Research Article

Fine-scale recombination landscapes between a freshwater and marine population of threespine stickleback fish

Alice F. Shanfelter[1], Sophie L. Archambeault[2,3], Michael A. White[1*]

[1]Department of Genetics, University of Georgia, Athens, GA, 30602, USA
[2]Institute of Ecology and Evolution, University of Bern, 3012 Bern, Switzerland
[3]Graduate Program in Molecular and Cellular Biology, University of Washington, Seattle, WA, 98195, USA

*Author for Correspondence: Michael White, Department of Genetics, University of Georgia, Athens, USA, Voice: 706-542-2464, Fax: 706-542-3910, whitem@uga.edu

Data deposition: Raw sequences are deposited in NCBI's Short Read Archive, reference number SRP137809 (https://submit.ncbi.nlm.nih.gov/subs/sra/SUB3748706/overview).

Short title: Recombination hotspots in threespine stickleback fish

Abstract

Meiotic recombination is a highly conserved process that has profound effects on genome evolution. Recombination rates can vary drastically at a fine-scale across genomes and often localize to small recombination "hotspots" with highly elevated rates surrounded by regions with little recombination. Hotspot targeting to specific genomic locations is variable across species. In some mammals, hotspots have divergent landscapes between closely related species which is directed by the binding of the rapidly evolving protein, PRDM9. In many species outside of mammals, hotspots are generally conserved and tend to localize to regions with open chromatin such as transcription start sites. It remains unclear if the location of recombination hotspots diverge in taxa outside of mammals. Threespine stickleback fish (*Gasterosteus aculeatus*) are an excellent model to examine the evolution of recombination over short evolutionary timescales. Using an LD-based approach, we found recombination rates varied at a fine-scale across the genome, with many regions organized into narrow hotspots. Hotspots had divergent landscapes between stickleback populations, where only ~15% were shared, though part of this divergence could be due to demographic history. Additionally, we did not detect a strong association of PRDM9 with recombination hotspots in threespine stickleback fish. Our results suggest fine-scale recombination rates may be diverging between closely related populations of threespine stickleback fish and argue for additional molecular characterization to verify the extent of the divergence.

Key words: Recombination, Threespine Stickleback, Linkage-Disequilibrium, Recombination Hotspots

Introduction

Meiotic recombination is a highly-conserved process across a broad range of taxa (de Massy 2013; Petes 2001). Recombination creates new allelic combinations by breaking apart haplotypes (Coop and Przeworski 2007; Otto and Lenormand 2002), promotes the proper segregation of chromosomes during meiosis in many species (Davis and Smith 2001; Fledel-Alon et al. 2009; Kaback et al. 1992; Mather 1936), and has a pronounced impact on the evolution of genomes (Mugal et al. 2015; Webster and Hurst 2012). In many species, meiotic recombination occurs in small 1-2 kb regions called recombination "hotspots" which are surrounded by large genomic regions with little to no recombination (Barton et al. 2008; Baudat et al. 2010; Hellsten et al. 2013; Jeffreys et al. 1998; McVean et al. 2004; Myers et al. 2005; Steiner et al. 2002).

In most species, hotspot location is highly conserved over long evolutionary timescales (Kawakami et al. 2017; Lam and Keeney 2015; Singhal et al. 2015; Tsai et al. 2010). For example, finches share upwards of 73% of hotspots across 3 million years of evolution (Singhal et al. 2015) while species of *Saccharomyces* share 80% of hotspots over 15 million years of evolution (Lam and Keeney 2015). Evolutionarily conserved hotspots are often localized around regions of open chromatin such as transcription start sites (TSSs) and CG-rich regions (i.e. CpG islands) in vertebrates (Auton et al. 2013; Kawakami et al. 2017; Lee et al. 2004; Pan et al. 2011; Pokholok et al. 2005; Singhal et al. 2015; Tischfield and Keeney 2012). This localization pattern is thought to be due the opportunistic nature of Spo11, a meiosis specific protein which initiates recombination by creating double stranded breaks at regions of open chromatin (Celerin et al. 2000; Ohta et al. 1994; Pan et al. 2011).

A notable exception to strong conservation of recombination hotspots has been documented in mammals, where hotspot location evolves rapidly between closely related species or even between populations (Baker et al. 2015; Brick et al. 2012; Pratto et al. 2014; Smagulova et al. 2016; Stevison et al. 2015). Contrary to the pattern observed in conserved systems, rapidly evolving hotspots typically form away from functional genomic elements and are localized by the zinc finger histone methyltransferase protein, PRDM9 (Baker et al. 2015; Baudat et al. 2010; Billings et al. 2013; Brick et al. 2012; McVean et al. 2004; Myers et al. 2005; Myers et al. 2010; Myers et al. 2008; Parvanov et al. 2010; Powers et al. 2016; Pratto et al. 2014). PRDM9 contains multiple DNA-binding zinc fingers that are under strong positive selection, leading to divergent

hotspot localization between closely related species (Baker et al. 2015; Billings et al. 2013; Brick et al. 2012; Myers et al. 2010; Parvanov et al. 2010). Though rapidly evolving hotspots have only been documented in some mammals, positive selection is acting on the zinc finger domain of PRDM9 orthologs in many non-mammalian species (Baker et al. 2017; Oliver et al. 2009). This raises the intriguing possibility that some species outside of mammals may also have rapidly evolving hotspots. It is also possible that PRDM9 is not necessary for rapid evolution of hotspots in other species and that other mechanisms could lead to the evolution of fine-scale rates of recombination over short timescales.

Threespine stickleback fish (*Gasterosteus aculeatus*) are an excellent system to study the evolution of fine-scale recombination rates. Multiple populations of threespine stickleback fish have independently adapted to freshwater environments from marine ancestors in the last 10-15 thousand years (Bell and Foster 1994; Orti et al. 1994), providing the opportunity to study the parallel evolution of hotspots in well-characterized populations across the Northern Hemisphere (Bell and Foster 1994; Ostlund-Nilsson et al. 2007; Wootton 1976). Broad-scale recombination rates have been examined in threespine stickleback using genetic crosses (Glazer et al. 2015; Peichel et al. 2001; Roesti et al. 2013; Sardell et al. 2018), but fine-scale recombination rates have not been estimated due to low marker density.

Fine-scale recombination rates can be estimated through a variety of approaches. Recombination rates can be directly measured through genetic linkage maps (Broman et al. 1998; Campbell et al. 2016; Drouaud et al. 2006; Marand et al. 2017) or though sperm genotyping (Baudat and de Massy 2007; Guillon and de Massy 2002; Jeffreys et al. 2001). Both methods require a large number of progeny or sperm and a high density of genetic markers to capture a sufficient number of crossovers. Recombination rates can also be indirectly measured by identifying the binding sites of proteins that initiate double strand breaks (Pratto et al. 2014; Smagulova et al. 2011) as well as repair double strand breaks through homologous recombination (Dumont and Payseur 2011; Froenicke et al. 2002). Another broadly used approach estimates recombination rates from patterns of linkage disequilibrium (LD) in populations, providing a historical measure of meiotic crossovers over multiple generations (Chan et al. 2012; McVean et al. 2004; Myers et al. 2005; Wall and Stevison 2016). LD-based methods are able to estimate rates at a fine-scale, but rate estimation can be biased by demographic history (e.g. bottlenecks, population expansions, population sub-structure, etc.)

(Dapper and Payseur 2017; Johnston and Cutler 2012), increasing false negative and false positive rates when calling recombination hotspots (Dapper and Payseur 2017). Despite the higher error rates, many recombination hotspots identified through LD-based methods have been validated using other approaches (Jeffreys et al. 2005; Morgan et al. 2017; Myers et al. 2006).

Here, we used an LD-based approach to estimate genome-wide recombination rates in a marine (Puget Sound) and freshwater (Lake Washington) population of threespine stickleback fish. We found recombination landscapes varied at a fine-scale between the two populations, often organized into recombination hotspots. We found most recombination hotspots were not shared between populations. We describe how the complex demographic histories of threespine stickleback fish populations (Bell and Foster 1994; Ferchaud and Hansen 2016; Hohenlohe et al. 2010; Liu et al. 2016) may influence the overall distribution of recombination hotspots and argue that the patterns we observe may not be completely driven by population bottlenecks. Additionally, we found little evidence that threespine stickleback hotspots are associated with PRDM9 binding, indicating hotspots are likely localized by a different mechanism.

Materials and Methods

Whole genome sequencing and assembly

Genomic DNA was extracted from caudal tail clips of 13 female and 12 male fish collected from Lake Washington (freshwater population; Washington, USA) and 18 female and 6 male fish collected from Northern Puget Sound (marine population; Washington, USA) using a standard phenol-chloroform extraction. Paired-end libraries were prepared using the Illumina TruSeq kit and were size-selected to target 400 bp fragments. Libraries were multiplexed and sequenced on Illumina NextSeq lanes for 300 cycles (Georgia Genomics and Bioinformatics Core, University of Georgia). Residual adapter sequences and low quality regions were trimmed from the sequencing reads using Trimmomatic (v0.33) with the following parameters: PE –phred 33 slidingwindow:4:20. Trimmed reads were aligned to the revised threespine stickleback genome assembly (supplemental file S5, https://datadryad.org/resource/doi:10.5061/dryad.q018v/1) (Glazer et al. 2015) using Bowtie2 (v2.2.4, default parameters) (Langmead and Salzberg 2012). With these parameters, the average alignment rate for Lake Washington was 94.2% and 87.3% for Puget Sound. Reads with a mapping PHRED quality score of 20 or less were removed from the analysis (Samtools, v1.2.0, default parameters) (Li et al. 2009). For

Puget Sound, four female individuals had 5x or lower sequencing coverage and were removed from the analysis. After removing poorly aligned reads and low coverage individuals, the average read coverage across all individuals in each population was 17x and 22x for Lake Washington and Puget Sound, respectively.

Two outgroup species were used to infer ancestral allele states and to estimate mutation matrices for each population (see Estimation of Recombination Rates). Whole-genome Illumina sequences for one female ninespine stickleback fish (*Pungitius pungitius*, DRX012173) (White et al. 2015) and one female blackspotted stickleback fish (*Gasterosteus wheatlandi*, DRX012174) (Yoshida et al. 2014) were aligned to the revised threespine stickleback genome assembly (Glazer et al. 2015) using Bowtie2 (v2.2.4). Less stringent alignment parameters were used to allow for greater sequence divergence between threespine stickleback and each outgroup (-D 20 –R 3 –N 1 –L 20 –I S,1,0.50 -rdg 3,2 -rfg 3,2 -mp 3). The overall alignment rate of *P. pungitius* was 46.0% whereas the overall alignment rate of *G. wheatlandi* was 74.2%. The higher alignment rate of *G. wheatlandi* is consistent with *G. wheatlandi* sharing a more recent common ancestor with *G. aculeatus* (Kawahara et al. 2009).

