RECOMBINATION RATE HETEROGENEITY COMPARTMENTALIZES GENOMIC VARIATION AND FACILITATES ADAPTIVE EVOLUTION IN THE THREESPINE STICKLEBACK

by

JOHNATHAN G. CRANDALL

A THESIS

Presented to the Department of Biology
and the Robert D. Clark Honors College
in partial fulfillment of the requirements for the degree of
Bachelor of Science

June 2017

An Abstract of the Thesis of

Johnathan G. Crandall for the degree of Bachelor of Science
in the Department of Biology to be taken June 2017

Title: Recombination Rate Heterogeneity Compartmentalizes Genomic Variation and Facilitates Adaptive Evolution in the Threespine Stickleback

Approved: _____

Dr. William A. Cresko

Adaptation to new environments is a driver of biodiversity, and often involves the coordinated action of many genomic loci that contribute to the fitness of an organism. Although adaptation has been extensively studied, an unanswered question is whether the rate or extent of adaptation depends not only upon what genes are involved but on their organization in the genome. To address this question we use the threespine stickleback (*Gasterosteus aculeatus*) as an evolutionary model system. This small fish inhabits coastal habitats in the northern hemisphere. Oceanic stickleback have repeatedly colonized new freshwater environments, resulting in rapid bouts of adaptive evolution involving parallel changes at the phenotypic and genomic levels. By investigating the structure of the stickleback genome in relation to adaptive evolution, we aim to determine factors that allow this quick and sustained adaptation to novel environments.

Gene flow between divergent populations breaks up associations between loci involved in adaptation due to the homogenizing action of meiotic recombination. Therefore, genomic architecture that isolates adaptive genomic regions may evolve. I

characterized the recombination landscape of three distinct stickleback lineages by creating genome-wide genetic maps of divergently evolved and hybrid stickleback. We compared these genetic maps with molecular and population genetic statistics to determine whether genomic patterns of divergence might be influenced by recombination rate variation. We find that recombination rate varies extensively across the stickleback genome, and importantly that a hybrid ocean-freshwater stickleback displays strikingly unique patterns of recombination and that genomic islands of divergence are inherited as compact genetic units, indicating that adaptive loci maintain their associations. Our results give insight into how non-random genomic organization can encourage rapid adaptation to novel environments.

Acknowledgements

I would like to extend my sincerest thanks to Bill Cresko for his guidance, support, and wisdom; to Thom Nelson for unparalleled mentorship, extreme patience, good humor, encouragement, many hours of editing, and for pushing me to strive for excellence in every aspect of my work; and to all members of the Cresko and Streisfeld labs for encouragement, informative discussion, useful feedback, and help with all aspects of my research, from molecular biology to statistics. I've been very lucky to have had the opportunity to be a part of such a high-achieving and fun group. I would also like to thank my family for their unwavering support in every aspect of my life, not least my studies.

Finally, I have been the beneficiary of generous funding which has allowed me to pursue my research. I would like to acknowledge the University of Oregon Department of Biology and Office of the Vice President for Research and Innovation, as well as the Graduate Evolutionary Biology and Ecology Students group for their support. I would also like to thank the Summer Program for Undergraduate Research at the University of Oregon, and in particular, Peter O'Day, for his tireless support and advocacy.

Thank you to all of the above for believing in and facilitating undergraduate research, and helping ensure excellence in the next generation of scientists.

Table of Contents

List of Figures

List of Tables

Introduction

Organisms evolve to adapt to environmental variation

Our natural world contains breathtaking biodiversity. The variation in the appearance of organisms mirrors the variability in ecological habitats. Across environments, organisms generally appear well-suited to survive given the specific circumstances of their surroundings (Darwin, 1859). This adaptation of organisms to their environments is thought to be the result of evolution by natural selection, a process first described by Darwin (Darwin, 1859). During evolution by natural selection individuals with morphological and behavioral characteristics (known collectively as phenotype) that cause them to be better suited to their environment are at a competitive advantage (Darwin, 1859; Fisher, 1930; Orr, 2004). These individuals survive and reproduce preferentially by dint of their advantageous phenotypes, and pass those beneficial characteristics to their offspring.

When a species' range encompasses more than one habitat, evolution by natural selection becomes a main driver of biodiversity (Darwin, 1859; Nosil, 2012). This is because natural selection will favor different phenotypes in different environments. Over time, populations of the same species that inhabit different environments can evolve divergently to the extent that they become separate species (the process of speciation). The evolution of more fit phenotypes in an environment is termed adaptation; intuitively, the process of natural selection will drive adaptation. Although adaptation has been the subject of intense study since the advent of the field of evolutionary biology, much of this process remains the subject of significant debate, and is the target of continued investigation (Flaxman *et al*., 2014; Lackey and

Boughman, 2016). One such issue, which is the subject of this study, is how divergent adaptation can occur on local scales, when populations adapting to very different habitats continue to interbreed. Interbreeding of individuals can lead to gene flow, or the movement of genetic variation among populations, which is understood to impede the process of adaptation by natural selection. The interplay of natural selection and gene flow, particularly as it affects patterns of genetic variation across the genome, is still not completely understood.

The genome-to-phenome map may be complex

Evolution by natural selection can only proceed if advantageous variation is also heritable (Darwin, 1859). Heritable variation in phenotype is dictated by genotype, which is the specific genetic composition of an organism. Genetic information has a physical basis in molecules of deoxyribonucleic acid (DNA). DNA is a linear molecule and stores information in sequences of the chemical residues adenine, thymine, cytosine, and guanine, which are known by their single-letter abbreviations A, T, C, and G. The sum total of all the DNA contained within a cell, which often includes multiple molecules of DNA known as chromosomes, comprises the genome. While loci on the same chromosome are therefore physically linked, different chromosomes are independent of each other. In the traditional Mendelian model, one gene (chromosomal locus) controls one trait, and each gene (and therefore phenotype) is inherited independently of all others (Mendel, 1865; Castle, 1903). However, phenotypes are often influenced by variation at many genomic loci, which may reside on the same or different chromosomes (e.g. Lango Allen *et al*., 2010; Miller *et al*., 2014); this is known as a polygenic basis. This presents a problem for natural selection. If all alleles

contributing to a phenotype are inherited independently, natural selection must act on many dozens, hundreds, or potentially even thousands of loci to influence variation in just one adaptive phenotype.

The spatial distribution of loci within the genome, known as the genomic architecture, is therefore highly important: if multiple alleles (variations in DNA sequence at a locus) contribute to an adaptive phenotype but are all located on different chromosomes, the likelihood that all will be inherited together is low. In contrast, if all are arranged close together on the same chromosome, they are far more likely to be co-inherited between generations. Clustered alleles that are frequently co-inherited act more like simple Mendelian loci: they are essentially one chromosomal locus imparting one phenotype, and are passed down through generations together. Because natural selection acts on phenotype, it can work far more efficiently on sets of adaptive alleles that are tightly clustered, as these present one target for natural selection, rather than many targets dispersed throughout the genome.

Recombination is an important process for adaptation

Multicellular organisms such as mammals, birds, and fish are diploid and therefore contain two copies of each piece of DNA, one obtained from the mother and the other from the father. Copies of the same chromosome are termed ‘homologous,’ and can vary in sequence between the two copies. Although homologous chromosomes are copies of the same strand of DNA, they are independent molecules and not physically linked. A locus with a different version of a gene on each copy of DNA is termed heterozygous, while one with the same genetic code on each homologous chromosome is homozygous. In diploid organisms, the physical linkage between loci on

the same strand of DNA can be broken by recombination, which occurs between homologous chromosomes during meiosis (gamete formation) (Figure 1).

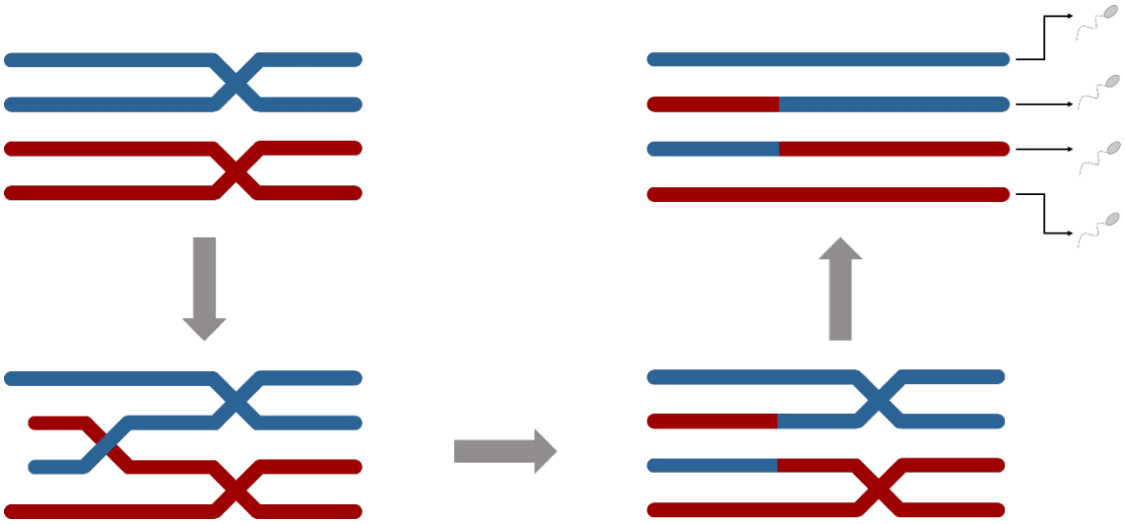

Figure 1: Meiotic recombination (simplified).

In diploid organisms, recombination, or 'crossing over,' occurs between homologous duplicated chromosomes during meiosis. In the above figure, the different colors of the two duplicated chromosomes indicate identifiable differences in sequence (i.e. polymorphisms). Clockwise, from top left: 1. Duplicated chromosomes align during meiosis. 2. Homologous sister chromatids physically cross over in preparation for recombination. 3. Recombination breakpoints are established and resolved in the chromosome arm, and the crossed over portions are exchanged. 4. Duplicated chromatids separate and meiosis completes with one haploid copy of the chromosome packaged in each gamete (sperm cells pictured). One crossover event between homologous chromosomes results in two recombinant copies, with potential new genotype combinations, and two original copies.

Recombination is required for proper chromosomal disjunction during meiosis, and so a lack of recombination across an entire chromosome in an individual can result in inviable offspring (Hassold and Hunt, 2001; Baker *et al.*, 1976; Mather, 1938). In addition, recombination generates variation in natural populations (Hurles, 2005; Noor, 2008), in part by homogenizing existing genetic variation to create novel combinations

of genotypes. Sexually-reproducing organisms benefit from this shuffling of genetic information, as new genotype combinations created by recombination may impart a novel phenotype that is even better-suited to its environment. This process is especially important during adaptation, when genetic variation is needed to establish genotypes suited to a new or changing habitat. The genetic variation produced by recombination during sexual reproduction is beneficial, or even crucial, to organisms adapting in response to selective pressures in their environment (Morran *et al.*, 2011).

However, recombination can also inhibit adaptation. Where environments differ, the organisms that inhabit them evolve divergently. Once adaptive genotypes have become established in divergent populations of the same species, gene flow (the transfer of genetic information resulting from migration and mating) between populations can significantly retard sustained adaptation (Lenormand, 2002). In this case, recombination in hybrid individuals resulting from interbreeding between the two populations yields strands of DNA containing mixtures of genotypes suited to either environment (Nosil, 2012; Martin *et al.*, 2013; Kirkpatrick and Barton, 2006). Recombination therefore has significant implications for the process of adaptation, especially when adaptation proceeds on local scales with no physical barriers to gene flow between divergent populations. Because divergent adaptation often occurs in the context of gene flow between populations, the maladaptive influence of recombination is a substantial barrier to natural selection (Lenormand, 2002; Andrew and Rieseberg, 2013). This presents an intriguing problem, then: how can divergent adaptation succeed in the face of gene flow?

How does recombination correspond to divergence?

The spatial organization of genomic loci can be measured in two different yet complementary ways: physical and genetic distance. Physical distance, the more intuitive of the two, describes the difference between loci with regards to the structure of DNA in terms of total number of intervening base pairs. Loci with a greater physical distance between them are separated by more DNA; i.e., they are farther apart on a chromosome. Genetic distance, on the other hand, measures the amount of recombination between two loci by observing how often alleles at the two loci are inherited together. Loci that are farther apart in physical distance are more likely to have recombination occur between them; that is, they are less likely to be inherited together. For this reason, genetic distance between two loci is often used as a proxy for physical distance on a chromosome.

The relationship between physical and genetic distance is not one-to-one, however. Recombination can be biased to 'hotspots' across the genome, where the majority of crossover events take place (Lange *et al.*, 2016). In contrast, some genomic regions experience far less recombination across the same physical distance. Evidence from natural populations has indicated that such local variation in recombination rate may be vital for adaptation in the face of gene flow (Via, 2012; Nachman and Payseur, 2012; Burri *et al.*, 2015; Flaxman *et al.*, 2014). Genomic regions with low recombination rate will tend to resist breakup of adaptive allelic combinations. Theory predicts that, where divergent populations experience gene flow, natural selection could favor mechanisms that prevent recombination across groups of adaptive alleles (Kirkpatrick and Barton, 2006; Ortiz-Barrientos *et al.*, 2016).

The evolution of low recombination rates can be achieved through different means (Ortiz-Barrientos *et al.,* 2016). One specific and widespread example is that of chromosomal inversions (Dobzhansky and Epling, 1948; Glazer *et al*., 2015; Bansal *et al.*, 2007; Corbett-Detig and Hartl, 2012). Just as chromosomes can have variation in DNA sequence within or between individuals, so too can they vary in structure. An inversion is a mutational event where a portion of a chromosome is physically reversed relative to the standard orientation. Chromosomal inversions are common in nature, and different orientations of the same chromosomal region can persist in populations (Bansal *et al.*, 2007; Corbett-Detig and Hartl, 2012; Jones *et al.*, 2012; Kirkpatrick and Barton, 2006). When an individual is heterozygous for an inversion on a pair of homologous chromosomes, recombination is suppressed across the inverted region through a complex mechanical process (Griffiths *et al.*, 2000; Dobzhansky and Epling, 1948; Glazer *et al*., 2015). Chromosomal inversions can therefore have adaptive importance (e.g. Lee *et al.*, 2017; Lee *et al.*, 2016; Lowry and Willis, 2010). If the inverted region contains adaptive alleles in divergent populations, and a different chromosomal orientation exists in each population, then gene flow and recombination cannot homogenize variation within the inversion (Kirkpatrick and Barton, 2006). Investigating the interplay between patterns of recombination and patterns of divergence therefore presents an opportunity to further our understanding of mechanisms by which adaptation can be maintained in the presence of gene flow. If natural populations exhibit a reduction of recombination across divergent genomic regions, this could indicate a genomic mechanism by which adaptive variation is preserved within and between divergent populations.

The threespine stickleback is a natural model of adaptive evolution

A natural system for the study of adaptive evolution is the threespine stickleback (*Gasterosteus aculeatus)*. This small fish is distributed across coastal regions of the northern hemisphere, including the Pacific coast of Oregon and Alaska (Bell and Foster, 1994). Oceanic stickleback populations have repeatedly colonized and adapted to freshwater environments, giving rise to numerous freshwater populations (Bell and Foster, 1994; McKinnon and Rundle, 2002; Hohenlohe *et al.*, 2010; Lescak *et al.*, 2015; Colosimo *et al.*, 2004; Jones *et al.*, 2012). In some cases, this evolution proceeds on timescales of decades, rather than centuries or millennia (Lescak *et al.*, 2015). Independent oceanic-to-freshwater transitions are striking in their parallelism of phenotypic evolution: derived freshwater populations consistently display similar changes in phenotype, brought about by specific selective pressures of freshwater environments (Figure 2) (Bell and Foster, 1994). Correspondingly, the same genomic regions have been implicated in facilitating adaptation across many derived freshwater populations: comparisons between different freshwater-oceanic population pairs show that similar genomic regions tend to be differentiated in independent freshwater populations (Hohenlohe *et al.*, 2010).

Evidence from natural stickleback populations has shown that certain genomic regions are more divergent than others between freshwater and oceanic populations (Hohenlohe *et al*., 2010; Jones *et al.*, 2012). A common method for identifying genomic locations influenced by natural selection involves sampling variation in DNA sequence across the genome, and calculating the differences between populations at the allelic level. Genomic regions that display elevated levels of divergence relative to the

background may be involved in adaptation to each environment, as natural selection has caused different alleles to be favored in each population. Such ‘genome scans’ of divergence between many ancestral oceanic and derived freshwater stickleback population pairs uncover heterogenous levels of differentiation across the genome (Figure 3), with similar regions involved in many adaptive ocean-to-freshwater transitions (Hohenlohe *et al.*, 2010; Jones *et al.*, 2012). These findings indicate that alleles important for adaptation may be spatially restricted to these regions, rather than distributed at random throughout the genome, and invite detailed investigation of the genomic processes shaping patterns of divergence (Hohenlohe *et al*., 2012).

