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

Thomas C. Nelson[1,2], Johnathan G. Crandall[1], Catherine M. Ituarte[1], Julian M Catchen[1,3], and William A. Cresko[1]*

[1]Institute of Ecology and Evolution, University of Oregon, Eugene, OR
[2]Division of Biological Sciences, University of Montana, Missoula, MT
[3]Department of Animal Biology, University of Illinois at Urbana-Champaign, Urbana, IL

*To whom correspondence should be addressed: wcresko@uoregon.edu

Abstract

The outcome of selection on genetic variation depends both on the geography of individuals and populations, and the 'geography' of loci within the genome. In the threespine stickleback, spatially variable selection between marine and freshwater habitats has had a significant and heterogeneous impact on patterns of genetic variation across the genome. 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 sampling 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. We find that divergent genomic regions co-localize with regions of reduced recombination, but that this pattern varies across different genetic maps. Linkage to loci under divergent selection had opposing effects on variation on chromosomes from marine and freshwater fish, and we attribute this effect to differences in population structure between the two habitats. These findings advance our understanding of how the recombination landscape facilitates and structures genomic divergence, and how the structuring of populations across geography can change the outcomes of selection.

Introduction

Natural populations harbor abundant genetic variation that is distributed heterogeneously across geography and across genomes. But what evolutionary genetic processes best explain these patterns? Classical population genetics theory was developed primarily by considering the effects of evolutionary processes on one or a small number of loci (HALDANE 1930; WRIGHT 1943; LEWONTIN 1974; GILLESPIE 2004). 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). While 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), in nature selective pressures tend to vary across time and space (CLAUSEN *et al.* 1941; ENDLER 1977; LENORMAND *et al.* 1999). Temporally fluctuating selection can maintain selected 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). Genetic introgression from one species to another can even fuel adaptive diversification (STANKOWSKI AND STREISFELD 2015; WALLBANK *et al.* 2016; JONES *et al.* 2018).

Despite this growing understanding of the general importance of genomic and geographic context on structuring genetic variation, precise understanding of how they interact to produce heterogeneous patterns of genetic variation is still lacking. Theoretical studies on how population subdivision impacts overall levels of variation make useful but somewhat contradictory predictions. Depending on model specifications and assumptions, population structure can either increase (CHARLESWORTH *et al.* 2003) or decrease (WHITLOCK AND BARTON 1997) levels of standing variation in a metapopulation. Therefore, empirical studies that examine the joint effects of variation in selection regimes, geographic context, and recombination landscape on genomic patterns of genetic variation in the wild can be helpful.

Here, we investigate how selection, linkage, and population structure 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, even in as little as 50 years (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) that is likely maintained by significant bidirectional gene flow.

Previously, we found that sequence variation in most genomic regions associated with local habitats is much older than variation across the rest of the genome (NELSON AND CRESKO 2018, and see Colosimo et al. 2005, Roesti et al. 2015). Here we extend our previous work by combining population genomic sampling of wild-caught marine and freshwater stickleback with genetic maps of fish derived from those same populations. We use these data to examine how linkage to loci under selection affects genetic variation across the stickleback genome, and how that effect varies among populations and ecotypes. We complement these empirical findings with forward-time simulations using SLiM (HALLER AND MESSER 2017) that recapitulate the genomic patterns of stickleback genetic variation in a parameter space that closely matches the geography and demography of stickleback in the wild.

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 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

152 read trimming, rather than strict removal, if quality decreased toward the end of the first-end read.
153 Overlapping read pairs were then merged using fastq-join (ARONESTY 2011), allowing for up to 25%
154 of bases in the overlapping region to mismatch, and the resulting contigs were trimmed to 350 bp.
155 Any read pairs that failed to merge, or were shorter than 350 bp, were removed from further analysis.
156 This step was required for processing reads through the Stacks pipeline. Below, we use the naming
157 conventions of Baird *et al.* (2008): A "RAD tag" refers to sequence generated from a single end of a
158 restriction site and the pair of RAD tags sequenced at a restriction site comprises a "RAD locus".

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

167 We used the program PHASE (STEPHENS *et al.* 2001; SCHEET AND STEPHENS 2006) to
168 combine sequence information from both RAD tags at a ***PstI*** cut site and generate phased
169 haplotypes at each RAD locus. We wrote custom Python scripts to identify all unique haplotypes at
170 each of a pair RAD tags and code them as alleles at a single, multiallelic locus. We required that
171 each individual included in this analysis was genotyped at both RAD tags. Loci containing individuals
172 only genotyped at a single RAD tag were removed from further analysis. RAD haplotypes at each
173 locus represent 696 bp of contiguous genomic sequence, giving us high-quality estimates of
174 sequence diversity and divergence even with our relatively small population-level sample sizes (NEI
175 1987, chapter 13; CRUICKSHANK AND HAHN 2014; NELSON AND CRESKO 2018). Sequencing of wild
176 stickleback resulted in 57,992 RAD loci distributed across all 21 threespine stickleback
177 chromosomes.

178 *Population genetic statistics*

179 The scripting language R version 3.5 (R Core Team 2016) was used for all downstream data
180 analysis. We estimated differentiation among threespine stickleback populations (all pairwise
181 combinations) and among ecotypes (combined freshwater ponds versus Rabbit Slough) using a
182 haplotype-based F_{ST} (equation 3 in Hudson, Slatkin et al. 1992) (HUDSON *et al.* 1992) implemented
183 in the R package 'PopGenome' (PFEIFER *et al.* 2014). We calculated π per site within and among

184 populations and absolute sequence divergence (d_{XY}) at each RAD locus by calculating pairwise
185 distances between all RAD haplotypes with the R package 'ape' (PARADIS *et al.* 2004; POPESCU *et*
186 *al.* 2012).

187 Previously, we detected patterns of reciprocal monophyly between marine and freshwater
188 haplotypes using maximum clade credibility trees generated in BEAST v1.7 (DRUMMOND AND
189 RAMBAUT 2007; DRUMMOND *et al.* 2012; NELSON AND CRESKO 2018; SUCHARD *et al.* 2018). Tree
190 topologies for all RAD loci were inferred from MCMC runs of 10^6 states with 10% burnin periods. We
191 used blanket priors and parameters across all RAD loci, including a coalescent tree prior and the
192 GTR+Γ substitution model. Monophyly of haplotypes from each population (RS, BT, BP) and each
193 habitat (marine, freshwater) was assessed using the is.monophyletic() function of the R packages
194 'ape'. Here, we use topological classifications that we inferred previously and designate gene trees
195 with reciprocally monophyletic marine and freshwater haplogroups (1,129 of 57,992 RAD loci) as
196 'divergent' loci.

197 *Laboratory crosses and genetic mapping*

198 To compare how heterogenous genomic divergence on the physical map is distributed
199 across the genetic map, we generated mapping families from laboratory lines of fish derived from
200 the Boot Lake and Rabbit Slough populations, as well as an F_1 hybrid female produced from a cross
201 between a Boot Lake female and a Rabbit Slough male. These crosses allowed us to examine
202 variation in the recombinational landscape within and among chromosomes and across multiple
203 genetic backgrounds. All maps were constructed using a pseudo-testcross design, which takes
204 advantage of existing heterozygosity in outbred individuals without the need to generate inbred lines
205 or F_1 mapping parents. To generate mapping families from Boot Lake and Rabbit Slough lines, we
206 manually crossed unrelated individuals from each line. We mapped the F_1 hybrid female by
207 backcrossing it to a Boot Lake male. All progeny were raised to 14 days post-fertilization, euthanized
208 with MS-222 (Sigma Aldrich), and fixed in 95% ethanol. We extracted genomic DNA from whole
209 progeny and from pectoral and caudal fins from all parents using proteinase K digestion (Qiagen)
210 followed by DNA purification with SPRI beads.

211 RAD genotyping of progeny and parents was used to identify segregating haplotypes for use
212 as genetic markers. We used the original single-digest RAD-seq protocol (BAIRD *et al.* 2008) with the
213 restriction enzyme SbfI, which cuts genomic DNA at approximately 22,000 sites in the stickleback
214 genome. All RAD libraries were multiplexed by family at equimolar ratios except for

215 mapping parents, which were multiplexed at twice the concentration of the progeny to increase
216 sequencing depth. The Boot Lake mapping family was sequenced in two lanes on an Illumina
217 HiSeq2000. The Rabbit Slough family was sequenced on a single HiSeq2000 lane. The F_1 family
218 was sequenced in one lane on an Illumina HiSeq2500.

219 RAD-seq data from all mapping crosses were processed with the Stacks analysis pipeline
220 (CATCHEN *et al.* 2013b). We demultiplexed and quality filtered sequences with process_shortreads
221 and aligned them to the stickleback reference genome with GSNAP (WU AND NACU 2010). We used
222 ref_map.pl to identify RAD tags and call genotypes. The Stacks component program genotypes was
223 used identify segregating markers for export to genetic mapping software. We specified a minimum
224 coverage of 3x to call individual genotypes and required that a marker be genotyped in at least 50%
225 of progeny.

226 Below, we present data for a single parent from each mapping cross (Tables I and II). By
227 conducting pseudo-testcrosses, we identified polymorphic RAD markers segregating in all six
228 mapping parents. However, to investigate the genome-wide recombination landscape, as well as
229 relationships between recombination rate and natural levels of polymorphism and divergence, we
230 restricted our analysis to mapping parents for which we observed segregating markers on all 21
231 chromosomes with no gaps of more than 1 megabase (Mb) between adjacent markers. One parent
232 from each cross fit these criteria. Thus, we present data for female parents from the Boot Lake and
233 F_1 crosses and the male parent from the Rabbit Slough cross. In our results, we focus on the effects
234 of genotype on relative rates of recombination within and among mapped individuals, rather than
235 absolute rates, because population-of-origin was confounded with sex.

236 We estimated genome-wide recombination rates between RAD markers under the
237 assumption of collinearity between all genetic maps and the stickleback reference genome (GLAZER
238 *et al.* 2015) (with some exceptions, see below). We used the mapping software Lep-MAP2 (RASTAS

Table I. Sequencing statistics of libraries for each mapping family. Coverage is the mean coverage per RAD locus across all progeny.

