Received: 13 February 2017 | Revised: 2 June 2017 | Accepted: 13 June 2017

DOI: 10.1111/mec.14226

FROM THE COVER

WILEY MOLECULAR ECOLOGY

Gene flow and selection interact to promote adaptive divergence in regions of low recombination

Kieran Samuk[1] | Gregory L. Owens[2] | Kira E. Delmore[3] | Sara E. Miller[4] | Diana J. Rennison[5] | Dolph Schluter[1]

[1]Department of Zoology, Biodiversity Research Centre, University of British Columbia, Vancouver, BC, Canada

[2]Department of Botany, University of British Columbia, Vancouver, BC, Canada

[3]Max Planck Institute for Evolutionary Biology, Plön, Germany

[4]Department of Neurobiology and Behavior, Cornell University, Ithaca, NY, USA

[5]Institut fur Okologie und Evolution, Universitat Bern, Bern, Switzerland

Correspondence
Kieran Samuk, Biology, Duke University, Durham, NC, USA.
Email: ksamuk@gmail.com

Funding information
Natural Sciences and Engineering Research Council of Canada, Grant/Award Number: RGPIN 93037-11

Abstract

Adaptation to new environments often occurs in the face of gene flow. Under these conditions, gene flow and recombination can impede adaptation by breaking down linkage disequilibrium between locally adapted alleles. Theory predicts that this decay can be halted or slowed if adaptive alleles are tightly linked in regions of low recombination, potentially favouring divergence and adaptive evolution in these regions over others. Here, we compiled a global genomic data set of over 1,300 individual threespine stickleback from 52 populations and compared the tendency for adaptive alleles to occur in regions of low recombination between populations that diverged with or without gene flow. In support of theory, we found that putatively adaptive alleles (F_{ST} and d_{XY} outliers) tend to occur more often in regions of low recombination in populations where divergent selection and gene flow have jointly occurred. This result remained significant when we employed different genomic window sizes, controlled for the effects of mutation rate and gene density, controlled for overall genetic differentiation, varied the genetic map used to estimate recombination and used a continuous (rather than discrete) measure of geographic distance as proxy for gene flow/shared ancestry. We argue that our study provides the first statistical evidence that the interaction of gene flow and selection biases divergence toward regions of low recombination.

KEYWORDS
adaptation, gene flow, genomics, hybridization, population genetics—empirical

1 | INTRODUCTION

Understanding the genetic basis of adaptation is a fundamental goal of evolutionary biology. Yet, we still know little about the myriad interacting factors that determine the number, genomic location and effect size of loci underlying adaptive traits. Recent work suggests that interactions between two common evolutionary forces, natural selection and gene flow, may profoundly shape where adaptation occurs in the genome (Aeschbacher, Selby, Willis, & Coop, 2017; Kirkpatrick & Barton, 2006; Nachman & Payseur, 2012; Noor & Feder, 2006; Yeaman & Whitlock, 2011). When divergent selection and gene flow co-occur (hereafter "DS-GF"), hybridization between migrant and local individuals breaks down positive linkage disequilibrium (LD) between sets of locally adapted alleles, impeding adaptation (Kirkpatrick & Barton, 2006; Nachman & Payseur, 2012; Sousa & Hey, 2013). This decay of positive LD can be slowed if locally adapted alleles are tightly genetically linked, for example physically close on the same chromosome or occurring together in a region of low recombination (Navarro & Barton, 2003; Noor, Grams, Bertucci, & Reiland, 2001; Rieseberg, 2001; Yeaman & Whitlock, 2011). Accordingly, theory predicts that DS-GF will drive a tendency for locally adapted alleles to be tightly linked in the genome, either by physical proximity or by colocalization in regions of low recombination (Aeschbacher et al., 2017; Bürger & Akerman, 2011; Yeaman & Whitlock, 2011).

Recent studies have offered mixed support for this prediction. Roesti, Moser, and Berner (2013) and Marques et al. (2016) both

report that parapatric pairs of stickleback ecotypes exhibit elevated divergence in regions of low recombination suggesting that gene flow and selection may interact as predicted. In contrast, Renaut et al. (2013) and Burri et al. (2015) found no relationship between gene flow, selection and recombination in sunflowers and flycatchers, respectively.

However, definitively testing the prediction that gene flow and selection interact to promote divergence in regions of low recombination requires a system in which we can carry out replicated comparisons of the genomic distribution of adaptive alleles between populations with and without gene flow, and populations with and without divergent selection. This has not yet been possible, as previous studies have focused on a small number of populations (Marques et al., 2016; Renaut et al., 2013; Roesti et al., 2013). It is also necessary to disentangle the effects of selection and gene flow from other processes that can generate clustering of adaptive alleles. For example, linked selection—hitchhiking and background selection—is widely known to cause clustering of diverged loci (e.g., a single adaptive allele and surrounding linked neutral alleles), an effect that is amplified in regions of low recombination even in the absence of gene flow (Charlesworth, 2012; Cutter & Payseur, 2013). In addition, recombination may itself be mutagenic, which would result in decreased rates of divergence in regions of low recombination (Hairston, Ellner, Geber, Yoshida, & Fox, 2005; Nachman & Payseur, 2012). Isolating the effects of these various processes has thus far proved challenging (Burri et al., 2015; Renaut et al., 2013).

To approach this problem, we assembled a large population genomic data set derived from threespine sticklebacks (*Gasterosteus aculeatus*) from across the Northern Hemisphere (Figure S1, Table S1). Threespine sticklebacks are a Holarctic species of fish that have evolved into a variety of unique forms over the last 10,000 years (McKinnon & Rundle, 2002). Notably, the various forms of stickleback have evolved repeatedly in the presence and absence of gene flow (McKinnon & Rundle, 2002). This allows for statistical comparisons of the genomic distribution of adaptive alleles among groups of population pairs experiencing varying levels of divergent selection and gene flow. Here, we focused on comparing population pairs in which divergent selection occurs in the face of gene flow to population pairs experiencing selection alone, gene flow alone, or neither. Using this approach, we tested the theoretical prediction that when divergent selection and gene flow co-occur, adaptive alleles are more likely to fix in regions of low recombination and/or occur in tightly linked clusters throughout the genome.

2 | METHODS

2.1 | GitHub repository

The code used to generate our data set and perform the analyses described here is available on GitHub at https://github.com/ksamuk/gene_flow_linkage. All scripts were written in PERL or R 3.2.2 (R Core Team 2015).

2.2 | Data sources

The stickleback population genomic data sets used in this study came from two sources: online databases and new data from two of the authors. During the period from May to July 2014, we periodically searched the Short Read Archive (SRA), the European Nucleotide Archive (ENA) and the Databank of Japan Sequence Read Archive (DRA) for "threespined/three-spined/threespine/three-spine stickleback," "stickleback," "Gasterosteous aculeatus." We also searched for stickleback population genetic studies on Google Scholar using the same terms as above, with the inclusion of "genomic," "genome scan," "population genetic" and "genetics," and examined them for SRA/ENA/DRA accession numbers. Detailed information for all the populations included in the study is shown in Table S1 (Catchen et al., 2013; Chain et al., 2014; Feulner et al., 2015; Hohenlohe et al., 2010; Roesti, Hendry, Salzburger, & Berner, 2012; Yoshida et al., 2014).

In addition to previously published data, we prepared three new data sets from benthic/limnetic, freshwater lake and white/marine populations from various locations in Canada. The libraries for these data sets were prepared using a mix of Genotyping-by-Sequencing method of (Elshire et al., 2011) and whole-genome genomic DNA (TruSeq DNA PCR-Free Library Preparation Kit, Illumina, California). The collection locations and sequencing methods are listed in Table S1. The resultant GBS libraries were sequenced at the University of British Columbia Biodiversity Sequencing Centre, and the whole-genome libraries were sent for sequencing to Genome Quebec. Sequencing was performed on an Illumina Hi-Seq 2000 at both facilities. These data sets are available on the SRA (see Table S1).

2.3 | Variant identification and processing

We identified variants using a standard, reference-based bioinformatics pipeline (see Github code repository for details). After demultiplexing, we used TRIMMOMATIC v0.32 (Bolger, Lohse, & Usadel, 2014) to filter low-quality sequences and adapter contamination. We then aligned reads to the stickleback reference genome (BROAD S1), (Jones et al., 2012) using BWA v0.7.10 (Li & Durbin, 2010), followed by realignment with STAMPY v1.0.23 (Lunter & Goodson, 2011). We then followed the GATK v3.3.0 (DePristo et al., 2011; McKenna et al., 2010) best practices workflow, excluding the MarkDuplicates step for reduced representation data sets (RAD and GBS). We realigned reads around indels using RealignTargetCreator and IndelRealigner, identified variants in individuals using the HaplotypeCaller, and called SNPS in each data set using GenotypeGVCFs. The results were sent to a VCF file containing all variant and invariant sites and converted to tabular format.

