Dear PCI review board,

Thank you for your time. We have revised this MS and believe that it has benefited greatly from your comments and suggestions. The results of the main table were better visualized in two new figures and all other figures were modified for increased clarity. The model parameters have been better described as has the distinction between linkage and mutation with respect to differences in mutational input. Please see our point by point response (in bold) to your reviews below, where the line numbers we provide refer to the new manuscript: https://doi.org/10.1101/656413

Best Regards,

Jobran Chebib and Frederic Guillaume

Response to anonymous reviewer 1 comments:
In this manuscript, the authors use individual-based simulations to explore to what extent closely linked loci affecting two separate traits differ from a single pleiotropic locus that jointly affects these traits, in terms of (i) the genetic correlation that is maintained between these traits at a balance between mutation and stabilizing selection (+drift), and (ii) the likelihood to correctly infer the genetic basis of quantitative traits in a GWAS. This study is partly motivated by verbal claims that fully linked loci should be equivalent to a single pleiotropic locus, since they cannot evolve independently from one another. The authors investigate this hypothesis in some detail, from the perspectives of both theoretical evolutionary biology and statistical genetics. I was quite interested to read this manuscript, which addresses a simple fundamental question that seems to not have received a fully satisfying answer yet.
The bulk of the paper focuses on the maintenance of genetic correlations between traits (with 7 out of 8 figures), for which the simulation results are compared to a prediction from Lande (1984) for fully linked loci as a reference. The authors perform an exhaustive analysis, investigating the effect of several key parameters: selection strength and amount of correlative selection, mutation rate and phenotypic effects, linkage, and migration. My main comment would be that, even though it is of course important to assess the influence of all parameters, it is also easy for the reader to get lost, especially since the presentation is mostly descriptive, and some explanation would seem to be required at multiple points. Below is a list of suggestions:

1 - First, I think the authors should explain a bit better what makes a difference between two fully linked loci and a single pleiotropic locus. They have identified an important difference that concerns mutation rates and effects, but I don't find their explanation sufficient. For instance, they write in the conclusion: "Without high mutation rates, the ability to create genetic covariance between linked loci is highly diminished because the combined likelihood of mutations in each linked loci with both mutational effects in the same direction is low." (466-469). I think this argument should be made more explicit earlier in the paper, as it"s central to the process the authors are investigating here. The joint distribution of mutations effects on two traits that can mutate via two (fully linked) loci is ff(xx 1 , xx 2) = [uu 1 (1 ? uu 2)ff 1 (xx 1)??(xx 2) + uu 2 (1 ? uu 1)ff 2 (xx 2)??(xx 1) + uu 1 uu 2 ff 1 (xx 1)ff 2 (xx 2)] / [1 ? (1 ? uu 1)(1 ? uu 2)]
where uu ii is the mutation rate at locus i, ff ii (xx ii) the distribution of mutation effects on trait i at locus i, and ??(xx) is a Dirac Delta function, which equals 0 except at x, and integrates to 1. For Gaussian distributions of mutation effects as here, ff 1 (xx 1)ff 2 (xx 2) in the last term is a bivariate Gaussian with no correlation, similar to the effect of a pleiotropic locus without mutational correlation (as modeled here). However, this last term rapidly vanishes with small mutation rates, such that we"re only left with a "cross" distribution, where all the probability mass is on mutational variation on one trait with the other trait fixed at 0. A Dirac delta is not convenient to plot as it goes to infinity, but here"s how this looks on simulated draws with

uu = uu 1 = uu 2 = 0.001: So the point is really that there is little opportunity for joint change in both traits, which make it difficult to select for genetic correlation. In comparison, the difference between correlated vs uncorrelated mutation effects at a pleiotropic is very mild, as illustrated below (with mutational correlation = 0, 0.5 and 0.9):
The authors agree that the importance of the difference between mutation rate in linked and pleiotropic loci should be made more explicit and earlier in the MS. We believe that a verbal description of the distinction clarifies this point in the introduction starting on Line 61.

