Dear Matteo Fumagalli,

We have now revised our manuscript following the suggestions from the first round of revision. A more detailed discussion of each issue is provided below. As a summary: no new simulations have been included in the work, but we have expanded the discussion about model assumptions. Confidence intervals for the parameter estimates are now provided by DriftTest (through an approximate bootstrap approach) and are reported for the population of *Medicago truncatula* studied. The code has been documented, adding a readme file with instructions on how to run them and descriptions of the input and output. We have revised the text for typos and grammar.

Please consider the new version for recommendation by PCI Evolutionary Biology.

Sincerely,

Miguel Navascués, on behalf of all authors.

Note: page and line numbers refer to the version of the manuscript with highlighted changes.

Reply to comments by Matteo Fumagalli

"[...]
The main point raised by all reviewers is on the assumption of no population structure. While Authors acknowledge and discuss this issue, I believe the study will greatly improved if Authors provide more intuition of how much population structure / discontinuity / metapopulations would affect their results (e.g. estimation of parameters and power to detect selection). I am not advocating for additional large-scale simulations but for more specific discussion on potential limitations for not including more complex but realistic scenarios (e.g. change of Ne, variation in recombination rate, linked selection on deleterious alleles)."

We agree that many population genetic processes have an effect on the methods for detecting loci under selection and they need to be addressed in the development of new methods and their application to real data. The purpose of this work is to characterize one of these processes (selfing) to better understand its effects and the potential solutions. This is why it was studied without other confounding factors: to study the specific problems of selfing. It was not our intention to suggest that the assumptions made in our simulations would be widely applicable to any data set. However, the previous version of the manuscript did not seem to convey properly this message so we have revised it as suggested (page 10, lines 818-870). More specific concerns of the Reviewers regarding this point are discussed below.

"The text on the methodology should be clarified, as pointed out by Reviewers. This is an important aspect to avoid readers having to extensively look at cited papers to understand the methodology. I also found it difficult to understand, for instance, when Authors used unlinked or linked SNPs in different analyses.
Another point raised by one Reviewer is how to evaluate the statistical uncertainty of parameters' estimates. Likewise, there are some concerns on the use of the arbitrary threshold of 0.05 for minimum global MAF. The text should be either clarified or additional results varying this threshold should be presented."

The text on the methodology has been clarified following the suggestions from the Reviewers (see details below). The other methodological points have also been addressed and are also discussed below.

"I have an additional comment. I appreciate the discussion on how to design sequencing experiments in light of these results. I'd like to see future directions and ideas for improving the detection of selection for selfing to be elaborated a bit more carefully. For instance, Authors briefly mentioned ABC and as such I wonder whether Authors have more precise thoughts on which aspect of the methodology could be improved to achieve higher power (e.g. use of more features than single allele frequencies? Different inferential framework such as ABC or ML?). This doesn't have to be too extensive tough."

We have expanded this part of the discussion to better described our view for future developments based on the results from this work (page 10, lines 840-845, 863-867, 871-886).

"[…] Should the estimates of Ne on real data be presented in the results as first instance?"

We think that is makes a more simple and structured text to report together all results from *Medicago truncatula*.

"As beneficial alleles are randomly assigned to a position, is there any border effect if such sites are too close to one of the extremities of the simulated region (e.g. for Fig. 3)?"

The only "border effect" that we observe in our simulations is the reduction on the the number of polymorphisms affected by hitchhiking. In Fig. 3, loci from both sides of the locus under selection are used to calculate the signal of selection. A simulation in which the locus under selection is in the very extreme of the chromosome would contribute only around half of SNPs for the estimates presented in Fig 3, compared to a simulation with the locus in the middle of the chromosome. Visual inspection of Manhattan plots for each simulation did not suggest any other border effect.

"I appreciate that all scripts are provided but the documentation is rather thin and as such it is possible but unnecessarily laborious to replicate all analyses reported herein."

We have worked on improving the documentation of the code so it can be reused. Versions with the documentation are now available both in Zenodo and GitHub (see references in the article).

"Were the parameters of simulations chosen to match any organism of interest? I believe some reference that these values (mutation and recombination rate, Ne) are what expected in nature."

The text has been expanded to explain the choice of parameter values for simulations (page 4, line 361-363). Note that some of the parameters are chosen to produce strong sweeps, rather than based on any particular organism. Parameter values were chosen to better contrast the effect of the reduction of power due to selfing. Under weaker sweeps, this reduction of power would have been more difficult to visualize. This is also discussed as part of our reply to the one of Reviewer 1's comments (see below).