2.4 | Calculation of divergence metrics

Our final data set included individuals from 56 unique populations. As there was no a priori reason to select only a subset pairs of populations in the analysis, we instead performed all possible pairwise

comparisons. We employ an unbiased significance testing method to overcome redundant use of populations in multiple pairs (see permutation test).

For each of the 1,128 pairwise comparisons, we calculated two divergence metrics: Weir and Cockerham's F_{ST} (Weir & Cockerham, 1984) and Nei's d_{XY} (Nei, 1987). We calculated F_{ST} at two scales: first, at each individual shared SNP; and second, averaged within 75-kilobase pair (kbp) windows. For all SNPs, we required: a minor allele frequency of at least 0.05 and coverage in at least five individuals per population. For windowed analysis, we required that windows contain at least three variable sites genotyped in at least five individuals per population. The distribution of total sequenced and total variable sites for all the comparisons is shown in Figure S10.

Window-averaged F_{ST} values were calculated by dividing the sum of the numerators of all SNP-wise F_{ST} estimates within a given window by the sum of their denominators. We calculated d_{XY} in 75-kbp windows, including all shared variant and invariant sites in the window. We required d_{XY} windows to contain more than 500 shared sequenced sites (i.e., nucleotides with a genotype call in both populations), because we found that the variance in d_{XY} greatly increases below this threshold. After calculating F_{ST} or d_{XY}, we classified SNPs and windows exhibiting extreme values as "outliers," defined as those in the 95th percentile or higher of F_{ST} or d_{XY}. Note, only d_{XY} window "outliers" were used because individual site d_{XY} scores are uninformative. All calculations were performed using CUSTOM PERL and R scripts (see Github repository).

2.5 | Classification of populations

For populations with multiple individuals (48 of the 56), we classified all pairwise comparisons between our 48 populations (n = 1,128 comparisons) along two axes: ecology and gene flow. We scored populations as ecologically "divergent" or "parallel" based on whether they (i) inhabited different ecosystems or ecological niches and/or (ii) had been directly identified by previous authors as ecologically divergent (Figure S1, see Table S1 for details). The correlation between divergent selection and ecology in stickleback is extremely well supported (Hendry, Bolnick, Berner, & Peichel, 2009; McKinnon & Rundle, 2002; Schluter, 1993) and while the strength of divergent selection may vary among comparisons, we believe this is a reasonable proxy.

Second, we scored whether there has been opportunity for gene flow between populations ("gene flow"/"allopatry"), based on geographic distance and barriers. This is a common assumption in comparative studies, and there is strong empirical evidence that this is a reasonable assumption for threespine sticklebacks. Extensive previous work suggests that nearby stickleback populations often interbreed (Hendry et al., 2009; Marques et al., 2016). This interbreeding leads to gene flow, as complete reproductive isolation is extremely rare among stickleback populations (Hendry et al., 2009; McKinnon & Rundle, 2002). Indeed, even the most highly differentiated populations (e.g., benthic vs. limnetic) experience ongoing gene flow (Gow, Peichel, & Taylor, 2006). In some cases, gene flow between nearby populations is opposed by divergent selection, limiting the number of loci affected by gene flow, although still allowing substantial gene flow in much of the genome (Jones et al., 2012; Roesti et al., 2012). In the cases where contemporary gene flow is unlikely (e.g., isolated lakes), the young age of post-glacial lakes (~5–10 kya) makes past gene flow between physically isolated but nearby populations likely (Schluter & Conte, 2009). Thus, the use of geographic isolation as a proxy for the opportunity (present or in last 5–10 ky years) for gene flow is likely reasonable for this species.

We thus considered any populations within 500 km of one another as having the potential for gene flow. We calculated geographic distance (great circle distance) between all pairs of populations using the function "earth.dist" from the R package FOSSIL (Vavrek, 2011). Note that this classifier is conservative, as it likely causes populations that are largely allopatric (DS-Allopatry) to be classified as DS-GF, decreasing the power to detect a difference between regimes.

Note that for both classification schemes, we are not assuming a perfect, discrete mapping of selection and gene flow onto individual populations. We only assume that when considered together, populations in each category will tend to exhibit greater (or less) gene flow and/or divergent selection. In total, our classification scheme resulted in the following number of comparisons: 130 divergent selection with gene flow, 31 parallel selection with gene flow, 113 parallel selection with gene flow and 821 divergent selection in allopatry.

2.6 | Addition of genomic variables

We measured three genomic variables in each 75-kbp window in the divergence data set with: recombination rate, mutation rate and gene density. Recombination rates (cM/MB) were obtained from a previously published high-density genetic map (Roesti et al., 2013). Where windows overlapped regions with different estimates of recombination rate, we assigned them an average of the two rates weighted by the degree of overlap.

We obtained estimates of mutation rate by estimating the synonymous substitution rate (d_S) in a phylogenetic framework. For neutral sites, d_S is an estimator of the primary mutation rate (Wielgoss et al., 2011). To do this, we used the R (version 3.2.2) package BIOMART to obtain a list of all annotated *G. aculeatus* coding DNA sequences (CDS) from ENSEMBL. For each *G. aculeatus* CDS, we queried ENSEMBL for all homologous CDS from three other fish species: *Xiphophorus maculatus*, *Poecilia formosa* and *Oreochromis niloticus*. These species all have identical estimated divergence times from *G. aculeatus* (150 MYR). We aligned each set of homologous coding sequences using PRANK (Löytynoja & Goldman, 2008) and analysed the output using PAML (Branch model 2) to estimate d_S trees. We excluded trees with fewer than three species, in order to ensure that lineage-specific artefacts did not bias d_S estimates. We also excluded any individual branches where d_S exceeded 5 standard deviations of the distribution of the d_S values from all branches of every tree (values exceeding this threshold were categorically the

result of bad alignments). After filtering d_S trees, we used the R package APE (Paradis, Claude, & Strimmer, 2004) to calculate the mean pairwise branch distance between *G. aculeatus* and each other species in the tree. Because the other three species all have identical divergence times from *G. aculeatus*, this results in a single normalized value of d_S for each coding sequence. After obtaining all the mutation rate estimates, we assigned them to 75-kbp windows in the divergence data sets by averaging the d_S estimates for genes in each window (if any), weighted by the degree of overlap for each gene.

Estimates of gene density (number of genes overlapping the window) were calculated by querying ENSEMBL (Kautt, Elmer, & Meyer, 2012) for the physical position of all genes in the stickleback genome using BIOMART (Yang, 2007). We then wrote a custom R script (see Github repository) to count the number of genes in each 75-kbp window along the reference genome.

2.7 | Tendency for adaptive divergence in regions of low recombination

To quantify the tendency for outliers to occur in regions of low recombination in each comparison, we employed a linear modelling approach. Using the 75-kbp windows as data points, we fit a logistic regression model to each comparison data set using the following form: outlier status = recombination rate + mutation rate + gene density, where outlier status is 1 if a window is an outlier (>95th percentile) and 0 otherwise. We performed separate model fits for F_{ST} and d_{XY} outliers. We also fit models of the same type using mean intrapopulation heterozygosity (H_S) as the response variable to assess its role in driving any patterns of increased divergence.

We fit these models in R (version 3.2.2) using the generalized linear model function "glm." Prior to model fitting, we filtered out pairwise population comparisons with fewer than 100 75-kbp windows represented to ensure convergence of the linear models. To assess statistical significance of the model fits, we extracted the regression coefficient for the recombination rate term from each model, representing the slope of the relationship between outlier occurrence and recombination rate. The steepness of the slope coefficients estimates the tendency for outliers to occur in regions of low recombination, controlling for the effects of mutation rate and gene density.

2.8 | Permutation tests

To test the hypothesis that adaptation with gene flow favours divergence in regions of low recombination, we employed a permutation test to assess whether the slopes from the models described above differed significantly between populations differing in divergent selection and gene flow. To do this, we randomly shuffled regime assignments of all the populations and estimated the mean low recombination outlier tendency (the grouped mean of the regression coefficients from above) for each regime in 10,000 permutations. This generated a null distribution of mean slopes for each regime, accounting for sample size differences between categories (Figure S2). We then calculated a two-sided *p* value for each empirical mean by the computing the fraction of samples in the tail of the null distribution greater than the observed value and multiplying by two. Note this method of analysis also employed elsewhere throughout the study (referred to as a "permutation test" wherever it was applied).

2.9 | Clustering vs. geographic distance and overall divergence

To ensure our results were not influenced by our discrete geographic categorization scheme, we examined how the tendency for F_{ST} outliers to occur in regions of low recombination varied with pairwise geographic distance. To do this, we regressed the low recombination outlier tendency (regression coefficients from above) on geographic distance between populations using the R function "lm." The linear model was of the form recombination bias = distance + ecology + distance * ecology (interaction). We then assessed significance of the model terms using a permutation test similar to the one previously described.

