

## Round #2

---

by Guillaume Achaz, 2020-04-29 19:19

Manuscript: <http://arxiv.org/abs/1909.11490>

### Minor revisions

Dear authors,

The two referees and myself have carefully read the revised version of your manuscript. All the three of us agree and recognize the great improvement of this version, compared to the previous one. The ms really nows conveys a more balanced and precise review of the interactions of the different processes at stake in the patterns of molecular diversity. Although the improvement of this new version is impressive, I believe that there is still some room for extra minor adjustments. Please read carefully the comments of the two reviewers on this new version and provide a point by point response.

> We thank the two reviewers as well as the recommender for their comments. We have responded point by point to the comments by expanding some sections and reducing others. We have clarified some definitions and modified Figure 1 as suggested to illustrate what is going on at the population level.

I also would like to suggest on top several minor points that the authors may want to consider. These points can help clarifying or illustrating the scientific content of this version.

1) Effective population size - its first occurrence is, I think on p3.  $N_e$  is never defined and this lacks, as there are multiple ways to define and/or to understand  $N_e$ . I would encourage the authors to define  $N_e$  at its first occurrence, or to avoid this term entirely. I suspect that on p3 the authors meant some sort of "neutral"  $N_e$  (e.g. harmonic mean of population size), but some people use the term  $N_e$  to refer to "level of diversity" (measured through  $\pi$  or any other statistics). As (linked) selection will greatly affect the diversity, replacing "diversity" by " $N_e$ " may lead to very confusing arguments. I know this is a long standing debate, getting back to at least Gillespie (and his articles on the meaning of  $N_e$ ).

>Thank you for pointing this out, we have now defined  $N_e$  here according to Wright's definition (1931), as this is the simplest explanation in terms of making the point of selection strength depending on  $N_e$ .

2) I am not a big fan of Figure 1 as it is now. Can the author work a bit to clarify it? I suspect that having more than 2 chromosomes/sequences per sample will be helpful.

> We have changed Figure 1 to represent what is going on at the population level. We now have 6 sequences per region/mechanism. We hope the changes clarify the 3 mechanisms. We changed the legend accordingly.

3) For meiotic drive, may I suggest the reading of: Henikoff S, Ahmad K, Malik HS: The centromere paradox: stable inheritance with rapidly evolving DNA. Science 2001, 293:1098-1102.

> Thank you for the reference. We have expanded the segregation distortion paragraph by including centromere drive evolution as well as the case of Wolbachia mentioned by Benoit Nabholz.

4) If I am not too confused, there is a major difference between gBGC and selective sweep. In the former, there is no room for hitchhiking, so the neutral surrounding diversity is not affected. It may be worth mentioning it.

>That is correct. We now mention this clearly in the BGC paragraph.

5) Finally, there is no Acknowledgments section in your article. Is this on purpose?

>We have added an Acknowledgments paragraph at the end of the manuscript.

## Reviews

*Reviewed by anonymous reviewer, 2020-04-07 15:19*

The authors have done an excellent job responding to the prior round of review in my opinion, and I believe the revised manuscript is much improved. I have a few remaining mostly minor comments that I think would improve the manuscript, including a few minor but I believe essential revisions. However, in general I think the major concerns from the previous round have been addressed.

Minor but essential revisions:

1. There are still a few places where I think definitions need to be clarified.
  - o First, I think that many of the demographic processes the authors discuss would often be considered special cases of genetic drift, as the patterns that arise under these demographic scenarios are still the consequence of random sampling of individuals instead of fitness differences. Of course, a major thrust of this review is to outline the non-adaptive but still non-equilibrium processes that can cause departures from a standard neutral model. But I think an issue remains that the null model here is not carefully defined (e.g., on bottom of page 5 is isn't really clear what the authors mean by 'nor are they purely subject to drift'). It seems that the authors are thinking about something like a standard Wright Fisher model as the null, and then distinguishing between 'non-adaptive' and 'adaptive' departures from that null. This of course has been the subject of many, many papers, but the strength of this review is bringing together disparate topics that have not often been treated together.

>This is a fair point by the reviewer, and I think our text a bit mis-emphasizes what we mean to lead to this comment. It is true that things

like gene surfing are just a special case of genetic drift across geographic space, but the outcome of this and the other non-adaptive processes we mention leave signatures different from what one would expect from pure and simple genetic drift in one population. This is what we try to emphasize (as many other studies as well), that simple drift cannot be the only thing disproved to identify selection, but also these other non-adaptive processes. We have tried to clarify this throughout the text, but in particular in the same paragraph at the bottom of page 5 mention that these processes differ from pure/simple drift.