"Please provide a citation when introducing the equation Ne=(2-sigma)N/2."

References added (page 3, line 295).

"Finally, please also address all minor issues raised and check your text carefully for typos as I was able to spot a few (e.g. "?." on page 8, line 9)."

We have revised to text and several typos (hopefully all) had been corrected.

Reply to comments by anonymous reviewer 1

"[…]
One of the main conclusions is that in partial selfing species, it is easier to detect sweeps from standing variation. However, the authors applied a MAF filter, discarding sites where the average minor allele frequency between the two time points was less than 0.05. Could this conclusion might simply reflect the fact that sweeps from new mutations with initial frequency of 1/2Ne did not have enough time in 25 generation to reach an average frequency larger than 0.05? Another potential related problem is that the test is based on the single site distribution of FST under the null hypothesis of drift. It is known that the maximum value of FST estimators depends on the minor allele frequencies, and hence for rarer alleles we expect lower bounds for the maximum FST value (e.g. Jakobsson et al. 2013 Genetics doi:10.1534/genetics.112.144758). Can these results reflect such limitations with single site FST-based estimators? The authors justify the choice of the MAF filter based on deviations from a uniform distribution of p-values under the null

hypothesis. However, they used an arbitraty threshold of 0.05. Given that it can the impact the detection of sweeps from new mutations, I think that the effects of such filtering options need to be better explored (e.g. MAF=1/2Ne, MAF=0.01).”

Beneficial allele frequency trajectories (e.g. Figure A: a. sweeps from new mutations, b. sweeps from standing variation, initial allele frequency 0.1) were recorded in every simulation and the presence of a selective sweep verified: the situation described by Reviewer 1 was not observed among the simulations presented in Fig. 3a. The resulting allele frequency trajectories were not reported because indirect evidence of the action of selection is presented through other results (e.g. Fig. S5 shows higher F_{ST} in simulations with selection on new mutation and different levels of selfing compared to neutral expectations).

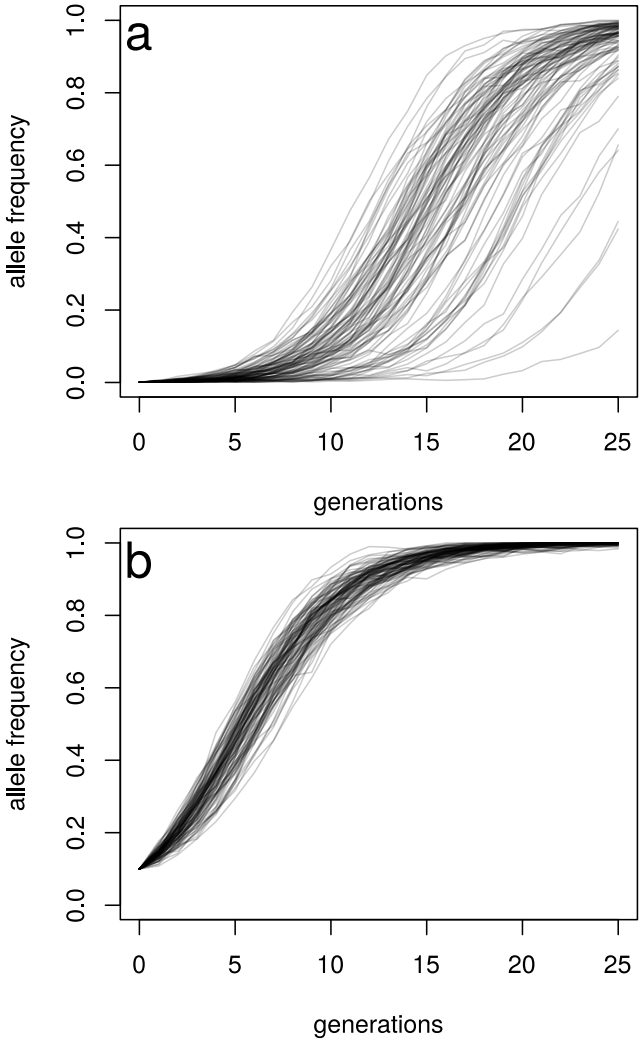

Figure A: Allele frequency trajectories for 100 simulation replicates of scenarios of adaptation from new mutation (a) and from standing variation (b).