The results of Burri et al. (2015) and Roesti et al. (2013) suggest that the tendency for F_{ST} outliers to occur in regions of low recombination may be highest at intermediate levels of overall genetic divergence (F_{ST} = 0.3–0.5). Overall F_{ST} thus represents a potential source of bias, as our use of geographic distance as a proxy for gene flow is naturally confounded with overall F_{ST}—with isolation by distance, more distant populations will have higher divergence, all else being equal. To test if this may have influenced our results, we examined the correlation between low-recombination clustering tendency and overall F_{ST}. To obtain overall F_{ST} estimates between each pair of populations, we divided the sum of the numerator terms by the sum of the denominator terms of all locus-specific F_{ST} values for each pair (Weir & Cockerham, 1984). This yielded a single average F_{ST} value for each pair of populations. We then employed the same approach as the analysis of distance, with a linear model the form recombination bias = F_{ST} + ecology + F_{ST} * ecology (interaction). We assess the significance of this difference again via permutation test (see code supplement).

2.10 | Increased clustering of outlier SNPs

To test the hypothesis that adaptation with gene flow favours clustering (reduced genetic map distance) between outlier SNPs, we used two metrics of clustering: nearest neighbour map distance between outliers (NND) and the coefficient of variation in map distance between consecutive outliers. Both of these metrics were calculated using the SNP-level data.

We first asked: Do map distances between nearest-neighbour outlier loci differ significantly from the expected map distances of identical numbers of nearest-neighbour SNPs? This approach was designed to account for disparities in SNP density that might occur due to differences in sequencing outcomes between our various data sets. To do this, we first partitioned each SNP data set by chromosome. Then, for each chromosome we identified the number of outlier loci using the previously described method. We then drew

10,000 samples of random SNPs from each chromosome equal to the number of outliers on that chromosome and calculated the mean map distance between each SNP and its nearest neighbour in the random sample. We then compared the empirical mean nearest neighbour map distance of outliers to this null distribution for each chromosome within each individual comparison data set. We then used permutation tests to compare (i) the proportion of chromosomes that were significantly overclustered and (ii) the difference between the average NND between outliers and the average NND expected between SNPs, in units of standard deviations, between the four selection and geneflow regimes.

In addition to the resampled approach, we also computed a coefficient of variation: the ratio of the standard deviation in map distances between consecutive SNP on the chromosome divided by the mean distance. Values exceeding one are indicative of overdispersion (clustering), whereas values below one suggest underdispersion (uniformity of distances). We calculated the coefficient of variation for outliers on each chromosome and computed the mean for all chromosomes containing outliers for each comparison. We then used a permutation test as described above to compare the means of this quantity among geneflow/selection regimes.

2.11 | Whole-genome data collection

We obtained whole-genome sequences from single individuals from a total of nine stickleback populations. One of these is the reference genome, derived from a freshwater individual from Bear Paw Lake, Alaska (Jones et al., 2012). Four were individuals collected from two pairs of populations that have diverged into benthic and limnetic ecotypes from Paxton and Priest Lake on Texada Island in BC, Canada. These two pairs of populations (one limnetic and one benthic in each lake) have diverged from each other in the face of gene flow (Taylor & McPhail, 2000), making them "DS-GF" populations in our classification scheme. The remaining five were collected from freshwater lakes with a single, nondiverged stickleback population—Hoggan, Bullock, Trout, Cranby and Stowell lakes (Miller, 2016). These latter populations diverged from the marine ancestor in allopatry—that is, they are "DS-Allopatry" populations in our scheme. DNA from these individuals was extracted via phenol–chloroform extraction, and whole-genome library preparation carried out using Nextera DNA Library Prep Kits (Illumina Inc.). All populations were sequenced on an Illumina HiSeq 2000 in the University of British Columbia Biodiversity Sequencing Facility.

2.12 | Whole-genome d_{XY} calculation and analysis

We used the GATK best practices workflow described above to call variants on the eight populations above (not including the reference). We emitted VCF files containing all variant and invariant sites for each population. We then computed d_{XY} in 75,000-base pair windows using the method described previously (see "Calculation of Divergence Metrics" above; code available in repository). For the two pairs of DS-GF populations (Paxton and Priest), we computed d_{XY} between sympatric populations within each lake. For the remaining DS-Allopatry populations, we computed d_{XY} between each population and a marine population (Bear Paw Lake, i.e., the reference genome). We allowed for missing sites, and for windows with no variable sites. Prior to analysis, we inspected relationships between the number of genotyped sites in each window and d_{XY}. We found that the variance in d_{XY} was highly inflated in windows containing fewer than 7,500 genotyped sites (variant and invariant). We thus excluded all windows with less than 7,500 sites (of 75,000) from the analysis. As before, we classified windows with d_{XY} values exceeding the 95th percentile as "outlier windows."

We used a generalized linear mixed model (GLMM) to test if the relationship between d_{XY} outlier status (0,1) and recombination differed between DS-GF pairs and DS-Allo pairs. We used the function "glmer" in the R package LME4 (Bates, Mächler, Bolker, & Walker, 2015) to fit a GLMM of the following form: d_{XY} outlier status = recombination rate + regime + comparison (random effect). Outlier status was a binary variable, and we thus used a binomial error function (i.e., a logistic regression). We then refit the model, but included an interaction term: recombination rate × regime. We then compared the fit of the latter model to the simpler model using a likelihood ratio test, implemented via the R function "ANOVA."

3 | RESULTS

3.1 | Population genomic data set

We obtained DNA sequences from databases and generated new genomic data for 20 populations. The combined data set included genomic data from 1,356 individuals from 52 unique populations, each belonging to one of seven described ecotypes: oceanic, lake, stream, benthic, limnetic, white and Sea of Japan (Figure S1, Table S1). The genomic data were a mixture of restriction amplified digest (RAD), genotyping-by-sequencing (GBS) and whole-genome resequencing data sets. We used a single bioinformatics pipeline to standardize the identification of single nucleotide polymorphisms (SNPs) across all study populations (see Methods). Using a variety of criteria (see Methods), we classified each pair of populations into four discrete "evolutionary regimes": divergent selection with gene flow (DS-GF), divergence selection in allopatry (DS-Allo), parallel selection with gene flow (PS-GF) and parallel selection in allopatry (PS-Allo).

3.2 | Localizing candidates for adaptive divergence

In accordance with previous work, we found a general pattern of divergence being higher in regions of low recombination (Figure 1). We identified adaptively differentiated regions of the genome by separately locating SNPs and 75-kilobase pair windows that exhibited unusually high levels of genetic divergence in each pairwise comparison. For all loci (SNPs or windows), we used two metrics of divergence: F_{ST} and d_{XY}, each analysed separately. We considered loci with divergence scores larger than the 95th percentile of the

