# 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 97401
[2]Division of Biological Sciences, University of Montana, Missoula, MT 59812
[3]Department of Animal Biology, University of Illinois at Urbana-Champaign, Urbana, IL 61801

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

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

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 of 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: near loci under selection, marine chromosomes are depauperate of variation while these same regions among freshwater genomes are the most genetically diverse. Forward genetic simulations recapitulate this pattern when different selective environments also differ 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 (DOBZHANSKY AND QUEAL 1938; HUBBY AND LEWONTIN 1966; LEWONTIN AND HUBBY 1966). A recent wave of discoveries reveals that genetic variation is distributed heterogeneously across genomes as well (TURNER *et al.* 2005; BEGUN *et al.* 2007; HOHENLOHE *et al.* 2010; ELLEGREN *et al.* 2012). Are these patterns related? For example, does the manner in which evolutionary processes play out across the geography of organisms influence heterogeneous genomic patterns of genetic diversity? With ever-increasing access to DNA sequence data, evolutionary biologists have come to recognize that selection strongly influences patterns of linked neutral variation (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; HERMISSON AND PENNINGS 2005; NIELSEN *et al.* 2005; CUTTER AND PAYSEUR 2013). 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; VAN BELLEGHEM *et al.* 2018), plants (CLAUSEN *et al.* 1941; ANGERT *et al.* 2018; TAVARES *et al.* 2018), and vertebrates (HOEKSTRA *et al.* 2004; JONES *et al.* 2018). Consequently,

86 gene flow can strongly influence the geographic distribution of genetic variation among populations
87 (WRIGHT 1931; SLATKIN 1985; SLATKIN 1991) and across species boundaries (PARDO-DIAZ *et al.*
88 2012; FONTAINE *et al.* 2015). Although genetic introgression from one species to another may
89 impede neutral divergence, it is now recognized as an important source of adaptive variation
90 (STANKOWSKI AND STREISFELD 2015; WALLBANK *et al.* 2016; MEIER *et al.* 2017; JONES *et al.* 2018).
91 Even in the absence of selection, population subdivision alone may affect the overall
92 abundance of neutral genetic variation within a species. Theoretical models provide useful – but at
93 times contradictory – predictions on this matter, with subdivision either increasing or decreasing
94 expected coalescence times depending on model specifications and assumptions. For example,
95 Slatkin (SLATKIN 1991) generalized the island model to demonstrate that population structure
96 increases coalescence times ($T_T$) in a structured population by:

97 $$T_T = T_0 + \frac{(d-1)^2}{2dm} \quad (1)$$

98 Where d is the number of demes, m is the migration rate, and $T_0$ is the expected coalescence time
99 with no population structure (using the nomenclature of CHARLESWORTH *et al.* 2003). Whitlock and
100 Barton (1997), however, used a similar model but allowed deme extinction and variable
101 contributions of each deme to the total population. The result was a decrease in effective
102 population size ,$N_e$, with increasing subdivision, and thus a decrease in expected levels of genetic
103 variation.
104 Despite this growing understanding of the general importance of genomic organization of
105 selected loci and geographic context of organisms in the wild on the structuring genetic variation,
106 precise understanding of how they *interact* to produce heterogeneous patterns of genetic variation
107 across the genome is still lacking. Therefore, empirical and simulation studies that examine the joint
108 effects of selection regime, geographic context, and recombination are necessary for fully
109 understanding the effects of these biological factors on genomic patterns of genetic variation in the
110 wild.
111 Here, we investigate how multiple evolutionary forces interact to shape the genomic and
112 geographic distribution of genetic variation in the threespine stickleback fish, ***Gasterosteus***
113 ***aculeatus***. The threespine stickleback is distributed holarctically in coastal marine, brackish and
114 freshwater habitats. The large marine population has repeatedly given rise to derived freshwater
115 populations, resulting in phenotypic divergence and parallel evolution throughout the species range
116 (BELL AND FOSTER 1994; CRESKO *et al.* 2004; JONES *et al.* 2012). When marine stickleback invade
117 freshwater habitats, more than a quarter of the genome can rapidly respond to the action of

118 divergent selection (HOHENLOHE *et al.* 2010; TEREKHANOVA *et al.* 2014; BASSHAM *et al.* 2018). This
119 process largely uses standing genetic variation (COLOSIMO *et al.* 2005; SCHLUTER AND CONTE 2009;
120 ROESTI *et al.* 2015; SAMUK *et al.* 2017) that can be found ubiquitously at low frequency in marine
121 populations (BASSHAM *et al.* 2018) much of which has likely been maintained for millions of years by
122 significant bidirectional gene flow (CALDERA AND BOLNICK 2008; KITANO *et al.* 2008; BERNER *et al.*
123 2009; NELSON AND CRESKO 2018).
124 While previous work has often focused on the genomic targets of selection, we focus here
125 on the processes that maintain and structure variation in regions of the genome physically linked to
126 those targets. We use a combination of population genomics of wild stickleback, forward-time
127 simulations using SLiM (HALLER AND MESSER 2017), and dense genetic maps to support a model
128 whereby differences in population structure between marine and freshwater habitats has led to
129 divergent outcomes of molecular evolution at loci linked to adaptive variants. Our results provide
130 new links between theoretical and empirical evolutionary genetics, new tools for future work in the
131 stickleback system, and new perspectives on the maintenance of genetic variation.
132

133 **Methods**

134 *Study populations and natural genetic variation*

135 Wild threespine stickleback were collected from Rabbit Slough (N 61.5595, W 149.2583),
136 Boot Lake (N 61.7167, W 149.1167), and Bear Paw Lake (N 61.6139, W 149.7539). Rabbit
137 Slough (RS) is an offshoot of the Knik Arm of Cook Inlet and is known to be populated by
138 anadromous populations of stickleback that are stereotypically marine in phenotype and genotype
139 (Figure 1; CRESKO *et al.* 2004). Boot Lake (BT) and Bear Paw Lake (BP) are both shallow lakes
140 formed during the end-Pleistocene glacial retreat approximately 12 thousand years ago. Fish were
141 collected in the summers of 2009 (RS), 2010 (BP), and 2014 (BT) using wire minnow traps and
142 euthanized ***in situ*** with Tricaine solution. Euthanized fish were immediately fixed in 95% ethanol
143 and shipped to the Cresko Laboratory at the University of Oregon (Eugene, OR, USA).
144 We generated restriction site-associated DNA (RAD) libraries of five fish each from RS and
145 BT and four fish from BP. Genomic DNA was isolated from ethanol-preserved fin clips by
146 proteinase K digestion followed by DNA extraction with solid phase reversible immobilization (SPRI)
147 beads. We created RAD libraries using the single digest and shearing method of Baird et al (BAIRD
148 *et al.* 2008) with the modifications of Nelson and Cresko (NELSON AND CRESKO 2018). Genomic
149 DNA from each fish was digested with *PstI*-HF (New England Biolabs) and ligated to Illumina P1

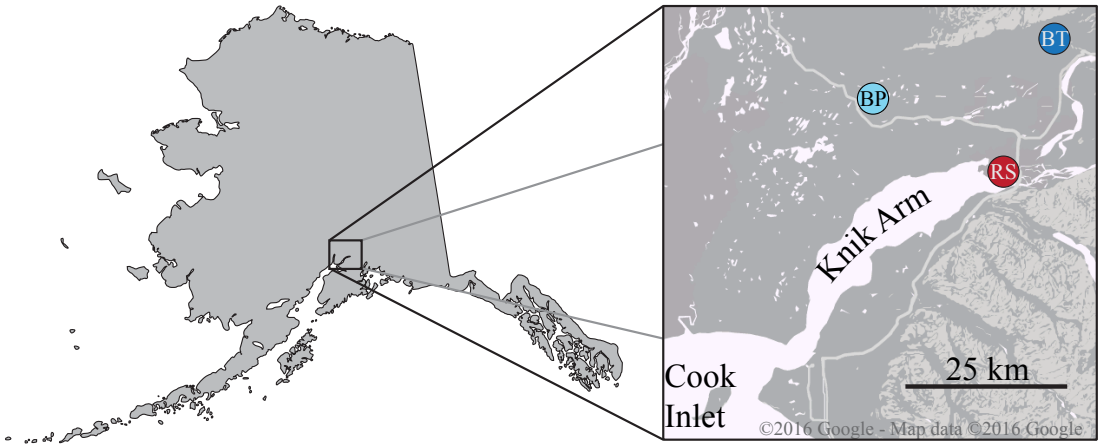



**Figure 1. Sampling locations of marine and freshwater stickleback populations from southcentral Alaska.** Inset: Geography of Cook Inlet near Anchorage, AK. Marine stickleback were sampled from Rabbit Slough (RS, red). Freshwater stickleback were sampled from Boot Lake (BT, dark blue) and Bear Paw Lake (BP, light blue). All areas in the inset were glaciated during the Late Wisconsin glaciation until ~10 thousand years ago.

150 adaptors with 6 bp inline barcodes. All barcodes differed by at least two positions, allowing for
151 recovery of sequence reads with single errors in the barcode sequence. Ligated samples were
152 then multiplexed at approximately equimolar concentrations and mechanically sheared via
153 sonication to a fragment range of ~200-800 bp. Sheared DNA was size selected by extraction
154 from a 1.25% agarose gel to generate a narrow insert size range of 425 bp to 475 bp. This size
155 range allowed consistent overlap of paired-end Illumina reads for the construction of local contigs
156 surrounding restriction enzyme cut sites. We then ligated Illumina P2 adaptors to the size-selected
157 libraries and amplified P1/P2-adapted fragments with 12 cycles of PCR using Phusion-HF
158 polymerase (New England Biolabs). RAD libraries were then sequenced in a single lane on an
159 Illumina HiSeq 2500 to generate paired-end 250-bp sequence reads. All libraries generated for this
160 study were sequenced at the University of Oregon's Genomics and Cell Characterization Core
161 Facility (GC3F: http://gc3f.uoregon.edu).
162 We used the Stacks analysis pipeline to process RAD sequence read pairs and call SNPs
163 (CATCHEN *et al.* 2011; CATCHEN *et al.* 2013b). Raw reads were first demultiplexed without quality
164 filtering using process_radtags, and then quality filtered using process_shortreads. This allowed for
165 read trimming, rather than strict removal, if quality decreased toward the end of the first-end read.
166 Overlapping read pairs were then merged using fastq-join (ARONESTY 2011), allowing for up to 25%
167 of bases in the overlapping region to mismatch, and the resulting contigs were trimmed to 350 bp.

168 Any read pairs that failed to merge, or were shorter than 350 bp, were removed from further
169 analysis. This step was required for processing reads through the Stacks pipeline. Below, a "locus"
170 refers to the combined sequence assembled from a single restriction site in the stickleback
171 genome.

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

180 We used the program PHASE (STEPHENS *et al.* 2001; SCHEET AND STEPHENS 2006) to
181 combine sequence information from both RAD tags at a ***PstI*** cut site and generate phased
182 haplotypes at each RAD locus. We wrote custom Python scripts to identify all unique haplotypes at
183 each pair of RAD tags and code them as alleles at a single, multiallelic locus. We required that
184 each individual included in this analysis was genotyped at both RAD tags. Loci containing
185 individuals only genotyped at a single RAD tag were removed from further analysis. RAD
186 haplotypes at each locus represent 696 bp of contiguous genomic sequence, giving us high-
187 quality estimates of sequence diversity and divergence even with our relatively small population-
188 level sample sizes (NEI 1987, chapter 13; CRUICKSHANK AND HAHN 2014; NELSON AND CRESKO
189 2018). Sequencing of wild stickleback resulted in 57,992 RAD loci distributed across all 21
190 threespine stickleback chromosomes, averaging 7,514 bp between adjacent RAD loci.

191 *Population genetic statistics*

192 The scripting language R version 3.5 (R Core Team 2016) was used for all downstream
193 data analysis. We estimated differentiation among threespine stickleback populations (all pairwise
194 combinations) and among ecotypes (combined freshwater ponds versus RS) using a haplotype-
195 based $F_{ST}$ (equation 3 in HUDSON *et al.* 1992) implemented in the R package 'PopGenome'
196 (PFEIFER *et al.* 2014). We calculated π per site within and among populations and absolute
197 sequence divergence ($d_{XY}$) at each RAD locus by calculating pairwise distances between all RAD
198 haplotypes with the R package 'ape' (PARADIS *et al.* 2004; POPESCU *et al.* 2012).

199 Previously, we detected patterns of reciprocal monophyly between marine and freshwater
200 haplotypes using maximum clade credibility trees generated in BEAST v1.7 (DRUMMOND AND
201 RAMBAUT 2007; DRUMMOND *et al.* 2012; NELSON AND CRESKO 2018; SUCHARD *et al.* 2018). Tree
202 topologies for all RAD loci were inferred from MCMC runs of $10^6$ states with 10% burnin periods.
203 We used blanket priors and parameters across all RAD loci, including a coalescent tree prior and
204 the GTR+Γ substitution model. Monophyly of haplotypes from each population (RS, BT, BP) and
205 each habitat (marine, freshwater) was assessed using the is.monophyletic() function of the R
206 packages 'ape'. Here, we use topological classifications that we inferred previously and designate
207 gene trees with reciprocally monophyletic marine and freshwater haplogroups (1,129 of 57,992
208 RAD loci) as 'divergent' loci (Figure 2B, see also NELSON AND CRESKO 2018).

209 *Forward simulations using SLiM*

210 We used forward simulations implemented in SLiM (HALLER AND MESSER 2017) to model
211 the effects of selection, linkage, and population structure in a manner that reflects the stickleback
212 metapopulation structure (Figure 2). We simulated a metapopulation of 2000 diploid individuals and
213 a genome consisting of a single chromosome with a genetic length of ten centiMorgans (cM). The
214 chromosome contained 50 kb of freely recombining sequence (recombination rate $1x10^{-6}$ per bp)
215 on either side of a 2 kb nonrecombining 'core' containing the locally adaptive locus. Per-base
216 mutation rate was kept constant at $5x10^{-7}$, resulting in a population-scaled mutation rate ($4N\mu$ =
217 0.004) aligned with genome-wide estimates of genetic diversity in stickleback (HOHENLOHE *et al.*
218 2010; NELSON AND CRESKO 2018). The general form of the simulations was as follows:

219 1. Initialize population $H_1$
220 a. $N_{H1}$ = 1000 diploids
221 b. evolve for 10,000 generations
222 2. Create *k* new populations, collectively "$H_2$"
223 a. each of size 1000/*k*
224 b. set migration rate m to $H_1$
225 c. evolve for 10,000 generations
226 3. Introduce mutation as single copy into $H_2$
227 a. Deleterious in $H_1$, advantageous in $H_2$
228 b. Run selection for [4,10,20]N generations
229 c. (If locally adaptive mutation is lost, go to [3])
230 4. End simulation

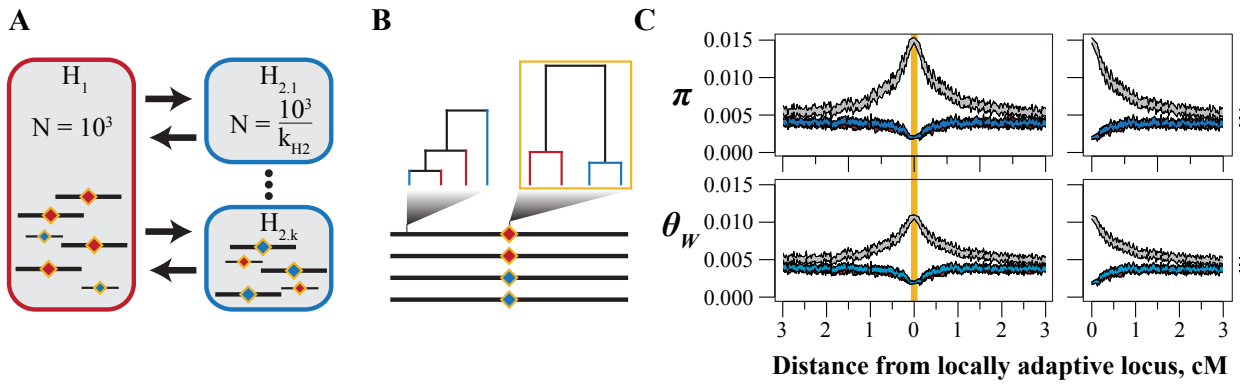



**Figure 2. The structure of simulations to test the effects of population structure on genetic variation in a model of local adaptation. A:** Habitat $H_1$ (red) (: red, $H_2$: blue) consists of a single deme while habitat $H_2$ (blue) consists of *k* demes. We adjust the size of each deme in $H_2$ so that the total census size of each habitat is constant across simulations. Diamonds represent a single locus with locally adaptive alleles. **B:** We assessed genetic diversity at the locally adaptive locus and in non-overlapping windows. We used the pattern of reciprocal monophyly (outlined in gold) to classify RAD loci as 'divergent' in the threespine stickleback population genomic dataset. (See Methods and NELSON AND CRESKO [2018]). **C:** Genetic diversity within and among allelic classes in a simulation where $k_{H2}$ = 1. Shown are 95% confidence intervals of nucleotide diversity (π) and Watterson's θ after 10N generations of selection. Because patterns of variation are symmetrical about the locus under selection, we present 'folded' curves (panel C, right) throughout this work.

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 $N_{H1}$ = 1000 and no sequence variation. We evolved this population for $10N_{H1}$ (10,000) generations. The second burn-in of $10N_{H1}$ generations began by creating $k_{H2}$ new populations, each of size $N_{H1}/k_{H2}$, 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 $H_1$, and each new population, now collectively referred to as habitat $H_2$. 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 $H_2$ that is advantageous in $H_2$ and deleterious in $H_1$. Progression through the selection phase was conditional on establishment of the locally adaptive mutation: if the

246 mutation was lost from the metapopulation, the simulation restarted at the end of the second burn-
247 in.
248 Genetic diversity among chromosomes carrying $H_1$- and $H_2$-adaptive alleles was assessed
249 at the end of the selection phase. We sampled 10 chromosomes of each allelic state and
250 calculated nucleotide diversity ($\pi$) and Watterson's $\theta$ ($\theta_W$) within and among chromosomes of the
251 two allelic states in non-overlapping 250 bp (0.025 cM) bins. We include results from $\theta_W$ because
252 this statistic is based on the number of segregating sites and is therefore more reflective of the
253 abundance of sequence variation and less affected by shifts in allele frequencies. We fold the
254 calculations of genetic diversity about the locally adaptive locus to emphasize the effects of
255 evolutionary forces on $\pi$ at a given distance from the locus (Figure 2C). These summaries mirror
256 our estimates of genetic variation on stickleback chromosomes and results from theoretical work
257 on the effects of local adaptation on linked variation (CHARLESWORTH *et al.* 1997).

258 *Laboratory crosses and genetic mapping*

259 To compare how heterogenous genomic divergence on the physical map is distributed
260 across the genetic map, we generated mapping families from laboratory lines of fish derived from
261 the Boot Lake (BT) and an $F_1$ hybrid female (hereafter $F_1$) produced from a cross between a BT
262 female and a Rabbit Slough (RS) male. These crosses allowed us to examine variation in the
263 recombinational landscape within and among chromosomes in distinct, evolutionarily relevant
264 genetic backgrounds. We also generated a genetic map of a RS-derived male (Figure S1) but,
265 because ecotype was confounded with sex (SARDELL *et al.* 2018), we limit our discussion to the BT
266 and $F_1$ maps.
267 All maps were constructed using a pseudo-testcross design, which takes advantage of
268 existing heterozygosity in outbred individuals without the need to generate inbred lines or $F_1$
269 mapping parents. To generate the BT mapping family, we manually crossed unrelated lab-reared
270 individuals. We mapped the $F_1$ female by backcrossing it to a BT male. All progeny were raised to
271 14 days post-fertilization, euthanized with MS-222 (Sigma Aldrich), and fixed in 95% ethanol. We
272 extracted genomic DNA from whole progeny and from pectoral and caudal fins from all parents
273 using proteinase K digestion (Qiagen) followed by DNA purification with SPRI beads.
274 RAD genotyping of progeny and parents was used to identify segregating haplotypes using
275 a RAD-seq protocol similar to that described previously but using the restriction enzyme *SbfI*. RAD-
276 seq data from all mapping crosses were processed with the Stacks analysis pipeline (CATCHEN *et*
277 *al.* 2013b). We demultiplexed and quality filtered sequences with process_shortreads and aligned

278 them to the stickleback reference genome with GSNAP (WU AND NACU 2010). We used ref_map.pl
279 to identify RAD tags and call genotypes. The Stacks component program genotypes was used to
280 identify segregating markers for export to genetic mapping software. We specified a minimum
281 coverage of 3x to call individual genotypes and required that a marker be genotyped in at least
282 50% of progeny. Below, we use the term 'RAD marker' to refer to a RAD tag with segregating
283 haplotypes.