Jakobsson *et al*. (2013) show that the maximum value of F_{ST} depends on the allele frequency across populations, *M*. Intermediate *M* values will have high maximum F_{ST} values while very low or very high *M* will have much more restricted range of F_{ST} values. In our simulations, a beneficial allele starting at a low frequency (1/2*N* for new mutations) reaches high frequencies after 25 generations (Figure A), so large values of F_{ST} are obtained (*M* value across temporal samples is intermediate). Therefore, this is not the reason for a weaker signal of selection in the case of sweeps from new mutation on a highly selfing population.

We disagree, for the reasons exposed above, that the MAF filter threshold had an impact on the lower power for detection of selection from new mutations. Simulations with different values for the filter threshold do not change substantially the results (see Figure B, note that it is not the same scale as in the figures in the manuscript). The MAF filter removes loci that start at very low allele frequency and get lost or remain at very low frequency, and would have a high *p*-values if tested. The main effect is to have less loci and to get a more uniform distribution, but *p*-value of the remaining loci do not differ substantially. We acknowledge that the chosen MAF threshold value is somehow arbitrary. The value will determine how many loci are lost for the test and how well the correction for the *p*-value distribution is done. In practice, however, there is no need for a fine tuning of this threshold.

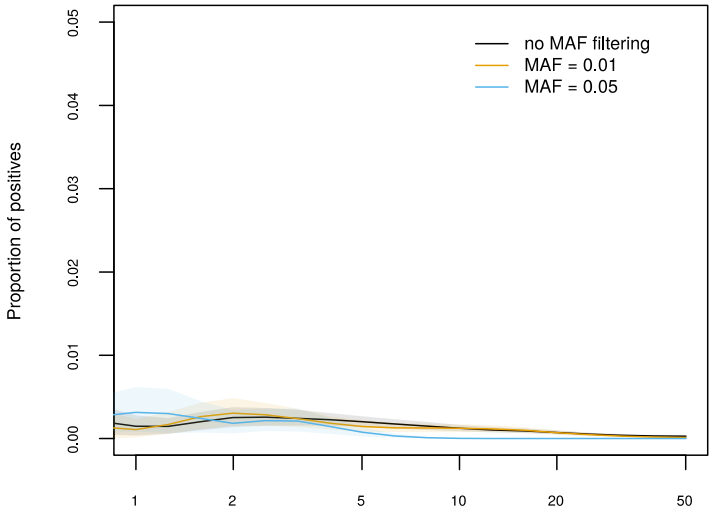

Figure B: Selection footprint on the selected chromosome using different values for MAF filter. Selfing rate σ=0.95, τ=25, s=0.5, adaptation from new mutation. x-axis: distance to selected locus in cM

"The authors extend previous methods based on a two step approach, common in genome scans based on outlier loci detection. First, based on the entire dataset they estimate the effective population size (Ne) and selfing rate. Second, conditional on those estimates, they obtain the single site distribution of FST under neutrality and compute a p-value for each site. The authors use point estimates of Ne (based on FST) and of selfing (based on FIS) in the first step, ignoring the uncertainty on FST and FIS estimates. When analysing the real dataset from M. truncatula no evidence for sweeps was detected. The explanation was the very small Ne (~40) and a very large selfing rate (~0.97). How much uncertainty is there on the estimates of Ne and selfing in the real data? How would be the conclusion affected by accounting for the uncertainty? I think this needs to be further investigated, e.g. by considering simulations done with values within the range of the confidence interval for FST and FIS, which could be obtained with resampling methods."

Confidence intervals for the effective population size and selfing rate have been obtained for *Medicago truncatula* following a bootstrap approach as suggested by Reviewer 1. Taking the more favorable values from these intervals (higher effective population size and lower selfing rate) as values for the test had very little impact on the results (see Figure C: *p*-values calculated for each locus in the *Medicago truncatula* data). We have incorporated the estimation of the confidence interval into the manuscript (page 6, lines 463-466, and page 7, lines 569-575).

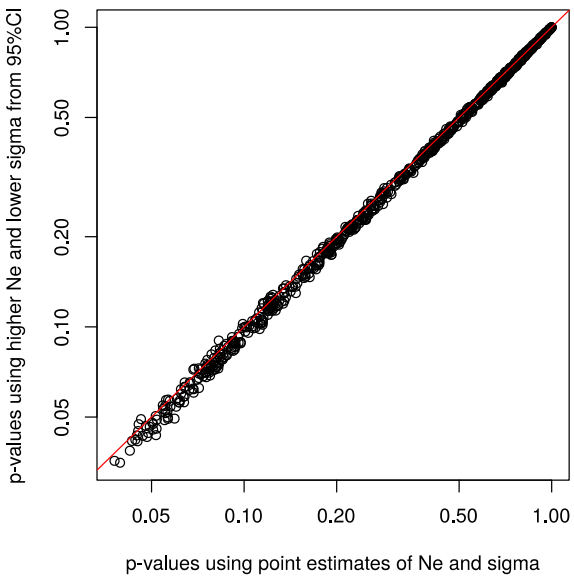

Figure C: Results of the Fst outlier test on data from Medicago truncatula using as a null model point estimates of Ne and sigma (x-axis) and "favorable values" from the 95%CI (y-axis))

