**Selection, linkage, and population structure interact to shape genetic variation among threespine stickleback genomes**

Thomas C. Nelson1,2, Johnathan G. Crandall1, Catherine M. Ituarte1, Julian M Catchen1,3, and William A. Cresko1\*

1Institute of Ecology and Evolution, University of Oregon, Eugene, OR 97401

2Division of Biological Sciences, University of Montana, Missoula, MT 59812

3Department of Animal Biology, University of Illinois at Urbana-Champaign, Urbana, IL 61801

\*To whom correspondence should be addressed: [wcresko@uoregon.edu](mailto:wcresko@uoregon.edu)

DNA sequence are available under NCBI BioProjects PRJNA429207 and PRJNAXXXXXX and SRA BioSamples SAMN10498162-10498548. Processed data are available on FigShare.

Nitpicky To-do’s:

• decide and be consistent about, e.g., “Boot Lake” vs BT

• check all figure/table numbers

• create ANOVA summary tables for ecotype\*linkage interaction (π and θ)

Running title: Linked variation in sticklebacks

Keywords: Population genomics, linked selection, recombination, adaptation, threespine stickleback, *Gasterosteus aculeatus*

Corresponding Author:

William A Cresko

312 Pacific Hall

5289 University of Oregon

Eugene, OR 97403-5289

(541) 346-4779

wcresko@uoregon.edu

# Abstract

The outcome of selection on genetic variation depends on the geographic organization of individuals and populations as well as the syntenic organization of loci within the genome. Spatially variable selection between marine and freshwater habitats has had a significant and heterogeneous impact on patterns of genetic variation across the genome on threespine stickleback fish. When marine stickleback invade freshwater habitats, more than a quarter of the genome can respond to divergent selection, even in as little as 50 years. This process largely uses standing genetic variation that can be found ubiquitously at low frequency in marine populations, can be millions of years old, and is likely maintained by significant bidirectional gene flow. Here, we combine population genomic data of marine and freshwater stickleback from Cook Inlet, Alaska, with genetic maps of stickleback fish derived from those same populations to examine how linkage to loci under selection affects genetic variation across the stickleback genome. Divergent selection has had opposing effects on linked genetic variation on chromosomes from marine and freshwater stickleback populations: marine chromosomes are depauperate of variation nearby loci under selection while these same regions among freshwater genomes are the most genetically diverse. Forward genetic simulations recapitulate this pattern when different selective environments also vary in population structure. Lastly, dense genetic maps demonstrate that the interaction between selection and population structure may impact large stretches of the stickleback genome. These findings advance our understanding of how the structuring of populations across geography influences the outcomes of selection, and how the recombination landscape broadens the genomic reach of selection.­

# Introduction

Biologists have long known that natural populations harbor abundant genetic variation that is distributed heterogeneously across geography (cite). A recent wave of discoveries reveals that genetic variation is distributed heterogeneously across genomes as well (cite). Are these patterns related? For example, does the manner in which evolutionary processes play out across the geography of organisms influence the heterogeneous genomic patterns of genetic diversity? With the evolutionary genomics revolution the role of linked selection is being increasingly appreciated (Charlesworth *et al.* 1993; Hahn 2008; Langley *et al.* 2012; Schrider and Kern 2017). Linked positive or purifying selection can affect genetic variation far beyond the causal mutations to the surrounding genomic neighborhood (Elyashiv *et al.* 2016; Schrider and Kern 2017). Indeed, persistent and pervasive linked selection can structure genetic variation during speciation (Burri *et al.* 2015), leading to predictable genome-wide genetic divergence across multiple speciation events (Stankowski *et al.* 2018). Consequently, any process that can influence linkage across the genome, such as variation in recombination, will affect the outcome of linked selection on patterns of genetic variation(Begun and Aquadro 1992; Charlesworth *et al.* 1993; Gillespie 2000).

Interacting with the genomic context of physical linkage is the geographic context over which evolutionary processes play out (Charlesworth *et al.* 1997; Lenormand 2002; Stankowski *et al.* 2017). A single selective sweep in one population can eliminate genetic variation at a selected locus and linked variants (Maynard Smith and Haigh 1974; Fay and Wu 2000; Nielsen *et al.* 2005). However, in nature selective pressures tend to vary across time and space in ways that can maintain variation. (Clausen *et al.* 1941; Endler 1977; Lenormand *et al.* 1999). Temporally fluctuating selection can maintain alleles at intermediate frequencies, as has been observed in *Drosophila melanogaster* (Bergland *et al.* 2014) and the yellow monkeyflower, *Mimulus guttatus* (Troth *et al.* 2018). Local adaptation in a geographically structured species can lead to the partitioning of variation among populations and the maintenance of alternatively adaptive alleles (Charlesworth *et al.* 2003; Wallbank *et al.* 2016; Nelson and Cresko 2018) as has been observed in insects (Mettler *et al.* 1977; Nosil and Crespi 2006), plants (Clausen *et al.* 1941; Angert *et al.* 2018), and vertebrates (Hoekstra *et al.* 2004; Jones *et al.* 2018). Consequently, gene flow can strongly influence the geographic distribution of genetic variation among populations (Wright 1931; Slatkin 1985) and across species boundaries (Pardo-Diaz *et al.* 2012; Fontaine *et al.* 2015). Although genetic introgression from one species to another may impede neutral divergence, it can actually fuel adaptive diversification (Stankowski and Streisfeld 2015; Wallbank *et al.* 2016; Jones *et al.* 2018).

Even in the absence of selection, population subdivision alone may affect the overall abundance of neutral genetic variation within a species. Theoretical models provide useful - but at times contradictory - predictions on this matter, with subdivision either increasing or decreasing expected coalescence times depending on model specifications and assumptions. For example, Slatkin (**CITE**) generalized the island model to demonstrate that population structure increases coalescence times (TT) in a structured population by:

Where d is the number of demes, m is the migration rate, and T0 is the expected coalescence time with no population structure (using the nomenclature of **CITE**). Whitlock and Barton (**CITE**), however, used a similar model but allowed deme extinction and variable contributions of each deme to the total population. The result was a decrease in effective population size, Ne, with increasing subdivision, and thus a decrease in expected levels of genetic variation.

Therefore, integrated simulation models paired with empirical studies that examine the joint effects of variation in selection regimes, geographic context, and recombination landscape are necessary for fully understanding the effects of these biological factors on genomic patterns of genetic variation in the wild.

Here, we investigate how multiple evolutionary forces interact to shape the genomic and geographic distribution of genetic variation in the threespine stickleback fish, *Gasterosteus aculeatus*. The threespine stickleback is distributed holarctically in coastal marine, brackish and freshwater habitats. The large marine population has repeatedly given rise to derived freshwater populations, resulting in phenotypic divergence and parallel evolution throughout the species range (Bell and Foster 1994; Cresko *et al.* 2004; Jones *et al.* 2012). When marine stickleback invade freshwater habitats, more than a quarter of the genome can respond to the action of divergent selection (Bassham *et al.* 2018). This process largely uses standing genetic variation (Colosimo *et al.* 2005; Schluter and Conte 2009; Roesti *et al.* 2015; Samuk *et al.* 2017) that can be found ubiquitously at low frequency in marine populations (Bassham *et al.* 2018) much of which has likely been maintained for millions of years by significant bidirectional gene flow.