Family	raw reads	filtered reads	aligned reads	mean coverage
Boot Lake	322871980	229263868	196451426	27.77
Rabbit Slough	188240315	110264451	95689374	25.21
F_1 hybrid	229219526	181462508	78349244	18.75
Total	740331821	520990827	370490044	24.72

239 *et al.* 2015) to estimate map positions of RAD markers with the marker order fixed to the aligned
240 positions on the reference genome. Known marker orders increased throughput of mapping
241 iterations and reduced the impact of genotyping errors on recombination rate estimation. While fixed
242 marker orders do not explicitly allow the detection of structural variation among genomes — such
243 as chromosomal inversions that are known to exist among stickleback populations — discrepancies
244 in the estimated map do provide indirect, and correctable, evidence of changes to map order (Figure
245 S1). For example, inversions in genetic map order relative to the reference order spuriously inflate
246 recombination rates between markers closely flanking inversion breakpoints when inverted segments
247 are forced into the wrong orientation. This is because genetic map distances are estimated
248 independently while the physical distance between markers is drastically underestimated. The
249 observed jumps in map distance on either side of the inversion are equal to each other and to the
250 total map length of the inversion. Reversing the marker order within the inversion removes these
251 artifacts and reduces the overall map length of the region.

252 *Recombination-polymorphism correlations*

253 We employed two methods to investigate the relationships between the recombinational
254 landscape and patterns of polymorphism within and among natural populations. To quantify
255 genome-wide correlations among recombination rate and population genetic statistics, we divided
256 the stickleback genome into non-overlapping windows and calculated average recombination rates
257 (in centiMorgans per megabase, cM/Mb), sequence diversity (π_{BT}, π_{RS}, and π), and genetic
258 divergence (F_{ST}, d_{XY}) in each window. Below, we show results from 1000-kb genomic windows. RAD
259 marker density made estimates of local recombination rates less reliable and more variable at smaller
260 window sizes (Figures S3 and S4). We used nonparametric correlations to test for correlations
261 between variables because the distributions of most variables lacked normality even using standard
262 data transformations. Below we present Spearman's rank order correlations. Kendall's tau and
263 parametric linear models gave qualitatively similar results.

264 Genomic heterogeneity exists not only in the proportion of the genome affected by divergent
265 selection but also in how genetic variation and divergence are clustered within the genome. Marine-
266 freshwater genomic divergence in stickleback is clustered into few, large regions that can
267 encompass most of the length of a chromosome. We sought to directly compare the genomic
268 distributions of population genetic statistics along the physical genome and on the three genetic
269 maps we constructed. We used a windowing approach that allowed direct comparisons across
270 maps despite differing numbers and distributions of markers among genetic maps. Using the R

271 package 'ksmooth', we binned each chromosome into equally sized intervals, the number of which
272 we set equal to the number of segregating RAD markers on the genetic map with fewest markers.
273 For each interval, we calculated genetic position from each laboratory cross and F_{ST} between Rabbit
274 Slough and Boot Lake populations (HOHENLOHE *et al.* 2010; NELSON AND CRESKO 2018) within a 250-
275 kb normally distributed kernel. Using this windowing approach, we were able to make direct
276 comparisons between recombination rates within and among genetic maps from different genetic
277 backgrounds and patterns of polymorphism in the populations from which the laboratory lines were
278 derived.

279 *Forward simulations*

280 We modeled the effects of selection, linkage, and population structure using forward
281 simulations implemented in SLiM (HALLER AND MESSER 2017). We simulated a metapopulation of
282 2000 diploid individuals and a genome consisting of a single chromosome with a genetic length of
283 one Morgan (physical length 10^5 bp; recombination rate $1x10^{-5}$ per bp). The metapopulation was
284 divided into two habitat types analogous to marine and freshwater stickleback habitats, which we
285 refer to as H_1 and H_2, respectively (Figure 1). Across all simulations, habitat H_1 remained a single
286 panmictic population of size $N_H = 1000$. Habitat H_2, in contrast, consisted of k subpopulations, each
287 of size N_H/k. We varied k between 1 (no substructure) and 5. Importantly, subpopulations of H_2
288 exchanged migrants only with H_1. We chose this constraint on our simulations because freshwater
289 stickleback populations are thought to most often be derived from marine ancestors, and because

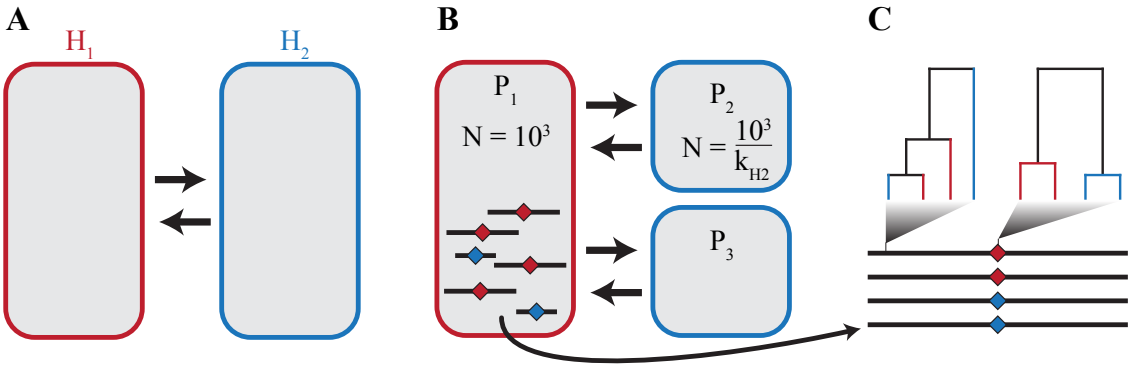


Figure 1. The structure of simulations to test the effects of population structure on genetic variation in a model of local adaptation. **A:** each habitat (H_1: red, H_2: blue) consists of a single deme ($k = 1$). **B:** Habitat H_2 was also simulated consisting of $k_{H2} = 2$ and $k_{H2} = 5$ demes (not shown). Chromosomes of both allelic state were sampled from H_1 to simulate founding variation available to a new population. N = population size of each habitat. **C:** Expected genealogies at the divergently adaptive locus and at a distal locus on the chromosome. We used the pattern of reciprocal monophyly seen in C, top right, to classify RAD loci as 'divergent' in the threespine stickleback population genomic dataset. (See Methods and NELSON AND CRESKO [2018]).

290 gene flow among freshwater populations primarily occurs through the marine population, especially
291 across broader geographic scales. Migration rates were constant and symmetrical between all pairs
292 of populations exchanging migrants.

293 At the beginning of each simulation, we introduced a locally adaptive allele to a locus at the
294 center of the chromosome (position 50 kb, 50 cM). Fitness effects of the alternative alleles at the
295 locus were additive, and the strength of selection was symmetrical among habitat types. To ensure
296 establishment of both alleles at the locus across all simulations, we introduced the H_2-adaptive allele
297 at a frequency of 0.25 into all populations of habitat H_2. Because we introduced this allele at the
298 beginning of the simulation, all chromosomes carrying that allele were identical at the time of
299 introduction.

300 Genetic diversity among chromosomes carrying H_1- and H_2-adaptive alleles was assessed
301 after 10,000 generations. We designed our sampling after stickleback evolutionary history, where
302 genetic variation available to new freshwater populations is sampled from the marine population, by
303 sampling chromosomes exclusively from habitat H_1. We then calculated nucleotide diversity (π) within
304 and among chromosomes of the two allelic states in non-overlapping 1000 bp (1 cM) bins. Below,
305 we fold the calculations of nucleotide diversity about the locally adaptive locus to emphasize the
306 effects of evolutionary forces on π at a given distance from the locus. These summaries mirror our
307 estimates of genetic variation on stickleback chromosomes and theoretical work (CHARLESWORTH *et*
308 *al.* 1997).

309

310 **Results**

311 *High density mapping crosses enable comparison of recombination landscapes*

312 RAD sequencing of progeny used to create the three genetic maps generated 740,331,821
313 raw single-end sequence reads distributed among 385 individuals. Of these, 520,990,827 (70.4%)
314 passed quality filtering, and 370,490,044 (71.1%) of those were unambiguously aligned to the
315 reference genome. Sequencing results of the families analyzed for each of the three genetics maps
316 are summarized in Table I.

317 All mapping crosses resulted in over 6000 segregating RAD markers distributed across all
318 21 stickleback chromosomes (Table II, Figure 2, Suppl. Figure S2). The Boot Lake female, mapped
319 using the largest family of 189 progeny, had the longest overall map length of 1694 cM among 6041
320 segregating markers and a genome-wide average recombination rate of 3.93 cM/Mb. The

321 Rabbit Slough male was mapped using the fewest number of offspring (n = 94) but the largest
322 number of segregating markers (7049). It had the shortest estimated map length of 1047 cM,
323 resulting in an average recombination rate of 2.43 cM/Mb. The F_1 hybrid genetic map was
324 intermediate in length, and also had an intermediate number of progeny (102) and markers (6669)
325 used in mapping. The total length of the hybrid map was 1406 cM with a mean recombination rate
326 of 3.26 cM/Mb.

327 Patterns of recombination were generally consistent across the three genetic maps (Figure
328 2, Suppl. Figure S2). As has been described previously (ROESTI *et al.* 2013; GLAZER *et al.* 2015),
329 recombination across the larger metacentric chromosomes in the stickleback genome was biased
330 toward the ends of the chromosomes, with very little recombination occurring across much of their
331 centers (e.g chromosomes 4 and 7). Recombination rates across a number of the smaller
332 chromosomes, in contrast, was typically highest on one chromosome end but this likely also reflects
333 a trend against recombination near the centromere.

