also removed that had aberrant allelic ratios (n=25 and 60 in the FTC and BEPA crosses, respectively). To remove markers with aberrant allele ratios, each marker was tested for deviation from the predicted 1:2:1 allele ratio by a chi-square test. Multiple adjacent markers with allelic distortion likely indicate a true allelic distortion of that genomic region, whereas a single marker with allelic distortion likely indicates genotyping error. Therefore, markers showing significant allelic distortion ($p<0.05$) without any adjacent makers with allelic distortion were removed. Genetic linkage maps were created with JoinMap 4.0 (Kyazma) using regression mapping with default settings.

Calling sex chromosome genotypes

The stickleback sex chromosome (chromosome 19) consists of a small pseudo-autosomal region and a large region that behaves like a sex chromosome (Peichel et al. 2004; Roesti et al. 2013). The cutoff for the boundary between these regions was chosen at 2.41 Mb, based on levels of sequencing coverage in male and female F2s. Genotypes for the pseudo-autosomal region were determined with the same method as the autosomal chromosomes, whereas a separate pipeline was used for the rest of the sex chromosome. First, reads that mapped to the X chromosome but not the Y chromosome were identified by sieving for SNPs that had an approximately 2:1 ratio of females to male coverage (SNPs that had female:male average RPMM ratios between 1.5 and 3.5). Female F2s should have a 3:1 ratio of X chromosomes originating from their grandmother:grandfather, whereas male F2s should have a 1:1 ratio. Therefore, a 3:1 ratio of alleles was filtered for in female F2s (average RPMM between 2 and 6) and a 1:1 ratio was filtered for in male F2s (average RPMM between 0.2 and 4). Finally, since all male F2s have a Y chromosome that originated from their grandfather, male F2

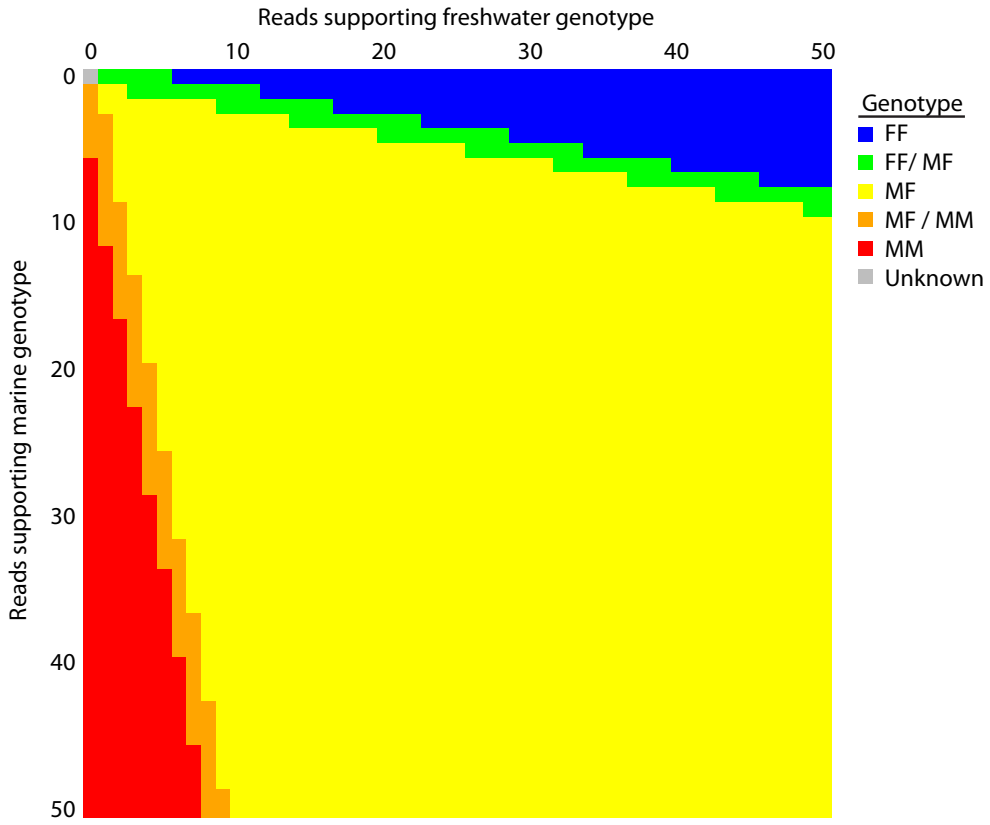


Figure 4.2: Genotype calling from marine and freshwater read counts
Reads were assumed to be independent and have a 1% error rate. M=marine allele, F=freshwater allele. The FF/MF and MF/MM genotypes indicate that there is a >95% probability that the genotype is not the third possible genotype, but that there is not a >95% certainty for any single genotype call. These genotypes were frequently called for bins that spanned recombination breakpoints.

genotypes that were genotyped as homozygous for their grandmother's allele were converted to heterozygous.

Using read correlations to anchor scaffolds

In addition to the linkage maps made with JoinMap, a second more sensitive, but less precise, method was used to further determine the genomic location of scaffolds. This method enabled the mapping of even more scaffolds, but did not reveal the scaffolds' precise genomic location and orientation. Every pair of markers was considered ("marker 1" and "marker 2"). Fish genotyped as homozygous marine or homozygous freshwater for marker 1 were examined. For these fish, raw read counts for marker 2 were summed and were scored as concordant or discordant with the marker 1 genotype. If the total concordance percentage was above 95%, marker 2 was considered linked to marker 1 with ~<5 cM distance. In 300 out of 302 cases (99.3%) where a scaffold mapped to a

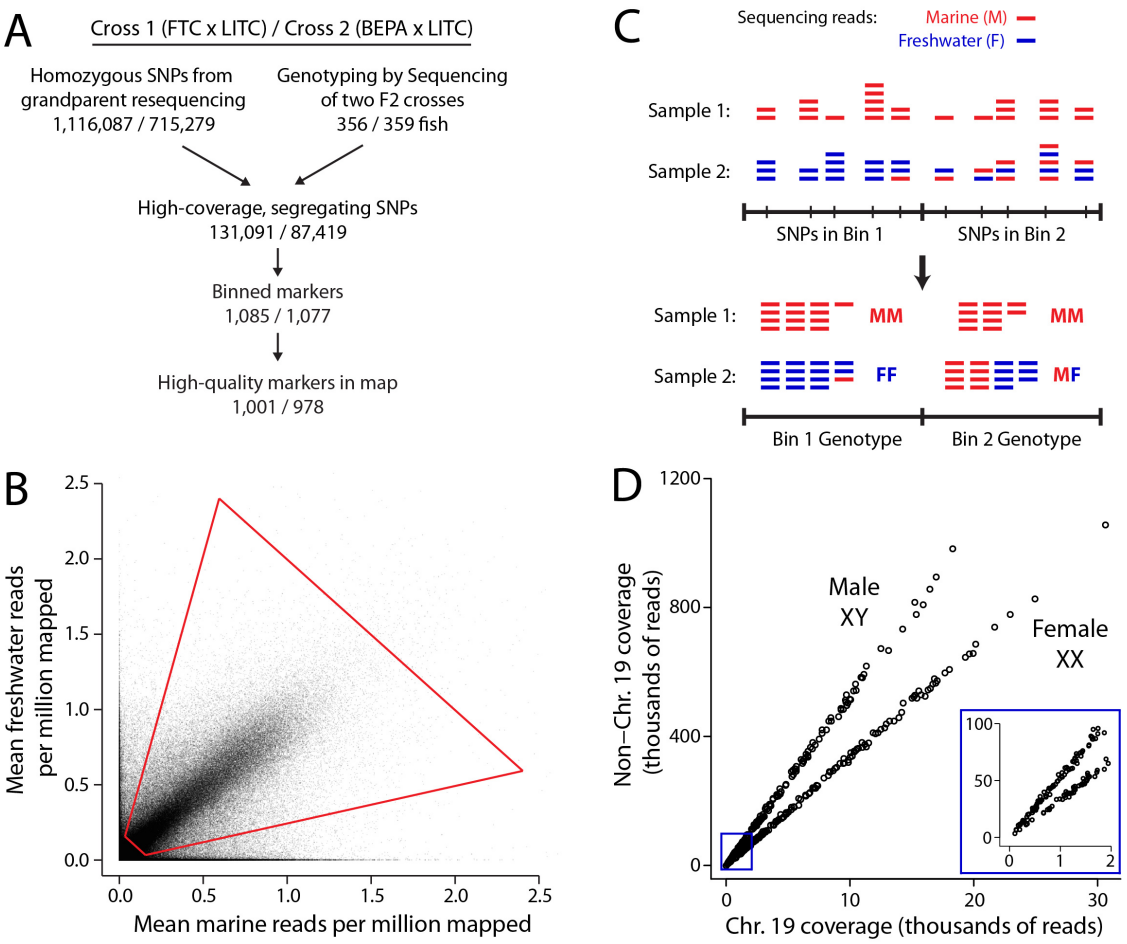


Figure 4.3: Genotyping-by-Sequencing Approach
(A) Flowchart of Genotyping-by-Sequencing (GBS). For each cross, the two grandparents were resequenced to determine homozygous SNP differences, which were filtered for high coverage levels and expected allele ratios in F2s (see Methods). (B) Sieve for high coverage, segregating SNPs. For each SNP, the mean number of mapped reads supporting the marine and freshwater alleles, normalized for the number of millions of reads mapped per sample, is displayed. Data is shown for the FTC x LITC cross. Sieve is shown with red quadrilateral: freshwater allele frequency between 0.2 and 0.8, and total coverage between 0.2 and 3. (C) Diagram of binning approach. Low-coverage sequencing generated read pileup at a large number of SNPs. For each F2, SNPs were binned by counting the total number of marine and freshwater reads within the bin and determining a genotype from the pooled counts. Sample 2 illustrates a case where a recombination breakpoint is near the boundary between two bins and Bin 1 containing the breakpoint is considered to have the FF genotype. Alternatively, bins containing recombination breakpoints also frequently were called with uncertain MF/FF or MM/MF genotypes (Figure 4.2). (D) Calling sex from sex chromosome (chromosome 19) coverage. Females (XX) have approximately equal sex chromosome and autosome coverage levels, whereas males (XY) have approximately half the coverage level on the sex chromosome compared to the autosomes. Data is shown for the FTC cross. Inset: zoom-in of low-coverage samples showing that female and male fish can still be distinguished.

chromosome by this method in both crosses, the chromosome was the same. The other two cases were likely due to reads mapping to repetitive genomic segments.

Sex determination

The sex of each F2 fish was determined by calculating the ratio of chromosome 19 (sex chromosome) coverage levels to non-chromosome 19 coverage levels. For females (XX) and males (XY), the theoretical ratios are 1 and 0.5, respectively, with a cutoff of 0.75. Indeed, in each cross the ratio cleanly formed two groups. However, there was slightly more sex chromosome coverage than expected in each cross. Therefore cutoff ratios in each cross were determined manually based on the empirical ratio distribution (0.85 in the FTC cross and 1.15 in the BEPA cross). Fish above and below the cutoff ratio were assigned as female and male, respectively.

Creating a consensus scaffold map

For each cross, the mean genetic positions of each scaffold in the linkage map were compared to create a scaffold map. In cases where multiple markers from the same scaffold were in the linkage map, scaffold orientation was determined by whether correlation between the physical and genetic positions of these markers was significantly positive or negative in a linear regression. A consensus scaffold map was created by merging the largely identical scaffold maps from the two crosses. In cases where the maps disagreed in marker order, the consensus map used the order from the map with the fewest discordant genotypes.

Fine mapping of recombinant breakpoints

Since the markers used for linkage mapping were based on binning multiple SNPs together, a complementary approach was used to fine-map recombinant breakpoints with a Hidden Markov Model (approach adapted from Andolfatto et al. 2011). SNP genotypes were binned into 10 kb bins, near the average density of SNPs used in the study. For each fish, each bin was assigned a raw genotype of M (only marine reads), F (only freshwater reads), B (both marine and freshwater reads), or X (no data). These raw genotypes were used as the observed data in a Hidden Markov Model with hidden states M (marine), F (freshwater), or H (heterozygous). The model was trained with the Baum-Welch algorithm on 1 million data points. For each fish and bin, the model was used to calculate the probability of each hidden state with the forward-backward algorithm. Details on the trained parameters of the model are presented in **Figure 4.9D**. Locations within a scaffold where the probability switched from over 99% probability of state A to over 99% probability of state B were considered to be the boundaries of a recombination event.

Phenotyping

To stain bones, fish were fixed for 1-2 days in 10% neutral buffered formalin or 5-7 days in 4% paraformaldehyde in 1xPBS, washed with water overnight, stained overnight with 0.008% Alizarin Red S in 1% potassium hydroxide, destained in water overnight, and cleared in a 0.25% potassium hydroxide, 50% glycerol solution.

Lateral plate number and gill raker length were measured from Alizarin-stained fish. The average of plate counts on the left and right sides was used for QTL mapping.

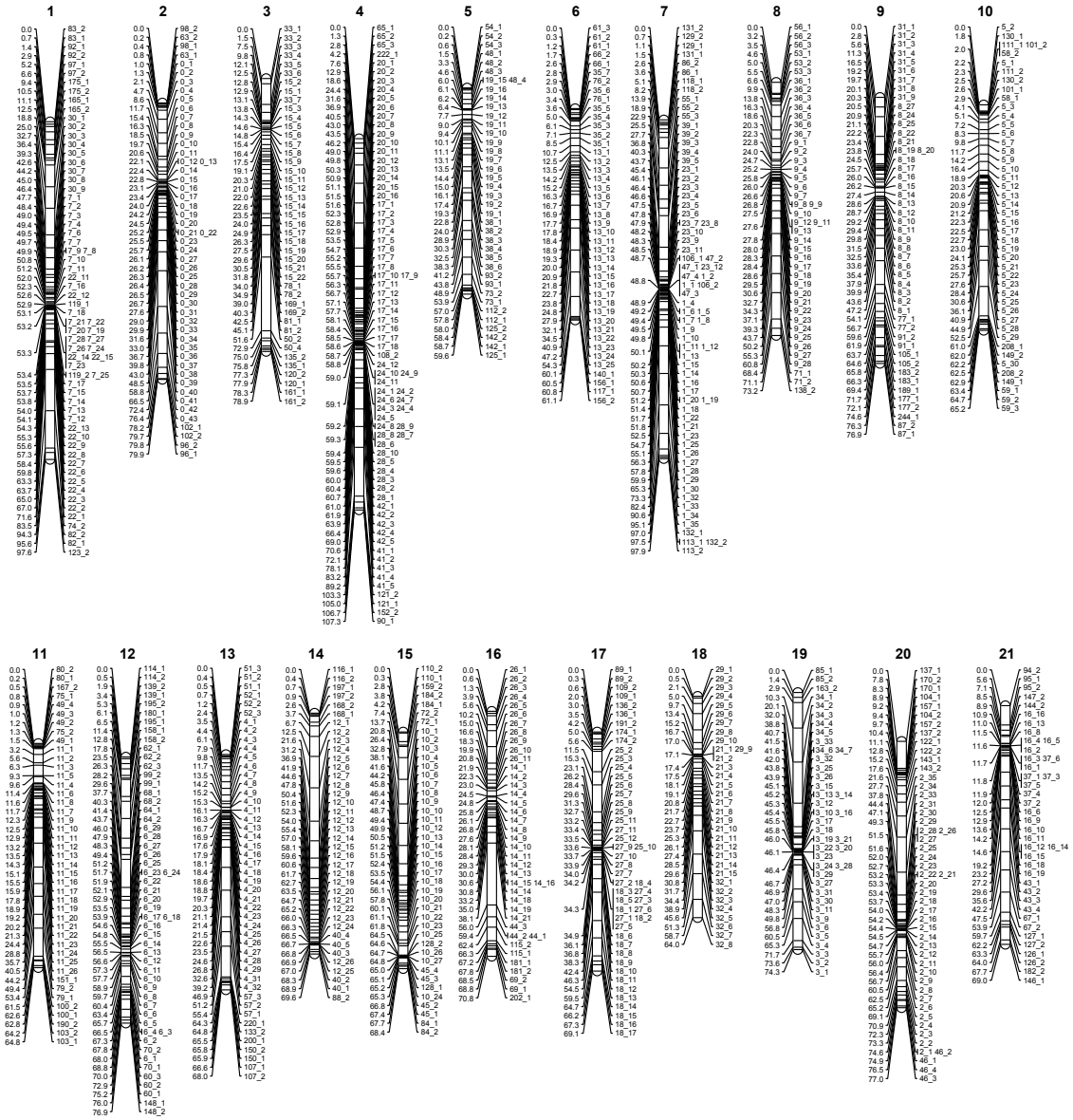


Figure 4.4: Linkage map from FTC x LITC cross
Diagram of linkage map generated from the FTC x LITC cross. For each chromosome, genetic position of each marker is shown on the left in centiMorgans, and marker name is shown on the right.

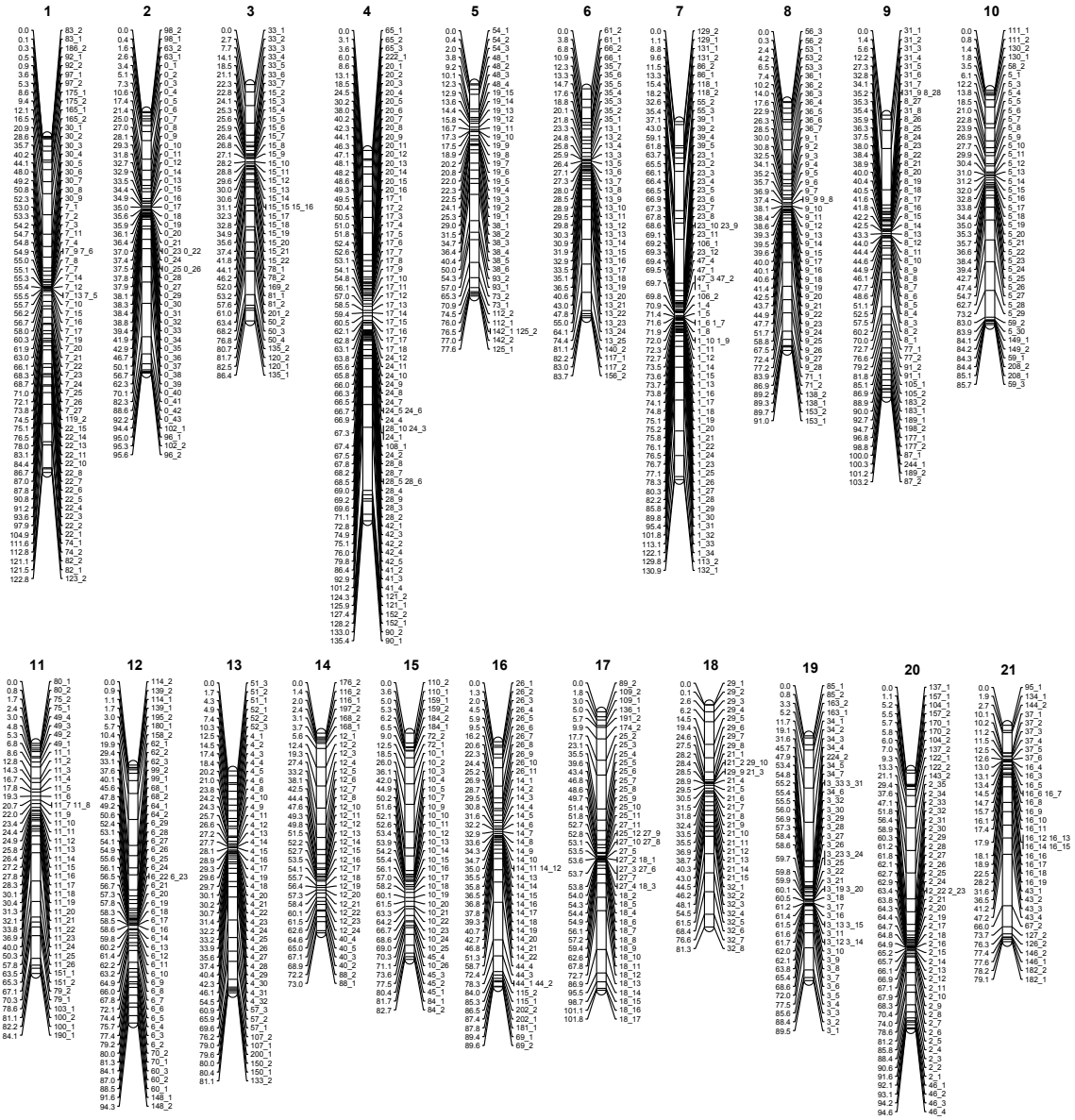


Figure 4.5: Linkage map from BEPA x LITC cross
Diagram of linkage map generated from the BEPA x LITC cross. For each chromosome, genetic position of each marker is shown on the left in centiMorgans, and marker name is shown on the right.

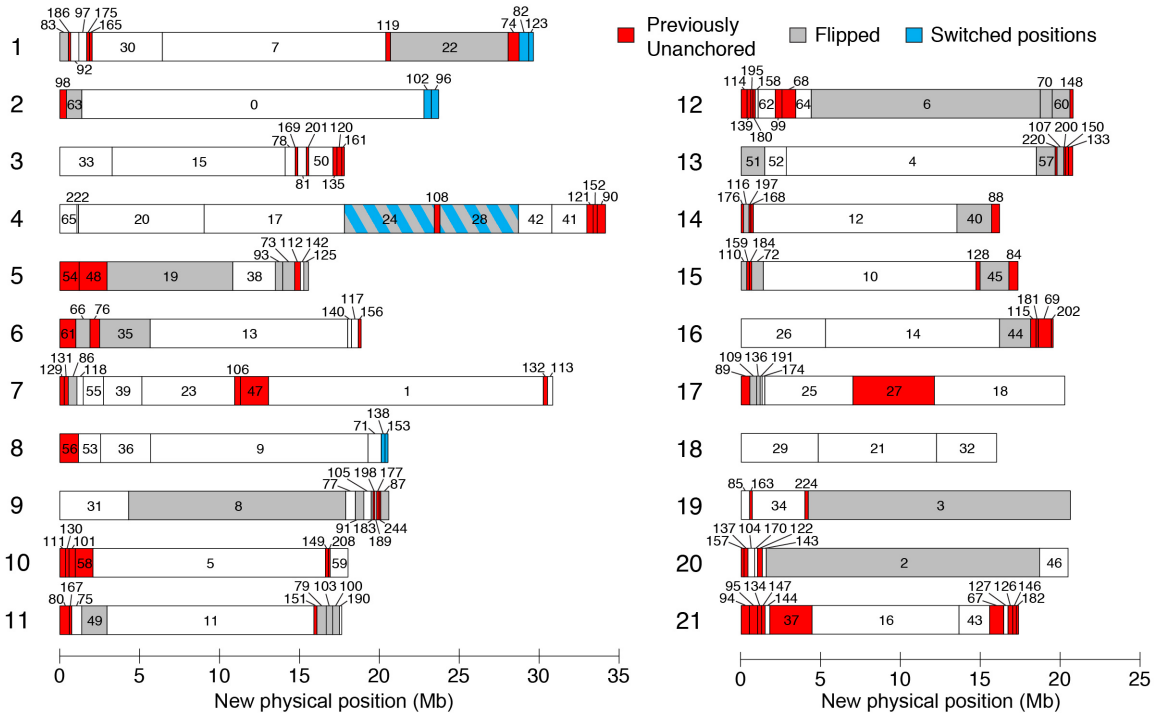


Figure 4.6: A revised map of stickleback scaffold order and orientation
Consensus scaffold map from the two crosses. Chromosomes are numbered on the left, and scaffolds numbered, with previously unanchored scaffolds colored red, previously anchored scaffolds whose orientation has been flipped colored grey, and scaffolds that have switched positions colored blue. Figure style adapted from Roesti et al. 2013.