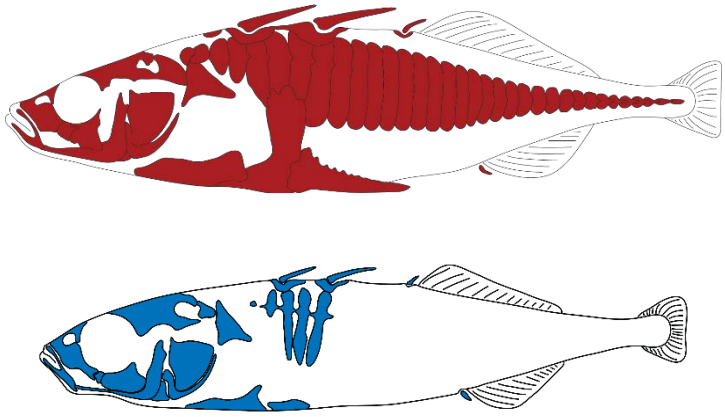

Figure 2: Typical morphologies of marine and freshwater threespine stickleback.

Line drawings (adapted from M. Currey) of wild-caught stickleback from the two populations used in this study, representative of typical ancestral oceanic (top) and derived freshwater (bottom) morphologies. Bony structures related to feeding and predator defense are colored to highlight the differences in phenotype.

The stickleback system can be used to study the interplay of recombination and adaptation

In many cases, freshwater and oceanic stickleback population pairs have no physical barriers to prevent gene flow between divergent populations. Despite this connectivity, freshwater and oceanic populations retain high divergence in key genomic regions (Hohenlohe *et al.*, 2010; Jones *et al.*, 2012). The system is therefore ideally suited for studying the influence of recombination rate variation on local adaptation in the presence of gene flow. With large regions of the stickleback genome implicated in the adaptation to freshwater environments, how are specific adaptive combinations of alleles (known as haplotypes) maintained? One peak of high differentiation between freshwater and oceanic stickleback is co-localized with a region of chromosome 21 that is inverted between freshwater and oceanic populations (Jones *et al*., 2012; Glazer *et al.*, 2015); however, this inversion accounts for only a small portion of the genomic divergence observed (Figure 3). This observation invites the question: do more broad-scale reductions in recombination play a role in facilitating and maintaining divergence between stickleback populations in the face of gene flow?

Previous work has indicated broad-scale variation in recombination across the stickleback genome, suggesting that regions of divergence between freshwater and oceanic populations may reside in low-recombination rate genomic areas (Hohenlohe *et al.*, 2012; Roesti *et al.*, 2013; Glazer *et al*., 2015). Here, we expand upon these studies by analyzing a previously-documented freshwater-oceanic population pair in Alaska's Cook Inlet (Hohenlohe *et al*., 2010). We combine a fine-scale characterization of recombination across three different stickleback genetic backgrounds (freshwater,

ocean, and F_1 freshwater-ocean hybrid) with population genetic sampling to assess patterns of freshwater-oceanic divergence across the genome. In doing so, we assess whether recombination between and within populations serves as a facilitator for maintaining genomic divergence, or whether it presents a barrier to adaptation. By investigating mechanisms by which divergence can proceed in the context of gene flow, we provide new insight into the maintenance of adaptive genomic variation. With this, we gain an increased understanding of the complex processes driving the origin of species and the generation of biodiversity.

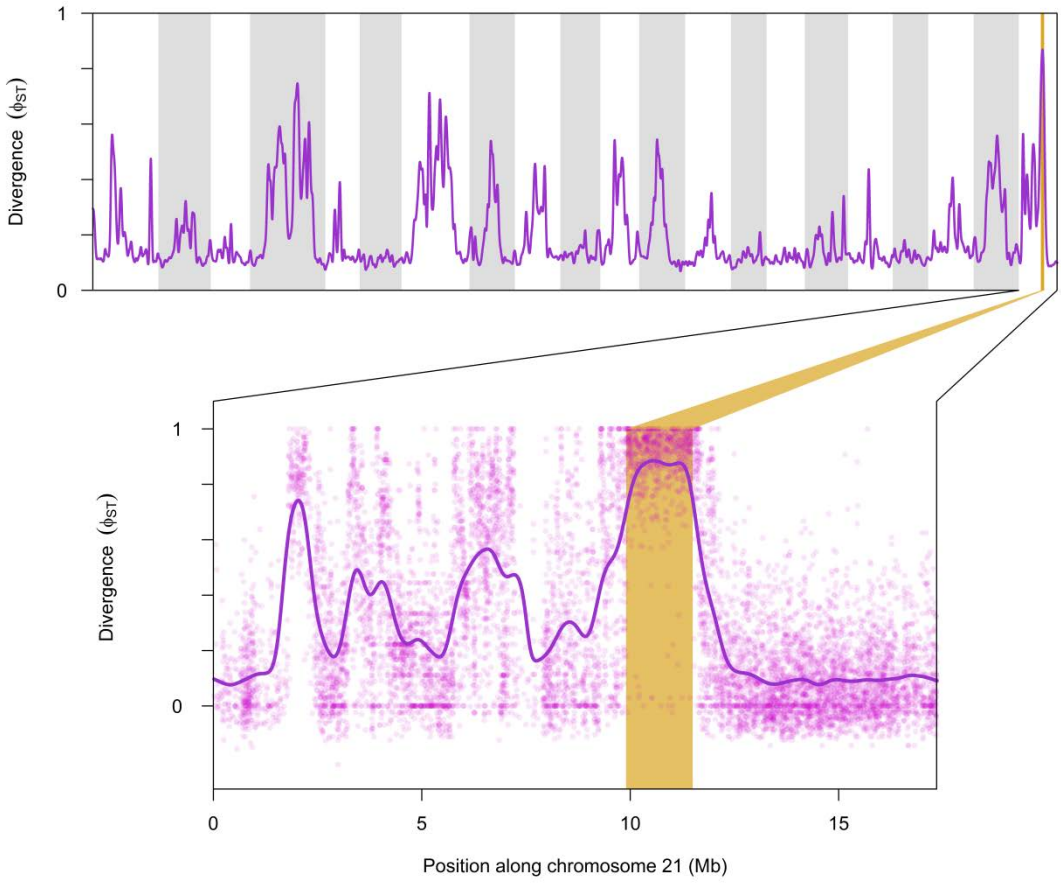

Figure 3: Differentiation across the stickleback genome is heterogeneous.

Top panel: a genome scan of divergence between the freshwater-oceanic population pair used in this study. A quantification of genetic divergence between populations (Φ_{ST}) is plotted against the stickleback genome (alternating bands of color demarcate chromosomes). $\Phi_{ST} = 0$ indicates that the populations are genetically identical at the tested locus; $\Phi_{ST} = 1$ indicates that the populations are completely genetically distinct. Divergence is highly heterogenous across the stickleback genome, with 'islands' of divergence clustered on some chromosomes. Note in particular the breadth of the divergence peaks on chromosomes 4 and 7 (two of the largest chromosomes in the genome).

Bottom panel: One of the highest divergence peaks of the stickleback genome co-localizes with a chromosomal region that is inverted in orientation between some freshwater and oceanic populations (highlighted in yellow). Pink dots indicate individual Φ_{ST} estimates plotted against chromosomal position; a kernel-smoothed regression (bandwidth = 1 Mb) is plotted as a line.

Experimental design

Genetic mapping

To characterize the genomic patterns of recombination in stickleback, this study uses genetic mapping. Genetic mapping detects recombination using markers on a piece of DNA. At its genesis, genetic mapping used as markers genotypes that would impart distinguishable phenotypes, so that the presence or absence of a marker in an individual could be visually assessed (Morgan, 1911; Sturtevant, 1913). Today, we use variations in DNA sequence as markers (e.g. Botstein *et al.*, 1980; Roesti *et al*., 2013; Glazer *et al.*, 2015; Rastas *et al*., 2016). When an individual is heterozygous at a genomic locus, it can serve as a genetic marker because we can track which allele is co-inherited with alleles from other heterozygous sites along the same chromosome. For example, take a diploid individual with two heterozygous loci on a chromosome (Figure 4). On one copy of the chromosome, the individual has alleles A – B; on the other copy, alleles a – b. Because loci A and B are heterozygous, the different alleles (uppercase and lowercase) are distinct and can be identified. When this individual reproduces, only one copy of the chromosome will be passed on to the next generation; the offspring will inherit either alleles A – B or alleles a – b. However, if during meiosis a recombination event occurs between locus A and locus B, the resultant DNA copies that could be inherited by the progeny would be A – b and a – B. By sequencing the DNA of both parents and progeny of a genetic cross, it is thus possible to identify markers in the parent, find the same markers in the many offspring, and determine the frequency with which markers are inherited together. The probability of two markers not being co-inherited (i.e., a recombination event occurring between the two) is termed “genetic

distance," and is measured in centimorgans (cM), where 1 cM is equal to a 1% probability of a recombination event in a single generation. For markers 1 cM apart, a recombination is observed in one out of every one hundred offspring. Each progeny of a genetic cross represents an individual meiosis from each parent, and carries within its genetic information evidence of many recombination events in the parent. In genetic mapping, therefore, the genotypes of the progeny are examined in order to map (assess the recombination in) the genomes of the parents. Because of this, genetic maps are inherently limited by the number of progeny used in constructing the map.

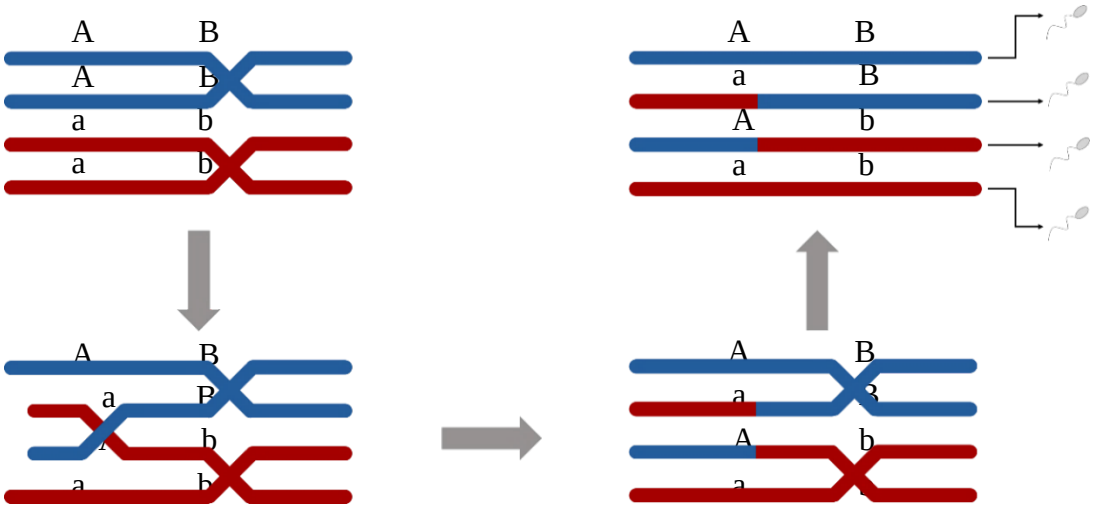

Figure 4: Mapping of genetic markers.

Recombination schematic as in Figure 2, illustrating how heterozygous genetic markers can be tracked from parents through to progeny to assess recombination between genomic loci. Here, the individual is heterozygous for alleles at two loci (A/a and B/b).

RAD-seq

Genetic mapping requires a suite of well-distributed genetic markers which can be used to finely localize recombination events. However, whole-genome sequencing of large numbers of individuals is often cost-prohibitive. In this project, I make use of a

reduced-representation genome sampling technique known as restriction site-associated DNA sequencing, or RAD-seq. This technique samples small portions of DNA sequence at random locations throughout the genome, to generate datasets amenable for genetic mapping, population genetics, or other analyses. First developed at the University of Oregon (Baird *et al.*, 2008), RAD-seq makes use of restriction enzymes, which physically cut DNA at specific sequences that are distributed more or less randomly throughout the genome. Genomic DNA from members of a population or genetic cross is treated with a restriction enzyme, which cuts at the same sites across the genomes of each individual. Through a series of molecular biological reactions, modifications are made to the extracted DNA from each individual, including the addition of synthetic DNA strands containing unique barcodes for identification. With the DNA from each individual thus tagged, all samples are multiplexed and sequenced together on a next-generation sequencing platform (i.e. Illumina HiSeq 2500). This process results in short sequenced reads of DNA extending outward in either direction from each restriction enzyme cut site. RAD-seq yields thousands of markers at the same loci across the genomes of many individuals, which can be used for a variety of purposes. Here, I use RAD-seq both for population genetic analyses and genetic mapping.

Using the Stacks analysis pipeline (Catchen *et al.*, 2013), short sequencing reads from RAD-seq are compared across individuals to identify single nucleotide polymorphisms (SNPs), heterozygous sites shared between individuals. A SNP (or number of SNPs) at a given locus can be used as a mappable genetic marker: when a parent is heterozygous for a marker, recombination events can be inferred by examining

the alleles inherited from the parents by the cross progeny. Observed recombination events between markers can be translated into genetic distance (cM) using Kosambi's mapping algorithm (Kosambi, 1944).

Mapping crosses

This project makes use of a hybrid backcross and two within-line crosses to generate mapping families (Figure 5). This approach allows characterization of and comparison between the recombination landscapes in three distinct genetic backgrounds (freshwater, oceanic, and hybrid). Freshwater and oceanic stickleback lines are kept at the University of Oregon's stickleback facility. Previously, freshwater fish from Boot Lake, Alaska, and oceanic fish from Rabbit Slough, Alaska had been crossed to create lines of F_1 hybrids. I backcrossed a female hybrid stickleback with a Boot Lake male stickleback to create backcross progeny. Mapping families from crosses between Boot Lake fish and between Rabbit Slough fish were likewise previously generated.

Population genetics

To investigate the genomic patterns of freshwater-oceanic divergence in stickleback, I make use of a population genetic dataset generated by T.C. Nelson. Five wild fish were sampled from each of the two populations in this study (Rabbit Slough and Boot Lake) and sequenced using a RAD-seq protocol designed to generate dense genomic sampling. At each of these loci, population genetic statistics quantifying divergence between and polymorphism within populations could then be calculated. This approach allows a direct assessment of the relationship between genomic divergence and recombination patterns within and between these populations.

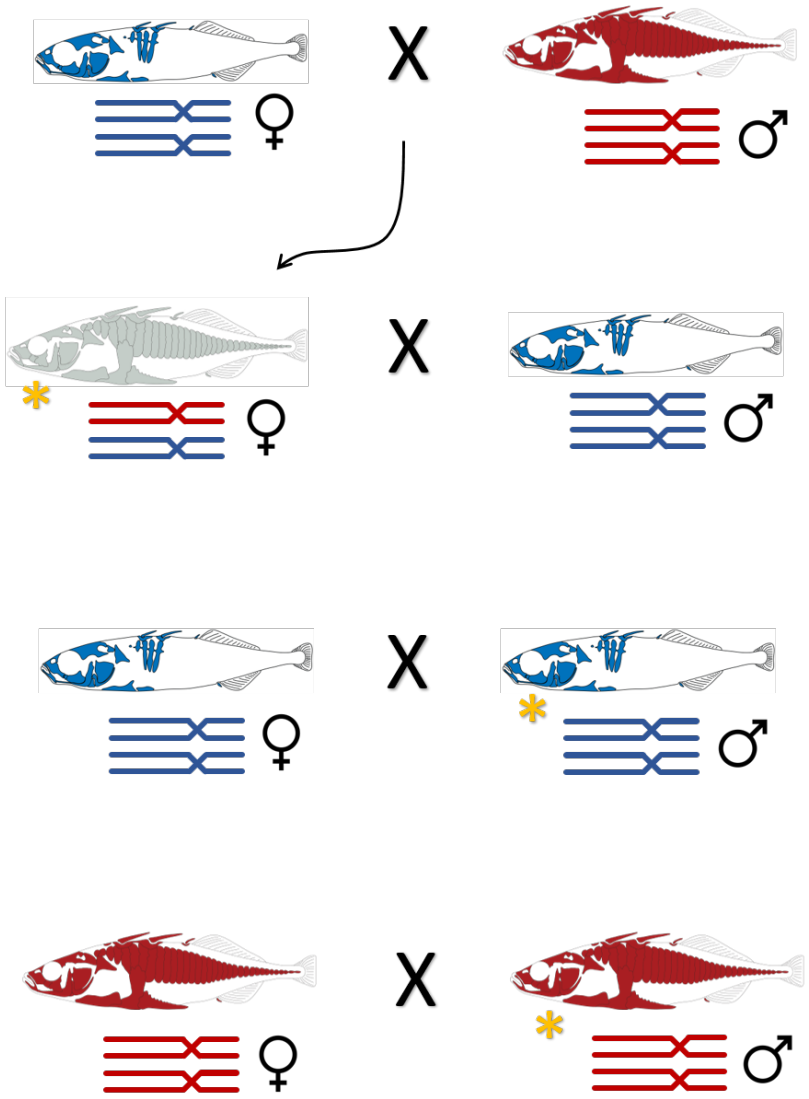

Figure 5: Cross scheme for genetic mapping families.