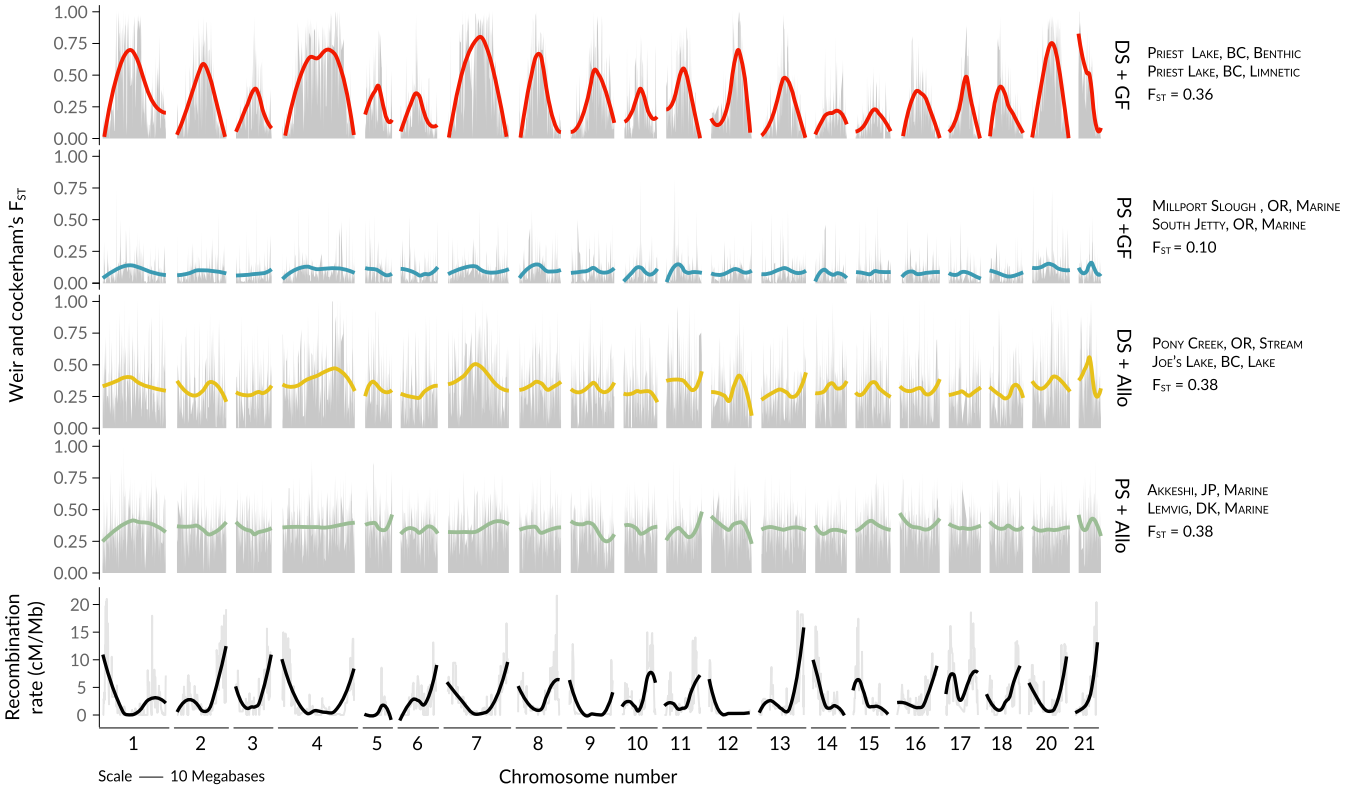

FIGURE 1 Representative plots of genomewide F_{ST} between single pairs of populations from four geneflow and selection regimes. Each coloured line represents a LOESS smooth of F_{ST} vs. chromosomal position for a single chromosome (numbered along bottom). Raw F_{ST} (calculated in 75,000-base pair windows) is depicted in grey behind each smoothed line. Line colour corresponds to geneflow and selection regime (labelled on the right side of the plot). Below the main plots, recombination rate estimates (black lines) from Roesti et al. (2013) are shown for each chromosome. Population pairs were chosen on the basis of similarity in overall F_{ST} and coverage of genomic data. Detailed additional statistics (diversity, d_{XY}, d_S, etc.) for each representative comparison are provided in Figures S6–S9

total distribution to be putatively adaptive loci. While other forces may have caused divergence at these loci, loci subject to divergent selection should be enriched in this set (Narum & Hess, 2011). For convenience, we refer to the loci hereafter as "outlier SNPs" and "outlier windows." For each window, we also estimated mutation rates using a phylogenetic approach and obtained estimates of gene density for each window from the stickleback reference genome annotations via the ENSEMBL database (Jones et al., 2012).

3.3 | Divergence in regions of low recombination

For each pairwise comparison, we used logistic regression to fit outlier status of windows (outlier vs. nonoutlier) to their estimated rates of recombination, while controlling for mutation rate and gene density. The slopes of these regressions were then compared among the four geneflow/selection regimes using a permutation test (see Methods).

In agreement with previous work (Marques et al., 2016; Noor & Bennett, 2009; Renaut et al., 2013; Roesti et al., 2013), we found that F_{ST} outlier windows occurred most often in regions of low recombination, even between allopatric populations and between populations inhabiting similar environments (Figure 2). However, as predicted, this tendency was significantly more extreme in DS-GF comparisons compared to other evolutionary regimes (Figure 2; Figure S2, permutation test on difference in correlation coefficients between regimes: two-sided $p = .0002$). The result remained significant after re-analysis using a window size of 150 kb (permutation test, $p < .0002$) and when recombination rates were estimated using a genetic map derived from North American stickleback populations (Glazer, Killingbeck, Mitros, Rokhsar, & Miller, 2015; permutation test, $p < .0024$).

d_{XY} outliers also showed a tendency (albeit nonsignificant) to occur most often in regions of low recombination (Figure S2; permutation test: two-sided $p = .475$). However, our estimates of d_{XY} from GBS/RAD data set had considerable levels of noise, likely due to low marker density in the 75 kb windows. We thus repeated the d_{XY} analysis, but restricted the analysis to whole genome data sets from a subset of populations (see Methods). Using this reduced data set and 75-kb windows, we found that the relationship between d_{XY} (both outlier status and mean d_{XY}) and recombination was negative in DS-GF comparison and positive in DS-Allo comparisons (Figure 3). This difference in slopes between regimes was highly significant (likelihood ratio test: $\chi^2_2 = 28.85$, $p = 5.41 \times 10^{-5}$). Thus, DS-GF comparisons exhibited unusually high levels of both relative and absolute divergence in regions of low recombination.

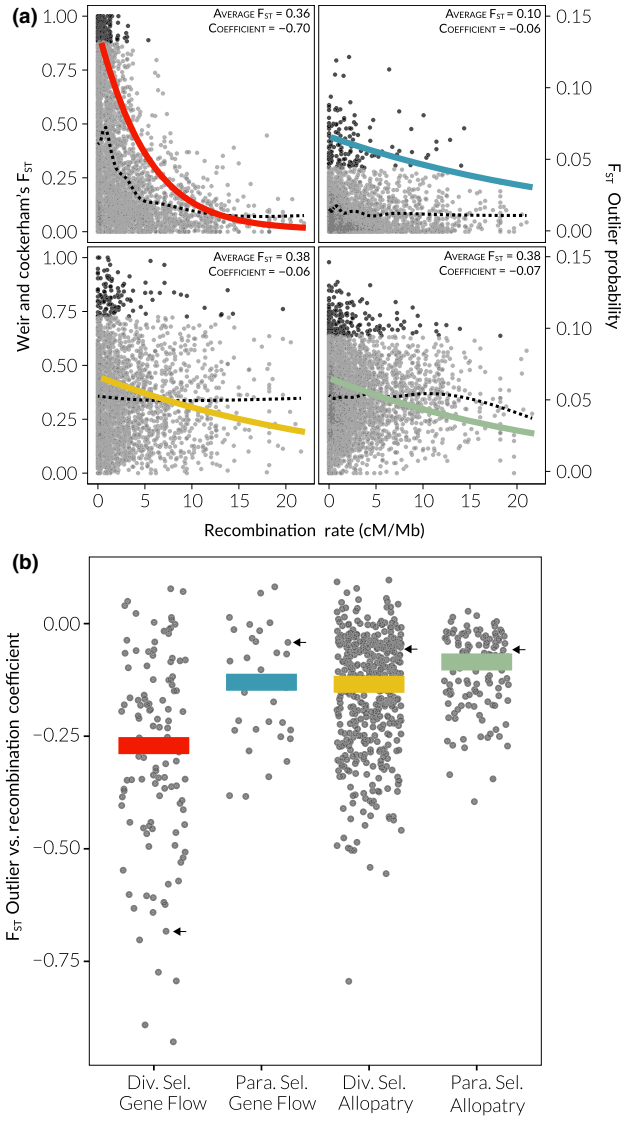

FIGURE 2 Patterns of low recombination bias among the four geneflow and selection regimes. (a) Representative logistic regressions of outlier status (0,1) against recombination rate. Each panel corresponds to a population pair shown in Figure 1 (clockwise from top left: DS-GF, PS-GF, PS-Allo, DS-Allo). Coloured lines depict logistic regression fits of outlier status (black = outliers, grey = nonoutliers), while dotted lines depict LOESS smooths of raw F_{ST}. Regression fits are corrected for variation in mutation rate and gene density. (b) Individual logistic regression coefficients for all pairwise comparisons (points) in each geneflow/selection regime. Coloured horizontal lines indicate means. Increasingly negative coefficients indicate a stronger bias for outliers to occur in the regions of low recombination. Black arrows indicate the coefficient of each representative comparison used in Figure 1 and panel (a) above

3.4 | Ruling out potential sources of bias

3.4.1 | Discretization of geographic distance

The use of a continuous measure of geographic distance led to qualitatively similar results for both F_{ST} and d_{XY} (Figure S5). The tendency for outliers of any type to occur in regions of low recombination was inversely correlated with geographic distance, but only when populations exhibited divergent adaptation (Figure S5; permutation test on differences in divergent vs. parallel slopes: two-sided p = .0002).

3.4.2 | Differences in genomewide F_{ST}

Previous studies have reported that the relationship between divergence and recombination might scale with genomewide divergence (Burri et al., 2015; Lowry, Modliszewski, Wright, Wu, & Willis, 2008). However, we found that the tendency for F_{ST} outlier windows to occur in regions of low-recombination was negatively associated with genomewide F_{ST} (Figure 4, permutation test on correlation, two-sided p = .0001). This suggests that the correlation between geography (as a proxy for gene flow) and F_{ST} in our data set likely biased our results in the *opposite* direction of our findings: as a regime, DS-GF had the greatest number of low-F_{ST} comparisons (Figure 4, red points). Further, we found that if we restricted our analyses in Figure 2 to comparisons in which genomewide F_{ST} is in the range shared across all regimes (0.185–0.675), the tendency for DS-GF comparisons to exhibit more F_{ST} outliers in regions of low recombination remained significant (Figure S4, permutation test: two-sided p = .0002). Moreover, when analysed in a similar fashion, the enrichment of d_{XY} outliers in regions of low recombination in DS-GF populations was also significant (Figure S4, permutation test: two-sided p = .0002).