284 Below, we present genetic maps for the female parent from each cross (Table 1; Table S1).
285 By conducting pseudo-testcrosses, we identified polymorphic RAD markers segregating in all
286 mapping parents. However, to investigate the genome-wide recombination landscape, as well as
287 relationships between recombination rate and natural levels of polymorphism and divergence, we
288 restricted our analysis to parents of the same sex for which we observed segregating markers on
289 all 21 chromosomes with no gaps of more than 1 mega-base pairs (Mbp) between adjacent
290 markers.

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

297 While fixed marker orders do not explicitly allow the detection of structural variation among
298 genomes — such as chromosomal inversions that are known to exist among stickleback
299 populations — discrepancies in the estimated map do provide indirect, and correctable, evidence
300 of changes to map order (Figure S2). For example, inversions in genetic map order relative to the

**Table 1. 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 (BT) | 189 | 6041 | 2255 | 0.993 | 5.20 |
| BT subset | 94 | 6169 | 2396 | 0.993 | 5.52 |
| $F_1$ hybrid | 102 | 6669 | 1878 | 0.995 | 4.33 |
| $F_1$ subset | 94 | 7678 | 1958 | 0.995 | 4.51 |

301 reference order spuriously inflate recombination rates between markers closely flanking inversion
302 breakpoints when inverted segments are forced into the wrong orientation. This is because genetic
303 map distances are estimated independently while the physical distance between markers is
304 drastically underestimated. The observed jumps in map distance on either side of the inversion are
305 equal to each other and to the total map length of the inversion. Reversing the marker order within
306 the inversion removes these artifacts and reduces the overall map length of the region. We
307 observed and corrected this artifact for a single previously known inversion on chromosome 21
308 (Figure S1, Figure S2, JONES *et al.* 2012). Two other known inversions on chromosomes 1 and 11
309 were too small to create artifacts in our data.

310 *Recombination-polymorphism correlations*

311 We employed three methods to investigate the relationships between the recombinational
312 landscape and patterns of polymorphism within and among natural populations. To compare our
313 results to other studies in stickleback (SAMUK *et al.* 2017), we quantified genome-wide correlations
314 among recombination rate and population genetic statistics. We divided the stickleback genome
315 into non-overlapping windows and calculated average recombination rates (in centiMorgans per
316 mega-base pair, cM/Mbp), sequence diversity ($\pi_{BT}$, $\pi_{RS}$, and $\pi$), and genetic divergence ($F_{ST}$, $d_{XY}$) in
317 each window. We calculated $F_{ST}$ between RS and BT as a measure of recent genetic differentiation
318 and $d_{XY}$ between RS and the combined freshwater populations as a measure of long-term genetic
319 divergence (NELSON AND CRESKO 2018). RAD marker density made estimates of local
320 recombination rates less reliable and more variable when using small window sizes (e.g. 100-kbp
321 windows; Figures S3 and S4). As a result, we show results from 1-Mbp genomic windows. We
322 used nonparametric correlations to test for correlations between variables because the
323 distributions of most variables lacked normality even using standard data transformations. Below
324 we present Spearman's rank order correlations. Kendall's tau and parametric linear models gave
325 qualitatively similar results (Table S2).

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

333 the R package 'ksmooth', we binned each chromosome into equally sized intervals, the number of
334 which we set equal to the number of segregating RAD markers on the genetic map with fewest
335 markers. For each interval, we calculated genetic position from each laboratory cross and $F_{ST}$
336 between RS and BT populations (HOHENLOHE *et al.* 2010; NELSON AND CRESKO 2018) within a 250-
337 kbp normally distributed kernel. We also imputed the genetic positions of all divergent loci using
338 the lm() function in R: we found the nearest flanking RAD markers in each mapping cross and
339 predicted the genetic position of the divergent locus assuming a constant recombination rate
340 between the flanking markers. With these approaches we were able to make direct comparisons
341 between recombination within and among genetic maps from different genetic backgrounds and
342 patterns of polymorphism in the populations from which the laboratory lines were derived.

343 *Data availability*

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

348 *Animal care and compliance*

349 Treatment of animals followed protocols approved by the University of Oregon Institutional Animal
350 Care and Use Committee (IACUC).

351

352 **Results**

353 *Chromosomes from freshwater but not marine stickleback contain abundant linked*
354 *variation*

355 To examine patterns of variation linked to loci under selection, we first partitioned RAD loci
356 from the population genomics dataset into those with evidence of complete marine-freshwater
357 lineage sorting ('divergent', with allelic states 'marine' and 'freshwater', see Figure 2B), those on
358 the same chromosome as a divergent locus ('linked'), and those on chromosomes without
359 divergent loci ('unlinked'). At divergent loci, average sequence diversity was partitioned almost
360 entirely among chromosomes carrying marine and freshwater alleles; overall π averaged 0.0067
361 per site (Figure 3A, black-filled diamond, genome-wide mean = 0.0042); average $d_{XY}$ between
362 marine and freshwater alleles was 0.0124, nearly three-fold higher than the genome-wide average
363 (0.0044). In contrast, and as expected at locally adaptive loci (Figure 2C), π was substantially

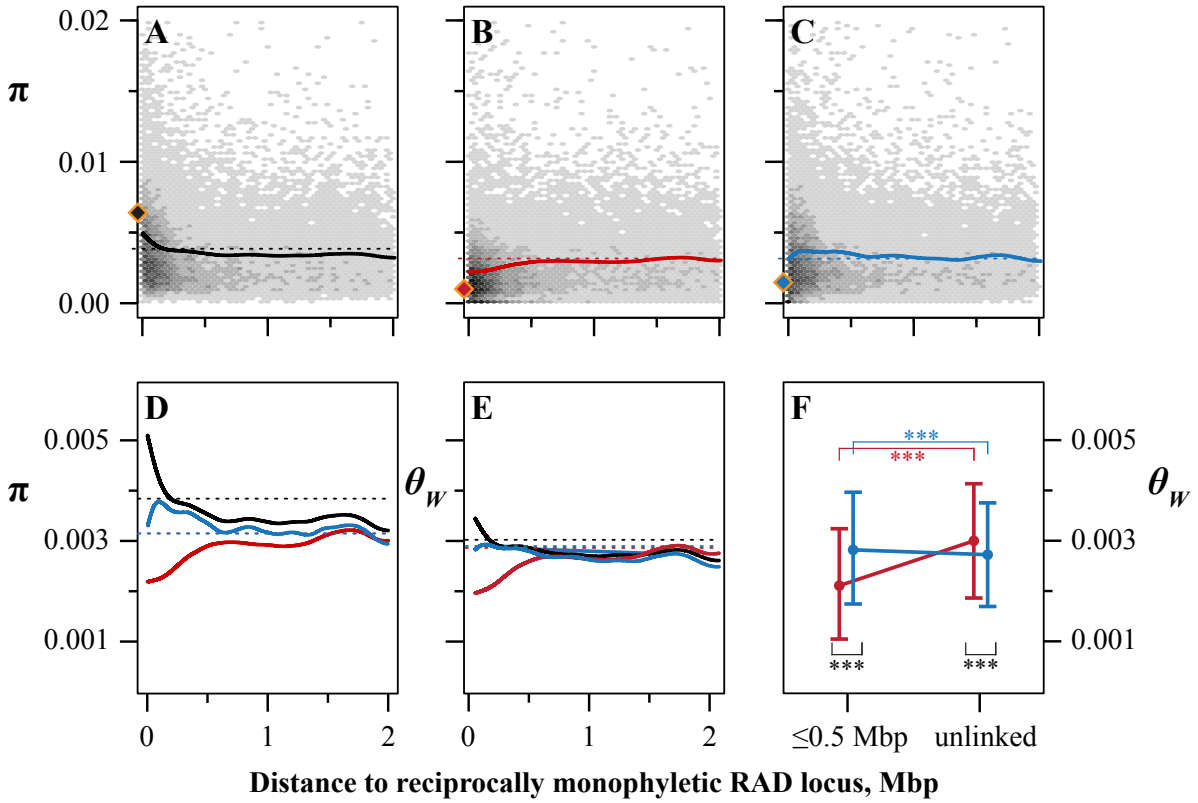



**Figure 3. Patterns of sequence variation differ on marine and freshwater stickleback chromosomes. A-C:** Genome-wide sequence diversity (π) as a function of the distance from a RAD locus with reciprocally monophyly of marine and freshwater haplotypes. Diamonds show average sequence diversity among 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:** Sequence diversity in the combined dataset (marine [RS] + freshwater [BT and BP]). **B:** Sequence diversity on chromosomes carrying marine haplotypes. **C:** Sequence diversity on chromosomes carrying freshwater haplotypes. **D:** Smoothed splines as in A-C. **E:** 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. *** $p < 0.001$. **F:** Splines as in **A-D** of Watterson's θ.

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 (Figure3C, blue-filled diamond).

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

372 divergent loci were nearly as polymorphic as those directly impacted by divergent selection. As
373 expected under local adaptation, π among chromosomes sampled from marine RS stickleback
374 was lowest in close proximity to divergent loci (Figure 3B), and a substantial fraction (12%) of RAD
375 loci within 250 kb of a divergent locus showed no variation at all on these chromosomes ($\pi_{RS} = 0$;
376 1651/13762 loci).

377 In stark contrast, diversity among freshwater chromosomes showed a proximity effect that
378 was distinct from either RS or the combined populations (Figure 3C-E). Rather than being lowest
379 near divergent loci, diversity actually increased with proximity to a divergent locus, peaking
380 approximately 200 kbp away on average before reversing direction. This pattern persisted using
381 either π or Watterson's theta ($\theta_W$), indicating that the density of segregating sites on freshwater
382 chromosomes is highest near divergent loci and that the signal we observe is not simply due to a
383 greater abundance of intermediate frequency alleles. We also note that this increase in diversity
384 qualitatively persisted within both freshwater populations individually (Figure S5).

385 Linkage to divergent loci, therefore, was associated with opposing effects on genetic
386 variation in stickleback: decreasing it among marine chromosomes while increasing variation
387 among freshwater chromosomes (Figure 3E). Chromosomes with no evidence of divergent
388 selection had a slightly higher density of segregating sites in RS than the combined freshwater
389 populations (Figure 3E, Table 2), though π was indistinguishable (Figure S6). However, within 500
390 kbp of a divergent locus, genetic diversity (π and $\theta_W$) was greater among freshwater populations
391 than in RS (Figure 3E, Table 2, Figure S6; population*proximity interaction $p \leq 10^{-10}$).

392 *Population structure maintains linked variation on simulated chromosomes*

393 To determine how population structure within the freshwater habitat influences patterns of
394 linked variation in stickleback, we conducted forward simulations of chromosomes under a model

**Table 2. ANOVA results for proximity-by-ecotype analyses.** F statistics are from ANOVAs on untransformed values, ***$p \leq 10^{-10}$. For Tukey's Honest Significant Difference (HSD), '+' and '-' indicate an increase or decrease in diversity, respectively, compared to variation at unlinked loci on marine (RS) chromosomes. 'linked' indicates loci ≤ 500 kb from a divergent locus, as in Figure 3.

| | $F_{1,34068}$ | | | Tukey's HSD vs $RS_{unlinked}$ | | |
|---|---|---|---|---|---|---|
| | ecotype | linkage | ecotype*linkage | $RS_{linked}$ | $FW_{linked}$ | $FW_{unlinked}$ |
| **π** | 858.8*** | 62.5*** | 382.8*** | - | + | = |
| **$\theta_W$** | 343.4*** | 149.6*** | 358.9*** | - | - | - |

395 of local adaptation with migration using SLiM v2.6 (HALLER AND MESSER 2017, Figure 2). In
396 simulations with two habitat types, each composed of a single, panmictic population (Figure 4,
397 column 1), total genetic diversity was highest at and adjacent to the locally adaptive locus (Figure
398 2C). Within allelic classes, diversity was lowest in proximity to the locally adaptive locus and
399 recovered equally within both allelic classes with increasing recombinational distance from the
400 selected locus (Figure 2C; Figure 4). This scenario is essentially the same as that presented by
401 Charlesworth, Nordborg, and Charlesworth (1997) and the results are qualitatively the same.
402 Adding to these simulations population structure informed by the stickleback system led to
403 previously undocumented results. When the second habitat ($H_2$) – representing the freshwater

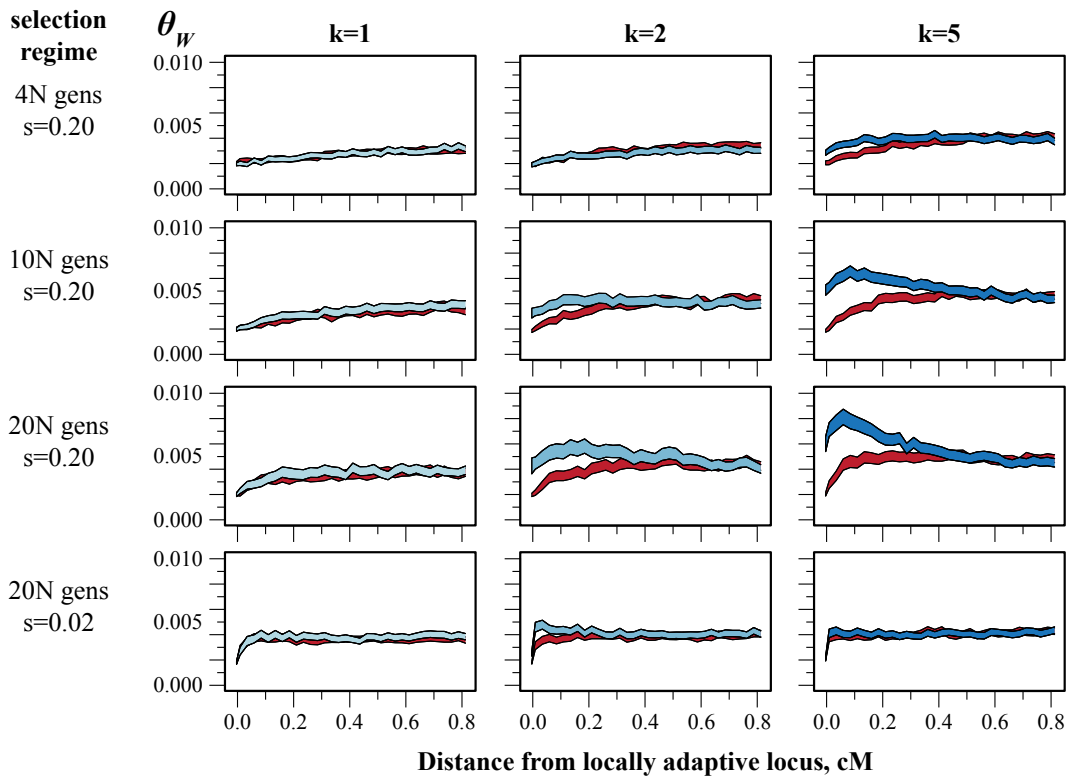



**Figure 4. Asymmetric population structure generates asymmetric patterns of linked variation on simulated chromosomes.** Simulations were performed as described in Methods and Figure 2. Bands are 95% confidence intervals of Watterson's θ at a given distance from a locally adaptive locus (red: $\theta_W$ on chromosomes carrying $H_1$-adaptive allele, blue: $\theta_W$ on chromosomes carrying $H_2$-adaptive allele). Columns show the effect of increasing population structure (*k*) in habitat $H_2$. Rows show the effect of increasing the length of the selection phase of the simulations. Rows 1-3 show results under strong selection (*s* = 0.20). Row 4 shows results for simulations identical to row 3 but with moderate selection (*s* = 0.02). All simulations were performed using a migration rate, *m*, of one migrant per generation between habitat $H_1$ and each population of habitat $H_2$. Total diversity (see Figure 2) is excluded to highlight variation within each allelic class.

404 stickleback habitat – consisted of two or more demes (Figure 2A; Figure 4, columns 2 and 3) the
405 additional structure maintained greater variation exclusively on $H_2$-adaptive ($H_2^+$) chromosomes, an
406 effect dependent on the duration and strength of selection and the number of demes. Under
407 strong selection (s = 0.20), population structure had a barely noticeable effect on variation until the
408 length of the selection phase was on the order of the neutral coalescence time (~$9N_H$ generations,
409 SLATKIN 1991; CHARLESWORTH *et al.* 2003). Beyond this point, any amount of population structure
410 resulted in higher levels of variation on $H_2^+$ than $H_1^+$ chromosomes within ~0.2 cM from the
411 selected locus. Moderate selection (s=0.02) was generally ineffective at altering among-habitat
412 levels of variation, though we did observe a modest effect on $H_2^+$ chromosomes immediately
413 adjacent to the selected locus.

414 Increasing population structure in $H_2$ led to substantial levels of variation near the selected
415 locus (Figure 6, column 3; Figure S7). When $H_2$ consisted of five demes, genetic variation on $H_2^+$
416 chromosomes was greatest ~0.1 cM from the selected locus even though levels of variation
417 among allelic states remained indistinguishable in distal regions of the chromosome. Greater
418 subdivision of $H_2$ first accentuated the accumulation of variation ($5 \leq k \leq 15$, Figure S7) but
419 eventually attenuated it when $k \geq 20$. While we did fully not explore this attenuation, we note that
420 we held both *N* and *m* (the average number of migrants per generation) constant across
421 simulations; higher values of *k* were therefore accompanied by greater total migration between
422 habitats and stronger within-deme drift in $H_2$.

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

430 *The recombination landscape varies among individuals*

431 RAD sequencing of both mapping crosses resulted in over 6000 segregating markers, with
432 markers averaging 70 kbp apart on the BT map and 56 kbp apart on the $F_1$ map (Table 1). Mean
433 per-locus sequencing depth was 35x and 23x for the BT and $F_1$ families, respectively (range:
434 BT=[28x,57x], $F_1$=[10x,90x]; Table S1). RAD markers were consistently genotyped in over 90% of
435 progeny (BT: mean=98%, range=[95%,99%]; $F_1$: mean=91%, range=[57%,99%]).

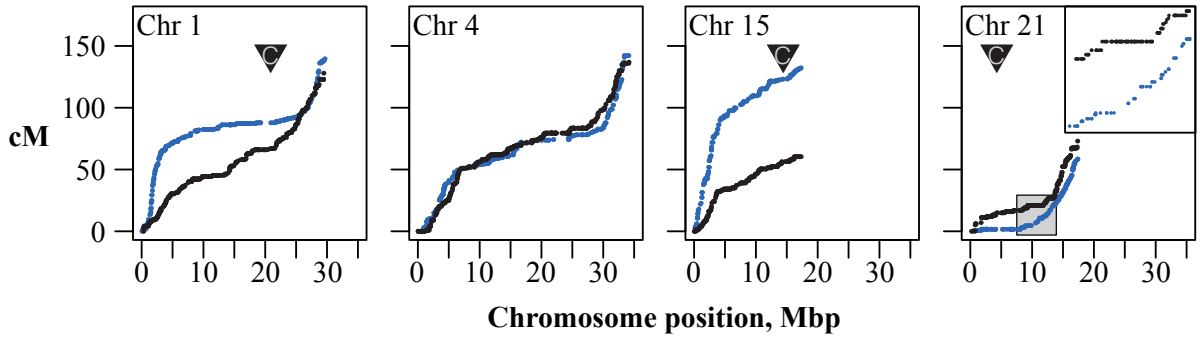



**Figure 5. The recombination landscape varies within and among genetic maps of BT and $F_1$ hybrid threespine stickleback.** Each point represents a RADseq-based marker segregating within a mapping family. Blue: BT, black: $F_1$. Inverted triangles represent the locations of centromeric repeats, where known, from SARDELL ET AL (2018). The inset on chromosome 21 shows suppression of recombination due to inversion heterozygosity in the $F_1$. Genetic maps for all chromosomes and for the marine (RS) map are given in Figure S1.

Patterns of recombination were generally consistent between the genetic maps (Figure 5, Figure S1). 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 (Figure 5, chromosome 15; Figure S1).

We also observed stark differences between genetic maps. As expected, recombination in the hybrid map was completely suppressed within and immediately surrounding an inversion on chromosome 21 but this region recombined freely in the collinear BT map (length 4.3 cM, Figure 5). Conversely, on chromosome 1 the BT map showed bias in recombination toward chromosome ends expected on larger chromosomes. This bias was almost entirely absent in the hybrid map, with recombination occurring steadily throughout the chromosome. In both cases, the total map lengths of each chromosome were similar (chromosome 1: BT=129 cM, $F_1$=137.5 cM; chromosome 21: BT=61.0 cM; $F_1$=75.7 cM), indicating that these differences were due to differences in the distribution but not the number of crossovers.

