1

2 **Title:** Background selection and the statistics of population differentiation:
3 consequences for detecting local adaptation

4

5 **Authors:** Remi Matthey-Doret¹ and Michael C. Whitlock

6

7 **Affiliation:** Department of Zoology and Biodiversity Research Centre, University of
8 British Columbia, Vancouver, British Columbia V6T 1Z4, Canada

9

10 ¹**Corresponding author:** matthey@zoology.ubc.ca

11

12

13 **Abstract**

14 Background selection is a process whereby recurrent deleterious mutations cause a
 15 decrease in the effective population size and genetic diversity at linked loci. Several
 16 authors have suggested that variation in the intensity of background selection could
 17 cause variation in F_{ST} across the genome, which could confound signals of local
 18 adaptation in genome scans. We performed realistic simulations of DNA sequences,
 19 using parameter estimates from humans and sticklebacks, to investigate how
 20 variation in the intensity of background selection affects different statistics of
 21 population differentiation. We show that, in populations connected by gene flow,
 22 Weir & Cockerham's (1984) estimator of F_{ST} is largely insensitive to locus-to-locus
 23 variation in the intensity of background selection. Unlike F_{ST} , however, d_{XY} is
 24 negatively correlated with background selection. We also show that background
 25 selection does not greatly affect the false positive rate in F_{ST} outlier studies. Overall,
 26 our study indicates that background selection will not greatly interfere with finding
 27 the variants responsible for local adaptation.

28

29 **Introduction**

30 Natural selection affects patterns of genetic diversity throughout the genome. How
 31 selection affects genetic diversity on a single isolated locus is relatively easy to
 32 model; however, when a large number of linked loci are considered, interactions
 33 between evolutionary pressures at different sites render the task of modelling much
 34 more difficult.

35 Maynard Smith & Haigh (1974) recognized the influence of selection on linked
 36 neutral sites, proposing that strong positive selection could reduce genetic diversity
 37 at nearby sites. This process is now referred to as a ‘selective sweep’. Much later,
 38 Charlesworth et al. (1993) proposed that deleterious mutations could also affect
 39 genetic diversity at nearby sites, because some haplotypes would be removed from
 40 the population as selection acts against linked deleterious alleles. They named this
 41 process background selection (BGS). Both selective sweeps and background
 42 selection affect genetic diversity; they both reduce the effective population size of
 43 linked loci. Empirical evidence of a positive correlation between genetic diversity
 44 and recombination rate has been reported in several species (Cutter and Payseur,
 45 2013), including *Drosophila melanogaster* (Begun & Aquadro, 1992; Elyashiv et al.,
 46 2016), humans (Spencer et al., 2006), collared flycatchers, hooded crows and
 47 Darwin’s finches (Dutoit et al., 2017; see also Vijay et al., 2017).

48

BGS is also expected to affect F_{ST} (Charlesworth et al., 1997; Cutter & Payseur, 2013; Cruickshank & Hahn, 2014; Hoban et al., 2016). At low effective population size, different populations may randomly fix different alleles, increasing F_{ST} , while at high population size, allele frequency changes less through time and different populations are more likely to have comparable allele frequencies, keeping F_{ST} low. This negative relationship between effective population size N_e and F_{ST} is captured in Wright's classical infinite island result; $F_{ST} = \frac{1}{1+4N_e(m+\mu)}$ (Wright, 1943). One might therefore expect that loci under stronger BGS would show higher F_{ST} .

Many authors have also argued that, because BGS reduces the within-population diversity, it should lead to high F_{ST} (Cutter & Payseur, 2013; Cruickshank & Hahn, 2014; Hoban et al., 2016). Expressed in terms of heterozygosities, $F_{ST} = \frac{H_T - H_S}{H_T} = 1 - \frac{H_S}{H_T}$, where H_T is the expected heterozygosity in the entire population and H_S is the average expected heterozygosity within subpopulations (H_S and H_T are also sometimes called π_S and π_T ; e.g. Charlesworth, 1998). All else being equal, a decrease of H_S would indeed lead to an increase of F_{ST} . However, all else is not equal; H_T is also affected by BGS (Charlesworth et al., 1997). Therefore in order to understand the effects of BGS on F_{ST} , we must understand the relative impact of BGS on both H_S and H_T .

Performing numerical simulations, Charlesworth et al. (1997) report that BGS reduces the within population heterozygosity H_S slightly more than it reduces the total heterozygosity H_T , causing a net increase in F_{ST} . The effect on F_{ST} reported is

quite substantial, but, importantly, their simulations were not meant to be realistic.

The authors highlighted their goal in the methods:

“The simulations were intended to show the qualitative effects of the various forces studied [...], so we did not choose biologically plausible values [...]. Rather, we used values that would produce clear-cut effects”.

For example, talking about their choice for the deleterious mutation rate of 8×10^{-4} per site:

“This unrealistically high value was used in order for background selection to produce large effects [...]

Much of the literature on the effect of BGS on F_{ST} is based on the results in Charlesworth et al. (1997), even though they only intended to show proof of concept (see also Zeng & Charlesworth, 2011 and Zeng & Corcoran, 2015). They did not attempt to estimate how strong of an effect BGS has on F_{ST} in real genomes.

It is important to distinguish two separate questions when discussing the effect of BGS on F_{ST} ; 1) How does BGS affect the average genome-wide F_{ST} ? and 2) How does locus-to-locus variation in the intensity of BGS affect locus-to-locus variation in F_{ST} ? The second question is of particular interest to those trying to identify loci under positive selection (local selection or selective sweep). Locus-to-locus variation in F_{ST}

potentially could be confounded with the F_{ST} peaks created by positive selection. In this paper, we focus on this second question.

The identification of loci involved in local adaptation is often performed via F_{ST} outlier tests (Lotterhos & Whitlock, 2014; Hoban et al., 2016). Other tests exist to identify highly divergent loci such as cross-population extended haplotype homozygosity (XP-EHH; Sabeti et al., 2007), comparative haplotype identity (Lange & Pool, 2016), cross-population composite likelihood ratio (XP-CLR; Chen et al., 2010). F_{ST} outlier tests, such as FDist2 (Beaumont & Nichols, 1996), BayeScan (Foll & Gaggiotti, 2008) or FLK (Bonhomme et al., 2010), look for genomic regions showing particularly high F_{ST} values to find candidates for local adaptation. If BGS can affect F_{ST} unevenly across the genome, then regions with a high intensity of BGS could potentially have high F_{ST} values that could be confounded with the pattern caused by local selection (Charlesworth et al., 1997; Cruickshank & Hahn, 2014). BGS could therefore inflate the false positive rate when trying to detect loci under local selection.

The potential confounding effect of BGS on signals of local adaptation has led to an intense effort trying to find solutions to this problem (Bank et al., 2014; Huber et al., 2016). Many authors have understood from Cruickshank and Hahn (2014) that d_{XY} should be used instead F_{ST} in outlier tests (e.g. McGee et al., 2015; Yeaman, 2015; Whitlock & Lotterhos, 2015; Brousseau et al., 2016; Picq et al., 2016; Payseur & Rieseberg, 2016; Hoban et al., 2016; Vijay et al., 2017; see also Nachman & Payseur,

2012). F_{ST} is a measure of population divergence relative to the total genetic diversity, while d_{XY} is an absolute measure of population divergence defined as the probability of non-identity by descent of two alleles drawn in the two different populations averaged over all loci (Nei, 1987; Nei, 1987 originally called it D_{XY} but, here, we follow Cruickshank and Hahn's, 2014 terminology by calling it d_{XY}). The argument is that because F_{ST} is a measure of divergence relative to the genetic diversity and d_{XY} an absolute measure of divergence and because BGS reduces genetic diversity, then BGS must affect F_{ST} but not d_{XY} , a claim that we will investigate in this paper.