334 *The recombination landscape varies among individuals*

335 We observed important differences among genetic maps. As expected, recombination in the
336 hybrid map was completely suppressed within and immediately surrounding an inversion on
337 chromosome 21 (Figure 2). In contrast, in the genetic maps made from BT and RS fish separately
338 recombination occurred steadily across this region, as expected because each population was
339 homozygous for opposite orientations of the inversion. The resolution of our RAD-based genetic
340 maps was not sufficiently high to be able to observe any recombination within the smaller marine-
341 freshwater-diagnostic inversions on chromosomes 1 and 11 (400 kb and 500 kb in length,
342 respectively). We also found the expected suppression of recombination across the X chromosome
343 (chromosome 19) in the male RS map, with the exception of the pseudoautosomal region that
344 comprises the first ~2.5 Mb of chromosome 19 (Suppl. Figure S2).

Table II. Genetic map statistics for the three maps used in this study. *map length given in centiMorgans (cM). †Genomic coverage is a percentage of the reference genome assembled onto chromosomes (436.6 Mb).

Family	progeny	markers	map length*	genomic coverage†	cM/Mb
Boot Lake	189	6041	1694	0.986	3.93
Rabbit Slough	94	7049	1047	0.987	2.43
F_1 hybrid	102	6669	1406	0.988	3.26

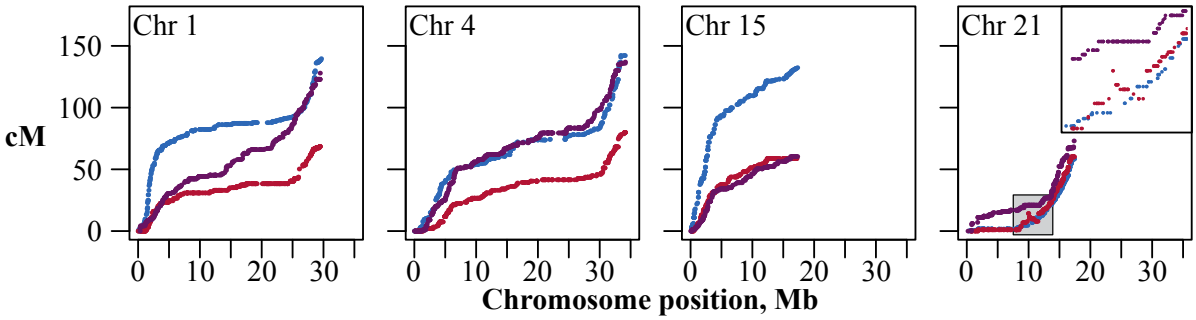


Figure 2. The recombination landscape varies within and among genetic maps of marine (RS), freshwater (BT), and F_1 hybrid threespine stickleback. Each point represents a RADseq-based marker segregating within a mapping family. Red: RS, blue: BT, purple: F_1 hybrid. Genetic maps for all chromosomes are given in supplemental figure S2.

345 Patterns of recombination also varied in unexpected ways, most notably in the hybrid map.
346 As noted above, the hybrid map was intermediate in length compared to the BT and RS maps, but
347 this pattern rarely held at a finer genomic scale. Across chromosome 4, both the pattern and the
348 total amount of recombination on the hybrid map was nearly identical to the BT map. The pattern of
349 the hybrid and BT maps being similar to one another, and different from RS, was also clearly the
350 case for chromosomes 12 and 14. The opposite was also apparent on different chromosomes, with,
351 for example, recombination in the hybrid on chromosome 15 closely matching the RS map, and both
352 being different from BT.

353 On some chromosomes, the distribution of crossovers on the hybrid map was qualitatively
354 different from either the BT or RS map. On chromosome 1 in both the BT and RS maps,
355 recombination was biased toward the ends of the chromosomes as it is along the other large
356 chromosomes. However, recombination rates remained fairly constant across the entire
357 chromosome in the hybrid map, with the result that this linkage group in the hybrid map was nearly
358 equal in length to the BT linkage group. This is despite very high local recombination rates toward
359 both ends of the BT map. Similar patterns, though more subtle, were also evident on chromosomes
360 9 and 21 (Suppl. Figure S2).

361 *Genomic regions of highest divergence represent a small fraction of the genetic map*

362 Genomic regions of highest divergence were compressed into proportionally smaller regions
363 of the genetic maps (Figure 3). For instance, some of the largest regions of divergence we observed
364 — including those surrounding ***eda***, the major effect locus for lateral plate number (COLOSIMO *et al.*
365 2005), and the chromosome 21 inversion — had average F_{ST} values in excess of 0.4 (Figure 4).
366 Regions above this threshold spanned 8.1% of the physical genome (35.3 Mb) but only 3.3% of the

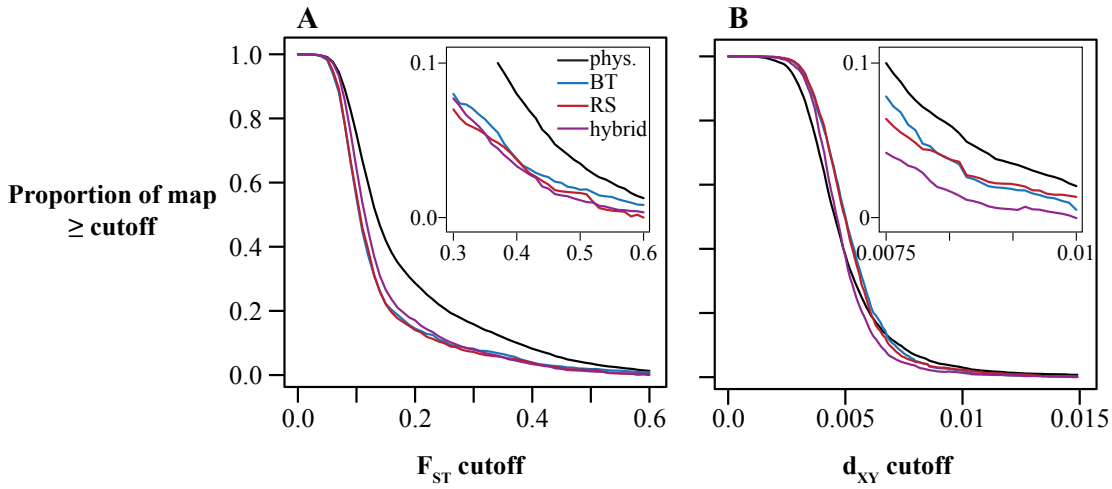


Figure 3. Genetic divergence comprises a smaller proportion of the genetic map than the physical map. Each line represents the proportion of the overall map length (Mb or cM) taken up by windows of genetic divergence (F_{ST} or d_{XY}) greater than or equal to a given value. Black: physical map (reference genome), red: RS genetic map, blue: BT genetic map, purple: F_1 hybrid genetic map.

367 F_1 hybrid genetic map (10 cM). In contrast to F_{ST}, which is a measure of relative divergence, the
368 proportion of the genome in regions of high absolute divergence (as measured by d_{XY}) was less
369 affected by the map on which it was measured. While the genetic maps did have the smallest
370 proportion of their total length in regions of high d_{XY} (Fig 3B, inset), this effect was subtle compared
371 to that seen for F_{ST}.

372 The compression of divergence on the genetic maps did not occur equally on all three maps,
373 as the genome-wide pattern may suggest. Chromosomal inversions specifically suppress
374 recombination in hybrids and, on chromosome 21, genomic divergence was only compressed on
375 the hybrid map (Figure 4). The region of highest divergence on chromosome 21 recombines steadily

Table III. Genome-wide correlations between recombination rate and genetic diversity and divergence. Statistics were calculated as means of 1000-kb non-overlapping bins. Correlation coefficients are given as Spearman's ρ. *p < 0.05, **p < 0.01, ***p < 0.001.

genetic map	π_{RS}	π_{FW}	π	d_{XY}	F_{ST}	D_a
BT	0.450***	0.349***	0.261***	0.154*	-0.431***	-0.178**
RS	0.489***	0.223***	0.225***	0.146*	-0.365***	-0.128*
F_1	0.483***	0.243***	0.146*	0.053	-0.406***	-0.211**

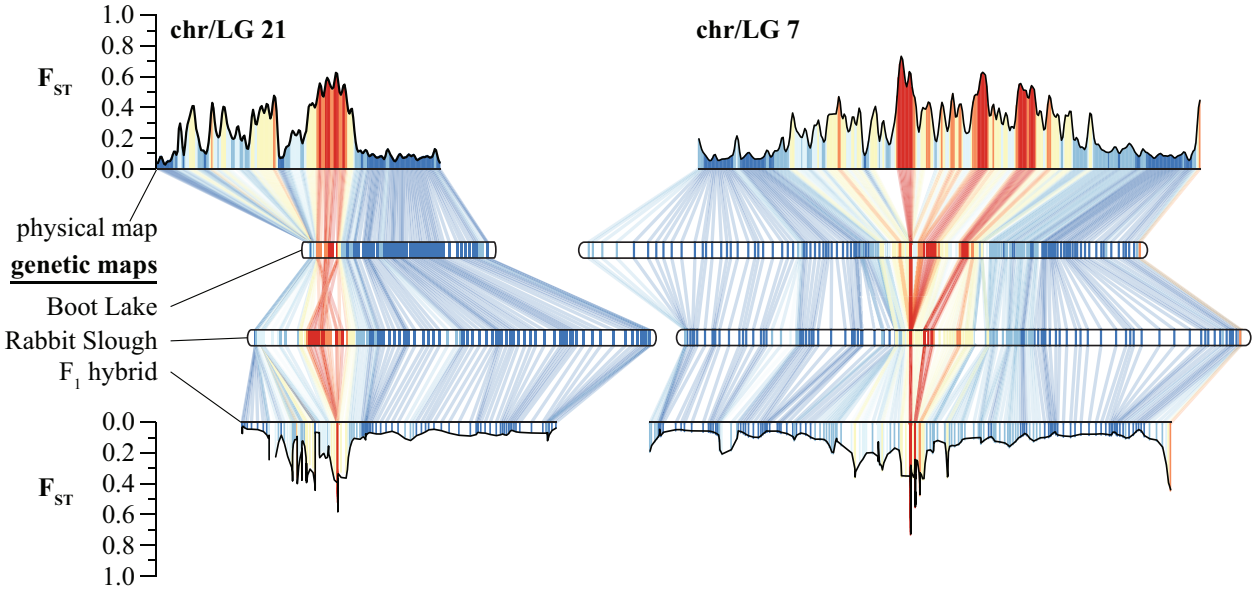


Figure 4. The recombination landscape simplifies the genetic architecture of divergence. F_{ST} scans were performed on the physical map (top) and the three genetic maps (middle and bottom). Lines connect evenly-spaced windows on the physical map to the imputed positions on each genetic map and are colored by the kernel-averaged F_{ST} for that window.

in both the BT and RS maps but, as expected, it collapses to a single locus on the hybrid genetic map. A similar effect was seen across a large collinear region of divergence on chromosome 7. On the physical map, chromosome 7 contains three peaks of strong divergence between BT and RS spanning over 9 Mb. All three peaks are clearly separated on the BT map, though they comprise a relatively smaller region than on the physical map. On the RS map, we observed no recombination between two of these peaks and the combined span of the region was further reduced. The greatest compression of this divergence occurred on the hybrid map; despite an overall map length that was intermediate to the other two maps, an 8.5-Mb region across the center of the chromosome — containing nearly all genomic divergence between BT and RS — collapsed to a single locus (span: 0 cM, position 57.7 cM on the map).

