Reviewer 1

1. I was dissatisfied by the fact that this study solely uses simulations, while substantial existing theory on BGS could be paired with the structured coalescent under the IM model to explore a much wider parameter space. I am not against simulations, as these are necessary to model complex genomic architectures and demographies. However, without a connection to well-understood marginal cases, I find it difficult to assess the simulated outcomes. The motivation behind the fairly complex simulations done here is to simulate 'realistic' scenarios. However, I am afraid that this 'realism' is compromised by crucial assumptions that should be better supported before they can be used to justify the complexity of the simulations. This was a general impression. I refer to specific issues below.

We have now increased the breadth of conditions simulated, in response to the reviewer's other comments. We believe that simulations are actually the appropriate approach to a topic as potentially complicated as this, given that the existing theory neglects many aspects of biology which conceivably could be important, such as linkage disequilibrium between selected loci, changes to effective migration, etc. Modifying theory to account for these other issues would be difficult. In any case, we believe that these simulations will be very useful to the community, as we currently only have misleading results based on specifically unrealistic parameter values to work from.

2. To what extent can the findings obtained here for humans and stickleback be extrapolated to other species? How generally do the findings apply to sticklebacks, even? I presume that e.g. marine vs freshwater ecotypes, and to some extent lake vs stream populations, have quite different demographic histories. Since demographic history may affect the efficacy of selection (including BGS, see e.g. Torres et al. 2017), I am concerned about the generality of the conclusions. As these conclusions are currently stated quite generally, I suggest the authors make it more clear that they are based on two exemplary genomes and a single, simple demographic scenario.

As we mention in response to the other reviewer, our simulations are not meant to mimic the specific biology of any particular organism. We use the human and stickleback recombination maps to give realism to the distribution of gene density and recombination rates. Having said that, the results in the paper are not based on only two scenarios, but nine, varying migration rates and recombination rates, as well as recombination maps and the distribution of selection coefficients. In any case, as we mention above, previous papers used single sets of unrealistic parameter values, which were chosen to produce an effect, and the results are strongly different when those parameters are returned to their more realistic values even across this spectrum of variation.

3. The authors have written an apparently new software (SimBit) for their simulations. While they state that results were double-checked against SFS_code (Hernandez 2008), they do not show these checks (l. 152–155). I suggest the authors show these comparisons, as it would make the verification of their results more transparent.

Good point! We performed more simulations and comparing them with previous papers as well as with SLiM and Nemo, and now present them in an Appendix.

To verify the accuracy of SimBit, we performed more simulations to compare our results with Zeng and Corcoran (2015). We also repeated the same simulations with both SLiM and Nemo and got consistent results. Interestingly, these simulations highlighted an interesting consequence on whether F_{ST} is computed from coalescence trees or from genetic diversity data. We have added the results of these extra simulations in appendix A.

In addition, we also demonstrate that we get the same results as Charlesworth et al. (1997).

4. Sampling of genomic architectures uniformly w.r.t. the physical map (l. 171 ff) likely results in a non-uniform representation of the recombination landscape. If a small proportion of the genome exhibits low recombination, low-recombination regions – which are expected to show stronger signals of BGS – will be underrepresented. Therefore, the sampling regime currently implemented by the authors may fail to detect stronger effects

of BGS simply because the weighting is uniform w.r.t. to the physical map, but not the linkage map. From a practical point of view, if BGS were to create a false signal of local adaptation only in low-recombination regions, this might be enough to produce false positives in an empirical study at a rate that leads to wrong conclusions. To address this concern, I suggest the authors complement their existing analyses with separate analyses for, say, the 5% of the genome with the highest and lowest recombination rates, respectively.

We indeed sampled uniformly from the physical map. This choice of sampling aimed at producing a distribution of sequences selected that resemble most what would be present in most empirical data sets. Indeed, weighting by the inverse of recombination rate as suggested by the reviewer would over-sample the low-recombination regions and bias the results. Note that we have also performed simulations in which recombination was set to 0 (*No Recombination* treatment) which robustly shows the outcome of no recombination regions (and there is indeed still no large effect of BGS on F_{ST} there).

