

Summary

Overall, I think this is an interesting study that leverages existing theory alongside computational tools, yielding promising results. Ultimately this work should be published, but I have several minor comments and I believe that the impact of the paper would be strengthened by addressing them. Most of these comments can be grouped into requests to 1) elaborate on points the authors made and their connection to population genetic theory and 2) make alterations to the figures to increase their clarity.

While I am familiar with mathematical models of molecular evolutionary dynamics, I am less familiar with models of selective sweeps. I have attempted to provide useful feedback and questions regarding sweeps and their statistics throughout this review.

Major comments

- The authors model the evolution of a recombining population with a seedbank as a diploid hermaphroditic population. While it may be outside the scope of the manuscript, it's worth considering that recombination can be modeled in a more general manner where there is some probability of recombination per-generation in a haploid population. Could the authors describe the extent that their model and results can be extended to haploid populations?
- Line 39: I don't see why the word "species" is needed here.
- Line 200: It would be useful for the unfamiliar reader if the authors made note that the number of generations they chose for the burn-in corresponds to the coalescent timescale. My understanding is that this is the timescale of the Kingman coalescent, if so or if not, please note it in the text.
- Line 204-206: Is there any existing theory that the authors could reference for the order-of-magnitude timescale required for signatures of a selective sweep to become detectable?
- Line 219-220: While I understand that an appropriate range of selection coefficients can be hard to justify, selection coefficients ranging from 0.1-10 seems too high. Even with the lowest value of the empirical population size (ignoring effective size) of 1,000, the population-scaled strength of selection is much larger than one for all parameter combinations ($|N*s| \gg 1$). I would think that it would be useful to examine signatures of selective sweeps with and without fixation as the population-scaled strength of selection (calculated using the effective population size) increases starting below 1 (drift dominates). There may be some nuance I'm missing, but as a general pop gen reader this is the range of parameter values I'd be looking for.

- Line 241–242: It would be helpful for the reader if the authors described the Omega statistic, how it works, and any additional parameter settings the authors may have chosen.
- Line 271-274: This was a very helpful point, thank you for making it.
- The strength of selection is not necessarily the key parameter, but rather the population-scaled strength of selection. Could the authors discuss that quantity in this sentence instead?
- Line 345-348: This was explained well.
- Fig. 2: a) The recombination rates should be listed in a legend inside the figure. And what do the boxplots represent? Are they quantile plots? If so, given that these are simulation results, it may be more useful for the reader if the spread around the mean was plotted as 95% confidence intervals. b) “ r^2 ” should be “ r^2 ” and there should be an equal sign for “ b ” in the figure legend. And what are distance bins? Should these be thought of as base pairs?
- Fig3: a) I do not understand the x axis. Are these selection coefficients ranging from 1 to 5? If so, this is the plot where I’d like to see the population scaled strength of selection ($|N_e*s|$) over a logarithmic range of values with $|N_e*s| = 1$ as a midpoint. The shaded areas around the lines are not described in the legend and it would be helpful if the parameter b was included in the figure legend (e.g, $b=0.25$, $b=0.35$, etc). b) It would be helpful if the y axis was on a log base 10 scale. Also, “Germination rate, $1/b$ ” should be written on the x-axis. Again, it would be helpful if “ s ” was included on the figure legend (e.g., $s=0.01$, $s=0.1$),. C) This is a useful plot, but I think it would drive the take-home point home if a horizontal line at $y=1$ was included as a null as dormancy decreases. Also, “Germination rate, $1/b$ ” should be written on the x-axis.
- Fig. 4: These are nice plots, but they need some tweaking to help the reader. What does the shaded area around a line represent? Are these standard errors or 95% CIs. All parameters in the legends should have equal signs. And “ sc ” is used in the figure, whereas “ s ” is used in the legend. Are these the selection coefficients?
- Fig. 5: The y axis could be more informative. Could it be labelled “Probability of detecting a sweep”? it would be helpful if the parameter under “model” was included in the figure legend with an equal sign.