SNP genotyping

Single nucleotide polymorphisms (SNPs) were genotyped in each threespine stickleback population and outgroup species independently following the GATK best practices for SNP discovery for whole genome sequences (v3.6) (Van der Auwera et al. 2013). PCR duplicates were removed using MarkDuplicates (REMOVE_DUPLICATES=true). Regions around insertions or deletions (indels) were realigned with RealignerTargetCreator (default parameters) and IndelRealigner (default parameters). Variants were called for each individual using HaplotypeCaller (genotyping mode DISCOVERY). Joint genotyping (GenotypeGVCFs, default parameters) was completed by pooling all individuals for each population. Low-quality SNPs were filtered from the data set using vcftools (v0.1.12b) (Danecek et al. 2011) with the following filters: removing all sites with more than two alleles, removing sites where genotype data was missing among individuals, removing sites where the population mean depth coverage was less than half or greater than twice the average coverage for each population (Lake Washington: 8x – 24x read depth coverage; Puget Sound: 11x - 44x read depth coverage), and removing sites with a genotype quality score less than 30. Singletons and sites fixed for the alternate allele across all

individuals in a population were also removed. After filtering, the Lake Washington population had 5,054,729 SNPs genome-wide (11 SNPs/kb) and the Puget Sound population had 4,142,876 SNPs (9 SNPS/kb) genome-wide (prior to filtering Lake Washington had 11,937,220 SNPs and Puget Sound had 11,070,421 SNPs). For the outgroup species, *P. pungitius* and *G. wheatlandi*, low-quality SNPs were excluded by removing variants with a genotyping quality score less than 30 or a read depth less than two, resulting in 13,691,521 SNPs genome-wide in *G. wheatlandi* (16,783,618 SNPs prior to filtering) and 7,791,420 in *P. pungitius* (26,173,287 SNPs prior to filtering).

Haplotype phasing

Each chromosome was phased independently with SHAPEIT (v2.r837), a read-aware phasing tool (Delaneau et al. 2013). Phase-informative reads with two heterozygous SNPs on the same read were identified to assist with the estimation of haplotypes. Phase-informative reads had a mapping quality score greater than 20. Convergence of the MCMC algorithm was estimated by examining switch error rates between individual runs. A low switch error rate would indicate that the MCMC phasing runs have converged on a similar haplotype configuration. Switch error was measured using vcftools (v0.1.12b) using –diff-switch-error (Danecek et al. 2011). A low switch error was achieved within a reasonable run time with the following SHAPEIT parameters: --main 2000 --burn 200 --prune 210 --states 1000 (average switch error between phasing runs: 0.824% for Lake Washington and 1.26% for Puget Sound). All other parameters were left at the default values.

Estimation of recombination rates

Recombination rates were estimated with LDHelmet (v1.7) (Chan et al. 2012). LDHelmet estimates historical recombination rates from population data by analyzing patterns of linkage disequilibrium across phased individuals. The ancestral allele state was defined for every SNP in each threespine stickleback population by comparing to the allele present in the two outgroup species. An ancestral allele state could not be assigned if a polymorphism was segregating among the outgroup species. Therefore, SNPs were only assigned an ancestral state if *P. pungitius* and *G. wheatlandi* were homozygous for the same allele. The ancestral allele was assumed to be the nucleotide carried by *P. pungitius* and *G. wheatlandi*, and was assigned a prior

probability of 0.91. To allow for uncertainty in the ancestral allele state, the other three possible nucleotides were assigned prior probabilities of 0.03. If the ancestral allele state could not be inferred, the prior probability of each nucleotide being the ancestral allele was set as the overall frequency of that particular nucleotide on the chromosome. Nucleotide frequencies were empirically determined from all sites on a threespine stickleback chromosome where *P. pungitius* and *G. wheatlandi* had read coverage that passed the filtering scheme. Mutation matrices were estimated for each population separately. For every position where an ancestral allele state could be inferred, the total number of each type of mutation away from the ancestral allele was quantified. A normalized 4x4 mutation matrix was generated for each chromosome as previously described (Chan et al. 2012). The ancestral allele state and mutation matrices were generated using a custom Perl script.

Each LDHelmet module was run using the following parameters. Custom Python scripts were used to create the SNP sequence and SNP position input files. Full FASTA sequence were created using vcf2fasta from vcflib (available at https://github.com/vcflib/vcflib). Haplotype configuration files were created for each chromosome with the find_confs module using a window size of 50 SNPs (-w 50). Likelihood tables were created using table_gen with the recommended grid of population scaled recombination rates per base pair (ρ/bp) (-r 0.0 0.1 10.0 1.0 100.0). Watterson's θ was estimated using a custom Python script with the R package, PopGenome (Pfeifer et al. 2014), where Watterson's θ was calculated in 2 kb regions with a sliding window of 1 kb and all windows were averaged together. To maintain a reasonable computational time, a single representative likelihood lookup table was generated for the autosomes of each population from chromosome one, using the average Watterson's θ between Lake Washington and Puget Sound (-t 0.002). Although Watterson's θ was different between the Lake Washington and Puget Sound populations, previous studies have determined that small changes to parameters such as Watterson's θ do not affect the final likelihoods (Auton and McVean 2007; McVean et al. 2004). Separate likelihood tables were created for the pseudoautosomal region of the sex chromosomes (chromosome 19). Padé coefficient files were created using the module pade with a Watterson's θ of 0.002 and the recommended 11 padé coefficients (-t 0.002 –x 11). The module rjmcmc was run for 1 million iterations with 100,000 burn in iterations, a block penalty of 10, and a window size of 50 SNPs (-w 50 –b 10 –burn_in 100000 –n 1000000). Population-scaled recombination rates were extracted from the rjMCMC

run with the post_to_text module. Recombination rates were reported in ρ/bp where ρ is a population scaled recombination rate ($4N_er$).

Correlation with genetic maps

Population-scaled recombination rates were compared with recombination rates estimated from a high-density genetic linkage map (Glazer et al. 2015). Recombination rates from LDHelmet were converted from ρ/bp to cM/Mb as previously described (Smukowski Heil et al. 2015). Briefly, the recombination rate (cM/Mb) was calculated between every pair of adjacent markers in the genetic map and a chromosome-wide recombination rate was calculated as the average among the regions. The average LD-based recombination rate (ρ/Mb) was computed in the same individual regions of a chromosome in Lake Washington and Puget Sound by averaging the per bp rho estimate across the total length of the region (ρ/Mb). A single conversion factor was calculated for each chromosome. Each conversion factor was calculated by dividing the average linkage map recombination rate for a chromosome (in cM/Mb) by the average LD-based recombination rate (ρ/Mb) for that chromosome.

Identification of recombination hotspots

Recombination hotspots were defined using a sliding window approach. In each window, the average recombination rate within a 2 kb window was compared to the average recombination rate from the 40 kb regions flanking either side of the 2 kb window. Hotspots were defined as the 2 kb regions that had a 5-fold or higher recombination rate relative to the mean recombination rate in the flanking background regions. The 2 kb windows iterated forward in 1 kb increments. If multiple hotspots were found within a 5 kb region, only the hotspot with the highest rate was retained. Misassemblies in the reference genome could generate false hotspots. To limit this, all hotspots that spanned a contig boundary in the reference genome were removed (384 hotspots out of 4,349 total hotspots). Hotspots were considered shared between populations if the midpoints of the two hotspots were within 3 kb of each other. Random permutations were used to calculate the expected amount of hotspot overlap between Lake Washington and Puget Sound. 10,000 random permutations were drawn from the genome totaling the number of 2 kb hotspots for each population. Recombination hotspots were identified and filtered using custom Perl and Python scripts.

280

281 **Genetic variation within and between populations**

282 Within population nucleotide diversity (π) and Tajima's D were calculated separately for
283 each chromosome. To capture rare variants, previously excluded singletons were included in the
284 analysis. Nucleotide diversity and Tajima's D were calculated using the R package, PopGenome
285 (Pfeifer et al. 2014) and a custom Python script. Nucleotide diversity was calculated
286 between populations by combining SNP variants among all individuals in each population.
287 Population structure was estimated between Lake Washington and Puget Sound using
288 FastStructure (v1.0) (Raj et al. 2014). For this analysis, SNPs from Lake Washington and Puget
289 Sound were merged using vcftools (Danecek et al. 2011) and only biallelic sites with no missing
290 data were retained. The sex chromosomes (chromosome 19) were also excluded. The final SNP
291 dataset was composed of 4,113,937 SNPs. Three trials were completed at K values of 1, 2, and 3.
292 These K values were chosen to differentiate scenarios where Lake Washington and Puget Sound
293 were one panmictic population (K=1) or Lake Washington and Puget Sound were two distinct
294 populations (K=2). A K of 3 was chosen to identify any hidden population structure within either
295 population. The model that best explained the population structure was determined using
296 chooseK.py (Raj et al. 2014) and the structure plot was visualized using distructK.py (Raj et al.
297 2014).

298

299 **Estimation of demographic history**

300 Demographic history can affect LD-based estimates of recombination rates (Dapper and
301 Payseur 2017; Johnston and Cutler 2012). To determine whether the demographic history of
302 threespine stickleback fish could influence the ability to detect recombination hotspots, hotspots
303 were assayed in simulated haplotypes with known recombination profiles and demographic
304 histories. Demographic histories used in the simulations were based on the estimated histories of
305 Lake Washington and Puget Sound, modeled using a Pairwise Sequentially Markovian
306 Coalescent (PSMC) process with default parameters (Li and Durbin 2011; Liu and Hansen
307 2017). PSMC was run on one female from Lake Washington and one female from Puget Sound.
308 Confidence intervals were estimated on 100 bootstrap replicates. Demographic histories were
309 visualized using psmc_plot.pl (Li and Durbin 2011).

310

311 **Simulations using estimated demographic histories**

312 Using the demographic histories estimated with PSMC, 250 kb haplotypes with four 2 kb
313 recombination hotspots were simulated using the program fin, part of the LDHat software
314 package (Auton and McVean 2007; McVean et al. 2004). The hotspots were place 50 kb apart at
315 75, 125, 175, and 225 kb. The background recombination rate was set at 0.03 ρ/kb. Hotspots had
316 varied intensities from 2 to 20 times the background rate, set at 0.06 ρ/kb, 0.15 ρ/kb, 0.3 ρ/kb,
317 and 0.6 ρ/kb. One scenario simulated a constant effective population size, with 500 sequences,
318 40 haplotypes each, with an average Watterson's θ of 0.00355, the average between Lake
319 Washington and Puget Sound (–nsamp 40 –len 250000 –theta 0.00355). For both populations, a
320 bottleneck was simulated 8,000 generations ago (Puget Sound: t = 0.029, theta = 0.0036; Lake
321 Washington: t = 0.022, theta = 0.0035). Two bottleneck strengths were simulated by setting the
322 probability that a lineage coalesces to 10% or 90% (s = 0.1, 0.9). Overall, hotspot sharing
323 between simulated Lake Washington and simulated Puget Sound populations was quantified by
324 examining all pairwise comparisons between populations and bottleneck strengths. The first
325 hotspot simulated should not be called using our method as it falls below our cutoff, but can
326 provide information about how hotspots that fall below our cutoff affect hotspot calling. The
327 number of false positive and false negative hotspots were calculated using custom Python scripts.
328
329 **Location of hotspots around transcription start sites**

330 Transcript annotations from Ensembl (build 90) were lifted to the revised threespine
331 stickleback genome assembly (Glazer et al. 2015) by aligning each transcript using BLAT (v36,
332 default parameters) (Kent 2002). Aligned transcripts were only retained if the entire transcript
333 aligned to the revised genome assembly. Transcript start sites (TSSs) consisted of a 2 kb region,
334 centered at the start position of the transcript. A hotspot was considered overlapping with a TSS
335 if the midpoint of the hotspot overlapped with any part of a 2 kb TSS region. Enrichment of
336 hotspots in TSSs were compared against 10,000 random permutations. 2 kb regions were
337 randomly drawn across the genome, totaling the number of hotspots identified in each
338 population. TSS annotation filtering, overlap of hotspots with TSSs, and random permutations
339 were completed using custom Python scripts.
340
341 **GC-Biased Substitutions**

GC to AT and AT to GC substitutions were quantified within 2 kb regions of the genome that had recombination rates in the top and bottom 5% as well as within all 2 kb recombination hotspots. The top 5% of recombination rates captures regions of the genome that may broadly have high recombination rates and not contain recombination hotspots. The top 5% of recombination rates includes 96 hotspots for Lake Washington (out of 1,627 hotspots) and 314 hotspots for Puget Sound (out of 2,338 hotspots). The equilibrium GC content was calculated as the proportion of AT to GC substitutions out of the total pool of substitutions (AT to GC and GC to AT) (Meunier and Duret 2004; Singhal et al. 2015; Sueoka 1962). To increase the total number of sites available for the analysis, the ancestral allele state was inferred using only *G. wheatlandi*, rather than requiring a matching ancestral allele in both *G. wheatlandi* and *P. pungitius*. Because CpG sites can have higher mutation rates (Fryxell and Moon 2005; Weber et al. 2014), all consecutive CG sites in the ancestral sequence were removed from the analysis.