3.4.3 | Differences in heterozygosity vs. recombination among regimes

Intrapopulation heterozygosity (H_S) was generally lower in regions of low recombination (as expected from linked selection in general), but DS-GF comparisons did not exhibit unusually low levels of heterozygosity in these regions (Figure S2; permutation test: two-sided p = .755). This suggests that the tendency for outliers to occur more often in regions of low recombination in DS-GF comparisons is not an artefact of reduced diversity in those specific comparisons.

3.5 | Clustering of outlier SNPs

In addition to our windowed analyses, we performed a separate analysis to test if individual outlier SNPs from DS-GF comparisons were more clustered than outlier SNPs in other regimes. To do this, we calculated (i) the nearest neighbour distance in centimorgans (cM) between outlier SNPs relative to nearest neighbour distance between all SNPs; and (ii) the coefficient of variation of genetic distances (in cM) between outlier SNPs. Importantly, these clustering metrics are not biased by variation in SNP density among genomic regions and thus are not biased by differences in sequencing coverage.

DS-GF population pairs showed more clustering of F_{ST} outlier SNPs than population pairs in other geneflow/selection regimes (Figure S4). Specifically, DS-GF outlier SNPs were on average

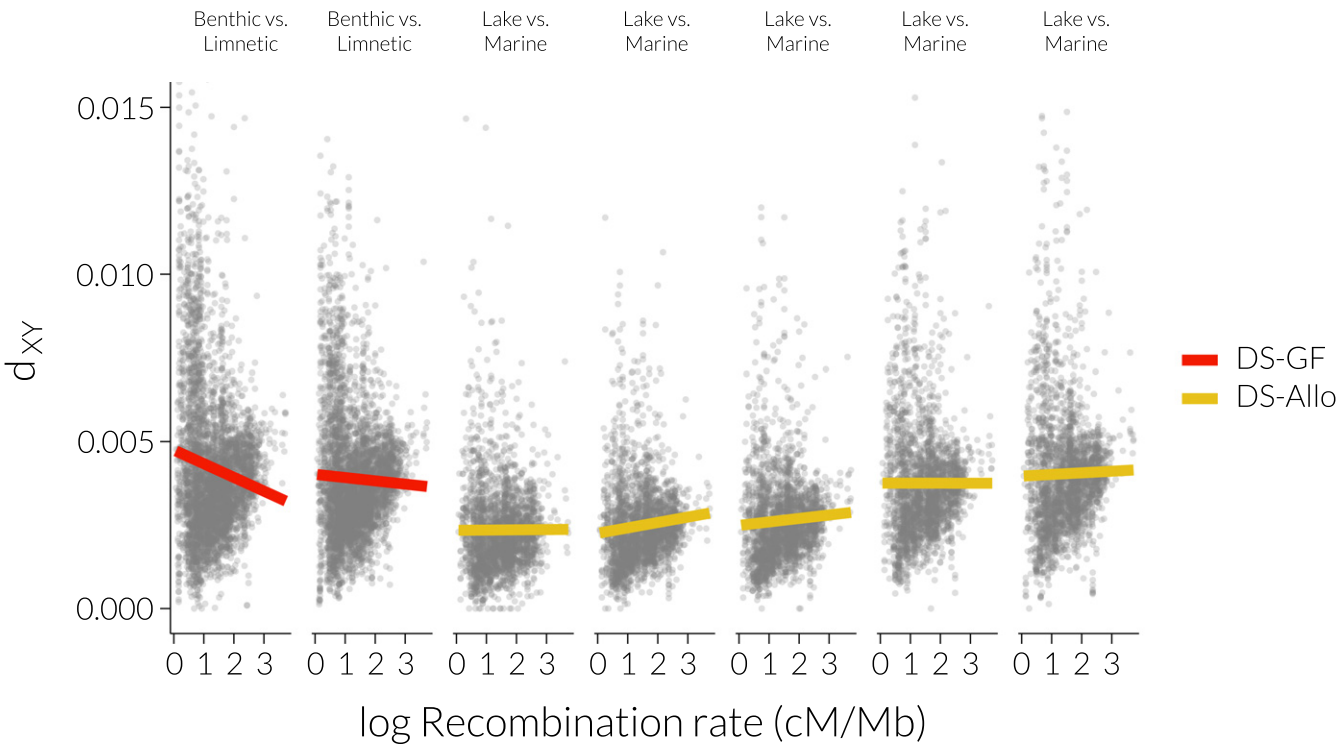

FIGURE 3 The relationship between recombination rate and d_{XY} estimated from whole-genome sequence from seven pairs of stickleback populations. Each panel depicts the relationship between recombination rate and d_{XY} in a single population, calculated by comparing the whole-genome sequences of two individuals. Each point represents the value of d_{XY} in a single 1,000-bp window. Points have been randomly down-sampled by a factor of 100 to aid in visualization. Coloured lines represent lines of best fit. DS-GF (red) comparisons represent d_{XY} between two sympatric populations (a single benthic/limnetic pair), whereas DS-Allopatry (yellow) comparisons represent dxy between two allopatric populations (solitary lake vs. marine). Values on the *x*-axis were transformed via log(value + 1)

approximately one standard deviation closer together in map distance than expected on the basis of overall SNP density (Figure S4, permutation test: two-sided $p < .0001$). Coefficients of variation for the distance between F_{ST} outlier SNPs showed similar results (Figure S4, permutation test: two-sided $p < .0001$), again indicating the highest levels of clustering in DS-GF comparisons.

4 | DISCUSSION

The role of gene flow in shaping the course of evolution remains a key topic in modern evolutionary genetics. Here, we found that in stickleback populations experiencing divergent selection in the face of gene flow (DS-GF), signatures of adaptation are unusually frequent in regions of low recombination. This finding is consistent with theory predicting that maladaptive gene flow favours genetic clustering of adaptive alleles (Aeschbacher et al., 2017; Bürger & Akerman, 2011; Yeaman & Whitlock, 2011).

This finding has several key implications for our understanding of the genetics of adaptation. First, we provide key support for theoretical predictions (Aeschbacher et al., 2017; Nachman & Payseur, 2012; Navarro & Barton, 2003; Yeaman & Whitlock, 2011) that DS-GF should exhibit unique patterns of genomic divergence. Testing these predictions has been a major challenge, because it is difficult to control for, or rule out the effects of other evolutionary processes—divergent selection per se being the most important (see below). Given that gene flow and selection often co-occur in nature, our results imply that the relative strengths of these processes are likely an important determinant of the genomic architecture of adaptation in general (Feder, Egan, & Nosil, 2012; Nosil, Harmon, & Seehausen, 2009; Schluter & Rambaut, 1996). Second, our results suggest that by constraining where divergence can occur, gene flow may cause the "usable area" of the genome to become effectively smaller. This may represent a general constraint on adaptation and could be an important contribution to our ability to explain and predict where adaptation occurs in the genome. Another key implication of this constraint is that by limiting the useable areas of the genome, gene flow may indirectly increase the probability that the same loci will be reused during phenotypic evolution in general. Thus, we might predict that pairs of DS-GF populations (perhaps even ones where selective pressures are different) should display unusual levels of concordance in the loci involved in divergence and that these loci will occur in regions of low recombination. Interestingly, many QTLs involved in parallel adaptation in sticklebacks localize to regions of low recombination in the genome (Noor, Cunningham, & Larkin, 2001; Peichel & Marques, 2017).

