

# Review of manuscript MEC-19-0061 by Matthey-Doret & Whitlock

## Summary

The authors present a simulation study showing that background selection (BGS) has little effect on locus-by-locus variation in  $F_{ST}$  in organisms with a recombination map and a distribution of genic regions that resemble those in humans or sticklebacks. The authors assumed a scenario of a single population that splits into two populations that are connected by varying degrees of migration (including a scenario of complete isolation). The authors explored treatments in which they varied the population size, selection pressure, mutation rate, and recombination rate. They also included a scenario of very strong BGS that was used by Charlesworth et al. (1997) to illustrate that BGS can lead to locus-by-locus variation in  $F_{ST}$ .

Compared to the first submission (which I also reviewed), the authors have made substantial improvements and addressed many of my earlier concerns. However, I still have a number of concerns and a few minor comments that should be relatively straight-forward to address, but that need to be fixed before this article seems ready for publication.

Overall, I think that this paper will make an important contribution to the field, mostly because it highlights that our understanding of the effect of BGS on neutral genetic variation is less complete than many may have thought. In that sense, I suspect that the article will trigger some theoretical follow-up studies.

## Major issues

1) The authors now more clearly state that they are interested in locus-by-locus variation in  $F_{ST}$  as compared to genome-wide patterns (l. 80–84). However, in a part of the Discussion the authors use their results to argue about genome-wide patterns, in particular about the correlation between  $F_{ST}$  and recombination rate (see l. 558–578). I think this is inconsistent, in particular because the authors do not report the correlation between recombination (and/or gene density) and  $F_{ST}$  observed in their simulations. At least in the case of humans, where there is only limited evidence for hard selective sweeps and genome-wide patterns of local adaptation, it seems remarkable that the authors suggest that a genome-wide positive correlation between  $F_{ST}$  and recombination rate should be driven by positive, rather than BGS. Along these lines, I think the authors should mention again in their concluding paragraph (l. 588–592) that the scope of their findings is restricted to locus-by-locus variation in  $F_{ST}$ .

2) For the  $F_{ST}$  outlier tests, the authors implemented a variant of FDist2, which assumes an island model of migration at equilibrium to produce the distribution of  $F_{ST}$  under the null model of no BGS. The authors say that the island model at equilibrium matches well with the demographic scenario simulated by the authors. However, this is only partly true. The No Migration treatment strongly violates the island model (see the greatly inflated false positive rate in Figure S4 for long times after the split). Even in the other treatments, the observed false positive rate significantly deviates from the expected value of 0.05 under neutrality in many cases. Can the authors explain this observation?

3) I am concerned that the authors have selectively moved the results that do not support their main argument to the SI. This holds for the CNC97 treatment in the case of both the dynamics of  $F_{ST}$ ,  $d_{XY}$ , and  $H_S$  as well as the false positive rate (compare Figures 2 and 4 to Figures S2 and S4). And it holds for the No Recombination and No Migration scenarios in the case of the dynamics of  $F_{ST}$ ,  $d_{XY}$ , and  $H_S$  (compare Figure 2 and Figure S2) and the false positive rate (compare Figure 4 and Figure S4), respectively. I suggest the authors present all scenarios in the main text.

4) The authors seem to make a fundamental distinction between  $F_{ST}$  estimated from allelic states (i.e. heterozygosity) and  $F_{ST}$  estimated from coalescence times (l. 501; Appendix A). Ignoring the fact that the latter can usually not be directly observed (unless in simulations), I found this distinction confusing,

at least if one considers  $F_{ST}$  at neutral sites, assumes the infinite sites model of mutation, and ignores the issue of averaging across loci. The more important distinction seems to be between different *definitions* of  $F_{ST}$  (cf. Charlesworth 1998, Eq. 3). For a given definition of  $F_{ST}$ , the estimates based on allelic states and coalescence times are identical in expectation for sites that are not directly under selection, and so I suggest the authors revise line 501 and the respective position in Appendix A accordingly. The authors are aware of the alternative definitions of  $F_{ST}$  (e.g. l. 346–347), but nevertheless seem to be inconsistent in their use of these definitions: the formulae shown in lines 51, 323, and 493 are based on the definition of  $F_{ST}$  as an analogon to Nei’s (1973)  $G_{ST}$  (Eqs. 3d and 4d in Charlesworth 1998), but in lines 346 to 347 the authors say that they focus on Weir & Cockerham’s (1984) definition (Eqs. 3a and 4a in Charlesworth 1998).