DNA motif identification

MEME (v4.11.0) was used to identify novel DNA motifs enriched in hotspots and matched coldspots (Bailey and Eklan 1994). Each hotspot was matched to a randomly selected 2 kb coldspot, which was located at least 25 kb from any identified hotspot, contained a GC nucleotide content that was within 2% of the hotspot after removing ancestral CpG sites (GC-matched), and had a mean recombination rate that was less than half the background recombination rate of the population (Lake Washington: less than 0.017 ρ/bp; Puget Sound: less than 0.035 ρ/bp). MEME ignored motif occurrences if they were present in a hotspot multiple times (-mod zoops). This was to prevent the reporting of repetitive motifs. MEME was run separately for each chromosome and population and was completed when 50 motifs were identified (-nmotifs 50). Motif identification was conducted separately for shared hotspots and population-specific hotspots.

The DNA-binding protein, PRDM9, is important for localizing recombination hotspots in mammals (Baker et al. 2015; Baudat et al. 2010; Billings et al. 2013; Brick et al. 2012; Myers et al. 2010; Myers et al. 2008; Parvanov et al. 2010; Powers et al. 2016; Pratto et al. 2014). To determine if any PRDM genes had a role in localizing hotspots in threespine stickleback fish, FIMO (v4.11.0, default parameters) (Grant et al. 2011) was used to scan hotspot sequences for the predicted DNA binding motifs for each of the 11 annotated PRDM genes in the threespine

stickleback genome (Ensembl, build 90). DNA binding motifs for each PRDM protein were predicted using the Cys_2His_2 zinc finger prediction tool, Predicting DNA-binding Specificities for the Cys_2His_2 Zinc Finger Proteins (Persikov et al. 2009; Persikov and Singh 2014). Predicted zinc finger domains were included if the HMMER bit score for the zinc fingers was 17.7 or higher (Persikov et al. 2009; Persikov and Singh 2014). To determine the expected number of occurrences of a motif of the same length and GC composition in hotspots, the PRDM motifs were shuffled 100 separate times. FIMO was run on the shuffled motifs to create a null distribution. Motifs were shuffled using a custom python script.

Results

Genetic differentiation between Lake Washington and Puget Sound

Freshwater populations of threespine stickleback fish frequently exhibit signs of past bottlenecks, consistent with their colonization from marine ancestors ~10-15 thousand years ago (Bell and Foster 1994; Ferchaud and Hansen 2016; Hohenlohe et al. 2010; Liu et al. 2016). Given the recent divergence and the close geographic proximity between Lake Washington (freshwater) and Puget Sound (marine), we first examined whether these two populations were genetically distinct. Using FastStructure, a two population model was the most highly supported (marginal likelihood = -0.834, Supplemental Figure 1).

Within each population, we explored whether there were signatures of past bottleneck events. The average nucleotide diversity within both populations was similar (Lake Washington: 0.003; Puget Sound: 0.003), whereas the genome-wide average nucleotide diversity between populations was 0.004. The nucleotide diversity values we calculated are similar to previously reported values for other marine and freshwater stickleback populations (Guo et al. 2015; Hohenlohe et al. 2010; Kitano et al. 2007). Both populations had negative Tajima's D values (Tajima 1989), consistent with an excess of rare variants from a recent population expansion (Lake Washington: -0.422; Puget Sound: -0.723).

The demographic histories of Lake Washington and Puget Sound were estimated using Pairwise Sequentially Markovian Coalescent (PSMC) models (Figure 1). Puget Sound experienced a bottleneck from around 18,000 years ago until about 8,000 years ago where the effective population size decreased to 74,250 ± 1,259 individuals (starting N_e = 132,700 ± 796) while Lake Washington experienced a small bottleneck around the same time where the effective

404 population size decreased to 91,760 ± 1,960 individuals (starting N_e = 129,138 ± 897) (Figure 1).
405 Both populations have had a constant effective population size for the last ~5,000 years. Puget
406 Sound has a larger effective population size than Lake Washington, matching the expected
407 pattern of marine populations having larger effective population sizes than freshwater
408 populations (DeFaveri and Merila 2015; Gow et al. 2006; Makinen et al. 2006).
409
410 **Fine-scale estimation of recombination rates across the genome**
411 Using a dense set of SNP markers from whole-genome sequencing, we estimated
412 recombination rates across the genomes of Lake Washington and Puget Sound threespine
413 stickleback fish. The average genome-wide population recombination rate in Lake Washington
414 was half of the rate observed in Puget Sound (Lake Washington: 0.035 ρ/bp; Puget Sound: 0.072
415 ρ/bp; Wilcoxon Rank Test; $p < 0.001$, Supplemental Table 1). Despite having an overall lower
416 genome-wide recombination rate in Lake Washington, recombination rates were largely
417 conserved at broad scales between the two populations. We observed a highly significant
418 positive correlation of recombination rates between the populations at the scale of 500 kb
419 windows (Spearman's Rank Correlation; $r = 0.931$, $p < 0.001$; Figure 2; Supplemental Figure 2).
420 Additionally, recombination rates were lower at the center of chromosomes (center 25% of all
421 chromosomes) and significantly higher at the ends of the chromosomes (terminal 25% of all
422 chromosomes) for both populations (Wilcoxon Rank Test; Lake Washington: ends of
423 chromosomes = 0.069 ρ/bp, center of chromosomes = 0.009 ρ/bp, $p < 0.001$; Puget Sound: ends
424 of chromosomes = 0.108 ρ/bp, center of chromosomes = 0.016 ρ/bp, $p < 0.001$; Figure 2). Rate
425 differences at chromosome ends have been documented in other populations of threespine
426 stickleback (Glazer et al. 2015; Roesti et al. 2013; Sardell et al. 2018) as well as across a wide-
427 range of other animals, plants, and fungi (Barton et al. 2008; Berner and Roesti 2017; Broman et
428 al. 1998; See et al. 2006).
429 To determine whether the broad-scale recombination rates we estimated from LD-based
430 methods are concordant with recombination rates measured from linkage mapping, we compared
431 the rates from Lake Washington and Puget Sound with the rates estimated from a genetic linkage
432 map from a freshwater female and a marine male (Glazer et al. 2015). We found a significant
433 positive correlation between recombination rates in both populations and the linkage map
434 (Spearman's Rank Correlation; Lake Washington: $r = 0.830$, $p < 0.001$; Puget Sound: $r = 0.810$,

435 $p < 0.001$; Figure 3). These data indicate that broad-scale changes are conserved across multiple
436 populations of threespine stickleback fish and confirm that the recombination rates estimated
437 from LD-based methods largely parallel the rates observed from genetic linkage maps. Although
438 broad-scale (Mb) recombination rates tend to be conserved over longer evolutionary timescales
439 (Fledel-Alon et al. 2009; Kong et al. 2002; Serre et al. 2005; Stevison et al. 2015), fine-scale (kb)
440 rates within chromosomes can rapidly evolve (Barton et al. 2008; Hellsten et al. 2013; McVean
441 et al. 2004; Myers et al. 2005). In many organisms, recombination is organized locally into
442 narrow regions of very high rates (i.e. "hotspots"), surrounded by regions of little to no
443 recombination (i.e. "coldspots") (Baudat et al. 2010; Jeffreys et al. 1998; Steiner et al. 2002).
444 Consistent with this, we found highly variable fine-scale recombination rates across individual
445 chromosomes in both Lake Washington and Puget Sound (Figure 4; Supplemental Figures 3 and
446 4).
447
448 **Divergent hotspot locations between populations of threespine sticklebacks**
449 Using a sliding-window approach, we identified 2,338 hotspots in Puget Sound and 1,627
450 hotspots in Lake Washington. Strikingly, only 312 of these hotspots were shared between
451 populations (13.3% of hotspots in Puget Sound and 19.2% of hotspots in Lake Washington).
452 This lack of hotspot overlap between Lake Washington and Puget Sound may, in part, be due to
453 hotspots falling just below the hotspot threshold. To investigate this, we looked for any increase
454 in recombination rate in locations where hotspots were present in one population, but absent in
455 the other. We found little evidence of a localized increase in recombination rate in these regions.
456 Recombination rates were close to the background rate in the population where hotspots were
457 deemed absent (Figures 5A and 5B). This pattern was even more apparent when shared hotspots
458 were removed from the analysis (Figures 5C and 5D). The small degree of overlap we observed
459 in hotspots between the populations was much greater than what would be expected from chance
460 alone (10,000 random permutations; $p < 0.001$; Supplemental Figure 5), indicating much of the
461 hotspot overlap likely represents shared ancestry.
462
463 **Increased recombination rate in the pseudoautosomal region**
464 Genetic recombination between sex chromosomes is restricted to the pseudoautosomal
465 region (PAR), where rates of recombination can be orders of magnitude above genome-wide

averages (Otto et al. 2011; Wright et al. 2016). In threespine stickleback, crossing over between the X and Y chromosomes is restricted to a ~2.5 Mb PAR (Peichel et al. 2004; Roesti et al. 2013; White et al. 2015; Yoshida et al. 2014). Because of the potential for high rates of crossing over in the PAR, we estimated population-scaled recombination rates for this region independently from the autosomes. The average recombination rate in the PAR was 0.232 ρ/bp for Puget Sound and 0.129 ρ/bp for Lake Washington. These rates were significantly higher than the average recombination rate across the autosomes (Lake Washington autosome average rate: 0.035 ρ/bp, $p < 0.001$; Puget Sound autosome average rate: 0.072 ρ/bp, $p < 0.001$; Figure 6). Although we observed some fine-scale variation in recombination rates across the PAR (Figure 6), we identified very few hotspots, which may be due to the increased background recombination rate across the PAR.