Note that the analyses presented here were not designed to detect changes in genome structure or the modification of

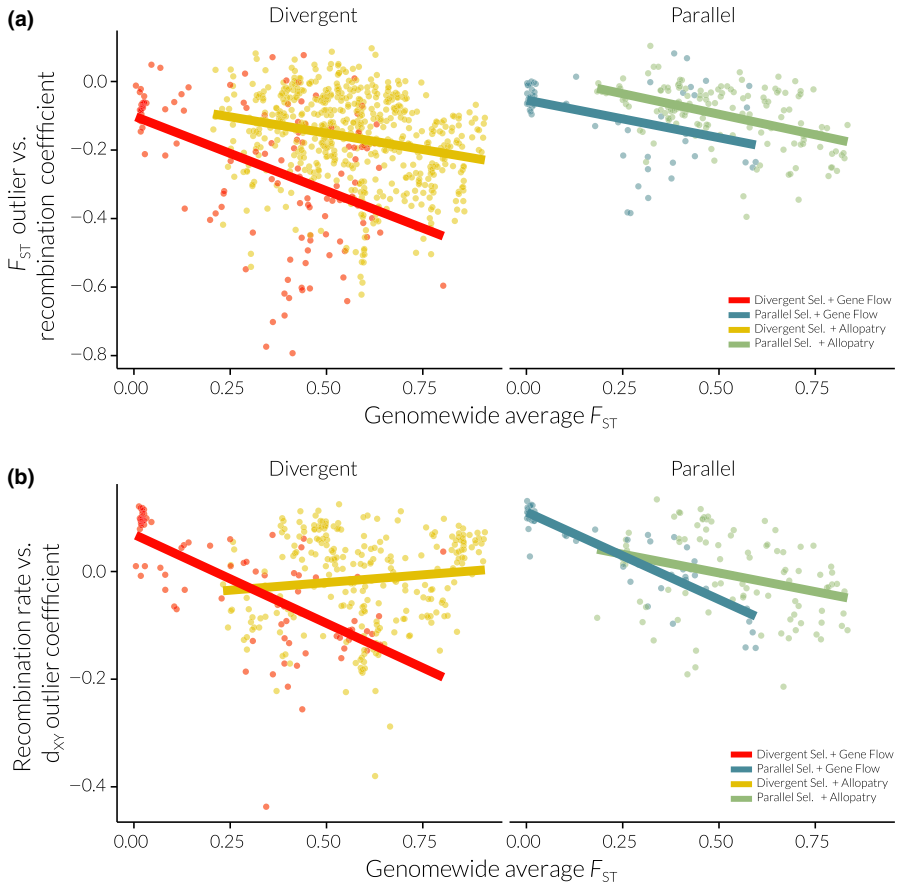

FIGURE 4 The relationship between the tendency for divergence outliers to occur in regions of low recombination (*y*-axis) and overall genetic divergence (*x*-axis) when measured for (a) the F_{ST} outliers and (b) d_{XY} outliers. *Y*-axis values are regression coefficients derived by performing logistic regressions of outlier probability vs. recombination rate for 75-kb genomic windows in each comparison. *X*-axis values are averages of F_{ST} at all loci across the genome for each comparison. Each point represents a single comparison of two populations. Colours indicate different geneflow + selection regimes, with divergent and parallel selection separated for clarity in each of (a) and (b)

recombination rate among populations. We assume that recombination rates are highly conserved between threespine stickleback populations. This is likely a reasonable assumption given that (i) recombination maps are highly similar among threespine stickleback populations from Europe and the United States (Glazer et al., 2015; Roesti et al., 2013), and (ii) homologous chromosomes in the distantly related nine-spine stickleback show very similar patterns of recombination (Rastas, Calboli, Guo, Shikano, & Merilä, 2016). While modification of recombination can be important in some systems, our results pertain to the (likely far more common) scenario in which many loci with potentially varying linkage relationships underlie adaptation and DS-GF favours genetic architectures in which adaptive alleles are tightly linked over other architectures (Yeaman & Whitlock, 2011).

It should be noted, however, that recombination-altering structural variants such as chromosomal inversions likely play an important role in adaptation in sticklebacks (Roesti, Kueng, Moser, & Berner, 2015). However, we were not able to systematically investigate their effects in the context of this study. That said, given the apparent conservation of broadscale recombination rates among populations, the patterns we observed here are unlikely to be driven by segregating chromosomal inversions (which would likely decrease recombination and increase divergence, but only in the particular populations in which they are polymorphic).

4.1 | The costs of low recombination

By definition, loci in regions of low recombination have increased linkage with all nearby loci. We have argued this linkage can facilitate the formation (or prevent the breakdown) of clusters of adaptive alleles, which are more likely to persist in the face of gene flow. However, low recombination also makes it more difficult to (i) establish LD between adaptive alleles that arise on different backgrounds and (ii) break down LD among adaptive alleles and deleterious alleles that happen to arise nearby (the Hill-Robertson effect; Barton, 2010). What then, is happening in the case of DS-GF populations? One possibility is that recombination is still sufficiently common in regions of low recombination to mitigate Hill-Robertson effects. This would imply that the extent of adaptation in regions of low recombination is a complex balance between selection, migration, recombination and the rate of deleterious mutation (Bürger & Akerman, 2011; Marques et al., 2016; Yeaman & Whitlock, 2011). Another

possibility is that the cumulative selective effects of a block of linked adaptive alleles are large enough to negate all but the strongest deleterious mutations. This latter scenario would imply that the (putatively adaptive) clusters of linked alleles are gradually accumulating weakly deleterious alleles, and thus may eventually decay (Kirkpatrick, 2016).

4.2 | Heterogeneous genomic divergence

Our findings also suggest that the patterns of heterogeneous genomic divergence observed in many speciation studies (Feder et al., 2012; Marko & Hart, 2011) may be partly a product of the interaction between gene flow and selection. Explaining this phenomenon has become a major question in speciation genetics, and many recent studies have shown that patterns of heterogeneous divergence in the genome are correlated with recombination rate (Burri et al., 2015; Renaut et al., 2013; Roesti et al., 2013). The association between diversity, divergence and recombination is widely thought to be the result of linked selection, that is background selection and hitchhiking (Charlesworth, 2012). Our results support the general negative association between recombination rate and both diversity and divergence (probably generated by background selection) and further suggest this relationship can be shaped by the effects of divergent selection (presumably through hitchhiking) and gene flow (through the decay of divergence in regions of high recombination and/or favouring linkage between adaptive alleles).

Interestingly, previous work (Burri et al., 2015; Renaut et al., 2013) found no relationship between gene flow and patterns of genomic divergence. One reason for this may simply be power: our data set had many individuals and populations and included pairs of populations across the speciation continuum (in terms of magnitude and time of divergence, geography and type of selection). In the case of Burri et al. (2015), there also appears to be limited amounts of actual introgression between flycatcher populations (although hybridization occurs), weakening any potential pattern.

Although most stickleback populations are less than 10,000 years old, the stickleback metapopulation has repeatedly cycled between adapting to freshwater environments during interglacial periods, followed by extinction of these populations during glacial periods (Hendry et al., 2009; Taylor & McPhail, 2000). However, gene flow between freshwater and marine populations has likely allowed ancient freshwater haplotypes to persist in marine populations throughout this process (Schluter & Conte, 2009). The persistence of these linked blocks of alleles may be, in part, due to their localization in regions of low recombination (e.g., the EDA locus involved in lateral plate formation is located in a region of low recombination on chromosome IV) (Colosimo, 2005). It may be the case that the "reselection" of these linked blocks in multiple populations from standing variation also contributes to the correlation between divergence and recombination in stickleback populations.

A potentially important pattern that emerged from our data set is that population pairs with low divergence ($F_{ST} < 0.1$) generally exhibited much weaker relationships between divergence and recombination (although DS-GF pairs in this category was still more biased on average than PS-GF pairs). This might imply that the accumulation of divergence in regions of low recombination requires a "build-up" phase, after which a tipping point is reached, generating a larger-scale pattern of divergence in regions of low recombination. This process might be akin to the "genomewide-congealing" process described by Flaxman, Wacholder, Feder, and Nosil (2014), but further work is needed to dissect the exact time course of genomic divergence.

4.3 | The effect of divergent selection

Widespread divergent selection alone is predicted to generate a correlation between recombination rate and genomic divergence across the genome via the effects of hitchhiking (Barton, 2010). This effect results in a detection bias for adaptation in regions of low recombination, particularly in reduced representation data sets, such as the RAD and GBS data sets we analysed here (Lowry et al., 2017). Our data support this idea: all "divergent selection" comparisons (DS-GF and DS-Allo) show increased divergence in regions of low recombination (e.g., Figure 2b, red and yellow lines). However, the divergence–recombination correlation is significantly *more* negative in DS-GF populations, which we interpret as a unique joint effect of gene flow and divergent selection. Note that this pattern held when the analysis was restricted to whole-genome data (Figure 3), suggesting that low marker density is not the sole source of the DS-GF low-recombination bias (although likely a contributor). Interestingly, gene flow alone (e.g., parallel selection + gene flow, blue lines in Figures 2 and 4) appears insufficient for generating a divergence bias in regions of low recombination.

A potential alternate explanation for the increase in outlier density in regions of low recombination in DS-GF comparisons is that maladaptive gene flow (migration load) per se increases the strength of divergent selection (Lenormand, 2002). Stronger selection magnifies the scale of linked selection (i.e., the number of loci influenced), and this in turn could increase the negative correlation between recombination and divergence (Barton, 2010). We cannot completely rule out this alternative. However, several facts suggest that variation in the strength of selection is not the sole explanation for our results. For one, the increased divergence in regions of low recombination we observe in DS-GF populations is partly generated by a deficit of highly diverged loci in regions of high recombination (e.g., top right region of panels in Figure 2a). Stronger selection per se should not result in fewer divergent loci in regions of *high* recombination (Barton, 2010; Cutter & Payseur, 2013). Gene flow, on the other hand, is predicted to cause such a deficit, particularly when divergent selection is also acting (Aeschbacher et al., 2017; Tine et al., 2014; Yeaman & Whitlock, 2011). Second, because we took an "all-pairwise" approach for our F_{ST} analyses, the DS-Allopatry category also includes populations experiencing unusually strong directional selection. Thus, the effects of any population-specific selective sweeps were balanced between comparisons of regimes. It should be noted that the connection between gene flow and the strength of selection is by no means

well characterized—indeed under some circumstances, gene flow may actually decrease the strength of divergent selection (Rolshausen et al., 2015), and selection itself often alters the overall migration rate (Peterson, Hilborn, & Hauser, 2014).