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

We 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 1), 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

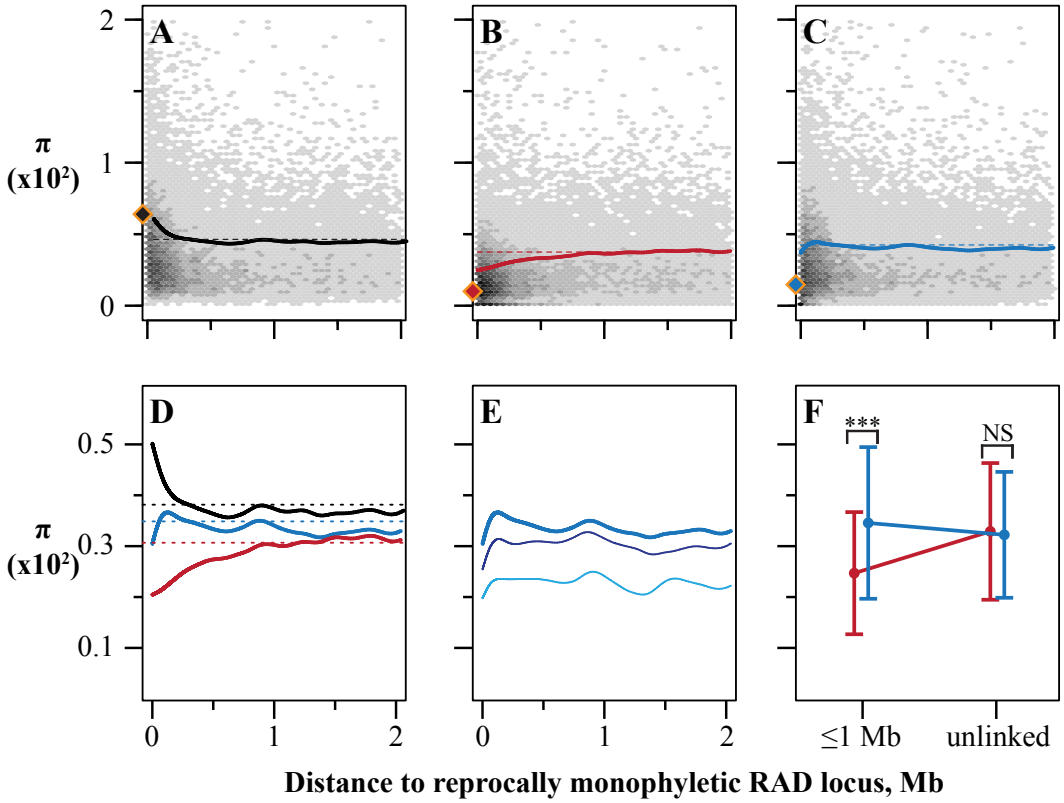


Figure 5. Patterns of linked variation differ on marine and freshwater stickleback chromosomes. **A-C:** Genome-wide sequence diversity (π) as a function of the distance from a RAD locus with reciprocally monophyly of marine and freshwater haplotypes. Diamonds show average sequence diversity among reciprocally monophyletic RAD loci. Density of RAD loci at x distance from the nearest reciprocally monophyletic locus with y level of diversity is shown in shades of gray. Bold lines are smoothed splines, dashed lines are genome-wide means. **A:** total sequence diversity. **B:** sequence diversity on chromosome carrying marine haplotypes. **C:** sequence diversity on chromosomes carrying freshwater haplotypes. Freshwater sequence diversity includes chromosomes sampled from BT as well as Bear Paw Lake, Alaska (BP). **D:** Smoothed splines as in A-C. **E:** Splines calculated from each freshwater population separately (BT: dark blue, BP: light blue). **F:** Sequence diversity of RAD loci within 1 Mb of a reciprocally monophyletic RAD locus and on chromosomes not carrying a reciprocally monophyletic locus ('unlinked'). Data are means ± 1 SD. *** $p < 0.001$.

averaged 0.0067 per site (Figure 5A, black-filled diamond, genome-wide mean = 0.0042). In contrast, and as expected at locally adaptive loci, π was substantially reduced within allelic states; π among marine alleles averaged 0.0012 per site (B, red-filled diamond) while average π among freshwater alleles at divergent loci was 0.0015 per site.

Patterns of linked variation extending away from divergent loci were striking and varied qualitatively between chromosomes sampled from marine and freshwater stickleback. Sequence diversity at linked loci was highly correlated with proximity to a divergent locus across all population partitions (Figure 5A-C, Spearman's rho: all p-values ≤ 10^{-10}). When all populations were combined,

403 π at linked loci increased sharply within approximately 250 kb from a divergent locus (Figure 5A).
404 Linked loci in close proximity to divergent loci were nearly as polymorphic as those directly impacted
405 by divergent selection.

406 In contrast, and as expected under local adaptation, π among chromosomes sampled from
407 the marine Rabbit Slough stickleback was lowest in close proximity to divergent loci, and a
408 substantial fraction (12%) of RAD loci with 250 kb of a divergent locus showed no variation at all on
409 these chromosomes (π_{RS} = 0; 1651/13762 loci). Freshwater populations, however, showed a more
410 complex proximity effect that was distinct from either Rabbit Slough or the combined populations
411 (Figure 5C-E). In data collected from both Boot Lake and Bear Paw Lake populations, as well as
412 when data from these populations combined, π actually increased in proximity to a divergent locus
413 and peaked at loci approximately 200 kb away from divergent loci before reversing direction. This
414 pattern qualitatively persisted within both freshwater populations individually as well as when Boot
415 Lake and Bear Paw Lake chromosomes were combined. Among chromosomes from freshwater
416 fish, therefore, we observed the most diversity genome-wide in those regions in close proximity, but
417 not directly adjacent, to a divergent locus.

418 We therefore observed a proximity-by-population interaction in levels of sequence variation
419 at linked versus unlinked loci in marine or freshwater populations (Figure 5F). While estimates of
420 genetic diversity were similar between the Rabbit Slough population and the combined freshwater
421 populations on chromosomes without evidence of divergent natural selection, sequence diversity
422 was higher among freshwater populations than in Rabbit Slough within 1 Mb of a divergent locus
423 (Figure 5F; ANOVA: population*proximity interaction $p \leq 10^{-10}$). On chromosomes with no evidence
424 of divergent natural selection, we found no evidence for differences in levels of variation between
425 Rabbit Slough and the combined freshwater populations (mean π_{RS} = 0.0032, mean π_{FW} = 0.0032;
426 Tukey's HSD, p = 0.53). However, among linked loci within 1 Mb of a divergent locus, π in the
427 combined freshwater populations was greater than in Rabbit Slough (Tukey's HSD, $p < 10^{-5}$), and at
428 least as high as π in unlinked regions (Tukey's HSD [ANOVA on untransformed π], $p < 10^{-5}$; Tukey's
429 HSD [ANOVA on log-transformed π], $p < 10^{-5}$).