Top row: A Boot Lake female fish (blue) was crossed with a Rabbit Slough male fish (red) to create a generation of F_1 hybrids (grey, second row). For genetic mapping, three crosses were made: a hybrid-Boot Lake backcross (second row), and within-lineage Boot Lake and Rabbit Slough crosses (third and fourth rows). Asterisks indicate which fish's genome from each cross was mapped in this project.

Methods

Fish rearing and crosses

Pure freshwater (Boot Lake) and oceanic (Rabbit Slough) lines were maintained in the Stickleback Facility at the University of Oregon. All work with stickleback complied with University of Oregon IACUC approved protocols. In 2014, a Boot Lake female was crossed with a Rabbit Slough male to create a line of F_1 hybrids. The hybrid line was transitioned to a long-daylight facility to simulate summer at ~9 months of age, to induce sexual maturation. I performed a series of within- and between-lineage crosses for the purposes of genetic mapping and viability scoring. All crosses for this study were performed *in vitro*: eggs were collected from the female fish, then both parents were euthanized in MESAB solution (300 mg/L Tricaine-S [Sigma-Aldrich], 1 M Tris-HCl pH=9 (Promega), buffered to pH=7 in tank water). Fin clips were taken from each parent for DNA extraction, and testes were harvested from the male by dissection; both soma were fixed for storage in 95% ethanol to allow for future DNA extraction, if necessary. Fin clips were likewise stored in 95% ethanol, at -4°C. Extracted testes were suspended in Stickleback Embryo Medium ("SEM"; 4 mg/mL Instant Ocean Aquarium Sea Salt Mixture, 0.15 mg/mL baking soda, NANOpure water ["npH_2O"; from Barnstead NANOpure Infinity, TOC <5ppb, resistivity ≥18.2 MΩ-cm], pH=7.5) and homogenized using a razor. The resulting sperm suspension was added dropwise over a clutch of eggs harvested from the female fish, and incubated at 25°C to facilitate fertilization. After several hours, eggs were individualized, cleaned, and inspected for fertilization. Fertilized eggs were incubated at 25°C in SEM for 12 days, with daily medium changes and discarding of unfertilized eggs and dead embryos.

Progeny of the hybrid backcross were euthanized at 12 days post-fertilization in MESAB solution and stored whole in 95% ethanol at -4°C. In 2011 and 2012, respectively, crosses were made from pure Rabbit Slough and Boot Lake lineages for the purpose of genetic mapping, with the parents and progeny processed similarly to as described.

DNA extractions

I extracted DNA from the parents and progeny of a cross for genetic mapping. Immediately prior to DNA extraction, individual larval stickleback and fin clips from the cross parents were rehydrated in nanopure water (npH_2O). Larvae were incubated at 55°C overnight, or for 4 hours with regular mixing, in 200 µL lysis solution (20 µL Proteinase K [Qiagen], 180 µL Buffer ATL [Qiagen]) until completely broken down. I extracted DNA from the lysis solutions per the manufacturer's instructions following precipitation with an equal volume of PEG solution (0.2 g/mL PEG-8000 [Sigma-Aldrich], 2.5 M NaCl [Fisher Scientific], npH_2O) and 20 µL DNA-binding beads (Thermo Scientific Sera-mag Speed Beads in 0.18 g/mL PEG-8000, 1 M NaCl, 10 mM Tris-HCl pH=8, 1 mM EDTA pH=8 [Sigma-Aldrich], npH_2O, 5.5x10-4% b.v. Tween 20 [Sigma-Aldrich]), using a Life Technologies DynaMag-96 side magnetic plate and two washes with fresh 80% ethanol (Sigma-Aldrich). DNA-binding and -eluting incubations were performed at 42°C to prevent flocculation of solutes. Samples were eluted in 75 µL Buffer EB (Qiagen) and treated with RNAse A (Thermo Scientific) to destroy residual RNA. Final DNA concentration of each sample was checked with an Invitrogen Quibit 2.0 fluorometer as per the manufacturer's instructions prior to RAD library construction.

RAD library construction and sequencing

I constructed two restriction site-associated DNA (RAD) libraries from progeny of the hybrid backcross, based on a modification of the protocol from Baird *et al*. (2008). ~300-500 ng DNA from each individual, including the parents of the cross, was normalized to standard volume in a multi-well plate and digested with restriction endonuclease SbfI-HF (New England Biolabs) per the manufacturer's instructions, in a total reaction volume of 50 or 60 µL containing enzyme, genomic DNA, Buffer EB, and NEB Buffer 4 (New England Biolabs). Restriction-site overhang-specific P1 adapters (modified Solexa adapter [Illumina]) containing unique, six-nucleotide barcodes (differing by at least three nucleotides) were ligated to the DNA from each individual in a 60 or 70 µL total reaction volume, containing inactivated SbfI digestions, NEB Buffer 2 (New England Biolabs), P1 adapter to final concentration of 4.2 nM, rATP (Promega) to final concentration of 1 nM, T4 DNA Ligase (Epicentre), and npH_2O. Individual barcoded samples were multiplexed together in equal DNA amounts (with the exception of the cross parents, which were included at double concentration to facilitate greater depth of sequencing) to a total DNA content of 1.6 ng. Multiplexes were sheared in a Bioruptor sonicator (Diagenode) in 100 µL aliquots 10 times for 30 seconds with 1.5 minutes rest between rounds, then recombined and purified using DNA-binding beads.

Sheared multiplexes were run on 1.25% agarose (Sigma-Aldrich), 0.5X TBE gels alongside GeneRuler 100bp DNA Ladder (Thermo Scientific), and size-selected for a 400-600bp range by excising the corresponding section of the gel containing the multiplex. Each multiplex was extracted from the gel using a Zymoclean Gel DNA recovery kit (Zymo Research) per the manufacturer's instructions (with one

modification: to improve representation of A- and T-rich sequences, gel slices were melted in the supplied buffer at room temperature on a rocker), and eluted in 20 µL Buffer EB. 5’ and 3’ overhangs created by shearing were converted to phosphorylated blunt ends to facilitate adapter ligation: gel-recovered and purified sample were incubated at room temperature for 30 minutes with 10X Blunting Buffer (New England Biolabs), dNTP mix (Thermo Scientific) to final concentration of 0.1 mM, and Blunt Enzyme Mix (New England Biolabs), in a final reaction volume of 26 µL. After purification using DNA-binding beads, sample will be eluted in 42 µL Buffer EB. To prepare DNA fragments for ligation to the P2 adapter, a 3’-dA overhang was added using Klenow 3’ to 5’ exo- polymerase. Eluate from the previous step was combined with 10X NEB Buffer 2, dATP (Fermentas) to final concentration of 0.2 mM, and Klenow (Epicentre) in a final reaction volume of 51 µL, and incubated at 37°C for 30 minutes before slowly cooling to room temperature.

Reaction product was purified using DNA-binding beads, and eluted in 44 µL Buffer EB in preparation for P2 adapter ligation. P2 adapters, containing a 3’-dT overhang, were ligated to the previously-created 3’-dA overhangs on DNA fragments by incubation of the previous reaction product with P2 adapter (modified Solexa adapter [Illumina]) in a final concentration of 0.2 mM, rATP in a final concentration of 2 mM, and T4 DNA Ligase, in a final reaction volume of 51 µL. Following a 1-hour incubation at room temperature, reactions were incubated at 4°C overnight, then purified using DNA-binding beads and eluted in 52 µL Buffer EB. After ligation of the P2 adapter to the DNA fragments in the multiplexed libraries, a test PCR amplification was performed using forward and reverse primers complementary to the P1 and P2 adapters,

so as to amplify only those fragments containing both adapters, and prepare them for hybridization to the flow cells of the Illumina sequencer. To check for robust amplification of the library, test PCRs were performed using 2.6-5.3 ng of RAD library as template, with Phusion HF buffer (Thermo Scientific), dNTP mix in a final concentration of 0.2 mM, 0.4 µM forward and reverse primer mix (modified Solexa Amplification primer mix [Illumina]), Phusion HF DNA Polymerase (Thermo Scientific), and npH_2O in a 25 µL total reaction volume. PCR reactions were subjected to 18 cycles of amplification (10 sec 98°C, 30 sec 65°C, 30 sec 72°C) and purified with DNA-binding beads.

The entire cleaned reaction was run on a 1% agarose, 0.5X TBE gel alongside GeneRuler 100bp ladder and unamplified library template in an equal volume to that amplified in the PCR reaction. Following confirmation of robust library amplification, another PCR was performed on the entire library, using 12 cycles of amplification so as to minimize bias in the proportions of amplified sequence. Reactions were purified using DNA-binding beads and eluted in 20 µL Buffer EB. A final size-selection was performed to eliminate free adapters and P1 dimer contaminants: cleaned PCR amplifications of RAD libraries were run on a 1.25% agarose, 0.5X TBE gel between lanes of 100bp ladder. The portion of the gel spanning a range of 385-585bp in the RAD library was excised with a razor, and DNA therefrom was recovered using the Zymoclean Gel DNA recovery kit as before. Gel extractions were eluted in 20 µL Buffer EB, and DNA concentration of each library was quantified a final time using a Quibit fluorometer. Sequencing was performed by the University of Oregon Genomics and Cell Characterization Core Facility on an Illumina HiSeq 2500 Ultra-High-

Throughput Sequencing System, using the single-read 100 setting. RAD-seq libraries from both the Rabbit Slough and Boot Lake incrosses were constructed previously, similar to as described, in 2011 and 2012, respectively.

Sequence processing and marker assembly

I analyzed raw sequences from all three crosses returned from the Genomics and Cell Characterization Core Facility using the Stacks 1.41 software pipeline (Catchen *et al.*, 2013). First, raw data was fed through the `process_radtags` module to de-multiplex raw reads, and reads were discarded which had been marked by Illumina's internal chastity filter as failing (`--filter_illumina`). Reads were only retained if they contained an intact barcode and restriction enzyme cutsite, and had average Illumina quality scores of 90% or higher within a 15bp-window slid along the read. De-multiplexed reads from each parent or cross progeny, identified by `process_radtags` using the unique barcode, were aligned to the stickleback reference genome (Jones *et al.*, 2012; Glazer *et al.*, 2015) using the GSnap Genomic Short-read Nucleotide Alignment Program (Wu and Watanbe, 2005) with the default settings. As a further quality filter, only those reads which were uniquely aligned were converted to the `.bam` format using SAMtools (Li *et al.*, 2009) for use in the rest of the Stacks pipeline. Each mapping family had a different number of progeny; the two larger datasets (hybrid and Boot Lake crosses) were down-sampled to achieve parity with the smallest (Rabbit Slough cross, n = 94 progeny). For the two larger families, the 94 progeny with the greatest depth of coverage were manually identified and retained for use in the pipeline. For each cross, the `ref_map.pl` module (which executes all of the components of the Stacks pipeline required to construct a reference genome-aligned

genetic map) was called using a chi square significance level of 0.1 to call heterozygous and homozygous genotypes in each individual (`--alpha 0.1`), and a minimum number of five identical raw reads was required to assemble a stack (`-m 5`). Each cross was processed as type CP (cross-pollination), in order to recognize all allelic segregation patterns. Finally, the `genotypes` program was called separately, to export genotypes and haplotypes at each locus. Within `genotypes`, automated corrections were enabled (`-c`); a minimum read depth of five in each individual to export a locus (`-m 5`) was required, as was the presence of a marker in at least half the mapping progeny for printing (`-r 47`). Output was written in the `.loc` format (`-o joinmap`) as cross type CP.

Genetic map construction

I constructed genetic maps for the entire genome of one parent fish of each cross using the Lep-MAP2 Linkage Map Construction Suite (Rastas *et al.*, 2016). As all markers were reference genome-aligned, a custom grouping map was constructed to assign markers to the chromosomes they mapped to on the stickleback reference genome. Each genotype locus file from the Stacks pipeline was converted to LINKAGE format (Lathrop et al., 1984), using the Lep-MAP2 `loc2linkage` script. To investigate recombination rate heterogeneity, a fixed marker order for each chromosome was constructed using the reference genome positions of the markers. Module `OrderMarkers` was called for each chromosome individually to evaluate fixed orders without making improvements to the map order (`improveOrder=0`), so as to estimate genetic distances in Kosambi centimorgans (`useKosambi=1`) between markers. Maps were produced at two estimated error rates, 0% and 1 (`minError=0,`

`0.01`), as per the developer's recommendation (based on the observation that for datasets with far more markers than progeny, map distances can become inflated due to genotyping errors). Each map produced was the product of 1,000,000 replications merged together (`numMergeIterations=1000000`), with duplicate markers retained in the computation (`removeDuplicates=0`); to increase the accuracy of genetic distance estimation, markers were only included in each computation if they were informative in the parent being mapped (`informativeMask=13, 23)`.

Despite filtering in the Stacks pipeline and an underlying hidden Markov model implemented in Lep-MAP2 to identify errors, some markers still showed signatures of genotyping error, inserting large, spurious distances in the genetic maps. To correct for this, each map was manually curated. Markers were dropped from the map one-by-one (`removeMarkers=`…) provided that a) they contributed 10% or more of the entire map distance of the chromosome, and b) their removal caused the genetic distance between the two markers most adjacent to the dropped marker to collapse (if the large genetic distance represented true recombination events, then dropping the marker would have no effect on the genetic positions of the two adjacent markers, as the recombination event would result in increased genetic distance for all markers beyond the recombination breakpoint; if marker removal did not result in a collapse of map distance, the marker was added back to the genetic map). Although this method required greater user input, it resulted in higher-confidence, higher-quality linkage maps.

Recombination rate analyses

For each mapped chromosome, I performed a series of calculations to investigate fine-scale recombination. All calculations were performed by custom scripts written in R (The R Project for Statistical Computing). For each pair of adjacent markers, I calculated the difference in cumulative genetic distance (ΔcM) as well as the difference in physical position (bp), to produce the standard metric centimorgans per megabase (cM/Mb), aligned against the physical midpoint of the marker pair. This yielded observations of recombination rate at thousands of points across the genome (N-21 data points per genome, where N = number of markers in final map). All calculations were performed using genetic distances from maps constructed assuming a 1% error rate in the datasets, to help guard against inflation of map distances by genotyping error, while simultaneously avoiding false suppression of legitimate recombinations (with one exception: on the heterogametic chromosome [19] of the male Boot Lake fish, the map was constructed using a 15% assumed error rate; on the heterogametic chromosome, sequence degeneration relative to the reference genome can result in inaccuracies in marker calling [e.g. Roesti *et al.*, 2012]). In order to further account for genotyping error without affecting global genetic distance estimates, I took two quality filtering steps in the calculation of recombination rate values: first, marker pairs which were exactly three basepairs apart on the reference genome were normalized to a cM/Mb value of 0. This was due to the supported inference that ostensibly-segregating polymorphisms three basepairs apart represent, in reality, sequencing or genotyping errors in the restriction enzyme cut site. Second, I created binned distributions of recombination rate values in 1 cM/Mb intervals for the entire

genome of each fish. Distributions were heavily weighted towards the low-recombination rate end, with a diminishingly small subset of large cM/Mb values contributing to a long tail. Taking the conservative assumption that these exceedingly high values represented further errors, the top 1% of recombination rate values were eliminated from each genome-wide dataset, resulting in further removal of false recombination events due to genotyping errors not recognized by the mapping program. Because a conservative approach was taken to filtering, maps and values may be sensitive to inflation due to genotyping error; however, it is unlikely that legitimate recombinations were excluded by this approach. To create smoothed values of recombination rate, I implemented Nadaraya-Watson kernel regression in R using the `ksmooth` function, with various smoothing bandwidths. For the estimation of recombination rates within genomic windows, I calculated the difference in physical and genetic position of the most-peripheral markers in each non-overlapping window slid along the length of each individual chromosome.

Sampling of wild populations

Five wild stickleback were collected each from Boot Lake, AK, and Rabbit Slough, AK. Live fish were immediately euthanized with MESAB as described previously, and fixed in 95% ethanol for transport. 500 ng DNA was extracted from each fish and incorporated into high-density RAD-seq libraries by T. C. Nelson. Libraries were sequenced by the University of Oregon Genomics and Cell Characterization Core facility using the paired-end 250 setting on an Illumina HiSeq 2500.

Population genetic statistics

Population genetic statistics were calculated from the Boot Lake and Rabbit Slough RAD-seq data both by hand and by the `populations` module of the Stacks pipeline. Raw reads were processed with the Stacks pipeline as described previously, then aligned to the reference genome using BBMap (Bushnell, 2014) with the most sensitive settings. `Populations` was executed requiring a marker to be genotyped in 80% of the sampled fish and present in both populations for export (`-r 0.8, -p 2`), with a minimum stack depth of three reads in an individual to report a locus (`-m 3`). Calculations of Phi$_{ST}$ (Φ_{ST}), a measure of relative sequence divergence, were taken from the `populations` module. Nucleotide diversity (π) was computed manually as in Nei (1987) using the `dist.dna` function of R package 'ape' (Paradis *et al.*, 2004) by T. C. Nelson. Φ_{ST} and π estimates were kernel smoothed with a bandwidth of 500 kb for genome scans, and averaged within 100 kb genomic windows for correlation testing.