"a) Page 3. Null distribution of drift. Please clarify in the text that when you obtain the null distribution of FST the underlying assumption is that all SNPs are independent, and that there is no account for linkage disequilibrium in the null distribution. This might be particularly relevant for selfing species where we expect linkage disequilibrium can have a genome-wide extent."

Each locus is tested separately, that does not mean that there is an assumption that SNPs are independent. Only information from the tested locus is used, so the amount of linkage disequilibrium has no influence on how the test is performed. We agree that "linkage disequilibrium can have a genome-wide extent" in predominantly selfing species and this is discussed throughout the paper.

"b) Page 3. Typo in the posterior of the Dirichlet for the genotype frequencies? I think that you mean that you use as a prior for the genotype frequencies, which I denote by the vector gamma0, a uniform Dirichelt D(gamma0, 1). By using information on the genotype counts at time 0 (K0), you get the posterior for the genotype frequencies, which will be described by a Dirichlet D(gamma0, 1+K0). In the text it seems that the posterior is a Dirichlet D(K0, 1), which I think is incorrect. Please clarify."

Typo corrected (page 3, line 247).

"c) Page 3. Equations of genotype frequencies. By using the FST and FIS estimators of Weir and Cockerham, I assume that the estimated FIS used to obtain the genotype frequencies (equations of Page 3) are based on the average FIS of the two time points. However, it is possible that the FIS would be different at time point 0 at the beginning of the sweep and at time point tau. Please clarify this in the text. Also, I am wondering whether the FIS estimated at time 0 and time tau could help distinguish the effect of selfing from the effect of selection, and hence increase the power of the selection test? Imagine that at time 0 you have a FIS0=0.9 and at time tau a FIStau=0.95. Assuming that the selfing rate was constant in the ancestral pop before the sweep, could the difference between the FIS estimates indicate the effects of selection on linked variation?"

Yes, we are assuming that selfing rate is constant and F_{IS} estimates comes from both samples, this has been clarified in the text (page 3, line 262). Single locus F_{IS} estimates are unlikely to yield much information on any temporal dynamic, particularly for very high levels of selfing in which most individuals are expected to be homozygous.

"Page 3. Simulations. The simulations assume a constant Ne. Given that the Ne before the sweep will affect the recombination events and standing genetic variation, and given that you find that higher historical recombination results in higher power to detect selection, could your method have increased power by simulating a larger Ne before the sweep? Indeed, selection might occur when migrants colonize new environments (founder event) and for some species it is plausible that the Ne during the sweep is lower than before."

As Reviewer 1 points out, scenarios with a large effective population size before the sweep are expected to have more historical recombination which could increase the probability to leave a local signal of selection on standing variation. We have incorporated this idea in the discussion. Note, however, that there are no assumptions on the method regarding the effective population size previous to the first sampling time.

"Page 5. Real data application. The description of the data is unclear – were the plants kept in the greenhouse since 1987? This would mean that several mutations could have occurred since then. I guess you mean that the seeds were maintained and then germinated. Please clarify that since the plants are annual you assume a generation time of 1 year. Is that correct?"

Plants were not kept in a greenhouse since 1987. Seeds were stored an germinated in 2011. The text has been modified to clarify this point (page 5, lines 418-421).

"Page 9, Arabidopsis thaliana results. "Based on the simulation results we present here, we can assume that this population has been adapting from standing variation over a short period of time (eight generations) rather than from many new mutations occurring during (or shortly before) the studied period." Where do these results come from? I could not find the description of the data for Arabidopsis and the eight generations. Does this refer to M. truncatula results? Please clarify."

Arabidopsis truncatula results are from another study. We have revised the text to make it more explicit (page 9, line 775).

"Fig. S2 shows the uniform distribution of p-values without selfing. What is the distribution of p-values with selfing? This should just be a re-scaling in Ne, but I think it would be important to see if with selfing the p-values also follow a uniform distribution."

