

Review of MEC-18-0562 Remi & Whitlock

Summary

In their manuscript "Background selection and the statistics of population differentiation: consequences for detecting local adaptation", Remi & Whitlock study the effect of variation in the intensity of background selection (BGS) along the genome, and address whether this variation affects the detectability of local adaptation. Accepting that BGS can affect genome-wide average patterns of genetic diversity, the authors focus on small-scale (locus-to-locus) variation in the intensity of BGS as it supposedly arises in real genomes of natural populations. This question is of great interest to the field, as sensible interpretation of patterns of genomic variation requires a detailed understanding of the joint signals of various evolutionary forces, and of the confounding effects individual forces have on the genomic signature of other forces. Remi & Whitlock use their own software to perform forward simulations of an ancestral population that splits into two populations t generations ago. These two daughter populations then may or may not exchange gene flow at a rate m . As such, this is the classic and well-studied isolation-with-migration (IM) model. To make these simulations 'realistic', the authors simulate subsamples of the human and stickleback genome. Remi & Whitlock track commonly used summary statistics of genetic diversity within and between populations, including the pairwise nucleotide difference d_{xy} and the relative population differentiation F_{ST} . The authors compare two estimators of F_{ST} with different statistical properties, and find that the one introduced by Weir & Cockerham (1984) is largely insensitive to variation in BGS intensity if populations are connected by gene flow. Using their simulations under BGS and neutrality, Remi & Whitlock also find that the false positive rate of an F_{ST} outlier test aimed at detecting loci under divergent positive selection is not much affected by BGS.

While Remi & Whitlock address a very important issue in population genomics, the manuscript in its current stage falls short of providing the necessary evidence to justify some of the stronger claims that are made. I feel that more justification is required, as otherwise the work presented here is at risk of being misinterpreted. Along these lines, I would request that the authors embed their hypotheses, findings, and interpretations more carefully in the recent literature about the effects of various forms of selection at linked sites, and use analytical results for simple marginal cases of their scenario to support their simulation results. Unfortunately, the lack of evidence is paired with a number of potential technical flaws (at

least, I found it hard to convince myself of the contrary based on the material presented) and some apparent misinterpretations of previous work. Overall, I feel that this amounts to a major revision at least, and requires substantial efforts in any case.

Major issues

- 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.
- 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.
- 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.
- 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.

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 lower-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.

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 B 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), respectively (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 selection 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.

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 $m \approx 0.0012$, F_{ST} 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 expected 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.

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 I. 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.

Minor comments

C: Comment

Q: Question

S: Suggested change

R: Requested change

Introduction

- [I. 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.'

- [l. 51–53] S: '..., increasing F_{ST} ,...' -> '..., which increases F_{ST} ,...'; S: '..., keeping F_{ST} ,...' -> '..., which keeps F_{ST} ,...'
- [l. 55] C: The mutation (μ) and migration (m) rates are not defined at this stage.
- [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. S: 'Locus-to-locus variation in F_{ST} potentially could...' -> 'Locus-to-locus variation in F_{ST} due to BGS potentially could...'
- [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).
- [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.
- [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 d_{xy} but potentially enhances the signal of local adaptation on F_{ST} , 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.
- [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.

Methods

- [l. 136–137] C: I missed variation in the strength of selection (i.e. the selection coefficient) as a reason for variation in BGS
- [l. 142] R: If data are not shown, there should be at least a reference.
- [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, Bhérier 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.
- [l. 147–148] S: '..., hence artificially [...] of BGS.' -> '..., which might bias the simulated

variance in the intensity of BGS.'

- [l. 148] S: 'fine scaled' -> 'fine-scaled'
- [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.
- [l. 179] S: '...plus the recombination rate present in the focal region of...' -> '...plus the map distance covered by the focal region of...'
- [l. 182] S: 'with' -> 'as'
- [l. 184] R: It was not clear what exactly is meant by 'in blocks of up to 100 nucleotides'. What criteria determine the block size? Please clarify
- [l. 210–212] R: '...we used multiplicative dominance,...' -> '...we assumed multiplicative fitness interactions among alleles,'. Q: How does this regime improve the performance of the simulations as compared to, e.g., additive interactions?
- [l. 219] R: 'quasi lethal' -> 'quasi-neutral'
- [l. 220, 221] R: 'selective coefficients' -> 'selection coefficients'
- [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...'
- [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 restriction 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 disproportionately underrepresented in low-recombination regions under the current simulation regime.
- [l. 239] R: Insert 'the' after 'explored'.
- [l. 243, 246] R: Italicise 'm' and 'N'.
- [l. 251] R: Insert 'the' after 'where'
- [l. 252] S: 'set at zero' -> 'set to zero'.
- [l. 253] R: '...results Charlesworth...' -> '...results of Charlesworth...'
- [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.
- [l. 285] C: Replace full stop after 'site' by ', and'.
- [l. 290–291] S: As far as I know, the B statistics from McVicker et al. (2009) are available, and so I suggest you compare your B values with those.
- [l. 301–302] R: The denominator in the formula for m should be changed to $1 - F_{ST}$.
- [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? If so, then I am worried that the migration rate estimated as $m = (1 - F_{ST}) / (8N F_{ST})$ 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.

- [l. 311–312] Q: Did the authors assess the effect of the filter for a minimum minor allele frequency?
- [l. 326] S: 'estimate of' -> 'estimator of'
- [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.

Results

- [l. 358] Q: Standard errors of what?
- [l. 407] S: Drop the commas.
- [l. 409] Q: How can these high false positive rates for the 'No Migration' scenario be explained?

Discussion

- [l. 415] C: As explained in detail above, I am unsure as to whether the parameter combinations are really that 'realistic'.
- [l. 418] S: 'over' -> 'above'
- [l. 435–438] C: This statement dit not make sense to me.
- [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.
- [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.
- [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 .

- [l. 456] S: 'yield to' -> 'result in'
- [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?
- [l. 479–482] C: I found this part not very conclusive.
- [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 populations, 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.
- [l. 490–496] C: These considerations seem correct under an isolation model, but in a symmetric model with migration $d_{xy} = 4N\mu + 2\mu/(2m)$ 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.
- [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.
- [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} .
- [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 $s \leq 0.07$, 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.
- [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. I am therefore not sure that the current argumentation about the effect of autocorrelation of mutation rates holds.

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

References

- [l. 573] R: Italicise 'D. melanogaster'.
- C: The page range seems to be missing or incomplete for a number of references.
- [l. 729–730] C: This paper has been published in the meantime.
- [l. 757–758] R: 'acids research' -> 'Acids Research'

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.

Supporting material

- [Fig. S2] Q: Is the reference to Hudson & Kaplan (1995), rather than (1996)?
- [Tables S1–S5] C: I may have missed it, but R does not seem to be defined.