Demographic history may not completely account for hotspot divergence

To explore how changes in past effective population size (N_e) may have affected our ability to detect hotspots, we simulated haplotypes with known demographic histories that followed the demographic histories we estimated from Lake Washington and Puget Sound, along with a known distribution of recombination hotspots. If the minimal hotspot overlap we observed between populations of threespine stickleback fish was because of high false positive and false negative rates induced by demographic history, we would expect hotspots to be incorrectly called to a similar degree in the bottleneck simulations. Both bottleneck strengths exhibited elevated false positive and false negative rates compared to the control simulation, with the highest false positive and false negative rates under the strong bottleneck scenario (Supplemental Table 2). To determine the overall effect of elevated error rates on determining the number of shared hotspots between populations, we compared the simulated Lake Washington haplotypes to the simulated Puget Sound haplotypes from both bottleneck scenarios. Despite the elevated error rates, hotspot sharing was higher between the simulated populations than the observed number of hotspots shared between actual Lake Washington and Puget Sound populations for the weak bottleneck (weak bottleneck: Lake Washington: 59.7%; Puget Sound: 55.2%; actual Puget Sound shared hotspots: 13.3%; actual Lake Washington shared hotspots: 19.2%). This indicates that a weak bottleneck in both populations is not sufficient to drive the high degree of hotspot divergence we observed. However, if the bottleneck strength was very high (s=0.9) in both populations,

497 elevated error rates in hotspot calling could result in a lack of hotspot overlap that mirrors the
498 divergence we observed between populations. In this simulation, there was a similar percent of
499 shared hotspots as observed in the actual populations (strong bottleneck: Lake Washington:
500 20.7%; Puget Sound: 19.8%; actual Puget Sound shared hotspots: 13.3%; actual Lake
501 Washington shared hotspots: 19.2%).
502 Based on the demographic histories we estimated, Lake Washington experienced a less
503 intense bottleneck than Puget Sound. We therefore also used simulations to explore the expected
504 hotspot overlap if only one of the populations experienced a strong bottleneck. If Puget Sound
505 experienced a strong bottleneck and Lake Washington experienced a weak bottleneck, 36.7% of
506 hotspots were shared in the simulated Lake Washington population and 20.5% of hotspots were
507 shared in the simulated Puget Sound population (actual Lake Washington shared hotspots:
508 19.2%, actual Puget Sound shared hotspots: 13.3%). Except for a scenario where both
509 populations underwent a severe bottleneck in the past, our simulations suggest that demographic
510 history alone is not sufficient to completely explain the divergence we observed in hotspot
511 location between populations.
512
513 **Hotspots are enriched around transcription start sites**
514 Hotspot localization in genomes varies among taxa. In yeast, birds, and some plants,
515 where hotspots are evolutionarily conserved, hotspots tend to be enriched within transcription
516 start sites (Kawakami et al. 2017; Pan et al. 2011; Singhal et al. 2015; Tischfield and Keeney
517 2012). In mammals with rapidly evolving hotspots, hotspots are typically located away from
518 genic regions (Brick et al. 2012; Brunschwig et al. 2012; Myers et al. 2005). We investigated
519 whether threespine stickleback fish hotspots mimic either of the patterns seen in other systems.
520 We found an enrichment of hotspots around TSSs, compared to random permutations of hotspots
521 (Lake Washington: 26% of hotspots fell within 3 kb of a TSS, $p < 0.034$; Puget Sound: 29% of
522 hotspots fell within 3 kb of a TSS, $p < 0.001$; Supplemental Figure 6). This pattern also held
523 when examining only population specific hotspots (Lake Washington: $p = 0.007$; Puget Sound: p
524 < 0.001; Supplemental Figure 7); however, shared hotspots were not enriched in TSSs compared
525 to random permutations (Lake Washington: $p = 0.370$; Puget Sound: $p = 0.827$; Supplemental
526 Figure 6). The lack of significant enrichment of shared hotspots around TSSs is likely due to the
527 small sample size. When we randomly drew samples from the population-specific hotspots that

were equal in size to the shared hotspot pools, there was no longer enrichment around TSSs (Lake Washington: $p = 0.947$; Puget Sound: $p = 0.808$).

Regions of high recombination exhibit GC-biased nucleotide substitution

Recombination leaves distinct signatures of nucleotide substitution across the genome (Duret and Arndt 2008; Mugal et al. 2015; Webster and Hurst 2012). Over time, the repair of heteroduplex DNA during meiosis favors the substitution of GC nucleotides over AT nucleotides, which increases the frequency of GC nucleotides, leading to GC-biased base composition (Lesecque et al. 2013; Marais 2003; Meunier and Duret 2004). Regions of the genome with higher recombination rates tend to have higher GC-biased base composition (Kawakami et al. 2017; Kong et al. 2002; Meunier and Duret 2004; Singhal et al. 2015). To determine whether regions of higher recombination rate showed signatures of GC-biased gene conversion, we calculated equilibrium GC content (Meunier and Duret 2004; Singhal et al. 2015; Sueoka 1962) in regions of the genome with the highest and lowest recombination rates (top and bottom 5%) as well as within recombination hotspots.

In both Lake Washington and Puget Sound, we detected a significantly higher equilibrium GC content in regions of the genome with a high recombination rate (top 5% of recombination rates among 2 kb windows) compared to regions of the genome with recombination rates in the bottom 5% (Table 1). Overall, these results indicate GC nucleotide composition is influenced by the historical recombination landscape across the threespine stickleback genome. Interestingly, although hotspots in both populations have locally elevated recombination rates, there was not a parallel increase in equilibrium GC content. Equilibrium GC content in population-specific hotspots and shared hotspots was not significantly different than regions of the genome with the lowest recombination rates (bottom 5%) (Table 1). Our results are consistent with recombination hotspots being more recently derived, where locally increased recombination rates have not yet had an effect on GC-biased nucleotide substitution.

PRDM genes are weakly associated with threespine stickleback recombination hotspots

Hotspots in many species are targeted to specific regions of the genome by DNA binding motifs (Baudat et al. 2010; Kon et al. 1997; Myers et al. 2008; Steiner et al. 2002). In species where PRDM9 targets recombination hotspots to specific regions of the genome, the zinc finger

559 domain of PRDM9 is typically under strong positive selection (Baker et al. 2015; Baudat et al.
560 2010; Billings et al. 2013; Myers et al. 2010; Oliver et al. 2009; Pratto et al. 2014) and the
561 protein contains functional KRAB and SSXRD domains (Baker et al. 2017). In Teleost fish, two
562 paralogs of PRDM9 have been identified, PRDM9α which contains all the protein domains and
563 PRDM9β which lacks the KRAB and SSXRD domains (Baker et al. 2017). Threespine
564 stickleback fish appear to have lost PRDM9α, but retain PRDM9β without the SSXRD and
565 KRAB domains. Consistent with a lack of function directing recombination hotspots, we did not
566 observe strong signatures of positive selection in the zinc finger domain of PRDM9β. We found
567 zero fixed differences between threespine and blackspotted stickleback for the PRDM9 ortholog.
568 There was one synonymous and one nonsynonymous mutation at moderate frequency in Lake
569 Washington and two synonymous and three nonsynonymous mutations at moderate frequency in
570 Puget Sound, indicating these mutations are likely not causing the population-specific
571 localization of hotspots we observed between Lake Washington and Puget Sound.
572 We also examined whether the predicted binding sites of any of the 11 previously
573 annotated PRDM genes in threespine stickleback fish were enriched in recombination hotspots.
574 Less than 14% of hotspots contained any of the predicted PRDM zinc finger binding domain
575 motifs (Supplemental Table 3). However, six of the motifs were significantly enriched in
576 hotspots, including PRDM9, when compared to scrambled motifs of the same size and GC
577 content (Supplemental Table 3), indicating PRDM genes could have some role in localizing a
578 subset of recombination hotspots. Outside of PRDM9 in mammals, multiple DNA binding motifs
579 assist with hotspot targeting in other systems such as *Schizosaccharomyces pombe* (Kon et al.
580 1997; Steiner et al. 2002). To see if other DNA motifs were targeting hotspots in threespine
581 stickleback fish, we searched for motifs enriched in hotspots. The most significant motifs
582 identified were simple mono- or di-nucleotide repeats which were present only in a subset of the
583 hotspots (Supplemental Figure 8). These repeats were not specific to hotspots as they were also
584 found in GC-matched coldspots.
585
586 **Discussion**
587 **Broad-scale recombination rates across the threespine stickleback genome**
588 At a broad scale, recombination rates across the threespine stickleback genome were
589 conserved between the two populations. This broad scale conservation of recombination rates is

a feature observed in many taxa (Fledel-Alon et al. 2009; Kong et al. 2002; Serre et al. 2005; Stevison et al. 2015) and may reflect the necessity of crossing over for the proper segregation of chromosomes during meiosis (Davis and Smith 2001; Fledel-Alon et al. 2009; Kaback et al. 1992; Mather 1936). Additionally, we observed differential rates of recombination associated with broad genomic regions that have been observed in other systems. First, we observed higher recombination rates towards the telomeres. In many species, the ends of chromosomes have higher rates of recombination (Barton et al. 2008; Berner and Roesti 2017; Kong et al. 2002; Roesti et al. 2013; Sardell et al. 2018), which is thought to be driven by male-specific localization of recombination (Broman et al. 1998; Moen et al. 2008; Singer et al. 2001). Our LD-based method estimates sex-averaged recombination rates, which does not allow us to test whether the pattern we observed around the ends of chromosomes is driven by males. However, sex-specific genetic linkage maps between the Japan Sea stickleback (*Gasterosteus nipponicus*) and the threespine stickleback (*G. aculeatus*) corroborate this pattern (Sardell et al. 2018). Second, we observed higher recombination rates in the pseudoautosomal region compared to the autosomes. Recombination rates in pseudoautosomal regions are often orders of magnitude above autosome-wide averages, as an obligate crossover should occur between the X and Y chromosomes in these small regions during every male meiosis (Hinch et al. 2014; Kauppi et al. 2012; Otto et al. 2011).

Overall, the genome-wide average recombination rate for Puget Sound was two-fold higher than in Lake Washington. Rate variation between populations or species can be driven by a number of processes. Structural variation (i.e. inversions, chromosomal rearrangements, and copy number variants) can contribute to rate variation among genomes. Indeed, recombination rates have been shown to vary across chromosomal regions due to segregating inversions between marine and freshwater populations of threespine stickleback (Glazer et al. 2015; Jones et al. 2012). Although structural variants could explain rate differences between Lake Washington and Puget Sound populations at a more localized level, they cannot explain the genome-wide rate differences we observed. Over longer evolutionary timescales, recombination rate also can evolve neutrally (Dumont and Payseur 2008), driving genome-wide rate variation between species. However, neutral divergence is likely not occurring at a pace that would alter genome-wide recombination rates between recently diverged populations of threespine stickleback fish. One plausible explanation for the observed rate differences is differences in

demographic history between the Lake Washington and Puget Sound populations. A larger effective population size could increase the population-scaled recombination rate (Burt 2000; Charlesworth 2009). In threespine stickleback, marine populations typically have a larger N_e than freshwater populations (DeFaveri and Merila 2015; Gow et al. 2006; Makinen et al. 2006), consistent with our observed pattern of a higher recombination rate in Puget Sound relative to Lake Washington.

Identifying hotspots using patterns of linkage disequilibrium

LD-based estimates of recombination rates can be affected by demographic processes that change patterns of linkage disequilibrium across the genome (Chan et al. 2012; Dapper and Payseur 2017; Johnston and Cutler 2012; McVean et al. 2004; Wall and Stevison 2016). The duration and timing of these events can have varying effects on hotspot identification, often reducing the power to detect hotspots and increasing the rate of errors (Dapper and Payseur 2017). Threespine stickleback fish have a complex history of bottleneck events and population expansions over the last 10-15 thousand years which vary across geographic regions (Bell and Foster 1994; Ferchaud and Hansen 2016; Hohenlohe et al. 2010; Liu et al. 2016; Orti et al. 1994). Based on simulations, demographic history likely has some role in the observed divergence in hotspot location between Lake Washington and Puget Sound populations, but it seems likely that population demography does not completely explain the pattern. Only in the scenario where both populations experienced a strong bottleneck do error rates rise high enough to mimic the observed divergence in hotspot location. However, our estimates of effective population size over time revealed that Lake Washington and Puget Sound did not experience similar fluctuations. Both populations began with effective population sizes that largely parallel those observed in other threespine stickleback fish populations (Liu and Hansen 2017; Ravinet et al. 2018). Puget Sound then experienced a larger population expansion roughly 18,000 years ago, followed with a decrease in population size at approximately 8,000 years ago. Lake Washington had a slight increase in population size, followed by a small bottleneck around the same time, but overall changes in effective population size were more stable in this population. Examination of where recombination hotspots are currently forming across the genome in Lake Washington and Puget Sound would help confirm the patterns we observed. Surveys of double strand break hotspots (Pratto et al. 2014; Smagulova et al. 2011) or crossover breakpoints in genetic crosses

(Broman et al. 1998; Campbell et al. 2016; Drouaud et al. 2006; Marand et al. 2017) would reveal the degree to which recombination hotspots are targeted to different genomic locations in these two populations.