While previous work has often focused on the genomic targets of selection, we focus here on the processes that maintain and structure variation in regions of the genome physically linked to those targets. We use a combination of population genomics of wild stickleback, forward-time simulations using SLiM (Haller and Messer 2017), and dense genetic maps to support a model whereby differences in population structure between marine and freshwater habitats has led to divergent outcomes of molecular evolution at loci linked to adaptive variants. Our results provide new links between theoretical and empirical evolutionary genetics, new tools for future work in the stickleback system, and new perspectives on the maintenance of genetic variation.

# Methods

## Study populations and natural genetic variation

Wild threespine stickleback were collected from Rabbit Slough (N 61.5595, W 149.2583), Boot Lake (N 61.7167, W 149.1167), and Bear Paw Lake (N 61.6139, W 149.7539). Rabbit Slough is an offshoot of the Knik Arm of Cook Inlet and is known to be populated by anadromous populations of stickleback that are stereotypically marine in phenotype and genotype (Cresko *et al.* 2004). Boot Lake and Bear Paw Lake are both shallow lakes formed during the end-Pleistocene glacial retreat approximately 12 thousand years ago. Fish were collected in the summers of 2009 (Rabbit Slough), 2010 (Bear Paw Lake), and 2014 (Boot Lake) using wire minnow traps and euthanized *in situ* with Tricaine solution. Euthanized fish were immediately fixed in 95% ethanol and shipped to the Cresko Laboratory at the University of Oregon (Eugene, OR, USA).

We generated restriction site-associated DNA (RAD) libraries of five fish each from Rabbit Slough and Boot Lake and four fish from Bear Paw Lake. Genomic DNA was isolated from ethanol-preserved fin clips by proteinase K digestion followed by DNA extraction with solid phase reversible immobilization (SPRI) beads. We created RAD libraries using the single digest and shearing method of Baird et al (Baird *et al.* 2008) with the modifications of Nelson and Cresko (Nelson and Cresko 2018). Genomic DNA from each fish was digested with PstI-HF (New England Biolabs) and ligated to Illumina P1 adaptors with 6 bp inline barcodes. All barcodes differed by at least two positions, allowing for recovery of sequence reads with single errors in the barcode sequence. Ligated samples were then multiplexed at approximately equimolar concentrations and mechanically sheared via sonication to a fragment range of ~200-800 bp. Sheared DNA was size selected by extraction from a 1.25% agarose gel to generate a narrow insert size range of 425 bp to 475 bp. This size range allowed consistent overlap of paired-end Illumina reads for the construction of local contigs surrounding restriction enzyme cut sites. We then ligated Illumina P2 adaptors to the size-selected libraries and amplified P1/P2-adapted fragments with 12 cycles of PCR using Phusion-HF polymerase (New England Biolabs). RAD libraries were then sequenced in a single lane on an Illumina HiSeq 2500 to generate paired-end 250-bp sequence reads. All libraries generated for this study were sequenced at the University of Oregon’s Genomics and Cell Characterization Core Facility (GC3F: <http://gc3f.uoregon.edu>).

We used the Stacks analysis pipeline to process RAD sequence read pairs and call SNPs (Catchen *et al.* 2011; Catchen *et al.* 2013b). Raw reads were first demultiplexed without quality filtering using process\_radtags, and then quality filtered using process\_shortreads. This allowed for read trimming, rather than strict removal, if quality decreased toward the end of the first-end read. Overlapping read pairs were then merged using fastq-join (Aronesty 2011), allowing for up to 25% of bases in the overlapping region to mismatch, and the resulting contigs were trimmed to 350 bp. Any read pairs that failed to merge, or were shorter than 350 bp, were removed from further analysis. This step was required for processing reads through the Stacks pipeline. Below, we use the naming conventions of Baird *et al.* (2008): A “RAD tag” refers to sequence generated from a single end of a restriction site and the pair of RAD tags sequenced at a restriction site comprises a “RAD locus”.

All polymorphisms were called relative to the threespine stickleback reference genome v1.0 (Jones *et al.* 2012), using the updated scaffolding of Glazer, et al. (2015). Trimmed contigs were aligned to the reference using bbmap with the most sensitive settings (‘vslow=t’; http://jgi.doe.gov/data-and-tools/bbtools/). We then used the Stacks core pipeline to identify read stacks, call SNPs, and identify alleles and haplotypes based on genomic alignment (pstacks and cstacks); find homologous RAD tags across individuals (sstacks); and catalog biologically plausible haplotypes based on within- and among-individual haplotype variation (populations). We required that a RAD tag be present in all three populations and in at least four fish in each population.

We used the program PHASE (Stephens *et al.* 2001; Scheet and Stephens 2006) to combine sequence information from both RAD tags at a *PstI* cut site and generate phased haplotypes at each RAD locus. We wrote custom Python scripts to identify all unique haplotypes at each of a pair RAD tags and code them as alleles at a single, multiallelic locus. We required that each individual included in this analysis was genotyped at both RAD tags. Loci containing individuals only genotyped at a single RAD tag were removed from further analysis. RAD haplotypes at each locus represent 696 bp of contiguous genomic sequence, giving us high-quality estimates of sequence diversity and divergence even with our relatively small population-level sample sizes (Nei 1987, chapter 13; Cruickshank and Hahn 2014; Nelson and Cresko 2018). Sequencing of wild stickleback resulted in 57,992 RAD loci distributed across all 21 threespine stickleback chromosomes.

## Population genetic statistics

The scripting language R version 3.5 (R Core Team 2016) was used for all downstream data analysis. We estimated differentiation among threespine stickleback populations (all pairwise combinations) and among ecotypes (combined freshwater ponds versus Rabbit Slough) using a haplotype-based FST (equation 3 in Hudson, Slatkin et al. 1992) (Hudson *et al.* 1992) implemented in the R package ‘PopGenome’ (Pfeifer *et al.* 2014). We calculated π per site within and among populations and absolute sequence divergence (dXY) at each RAD locus by calculating pairwise distances between all RAD haplotypes with the R package ‘ape’ (Paradis *et al.* 2004; Popescu *et al.* 2012).

Previously, we detected patterns of reciprocal monophyly between marine and freshwater haplotypes using maximum clade credibility trees generated in BEAST v1.7 (Drummond and Rambaut 2007; Drummond *et al.* 2012; Nelson and Cresko 2018; Suchard *et al.* 2018). Tree topologies for all RAD loci were inferred from MCMC runs of 106 states with 10% burnin periods. We used blanket priors and parameters across all RAD loci, including a coalescent tree prior and the GTR+Γ substitution model. Monophyly of haplotypes from each population (RS, BT, BP) and each habitat (marine, freshwater) was assessed using the is.monophyletic() function of the R packages ‘ape’. Here, we use topological classifications that we inferred previously and designate gene trees with reciprocally monophyletic marine and freshwater haplogroups (1,129 of 57,992 RAD loci) as ‘divergent’ loci **(NelsonCresko, and Figure 1C)**.

