

Reviewed by anonymous reviewer, 2021-01-06

16:16

I want to thank the authors for improving a lot their manuscript. I highly appreciate your effort.

I hope you agree that your second version is more readable and precise than the first one.

In my case, almost all the comments I had have been answered.

- You changed the abstract
- You showcase an example
- you clarify gene conversion term
- the importance of considering a circular genome etc...

We thank the reviewer for the comments, and agree that the manuscript's clarity has been improved.

Two major suggestions:

However, I would like to suggest focusing on two critical aspects of the paper:

1. the figure legends are better, but they do not directly convey the entire message.

For example when I see Figure 1 I am not sure if rescaling factors have also been applied in FastSimBac and ms.

Answer: Rescaling simulations only applies to forward-in-time simulations. We clarified this in the legend and in the main text (result section). If there are other figure legends or captions that require additional clarification, we would welcome that feedback; of course we have already made them as clear as we can, from our perspective.

Would be a big plus to include a summary table of your results.

Answer: Following the reviewer's recommendation, we added at the beginning of the Method section, a summary table of the different simulation methods used in the paper with a brief comment and example for each.

1. It's a bit of a pain to read through the supplementary figures. Low quality, tiny labels. It's not easy to go through them.

Answer: We increased the font size in supplementary figures S5 to S7, as indeed labels were small. As for the quality, we are unsure as to what the reviewer is referring to as the figures are vector-based and remain sharp at infinite zoom; perhaps they were converted to bitmaps by the reviewer's PDF-viewing software, or by the journal's PDF assembly process? In any case, they should be full quality in the final typeset PDF.

One minor suggestion:

3. This is a minor suggestion: when you show the code on page 6 and 7 avoid better this format.

Use a grey shaded area similar to StackOverflow when you present a code case, so as the reader can copy paste and test the code easily. That's an aesthetic suggestion.

Answer: For the coding snippet, we followed the aesthetic of the SLiM's manual code snippets so people will be less confused when going from the article to the manual or vice versa.

The code can already be copy-pasted easily. The line numbers may or may not be copied depending on the pdf reader, but in any case can be removed easily by vertical selection, which is a widespread feature of coding editors. Thus we are keeping the line numbers since they can be used for pedagogical descriptions and in-line text references.

In any case,

Thank you for your effort.

Happy new year

We thank again the reviewer for the various inputs provided through the different rounds of revision. Happy new year !

Reviewed by anonymous reviewer, 2020-12-18 10:58

I have taken a look at the authors' responses to my original comments and found them sound.

We thank the reviewer for their comments throughout the reviewing process.

Reviewed by anonymous reviewer, 2021-01-06 07:39

This revised manuscript is partly incorporating the reviewer's suggestions. Most of my points were taken care of, however, I see room for improvement for the description of the simulation, the clarity of the figure, and the explanation of the burn-in phase. My largest remaining concern is the discussion of selected recombination events. Here are my detailed points:

An overview figure is provided, however it is only in the supplement and it does not visualise the simulation parameters (N_e , generations, ρ , t_{tractlen} , genomesize, h_{gtrate}).

Answer: We think that reviewers did not receive the version 3 of the manuscript, which was made following comments from the recommender (editor). This version contains small improvements, including additional labels in supplementary figure S1 (with addition of labels for graphical elements such as Ne, generation, tree, mutation, Bacteria, chromosome). We had previously thought about including the simulation parameters in the figure, however this leads to an overloaded and cumbersome scheme. We are now adding a discussion of them in the figure legend to highlight the link with the main text, as well as back-referencing to the code snippets when appropriate.

We believe that attempting to depict all of those parameters in the figure would prove more confusing than helpful, since there is no common, standard graphical vocabulary for presenting such model details in a diagram. Given that the potential reader may range from a wet lab microbiologist to a theoretical population geneticist, it seems wise to keep the figure as a simple, high-level overview of the design, and explain all of the complicated details in the text.

We also think that there are already many interruptions to the main text, with code snippets and other figures. For these reasons we would prefer keeping the figure in the supplementary material. Besides, the figure will not be far away since PCI journals do not reformat the pdf and thus the supplementary figures will not be in a different document. We added the possibility to click on figure names (and references in general) to jump directly to the corresponding element to reduce friction as much as possible.