Whether BGS can affect genome-wide F_{ST} under some conditions is not in doubt (Charlesworth et al., 1997), but whether locus-to-locus variation in the intensity of BGS present in natural populations substantially affects variation in F_{ST} throughout the genome is very much unknown. Empirically speaking, it has been very difficult to measure how much of the genome-wide variation in genetic diversity is caused by BGS, as opposed to selective sweeps or variation in mutation rates (Cutter & Payseur, 2013; see also attempts in humans by Cai et al., 2009 and McVicker et al. 2009). We are therefore in need of realistic simulations that can give us more insight into how BGS affects genetic diversity among populations and how it affects the statistics of population divergence.

In this article, we investigate the effect of BGS in structured populations with realistic numerical simulations using parameter estimates from humans and

stickleback. Our two main goals are 1) to quantify the impact of locus-to-locus variation in the intensity of BGS on d_{XY} (Nei, 1987) and F_{ST} (Weir & Cockerham, 1984) and 2) to determine whether BGS inflates the false positive rate of F_{ST} outlier tests.

Methods

Our goal is to perform biologically plausible simulations of the local genomic effects of background selection. BGS is expected to vary with gene density, mutation rate and recombination rate across the genome. We used data from real genomes to simulate realistic covariation in recombination rates and gene densities. We chose to base our simulated genomes on two eukaryote genomes, sticklebacks and humans, because these two species have attracted a lot attention in studies of local adaptation and because sticklebacks have a variance in recombination rate which is almost 15 times higher than humans (data not shown), allowing us to test vastly different types of eukaryotic genomes. The recombination rate variation in humans is extremely fine scale, but it presents the potential issue that it is estimated from linkage disequilibrium data. As selection causes linkage disequilibrium to increase, estimates of recombination rate at regions under strong selection may be underestimated, hence artificially increasing the simulated variance in the intensity of BGS. Although the recombination map for stickleback is much less fine scaled, the estimates are less likely to be biased as they are computed from pedigrees.

Our simulations are forward in time and were performed using the simulation platform SimBit version 3.69. The code and user manual are available at <https://github.com/RemiMattheyDoret/SimBit>. To double check our results, we also ran some simulations with SFS_code (Hernandez, 2008), confirming that we get consistent distributions of genetic diversity and of F_{ST} among simulations (results not shown). Generations are non-overlapping, individuals are hermaphrodites, mating is random within patches and selection occurs before dispersal.

Genetics

For each simulation, we randomly sampled a sequence of about 10 cM coming either from the stickleback (*Gasterosteus aculeatus*) genome or from the human genome (see treatments below) and used this genomic location to determine the recombination map and exon locations for a simulation replicate. For the stickleback genome, we used the gene map and recombination map from Roesti et al. (2013). Ensembl-retrieved gene annotations were obtained from Marius Roesti. For the human genome, we used the recombination map from The International HapMap Consortium (2007) and the gene positions from NCBI and positions of regulatory sequences on Ensembl (Zerbino et al., 2017). We excluded sex chromosomes to avoid complications with haploid parts of the genomes. As estimates of mutation rate variation throughout the genome are very limited, we assumed that the haploid mutation rate varies from site to site following an exponential distribution with mean of 2.5×10^{-8} per generation (Nachman & Crowell, 2000).

171 More specifically, we first randomly sampled a sequence of 10^5 nucleotides, which
172 we will refer to as the focal region. All of the statistics (defined under the section
173 *Statistics* below) are calculated only on the focal region of each simulation.
174 Nucleotides that occur in locations determined to be exons in the sampled genomic
175 map are subject to selection (see *Selection*), while all other nucleotides are assumed
176 to be neutral. The focal region itself contained on average ~ 0.44 genes for the
177 human genome and ~ 3.15 genes for the stickleback genome.

178 We simulated a 5 cM region on each side of the focal region (resulting in a window
179 of 10 cM plus the recombination rate present in the specific focal region of 10^5 sites)
180 in order to capture the local effects of background selection. In these 10 cM flanking
181 regions, we only tracked exons. In the nearest 1 cM on each side of the focal region,
182 as well within the focal region, we individually simulated each nucleotide with a bi-
183 allelic locus. On the remaining outer 4 cM, to improve the speed and RAM usage of
184 the process, we tracked the number of mutations in blocks of up to 100 nucleotides.
185 For these blocks, we tracked only the number of mutations but not their location
186 within the block. Ignoring recombination within a block likely had little effect on the
187 results because the average recombination distance between the first and last site of
188 a block is of the order of 10^{-6} cM. The expected number of segregating sites within a
189 block is $4N\mu \sum_{i=1}^{2N-1} (1/i)$, which for a mutation rate per block of 10^{-6} and a
190 population size of $N = 10,000$ is ~ 0.42 . The probabilities of having more than one
191 mutation and more than two mutations (based on a Poisson approximation) are

therefore only approximately 6.7% and 0.9%, respectively. Overall, the level of approximation used is very reasonable.

Selection

As we are interested in the effect of BGS, we modelled the effects of purifying selection against novel deleterious mutations. Each nucleotide in the exons (and regulatory sequences for the human genome) is subject to purifying selection with a selection coefficient against mutant alleles determined by a gamma distribution described below. For focal regions that include exons, statistics are computed over a sequence that is at least partially under direct purifying selection.

To create variance in selection pressures throughout the genome, each exon (and regulatory sequence for the human genome) has its own gamma distribution of heterozygous selection coefficients s . The mean and variance of these gamma distributions are drawn from a bivariate uniform distribution with correlation coefficient of 0.5 (so that when the mean is high, so is the variance) bounded between 10^{-8} and 0.2 for both the mean and the variance. These bounds were inspired by the methodology used in Gilbert et al. (2017). The gamma distributions are bounded to one. Figure S1 shows the overall distribution of selection coefficient s , with 2% of mutations being lethal and an average deleterious selection coefficient for the non-lethal mutations of 0.07. To improve the performance of our simulations, we used multiplicative dominance, where the fitness of heterozygotes is at locus i is $1 - s_i$ and the fitness of the double mutant is $(1 - s_i)^2$.

As a consequence of our parameter choices, our genome-wide deleterious mutation rate was about 1.6 in sticklebacks and about 3 in humans. 9.4% of the stickleback genome and 2.6% of the human genome was under purifying selection. For comparison, the genome-wide deleterious mutation rate is estimated at 2.2 in humans (Keightley, 2012) and 0.44 in rodents (Keightley & Gaffney, 2003). To our knowledge, there is currently no such estimation for sticklebacks. Note however that the above estimates cannot reliably detect mutations that are quasi lethal ($s \ll 1/2N$). By our distribution of selective coefficients, 49% of all deleterious mutations have a heterozygote selective coefficient lower than $1/2N_e$ when $N_e = 1,000$ (42% when $N_e = 10,000$).

It is worth noting however that, in rodents, about half of the deleterious mutation rate occurs in non-coding sequences (Keightley & Gaffney, 2003). Our simulations using human genome had all exons and all regulatory sequences under purifying selection. With our simulations based on the stickleback genome, however, only exons were under purifying selection. It is therefore possible that we would have over-estimated the deleterious mutation rate in gene-rich regions and underestimated the deleterious mutation rate in other regions, especially in stickleback. This would artificially increase the locus-to-locus variation in the intensity of BGS in our simulations.

Demography

In all simulations, we started with a burn-in phase with a single population of N diploid individuals, lasting $5 \times 2N$ generations. The population was then split into two populations of N individuals each with a migration rate between them equal to m . After the burn-in phase, each simulation was run for $5 \times 2N$ more generations for a total of $10 \times 2N$ generations.

Treatments

We explored presence and absence of deleterious mutations over two patch sizes, three migration rates, and two genomes.