To phenotype gill raker length, branchial skeletons were dissected out of fish and mounted flat on bridged coverslips as described (Miller et al. 2014). Measurements were obtained by acquiring digital images of left side row 1 ventral gill rakers on a Leica M165 microscope, and tracing a line segment from the gill raker base to tip in imageJ (Schneider et al. 2012). Three gill raker lengths were measured on the first ceratobranchial: lateral (2nd gill raker from end near ventral/dorsal joint), middle (middle of ceratobranchial), and medial (2nd gill raker from end near midline).

Phenotype processing

For QTL mapping, plate number and gill raker length phenotypes were tested for an association with standard length and sex by linear regression in R (www.r-project.org) and corrected for size, sex, and/or log transformed, when appropriate. When association with standard length was significant ($p<0.05$), residuals were taken from a linear regression, then back-transformed to values expected for a 40 mm (marine x freshwater F2s) or 50 mm (lab-reared incross) standard length fish. When correlations with standard length and sex were both significant in a linear model, both were corrected for. When log

	Scaffolds		Length (Mb)		Genes	
Chr	Original	Revised	Original	Revised	Original	Revised
1	8	13	28.18	29.63	1262	1328
2	4	5	23.29	23.7	861	907
3	5	10	16.79	17.8	934	1004
4	8	12	32.63	34.14	1329	1410
5	6	9	12.25	15.56	733	861
6	5	8	17.08	18.85	721	760
7	7	12	27.93	30.84	1320	1481
8	6	7	19.36	20.53	885	924
9	9	11	20.24	20.58	1012	1016
10	2	8	15.66	18.03	816	931
11	7	10	16.7	17.64	1060	1108
12	6	13	18.4	20.76	1007	1138
13	6	9	20.08	20.74	971	1014
14	3	7	15.24	16.17	739	792
15	5	8	16.19	17.32	778	823
16	3	7	18.11	19.52	803	864
17	6	8	14.6	20.25	702	1064
18	4	3	16.28	15.99	764	739
19	3	5	20.24	20.61	1046	1086
20	5	8	19.73	20.45	934	990
21	5	13	11.71	17.35	464	614
Total	113	186	400.7	436.45	19141	20854

Table 4.4: Summary of improved genome assembly
A comparison of the original genome assembly and the revised scaffold order presented in this study. The number of scaffolds, the physical size in megabases (Mb), and the number of Ensembl-predicted genes (Jones et al. 2012) are compared.

transformation improved the normality of the residuals (i.e. by resulting in an Anderson-Darling test changing from $p<0.05$ to $p>0.05$), log transformation was performed.

QTL mapping

For QTL mapping, plate number and gill raker length phenotypes were tested for an association with standard length and sex by linear regression in R (www.r-project.org) and corrected for size, sex, and/or log transformed, when appropriate (see Supplemental Methods). QTL mapping was performed in R/qtl (Broman and Sen 2009; Broman et al. 2003). Initial QTL mapping was performed with *scanone* with Haley-Knott regression. Trait-specific genome-wide significance thresholds with an α of 0.05 were calculated with 1,000 permutations. In cases where multiple significant QTL affected a phenotype,

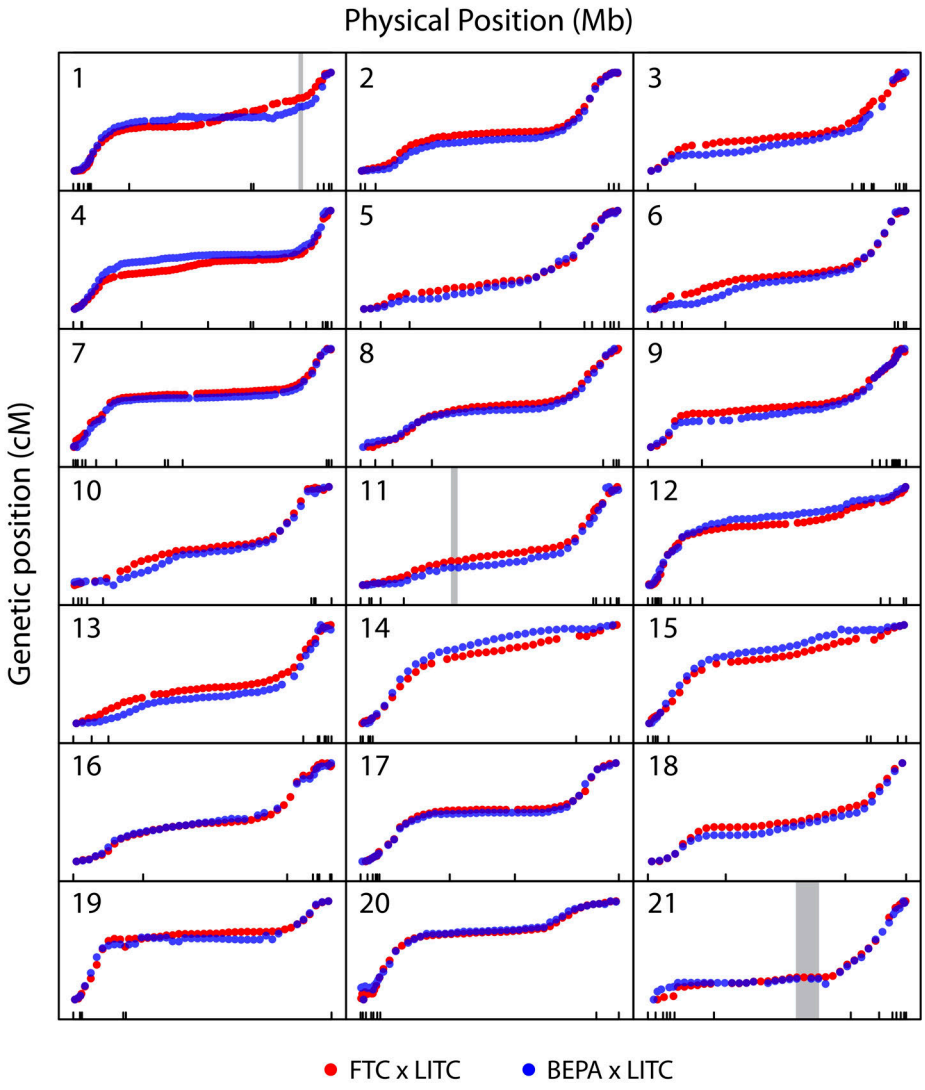


Figure 4.7: Similar genome-wide recombination patterns in both crosses
Plots of genetic vs. physical position for each chromosome. Plots are scaled to have constant width and height for each chromosome. Markers from the FTC and BEPA crosses are plotted in blue and red, respectively. Physical position is according to the revised scaffold map, and tick marks along the x-axis indicate scaffold boundaries. Most chromosomes have highly similar regions of high and low recombination rates between the two crosses. The positions of the three marine/freshwater inversions on chromosomes 1, 11, and 21 reported in Jones et al. 2012, are indicated with gray rectangles. For a closer zoom-in of these inversions, see Figure 4.9 which shows that no recombination events were detected within the three inversions.

multiple QTL mapping was performed with *stepwiseqtl,* QTL peak markers and LOD plots calculated with *refineqtl*, and peak LOD scores and percent variance explained values calculated with *fitqtl*, adjusting for the effect of other QTL underlying the phenotype when appropriate.

RESULTS

Two genome-wide linkage maps

To build genome-wide linkage maps in two marine x freshwater stickleback F2 crosses, we used a binned GBS approach (modified from Elshire et al. 2011, outlined in Figure 1A). Sticklebacks from two independently derived freshwater populations (Fish Trap Creek—"FTC" and Bear Paw Lake—"BEPA") were crossed to fish from a single marine population (Little Campbell River—"LITC"). These two F2 crosses are hereafter called the "FTC" and "BEPA" crosses. F2 fish (n=358 and 361 in the FTC and BEPA crosses, respectively) were sequenced with GBS, multiplexing up to 384 samples in a single Illumina lane (**Table 4.2, Figure 4.1, Table 4.3**). SNPs were phased using grandparent resequencing and filtered for those that had proper allele ratios and coverage levels, resulting in 131,091 and 87,419 high-quality, segregating SNPs in the FTC and BEPA crosses (**Figure 4.3A-B**). To generate high-quality genotypes, we binned together multiple low-coverage SNPs into high-coverage binned markers (referred to as "markers" in this study), using equal size bins of at most 500 kb (**Figure 4.3C**). Linkage maps were made with 1,001 and 978 markers in the FTC and BEPA crosses (**Figure 4.3A, 4.4, 4.5**). These maps had missing genotype rates of 1.9% (FTC) and 1.8% (BEPA). The sex of each F2 fish was determined from sequencing coverage levels of the sex chromosome (**Figure 4.3D**).

Improvements to stickleback genome assembly

The stickleback reference genome assembly (Jones et al. 2012) contains 113 anchored and 1,822 unanchored scaffolds, which comprise 86.8% and 13.2% of the genome assembly, respectively. In both linkage maps, all previously anchored scaffolds mapped to their originally assigned chromosome. In addition, in all cases where a scaffold mapped to a chromosome in both crosses, the scaffold mapped to the same chromosome. Combining the two linkage maps, we generated a consensus scaffold map containing 186 scaffolds, which differed from the genome assembly in 153 places. These differences consisted of 78 previously unanchored scaffolds that were newly anchored in the genome (comprising 36.1 Mb), 40 inversions of previously anchored scaffolds (113.3 Mb), and 4 rearrangements of previously anchored scaffolds (12.8 Mb) (**Figure 4.6**). In the consensus scaffold map, the 186 total scaffolds comprised 94.6% of the total assembly sequence and included the largest 124 scaffolds. Based on the linkage map positions of markers within each scaffold, an orientation was determined for 166 of the 186 scaffolds (436.2 Mb). The anchored scaffolds in the consensus scaffold map contained 9% more bases and Ensembl-predicted genes (Jones et al. 2012) than in the original assembly (**Table 4.4**).

We developed a second, more sensitive method to map unanchored scaffolds by examining the correlation of read counts for every pair of markers (see Methods). With this read correlation method, 538 scaffolds (96.9% of total assembly sequence), including 352 scaffolds not mapped by the first method, mapped to within approximately 5 centiMorgans (cM) of a marker in the consensus scaffold map. Compared to the linkage map-based assembly, an additional 10.8 Mb and 490 Ensembl-predicted genes were linked to a chromosome.

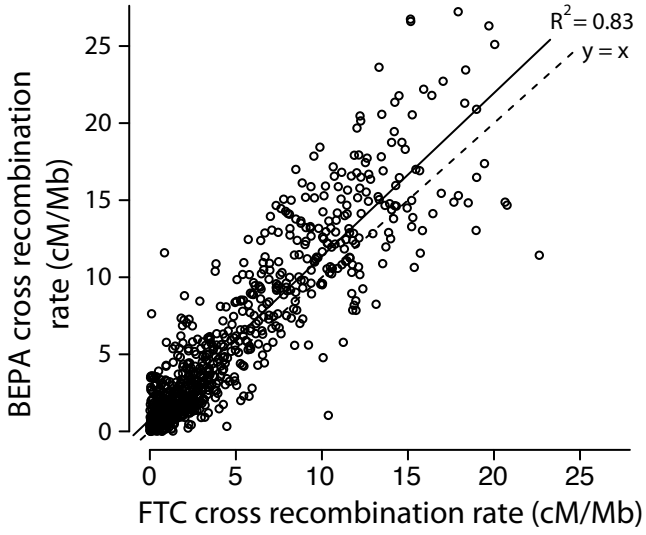
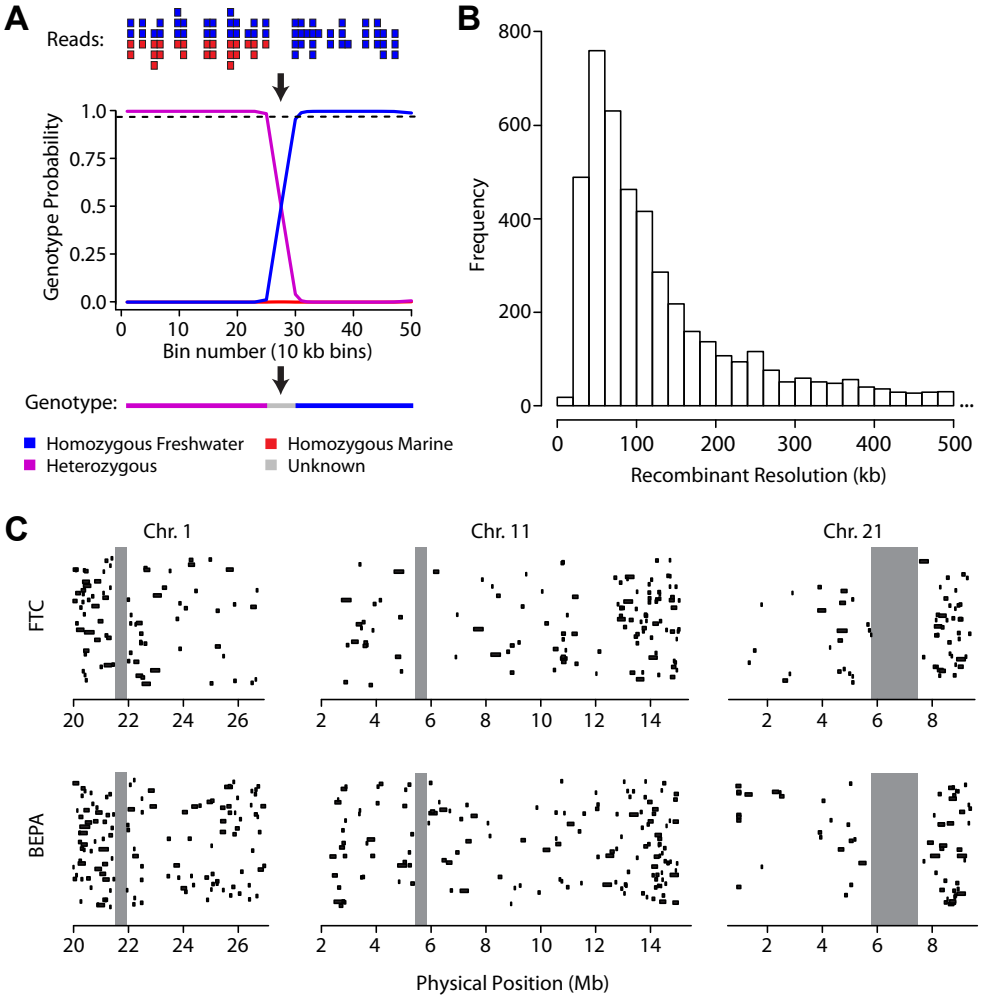


Figure 4.8: High correlation of recombination rates
Recombination rate in the FTC cross plotted against recombination rate in the BEPA cross. Each point is a marker. Recombination rates were calculated by taking the local slope of the loess-smoothed genetic vs. physical plots shown in **Figure 4.7**. Recombination rates are highly correlated ($R^2 = 0.83$). Solid line: best fit line from a linear regression; dotted line: y=x. Total genetic map size was smaller in the FTC cross (1569.9 cM) compared to the BEPA cross (1963.3 cM). Most points were displaced from the y=x line, indicating a higher recombination rate throughout the genome in the BEPA cross.

To determine whether large-scale genomic rearrangements or patterns of recombination rates differ between freshwater populations, we examined the genome-wide patterns of recombination. The two crosses did not indicate any large-scale differences in genomic structure and had strikingly similar patterns of recombination across the genome, with similar regions of high and low recombination (**Figure 4.7**). The BEPA cross had an elevated overall recombination rate relative to the FTC cross, with a total map size of 1570 cM and 1963 cM in the FTC and BEPA crosses, respectively. This difference was due to an elevated rate of recombination throughout the genome (**Figure 4.8**). Consistent with a previous study (Roesti et al. 2013), most chromosomes appeared to have suppressed recombination in the middle, with ends of chromosomes having higher rates of recombination (**Figure 4.7**). The pattern of recombination within each chromosome correlated partially with previously described chromosome morphologies (Urton et al. 2011). For example, as predicted, recombination rates were high on both ends of metacentric chromosome 7 and recombination occurred mostly on one end of telocentric chromosome 15. However, some chromosomes did not match predictions (e.g. recombination occurred mostly on one end of metacentric chromosomes 14 and 21). As expected, in both crosses recombination was completely suppressed in three previously described (Jones et al. 2012) marine/freshwater inversions on chromosomes 1, 11, and 21 (**Figure 4.9**).



D

Start Probabilites:

M	0.25
H	0.50
F	0.25

Transition Probabilites:

Hidden state (true genotype)	M	H	F
M	0.997	0.002	0.001
H	0.001	0.998	0.001
F	0.001	0.002	0.997

Emission Probabilities:

Hidden state (true genotype)	M	B	F	X
M	0.468	0.011	0.001	0.520
H	0.082	0.317	0.085	0.516
F	0.001	0.010	0.467	0.522

Figure 4.9 (previous page): Fine-mapping recombinant breakpoints with a Hidden Markov Model

(A) Reads were binned into small 10 kilobase (kb) bins. For each bin, the probability of each genotype was determined, and boundaries of the recombination breakpoint were set at the 99% cutoff. (B) Histogram of recombination breakpoint resolution in the FTC x LITC cross. Median resolution was 89 kb. 539 additional recombinants (not shown) had a resolution of greater than 500 kb. (C) Diagram of high-resolution (<350 kb) recombination breakpoints near three marine/freshwater inversions. Each line represents the 99% confidence interval for a recombination breakpoint in a single fish and the grey rectangles represent the previously published marine/freshwater inversion boundaries (Jones et al. 2012). Data is presented for the entire scaffold containing each inversion (scaffold 22 on chromosome 1, scaffold 11 on chromosome 11, and scaffold 16 on chromosome 21). Physical position coordinates from the original assembly are used. No recombination events are observed within these three inversion intervals in each cross. (D) Trained Hidden Markov Model parameters used for fine-mapping recombination breakpoints. SNP genotypes were binned into 10 kb bins. For each fish, each bin was assigned a raw genotype of M (only marine reads), F (only freshwater reads), B (both marine and freshwater reads), or X (no data). These raw genotypes were used as the observed data in a Hidden Markov Model with hidden states M (marine), F (freshwater), or H (heterozygous). The model was trained with the Baum-Welch algorithm on 1 million data points.

QTL mapping of lateral plate reduction

Like most freshwater populations, freshwater FTC and BEPA fish have evolved reduced lateral plates (Cresko et al. 2004; Hagen and Gilbertson 1972). Plate reduction is typically controlled by a large-effect quantitative trait locus (QTL) on chromosome 4 in both Pacific Northwest and Alaskan freshwater populations, including BEPA (Colosimo et al. 2004; Cresko et al. 2004). This QTL has been shown to be a regulatory haplotype of the *Ectodysplasin* (*Eda*) gene (Colosimo et al. 2005; O'Brown et al. 2015). Lab-reared FTC and BEPA fish were both low-plated (**Figure 4.10A**). As a positive control to validate the GBS linkage maps, we mapped lateral plate number QTL in both crosses. As expected, in both crosses a near-Mendelian QTL on chromosome 4 (percent variance explained of 97.8% and 95.7% in the FTC and BEPA crosses, respectively) controlled lateral plate number (**Figure 4.11A-B, Table 4.5**). The chromosome 4 QTL was largely recessive in each cross; all marine homozygotes and heterozygotes had more than 15 plates and all freshwater homozygotes had fewer than 15 plates (**Figure 4.11C-D**). In contrast with Berner et al. 2014, we did not find a double QTL peak on chromosome 4 in either cross (**Figure 4.11A-B**), consistent with a single underlying genetic locus.