2- Another point about mutation is that, as the mutation rate decreases, the Kimura-Lande Gaussian approximation used here is expected to be rapidly replaced by Gillespie's house of cards approximation. In fact, this HoC approximation seems to work fine over all the parameter range explored here, judging by figure S3. So it would seem useful to also report predictions for the genetic correlation based on this approximation, if they exist (perhaps from the Gillespie 1985 multivariate paper cited in the ms).
The authors agree that the HoC approximation predicts the level of genetic variation at equilibrium better at lower mutation rates but we were unable to find any HoC predictions for genetic correlation / covariance expected at equilibrium due to completely linked pairs of loci affecting different tratis, despite an extensive literature search. The lack of HoC expectation has been made explicit for the readers starting on Line 126.

3- Regarding the influence of migration, I also think more can be said, in particular about the fact that its "effect on genetic correlation is still observed when there is no correlational selection on the traits in the source population" (272-274). This occurs because of the difference in mean phenotypes between the two populations, coming from the fact they are selected for their different optima. The phenotype distribution in the focal patch is a mixture between the distributions of residents and migrants (and their subsequent crosses: F1, backcrosses, etc). From the law of the total covariance, the covariance of a mixture includes a component caused by differences in means of the underlying distributions, even when their covariances themselves are identical. Here, neglecting the contribution of later generations of crosses between residents and migrants, the local covariance matrix should be inflated by mm(1 ? mm)(zz rr ? zz mm)(zz rr ? zz mm)T ,with zz rr and zz mm the mean (bivariate) phenotypes of residents and migrants, respectively. The direction of the genetic correlation produced by this effect depends on the direction of phenotypic divergence between residents and migrants for the two traits. In fact, under strong differentiation this term may have a stronger effect on the genetic correlation than the genetic correlation in migrants themselves (here caused by correlational selection in their population of origin). This is worth discussing and exploring a bit further in your context of linked vs pleiotropic genes.
The reviewer is correct that the effect of migration on genetic correlation (even in the absence of correlational selection) is due to a difference in phenotypic means in the two populations. These results and their implications are more throughly investigated in Guillaume and Whitlock 2007, and therefore, not repeated here beyond their corroboration of those results. As for the differential effect of genetic architectures investigated in this paper, no effect was observed, and therefore little discussion was included.

4 - I have trouble grasping at the main result with respect to GWAS. Here there were no neutral markers in the simulations, so all loci were truly causal (although they may be causal only with respect to one of the traits). This means that the recombination rate between adjacent loci has two effects: it influences (i) the evolution of genetic correlation between traits, and (ii) the association between a non-casual locus and a causal one (with respect to a given trait). This should be written more explicit in the paper, as it bears on the interpretation of the results. From Table 1 and Figure S1, the main result seems to be that, for non-pleiotropic genes, the rate of false positives increases in proportion to the genetic correlation between traits, and for a given genetic correlation depends to a lesser extent on the recombination rate. This makes

sense, since genetic correlations comes entirely from LD in this case (the total genetic covariance is just the sum of LD multiplied by allelic effects at pairs of loci), and LD is also what causes association between a trait and a non-causal locus. So by setting the level of genetic correlation, the authors indirectly control the mean LD between adjacent loci (fig S2), which is also the mean LD between a causal and non-local locus for a given trait. It seems that this argument could be made a bit more explicit in the paper, making use of well-known mathematical formulas. You may also want to simulate/discuss what would be the difference with a purely neutral marker. My guess is that the probability of false positives would then mostly depend on recombination distance, because only this will influence LD in this case, not the genetic correlation between traits.
The authors agree that the main results of the GWAS could be made more explicit and their interpretations could be expanded to include a further discussion on LD as the source of genetic correlation in these simulations and a comparison with linked neutral loci. We have added this into both the results (Lines 298-299 and in Figures 8 and 9) as well as the discussion (Lines 407-417).