An additional simulation of bacteria under antibiotics on a Petri dish is now included. The choice of parameters is quite arbitrary, e.g., the antibiotic is reducing the fitness only to 0.47, I guess lower values are more realistic. However, the simulation should simply display an example of application and it serves this purpose. Nevertheless, method's details are still missing. How does the spatial model work, how is the neighbourhood of a bacterium defined? I guess this is a standard model, so references would help here a lot. I am also missing the information on recombination rate and tract length for this simulation.

Answer: The choice of parameters is indeed arbitrary and only provides an example of what is possible. Here, they were chosen for illustrating an interesting behaviour (no resistance would have emerged with an excessively strong fitness cost and no colonies would have survived to the antibiotic). However, one could argue that low concentration of antibiotic can lead to a similar reduction in fitness, so despite being arbitrary, these parameters remain realistic.

All the underlying mechanisms for the spatial model are described extensively in the SLiM manual. However the reviewer is right in that we had forgotten to reference the manual in the article. Thus, we extended the general description and now refer to the manual for the specifics. We also included details on the recombination process.

I had provided several references for the discussion of realistic recombination tract lengths. Nevertheless, the authors decided to not discuss the range of recombination tract lengths in the manuscript and point to their simulations of length 1220bp, 12,200bp and 122,000bp. I had also pointed out that their initial choice of a recombination tract length of 122kbp is based on "selected" recombination events. Thus, this estimate is not a good choice for simulations which should be based on unselected events. The authors ignored this point in their answer. I understand, that at this point of the manuscript it is not feasible to repeat the simulations, but the discussion of unselected vs. selected recombination events should at least be mentioned in the discussion. Otherwise the reference to the *S. agalactiae* length is misleading.

Answer: It appears that we did not understand fully the reviewer's previous comment. We apologize for that.

Concerning our estimation of 122kb, we are aware that this mean tract length is much larger than it is for other bacteria. Recombination in *S. agalactiae* is thought to occur through conjugation and not by transformation or transduction. Experimental analysis of recombination tract length confirms that in *S. agalactiae*, tracts can indeed be much larger than in bacteria where it occurs through transformation. In this paper (Brochet et al, 2008), the authors observe that during conjugation assays, recombination of large fragments occurs (despite being conservative by considering the lower bound of plausible sizes for each observed fragment). These recombined fragments represent a subset of the total fragments since they are observed thanks to a marker and could thus be called "selected". Yet, because this marker was uniformly located at different positions along the genome, we expect selected and unselected fragments to follow the same length distribution. In this experiment the bias that could arise would be due to the marker size; however, this size represents a small fraction of the transferred region (1 gene over 14 to 330kb). Moreover, in another paper, (Almeida et al. 2017), a few tracts of recombination are identified, and appear to be much larger than what the experimental assay showed. This is coherent with the first experiment for the reasons that the reviewer discussed: smaller recombined fragments are less likely to bring new mutations and thus are harder to detect.

Overall, our estimate appears realistic for *S. agalactiae*.

However, we agree that the various sizes of recombination tract length could be better discussed. In fact, this wide range of sizes motivated the section describing the impact of the tract length parameter (Figures 5 and 6). We added a couple of sentences in this direction, in the section "Impact of recombination".

The authors added more information on the burn-in phase and also display it in the figure. As I understand, in the WF model, the burn-in has to be run before the SLiM simulation, however, with the nonWF model it is possible to run it afterwards only on the individuals that have descendants (as displayed in Fig. 8A). However, how can SLiM be run with selection if the diversity and the mutations are not yet clear at the start of the simulation? How is the fitness of the individuals known? I think I am missing a piece of information here.

Answer: The reviewer is right: in the WF model burn-in has to be run before SLiM, which means that instead of starting with a population of identical individuals it will start with a population at mutation-drift equilibrium. In the nonWF model with selection, the forward part simulates only the alleles at loci under selection (which are created during the forward simulation, and thus do not need to be supplied by the ancestral evolutionary dynamics), and this will define the fitness of the individuals. Because we record the tree sequence, we have the entire history of every genome segment in the sample at the end of the simulation. The shape of the tree is impacted by recombination, selection, migration, demography, etc... After recapitation, where the tree coalesces fully, one can overlay neutral mutations, the same way it is done with coalescent simulator, where basically, the tree is simulated first, and then neutral mutations are added on top of it. We added a sentence in the method section of the spatial model to clarify this process.

Overlaying neutral mutations in this way is provably identical to tracking neutral mutations during forward simulation; the power of this technique is a big reason why tree-sequence recording is becoming so popular. For further background on this, the cited papers by Kelleher et al. and Haller et al. go into much greater detail on mutation overlay and recapitation.