To test the resolution of the linkage maps, we used a Hidden Markov Model on the raw allele counts for each SNP to fine-map recombination breakpoints (**Figure 4.9A**). This method enabled us to fine-map recombinant breakpoints to a median resolution of 89 kb (**Figure 4.9B**). With the fine-mapped recombination breakpoints, we identified 10 heterozygous/homozygous freshwater recombinant animals in the two crosses that recombined within a 1 Mb interval surrounding *Eda*. These recombinant animals defined a 199.8 kb interval that perfectly correlates with plate number in both crosses (**Figure 4.11C-E**). This interval contains 17 Ensembl-predicted genes, including *Eda,* as well as an intergenic SNP recently shown to affect a lateral plate enhancer (O'Brown et al. 2015). No additional QTL were detected upon conditioning on *Eda* genotype in a single model (data not shown). Because *Eda* heterozygotes had the most variance of any genotypic

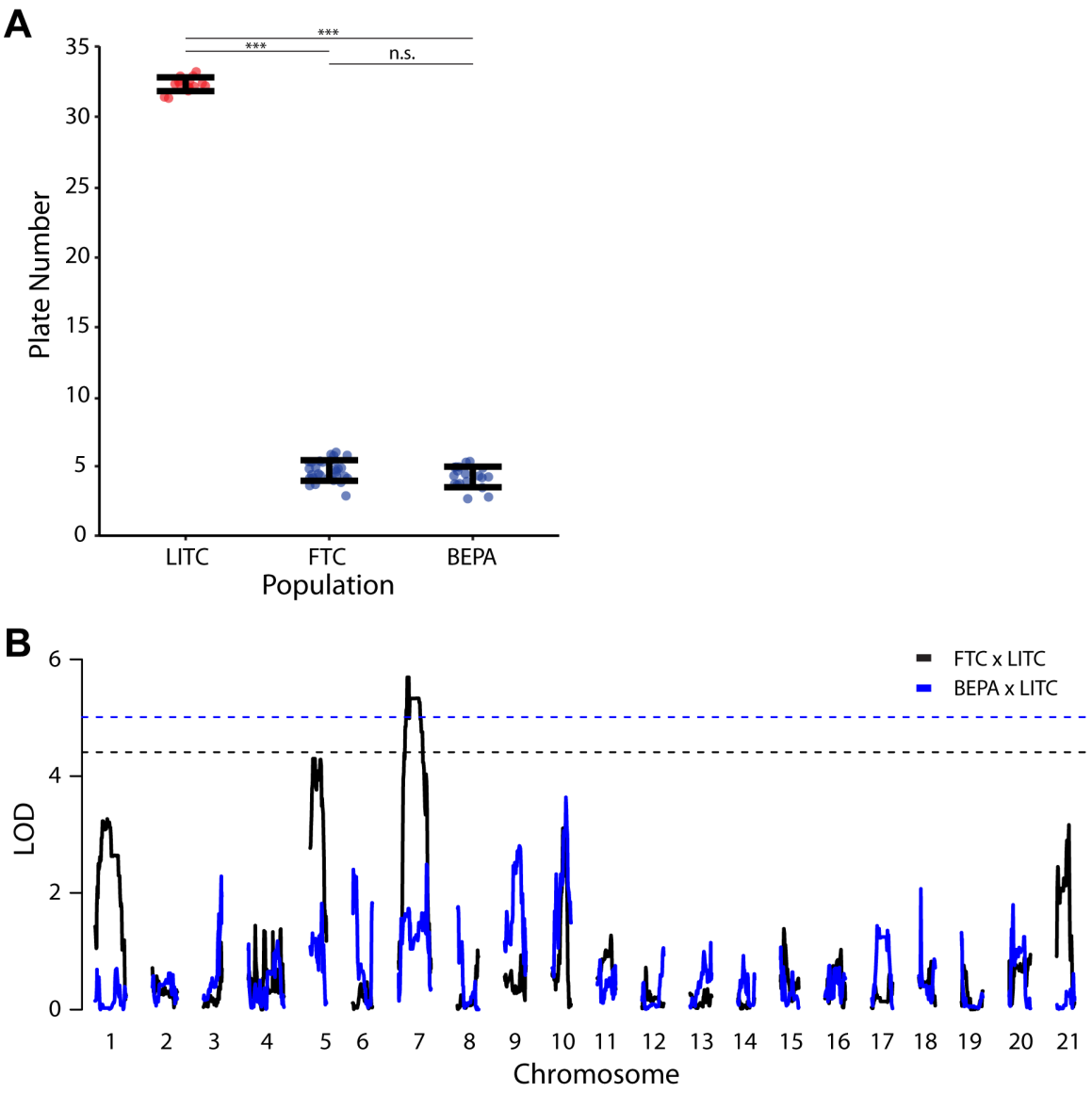


Figure 4.10: QTL mapping of lateral plate modifiers
(A) Lateral plate reduction in lab-reared FTC and BEPA freshwater fish (blue) compared to LITC marine fish (red). *** indicates $p<10^{-10}$, n.s. = not significant by Tukey's HSD test. Average of left and right side plates is shown. LITC, FTC, and BEPA fish had 32.3+/- 0.5, 4.9+/- 0.7, and 4.4+/- 0.7 plates and sample sizes of 15, 30, and 19, respectively. (B) Manhattan plot of QTL mapping of lateral plate modifiers in the FTC (black) and BEPA (blue) crosses. Mapping was performed in fish heterozygous for *Eda* (n=184/194 in the FTC/BEPA crosses). One QTL was detected in the FTC cross but no significant QTL were detected in the BEPA cross. $\alpha = 0.05$ significance levels based on 1,000 permutations are shown as dotted lines. LOD is shown as a function of adjusted physical position.

class (**Figure 4.11C-D**), we mapped plate number in *Eda* heterozygotes (as in Colosimo et al. 2004) and detected one modifier QTL on chromosome 7 in the FTC cross but no significant modifier QTL in the BEPA cross (**Figure 4.10B, Table 4.5**).

QTL mapping of gill raker length

We previously discovered a strikingly high degree of modularity of skeletal evolution in sticklebacks, consistently observed across a variety of axial and craniofacial skeletal traits (Miller et al. 2014). To test the hypothesis that gill raker length is also under modular genetic control, we examined gill rakers at three positions, located at lateral, middle, and medial points of the anterior-most ceratobranchial bone (**Figure 4.12A**). We observed a modular reduction of gill raker length in lab-reared FTC and BEPA fish, with strongest length reductions in the lateral and middle domains (**Figure 4.13**). QTL controlling gill raker length were detected on chromosomes 1, 4, 10, 11, and 20 in the FTC cross and chromosomes 16, 19, and 20 in the BEPA cross (**Figure 4.12B-C, Table 4.5**). The peak marker on chromosome 4 in the FTC cross was 17_8, a bin containing *Eda*. While one QTL (chromosome 16 in the BEPA cross) had effects on lateral, middle, and medial gill raker lengths, most QTL were surprisingly modular, with significant effects on only one gill raker length (**Table 4.5**). Most (6/8) of the gill raker length QTL were concordant with the direction of evolutionary change (freshwater allele yielding shorter gill rakers), consistent with gill raker length being under strong natural selection. QTL on chromosome 20 controlling lateral gill raker length were detected in both crosses, but had non-overlapping 1.5 LOD intervals. Overall, no QTL with overlapping 1.5 LOD intervals were observed in both crosses. Thus, unlike lateral plate reduction, the convergent evolution of gill raker length reduction has occurred via distinct genetic bases in these two freshwater populations.

DISCUSSION

High-quality linkage maps from binned GBS

The power of next generation sequencing has revolutionized high throughput genotyping, beginning in 2008 with the RAD-seq approach (Baird et al. 2008). RAD-seq was rapidly applied to a variety of model and non-model organisms (Narum et al. 2013; Rowe et al. 2011). An extension of RAD-seq, the simpler and cheaper genotyping-by-sequencing (GBS) method was published in 2011 (Elshire et al. 2011). In fish, RAD-seq or GBS has been used to build genome-wide linkage maps in stickleback (Roesti et al. 2013; this study), salmon (Gonen et al. 2014; Limborg et al. 2014), and Mexican tetra (Carlson et al. 2015), in each case successfully building a map with roughly a marker per centiMorgan. Differences in these linkage maps appear largely attributable to the details of the cross design (e.g. number of F2 fish genotyped), sequencing depth (e.g. number of lanes sequenced and whether single or paired end reads were sequenced), and/or genome assembly used to align reads (e.g. size of genome). For example, our threespine stickleback maps presented here are larger in total genetic distance than those published in Roesti et al., 2013, however we analyzed over twice as many F2 fish, sequenced paired end reads vs. single end reads, and generated more total sequence than in this earlier

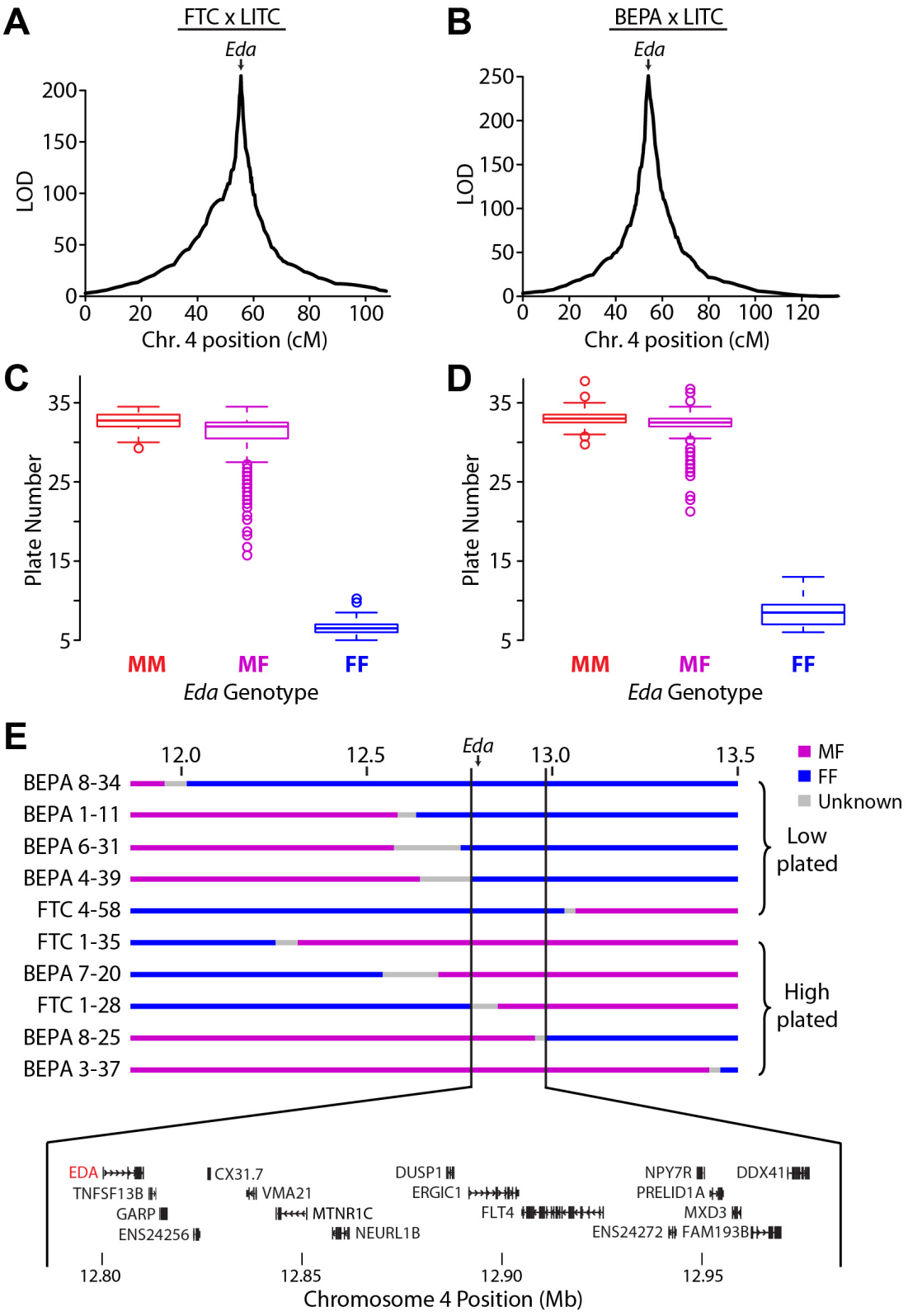

A
FTC x LITC
Eda
LOD
Chr. 4 position (cM)
B
BEPA x LITC
Eda
LOD
Chr. 4 position (cM)
C
Plate Number
MM
MF
FF
Eda Genotype
D
Plate Number
MM
MF
FF
Eda Genotype
E
Eda
MF
FF
Unknown
BEPA 8-34
BEPA 1-11
BEPA 6-31
BEPA 4-39
FTC 4-58
FTC 1-35
BEPA 7-20
FTC 1-28
BEPA 8-25
BEPA 3-37
Low plated
High plated
EDA
TNFSF13B
GARP
ENS24256
CX31.7
VMA21
MTNR1C
NEURL1B
DUSP1
ERGIC1
FLT4
NPY7R
PRELID1A
MXD3
ENS24272
DDX41
FAM193B
Chromosome 4 Position (Mb)

Figure 4.11 (previous page): Lateral plate reduction is controlled by a near-Mendelian locus containing *Eda*
(A-B) QTL mapping of plate number in the FTC (A) and BEPA (B) crosses. A main large-effect QTL was found in both crosses on chromosome 4 with a single peak at *Ectodysplasin (Eda)*. cM = centiMorgans. (C-D) Boxplots showing association between *Eda* genotype and lateral plate number in the FTC (C) and BEPA (D) crosses. M=marine, F=freshwater. The QTL is recessive, with all MF and MM fish having over 15 plates and all FF fish having fewer than 15 plates. (E) Fine-mapping the chromosome 4 lateral plate QTL with MF/FF recombinants in both crosses. Genotype at a 199.8 kb interval perfectly correlates with whether plate number is low (<15 plates per side) or high (>15 plates). This interval contains the coding regions of 17 Ensembl-predicted genes, including *Eda* and recently identified intergenic regulatory mutations of *Eda* (O'Brown et al. 2015). ENS24256 and ENS24272 refer to ENSGACT00000024256 and ENSGACT00000024272, respectively.

study. In contrast, our stickleback maps are smaller than the total genetic length of a recently published Mexican tetra map, likely in part due to Mexican tetra having twice as large a genome as stickleback. One methodological difference between our maps and these other fish linkage maps is we employed a binned approach, binning SNPs to generate genetic markers, similar to as was successfully employed in corn (Li et al. 2015; Chen et al. 2014).

This study utilized a binned Genotyping-by-Sequencing approach to generate high quality genotypes. First, a large number of SNPs (about 100,000 per cross) were sequenced to a low level of sequencing coverage (about 1.5X per sample). Then, multiple SNPs were binned together to form around 1,000 high-coverage (about 150X) markers. This approach contrasts with other reduced representation approaches (Baird et al. 2008; Elshire et al. 2011; Peterson et al. 2012; but see Andolfatto et al. 2011), which target a smaller number of SNPs (typically 1,000-5,000) at high coverage (>20X). The tradeoff of coverage vs. SNP number can easily be controlled through the choice of restriction enzymes as well as the degree of size selection of the library (Peterson et al. 2012). This study used ApeKI, which cuts a 5 bp restriction site that occurs frequently in the genome, and no library size selection to target a large number of SNPs. Individual SNPs can be biased towards one allele, have mapping or genotyping errors, and have variable coverage levels. Thus, binning a large number of low-coverage SNPs together resulted in a robust and reliable set of genotypes. The binning approach in this study also enabled the use of the same bins of markers for direct comparison between crosses. Several quality control steps appeared to be crucial to generate high-quality linkage maps, including dropping SNPs and markers that deviated from expected allele ratios, dropping low-coverage samples and markers, and using a separate computational pipeline to generate sex chromosome genotypes.

It is unlikely that the order and orientation of scaffolds presented in this study is completely correct or universal to all sticklebacks. However, there are several reasons to believe that most of the revised scaffold orders and orientations in this study are correct and typical stickleback features, rather than individual polymorphisms for genomic rearrangements. First, the vast majority of the scaffold orders and orientations were supported by multiple markers from both crosses, which derived from freshwater fish from two independently derived populations, including one fish from the same population as the original reference genome (BEPA). Second, the 31 changes to the stickleback

Trait	Cross	Chr	n	LOD	PVE	1.5 LOD interval			Trait mean +/- standard error		
						Left	Peak	Right	MM	MF	FF
Plate #	FTC	4	356	214.1	97.8	17_7	17_8	17_10	32.7 +/- 0.3	30.7 +/- 0.2	6.7 +/- 0.3
Plate #	BEPA	4	359	244.9	95.7	17_8	17_9	17_10	33 +/- 0.3	32.1 +/- 0.2	8.8 +/- 0.2
Plate # (*Eda* hets)	FTC	7	184	5.7	13.3	39_5	23_7	1_23	32 +/- 0.5	31.1 +/- 0.3	28.4 +/- 0.5
Lat. GR length	FTC	10	246	3.8	6.3	5_19	5_24	5_29	1025 +/- 13	1032 +/- 9	973 +/- 14
Lat. GR length	FTC	20	246	4.6	9.3	2_6	2_3	46_3	1039 +/- 13	1032 +/- 9	967 +/- 12
Mid. GR length	FTC	1	251	5.5	7.4	30_2	7_10	22_5	968 +/- 13	955 +/- 9	903 +/- 12
Mid. GR length	FTC	4	251	4.8	7.6	20_8	17_8	42_4	987 +/- 12	938 +/- 9	902 +/- 14
Mid. GR length	FTC	11	251	4.8	7.5	11_11	11_18	11_23	984 +/- 12	943 +/- 8	893 +/- 14
Lat. GR length	BEPA	16	298	4.5	6.3	44_4	44_3	115_1	1152 +/- 14	1130 +/- 10	1078 +/- 13
Lat. GR length	BEPA	20	298	4.6	6.5	2_31	2_9	2_7	1078 +/- 13	1139 +/- 10	1139 +/- 15
Mid. GR length	BEPA	16	300	6.2	9.1	44_4	44_3	115_2	1054 +/- 14	1031 +/- 9	955 +/- 14
Med. GR length	BEPA	16	301	4.3	6.0	44_4	44_3	115_2	700 +/- 13	676 +/- 9	620 +/- 13
Med. GR length	BEPA	19	301	4.4	6.1	3_7	3_2	3_1	627 +/- 13	661 +/- 10	706 +/- 12

Table 4.5: QTL identified in this study
Lat.=lateral, mid.=middle, med.=medial (see **Figure 4.12A**), LOD=log of the odds, PVE=percent variance explained, M=marine allele, F=freshwater allele. The intervals for the chromosome 4 plate number QTL were further defined through fine mapping of recombinant breakpoints (**Figure 4.11E**). The FTC chromosome 7 plate number QTL was identified by mapping plate number in *Eda* heterozygotes; PVE refers to residual variance in *Eda* heterozygotes, not total plate number variance. LOD significance thresholds (α = 0.05) were 3.85/3.92 for plate number and 3.73/3.90 for gill raker length in the FTC/BEPA crosses, and 4.41 for plate number in *Eda* heterozygotes in the FTC cross. See **Figures 4.10, 4.11A-B, and 4.12B-C** for lodplots of the QTL. Gill raker lengths are in microns.

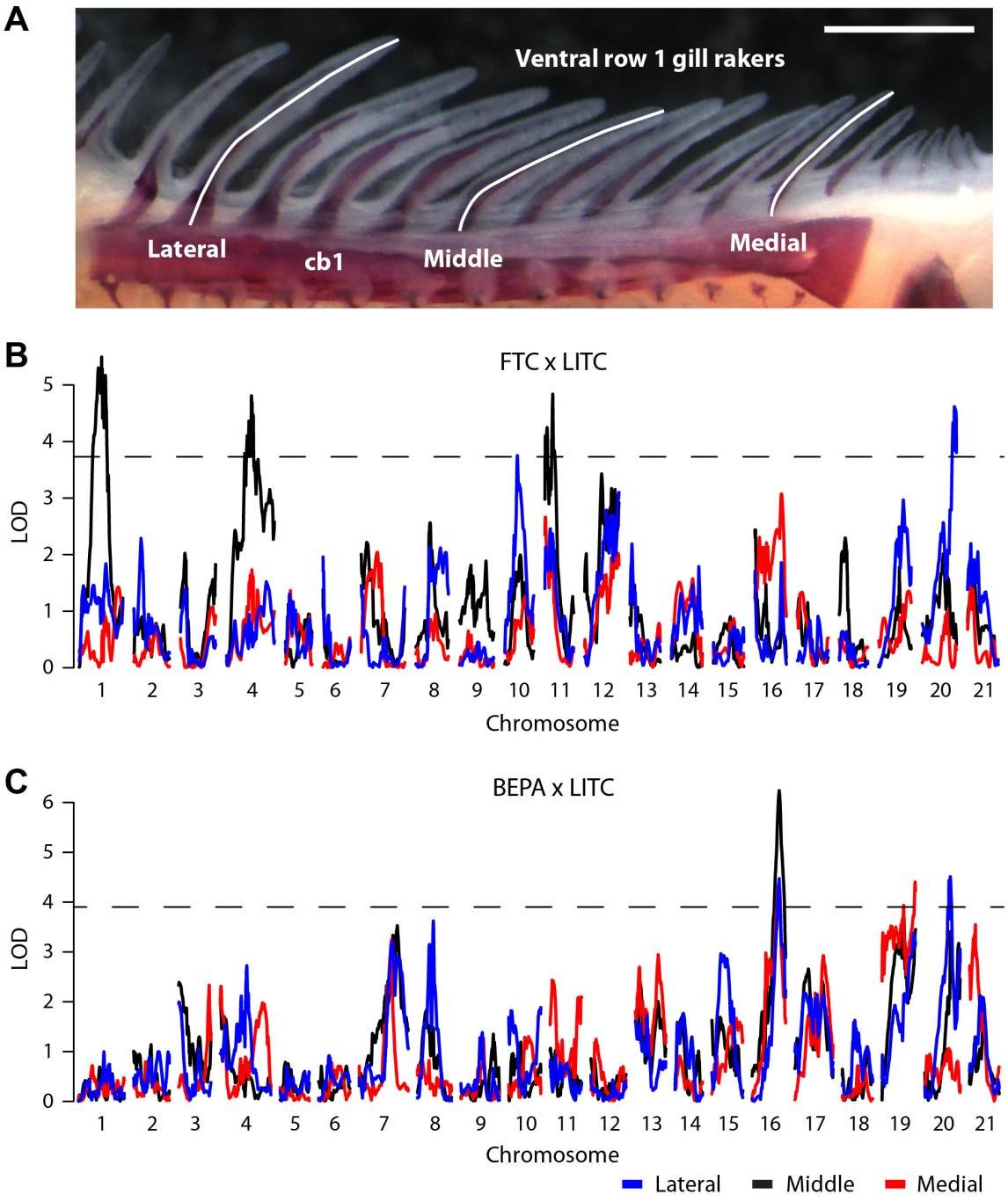


Figure 4.12: Genetic mapping of gill raker length reduction
(A) Lengths of three ventral row 1 gill rakers were measured at lateral, middle, and medial positions. cb1=ceratobranchial 1. Scale bar = 500 um. (B-C) Manhattan plots of QTL mapping of gill raker length in the FTC cross (B) and the BEPA cross (C). Three gill raker lengths were mapped: lateral (blue), middle (black), and medial (red). LOD is shown as a function of genetic position. The significance thresholds ($\alpha = 0.05$) are shown with a dotted line.

genome assembly identified in a cross of (geographically distant) European sticklebacks by Roesti et al. 2013 were all detected by this study. Third, two of the scaffold changes identified in this study (on chromosomes 19 and 20) are consistent with previous cytogenetic evidence (Urton et al. 2011). The expanded set of assembly changes identified in this study (78 newly anchored scaffolds, 40 reoriented scaffolds, and 4 scaffold rearrangements) should further aid efforts to understand the evolutionary dynamics of the stickleback genome and discover the genes underlying adaptive phenotypes in sticklebacks.