Minor comments:
Figures 1-7: why not also plot the expected minimum correlation under no linkage in these graphs, since this was also predicted by Lande (1984)?
The authors had versions of these figures with Lande's expected minimums included but it made the figures less readable, especially when multiple parameter values were being plotted. In the end, they were left out for clarity.

Table 1: it is unclear at first what distinguishes the first 4 lines from the 4 lines below; the difference is the genetic correlation between traits (modified by tuning correlational selection), but this should appear somewhere in the legends or table captions, rather than just in the methods.
This distinction has been included in Table 1 as suggested by the reviewer.

3 : "traits do not act independently of one another" is not really clear: traits don't necessarily "act" on anything. I would just write that traits are not independent, or do not to vary independently.
This has been changed to read “vary independently” on Line 3.

5-6: "reaching their respective optimal potentials" reads awkward. What"s the optimal potential for a trait? Do you mean the optimum trait value favored by natural selection? If so just write this instead.
This has been changed to read “optimum trait values favored by natural selection” on Lines 5-6.

61-63: "recombination can break up associations between alleles at linked loci, reducing genetic correlations between the traits they affect, but the same cannot occur with a pleiotropic locus". Well, in the longer run, duplication followed by sub- or neo-functionalization can alleviate potential costs of pleiotropy. This is worth mentioning somewhere, as it has been discussed as a mechanism leading to the evolution of modularity starting from pleiotropic genes (eg Wagner et al 2007 Nature Reviews Genetics)
This has been mentioned with the above citation included on Line 68.

74-76: "The detriment of pleiotropic effects is exacerbated when increasing the strength of selection or the correlational selection between traits". However, increasing correlational selection can rotate the fitness surface, in effect favoring pleiotropic effects causing joint change in multiple traits. How can you reconcile these two apparently contradictory statements?
This has been clarified on Line 83 by specifying the conditions under which this is true, “unless pleiotropic effects are aligned with the fitness surface created by correlational selection” on Line 82.

98-101: "From these equations we see that [...] genetic correlation [...] depends on [...] the mutational inputs (mutation rates and mutational variances)". This sentence is wrong as stated, since the equation just above (eq. 3) precisely shows that the genetic correlation does not depend on these

mutational properties, but only on correlational selection rho. That mutational properties may actually matter is because assumptions leading to eq. (3) don't always hold, which you explore here, but this is another issue.
This has been revised to read “genetic covariance” to make the included statement true on Lines 107-109.

175: I imagine that rho_omega is the factor that multiplies omega^2 two lines above, but please make this explicit.
This has been made explicit by including “and off-diagonal set to omega^2 times rho_omega (where the correlational selection, rho_omega = 0.5 or 0.9” on Lines 191-192.

182-183: This is another clear difference between pleiotropy and linkage: fully linked non-pleotropic loci have uncorrelated effects on the two traits, while pleiotropic genes may have correlated effects (even though here you assume a covariance of 0).
The authors agree and have made this more explicit in Lines 61-65.

242-244: "A decrease in correlational selection [...] has a larger effect on maintaining strong genetic correlations than a decrease in selection strength". This sentence seems wrong as stated. Why not just write that with the parameters you chose, a decrease in correlational selection reduces genetic correlation more than does a decrease in the strength of selection? (your sentence seems to mean the reverse, ie that higher genetic correlation is maintained under reduced correlational selection). Note however that these are difficult to compare, since they concern different parameters, which are expected to have different influences on on genetic correlations anyway based on analytical predictions, eg your eq. (3).
The authors agree that this statement can be re-worded for clarity and has been changed to “The genetic correlation between the traits decreases with reduction in all four factors tested and for all genetic architectures, with the coefficient of correlational selection (rho_omega) having the strongest effect, as expected from equation 3” on Line 265-268.

248-249: Perhaps you should refer here to figure S3 to convince the reader that that the equilibrium has been reached after 10000 generations at mutation rates below 10^-3. Note also that the mutation rate is not provided in figs 3,4, I imagine it it also 10^-3 as in fig 2?
We have changed Line 250 to refer to Figure 2 to show that equilibrium has been reached. We have also included the mutation rate in the caption of the relevant figures (now Figures 4 and 5) as suggested.