Further confusion arises in Appendix A, where the authors provide coalescent-derived formulae for both Weir & Cockerham’s and Nei’s  $F_{ST}$  without saying which is which and without saying why two versions are needed. In addition, the authors suggest that Zeng & Corcoran (2015) used the “wrong neutral expectation” for  $F_{ST}$ . I agree that there is a difference between the definition of  $F_{ST}$  that Zeng & Corcoran used,  $F_{ST} = (\bar{T}_B - \bar{T}_S)/\bar{T}_B$  (see their Eq. 1), and the reference they give (Slatkin 1991; see his Eqs. 8 and 26). Slatkin’s (1991) definition is  $F_{ST} = (\bar{T}_B - \bar{T}_S)/(\bar{T}_B + \bar{T}_S)$ , where  $\bar{T}_S$  and  $\bar{T}_B$  are the expected coalescence times for two lineages sampled from the same and from different demes, respectively. Saying that Zeng & Corcoran (2015) used the wrong definition nevertheless reaches too far. They may cite the wrong reference, but what they use is consistent with  $F_{ST}$  as defined by Weir & Cockerham (1984) and several other authors in the relevant case of two demes (see Eq. 3a in Charlesworth 1998). This observation explains why the top red and the bottom black horizontal lines in Figure A2 of Appendix A coincide. But it also shows that the authors here seem to be estimating  $F_{ST}$  according to Nei’s definition, not to the one by Weir & Cockerham as claimed (compare bottom red horizontal line to simulated values in Figure A2 for the scenario with 13k neutral loci).

Related to the previous point, what confused me in Appendix A is that the authors think they can explain the good match between the upper black horizontal line in Figure A2 ( $F_{ST}$  by Zeng & Corcoran (2015) under negative selection) and the  $F_{ST}$  that the authors computed from coalescent times simulated using SLiM under negative selection: Either the authors have also used the same (“wrong”) definition of  $F_{ST}$  as Zeng & Corcoran (2015), or the supposed good agreement is actually not true.

5) Related to concern 4) above, it is not surprising that  $F_{ST}$  estimated from allelic states exclusively at sites under *direct* purifying selection strongly deviates from  $F_{ST}$  estimated from coalescence times at these sites (see Figure A2). Mutation and coalescence are not independent at sites under direct selection, but they are under the model of BGS considered here at neutral sites linked to sites under selection, even if linkage is complete. This is because deleterious mutations occur independently from the allelic state at linked neutral sites, and so on average and over sufficient time, there is no correlation (but increased variance in coalescence times at the neutral sites). It is true that  $F_{ST}$  can only be estimated from allelic states in empirical studies, but it is questionable whether the majority of sites in empirical studies is under *direct* selection. A substantial proportion of sites may be under the effect of some kind of selection *at linked sites*, but for these, the majority of available theory suggests that an appropriately scaled coalescent process provides a very accurate approximation for the evolution at these neutral sites. From my understanding, Zeng & Corcoran’s (2015) theory only applies to these linked neutral sites, and Zeng & Corcoran’s estimates of  $F_{ST}$  were obtained only from neutral sites (specifically, from their genealogies, without simulation of neutral mutations).

Overall, Figure A2 seems to illustrate two main results: First, Zeng & Corcoran’s (2015) model for the effect of BGS on  $F_{ST}$  at linked neutral sites performs well (compare upper black horizontal line to light blue point and whiskers). Second, BGS can cause a strong increase in  $F_{ST}$  at tightly linked neutral loci compared to neutrality (compare yellow and brown points and whiskers of the scenario with 26k neutral and selected loci to the points and whiskers of the scenario with 13k neutral loci). It was not entirely clear to me how these observations would support the authors’ main conclusions from the main text.

It is great that the authors justify the implementation of their own simulator in Appendix A and that

they compare SimBit to other common simulators. Other than that, though, Appendix A introduces just as much confusion as clarification in its current form.

## Minor comments

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

### Abstract

- [l. 21] C: Please clarify if indeed Weir & Cockerham’s (1984) estimator of  $F_{ST}$  was used, and not the  $G_{ST}$  analogon by Nei (1973). See comment 4) above and minor comments below.