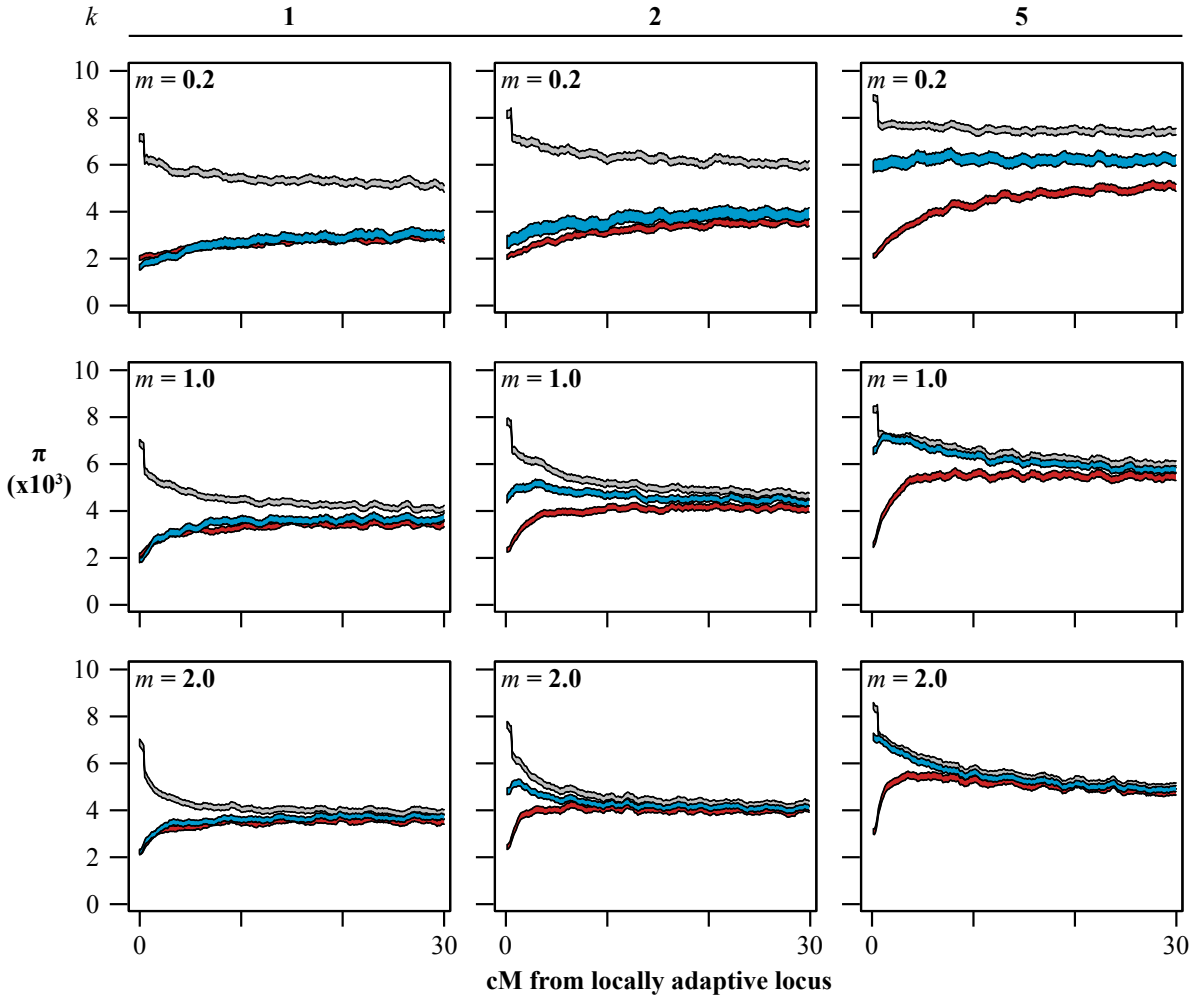


Figure 6. Asymmetric population structure generates asymmetric patterns of linked variation on simulated chromosomes. Simulations were performed as described in Methods and Figure 1. Bands are 95% confidence intervals of nucleotide diversity (π) at a given distance from a locally adaptive locus (gray: π_{total}, red: π_1, blue: π_2). Columns show the effect of increasing population structure (k) in habitat H_2. Rows show the effect of increasing migration rates (m, in number of migrants per generation) between habitat H_1 and populations of habitat H_2.

Population structure maintains linked variation on simulated chromosomes

To explore the possibility that population structure within the freshwater habitat influences patterns of linked variation, we conducted forward simulations of chromosomes under a model of migration-selection balance using SLiM (Figures 1 and 6). We first simulated two habitat types, each composed of a single, panmictic population (Figure 6, column 1), analogous to the model presented by Charlesworth, Nordberg, and Charlesworth (1997). As expected, overall nucleotide diversity was highest within 10 cM of the locally adaptive locus, while diversity within allelic classes dropped precipitously and symmetrically. When migration rates were on the order of a single migrant per

438 generation or less (Figure 6, rows 1 and 2), genetic variation was significantly partitioned among
439 habitats, even distal to the locally adaptive locus (95% confidence intervals: π_{total} vs $\pi_{H1,H2}$, $p \leq 0.05$).
440 When habitat 2 consisted of two or five demes (Figure1; Figure 6, columns 2 and 3) the
441 additional structure altered patterns of linked variation in two ways. First, nucleotide diversity was
442 higher overall, both within and among allelic classes, compared to simulations without substructure
443 (two-way ANOVA: substructure main effect $F_{2,4365} = 496.1$, $p \leq 1\text{x}10^{-10}$). Notably, this effect occurred
444 without any increase in the census size of the metapopulation. Second, while nucleotide diversity
445 decreased adjacent to the locally adaptive locus among chromosomes carrying H_1-adaptive alleles,
446 diversity on chromosomes carrying H_2-adaptive alleles was highest adjacent to the locally adaptive
447 locus when migration rates were at least $m = 1$ (Figure 6, columns 2 and 3). When habitat H_2
448 consisted of two demes, diversity on chromosomes carrying H_2-adaptive alleles was intermediate to
449 the low levels of diversity on H_1 chromosomes and the higher levels of diversity observed among all
450 chromosomes. This pattern was consistent with our observations of diversity among threespine
451 stickleback chromosomes. Further subdivision of habitat 2 (Figure 6, column 3) resulted in
452 chromosomes carrying H_2-adaptive alleles also carrying nearly all the linked variation present in the
453 metapopulation. Except in the window that contained the adaptive locus (0 cM), we never saw
454 evidence that π on H_1 chromosomes was reduced compared to π across the entire system
455 (overlapping 95% confidence intervals, π_{total} vs π_{H2}).
456

457 **Discussion**

458 *A variable recombination landscape simplifies the architecture of divergence*

459 The recombination landscape has profound effects on genome-wide patterns of genetic
460 variation. Linkage to or among selected sites is now thought to affect many, if not most, variable
461 sites in the genome (SCHRIDER AND KERN 2017; KERN AND HAHN 2018), and suppression of
462 recombination among the genomes of diverging populations and species appears in large part to
463 determine patterns of genetic differentiation and divergence (BURRI *et al.* 2015; VIJAY *et al.* 2017;
464 STANKOWSKI *et al.* 2018). In the threespine stickleback, two related patterns point to the importance
465 of the recombination landscape in structuring the genetics of divergence. First, genetic differentiation
466 among phenotypically divergent populations is known to accumulate in regions of low recombination
467 (ROESTI *et al.* 2013). Second, quantitative trait locus (QTL) mapping studies consistently find highly
468 pleiotropic QTL associated with clusters of markers in the centers of relatively few linkage groups
469 (ALBERT *et al.* 2008; MILLER *et al.* 2014; MARQUES *et al.* 2016). This finding could be explained by the

470 presence of a one or a small number of pleiotropic loci of large effect in each region, or clusters of
471 linked loci.

472 By combining population genomics with multiple mapping crosses we are able to
473 corroborate and further our understanding in stickleback fish and other organisms in two distinct
474 ways. First, genomic divergence that occurs in the centers of the largest chromosomes occurs within
475 regions of consistently low recombination across all three genetic maps (freshwater, marine, and F_1
476 hybrid fish). The negative association between divergence and recombination rate is a common
477 finding across systems (BURRI *et al.* 2015; MARTIN *et al.* 2016) and our results further demonstrate
478 that stickleback are no exception to this rule both at the local (i.e. specific genomic regions, Figure
479 4) and genome-wide scales (Figure 3). In addition, by directly comparing the genetic maps of purely
480 marine and freshwater fish, our results do not indicate the presence of any large or complex structural
481 variation between marine and freshwater stickleback genomes associated with the largest regions
482 of genomic divergence.

483 Second, although there was general congruence among the genetic maps, we find that the
484 recombination landscape varied unexpectedly and in evolutionarily relevant ways. Within the large
485 regions of divergence at the centers of chromosomes, low recombination ranged from being
486 universal across the three cross types to genotype-specific (Figure 2, chromosomes 4 and 1,
487 respectively), suggesting different genetic bases to the low recombination rates across these large
488 genomic regions. Interestingly, recombination rates in regions of divergence were often lowest in the
489 F_1 hybrid map (Figure 4). While this was expected within chromosomal inversions, where the
490 alternative homozygotes demonstrated steady recombination throughout the inverted region, this
491 was also the case on chromosome 7 despite the apparent lack of any large scale, simple structural
492 variation (Figures 2 and 4) although clusters of smaller structural variants that could not be detected
493 in our maps could also reduce recombination. On the hybrid map, this entire region collapsed to a
494 region inherited essentially as a single Mendelian locus.

495 The specific reductions in recombination on the hybrid map are suggestive of adaptive
496 evolution of the recombination landscape itself and, if true, could have profound implications for
497 adaptation and genomic divergence in stickleback. For example, inversions are advantageous when
498 recombination between multiple linked alleles that contribute to fitness (either in an additive or
499 epistatic fashion) is maladaptive (KIRKPATRICK AND BARTON 2006). The three previously identified
500 inversions on chromosomes 1, 11 and 21 (JONES *et al.* 2012) are associated with divergence
501 between the freshwater and marine habitats and occur in regions of the genome that readily

502 recombine in the chromosomal homozygotes (BT and RS maps). Had inversions not evolved, gene
503 flow and recombination among marine and freshwater populations may have been strong enough
504 to limit adaptive divergence (LENORMAND 2002; YEAMAN AND WHITLOCK 2011; AESCHBACHER *et al.*
505 2017).

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

514 Our observations of coincidence of reduced recombination and genomic divergence may
515 help explain the well documented pattern repeatable and rapid adaptive divergence in stickleback
516 (BELL AND FOSTER 1994; LESCAK *et al.* 2015) that is largely the result of the reuse of standing genetic
517 variation (COLOSIMO *et al.* 2005; SCHLUTER AND CONTE 2009; TEREKHANOVA *et al.* 2014; ROESTI *et al.*
518 2015; BASSHAM *et al.* 2018). Because freshwater populations are typically founded by marine
519 stickleback, the sources of standing genetic variation are low frequency variants in the marine
520 population that nearly always exist in a heterozygous state with a marine genome. Our results
521 suggest that the recombination landscape may therefore facilitate the maintenance of freshwater
522 haplotypes during their transit through the marine environment by reducing recombination even in
523 collinear genomic regions. Multi-megabase haplotypes — potentially containing many alleles
524 contributing to local adaptation — then have a higher probability of being selected in concert when
525 a new freshwater population is established.