## Forward simulations using SLiM

We used forward simulations implemented in SLiM (Haller and Messer 2017) to model the effects of selection, linkage, and population structure in a manner that reflects the stickleback metapopulation structure. We simulated a metapopulation of 2000 diploid individuals and a genome consisting of a single chromosome with a genetic length of ten centiMorgans (cM). The chromosome contained 50 kb of freely recombining sequence (recombination rate 1x10-6 per bp) on either side of a 2 kb nonrecombining ‘core’ containing the locally adaptive locus. Per-base mutation rate was kept constant at 5x10-7, resulting in a population-scaled mutation rate (4Nμ = 0.004) aligned with genome-wide estimates of genetic diversity in stickleback (**CITE**). The general form of the simulations was as follows:

1. Initialize a population, H1, of size 1000, and evolve for 10,000 generations
2. Create *k* new populations, each of size 1000/*k*, and evolve for 10,000 generations
3. Introduce locally adaptive mutation
4. Run selection for [4,10,20]N generations (If locally adaptive mutation is lost, go to [3])
5. Sample chromosomes and end simulation

Simulations included three phases: (1) a burn-in without population structure, (2) a burn-in after creating population structure, and (3) a selection phase contingent on establishment of a locally adaptive allele.

The first burn-in began with a single panmictic population of size NH1 = 1000 and no sequence variation. We evolved this population for 10NH1 (10,000) generations. The second burn-in of 10NH1 generations began by creating *k*H2 new populations, each of size NH1/kH2, by sampling variation from the existing population (Figure 2), where *k*H2 ranged from one to twenty-five. We set bidirectional migration rates equivalent to one migrant per generation between the existing population, now designated habitat H1, and each new population, now collectively referred to as habitat H2. We chose this migration structure to reflect the stickleback metapopulation, where freshwater stickleback populations are thought to be derived most often from marine ancestors, and because gene flow among freshwater populations primarily occurs through the marine population, especially across broader geographic scales.

To start the selection phase, we introduced a mutation at the center of the ‘core’ of a single chromosome in habitat H2 that is advantageous in H2 and deleterious in H1. Progression through the selection phase was conditional on establishment of the locally adaptive mutation: if the mutation was lost from the metapopulation, the simulation restarted at the end of the second burn-in.

Genetic diversity among chromosomes carrying H1- and H2-adaptive alleles was assessed at the end of the selection phase. We sampled 10 chromosomes of each allelic state and calculated nucleotide diversity (π) and Watterson’s θ (θW) within and among chromosomes of the two allelic states in non-overlapping 250 bp(0.025 cM) bins. We include results from θW because this statistic is based on the number of segregating sites and is therefore more reflective of the abundance of sequence variation and less affected by shifts in allele frequencies. We fold the calculations of genetic diversity about the locally adaptive locus to emphasize the effects of evolutionary forces on π at a given distance from the locus (Figure 2C). These summaries mirror our estimates of genetic variation on stickleback chromosomes and results from theoretical work on the effects of local adaptation on linked variation (Charlesworth *et al.* 1997).

## Laboratory crosses and genetic mapping

To compare how heterogenous genomic divergence on the physical map is distributed across the genetic map, we generated mapping families from laboratory lines of fish derived from the Boot Lake and an F1 hybrid female produced from a cross between a Boot Lake female and a Rabbit Slough male. These crosses allowed us to examine variation in the recombinational landscape within and among chromosomes in distinct, evolutionarily relevant genetic backgrounds. We also generated a genetic map of a Rabbit Slough-derived male (**Figures SXX**) but, because ecotype was confounded with sex, we limit our discussion to the Boot Lake and hybrid maps.

All maps were constructed using a pseudo-testcross design, which takes advantage of existing heterozygosity in outbred individuals without the need to generate inbred lines or F1 mapping parents. To generate the Boot Lake mapping family, we manually crossed unrelated lab-reared individuals. We mapped the F1 hybrid female by backcrossing it to a Boot Lake male. All progeny were raised to 14 days post-fertilization, euthanized with MS-222 (Sigma Aldrich), and fixed in 95% ethanol. We extracted genomic DNA from whole progeny and from pectoral and caudal fins from all parents using proteinase K digestion (Qiagen) followed by DNA purification with SPRI beads.

RAD genotyping of progeny and parents was used to identify segregating haplotypes using a RAD-seq protocol similar to that described previously but using the restriction enzyme *SbfI*. RAD-seq data from all mapping crosses were processed with the Stacks analysis pipeline (Catchen *et al.* 2013b). We demultiplexed and quality filtered sequences with process\_shortreads and aligned them to the stickleback reference genome with GSNAP (Wu and Nacu 2010). We used ref\_map.pl to identify RAD tags and call genotypes. The Stacks component program genotypes was used identify segregating markers for export to genetic mapping software. We specified a minimum coverage of 3x to call individual genotypes and required that a marker be genotyped in at least 50% of progeny.

Below, we present data for the female parent from each mapping cross (Tables S1 and 1). By conducting pseudo-testcrosses, we identified polymorphic RAD markers segregating in all mapping parents. However, to investigate the genome-wide recombination landscape, as well as relationships between recombination rate and natural levels of polymorphism and divergence, we restricted our analysis to parents of the same sex for which we observed segregating markers on all 21 chromosomes with no gaps of more than 1 megabase (Mb) between adjacent markers.

We estimated genome-wide recombination rates between RAD markers under the assumption of collinearity between all genetic maps and the stickleback reference genome (Glazer *et al.* 2015) (with some exceptions, see below). We used the mapping software Lep-MAP2 (Rastas *et al.* 2015) to estimate map positions of RAD markers with the marker order fixed to the aligned positions on the reference genome. Known marker orders increased throughput of mapping iterations and reduced the impact of genotyping errors on recombination rate estimation.

While fixed marker orders do not explicitly allow the detection of structural variation among genomes — such as chromosomal inversions that are known to exist among stickleback populations — discrepancies in the estimated map do provide indirect, and correctable, evidence of changes to map order (Figure S1). For example, inversions in genetic map order relative to the reference order spuriously inflate recombination rates between markers closely flanking inversion breakpoints when inverted segments are forced into the wrong orientation. This is because genetic map distances are estimated independently while the physical distance between markers is drastically underestimated. The observed jumps in map distance on either side of the inversion are equal to each other and to the total map length of the inversion. Reversing the marker order within the inversion removes these artifacts and reduces the overall map length of the region.

## Recombination-polymorphism correlations

We employed three methods to investigate the relationships between the recombinational landscape and patterns of polymorphism within and among natural populations. To quantify genome-wide correlations among recombination rate and population genetic statistics, we divided the stickleback genome into non-overlapping windows and calculated average recombination rates (in centiMorgans per megabase, cM/Mb), sequence diversity (πBT, πRS, and π), and genetic divergence (FST, dXY) in each window **(table…)**. RAD marker density made estimates of local recombination rates less reliable and more variable when using small window sizes (**e.g. 100-kb windows;** Figures S3 and S4). As a result, we show results from 1000-kb genomic windows. We used nonparametric correlations to test for correlations between variables because the distributions of most variables lacked normality even using standard data transformations. Below we present Spearman’s rank order correlations. Kendall’s tau and parametric linear models gave qualitatively similar results.