We observed little difference in the two maps whose SNPs were phased by sequencing the cross grandparents to high (60X, FTC) versus low (6X, BEPA) coverage. We suspect that sequencing grandparents with GBS, as opposed to full genome sequencing, could further reduce costs. Several cost-saving factors, including halving reagent volumes during library creation and barcoding 384 samples together, did not result in a significant decrease in genotyping quality. In library 7, we used 96 barcoded adapters and 4 index primers to multiplex 384 samples in a single lane of Illumina sequencing. The high average marker coverage (about 50X) and high genotyping success rate with 384 barcoded samples (<3% genotyping fail rate, see **Table 4.2**) suggest that more samples could be multiplexed together. For instance, 48 barcoded adapters and 16 index primers could allow barcoding of 768 samples while requiring fewer unique primers.

Genetic mapping of lateral plate reduction

QTL mapping of armor and trophic traits demonstrated the power of dense GBS-generated genome-wide linkage maps to detect QTL of large and small effect, as well as tested several hypotheses about the genetic basis of these adaptive skeletal changes. As expected from previous genetic studies of lateral plate number (Colosimo et al. 2004; Cresko et al. 2004), a large-effect QTL on chromosome 4 controlled plate number in both crosses. In contrast to a previous study which reported multiple QTL peaks on chromosome 4 for lateral plate number, perhaps suggesting multiple underlying chromosome 4 loci (Berner et al. 2014), in this study the chromosome 4 QTL in each cross had a clear single peak at *Eda*. A 17-gene minimal genomic interval included *Eda* and a recently identified intergenic lateral plate enhancer with a polymorphic SNP that affects enhancer activity (O'Brown et al. 2015). Previous mapping of plate modifier QTL in *Eda* heterozygotes identified three plate number modifier QTL in a similarly sized cross with the Paxton benthic (PAXB) population (Colosimo et al. 2004). With an identical mapping approach, we detected fewer modifier QTL (1 and 0 in the FTC and BEPA crosses, respectively). This difference in genetic architecture might be due to differences in the extent of plate reduction in the freshwater populations used in the studies (mean of 0.3 plates in PAXB (McPhail 1992) vs. 4.9 in FTC and 4.4 in BEPA). Intriguingly, the chromosome 7 modifier QTL detected in the FTC cross overlaps one of the 3 previously detected plate modifier QTL (Colosimo et al. 2004). This chromosome 7 modifier QTL might be re-used along with *Eda* in multiple freshwater populations to reduce plate number.

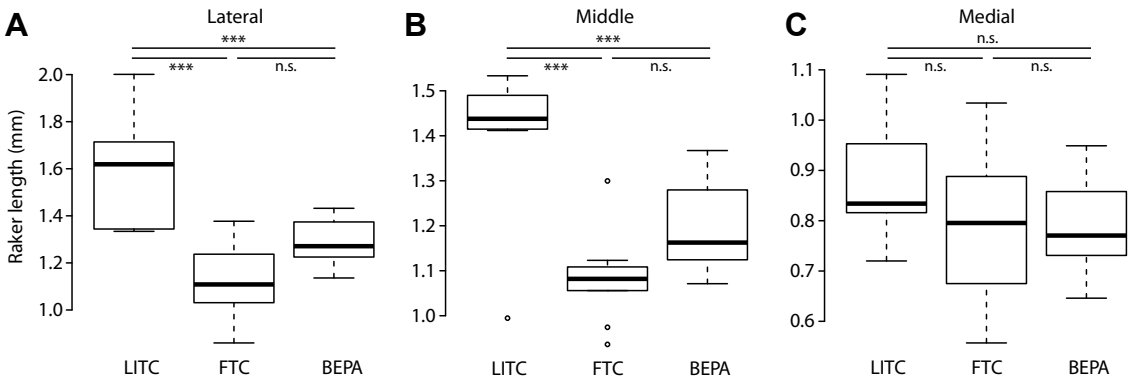


Figure 4.13: Convergent evolution of freshwater gill raker length reduction
Gill raker lengths were measured in three domains: lateral (A), middle (B), and medial (C) in marine (LITC) and freshwater (FTC and BEPA) lab-reared fish. See **Figure 4.12A** for a diagram of gill raker length measurements. Gill raker lengths were back transformed to values expected for a 50 mm standard length fish. *** indicates $p<0.001$, n.s. = not significant by Tukey's HSD test. n=10 per population.

Genetic mapping of gill raker length reduction

We also identified eight new QTL controlling the classic adaptive trait of gill raker length (Schluter 2000). Motivated by our previous finding of pervasive modularity in the evolution of serially homologous axial and craniofacial skeletal elements (Miller et al. 2014), we hypothesized that gill raker length might also be genetically controlled in a modular fashion. Indeed, gill raker lengths measured at different mediolateral locations had surprisingly different genetic architectures, indicating complex modularity of this trait. Therefore, gill raker lengths at different positions might not be directly comparable in ecological studies. This modularity might reflect differences in retaining different types of prey with gill rakers of different lengths along the mediolateral axis. Developmental timing might contribute to this genetic modularity, as gill rakers form during embryonic development in a wave from lateral to medial (Glazer et al. 2014). Intriguingly, the chromosome 4 gill raker length QTL in the FTC cross has a peak marker bin that contains *Eda*. The EDA pathway, in addition to its role in plate development, is intimately involved in gill raker development in zebrafish and sticklebacks (Glazer et al. 2014; Harris et al. 2008). However, in the BEPA cross, a chromosome 4 plate number QTL, but not a gill raker length QTL, was detected. Therefore, if *Eda* is contributing to the FTC gill raker length QTL, there is likely different linked regulatory variation of *Eda* in FTC compared to BEPA.

We previously identified an enrichment of skeletal QTL on chromosomes 4, 20 and 21 (Miller et al. 2014), and our findings here add gill raker length as yet another skeletal trait controlled by two of these three trait clusters, as gill raker length mapped to chromosome 20 in both crosses, and chromosome 4 in the FTC cross. Linked chromosome 4 and 20 alleles promoting reduction of gill raker length (this study) and gill raker number (Glazer et al. 2014; Miller et al. 2014) might promote coevolution of these phenotypes in freshwater environments. In sticklebacks, a predictable, shared genetic

basis has been found to underlie the convergent evolution of several evolved phenotypes (Chan et al. 2010; Colosimo et al. 2005; Colosimo et al. 2004; Glazer et al. 2014; Miller et al. 2007a). In contrast to these studies, we detected no overlapping gill raker length QTL in the two crosses. In addition, none of the gill raker length QTL in this study overlap two previously reported QTL from a European lake x stream cross (Berner et al. 2014). Thus, unlike several other stickleback phenotypes, different loci appear to underlie the convergent evolution of gill raker length in different populations.

CONCLUSIONS

This work used a binned Genotyping-by-Sequencing approach to build dense linkage maps of sticklebacks, which were used for genome assembly improvement and QTL mapping of two ecologically important traits. The revised genome assembly provides a more accurate understanding of the structure of the stickleback genome, which should aid efforts to map genes controlling stickleback phenotypes and understand genomic dynamics during stickleback evolution. The genetic mapping of distinct QTL controlling gill raker length in two crosses illustrates that, in contrast to several prominent cases in sticklebacks (Chan et al. 2010; Colosimo et al. 2005; Miller et al. 2007b), a non-parallel genetic basis is sometimes used in cases of repeated phenotypic evolution.

ACKNOWLEDGEMENTS

Dan Rokhsar for initially conceived of the idea to use GBS genotypes to map stickleback QTL and provided invaluable guidance with the analysis. Emily Killingbeck tirelessly made many GBS libraries, generating most the raw data analyzed in this study, and Therese Mitros provided computational assistance. Craig Miller provided guidance in designing experiments, analyzing data, and writing the manuscript.

In addition to the coauthors listed above, I thank David Kingsley for his input and support during the beginning of this study, Mike Bell for Bear Paw lake fish, Alex Pollen for assistance collecting Fishtrap Creek and Little Campbell River fish, the Semiahmoo First Nation for permission to collect sticklebacks from the Little Campbell River, Patrick Lee and Anthony Lee for expert assistance in crossing and raising fish, Angela Lam, Priscilla Erickson, and Nihar Patel for DNA isolation and phenotyping assistance, Blair Gainous and Allison Quan for performing pilot experiments, and Jessica Lyons, Megan Hall, and Chris Martin for helpful discussions. This work was supported in part by the NIH (R01-DE021475 to CTM and a Predoctoral Training Grant 5T32GM007127 to AMG) and the NSF Graduate Research Fellowship Program (AMG).

REFERENCES

Andolfatto, P., D. Davison, D. Erezyilmaz, T.T. Hu, J. Mast *et al.*, 2011 Multiplexed shotgun genotyping for rapid and efficient genetic mapping. Genome Res. 21: 610-617.

Arnegard, M.E., M.D. McGee, B. Matthews, K.B. Marchinko, G.L. Conte *et al.*, 2014 Genetics of ecological divergence during speciation. Nature 511: 307-311.

Baird, N.A., P.D. Etter, T.S. Atwood, M.C. Currey, A.L. Shiver *et al.*, 2008 Rapid SNP discovery and genetic mapping using sequenced RAD markers. PLoS One 3: e3376.

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

Berner, D., D. Moser, M. Roesti, H. Buescher, and W. Salzburger, 2014 Genetic architecture of skeletal evolution in European lake and stream stickleback. Evolution 68: 1792-1805.

Berner, D., M. Roesti, A.P. Hendry, and W. Salzburger, 2010a Constraints on speciation suggested by comparing lake-stream stickleback divergence across two continents. Mol. Ecol. 19: 4963-4978.

Berner, D., W.E. Stutz, and D.I. Bolnick, 2010b Foraging trait (co)variances in stickleback evolve deterministically and do not predict trajectories of adaptive diversification. Evolution 64: 2265-2277.

Broman, K.W., and S. Sen, 2009 *A Guide to QTL Mapping with R/qtl*. Dordrecht: Springer.

Broman, K.W., H. Wu, S. Sen, and G.A. Churchill, 2003 R/qtl: QTL mapping in experimental crosses. Bioinformatics 19: 889-890.

Carlson, B.M., S.W. Onusko, and J.B. Gross, 2015 A High-Density Linkage Map for Astyanax mexicanus Using Genotyping-by-Sequencing Technology. G3-Genes Genomes Genetics 5: 241-251.

Chan, Y.F., M.E. Marks, F.C. Jones, G. Villarreal, Jr., M.D. Shapiro *et al.*, 2010 Adaptive evolution of pelvic reduction in sticklebacks by recurrent deletion of a *Pitx1* enhancer. Science 327: 302-305.

Chen, Z.L., B.B. Wang, X.M. Dong, H. Liu, L.H. Ren *et al.*, 2014 An ultra-high density bin-map for rapid QTL mapping for tassel and ear architecture in a large F-2 maize population. BMC Genomics 15.

Colosimo, P.F., K.E. Hosemann, S. Balabhadra, G. Villarreal, Jr., M. Dickson *et al.*, 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 *et al.*, 2004 The genetic architecture of parallel armor plate reduction in threespine sticklebacks. PLoS Biol. 2: 635-641.

Cresko, W.A., A. Amores, C. Wilson, J. Murphy, M. Currey *et al.*, 2004 Parallel genetic basis for repeated evolution of armor loss in Alaskan threespine stickleback populations. Proc. Natl. Acad. Sci. USA 101: 6050-6055.

Day, T., J. Pritchard, and D. Schluter, 1994 A comparison of two sticklebacks. Evolution 48: 1723-1734.

Elshire, R.J., J.C. Glaubitz, Q. Sun, J.A. Poland, K. Kawamoto *et al.*, 2011 A robust, simple genotyping-by-sequencing (GBS) approach for high diversity species. PLoS One 6: e19379.

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.

Gonen, S., N.R. Lowe, T. Cezard, K. Gharbi, S.C. Bishop *et al.*, 2014 Linkage maps of the Atlantic salmon (Salmo salar) genome derived from RAD sequencing. BMC Genomics 15.

Hagen, D.W., 1973 Inheritance of numbers of lateral plates and gill rakers in *Gasterosteus aculeatus*. Heredity (Edinb) 30: 303-312.

Hagen, D.W., and L.G. Gilbertson, 1972 Geographic variation and environmental selection in *Gasterosteus aculeatus* L. in the Pacific Northwest, America. Evolution 26: 32-51.

Harris, M.P., N. Rohner, H. Schwarz, S. Perathoner, P. Konstantinidis *et al.*, 2008 Zebrafish *eda* and *edar* mutants reveal conserved and ancestral roles of ectodysplasin signaling in vertebrates. PLoS Genet. 4: e1000206.

Hatfield, T., 1997 Genetic divergence in adaptive characters between sympatric species of stickleback. Am. Nat. 149: 1009-1029.

Hohenlohe, P.A., S. Bassham, P.D. Etter, N. Stiffler, E.A. Johnson *et al.*, 2010 Population genomics of parallel adaptation in threespine stickleback using sequenced RAD tags. PLoS Genet. 6: e1000862.

Jones, F.C., M.G. Grabherr, Y.F. Chan, P. Russell, E. Mauceli *et al.*, 2012 The genomic basis of adaptive evolution in threespine sticklebacks. Nature 484: 55-61.

Li, C.H., Y.X. Li, Y.S. Shi, Y.C. Song, D.F. Zhang *et al.*, 2015 Genetic Control of the Leaf Angle and Leaf Orientation Value as Revealed by Ultra-High Density Maps in Three Connected Maize Populations. PLoS One 10.

Limborg, M.T., R.K. Waples, J.E. Seeb, and L.W. Seeb, 2014 Temporally Isolated Lineages of Pink Salmon Reveal Unique Signatures of Selection on Distinct Pools of Standing Genetic Variation. Journal of Heredity 105: 741-751.

Liu, J., T. Shikano, T. Leinonen, J.M. Cano, M.H. Li *et al.*, 2014 Identification of major and minor QTL for ecologically important morphological traits in three-spined sticklebacks (*Gasterosteus aculeatus*). G3 (Bethesda) 4: 595-604.

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.

Miller, C.T., S. Beleza, A.A. Pollen, D. Schluter, R.A. Kittles *et al.*, 2007a *cis*-Regulatory changes in *Kit ligand* expression and parallel evolution of pigmentation in sticklebacks and humans. Cell 131: 1179-1189.

Miller, C.T., S. Beleza, A.A. Pollen, D. Schluter, R.A. Kittles *et al.*, 2007b 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 *et al.*, 2014 Modular skeletal evolution in sticklebacks Is controlled by additive and clustered Quantitative Trait Loci. Genetics 197: 405-420.

Narum, S.R., C.A. Buerkle, J.W. Davey, M.R. Miller, and P.A. Hohenlohe, 2013 Genotyping-by-sequencing in ecological and conservation genomics. Mol. Ecol. 22: 2841-2847.
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 4: e05290.
Peichel, C.L., J.A. Ross, C.K. Matson, M. Dickson, J. Grimwood *et al.*, 2004 The master sex-determination locus in threespine sticklebacks is on a nascent Y chromosome. Current Biology 14: 1416-1424.
Peterson, B.K., J.N. Weber, E.H. Kay, H.S. Fisher, and H.E. Hoekstra, 2012 Double digest RADseq: an inexpensive method for *de novo* SNP discovery and genotyping in model and non-model species. PLoS One 7: e37135.
Reimchen, T.E., 1992 Injuries on stickleback from attacks by a toothed predator (Oncorhynchus) and implications for the evolution of lateral plates. Evolution 46: 1224-1230.
Roesti, M., D. Moser, and D. Berner, 2013 Recombination in the threespine stickleback genome--patterns and consequences. Mol. Ecol. 22: 3014-3027.
Rowe, H.C., S. Renaut, and A. Guggisberg, 2011 RAD in the realm of next-generation sequencing technologies. Mol. Ecol. 20: 3499-3502.
Schluter, D., 2000 *The Ecology of Adaptive Radiation*. Oxford: Oxford University Press.
Schluter, D., and J.D. McPhail, 1992 Ecological character displacement and speciation in sticklebacks. Am. Nat. 140: 85-108.
Schneider, C.A., W.S. Rasband, and K.W. Eliceiri, 2012 NIH Image to ImageJ: 25 years of image analysis. Nat. Methods 9: 671-675.
Stern, D.L., and V. Orgogozo, 2009 Is genetic evolution predictable? Science 323: 746-751.
Theis, A., F. Ronco, A. Indermaur, W. Salzburger, and B. Egger, 2014 Adaptive divergence between lake and stream populations of an East African cichlid fish. Mol. Ecol. 23: 5304-5322.
Urton, J.R., S.R. McCann, and C.L. Peichel, 2011 Karyotype differentiation between two stickleback species (Gasterosteidae). Cytogenet. Genome Res. 135: 150-159.
Wagner, G.P., M. Pavlicev, and J.M. Cheverud, 2007 The road to modularity. Nat. Rev. Genet. 8: 921-931.
Wark, A.R., M.G. Mills, L.H. Dang, Y.F. Chan, F.C. Jones *et al.*, 2012 Genetic architecture of variation in the lateral line sensory system of threespine sticklebacks. G3 (Bethesda) 2: 1047-1056.

Chapter 5:
Induced mutations in *Fgf20* phenocopy evolved changes in gill raker spacing

ABSTRACT

What genes underlie evolved changes in nature, and whether convergently evolved traits have a predictably similar genetic basis remain two fundamental questions in biology. In order to adapt to novel diets, hundreds of populations of independently derived freshwater stickleback fish (*Gasterosteus aculeatus*) have convergently evolved a quantitative reduction in gill raker number. Previous work determined that gill raker reduction has a highly polygenic basis, with over 20 Quantitative Trait Loci (QTL) contributing to the trait. However, the extent of parallel genetic architecture underlying the convergent evolution of this trait was largely unknown. Using genome-wide genotypes generated from Genotyping-by-Sequencing, gill raker number and spacing were mapped in two large marine by freshwater F2 crosses of over 350 fish each. Overlapping QTL on chromosomes 4, 16, and 20 were found in both crosses, as well as 14 additional unique QTL, suggesting a combination of parallel and non-parallel genetic changes underlying the convergent evolution of raker reduction. Using genetic crosses with recombinant chromosomes, the chromosome 4 and 20 QTL from these two freshwater populations were fine-mapped, substantially narrowing the interval sizes. Surprisingly, the resultant fine-mapped chromosome 4 QTL intervals from the two freshwater populations mapped to non-overlapping regions, indicating a non-parallel genetic basis. An excellent candidate gene, *Fibroblast Growth Factor 20 (Fgf20)* was located within one of the fine-mapped QTL. Induced loss of function mutations in *Fgf20* resulted in a decrease in gill raker number through an increase in raker spacing, similar to the evolved phenotype. Induced mutations in a second gene, *Smad5*, also resulted in a raker phenotype; however *Smad5* was excluded from the fine-mapped QTL intervals. No coding changes in *Fgf20* likely cause the phenotype, suggesting a model in which *cis*-regulatory mutations in *Fgf20* contribute to evolved gill raker reduction.

INTRODUCTION

Understanding the genetic basis of evolutionary changes is a longstanding goal in biology. For example, when the same phenotype evolves convergently in independent lineages, is the same genetic basis used? The threespine stickleback (*Gasterosteus aculeatus*) is an excellent model system to interrogate the genetic basis of evolutionary changes. Sticklebacks have undergone a dramatic adaptive radiation, in which ancestral marine sticklebacks have colonized and adapted to freshwater environments throughout the Northern Hemisphere (Bell and Foster 1994). One of the most characteristic freshwater adaptations is the evolution of changes to gill rakers, periodic epithelial appendages in a fish's throat that determine what size prey fish can efficiently eat. In response to a decreased prevalence of planktonic prey, freshwater sticklebacks have repeatedly evolved fewer, shorter gill rakers (Gross and Anderson 1984). The reduction in gill raker number is due to a concomitant increase in the spacing between gill rakers (Glazer et al. 2014). These gill raker changes are largely heritable and not due to phenotypic plasticity (Day et al. 1994; Hagen 1973; Schluter 1996).

Previously, we used genome-wide linkage mapping to map 23 Quantitative Trait Loci (QTL) controlling gill raker number and spacing differences in a Paxton Benthic freshwater (PAXB) x Japanese Marine (JAMA) F2 cross, with the two largest effect QTL mapping to chromosomes 4 and 20 (23 and 25 percent variance explained, respectively, Miller et al. 2014). We subsequently identified chromosome 4 and 20 QTL in crosses using three independently derived freshwater populations (PAXB, Fish Trap Creek—FTC, and Bear Paw Lake—BEPA), each crossed to the Little Campbell (LITC) marine population (Glazer et al. 2014). For both chromosome 4 and 20, the 1.5 LOD intervals of the QTL overlapped in all three crosses, and largely mapped to the same regions, with the exception of the BEPA chromosome 4 QTL, which overlapped the FTC and PAXB QTL to a lesser extent. In addition to genomic position, the chromosome 4 and 20 QTL in the different crosses shared many features, including additivity, modularity, and an effect on both gill raker number and spacing, suggesting that the same underlying gene and/or mutation(s) might be used. The intervals were broad and contained several candidate genes, including *Fgf20, Fgf4,* and *Smad5* on chromosome 4 and *Hey1* and *Gsk3a* on chromosome 20.

Recent developments in using high-throughput sequencing allow economical high-throughput genotyping of hundreds of samples simultaneously (Baird et al. 2008; Rowe et al. 2011; Narum et al. 2013). Recently we developed a Genotyping-by-Sequencing approach to make genome-wide genetic linkage maps in the FTC x LITC and BEPA x LITC crosses (Glazer et al. 2015, in press). Using these maps, we demonstrated that the convergent evolution of shorter gill rakers in the FTC and BEPA freshwater populations had a non-parallel genetic basis, with different QTL present in the two crosses. Although previous studies have mapped gill raker number in F2 genetic crosses using one gill raker reduced freshwater population (Miller et al. 2014; Arnegard et al. 2014; Peichel et al. 2001), the genome-wide architecture of gill raker number in multiple freshwater populations has not been compared to determine whether a parallel genetic basis underlies the convergent evolution of this trait.

Gill rakers are epithelial appendages, a class of organs that form from placodes, transient embryonic thickened epithelia that bud inwards or outwards to form an epithelial appendage (reviewed in Chuong 1998; Chuong et al. 2013). The Fibroblast Growth Factor (FGF) pathway is intimately involved in patterning different epithelial appendages in diverse vertebrates, and perturbations of FGF signaling affect epithelial appendage spacing (Jung et al. 1998; Pispa and Thesleff 2003; Thesleff 2003; Wells et al. 2012). In carp, mutants for an FGF receptor *Fgfr1* have fewer scales and quantitative reductions in gill raker number (Golovinskaya 1940; Rohner et al. 2009). The extent to which FGF pathway genes affect gill raker patterning and evolution in sticklebacks is unknown. Understanding the developmental and genetic mechanisms underlying stickleback gill raker evolution might further shed light on general principles of epithelial appendage evolution.