### *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, Table S2, Appendix 1). 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 $F_{ST}$ values

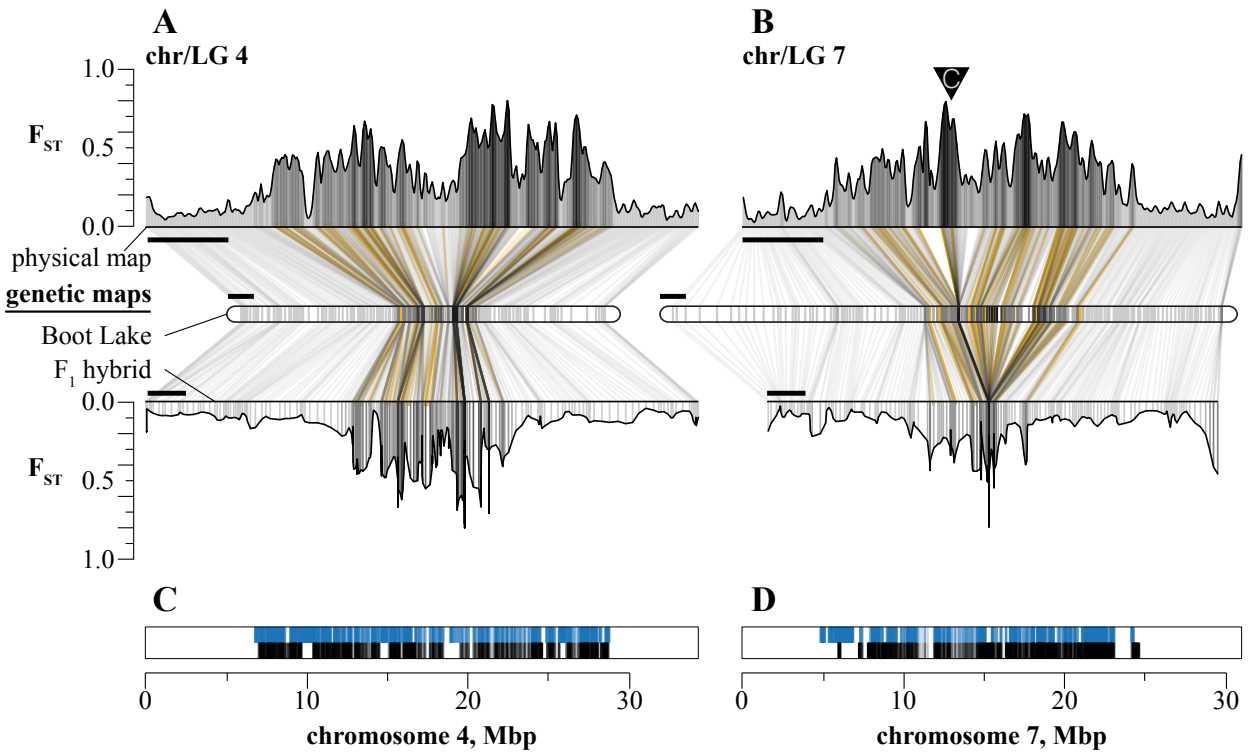



**Figure 6. The recombination landscape extends to genomic reach of divergent selection. A-B:** RS-BT $F_{ST}$ scans were performed on the physical map (top, scale bars = 5 Mbp) and imputed onto the freshwater and $F_1$ hybrid genetic maps (middle and bottom, scale bars = 10 cM) for chromosomes 4 and 7. 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. Positions of divergent RAD loci are shown in gold. Inverted triangle represents the position of centromeric repeats (SARDELL ET AL 2018). **C-D:** physical positions of all RAD loci within 0.2 cM from a divergent locus on the freshwater map (blue lines) and the $F_1$ hybrid map (black lines).

in excess of 0.4. Regions above this threshold spanned 8.1% of the physical genome (35.3 Mbp) but only 3.7% of the BT genetic map and 3.3% of the $F_1$ hybrid genetic map (Figure S8). On chromosome 4, the large central regions of differentiation are similarly compressed on both genetic maps but this was not always the case. On chromosome 21, recombination was suppressed – and genomic differentiation compressed – only on the hererokaryotypic $F_1$ map (Figure S1). 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 S1). On the physical map, chromosome 7 contained three peaks of strong differentiation between BT and RS spanning over 9 Mbp. 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).

468 Because our simulations suggested that population structure could interact with selection
469 to maintain variation, but only with tight linkage (≤0.2 cM), we imputed the positions of RAD loci
470 from the population genomic dataset onto the BT and hybrid genetic maps. Based on average
471 genome-wide recombination rates, 0.2 cM equates to less than 50 kbp of sequence space (BT
472 map: 36 kbp; $F_1$ map: 44 kbp) and includes ~6% of sequenced loci. However, the clustering of
473 divergent loci in regions of low recombination resulted in ~20% of all loci in the dataset fitting this
474 linkage criterion (BT map: 19.5%; $F_1$ map: 20.4%; Figure 6C,D; Figure S9). These tightly linked loci
475 averaged over 200 kbp away from a divergent locus (BT map: 219 kbp; $F_1$ map: 256 kbp) and
476 included loci over 2 Mbp away (max, BT map: 2.3 Mbp; max, $F_1$ map: 2.9 Mbp).
477

478 **Discussion**

479 *Asymmetric population structure maintains asymmetries in patterns of linked genetic*
480 *variation*

481 The patterns of linked variation that we document here highlight not only the importance of
482 selection but also the structure of the threespine stickleback metapopulation in maintaining genetic
483 variation. Throughout the species range, the marine population is remarkably uniform
484 phenotypically, which we now know is reflected genetically by minimal isolation-by-distance over
485 thousands of kilometers (BELL AND FOSTER 1994; CATCHEN *et al.* 2013a; DEFAVERI *et al.* 2013). This
486 large population with few barriers to migration is contiguous with thousands of freshwater lake and
487 stream populations that are more clearly isolated by geography. While the influence of freshwater
488 populations on the evolution of the species was once unclear (BELL AND FOSTER 1994), it is now
489 evident that gene flow between freshwater and marine stickleback populations is common and
490 may facilitate adaptation through the indirect sharing of alleles among freshwater populations
491 (SCHLUTER AND CONTE 2009).

492 Our findings suggest that the structuring of the freshwater habitat acts as a reservoir of
493 genetic variation in the species, and that this effect is particularly potent in regions of the genome
494 experiencing divergent selection. In genomic regions that are recombinationally distant to targets of
495 divergent selection, gene flow among habitats homogenizes variation such that levels of genetic
496 variation are similar (Figure 3) and differentiation is low (Figure 6) across habitats. At divergent loci
497 themselves, selection has maintained a small number of haplotypes in each habitat and variation is
498 partitioned almost entirely among habitats (Figure 2, Figure 3, NELSON AND CRESKO 2018).
499 However, in the substantial fraction of the genome that is tightly linked to divergent loci, we

500 hypothesize that selection has reduced the effective migration rate between alternative habitats
501 such that the geographic structuring of the freshwater habitat becomes a significant factor
502 affecting levels of genetic variation. Because migration among freshwater populations occurs
503 primarily through the marine environment, divergent selection accentuates the effect of population
504 structure in the freshwater habitat.

505 Forward simulations support this hypothesis. We found that under realistic migration rates
506 (CALDERA AND BOLNICK 2008; BERNER *et al.* 2009) and selection coefficients (BARRETT *et al.* 2008;
507 KITANO *et al.* 2008), such asymmetric population subdivision ***by itself*** provides a simple explanation
508 for the counterintuitive observation that smaller, isolated populations contain abundant linked
509 variation (Figure 3, Figure 4). The degree to which population structure increased levels of diversity
510 depended both on the amount of substructure and the length of time over which selection acted.
511 While even minimal substructure eventually resulted in patterns of diversity mirroring those we
512 observed in stickleback, this effect required a selection phase on the order of the expected
513 coalescence time of the metapopulation. This follows from the fact that selection and population
514 structure protect variants from fixation or loss at adaptive and linked loci, respectively. Our findings
515 generally agree with expectations from Slatkin's (1991) model, where limited migration among
516 demes increases the expected coalescence times of chromosomes sampled from a
517 metapopulation. Nearby an alternatively adaptive locus, selection accentuates this effect by
518 reducing effective migration rates among demes of alternative habitats.

519 Our empirical and simulated results, in combination with previous findings in threespine
520 stickleback (REIMCHEN 1994; KAEUFFER *et al.* 2012; ROESTI *et al.* 2014), have important implications
521 for the long-term evolution of the species. Freshwater populations are ecologically diverse (BELL
522 AND FOSTER 1994; MCKINNON AND RUNDLE 2002; KAEUFFER *et al.* 2012; LEAVER AND REIMCHEN
523 2012; REIMCHEN *et al.* 2013) and the relative impacts of parallel versus non-parallel selection
524 pressures among freshwater populations is actively debated (KAEUFFER *et al.* 2012; STUART *et al.*
525 2017). While it is possible that multifarious selection pressures also contribute to the maintenance
526 of more genetic diversity on chromosomes from freshwater fish, our model does not require
527 additional selection to maintain linked variation. Furthermore, the freshwater pond populations
528 studied here are ecologically similar and geographically proximal (CRESKO *et al.* 2004), so it is likely
529 that parallel selection on a common pool of variation is largely responsible for the patterns of
530 variation we observe. Put another way, we find no need to invoke more complex selection regimes
531 in order to explain the patterns of variation we observe. Also studying ecologically similar

532 freshwater populations, Roesti et al (2014) found that parallel adaptation led to increases in $F_{ST}$
533 among freshwater populations near peaks of marine-freshwater divergence, and inferred that
534 sweeps of common SNPs on different genetic backgrounds led to this pattern. We have shown
535 these linked genomic regions to be not only the most heavily partitioned but also the most diverse
536 in the genome (Figure 3F, Figure S6). We find it likely that selection and gene flow happening over
537 shorter evolutionary timescales (tens to thousands of years), like those investigated by Roesti et al.
538 (2014), have occurred throughout the evolutionary history of this species and have long-lasting,
539 collective effects on patterns of genomic variation over the course of millions of years.

540 *A variable recombination landscape increases the genomic reach of selection and*
541 *simplifies the architecture of divergence*

542 In addition to the consequences of asymmetric population structure in maintaining genetic
543 variation unevenly across geography, the recombination landscape also can have profound effects
544 on heterogenous genome-wide patterns of genetic variation. Linkage to or among selected sites is
545 now thought to affect many, if not most, variable sites in the genome (SCHRIDER AND KERN 2017;
546 KERN AND HAHN 2018), and suppression of recombination among the genomes of diverging
547 populations and species appears in large part to determine patterns of genetic differentiation and
548 divergence (BURRI *et al.* 2015; SAMUK *et al.* 2017; VIJAY *et al.* 2017; STANKOWSKI *et al.* 2018). In the
549 threespine stickleback, genetic differentiation among phenotypically divergent populations is known
550 to accumulate in regions of low recombination (ROESTI *et al.* 2013; MARQUES *et al.* 2016; SAMUK *et*
551 *al.* 2017). Our work adds an important outcome of this phenomenon: low recombination adjacent
552 to targets of divergent selection extends selection's reach into linked regions, maintaining genetic
553 variation and structuring it among chromosomes with divergent evolutionary histories.

554 The accumulation of genetic diversity adjacent to targets of selection required tight linkage
555 in our simulations, potentially limiting the relevance of this model in explaining patterns of genetic
556 variation. However, because adaptive divergence has occurred principally in regions of low
557 recombination, a substantial fraction of the genome is likely near enough (~0.2 cM) to divergent loci
558 for our model to explain the accumulation of variation on freshwater chromosomes. The absolute
559 genetic distance through which divergent selection can influence linked variation is proportional to
560 the strength of selection (e.g. Figure 4) and inversely proportional to the migration rate (e.g. Figure
561 S10, CHARLESWORTH *et al.* 1997). While we did not fully explore either parameter space, our
562 chosen values are congruent with empirical estimates from stickleback. Estimates of migration
563 rates vary widely across studies of stickleback, from nearly zero to rates that could swamp

564 divergent selection (CALDERA AND BOLNICK 2008; BERNER *et al.* 2009). Estimated selection
565 coefficients on lateral plate armor are consistently above 0.2 (BARRETT *et al.* 2008; KITANO *et al.*
566 2008), and the total effect of selection on a locus depends on the total number of nearby selected
567 sites (i.e. the 'selection density': AESCHBACHER *et al.* 2017). Given the rate of adaptation observed
568 in recently colonized freshwater habitats (LESCAK *et al.* 2015; BASSHAM *et al.* 2018) combined with
569 generally low recombination rates, the effect of selection may at times extend across entire
570 chromosomes.

571 Our results also provide general insight into the variability and evolution of the
572 recombination landscape. First, genomic divergence that occurs in the centers of the largest
573 chromosomes (HOHENLOHE *et al.* 2010; JONES *et al.* 2012; BASSHAM *et al.* 2018; NELSON AND
574 CRESKO 2018) occurs within regions of consistently low recombination across maps from different
575 genetic backgrounds(HOHENLOHE *et al.* 2012; ROESTI *et al.* 2013; SAMUK *et al.* 2017). The negative
576 association between divergence and recombination rate is a common finding across systems
577 (NACHMAN 2002; CARNEIRO *et al.* 2009; GERALDES *et al.* 2011; CUTTER AND PAYSEUR 2013; BURRI *et*
578 *al.* 2015; AESCHBACHER *et al.* 2017), and can result from both negative (BURRI *et al.* 2015;
579 STANKOWSKI *et al.* 2018) and positive selection (AESCHBACHER *et al.* 2017; SAMUK *et al.* 2017)
580 between allopatric populations (BURRI *et al.* 2015) and those currently exchanging genes
581 (CARNEIRO *et al.* 2009; AESCHBACHER *et al.* 2017). Our results further demonstrate that stickleback
582 are no exception to this rule and that divergent selection in the face of (historical) gene flow has
583 generated this relationship at both at the local (i.e. specific genomic regions, Figure 6) and
584 genome-wide scales (Table S2, Figure S8). Second, although these results generally hold using
585 both intra- and inter-population genetic maps, we find that recombination rates in regions of
586 divergence were often lowest in the $F_1$ hybrid map (Figure 5; Figure 6B; Figure S8). While this was
587 expected within chromosomal inversions, where the alternative homozygotes demonstrated steady
588 recombination throughout the inverted region, this was also the case on chromosome 7 despite
589 the apparent lack of any large scale, simple structural variation (Figures 2 and 4) although clusters
590 of smaller structural variants that could not be detected in our maps could also reduce
591 recombination. On the hybrid map, this entire region collapsed to a region inherited essentially as a
592 single Mendelian locus.

593 The specific reductions in recombination on the hybrid map are suggestive of the evolution
594 of the recombination landscape itself and, if true, could have profound implications for adaptation
595 and genomic divergence in stickleback. For example, inversions are advantageous when

596 recombination between multiple linked alleles that contribute to fitness (either in an additive or
597 epistatic fashion) is maladaptive (KIRKPATRICK AND BARTON 2006). The three previously identified
598 inversions on chromosomes 1, 11 and 21 (JONES *et al.* 2012) are associated with divergence
599 between the freshwater and marine habitats and occur in regions of the genome that readily
600 recombine in chromosomal homozygotes (Figure 4, Figure S1, Figure S2). Had inversions not
601 evolved, gene flow and recombination among marine and freshwater populations may have been
602 strong enough to prevent adaptive divergence in these genomic regions (LENORMAND 2002;
603 YEAMAN AND WHITLOCK 2011; AESCHBACHER *et al.* 2017).
604 In contrast, our work and that of others (ROESTI *et al.* 2013; GLAZER *et al.* 2015) has shown
605 that megabase pair-scale inversions have not evolved across the largest regions of divergence in
606 the genome. In these regions, low recombination rates — which may themselves have evolved
607 adaptively or been ancestral platforms for adaptive divergence — combined with strong selection
608 has been effective at maintaining allelic combinations across megabase pairs of genomic space.
609 These are also regions of exceptional sequence divergence between marine and freshwater
610 chromosomes (Table S2; SAMUK *et al.* 2017; NELSON AND CRESKO 2018). Sequence divergence
611 alone may limit double-strand break resolution as crossovers in these regions (MODRICH AND LAHUE
612 1996; OPPERMAN *et al.* 2004; LI *et al.* 2006), biasing crossovers toward regions of lower sequence
613 divergence. Explicit tests of this hypothesis, for example with $F_3$ mapping families where the effects
614 of marine-freshwater hetero- and homozygosity in specific genomic regions can be directly
615 assessed, will be a productive avenue of research.
616 The coincidence of reduced recombination and genomic divergence may help explain the
617 well documented pattern of repeatable and rapid adaptive divergence in stickleback (BELL AND
618 FOSTER 1994; LESCAK *et al.* 2015) that is largely the result of the reuse of standing genetic variation
619 (COLOSIMO *et al.* 2005; SCHLUTER AND CONTE 2009; DEAGLE *et al.* 2012; TEREKHANOVA *et al.* 2014;
620 ROESTI *et al.* 2015; MARQUES *et al.* 2016; BASSHAM *et al.* 2018). Because freshwater populations
621 are typically founded by marine stickleback, the sources of standing genetic variation are low
622 frequency variants in the marine population that nearly always exist in a heterozygous state with a
623 marine genome (BASSHAM *et al.* 2018). Our results suggest that the recombination landscape may
624 therefore facilitate the maintenance of freshwater haplotypes during their transit through the marine
625 environment by reducing recombination even in collinear genomic regions. Multi-megabase
626 haplotypes — potentially containing many alleles contributing to local adaptation — then have a
627 higher probability of being selected in concert when a new freshwater population is established.

628 *Conclusions*

629 Divergent natural selection is a powerful force for the maintenance of genetic variation in nature.
630 When local adaptation plays out on variable geographic and recombinational landscapes, we find
631 that the effect of selection on genetic variation are amplified and shape patterns of linked variation
632 in unexpected ways. Here we document using simulation and empirical studies that asymmetric
633 population subdivision among habitats in stickleback leads to an overall greater maintenance of
634 diversity in freshwater as compared to the larger ancestral marine population. Furthermore, we
635 hypothesize that the stickleback recombinational landscape is the product of repeated adaptive
636 evolution that transforms a large genomic architecture of marine-freshwater divergence into a
637 much more simplified genetic architecture. If true, two consequences are the efficient maintenance
638 of adaptive divergence in the face of gene flow, and widespread linked selection that eliminates
639 genetic variation in the panmictic marine population while maintaining genetic diversity in the
640 freshwater habitat. The interaction of selection, recombination, and population structure has turned
641 small, isolated freshwater stickleback populations into primary reservoirs of standing genetic
642 variation that may potentiate future evolutionary change. More broadly, our results underscore how
643 integrated studies of evolutionary and genetic processes can yield exciting, unexpected patterns
644 and deeper understanding when we jointly consider the processes and their interactions with one
645 another.
646
647
648
649
650
651
652
653
654
655
656
657
658
659

## 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, Nadia Singh, Peter Ralph, Lila Fishman, Madeline Chase, Emily Beck, Kristin Alligood and the evolutionary genetics group at the University of Montana for advice for useful discussions throughout the development of this project. And thanks to Katie Peichel, David Begun and three 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 U S A 114**:** 7061-7066.

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.

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

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.

Begun, D. J., A. K. Holloway, K. Stevens, L. W. Hillier, Y. P. Poh *et al.*, 2007 Population genomics: whole-genome analysis of polymorphism and divergence in Drosophila simulans. PLoS Biol 5**:** e310.

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.

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

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.

Caldera, E. J., and D. I. Bolnick, 2008 Effects of colonization history and landscape structure on genetic variation within and among threespine stickleback (Gasterosteus aculeatus) populations in a single watershed. Evol Ecol Res 10**:** 575-598.

Carneiro, M., N. Ferrand and M. W. Nachman, 2009 Recombination and speciation: loci near centromeres are more differentiated than loci near telomeres between subspecies of the European rabbit (Oryctolagus cuniculus). Genetics 181**:** 593-606.

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

Catchen, J., P. A. Hohenlohe, S. Bassham, A. Amores and W. A. Cresko, 2013b 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-Genes Genomes Genetics 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 Systemat 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 U S A 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.

Cutter, A. D., and B. A. Payseur, 2013 Genomic signatures of selection at linked sites: unifying the disparity among species. Nat Rev Genet 14**:** 262-274.