Genomic heterogeneity exists not only in the proportion of the genome affected by divergent selection but also in how genetic variation and divergence are clustered within the genome. Marine-freshwater genomic divergence in stickleback is clustered into few, large regions that can encompass most of the length of a chromosome. We sought to directly compare the genomic distributions of population genetic statistics along the physical genome and on the genetic maps we constructed. We used a windowing approach that allowed direct comparisons across maps despite differing numbers and distributions of markers among genetic maps. Using the R package ‘ksmooth’, we binned each chromosome into equally sized intervals, the number of which we set equal to the number of segregating RAD markers on the genetic map with fewest markers. For each interval, we calculated genetic position from each laboratory cross and FST between Rabbit Slough and Boot Lake populations (Hohenlohe *et al.* 2010; Nelson and Cresko 2018) within a 250-kb normally distributed kernel. We also imputed the genetic positions of all divergent loci using the lm() function in R: we found the nearest flanking RAD markers in each mapping cross and predicted the genetic position of the divergent locus assuming a constant recombination rate between the flanking markers. With these approaches we were able to make direct comparisons between recombination within and among genetic maps from different genetic backgrounds and patterns of polymorphism in the populations from which the laboratory lines were derived.

## Data availability

Raw sequence data will be made available on NCBI under BioProjects PRJNA429207 and PRJNAXXXXXX. RAD sequences for mapping families are available on the Sequence Read Archive, BioSamples SAMN10498162-10498548. Scripts and processed data will be available on FigShare and are immediately available on GitHub at <https://github.com/thomnelson/linkedvariation>.

## Animal care and compliance

# Results

## Chromosomes from freshwater but not marine stickleback contain abundant linked variation

To examine patterns of variation linked to loci under selection, we first partitioned RAD loci from the population genomics dataset into those with evidence of complete marine-freshwater lineage sorting (‘divergent’, with allelic states ‘marine’ and ‘freshwater’, see Figure 2B), those on the same chromosome as a divergent locus (‘linked’), and those on chromosomes without divergent loci (‘unlinked’). At divergent loci, average sequence diversity was partitioned almost entirely among chromosomes carrying marine and freshwater alleles; overall π averaged 0.0067 per site (Figure 3A, black-filled diamond, genome-wide mean = 0.0042); average dXY between marine and freshwater alleles was 0.0124, nearly three-fold higher than the genome-wide average (0.0044). In contrast, and as expected at locally adaptive loci (Figure 2C), π was substantially reduced within allelic states; π among marine alleles averaged 0.0012 per site (Figure3­B, red-filled diamond) while average π among freshwater alleles at divergent loci was 0.0015 per site.

Patterns of linked variation among habitats and on marine chromosomes followed the expectations from simulated local adaptation (Figure 2C, Figure 3A,B). Sequence diversity at linked loci was highly correlated with proximity to a divergent locus (Figure 3A-C, Spearman’s ρ: all p-values ≤ 10-10). When all populations were combined, π at linked loci decreased sharply in the first approximately 250 kb from a divergent locus (Figure 3A,D), and linked loci in closest proximity to divergent loci were nearly as polymorphic as those directly impacted by divergent selection. As expected under local adaptation, π among chromosomes sampled from marine Rabbit Slough stickleback was lowest in close proximity to divergent loci (Figure 3B), and a substantial fraction (12%) of RAD loci with 250 kb of a divergent locus showed no variation at all on these chromosomes (πRS = 0; 1651/13762 loci).

In stark contrast, diversity among freshwater chromosomes showed a proximity effect that was distinct from either Rabbit Slough or the combined populations (Figure 3C-E). Rather than being lowest near divergent loci, diversity actually increased with proximity to a divergent locus, peaking approximately 200 kb away on average before reversing direction. This pattern persisted using either π or Watterson’s theta (θW), indicating that the density of segregating sites on freshwater chromosomes is highest near divergent loci and that the signal we observe is not simply due to a greater abundance of intermediate frequency alleles. We also note that this increase in diversity qualitatively persisted within both freshwater populations individually **(Suppl Fig)**.

Linkage to divergent loci, therefore, was associated with opposing effects on genetic variation in stickleback: decreasing it among marine chromosomes while increasing variation among freshwater chromosomes (Figure 3E). Chromosomes with no evidence of divergent selection had a slightly higher density of segregating sites in the Rabbit Slough population than the combined freshwater populations (Figure 3E), though π was indistinguishable (**Figure SX**). However, within 500 kb of a divergent locus, genetic diversity (π and θW) was greater among freshwater populations than in Rabbit Slough (Figure 3F; ANOVA: population\*proximity interaction p ≤ 10-10).

## Population structure maintains linked variation on simulated chromosomes

To determine how population structure within the freshwater habitat influences patterns of linked variation in stickleback, we conducted forward simulations of chromosomes under a model of local adaptation with migration using SLiM v2.6 (**CITE &** Figures 1 and 4). In simulations with two habitat types, each composed of a single, panmictic population (Figure 4, column 1), total genetic diversity was highest at and adjacent to the locally adaptive locus (Figure 2C). Within allelic classes, diversity was lowest in proximity to the locally adaptive locus and recovered equally within both allelic classes with increasing recombinational distance from the selected locus (Figure 2C; Figure 4). This scenario is essentially the same as that presented by Charlesworth, Nordborg, and Charlesworth (1997) and the results are qualitatively the same.

Adding to these simulations population structure informed by the stickleback system led to previously undocumented results. When the second habitat (H2) – representing the freshwater stickleback habitat – consisted of two or more demes (Figure 2A; Figure 4, columns 2 and 3) the additional structure maintained greater variation exclusively on H2-adaptive (H2+) chromosomes, an effect dependent on the duration and strength of selection and the number of demes. Under strong selection (s = 0.20), population structure had a barely noticeable effect on variation until the length of the selection phase was on the order of the neutral coalescence time (~9NH generations, **CITE SLATKIN, CC&B**). Beyond this point, any amount of population structure resulted in higher levels of variation on H2+ than H1+ chromosomes within ~0.2 cM from the selected locus. Moderate selection (s=0.02) was generally ineffective at altering among-habitat levels of variation, though we did observe a modest effect on H2+ chromosomes immediately adjacent to the selected locus.

Increasing population structure in H2 led to substantial levels of variation near the selected locus (Figure 6, column 3; Figure **SX**). When H2 consisted of five demes, genetic variation on H2+ chromosomes was greatest ~0.1 cM from the selected locus even though levels of variation among allelic states remained indistinguishable in distal regions of the chromosome. Greater subdivision of H2 first accentuated the accumulation of variation (5 ≤ *k* ≤ 15, Figure SX) but eventually attenuated it when *k* ≥ 20. While we did fully not explore this attenuation, we note that we held both *N* and *m* (the average number of migrants per generation) constant across simulations; higher values of *k* were therefore accompanied by greater total migration between habitats and stronger within-deme drift in H2.