347-350: what you describe here is the Bulmer effect (Bulmer 1971), whereby negative LD reduces genetic variance under stabilizing selection. However, it's a bit unclear how you here extend this to the correlation between traits. Note also that this build-up of negative LD also occurs in the model by Lande (1976 for a single trait, and 1980, 1984 for multiple traits), so it does not seem to really distinguish strong from weak selection.
After reconsideration the authors agree with the reviewer that this statement does not necessarily hold up with regards to the correlation between traits and it has been removed.

Response to Reviewer 2 (Pär Ingvarsson) comments:
Chebib and Guillaume investigate how linkage and pleiotropy contribute to the genetic correlation between quantitative traits. They use computer simulations to evaluate how a number of parameters, such as mutation rate, linkage and strength of stabilising and correlational selection affect genetic correlations between two quantitative traits. The results show that pleiotropy in most cases maintain a stronger genetic correlation than linkage, unless causal loci are in complete linkage. They also explore how linkage and pleiotropy affect the ability to detect causal loci in a GWAS setting.
Overall the results enhance our understanding of how linkage and pleiotropy affect the genetic correlation between quantitative traits and will be useful for interpreting results from GWAS studies of correlated traits. I only have a

few comments that I think will help make the presentation more clear and aid with the interpretation of the results presented in the paper.
Major comments:
The model formulation of the selection is rather vaguely described. The strength of selection is described by the ?2 parameters and the correlational selection is given as ?2"0.5 or ?2"0.9 (line 173). However later in the manuscript, correlational selection is only described using parameter ?w. I think it would be good to include the parameter ?w in the formulation of correlational selection this more explicit. It took me a while to realise that the ?w parameter actually refers to the 0.5 and 0.9 values in the formulation on line 173.
This has been made explicit by including "off-diagonal set to omega^2 times rho_omega (where the correlational selection, rho_omega = 0.5 or 0.9" on Lines 191-192.

The presentation of the GWAS results in Figure 8 is a little intuitive in light of how the simulations were set up (as presented in Figure 1). Under the linkage model, pairs of loci are linked with one locus affecting trait 1 and one lcous affecting trait 2) and individual pairs are unlinked but when GWAS results are presented in Figure 8, the loci are plotted based on whether they affect trait 1 or trait 2. I assume this is done for visualisation purposes but is a little intuitive given the description given in Figure 1. Also, this is never explicitly mentioned in the text or the figure legend.
Also, the use of grayscale in all figures is sometimes not enough to clearly distinguish different parameters (especially when error bars are small). Using different grayscales in combination with different plotting symbols (e.g. circles, triangles and diamonds) would make it easier to distinguish between different parameters in the figures. Also, making the symbols in the plots a little bit larger (like in Figure 7) would also help.
The authors agree that these figures could have been presented in a clearer manner and the following suggestion from the reviewer have been included in the revised manuscript. The order of the loci in the GWAS analysis plot was indeed formatted for better visualization and this has been specified in the caption. Different plotting symbols have been incorporated along with gray scale to make Figures 3-9 more clear, and the data point sizes have been increased in cases where it led to more clarity.

Minor comments:
Line 25: "leverage the explosion in genomic sequencing" sounds a little dangerous, I reformulate to something like "...leverage the rapid development in genome sequencing technologies"
This has been changed to "rapid increase in genomic sequencing" on Lines 25.

Line 26: "on the size of the effect" - of what? Do you mean effect size of alleles?
This has been changed to "rapid increase in genomic sequencing" on Lines 26.