5. I did not see any obvious justification for why mutation rate should vary according to an exponential distribution along the genome (l. 167–170). In fact, I think that this assumption might introduce far too much variation and drive some of the observations. Specifically, because of the strong leptocurtic shape of the exponential, the current simulation scheme might assign too much weight to lover-than-average mutation rates, and so underestimate the effects of BGS. To address this concern, I suggest the authors repeat some of their simulations with a constant mutation rate and compare the results to the ones obtained with an exponential distribution.

We have now simulated a new treatment with constant mutation rate. The results are consistent with other treatments. See Figures 2 and 4.

6. The authors assumed a distribution of selection coefficients that amounts to an average selection coefficient per non-lethal deleterious mutation of 0.07 (l. 208–212). This seems very high and is more than a magnitude higher than what McVicker et al. (2009) inferred for *conserved* exonic sites (0.0025) in humans. I therefore suspect that

deleterious mutations are so rapidly removed from the simulated genomic regions that the effect of BGS on (at least the human) genome is strongly underestimated (except for the scenario without recombination). The reduction in effective population size (or *B*) due to BGS scales inversely with the selection coefficient *s*: in other words, the higher *s* the higher *B*. Indeed, Remi & Whitlock find that the mean across their simulated focal regions is much higher (0.975) than the one inferred by McVicker et al. (2009) if selection is restricted to conserved exonic sites (90% CI ranging from 0.74 to 0.81).

If I assume a genomic region of length 10 Mb spanning a total map length of 1 cM, which roughly corresponds to a genomic region simulated by the authors, and that 1.75% of the sites in this region are exonic with a deleterious mutation rate of 2.5×10^{-8} per exonic site, then the *B* statistic calculated using Eq. 8 of Hudson & Kaplan (1995) [HK95] roughly connects the value pairs (*s*, *B*) pertaining to Remi & Whitlock and McVicker (2009), repsectively (see Fig. 1 [MEC_Remi&Whitlock_fig1.png]). In other words, by simply changing the distribution of selection coefficient such that the mean is ~0.0025 rather than 0.07, but keeping the other parameters constant, the authors might obtain more plausible BGS simulations.

If, instead of Eq. 8 of HK95, I use the equation on l. 283 given by the authors, then the mean seleciton coefficient would have to be reduced from 0.07 to at least about 0.015 in order to reproduce the mean *B* found by McVicker et al. (2009) (Fig. 2 [MEC_Remi&Whitlock_fig2.png]). Overall, since the mean selection coefficient used in the simulations is about an order of magnitude too high, and the mean *B* predicted by these simulations is about 20% below previous estimates of the mean *B* in humans, I am concerned that simulations might not be appropriately representing the pressure of purifying selection in the human genome.

We have now simulated a new treatment with constant lower selection pressures as suggested by the reviewers. The results are consistent with other treatments. Because we use a distribution of *s*, the mean of *s* should have little effect on the results, because large effect loci contribute little to BGS and are relatively rare. Removing them from the distribution should cause little change.

7. The authors assumed migration rates of $m \in \{0, 0.005, 0.05\}$ and population sizes of $N \in \{1000, 10\,000\}$. In the "Default" and "Human" scenarios studied by the authors, the scaled migration rate is $M = 4Nm = 4{\times}1000{\times}0.005 = 20$, which seems very high for humans, at least on an inter-continental scale. Akey et al. (2002) and others (Elhaik 2012 PLoS One) reported average F_{ST} estimates on the order of 10%, which is an order of magnitude higher than the F_{ST} for the "Default" and "Human" scenarios reported by the authors in Fig. 1 of their manuscript (left column). In the limit of high migration, BGS has a decreasing effect on F_{ST}, and so if m is chosen too high, then there is little scope for detecting BGS at all.

To illustrate this, I have plotted the equilibrium $F_{ST} = 1/(1 + 8BNm)$ (note that this applies only to the later stages of the simulations, i.e. a large enough time since the population split, and so likely represents an overestimate) as a function of $M = 4Nm$ in Fig. 3 for three values of B. The vertical grid line indicates $M = 20$, and the dashed and dot-dashed horizontal grid lines represent an F_{ST} of 0.02 and 0.1, respectively. My impression is that the combination of an inflated selection coefficient (and, hence, an inflated mean B; see point **6.** above) with too high a migration rate m explains why the authors find no local signal of BGS. In contrast, with a value for $m \approx 0.0012$, would be on the order of 0.1. Then, combined with the average B inferred by McVicker et al. (2009) of about 0.775, the effect of BGS on F_{ST} would be expcected to be on the order of a 20% reduction compared to $B = 1$ (compare black to blue curves in Fig. 3 [MEC_Remi&Whitlock_fig3.png]). These considerations apply to the *average*; regions of low recombination may well exhibit a much stronger increase in F_{ST}. I am not an expert in stickleback demographic history, but presume that similar arguments could be made there.