- A second issue is using the word 'hitchhiking' to refer to all linked selection, which I think is non-standard in the field and may be confusing. Following Maynard Smith and Haigh (Genet Res. 1974), it is preferable to limit hitchhiking to the case of neutral alleles whose frequency changes due to a selective sweep (of some kind) and use the more general term linked selection (e.g. <https://www.nature.com/articles/nrg3425>) to refer to the general case.

>Thank you for pointing this out; it was an oversight on our part to conflate hitchhiking with linked selection in several spots. We have corrected this for all instances where we used hitchhiking to refer to something other than a neutral allele sweeping to fixation when linked to a beneficial allele.

2. Table 1. I am not sure the benefit of including the middle three columns in this table, since the answers are yes for all processes discussed in this paper, so it is not clear that listing this information repeatedly in the table is a valuable use of space.

>That is correct. We removed this column from the table.

3. The paragraphs in sections 3 and 4 vary considerably in their depth and detail, particularly with regard to citations. I am impressed with how the authors have managed to pull together a huge amount of disparate topics for this review, but some of the value is a bit lost when certain sections (that have a very deep literature behind them) are shallowly referenced. In particular, sections 3.1 and 3.2 have been the subject of extensive empirical and theoretical work, as well as the topic of numerous reviews. Yet, in section 3.1 only three papers are cited, all empirical work on humans; similarly only two empirical papers are cited for section 3.2 In contrast, section 3.3 provides a deep context for further reading. I think, because many of the topics covered in this review have been discussed extensively before, it would be particularly valuable for the authors to provide more context of the previous literature. In this way, this review could almost serve as a one stop shop as a introduction to a bunch of related topics. This similarly comes up in the section on linked selection.

>We thank the reviewer for this feedback, and similarly to the same feedback from reviewer 2. We have shortened slightly section 3.3 (see response to reviewer 2 below) and have now expanded the discussion in 3.1 and 3.2 to also include more references and better context for the relevance of these population characteristics

or processes as well as useful methods when dealing with them. We have also now expanded the context in the linked selection paragraph by mentioning a review written by Isabel Gordo and Brian Charlesworth (2001) on genetic linkage and molecular evolution as well as two of the first papers/books considering the effect of genetic linkage on genetic evolution (Fisher 1930 and Muller 1932).

I also have some comments that I think would improve the manuscript, but largely involve stylistic concerns, so I list these solely for the author's consideration.

1. I think the introduction could even more strongly set up the central focus of the review, which I take to be "highlight[ing]...the major evolutionary processes that can change allele frequencies in ways that mimic signatures of selection." This is of course an important and perennially interesting topic, and throughout the review the authors cite a number of examples where consideration of non-adaptive forces have been important to correct interpretation of population genetic signatures. However in the introduction, the focus is at least in my reading more strongly on the question of 'what fraction of sites in the genome are free from selection', which isn't really strongly touched upon during the rest of the manuscript (although of course is related). This is largely a stylistic concern so I totally understand if the authors like the intro the way it is, I just think the review could have a more focused impact with greater emphasis on the goal of a summary / overview of all the ways data can trick us into thinking selection is acting when it isn't.

>In response to major comment 1 above, we have improved the emphasis of this point at the end of the intro (and agree with the reviewer here about our emphasis within this manuscript). Though we do think the importance of this topic arises out of the understanding of selection and drive and the relative importance of these processes as well as the fraction of the genome subject to these processes, so we largely keep the flow of the introduction the same.

2. I still struggle with the way that linked selection is incorporated into this manuscript. It is certainly the case that some methods, especially for inference of demographic models, assume that sites are free of the influence of selection, as the authors discuss on the top of page 3. But in general, linked selection is not, in my opinion, a process that 'can change allele frequencies in ways that mimic signatures of selection' -- linked selection is a byproduct of selection itself. This is particularly the case for certain subsets of selection -- e.g., a completed sweep is generally invisible to detection in the absence of the signature left behind in linked neutral alleles. So I feel like linked selection really doesn't fit with the other topics that focus on 'things that are like selection but aren't', and including it in this way is confusing. However, I get this may be a semantic argument that me and the authors simply disagree on, so I am not belaboring the point.