Hotspot evolution in freshwater and marine threespine stickleback populations

Of the 3,965 hotspots between Lake Washington and Puget Sound, only ~15% of hotspots are shared, indicating many of the hotspots are recently derived within populations of threespine stickleback fish. Consistent with the recent evolution of hotspots, we did not observe an elevated equilibrium GC content in these regions. One possible model is that recombination hotspots can shift over short evolutionary timescales among regions of the genome that are susceptible to homologous recombination, such as regions of accessible chromatin. Both evolutionarily conserved and rapidly evolving hotspots tend to locate to regions of accessible chromatin (Lam and Keeney 2015; Ohta et al. 1994; Pan et al. 2011; Tischfield and Keeney 2012) or regions with histone 3 lysine 4 trimethylation (H3K4me3) (Auton et al. 2013; Baker et al. 2015; Marand et al. 2017; Smagulova et al. 2011).

In taxa where hotspots are evolutionarily conserved, hotspots are highly enriched around TSSs (Auton et al. 2013; Kawakami et al. 2017; Pan et al. 2011; Singhal et al. 2015; Tischfield and Keeney 2012).This pattern could be due to either higher selective constraints at TSSs or the chromatin structure at TSSs. TSSs are often under purifying selection and if a genomic feature, like a DNA motif, is targeting hotspots to these regions, these features would also be preserved through purifying selection, maintaining the location of the hotspot (Kawakami et al. 2017; Lam and Keeney 2015; Singhal et al. 2015; Tsai et al. 2010). On the other hand, an open chromatin conformation could be driving this pattern. TSSs and the surrounding regions must be accessible for transcription to occur while also providing sites for Spo11 to bind, initiating recombination (Lee et al. 2004; Pokholok et al. 2005) as Spo11 will create double strand breaks at any sites with accessible chromatin (Celerin et al. 2000; Ohta et al. 1994; Pan et al. 2011). In Lake Washington and Puget Sound populations, we found some enrichment of hotspots at TSSs (Lake Washington: 26% of hotspots fell within 3 kb of a TSS; Puget Sound: 29% of hotspots fell within 3 kb of a TSS) which is similar to hotspot enrichment around TSS in taxa that do not have a functional PRDM9 protein. In birds and dogs, for example, ~20-30% of hotspots overlap with TSSs (Auton et al. 2013; Kawakami et al. 2017; Singhal et al. 2015). Additional characterization is needed to

683 determine if hotspots in threespine stickleback are occurring in regions of the genome that are
684 already open due to transcription or if there is a mechanism that creates accessible chromatin
685 specifically for double strand break formation, like what is believed to occur with PRDM9 in
686 mammalian species (Diagouraga et al. 2018; Hayashi et al. 2005; Powers et al. 2016).
687 In some mammalian systems, positive selection acting on the zinc finger binding domain
688 of PRMD9 has led to multiple distinct DNA binding motifs between closely related species
689 (Baudat et al. 2010; Myers et al. 2010; Myers et al. 2008; Pratto et al. 2014). This leads to a rapid
690 evolution of hotspot localization (Baker et al. 2015; Brick et al. 2012; Pratto et al. 2014;
691 Smagulova et al. 2016; Stevison et al. 2015). Typically, ~40% of hotspots will contain a PRDM9
692 motif in mouse and humans (Baudat et al. 2010; Myers et al. 2008). In threespine stickleback
693 fish, we found that less than 14% of hotspots had PRMD9 or any other PRDM motifs, contrary
694 to what we would expect if PRDM9 was controlling hotspot location in threespine stickleback.
695 In addition, we did not find any other DNA motifs enriched in hotspots that would indicate a role
696 of an alternative DNA-binding protein that could localize hotspots. Additional characterization is
697 needed to understand what genomic features could be targeting hotspots, leading to the distinct
698 fine-scale recombination landscapes observed between populations of threespine stickleback
699 fish.
700

701 **Acknowledgements**

702 This work was supported by the Office of the Vice President of Research at the
703 University of Georgia (M.A.W); and the National Institute of Health (T32GM007103 to A.F.S).
704 We thank Robert Schmitz for help with library preparations and Catherine Peichel for providing
705 population samples. Sequencing was conducted at the Georgia Genomics and Bioinformatics
706 Core at the University of Georgia. Raw sequences are deposited in NCBI's Short Read Archive,
707 reference number SRP137809 (https://submit.ncbi.nlm.nih.gov/subs/sra/SUB3748706/overview)
708 and custom scripts can be found on Open Science Framework under A.F.S.
709 (https://osf.io/dezug/).

References

Auton A, McVean G 2007. Recombination rate estimation in the presence of hotspots. Genome Res 17: 1219-1227. doi: 0.1101/gr.6386707

Auton A, et al. 2013. Genetic recombination is targeted towards gene promoter regions in dogs. PLoS Genet 9: e1003984. doi: 10.1371/journal.pgen.1003984

Bailey TL, Eklan C 1994. Fitting a mixture model by expectation maximization to discover motifs in biopolymers. Proceedings of the Second International Conferene on Intelligent Systems for Molecular Biology: 28-36.

Baker CL, et al. 2015. PRDM9 drives evolutionary erosion of hotspots in *Mus musculus* through haplotype-specific initiation of meiotic recombination. PLoS Genet 11: e1004916. doi: 10.1371/journal.pgen.1004916

Baker Z, et al. 2017. Repeated losses of PRDM9-directed recombination despite the conservation of PRDM9 across vertebrates. Elife 6: e24133. doi: 10.7554/eLife.24133

Barton AB, Pekosz MR, Kurvathi RS, Kaback DB 2008. Meiotic recombination at the ends of chromosomes in *Saccharomyces cerevisiae*. Genetics 179: 1221-1235. doi: 10.1534/genetics.107.083493

Baudat F, et al. 2010. PRDM9 is a major determinant of meiotic recombination hotspots in humans and mice. Science 327: 836-840.

Baudat F, de Massy B 2007. Regulating double-stranded DNA break repair towards crossover or non-crossover during mammalian meiosis. Chromosome Res 15: 565-577. doi: 10.1007/s10577-007-1140-3

Bell MA, Foster SA. 1994. The evolutionary biology of the threespine stickleback: Oxford ; New York : Oxford University Press, 1994.

Berner D, Roesti M 2017. Genomics of adaptive divergence with chromosome-scale heterogeneity in crossover rate. Mol Ecol 26: 6351-6369. doi: 10.1111/mec.14373

Billings T, et al. 2013. DNA binding specificities of the long zinc-finger recombination protein PRDM9. Genome Biol 14: R35. doi: 10.1186/gb-2013-14-4-r35

Brick K, et al. 2012. Genetic recombination is directed away from functional genomic elements in mice. Nature 485: 642-645. doi: 0.1038/nature11089

Broman KW, et al. 1998. Comprehensive human genetic maps: individual and sex-specific variation in recombination. Am J Hum Genet 63: 861-869.

Brunschwig H, et al. 2012. Fine-scale maps of recombination rates and hotspots in the mouse genome. Genetics 191: 757-764. doi: 10.1534/genetics.112.141036

Burt A 2000. Perspective: Sex, recombination, and the efficacy of selection - was Weismann right? Evolution 54: 337-351.

Campbell CL, et al. 2016. A pedigree-based map of recombination in the domestic dog genome. G3-Genes Genom Genet 6: 3517-3524. doi: 10.1534/g3.116.034678

Celerin M, et al. 2000. Multiple roles of Spo11 in meiotic chromosome behavior. EMBO J 19: 2739-2750.

Chan AH, Jenkins PA, Song YS 2012. Genome-wide fine-scale recombination rate variation in *Drosophila melanogaster*. PLoS Genet 8: e1003090. doi: 10.1371/journal.pgen.1003090

Charlesworth B 2009. Fundamental concepts in genetics: effective population size and patterns of molecular evolution and variation. Nat Rev Genet 10: 195-205. doi: 10.1038/nrg2526

Coop G, Przeworski M 2007. An evolutionary view of human recombination. Nat Rev Genet 8: 23-34. doi: 10.1038/nrg1947

Danecek P, et al. 2011. The variant call format and VCFtools. Bioinformatics 27: 2156-2158. doi: 10.1093/bioinformatics/btr330

Dapper AL, Payseur BA 2017. Effects of demographic history on the detection of recombination hotspots from linkage disequilibrium. Mol Biol Evol 35: 335-353. doi: 10.1093/molbev/msx272

Davis L, Smith GR 2001. Meiotic recombination adn chromosome segregation in *Schizosaccharomyces pombe*. Proc Natl Acad Sci U S A 98: 8395-8402.

de Massy B 2013. Initiation of meiotic recombination: how and where? Conservation and specificities among eukaryotes. Annu Rev Genet 47: 563-599. doi: 10.1146/annurev-genet-110711-155423

DeFaveri J, Merila J 2015. Temporal stability of genetic variability and differentiation in the three-spined stickleback (*Gasterosteus aculeatus*). PLoS One 10: e0123891. doi: 10.1371/journal.pone.0123891

Delaneau O, et al. 2013. Haplotype estimation using sequencing reads. Am J Hum Genet 93: 687-696. doi: 10.1016/j.ajhg.2013.09.002

Diagouraga B, et al. 2018. PRDM9 methyltransferase activity is essential for meiotic DNA double-strand break formation at its binding sites. Mol Cell 69: 853-865. doi: 10.1016/j.molcel.2018.01.033

Drouaud J, et al. 2006. Variation in crossing-over rates across chromosome 4 of *Arabidopsis thaliana* reveals the presence of meiotic recombination "hot spots". Genome Res 16: 106-114. doi: 10.1101/gr.4319006

Dumont BL, Payseur BA 2008. Evolution of the genomic rate of recombination in mammals. Evolution 62: 276-294. doi: 10.1111/j.1558-5646.2007.00278.x

Dumont BL, Payseur BA 2011. Genetic analysis of genome-scale recombination rate evolution in house mice. PLoS Genet 7: e1002116. doi: 10.1371/

Duret L, Arndt PF 2008. The impact of recombination on nucleotide substitutions in the human genome. PLoS Genet 4: e1000071. doi: 10.1371/

Ferchaud A-L, Hansen MM 2016. The impact of selection, gene flow and demographic history on heterogeneous genomic divergence: three-spine sticklebacks in divergent environments. Mol Ecol 25: 238-259. doi: 10.1111/mec.13399

Fledel-Alon A, et al. 2009. Broad-scale recombination patterns underlying proper disjunction in humans. PLoS Genet 5: e1000658. doi: 10.1371/journal.pgen.1000658

Froenicke L, Anderson LK, Wienberg J, Ashley T 2002. Male mouse recombination maps for each autosome identified by chromosome painting. Am J Hum Genet 71: 1353-1368.