The interaction we observed between selection and geographic structure in simulated populations closely mimicked the patterns found in stickleback, but the extent of this effect seemed limited. Even in the case of strong selection (s=0.20) and substantial population structure, increases in genetic diversity extended less than 0.5 cM away from the locus under selection; maximal levels of variation required even tighter linkage. We therefore examined how recombination rate varies across stickleback genomes to better understand the extent to which selection may influence linked genetic variation.

## The recombination landscape varies among individuals

RAD sequencing of progeny used to create the three genetic maps generated 740,331,821 raw single-end sequence reads distributed among 385 individuals. Of these, 520,990,827 (70.4%) passed quality filtering, and 370,490,044 (71.1%) of those were unambiguously aligned to the reference genome. Both mapping crosses resulted in over 6000 segregating markers (Table 1). Per-locus sequencing depth averaged 23X and 15X for the Boot Lake and hybrid mapping families, respectively. Sequencing results of the families analyzed for each map are summarized in Table S1.

Patterns of recombination were generally consistent between the genetic maps (Figure 5, Suppl. Figure S2). As has been described previously (Roesti *et al.* 2013; Glazer *et al.* 2015), recombination on the larger metacentric chromosomes in the stickleback genome (e.g chromosomes 4 and 7) was biased toward chromosome ends, with little recombination occurring across central, presumably centromeric, regions. Recombination rates across a number of the smaller chromosomes, in contrast, was typically highest toward one end; this likely also reflects a trend against recombination near the centromere (Figure 5, chromosome 15).

We also observed important differences between genetic maps. As expected, recombination in the hybrid map was completely suppressed within and immediately surrounding an inversion on chromosome 21 but occurred freely in the collinear BT map (Figure 5). On chromosome 1, the BT map showed the bias toward chromosome ends expected on larger chromosomes but this bias was mostly absent in the hybrid map.

## Much of the genome is tightly linked to loci under divergent selection

Genomic regions of greatest differentiation were compressed into proportionally smaller regions of the genetic maps (Figure 6). Some of the largest regions of differentiation we observed — including those surrounding *eda*, the major effect locus for lateral plate number (Colosimo *et al.* 2005), and the chromosome 21 inversion — had average FST values in excess of 0.4. Regions above this threshold spanned 8.1% of the physical genome (35.3 Mb) but only 3.7% of the BT genetic map and 3.3% of the F1 hybrid genetic map (**Figure SX**). On chromosome 4, the large central regions of differentiation are similarly compressed on both genetic maps but this was not always the case. Chromosomal inversions specifically suppress recombination in heterokaryotypes. On chromosome 21, recombination was suppressed – and genomic differentiation compressed – only on the hybrid map (**Figure SX**). A similar effect was seen across a large region of differentiation on chromosome 7 (Figure 6B) despite this chromosome being collinear between marine and freshwater stickleback genomes (Figure **SX**). On the physical map, chromosome 7 contained three peaks of strong differentiation between BT and RS spanning over 9 Mb. All three peaks were clearly separated on the BT map, though they comprise a relatively smaller region than on the physical map, but collapsed to a single locus on the hybrid map (span: 0 cM, position 57.7 cM on the map).

Because our simulations suggested that population structure could interact with selection to maintain variation, but only with tight linkage (≤0.2 cM), we imputed the positions of RAD loci from the population genomic dataset onto the BT and hybrid genetic maps. Based on average genome-wide recombination rates, 0.2 cM equates to less than 50 kb of sequence space (BT map: 36 kb; F1 map: 44 kb) and includes ~6% of sequenced loci. However, the clustering of divergent loci in regions of low recombination resulted in ~20% of all loci in the dataset fitting this linkage criterion (BT map: 19.5%; F1 map: 20.4%; Figure 6C,D; **Figure SX**). These tightly linked loci averaged over 200 kb away from a divergent locus (BT map: 219 kb; F1 map: 256 kb) and included loci over 2 Mb away (max, BT map: 2,346 kb max; max, F1 map: 2,900 kb).

# Discussion

ing of theacts as athat this effectpotentIn genomic regions that are recombinationally distant to targets of divergent selection, gene flow among habitats homogenizes variation such that levels of genetic variation are similar (Figure 3) and differentiation is low (Figure 6) across habitats. At divergent loci themselves, selection has maintained a small number of haplotypes in each habitat and variation is partitioned almost entirely among habitats (Figure 1, Figure 3, NELSONCRESKo). However, in the substantial fraction of the genome that is tightly linked to divergent loci, we hypothesize that selection has reduced the effective migration rate between alternative habitats such that the geographic structuring of the freshwater habitat becomes a significant factor affecting levels of genetic variation.

. Put another way, we find no need to invoke more complex selection regimes in order to explain the patterns of variation we observe

Our novel findings align but extend previous population genomic studies in threespine stickleback.

## A variable recombination landscape simplifies the architecture of divergence and increases the genomic reach of linked selection

In addition to the consequences of asymmetric population structure in maintaining genetic variation unevenly across geography, the recombination landscape also can have profound effects on heterogenous genome-wide patterns of genetic variation. Linkage to or among selected sites is now thought to affect many, if not most, variable sites in the genome (Schrider and Kern 2017; Kern and Hahn 2018), and suppression of recombination among the genomes of diverging populations and species appears in large part to determine patterns of genetic differentiation and divergence (Burri *et al.* 2015; Vijay *et al.* 2017; Stankowski *et al.* 2018). In the threespine stickleback, two related patterns point to the importance of the recombination landscape in structuring the genetics of divergence. First, genetic differentiation among phenotypically divergent populations is known to accumulate in regions of low recombination (Roesti *et al.* 2013). Second, quantitative trait locus (QTL) mapping studies consistently find highly pleiotropic QTL associated with clusters of markers in the centers of relatively few linkage groups (Albert *et al.* 2008; Miller *et al.* 2014; Marques *et al.* 2016). This finding could be explained by the presence of a one or a small number of pleiotropic loci of large effect in each region, or clusters of linked loci.

By combining population genomics with multiple mapping crosses we are able to corroborate and further our understanding in stickleback fish and other organisms in two distinct ways. First, genomic divergence that occurs in the centers of the largest chromosomes occurs within regions of consistently low recombination across all three genetic maps (freshwater, marine, and F1 hybrid fish). The negative association between divergence and recombination rate is a common finding across systems (Burri *et al.* 2015; Martin *et al.* 2016) and our results further demonstrate that stickleback are no exception to this rule both at the local (i.e. specific genomic regions, Figure 4) and genome-wide scales (Figure 3). In addition, by directly comparing the genetic maps of purely marine and freshwater fish, our results do not indicate the presence of any large, obvious structural variation between marine and freshwater stickleback genomes associated with the largest regions of genomic divergence in addition to three large inversions identified previously (Jones 2012).