Deagle, B. E., F. C. Jones, Y. F. Chan, D. M. Absher, D. M. Kingsley *et al.*, 2012 Population genomics of parallel phenotypic evolution in stickleback across stream-lake ecological transitions. Proc R Soc B-Biol Sci 279**:** 1277-1286.

Defaveri, J., T. Shikano, Y. Shimada and J. Merilä, 2013 High degree of genetic differentiation in marine three-spined sticklebacks (Gasterosteus aculeatus). Mol Ecol 22**:** 4811-4828.

Dobzhansky, T., and M. L. Queal, 1938 Genetics of Natural Populations. II. Genic Variation in Populations of Drosophila Pseudoobscura Inhabiting Isolated Mountain Ranges. Genetics 23**:** 463-484.

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.

Ellegren, H., L. Smeds, R. Burri, P. I. Olason, N. Backström *et al.*, 2012 The genomic landscape of species divergence in Ficedula flycatchers. Nature**:** 1-5.

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, pp. Princeton University Press, Princeton, NJ.

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.

Geraldes, A., P. Basset, K. L. Smith and M. W. Nachman, 2011 Higher differentiation among subspecies of the house mouse (Mus musculus) in genomic regions with low recombination. Mol Ecol 20**:** 4722-4736.

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

Glazer, A. M., E. E. Killingbeck, T. Mitros, D. S. Rokhsar and C. T. Miller, 2015 Genome Assembly Improvement and Mapping Convergently Evolved Skeletal Traits in Sticklebacks with Genotyping-by-Sequencing. G3-Genes Genomes Genetics 5**:** 1463-1472.

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

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

Hermisson, J., and P. S. Pennings, 2005 Soft sweeps: Molecular population genetics of adaptation from standing genetic variation. Genetics 169**:** 2335-2352.

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, M. Currey and W. A. Cresko, 2012 Extensive linkage disequilibrium and parallel adaptive divergence across threespine stickleback genomes. Philos Trans R Soc Lond B Biol Sci 367**:** 395-408.

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.

Hubby, J. L., and R. C. Lewontin, 1966 A molecular approach to the study of genic heterozygosity in natural populations. I. The number of alleles at different loci in Drosophila pseudoobscura. Genetics 54**:** 577-594.

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

794 Jones, F. C., M. G. Grabherr, Y. F. Chan, P. Russell, E. Mauceli *et al.*, 2012 The genomic basis
795 of adaptive evolution in threespine sticklebacks. Nature 484**:** 55-61.
796 Jones, M. R., L. S. Mills, P. C. Alves, C. M. Callahan, J. M. Alves *et al.*, 2018 Adaptive
797 introgression underlies polymorphic seasonal camouflage in snowshoe hares. Science
798 360**:** 1355-1358.
799 Kaeuffer, R., C. L. Peichel, D. I. Bolnick and A. P. Hendry, 2012 Parallel and nonparallel
800 aspects of ecological, phenotypic, and genetic divergence across replicate population
801 pairs of lake and stream stickleback. Evolution 66**:** 402-418.
802 Kern, A. D., and M. W. Hahn, 2018 The neutral theory in light of natural selection. Mol Biol Evol
803 35**:** 1366-1371.
804 Kirkpatrick, M., and N. Barton, 2006 Chromosome inversions, local adaptation and speciation.
805 Genetics 173**:** 419-434.
806 Kitano, J., D. I. Bolnick, D. A. Beauchamp, M. M. Mazur, S. Mori *et al.*, 2008 Reverse evolution
807 of armor plates in the threespine stickleback. Curr Biol 18**:** 769-774.
808 Langley, C. H., K. Stevens, C. Cardeno, Y. C. Lee, D. R. Schrider *et al.*, 2012 Genomic variation
809 in natural populations of Drosophila melanogaster. Genetics 192**:** 533-598.
810 Leaver, S. D., and T. E. Reimchen, 2012 Abrupt changes in defence and trophic morphology of
811 the giant threespine stickleback (Gasterosteus sp.) following colonization of a vacant
812 habitat. Biol J Linn Soc 107**:** 494-509.
813 Lenormand, T., 2002 Gene flow and the limits to natural selection. Trends in Ecology &
814 Evolution 17**:** 183-189.
815 Lenormand, T., D. Bourguet, T. Guillemaud and M. Raymond, 1999 Tracking the evolution of
816 insecticide resistance in the mosquito Culex pipiens. Nature 400**:** 861-864.
817 Lescak, E. A., S. L. Bassham, J. Catchen, O. Gelmond, M. L. Sherbick *et al.*, 2015 Evolution of
818 stickleback in 50 years on earthquake-uplifted islands. Proc Natl Acad Sci U S A 112**:**
819 E7204-7212.
820 Lewontin, R. C., and J. L. Hubby, 1966 A molecular approach to study of genic heterozygosity
821 in natural populations .II. Amount of variation and degree of heterozygosity in natural
822 populations of Drosophila pseudoobscura. Genetics 54**:** 595-609.
823 Li, L., M. Jean and F. Belzile, 2006 The impact of sequence divergence and DNA mismatch
824 repair on homeologous recombination in Arabidopsis. Plant J 45**:** 908-916.
825 Marques, D. A., K. Lucek, J. I. Meier, S. Mwaiko, C. E. Wagner *et al.*, 2016 Genomics of rapid
826 incipient speciation in sympatric threespine stickleback. PLoS Genet 12**:** e1005887.
827 Maynard Smith, J., and J. Haigh, 1974 The hitch-hiking effect of a favourable gene. Genetical
828 Research 23**:** 23-35.
829 McKinnon, J. S., and H. D. Rundle, 2002 Speciation in nature: the threespine stickleback
830 model systems. TREE 17**:** 480-488.
831 Meier, J. I., D. A. Marques, S. Mwaiko, C. E. Wagner, L. Excoffier *et al.*, 2017 Ancient
832 hybridization fuels rapid cichlid fish adaptive radiations. Nat Comm 8**:** 14363.
833 Mettler, L. E., R. A. Voelker and T. Mukai, 1977 Inversion clines in populations of Drosophila
834 melanogaster. Genetics 87**:** 169-176.
835 Modrich, P., and R. Lahue, 1996 Mismatch repair in replication fidelity, genetic recombination,
836 and cancer biology. Ann Rev Biochem 65**:** 101-133.
837 Nachman, M. W., 2002 Variation in recombination rate across the genome: evidence and
838 implications. Curr Opinion Genet Devel 12**:** 657-663.
839 Nei, M., 1987 *Molecular evolutionary genetics*. Columbia university press.
840 Nelson, T. C., and W. A. Cresko, 2018 Ancient genomic variation underlies repeated ecological
841 adaptation in young stickleback populations. Evol Lett 2**:** 9-21.

842 Nielsen, R., S. Williamson, Y. Kim, M. J. Hubisz, A. G. Clark *et al.*, 2005 Genomic scans for
843 selective sweeps using SNP data. Genome Res 15**:** 1566-1575.
844 Nosil, P., and B. J. Crespi, 2006 Experimental evidence that predation promotes divergence in
845 adaptive radiation. Proc Natl Acad Sci U S A 103**:** 9090-9095.
846 Opperman, R., E. Emmanuel and A. A. Levy, 2004 The effect of sequence divergence on
847 recombination between direct repeats in Arabidopsis. Genetics 168**:** 2207.
848 Paradis, E., J. Claude and K. Strimmer, 2004 APE: Analyses of phylogenetics and evolution in
849 R language. Bioinformat 20**:** 289-290.
850 Pardo-Diaz, C., C. Salazar, S. W. Baxter, C. Merot, W. Figueiredo-Ready *et al.*, 2012 Adaptive
851 introgression across species boundaries in Heliconius butterflies. PLoS Genet 8.
852 Pfeifer, B., U. Wittelsburger, S. E. Ramos-Onsins and M. J. Lercher, 2014 PopGenome: an
853 efficient Swiss army knife for population genomic analyses in R. Mol Biol Evol 31**:** 1929-
854 1936.
855 Popescu, A. A., K. T. Huber and E. Paradis, 2012 ape 3.0: New tools for distance-based
856 phylogenetics and evolutionary analysis in R. Bioinformat 28**:** 1536-1537.
857 Rastas, P., F. C. Calboli, B. Guo, T. Shikano and J. Merila, 2015 Construction of ultradense
858 linkage maps with Lep-MAP2: stickleback F2 recombinant crosses as an example.
859 Genome Biol Evol 8**:** 78-93.
860 Reimchen, T. E., 1994 Predators and morphological evolution in threespine stickleback, pp.
861 240-276 in *The Evolutionary Biology of the Threespine Stickleback*, edited by M. A. Bell
862 and S. A. Foster. Oxford University Press, New York.
863 Reimchen, T. E., C. Bergstrom and P. Nosil, 2013 Natural selection and the adaptive radiation
864 of Haida Gwaii stickleback. Evol Ecol Res 15**:** 241-269.
865 Roesti, M., S. Gavrilets, A. P. Hendry, W. Salzburger and D. Berner, 2014 The genomic
866 signature of parallel adaptation from shared genetic variation. Mol Ecol 23**:** 3944-3956.
867 Roesti, M., B. Kueng, D. Moser and D. Berner, 2015 The genomics of ecological vicariance in
868 threespine stickleback fish. Nat Comm 6**:** 8767.
869 Roesti, M., D. Moser and D. Berner, 2013 Recombination in the threespine stickleback genome
870 - Patterns and consequences. Mol Ecol 22**:** 3014-3027.
871 Samuk, K., G. L. Owens, K. E. Delmore, S. E. Miller, D. J. Rennison *et al.*, 2017 Gene flow and
872 selection interact to promote adaptive divergence in regions of low recombination. Mol
873 Ecol 26**:** 4378-4390.
874 Sardell, J. M., C. D. Cheng, A. J. Dagilis, A. Ishikawa, J. Kitano *et al.*, 2018 Sex Differences in
875 Recombination in Sticklebacks. G3-Genes Genomes Genetics 8**:** 1971-1983.
876 Scheet, P., and M. Stephens, 2006 A fast and flexible statistical model for large-scale
877 population genotype data: applications to inferring missing genotypes and haplotypic
878 phase. Am J Hum Genet 78**:** 629-644.
879 Schluter, D., and G. L. Conte, 2009 Genetics and ecological speciation. Proc Natl Acad Sci U S
880 A 106**:** 9955-9962.
881 Schrider, D. R., and A. D. Kern, 2017 Soft sweeps are the dominant mode of adaptation in the
882 human genome. Mol Biol Evol 34**:** 1863-1877.
883 Slatkin, M., 1985 Gene flow in natural populations. Ann Rev Ecol Systemat 16**:** 393-430.
884 Slatkin, M., 1991 Inbreeding coefficients and coalescence times. Genet Res 58**:** 167-175.
885 Stankowski, S., M. A. Chase, A. M. Fuiten, P. L. Ralph and M. A. Streisfeld, 2018 The tempo of
886 linked selection: Rapid emergence of a heterogeneous genomic landscape during a
887 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 R 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. American Journal of Human Genetics 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.

Tavares, H., A. Whibley, D. L. Field, D. Bradley, M. Couchman *et al.*, 2018 Selection and gene flow shape genomic islands that control floral guides. Proc Natl Acad Sci U S A 115**:** 11006.

Team, R. C., 2016 R: A language and environment for statistical computing, pp. 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**:** e1004696.

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.

Turner, T. L., M. W. Hahn and S. V. Nuzhdin, 2005 Genomic islands of speciation in Anopheles gambiae. PLoS Biol 3**:** 1572-1578.

Van Belleghem, S. M., C. Vangestel, K. De Wolf, Z. De Corte, M. Möst *et al.*, 2018 Evolution at two time frames: Polymorphisms from an ancient singular divergence event fuel contemporary parallel evolution. PLoS Genet 14**:** e1007796.

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.

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

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

## Supplementary Tables and Figures

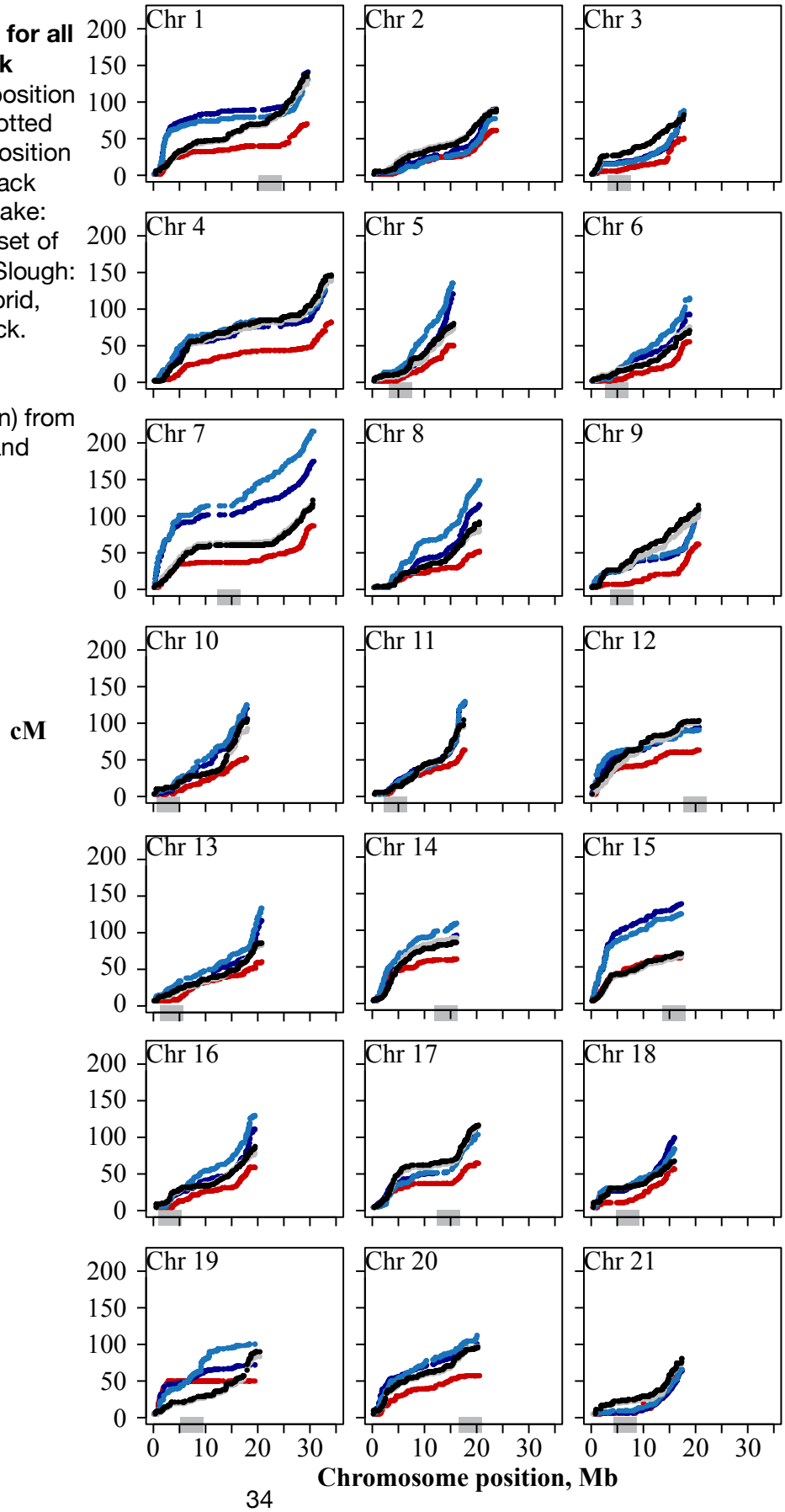
**Table S1. Sequencing statistics of libraries for mapping families in the main text.** *Coverage is the mean sequencing depth per RAD marker across the 94 progeny with highest coverage. These progeny were used to estimate recombination rates in the main text. †Missing data calculated as the average percent of RAD markers per progeny with zero coverage.

| Family | reads: raw | filtered | aligned | mean coverage* | % missing† |
|---|---|---|---|---|---|
| Boot Lake | 322871980 | 229263868 | 196451426 | 35.00 | 1.90 |
| Rabbit Slough | 188240315 | 110264451 | 95689374 | 24.38 | 4.71 |
| $F_1$ hybrid | 229219526 | 181462508 | 78349244 | 22.52 | 8.55 |
| *Total* | *740331821* | *520990827* | *370490044* | *27.30* | *5.05* |

**Table S2. Genome-wide correlations between recombination rate and genetic diversity and divergence.** Statistics were calculated as means of 1000-kb non-overlapping bins. $\pi_{RS}$: sequence diversity in the marine RS population. $\pi_{FW}$: sequence diversity in the combined freshwater populations. Correlation coefficients are given as Spearman's ρ. Characters in parentheses indicate significant correlations at $p < 0.01$: *linear model; ρ = Spearman's ρ, τ = Kendall's τ.

| map | $\pi_{RS}$ | $\pi_{FW}$ | $\pi$ | $d_{XY}$ | $F_{ST}$ |
|---|---|---|---|---|---|
| **BT** | 0.450 (*,ρ,τ) | 0.349 (*,ρ,τ) | 0.261 (*,ρ,τ) | 0.154 (ρ,τ) | -0.431 (*,ρ,τ) |
| **RS** | 0.489 (*,ρ,τ) | 0.223 (*,ρ,τ) | 0.225 (*,ρ,τ) | 0.146 (ρ,τ) | -0.365 (ρ,τ) |
| **$F_1$** | 0.483 (*,ρ,τ) | 0.243 (*,ρ,τ) | 0.146 (ρ,τ) | 0.053 | -0.406 (*,ρ,τ) |

**Figure S4. 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. Boot Lake: dark blue; Boot Lake, subset of 94 progeny: blue; Rabbit Slough: red; $F_1$ hybrid: gray; $F_1$ hybrid, subset of 94 progeny: black. Gray boxes indicate the approximate positions of centromeres (where known) from Cech and Peichel (2015) and Sardell et al (2018).



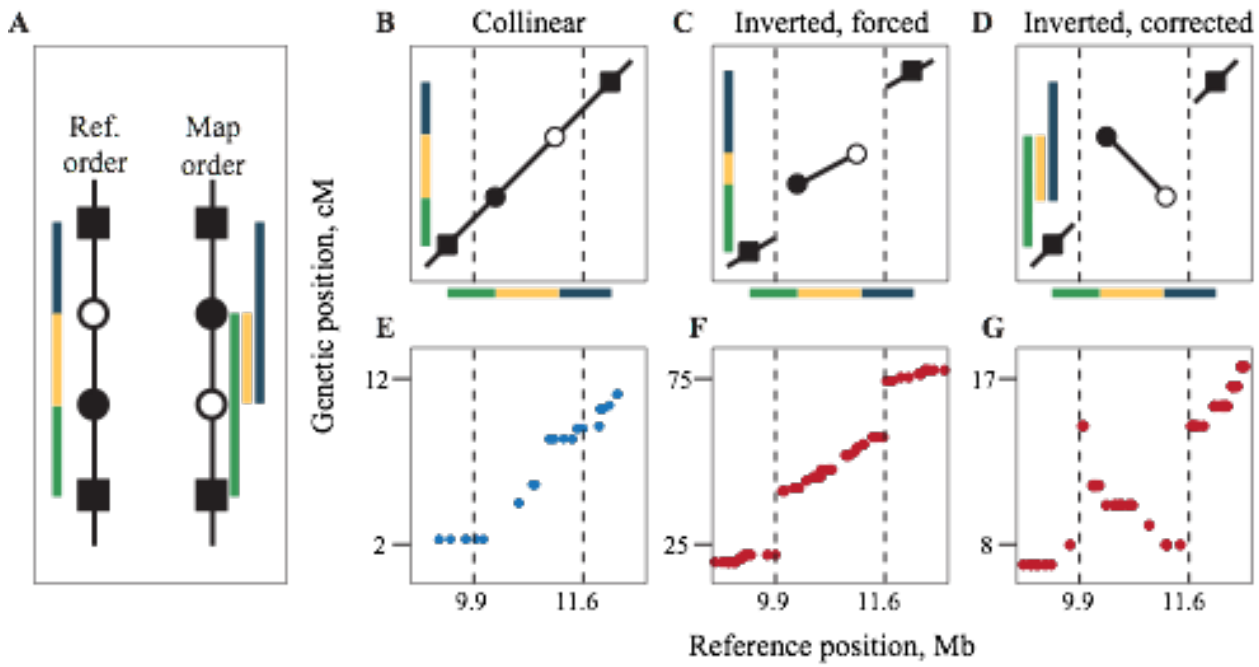



**Figure S2. Reference-guided genetic map construction allows identification and correction of chromosomal inversions.** A: Inversion polymorphisms create discrepancies between the reference physical genome and a genetic map. Colored bars indicate true map distances between genetic markers inside and on either side of the inversion. B: When the physical and genetic maps are collinear, 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.