Fryxell KJ, Moon W-J 2005. CpG mutation rates in the human genome are highly dependent on local GC content. Mol Biol Evol 22: 650-658. doi: 10.1093/molbev/msi043

Glazer AM, et al. 2015. Genome assembly improvement and mapping convergently evolved skeletal traits in sticklebacks with genotyping-by-sequencing. G3-Genes Genom Genet 5: 1463-1472. doi: 10.1534/g3.115.017905

Gow JL, Peichel CL, Taylor EB 2006. Contrasting hybridization rates between sympatric three-spined sticklebacks highlight the fragility of reproductive barriers between evolutionarily young species. Mol Ecol 15: 739-752. doi: 10.1111/j.1365-294X.2006.02825.x

Grant CE, Bailey TL, Noble WS 2011. FIMO: scanning for occurrences of a given motif. Bioinformatics 27: 1017-1018. doi: 10.1093/bioinformatics/btr064

Guillon H, de Massy B 2002. An initiation site for meiotic crossing-over and gene conversion in the mouse. Nat Genet 32: 296-299. doi: 10.1038/ng990

Guo B, et al. 2015. Population genomic evidence for adaptive differentiation in Baltic Sea three-spined sticklebacks. BMC Biol 13: 19. doi: 10.1186/s12915-015-0130-8

Hayashi K, Yoshida K, Matsui Y 2005. A histone H3 methyltransferase controls epigenetic events required for meiotic prophase. Nature 438: 374-378. doi: 10.1038/nature04112

Hellsten U, et al. 2013. Fine-scale variation in meiotic recombination in *Mimulus* inferred from population shotgun sequencing. PNAS 110: 19478-19482. doi: 10.1073/pnas.1319032110

Hinch AG, et al. 2014. Recombination in the human pseudoautosomal region PAR1. PLoS Genet 10: e1004503. doi: 10.1371/journal.pgen.1004503

Hohenlohe PA, et al. 2010. Population genomics of parallel adaptation in threespine stickleback using sequenced RAD tags. PLoS Genet 6: e1000862. doi: 10.1371/journal.pgen.1000862

Jeffreys AJ, Kauppi L, Neumann R 2001. Intensely punctate meiotic recombination in the class II region of the major histocompatibility complex. Nat Genet 29: 217-222.

Jeffreys AJ, Murray J, Neumann R 1998. High-resolution mapping of crossovers in human sperm defines a minisatellite-associated recombination hotspot. Mol Cell 2: 267-273.

Jeffreys AJ, et al. 2005. Human recombination hot spots hidden in regions of strong marker association. Nat Genet 37: 601-606. doi: 10.1038/ng1565

Johnston HR, Cutler DJ 2012. Population demographic history can cause the appearance of recombination hotspots. Am J Hum Genet 90: 774-783. doi: 10.1016/j.ajhg.2012.03.011

Jones FC, et al. 2012. The genomic basis of adaptive evolution in threespine sticklebacks. Nature 484: 55-61. doi: 10.1038/nature10944

Kaback DB, Guacci V, Barber D, Mahon JW 1992. Chromosome size-dependent control of meiotic recombination. Science 256: 228-232.

Kauppi L, Jasin M, Keeney S 2012. The tricky path to recombining X and Y chromosomes in meiosis. Ann N Y Acad Sci 1267: 18-23. doi: 10.1111/j.1749-6632.2012.06593.x

Kawahara R, et al. 2009. Stickleback phylogenies resolved: evidence from mitochondrial genomes and 11 nuclear genes. Mol Phylogenet Evol 50: 401-444. doi: 10.1016/j.ympev.2008.10.014

Kawakami T, et al. 2017. Whole-genome patterns of linkage disequilibrium across flycatcher populations clarify the causes and consequences of fine-scale recombination rate variation in birds. Mol Ecol 26: 4158-4172. doi: 10.1111/mec.14197

Kent WJ 2002. BLAT - The BLAST-like alignment tool. Genome Res 12: 656-664. doi: 10.1101/

Kitano J, Mori S, Peichel CL 2007. Phenotypic divergence and reproductive isolation between sympatric forms of Japanese threespine sticklebacks. Biol J Linn Soc 91: 671-685.

Kon N, et al. 1997. Transcription factor Mts1/Mts2 (Atf1/Pcr1, Gad7/Pcr1) activates the *M26* meiotic recombination hotspot in *Schizosaccharomyces pombe*. Proc Natl Acad Sci U S A 94: 13765-13770.

Kong A, et al. 2002. A high-resolution recombination map of the human genome. Nat Genet 31: 241-247. doi: 10.1038/ng917

Lam I, Keeney S 2015. Nonparadoxical evolutionary stability of the recombination initiation landscape in yeast. Science 350: 932-937.

Langmead B, Salzberg SL 2012. Fast gapped-read alignment with Bowtie 2. Nature Methods 9: 357-359. doi: 0.1038/NMETH.1923

Lee C-K, et al. 2004. Evidence for nucleosome depletion at active regulatory regions genome-wide. Nat Genet 36: 900-905. doi: 10.1038/ng1400

Lesecque Y, Mouchiroud D, Duret L 2013. GC-biased gene conversion in yeast is specifically associated with crossovers: molecular mechanisms and evolutionary significance. Mol Biol Evol 30: 1409-1419. doi: 10.1093/molbev/mst056

Li H, Durbin R 2011. Inference of human population history from individual whole-genome sequences. Nature 475: 493-496. doi: 10.1038/nature10231

Li H, et al. 2009. The sequence alignment/map format and SAMtools. Bioinformatics 25: 2078-2079. doi: 10.1093/bioinformatics/btp352

Liu S, Hansen MM 2017. PSMC (pairwise sequentially Markovian coalescent) analysis of RAD (restriction site associated DNA) sequencing data. Mol Ecol Resour 17: 631-641. doi: 10.1111/1755-0998.12606

Liu S, Hansen MM, Jacobsen MW 2016. Region-wide and ecotype-specific differences in demographic histories of threespine stickleback populations, estimated from whole genome sequences. Mol Ecol 25: 5187-5202. doi: 10.1111/mec.13827

Makinen HS, Cano JM, Merila J 2006. Genetic relationships among marine and freshwater populations of the European three-spined stickleback (*Gasterosteus aculeatus*) revealed by microsatellites. Mol Ecol 15: 1519-1534. doi: 10.1111/j.1365-294X.2006.02871.x

Marais G 2003. Biased gene conversion: implications for genome and sex evolution. Trends Genet 19: 330-338. doi: 10.1016/s0168-9525(03)00116-1

Marand AP, et al. 2017. Meiotic crossovers are associated with open chromatin and enriched with *Stowaway* transposons in potato. Genome Biol 18: 203. doi: 10.1186/s13059-017-1326-8

Mather K 1936. The determination of position in crossing-over. J Genet 33: 207-235.

McVean GAT, et al. 2004. The fine-scale structure of recombination rate variation in the human genome. Science 304: 581-584.

Meunier J, Duret L 2004. Recombination drives the evolution of GC-content in the human genome. Mol Biol Evol 21: 984-990. doi: 10.1093/molbev/msh070

Moen T, et al. 2008. A linkage map of the Atlantic salmon (*Salmo salar*) based on EST-derived SNP markers. BMC Genomics 9: 223. doi: 10.1186/1471-2164-9-223

Morgan AP, et al. 2017. Structural variation shapes the landscape of recombination in mouse. Genetics 206: 603-619. doi: 10.1534/genetics.116.197988

Mugal CF, Weber CC, Ellegren H 2015. GC-biased gene conversion links the recombination landscape and demography to genomic base composition: GC-biased gene conversion drives genomic base composition across a wide range of species. Bioessays 37: 1317-1326. doi: 10.1002/bies.201500058

Myers S, et al. 2005. A fine-scale map of recombination rates and hotspots across the human genome. Science 310: 321-324.

Myers S, et al. 2010. Drive against hotspot motifs in primates implicates the PRDM9 gene in meiotic recombination. Science 327: 876-879.

Myers S, et al. 2008. A common sequence motif associated with recombination hot spots and genome instability in humans. Nat Genet 40: 1124-1129. doi: 10.1038/ng.213

Myers S, et al. 2006. The distribution and causes of meiotic recombination in the human genome. Biochem Soc T 34: 526-530.

Ohta K, Shibata T, Nicolas A 1994. Changes in chromatin structure at recombination initiation sites during yeast meiosis. EMBO J 13: 5754-5763.

Oliver PL, et al. 2009. Accelerated evolution of the *Prdm9* speciation gene across diverse metazoan taxa. PLoS Genet 5: e1000753. doi: 10.1371/journal.pgen.1000753

Orti G, Bell MA, Reimchen TE, Meyer A 1994. Global survey of mitochrondrial DNA sequences in the threespine stickleback: evidence for recent migrations. Evolution 48: 608-622.

Ostlund-Nilsson S, Mayer I, Huntingford FA. 2007. The Biology of the Threespine Stickleback. Boca Raton, FL: CRC Press.

Otto SP, Lenormand T 2002. Resolving the paradox of sex and recombination. Nat Rev Genet 3: 252-261. doi: 10.1038/nrg761

Otto SP, et al. 2011. About PAR: the distinct evolutionary dynamics of the pseudoautosomal region. Trends Genet 27: 358-367. doi: 10.1016/j.tig.2011.05.001

Pan J, et al. 2011. A hierarchical combination of factors shapes the genome-wide topography of yeast meiotic recombination initiation. Cell 144: 719-731. doi: 10.1016/j.cell.2011.02.009

Parvanov ED, Petkov PM, Paigen K 2010. *Prdm9* control activation of mammalian recombination hotspots. Science 327: 835.

Peichel CL, et al. 2001. The genetic architecture of divergence between threespine stickleback species. Nature 414: 901-905.

Peichel CL, et al. 2004. The master sex-determination locus in threespine sticklebacks is on a nascent Y chromosome. Curr Biol 14: 1416-1424. doi: 10.1016/j.cub.2004.08.030

Persikov AV, Osada R, Singh M 2009. Predicting DNA recognition by Cys2His2 zinc finger proteins. Bioinformatics 25: 22-29. doi: 10.1093/bioinformatics/btn580

Persikov AV, Singh M 2014. De novo prediction of DNA-binding specificities for Cys2His2 zinc finger proteins. Nucleic Acids Res 42: 97-108. doi: 10.1093/nar/gkt890

Petes TD 2001. Meiotic Recombination Hotspots and Coldspots. Nat Rev 2: 360-369.

Pfeifer B, Wittelsburger U, Ramos-Onsins SE, Lercher MJ 2014. PopGenome: An efficient Swiss army knife for population genomic analyses in R. Mol Biol Evol 31: 1929-1936. doi: 10.1093/molbev/msu136

Pokholok DK, et al. 2005. Genome-wide map of nucleosome acetylation and methylation in yeast. Cell 122: 517-527.

Powers NR, et al. 2016. The meiotic recombination activator PRDM9 trimethylates both H3K36 and H3K4 at recombination hotspots *in vivo*. PLoS Genet 12: e1006146. doi: 10.1371/journal.pgen.1006146

Pratto F, et al. 2014. Recombination initiation maps of individual human genomes. Science 346: 1256442. doi: 10.1126/science.1256442

Raj A, Stephens M, Pritchard JK 2014. fastSTRUCTURE: variational inference of population structure in large SNP data sets. Genetics 197: 573-589. doi: 10.1534/genetics.114.164350

Ravinet M, et al. 2018. The genomic landscape at a late stage of stickleback speciation: High genomic divergence interspersed by small localized regions of introgression. PLoS Genet 14: e1007358. doi: 10.1371/journal.pgen.1007358

Roesti M, Moser D, Berner D 2013. Recombination in the threespine stickleback genome--patterns and consequences. Mol Ecol 22: 3014-3027. doi: 10.1111/mec.12322

Sardell JM, et al. 2018. Sex differences in recombination in sticklebacks. G3-Genes Genom Genet. doi: 10.1534/g3.118.200166

See DR, et al. 2006. Gene evolution at the ends of wheat chromosomes. Proc Natl Acad Sci U S A 103: 4162-4167.

Serre D, Nadon R, Hudson TJ 2005. Large-scale recombination rate patterns are conserved among human populations. Genome Res 15: 1547-1552. doi: 10.1101/

Singer A, et al. 2001. Sex-specific recombination rates in zebrafish (*Danio rerio*). Genetics 160: 649-657.

Singhal S, et al. 2015. Stable recombination hotspots in birds. Science 350: 928-932.