Response to Kathleen Lotterhos comments:
Both reviewers point out the merit of this simulation study, which tests verbal arguments that linked loci should behave similarly to a single pleiotropic locus. Both reviewers suggested clarifications to the text and/or extensions to the mathematics, with which I agree are necessary before this manuscript would be recommended. Specifically, clarification is needed for the model parameters throughout the manuscript, the figures need to be presented more clearly, and the explanation for the difference between two fully linked loci and a single pleiotropic locus needs to be made more explicit earlier in the paper. Reviewer 2 points out that there may be some important differences in the the joint distribution of mutational effects, and this need to be clarified in the manuscript. This reviewer also points out how the influence of migration may be predicted from the law of total covariance, which is worth incorporating into the manuscript.
See above responses to reviewer's comments.

Also, clarification on the demography is needed. Is this an island-mainland model? Or a 2-patch model with asymmetrical migration? What is the population size in each patch?
This has been clarified as "To examine the effects of migration from a source population on genetic correlation between traits, additional sets of simulations were run with uni-directional migration from a second population (as in an island-mainland model with each population consisting of 5000 individuals) with backward migration rates (*m*) of 0.1, 0.01, and 0.001." in Lines 206-209.

Both reviewers point out that the GWAS results are not clearly presented and I agree. Major revisions will be needed in this section if the paper is going to earn a recommendation. Firstly, it seems strange to do a GWAS analysis only on causal loci and not to include simulated neutral loci for the calculation of false positive rates. Second, it is below the standard of the field to conduct a GWAS without a correction for population structure. If structure is corrected for in the model it is unclear in the manuscript, and if it is not then false positive rates could be inflated. In the context of fully linked loci, "false positives" are linked loci that have an effect on a different trait other than the one being analyzed, so they are not truly neutral and this needs to be clarified. Finally, the presentation of results in Figure 8 is not intuitive, especially for the linked architectures - is locus 1 linked to locus 121 on the same linkage group? Linkage architectures should still have Type II error rates reported (even if these are zero) in Table 1. It"s hard to figure out what the main message from Table 1 is, so a figure here might be warranted.
The GWA results section has been changed significantly by replacing the Table with 2 figures that are intended to visualize the relevant data. "Figure 8" and Table 1 showing GWA results have been moved to the supplementary material section and their captions have been revised for clarification. No correction for population structure was included in the GWA analysis because there is only a single, large, randomly-breeding population (clarified on Line 227). Also, we agree that "false positives" are not neutral and this has been clarified and discussed in Lines 298-299 and 407-417.

Overall, I agree with the reviewer that said it"s easy to get lost in the results, especially in Figures 3-6. Streamlining the message would strengthen the manuscript.
The figures have been revised and the main results have been provided to the reader first (Figures 2 and 3). The main message has been made clearer throughout the MS.

Additional requirements of the managing board:
As indicated in the 'How does it work?" section and in the code of conduct, please make sure that:
-Data are available to readers, either in the text or through an open data repository such as Zenodo (free), Dryad or some other institutional repository. Data must be reusable, thus metadata or accompanying text must carefully describe the data.
These have been included in the section "Data Archival" that includes a Zenodo link: https://zenodo.org/record/3370185\#collapseTwo
-Details on quantitative analyses (e.g., data treatment and statistical scripts in R, bioinformatic pipeline scripts, etc.) and details concerning simulations (scripts, codes) are available to readers in the text, as appendices, or through an open data repository, such as Zenodo, Dryad or some other institutional repository. The scripts or codes must be carefully described so that they can be reused.
These have been included in the section "Data Archival" that includes a Sourceforge link: https://sourceforge.net/projects/nemo2/files/Publications-Code/ChebibGuillaume-PleiotropyOrLinkage-2019/
-Details on experimental procedures are available to readers in the text or as appendices.
These are included in the text.
-Authors have no financial conflict of interest relating to the article. The article must contain a "Conflict of interest disclosure" paragraph before the reference section containing this sentence: "The authors of this preprint

declare that they have no financial conflict of interest with the content of this article." If appropriate, this disclosure may be completed by a sentence indicating that some of the authors are PCI recommenders: "XXX is one of the PCI XXX recommenders."

A conflict of interest statement has been included.