Our label for the so-called ”human” treatment was unfortunate and misleading, as we have now changed it. That label merely referred to the recombination map we used in that case, and was not intended to mean that we tried to mimic any other aspects of human biology. We have changed the label to clarify this. Please note that we also varied migration rate in other treatments.

8. I was uncertain about how the migration rate m was determined for the FDist2 simulations. Depending on the answer of the authors to my question related to l. 301–310 (see below), the estimated false positive rates might not be appropriate, and this could explain why scenarios with and without BGS showed similar false positive rates except for a few extreme scenarios.

We have clarified that (see paragraph starting at line 322) and have rerun the entire Fdist2 procedure to confirm that we get the same results. We also considered different methods to estimate the migration rate to make sure that we consistently see no difference between treatments with and without BGS. We show the one method for which the observed mean F_{ST} of the Fdist2 simulations are most similar to the mean F_{ST} of the SimBit simulations.

Minor Comments

C: Comment
Q: Question
S: Suggested change
R: Requested change

Introduction

[l. 42–43] R: '...reduce the effective population size of linked loci.' -> '...reduce the effective population size and distort the site frequency spectrum (SFS) of linked loci.'

Edited

[l. 51–53] S: '..., increasing F_{ST},...' -> '..., which increases F_{ST},...'; S: '..., keeping F_{ST},...' -> '..., which keeps F_{ST},...'

Edited

[l. 55] C: The mutation (μ) and migration (m) rates are not defined at this stage.

Edited

[l. 83–89] C: The first question would seem equally important in the context of identifying loci under positive selection, as an increased background level of F_{ST} also alters the power to detect outliers.

Indeed, larger F_{ST} would change the power of outlier approaches, but this reduces the probability of both false and true positives. We focus on the purported source of bias in the outlier approach caused by BGS.

S: 'Locus-to-locus variation in F_{ST} potentially could...' -> 'Locus-to-locus variation in F_{ST} due to BGS potentially could...'

Edited

[l. 89] S: Perhaps briefly introduce the main reasons for why BGS might vary along the genome already here: i) variation in the intensity of selection (selection coefficient, density of targets of selection and hence mutation rate), ii) variation in recombination rate, iii) effects of other modes of selection at linked sites (e.g. a selective sweep, Hill-Robertson interference).

Good point! Edited at line 75.

[l. 103–105] S: See Aeschbacher et al. (2017) PNAS for a two-step procedure that avoids the confounding effects of BGS on the inference of divergent selection in the face of gene flow.

Edited

[l. 113–117] C: As far as my recollection of the existing literature goes, it is not claimed that d_{xy} should not depend on BGS, but that, as a function of recombination, it should do so in an opposite direction to the effect of divergent selection in the face of gene flow. The argument is then rather that, because BGS reduces the signal of local adaptation on F_{ST} but potentially enhances the signal of local adaptation on d_{xy}, from a conservative point of view, the former (d_{xy}) should be preferred over the latter (F_{ST}) as a means of identifying divergent selection. I might be missing specific references supporting the authors' claim, though. In this case, I suggest these references be added.

See Cruickshank and Hahn (2014), box 1, figure 1 as well as the last paragraph of page 3137 that ends at page 3138, where the claims are made that "d_{XY} is independent of the levels of diversity within the two populations" and d_{xy} "is not affected by variation in levels of current polymorphism" and other similar statements. See also Hoban et al. (2016). References added in the text.

[l. 123–125] S: I missed a citation of Elyashiv et al. (2016) here, as they provide a nice example of how a joint model of BGS and positive selection substantially increases the fit to observed, small-scale genomic variation in diversity in *Drosophila* as compared to models with only one mode of selection.