Genetic map scans

To plot divergence across the genetic map, I used the physical genome as an intermediary. I kernel smoothed both Φ_{ST} and genetic position in the hybrid map with regards to the physical genome, using a bandwidth of 500 kb and evaluating the fit at the same genomic coordinates.

Results and Discussion

Maintenance of freshwater-oceanic divergence is not a result of intrinsic reproductive isolation

Intrinsic reproductive isolation could maintain divergence between populations without physical barriers to gene flow. In addition, the presence of a chromosomal inversion with different orientations between populations can result in sexual isolation (Lenormand, 2002). To test for the presence of intrinsic incompatibilities between divergent stickleback populations which could prevent gene flow, I performed a series of within- and between-lineage laboratory crosses (Figure 6). For each cross, I counted the number of eggs in the clutch prior to fertilization, and the number of eggs in each cross that successfully hatched into healthy-looking larval stickleback by 11 days post-fertilization. A total of 36 crosses were performed, and 2,379 eggs scored. As expected, I found no compelling indication of strikingly lower viability in the backcrosses than in the within-lineage crosses. One backcross (hybrid female, Rabbit Slough male) did display an extremely low success rate, likely due to poor egg quality; this event was not representative of the rest of the cross series. The backcross family with the largest number of progeny (hybrid female, Boot Lake male) was used for the genetic mapping study.

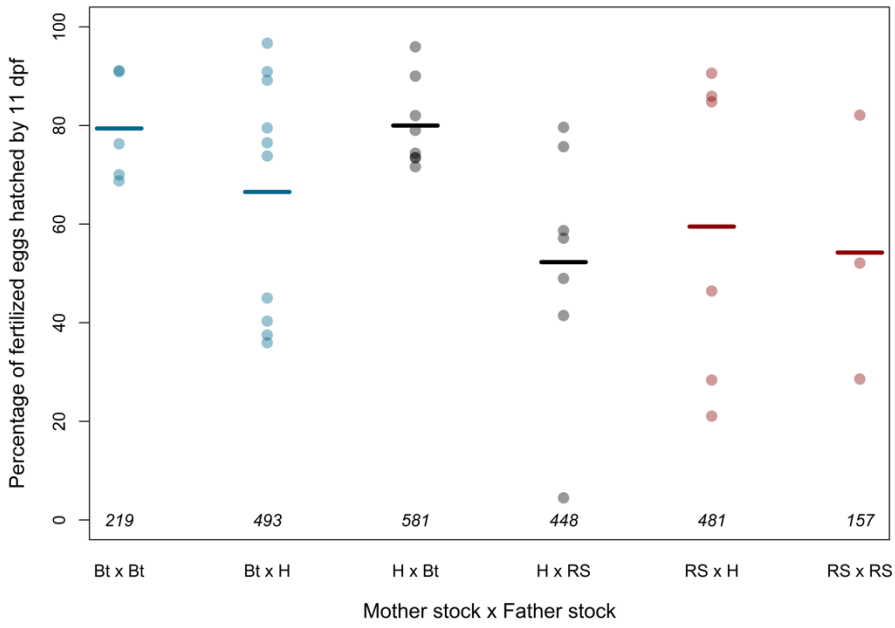

Figure 6: Hatching success of within- and between-lineage lab crosses.

A series of laboratory crosses was made between stocks derived from wild-caught stickleback ("RS," Rabbit Slough, "Bt," Boot Lake) and stocks of F_1 hybrid stickleback ("H," Boot Lake and Rabbit Slough parentage). Circles represent individual crosses and bars the mean for the series; italic numbers indicate total number of eggs scored for each cross type.

Recombination in the stickleback genome is lineage- and chromosome-specific

Local reductions in recombination in a hybrid (for example, across a heterozygous chromosomal inversion) are thought to be instrumental in facilitating adaptation in the presence of gene flow (Kirkpatrick, 2010; Hohenlohe *et al*., 2012; Jones *et al.*, 2012; Roesti *et al*., 2015). To compare recombination in the F_1 hybrid to that in the parental lines, and to determine the degree of conservation in the recombination patterns of the divergently-adapted freshwater and oceanic stickleback, I created genome-wide genetic maps for three stickleback as described in Methods

(Figures 8, 9, and 10, statistics in Table 1 and Figure 7). The hybrid, Boot Lake, and Rabbit Slough libraries were sequenced to an average read depth at each RAD locus of 24X, 34.7X, and 26.8X, respectively (Figure 7). After the quality filtering steps described, final genetic maps contained between 6169 and 7678 SNP markers, and ranged in length from 1206.2 to 1996 Kosambi cM; in all maps, markers covered roughly 98.7% of the reference genome (Table 1). Although all mapping families were normalized to equal size, the Rabbit Slough genetic map was strikingly shorter in length than the others. Overall, our genetic maps spanned the range of previously-described lengths in threespine stickleback (Roesti *et al*., 2013; Glazer *et al*., 2015), yet contained three to seven times the markers in each map. Marker densities were thus sufficient to facilitate investigations of fine-scale recombination: the Boot Lake map, with the fewest markers, displayed an average distance between adjacent markers of 70.8 kb and 0.32 cM (Figure 7). In each map, I found recombination rates to generally be low near the centers of chromosomes, and elevated in the chromosomal peripheries, consistent with previous work and predictions (Roesti *et al*., 2013; Glazer *et al*., 2015). Broad-scale patterns of recombination were generally conserved between lines, albeit with the F_1 hybrid displaying unique distributions of crossovers in certain genomic regions, such as chromosomes 1, 7, and 21 (Figure 9). Consistent with our expectations, the hybrid displayed suppressed recombination across the heterozygous inversion on chromosome 21 (Figure 10). Variation in chromosome-wise recombination rate both between and within fish was striking, with the former largely resulting from global differences in map length (Table 2).

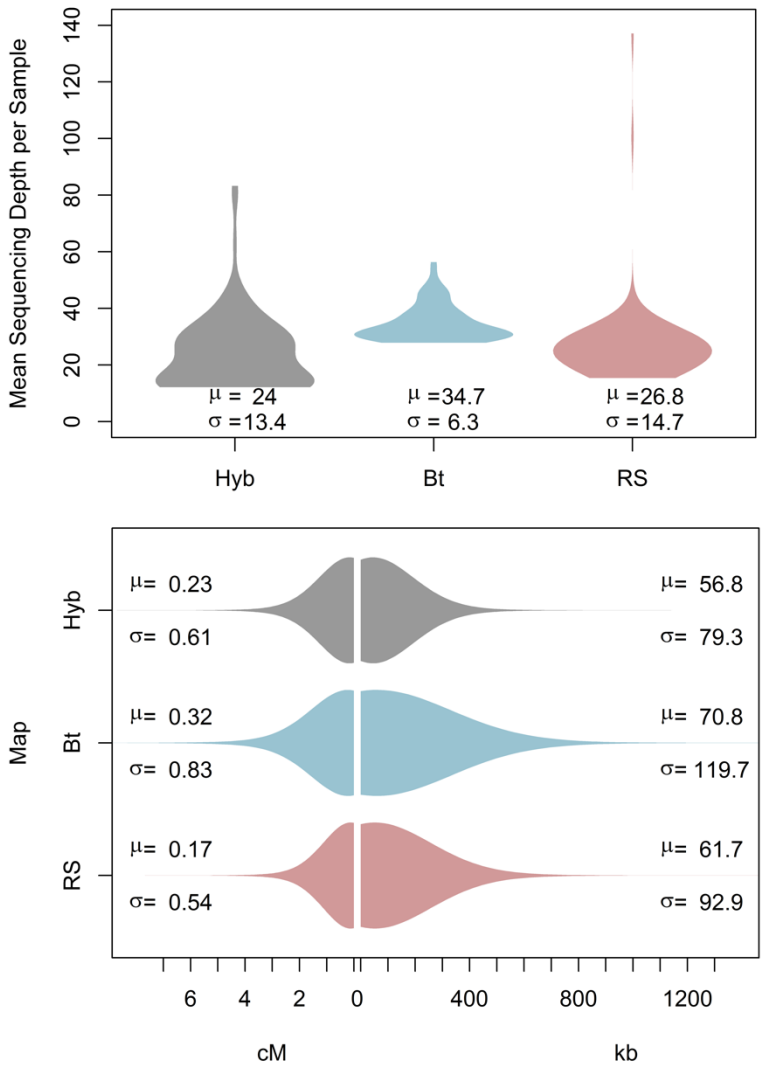

Figure 7: Statistics for genetic map datasets.

Kernel-density plots of sequencing depth for samples in the three RAD-seq datasets (top panel, *n* = 96 in each), and of physical and genetic distances between adjacent markers in each genetic map (bottom panel). Mean (μ) and standard deviation (σ) are displayed for each. “Hyb,” hybrid fish; “Bt,” Boot Lake fish; “RS,” Rabbit Slough fish.

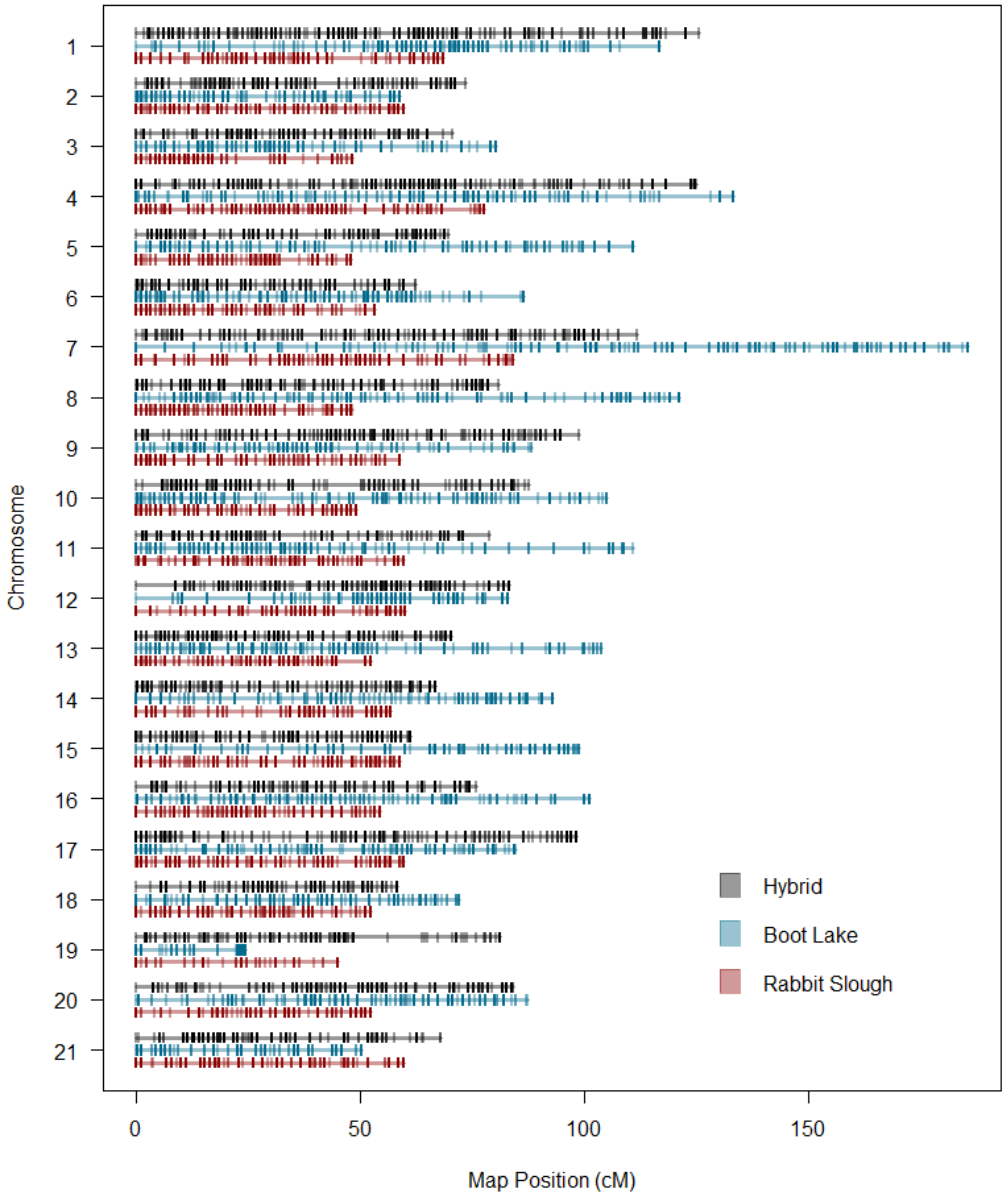

Figure 8: Full genetic maps from each line.

Genetic maps are displayed grouped by chromosome and colored according to mapping family. Marker positions are given by ticks. Genetic maps were generally longest in the Boot Lake fish, and shortest in the Rabbit Slough fish. Genetic map statistics are provided in Table 1.

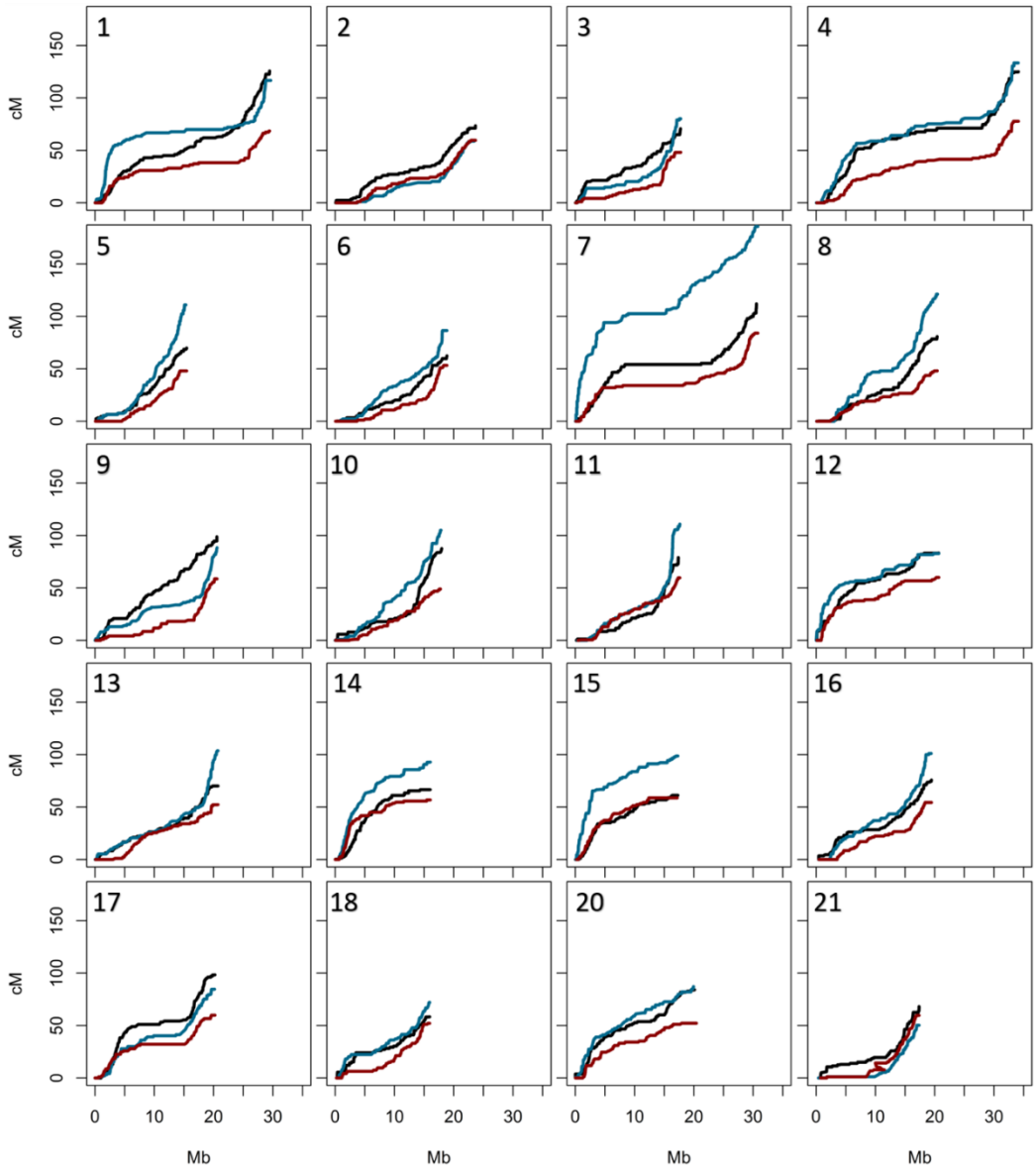

Figure 9: The recombination landscape of three stickleback lineages.

