Manuscript ID MER-15-0232

Dear Editor,

Thank you for the careful review of our paper and for the improvements that you have suggested. We are especially grateful to the three reviewers and the associate editor for their time and efforts in evaluating our manuscript. We wish to submit a revised version of our manuscript to *Molecular Ecology Resources.* Our revision addresses the concerns raised by the reviewers.

Here is a summary of the main changes:

1) Following Reviewer 1's comments, we added a new paragraph to the Material and Methods section summarizing the main principle of our algorithm and avoiding technical terms.

2) We also included an Appendix section that describes the technical details of the algorithm and of its internal parameters.

3) We performed additional simulations providing further evidence that the power of the tests proposed in our manuscript was close the optimal power expected for those data. Thus, the reason why the power of tests was quite low in our study was explained by the difficulty of the simulated data, and not by some poor properties of our test statistic.

We answered the questions of the reviewers, and provided a detailed description of our changes below. Our responses are colored in blue.

Please let us know if you need any further clarification of these changes. We greatly appreciate your time and assistance in preparing our manuscript for publication, and look forward to hearing from you.

Sincerely,

O. Francois on behalf of all authors

Reviewer Comments to Author: Reviewer: 1

Recommendation: Accept, minor revisions

Comments:

In this well written manuscript, Caye and colleagues develop and implement a fast algorithm to infer spatial population structure. While the presented algorithm does not infer quantities that could not be inferred by existing approach, the presented algorithm is about one order of magnitude faster. Beyond the valid intellectual entertainment, such improvements are urgently needed to render common population genetics approaches ready for the large data sets currently produced thanks to the technical revolution in sequencing technologies.

In addition, to the discussed speed-up, the authors also propose to use the inferred allele frequencies in the different populations to perform an outlier test to identify loci that show extreme Fst values and are hence candidates for loci under selection. This is surely a nice feature, even if the reported power seems a bit low.

A: Thank you for acknowledging the usefulness of our approach. The reason why the reported power was low was because loci under selection were difficult detect by any method for our simulation settings (see Answer to Reviewer 2).

To show this, we performed additional simulations considering no admixture (an easier context). For those data, standard tests based on allele frequency differentiation achieved only slightly better performances than the tests presented here. Thus, the tests implemented in TESS3 have power close to optimal value.

Overall, I judge this contribution very important and for sure worthy of a publication in Molecular Ecology Ressoruces (or even Molecular Ecology, after all). And I have only very few comments I hope the authors can address quickly.

Firstly, I was wondering about the a comparison of the proposed outlier detection method to existing programs, in particular BayeScan. It is my feeling that a method such as the one proposed here should outperform methods that ignore admixture when estimating population allele frequencies (such as BayesScan). Since this is potentially a strong selling point for the algorithm proposed here, I was a bit surprised the authors do not even discuss their method in detail nor in comparison to existing approaches. However, I think the manuscript would benefit a lot from such a discussion, as well as from a simple simulation study showing that their method outperforms existing methods that require strict population assignment.

A: We feel that a direct comparison with programs such as Bayescan of fdist would be quite difficult to run. TESS3 was designed for performing neutrality tests using individual-based sampling designs, for which no populations are defined a priori. In contrast, population-based tests such as Bayescan or fdist require population samples. For example, the data analysis presented for *A. thaliana* would be difficult to reproduce using Bayescan or fdist.

It is likely that TESS3 will outperform population-based tests when applied to continuous populations or to admixed population samples. We also expect bayescan and fdist to be better than TESS3 for diverging population samples. Evaluating the relative roles of geography, demographic history and sampling strategies require a detailed analysis that would be beyond the scope of our computer note.

