

This is an overall very clear and well written manuscript that addresses an important topic in population genetics inference. It complements previous work that have assessed the performance of SMC based methods in the absence of panmixia by exploring the consequences of other common violations to model assumptions. Not only are the tests performed useful for readers interested in application of these methods, but the interpretation of the results is also interesting to readers focused on more theoretical aspects. I find the suggestion to use the transition matrix between pairs of genealogies as a summary statistic to be particularly exciting!

Major comments and suggestions:

I don't have any major comments.

Minor comments and suggestions:

It could be very useful further explore the case of simulated flat demography. Since variation in the recombination rate can cause spurious waves in the inference (and virtually all species have a heterogeneous recombination map), it would be nice to show if/when one can trust "small" wiggles in the N_e curve. If simulations show that it is not uncommon to see an artifactual > 2 fold expansion/contraction, then guidelines specific to these cases (eg, when can I reject the "null" model of constant N_e) would be appreciated by the community.

Section 2.1.6 could be written in a more clear way (e.g., link time windows to hidden states).

Like 149: unless I am mistaken the number of hidden states is not necessarily proportional to the number of parameters in SMC++ since they use cubic splines interpolation.

Lines 298 and 304: Be more specific. Which percentage of the genome is removed? Are all fragments of the same size? Which size?

Line 335: This is overall a very instructive paragraph, and it seems to me that the CV is computed over the entries of a single transition matrix, but are there not replicates for each scenario? Could it be worth to explore the CV of particular entries over different replicates?

Line 432: Somewhat unclear writing, I don't understand why they would not be accounted for as missing data (eg, emission probability = 1 as in PSMC)?

Typos and other small suggestions:

Line 96: "hypothesis" -> "hypotheses" (supposedly this is plural)?

Line 101: "simulates" -> "simulated"

Line 129, 133, 134: "genealogy" -> "Ancestral Recombination Graph"?

Line 163: "does" -> "do"

Lines 165 - 168: a bit of an over-simplified picture since this explanation does not take recombination rate (and past N_e) into account. It could seem to the reader that the inferred hidden state at a homozygous site is always the shortest genealogy whereas for a heterozygous site it would be the tallest genealogy (ie, it sounds like the hidden states are completely free to change from one site to the next).

Line 169: the TMRCA does not necessarily change after a recombination event in the SMC'

Line 249: add "exponential" to be more precise?

Line 316: avoid the term "perfect fit".

Line 363: “find” sounds a bit confusing. Does it refer to the occurrence of hidden states (unknown truth) or the inference procedure?

Line 363: “which” -> “with”?

Line 390: “MSMC is able to retrieve the correct recombination rate [...]”. I can see it’s closer to the real value, but it’s hardly “correct”.

Line 430: “The smaller the sequences that are removed, the more rho/theta is over-estimated.” -> “For a fixed amount of missing data, the smaller the sequences that are removed, the more rho/theta is over-estimated.” or something like that?

Lines 444, 448: avoid using “perfect”.

Line 457, 460: “genealogy” -> “genealogies”

Lines 458 - 460: “Our results suggest that whole genome polymorphism data can be summarized in a transition matrix based on the SMC theory to estimate demographic history.” The term “demographic history” could potentially be interpreted in the broad sense (eg, including structure and migration), but this has only been shown for the panmitic case. Maybe rephrase?

Line 468: “genealogy” -> “genealogies”

Line 472: “hidden states” -> “transitions”

Line 492: This is an important sentence, would be important to clarify how missing data is handled in the Results section.