526 *Asymmetric population structure maintains asymmetries in patterns of linked genetic*
527 *variation*

528 The patterns of linked variation that we document here highlight not only the importance of
529 selection but also of the precise structure of the threespine stickleback metapopulation in maintaining
530 genetic variation. Throughout the species range, the marine population is remarkably uniform
531 phenotypically, which we now know is reflected genetically by minimal isolation-by-distance over
532 thousands of kilometers (BELL AND FOSTER 1994; CATCHEN *et al.* 2013a). This large population with
533 few barriers to migration is contiguous with thousands of freshwater lake and stream populations

534 that are more clearly isolated by geography. While the influence of freshwater populations on the
535 evolution of the species was once unclear (BELL AND FOSTER 1994), it is now evident that gene flow
536 between freshwater and marine stickleback populations is common and may facilitate adaptation
537 through the indirect sharing of alleles among freshwater populations (SCHLUTER AND CONTE 2009).

538 Our findings suggest that the more highly structured freshwater habitat is actually the primary
539 reservoir of genetic variation in the species, and this diversity is particularly concentrated in regions
540 of the genome experiencing divergent selection. In these regions of the genomes, chromosomes
541 carrying locally adapted marine and freshwater haplotypes had strikingly different patterns of
542 ancestry. We suggest that this is the consequence of asymmetric population structure among
543 stickleback populations in marine and freshwater habitats (Figure 6). Sequence diversity on marine
544 chromosomes increased steadily away from loci under divergent selection (Figure 5B). In stark
545 contrast, these same regions were the most diverse genomic regions we observed on freshwater
546 chromosomes (Figure 5C-F).

547 We hypothesize that these genomic patterns are the result of greater population structure
548 among freshwater populations in combination with the 'protection' against recombination with
549 marine chromosomes when in marine habitats. Theory shows that population structure itself can
550 increase inter-population coalescence times relative to panmixia (CHARLESWORTH *et al.* 2003). We
551 suggest that selection has naturally partitioned the genome into less- and more-structured regions
552 based on proximity to loci under divergent selection. The increased diversity on freshwater
553 chromosomes adjacent to divergently adapted loci is the result of the geographic structuring of
554 variation among semi-isolated freshwater populations (Figure 6). This signature is detected only in
555 those genomic regions also affected by divergent selection because selection itself has reduced the
556 effective migration and recombination rates between marine and freshwater populations in those
557 genomic regions. Because migration among freshwater populations occurs primarily through the
558 marine environment, divergent selection accentuates the effect of population structure in the
559 freshwater habitat.

560 The forward simulations we performed in SLiM support this hypothesis. We found that under
561 realistic migration rates, such asymmetric population subdivision ***by itself*** provides a simple
562 explanation for the counterintuitive observation that smaller, isolated populations contain abundant
563 linked variation. Additional asymmetric population structure increased levels of variation in the system
564 as a whole. Simulations including within-habitat population structure resulted in higher nucleotide
565 diversity across the entire chromosome both within and among allelic classes, and a strong peak in

566 diversity near the locally adaptive locus. These findings generally agree with expectations from the
567 structured coalescent (SLATKIN 1987; HUDSON 1990; CHARLESWORTH *et al.* 2003), where limited
568 migration increases the expected coalescence times of chromosomes sampled from different demes
569 because the chromosomes must be found in the same deme in order to coalesce. Nearby an
570 alternatively adaptive locus, selection accentuates this effect by reducing effective migration rates
571 among demes of alternative habitats.

572 Our empirical and simulated results, in combination with previous findings in threespine
573 stickleback (REIMCHEN 1994; KAEUFFER *et al.* 2012; ROESTI *et al.* 2014), have important implications
574 for the long-term evolution of the species. Freshwater populations are ecologically diverse (BELL AND
575 FOSTER 1994; KAEUFFER *et al.* 2012; LEAVER AND REIMCHEN 2012) and the relative impacts of parallel
576 versus non-parallel selection pressures among freshwater populations is actively debated (KAEUFFER
577 *et al.* 2012; STUART *et al.* 2017). While it is possible that multifarious selection pressures also
578 contribute to the maintenance of more genetic diversity on chromosomes from freshwater fish, our
579 model does not require additional selection to maintain linked variation. Furthermore, the freshwater
580 pond populations studied here are ecologically similar and geographically proximal (CRESKO *et al.*
581 2004), so it is likely that parallel selection on a common pool of variation has produced the patterns
582 of variation we observe. Also studying ecologically similar freshwater populations, Roesti et al (2014)
583 found that parallel adaptation led to increases in F_{ST} among freshwater populations near peaks of
584 marine-freshwater divergence, and inferred that sweeps of common SNPs on different genetic
585 backgrounds led to this pattern. We have shown these linked genomic regions to be not only the
586 most heavily partitioned but also the most diverse in the genome. We find it likely that selection and
587 gene flow happening over shorter evolutionary timescales (tens to thousands of years), like those
588 investigated by Roesti et al. (2014), have occurred throughout the evolutionary history of this species
589 and have long-lasting, collective effects on patterns of genomic variation over the course of millions
590 of years.
591
592
593
594
595

596 *Conclusions*

597 Divergent natural selection is a powerful force for the maintenance of genetic variation in nature.
598 When local adaptation plays out on variable recombinational and geographic landscapes, we find
599 that its effects on genetic variation are amplified and shape patterns of linked variation in unexpected
600 ways. We hypothesize that the stickleback recombinational landscape is the product of adaptive
601 evolution that transforms a large genomic architecture of marine-freshwater divergence into a much
602 more simplified genetic architecture. If true, two consequences are the efficient maintenance of
603 adaptive divergence in the face of gene flow, and widespread linked selection that eliminates genetic
604 variation in the panmictic marine population while maintaining genetic diversity in the freshwater
605 habitat. The interaction of selection, recombination, and population structure has turned small,
606 isolated freshwater stickleback populations into the primary reservoirs of standing genetic variation
607 that potentiate future evolutionary change. More broadly, our results underscore how integrated
608 studies of evolutionary and genetic processes can yield exciting, unexpected patterns and deeper
609 understanding when we jointly consider the processes and their interactions with one another.

610

611 **Data availability**

612 Raw sequence data will be made available on NCBI under BioProjects PRJNA429207 and
613 PRJNAXXXXXX. Scripts and processed data will be available on FigShare and are immediately
614 available on GitHub at https://github.com/thomnelson/linkedvariation.

615

616 Literature Cited

617 Aeschbacher, S., J. P. Selby, J. H. Willis and G. Coop, 2017 Population-genomic inference of the
618 strength and timing of selection against gene flow. *Proc. Natl. Acad. Sci. USA* **114:** 7061-
619 7066.
620 Albert, A. Y. K., S. Sawaya, T. H. Vines, A. K. Knecht, C. T. Miller *et al.*, 2008 The genetics of
621 adaptive shape shift in stickleback: Pleiotropy and effect size. *Evolution* **62:** 76-85.
622 Angert, A. L., M. Bayly, S. N. Sheth and J. R. Paul, 2018 Testing Range-Limit Hypotheses Using
623 Range-Wide Habitat Suitability and Occupancy for the Scarlet Monkeyflower (Erythranthe
624 cardinalis). *Am. Nat.* **191:** E76-E89.
625 Aronesty, E., 2011 ea-utils: Command-line tools for processing biological sequencing data.
626 Baird, N. A., P. D. Etter, T. S. Atwood, M. C. Currey, A. L. Shiver *et al.*, 2008 Rapid SNP discovery
627 and genetic mapping using sequenced RAD markers. *PLoS One* **3:** e3376.
628 Bassham, S., J. Catchen, E. Lescak, F. A. von Hippel and W. A. Cresko, 2018 Repeated selection
629 of alternatively adapted haplotypes creates sweeping genomic remodeling in stickleback.
630 *GENETICS* **209**: 921-939.
631 Begun, D. J., and C. F. Aquadro, 1992 Levels of naturally occurring DNA polymorphism correlate
632 with recombination rates in D. melanogaster. *Nature* **356:** 519-520.