Genetic maps (with individual marker positions omitted) for each fish are displayed across the autosomes (chromosome number indicated in the top left of each panel). For each chromosome, lines connect the physical (x-axis, Mb) and genetic (y-axis, cM) positions of markers in each map, with the hybrid plotted in black, Boot Lake in blue, and Rabbit Slough in red. There was a general concordance in recombination landscape between maps on each chromosome, with the hybrid fish occasionally displaying unique deviations (i.e. chromosomes 1, 7, 9, 21).

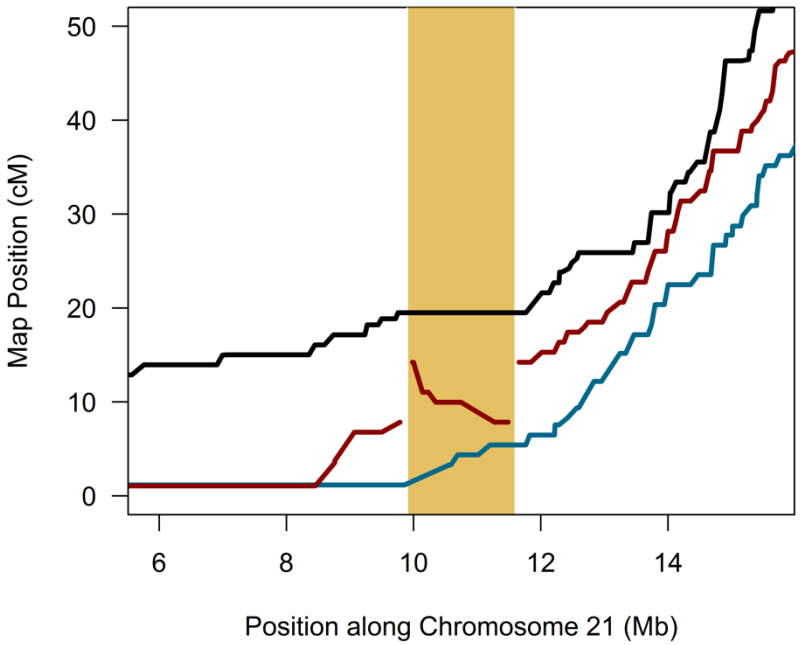

Figure 10: Recombination patterns across the chromosome 21 inversion.

Genetic maps for each fish (Boot Lake, blue, Rabbit Slough, red, hybrid, black) are displayed for a portion of chromosome 21, plotted by physical and genetic position. The yellow bar highlights the position of a chromosomal inversion; the Rabbit Slough and Boot Lake fish possess different orientations. The oceanic form is inverted relative to the stickleback reference genome, so genetic distance seems to decrease as physical position increases in the Rabbit Slough map. In the hybrid, recombination is suppressed between the two forms of the inversion, but elevated elsewhere on the chromosome, consistent with described phenomena in inversion heterozygotes (Schultz and Redfield, 1951).

Chromosome	Length (Mb)[1]	Map Length (cM)	Marker Number	Marker Coverage (% physical chromosome length)	Average Recombination Rate (cM/Mb)[†]	Map
1	29.63	125.7	526	98.29	4.32	Hybrid
		116.7	402	99.80	3.95	Boot Lake
		68.5	487	99.31	2.33	Rabbit Slough
2	23.70	73.5	375	99.48	3.12	Hybrid
		58.7	304	97.53	2.54	Boot Lake
		59.7	373	99.12	2.54	Rabbit Slough
3	17.80	70.7	332	99.57	3.99	Hybrid
		80.2	284	98.55	4.57	Boot Lake
		48.2	316	99.49	2.72	Rabbit Slough
4	34.14	125.0	684	99.75	3.67	Hybrid
		133.5	368	99.77	3.92	Boot Lake
		77.8	419	99.64	2.29	Rabbit Slough
5	15.56	69.8	307	98.24	4.57	Hybrid
		111.0	269	97.32	7.33	Boot Lake
		48.0	275	98.94	3.12	Rabbit Slough
6	18.85	62.4	289	98.87	3.35	Hybrid
		86.5	268	99.76	4.60	Boot Lake
		53.3	327	99.76	2.83	Rabbit Slough
7	30.84	112.0	555	98.98	3.67	Hybrid
		185.8	370	99.70	6.04	Boot Lake
		84.1	412	99.55	2.74	Rabbit Slough
8	20.53	80.8	378	98.35	4.00	Hybrid
		121.2	323	99.44	5.94	Boot Lake
		48.1	328	99.06	2.37	Rabbit Slough
9	20.58	98.8	358	99.22	4.84	Hybrid
		88.4	251	99.87	4.30	Boot Lake
		58.7	352	99.32	2.87	Rabbit Slough
10	18.03	87.6	305	99.59	4.88	Hybrid
		105.0	256	99.04	5.88	Boot Lake
		49.0	311	98.24	2.77	Rabbit Slough
11	17.64	79.0	330	97.46	4.59	Hybrid
		110.9	273	97.84	6.43	Boot Lake
		59.7	291	97.25	3.48	Rabbit Slough
12	20.67	83.4	393	98.99	4.06	Hybrid
		82.9	285	99.50	4.01	Boot Lake

		60.1	308	99.02	2.92	Rabbit Slough
13	20.74	70.2	375	99.90	3.39	Hybrid
		103.8	276	99.90	5.01	Boot Lake
		52.3	355	99.60	2.53	Rabbit Slough
14	16.17	66.7	316	98.93	4.17	Hybrid
		93.0	233	99.09	5.80	Boot Lake
		56.8	346	99.05	3.55	Rabbit Slough
15	17.32	61.3	279	99.63	3.55	Hybrid
		98.7	258	99.04	5.76	Boot Lake
		58.8	297	98.03	3.46	Rabbit Slough
16	19.52	75.8	322	97.45	3.99	Hybrid
		101.2	281	97.25	5.33	Boot Lake
		54.4	294	97.30	2.86	Rabbit Slough
17	20.25	98.4	325	99.12	4.90	Hybrid
		84.6	305	99.54	4.20	Boot Lake
		59.8	333	99.87	2.96	Rabbit Slough
18	15.99	58.3	250	97.48	3.74	Hybrid
		72.1	262	97.45	4.63	Boot Lake
		52.2	284	98.36	3.32	Rabbit Slough
19	20.61	81.3	255	97.70	4.04	Hybrid
		24.3*	406*	94.20*	1.25*	Boot Lake
		44.9*	378*	94.16*	2.31*	Rabbit Slough
20	20.45	84.2	360	98.24	4.19	Hybrid
		87.2	261	97.32	4.38	Boot Lake
		52.2	283	99.81	2.56	Rabbit Slough
21	17.35	68.1	364	95.85	4.09	Hybrid
		50.2	234	97.71	2.96	Boot Lake
		59.7	280	95.85	3.59	Rabbit Slough
Whole Genome	436.46	1732.8	7678	98.69	4.05	Hybrid
		1996.0	6169	98.66	4.71	Boot Lake
		1206.2	7049	98.70	2.86	Rabbit Slough

Table 1: Statistics from each genetic map.

[1]Lengths from Glazer *et al.* (2015) genome assembly. *Maps from male (heterogametic) fish; recombination is restricted to the ~2 Mb pseudo-autosomal region (Roesti et al., 2013). [†]Recombination rate expressed as the map length divided by the physical distance spanned by markers in the map.

	Group	μ (cM/Mb)†	σ (cM/Mb)†
Each chromosome, across all 3 maps	1	3.53	1.06
	2	2.73	0.33
	3	3.76	0.95
	4	3.29	0.88
	5	5.01	2.14
	6	3.59	0.91
	7	4.15	1.70
	8	4.1	1.79
	9	4	1.02
	10	4.51	1.59
	11	4.83	1.49
	12	3.66	0.64
	13	3.64	1.26
	14	4.51	1.17
	15	4.26	1.30
	16	4.06	1.24
	17	4.02	0.98
	18	3.89	0.67
	19	2.53	1.41
	20	3.71	1.00
	21	3.55	0.57
Each genetic map, across all 21 chromosomes	Hyb	4.05	0.51
	Bt	4.71	1.40
	RS	2.86	0.42

Table 2: Variation in chromosome-wise recombination within and between genetic maps.

Mean (μ) and standard deviation (σ) in recombination rate are shown for groupings of each chromosome in all genetic maps, and all chromosomes in each genetic map. "Hyb," hybrid map; "Bt," Boot Lake map; "RS," Rabbit Slough map. †Recombination rate on each chromosome expressed as map length divided by the physical distance spanned by markers. Average recombination rates between chromosomes in the Boot Lake map varied greatly, and were less variable in the Rabbit Slough and hybrid maps. Interestingly, the greatest variation in average recombination rate between maps was observed on chromosome 5, which boasts only a small peak of divergence (Figure 3).

To compare local recombination rates between fish, I calculated recombination rates in non-overlapping 1 Mb genomic windows for each map. Genome-wide recombination rates from each genetic map were then scaled relative to the maximum value, to account for differences in average recombination among maps; below, I refer to this metric as 'recombination intensity,' which ranges from 0 (no recombination) to 1 (highest rate for the map). A subtraction of recombination intensities between maps in each homologous genomic window was performed, to identify regions of elevated and decreased recombination in the hybrid relative to each parental line (Figure 11). I found that the hybrid stickleback displayed strikingly elevated recombination intensity in many genomic regions, and decreased recombination in only a few. This was surprising, as we expected to see extensive suppression of recombination in the hybrid relative to the other lines. To quantify the degree of concordance between the recombination landscapes in the Boot Lake and Rabbit Slough lines, I performed Kendall's Rank Correlation test on the inter-line recombination rate comparisons in each genomic window (Figure 12). Generally, genomic regions where the hybrid displayed relatively more or less recombination were conserved in both the Rabbit Slough and Boot Lake lines, highlighting the unique patterns of recombination in the F_1 hybrid and the conservation of recombination landscape, if not overall genetic map length, between the divergent Boot Lake and Rabbit Slough fish.

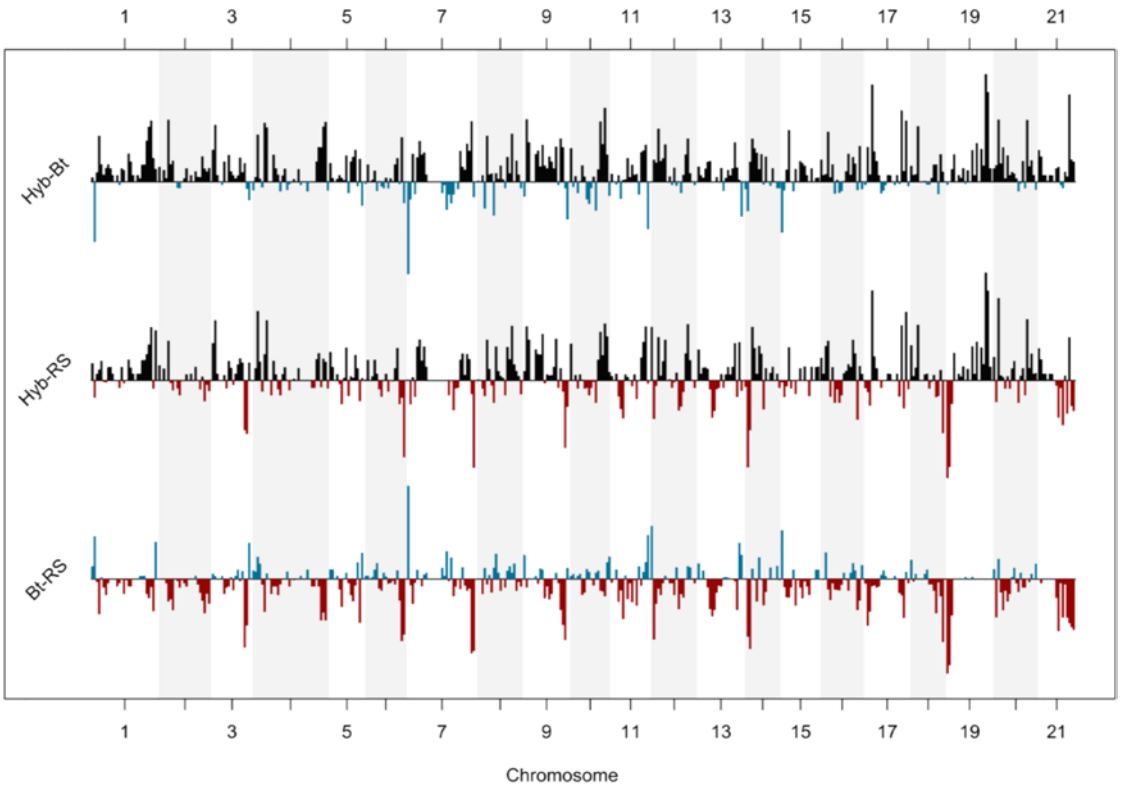

Figure 11: Lineage-specific variation in recombination intensity across the genome.

Pairwise comparisons of relative recombination rate (recombination intensity) between maps. Recombination rate (cM/Mb) was calculated in non-overlapping 1 Mb windows across the genome, then scaled to the maximum value of the map. For each pairwise comparison, a simple subtraction was performed: hybrid map – Boot Lake map (top row), hybrid map – Rabbit Slough map (middle), and Boot Lake map – Rabbit Slough map (bottom). Colored bars indicate the magnitude of the local difference in relative recombination between the two maps. A value of 0 (no colored bar plotted) indicates equal recombination intensities between maps (not necessarily a value of 0 cM/Mb). Relative recombination intensity was found to be higher in the hybrid compared to the pure lines in conserved genomic regions (see Figure 12).

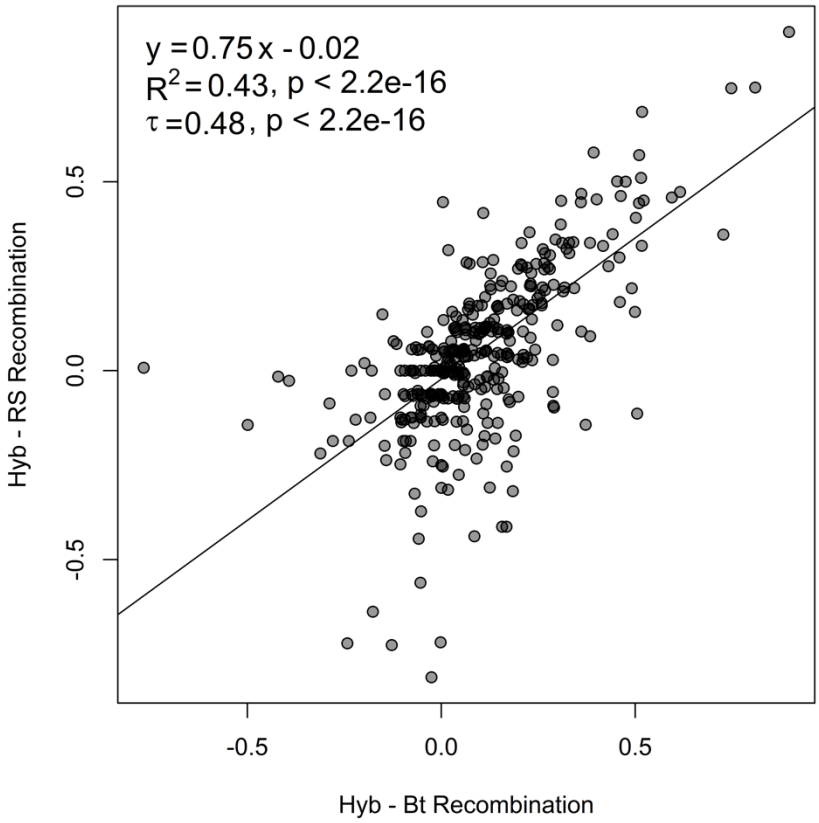

Figure 12: Genomic regions of recombination intensity variation are conserved.

Non-overlapping 1-Mb genomic windows are plotted by values from inter-map recombination rate intensity comparisons, as described in the legend of Figure 11. Briefly: the x-axis measures the difference in relative recombination rate in the hybrid and Boot Lake maps, and the y-axis measures the same metric between the hybrid and Rabbit Slough maps. Recombination intensity is relatively conserved between the two divergent lines. Displayed are regression equation, R^2, and p-value, as well as Kendall's Rank Correlation τ and p-value.

The stickleback genome displays extensive fine-scale recombination rate heterogeneity

The patterns of recombination we observed in the three genetic maps indicated the presence of extensive recombination rate heterogeneity not only between chromosomes, but also *within* each chromosome of each fish, especially in the hybrid. To investigate the extent of this heterogeneity, I calculated the mean and standard deviation of recombination rates in 1-Mb windows within each chromosome (Figure 13). I found striking recombination rate heterogeneity within chromosomes in all three genetic backgrounds. In all three maps, the majority of recombination events were clustered into relatively compact genomic regions, generally at the chromosomal peripheries. However, the distribution of high recombination-rate windows differed somewhat in the F_1 hybrid, with some recombination occurring in broader regions across many chromosomes (consistent with the results in Figure 11, where I found increased relative recombination rates in many genomic regions in the hybrid).

