

Paris, July 7, 2023

Dear Frédéric,

Thank you for handling and reviewing our manuscript for PCI Evolutionary Biology. We are happy to send a revised version for your consideration.

I am sorry that it took us so long to send you this revised version. We found your comments, and those of the reviewers, quite helpful and we have followed most suggestions. You can find a point-by-point reply to your comments below, as well as comparison the original and revised manuscript in a separate file.

With my best regards,

Henrique

Reply to Editor:

Thank you for submitting this very interesting manuscript to *PLoS Evolutionary Biology*. I have received two reviews of your manuscript. Both find interest in publishing your findings, with one very enthusiastic reviewer. I do also find value in the reported data and analyses. There are nevertheless some major points that have been raised by both reviewers, and myself, which need to be addressed in a revision of your work. The suggested revisions will necessitate some complementary analyses on the effect of sampling of the inbred lines in the CA populations and some additional effort with streamlining and focusing of the text, especially in the Results and Discussion sections.

Reply: We have followed your suggestions and those of the reviewers, please see next replies and also the comparison file between the original and the revised versions of the manuscript.

I would like you to pay particular attention to the comments of reviewer 2. I do also find that the marked difference in G-matrix size between A6140 and CA1-3 are suspicious. Please carefully address the reviewer's point here. I'd suggest that you use the down-sampling of A6140 inbred lines to evaluate the effect of smaller number of inbred lines in CA1-3 on the reduction of the trace of G-matrices in CA1-3.

Reply: We thank you and reviewer 2 for this comment. Indeed, the reduction in G of the CA1-3 populations, particularly at generation 50, is odd. We have down-sampled the A6140 inbred lines to a number similar to the average number of lines from the CA1-3 populations. Results can be found in the supplementary Figure S8. We don't observe a reduction in variance, though the error of the estimates increases. Given this analysis, we don't think that sampling a limited number of lines explain our observations.

We went back to the log books of data collection and revised and re-analyzed all the raw data. There is nothing particular that we could find about the CA1-3 populations during experimental evolution or inbred line derivation. However, phenotyping of the CA[1-3]50 lines was done almost exclusively by a unique experimenter, at different calendar dates, when compared with A6140 or CA[1-3]100 lines (detailed in the methods). We find that mean body size is consistently smaller in CA[1-3]50 lines than in other lines, but cannot explain why the variance estimates of transition rates are affected. We suspect that the height of the camera during recording was different for these lines. Unfortunately, we could not statistically account for these scaling effects and, as a consequence, we no longer present the results of the CA[1-3]50 populations together with those of the A6140 and CA[1-3]100 populations. To show divergence we rely on a comparison between A6140 and CA[1-3]100 only (Figures 3 and 5), and to show differentiation between replicate populations we present the eigentensor analysis of the CA[1-3]50 populations, as before, but separately from the other populations (Figure S10).

To address the question of the reduction in variance in the CA[1-3]100 populations, we now present an analysis of A6140 and CA[1-3]100 lines that were measured by the same experimenter, at similar calendar dates (it therefore does not include all the A6140 lines). This new data and analysis show a similar reduction in variance (Figure S9), as is the case when all A6140 lines are used (Figure 3). The reduction in variance is thus reproducible.

Similarly, the discussion is in part off topic when discussing un-observed increase in variance relative to ancestral population. I did not find any evidence reported in your manuscript about the increase of variance in y_5 - y_6 as suggested in lines 496-498. In fact, reduction of variance along those axes of selection is reported previously, lines 475-476. The following discussion start with "If true" (line 498), speculation that doesn't seem supported by your data.

Reply: We have eliminated this paragraph and greatly reduced the discussion on the maintenance of genetic variation despite stabilizing selection. Admittedly, we have no power to say that there is maintenance of genetic variation above that expected by drift for any selection axis. We nonetheless would like to keep some of the discussion on the problems with estimating selection and how genetic variance might maintained by balancing selection at the relevant QTL (for which we have some independent evidence).

The first reviewer points to the lack of a better discussion on the implications of your study for the field. I would add to this that the manuscript is in general rather dry and technically oriented which will be a deterrent for a more generalist audience. I would strongly suggest to bring in more biological examples or interpretation of some of the methods, especially those dealing with canonical analyses of covariance matrices. Some basic mention of the biological meaning of vectors in phenotypic space should be added.

Reply: We have provided more detail, particularly on the biological significance of genetic variance and selection axis, while generalizing some of the results to wild populations (last paragraph discussion). We have added two new tables, with the eigen decomposition of the ancestral G matrix and the gamma matrix, and further discuss these results (line 588 paragraph).

I would add that I do not find strong evidence for "squashed stabilizing selection" in your manuscript as mentioned on lines 469-470. At best, it suggests such possibility, but my understanding is that squashed stabilizing selection requires some form of frequency-dependent selection. More analysis or justification for a statement about whether your finding "provide evidence" of such selection regime are necessary.