We do not have a full factorial design. We considered a basic design and explored variations from this design. The basic design had a population size per patch of $N = 1000$, a migration rate of $m = 0.005$ and used the stickleback genome for its recombination map and gene positions. As deviations from this basic design, we explored modification of every variable, one variable at a time. The *Large N* treatment has $N = 10000$. The *Human* treatment uses the human genome for gene positions, regulatory sequences and recombination map. The treatments *No Migration* and *High Migration* have migration rates of $m = 0$ and $m = 0.05$, respectively.

To test the robustness of our results and because it may be relevant for inversions, we also had unrealistic simulations where recombination rate for the entire genome was set at zero. As a check against previous work, we qualitatively replicated the

results Charlesworth et al. (1997) by performing simulations with similar assumptions as they used. We named this treatment *CNC97*. In our *CNC97* simulations, $N=2000$, $m=0.001$, and 1000 loci were all equally spaced at 0.1 cM apart from each other with constant selection pressure with heterozygotes having fitness of 0.98 and double homozygotes fitness of 0.9 and constant mutation rate $\mu = 0.0004$.

In all treatments (except *Large N*), we performed 4000 simulations; 2000 simulations with BGS and 2000 simulations without selection (where all mutations were neutral). For *Large N*, simulations took more memory and more CPU time. We therefore could only perform 2000 simulations for *Large N*; 1000 simulations with background selection and 1000 simulations without selection. That represents a total of 26,000 simulations for 7 treatments. A full list of all treatments can be found in table 1.

We set the generation 0 at the time of the split. The state of each population was recorded at the end of the burn-in period (generation -1) and at generations $0.001 \times 2N$, $0.05 \times 2N$, $0.158 \times 2N$, $1.581 \times 2N$ and $5 \times 2N$ after the split. For $N=1000$, the sampled generations are therefore -1, 2, 100, 316, 3162 and 10000.

Predicted intensity of Background Selection

In order to investigate the locus-to-locus correlation between the predicted intensity of BGS and various statistics, we computed B , a statistic that approximates

the expected ratio of the coalescent time with background selection over the coalescent time without background selection ($B = \frac{T_{BGS}}{T_{neutral}}$). B quantifies how strong BGS is expected to be for a given simulation (Nordborg et al. 1996). A B value of 0.8 means that BGS has caused a drop of genetic diversity of 20% compared to a theoretical absence of BGS. Lower B values indicate stronger BGS.

Both Hudson & Kaplan (1995) and Nordborg et al. (1996) have derived theoretical expectations for B . We applied both methods and found that the predictions of the two formulas are highly correlated (Figure S2). Because the Hudson & Kaplan (1995) approach has been more popular in the literature (cited almost twice as often), we show only the B values computed following Hudson & Kaplan (1995):

$$B = \exp\left(-\sum_i \frac{u_i s_i}{(s_i + r_i(1 - s_i))^2}\right)$$

where r_i is the recombination rate between the focal site and the i^{th} site under selection, and s_i is the heterozygous selection coefficient at that site. u_i is the mutation rate at the i^{th} site. By this formula, B is bounded between 0 and 1, where 1 means no BGS at all and low values of B mean strong BGS. We computed B for all sites in the focal region and report the average B for the region.

For the stickleback genome, B values ranged from 2×10^{-6} to 1.0 with a mean of 0.937 (Figure S2). For the human genome, B values ranged from 0.45 to 1.0 with a mean at 0.975. There is indeed less variability and much fewer extremely low B

values in the human genome. In the unrealistic *No Recombination* treatment, B values range from 0.00003 to 0.84 with a mean of 0.17.

F_{ST} outlier tests

In order to know the effect of BGS on outlier tests of local adaptation, we used a variant of FDist2 (Beaumont & Nichols, 1996). We chose FDist2 because it is a simple and fast method for which the assumptions of the test match well to the demographic scenario simulated here. Because the program FDist2 is not available through the command line, we rewrote the FDist2 algorithm in R and C++. Source code can be found at <https://github.com/RemiMattheyDoret/Fdist2>.

Our FDist2 procedure is as follows; first, we estimated the migration rate from the average F_{ST} ($m = \frac{F_{ST}-1}{8 \cdot F_{ST} \cdot N}$; Charlesworth, 1998) and then running 50000 simulations each lasting for 50 times the half-life to reach equilibrium F_{ST} given the estimated migration rate (Whitlock, 1992). For each SNP, we then selected the subset of FDist2 simulations for which allelic diversity was less than 0.02 away from the allelic diversity of the SNP of interest. The P -value is computed as the fraction of FDist2 simulations within this subset having a higher F_{ST} than the one we observed. The false positive rate is then defined as the fraction of neutral SNPs for which the P -value is lower than a given α value. The α values explored are 0.1, 0.05, 0.01, 0.001, 0.0001 and 0.00001.

For the outlier tests, to avoid issues of pseudo-replication, we considered only a single SNP per simulation whose minor allele frequency is greater than 0.05. Then, we randomly assembled SNPs from a given treatment into groups of 500 SNPs to create the data file for FDist2. We have 4000 simulations (2000 with BGS and 2000 without BGS) per treatment (*Large N* is an exception with only 2000 simulations total), which allowed 8 independent false positive rate estimates per treatment (4 estimates with BGS and 4 without BGS). In each treatment, we tested for different false positive rate with and without BGS with both a Welch's *t*-test and a Wilcoxon test.

Statistics

F_{ST} and d_{XY} are both measures of population divergence. In the literature there are several definitions of F_{ST} , and we also found potential misunderstanding about how d_{XY} is computed. We want to clarify here these definitions and what we mean when we use the terms F_{ST} and d_{XY} .

There are two main estimators of F_{ST} in the literature; G_{ST} (Nei, 1973) and θ (Weir & Cockerham, 1984). In this article, we focus on θ as an estimate of F_{ST} (Weir & Cockerham, 1984). There are also two methods of averaging F_{ST} over several loci. The first method is to simply take an arithmetic mean over all loci. The second method consists at calculating the sum of the numerator of θ over all loci and dividing it by the sum of the denominator of θ over all loci. Weir and Cockerham (1984) showed that this second averaging approach has lower bias than the simple

arithmetic mean. We will refer to the first method as the “average of ratios” and to the second method as “ratio of the averages” (Reynolds et al. 1983; Weir & Cockerham, 1984) . In this article, we use F_{ST} as calculated by “ratio of the averages”, as advised by Weir and Cockerham (1984). To illustrate the effects of BGS on the biased estimator of F_{ST} , we also computed F_{ST} as a simple arithmetic mean (“average of the ratio”), and we will designate this statistic with a subscript $F_{ST} \text{ (average of ratios)}$.

d_{XY} is a measure of genetic divergence between two populations X and Y. Nei (1987) defined d_{XY} as

$$d_{XY} = \frac{\sum_{l=1}^L \left(1 - \sum_{k=1}^{A_l} x_{l,k} y_{l,k} \right)}{L}$$

where L is the total number of sites, A_l is the number of alleles at the l^{th} site and $x_{l,k}$ and $y_{l,k}$ are the frequency of the k^{th} allele at the l^{th} locus in the population X and Y respectively.

Some population genetics software packages (e.g., EggLib; De Mita and Siol, 2012) average d_{XY} over polymorphic sites only, instead of averaging over all sites, as in Nei's (1987) original definition of d_{XY} . This measure averaged over polymorphic sites only will be called d_{XY-SNP} ; otherwise, we use the original definition of d_{XY} by Nei (1987).