Citation added here too

Methods

Added

[l. 136–137] C: I missed variation in the strength of selection (i.e. the selection coefficient) as a reason for variation in BGS

clarified

[l. 142] R: If data are not shown, there should be at least a reference.

Edited. Simulated the same scenarios than in Zeng and Corcoran (2015) and reproduced these simulations with both SLiM and Nemo, too. See Appendix A.

[l. 143-145] C: Humans are one of the few mammal species for which a non-LD based recombination map *is* available! See e.g. Kong et al. (2002) Nat. Genet., Kong et al. (2010) Nature, Bhérer et al. (2017) Nat. Commun. I wonder why the authors did not use these resources in order to avoid the bias mentioned on l. 145–148.

We started the project considering that using one map estimated from LD data and one map estimated from pedigree data would be a good idea to avoid potential systematic bias from any method of estimation.

[l. 147–148] S: '..., hence artificially [...] of BGS.' -> '..., which might bias the simulated variance in the intensity of BGS.'

Edited

[l. 148] S: 'fine scaled' -> 'fine-scaled'

Edited

[l. 168–170] R: From the current formulation, it is not clear whether the reference to Nachman & Crowell (2000) applies to the mean of 2.5×10^{-8} only or also to the assumption of the exponential distribution (only the former seems justified). Please clarify.

Clarified

[l. 179] S: '...plus the recombination rate present in the focal region of...' -> '...plus the map distance covered by the focal region of...'

Edited

[l. 182] S: 'with' -> 'as'

Edited

[l. 184] R: It was not clear what exaclty is meant by 'in blocks of up to 100 nucleotides'. What criteria determine the block size? Please clarify

This is now explained in the lines that follow.

[l. 210–212] R: '...we used multiplicative dominance,...' -> '...we assumed multiplicative fitness interactions among alleles,'.

Edited

Q: How does this regime improve the performance of the simulations as compared to, e.g., additive interactions?

SimBit tracks the fitness effects of each sequence of the genome and, if no recombination has happened in these sequences, then we can just multiply the stored fitness effect of the entire sequence with the fitness effect of the entire sequence at the other haplotype without having to loop through each site at every generation. It hence greatly speeds up fitness calculation. This trick is a big part of what makes SimBit so much faster than other existing pop gen simulators (see Appendix A).

[l. 219] R: 'quasi lethal' -> 'quasi-neutral'

Edited

[l. 220, 221] R: 'selective coefficients' -> 'selection coefficients'

Edited

[l. 223–225] S: '...mutation rate occurs in non-coding...' -> '...mutations occur in non-coding...'; S: '...simulations using human genome...' -> 'simulations of the human genome...

Edited with slight change

[l. 224–231] C: Given the growing evidence that purifying selection not only acts on coding regions in the human and stickleback genomes, it may seem appropriate to relax the restriciton of deleterious mutations only occurring at exonic (and regulatory sequences). While I agree that relaxing this constraint might reduce the variation in the intensity of BGS at a small scale, the positive correlation between gene density and recombination rate (at least in humans) might cause purifying selection to be disproportionally underrepresented in low-recombination regions under the current simulation regime.

This assumption is only made for the stickleback genome and not when using the human genome.

[l. 239] R: Insert 'the' after 'explored'.

Edited

[l. 243, 246] R: Italicise 'm' and 'N'.

Edited

[l. 251] R: Insert 'the' after 'where'

Edited

[l. 252] S: 'set at zero' -> 'set to zero'.

Edited

[l. 253] R: '...results Charlesworth...' -> '...results of Charlesworth...'.

Edited

[l. 283] C: The formula for B does not seem identical (at least at first sight) to any equation in Hudson & Kaplan (1995). Please clarify.

Indeed, the mistake is related to a failure from our side to realize that Hudson & Kaplan (1995) define μ as the diploid mutation rate, unusually. Thank you for noticing it! We have recomputed our *B* values and have edited the formula. The new results are similar to the previous one. The match between our mean *B* values and those found in previous studies has been improved by actually using the corrected formula.

[l. 285] C: Replace full stop after 'site' by ', and'.

Edited

[l. 290–291] S: As far as I know, the B statistics from McVicker et al. (2009) are availble, and so I suggest you compare your B values with those.