>We changed the end of the intro to encompass linked selection : "In this review we highlight both the importance of considering the genomic, biotic, and abiotic context in which new mutations occur and the major evolutionary processes that can change allele frequencies, creating a major confounding factor for evolutionary inference of natural selection."

3. Personally, I would leave out the section on TEs, although I'm sure a TE biologist would disagree. To me this section doesn't really fit in the framework since the issue with TEs is generally that they are a hard-to-study kind of mutation, not that they are a process that mimics selection (although intra-genomic TE expansion can have its own complications of course).

> We have decided to leave the TE paragraph in as this topic is largely under-appreciated in the field of population genomics (except for TE biologist as the reviewer mentioned), even though TE dynamics probably have more than a strong impact on genome evolution. Yet, to improve the flow of this section with the remainder of the manuscript, we emphasize how TEs are so little studied that they may perhaps be harboring a lot of information or creating many biases in inference of selection or estimation of demographic parameters, if they are at random changing polymorphisms that we are not well aware of. Inversions, in particular, may be unique non-recombining sections of genomes that harbor chunks of beneficial alleles, or mask large chunks of deleterious variation – an area ripe for further study.

Finally, as an aside, I hope the authors are holding up in these difficult times. Of course as a review paper revisions are not dependent on access to labs, but even so I realize it can even be harder to search for literature when off campus (depending on individual circumstances), and I am mindful of trying not to ask for too much revision at this stage. I hope the authors (and editor/other reviewers) will take my comments in that spirit and focus on what they see as the highest value improvements.

*Reviewed by Benoit Nabholz, 2020-04-12 15:09*

I would like to congratulate the authors on their work. The manuscript is much improved. The manuscript is now more balanced, has more sections and appears more thorough. As an example, the number of citations has increase from 59 to 104. I think that most, if not all, readers will learn something out of this manuscript now. I think that it could be recommended by PCI Evol Biol.

I have few minor comments:

- 1) In the new section “Segregation distortion”, the authors should consider to include *Wolbachia* that could drive fixation of mitochondrial haplotype by linkage (Hurst and Jiggins 2005).

> Indeed, that is a well-studied example and we thank the reviewer for pointing this oversight out. We now mention how *Wolbachia* impacts the evolution of host mitochondria and we refer to the paper of Hurst and Jiggins, 2005 as well as the paper of Cariou, Duret and Charlat, 2017 which shows that *Wolbachia* reduce the efficacy of selection by reducing the effective population size of host mitochondria at a large evolutionary scale.

- 2) P4. Temporal variation in selection could also exist in addition to spatial variation. See for example (Bergland et al. 2014; Wittmann et al. 2017).

> We entirely agree with the reviewer here. Unless we are looking at a different line (it is a shame that ArXiv requests line numbers to be removed), our sentence reads “Equally complicating is the individual-level scenario where fitness is dependent on the local community of organisms and whatever traits they exhibit (e.g. frequency-dependent selection) or on the abiotic local conditions exerted by variable environments on the phenotype (spatially or temporally varying selection).” We include here both the possibility of temporally or spatially varying selection playing a role, as both can have major impacts on evolution.

3) P8. The two sentences “Additionally, the process of migration into populations or admixture among species can create an influx of novel genetic material. Even if fully neutral, the presence of such heterozygosity in the population leaves a signal indicative of either adaptive processes (e.g. balancing selection) or non-adaptive processes (e.g. secondary contact).” need references.

> We have slightly broadened the meaning in the non-adaptive description, to include the mention of gene flow among structured populations, and now cite “Molecular Population Genetics” textbook by Matthew Hahn, which thoroughly covers this topic at length.

4) The section of “Gene surfing” still has too many references compare to the other section. P8, 16 references were included in three lines (ref 47 to 63). This is far more than for any other phenomenon described in the manuscript. The authors should make the effort to include more reference elsewhere of to reduce the number of references for this section.

> We have shortened this section to now leave out the discussion of the debate surrounding gene surfing, as this was what demanded the manifold numbers of citations. We think this is an appropriate change, as the emerging evidence since that debate arose (previously discussed in this section and now removed) has increasingly settled this debate into expansion load and gene surfing being important population genetic processes, so it is only tangential for the purpose of the current manuscript to mention the debate and the copious examples within.

In response to reviewer 1, we have also added some to the other sections, so that there is more balance among the “population-level processes” sections.