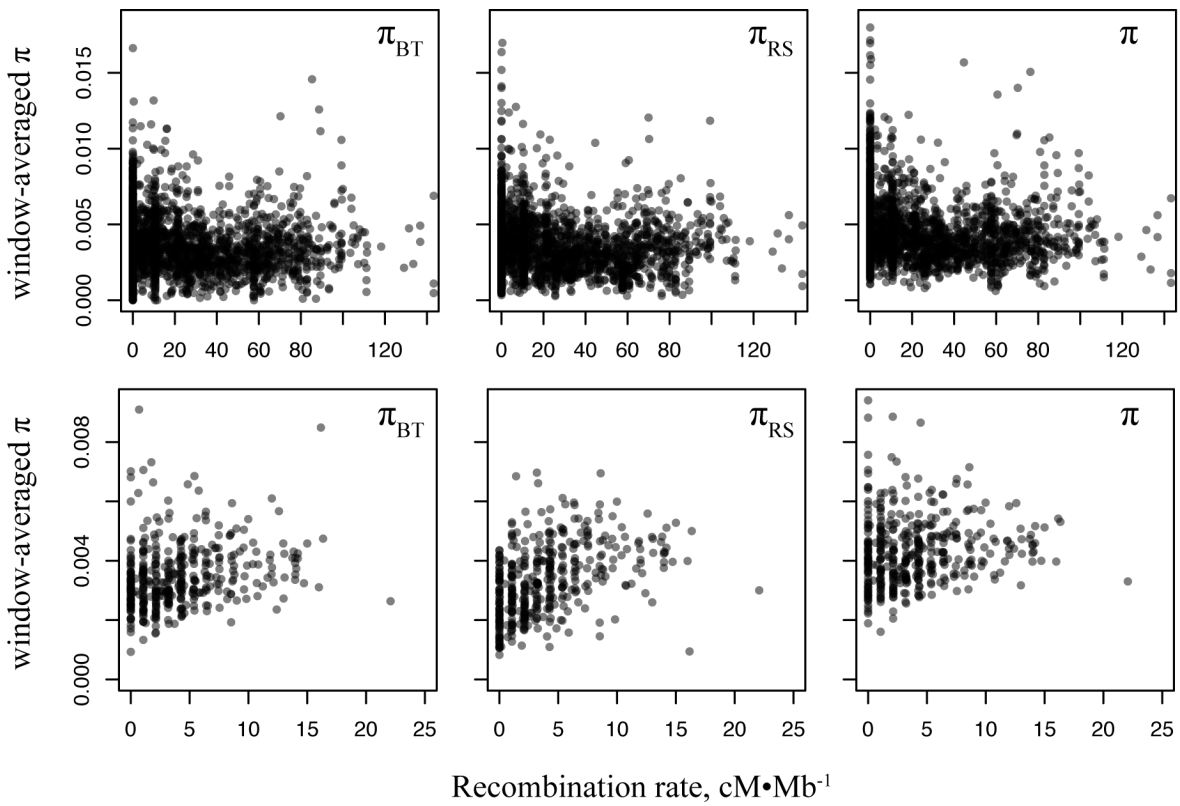



**Figure S3. Window-based estimates of recombination rate (cM/Mb) and sequence diversity using (top) 100-kb non-overlapping windows and (bottom) 1-Mb non-overlapping windows.** Note the difference in axis scale with different window sizes.

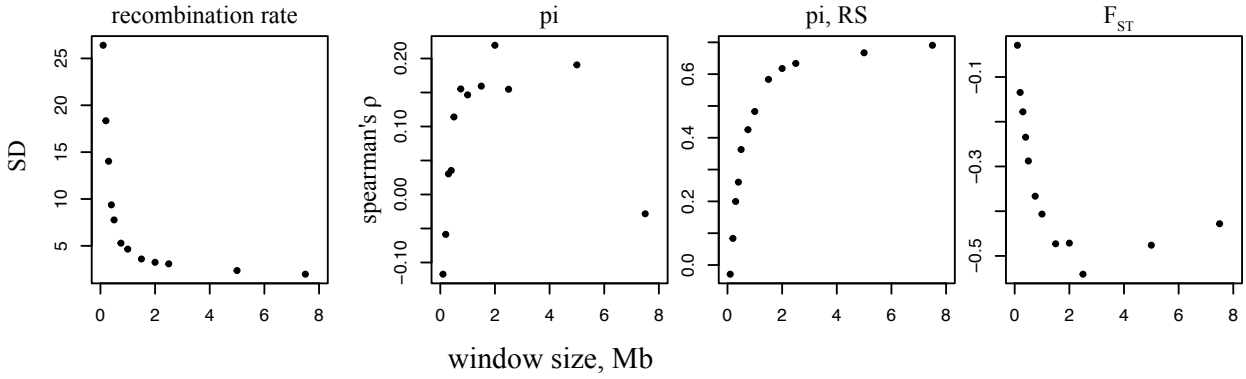



**Figure S4. Variation in recombination rate and covariation with population genetic statistics at different genomic window sizes.** All points are genome-wide summaries of recombination rate (cM/Mb) and correlation (as Spearman's ρ) with population genetic variation and differentiation at a given window size using non-overlapping genomic windows. SD: standard deviation; π: sequence diversity; $F_{ST}$: differentiation between Rabbit Slough and Boot Lake populations.

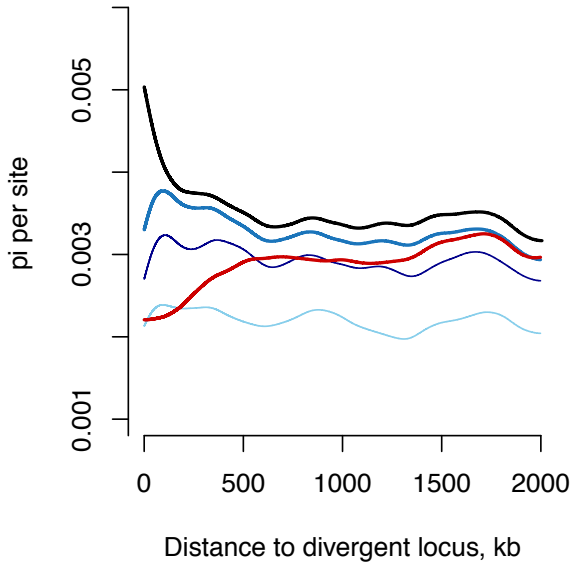



**Figure S5. Sequence diversity as a function of distance from divergent loci in all populations.** All genomes: black; Rabbit Slough: red; combined freshwater: middle blue; Boot Lake: dark blue; Bear Paw Lake: light blue.

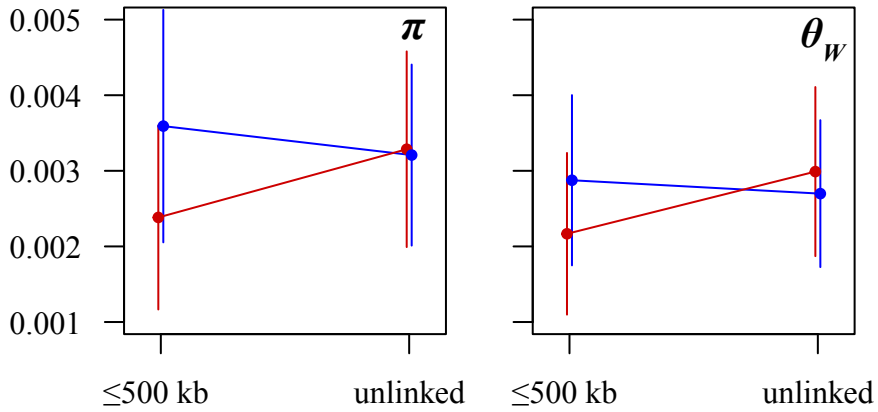



**Figure S6. Interactions between genetic diversity and distance from divergent loci for both π and $\theta_W$.** Rabbit Slough chromosomes: red; combined freshwater chromosomes: blue. All comparisons are significant (Tukey's HSD: $p \leq 0.01$) except Rabbit Slough and freshwater π at unlinked loci.

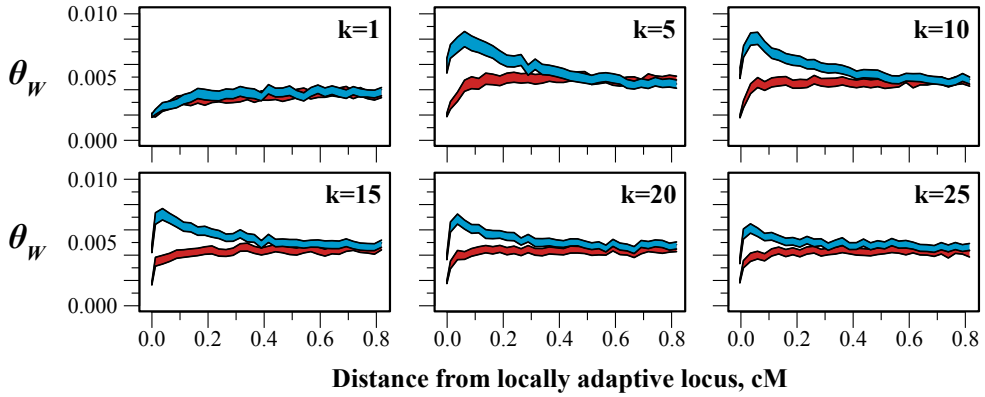



**Figure S7. Effects of population subdivision linked genetic variation.** All simulations were run as in the main text, Figure 4: s = 0.20; m = 1 migrant/generation; $20N_H$ generations of selection.

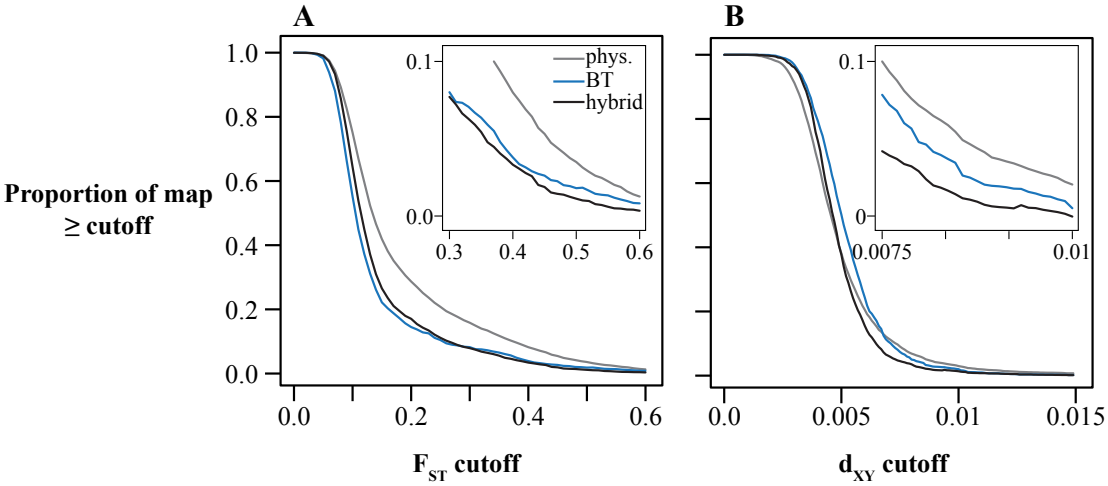



**Figure S8. Genetic divergence comprises a smaller proportion of the genetic map than the physical map.** Each line represents the proportion of the overall map lengths (Mb or cM) taken up by windows of genetic divergence ($F_{ST}$ or $d_{XY}$) greater than or equal to a given value. Gray: physical map; blue: Boot Lake genetic map; black: $F_1$ hybrid genetic map.

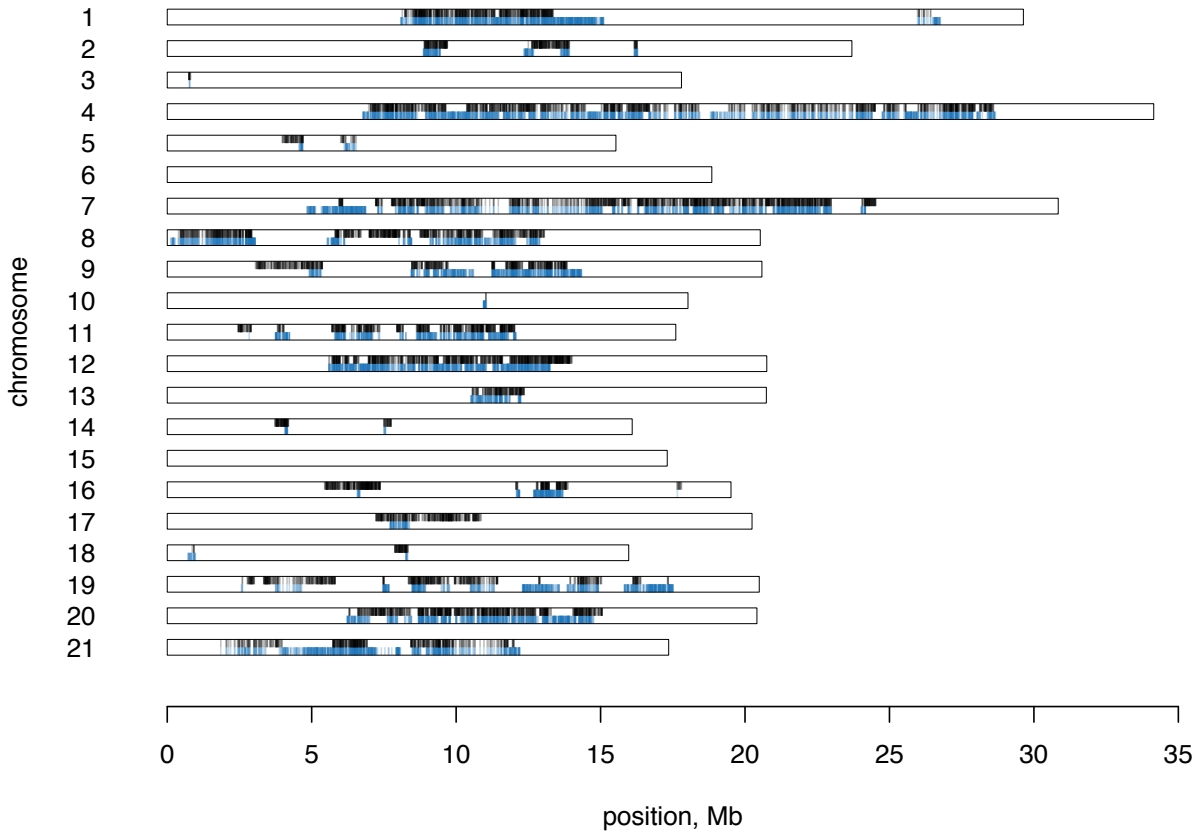



**Figure S9. Locations of all RAD loci in the population genomics dataset with imputed genetic positions within 0.2 cM of a divergent locus.** $F_1$ hybrid genetic map: black; Boot Lake genetic map: blue.

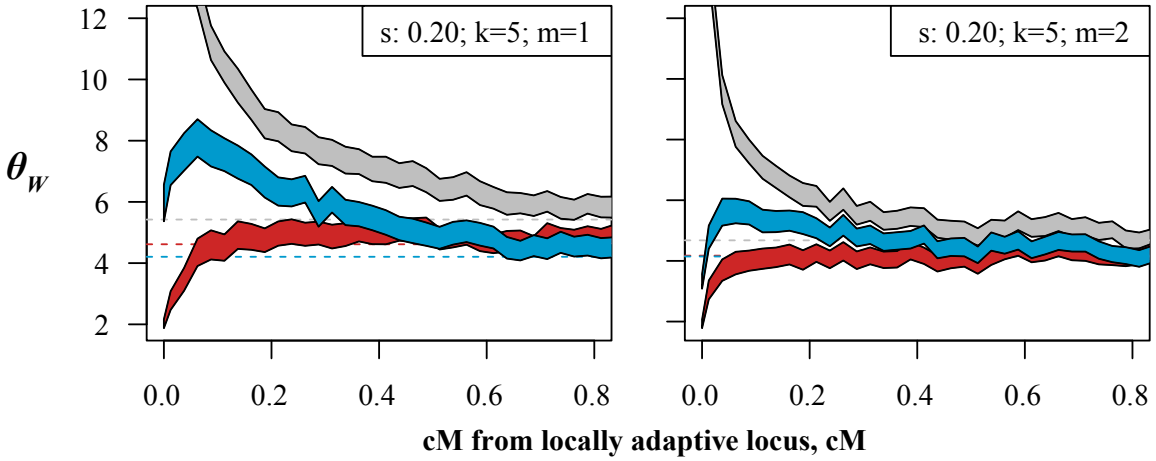
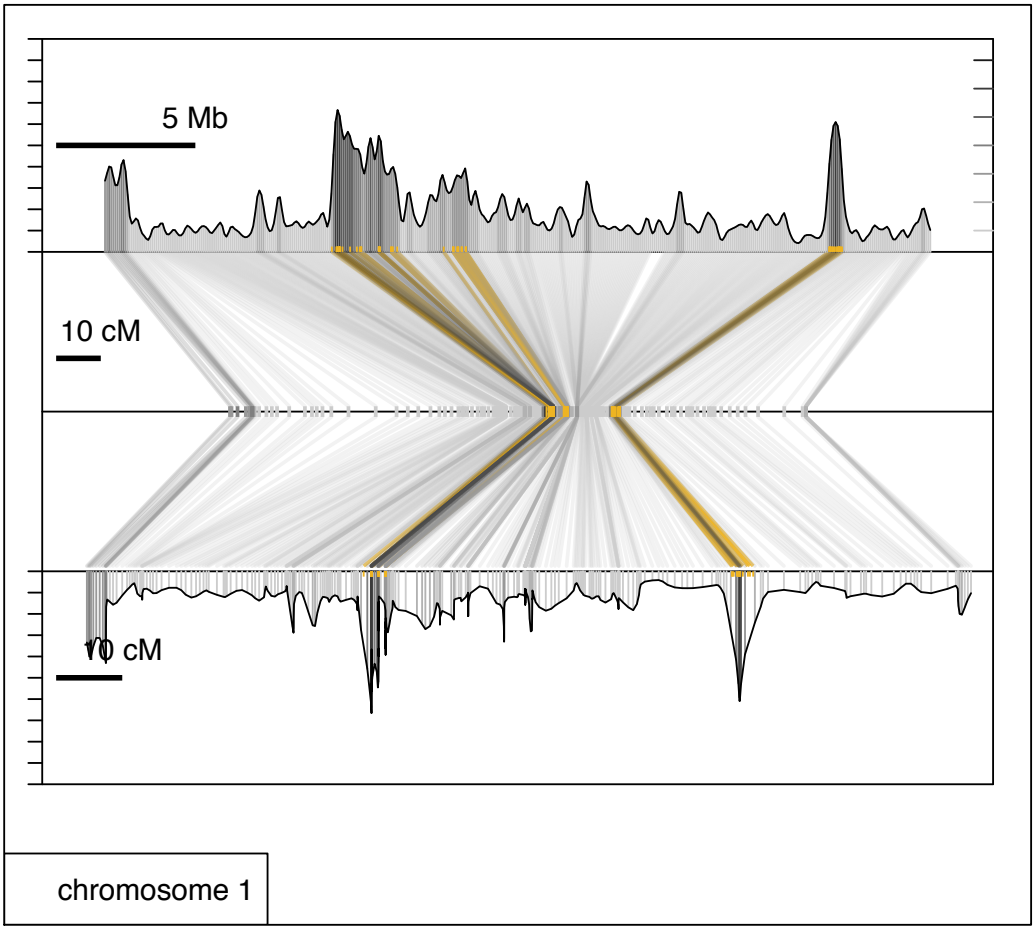
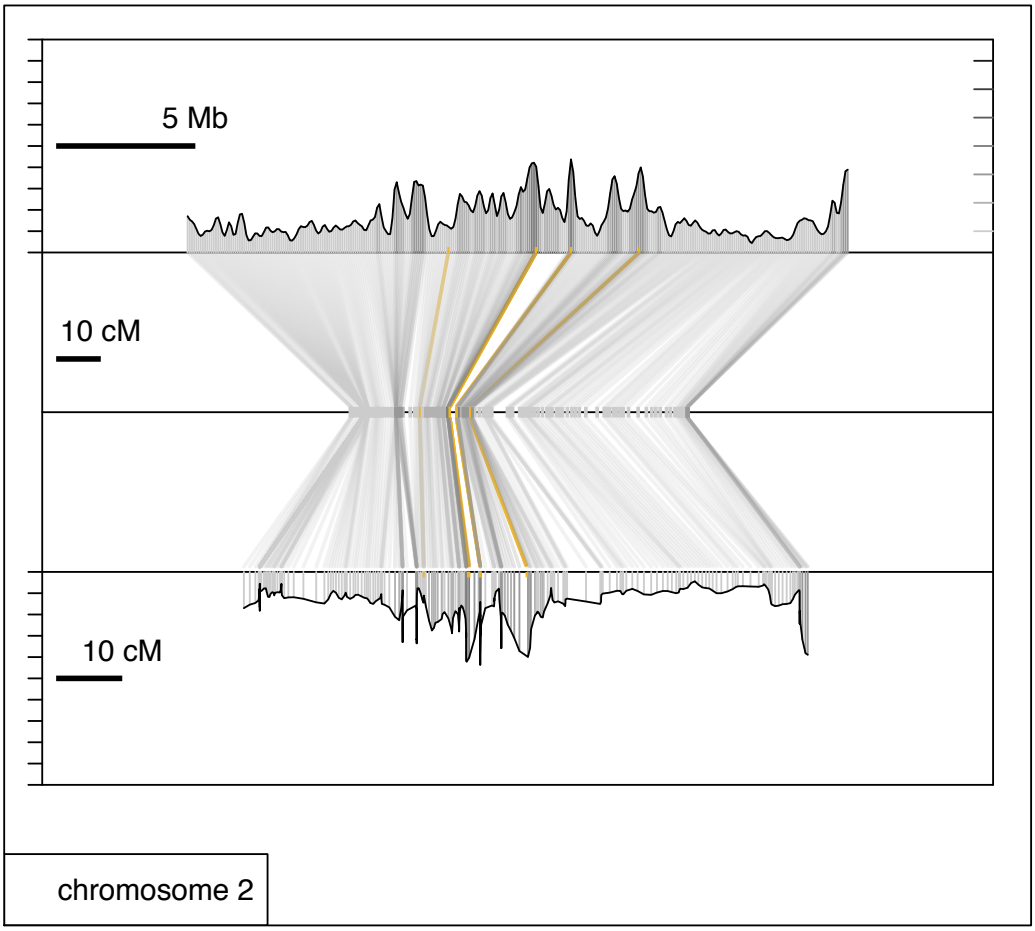