Genome editing technologies such as Clustered Regularly Interspaced Short Palindromic Repeats (CRISPR) and Transcription Activator-Like Effector Nucleases (TALENs) allow the ability to make targeted double strand breaks to make directed mutations in a gene of interest (Gaj et al. 2013). These double strand breaks, when repaired by the Non-Homologous End Joining pathway, can result in small insertions or deletions, which can be used to generate nonsense mutations in a gene of interest. These methods work especially efficiently in fish, where external fertilization allows collection of one-celled zygotes that can be injected with these genome editing reagents (e.g. Bedell et al. 2012; Erickson et al. 2015). Thus, reverse genetic approaches that directly test the function of candidate genes are now possible in sticklebacks.

Here we perform genome-wide linkage mapping of gill raker number and spacing in two marine x freshwater F2 genetic crosses (the FTC x LITC and BEPA x LITC crosses) to test whether the genome-wide architecture of convergent evolution of gill raker number and spacing has evolved through parallel genetic changes. We then use recombinant chromosomes to fine-map the chromosome 4 and 20 raker QTL in both crosses, narrowing the intervals substantially. Finally, we test whether TALEN-induced mutations in an outstanding candidate gene, *Fgf20*, phenocopy the evolved changes in gill raker patterning.

RESULTS

To test whether parallel genetic changes underlie the convergent evolution of gill raker patterning, gill raker number and spacing were measured in over 350 F2 fish from each of two large marine x freshwater crosses (the FTC x LITC and BEPA x LITC crosses). In both crosses, Genotyping-By-Sequencing (GBS) was used to efficiently identify genome-wide genotypes (Glazer et al. 2015, in press) for quantitative trait loci (QTL) mapping. Previously we identified 23 gill raker number and spacing QTL, most of which had highly modular effects, and only affected a subset of possible gill raker domains (Miller et al. 2014). Therefore, raker number and spacing were measured for all nine rows of ventral gill rakers, as well as dorsal row 1 (**Figure 5.1A**). Using genome-wide linkage mapping, we mapped QTL in both the FTC and BEPA crosses and filtered the QTL for non-overapping QTL by the largest effect QTL by LOD score, as described (Miller et al. 2014). We identified 9 QTL in the BEPA cross and 11 QTL in the FTC cross

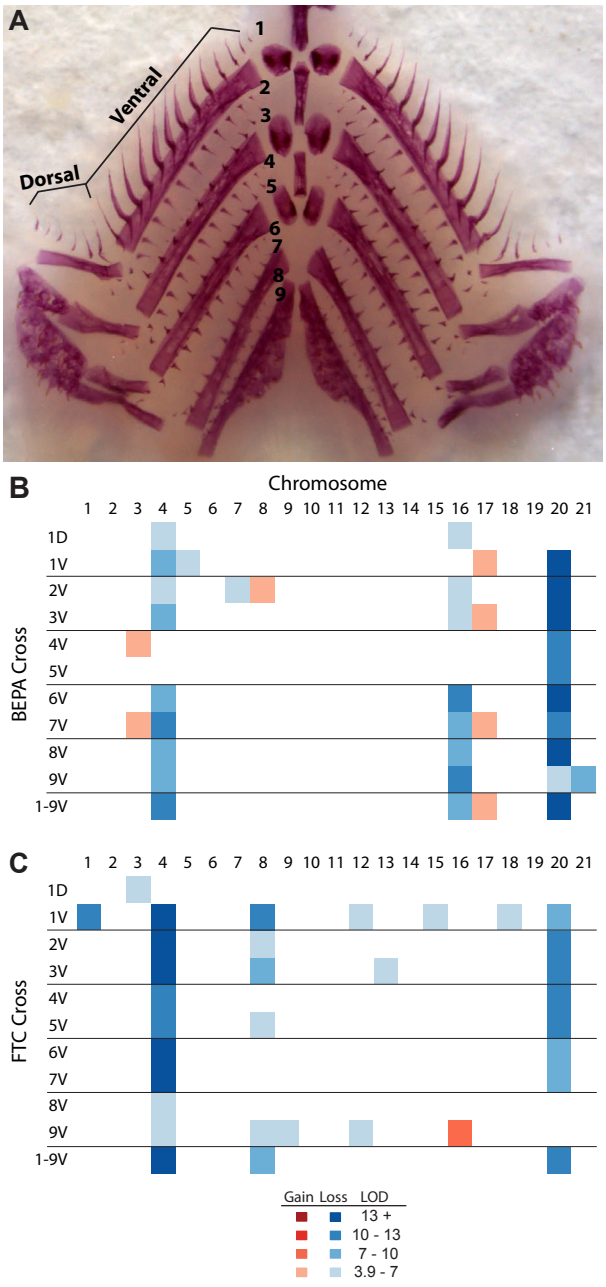


Figure 5.1: Genome-wide mapping of gill raker reduction
(A) Stickleback gill raker domains. The flat-mounted branchial prep is from an Alizarin red-stained FTCxLITC F_2. Gill raker rows 1-9 are numbered, and ventral (rows 1-9) and dorsal (row 1) gill rakers are indicated. (B-C) Heatmap of gill raker number and spacing QTL in the BEPA (B) and FTC (C) crosses. D is dorsal, V is ventral, LOD is logarithm of the odds, a measure of statistical significance. If both raker number and spacing QTL were present for a given domain, the higher LOD QTL was shown. "Gain" (red) indicates that the freshwater allele of the QTL promotes more gill rakers or denser raker spacing and "loss" (blue) indicates that the freshwater allele of the QTL promotes fewer gill rakers or wider raker spacing (the direction of evolutionary change). Additional QTL statistics are presented in **Table 5.1**.

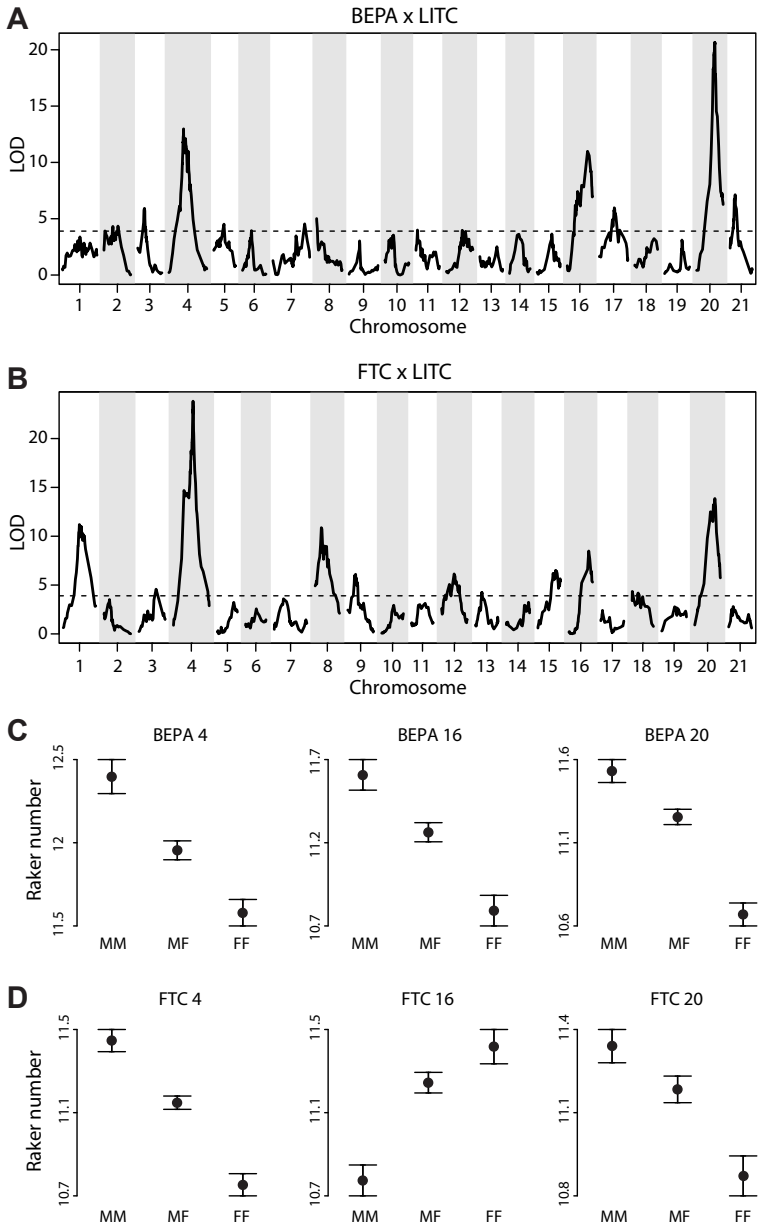


Figure 5.2: Three overlapping and 14 non-overlapping gill raker QTL
(A-B) Manhattan plot of gill raker number and spacing in the BEPA (A) or FTC (B) crosses. LOD is logarithm of the odds. For each chromosome, LOD scores for the phenotype with the maximum peak LOD score is plotted. Additional QTL statistics are presented in **Table 5.1**. 9 and 11 significant QTL map in the BEPA and FTC crosses, respectively, including overlapping QTL on chromosome 4, 16 and 20. (C-D) Phenotype by genotypic class for the chromosome 4, 16, and 20 QTL in the BEPA (C) and FTC (D) crosses. M is the marine allele, and F is the freshwater allele. All QTL are "loss" QTL, where the freshwater allele promotes fewer gill rakers, except for the FTC chr. 16 QTL, where the freshwater allele promotes more gill rakers. For each QTL, the gill raker number phenotype with the highest LOD score is shown (see **Table 5.1**).

(**Figure 5.1B-C, Figure 5.2A-B, Table 5.1**). In both crosses, the largest effect QTL mapped to chromosomes 4 and 20, confirming the microsatellite marker-based QTL previously identified in these crosses (Glazer et al. 2014). Each cross also contained a QTL on chromosome 16 with adjacent peak markers (44_3 in the BEPA cross and 44_2 in the FTC cross, **Table 5.1**). However, the direction of effect of the chromosome 16 QTLs differed between the two crosses. In the BEPA cross, freshwater alleles at the chr. 16 QTL promoted fewer rakers whereas in the FTC cross, freshwater alleles promoted more rakers (**Figure 5.2C-D**). The remaining 14 modifier QTL mapped to different chromosomes in the two crosses or had non-overlapping 1.5 LOD intervals, revealing significant differences in the quantitative genetic architecture of gill raker reduction in these two populations (**Figure 5.2A-B**). Three QTL in the FTC cross (on chromosomes 1, 15, and 18) in the FTC cross had specific effects on row 1 gill raker number (**Figure 5.1C**), consistent with an enhanced reduction of gill raker number in row 1 in FTC (Glazer et al. 2014). Overall, 6 of 9 QTL in the BEPA cross and 10 of 11 QTL in the FTC cross were in the "loss" direction, with derived freshwater alleles causing fewer or more widely spaced gill rakers (**Table 5.1**). The enrichment (16 of 20) of total QTL in the "loss" direction of evolutionary change was significantly different than would be expected by chance (p=0.012, two-tailed binomial test).

The two largest effect QTL in each cross mapped to chromosomes 4 and 20. Using a small number of microsatellites, we previously mapped the chromosome 4 QTL to 17.9 Mb and 17.7 Mb intervals in the BEPA and FTC crosses, respectively (Glazer et al. 2014, Figure 5.3). Although the 1.5 LOD intervals in the two crosses overlapped, the peak markers were different and the region of overlap was small, suggesting that the underlying genes might be different. With the much denser GBS-based genotyping of these QTL, we further narrowed the QTL 1.5 LOD intervals to 6.3 Mb and 12.0 Mb (**Figure 5.3**). These intervals still overlapped, but to an even lesser extent (1.5 Mb of overlap). Several candidate genes with known roles in developmental pathways were located within the BEPA interval (including *Eda*, *Msx2*, *Fgf20*, and *Smad5*) and the FTC interval (including *Fgf20*, *Smad5*, *Fgf4*, and *Wnt7b*). To further resolve the genomic locations of these QTL, we performed a series of recombinant mapping experiments using inbred lines generated from the BEPA and FTC F2 crosses. In subsequent generations, fish were identified that had chromosomes that were recombinant within the QTL intervals. These fish were bred to generate recombinant test crosses, which asked whether each recombinant chromosome behaved like a marine or freshwater chromosome, and therefore whether the QTL localized to the left or the right of the recombinant breakpoint. Three recombinant cross schemes were used (**Figure 5.5**). For each cross, we performed two statistical tests to distinguish between the recombinant=marine and recombinant=freshwater models, a test using the Akaike Information Criteria (AIC), and a likelihood ratio test. For each recombinant cross, the mapping result was statistically supported by both methods (**Tables 5.2, 5.3**). We analyzed five chromosome 4 recombinants in each cross, which narrowed the QTL intervals to a 0.69 Mb and 3.1 Mb interval in the BEPA and FTC crosses, respectively. These newly-resolved intervals did not overlap, indicating a separate genetic basis of the chromosome 4 QTL in the two crosses. The fine-mapping excluded the coding regions of most of the chromosome 4

Cross	Phenotype	Chr.	LOD	PVE	Peak		1.5 LOD Interval		Phenotype Mean (Standard Error)			Direction	Dominance/Effect		
					Marker	Position	Marker	Position	MM	MF	FF		d	a	d/a
BEPA	Lr7Avgsp	3	4.9	4.9	15_7	26.4	33_3 - 15_19	7.7-34.9	227.1 (1.8)	225.0 (1.3)	218.8 (1.8)	gain	2.04	-4.16	-0.49
BEPA	Ar7V	4	13	12.8	17_6	52.4	17_1 - 17_10	49.5-54.8	12.39 (0.09)	12.00 (0.05)	11.67 (0.07)	loss	-0.03	-0.359	0.08
BEPA	Ar1V	5	4.5	4.4	38_4	36.4	19_3 - 38_5	22.5-40.4	12.15 (0.08)	11.93 (0.06)	11.73 (0.08)	loss	-0.009	-0.208	0.05
BEPA	Lr2Avgsp	7	4.5	5.2	1_33	113.2	1_32 - 113_2	101.9-129.9	240.2 (2.3)	247.5 (1.4)	255.1 (2.2)	loss	-0.17	7.47	-0.02
BEPA	Ar2V	8	5	4.6	56_3	0	56_3 - 53_2	0-4.2	11.34 (0.10)	11.74 (0.07)	11.85 (0.09)	gain	0.145	0.254	0.57
BEPA	Ar6V	16	11.2	9.9	44_3	72.4	44_4 - 115_2	58.7-84	11.53 (0.08)	11.23 (0.05)	10.82 (0.08)	loss	0.05	-0.353	-0.14
BEPA	Lr1Avgsp	17	6	7	18_3	53.8	25_8 - 18_10	48.6-62.6	269.2 (2.1)	263.4 (1.5)	253.9 (1.9)	gain	1.84	-7.65	-0.24
BEPA	Ar1.9V	20	20.7	19.3	2_15	65.2	2_29 - 2_13	60.3-66.1	11.44 (0.06)	11.20 (0.04)	10.69 (0.06)	loss	0.135	-0.378	-0.36
BEPA	Ar9V	21	7.1	6.9	16_12	18.1	16_5 - 16_19	13.6-28.3	11.38 (0.09)	11.19 (0.06)	10.72 (0.10)	loss	0.138	-0.33	-0.42
FTC	Ar1V	1	11.2	8.7	7_2	48.4	30_8 - 22_7	45-58.4	11.28 (0.09)	10.89 (0.06)	10.44 (0.08)	loss	0.033	-0.421	-0.08
FTC	Lr1EAvgsp	3	4.6	6.8	50_2	51.6	81_1 - 50_4	42.5-72.9	211.8 (3.1)	217.2 (2.0)	231.3 (3.2)	loss	-4.29	9.75	-0.44
FTC	Ar1.9V	4	23.8	23.1	24_12	58.8	17_14 - 28_3	57.7-60	11.43 (0.05)	11.15 (0.03)	10.78 (0.05)	loss	0.044	-0.322	-0.14
FTC	Ar1V	8	10.9	8.4	36_5	18.6	36_4 - 36_7	16.3-22.3	11.37 (0.09)	10.81 (0.06)	10.51 (0.08)	loss	-0.127	-0.431	0.29
FTC	Ar9V	9	6.1	6.2	8_21	23.4	31_5 - 8_8	16.5-31.5	11.42 (0.1)	11.31 (0.07)	10.85 (0.09)	loss	0.177	-0.282	-0.63
FTC	Lr1Avgsp	12	4.4	3.7	68_1	37.7	99_1 - 6_26	29.6-49.4	258.4 (2.0)	262.5 (1.5)	269.5 (2.3)	loss	-1.43	5.55	-0.26
FTC	Ar3V	13	4.2	3.5	4_21	19.7	4_11 - 4_31	16.1-32.6	10.58 (0.06)	10.61 (0.05)	10.21 (0.08)	loss	0.22	-0.183	-1.20
FTC	Lr1Avgsp	15	6.4	5.5	45_2	66.8	10_23 - 84_2	61.8-68.4	256.7 (2.2)	262.7 (1.5)	269.6 (2.2)	loss	-0.44	6.43	-0.07
FTC	Ar9V	16	8.5	8.8	44_2	59.4	44_3 - 115_1	56-66.3	10.75 (0.09)	11.32 (0.06)	11.53 (0.1)	gain	0.185	0.394	0.47
FTC	Ar1V	18	4.1	3	32_2	31.7	29_7 - 32_4	15.2-38.9	11.11 (0.09)	10.85 (0.06)	10.67 (0.09)	loss	-0.044	-0.216	0.20
FTC	Lr1.9Avgsp	20	12.4	10.9	2_11	56	2_33 - 2_6	37.8-65.2	217.9 (1.3)	221.9 (1.0)	230.8 (1.5)	loss	-2.39	6.44	-0.37

Table 5.1: QTL statistics

For each chromosome, the gill raker number or spacing QTL with the highest LOD score is shown. LOD is logarithm of the odds, PVE is percent variance explained, M is the marine allele, F is the freshwater allele, d is the dominance coefficient, a is the additive coefficient, d/a is the dominance (see Methods). “Gain” indicates that the freshwater allele of the QTL promotes more gill rakers or denser raker spacing and “loss” indicates that the freshwater allele of the QTL promotes fewer gill rakers or wider raker spacing (the direction of evolutionary change). Phenotypes were adjusted for standard length and back transformed to a mean standard length of 40 mm. Positions of markers are in centiMorgans.

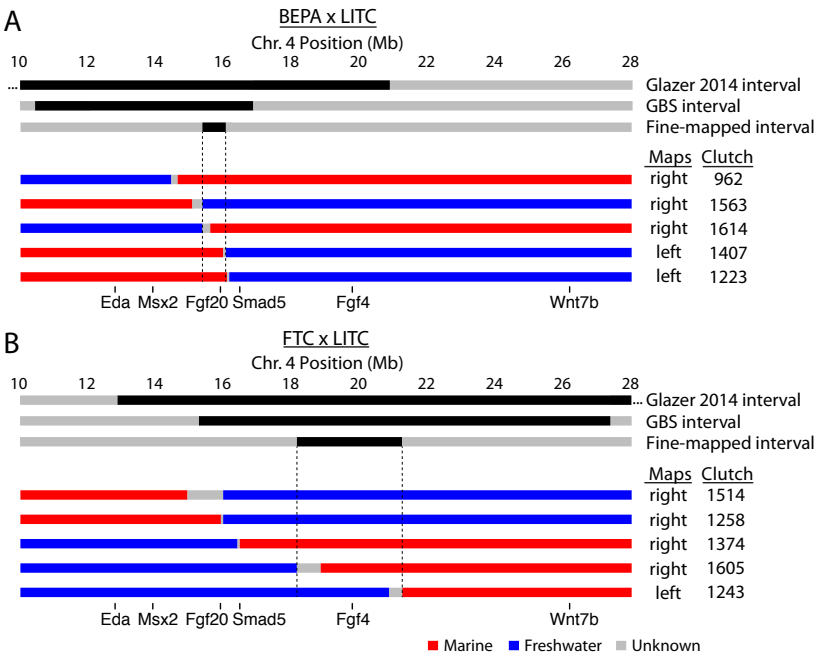


Figure 5.3: Fine mapping of chromosome 4 QTL
Fine mapping results from the BEPA (A) or FTC (B) crosses are shown. Five recombinant chromosomes were analyzed for each QTL. For each recombinant chromosome, the QTL was mapped to the left or right of the recombinant breakpoint. Chromosome 4 position is adjusted for the genome misassembly from 17.82 to 28.36 Mb (Glazer et al. 2014; Glazer et al. 2015). For additional statistics on the recombinant crosses, see **Tables 5.2** and **5.3**.

candidate genes. Excitingly, *Fibroblast Growth Factor 20 (Fgf20)* and *Fibroblast Growth Factor 4 (Fgf4)* remained within the QTL intervals in the BEPA and FTC crosses, respectively.

Next, we performed recombinant fine-mapping of the chromosome 20 QTL in both the FTC and BEPA crosses. We analyzed two recombinant chromosomes in the BEPA cross and three recombinant chromosomes in the FTC cross (**Figure 5.4**). Combined, these genetic crosses fine-mapped the BEPA QTL to a 2.3 Mb interval and the FTC QTL to a 5.9 Mb interval. These intervals overlapped, further supporting a similar genetic basis in the two crosses. The overlapping interval was 0.90 Mb in size and contained a Hox cluster and *Hey1*, a gene in the Notch pathway.