4.4 | Caveats

The main strength of the approach we applied here was that it allowed for replication within each geneflow/selection regime, which is necessary for examining statistical differences between regimes in their recombination bias. However, the number of comparisons involved (1,000+) also created computational bottlenecks, which precluded using more sophisticated methods for detecting natural selection and gene flow (Aeschbacher et al., 2017). Further, we do not have detailed knowledge of the demographic history and historical rates of introgression between any of the populations studied here. Both of these factors are known to affect patterns of divergence and can potentially alter the relationship between divergence and recombination (Tine et al., 2014). It is possible that the steep relationship we observed in DS-GF populations between divergence and recombination rate was a result of an unusual demographic or introgression history that was somehow confounded with our current classification of population pairs based on geography and divergent selection. For example, DS-GF comparisons may be enriched for populations that have experienced a period of allopatry, followed by the resumption of gene flow (secondary contact). However, this would still imply that divergent selection and gene flow interact to favour divergence in regions of low recombination, because loci not experiencing divergent selection should still flow freely between populations.

Finally, a major improvement to our approach here would be the incorporation of quantitative estimates of the strength of divergent selection—perhaps using measures of ecological differentiation as a proxy. Recent studies have suggested that there is a great deal of variance in the degree of divergence and parallelism among stickleback ecotype pairs (Oke, Rolshausen, & LeBlond, 2017), and a framework for incorporating said differences would be highly desirable. Thus, while the mechanistic details behind the patterns we describe here are still unclear, we hope our study stimulates further studies of the relationship between gene flow, selection and recombination in shaping patterns of divergence.

ACKNOWLEDGEMENTS

We are very grateful to the Semiahmoo (BC) and Waycobah (NS) First Nations for granting us land access and guidance for the collection of fish used directly or indirectly in this study. We also are indebted to the threespine stickleback research community, whose body of work made this study possible. A.L. Ferchaud, M. Roesti, M. Ravinets, J. Kitano and T. Veen helped in obtaining the data sets used in this study. G. Blackburn, M. Whitlock, L. Rieseberg, S. Yeaman, A. Geraldes, M. Roesti, M. Noor and K. Ostevik and six anonymous reviewers provided key comments on ideas presented here. WestGrid (Compute Canada) provided computational resources used in this project. This work was supported by a Natural Sciences and Engineering Research Council (NSERC) Discovery Grant to DS. KS, GO, DR and KD were additionally supported by NSERC graduate doctoral scholarships.

DATA ACCESSIBILITY

Published genomic datasets: The original study references and accession numbers are listed in Table S1. New genomic datasets: All new data sets are available on the SRA, as listed in Table S1. Analysis code and processed data: https://github.com/ksamuk/gene_flow_linkage. Sequenced reads for the two new data sets provided here are deposited on the NCBI Sequence Read Archive (see Table S1 for accession details).

AUTHOR CONTRIBUTIONS

The authors made the following contributions to the work presented here. K.S., K.D., S.M., G.O., D.R., D.S. project conception and development; K.S., K.D., S.M., G.O., D.R. developed genomic pipeline; K.S., D.R. and G.O. field and laboratory work for new data sets; K.S., G.O., D.S. performed the statistical analysis with input from K.D., S.M. and D.R.; K.S. wrote the paper with input from D.S. and the other authors.

REFERENCES

Aeschbacher, S., Selby, J. P., Willis, J. H., & Coop, G. (2017). Population-genomic inference of the strength and timing of selection against gene flow. *Proceedings of the National Academy of Sciences*, *114*, 7061–7066.

Barton, N. H. (2010). Genetic linkage and natural selection. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, *365*, 2559–2569.

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, *67*, 1–48.

Bolger, A. M., Lohse, M., & Usadel, B. (2014). Trimmomatic: A flexible trimmer for Illumina sequence data. *Bioinformatics*, *30*, 2114–2120.

Bürger, R., & Akerman, A. (2011). The effects of linkage and gene flow on local adaptation: A two-locus continent-island model. *Theoretical Population Biology*, *80*, 272–288.

Burri, R., Nater, A., Kawakami, T., Mugal, C. F., Olason, P. I., Smeds, L., … Hogner, S. (2015). Linked selection and recombination rate variation drive the evolution of the genomic landscape of differentiation across the speciation continuum of Ficedula flycatchers. *Genome Research*, *25*, 1656–1665.

Catchen, J., Bassham, S., Wilson, T., Currey, M., O'Brien, C., … Cresko, W. A. (2013). The population structure and recent colonization history of Oregon threespine stickleback determined using restriction-site associated DNA-sequencing. *Molecular Ecology*, *22*, 2864–2883.

Chain, F. J., Feulner, P. G. D., Panchal, M., Eizaguirre, C., Samonte, I. E., Kable, M., … Reusch, T. B. (2014). Extensive copy-number variation of young genes across stickleback populations (J Zhang, Ed). *PLoS Genetics*, *10*, e1004830.

Charlesworth, B. (2012). The role of background selection in shaping patterns of molecular evolution and variation: Evidence from variability on the Drosophila X chromosome. *Genetics*, *191*, 233–246.

Colosimo, P. F. (2005). Widespread parallel evolution in sticklebacks by repeated fixation of Ectodysplasin alleles. *Science*, *307*, 1928–1933.

Cutter, A. D., & Payseur, B. A. (2013). Genomic signatures of selection at linked sites: Unifying the disparity among species. *Nature Reviews Genetics*, *14*, 262–274.

DePristo, M. A., Banks, E., Poplin, R., Garimella, K. V., Maguire, J. R., Hartl, C., … Daly, M. J. (2011). A framework for variation discovery and genotyping using next-generation DNA sequencing data. *Nature Genetics*, *43*, 491–498.

Elshire, R. J., Glaubitz, J. C., Sun, Q., Poland, J. A., Kawamoto, K., Buckler, E. S., & Mitchell, S. E. (2011). A robust, simple genotyping-by-sequencing (GBS) approach for high diversity species (L Orban, Ed). *PLoS ONE*, *6*, e19379.

Feder, J. L., Egan, S. P., & Nosil, P. (2012). The genomics of speciation-with-gene-flow. *Trends in Genetics*, *28*, 342–350.

Feulner, P. G. D., Chain, F. J., Panchal, M., Huang, Y., Eizaguirre, C., Kable, M., … Milinski, M. (2015). Genomics of divergence along a continuum of parapatric population differentiation (J Zhang, Ed). *PLoS Genetics*, *11*, e1004966.

Flaxman, S. M., Wacholder, A. C., Feder, J. L., & Nosil, P. (2014). Theoretical models of the influence of genomic architecture on the dynamics of speciation. *Molecular Ecology*, *23*, 4074–4088.

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

Gow, J. L., Peichel, C. L., & Taylor, E. B. (2006). Contrasting hybridization rates between sympatric three-spined sticklebacks highlight the fragility of reproductive barriers between evolutionarily young species. *Molecular Ecology*, *15*, 739–752.

Hairston, N. G. Jr, Ellner, S. P., Geber, M. A., Yoshida, T., & Fox, J. A. (2005). Rapid evolution and the convergence of ecological and evolutionary time. *Ecology Letters*, *8*, 1114–1127.

Hendry, A. P., Bolnick, D. I., Berner, D., & Peichel, C. L. (2009). Along the speciation continuum in sticklebacks. *Journal of Fish Biology*, *75*, 2000–2036.

Hohenlohe, P. A., Bassham, S., Etter, P. D., Stiffler, N., Johnson, E. A., & Cresko, W. A. (2010). Population genomics of parallel adaptation in threespine stickleback using sequenced RAD tags. *PLoS Genetics*, *6*, e1000862.

Jones, F. C., Grabherr, M. G., Chan, Y. F., Russell, P., Mauceli, E., Johnson, J., … Kingsley, D. M. (2012). The genomic basis of adaptive evolution in threespine sticklebacks. *Nature*, *484*, 55–61.

Kautt, A. F., Elmer, K. R., & Meyer, A. (2012). Genomic signatures of divergent selection and speciation patterns in a "natural experiment," the young parallel radiations of Nicaraguan crater lake cichlid fishes. *Molecular Ecology*, *21*, 4770–4786.

Kirkpatrick, M. (2016). The evolution of genome structure by natural and sexual selection. *Journal of Heredity*, *108*, 3–11.