Within each chromosome, variation in recombination rate tended to be highest in the hybrid and lowest in the Rabbit Slough map, while mean recombination rate values were generally highest in the Boot Lake map and lowest in the Rabbit Slough map. This indicates that although the hybrid did not have the most recombination, it had the highest discrepancy in low- and high-recombination rate regions across chromosomes. In particular, the hybrid map had fewer genomic windows with low recombination rates (0-1 cM/Mb) than either of the other maps, and a higher number of windows with intermediate recombination rates (Figure 14).

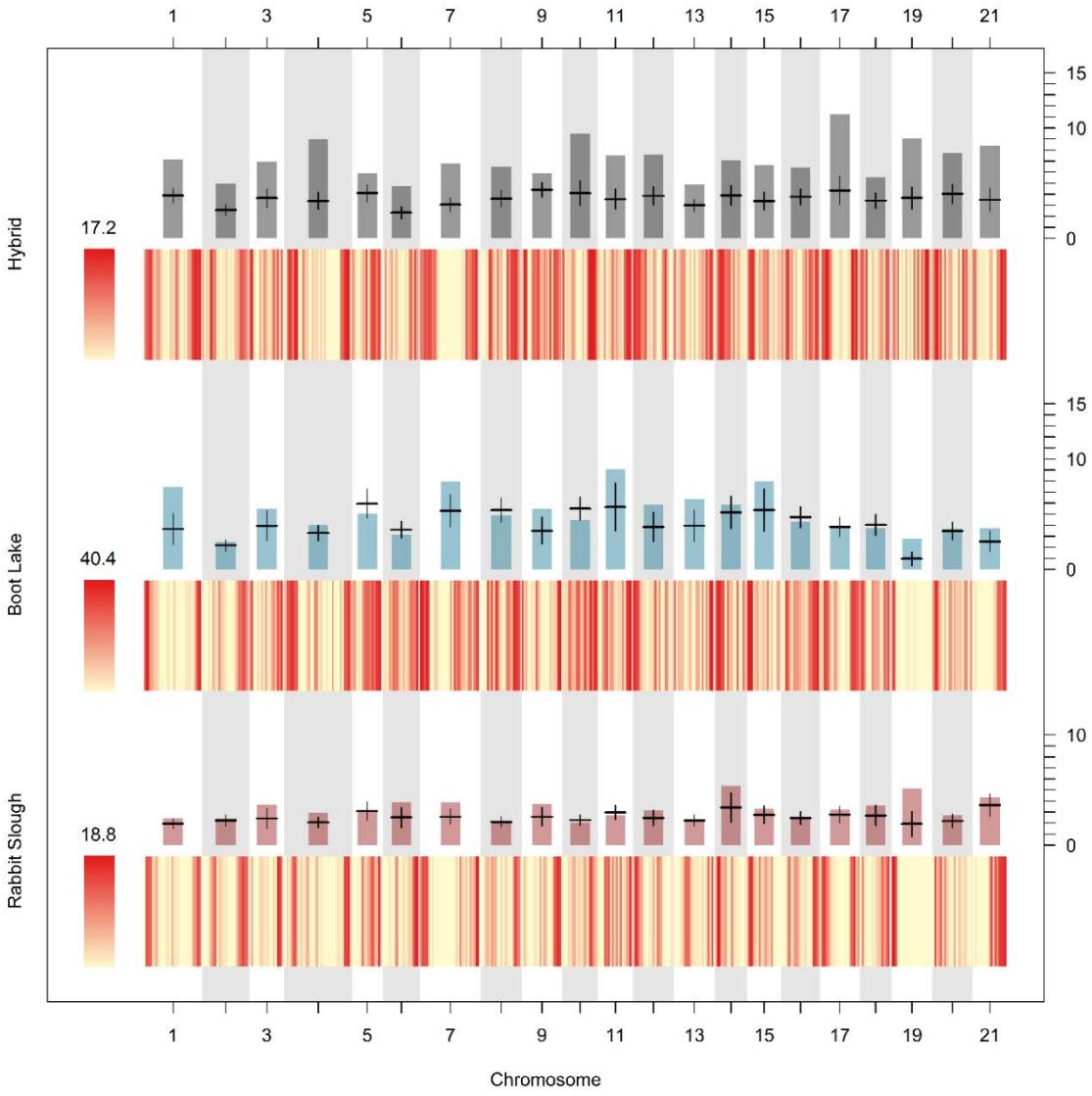

Figure 13: Recombination rate heterogeneity across the stickleback genome.

For each genetic map, heatmaps of recombination rates (cM/Mb) in non-overlapping 1 Mb windows are plotted, with colors indicating relative value (ranging from yellow = 0 cM/Mb to dark red = maximum map value, indicated above heatmap scale bars). Above heatmaps, bars and whiskers indicate the mean recombination rate value and standard error for the chromosome across all windows, and colored rectangles display the standard deviation, both plotted relative to the right y-axis (in cM/Mb).

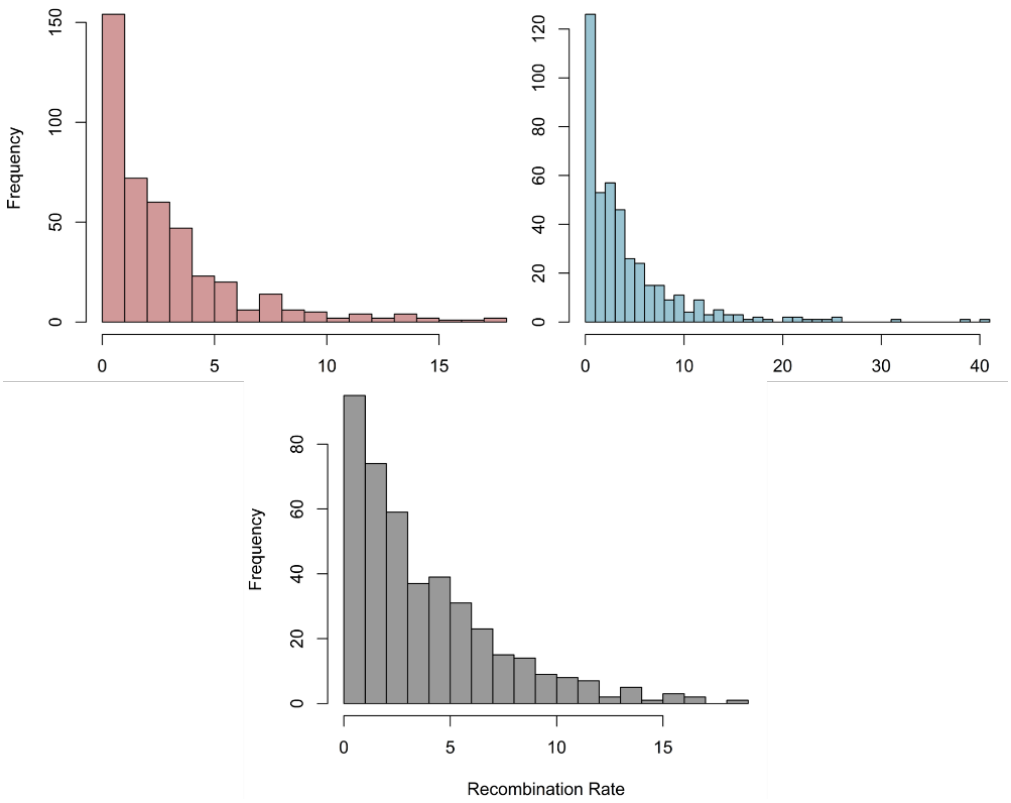

Figure 14: Distributions of recombination rates in each genetic map.

Histograms of recombination rates within 1-Mb genomic windows are shown for all genetic maps (Rabbit Slough, red; Boot Lake, blue; hybrid, grey). All three distributions are heavily skewed towards low recombination rates, but the hybrid map had more windows with intermediate recombination rates than either the Boot Lake or Rabbit Slough maps, and fewer windows with very low recombination rates.

Low recombination is correlated with high divergence

To investigate fine-scale recombination rate heterogeneity in the stickleback genome, I calculated recombination rates between all adjacent markers in the F_1 hybrid genetic map as described in Methods. I performed a genome scan of recombination rate in the hybrid map by plotting smoothed centimorgans per megabase with respect to the physical genome. We observed that on a number of chromosomes, particularly those displaying high levels of freshwater-ocean divergence, recombination in the hybrid is

negatively correlated with divergence; genomic islands of differentiation were predominantly clustered in regions of low or no recombination, while adjacent increases in recombination rate seemed to immediately erode between-population divergence (Figure 15). To quantify the relationship between recombination and population genetic statistics, I calculated average recombination rate, freshwater-oceanic divergence (Φ_{ST}), and nucleotide diversity between (π) and within (π_{Bt}, π_{RS}) each population in non-overlapping 100 kb windows. For each comparison, I calculated Kendall's Rank Correlation, excluding genomic windows with missing information. Consistent with expectations, we found recombination in the F_1 hybrid to be negatively correlated with divergence, and positively correlated with nucleotide diversity (Figure 16, Table 3). The positive correlation between recombination and nucleotide diversity, especially within populations, illustrates the power of recombination to generate variation (e.g. Hurles, 2005; Noor, 2008). Interestingly, genomic windows with a recombination rate of 0 cM/Mb in the hybrid map spanned the entire range of freshwater-oceanic divergence values; the negative correlation between recombination and Φ_{ST} was driven in part by a tendency for low-divergence genomic windows to display high recombination, not purely by a lack of recombination in high-divergence windows. This is consistent with our finding that large portions of the hybrid genome had little or no recombination. Thus, the divergence landscape in the stickleback genome seems to be driven, at least in part, by the homogenizing effects of recombination at certain genomic regions. Because a lack of recombination is not exclusive to high-divergence genomic regions, but high-divergence regions almost exclusively have little or no recombination (Figure 16), our

results indicate the importance of the spatial organization of adaptive loci within the stickleback genome.

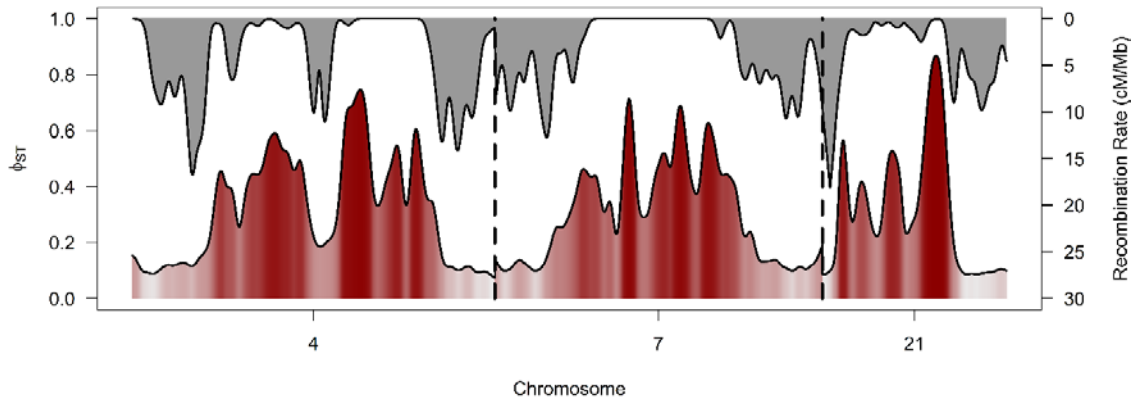

Figure 15: Recombination rate and molecular divergence appear to have a negative relationship across differentiated chromosomes.

We observed that genomic regions of low recombination rate in the hybrid stickleback (gray, inverted; plotted against right y-axis) tended to exhibit high freshwater-oceanic divergence (red, plotted against left y-axis) across many chromosomes (4, 7, and 21 displayed). Both metrics are smoothed here with a bandwidth of 500 kb. Note the striking lack of co-localization between peaks of recombination rate and divergence.

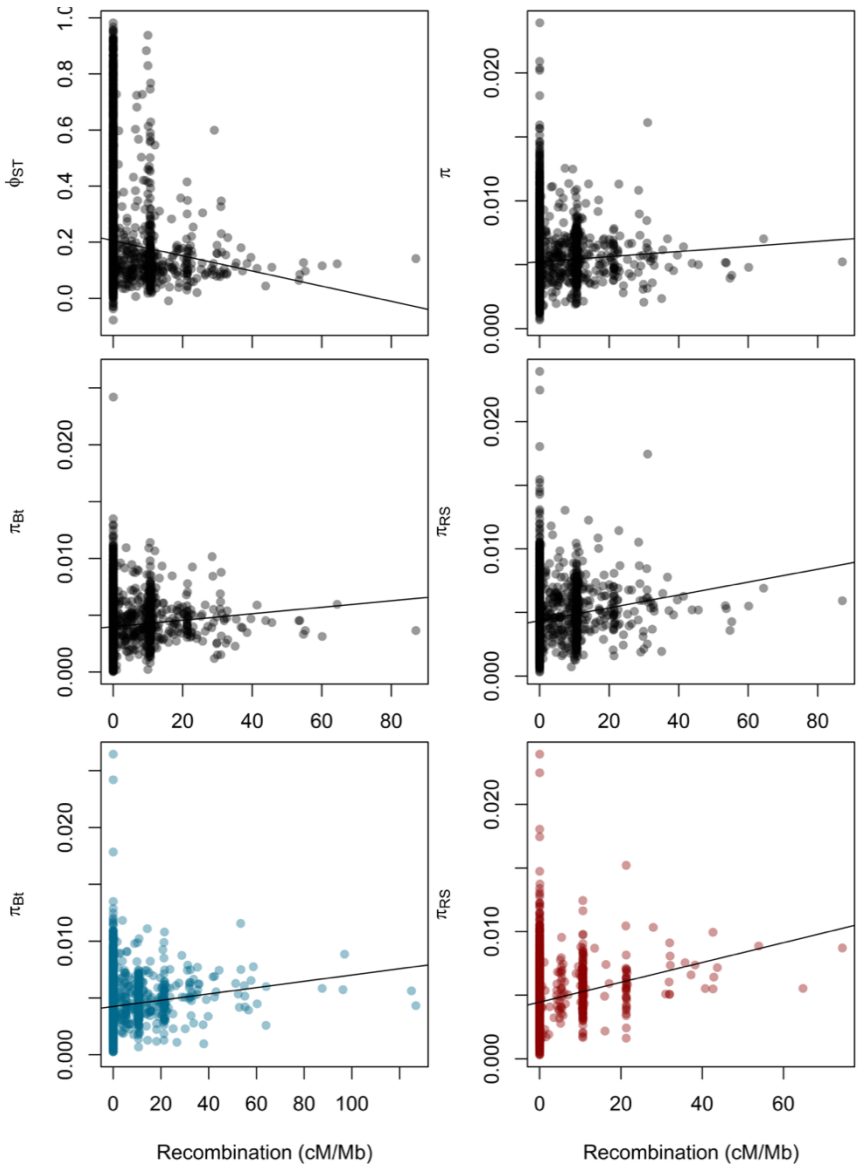

Figure 16: Statistical correlations of recombination and population genetic statistics.

Recombination in the hybrid stickleback is negatively correlated with freshwater-oceanic divergence (Φ_{ST}), and positively correlated with nucleotide diversity (π) within and between populations. Recombination in Boot Lake and Rabbit Slough fish is likewise correlated with nucleotide diversity within each population. 100 kb non-overlapping genomic windows are plotted by mean divergence or nucleotide diversity value and recombination rate in the hybrid (top four panels, grey points), Rabbit Slough (bottom right, red points), or Boot Lake (bottom left, blue points) genetic map. A regression is fitted to each scatterplot to illustrate the trend.

	Recombination rate (cM/Mb)		
	Hybrid map	Boot Lake map	Rabbit Slough map
Φ_{ST}	$\tau = -0.0621$ $p = 7.18 \times 10^{-6}$	———	———
π	$\tau = 0.0867$ $p = 3.67 \times 10^{-10}$	———	———
π_{Bt}	$\tau = 0.11$ $p = 3.55 \times 10^{-15}$	$\tau = 0.127$ $p = <2.2 \times 10^{-16}$	———
π_{RS}	$\tau = 0.144$ $p = <2.2 \times 10^{-16}$	———	$\tau = 0.153$ $p = <2.2 \times 10^{-16}$

Table 3: Correlations of recombination with population genetic statistics.

For each panel in Figure 16, Kendall's Rank Coefficient τ and p-value are given for the compared metrics assessed in 100 kb genomic windows.

Mutations that limit recombination within adaptive genomic regions may be favored by natural selection as they arise, as they help maintain haplotypes tailored to different environments in the face of gene flow (Kirkpatrick and Barton, 2006; Ortiz-Barrientos *et al.*, 2016). However, our characterization of the recombination landscape of the stickleback genome indicates that certain regions display a natural predisposition to low recombination rates, as evidenced by the presence of long, non-recombining tracts in all maps. This observation is supported by the findings of other groups (Roesti *et al.*, 2013; Hohenlohe *et al.*, 2012; Glazer *et al.*, 2015). Some genomic regions—including those implicated in adaptation—display local reductions in recombination rate

not unique to the hybrid fish. Whether or not these genomic regions evolved low recombination rates in order to segregate adaptive variation, if they happen to harbor suites of genes important for adaptation, divergence can be facilitated and maintained in the face of gene flow by the extensive physical linkage within the region.

