Dear Fred,

Thank you and the reviewers for the positive and constructive evaluation of the revisions. I carefully addressed all issues raised and answered the remaining comments of reviewer 1 below in the sense that you indicated.

Best regards,

Markus

Decision for round #2 : *Revision needed*
Dear Markus,

Thank you for submitting your revised manuscript on popGWAS. The improvements are significant and I am willing to go on with recommendation pending some minor revisions. I attach a commented version of your manuscript where I have highlighted typos and places where corrections may be needed. I would also like to give you the opportunity to reply to the criticisms of reviewer 1. Many points raised by the reviewer are actually addressed in your discussion, I thus do not expect additional explanations for those points.

I would also like to ask you to clarify the iGWAS part as much as you can. Information on iGWAS in the supplementary material is missing and it would be informative for the readers to know more about the impact of population structure/relatedness on p-value inflation. I hope this would not require much more work from you.

I'd also recommend to avoid using "phenotypic plasticity" when referring to the "environmental variance" parameter. The two are used in the manuscript but it would be preferable to use only the environmental variance to refer to the random noise added to the genotypic value when computing the phenotype. A plastic response is expected to be more directional/predictive response to environmental variation rather than just noise.

Thank you for your contribution.

Best wishes,
Fred

Download recommender's annotations
*by **Frédéric Guillaume**, 04 Mar 2025 15:20*
Manuscript: **doi: https://doi.org/10.1101/2024.06.12.598621**
version: 2

Review by Petri Kemppainen, 07 Feb 2025 07:30

While the manuscript has improved a lot and there is a lot to like, there still are some issues that need addressing.

While I'm still not entirely convinced by the simulations that are still run for a very small number of generations (but where the allele frequencies are now at least drawn from a beta distribution with alpha=beta=0.5, instead of a uniform distribution), I am willing to compromise on this. The main reason is that given the same sequencing effort, popGWAS always outperformed iGWAS in the same data sets, which is indeed the main selling point of this manuscript (but also iGWAS cannot be used for poolseq data). Thus, whatever shortcomings there are with the simulations, I cannot see how any of them could bias the results in favor of popGWAS.

However, the only information about this iGWAS I can find in the manuscript is on L.325: "… a traditional linear GWAS on a random sample of individuals from all subpopulations was performed for all simulation scenarios …". If this is the case, I am not surprised that it did not perform in par with popGWAS, since the iGWAS seemingly did not even attempt to control for any relatedness in the data for instance by including relatedness as a random effect, which is I consider to be a bare minimum (for instance using EMMAX).

- As suggested by the reviewer in his comments to the first manuscript version as the minimum, I tentatively performed genomic control by dividing the -log10p value by the slope of the qq-plot for iGWAS. However, since this is a linear operation (and the mean value for iGWAS was close to 1 anyway, see below), it had no influence on the number of SNPs passing a fixed quantile threshold. Comparability of iGWAS with popGWAS with regard to the criteria for choosing candidate SNPs seemed more important and genomic control was therefore not included in the final analyses. The suggested EMMAX approach is computationally prohibitive for simulation studies ("The EMMAX algorithm uses a variance component approach that can analyze GWAS datasets within hours" Citation from the EMMAX website).

Also, it was argued that popGWAS is not sensitive to p-value inflation (L.662-), but for comparison, the levels of p-value inflation were never reported for the iGWAS results nor any attempts to control this (if present). I'm still quite confident that popGWAS can outperform iGWAS but before these shortcomings are addressed, I cannot know for sure.

- I never claimed that iGWAS was suffering from p-value inflation (contrary to the FET-approach between extreme value populations where p-value inflation was huge). As detailed in the answers to the reviewer's comments and illustrated by an example, iGWAS performed so poorly that the -log10p distributions of the TPL were mostly not distinguishable from a random distribution (i.e. qq-slope ~1, because with the given design, the chance of a TPL to turn up in the upper 2% quantile was on average only slightly larger than expected by chance). Therefore power, but not p-value inflation was the issue with iGWAS. The comparative distribution of the qq-slopes for iGWAS plus PPV and predictive accuracy comparison is now reported in the Supplements and a sentence regarding the absence of p-value inflation included in the Results.

Furthermore, I'm not convinced that a mean lambda of 1.27 (ranging between 1-1.99, L.440) can be interpreted as no p-value inflation, since the threshold is generally assumed to be 1.1. This is not that much of an issue in the simulations, mainly because the performance was assessed by considering the proportion of true positive loci (TPL) above a given quantile, not a fixed threshold. However, in empirical data, the -log10(P) would, on average, have to be 1.27x higher, provided genomic control was used, which may be the difference between many TPL being significant or not in empirical data.

- See discussion on qq plots in the answer to reviewer comments; the expected lambda depends on the difference in underlying distributions (i.e. the power of the method to identify TPL) as well as the ratio between TPL and neutral loci.