Of these four fine-mapped QTL, the QTL with the smallest physical interval was the BEPA chromosome 4 QTL. The fine-mapped chromosome 4 BEPA interval contained 44 Ensembl-predicted genes (Jones et al. 2012), of which 33 had expression detected in larval branchial tissue by RNA-seq (**Table 5.4**). An excellent candidate gene, *Fgf20*, was located within the fine-mapped interval. To further test whether *Fgf20* is involved in gill raker patterning, we generated fish with mutated copies of *Fgf20* using TALENs (**Figure 5.6**). Marine LITC fish were injected with TALENs targeting exon 1 of *Fgf20*, and F_0 fish were identified that were mosaic for induced mutations in *Fgf20*. We bred two mosaically mutant F_0s to generate an F_1 cross that contained a mixture of

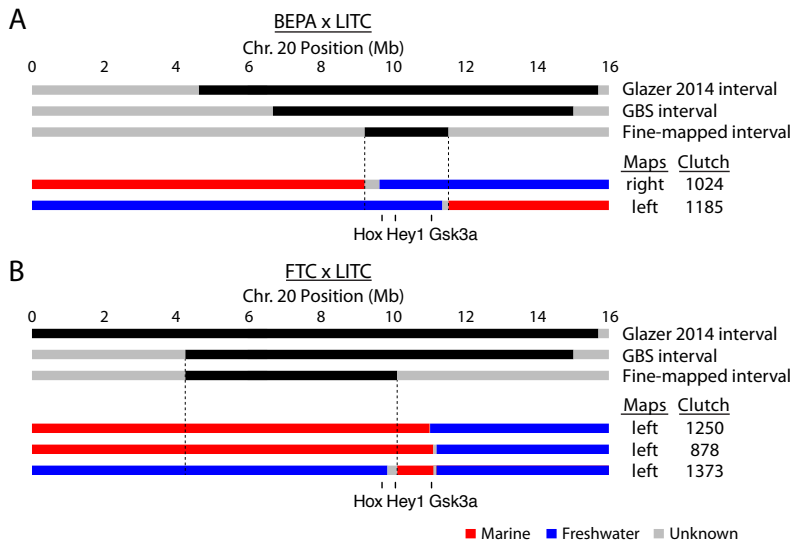


Figure 5.4: Fine mapping of chromosome 20 QTL
Fine mapping results from the BEPA (A) or FTC (B) crosses are shown. Two recombinant chromosomes were analyzed in the BEPA cross and 3 in the FTC cross. For each recombinant chromosome, the QTL was mapped to the left or right of the recombinant breakpoint. Clutch 1373 tested a double recombinant chromosome that had the same right breakpoint as clutch 878. The combined results of clutch 878 and 1373 imply that the interval maps to the left of the left breakpoint of clutch 1373. For additional statistics on the recombinant crosses, see **Tables 5.2** and **5.3**.

homozygous wild-type, heterozygous mutant, and homozygous mutant (transheterozygous for two different mutations) fish. Fragment analysis was used to identified several small insertion and deletion mutations near the predicted TALEN target site in the F_1s. Four mutations were validated by Sanger sequencing, and all four had the exact mutation sizes predicted by fragment analysis (**Figure 5.7A**). These four mutations were predicted to result in frameshifted FGF20 proteins that were missing most of the canonical FGF protein domain (**Figure 5.8**). Therefore, these mutations (and other small indels that were not multiples of 3) are likely loss of function alleles.

We analyzed the gill raker number and spacing phenotypes of the TALEN mutants, excluding fish with insertions or deletions that were multiples of 3. Excitingly, fish with TALEN mutations had quantitative reductions in gill raker number relative to their wild-type siblings (**Figure 5.7B-D**, P=0.007 by ANOVA). Heterozygous fish had intermediate gill rakers to homozygous wild-type or mutant fish, although only the homozygous mutant-homozygous wild-type comparison was significant (p=0.005 by Tukey's HSD test). The decrease in gill raker number was accompanied by a concomitant increase in gill raker spacing (P<0.05 by ANOVA), that also had approximately additive effects (**Figure 5.7E**). One non-synonomous coding mutation was present between the grandparents of the BEPA cross. The BEPA grandparent was heterozygous for a leucine→isoleucine mutation (**Figure 5.8**). However, both freshwater alleles of the chromosome 4 QTL had an effect on gill raker number (Glazer et al. 2014). In addition,

Clutch	Cross	Chr	n	Breakpoint	Cross		M-F	Likelihood ratio test		T test	Allele	Rec. phenotype	Maps
					Type	Scheme	AIC	M vs. M+F	F vs. M+F				
962	BEPA	4	86	14.45-14.64	A	FR x MF	3.7	0.4856	0.0396	N/A	F-M	M	Right
1563	BEPA	4	82	15.07-15.37	A	MR x MF	-6.2	0.0113	0.6087	N/A	M-F	F	Right
1614	BEPA	4	74	15.37-15.6	A	FR x MF	6.5	0.1351	0.0032	N/A	F-M	M	Right
1407	BEPA	4	79	15.98-16.06	B	MR x MM	N/A	N/A	N/A	0.471 (MM,MR)	M-F	M	Left
1223	BEPA	4	84	16.09-16.16	A	MR x MF	3.9	0.7142	0.0462	N/A	M-F	M	Left
1514	FTC	4	88	14.93-15.98	A	MR x MF	-13.7	0.0002	0.5916	N/A	M-F	F	Right
1258	FTC	4	84	15.91-15.98	A	FR x MF	-7.1	0.0057	0.4676	N/A	M-F	F	Right
1374	FTC	4	90	16.4-16.468	A	FR x MF	5.2	0.2227	0.0097	N/A	F-M	M	Right
1605	FTC	4	86	18.15-18.85	A	FR x MF	6.2	0.7111	0.0116	N/A	F-M	M	Right
1243	FTC	4	164	20.86-21.25	A	MR x MF	-8.1	0.0042	0.7543	N/A	F-M	F	Left
1024	BEPA	20	94	9.24-9.65	A	FR x MF	17.6	0.4825	2.07E-05	N/A	F_M	M	Right
1185	BEPA	20	86	11.39-11.56	A	FR x MF	19.5	3.02E-06	0.1291	N/A	F_M	F	Left
878	FTC	20	82	11.14-11.23	A	MR x MF	30.4	0.5504	2.97E-08	N/A	M-F	M	Left
1250	FTC	20	91	11.04-11.05	B	FR x MM	N/A	N/A	N/A	0.018 (MF,MR)	M-F	M	Left
1373	FTC	20	89	9.87-10.14*	C	RD x MF	7.9	0.003667	0.4619	N/A	F-M-F	F	Left

Table 5.2: Fine-mapping statistics
Statistics on each recombinant cross. Chromosome 4 position is adjusted for the genome misassembly from 17.82 to 28.36 Mb (Glazer et al. 2014; Glazer et al. 2015). AIC is Akaike Information Criteria, M is the marine allele, F is the freshwater allele, R is the recombinant allele, D is the double recombinant allele. Type refers to the recombinant cross scheme (see Methods and **Figure 5.5**). AIC refers to the difference in AIC between the freshwater and marine statistical models (positive numbers indicate better support negative numbers indicate better support for the freshwater model). The likelihood ratio test gives a P value for the additional fit of a marine+freshwater model compared to a marine only model or a freshwater only model. The M vs. MF column gives a P value for the significance of the freshwater model having an additional effect, and the F vs. MF column gives a P value for the significance of the marine model having an additional effect.

Clutch	Cross	Chr	Cross		Genotype				Adjusted Raker Number Mean (Standard Error)				n			
			Type	Scheme	A	B	C	D	A	B	C	D	A	B	C	D
962	BEPA	4	A	FR x MF	MF	FF	MR	FR	0.205 (0.272)	-0.465 (0.2)	0.563 (0.146)	-0.017 (0.188)	14	27	18	27
1563	BEPA	4	A	MR x MF	MM	MF	MR	FR	0.483 (0.213)	0.122 (0.221)	0.272 (0.219)	-0.666 (0.166)	19	19	19	25
1614	BEPA	4	A	FR x MF	MF	FF	MR	FR	-0.258 (0.194)	-0.527 (0.334)	0.5 (0.262)	0.157 (0.179)	21	11	16	26
1407	BEPA	4	B	MR x MM	MM	MR	*	*	-0.092 (0.147)	0.074 (0.169)	*	*	35	43	*	*
1223	BEPA	4	A	MR x MF	MM	MF	MR	FR	0.434 (0.172)	-0.392 (0.173)	0.155 (0.288)	-0.253 (0.249)	26	26	17	15
1514	FTC	4	A	MR x MF	MM	MF	MR	FR	0.61 (0.21)	0.127 (0.2)	-0.02 (0.194)	-0.712 (0.161)	22	21	23	22
1258	FTC	4	A	FR x MF	MF	FF	MR	FR	0.365 (0.255)	-0.203 (0.185)	0.372 (0.229)	-0.492 (0.158)	17	24	23	20
1374	FTC	4	A	FR x MF	MF	FF	MR	FR	0.149 (0.16)	-0.59 (0.174)	0.747 (0.293)	-0.18 (0.164)	30	23	17	20
1605	FTC	4	A	FR x MF	MF	FF	MR	FR	0.096 (0.143)	-0.528 (0.197)	0.626 (0.227)	-0.011 (0.292)	29	25	17	15
1243	FTC	4	A	MR x MF	MM	MF	MR	FR	0.332 (0.152)	-0.036 (0.135)	0.029 (0.141)	-0.664 (0.153)	53	43	42	26
1024	BEPA	20	A	FR x MF	MF	FF	MR	FR	-0.05 (0.164)	-0.901 (0.174)	-0.557 (0.185)	0.153 (0.193)	27	19	27	21
1185	BEPA	20	A	FR x MF	MF	FF	MR	FR	0.702 (0.221)	-0.222 (0.165)	0.631 (0.181)	-0.631 (0.181)	16	31	16	23
878	FTC	20	A	MR x MF	MM	MF	MR	FR	0.521 (0.122)	-0.987 (0.178)	0.594 (0.208)	-0.851 (0.16)	33	13	18	18
1250	FTC	20	B	FR x MM	MF	MR	*	*	-0.294 (0.142)	0.207 (0.144)	*	*	38	53	*	*
1373	FTC	20	C	RD x MF	MR	FR	MD	FD	0.555 (0.131)	0.002 (0.233)	-0.189 (0.204)	-0.411 (0.239)	24	19	24	22

Table 5.3: Additional fine-mapping statistics
Additional statistics on each recombinant cross. N is sample size, M is the marine allele, F is the freshwater allele, R is the recombinant allele, and D is the double recombinant allele. Type refers to the recombinant cross scheme (see Methods and **Figure 5.5**). Phenotypes were adjusted for standard length, the other QTL, and/or family, as appropriate, and z-scored to have a mean of 0 and standard deviation of 1. N is sample size.

this mutation was predicted to be tolerated by SIFT, with a score 0.19 (Kumar et al. 2009). Therefore, this coding mutation likely does not underlie the BEPA chromosome 4 gill raker QTL.

We also examined the phenotype of fish with TALEN-induced mutations in *Smad5*, a gene on chromosome 4 inside the original BEPA and FTC intervals. Marine Rabbit Slough (RS) fish were injected with TALENs targeting exon 1 of *Smad5* (**Figure 5.9**), and F_0 fish were identified that were mosaic for induced mutations in *Smad5*. F_0 fish were crossed to lab-reared RS fish to generate heterozygous F1s, which were incrossed to generate F2s with a mixture of wild-type, heterozygous, and homozygous mutant for a single 7 base pair deletion allele of *Smad5* (Figure 5.10A). This mutant allele is predicted to result in a frameshifted protein that lacks the MH2 domain and most of the MH1 domain (**Figures 5.10B, 5.12**). Fish that were homozygous mutant for *Smad5* had fewer gill rakers relative to wild-type or heterozygous mutant siblings ($P<0.0001$ by ANOVA and Tukey's HSD test), indicating a recessive phenotype. Homozygous mutant fish also had a recessive increase in gill raker spacing ($P= 0.03$ by ANOVA, 0.002 for homozygous mutant-wild type comparison by Tukey's HSD test). Homozygous mutant *Smad5* fish also had additional skeletal phenotypes. These fish had epibranchial 4 bones that were either fused with epibranchial 3 or absent (**Figure 5.11A-D**). They also had abnormal branchiostegal rays that were bifurcated and small opercles (**Figure 5.11E-H**). Finally, homozygous mutants were significantly smaller than wild-type fish or heterozygotes, suggesting decreased fitness ($P<0.0001$ by Tukey's HSD test, **Figure 5.11I**). Despite the gill raker phenotype of *Smad5* TALEN-induced mutants, *Smad5* was not inside the fine-mapped BEPA or FTC chromosome 4 QTL. *Smad5* was located 0.41 Mb and 1.68 Mb from the BEPA and FTC intervals, respectively. No non-synonymous coding changes in *Smad5* were detected between the grandparents of the BEPA or FTC crosses (**Figure 5.12**).

To investigate whether FGF signaling plays a conserved role in gill raker development, we asked whether previously described mutations in the FGF receptor *Fgfr1* in carp and zebrafish affect gill raker patterning. Previous work from the 1940s described a quantitative gill raker reduction phenotype in *mirror* carp, an allele that was recently shown to be loss of function mutation in *Fgfr1a1*. We analyzed a genetic cross segregating this carp mutant allele for *Fgfr1a1* and found homozygous mutants to have significant quantitative reductions in gill raker number ($P=0.009$, Student's T test, Figure 5.13). We also analyzed the zebrafish *spiegeldanio* (*spd*) allele, which is a loss of function mutation in the FGF receptor *Fgfr1a*, and found a statistically significant quantitative reduction in gill raker number as well ($P=0.04$, Student's T test, **Figure 5.13**). Together these results establish a conserved role of FGF signaling in gill raker development across fish.

DISCUSSION

To test whether the convergent evolution of gill raker reduction was due to parallel genetic changes, as well as to identify the genes underlying this classic adaptive trait, we mapped the genome wide genetic basis of gill raker number and spacing in two crosses with different independently derived freshwater populations. Three overlapping

Ensembl Gene ID	Location	Ensembl name	Expression Level
18534	15.373 - 15.51	tenm1	0.3
18539	15.512 - 15.515	sh2d1ab	8.4
18540	15.518 - 15.52	agtr2	0
18541	15.527 - 15.532	APOOL	41.7
18542	15.536 - 15.581	si:ch211-26b3.4	0
18543	15.585 - 15.588	HDX	6.8
18545	15.59 - 15.601	RPS6KA6	7.9
18547	15.61 - 15.611	POU3F4	0
18548	15.645 - 15.648	*	7.5
18550	15.649 - 15.654	*	0
18552	15.66 - 15.668	SH3BGRL	103.4
18556	15.669 - 15.672	*	118.9
18558	15.681 - 15.685	FAM46D (1 of 2)	0
18559	15.72 - 15.725	ITM2A	14.3
18571	15.726 - 15.728	GPR174	4.6
18572	15.736 - 15.737	p2ry10	8.4
18574	15.741 - 15.759	LPAR4	0.8
18578	15.741 - 15.757	*	0.8
18579	15.772 - 15.773	cysltr1	0.5
18580	15.794 - 15.801	*	2.5
18585	15.804 - 15.808	polr1d	76.2
18587	15.809 - 15.812	sybl1	30.1
18595	15.824 - 15.839	mtus1a	1.9
18598	15.84 - 15.843	pdgfrl	52.6
18601	15.845 - 15.852	slc7a2	0.8
18602	15.857 - 15.864	mtmr7a	3.4
18605	15.867 - 15.87	cnot7	5.4
18608	15.872 - 15.877	zdhhc2	1.8
18610	15.885 - 15.897	MICU3 (1 of 2)	1.4
18613	15.898 - 15.899	fgf20b	5.1
18614	15.917 - 15.918	her11	0
18615	15.92 - 15.921	her5	0
18616	15.924 - 15.928	pfdn6	161.3
18620	15.929 - 15.931	mmgt1	137.5
18623	15.933 - 15.942	ddx26b	3.3
18626	15.943 - 15.948	mospd1	26.3
18631	15.95 - 15.952	fam122b	6.3
22898	15.951 - 15.951	snoU109	0
18634	15.954 - 15.959	hprt1	43.3
18642	15.963 - 15.97	phf6	33
18647	15.972 - 15.975	cab39l1	9.4
21515	15.981 - 15.981	*	0
22231	15.981 - 15.981	*	0
18655	15.982 - 16.064	gpc3	0

Table 5.4: Genes in BEPA 4 interval and expression by RNAseq
The 44 Ensembl-predicted genes (Jones et al. 2012) in the BEPA chromosome 4 fine-mapped QTL interval are shown. Ensembl gene numbers are abbreviated. For example 18548 refers to ENSGACT00000018548. Expression level refers to Fragments Per Kilobase of exon per Million fragments mapped (FPKM) by RNAseq from 18 dpf FTC x LITC F_1 larval branchial skeletal tissue. 33 genes in the interval have detected expression in branchial tissue.

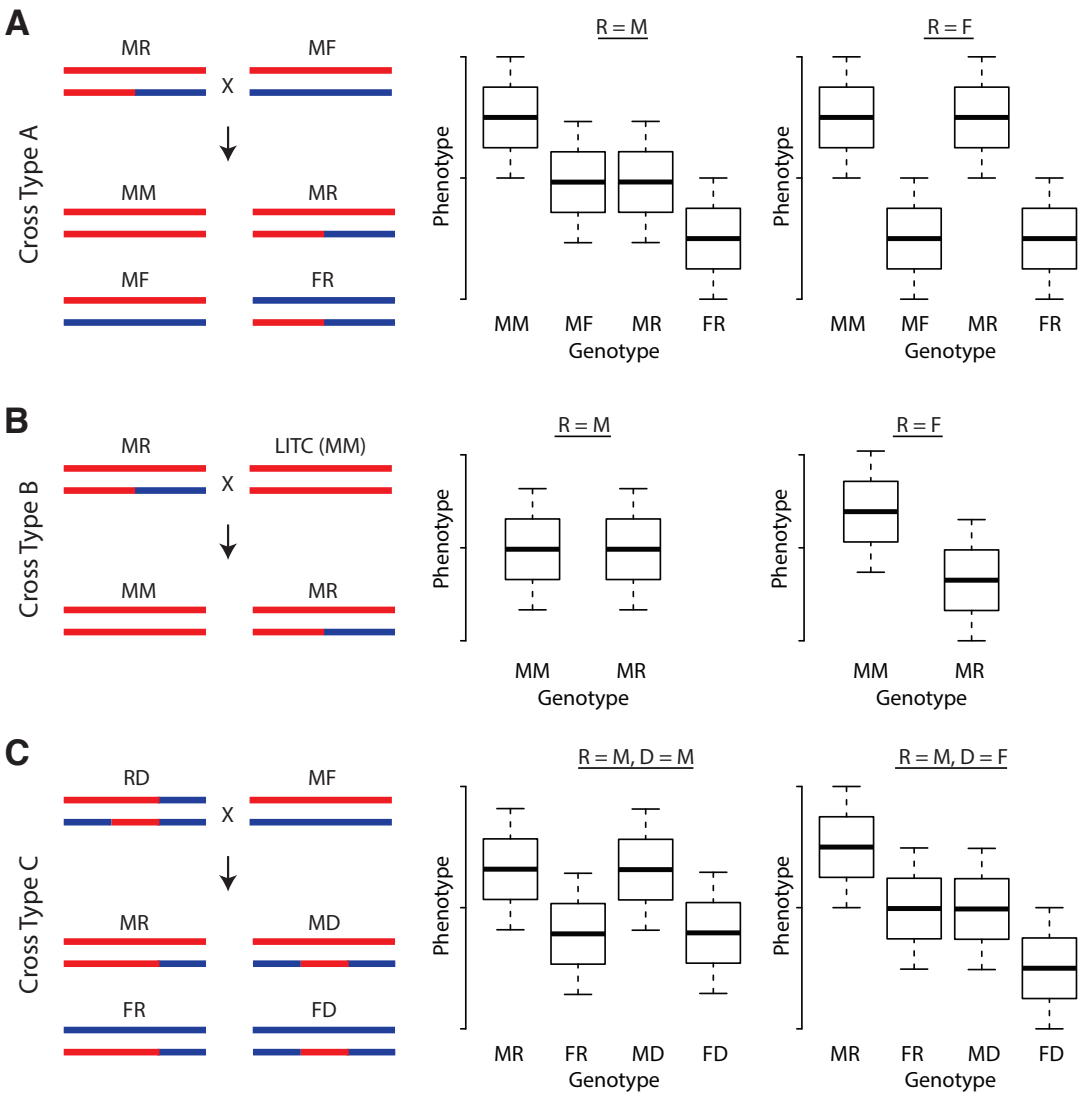


Figure 5.5: Diagram of recombinant crossing schemes
M is marine allele, F is freshwater allele, R is recombinant allele, D is double recombinant allele. Three recombinant crossing schemes were used. Type A consists of a single recombinant chromosome (MR or FR fish) crossed to MF fish. Cross Type B consists of a single recombinant chromosome (MR or FR fish) crossed to MM fish. Cross Type C consists of a double recombinant chromosome cross; a RD fish was crossed to MF fish. Phenotype predictions for the possible recombinant chromosome outcomes are shown.

medium-effect QTL mapped to chromosomes 4, 16, and 20, whereas six other non-overlapping modifier QTL mapped in the BEPA cross and eight other QTL mapped in the FTC cross. The shared chromosome 4, 16, and 20 QTL are the three largest-effect QTL in the BEPA cross and two of the three largest-effect QTL in the FTC cross, perhaps indicating that larger-effect QTL are more likely to be used repeatedly in evolution. The chromosome 16 QTL mapped to nearly identical locations in the two

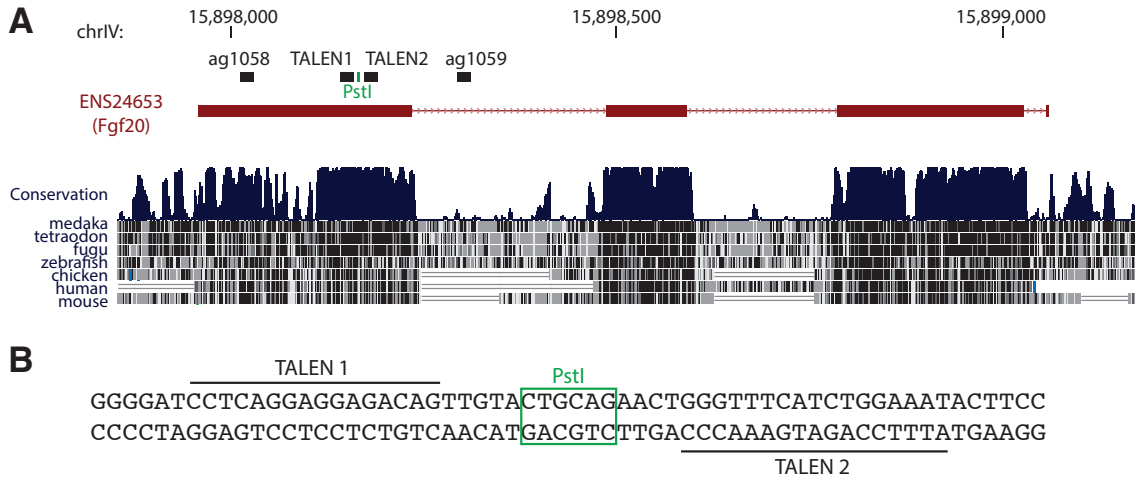


Figure 5.6: Targeting scheme for *Fgf20* TALEN
(A) TALENs were targeted against exon 1 of *Fgf20*. TALEN 1 and TALEN 2 binding sites are shown. A restriction digest assay was to detect mutations. Primers ag1058 and ag1059 amplified a PCR product and wild-type alleles were able to be cut by PstI. (B) Zoomed in view of TALEN binding sites and PstI cut site.

crosses, but had opposite directions of effect. One possibility for this finding is the presence of sign epistasis or allelic reversal, where the same allele has opposite effects depending on the genetic background (Aung et al. 2015; Orgogozo et al. 2006). Alternatively, the chromosome 16 QTL in the two crosses could be due to different, nearby genes or different alleles of the same gene. Several of the FTC QTL have specific modular effects on row 1 gill raker number and spacing, which might cause the extra raker reduction in FTC row 1 relative to other freshwater populations (Glazer et al. 2014). Overall 16 of the 20 total QTL promote fewer gill rakers or wider spacing, a significant bias in direction that is consistent with gill raker reduction being under strong natural selection.