Smagulova F, et al. 2016. The evolutionary turnover of recombination hot spots contributes to speciation in mice. Gene Dev 30: 266-280. doi: 10.1101/gad.270009

Smagulova F, et al. 2011. Genome-wide analysis reveals novel molecular features of mouse recombination hotspots. Nature 472: 375-378. doi: 10.1038/nature09869

Smukowski Heil CS, Ellison C, Dubin M, Noor MAF 2015. Recombining without hotspots: a comprehensive evolutionary portrait of recombination in two closely related species of *Drosophila*. Genome Biol Evol 7: 2829-2842. doi: 10.1093/gbe/evv182

Steiner WW, Schreckhise RW, Smith GR 2002. Meiotic DNA breaks at the *S.pombe* recombination hotspots *M26*. Mol Cell 9: 847-855.

Stevison LS, et al. 2015. The time scale of recombination rate evolution in great apes. Mol Biol Evol 33: 928-945.

Sueoka N 1962. On the genetic basis of variation and heterogeneity of DNA base composition. Proc Natl Acad Sci U S A 48: 582-592.

Tajima F 1989. Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. Genetics 123: 585-595.

Tischfield SE, Keeney S 2012. Scale matters: the spatial correlation of yeast meiotic DNA breaks with histone H3 trimethylation is driven largely by independent colocalization at promoters. Cell Cycle 11: 1496-1503. doi: 10.4161/cc.19733

Tsai IJ, Burt A, Koufopanou V 2010. Conservation of recombination hotspots in yeast. Proc Natl Acad Sci U S A 107: 7847-7852. doi: 10.1073/pnas.0908774107

Van der Auwera GA, et al. 2013. From FastQ data to high confidence variant calls: the Genome Analysis Toolkit best practices pipeline. Curr Protoc Bioinformatics 43: 11.10.11-33. doi: 10.1002/0471250953.bi1110s43

Wall JD, Stevison LS 2016. Detecting recombination hotspots from patterns of linkage disequilibrium. G3-Genes Genom Genet 6: 2265-2271. doi: 10.1534/g3.116.029587

Weber CC, et al. 2014. Evidence for GC-biased gene conversion as a driver of between-lineage differences in avian base composition. Genome Biol 15: 549-565.

Webster MT, Hurst LD 2012. Direct and indirect consequences of meiotic recombination: implications for genome evolution. Trends Genet 28: 101-109. doi: 10.1016/j.tig.2011.11.002

White MA, Kitano J, Peichel CL 2015. Purifying selection maintains dosage-sensitive genes during degeneration of the threespine stickleback Y chromosome. Mol Biol Evol 32: 1981-1995. doi: 10.1093/molbev/msv078

Wootton RJ. 1976. *The biology of the sticklebacks*: London ; New York : Academic Press.

Wright AE, Dean R, Zimmer F, Mank JE 2016. How to make a sex chromosome. Nat Commun 7: 12087. doi: 0.1038/ncomms12087

Yoshida K, et al. 2014. Sex chromosome turnover contributes to genomic divergence between incipient stickleback species. PLoS Genet 10: e1004223. doi: 10.1371/journal.pgen.1004223

Table 1. Mean equilibrium GC content (± SE)

	Lake Washington	Puget Sound
Top 5% of Recombination Rates	0.432 (±0.0006)[a]	0.430 (±0.0006)[a]
Bottom 5% of Recombination Rates	0.421 (±0.0005)[b]	0.415 (±0.0005)[b]
Population-Specific Hotspots	0.420 (±0.002)[b]	0.416 (±0.001)[b]
Shared Hotspots	0.422 (±0.004)[a,b]	0.417 (±0.004)[b]

[a,b]Groups significantly different within populations by Wilcoxon Rank Test; $p < 0.05$

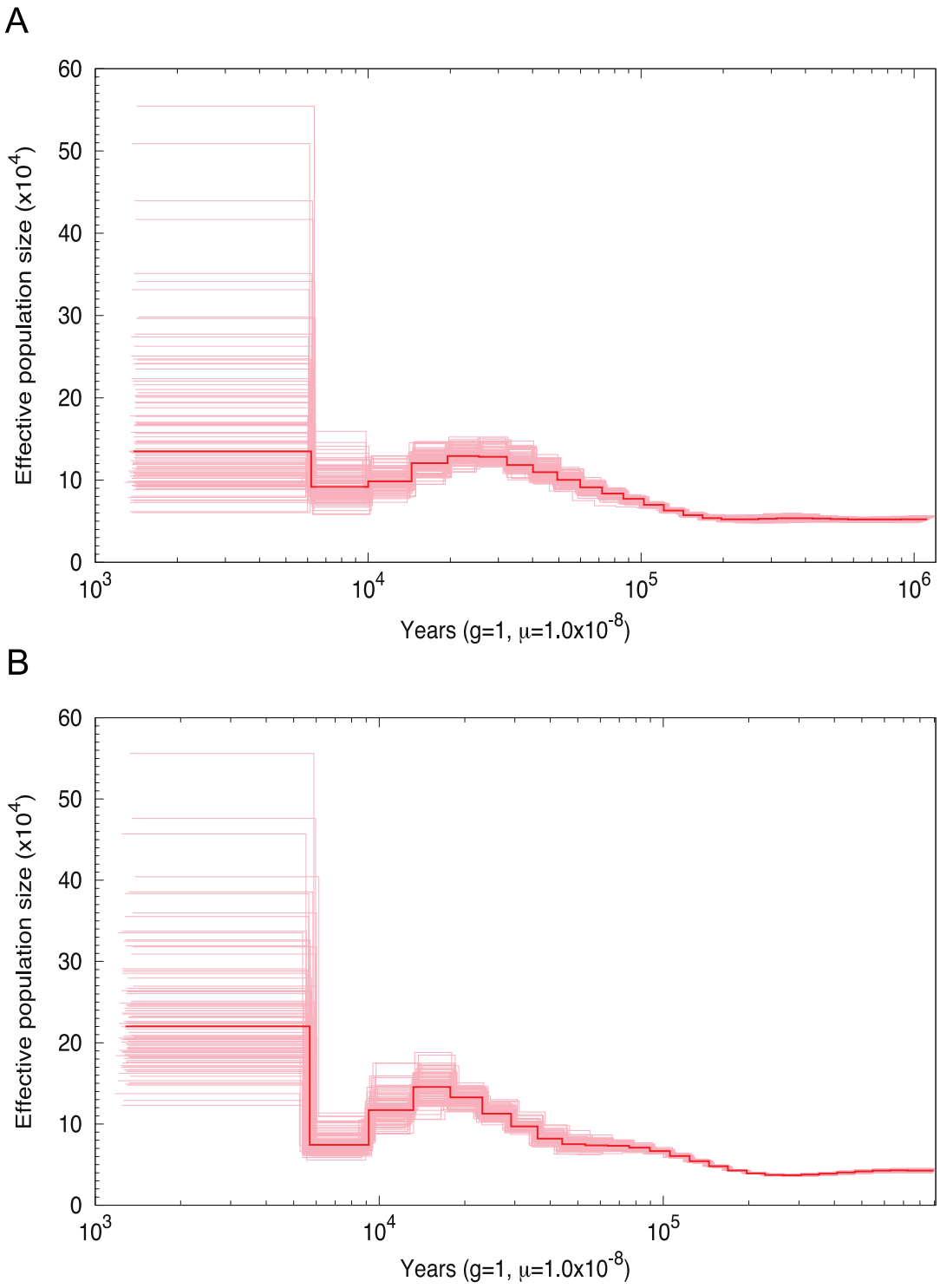

Figure 1. Lake Washington and Puget Sound have experienced past population bottlenecks. Demographic history for Lake Washington (A) and Puget Sound (B) was estimated using pairwise sequential Markov coalescent (PSMC) from a single female fish from each population. 100 bootstrap replicates around the estimated history are shown.

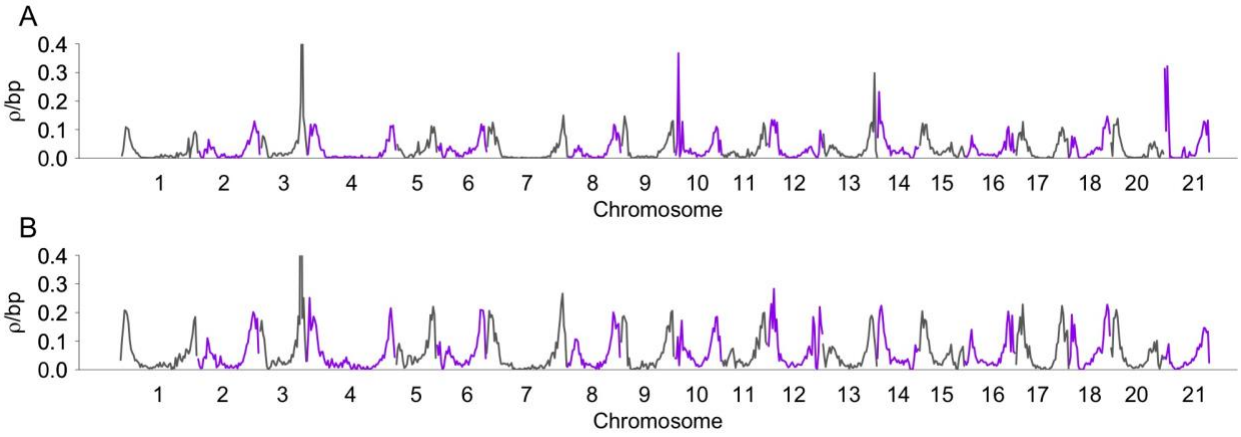

Figure 2. Recombination rates are similar at a broad scale in each population. Mean recombination rates were estimated using LDHelmet in non-overlapping 500 kb windows for each autosome in Lake Washington (A) and Puget Sound (B). Centromere positions are shown in Supplemental Figures 3 and 4. Transitions between gray and purple indicate different chromosomes.

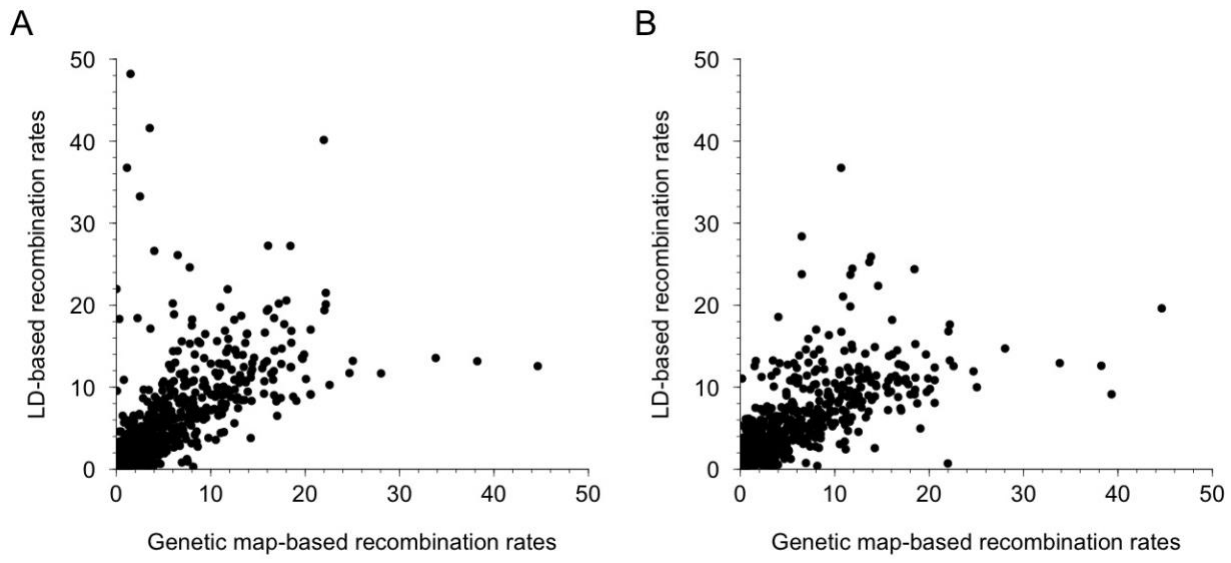

Figure 3. LD-based estimates of recombination rates are highly correlated with estimates from genetic linkage maps. Population-scaled recombination rates were converted to cM/Mb. There is a significant positive correlation in Lake Washington (Spearman's rank correlation; $r = 0.830$; $p < 0.001$) (A) and Puget Sound (Spearman's rank correlation; $r = 0.810$; $p < 0.001$) (B) between LD-based recombination rates and genetic map-based recombination rates.