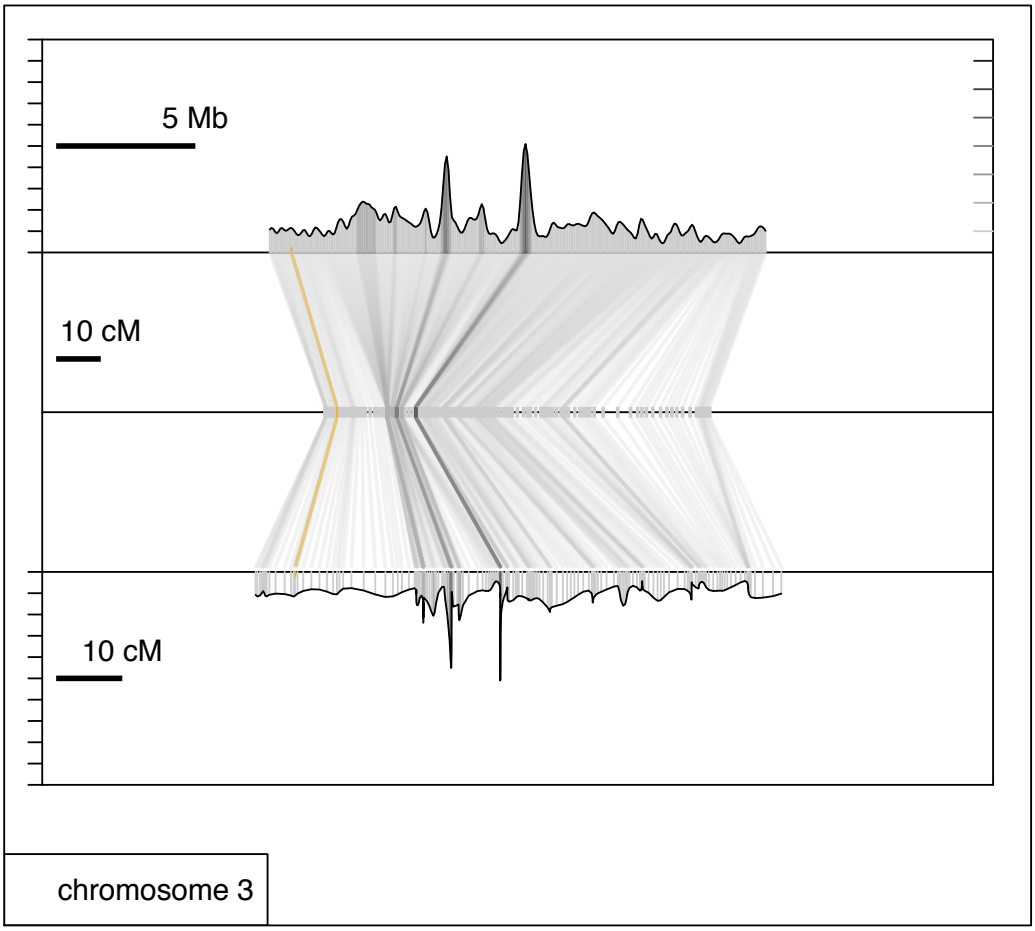

**Figure S10. Effect of migration rate on linked genetic variation.** All simulations were run for $20N_H$ generations of selection.

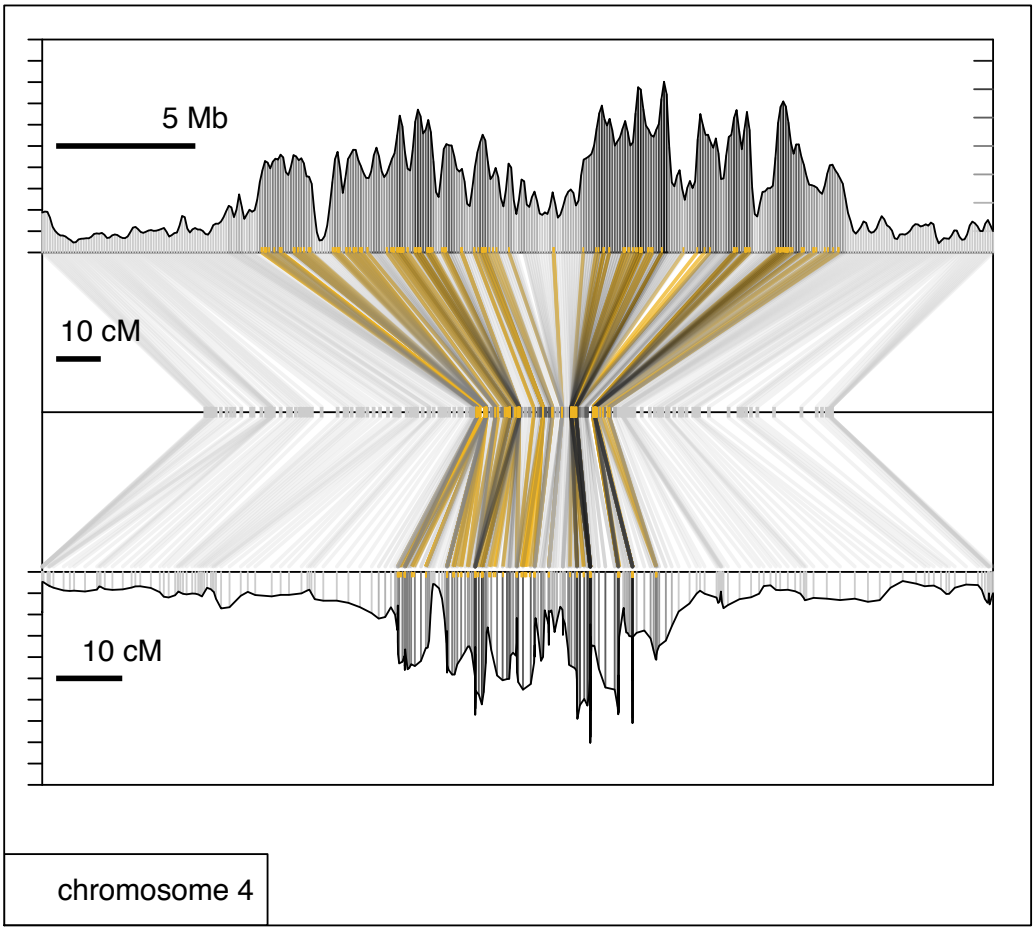
## Appendix 1. Physical-to-genetic map scans for all chromosomes

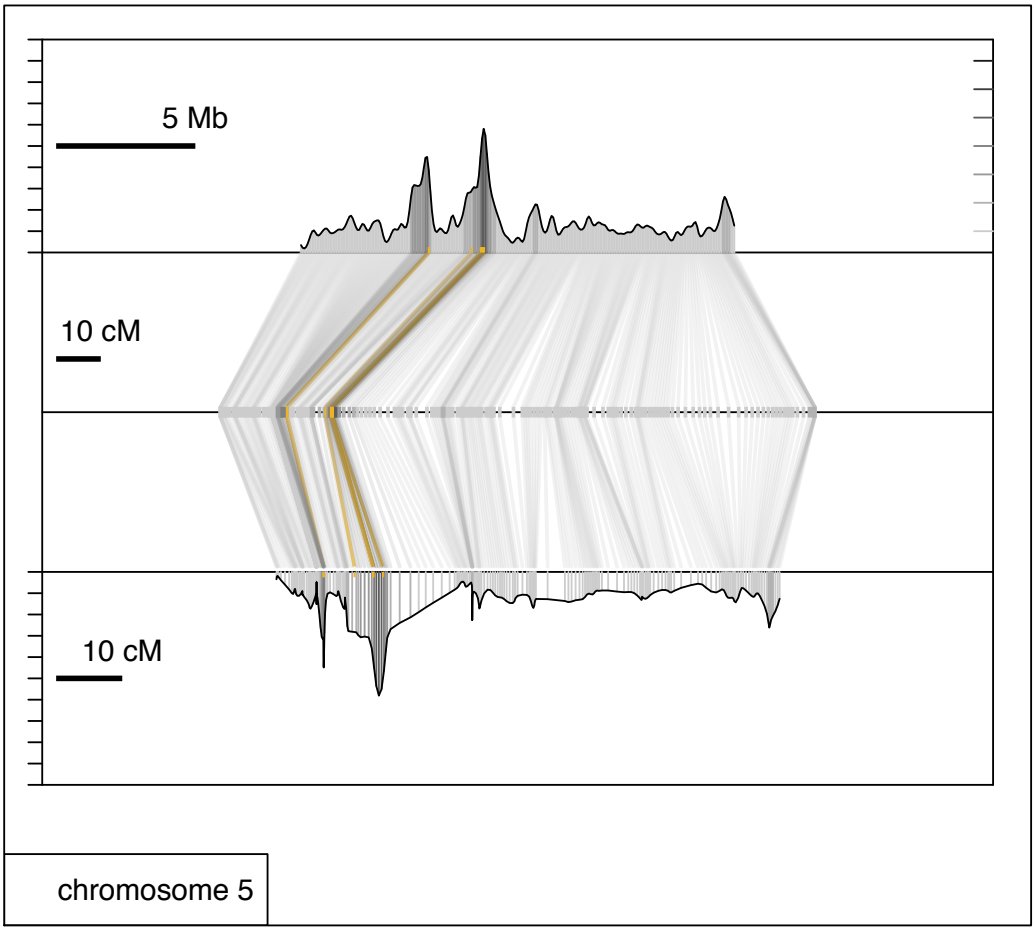
Maps plotted as per Figure 6 in the main text. Grayscale reflects window-averaged FST. Gold lines are positions of divergent loci (NELSON AND CRESKO 2018).