In both the initial microsatellite marker mapping (Glazer et al. 2014) and the GBS mapping presented here, the 1.5 LOD intervals of the chromosome 4 QTL in the BEPA and FTC cross, while overlapping, mapped to largely distinct regions. To further test whether these two chromosome 4 QTL were genetically distinct, we undertook a series of fine-mapping experiments that narrowed the intervals in each cross substantially. The two fine-mapped QTL did not overlap, indicating that a separate genetic basis, likely using different genes, underlies the chromosome 4 gill raker QTL in these two freshwater populations. This result serves as a cautionary note to conclude that similarly mapped QTL are identical. Other features of the chromosome 4 QTL further support that the two QTL are distinct. For example, while the FTC QTL had effects on all three domains of raker spacing, the BEPA QTL had a highly modular effect on the lateral domain. This effect might be due to developmental timing, as the lateral gill rakers are the first to form

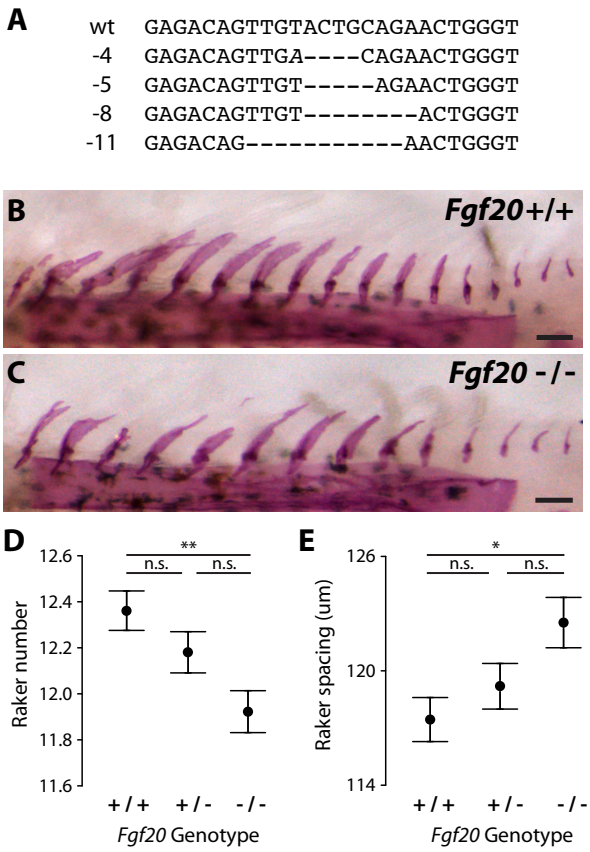


Figure 5.7: Loss of function mutations in *Fgf20* phenocopy the evolved raker reduction phenotype
(A) Wild-type and four mutant deletion alleles were Sanger sequenced. (B-C) Left side row 1 gill rakers in a *Fgf20* homozygous wild-type (B) and homozygous mutant (C) stickleback. Mutants have fewer gill rakers that are more widely spaced. These differences are quantified for row 1-9 average gill raker number (D) and row 1 gill raker spacing (E). n.s. is not significant, * P<0.05, ** P<0.01, n.s. is not significant, Tukey's HSD test. Scale bar is 100 microns.

during development (Glazer et al. 2014). In addition, the BEPA QTL had a significant effect on both dorsal and ventral raker number, whereas the FTC QTL only had significant effects on ventral raker number. The chromosome 20 QTL in the BEPA and FTC cross still overlapped in a region containing a *Hox* cluster and *Hey1*, a gene in the Notch pathway.

Fine-mapping of the BEPA chromosome 4 QTL narrowed the QTL to a 0.69 Mb interval that included an excellent candidate gene, *Fgf20*. If *Fgf20* underlies the BEPA 4 QTL, fish with induced mutations in *Fgf20* should have a gill raker phenotype. Indeed, TALEN-induced homozygous mutants had fewer gill rakers and wider gill raker spacing than wild-type siblings. The *Fgf20* mutant shares many phenotypic similarities with the evolved BEPA 4 raker QTL: (1) a quantitative decrease in gill raker number of 0.6-0.8

Figure 5.8: A coding mutation in *Fgf20* is not predicted to affect protein function
Multiple sequence alignment of the main wild-type allele (present in the marine LITC grandparent) and a variant allele (present as a heterozygote in the BEPA grandparent). The variant allele contains a single non-synonomous coding change, from leucine to isoleucine. This change is not predicted to affect protein function by SIFT. Also aligned are four TALEN-induced deletions. Each yields a frameshifted protein that results in premature truncation, losing most of the canonical FGF protein domain.

rakers per row, (2) an increase in gill raker spacing, and (3) an additive effect on gill raker number and spacing where heterozygotes have an intermediate phenotype. No other pleiotropic effects of the *Fgf20* mutation were observed, although comprehensive phenotyping of other traits was not performed. One non-synonomous coding mutation in *Fgf20* was present between the BEPA freshwater and LITC marine grandparents of the BEPA cross. However, this mutation is unlikely to underlie the QTL for two reasons. First, the BEPA grandparent was heterozygous for the mutation and both freshwater alleles had an effect on gill raker number (Glazer et al. 2014). Second, the mutation is not predicted to significantly effect protein function. Therefore, if *Fgf20* underlies the BEPA chromosome 4 QTL, the effect is likely due to *cis*-regulatory variation that affects *Fgf20*'s expression level and/or spatial pattern during development. Given the gill raker phenotype of the *Fgf20* mutant, *Fgf4* stands out as an intriguing candidate for underlying the FTC chromosome 4 QTL.

Previous work has suggested that FGF signaling might have a role in gill raker development. The *mirror* mutation in carp, recently identified as a loss of function allele of *Fgfr1a1*, was described in 1940 to have a quantitative reduction in gill raker number (Golovinskaya 1940; Rohner et al. 2009). This result was confirmed in this study, with a similar effect size (0.8 gill rakers/row in Golovinskaya 1940, 1.2 gill rakers/row in this study). We also found that mutations in zebrafish *Fgfr1a* also affect gill raker number. These results, combined with the stickleback *Fgf20* TALEN phenotype, imply that FGF signaling has a conserved role in gill raker patterning across fish species. FGF signaling

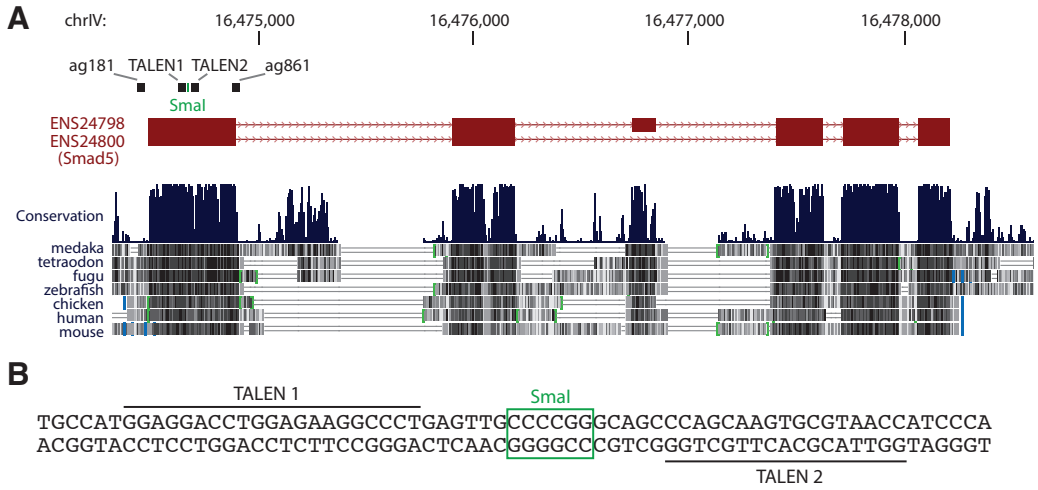


Figure 5.9: Targeting scheme for *Smad5* TALEN
(A) TALENs were targeted against exon 1 of *Smad5*. TALEN 1 and TALEN 2 binding sites are shown. A restriction digest assay was to detect mutations. Primers ag181 and ag861 amplified a PCR product and wild-type alleles were able to be cut by SmaI. (B) Zoomed in view of TALEN binding sites and SmaI cut site.

variation might play a broader role in the dramatic variation in gill raker number and spacing within and between species (Kahilainen et al. 2011; Magnuson and Heitz 1971; Noramly and Morgan 1998; Schluter 2000).

Gill rakers appear to have genetic and developmental similarities to other vertebrate epithelial appendages, a broad class of periodically patterned organs that include hair, teeth, feathers, sweat glands, and scales (reviewed in Chuong 1998; Chuong et al. 2013). FGF signaling plays a major role in the patterning of many epithelial appendages, including feathers, hair, scales, and teeth (Jung et al. 1998; Pispa and Thesleff 2003; Wells et al. 2012) FGF ligands are typically expressed in the placode and restricted from the inter-placode regions. FGF promotes denser appendage spacing, and manipulations to FGF affect epithelial appendage spacing. Intriguingly, *Fgf20* is the gene underlying the *scaleless* mutation in chickens, which results in a loss of feathers (Wells 2012). Therefore, *Fgf20* may play a conserved role across organisms in epithelial appendage patterning and evolution.

METHODS

Animal statement

Wild anadromous marine fish were collected from the Little Campbell River 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 in Washington under a fish scientific collection permit from the Washington

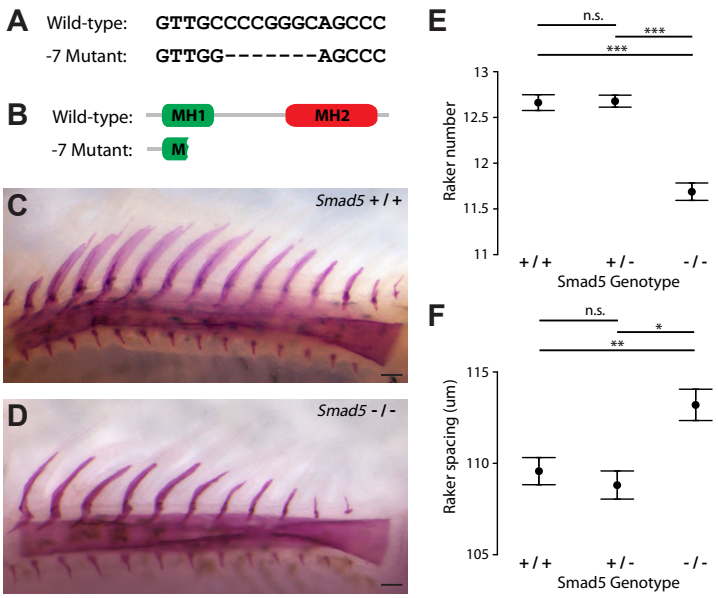


Figure 5.10: Loss of function mutation in *Smad5* results in gill raker number and spacing phenotype
(A) Wild-type and TALEN-induced 7 base pair deletion allele were Sanger sequenced. (B) Predicted truncated protein from 7 base pair deletion allele. (C-D) Left side row 1 gill rakers in a *Smad5* homozygous wild-type (C) and homozygous mutant (D) stickleback. Mutants have fewer gill rakers that are more widely spaced. These differences are quantified for row 1-9 average gill raker number (E) and row 1 gill raker spacing (F). n.s. is not significant, * P<0.05, ** P<0.01, ***P<0.001, n.s. is not significant, Tukey's HSD test.

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 number R330 and 13834).

Generation of F2 crosses
Two marine x freshwater F_1 crosses were previously described (Glazer et al. 2014; Glazer et al. 2015, in press). Briefly, a wild-caught male marine fish from the Little Campbell River (British Columbia, Canada,"LITC") was crossed to a wild-caught female freshwater fish from Fishtrap Creek (Washington state; "FTC") to generate the FTC cross. A male freshwater fish from Bear Paw Lake (Alaska, "BEPA", lab-reared offspring of wild-caught parents) was crossed to a wild-caught marine female LITC fish to produce the BEPA cross. F_1s were intercrossed to generate 360 and 363 F2 fish from the FTC and BEPA crosses, respectively. Fish with low genotype coverage (n = 2 from each cross) were removed from the analysis.

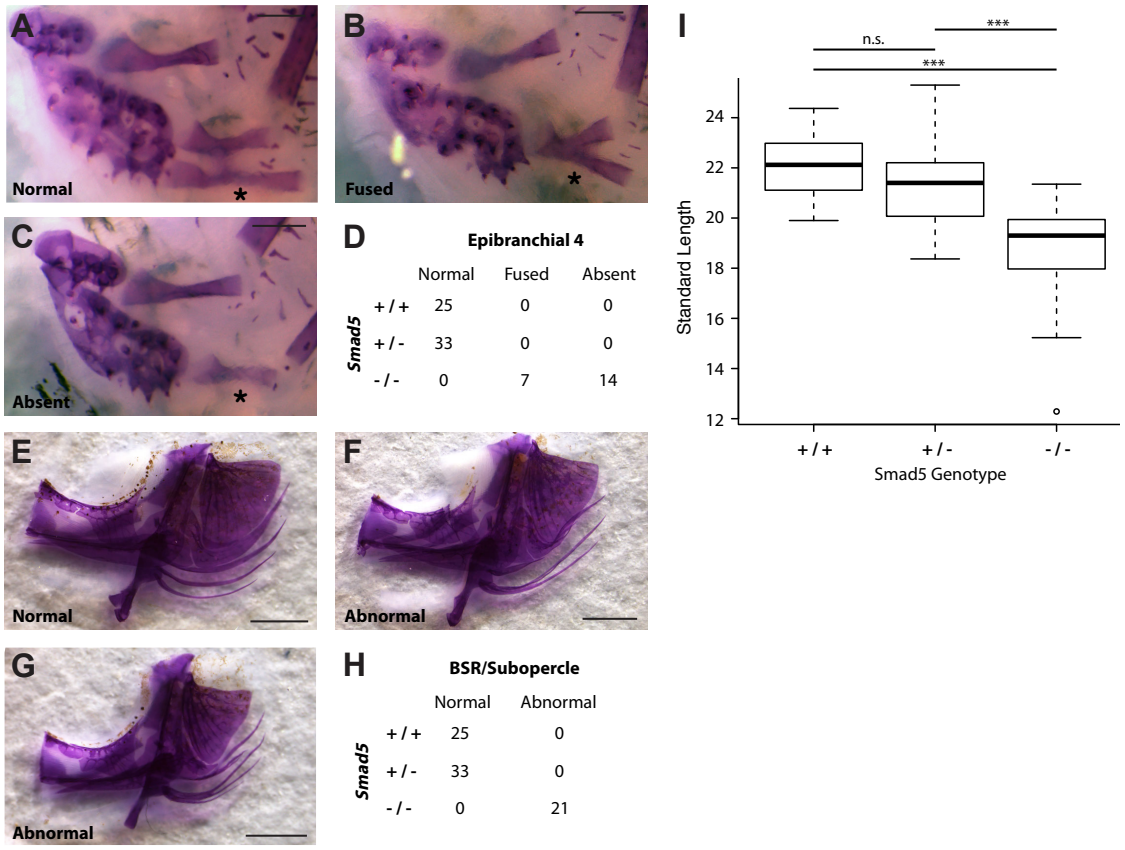


Figure 5.11: Additional *Smad5* skeletal phenotypes
(A-C) Pictures of epibranchial 4 (eb4) bone in homozygous wild-type (A) or homozygous mutant (B-C) fish. In *Smad5* homozygotes, Eb4 was either fused with eb3 (B) or absent (C). (D) Quantification of eb4 defects. (E-F) Pictures of opercle, subopercle, and branchistegal rays (BSRs) in homozygous wild-type (A) or homozygous mutant (B-C) fish. In Smad5 homozygotes, BSRs were bifurcated and subopercles and opercles were reduced in size. (H) Quantifications of defects. (I) *Smad5* mutants were significantly shorter compared to sibling heterozygotes or wild-types. *** P<0.001, n.s. is not significant, Tukey's HSD test.

Fish raising

Lab-reared fish were raised at 18°C in 110 liter (29 gallon) aquaria in a common brackish salinity (3 ppt, 10% ocean water) and fed a common diet of live *Artemia* nauplii and frozen *Daphnia* as fry and juveniles, and bloodworms and *Mysis* shrimp as adults. Adult F2s were raised to a mean size (standard length) of 38.0 mm and 42.1 mm in the FTC and BEPA crosses, respectively. Fish from F_1 crosses were intercrossed to generate F2 families ranging in size from 26-108 fish. Adult F2 fish (n= 360 and 363 from the FTC and BEPA crosses, respectively) were sequenced with GBS from 5 and 9 F2 families in the FTC and BEPA crosses, respectively. Fish from the *Fgf20* and *Smad5* TALEN crosses were raised to a mean size of at least 20 mm SL before phenotypic analysis, after which point gill raker number is approximately fixed (Glazer et al. 2014).

Figure 5.12: No coding mutations in Smad5
Multiple sequence alignment of the marine and freshwater alleles from the grandparents of the BEPA and FTC crosses. No non-synonymous coding differences were present. Also aligned is the 7 base pair TALEN-induced deletions. The allele yields a frameshifted protein that results in premature truncation, losing most of the canonical MH1 and all of the MH2 protein domain.

Phenotyping gill raker number and spacing

To stain bones, fish were fixed for 1-2 days in 10% neutral buffered formalin or 5-7 days in 4% paraformaldehyde in 1xPBS, washed with water overnight, stained overnight with 0.008% Alizarin Red S in 1% potassium hydroxide, destained in water overnight, and cleared in a 0.25% potassium hydroxide, 50% glycerol solution. Branchial skeletons were dissected and flat mounted on a bridged coverslip as previously described (Glazer et al. 2014). Ventral row 1-9 and dorsal row 1 gill raker counts were defined as Alizarin-positive gill rakers overlying the ceratobranchial 1 and epibranchial 1 bones, respectively. To determine gill raker spacing, images of flat-mounted branchial skeletons were taken on a Leica M165 dissecting microscope, the coordinates of the center of the base of each gill raker were determined in imageJ (Schneider et al. 2012), and these coordinates were processed in a custom Python script to determine spacing phenotypes. For each ventral row, the average of every space in a row ("Avgsp") was determined. In addition to gill raker number and spacing, *Smad5* TALEN mutants were also phenotyped for additional phenotypes, including opercle and branchiostegal ray defects and epibranchial 4 absence or fusion with epibranchial 3.

Phenotype processing

For QTL mapping, gill raker number and gill raker spacing phenotypes were tested for an association with standard length and sex (as determined by GBS, Glazer et al. 2015, in press) by linear regression in R (www.r-project.org) and corrected for size, sex, and/or log transformed, when appropriate. When association with standard length was significant ($p<0.05$), residuals were taken from a linear regression. When correlations with standard length and sex were both significant in a linear model, both were corrected

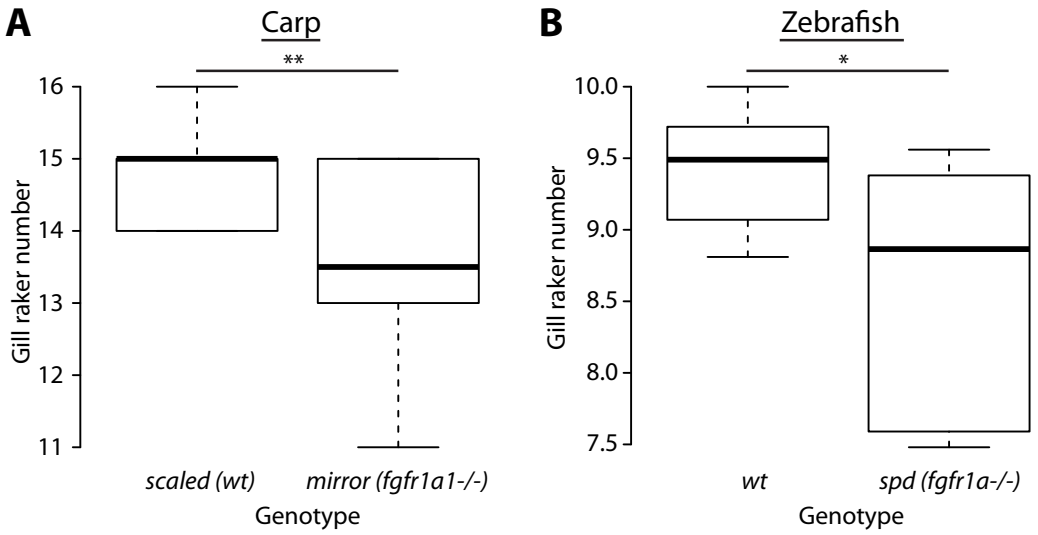


Figure 5.13: FGF signaling is required for proper gill raker patterning in carp and zebrafish (A) Comparison of gill raker number in *scaled* carp (wild-type and *Fgfr1a1*+/-) and *mirror* carp (*Fgfr1a1*-/-). Mirror carp had a significant reduction in row 1 ventral gill raker number (p=0.009, Student's T test). (B) Comparison of gill raker number in wild-type zebrafish and *speigeldanio* mutant (*spd*, *Fgfr1a* -/-) zebrafish. Spd mutants had a significant reduction in row 1-9 average gill raker number (p=0.04, Student's T test).

for. When log transformation improved the normality of the residuals (i.e. by resulting in an Anderson-Darling test changing from $p<0.05$ to $p>0.05$), log transformation was performed. Outliers greater than 4 standard deviations from the trait means were removed (fewer than 0.01% of all values). For QTL mapping, all phenotypes were z-scored to have a mean of 0 and standard deviation of 1. For calculating QTL mean phenotype values by genotypic class, phenotypes were adjusted for standard length in a linear regression and back transformed to values expected for a fish of length 40 mm.