### Introduction

- [l. 72] S: “show” → “provide a”

### Methods

- [l. 136] R: Insert “of” in front of “attention”.
- [l. 154] S: “get” → “obtain”
- [l. 180–181] C: Here it says that in the flanking regions you only tracked (simulated) exons, but in lines 196 to 199 and 201 to 202 it says that in the case of the human genetic map you also simulated selection at regulatory sites. Please clarify. It would seem best to adjust the statement in lines 180 to 181.
- [l. 209–210] C: If 2% of the mutations are lethal ( $s = 1$ ) and the average deleterious selection coefficient for the non-lethal mutations is 0.07, then the overall average would seem to be  $0.02 \times 1 + 0.98 \times 0.07 = 0.0886$ . The result is close to but not identical to the “mean heterozygous selection coefficient” of 0.1 reported in Table 1. Please clarify.
- [l. 214] R: Remove “is” after “heterozygotes”.
- [l. 222–227] R: Surround  $2N$  and  $2N_e$  by brackets for clarity.
- [l. 228–229] C: Related to the point made here, Pouyet et al. (2018, eLife) showed that up to 95% of the human genome might be affected by BGS, which suggests that purifying selection is acting in non-coding regions in humans as well.
- [l. 256–265] C: Comparing the CNC97 treatment to the other treatments, I think it is worth highlighting that deleterious mutations in the CNC97 treatment are partially recessive, whereas they are partially dominant in the other scenarios due to the assumption of multiplicative interactions.
- [l. 289–290] C: My understanding was that you computed  $B$  at each site in the focal region, accounting for deleterious mutations across both the focal and the flanking regions. Is this correct? If so, please make it clear.
- [l. 302] R: Remove “at” after “up to”.
- [l. 317–319] S: Please revise this sentence and, in particular, state that the island model in fact provides a bad approximation to the No Migration treatment.

## Results

- [l. 424–425] S: Please discuss why the false positive rate is far from the  $\alpha$  value in the No Migration treatment.
- [l. 427] R: “FPR” has not been defined. Remove the full stop after “by chance”.
- [l. 439–440] Q: Did you explore different sizes of the focal and flanking regions, and can you be sure that your choices are appropriate to catch local variation in  $F_{ST}$  due to BGS at all?

## Discussion

- [l. 446–448] C: This statement is inconsistent with lines 387 to 391, where the authors say that mean  $F_{ST}$  differs between the neutral and BGS scenario.
- [l. 500] S: “yield to”  $\rightarrow$  “yield”
- [l. 517–535] C: The discussion of  $d_{XY}$  here is flawed by the fact that  $d_{XY}$  is *not* equal to  $4N\mu + 2t\mu$  for the majority of the treatments considered here, except for the No Migration treatment. The appropriate expression is found in Eq. (11) of Wakeley (1996, J. Genetics). See also Wilkinson-Herbots (2008). Wakeley’s expression applies to all demographic scenarios simulated by the authors, and it accounts for the phase before the migration–drift equilibrium has been reached in the scenarios with  $M > 0$ . During this phase, it is not appropriate to distinguish between a “fixation term” and a “heterozygosity term”. At equilibrium,  $d_{XY}$  converges to  $4N\mu + \mu/m$  in the scenarios with migration, and there is no “fixation term”. Please revise this paragraph accordingly.

## Figures and tables

- [Fig. 2] S: As the authors mention in the main text, it is difficult to see when the 95% confidence intervals for the neutral and BGS scenarios overlap and when they do not. I wonder if the authors could add an asterisk or similar in cases of no overlap (as in Figure 4). R: Please fix the reference to the supporting figure (should be Figure S2 instead of Figure S3).

## Appendix A

- [Paragraph 2] R: Please fix the grammar in the first sentence (“simulations which results are”). C: The second sentence is confusing, as Figures A1 and A2 are arranged vertically (in fact, as separate figures), but referred to as “left” and “right” parts apparently.
- [Paragraph 4, last sentence] Q: What formula did you use to compute  $F_{ST}$  from coalescence times simulated with SLiM? See also major concern 4) above.
- [Paragraph 5, last sentence] Q: Should it say “in blue” instead of “in red”?

## Supporting material

- [Fig. S2] R: In the caption, “means”  $\rightarrow$  “mean”. S: Please improve the ratio of figure content to annotation. Labels of panel rows and columns, axis label, and legend text are too heterogeneous in size and rather large compared to the plotting area, which makes it hard to digest the information. Reduce the number of font types in the mix. Please fix the reference to the figure in the main text (should be Figure 2 instead of Figure 1). S: Perhaps avoid the expression “Human treatment” (compare to “Human genetic map” in the main text).