Genomic islands of divergence are tight genetic units

The stickleback genome displays broad genomic islands of divergence, whose vast size indicates their inherent genomic complexity. These islands encompass hundreds of predicted genes and in some cases span the majority of the length of the chromosome (i.e. chromosomes 4 and 7; see Figure 3). Under a model that assumes free recombination between every gene, all components of these genomic islands would need to be maintained independently by divergent selection for the divergence peaks to persist in the face of gene flow. As our results indicate, however, the relationship between the genetic and physical map is not one-to-one, especially in the F_1 hybrid. The presence of long non-recombining genomic tracts in the hybrid map prompted us to investigate the relationship between discrete genetic, rather than genomic, loci and genetic divergence between freshwater and oceanic stickleback. To do so, we scanned for levels of divergence along the genetic map of the F_1 hybrid, rather than the genome. We observed that recombination clearly compartmentalizes adaptive and neutral genomic regions in the hybrid: massive genomic islands of high divergence were consistently found to be compact genetic loci in the hybrid map (Figures 17 & 18). Conversely, freely-recombining regions exhibited low divergence.

Large, divergent genomic regions segregate as a small number of genetic loci between, and to an extent within, populations; however, genomic regions

compartmentalized into small genetic loci are not necessarily divergent (Figure 19), and thus, the high divergence observed across non-recombining regions of chromosomes 4, 7, and 21 is not an artifact of the lack of recombination, but rather an indication of adaptive importance. Our results illustrate the extent to which the non-random patterns of recombination in the stickleback genome are crucial for facilitating adaptation: genomic regions which potentially harbor suites of adaptive alleles are visible to natural selection as a small number of genetic units, rather than a multitude of discrete genomic loci. In contrast, some genomic regions that segregate as single genetic loci lack apparent adaptive importance; this presents the possibility that adaptive genomic elements are arranged in regions of low recombination rate serendipitously.

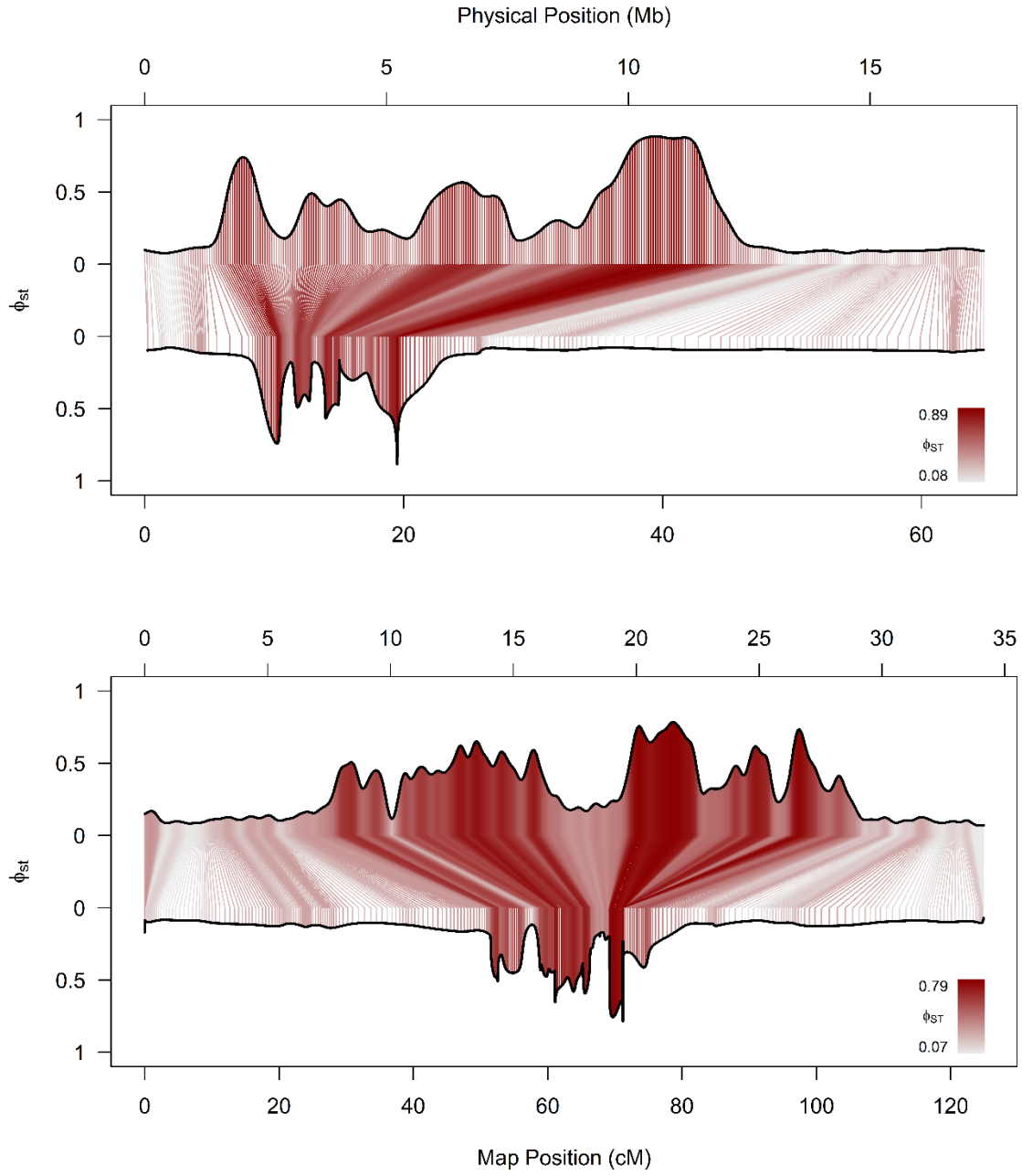

Figure 17: Recombination compartmentalizes adaptive genomic regions.

Physical- and genetic map-scans of divergence are displayed for chromosomes 21 (top panel) and 4 (bottom panel). Molecular divergence (Φ_{ST}) between Boot Lake and Rabbit Slough populations is plotted against position along the chromosome (top track) and position in the hybrid genetic map (bottom track, inverted). Shading lines are evenly spaced on the physical map, connect genomic loci to the corresponding position in the genetic map, and are colored according to local smoothed Φ_{ST} value (scaled to chromosome maximum and minimum). The number of connecting lines is equal to the number of markers on the chromosome in the genetic map. Local reductions in recombination rate cause large genomic regions to be compartmentalized into small regions of the genetic map. For example, an inversion on chromosome 21 between freshwater and oceanic populations (spanning roughly 9.9-11.6 Mb on the physical map) displaying high levels of differentiation collapses to a small genetic map region, as recombination is restricted between the two orientations when they are combined in the hybrid genome (see Figure 10).

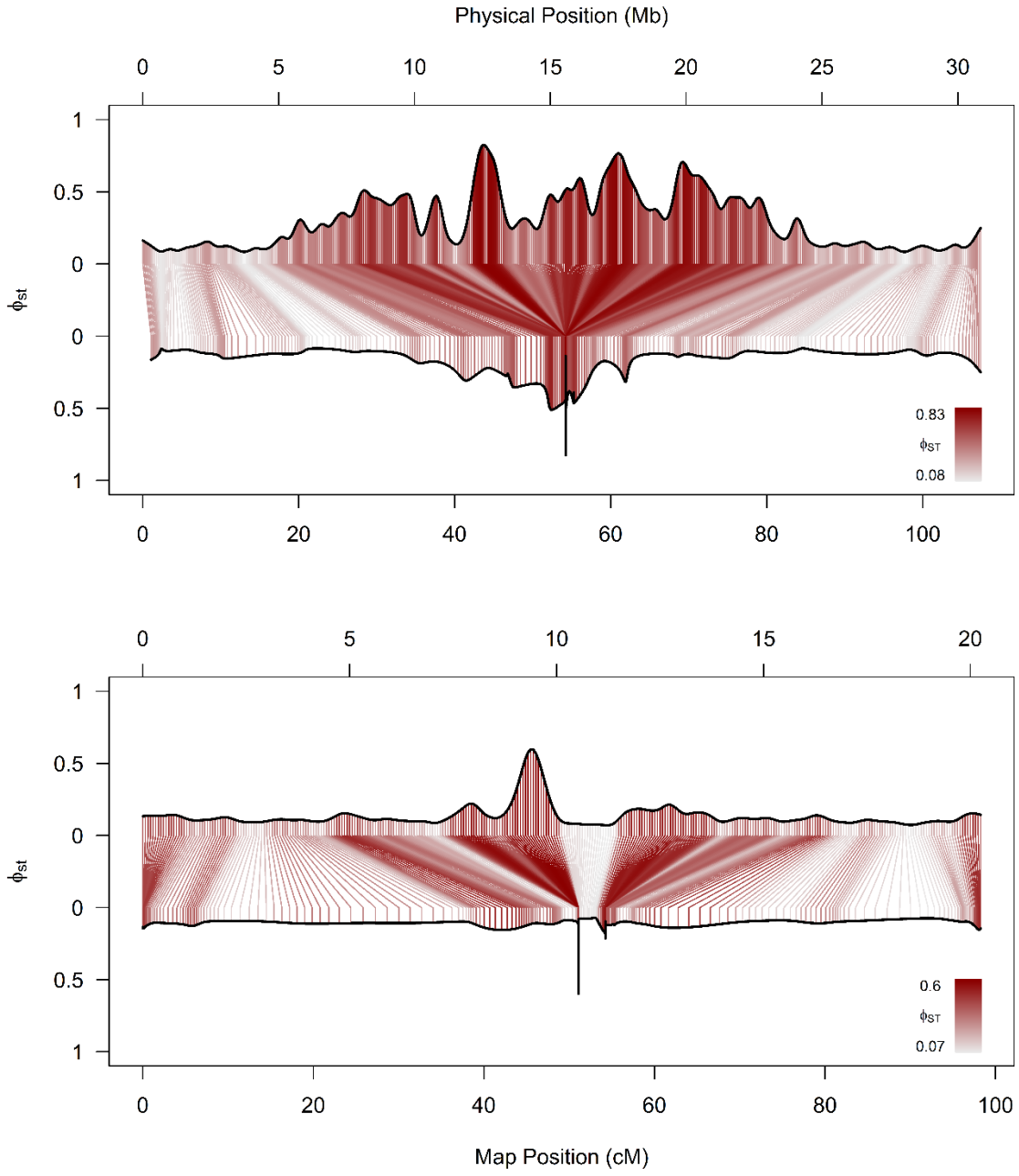

Figure 18: Compartmentalization of adaptive regions on chromosomes 7 and 17.

Physical- and genetic map-scans as in Figure 17, for chromosomes 7 (top panel) and 17 (bottom panel). On chromosome 7, a vast genomic region encompassing several divergence peaks is compartmentalized into very few genetic loci; similarly, on chromosome 17, the highest divergence peaks (spanning several megabases) share a single genetic locus.

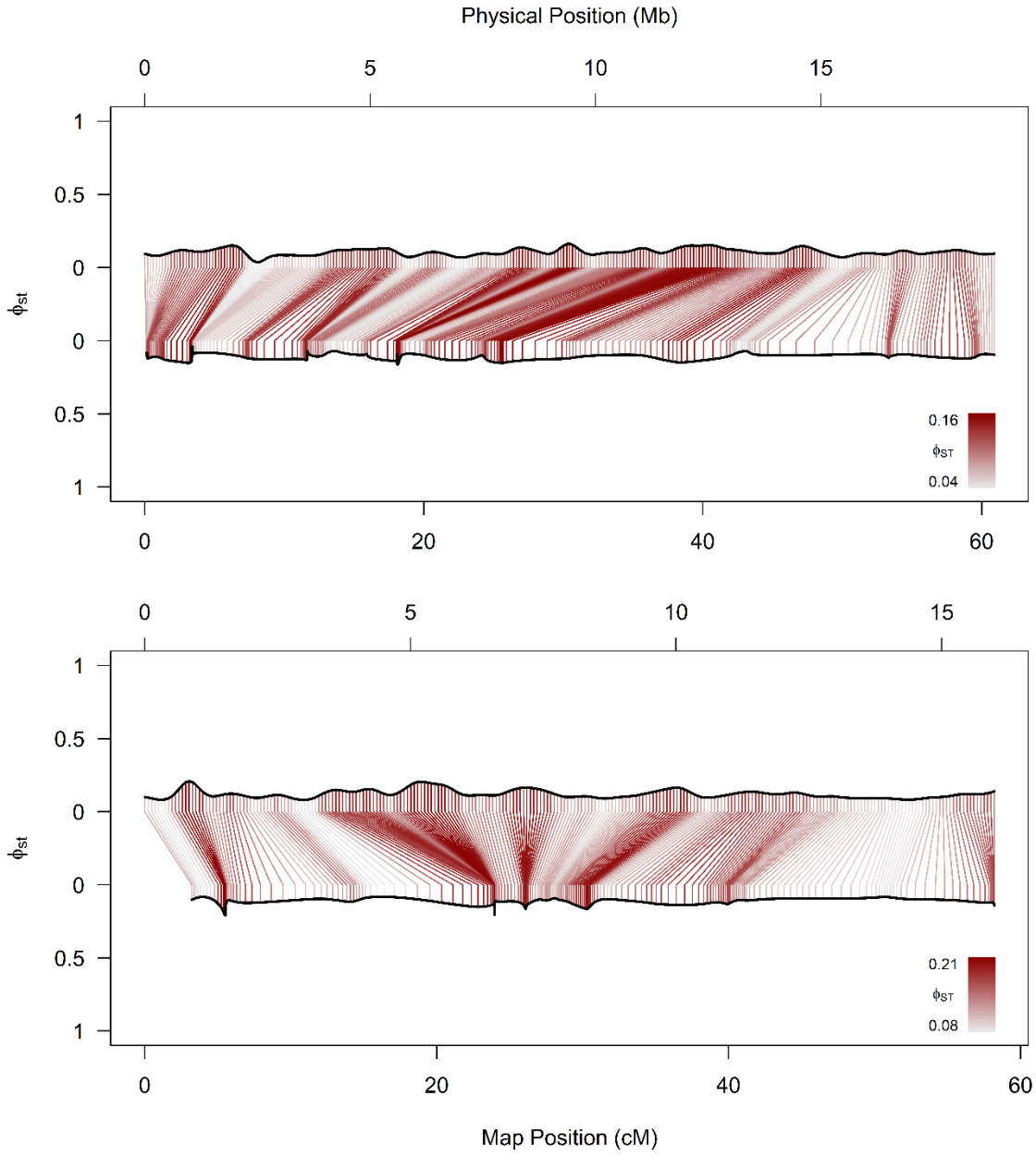

Figure 19: Compartmentalization by recombination does not predict divergence.

Physical- and genetic-map scans for chromosomes 6 (top panel) and 18 (bottom panel). On both, genomic regions that are small genetic loci are not highly differentiated. Although points of condensation do tend to represent local divergence maxima, overall differentiation is low. This illustrates that the high divergence of compartmentalized regions on other chromosomes is not simply a result of local reductions in recombination rate.

Conclusions

The stickleback recombination landscape is heterogeneous

The mechanisms maintaining adaptive divergence between closely related populations have historically been targets of study by evolutionary biologists, as theory predicts that gene flow should overwhelm natural selection (Lenormand, 2002). Here, we have generated a dataset amenable for investigating the role of recombination in facilitating adaptation in the threespine stickleback system. Using laboratory crosses, I created dense genetic maps to characterize the recombination landscape in three distinct genetic backgrounds, and coupled these with an analysis of divergence between natural freshwater and oceanic populations. The marker density of our genetic maps allowed fine-scale investigations of both recombination rate variation and the relationship between recombination and divergence. We found that broad-scale genomic patterns of recombination are generally conserved between the divergent populations in our study, as well as between stickleback in our system and others (Roesti *et al.*, 2013; Glazer *et al*., 2015). However, we also found a far greater level of recombination rate heterogeneity across the stickleback genome than previously described, with striking variation in recombination rate a) within individual chromosomes, b) between chromosomes within individuals, and c) between chromosomes across individuals of different genetic backgrounds.

Recombination is a mediator of adaptive genomic variation

The distinct patterns of recombination in the stickleback genome have facilitated rapid and sustained adaptive evolution in wild populations, due in part to the distinct

spatial organization of adaptive regions. The peaks of divergence across the stickleback genome are good evidence of adaptive evolution, and a subset of these have been linked to phenotypic variation thought to be ecologically important (Hohenlohe *et al.*, 2010; Jones *et al*., 2012; Miller *et al.*, 2014; Colosimo *et al.*, 2004). These peaks are broadly co-localized with regions of low, or no, recombination in fish from freshwater and marine genetic backgrounds and an inter-population F1 hybrid. It remains unclear whether low recombination rates in these regions evolved in response to selective pressure on the adaptive alleles contained within, or whether loci necessary for adaptive phenotypes in both freshwater and marine environments happen to serendipitously be located in regions of low recombination. In either case, the specific recombination patterns in the stickleback genome, which strikingly and effectively compartmentalize adaptive genomic variation, help facilitate and maintain divergence in the face of gene flow. Large, complex, adaptive genomic regions segregate as small Mendelian loci. In this manner, they present simple targets to selection, allowing rapid and sustained adaptation.