The statistics reported are the average F_{ST} , d_{XY} , and within population genetic

diversity $H_S = \sum_{l=1}^L \left(1 - \sum_{k=1}^{A_l} x_{l,k}^2 \right) / L$. For each treatment and at each generation, we

computed five independent Pearson correlation tests between B and F_{ST} , F_{ST} (average of

ratios), d_{XY} , d_{XY-SNP} and H_S . We compared our correlation tests with ordinary least

squares regressions and robust regressions (using M-estimators; Huber, 1964), and

the results were consistent.

Results

Correlations between the statistics H_S , F_{ST} , F_{ST} (average of ratios), d_{XY} , and d_{XY-SNP} and B , are

summarized in tables S1, S2, S3, S4 and S5, respectively. Figure 1 shows the means

and standard errors for the treatments *Default*, *High Migration*, *Large N*, *Human* and

No Migration. The same graphs for the treatments *No Recombination* and *CNC97* can

be found in Figure S3.

Genetic Diversity

Genetic diversity within populations (H_S) is very similar among the treatments

Default, *High Migration* and *Human* (around $H_S = 1.9 \times 10^{-4}$) but is about 1.9 times

lower in the *No Migration* treatment ($H_S = 1.0 \times 10^{-4}$) and about 10 times higher in

the *Large N* treatment ($H_S = 1.9 \times 10^{-3}$; Figure 1).

B is significantly correlated with genetic diversity within populations (H_S) for all

treatments using the stickleback genome (and at almost all generations) but not

with the *Human* treatment (table S1). Excluding the unrealistic treatments (*No Recombination* and *CNC97*), simulations with BGS have a genetic diversity 4% to 20% lower than simulations without BGS (Figure 1, right column and Figure S3, right column). In the *Human* treatment, there is no significant correlation of B and H_S . Note that Pearson's correlation coefficients between B and H_S are always very small even when the effect is highly significant. The largest R^2 observed in realistic simulations (excluding in the *No Recombination* and *CNC97* treatments) between B and H_S is $R^2 \approx 0.0121$.

Statistics of population divergence

Figure 2 shows the correlation between B and the statistics F_{ST} , d_{XY} and H_S for *Default* at the last generation. These graphs highlight the general tendencies of the treatments *Default*, *High Migration*, *Large N* and *No Recombination*. The strongest correlation with B is observed for the statistics d_{XY} ($P = 3.28 \times 10^{-5}$, $R = 0.093$) and H_S ($P = 3.1 \times 10^{-5}$, $R = 0.093$). In fact, the two statistics d_{XY} and H_S are very highly correlated ($P < 2.2 \times 10^{-16}$, $R = 0.99$). This high correlation explains the resemblance between the central and right graphs of figure 2. F_{ST} is not correlated with B ($P = 0.99$, $R = 10^{-4}$). All correlation tests between B and the statistics F_{ST} , F_{ST} (average of ratios), d_{XY} , d_{XY-SNP} can be found in Tables S2, S3, S4 and S5, respectively.

The *No Migration* treatment is an exception to the other treatments. F_{ST} is not significantly correlated with B at early generations but become slightly correlated as divergence rises to 0.6 and higher. d_{XY} shows an opposite pattern. d_{XY} is very

significantly correlated with B at early generations and seemingly independent of B at the last generation. Note that for both d_{XY} and F_{ST} , all correlation coefficients are always very small. The largest R^2 observed is $R^2=0.0121$ in realistic simulations (found for F_{ST} *No Migration* and for d_{XY} Large N ; Tables S2, S4) and $R^2=0.0256$ for the *No Recombination* treatment (found for d_{XY} ; Table S4).

As expected, in the *CNC97* simulations, there is a strong difference between simulations with BGS and simulations without BGS for all three statistics (F_{ST} , d_{XY} , and H_S) at all generations (Welch's t -tests; all $P < 2.2 \times 10^{-16}$; Figure S3).

F_{ST} calculated as advised by Weir and Cockerham (1984) was generally less sensitive to BGS than F_{ST} calculated as an average of ratios (compare tables S2 and S3). Figure S4 illustrates the sensitivity of F_{ST} (*average of ratios*) in the worst case, the *No Recombination* treatment. This sensitivity is driven largely by rare alleles and goes away when minor alleles below a frequency of 0.05 are excluded.

F_{ST} outlier tests

The observed false positive rate is relatively close to the α values except for *No Migration* (with and without BGS) and *CNC97* (with BGS). With the exception of treatments *No Migration* and *CNC97*, there is no significant difference in false positive rates between simulations with BGS and those without BGS for α of 0.05 (Figure 3). In the treatment, *No Migration*, the false positive rates between simulations with and without BGS are significantly different for the latter

generations. In this treatment, the false positive rate for both simulations with and without BGS are much higher than the α value of 0.05. Results remain very congruent for other α values (results not shown).

Discussion

Background selection reduces genetic diversity, both within and among populations, but the effect on F_{ST} is rather small. In simulations of interconnected populations with realistic parameters, F_{ST} is insensitive to BGS while the absolute measure of divergence d_{XY} is affected by BGS. On the other hand, in highly diverged populations unconnected by migration, BGS can have a greater impact on F_{ST} and a lower impact on d_{XY} . The effects of BGS, when observed, are always very small with R^2 never over 1.1%.

BGS impacts both total and within population genetic diversity. Excluding the treatments *No recombination* and *CNC97*, we observe that simulations with BGS have a genetic diversity (whether H_T or H_S ; H_T data not shown) 6% to 16% lower than simulations without BGS. Messer & Petrov (2013) simulated a panmictic population, looking at a sequence of similar length inspired from a gene-rich region of the human genome, and reported a similar decrease in genetic diversity. Under the *No Recombination* treatment, this average reduction of genetic diversity due to BGS is 53%. Although empirical estimates are very complex and can hardly disentangle BGS from selective sweeps, our results are also comparable with empirical estimates. Reduction in genetic diversity in humans between regions

under high BGS are estimated at 6% according to Cai et al. (2009) or 19-26% according to McVicker et al. (2009). In *Drosophila melanogaster*, where gene density is higher, the reduction in genetic diversity due to BGS is estimated at 36% when using Kim & Stephan (2000)'s methodology and is estimated at 71% reduction using a composite likelihood approach (Elyashiv et al., 2016) and is hence closer to our *No Recombination* treatment than to the other treatments. It is worth noting that, because we were interested in simulating variance among sites in effects of BGS, we only simulated local effects and therefore underestimate the expected genome-wide effect of BGS on H_T .

In contrast to measures of heterozygosity, F_{ST} was generally not significantly correlated with B . The only exception is for the *No Migration* treatment, where, after many generations, as the average F_{ST} becomes very high ($F_{ST} > 0.5$), we observe a slight, yet significant, negative correlation between the expected effects of BGS, B , and F_{ST} (intense BGS lead to high values of F_{ST}). This highlights that F_{ST} is not completely insensitive to BGS, but F_{ST} is largely robust to BGS.

Future research is needed to attempt a theoretical estimate of the genome-wide effect of BGS. Our work has been restricted to the stickleback and human genomes. While these two genomes are good representatives of many cases of eukaryotic genomes, they are not good representatives of more compact genomes such as bacterial genomes or yeasts. Our simulations used randomly mating diploid populations. Non-random mating, selfing, and asexual reproduction could also affect

our general conclusion, and potentially strongly increase the effects of BGS on F_{ST} (Charlesworth *et al.* 1997). We have explored two population sizes, but we could not explore population sizes of the order of a million individuals (like *Drosophila melanogaster*) and still realistically simulate such long stretch of DNA. It is not impossible that a much greater population size or a more complex demography could yield to BGS having a greater effect on F_{ST} than what we observed here (Torres *et al.* 2017).