QTL mapping

F2s were genotyped with a binned Genotyping-by-Sequencing approach as previously described (Glazer et al. 2015, in press). The identical genotype files as in (Glazer et al. 2015, in press) were used, consisting of 1,001 markers in the FTC cross and 978 markers in the BEPA cross. For QTL mapping, plate number and gill raker length phenotypes were tested for an association with standard length and sex by linear regression in R (www.r-project.org) and corrected for size, sex, and/or log transformed, when appropriate (see Supplemental Methods). QTL mapping was performed in R/qtl (Broman and Sen 2009; Broman et al. 2003). Initial QTL mapping was performed with *scanone* with Haley-Knott regression. Trait-specific genome-wide significance thresholds with an α of 0.05 were calculated with 1,000 permutations. In cases where multiple significant QTL affected a phenotype, multiple QTL mapping was performed with *stepwiseqtl,* QTL peak markers and LOD plots calculated with *refineqtl*, and peak LOD scores and percent

Primer	Sequence (5' to 3')	Description
ag181	ACAGGTCCTCCCTTTCACCT	Smad5 cut assay forward
ag861	GTGGACCTACCAGGACTCTCC	Smad5 cut assay reverse
ag1058	AGGTGGGGTCGCATTTCAT	Fgf20 cut assay forward
ag1059	TGCTTTCGGCAAATACTTTCT	Fgf20 cut assay reverse
ag1133	TGTAAAACGACGGCCAGTCACTT CCTACCTGGGGACAG	Fgf20 fragment assay (M13+forward)
ag1134	GACTGTGGTCCTTCCTCGTC	Fgf20 fragment assay reverse
ag1137	TGTAAAACGACGGCCAGTGGTG AAAAAGCTGAAGAAGAAGA	Smad5 fragment assay (M13+forward)
ag1138	CGGAAGACCTTTCCTGTGAG	Smad5 fragment assay reverse
Fam-M13	Fam-TGTAAAACGACGGCCAGT	Fragment 3-primer PCR

Table 5.5: Primers used in this study
Primer name, sequence from 5' to 3', and description are shown. Further description of each primer's usage is in the Methods.

variance explained values calculated with *fitqtl*, adjusting for the effect of other QTL underlying the phenotype when appropriate.

Fast DNA isolation for recombinant mapping and screening TALEN mutants
To isolate DNA for PCR, tissue (ground up embryos or fin tissue) was placed into a 96 well plate containing 50 ul of lysis buffer (10mM Tris—pH 8.3, 50 mM KCl, 1.5 mM $MgCl_2$). Samples were incubated at 94^o C for 20 minutes, then 2.5 uL 20 mg/ml Proteinase K (Ambion) was added, and samples were incubated at 55^o C for 30-60 minutes, then 94^o C for 20 minutes to denature the Proteinase K. The resulting liquid was then diluted 11-fold into water.

Recombinant mapping
Inbred lines from the BEPA x LITC and FTC x LITC F2 crosses were bred to study the chromosome 4 and chromosome 20 QTL in each cross. To preserve marine and freshwater alleles of these QTL, F2s from the BEPA x LITC and FTC x LITC crosses were randomly incrossed, and in subsequent generations marine and freshwater alleles of the chromosome 4 and 20 QTL were tracked. Fish were screened by genotyping a small piece of caudal fin tissue as previously described (Glazer et al. 2014) with microsatellite and/or indel markers. A complete list of chromosome 4 and 20 markers used in this study is presented in Table Sx.

F3s were screened for fish that had a non-recombined marine or freshwater allele for the entire 2 LOD intervals of the chr. 4 and/or chr. 20 raker QTL. When possible, markers that could distinguish the two grandparental marine alleles and/or two grandparental freshwater alleles were used. Based on these genotypes, lines of F4 and F5 fish were bred that 1) had non-recombined marine or freshwater alleles for the QTL intervals and 2) had only a single marine allele and single freshwater allele represented.

To maintain the marine and freshwater alleles of a QTL and identify marine-freshwater recombinants within a QTL interval, the following crossing scheme was used, where M represents the marine allele of a QTL and F represents the freshwater allele. Two MF fish were crossed together and their offspring screened as described above. The cross generated an approximate 1:2:1 ratio of MM:MF:FF fish, with x% recombinants, where x was the centiMorgan size of the QTL interval. Fish were screened with two markers on either end of the QTL region, and typically MF fish and recombinant fish were kept. Recombinant breakpoint locations were further mapped with additional markers within the QTL interval.

Three methods of recombinant crosses were used to perform recombinant fine-mapping. See Figure 5.5 for a diagram of the crossing schemes. *Cross Type A*: A fish with a recombinant QTL was crossed to a MF fish. *Cross Type B*: A fish with a recombinant QTL was crossed to a lab reared Little Campbell River marine fish. *Cross Type C*: Double recombinant chromosomes were analyzed. To generate double recombinant chromosomes, a single recombinant fish was bred to a MF fish and the offspring were screened using makers tiling the QTL interval to identify double recombinants. These double recombinants were then bred to a MF fish. For all three cross types, 70-200 offspring fish were raised to 30-40 days post fertilization. In some cases, multiple families of a similar cross type were examined, for example a single recombinant fish crossed to two different MF fish. Details on each recombinant cross are presented in Table x. Details on how statistical tests on the recombinant crosses was performed are in the Supplemental Methods.

Recombinant cross analysis

To analyze recombinant crosses, caudal fins from each fish were removed and genotyped as above with two markers on the chromosome of interest (one on either side of the recombinant breakpoint). Additionally, one marker on the other chromosome was genotyped (e.g. one marker within the chr. 20 QTL was genotyped when analyzing chromosome 4 recombinants). For each fish, bone was stained with Alizarin Red, branchial skeletons were dissected, and row 1-9 ventral gill raker number was determined on the left side or on both sides, as described (Glazer et al. 2014). All statistical analyses were performed in R. Raker number was tested for correlation with standard length, the genotype of the non-recombinant chromosome, and family (in cases of multiple half-sibling families) with a general linear model. Residuals from this model were taken and analyzed as a function of the genotype of the recombinant chromosome of interest. For Cross Type B, a two-tailed Student's T test was used to compare the two classes of fish to test for significant differences. For Cross Type A and Cross Type C, two models were considered: that the recombinant chromosome was genetically equivalent to a marine chromosome, and that the recombinant chromosome was genetically equivalent to a freshwater chromosome. Two new variables were created, a marine and a freshwater dummy variable. For the marine dummy variable, the MR and FR genotypes were turned into MM and MF genotypes, respectively. For the freshwater dummy variable, the MR and FR genotypes were turned into MF and FF genotypes, respectively. The MM, MF, and FF genotypes were then converted into numbers as follows: MM=1, MF=0, FF=-1. Two statistical methods were used to determine whether the relative fit of these two

models to the left side row 1-9 raker number (adjusted for standard length/other QTL as described above). First, the Akaike Information Criterion (AIC) was used to distinguish between the phenotype~dummy1 and phenotype~dummy2 models. The model with the lower AIC value was considered to be a better fit for the data. To determine the significance of AIC differences, simulations were performed with random normally distributed data, which determined the 5% significance threshold to be an AIC difference of 3.20 and a 1% significance threshold to be an AIC difference of 5.47. Second, a likelihood ratio test was performed comparing phenotype~dummy1+dummy2 to phenotype~dummy2 (giving a p value for the significance of adding model 1) and phenotype~dummy1+dummy2 to phenotype~dummy1 (giving a p value for the significance of adding model 2).

Mutating *Fgf20* and *Smad5* with TALENs

TALENs targeting exon 2 of *Fgf20* and exon 2 of *Smad5* were designed using the TALE-NT online tool (https://tale-nt.cac.cornell.edu/, Doyle et al. 2012). The *Fgf20* TALENs HD HD NG HD NI NN NN NI NN NN NI NN NI HD NI NN and NI NG NG NG HD HD NI NN NI NG NN NI NI NI HD HD HD were designed to bind to the DNA sequences CCTCAGGAGGAGACAG and ATTTCCAGATGAAACCC, respectively. The *Smad5* TALENs NN NN NI NN NN NI HD HD NG NN NN NI NN NI NI NN NN HD HD HD NG and NN NN NG NG NI HD NN HD NI HD NG NG NN HD NG NN NN were designed to bind to the DNA sequences GGAGGACCTGGAGAAGGCCCT and GGTTACGCACTTGCTGG, respectively. The TALENs were cloned using Golden Gate cloning into the pTAL3DD and pTAL3RR vectors as described in (Cermak et al. 2011). The plasmids were digested with NotI and RNA was transcribed from each TALEN plasmid using a SP6 transcription kit (New England Biolabs) and cleaned up with an RNeasy Mini kit (Qiagen). A mixture containing 400 ng TALEN 1 RNA, 400 ng TALEN 2 RNA and 1 ul phenol red injection dye in a total volume of 10 ul water was made. Approximately 5-10 nanoliters of this mixture was injected into fertilized 1 cell embryos from a lab-reared Little Campbell River (*Fgf20*) or Rabbit Slough (*Smad5*) marine line.

To detect TALEN-induced mutations, DNA was isolated from ground up embryos at 2 days post fertilization or from pieces of caudal fin tissue from fry at 6 weeks to 3 months post fertilization. Three methods were used to detect TALEN-induced mutations: (1) The target site for the TALEN-induced double-strand break overlapped a restriction enzyme cut site. A PCR product spanning the TALEN target site was assayed for whether it could be cleaved by the restriction enzyme. PCR products were amplified with a touchdown Phusion PCR at 98^{o} 120s, (98^{o} 10s, 65->55^{o} 30s, 72^{o} 15s)x10, (98^{o} 10s, 55^{o} 30s, 72^{o} 15s)x25. For *Fgf20*, primers ag1058 and ag1059 primers were used to amplify a ~294 bp band, and the wild type allele was cut by PstI into 150 and 144 bp. For *Smad5*, primers ag861 and ag181 were used to amplify a ~450 bp band, and the wild type allele was cut by SmaI into 214 and 236 bp. PCR products were assayed by gel electrophoresis, with cut PCR products indicating wild type alleles and uncut PCR products indicating mutant alleles. (2) Fragment analysis was used to assay the exact base pair size of a PCR product spanning the TALEN target site. A three primer "Poor Man's Genotyping" reaction was used (Schuelke 2000). For *Fgf20*, the primers FAM-M13,

ag1133, and 1134 amplified a 175 bp product for the wt allele, and for *Smad5* the primers Fam-M13, ag1137, and ag1138 ampified a 160 bp product for the wt allele. Fish heterozygous or homozygous for mutant sequences had different size amplicons that could be used to determine the insertion or deletion size by comparing the band size to the 175 bp or 160 bp of the wt allele. (3) Sanger sequencing of mutant heterozygotes (as assayed by approach 1 or 2) was performed to identify specific mutant sequences. The cut assay (approach 1) PCRs and digests were performed as above. The uncut (mutant) band was gel extracted (Qiagen Gel Extraction Kit) and Sanger sequenced with primer ag1058 (*Fgf20*) or ag181 (*Smad5*). In some cases, the yield from the PCR was low, so a second PCR was performed using the uncut band as template, and the digest, gel extraction, and sequencing were repeated.

For the *Fgf20* TALEN mutants, F_0 Little Campbell marine fish were assayed using the restriction digest assay, and fish with a high (>30%) of mutant alleles were crossed to each other. This generated F_1 offspring that were a mix of homozygous wild-type, heterozygous, and homozygous mutant (transheterozygous for two different induced mutations). The F_1 cross was genotyped with methods (1) and (2). For the *Smad5* TALEN mutants, F_0 Rabbit Slough marine fish were assayed using the restriction digest assay, and fish with a high (>30%) of mutant alleles were outcrossed to lab-reared Rabbit Slough fish. F_1 offspring were a mix of homozygous wild types and heterozygotes. F_1 offspring were screened using method (2) and heterozygotes were incrossed to generate F_2 clutches with a 1:2:1 ratio of wild-type: heterozygous mutant: homozygous mutant. The F2 cross was genotyped with methods (1) and (2). All primer sequences are presented in **Table 5.5**.

Identification of *Fgf20* and *Smad5* coding mutations

Whole genome resequencing of the marine and freshwater grandparents of the BEPA and FTC crosses was previously described (Glazer et al. 2015). The Ensembl-generated gene models of *Fgf20* and *Smad5* were examined and for each pair of grandparents, any non-synonomous coding mutations were identified. To determine the likely effect of a non-synonomous mutation in *Fgf20* on protein function, the SIFT algorithm was run on a single protein using a PSI-blast search (Kumar et al. 2009). For visual display, protein sequences of TALEN mutants and wild variants were aligned with ClustalW2 (Larkin et al. 2007).

Carp and zebrafish FGF receptor mutants

A cross of carp was analyzed between two fish heterozygous mutations in *Fgfr1a1*, and phenotypes of *mirror* fish (homozygous mutant) were compared to scaled fish (heterozygotes and homozygous wild-type). Fish were stained for bone as described above. Left side row 1 gill raker number was counted using a Leica DM2500 microscope. A Student's T-test was used to compare wild-type to *mirror* raker counts. Homozygous *speigeldanio (spd)* zebrafish and wild-type WIK zebrafish were compared. Left side row 1-9 raker numbers were counted with a Leica M165 microscope. A Student's T-test was used to compare wild-type to *spd* raker counts.

RNAseq of larval branchial tissue

Three 18 days post fertilization larvae from a FTC x LITC F1 cross were analyzed. Branchial skeletons were dissected and placed in TRI reagent (Ambion), and RNA was isolated. RNAseq Illumina sequencing libraries were made using a TruSeq kit (Illumina). Reads were aligned to the stickleback genome (Jones et al. 2012) with Bowtie and Fragments Per Kilobase of exon per Million fragments mapped (FPKM) expression levels were determined with Tophat (Langmead et al. 2009). The average of FPKM values from the three biological replicates is presented.

ACKNOWLEDGEMENTS

Craig Miller provided invaluable guidance on experiment design, data analysis, and writing the manuscript. Matt Harris provided the carp and zebrafish FGF receptor mutants. Kari Miller screened hundreds of fish to identify recombinants, analyzed many of the recombinant mapping crosses, and helped analyze the *Smad5* and *Fgf20* TALEN mutant phenotypes. Emily Killingbeck helped generate F2 GBS genotypes, measured gill raker spacing in the two F2 crosses, helped with the recombinant mapping experiments, and helped clone the *Fgf20* TALENs. Angela Lam phenotyped gill raker number in the FTC cross and analyzed one of the recombinant mapping crosses. Priscilla Erickson injected the *Fgf20* TALEN F_0 fish. Courtney French generated and analyzed the RNAseq data. Patrick Lee and Anthony Lee expertly maintained the sticklebacks. Brock Roberts provided TALEN reagents and advice. Thanks to Mike Bell for his generous gifts of Bear Paw lake fish, Alex Pollen for assistance collecting Fishtrap Creek and Little Campbell River fish, and the Semiahmoo First Nation for permission to collect sticklebacks from the Little Campbell River. This work was supported in part by the NIH (R01- DE021475) to CTM and a Predoctoral Training Grant 5T32GM007127 (AMG), the NSF-Graduate Research Fellowship Program (AMG), a March of Dimes Basil O'Connor Starter Scholar Award (CTM), and a Pew Scholar in Biomedical Sciences award supported by the Pew Charitable Trusts (CTM).

REFERENCES

Arnegard, M.E., M.D. McGee, B. Matthews, K.B. Marchinko, G.L. Conte *et al.*, 2014 Genetics of ecological divergence during speciation. Nature 511: 307-311.

Aung, T., M. Ozaki, T. Mizoguchi, R.R. Allingham, Z. Li *et al.*, 2015 A common variant mapping to CACNA1A is associated with susceptibility to exfoliation syndrome. Nat Genet 47: 387-392.

Baird, N.A., P.D. Etter, T.S. Atwood, M.C. Currey, A.L. Shiver *et al.*, 2008 Rapid SNP discovery and genetic mapping using sequenced RAD markers. PLoS One 3: e3376.

Bedell, V.M., Y. Wang, J.M. Campbell, T.L. Poshusta, C.G. Starker *et al.*, 2012 *In vivo* genome editing using a high-efficiency TALEN system. Nature 491: 114-118.

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

Broman, K.W., and S. Sen, 2009 *A Guide to QTL Mapping with R/qtl*. Dordrecht: Springer.

Broman, K.W., H. Wu, S. Sen, and G.A. Churchill, 2003 R/qtl: QTL mapping in experimental crosses. Bioinformatics 19: 889-890.

Cermak, T., E.L. Doyle, M. Christian, L. Wang, Y. Zhang *et al.*, 2011 Efficient design and assembly of custom TALEN and other TAL effector-based constructs for DNA targeting. Nucleic Acids Res 39: e82.

Chuong, C.-M., 1998 *Molecular Basis of Epithelial Appendage Morphogenesis*. Austin, TX: R.G. Landes.

Chuong, C.M., C.Y. Yeh, T.X. Jiang, and R. Widelitz, 2013 Module based complexity formation: periodic patterning in feathers and hairs. WIREs Dev. Biol. 2: 97-112.

Day, T., J. Pritchard, and D. Schluter, 1994 A Comparison of 2 Sticklebacks. Evolution 48: 1723-1734.

Doyle, E.L., N.J. Booher, D.S. Standage, D.F. Voytas, V.P. Brendel *et al.*, 2012 TAL Effector-Nucleotide Targeter (TALE-NT) 2.0: tools for TAL effector design and target prediction. Nucleic Acids Res 40: W117-122.

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

Gaj, T., C.A. Gersbach, and C.F. Barbas, 3rd, 2013 ZFN, TALEN, and CRISPR/Cas-based methods for genome engineering. Trends Biotechnol 31: 397-405.

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 in press.

Golovinskaya, K., 1940 Pleiotropic effect of scale genes in carp. Comptes Rendus De L Academie Des Sciences De L Urss 28: 533-536.

Gross, H.P., and J.M. Anderson, 1984 Geographic variation in the gillrakers and diet of European threespine sticklebacks, *Gasterosteus aculeatus*. Copeia: 87-97.

Hagen, D.W., 1973 Inheritance of Numbers of Lateral Plates and Gill Rakers in Gasterosteus-Aculeatus. Heredity 30: 303-312.
Jones, F.C., M.G. Grabherr, Y.F. Chan, P. Russell, E. Mauceli *et al.*, 2012 The genomic basis of adaptive evolution in threespine sticklebacks. Nature 484: 55-61.
Jung, H.S., P.H. Francis-West, R.B. Widelitz, T.X. Jiang, S. Ting-Berreth *et al.*, 1998 Local inhibitory action of BMPs and their relationships with activators in feather formation: implications for periodic patterning. Dev Biol 196: 11-23.
Kahilainen, K.K., A. Siwertsson, K.O. Gjelland, R. Knudsen, T. Bohn *et al.*, 2011 The role of gill raker number variability in adaptive radiation of coregonid fish. Evolutionary Ecology 25: 573-588.
Kumar, P., S. Henikoff, and P.C. Ng, 2009 Predicting the effects of coding non-synonymous variants on protein function using the SIFT algorithm. Nat Protoc 4: 1073-1081.
Langmead, B., C. Trapnell, M. Pop, and S.L. Salzberg, 2009 Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. Genome Biol 10: R25.
Larkin, M.A., G. Blackshields, N.P. Brown, R. Chenna, P.A. McGettigan *et al.*, 2007 Clustal W and Clustal X version 2.0. Bioinformatics 23: 2947-2948.
Magnuson, J.J., and J.G. Heitz, 1971 Gill Raker Apparatus and Food Selectivity among Mackerels, Tunas, and Dolphins. Fishery Bulletin of the National Oceanic and Atmospheric Administration 69: 361-&.
Miller, C.T., A.M. Glazer, B.R. Summers, B.K. Blackman, A.R. Norman *et al.*, 2014 Modular skeletal evolution in sticklebacks is controlled by additive and clustered quantitative trait Loci. Genetics 197: 405-420.
Narum, S.R., C.A. Buerkle, J.W. Davey, M.R. Miller, and P.A. Hohenlohe, 2013 Genotyping-by-sequencing in ecological and conservation genomics. Molecular Ecology 22: 2841-2847.
Noramly, S., and B.A. Morgan, 1998 BMPs mediate lateral inhibition at successive stages in feather tract development. Development 125: 3775-3787.
Orgogozo, V., K.W. Broman, and D.L. Stern, 2006 High-resolution quantitative trait locus mapping reveals sign epistasis controlling ovariole number between two Drosophila species. Genetics 173: 197-205.
Peichel, C.L., K.S. Nereng, K.A. Ohgi, B.L. Cole, P.F. Colosimo *et al.*, 2001 The genetic architecture of divergence between threespine stickleback species. Nature 414: 901-905.
Pispa, J., and I. Thesleff, 2003 Mechanisms of ectodermal organogenesis. Dev Biol 262: 195-205.
Rohner, N., M. Bercsenyi, L. Orban, M.E. Kolanczyk, D. Linke *et al.*, 2009 Duplication of fgfr1 permits Fgf signaling to serve as a target for selection during domestication. Curr Biol 19: 1642-1647.
Rowe, H.C., S. Renaut, and A. Guggisberg, 2011 RAD in the realm of next-generation sequencing technologies. Molecular Ecology 20: 3499-3502.
Schluter, D., 1996 Adaptive radiation along genetic lines of least resistance. Evolution 50: 1766-1774.
Schluter, D., 2000 *The ecology of adaptive radiation*. Oxford: Oxford University Press.

Schneider, C.A., W.S. Rasband, and K.W. Eliceiri, 2012 NIH Image to ImageJ: 25 years of image analysis. Nat. Methods 9: 671-675.
Schuelke, M., 2000 An economic method for the fluorescent labeling of PCR fragments. Nat. Biotechnol. 18: 233-234.
Thesleff, I., 2003 Epithelial-mesenchymal signalling regulating tooth morphogenesis. J Cell Sci 116: 1647-1648.
Wells, K.L., Y. Hadad, D. Ben-Avraham, J. Hillel, A. Cahaner *et al.*, 2012 Genome-wide SNP scan of pooled DNA reveals nonsense mutation in FGF20 in the scaleless line of featherless chickens. BMC Genomics 13: 257.