Second, although there was general congruence among the genetic maps, we find that the recombination landscape varied unexpectedly and in evolutionarily relevant ways. Within the large regions of divergence at the centers of chromosomes, low recombination ranged from being universal across the three cross types to genotype-specific (Figure 2, chromosomes 4 and 1, respectively), suggesting different genetic bases to the low recombination rates across these large genomic regions. Interestingly, recombination rates in regions of divergence were often lowest in the F1 hybrid map (Figure 4). While this was expected within chromosomal inversions, where the alternative homozygotes demonstrated steady recombination throughout the inverted region, this was also the case on chromosome 7 despite the apparent lack of any large scale, simple structural variation (Figures 2 and 4) although clusters of smaller structural variants that could not be detected in our maps could also reduce recombination. On the hybrid map, this entire region collapsed to a region inherited essentially as a single Mendelian locus.

The specific reductions in recombination on the hybrid map are suggestive of adaptive evolution of the recombination landscape itself and, if true, could have profound implications for adaptation and genomic divergence in stickleback. For example, inversions are advantageous when recombination between multiple linked alleles that contribute to fitness (either in an additive or epistatic fashion) is maladaptive (Kirkpatrick and Barton 2006). The three previously identified inversions on chromosomes 1, 11 and 21 (Jones *et al.* 2012) are associated with divergence between the freshwater and marine habitats and occur in regions of the genome that readily recombine in the chromosomal homozygotes (BT and RS maps). Had inversions not evolved, gene flow and recombination among marine and freshwater populations may have been strong enough to limit adaptive divergence (Lenormand 2002; Yeaman and Whitlock 2011; Aeschbacher *et al.* 2017).

A noteworthy conclusion from our data is that such large, simple inversions have not evolved across the largest regions of divergence in the genome. In these regions, low recombination rates — which may themselves have evolved adaptively or been ancestral platforms for adaptive divergence — combined with strong selection has been effective at maintaining allelic combinations across megabases of genomic space. Discovering the molecular mechanisms of such recombination rate reduction will be a productive avenue of research. Regardless of the molecular mechanisms, the packaging of variation across the genome into tightly linked (even Mendelian) loci simplifies the genetic architecture of adaptation and increases the impact of linked selection.

Our observations of coincidence of reduced recombination and genomic divergence may help explain the well documented pattern repeatable and rapid adaptive divergence in stickleback (Bell and Foster 1994; Lescak *et al.* 2015) that is largely the result of the reuse of standing genetic variation (Colosimo *et al.* 2005; Schluter and Conte 2009; Terekhanova *et al.* 2014; Roesti *et al.* 2015; Bassham *et al.* 2018). Because freshwater populations are typically founded by marine stickleback, the sources of standing genetic variation are low frequency variants in the marine population that nearly always exist in a heterozygous state with a marine genome. Our results suggest that the recombination landscape may therefore facilitate the maintenance of freshwater haplotypes during their transit through the marine environment by reducing recombination even in collinear genomic regions. Multi-megabase haplotypes — potentially containing many alleles contributing to local adaptation — then have a higher probability of being selected in concert when a new freshwater population is established.

## Conclusions

Divergent natural selection is a powerful force for the maintenance of genetic variation in nature. When local adaptation plays out on variable geographic and recombinational landscapes, we find that the effect of selection on genetic variation are amplified and shape patterns of linked variation in unexpected ways. Here we document using simulation and empirical studies that asymmetric population subdivision among habitats in stickleback leads to an overall greater maintenances of diversity in freshwater as compared to the larger ancestral marine population. Furthermore, we hypothesize that the stickleback recombinational landscape is the product of repeated adaptive evolution that transforms a large genomic architecture of marine-freshwater divergence into a much more simplified genetic architecture. If true, two consequences are the efficient maintenance of adaptive divergence in the face of gene flow, and widespread linked selection that eliminates genetic variation in the panmictic marine population while maintaining genetic diversity in the freshwater habitat. The interaction of selection, recombination, and population structure has turned small, isolated freshwater stickleback populations into the primary reservoirs of standing genetic variation that potentiate future evolutionary change. More broadly, our results underscore how integrated studies of evolutionary and genetic processes can yield exciting, unexpected patterns and deeper understanding when we jointly consider the processes and their interactions with one another.

# Author Contributions

TCN, JMC, and WAC conceived of and designed the study. TCN, JGC, CMI, and JMC performed mapping crosses and prepared sequencing libraries. TCN, JGC, and JMC performed linkage mapping. TCN performed population genomic analyses and conducted population genetic simulations. TCN and WAC wrote the paper.

# Acknowledgements

We thank Patrick Phillips, John Postlethwait, Kirstin Sterner, Matt Streisfeld, and members of the Cresko Laboratory for their guidance and advice. Thanks go to Sean Stankowski, Madeline Chase, Emily Beck, and Kristin Alligood for useful discussions during the development of this project. And thank you to Peter Ralph, Nadia Singh, and anonymous reviewers for their comments on and improvements to this manuscript. We acknowledge National Science Foundation awards NSF DEB 1501423 (W.A.C. and T.C.N.), NSF DEB 0919090 (W.A.C.), and National Institutes of Health award NIH T32GM007413 (T.C.N.).

**Literature Cited**

Aeschbacher, S., J. P. Selby, J. H. Willis and G. Coop, 2017 Population-genomic inference of the strength and timing of selection against gene flow. *Proc. Natl. Acad. Sci. USA* **114:** 7061-7066.

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

Angert, A. L., M. Bayly, S. N. Sheth and J. R. Paul, 2018 Testing Range-Limit Hypotheses Using Range-Wide Habitat Suitability and Occupancy for the Scarlet Monkeyflower (Erythranthe cardinalis). *Am. Nat.* **191:** E76-E89.

Aronesty, E., 2011 ea-utils: Command-line tools for processing biological sequencing data.

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

Bassham, S., J. Catchen, E. Lescak, F. A. von Hippel and W. A. Cresko, 2018 Repeated selection of alternatively adapted haplotypes creates sweeping genomic remodeling in stickleback. *GENETICS* **209**: 921-939.

Begun, D. J., and C. F. Aquadro, 1992 Levels of naturally occurring DNA polymorphism correlate with recombination rates in D. melanogaster. *Nature* **356:** 519-520.

Bell, M. A., and S. A. Foster, 1994 Introduction to the evolutionary biology of the threespine stickleback, pp. 1-27 in *The Evolutionary Biology of the Threespine Stickleback*, edited by M. A. Bell and S. A. Foster. Oxford University Press, New York.

Bergland, A. O., E. L. Behrman, K. R. O'Brien, P. S. Schmidt and D. A. Petrov, 2014 Genomic Evidence of Rapid and Stable Adaptive Oscillations over Seasonal Time Scales in Drosophila. *PLoS Genet.* **10**: e1004775.

Burri, R., A. Nater, T. Kawakami, C. F. Mugal, P. I. Olason *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 Res.* **25:** 1656-1665.

Catchen, J., S. Bassham, T. Wilson, M. Currey, C. O'Brien *et al.*, 2013 The population structure and recent colonization history of Oregon threespine stickleback determined using restriction-site associated DNA-sequencing. *Mo.l Eco.l* **22:** 2864-2883.

