Dear Recommender,

We thank you for your positive evaluation of our work and the constructive comments made by the referees. We carefully followed your recommendations and the reviewer's comments, and now propose a fully revised manuscript that contains a number of additional analyses. These include an assessment of the impact of phasing errors on recombination map reconstruction, and a consideration of demographic scenarios that depart from the assumptions of panmixia and demographic equilibrium. We did not explicitly model linked selection, but we argue that its expected effects can be understood as local demographic perturbations along the chromosome, which are largely covered by our new demographic analyses. These additions entailed a reorganization of the Results section of our manuscript, in which we now first present the effect of methodological parameters (sample size, LDhelmet settings, hotspot definition), and then the effect of biological parameters (*Ne*, demographic/admixture history, recombination/mutation rate ratio) on the accuracy of inferred recombination maps. Please find below a point-by-point response, where we explain in detail how your and the reviewers' comments were taken into account in this revision.
We hope that you will be satisfied with these changes and find our revised manuscript suitable for publications in the *Peer Community journal*.

Best regards

M. Raynaud, P.A. Gagnaire, and N. Galtier.

Recommender's comments:

1. Some figures are color coded although the color is not visible on the graph (Figure 3, Figure S3), Other Figures have unclear comparisons (Figure 5, the actual rate is hardly visible in blue) and some others may include labels additions for a quicker understanding of the multiple axis. Improve the presentation of the figure in its revised version.

The Figures have been re-designed to account for this comment.

2. "line 274, should the sensitivity be TPR?, line 275, FDR is a way to measure type I error, which is based on alternative hypotheses, although type I error is usually defined as FPR"

This has been clarified. We actually rely on the definition of these terms provided in their first occurrence in the Results section. The text has been modified as follows:
"The sensitivity (or TPR) of LDhelmet was medium to high, since depending on the SS and the BP used, between 29.4% and 52.7% of the simulated hotpots were inferred as such. [...] The proportion of false hotspot calls (FDR, *i.e.* the inferred hotspots corresponding to non-hotspot windows in the simulated maps)" ranged between… (p 10, l 269-274).

Reviewer #1 comments:

1. Based on empirical data and error rates, would it be possible to include some plausible scenarios of sequencing error? For accounting for uncertainty of phasing, can genotype data be simulated and subsequently phased before conducting analysis.

We did not consider sequencing errors, which is less and less of a problem as sequence coverage increases. Recombination map reconstruction is particularly unlikely to be heavily affected by sequencing errors since these are expected to generate false singletons, and singletons carry essentially no signal of linkage.
Phasing errors, in contrast, are clearly of concern. To address this, we first evaluated the ability of a widely-used statistical phasing method, SHAPEIT4, to correctly phase our simulated data. We found that the phasing error rate was of the order of 7%. Then we assessed the impact of phasing errors on the accuracy of inferred recombination maps, making the error rate vary from 0 to 10%. We found that LDhelmet was relatively robust to this problem, the TPR and FDR being decreased/increased by 3 to 20%, depending on the conditions and population replicate. This analysis is presented p 11, l 289-298 of the revised version.

2. There could be additional confounding factors when inferring fine-scale recombination patterns from LD, in particular the confounding between cross-overs and gene-conversions. Has any analyses been done in this regard?

Indeed, gene conversion is expected to contribute to the dissipation of LD in high-recombining regions, in addition to the well-known effect of crossing-overs. Gene conversion only applies very locally, at the scale of ~100 bp, but can be quite common - in mice for instance DNA double-strand breaks are resolved in non-crossing-over (i.e., gene conversion) ~10 times more frequently than in crossing-over (Cole et al 2014 Nat Genet 2014 46:1072). This implies that recombination rates and hotspots inferred from LD-based analysis presumably reflect the combination of crossing-overs and gene conversion. Distinguishing between the two processes, if possible, would be quite an interesting goal to pursue. A recent paper did address this problem (Setter et al. 2022 Genetics 222:iyac100), while not accounting for the small-scale heterogeneity of recombination rate - so, not really applicable to our study. We added a sentence in the manuscript discussing this point and citing Setter et al. 2022 (p 20, l 540-544).

3. Human populations can have complex demography with migration between populations. Can this further impact the accuracy of such inference methods or do we expect relative rates to be robust?

Clearly a sensible remark, knowing that LDhelmet relies on a model that assumes panmixia and mutation/recombination/drift equilibrium. Two recent papers analyzed the effect of fluctuating *Ne* (Dapper & Payseur 2018 Mol Biol Evol 35:335) and gene flow (Samuk & Noor 2022 G3 12:jkac236) on the estimation of recombination rate. These papers, however,

considered relatively simple underlying landscapes that contained either no hotspot, or hotspots separated by long segments of low-recombining DNA. Therefore, we decided to investigate the impact of demographic fluctuations and population structure on recombination map inference under our complex, human-like recombination landscapes. We found that both bottleneck and admixture scenarios degrade the performance of LDhelmet, to an extent that depends on conditions and parameters. These new analyses are presented p 12, l 319-328 of the revised version, and reinforce our overall message of caution.

4. Presence of natural selection in a region can bias recombination inference. How may this affect hotspot inference? In general, other than simple recombination landscapes, what other assumptions need to be met for hotspot inference to be accurate?