633 Bell, M. A., and S. A. Foster, 1994 Introduction to the evolutionary biology of the threespine
634 stickleback, pp. 1-27 in *The Evolutionary Biology of the Threespine Stickleback*, edited by
635 M. A. Bell and S. A. Foster. Oxford University Press, New York.
636 Bergland, A. O., E. L. Behrman, K. R. O'Brien, P. S. Schmidt and D. A. Petrov, 2014 Genomic
637 Evidence of Rapid and Stable Adaptive Oscillations over Seasonal Time Scales in
638 Drosophila. *PLoS Genet.* **10**: e1004775.
639 Burri, R., A. Nater, T. Kawakami, C. F. Mugal, P. I. Olason *et al.*, 2015 Linked selection and
640 recombination rate variation drive the evolution of the genomic landscape of differentiation
641 across the speciation continuum of *Ficedula* flycatchers. *Genome Res.* **25:** 1656-1665.
642 Catchen, J., S. Bassham, T. Wilson, M. Currey, C. O'Brien *et al.*, 2013 The population structure
643 and recent colonization history of Oregon threespine stickleback determined using
644 restriction-site associated DNA-sequencing. *Mo.l Eco.l* **22:** 2864-2883.
645 Catchen, J., P. A. Hohenlohe, S. Bassham, A. Amores and W. A. Cresko, 2013 Stacks: an
646 analysis tool set for population genomics. *Mol. Ecol.* **22:** 3124-3140.
647 Catchen, J. M., A. Amores, P. Hohenlohe, W. Cresko and J. H. Postlethwait, 2011 Stacks: building
648 and genotyping Loci de novo from short-read sequences. *G3 (Bethesda)* **1:** 171-182.
649 Charlesworth, B., D. Charlesworth and N. H. Barton, 2003 The effects of genetic and geographic
650 structure on neutral variation. *Ann. Rev. Ecol. Evol. Syst.* **34:** 99-125.
651 Charlesworth, B., M. T. Morgan and D. Charlesworth, 1993 The Effect of deleterious mutations on
652 neutral molecular variation. *GENETICS* **134:** 1289-1303.
653 Charlesworth, B., M. Nordborg and D. Charlesworth, 1997 The effects of local selection, balanced
654 polymorphism and background selection on equilibrium patterns of genetic diversity in
655 subdivided populations. *Genet. Res.* **70:** 155-174.
656 Clausen, J., W. M. Keck and W. M. Hiesey, 1941 Regional differentiation in plant species. *Am. Nat.*
657 **75:** 231-250.
658 Colosimo, P. F., K. E. Hosemann, S. Balabhadra, G. Villarreal, Jr., M. Dickson *et al.*, 2005
659 Widespread parallel evolution in sticklebacks by repeated fixation of *Ectodysplasin* alleles.
660 *Science* **307:** 1928-1933.
661 Cresko, W. A., A. Amores, C. Wilson, J. Murphy, M. Currey *et al.*, 2004 Parallel genetic basis for
662 repeated evolution of armor loss in Alaskan threespine stickleback populations. *Proc Natl*
663 *Acad Sci USA* **101:** 6050-6055.
664 Cruickshank, T. E., and M. W. Hahn, 2014 Reanalysis suggests that genomic islands of speciation
665 are due to reduced diversity, not reduced gene flow. *Mol. Ecol.* **23:** 3133-3157.
666 Drummond, A. J., and A. Rambaut, 2007 BEAST: Bayesian evolutionary analysis by sampling
667 trees. *BMC Evol. Biol.* **7:** 214.
668 Drummond, A. J., M. A. Suchard, D. Xie and A. Rambaut, 2012 Bayesian phylogenetics with
669 BEAUti and the BEAST 1.7. *Mol. Biol. Evol.* **29:** 1969-1973.
670 Elyashiv, E., S. Sattath, T. T. Hu, A. Strutsovsky, G. McVicker *et al.*, 2016 A Genomic map of the
671 effects of linked selection in *Drosophila*. *PLoS Genet.* **12**: e1006130.
672 Endler, J. A., 1977 *Geographic Variation, Speciation, and Clines*. Princeton University Press.
673 Fay, J. C., and C. I. Wu, 2000 Hitchhiking under positive Darwinian selection. *GENETICS* **155:**
674 1405-1413.
675 Fontaine, M. C., J. B. Pease, A. Steele, R. M. Waterhouse, D. E. Neafsey *et al.*, 2015 Mosquito
676 genomics. Extensive introgression in a malaria vector species complex revealed by
677 phylogenomics. *Science* **347:** 1258524.
678 Gillespie, J. H., 2000 Genetic drift in an infinite population: The pseudohitchhiking model.
679 *GENETICS* **155:** 909-919.

680 Gillespie, J. H., 2004 *Population genetics : a concise guide*. Johns Hopkins University Press,
681 Baltimore, Md.
682 Glazer, A. M., E. E. Killingbeck, T. Mitros, D. S. Rokhsar and C. T. Miller, 2015 Genome assembly
683 improvement and mapping convergently evolved skeletal traits in sticklebacks with
684 genotyping-by-sequencing. *G3 (Bethesda)* **5:** 1463-1472.
685 Hahn, M. W., 2008 Toward a selection theory of molecular evolution. *Evolution* **62:** 255-265.
686 Haldane, J. B. S., 1930 A mathematical theory of natural and artificial selection. (Part VI, Isolation.).
687 *Mathematical Proc. Cam. Phil. Soc.* **26:** 220-230.
688 Haller, B. C., and P. W. Messer, 2017 SLiM 2: flexible, interactive forward genetic simulations. *Mol.*
689 *Biol. Evol.* **34:** 230-240.
690 Hoekstra, H. E., K. E. Drumm and M. W. Nachman, 2004 Ecological genetics of adaptive color
691 polymorphism in pocket mice: geographic variation in selected and neutral genes. *Evolution*
692 **58:** 1329-1341.
693 Hohenlohe, P. A., S. Bassham, P. D. Etter, N. Stiffler, E. A. Johnson *et al.*, 2010 Population
694 genomics of parallel adaptation in threespine stickleback using sequenced RAD tags. *PLoS*
695 *Genet* **6:** e1000862.
696 Hudson, R. R., 1990 Gene genealogies and the coalescent process, pp. 1-44 in *Oxford Surveys in*
697 *Evolutionary Biology*, edited by D. J. Futuyma and J. Antonovics. Oxford University Press,
698 Oxford.
699 Hudson, R. R., M. Slatkin and W. P. Maddison, 1992 Estimation of levels of gene flow from DNA
700 sequence data. *GENETICS* **132:** 583-589.
701 Jones, F. C., M. G. Grabherr, Y. F. Chan, P. Russell, E. Mauceli *et al.*, 2012 The genomic basis of
702 adaptive evolution in threespine sticklebacks. *Nature* **484:** 55-61.
703 Jones, M. R., L. S. Mills, P. C. Alves, C. M. Callahan, J. M. Alves *et al.*, 2018 Adaptive
704 introgression underlies polymorphic seasonal camouflage in snowshoe hares. *Science* **360:**
705 1355-1358.
706 Kaeuffer, R., C. L. Peichel, D. I. Bolnick and A. P. Hendry, 2012 Parallel and nonparallel aspects of
707 ecological, phenotypic, and genetic divergence across replicate population pairs of lake
708 and stream stickleback. *Evolution* **66:** 402-418.
709 Kern, A. D., and M. W. Hahn, 2018 The neutral theory in light of natural selection. *Mol. Biol. Evol.*
710 **35:** 1366-1371.
711 Kirkpatrick, M., and N. Barton, 2006 Chromosome inversions, local adaptation and speciation.
712 *GENETICS* **173:** 419-434.
713 Langley, C. H., K. Stevens, C. Cardeno, Y. C. Lee, D. R. Schrider *et al.*, 2012 Genomic variation in
714 natural populations of *Drosophila melanogaster*. *GENETICS* **192:** 533-598.
715 Leaver, S. D., and T. E. Reimchen, 2012 Abrupt changes in defence and trophic morphology of
716 the giant threespine stickleback (*Gasterosteus* sp.) following colonization of a vacant
717 habitat. *Biol. J. Linn. Soc.* **107:** 494-509.
718 Lenormand, T., 2002 Gene flow and the limits to natural selection. *TREE* **17:** 183-189.
719 Lenormand, T., D. Bourguet, T. Guillemaud and M. Raymond, 1999 Tracking the evolution of
720 insecticide resistance in the mosquito *Culex pipiens*. *Nature* **400:** 861-864.
721 Lescak, E. A., S. L. Bassham, J. Catchen, O. Gelmond, M. L. Sherbick *et al.*, 2015 Evolution of
722 stickleback in 50 years on earthquake-uplifted islands. *Proc Natl Acad Sci USA* **112:**
723 E7204-7212.
724 Lewontin, R. C., 1974 *The genetic basis of evolutionary change*. Columbia University Press, New
725 York.
726 Marques, D. A., K. Lucek, J. I. Meier, S. Mwaiko, C. E. Wagner *et al.*, 2016 Genomics of rapid
727 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.

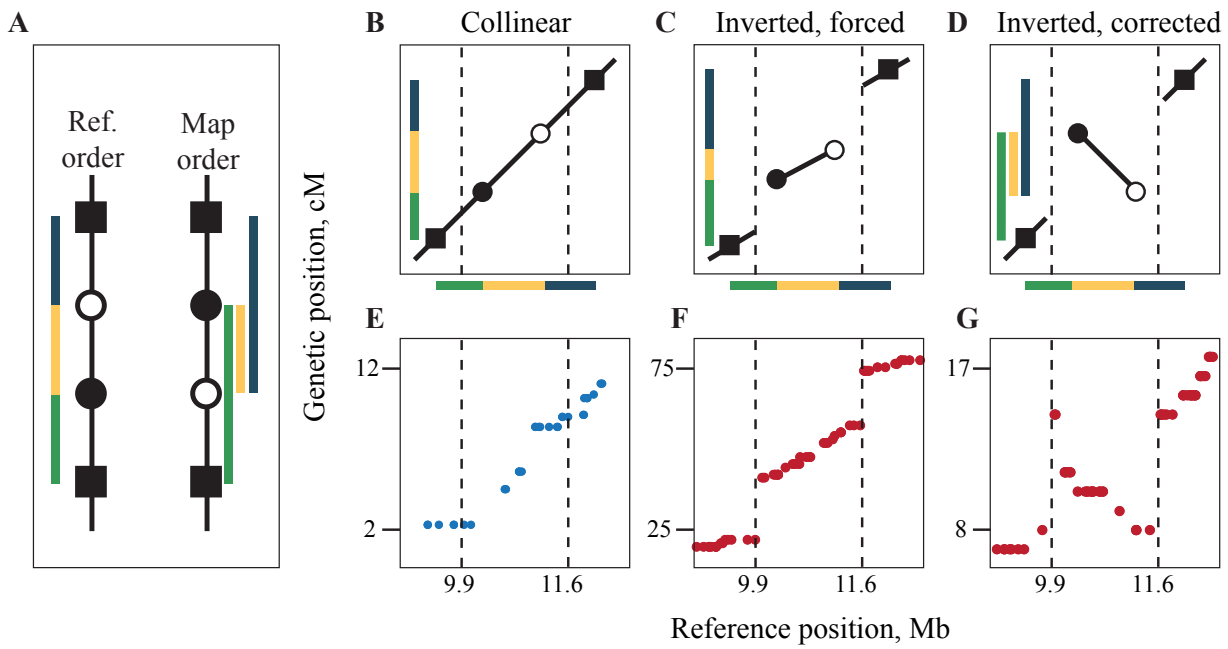
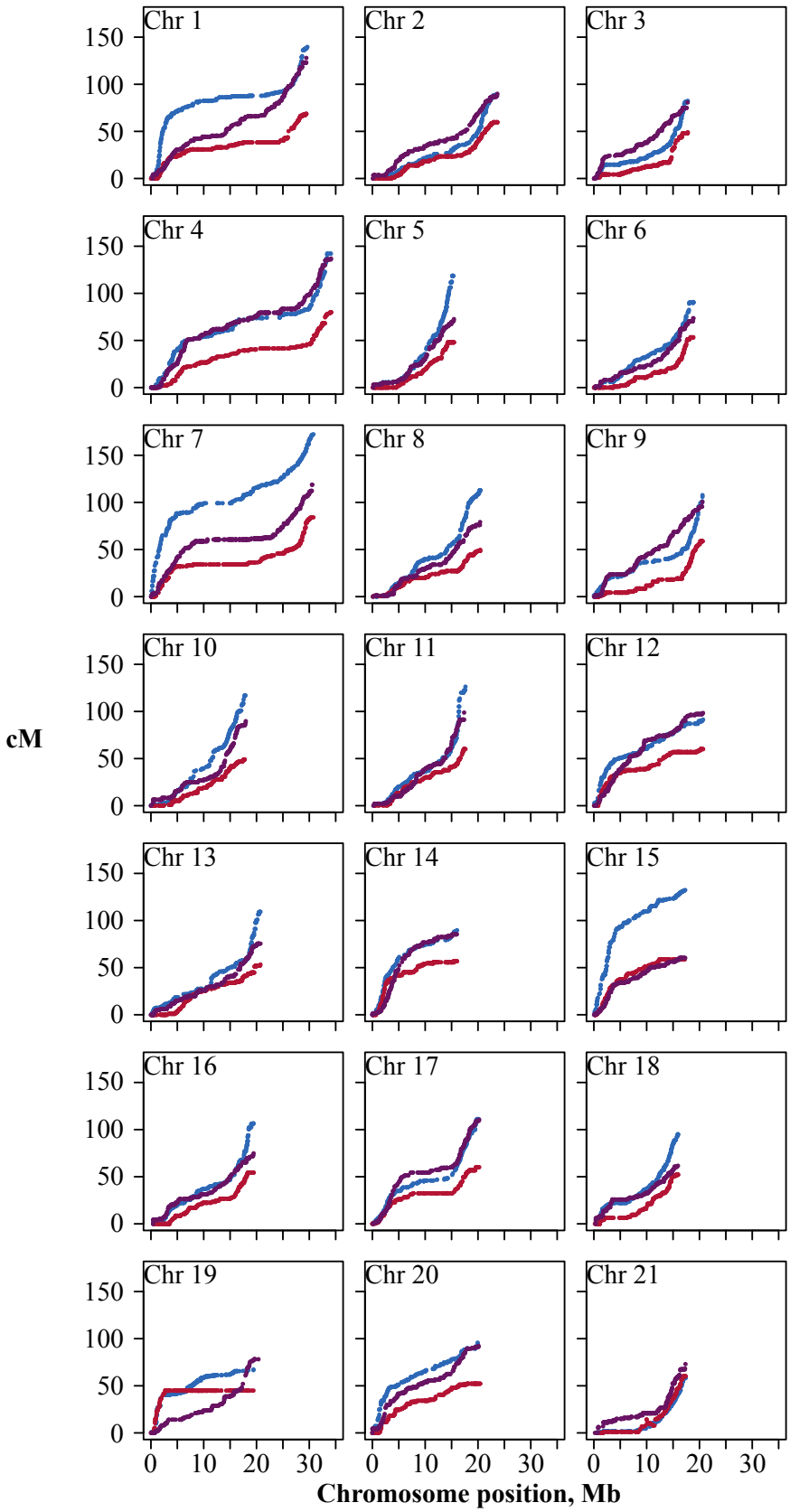