Our *B* values are specific to each simulation but not so specific to each sequence as selective coefficients are randomly assigned to each exon (following the specified DFEs). We therefore cannot attempt a meaningful correlation.

[l. 301–302] R: The denominator in the formula for m should be changed to $1 - F_{ST}$.

Edited

[l. 301–310] Q: It was not clear which 'average F_{ST} ' is referred to on l. 302. Am I right in assuming that the average is taken from the SimBit simulations under the respective scenario, i.e. including BGS in some cases?

Yes, you are right! Clarified at line 431

If so, then I am worried that the migration rate estimated as

$$m = (1 - F_{ST})/(8NF_{ST})$$

is affected by BGS in two ways, namely by the reduction in N_e and the reduction of m to an effective migration rate m_e, both due to BGS. This means that the average F_{ST} is higher in the BGS simulations, such that the m used for the FDist2 simulations with BGS is lower, and hence the P-value is inflated in the BGS scenarios.
Would this not lead to an underestimation of the false positive rate in the case of BGS w.r.t. the neutral scenario? In other words, if you calibrate your test separately for simulations with and without BGS, isn't it expected then that you see *no* difference in false positive rates? I realise that my reasoning may not be correct, and that the authors *do* see differences in the false positive rate between BGS and neutrality in the CNC97 scenarios.

The FDist2 procedure is blind as to whether or not there is BGS in the simulations. The fact that BGS affect both N_e and m_e would also be true in nature. If we were to somehow parametrize our FDist2 procedure in order to take into account the effect of BGS on m_e, then we would not make an accurate representation of the effect of BGS on the FPR of tests as they are applied.

[l. 311–312] Q: Did the authors assess the effect of the filter for a minimum minor allele frequency?

We are using a filter as the vast majority of authors do. We have not explored the effect of not using such filter on the FPR. We have however analyzed the F_{ST} distributions (and distributions of other statistics) with and without this filter.

[l. 326] S: 'estimate of' -> 'estimator of'

Edited

[l. 351] S: Since the relationship among the statistics is not always linear, I suggest using a Spearman rank correlation test instead of a Pearson correlation test.

We have used Spearman correlation tests on top of Pearson correlation coefficient, ordinary least squares regressions and, robust regressions (using M-estimators; Huber, 1964) and got consistent results. The main reason for our decision to report Pearson correlation test is because we wanted to report Pearson's correlation coefficient, as these correlation coefficients are more meaningful than the P.values for our study. To be 100% safe, we have now also computed permutation tests and have decided to report Pearson correlation coefficients alongside the P.values from these permutation tests.

Results

[l. 358] Q: Standard errors of what?

Edited

[l. 407] S: Drop the commas.

Edited

[l. 409] Q: How can these high false positive rates for the 'No Migration' scenario be explained?

Many of these SNPs have a F_{ST} of 1.0. We are reaching a point where, there is no way to tell two loci apart as the statistic used is saturated. We think that, per consequence of that, an F_{ST} outlier method has no way to tell apart what is and what is not under selection. We would also note that the migration rate is estimated from the average Weir and Cockerham F_{ST} coming from the one SNP per simulation. As we used a MAF threshold (5%), that might cause a mismatch.

We have rerun these Fdist2 simulations, exploring different methods to estimate the migration rate but keep getting similar results. We have increased our discussion on the No Migration results in the discussion. Also, as most the FPR variation was hard to distinguish in the graph caused by the high FPR values found in the treatments No Migration and CNC97, we moved these two treatments in supplementary figures. Rerunning these simulations lead to absence of significant FPR difference in the No Migration case leading a slight rewriting of our results.

Discussion

[l. 415] C: As explained in detail above, I am unsure as to whether the parameter combinations are really that 'realistic'.

See comments above

[l. 418] S: 'over' -> 'above'

Edited

[l. 435–438] C: This statement dit not make sense to me.

Clarified

[l. 439–444] C: I suspect that what is dicussed here strongly depends on the problematic choice of parameter values for *s* and *m*, as explained above. I therefore think that the conclusion on l. 443–444 is not justified.

We have used three very different parameter values for *m* and keep getting a consistent result. We have now added two new treatments; the *Constant μ* and the *Low selection pressure* treatment to investigate a greater genetic parameter space.