Secondly, I think the material and methods part is a strange mix of very detailed description and missing information. Specifically, the section on the geographically constrained least-squares estimates of ancestry coefficients is hard to read. I think it is very important to be specific and

present the methodology thoroughly. However, some aspects seem to be discussed in great detail, but then others are to sparse to follow. For example, on L129 the minimization criterion is given in full, yet any intuition for the algorithm or what this criterion is used for is lacking completely. I think this section should be rewritten in one of two possible ways: 1) Either, the detailed equations are moved to the appendix (or supp info) and the text is kept superficial with the reader being required to essentially read all the cited papers to follow the methods. 2) Or, the authors add more details and give some intuition about how these algorithms work and why the equations look like they are. After all, most readers of Molecular Ecology are not (yet?) familiar with most of the tool used here (in contrast to, say, MCMC methods). I personally prefer option 2, as I think this paper is essentially a methods paper, but the choice is left to the authors. Also, I'm aware that there is a genetics paper essentially detailing the method itself, and it can be well enough to just refer to specific paragraphs in this paper at appropriate locations.

A: Understanding all the details of the TESS3 algorithms would ask readers to be experts in matrix factorization methods. So, we have changed our manuscript, and have added a full paragraph explaining how the method works in principle. This paragraph attempts to avoid technical terms.

We also moved the technical aspects of the algorithm to an Appendix section, at the end of our manuscript. We prefer this format to an online supporting text, as we would like to see the details of our methods included in the final printed version of our manuscript.

Finally, I just have two very minor issues:

L75: While the speed-up is remarkable, it is not necessarily true to be “several orders of magnitude”. Actually, it is for most of the tested range only little over one order of magnitude faster.

L193: don't you mean “… for values of FST LARGER than ...”?

A: We included these changes to our text. Thank you for the careful review of our paper and for the improvements that you have suggested.

Reviewer: 2

Recommendation: Accept, minor revisions

Comments:

The authors provide a new version of the program TESS that can be used for estimating ancestry coefficients and locus specific ancestral genotype frequencies incorporating geographic coordinates of the study samples. The main novelty is a substantial increase in the speed of computation in contrast to the predecessor version, which was mainly achieved by the extension of a nonlinear matrix factorisation approach (e.g. Frichot et al. 2014) over a Bayesian approach (TESS 2.3 Durand et al.). Locus specific estimates about ancestral genotype frequencies can be provided in at least similar accuracies in comparison to the previous program TESS 2.3, which were shown by extensive coalescent simulations. The authors propose in addition an outlier test to ascertain locus specific

deviations from estimated ancestral admixture proportions. They apply their approach to a dataset of Arabidopsis thaliana for deviations from neutral expectation and discovered two reasonable candidate loci that appear to be involved in phenotypic differentiation between Western Europe and Northern Scandinavian plant populations.

Improvements regarding the speed of an algorithm can be very important when studying large sample sizes from NGS data. The performance gain is quite substantial and very impressive and even goes together a slight improved quality of ancestry estimation over the Bayesian version of TESS. I particularly like the idea of utilising TESS for the study of frequency changes between populations that could by caused (e.g. polygenic) adaptation. I believe that this could broaden the range of application and audience. Overall I find the paper already quite complete as it is. I have only a few minor comments and suggestions that might help to round out the manuscript.

A: Thank you for those comments. We were pleased to hear that you liked the approach presented in this study.

First, a few questions that could be addressed in addition in the manuscript:

(1) It might be important to mention that the approach is applicable on haplotypic data, which I assume is the case, such as in Friechot et al. (2014).

A: We included a sentence clarifying that TESS3 can handle any type of allelic markers, including SNPs, microsatellites, AFLP and haplotypic data.

(2) Speed performance is one desired attribute of a program, how about memory consumption?

A: The memory size required by the program is of order O(nL), the number of individuals times the total number of alleles. We did not mention this estimate because the number is not critical to current applications of TESS3, and most users have more RAM than necessary to run TESS3 for their data. For data sets of moderate size, eg 100 individuals and 100k SNPs, the program works fine on a standard workstation (and on laptops).

(3) How severely would the estimates of the algorithm be affected by violation of independence (linkage disequilibrium) among sites? If so, it could be mentioned that sites should be in linkage equilibrium.

A: The method doesn't make any assumption about the data. In particular, it does not require that the sites are in linkage equilibrium. For the *A. thaliana* example, the SNP were genotyped dense (every 300bp), and adjacent markers were in strong LD. This was visible from Figure 3B where the correlation between p-values can be observed at nearby SNPs.