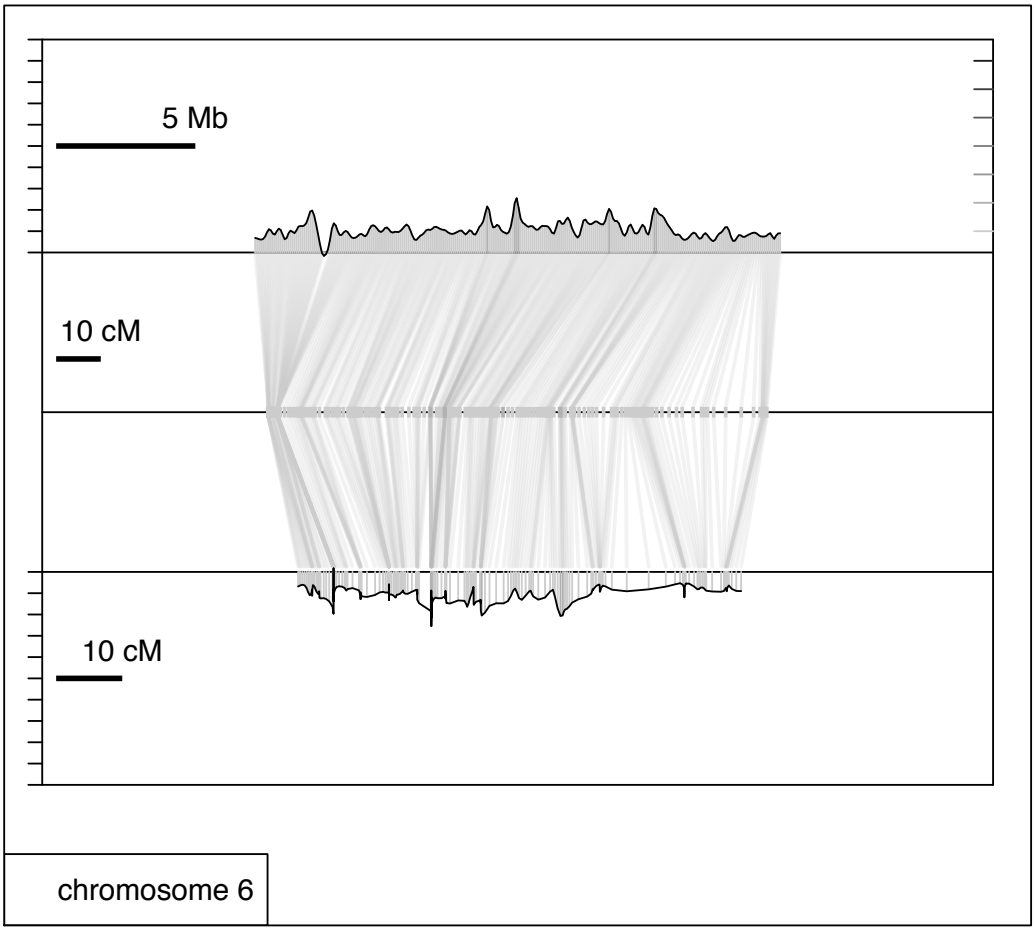


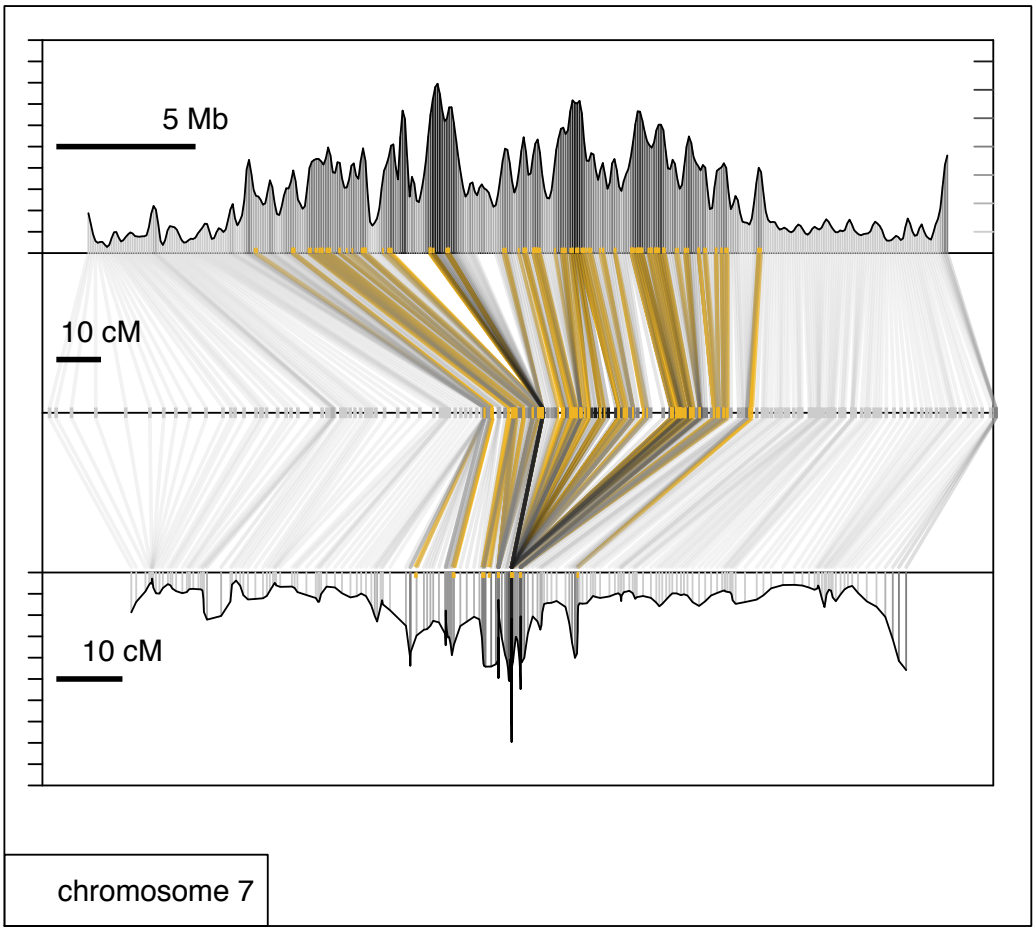

5 Mb
10 cM
10 cM
chromosome 2

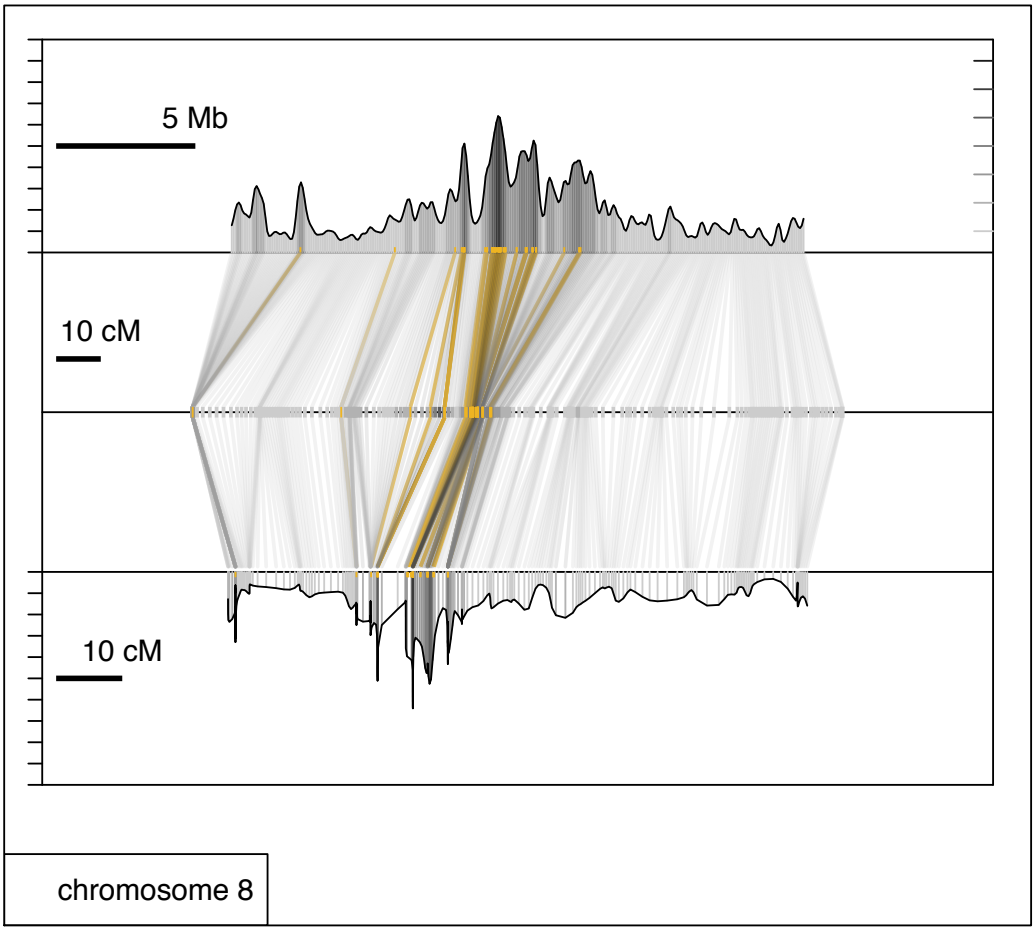

5 Mb
10 cM
10 cM
chromosome 3

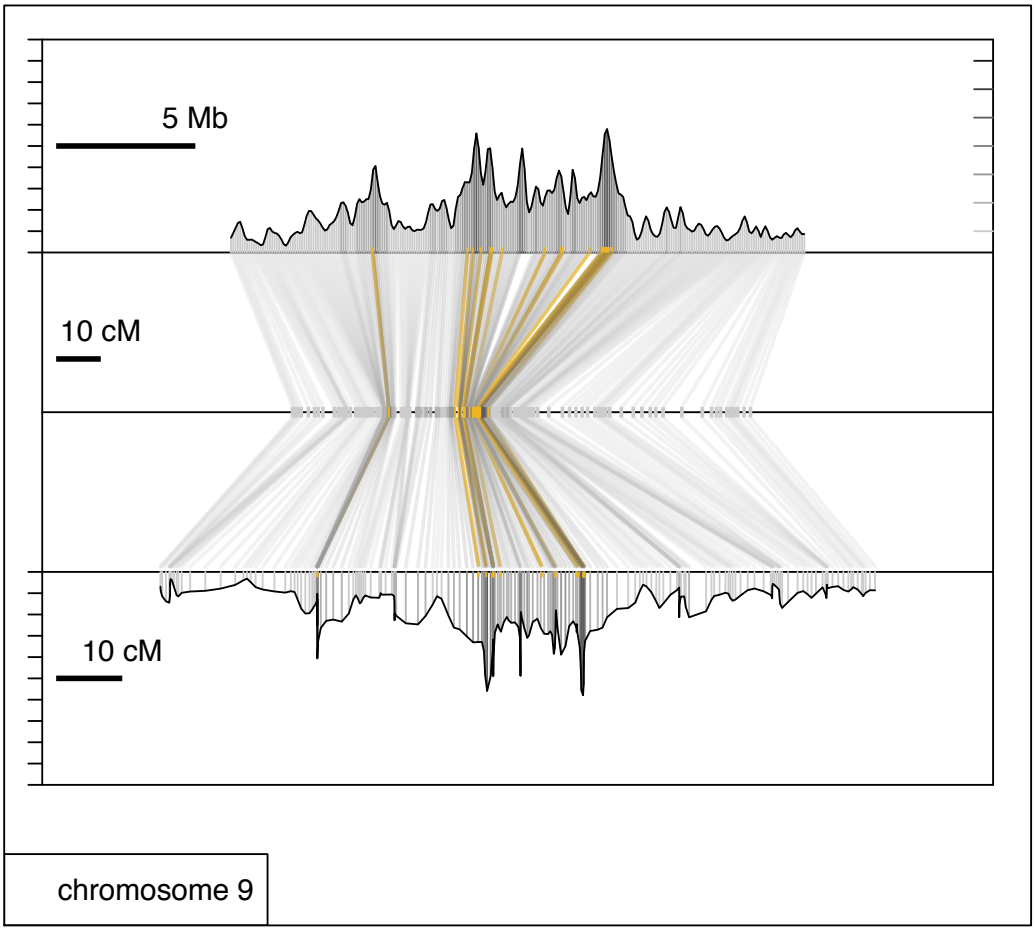

5 Mb
10 cM
10 cM
chromosome 4

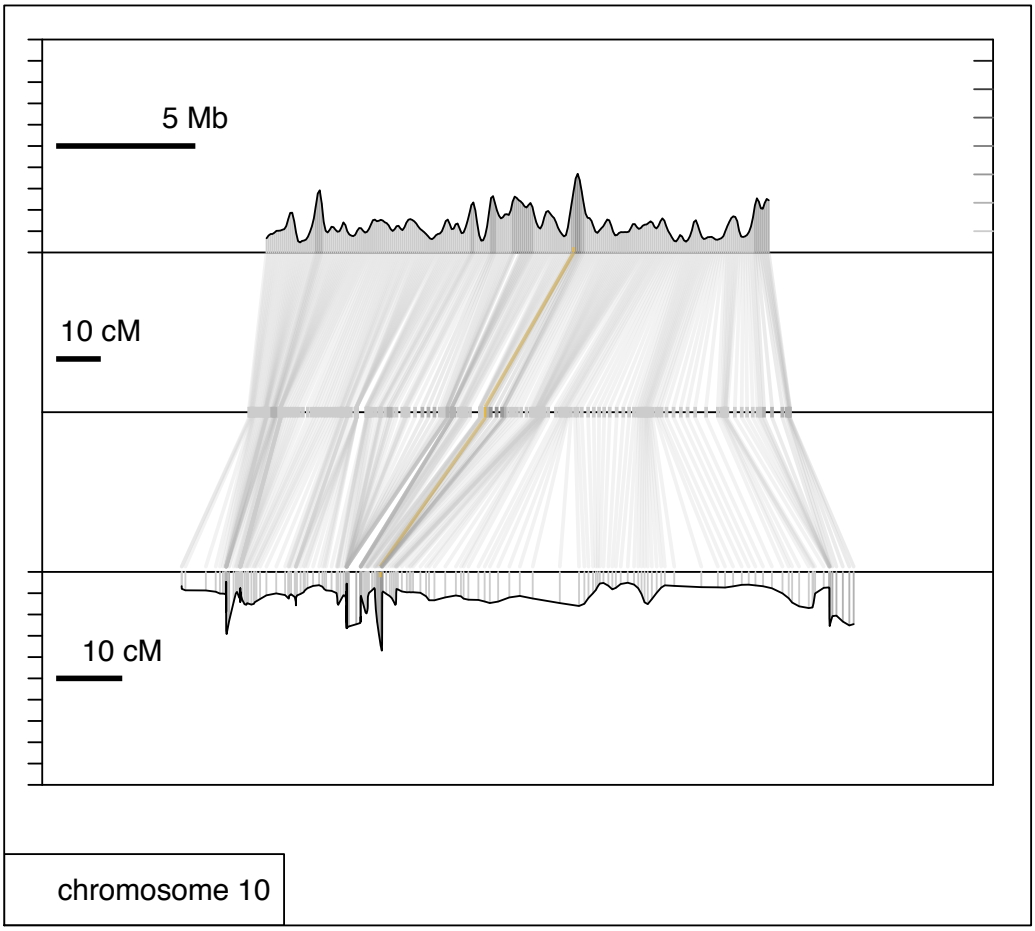

5 Mb
10 cM
10 cM
chromosome 5

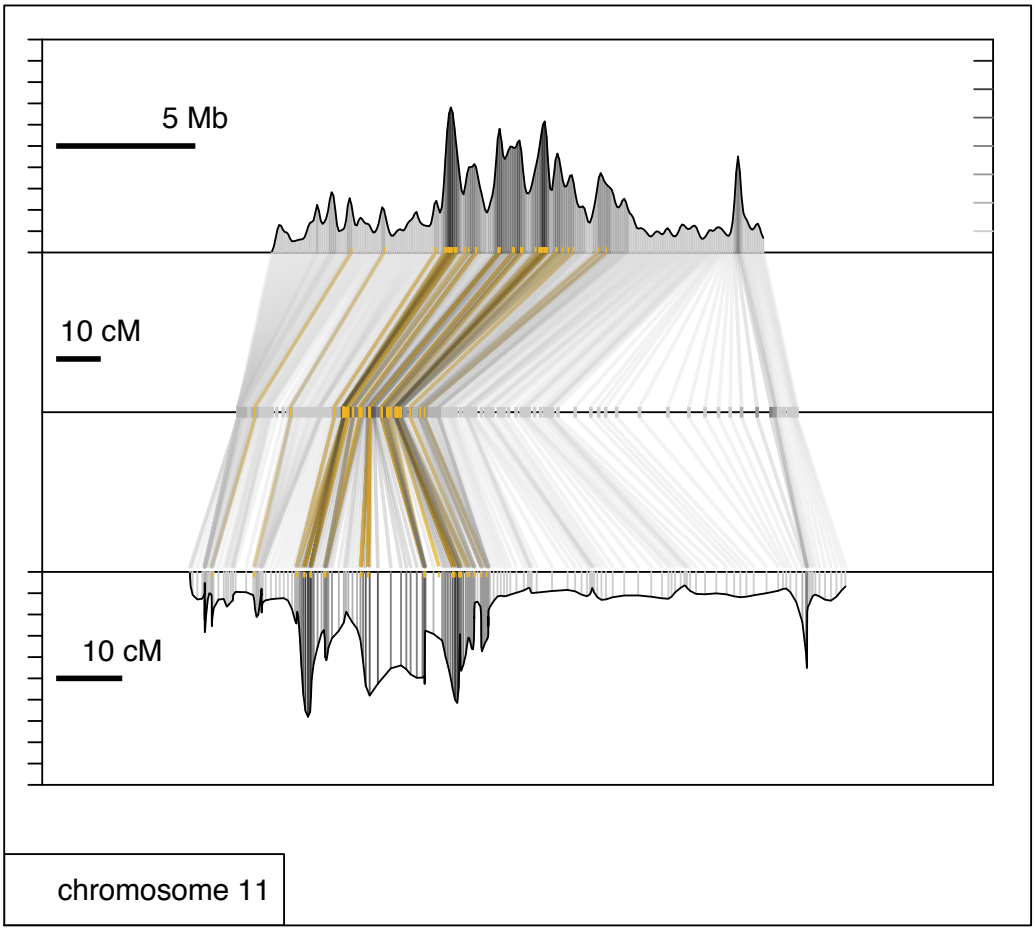

5 Mb
10 cM
10 cM
chromosome 6

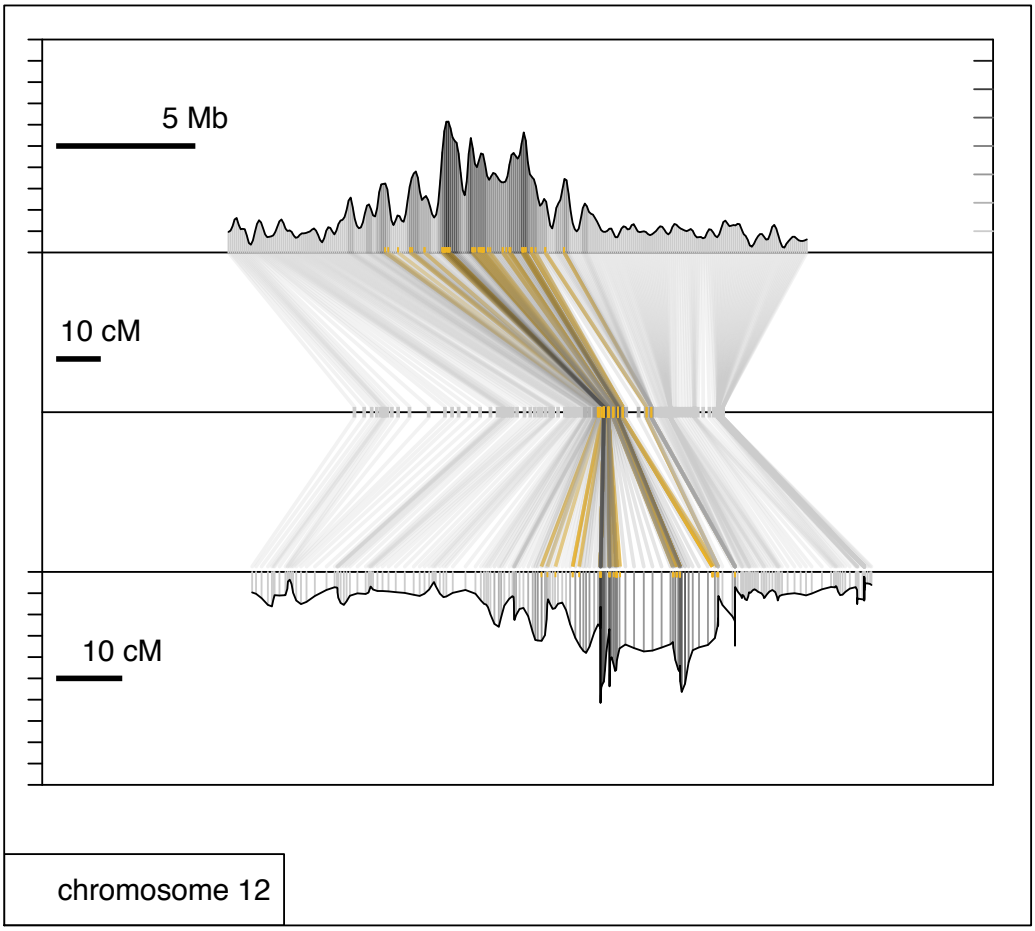

5 Mb
10 cM
10 cM
chromosome 7

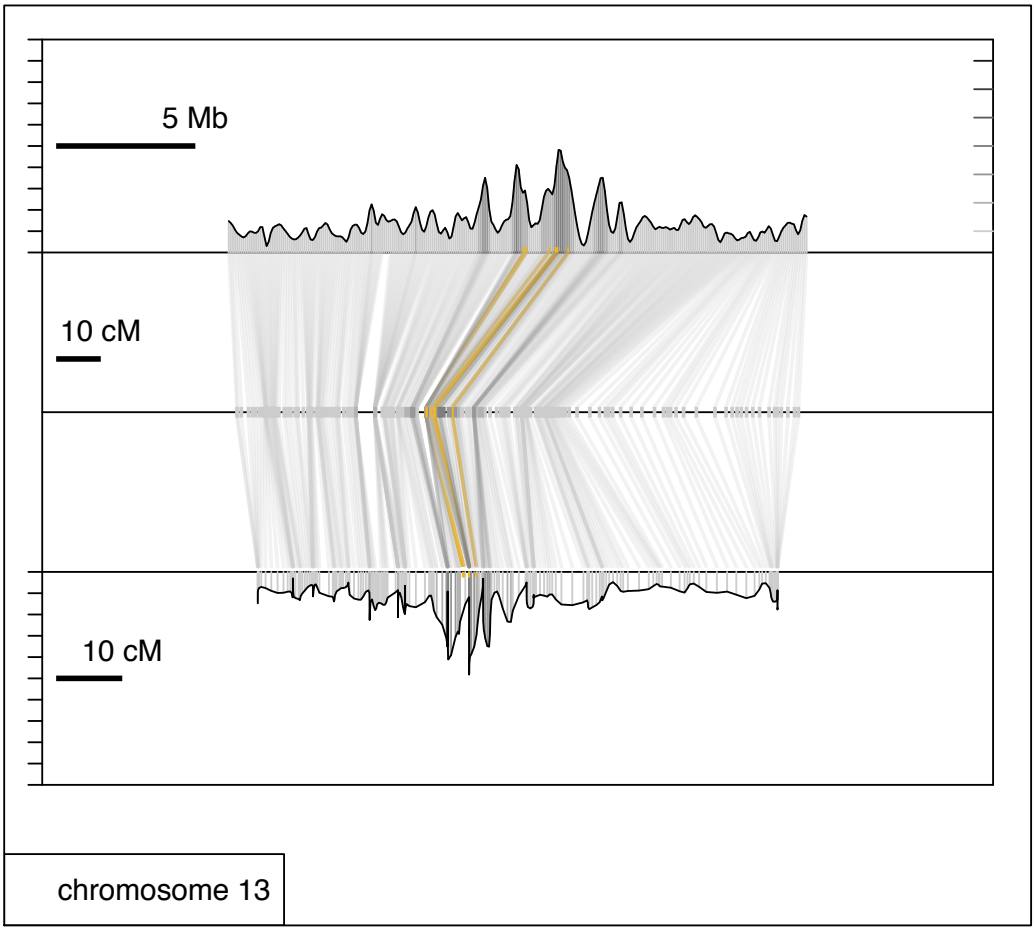

5 Mb
10 cM
10 cM
chromosome 8

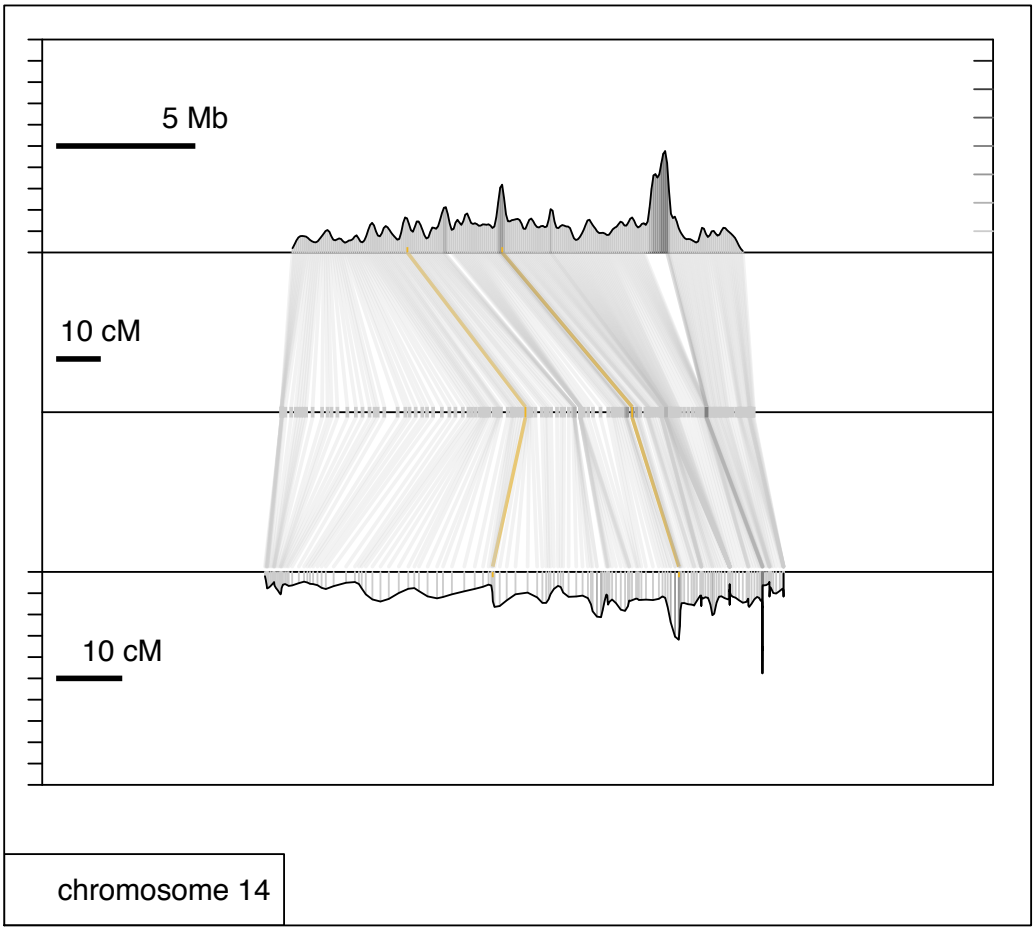

5 Mb
10 cM
10 cM
chromosome 9

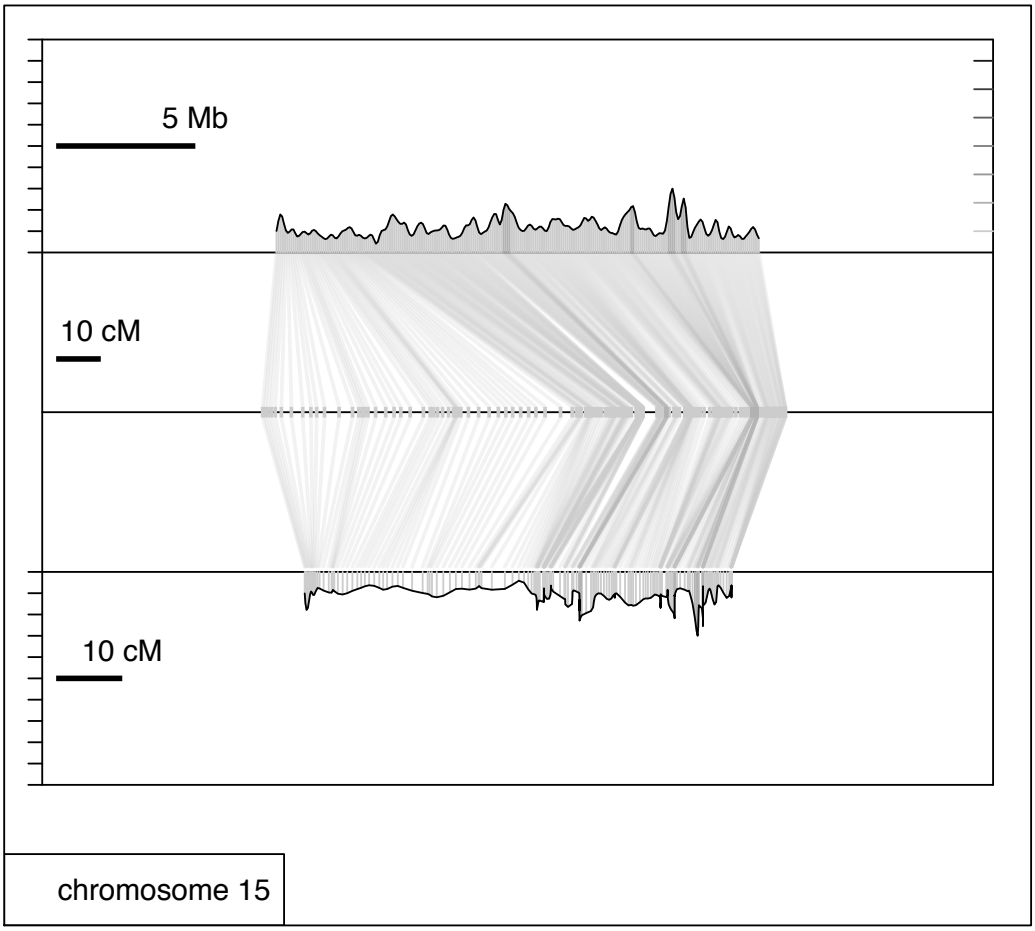