[l. 445–446] C: I agree that BGS needs to studies more, but given the growing existing literature on the genome-wide effects of BGS, I missed some references to this work.

We meant to refer to the effect of BGS on F_{ST}. It was unclear and has been clarified. On the effect of BGS on $F_{ST,}$ the literature is not that extensive. We made sure to recite the two most important papers in our opinion, Charlesworth et al. (1997) and Zeng and Corcoran (2015).

[l. 452–457] C: I see that simulating larger populations is hard, but couldn't the authors use a scaling argument to approximate the effect of larger *N*? As mentioned above, in the case of the 'Human' scenario, I think one does not need to invoke a drastically different pouplation history or *N* to explain the surprising findings by the authors. It may be sufficient to use more realistic parameter values for *s* and *m*.

We have now considered other values for *s* and *m*, as described above. Our goal with the paper was to consider moderate F_{ST} values as may typically be seen within a species; the resulting F_{ST} values near 0.05 meet this goal well.

[l. 456] S: 'yield to' -> 'result in'

Edited

[l. 464–467] C: This is another statement that did not make sense to me. If there are no local effects of BGS, how should there be global ones?

We do not claim that F_{ST} is completely insensitive to BGS but that it is robust to it. We tend to think that on a genome-wide scale F_{ST} would be affected by BGS, at least in

cases with low genomic rates of recombination, but performing realistic simulation of an entire genome would be extremely RAM consuming and we could not run such simulations.

[l. 479–482] C: I found this part not very conclusive.

We are not sure which part of this paragraph that the reviewer is unsure of.

[l. 489–490] C: I don't think that BGS does not affect the rate of fixation of mutations arising after the populations diverged. It does affect N_e, and so in a finite-sized population certainly also affects the fixation probability and rate of mutations. I think the point is more that after a complete separation of the two pouplations, there is no way two lineages in different demes can come together. From a backward-perspective, no coalescence can occur until time t if m = 0. This is why BGS makes no contribution to d_{xy} during the isolation period.

We have edited the sentence to make it clearer that we are referring to the rate of fixation of neutral mutations. This rate is unaffected by BGS, because the rate of substitution of neutral mutations is simply the mutation rate, independent of drift or hitchhiking.

[l. 490–496] C: These considerations seem correct under an isolation model, but in a symmetric model with migration $d_{xy} = 4N\mu + \frac{2\mu}{2m}$ at equilibrium, i.e. in the limit of large t. Therefore, if m is large enough, $4N\mu$ will remain large relative to μ/m, and so BGS can affect d_{xy} considerably over extended amounts of time.

We clarified that we were indeed talking about isolated populations.

[l. 497] C: The statement that d_{xy} becomes less sensitive to BGS when F_{ST} becomes more sensitive cannot apply generally. In a model without migration, F_{ST} converges to 1, and so cannot be sensitive to BGS anymore, while d_{xy} keeps increasing and remains, albeit weakly, sensitive to BGS.

The statement has been made by others before us (we add one more reference, at line 532, compared to the version we first submitted), we only meant to draw parallel between our simulations and previous results. We do not recall that previous works have explicitly addressed the extreme case when F_{ST} approaches one, so we made sure to briefly mention it at lines 533-535.

[l. 483–504] C: What is missing in the current discussion of F_{ST} and d_{xy} is that, in the long term and with non-zero migration, the effects of BGS and divergent selection on F_{ST} are totally confounded (m and N_e enter only via their product). In contrast, $1/m$ and N_e enter as separate terms into the equilibrium expression for d_{xy}, such that BGS at first approximation is expected to reduce d_{xy} via a reduction of N_e, and divergent selection is expected to increase d_{xy} via a reduction of the effective migration rate. This latter property of d_{xy} makes it preferable as a statistic, although in general d_{xy} is less sensitive to changes in m and N_e than is F_{ST}.

It’s not entirely clear to us what the reviewer means here. In principle, when looking for outliers caused by divergent selection, we want a value that is not sensitive to N and m

at all, but instead is sensitive to the additional effects of local selection. The real difficulty is how greatly d_{xy} is sensitive to mutation rate and other forces that influence the absolute amount of diversity, including variation in BGS.