Some have argued that, because BGS reduces the within population diversity, it should lead to high F_{ST} (Cutter & Payseur, 2013; Cruickshank & Hahn, 2014; Hoban *et al.*, 2016). All else being equal, this statement is correct. However, BGS reduces H_T almost as much as H_S (Figure 4). It is therefore insufficient to consider only one component, and we must consider the ratio of these two quantities captured by the definition of F_{ST} , $F_{ST} = 1 - \frac{H_S}{H_T}$. This ratio, as we have shown, appears to be relatively robust to BGS. While genome-wide BGS might eventually be strong enough to cause departures with F_{ST} values, it appears that locus-to-locus variation in the intensity of BGS is not strong enough to have much impact on F_{ST} as long as populations are not too highly diverged.

We also investigated the consequences of BGS on the widely-used but imperfect estimator, $F_{ST}(\text{average of ratios})$, for which F_{ST} measures for each locus are averaged to create a genomic average. It is well known that $F_{ST}(\text{average of ratios})$ is a biased way to average F_{ST} over several loci (Weir & Cockerham, 1984); however, its usage is

relatively common today. In our simulations, F_{ST} (average of ratios) is more affected by BGS than F_{ST} . Interestingly, F_{ST} (average of ratios) is most often higher with weaker BGS. The directionality of this correlation may seem unintuitive at first. To understand this discrepancy, remember that BGS affects the site frequency spectrum; we observed that BGS leads to an excess of loci with low H_T (results not shown but see Charlesworth et al., 1995; see also contrary expectation in Stephan, 2010). Loci associated with very low H_T also have low F_{ST} (figure S5), a well-known result described by Beaumont and Nichols (1996). As BGS creates an excess of loci with low H_T and loci with low H_T tend to have low F_{ST} , BGS can actually reduce F_{ST} (average of ratios). After filtering out SNPs with a minor allele frequency lower than 5%, most of the correlation between F_{ST} (average of ratios) and B is eliminated (Figure S4).

The absolute measure of divergence d_{XY} is sensitive to BGS. Regions of stronger BGS are associated with low d_{XY} . The effect, although significant, is of relatively small size. The expected d_{XY} for neutral loci is $d_{XY} = 4N\mu + 2t\mu$ (Nei, 1987), where t is the time in generation since the populations started to diverge. $4N\mu$ is the expected heterozygosity in the ancestral population (before splitting) and $2t\mu$ is the expected number of mutations fixed over time in either population since the population split. BGS does not affect the rate of fixation of mutations arising after the populations diverged, but BGS affects the expected heterozygosity. Therefore, BGS should affect d_{XY} by its effect on the expected heterozygosity, and this effect should be greater early in divergence when the $4N\mu$ term is large relative to the fixation term. This is consistent with the results of our simulations. This result is in agreement with Vijay

et al. (2017) who reported a strong correlation between H_S and d_{XY} when F_{ST} is low ($F_{ST} \approx 0.02$), but this correlation breaks down when studying more distantly related populations ($F_{ST} \approx 0.3$).

Interestingly, d_{XY} becomes less sensitive to BGS when F_{ST} becomes more sensitive. While our methodology does not allow us to test the efficiency of d_{XY} in outlier tests, it is possible that d_{XY} could be used for highly divergent lineages, but not in cases when divergence is relatively low. Cruickshank and Hahn (2014) suggested relying more on d_{XY} than F_{ST} for finding highly divergent loci. Their conclusion was based on analysis of a dataset involving highly divergent populations only (F_{ST} values range from about 0.38 to about 0.8). Based on our simulations, their conclusion is not valid when populations are not highly diverged.

As BGS also leads to a reduction of the number of polymorphic sites, BGS has an even stronger effect on d_{XY-SNP} than on d_{XY} (Figure S4). (The measure that we call d_{XY-SNP} is d_{xy} improperly calculated based only on polymorphic sites, as is done in some software packages.) This result highlights the importance of not blindly trusting the output of a given software package.

Outside the effect of BGS on N_e , there are at least two other possible factors that can potentially affect the correlation between B and μ : the effect of deleterious mutations on the effective migration rate and the auto-correlation of μ . Because most deleterious mutations are recessive (García-Dorado and Caballero, 2000; Peters et al., 2003; Shaw & Chang, 2006), the offspring of migrants, who enjoy an

515 increased heterozygosity compared to local individuals, will be at a selective
516 advantage. The presence of deleterious mutations therefore lead to an increase in
517 the effective migration rate (Ingvarsson & Whitlock, 2000). This increases the
518 effective migration rate and hence, leads to a decrease in F_{ST} .

519 As mutation rate is auto-correlated throughout the genome, neutral sequences
520 closely linked to sequences that frequently receive deleterious mutation are also
521 likely to experience frequent neutral mutations. As a high mutation rate leads to low
522 F_{ST} values ($F_{ST} \cong \frac{1}{1+4Ne(m+\mu)}$, Wright 1943), autocorrelation in mutation rate may
523 also impact the correlation between B and F_{ST} . This effect is likely to be negligible as
524 long as $m \gg \mu$.

525 Recently, evidence of a correlation between recombination rate and F_{ST} has been
526 interpreted as likely being caused by deleterious mutations rather than positive
527 selection, whether the divergence between populations is very high (e.g.
528 Cruickshank & Hahn, 2014), moderately high (Vijay et al., 2017) or moderately low
529 (Torres et al., 2017). Here we showed the BGS is unlikely to explain all of these
530 correlations. It can be hypothesized that positive selection (selective sweeps and
531 local adaptation) could be the main cause of this correlation.

532 McVicker et al. (2009) attempted an estimation of B values in the human genome
533 (see also Elyashiv et al., 2016). They did so using equations from Nordborg et al.
534 (1996). As there is little knowledge about the strength of selection throughout the

genome, to our understanding, this estimation of B values should be highly influenced by the effects of beneficial mutations as well as deleterious mutations. Torres et al. (2017) reused this dataset and found a slight association between B and F_{ST} among human lineages. It is plausible that this correlation between B and F_{ST} could be driven by positive selection rather than by deleterious mutations.

Our FDist2 analysis shows that the false positive rate does not differ in simulations with BGS or without BGS. The only exceptions concern the unrealistic *CNC97* treatment and the *No Migration* treatment after many generations (Figure 3). The average F_{ST} at the last generation of the *No Migration* treatment is greater than 0.8. With such high F_{ST} both the simulation without BGS and with BGS lead to very high false positive rates (0.45 without BGS and 0.5 with BGS). This difference in false positive rates is significant, but the observed high false positive rate should make it clear that for such highly diverged populations, F_{ST} outlier tests are not recommended in general.

Many authors (e.g. Nachman & Payseur, 2012; Cutter & Payseur, 2013; McGee et al., 2015; Yeaman, 2015; Whitlock & Lotterhos, 2015; Brousseau et al., 2016; Picq et al., 2016; Payseur & Rieseberg, 2016; Hoban et al., 2016) have raised concerns that BGS can strongly reduce our ability to detect the genomic signature of local adaptation. Our analysis shows that BGS is not a strong confounding factor to F_{ST} outlier tests of populations that are not too highly diverged.

Acknowledgments

Many thanks to Loren H. Rieseberg, Sarah P. Otto and Amy L. Angert for their help in discussing the design of the project and for feedback. Thanks to Sarah P. Otto, Darren E. Irwin, and Bret A. Payseur for helpful comments on the manuscript. We also thank Yaniv Brandvain and Tom Booker for their feedback and Marius Roesti for help with the Ensembl-retrieved gene annotations.

Funding

The work was funded by NSERC Discovery Grant RGPIN-2016-03779 to MCW and by the Swiss National Science Foundation via the fellowship Doc.Mobility P1SKP3_168393 to RMD.

References

Bank C, Ewing GB, Ferrer-Admettla A, Foll M, Jensen JD (2014) Thinking too positive? Revisiting current methods of population genetic selection inference. *Trends in Genetics*, **30**, 540–546.