Kirkpatrick, M., & Barton, N. (2006). Chromosome inversions, local adaptation and speciation. *Genetics*, *173*, 419–434.

Lenormand, T. (2002). Gene flow and the limits to natural selection. *Trends in Ecology & Evolution*, *17*, 183–189.

Li, H., & Durbin, R. (2010). Fast and accurate long-read alignment with Burrows-Wheeler transform. *Bioinformatics*, *26*, 589–595.

Lowry, D. B., Hoban, S., Kelley, J. L., Lotterhos, K. E., Reed, L. K., Antolin, M. F., & Storfer, A. (2017). Breaking RAD: An evaluation of the utility of restriction site-associated DNA sequencing for genome scans of adaptation. *Molecular Ecology Resources*, *17*, 142–152.

Lowry, D. B., Modliszewski, J. L., Wright, K. M., Wu, C. A., & Willis, J. H. (2008). The strength and genetic basis of reproductive isolating barriers in flowering plants. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *363*, 3009–3021.

Löytynoja, A., & Goldman, N. (2008). Phylogeny-aware gap placement prevents errors in sequence alignment and evolutionary analysis. *Science*, *320*, 1632–1635.

Lunter, G., & Goodson, M. (2011). Stampy: A statistical algorithm for sensitive and fast mapping of Illumina sequence reads. *Genome Research*, *21*, 936–939.

Marko, P. B., & Hart, M. W. (2011). The complex analytical landscape of gene flow inference. *Trends in Ecology & Evolution*, *26*, 448–456.

Marques, D. A., Lucek, K., Meier, J. I., Mwaiko, S., Wagner, C. E., Excoffier, L., & Seehausen, O. (2016). Genomics of rapid incipient speciation in sympatric threespine stickleback. *PLoS Genetics*, *12*, e1005887.

McKenna, A., Hanna, M., Banks, E., Sivachenko, A., Cibulskis, K., Kernytsky, A., … DePristo, M. A. (2010). The Genome Analysis Toolkit: A MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Research*, *20*, 1297–1303.

McKinnon, J. S., & Rundle, H. D. (2002). Speciation in nature: The threespine stickleback model systems. *Trends in Ecology & Evolution*, *17*, 480–487.

Miller, S. E. (2016). Intraguild predation is a mechanism of divergent selection in the threespine stickleback. Vancouver: University of British Columbia.

Nachman, M. W., & Payseur, B. A. (2012). Recombination rate variation and speciation: Theoretical predictions and empirical results from rabbits and mice. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, *367*, 409–421.

Narum, S. R., & Hess, J. E. (2011). Comparison of F(ST) outlier tests for SNP loci under selection. *Molecular Ecology Resources*, *11*(Suppl. 1), 184–194.

Navarro, A., & Barton, N. H. (2003). Accumulating postzygotic isolation genes in parapatry: A new twist on chromosomal speciation. *Evolution*, *57*, 447–459.

Nei, M. (1987). *Molecular evolutionary genetics*. New York, NY: Columbia University Press.

Noor, M. A. F., & Bennett, S. M. (2009). Islands of speciation or mirages in the desert? Examining the role of restricted recombination in maintaining species. *Heredity*, *103*, 439–444.

Noor, M. A., Cunningham, A. L., & Larkin, J. C. (2001). Consequences of recombination rate variation on quantitative trait locus mapping studies: Simulations based on the Drosophila melanogaster genome. *Genetics*, *159*, 581–588.

Noor, M. A. F., & Feder, J. L. (2006). Speciation genetics: Evolving approaches. *Nature Reviews Genetics*, *7*, 851–861.

Noor, M. A., Grams, K. L., Bertucci, L. A., & Reiland, J. (2001). Chromosomal inversions and the reproductive isolation of species. *Proceedings of the National Academy of Sciences*, *98*, 12084–12088.

Nosil, P., Harmon, L. J., & Seehausen, O. (2009). Ecological explanations for (incomplete) speciation. *Trends in Ecology & Evolution*, *24*, 145–156.

Oke, K. B., Rolshausen, G., & LeBlond, C. (2017). How parallel is parallel evolution? A comparative analysis in fishes *The American Naturalist*, *190*, 1–16.

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

Peichel, C. L., & Marques, D. A. (2017). The genetic and molecular architecture of phenotypic diversity in sticklebacks. *Philosophical transactions of the Royal Society of London. Series B, Biological Sciences*, *372*, 20150486.

Peterson, D. A., Hilborn, R., & Hauser, L. (2014). Local adaptation limits lifetime reproductive success of dispersers in a wild salmon metapopulation. *Nature Communications*, *5*, 3696.

R Core Team (2015). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, 2012. Retrieved from http://www.R-project.org

Rastas, P., Calboli, F. C. F., Guo, B., Shikano, T., & Merilä, J. (2016). Construction of ultradense linkage maps with Lep-MAP2: Stickleback F 2 recombinant crosses as an example. *Genome Biology and Evolution*, *8*, 78–93.

Renaut, S., Grassa, C. J., Yeaman, S., Moyers, B. T., Lai, Z., Kane, N. ... Rieseberg, L. H. (2013). Genomic islands of divergence are not affected by geography of speciation in sunflowers. *Nature Communications*, *4*, 1827.

Rieseberg, L. H. (2001). Chromosomal rearrangements and speciation. *Trends in Ecology & Evolution*, *16*, 351–358.

Roesti, M., Hendry, A. P., Salzburger, W., & Berner, D. (2012). Genome divergence during evolutionary diversification as revealed in replicate lake-stream stickleback population pairs. *Molecular Ecology*, *21*, 2852–2862.

Roesti, M., Kueng, B., Moser, D., & Berner, D. (2015). The genomics of ecological vicariance in threespine stickleback fish. *Nature Communications*, *6*, 8767.

Roesti, M., Moser, D., & Berner, D. (2013). Recombination in the threespine stickleback genome-patterns and consequences. *Molecular Ecology*, *22*, 3014–3027.

Rolshausen, G., Muttalib, S., Kaeuffer, R., Oke, K. B., Hanson, D., & Hendry, A. P. (2015). When maladaptive gene flow does not increase selection. *Evolution*, *69*, 2289–2302.

Schluter, D. (1993). Adaptive radiation in sticklebacks: Size, shape, and habitat use efficiency. *Ecology*, *74*, 699.

Schluter, D., & Conte, G. L. (2009). Genetics and ecological speciation. *Proceedings of the National Academy of Sciences of the United States of America*, *106*(Suppl 1), 9955–9962.

Schluter, D., & Rambaut, A. (1996). Ecological speciation in postglacial fishes [and discussion]. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *351*, 807–814.

Sousa, V., & Hey, J. (2013). Understanding the origin of species with genome-scale data: Modelling gene flow. *Nature Reviews Genetics*, *14*, 404–414.

Taylor, E. B., & McPhail, J. D. (2000). Historical contingency and ecological determinism interact to prime speciation in sticklebacks, Gasterosteus. *Proceedings of the Royal Society B: Biological Sciences*, *267*, 2375–2384.

Tine, M., Kuhl, H., Gagnaire, P.-A., Louro, B., Desmarais, E., Martins, R. S. T., ... Reinhardt, R. (2014). European sea bass genome and its variation provide insights into adaptation to euryhalinity and speciation. *Nature Communications*, *5*, 5770.

Vavrek, M. J. (2011). Fossil: palaeoecological and palaeogeographical analysis tools. *Palaeontologia Electronica*, *14*, 1T.

Weir, B. S., & Cockerham, C. C. (1984). Estimating F-statistics for the analysis of population structure. *Evolution*, *38*, 1358.

Wielgoss, S., Barrick, J. E., Tenaillon, O., Cruveiller, S., Chane-Woon-Ming, B., Medigue, C., ... Schneider, D. (2011). Mutation rate inferred from synonymous substitutions in a long-term evolution experiment with *Escherichia coli* (BJ Andrews, Ed). *G3: Genes, Genomes, Genetics*, *1*, 183–186.

Yang, Z. (2007). PAML 4: Phylogenetic analysis by maximum likelihood. *Molecular Biology and Evolution*, *24*, 1586–1591.

Yeaman, S., & Whitlock, M. C. (2011). The genetic architecture of adaptation under migration-selection balance. *Evolution*, *65*, 1897–1911.

Yoshida, K., Makino, T., Yamaguchi, K., Shigenobu, S., Hasebe, M., Kawata, M., ... Kitano, J. (2014). Sex chromosome turnover contributes to genomic divergence between incipient stickleback species (J Zhang, Ed). *PLoS Genetics*, *10*, e1004223.

SUPPORTING INFORMATION

Additional Supporting Information may be found online in the supporting information tab for this article.

How to cite this article: Samuk K, Owens GL, Delmore KE, Miller SE, Rennison DJ, Schluter D. Gene flow and selection interact to promote adaptive divergence in regions of low recombination. *Mol Ecol*. 2017;00:1–13. https://doi.org/10.1111/mec.14226