Figure S1. Reference-guided genetic map construction allows identification and correction of chromosomal inversions. A: Inversion polymorphisms create discrepancies between the reference physical genome and a genetic map. Colored bars indicate true map distances between genetic markers inside and on either side of the inversion. B: When the physical and genetic maps are collinear, everything's a-okay. Genetic map distances agree with the reference. C: An inverted genetic map forced into reference order results in a disjunct genetic map because crossover frequencies between markers inside and outside the inversion are inferred correctly. Note that the green and blue bars reflect the true distances between the markers involved but they are in the wrong order. D: Reversing the order of markers within the inversion preserved the relative map distances and reduced the overall map length. E-G: Genetic maps from the BT fish (E, collinear with reference) and the RS fish (F and G, inverted relative to reference) demonstrate this correction using empirical data.

Figure S2. Genetic maps for all 21 threespine stickleback chromosomes. Genetic position for each RAD marker is plotted against aligned physical position on the threespine stickleback reference genome. Blue: Boot Lake; Red: Rabbit Slough; Purple: F_1 hybrid.



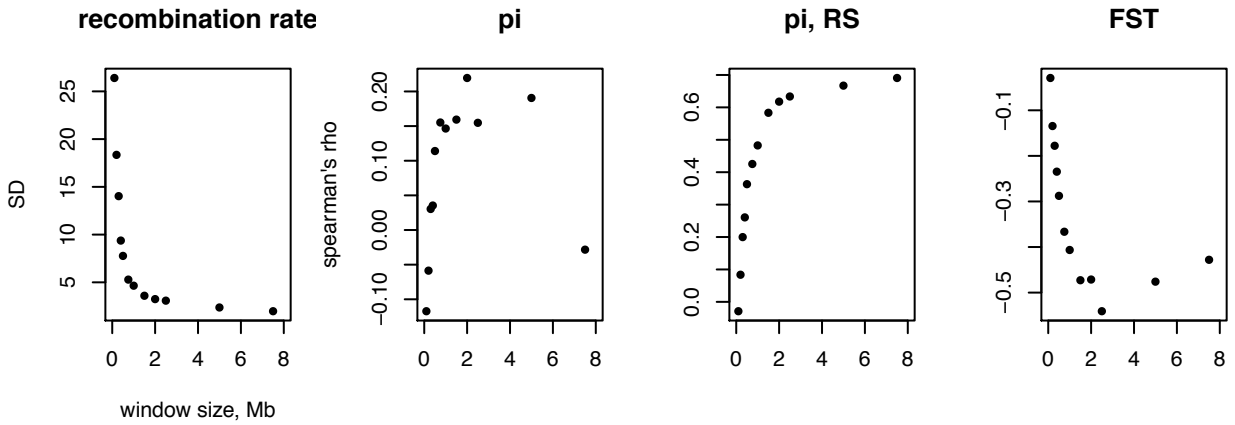
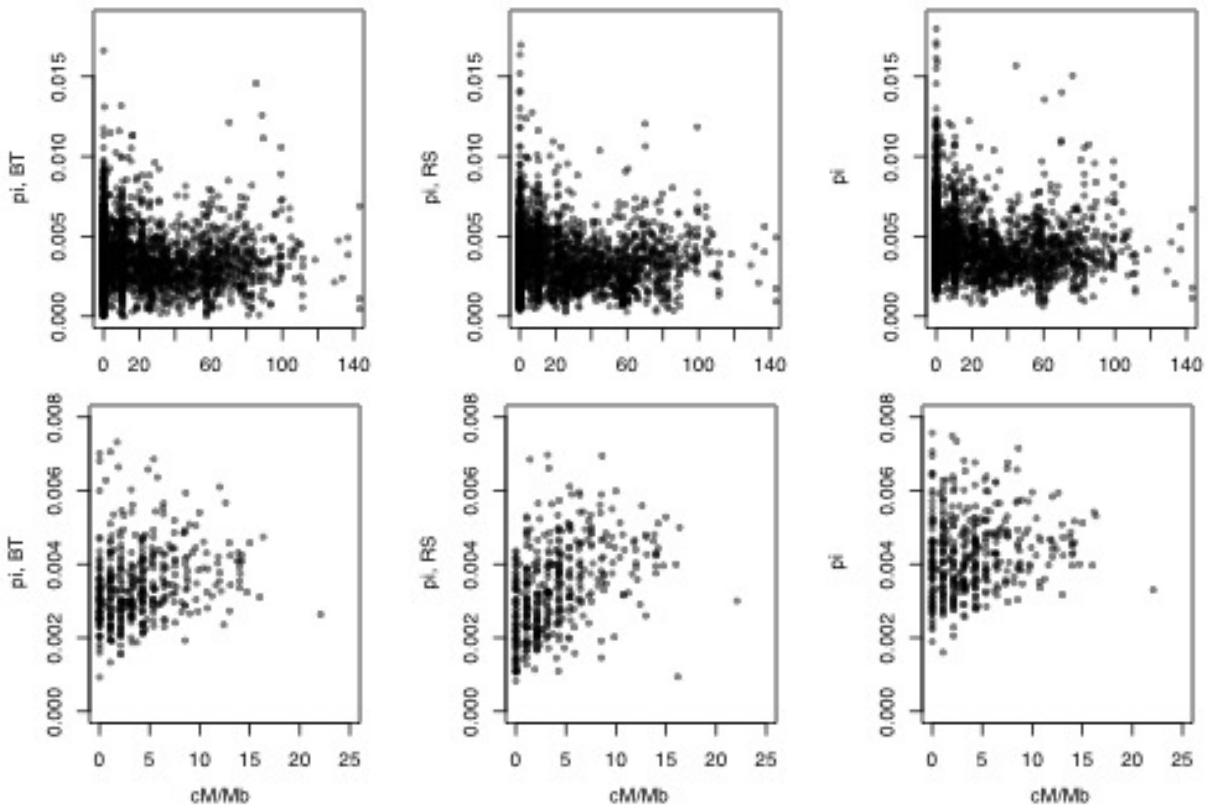


Figure S3. Variation in recombination rate and covariation with population genetic statistics at different genomic window sizes. All points are genome-wide summaries of recombination rate variation (cM/Mb) and covariation (shown as Spearman's ρ) with population genetic variation and differentiation at a given window size using non-overlapping genomic windows. SD: standard deviation; pi: sequence diversity across the entire population genomic dataset (Rabbit Slough, Boot Lake, Bear Paw Lake); pi, RS: sequence diversity among Rabbit Slough chromosomes; F_{ST}: differentiation between Rabbit Slough and Boot Lake population samples.



934 **Figure S4. Window-based estimates of recombination rate (cM/Mb) and sequence diversity**
935 **using (top) 100-kb non-overlapping windows and (bottom) 1000-kb non-overlapping**
936 **windows.** Note the difference in x- and y-axis scale with different window sizes.
937