Beaumont MA, Nichols RA (1996) Evaluating loci for use in the genetic analysis of population structure. *Proceedings of the Royal Society B: Biological Sciences*, **263**, 1619–1626.

573

574 Begun DJ, Aquadro CF (1992) Levels of naturally occurring DNA polymorphism
575 correlate with recombination rates in *D. melanogaster*. *Nature*, **356**, 519–520.

576

577 Bonhomme M, Chevalet C, Servin B *et al.* (2010) Detecting selection in population
578 trees: the Lewontin and Krakauer test extended. *Genetics*, **186**, 241–262.

579

580 Brousseau L, Postolache D, Lascoux M *et al.* (2016) Local adaptation in European
581 Firs assessed through extensive sampling across altitudinal gradients in southern
582 Europe. *Plos One*, **11**.

583

584 Cai JJ, Macpherson JM, Sella G, Petrov DA (2009) Pervasive hitchhiking at coding and
585 regulatory sites in humans. *PLoS Genetics*, **5**.

586

587 Charlesworth B, Morgan MT, Charlesworth D (1993) The effect of deleterious
588 mutations on neutral molecular variation. *Genetics*, **134**, 1289–1303.

589

590 Charlesworth B (1998) Measures of divergence between populations and the effect
591 of forces that reduce variability. *Molecular Biology and Evolution*, **15**, 538–543.

592

593 Charlesworth B, Nordborg M, Charlesworth D (1997) The effects of local selection,
594 balanced polymorphism and background selection on equilibrium patterns of
595 genetic diversity in subdivided populations. *Genetical Research*, **70**, 155–174.

596

597 Charlesworth D, Charlesworth B, Morgan MT (1995) The pattern of neutral
598 molecular variation under the background selection model. *Genetics*, **141**, 1619–
599 1632.

600

601 Chen H, Patterson N, Reich D (2010) Population differentiation as a test for selective
602 sweeps. *Genome Research*, **20**, 393–402.

603

604 Cruickshank TE, Hahn MW (2014) Reanalysis suggests that genomic islands of
605 speciation are due to reduced diversity, not reduced gene flow. *Molecular*
606 *Ecology*, **23**, 3133–3157.

607

608 Cutter AD, Payseur BA (2013) Genomic signatures of selection at linked sites:
609 Unifying the disparity among species. *Nature Reviews Genetics*, **14**, 262–274.

610

611 Dutoit L, Vijay N, Mugal CF *et al.* (2017) Covariation in levels of nucleotide diversity
612 in homologous regions of the avian genome long after completion of lineage
613 sorting. *Proceedings of the Royal Society B: Biological Sciences*, **284**, 20162756.

614

615 Elyashiv E, Sattath S, Hu TT *et al.* (2016) A genomic map of the effects of linked
616 selection in *Drosophila*. *PLOS Genetics*, **12**.

617

618 Foll M, Gaggiotti O (2008) A genome-scan method to identify selected loci
619 appropriate for both dominant and codominant markers: A Bayesian
620 perspective. *Genetics*, **180**, 977–993.

621

622 García-Dorado A, Caballero A (2000) On the average coefficient of dominance of
623 deleterious spontaneous mutations. *Genetics*, **155**, 1991–2001.

624

625 Gilbert KJ, Sharp NP, Angert AL *et al.* (2017) Local adaptation interacts with
626 expansion load during range expansion: Maladaptation reduces expansion load. *The*
627 *American Naturalist*, **189**, 368–380.

628

629 Hernandez RD (2008) A flexible forward simulator for populations subject to
630 selection and demography. *Bioinformatics*, **24**, 2786–2787.

631

632 Hoban S, Kelley JL, Lotterhos KE *et al.* (2016) Finding the genomic basis of local
633 adaptation: Pitfalls, practical solutions, and future directions. *The American*
634 *Naturalist*, **188**, 379–397.

635

636 Huber CD, Degiorgio M, Hellmann I, Nielsen R (2016) Detecting recent selective
637 sweeps while controlling for mutation rate and background selection. *Molecular*
638 *Ecology*, **25**, 142–156.

639

640 Huber PJ (1964) Robust estimation of a location parameter. *The Annals of*
641 *Mathematical Statistics*, **35**, 73–101.

642

643 Hudson RR, Kaplan NL (1995) Deleterious background selection with
644 recombination. *Genetics*, **141**, 1605–1617.

645

646 Ingvarsson PK, Whitlock MC (2000) Heterosis increases the effective migration
647 rate. *Proceedings of the Royal Society B: Biological Sciences*, **267**, 1321–1326.

648

649 International HapMap Consortium (2007) A second generation human haplotype
650 map of over 3.1 million SNPs. *Nature*, **449**, 851.

651

652 Keightley PD, Gaffney DJ (2003) Functional constraints and frequency of deleterious
653 mutations in noncoding DNA of rodents. *Proceedings of the National Academy of*
654 *Sciences*, **100**, 13402–13406.

655

656 Keightley PD (2012) Rates and Fitness Consequences of new mutations in
657 humans. *Genetics*, **190**, 295–304.

658

659 Kim Y, Stephan W (2000) Joint effects of genetic hitchhiking and background
660 selection on neutral variation. *Genetics*, **155**, 1415–1427.

661

662 Lange JD, Pool JE (2016) A haplotype method detects diverse scenarios of local
663 adaptation from genomic sequence variation. *Molecular Ecology*, **25**, 3081–3100.

664

665 Lotterhos KE, Whitlock MC (2014) Evaluation of demographic history and neutral
666 parameterization on the performance of F_{ST} outlier tests. *Molecular Ecology*, **23**,
667 2178–2192.

668

669 Maynard Smith J, Haigh J (1974) The hitch-hiking effect of a favourable
670 gene. *Genetical Research*, **23**, 23.

671

672 McGee MD, Neches RY, Seehausen O (2015) Evaluating genomic divergence and
673 parallelism in replicate ecomorphs from young and old cichlid adaptive
674 radiations. *Molecular Ecology*, **25**, 260–268.

675

676 McVicker G, Gordon D, Davis C, Green P (2009) Widespread genomic signatures of
677 natural selection in hominid evolution. *PLoS Genetics*, **5**.

678

679 Messer PW, Petrov DA (2013) Frequent adaptation and the McDonald-Kreitman
680 test. *Proceedings of the National Academy of Sciences*, **110**, 8615–8620.

681

682 Mita SD, Siol M (2012) EggLib: processing, analysis and simulation tools for
683 population genetics and genomics. *BMC Genetics*, **13**, 27.

684

685 Nachman MW, Crowell SL (2000) Estimate of the mutation rate per nucleotide in
686 humans. *Genetics*, **156**, 297–304.

687

688 Nachman MW, Payseur BA (2012) Recombination rate variation and speciation:
689 Theoretical predictions and empirical results from rabbits and mice. *Philosophical*
690 *Transactions of the Royal Society B: Biological Sciences*, **367**, 409–421.

691

692 Nei M (1973) Analysis of gene diversity in subdivided populations. *Proceedings of*
693 *the National Academy of Sciences*, **70**, 3321–3323.

694

695 Nei M (1987) *Molecular evolutionary genetics*. Columbia University Press, New York
696 (N.Y.).

697

698 Nordborg M, Charlesworth B, Charlesworth D (1996) The effect of recombination on
699 background selection. *Genetical Research*, **67**, 159.

700

701 Payseur BA, Rieseberg LH (2016) A genomic perspective on hybridization and
702 speciation. *Molecular Ecology*, **25**, 2337–2360.

703

704 Peters AD, Halligan DL, Whitlock MC, Keightley PD (2003) Dominance and
705 overdominance of mildly deleterious induced mutations for fitness traits in
706 *Caenorhabditis elegans*. *Genetics*, **165**, 589–599.