Catchen, J., P. A. Hohenlohe, S. Bassham, A. Amores and W. A. Cresko, 2013 Stacks: an analysis tool set for population genomics. *Mol. Ecol.* **22:** 3124-3140.

Catchen, J. M., A. Amores, P. Hohenlohe, W. Cresko and J. H. Postlethwait, 2011 Stacks: building and genotyping Loci de novo from short-read sequences. *G3 (Bethesda)* **1:** 171-182.

Charlesworth, B., D. Charlesworth and N. H. Barton, 2003 The effects of genetic and geographic structure on neutral variation. *Ann. Rev. Ecol. Evol. Syst.* **34:** 99-125.

Charlesworth, B., M. T. Morgan and D. Charlesworth, 1993 The Effect of deleterious mutations on neutral molecular variation. *GENETICS* **134:** 1289-1303.

Charlesworth, B., M. Nordborg and D. Charlesworth, 1997 The effects of local selection, balanced polymorphism and background selection on equilibrium patterns of genetic diversity in subdivided populations. *Genet. Res.* **70:** 155-174.

Clausen, J., W. M. Keck and W. M. Hiesey, 1941 Regional differentiation in plant species. *Am. Nat.* **75:** 231-250.

Colosimo, P. F., K. E. Hosemann, S. Balabhadra, G. Villarreal, Jr., M. Dickson *et al.*, 2005 Widespread parallel evolution in sticklebacks by repeated fixation of *Ectodysplasin* alleles. *Science* **307:** 1928-1933.

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

Cruickshank, T. E., and M. W. Hahn, 2014 Reanalysis suggests that genomic islands of speciation are due to reduced diversity, not reduced gene flow. *Mol. Ecol.* **23:** 3133-3157.

Drummond, A. J., and A. Rambaut, 2007 BEAST: Bayesian evolutionary analysis by sampling trees. *BMC Evol. Biol.* **7:** 214.

Drummond, A. J., M. A. Suchard, D. Xie and A. Rambaut, 2012 Bayesian phylogenetics with BEAUti and the BEAST 1.7. *Mol. Biol. Evol.* **29:** 1969-1973.

Elyashiv, E., S. Sattath, T. T. Hu, A. Strutsovsky, G. McVicker *et al.*, 2016 A Genomic map of the effects of linked selection in *Drosophila*. *PLoS Genet.* **12**: e1006130.

Endler, J. A., 1977 *Geographic Variation, Speciation, and Clines*. Princeton University Press.

Fay, J. C., and C. I. Wu, 2000 Hitchhiking under positive Darwinian selection. *GENETICS* **155:** 1405-1413.

Fontaine, M. C., J. B. Pease, A. Steele, R. M. Waterhouse, D. E. Neafsey *et al.*, 2015 Mosquito genomics. Extensive introgression in a malaria vector species complex revealed by phylogenomics. *Science* **347:** 1258524.

Gillespie, J. H., 2000 Genetic drift in an infinite population: The pseudohitchhiking model. *GENETICS* **155:** 909-919.

Gillespie, J. H., 2004 *Population genetics : a concise guide*. Johns Hopkins University Press, Baltimore, Md.

Glazer, A. M., E. E. Killingbeck, T. Mitros, D. S. Rokhsar and C. T. Miller, 2015 Genome assembly improvement and mapping convergently evolved skeletal traits in sticklebacks with genotyping-by-sequencing. *G3 (Bethesda)* **5:** 1463-1472.

Hahn, M. W., 2008 Toward a selection theory of molecular evolution. *Evolution* **62:** 255-265.

Haldane, J. B. S., 1930 A mathematical theory of natural and artificial selection. (Part VI, Isolation.). *Mathematical Proc. Cam. Phil. Soc.* **26:** 220-230.

Haller, B. C., and P. W. Messer, 2017 SLiM 2: flexible, interactive forward genetic simulations. *Mol. Biol. Evol.* **34:** 230-240.

Hoekstra, H. E., K. E. Drumm and M. W. Nachman, 2004 Ecological genetics of adaptive color polymorphism in pocket mice: geographic variation in selected and neutral genes. *Evolution* **58:** 1329-1341.

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

Hudson, R. R., 1990 Gene genealogies and the coalescent process, pp. 1-44 in *Oxford Surveys in Evolutionary Biology*, edited by D. J. Futuyma and J. Antonovics. Oxford University Press, Oxford.

Hudson, R. R., M. Slatkin and W. P. Maddison, 1992 Estimation of levels of gene flow from DNA sequence data. *GENETICS* **132:** 583-589.

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

Jones, M. R., L. S. Mills, P. C. Alves, C. M. Callahan, J. M. Alves *et al.*, 2018 Adaptive introgression underlies polymorphic seasonal camouflage in snowshoe hares. *Science* **360:** 1355-1358.

Kaeuffer, R., C. L. Peichel, D. I. Bolnick and A. P. Hendry, 2012 Parallel and nonparallel aspects of ecological, phenotypic, and genetic divergence across replicate population pairs of lake and stream stickleback. *Evolution* **66:** 402-418.

Kern, A. D., and M. W. Hahn, 2018 The neutral theory in light of natural selection. *Mol. Biol. Evol.* **35:** 1366-1371.

Kirkpatrick, M., and N. Barton, 2006 Chromosome inversions, local adaptation and speciation. *GENETICS* **173:** 419-434.

Langley, C. H., K. Stevens, C. Cardeno, Y. C. Lee, D. R. Schrider *et al.*, 2012 Genomic variation in natural populations of *Drosophila melanogaster*. *GENETICS* **192:** 533-598.

Leaver, S. D., and T. E. Reimchen, 2012 Abrupt changes in defence and trophic morphology of the giant threespine stickleback (*Gasterosteus* sp.) following colonization of a vacant habitat. *Biol. J. Linn. Soc.* **107:** 494-509.

Lenormand, T., 2002 Gene flow and the limits to natural selection. *TREE* **17:** 183-189.

Lenormand, T., D. Bourguet, T. Guillemaud and M. Raymond, 1999 Tracking the evolution of insecticide resistance in the mosquito *Culex pipiens*. *Nature* **400:** 861-864.

Lescak, E. A., S. L. Bassham, J. Catchen, O. Gelmond, M. L. Sherbick *et al.*, 2015 Evolution of stickleback in 50 years on earthquake-uplifted islands. *Proc Natl Acad Sci USA* **112:** E7204-7212.

Lewontin, R. C., 1974 *The genetic basis of evolutionary change*. Columbia University Press, New York.

Marques, D. A., K. Lucek, J. I. Meier, S. Mwaiko, C. E. Wagner *et al.*, 2016 Genomics of rapid incipient speciation in sympatric threespine stickleback. *PLoS Genetics* **12:** e1005887.

Martin, S. H., M. Most, W. J. Palmer, C. Salazar, W. O. McMillan *et al.*, 2016 Natural selection and genetic diversity in the butterfly *Heliconius melpomene*. *GENETICS* **203:** 525-541.

Maynard Smith, J., and J. Haigh, 1974 The hitch-hiking effect of a favourable gene. *Genet. Res.* **23:** 23-35.