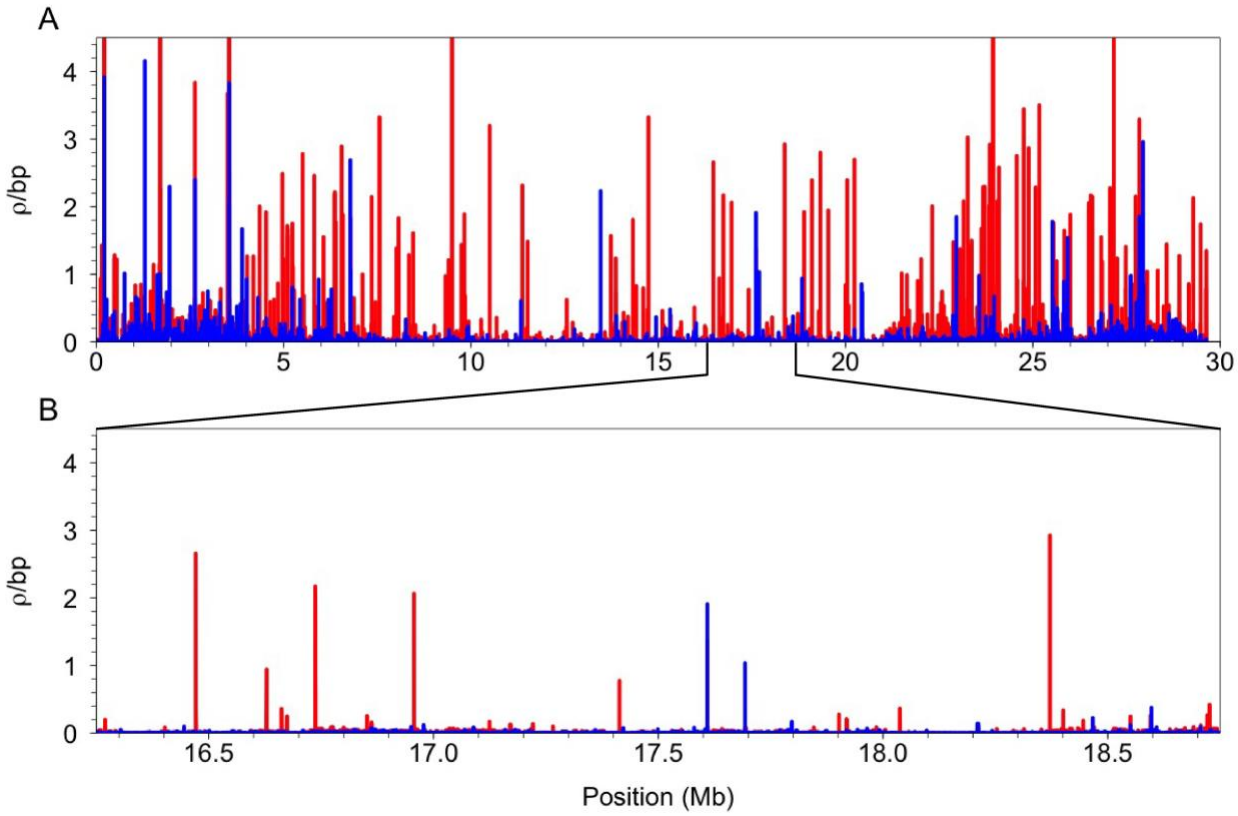

Figure 4. Recombination rates vary at a fine-scale across chromosome one. Population-scaled recombination rates across chromosome one are shown for Puget Sound (red) and Lake Washington (blue) (A). A subset of chromosome one is shown to highlight population-specific peaks of recombination across a narrow 2.5 Mb region (B). Only recombination rates below 4.5 ρ/bp are shown. The remaining chromosome plots are in supplemental figures 3 and 4.

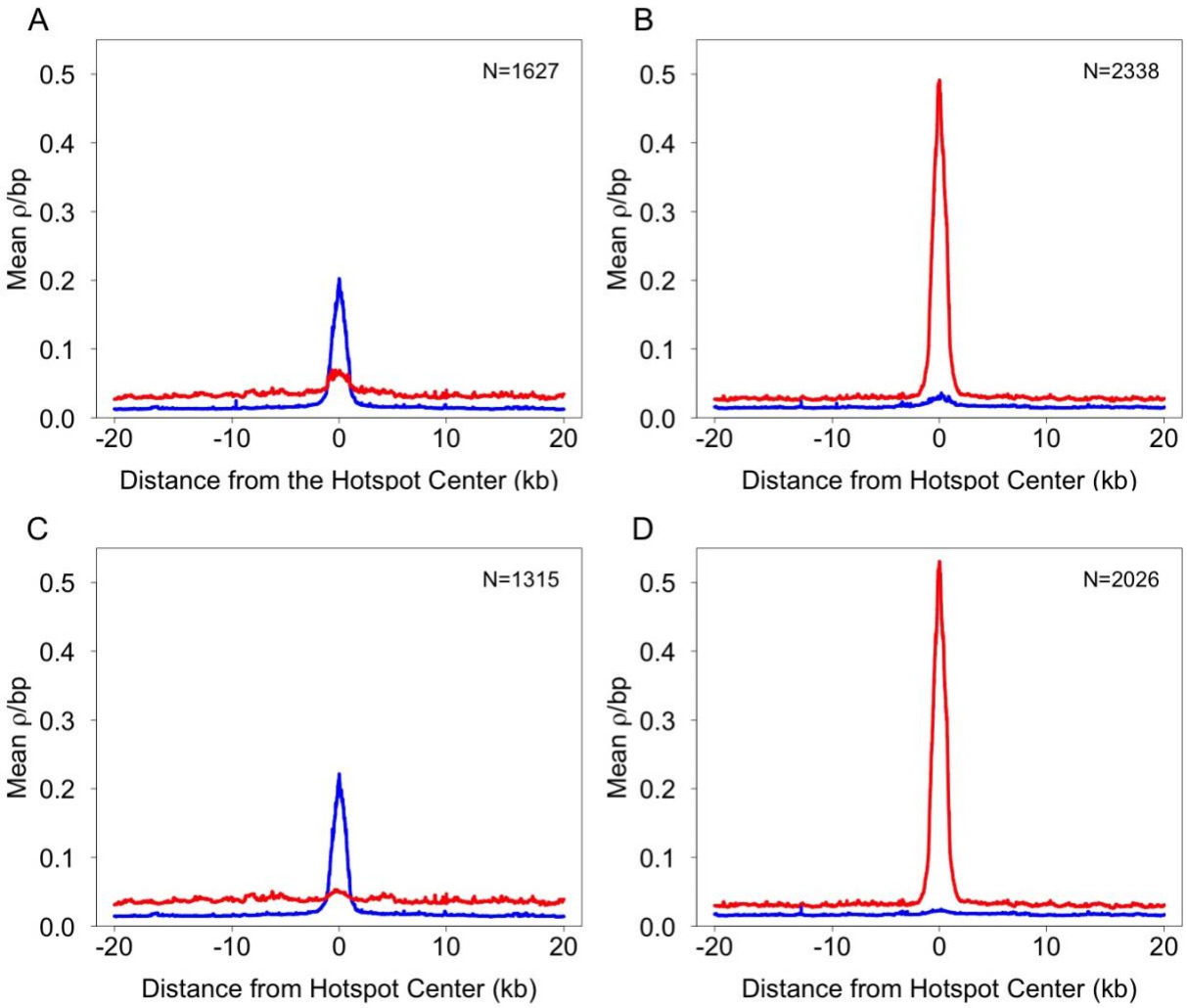

Figure 5. LD-based recombination rates around hotspots are population-specific. Mean recombination rates are shown across a 40 kb interval, flanking the center of hotspots. The mean recombination rate in shared and population-specific Lake Washington hotspots is higher in the Lake Washington population compared to the homologous regions in the Puget Sound population (A). The mean recombination rate in shared and population-specific Puget Sound hotspots are higher in the Puget Sound population compared to the homologous regions in the Lake Washington population (B). The pattern is more pronounced when shared hotspots are removed from the comparison, leaving only the population-specific hotspots (C and D). Puget Sound is shown in red and Lake Washington is shown in blue.

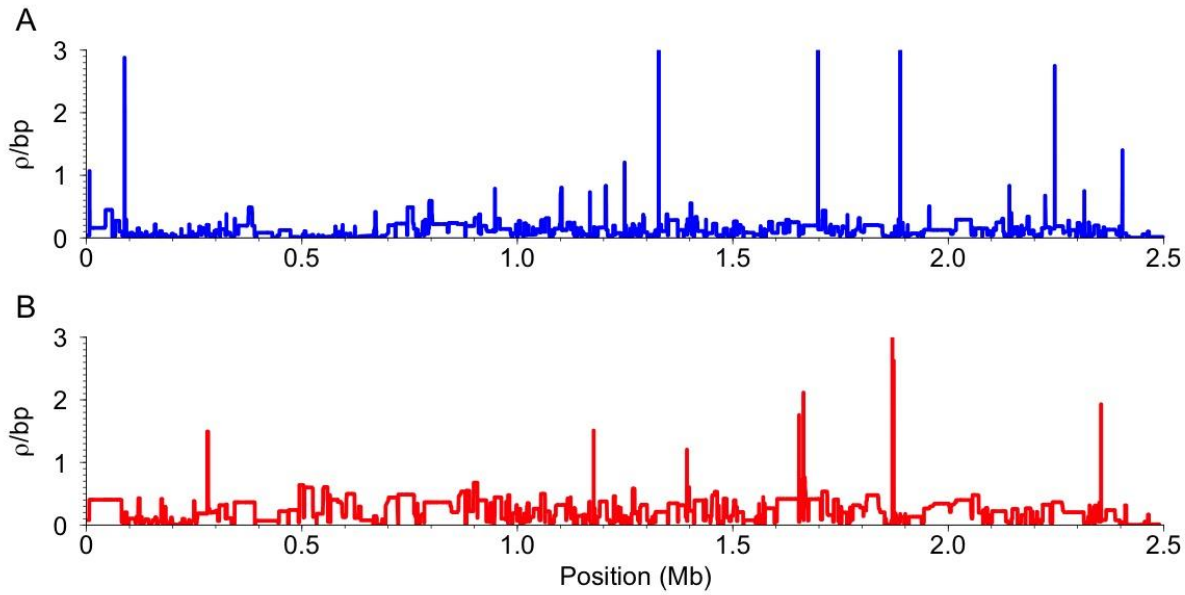

Figure 6. LD-based recombination rates are higher across the pseudoautosomal region (PAR). The PAR is the first ~2.5 Mb of linkage group 19. The Lake Washington (A) and Puget Sound (B) population-specific rates are shown separately. Overall, recombination rates are higher across the PAR than the autosomes (see Figure 4B) (Lake Washington PAR average: 0.129 ρ/bp; Lake Washington autosome-wide average: 0.035 ρ/bp; Puget Sound PAR average: 0.232 ρ/bp; Puget Sound autosome-wide average: 0.072 ρ/bp).