The effect of selfing on the distribution of *p*-values were shown in Fig. S7. Distributions are largely uniform but deviate from uniformity for small *p*-values, as discussed in the text.

"Fig S5 legend: Typo? Unclear what is the boxplot mentioned in the legend"

Yes, text has been removed as it described a previous version of the figure.

"Fig S7 legend: Typo. Please use symbol sigma rather than "sigma"."

Corrected.

There are a few other typos, especially in the results and discussion.

We have revised to text and several typos (hopefully all) had been corrected.

Reply to comments by anonymous reviewer 2

"[...]
1) I have a general concern for the manuscript in how there seems to be a misspecification of what a 'population' is, and so while the intent is to suggest that increases in selfing will generate problems for inferring selection with temporal sampling, one would expect such to be the case in almost any situation in which a misspecification of this sort arises. If it's not just a problem with misspecification, could one then expect the method to perform well in continuous

and discrete population structures? If the former is the case, then the utility of the approach is quite a bit larger than the present manuscript's focus."

The problems our work addresses are not just a matter of misspecification. Selfing can cause genome-wide hitchhiking effects that remove the local signal of selection. This is a question of biology not model misspecification.

"2) The authors suggest that limited, or 'effective', recombination results from selfing, but then this is the same effect that arises as the result of spatial population structure. So while population structure certainly increases gametic LD, it's not for a lack of recombination, and it's the same for the case of selfing. This then provides a reiteration of point 1) above."

The effects of population structure and selfing on recombination are not the same. In an outcrossing, structured population, the sweeping haplotype will arrive at each subpopulation and recombine with the different genetic backgrounds present in it. In the predominantly selfing population the sweeping haplotype can spread and get fixed without any recombination event with a different genetic background.

"3) It is generally assumed that selfing, or a lack of outcrossing, is invariably a process that limits adaptation. I guess I can generally agree with this point, in that at least empirical literature has provided limited evidence to the contrary. But then I wonder what role variation in distinct 'lineages' carrying particular 'ancenstral advantageous alleles at low frequency' which then 'leave the strongest local signal' might again be an issue with misidentifying where and how selection is acting, thereby driving a disproportionate deviation from biological reality? This point concerns how beneficial alleles are considered while the next point considers deleterious alleles.

The selection detected on "ancestral advantageous alleles at low frequencies" was on simulated data, for which we know "where and how selection is acting". We have not argued that to be the case for the results from empirical data study or the general case in nature. A positive result in the test presented in this work does not provide such kind of information.

"4) The effects of selection on linked sites is clearly important in understanding the distribution of genetic diversity across the genome. In a selfing species this might actually be even more important. Point 3) concerned the role of beneficial alleles but opposite might be even more important for general application of this method as there probably exists a general pattern of increased genetic load with increased selfing (though factors like Ne and Fst will clearly play a role in modulating just how much load exists). Given just how much constraint to positive selection which might arise through the effects of linked selection on deleterious alleles with high selfing/gametic LD I think it's important to include this factor into the simulation approach. To not do so would certainly limit the degree to which the methodological advance might find direct use in natural systems."

A discussion of the effects from background selection has been included (page 10, lines 855-870).

Reply to comments by Christian Huber

"[...]
1) Maybe cite and relate the results to the recent paper by Hartfield and Bataillon (G3, 2020) on sweep patterns in selfing populations, e.g. regarding the size of the sweep in inbreeding populations for selection on new mutations vs. standing variation."

Hartfield and Bataillon (2020) developed an analytical model that shows the effect of inbreeding on the diversity and site frequency spectrum of hitch-hiking neutral variation after a selective sweep (both from standing variation and new mutation). Many of their results are difficult to relate to our work, because they focus on the diversity after the sweep, while our approach focuses on the differences with an ancestral sample. Nevertheless, they also describe how selfing rate and selection model changes the distance at which the effect of hitch-hiking spreads, as pointed out by Christian Huber. We have incorporated this into the text (page 6, lines 518-523).

“2) The description of the new method was somewhat hard to follow and I had to look up the original paper by Frachon et al. (2017) to understand it. I think it could be improved, for example by explicitly providing the steps for computing the significance of an Fst value at a specific locus.”

We have added a summary of the steps to build the null distribution to make the description more clear (page 3, lines 212-221).

“3) The biggest assumption of the method, most likely violated in most selfing plant populations, is the assumptions of no population structure. This is discussed, but I wonder if the authors could provide more intuition on how such structure would affect their parameter estimation (Ne, Fis) and in general the performance of the method.”

Further discussion on this has been included (page 10, lines 818-840).