Mettler, L. E., R. A. Voelker and T. Mukai, 1977 *Inversion clines in populations of Drosophila melanogaster*. *GENETICS* **87:** 169-176.

Miller, C. T., A. M. Glazer, B. R. Summers, B. K. Blackman, A. R. Norman *et al.*, 2014 Modular skeletal evolution in sticklebacks Is controlled by additive and clustered quantitative trait loci. *GENETICS* **197:** 405-420.

Nei, M., 1987 *Molecular evolutionary genetics*. Columbia university press.

Nelson, T. C., and W. A. Cresko, 2018 Ancient genomic variation underlies repeated ecological adaptation in young stickleback populations. *Evolution Letters* **2:** 9-21.

Nielsen, R., S. Williamson, Y. Kim, M. J. Hubisz, A. G. Clark *et al.*, 2005 Genomic scans for selective sweeps using SNP data. *Genome Res.* **15:** 1566-1575.

Nosil, P., and B. J. Crespi, 2006 Experimental evidence that predation promotes divergence in adaptive radiation. *Proc. Nat. Acad. Sci. USA* **103:** 9090-9095.

Paradis, E., J. Claude and K. Strimmer, 2004 APE: Analyses of phylogenetics and evolution in R language. *Bioinformatics* **20:** 289-290.

Pardo-Diaz, C., C. Salazar, S. W. Baxter, C. Merot, W. Figueiredo-Ready *et al.*, 2012 Adaptive introgression across species boundaries in *Heliconius* butterflies. *PLoS Genet.* **8**:e1002752.

Pfeifer, B., U. Wittelsburger, S. E. Ramos-Onsins and M. J. Lercher, 2014 PopGenome: An efficient Swiss army knife for population genomic analyses in R. *Mol. Biol. Evol.* **31:** 1929-1936.

Popescu, A. A., K. T. Huber and E. Paradis, 2012 ape 3.0: New tools for distance-based phylogenetics and evolutionary analysis in R. *Bioinformatics* **28:** 1536-1537.

Rastas, P., F. C. Calboli, B. Guo, T. Shikano and J. Merila, 2015 Construction of ultradense linkage maps with Lep-MAP2: stickleback F2 recombinant crosses as an example. *Genome Biol*. *Evol*. **8:** 78-93.

Reimchen, T. E., 1994 Predators and morphological evolution in threespine stickleback, pp. 240-276 in *The Evolutionary Biology of the Threespine Stickleback*, edited by M. A. Bell and S. A. Foster. Oxford University Press, New York.

Roesti, M., S. Gavrilets, A. P. Hendry, W. Salzburger and D. Berner, 2014 The genomic signature of parallel adaptation from shared genetic variation. *Mol. Ecol.* **23:** 3944-3956.

Roesti, M., B. Kueng, D. Moser and D. Berner, 2015 The genomics of ecological vicariance in threespine stickleback fish. *Nat comm.* **6:** 8767.

Roesti, M., D. Moser and D. Berner, 2013 Recombination in the threespine stickleback genome - Patterns and consequences. *Mol. Ecol.* **22:** 3014-3027.

Samuk, K., G. L. Owens, K. E. Delmore, S. E. Miller, D. J. Rennison *et al.*, 2017 Gene flow and selection interact to promote adaptive divergence in regions of low recombination. *Mol. Ecol.* **26:** 4378-4390.

Scheet, P., and M. Stephens, 2006 A fast and flexible statistical model for large-scale population genotype data: applications to inferring missing genotypes and haplotypic phase. *Am. J. Human Genet.* **78:** 629-644.

Schluter, D., and G. L. Conte, 2009 Genetics and ecological speciation. *Proc Natl Acad Sci USA* **106** Suppl 1**:** 9955-9962.

Schrider, D. R., and A. D. Kern, 2017 Soft sweeps are the dominant mode of adaptation in the human genome. *Mol. Biol. Evol.* **34:** 1863-1877.

Slatkin, M., 1985 Gene flow in natural populations. *Ann. Rev. Ecol. Evol. Sys.* **16:** 393-430.

Slatkin, M., 1987 The average number of sites separating DNA sequences drawn from a subdivided population. *Theoret. Pop. Biol.* **32:** 42-49.

Stankowski, S., M. A. Chase, A. M. Fuiten, P. L. Ralph and M. A. Streisfeld, 2018 The tempo of linked selection: Rapid emergence of a heterogeneous genomic landscape during a radiation of monkeyflowers. *BioRxiv*.

Stankowski, S., J. M. Sobel and M. A. Streisfeld, 2017 Geographic cline analysis as a tool for studying genome-wide variation: a case study of pollinator-mediated divergence in a monkeyflower. *Mol. Ecol.* **26:** 107-122.

Stankowski, S., and M. A. Streisfeld, 2015 Introgressive hybridization facilitates adaptive divergence in a recent radiation of monkeyflowers. *Proc. Royal Soc. B: Biol. Sci.* **282:** 20151666.

Stephens, M., N. J. Smith and P. Donnelly, 2001 A new statistical method for haplotype reconstruction from population data. *Am. J. of Human Genet.* **68:** 978-989.

Stuart, Y. E., T. Veen, J. N. Weber, D. Hanson, M. Ravinet *et al.*, 2017 Contrasting effects of environment and genetics generate a continuum of parallel evolution. *Nat. Ecol. Evol.* **1:** 0158.

Suchard, M. A., P. Lemey, G. Baele, D. L. Ayres, A. J. Drummond *et al.*, 2018 Bayesian phylogenetic and phylodynamic data integration using BEAST 1.10. *Virus Evol.* **4**: vey016.

R Core Team, 2016 R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria.

Terekhanova, N. V., M. D. Logacheva, A. A. Penin, T. V. Neretina, A. E. Barmintseva *et al.*, 2014 Fast evolution from precast bricks: genomics of young freshwater populations of threespine stickleback *Gasterosteus aculeatus*. *PLoS Genet.* **10**: e1004694.

Troth, A., J. R. Puzey, R. S. Kim, J. H. Willis and J. K. Kelly, 2018 Selective trade-offs maintain alleles underpinning complex trait variation in plants. *Science* **361:** 475-478.

Vijay, N., M. Weissensteiner, R. Burri, T. Kawakami, H. Ellegren *et al.*, 2017 Genomewide patterns of variation in genetic diversity are shared among populations, species and higher-order taxa. *Mol. Ecol.* **26:** 4284-4295.

Wallbank, R. W., S. W. Baxter, C. Pardo-Diaz, J. J. Hanly, S. H. Martin *et al.*, 2016 Evolutionary novelty in a butterfly wing pattern through enhancer shuffling. *PLoS Biol*. **14:** e1002353.

Whitlock, M. C., and N. H. Barton, 1997 The effective size of a subdivided population. *GENETICS* **146:** 427-441.

Wright, S., 1931 Evolution in Mendelian populations. *GENETICS* **16:** 97-159.

Wright, S., 1943 Isolation by distance. *GENETICS* **28:** 114-138.

Wu, T. D., and S. Nacu, 2010 Fast and SNP-tolerant detection of complex variants and splicing in short reads. *Bioinformatics* **26:** 873-881.

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