Minor comments

- The authors occasionally use the first letter of an author’s first name in in-text citations. Is there a reason for this? If not, they should be removed from the revision. I noticed

this on lines 42, 84, and 170, for example. I did not count all the occurrences in the manuscript.

- Throughout the manuscript the authors use “e” notation to describe the orders-of-magnitude of numbers. Could they switch to base-10 notation (e.g., 10^{-3} instead of $1e-3$?).
- I may have missed it, but I didn’t see the simulated data repository mentioned in the manuscript. If it’s not there,
- Lines 24-26: Could bacteria be included here as well? I understand that bacteria are typically thought of as asexual, but recent research efforts point towards their rate of recombination being higher than previously thought (Garud, Good et al., 2019; Good, 2020; Sakoparnig et al., 2021).
- Lines 49: This sentence is somewhat unclear to me. Unless I’m misunderstanding something, existence of dormancy increases the true T_{MRCA} in a population, so it will increase the T_{MRCA} estimated from a sample from a population. “of a sample or population” implies exchangeability between a sample and the true population, which I don’t think is the authors’ intent.
- Lines 61-63: I think it’s worth mentioning that the weak seed bank model is also applicable to cases where the existence of a seed bank is experimentally imposed, where it is unfeasible for the observer to observe a system over the timescale required for the strong seed bank effect. This could include bacteria (e.g., Shoemaker et al., 2022).
- Lines 74-77: This was well said.
- Line 100: How do mutation rates work here? Seeds don’t reproduce, so they’re acquiring mutations at some rate per-unit time (days, years, etc.), whereas the above-ground plants are acquiring mutations at a rate of per-generation. Are there any rough estimates of how these rates compare when they have the same units?
- Line 227-228: Should it be either “frequency of one” or “size of $2N$ ”.
- Line 300-302: I’m unsure what the shaded areas are in the referenced figure, but if the authors resample the simulated data with replacement to obtain 95% CIs, would they fall within the theoretical prediction? This may allow you to confirm that stochasticity was a contributing factor.
- Line 329-332: “However, when the beneficial allele reaches fixation, the time to fixation” I do not understand the framing of this sentence. It seems to be implying that

there is some non-zero time to fixation once an allele becomes fixed (i.e., all individuals have the mutation).

- Line 333-334: “yielding the counter-intuitive result that dormancy enhances the efficiency of selection compared to genetic drift”. I don’t understand what “efficiency” means here. The plot in Fig. 3c examined the time to fixation with a seedbank relative to the time to fixation without a seed bank. This ratio seems to increase with germination rate for all selection coefficients, so I do not see how the notion of “efficiency” fits in here.
- Line 423: replace “>=” with “≥”
- Line 460: What are the CLR tests?
- Line 473: Should “reduce” be plural?
- Listing 1: Should “analysis of” be “analysis”?
- Fig. 1: Do the authors have a version of the figure with higher resolution? In its present form, the resolution contrasts with the resolution of the PDF, drawing the eye towards the difference.
- Appendix: In both figures in the appendix, it would be helpful if equal signs were included for the parameters. It is also unclear what the parameter “d” represents.

References

- Garud, N. R., Good, B. H., Hallatschek, O., & Pollard, K. S. (2019). Evolutionary dynamics of bacteria in the gut microbiome within and across hosts. *PLOS Biology*, *17*(1), e3000102.
<https://doi.org/10.1371/journal.pbio.3000102>
- Good, B. H. (2020). Linkage disequilibrium between rare mutations. *BioRxiv*, 2020.12.10.420042. <https://doi.org/10.1101/2020.12.10.420042>
- Sakoparnig, T., Field, C., & van Nimwegen, E. (2021). Whole genome phylogenies reflect the distributions of recombination rates for many bacterial species. *ELife*, *10*, e65366.
<https://doi.org/10.7554/eLife.65366>

Shoemaker, W. R., Polezhaeva, E., Givens, K. B., & Lennon, J. T. (2022). Seed banks alter the molecular evolutionary dynamics of *Bacillus subtilis*. *Genetics*, 221(2), iyac071.
<https://doi.org/10.1093/genetics/iyac071>