[l. 510–518] R: Because mutations are only partially recessive (and there certainly is no heterosis), I actually doubt that the effective migration rate is increased in the simulations done here. I rather suspect that the effective migration rate is decreased, especially because the multiplicative fitnesses are very close to the additive case for s ≤ 0.07, and so I do not see why immigrant alleles would profit from recessivity. Related to this point, Harris & Nielsen (2016) used simulations to show that the proportion of deleterious Neandertal-derived mutations in humans increases only if deleterious mutations are fully recessive. If they are partially recessive (h = 0.1), then they behave much like additive mutations and their proportion decreases over time.

This discussion point was general and does not relate to the specifics of our simulations. However, we will add here that we are referring here not to the increase of deleterious alleles in migrants, but instead of beneficial alleles, which may complement the deleterious effects of locally accumulated mutations. There is an extensive literature showing that this can have substantial positive effects on the effective migration rate.

[l. 519–524] Q: But would the deleterious mutation rate in such highly mutable regions not also be increased? Then, the increased strength of BGS might cancel the increase in the baseline mutation rate.

This is true, but the point of the paragraph is that mutation can also directly affect F_{ST}, rather than through the relatively weak effect on F_{ST} via BGS.

I am therefore not sure that the current argumentation about the effect of autocorrelation of mutation rates holds.

We merely said that this effect could impact the correlation. We did not claim that it would generate a particular direction of effect overall.

[l. 530–531] C: In the absence of strong evidence for strong and genome-wide signals of local adaptation and (hard) sweeps, I found this statement quite ambitious. At least for *Drosophila*, Elyashiv et al. (2016) nicely show that both BGS and selective sweeps are required to explain small-scale variation in genomic diversity.

This paragraph cited by the reviewer is referring to the patterns in F_{ST} caused by BGS, and is supported by the evidence that we present. Elyashiv et al. is about within population diversity, and our results also show an effect of BGS on within population diversity. (see the first sentence of the first paragraph of the discussion) and hence we do not contradict Elyashiv et al. (2016) (or many other papers). We have shown that BGS explains little to no variation in F_{ST}. We hence hypothesized (the sentence started with `It can be hypothesized`) that positive selection could be an important cause of F_{ST} variation.

We agree that despite that the sentence started with `It can be hypothesized`, it might sound a bit like we are claiming that positive selection explain much of the variation in F_{ST}. Also, it might have been unclear that we are talking about F_{ST} and not H_S. We hence edited the MS from

"Here we showed the BGS is unlikely to explain all of these correlations between F_{ST} and recombination rate. It can be hypothesized that positive selection (selective sweeps and local adaptation) could be the main cause of this correlation."

to

"Here we showed the BGS is unlikely to explain all of these correlations between F_{ST} and recombination rate. As positive selection has also been shown to have important effect on genetic diversity (Eyre-Walker & Keightley, 2009; Hernandez, Kelley, Elyashiv, Melton, & Auton, 2011; Macpherson, Sella, Davis, & Petrov, 2007; Sattath, Elyashiv, Kolodny, Rinott, & Sella, 2011; Wildman, Uddin, Liu, Grossman, & Goodman, 2003), it would be important to investigate whether positive selection (selective sweeps and local adaptation) could be an important driver of the correlations between F_{ST} and recombination rate. More research would be needed to investigate whether this is true."

[l. 549–552] C: Given the current version of the analyses, I am seriously concerned about whether this statement is valid.

We hope that the clarifications offered will convince you of the quality of our analysis and of how conservative we have been with this study.

References

[l. 573] R: Italicise 'D. melanogaster'.

Edited

C: The page range seems to be missing or incomplete for a number of references.

Edited

[l. 729–730] C: This paper has been published in the meantime.

Edited

[l. 757–758] R: 'acids research' -> 'Acids Research'

Edited

Figures and tables

[Fig. 3] C: I wonder if it would make sense to display the y axis on a log scale. I could not see the asterisks and dots that should indicate the significance based on the Welch t-test.

We think a log scale might just add some extra mental load to read the graph.
The asterisks were missing. Now added.

Supporting material
[Fig. S2] Q: Is the reference to Hudson & Kaplan (1995), rather than (1996)?

Indeed. Edited!

[Tables S1–S5] C: I may have missed it, but $\mathcal{R}$ does not seem to be defined.

The tables are referred to from the first paragraph of the results (as well as later in the text).