5 Mb
10 cM
10 cM
chromosome 10

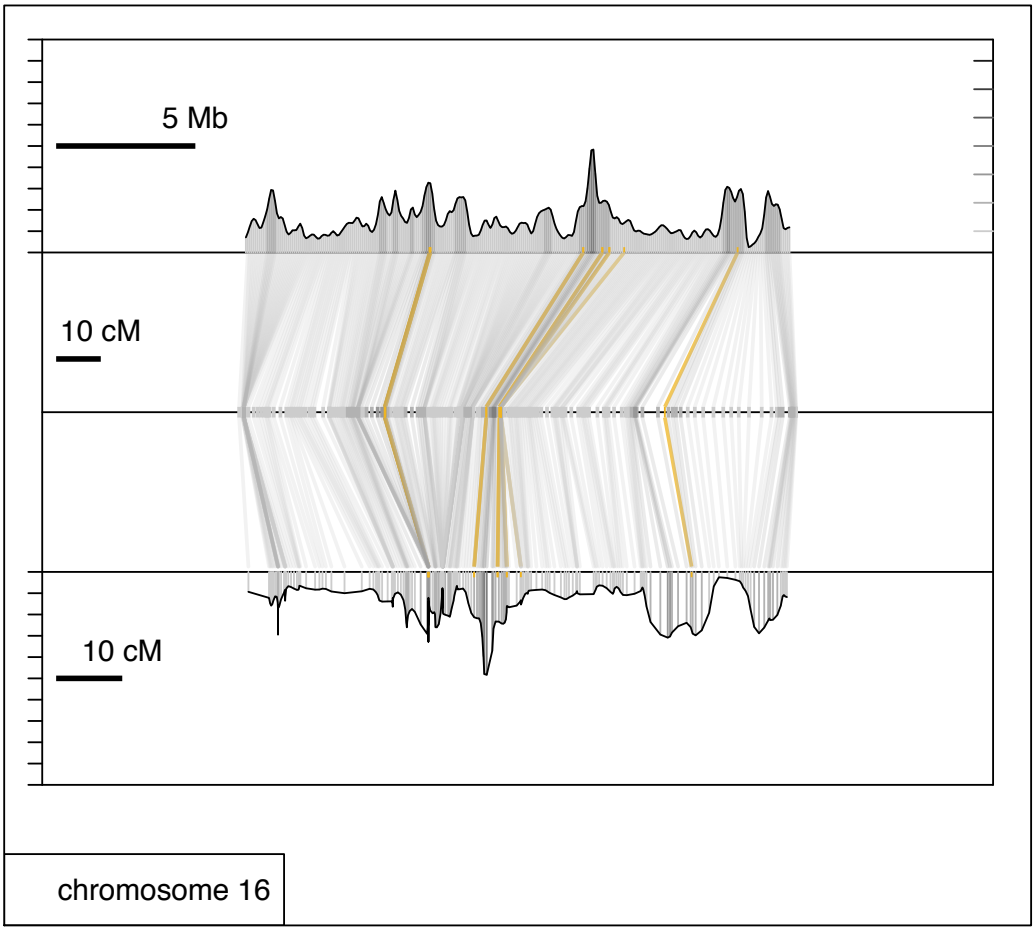

5 Mb
10 cM
10 cM
chromosome 11

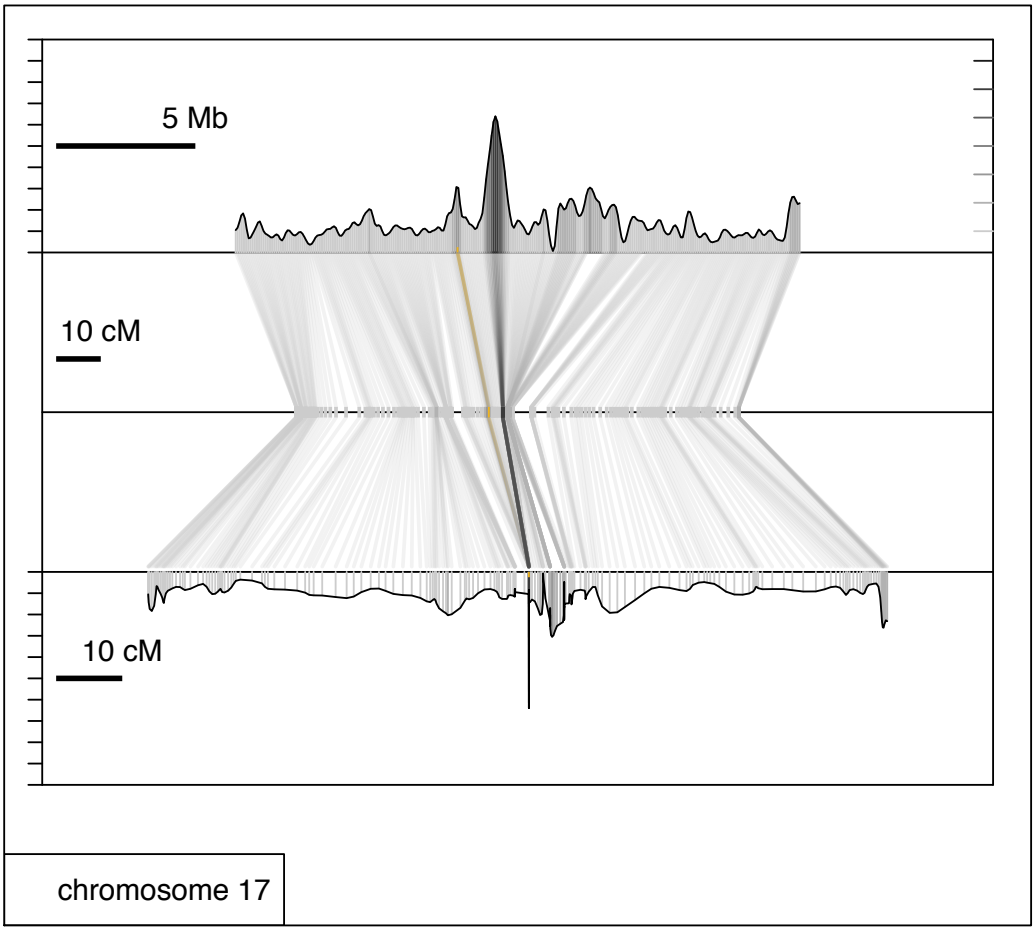

5 Mb
10 cM
10 cM
chromosome 12

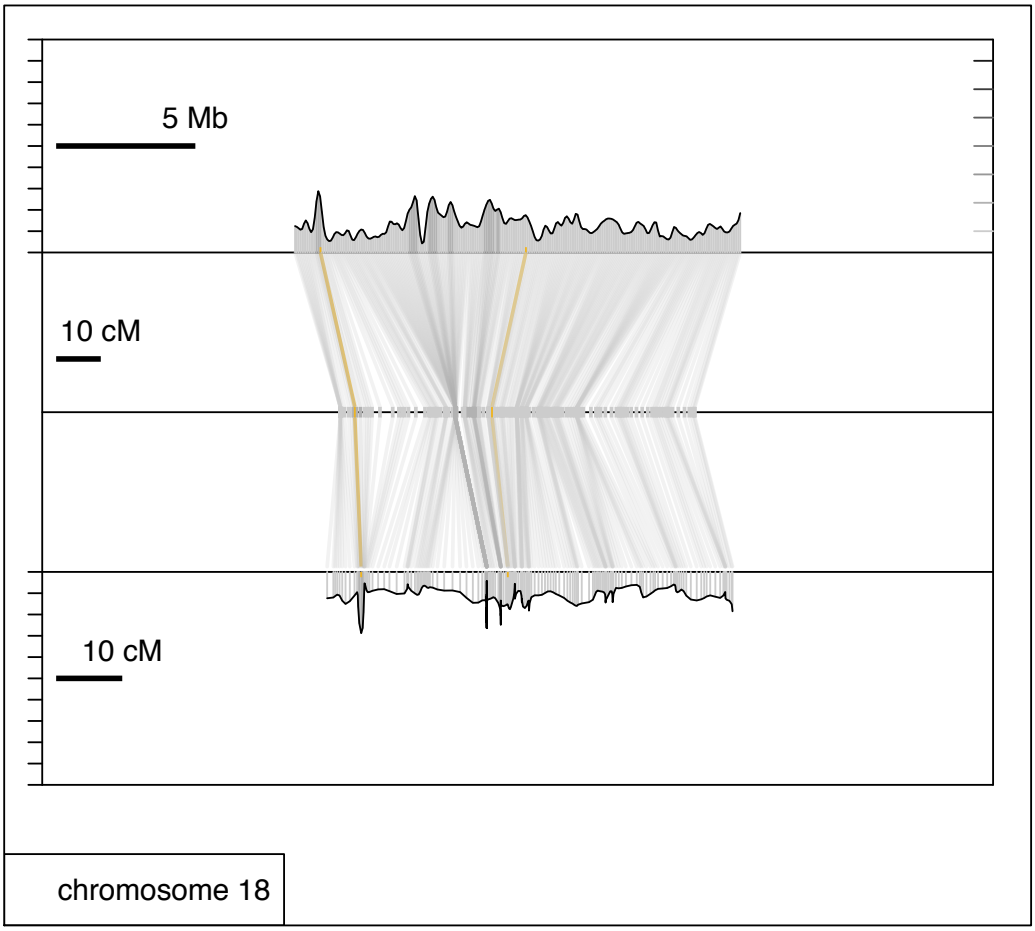

5 Mb
10 cM
10 cM
chromosome 13

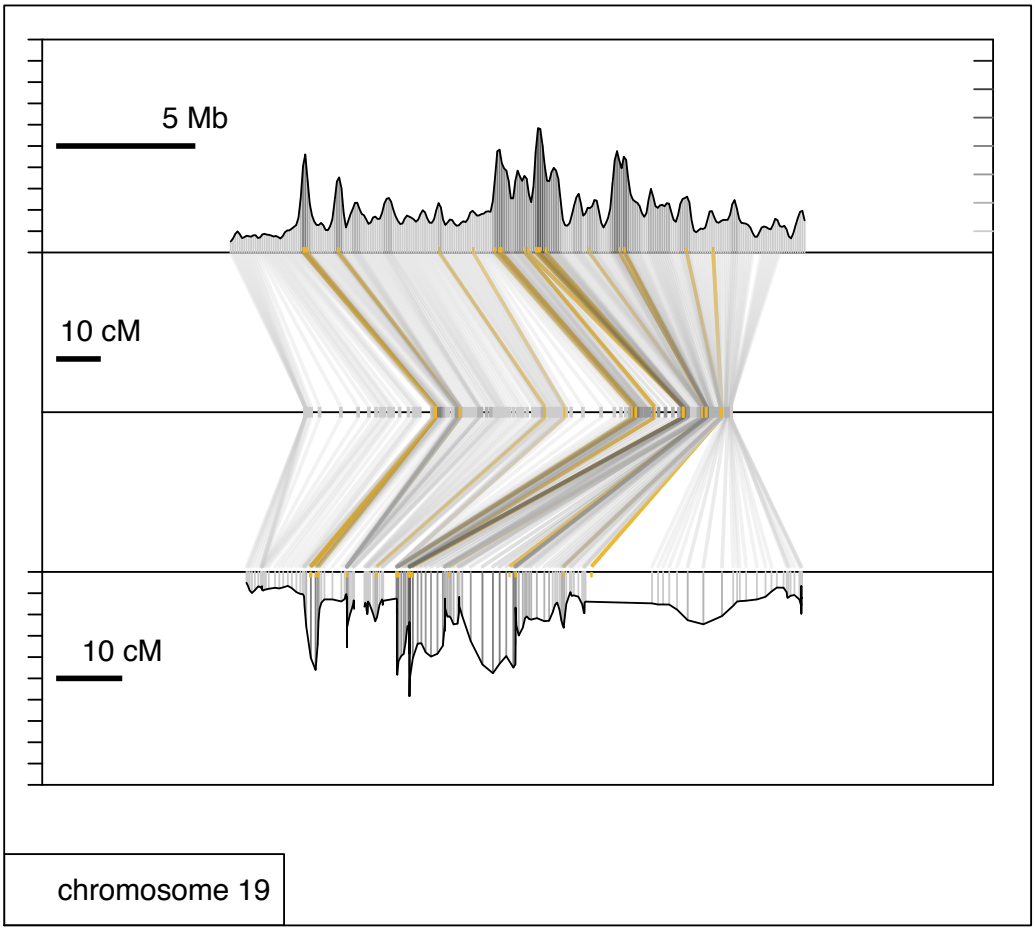

5 Mb
10 cM
10 cM
chromosome 14

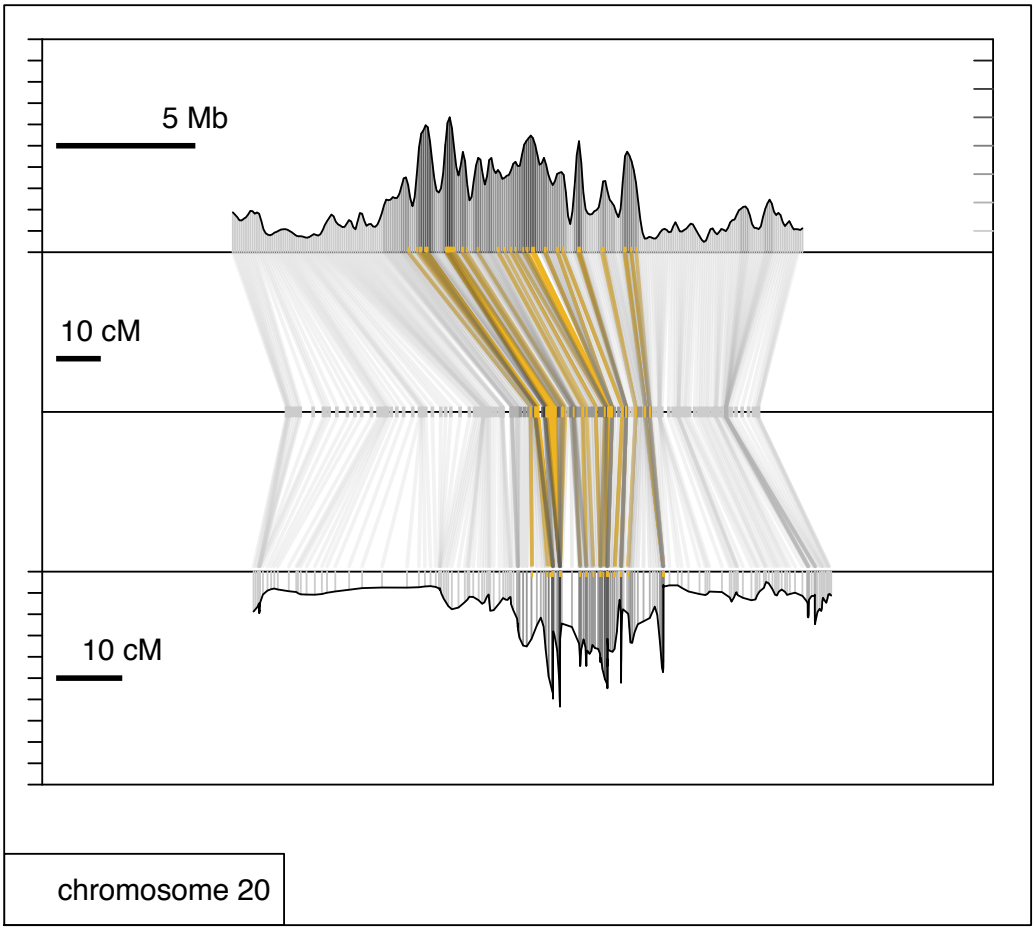

5 Mb
10 cM
10 cM
chromosome 15

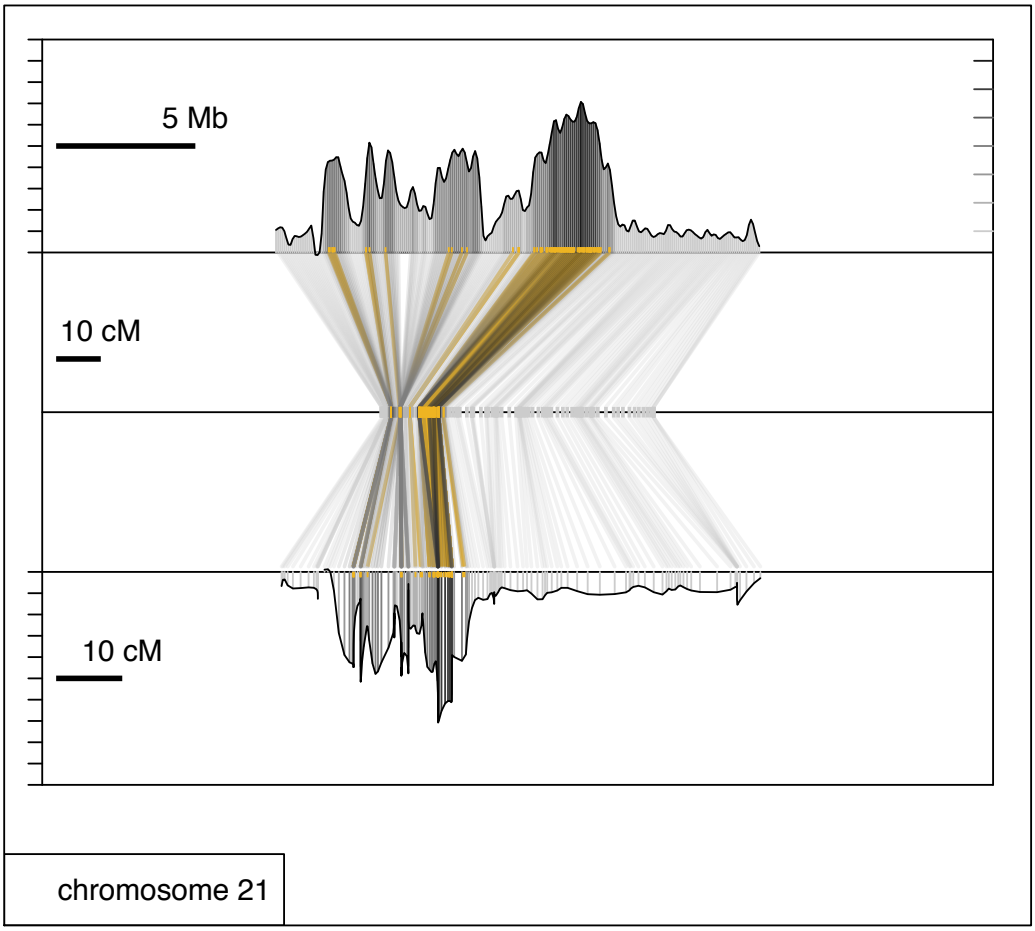

5 Mb
10 cM
10 cM
chromosome 16


5 Mb
10 cM
10 cM
chromosome 17


5 Mb
10 cM
10 cM
chromosome 18


5 Mb
10 cM
10 cM
chromosome 19


5 Mb
10 cM
10 cM
chromosome 20


5 Mb
10 cM
10 cM
chromosome 21