Linked selection is an additional level of complication, which entails a wide range of scenarios (directional selection, either positive or negative, balancing selection) and parameters (selection coefficients). Importantly, it should be recognized that the effect of linked selection is often hardly distinguishable from the effect of demographic fluctuations (e.g. Johri et al. 2021 Mol Biol Evol 38:2986). For instance, recurrent selective sweeps and a bottleneck both result in a decreased mean and an increased variance in coalescence time (e.g. Jensen et al. 2005 Genetics 170:1401), while background selection is often thought of as equivalent to among-loci variation in Ne (Zeng & Charlesworth 2011 Genetics 189:251). We therefore suggest that our analyses of the effect of demographic fluctuations cover to some degree the issue of linked selection. We added a sentence mentioning this analogy (p 18, l 454-457)

5. Structural variation such as inversions can also impact recombination inference in certain species like drosophilia. This may further contribute to uncertainty.

This is another interesting point. Although we did not specifically account for structural variants in our simulations, our scenario of recent admixture between two divergent populations mimics the most important characteristics of large inversions - the lack of recombination events between two haplogroups for a large fraction of their coalescent history. We thus specified in the text that our simulations under the admixture scenario provide insights into the effect of structural variants such as inversions on the inference of local recombination rates (p 18, l 454-457).

Reviewed by anonymous reviewer #2, 08 Jun 2022 00:55

Raynaud et al present a manuscript using simulations to test the performance and limitations of a commonly used method to infer recombination landscapes, LDhelmet. They find that maps produced with the method have good correlations with the true simulated maps, but there are limitations when the method is applied to detect hotspots. In particular,

LDhelmet tends to overestimate the local recombination rate. Additionally, they note that the method can find shared hotspots when there are no real shared hotspots. This result has implications, for example, for interpreting data from Shanfelter et al 2019, who used LDhelmet and found little overlap between marine and freshwater populations of three-spine stickleback. In general I found the manuscript interesting and well-written, but I have some suggestions which I hope will improve the manuscript.

Major comments:

· I found the scope of the study to be more limited that I expected. The authors focus on a single method published in 2012. While this method is widely used, there have been several additional methods published in the following years (which the authors cite in the introduction, line 62-63). I would be interested in understanding how other recently developed methods compare.

Indeed, it would have been of great interest to compare LDhelmet with some other widely used methods (i.e. LDhat, FastEPRR), including the most recent ones such as Pyrho, iSMC, ReLEARNN ... We choose to focus on LDhelmet since it has been the second mostly used method to infer population-based recombination landscapes over the last 15 years, after LDhat (over 87 papers identified to have build LD-based recombination landscapes, 47.1% used LDhat and 20.7% used LDhelmet). We added a Supplementary Figure S1 to our manuscript, cited in the Introduction page 4 line 107-108, illustrating the use of the different existing methods in the literature of population recombination landscapes. Our main objective was thus to help users to interpret their recombination maps generated with LDhelmet.
We did some simulations using LDhat varying the simulated effective size Ne (25000, 250000) and the sample size SS (10, 20). We found the same trends, in terms of correlations between simulated and inferred landscapes, TPR, FDR and hotspot sharing, but LDhat generates slightly less true positives and more false discoveries. We also found less hotspots shared between pairs of populations for low SS compared to LDhelmet.
While the various existing methods use different approaches (pairwise composite likelihood for LDhelmet and LDhat, machine learning for ReLEARNN …), we recall that they are all based on linkage disequilibrium information, so, are likely to be similarly affected to LDhelmet by the various factors we simulate.

· Additional evolutionary parameters: in a study like this, one has to make choices about which parameters to study. I agree with the authors that studying the impacts of effective population size, mutation rates, and recombination rates is important. However, I will suggest 2 additional factors that I think would be substantial benefits to the manuscript:

a. Selection: the impacts of both positive and negative selection on patterns of LD are well known. However, I wonder how these forces affect hotspot inference. The authors could implement simple simulations in SLiM (Haller et al 2019 MBE), which will output a VCF and should fit fairly smoothly into the pipeline the authors have already set up.

b. Demographic changes with large effects on LD: it is well known that bottlenecks and exponential growth will affect LD patterns. Given the results presented in the paper, I would expect that these would also affect inferences of recombination hotspots, but I would be interested in quantifying how much.

Thanks for this comment. Indeed demography and selection are known to affect patterns of LD, and we agree these are worth investigating in this study. We note that linked selection and a departure from demographic equilibrium often result in similarly distorted patterns of coalescence, to the point that only the most sophisticated methods can distinguish between the two kinds of effects (e.g. see Jensen et al. 2005 Genetics 170:1401, Zeng & Charlesworth 2011 Genetics 189:251, Johri et al. 2021 Mol Biol Evol 38:2986, among many other articles on the subject). For this reason, and also given the high dimensionality of the parameter space, we decided to focus our analysis on demographic effects, while suggesting that our analyses also cover in part the issue of linked selection. Specifically, we explored the effect of a population bottleneck (on one hand), and of a scenario of admixture between diverged populations (on the other hand), varying the time and strength of the considered events. We found that such departures from the mutation/drift equilibrium tended to degrade the performance of LDHelmet, to a far-from-negligible extent, thus enriching the existing literature on the subject (Dapper & Payseur 2018 Mol Biol Evol 35:335, Samuk & Noor 2022 G3 12:jkac236), which assumed simpler recombination landscapes and used distinct reconstruction methods.

Minor comments:

· Could the authors give some intuition about what the block penalty does?

We made clearer that the block penalty is a smoothing parameter that controls the level of autocorrelation of the estimated population recombination rate across neighboring windows. Increasing the block penalty will result in a smaller number of transitions from a low predicted rho to a high predicted rho, or the other way around, as one moves across the genome - so, longer segments sharing the same estimated rho. The genome-wide mean rho is not expected to be affected by the block penalty; rather a question of the scale at which rho is averaged locally.

· Line 338-339: 10-9 should be 10^-9

Corrected, thanks.