707

708 Picq S, Mcmillan WO, Puebla O (2016) Population genomics of local adaptation
709 versus speciation in coral reef fishes (*Hypoplectrusspp*, *Serranidae*). *Ecology and*
710 *Evolution*, **6**, 2109–2124.

711

712 Reynolds J, Weir BS, Cockerham CC (1983) Estimation of the coancestry coefficient:
713 Basis for a short-term genetic distance. *Genetics*, **105**, 767–779.

714

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

717

718 Sabeti PC, Varilly P, Fry B *et al.* (2007) Genome-wide detection and characterization
719 of positive selection in human populations. *Nature*, **449**, 913.

720

721 Shaw RG, Chang S-M (2006) Gene action of new mutations in *Arabidopsis*
722 *thaliana*. *Genetics*, **172**, 1855–1865.

723

724 Spencer CCA, Deloukas P, Hunt S *et al.* (2006) The influence of recombination on
725 human genetic diversity. *PLoS Genetics*, **2**.

726

727 Stephan W (2010) Genetic hitchhiking versus background selection: the controversy
728 and its implications. *Philosophical Transactions of the Royal Society B: Biological*
729 *Sciences*, **365**, 1245–1253.

730

731 Torres R, Szpiech ZA, Hernandez RD (2017) Human demographic history has
732 amplified the effects of background selection across the genome.

733

734 Vijay N, Weissensteiner M, Burri R *et al.* (2017) Genomewide patterns of variation in
735 genetic diversity are shared among populations, species and higher-order
736 taxa. *Molecular Ecology*, **26**, 4284–4295.

737

738 Weir BS, Cockerham CC (1984) Estimating F-statistics for the analysis of population
739 structure. *Evolution*, **38**, 1358.

740

741 Whitlock MC (1992) Temporal fluctuations in demographic parameters and the
742 genetic variance among populations. *Evolution*, **46**, 608.

743

744 Whitlock MC, Lotterhos KE (2015) Reliable detection of loci responsible for local
745 adaptation: Inference of a null model through trimming the distribution of F_{ST} . *The*
746 *American Naturalist*, **186**.

747

748 Wright S (1943) Isolation by distance. *Genetics*, **28**, 114–138.

749

750 Yeaman S (2015) Local adaptation by alleles of small effect. *The American*
751 *Naturalist*, **186**.

752

753 Zeng K, Corcoran P (2015) The effects of background and interference selection on
754 patterns of genetic variation in subdivided populations. *Genetics*, **201**, 1539–1554.

755

756 Zeng K, Charlesworth B (2011) The joint effects of background selection and genetic
757 recombination on local gene genealogies. *Genetics*, **189**, 251–266.

758

759 Zerbino DR, Achuthan P, Akanni W *et al.* (2017) Ensembl 2018. *Nucleic acids*
760 *research*, **46**, D754–D761.

761

762

Table 1: Summary of treatments. For all treatments but *CNC97*, the average mutation rate was set to 2.5×10^{-8} per site, per generation and the mean heterozygous selection coefficient to 0.1.

<i>Treatment</i>	<i>N</i>	<i>m</i>	<i>Genome</i>	<i>Recombination</i>	<i>BGS</i>
<i>Default</i>	<i>1000</i>	<i>0.005</i>	<i>Stickleback</i>	<i>Yes</i>	<i>Yes</i>
					<i>No</i>
<i>No Migration</i>	<i>1000</i>	<i>0</i>	<i>Stickleback</i>	<i>Yes</i>	<i>Yes</i>
					<i>No</i>
<i>High Migration</i>	<i>1000</i>	<i>0.05</i>	<i>Stickleback</i>	<i>Yes</i>	<i>Yes</i>
					<i>No</i>
<i>Large N</i>	<i>10000</i>	<i>0.005</i>	<i>Stickleback</i>	<i>Yes</i>	<i>Yes</i>
					<i>No</i>
<i>Human</i>	<i>1000</i>	<i>0.005</i>	<i>Human</i>	<i>Yes</i>	<i>Yes</i>
					<i>No</i>
<i>No Recombination</i>	<i>1000</i>	<i>0.005</i>	<i>Stickleback</i>	<i>No</i>	<i>Yes</i>
					<i>No</i>
<i>CNC97</i>	<i>2000</i>	<i>0.001</i>	<i>NA</i>	<i>Yes</i>	<i>Yes</i>
					<i>No</i>

Figure 1: Comparisons of means F_{ST} (left column), d_{XY} (central column), and H_S (right column) between simulations with (black) and without (grey) BGS. Similar graphs for the treatments *No Recombination* and *CNC97* are in figure S3. Error bars are 95% CI.

Figure 2: Correlation between B and F_{ST} , H_S , and d_{XY} for the last generation ($5 \times 2N$ generations after the split) of the *Default* treatment. Each grey dot is a single simulation where there is BGS. The large black dot is the mean of the simulations where BGS was artificially turned off. The P -values are computed from a Pearson's correlation test. P -values and R are computed on the simulations with BGS (grey dots) only.

Figure 3: Comparison of false positive rate (FPR) returned by FDist2 between simulations with BGS (black) and without BGS (grey) for all treatments by generation. The significance level is 0.05 and is represented by the horizontal dashed line. Significance based on a Welch's *t*-test is indicated with stars (2×10^{-16} '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1). Significance levels are the same with Wilcoxon tests.

Figure 4: Regressions of total (H_T ; upper line) and within (H_S ; lower line) population expected heterozygosity on the coefficient of BGS (B) for the last generation of the *Default* treatment. The two regression lines are not exactly parallel with H_S tending to H_T as B goes to low values (more intense BGS).

Table S1: Pearson's correlation tests for the association between the coefficient of background selection (B) and H_S . P -values and R are computed on the simulations with BGS only.

<i>Treatment</i>	<i>Generation after split</i>	<i>R</i>	<i>P</i>	<i>Bonferroni correction</i>
<i>Default</i>	1	0.085	1.4×10^{-4}	**
	5	0.083	1.9×10^{-4}	**
	158	0.082	2.4×10^{-4}	**
	1581	0.096	1.8×10^{-5}	***
	10000	0.093	3.1×10^{-5}	**
<i>High Migration</i>	1	0.056	0.01	
	5	0.054	0.02	
	158	0.056	0.01	
	1581	0.062	5.3×10^{-3}	
	10000	0.079	4.4×10^{-4}	*
<i>Human</i>	1	-0.036	0.11	
	5	-0.039	0.09	
	158	-0.047	0.04	
	1581	-0.067	2.7×10^{-3}	.
	10000	-0.036	0.11	
<i>Large N</i>	10	0.097	2.2×10^{-3}	.
	50	0.096	2.3×10^{-3}	.
	1580	0.091	0.4×10^{-2}	
	15810	0.087	5.6×10^{-3}	
	100000	0.11	5.8×10^{-4}	*
<i>No Migration</i>	1	0.088	7.7×10^{-5}	**
	5	0.088	8.4×10^{-5}	**
	158	0.095	1.9×10^{-5}	***
	1581	0.089	6.2×10^{-5}	**
	10000	0.09	5.9×10^{-5}	**
<i>No Recombination</i>	1	0.086	1.3×10^{-4}	**
	5	0.086	1.3×10^{-4}	**
	158	0.088	9.1×10^{-5}	**
	1581	0.1	4.9×10^{-6}	***
	10000	0.16	1.9×10^{-12}	***

Table S2: Pearson's correlation tests for the association between the coefficient of background selection (B) and F_{ST} . P -values and R are computed on the simulations with BGS only.