Furthermore, our results suggest that specific recombination patterns could not only maintain divergence between divergent stickleback populations, but also facilitate repeated adaptation to new environments. Oceanic stickleback are known to invade and rapidly adapt to freshwater habitats (Hohenlohe *et al.*, 2010; Lescak *et al*., 2015). If adaptive freshwater haplotypes can re-enter oceanic populations due to gene flow and resist breakup due to a lack of recombination, then these would represent a rich source of genetic variation. When oceanic populations invade new freshwater environments,

this adaptive variation—established over time in freshwater environments by natural selection—can be immediately recruited to enable rapid adaptation.

Furthermore, our results illustrate the importance of a detailed genetic map in understanding the complex genomics underlying adaptation. Recently, genetic maps have been used as a tool to assemble physical genomes, which are then investigated for signatures of adaptive evolution. Both the physical and genetic maps represent different aspects of biological reality, and our results underscore the lack of concordance between the two. Our approach of scanning across the genetic map for divergence allows us to view the genome as it is visible to natural selection. Instead of making the assumption that all genomic peaks of divergence are independent, we can observe the functional result of the spatial organization of adaptive loci, and the striking degree to which recombination rate heterogeneity segregates adaptive genomic variation.

Our work here shows that recombination with regards to adaptation must be viewed as a double-edged sword. Naturally, recombination poses a great impediment to the process of divergence-with-gene flow. But in the Alaskan stickleback system, distinct recombination patterns across different genetic backgrounds have directly influenced the evolutionary trajectories of natural populations. By compartmentalizing adaptive genomic variation and segregating large haplotypes as small genetic loci, recombination facilitates sustained and repeated adaptation. Although small in scope, this work sheds light on the complex process of adaptation, thereby furthering our understanding of how biodiversity is generated on our planet.

Glossary of terms

cM (centimorgan): Measure of genetic distance. Describes the number of times recombinations were observed between two markers; 1 cM equals a recombination in 1 out of 100 meioses. In this study, genetic distance was estimated using Kosambi's mapping algorithm (Kosambi, 1944) in the software package Lep-MAP2 (Rastas *et al.*, 2016).

F_1: First-generation.

Gamete: The sex cells; eggs in females, sperm in males.

Gene flow: the exchange of genetic information, in one or both directions, between populations (i.e. mating).

Genomic architecture: The distribution and functionality of important genomic elements, e.g.: "…the totality of non-random arrangements of functional elements in the genome," (Koonin, 2009); "…the genome-wide distribution and covariation of loci and genomic regions important for adaptation and reproductive isolation," (Hohenlohe *et al.*, 2012).

Nucleotide: One molecular letter of DNA; the fundamental unit of genes and the genome. There are four different nucleotides in DNA: Adenine, Thymine, Cytosine, and Guanine (A, T, G, and C). When measuring physical distance along chromosomes or genomes, nucleotides are commonly termed base pairs (bp); for reference, the Stickleback genome assembly is roughly 436.4 million base pairs long, and the human genome is over 3.2 billion base pairs long.
kb: "kilobase," one thousand base pairs (see 'nucleotide'). Measure of physical distance along a chromosome or genome.

Mb: "megabase," one million base pairs (see 'nucleotide'). Measure of physical distance along a chromosome or genome.

π (pi): π is a population genetic statistic which measures nucleotide diversity across loci between or within populations, thereby indicating levels of inter- or intra-population genetic variation. Calculated as in Nei, 1987.

Φ_{ST} (phi_{ST}): A population genetic statistic that measures relative genetic divergence between populations. Φ_{ST} compares allele frequencies in each population at each tested locus. Originally from Excoffier *et al.*, 1992, and Excoffier, 2001; calculated via the Stacks analysis pipeline (Catchen *et al.*, 2013).

Sequencing: Sequencing of DNA yields identification of the specific nucleotides in a strand of DNA and their order. Errors are inherent in the next-generation sequencing process, and so in order for high-confidence identification of the nucleotide at a given position, that position must be sequenced multiple times from many identical strands of DNA. The number of times a locus is sequenced is referred to as 'depth of coverage,' and is influenced by a number of factors. One sequencing run can yield only a finite number of DNA reads (e.g. Illumina HiSeq 2500 High Output yields 160-200 million DNA reads, as reported by the University of Oregon Genomics and Cell Characterization Core Facility), so depth of coverage at each locus depends on the number of loci being sequenced.

Bibliography

Andrew, R. L., and L. H. Rieseberg., 2013. “Divergence is Focused on Few Genomic Regions Early in Speciation: Incipient Speciation of Sunflower Ecotypes.” *Evolution* 67: 2468-82.

Baird, N. A., P. D. Etter, T. S. Atwood, M. C. Currey, A. L. Shiver, Z. A. Lewis, *et al.*, 2008. “Rapid SNP Discovery and Genetic Mapping Using Sequenced RAD Markers.” *PLoS ONE* 3(10): e3376. doi:10.1371.

Baker, B. S., A. T. C. Carpenter, M. S. Esposito, R. E. Esposito, and L. Sandler, 1976. “The genetic control of meiosis.” *Annual Review of Genetics* 10: 53-134.

Bansal, V., A. Bashir, and V. Bafna, 2007. “Evidence for large inversion polymorphisms in the human genome from HapMap data.” *Genome Research* 17(2): 219-30.

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

Bergey, C. M., L. Pozzi, T. R. Disotell, and A. S. Burnell., 2013. “A New Method for Genome-wide Marker Development and Genotyping Holds Great Promise for Molecular Primatology.” *International Journal of Primatology* 34(2): 303-14.

Botstein, D., R. L. White, M. Skolnick, and R. W. Davis, 1980. “Construction of a genetic linkage map in man using restriction fragment length polymorphisms.” *American Journal of Human Genetics* 32(3): 314-31.

Burri, R., A. Nater, T. Kawakami, C. F. Mugal, P. I. Olason, L. Smeds, *et al*., 2015. “Linked selection and recombination rate variation drive the evolution of the genomic landscape of differentiation across the speciation continuum of Ficedula flycatchers.” *Genome Research*: doi: 10.1101/gr.196485.115.

Bushnell, B., 2014. “BBMap: A Fast, Accurate, Splice-Aware Aligner.”

Castle, W. E., 1903. “Mendel’s Law of Heredity.” *Proceedings of the National Academy of Sciences* 38: 535-48.

Catchen, J, P. A. Hohenlohe, S. Bassham, A. Amores, and W. A. Cresko, 2013. “Stacks: an analysis tool set for population genomics.” *Molecular Ecology* 22(11): 3124-40.

Colosimo, P. F., C. L. Peichel, K. Nereng, B. K. Blackman, M. D. Shapiro, D. Schluter, *et al.*, 2004. “The Genetic Architecture of Parallel Armor Plate Reduction in Threespine Sticklebacks.” *PLoS Genetics* 2(5): 635-41.

Corbett-Detig, R. B., and D. L. Hartl, 2012. “Population Genomics of Inversion Polymorphisms in *Drosophila melanogaster*.” *PLoS Genetics* 8(12): e1003056. doi:10.1371/journal.pgen.1003056.

Coyne, J. A., and H. A. Orr, 2004. *Speciation*. Sinauer Associates, Inc.

Darwin, C. R., 1859. *On the origin of species by means of natural selection, or the preservation of favoured races in the struggle for life*. London: John Murray.

Dobzhansky, T., and C. Epling, 1948. “The Suppression of Crossing Over in Inversion Heterozygotes of *Drosophila pseudoobscura*.” *Proceedings of the National Academy of Sciences* 34: 137-41.

Excoffier, L., 2001. “Analysis of population subdivision.” *Handbook of statistical genetics*. Chichester: John Wiley & Sons.

Excoffier, L., P. E. Smouse, and J. M. Quattro, 1992. “Analysis of molecular variance inferred from metric distances among DNA haplotypes: application to human mitochondrial DNA restriction data.” *Genetics* 131: 479-91.

Fisher, R. A., 1930. *The Genetical Theory of Natural Selection*. Oxford at the Clarendon Press.

Flaxman, S. M., Wacholder, A. C., Feder, J. L., and Nosil, P., 2014. “Theoretical models of the influence of genomic architecture on the dynamics of speciation.” *Molecular Ecology* 23(16): 4074-88.

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: Genes | Genomes | Genetics* 5(7): 1463-72.

Glazer, A. M., E. E. Killingbeck, T. Mitros, D. S. Rokhsar, and C. T. Miller, 2015. Data from: “Genome Assembly Improvement and Mapping Convergently Evolved Skeletal Traits in Sticklebacks with Genotyping-by-Sequencing.” Dryad Digital Repository. http://dx.doi.org/10.5061/dryad.q018v.

Griffiths, A. J. F., J. H. Miller, D. T. Suzuki, R. C. Lewontin, and W. M. Gelbart. *An Introduction to Genetic Analysis* (7th Ed.). New York: W. H. Freedman.

Hassold, T., and P. Hunt, 2006. “To err (meiotically) is human: the genesis of human aneuploidy.” *Nature Reviews Genetics* 2: 280-91.

Hohenlohe, P. A., S. Bassham, P. D. Etter, N. Stiffler, E. A. Johnson, and W. A. Cresko, 2010. “Population Genomics of Parallel Adaptation in Threespine Stickleback using Sequenced RAD Tags.” *PLoS Genetics* 6(2): e1000862. doi:10.1371.

Hohenlohe, P. A., S. Bassham, M. Currey, and W. A. Cresko, 2012. “Extensive linkage disequilibrium and parallel adaptive divergence across threespine stickleback genomes.” *Philosophical Transactions of the Royal Society* 367: 395-408.

Hudson, R. R., M. Slatkin, and W. P. Maddison, 1902. “Estimation of levels of gene flow from DNA sequence data.” *Genetics* 132(2): 583-9.

Hurles, M., 2005. “How homologous recombination generates a mutable genome.” *Human Genomics* 2: doi: 10.1186/1479-7364-2-3-179.

Jones, F. C., M. G. Grabherr, F. C. Yingguang, P. Russell, E. Mauceli, J. Johnson, *et al.*, 2012. “The genomic basis of adaptive evolution in threespine sticklebacks.” *Nature* 484(7392): 55-61.

Kirkpatrick, M., 2012. “How and Why Chromosome Inversions Evolve.” *PloS Biology* 8(9): e1000501.

Kirkpatrick, M., and N. Barton, 2006. “Chromosome inversions, local adaptation, and speciation.” *Genetics* 173(1): 419-34.

Koonin, E. V., 2009. “Evolution of Genome Architecture.” *International Journal of Biochemistry and Cell Biology* 41(2): 298-306.

Kosambi, D. D., 1944. “The estimation of map distances from recombination values.” *Annuals of Eugenics* 12:172–5.

Lackey, A. C. R., and J. W. Boughman, 2016. “Evolution of reproductive isolation in stickleback fish.” *Evolution* doi: 10.1111/evo.13114.

Lange, J., S. Yamada, S. E. Tischfield, J. Pan, S. Kim, X. Zhu, *et al*., 2016. “The Landscape of Mouse Meiotic Double-Strand Break Formation, Processing, and Repair.” *Cell* 167(3): 695-708.

Lango Allen, H., K. Estrada, G. Lettre, S. I. Berndt, M. N. Weedon, F. Rivadeneira *et al*., 2010. “Hundreds of variants clustered in genomic loci and biological pathways affect human height.” *Nature* 467(7317): 832-8.

Lathrop, G.M., J. M. Lalouel, C. Julier, and J. Ott, 1984. “Strategies for multilocus linkage analysis in human.” *Proceedings of the National Academy of Sciences* 81: 3443–46.

Lee, C.-R., B. Wang, J. P. Mojica, T. Mandakova, K. V. S. K. Prasad, J. L. Goicoechea, *et al.*, 2017. “Young inversion with multiple linked QTLs under selection in a hybrid zone.” *Nature Ecology & Evolution:* doi: 10.1038/s41559-017-0119.

Lee, Y. W., L. Fishman, J. K. Kelly, and J. H. Willis, 2016. "A Segregating Inversion Generates Fitness Variation in Yellow Monkeyflower (*Mimulus guttatus*)." *Genetics* 202(4): 1473-84.

Lenormand, T., 2002. “Gene flow and the limits to natural selection.” *Trends in Ecology and Evolution* 17: 183-9.

Lescak, E. A., S. Bassham, J. Catchen, O. Gelmond, M. L. Sherbick, F. A. von Hippel, *et al.*, 2015. “Evolution of stickleback in 50 years on earthquake-uplifted islands.” *Proceedings of the National Academy of Sciences of the United States of America*: e7204–12. doi: 10.1073.

Li, H., B. Handsaker, A. Wysoker, T. Fennell, J. Ruan, N. Homer, *et al.*, 2009. “The Sequence alignment/map (SAM) format and SAMtools.” *Bioinformatics* 25: 2078-9.

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 Biology* 8(9): e1000500.

Martin, S. H., K. K. Dasmahapatra, N. J. Nadeau, C. Salazar, J. R. Walters, F. Simpson, *et al.*, 2013. “Genome-wide evidence for speciation with gene flow in Heliconius butterflies.” *Genome Research*: doi: 10.1101/gr.159426.113

Mather, K., 1938. “Crossing-over.” *Biological Reviews* 13: 252-92.

McKinnon, J. S., and H. D. Rundle, 2002. "Speciation in nature: the threespine stickleback model systems." *Trends in Ecology & Evolution* 17: 480-8.

Mendel, G., 1865. “Experiments in plant hybridization.”

Miller, C. T., A. M. Glazer, B. R. Summers, B. K. Blackman, A. R. Norman, M. D. Shapiro, *et al*., 2014. “Modular skeletal evolution in sticklebacks is controlled by additive and clustered quantitative trait Loci.” *Genetics* 197(1): 405-20.

Morgan, T. H., 1911. “Random Segregation Versus Coupling in Mendelian Inheritance.” *Science* 34(873): 384.

Morran, L. T., O. G. Schmidt, I. A. Gelarden, R. C. Parrish II, and C. M. Lively. “Running with the Red Queen: host-parasite coevolution selects for biparental sex.” *Science* 333(6039): 216-8.

Myles, S., 2013. “Improving fruit and wine: what does genomics have to offer?” *Trends in Genetics* 29(4): 190-6.

Nachman, M. W., and B. A. Payseur, 2012. “Recombination rate variation and speciation: theoretical predictions and empirical results from rabbits and mice.” *Philosophical Transactions of the Royal Society B: Biological Sciences* 376: 409-21.

Nei, M., 1987. *Molecular Evolutionary Genetics*. Columbia University Press, New York.

Noor, M. A., 2008. “Connecting recombination, nucleotide diversity and species divergence in Drosophila.” *Fly (Austin)* 2(5):255-6.

Nosil, P., 2012. *Ecological Speciation*. Oxford University Press, Oxford.

Orr, H. A., 2005. “The genetic theory of adaptation: a brief history.” *Nature Reviews Genetics* 6: 119-127

Ortiz-Barrientos, D, J. Engelstaedter, and L. H. Riesenberg, 2016. “Recombination Rate Evolution and the Origin of Species.” *Trends in Ecology & Evolution* 31(3): 226-36.

Paradis, E., J. Claude, and K. Strimmer, 2004. “APE: analyses of phylogenetics and evolution in R language.” *Bioinformatics* 20: 289-90.

Rastas, P, F. C. F. Calboli, B. Guo, T. Shikano, and J. Merilä, 2016. “Construction of Ultradense Linkage Maps with Lep-MAP2: Stickleback F2 Recombinant Crosses as an Example.” *Genome Biology and Evolution* 8(1): 78-93.

Roesti, M., B. Kueng, D. Moser, and D. Berner, 2015. “The genomics of ecological vicariance in threespine stickleback fish.” *Nature Communications* 6: doi:10.1038/ncomms9767.

Roesti, M., D. Moser, and D. Berner, 2013. “Recombination in the threespine stickleback genome - Patterns and consequences.” *Molecular Ecology* 22(11): 3014-27.

Schultz, J., and H. Redfield, 1951. "Interchromosomal effects on crossing over in Drosophila." *Cold Spring Harbor Symposia on Quantitative Biology* 16: 175-97.

Sturtevant, A. H., 1913. "The Linear Arrangement of Six Sex-Linked Factors in Drosophila, as shown by their mode of Association" *Journal of Experimental Zoology* 14 43-59.

Wu, T. D. and C. K. Watanabe, 2005. “GMAP: a genomic mapping and alignment program for mRNA and EST sequences.” *Bioinformatics* 21: 1859-75.