The impact of LD is to reduce the power of tests. Like PCA (Patterson et al. Plos Genetics 2006), the reduction in power corresponds to an "effective" number of independent SNPs, L'. We did not mention this point because estimates of L' are difficult to obtain. As we used low levels of expected FDR in our analysis (conservative tests), the risk to include false positives is then very small.

Second, a few minor line by line comments:

Equation numbering stopped at pg 5 ln 111 and could be carried on.

A: We changed this.

In formula on pg 7 above ln 130 the variables Id and type of product symbol wasn't defined.

A: We changed this, and defined the variables.

On page 11 line 241 the relation to machine learning in this context might not be clear to the reader and could be explained briefly (or omitted).

A: We used "matrix factorization" (MF) instead of "machine learning". Nonnegative MF is what we really applied here. MF methods are very popular in the machine learning community. This explains our previous sentence.

Thank you for the careful review of our paper and for the improvements that you have suggested.

Reviewer: 3

Recommendation: Accept, minor revisions

Comments:

Caye et al present a new version of TESS, as software used for the spatial analysis of population structure. The extension makes the application of the method feasible for large genomic data sets and, in addition, the authors add the option to detect outlier loci. The paper is well written and the method seems to work well. It will be a tool interesting to use for many in the field of spatial genetics/genomics. I only have some minor comments on the manuscript.

A: Thank you for your comments.

Since an important motivation for the development of TESS3 is running time, a fair comparison to TESS2.3 is important. The running time of TESS2.3 is highly dependent on the number of MCMC iterations. The authors chose 1000 iterations in all cases but the MCMC may have converged earlier in some scenarios. By picking a fixed number of iterations in advance, the authors kind of pre-set the running time for TESS2.3 while the running time of TESS3 depends on the particular data set. Another observation which the authors chose to not comment on is the fact that TESS3 needs more time than TESS2.3 for small numbers of loci and high numbers of K.

A: You're right. The MCMC program of TESS 2.3 reached its equilibrium state before 1,000 cycles (about 200 cycles were sufficient). But running the program longer than 200 cycles was necessary because the Monte Carlo estimates of Q were computed by averaging over a run. For large values of K (and n) we need to average over many cycles in order to obtain accurate estimates of Q.

A big issue with MCMC algorithms is that no reliable procedure warrants a pre-specified accuracy. Though they can't avoid local minima, minimization algorithm stopping rules are much clearer than for MCMC programs. Here we used a relatively small value for the number of cycles, close to the default value for the program. But remember that Pritchard et al. original suggestion for STRUCTURE was of order of 100k cycles.

We modified the discussion section and mentioned that TESS 2.3 is still preferable to TESS3 when the total number of variants is small.

The 4Nm=20 chosen for neutral markers in their outlier simulations seem extremely high which corresponds to an neutral Fst of almost zero. This might overestimate the power of the outlier approach since the difference between neutral and selected markers is pretty high.

A: For 4Nm = 20, the neutral markers were still difficult to separate from the non-neutral ones (4Nm = 1). This can be explained by the high variance of the Fst statistics computed from coalescent simulations. We actually observed that the distribution of Fst under island models with M=1 and M=20 overlapped substantially.

We ran additional simulations to provide evidence that, though it was low, the power of the test was close to the optimal value. We modified the result section.

I understand that authors like to refer to their previous research and self-citations are not uncommon. I think, however, that referring to the softwares LEA (line 258) and POPS (line 262) are unnecessary since the functionalities are covered by TESS and those tools are used in completely different contexts than TESS.

A: We modified the manuscript and cited the references at their right places.

LEA contains methods that are similar to the ones presented here, for choosing the number of factors and for running the tests of neutrality.

POPS was cited when we used Flora Jay's scripts to display Figure 3 and Figure 4 (the same approach was used in jay et al. 2012).

I think the legend of figure 3 should read as B instead of C.

A: We modified this. Thank you for the careful review of our paper and for the improvements that you have suggested.