<i>Treatment</i>	<i>Generation after split</i>	<i>R</i>	<i>P</i>	<i>Bonferroni correction</i>
<i>Default</i>	1	0.012	0.6	
	5	0.029	0.2	
	158	-0.015	0.49	
	1581	-0.021	0.35	
	10000	0.00014	0.99	
<i>High Migration</i>	1	-0.017	0.45	
	5	0.013	0.56	
	158	0.045	0.04	
	1581	-0.0023	0.92	
	10000	-0.0037	0.87	
<i>Human</i>	1	-0.012	0.6	
	5	0.033	0.14	
	158	-0.019	0.39	
	1581	0.028	0.22	
	10000	0.0077	0.73	
<i>Large N</i>	10	0.0084	0.79	
	50	-0.034	0.28	
	1580	-0.02	0.52	
	15810	-0.026	0.4	
	100000	-0.035	0.27	
<i>No Migration</i>	1	0.0061	0.78	
	5	0.028	0.21	
	158	-0.065	3.5×10^{-3}	
	1581	-0.053	0.02	
	10000	-0.11	6.3×10^{-7}	***
<i>No Recombination</i>	1	0.0058	0.8	
	5	-0.039	0.08	
	158	-0.0011	0.96	
	1581	-0.0048	0.83	
	10000	-0.0055	0.81	

Table S3: Pearson's correlation tests for the association between the coefficient of background selection (B) and F_{ST} (average of ratios). P -values and R are computed on the simulations with BGS only.

<i>Treatment</i>	<i>Generation after split</i>	<i>R</i>	<i>P</i>	<i>Bonferroni correction</i>
<i>Default</i>	1	0.032	0.15	
	5	0.065	3.6×10^{-2}	
	158	0.04	0.07	
	1581	0.028	0.20	
	10000	0.051	0.02	
<i>High</i>	1	-0.013	0.57	
	5	0.023	0.31	
	158	0.058	9.4×10^{-3}	
	1581	0.012	0.6	
	10000	0.041	0.07	
<i>Human</i>	1	-0.011	0.62	
	5	0.064	4.5×10^{-3}	
	158	0.022	0.32	
	1581	0.036	0.11	
	10000	0.031	0.17	
<i>Large N</i>	10	0.048	0.13	
	50	0.038	0.23	
	1580	0.038	0.23	
	15810	0.027	0.4	
	100000	0.04	0.2	
<i>No Migration</i>	1	0.014	0.54	
	5	0.073	1.1×10^{-3}	*
	158	0.022	0.32	
	1581	0.038	0.09	
	10000	-0.0012	0.96	
<i>No Recombination</i>	1	0.016	0.48	
	5	0.035	0.12	
	158	0.092	4.3×10^{-4}	**
	1581	0.097	1.5×10^{-5}	***
	10000	0.11	5.0×10^{-7}	***

Table S4: Pearson's correlation tests for the association between the coefficient of background selection (B) and d_{XY} . P -values and R are computed on the simulations with BGS only.

<i>Treatment</i>	<i>Generation after split</i>	<i>R</i>	<i>P</i>	<i>Bonferroni correction</i>
<i>Default</i>	1	0.085	1.35×10^{-4}	**
	5	0.084	1.35×10^{-4}	**
	158	0.082	2.46×10^{-4}	**
	1581	0.095	2.12×10^{-5}	***
	10000	0.093	3.28×10^{-5}	**
<i>High Migration</i>	1	0.056	0.01	
	5	0.054	0.02	
	158	0.056	0.01	
	1581	0.062	5.26×10^{-3}	
	10000	0.079	4.36×10^{-4}	*
<i>Human</i>	1	-0.036	0.11	
	5	-0.037	0.10	
	158	-0.047	0.03	
	1581	-0.067	3.0×10^{-3}	
	10000	-0.036	0.11	
<i>Large N</i>	10	0.097	2.2×10^{-3}	.
	50	0.096	2.3×10^{-3}	.
	1580	0.091	4.0×10^{-3}	
	15810	0.087	5.7×10^{-3}	
	100000	0.11	5.8×10^{-4}	*
<i>No Migration</i>	1	0.088	7.6×10^{-5}	**
	5	0.089	6.7×10^{-5}	**
	158	0.086	1.2×10^{-4}	**
	1581	0.052	0.02	
	10000	0.029	0.20	
<i>No Recombination</i>	1	0.086	1.3×10^{-4}	**
	5	0.085	1.6×10^{-4}	**
	158	0.088	8.8×10^{-5}	**
	1581	0.1	4.9×10^{-5}	***
	10000	0.16	2.3×10^{-12}	***

Table S5: Pearson's correlation tests for the association between the coefficient of background selection (B) and d_{XY-SNP} . P -values and R are computed on the simulations with BGS only.

<i>Treatment</i>	<i>Generation after split</i>	<i>R</i>	<i>P</i>	<i>Bonferroni correction</i>
<i>Default</i>	1	0.15	3.1×10^{-11}	***
	5	0.12	8.9×10^{-8}	***
	158	0.14	5.3×10^{-10}	***
	1581	0.16	2.8×10^{-13}	***
	10000	0.15	5.8×10^{-12}	***
<i>High Migration</i>	1	0.075	8.1×10^{-4}	*
	5	0.088	7.5×10^{-5}	**
	158	0.09	5.5×10^{-5}	**
	1581	0.12	5.7×10^{-8}	***
	10000	0.17	5.4×10^{-14}	***
<i>Human</i>	1	0.071	1.6×10^{-3}	.
	5	0.062	5.6×10^{-3}	.
	158	0.053	0.02	.
	1581	0.023	0.3	.
	10000	0.067	2.7×10^{-3}	.
<i>Large N</i>	10	0.22	6.4×10^{-12}	***
	50	0.21	1.2×10^{-11}	***
	1580	0.2	6.6×10^{-11}	***
	15810	0.19	2.2×10^{-9}	***
	100000	0.25	2.7×10^{-15}	***
<i>No Migration</i>	1	0.15	2.3×10^{-11}	***
	5	0.16	1.5×10^{-12}	***
	158	0.14	1.2×10^{-10}	***
	1581	0.048	0.03	.
	10000	-0.0014	0.95	.
<i>No Recombination</i>	1	0.18	6.3×10^{-16}	***
	5	0.16	4.0×10^{-13}	***
	158	0.18	2.7×10^{-15}	***
	1581	0.22	3.6×10^{-23}	***
	10000	0.3	6.6×10^{-42}	***

- 1 Figure S1: Overall distribution of selection coefficient in the heterozygotes. There
- 2 are 2% lethal mutations, and the average selection coefficient of the non-lethal
- 3 mutations is approximately 0.07.

- 4
- 5
- 6
- 7
- 8

Figure S2: Relationship between B values computed for all simulations using the stickleback genome using the methods of Hudson and Kaplan (1996) and Nordborg et al. (1997).

Figure S3: Comparisons of means F_{ST} (left column), d_{XY} (central column), and H_S (right column) between simulations with (black) and without (grey) BGS for all unrealistic treatments. Realistic treatments (*Default*, *No Migration*, *High Migration*, *Human* and *Large N*) are in Figure 1 in the main text. Error bars are 95% confidence intervals.

Figure S4: Correlations between B and F_{ST} , d_{xy} , F_{ST} (average of ratios), d_{xy-SNP} , and F_{ST} (average of ratios) after removing all loci that have minor allele frequency (MAF) lower than 0.05 (called MAF - F_{ST} (average of ratios)) for the treatment *No Recombination* only at the last generation ($5 \times 2N$ generations after the split). Each grey dot is a single simulation with BGS. The large black dot is the mean of all simulations without BGS. The P -values are computed from a Pearson's correlation test. P -values and R are computed on the simulations with BGS (grey dots) only.

Figure S5: Relationship between total heterozygosity H_T and F_{ST} on a site per site basis. Each dot represents a single site. The black line is a Local Polynomial Regression (LOESS). Data is a random subset from the *Default* treatment at the last generation.