Perhaps more importantly, SNPs are assumed to represent haplotypes, thus the QTN are assumed to be 100% correlated with the SNPs in the simulated data. This is rarely the case in and to ensure that at least one SNPs is 100% (or at least close to) associated with a QTN (as in the simulations), much higher SNP densities are needed in empirical data for a given level of "LD-structure". In other words, the ratio between QTN (or SNPs strongly associated with them) and neutral loci randomly sampled in the genome are likely to be much lower in empirical data compared to the simulations making corrections for multiple testing much more taxing (there will many more chances of neutral loci to also be as differentiated as SNPs highly correlated with the QTN or the QTN themselves).

- Yes, that's probably right but would result rather in lower qq-slopes (see above). The highest outliers are nevertheless most likely TPL.

This does not take anything away from the fact the popGWAS performed better than iGWAS (but see above), but could be a reason why popGWAS may not be as powerful in empirical data as implied in the manuscript, which should be addressed in the discussion.

- See discussion in the paragraph "Typical genome sizes of real species are no obstacle"

One reason p-value inflation was not an issue for popGWAS was the overall high levels of migration between sub-populations (a minimum of 5 migrants per generation), and in addition the simulations were not run for enough generations for population structuring to come into effect (L. 416). Indeed, it seemed that Fst did not exceed much beyond 0.07 in any of the simulations (L. 669), which I still consider to be very low for many species and populations in the wild, and a level that I do not think would lead to much p-value inflation in the iGWAS data either (which is why I would like to see them reported).

- The initial simulations before the revision were without migration, which was the major point of critique back then (and performance was comparable, if not better). As the results in Supplemental Figure 3B show, the differences between no migration and 5 migrants per generation are not dramatic for 30 generations.

Perhaps the biggest weakness of popGWAS in this context is the fact that it relies on some level of "parallelism" (as accurately pointed out on L. 668-692) so migration cannot be too low, which is perhaps the reason why lower levels of migration were not tested. And, as I already pointed out in the previous review, the fact that the simulation started from a common pool of alleles with selection being imposed on the populations before population structuring has reached its full effect gives higher chances of parallelism than would be expected for a given level of gene flow in the wild.

- See Discussion and previous answer; I think any GWAS can only detect the trait loci common to at least he majority of the populations in the analysis and therefore requires a certain level of gene-flow, respectively recency of divergence.

The fact that these the simulations are not at equilibrium is also not an "advantage" as explicitly implied on L. 613-617, since the kind of non-equilibrium that was present in the simulations are highly unrealistic relative to the types of non-equilibrium conditions that can be expected in the wild (population size changes, range expansions etc.).

- This is indeed a point and the impact of classical demographic scenarios in an explicitly spatial framework should be evaluated in a future study. Respective sentence added to the discussion.

Furthermore, contrary to what is implied (L.668), many of these limitations are not unique to popGWAS but are expected to affect iGWAS performance as well – QTL involved controlling the same traits across larger number of populations/geographic ranges will overall be more correlated with habitat regardless of how this association is tested.

- Of course, you're right, but at this point, I was already beyond the comparison of the two approaches (which were not the primary focus of the study anyway) and just discussing what is needed for a successful popGWAS application.

Ideally, I would therefore want to see that the simulations were run until migration-drift equilibrium BEFORE selection is imposed.

- As pointed out in the discussion, equilibrium populations are not a particularly realistic scenario and letting them first differentiate and then select them appears to me highly artificial.

Given this, the limits of popGWAS may be apparent already with the, in my opinion, "high" levels of migration so far tested. If not, I would also like to have a level of migration that gives an overall Fst of at least 0.1-0.2, though even higher levels are of course not rare in the wild.

- This was the case in first version of the manuscript and popGWAS did not perform very well above FST > 0.07. This is a limitation, but I guess one that does not limit the applicability of the method much. The spatial range of the study might be tailored to include only populations below the threshold. This might be a problem with some species, but genome-wide FSTs > 0.1 are often subspecies, if not sister species. Most species I've been working with on a genome-wide scale (mosquitos, midges, trees, ants, wildcats, Daphnia, land snails) showed F_{ST}s << 0.05 within species, while only F_{ST}s among (closely related) species were larger than 0.1. Moreover, such a large divergence questions a large common genomic basis for the trait in question.

I still do not understand what is meant by the "phenotypic plasticity" parameter (e.g. L. 505). Since "phenotypic plasticity" is not explained anywhere in the text, from the context I assume this is the same as environmental variance, but there is a large difference between these two – plasticity is predictable, whereas environmental variance is not. I think I mentioned this in the first review as well.

- This is right, I've missed to change it throughout the manuscript, so some remnants remained, sorry. Clarified.

Lastly, the idea that "…complex traits are influenced by a few dozen genes that are mechanistically directly involved in their expression, but often also by numerous, if not almost all other genes as well…" is important but not commonly known (and perhaps a bit controversial) and requires thus some more explanation (L. 45-47).

- In response, I added the following sentence: "if only by the use of common resources and machinery (omnigenic theory)" .

Review by anonymous reviewer 1, 05 Feb 2025 13:00

The author made a very thorough job addressing both reviewers comments. I think that the new simulations to compare popGWAS to stablished GWAS-like methods is a great contribution to the discussion regarding the strengths, weaknesses, and releveance of the proposed method.

I think the current manuscript is rigorous, clear, and will be an important contribution to the community.

- Thanks!