Reply: Correct. We have added more justification for frequency-dependent selection, based on our previous work on life-history evolution and the population genetics/genomics of the experiment (first paragraph discussion). We do not provide more analyses, as it is unclear how squashed stabilizing selection could be shown with the data we have.

Review by **Benoit Pujol**

This is one of the most interesting paper I was given to read in a long time! I found it very clear. The authors did a brilliant job at presenting complex mechanisms (and the complex methods used to evaluate them) in a way that is easily understandable. Understanding how and why stasis can occur at mutation-selection equilibrium after a history of adaptation without blocking other evolutionary processes that shape the diversity of populations such as divergence is crucial. This is important because the disconnection between the theory and the empirical results found in wild populations is widely explained by the action of this or this mechanism. We need to integrate this perspective on mechanisms influencing the evolutionary potential of populations to respond to selection in a broader evolutionary perspective of stasis.

Reply: We thank you for your enthusiastic support of our work.

Below you will find comments, many of them positive, and some suggestions. I have only one point that I'd really like the authors to take into account (see my last comment): the implication of this work in terms of contribution to the knowledge in the domain and how it participates to push forward the theory is not clearly stated. I'd like to see it addressed by the authors.

Reply: We have added a brief conclusion on how short-term evolution can impact long-term evolution (last paragraph of discussion). With the exception of numerical simulation studies, current analytical models of genetic differences between populations or species do not seem to consider within-population variation.

L30-31 I'm not a native English speaker and I trust your writing skills. Please check that this sentence is correct. I was in doubt because "important" came right after drift.

Reply: Corrected.

L42 These studies indicate... I totally agree with your point that a multivariate approach is necessary to understand stasis, and you have already mentioned that other mechanisms were at play (plasticity, indirect environmental covariance, etc.). Obviously, other mechanisms should be considered to and it's impossible to investigate all of them simultaneously. It is only possible to study them in isolated studies, even in a given study system. If you feel like saying a word about this, feel free to do so.

Reply: It is difficult to list them all while keeping readability, and opted to refer to your 2018 review paper where you list several different mechanisms.

L50 Aren't we missing a "to" btw "expected" and "facilitate"

Reply: Corrected.

L 66 Aren't we missing a comma after "occurred"

Reply: Corrected.

LI 74-86 I was missing something here between the conceptual state of art and the down to earth presentation of the experiment. I think that moving the last sentence of this paragraph: "We seek to once most adaptation has been achieved." At the beginning of the paragraph would help connecting these aspects.

Reply: Thank you.

LI173-174 Not sure to understand. Do you mean that they were not significant as covariables in a linear regression analysis and you therefore removed them before fitting an even fitter model? Or that you used the residuals of the linear regression as data to feed the model fitting procedure.

Reply: It is the second option, we've reworded the sentence.

I think that the differentiation approach btw males and hermaphrodites is correct. I might have preferred other (more direct) approaches but I can understand why the authors did it this way in this study system.

Reply: Given the number of samples to phenotype, it would have been cumbersome to obtain time series data. But we did extensively calibrate the model. One of the advantages of our modelling is that we can now use it in future assays/experiments.

LI226-229 "As we derived" Anyway, there is no "true" estimate of COVa since nobody can exclude unexpected indirect and/or environmental covariances, and you then explain that you estimate a broad-sense G matrix. You mention the assumptions but it is difficult to predict what bias it may generate in the results, if any. If you feel that it is justified, do not hesitate to add a few words about it.

Reply: As you write this is a difficult topic, though a classic one in quantitative genetics. We refer to chapter 3 of Kersey and Pooni (1996), where a discussion of how variances change with inbreeding can be found. Had we found a positive result (that the mean trait values were different when comparing the outbred populations with the inbred lines) then we feel that a more detailed discussion would be warranted.

L 259 please check whether "were fit with" is correct phrasing

Reply: Thank you.

L275 “this decomposition of describes” sounds like we’re missing a word here, please check

Reply: Corrected.

The results section globally reads well, but I can’t help thinking that it includes a lot of methodological aspects that belong to the methods rather than the results in the sections 4.1, 4.2, 4.3. I appreciated that you remind the reader (me in particular) what is done because it is a dense paper that includes complex methods but it would be good if the focus was ore on the results than the methods in the method section of the main text. This is very minor and the take home message of my evaluation of the result section is that I greatly appreciated that the authors took great care to make their results accessible to a broad readership.

Reply: We have maintained the methodological details in the results section, because, as you mention, we think it is difficult for the reader to understand all the complicated modeling we did.

LI 329-348 these elegant and easily understandable synthetic sentences about the lab culture read nicely. It nevertheless feels like I already read some of it in the methods and the Figure legend. Please check for direct redundancy.

Reply: Some readers might initially only briefly go over the methods section and will focus only the results. We believe it is important to summarize experimental design at the start of the results section.

There are many figures (when including the supplementary) to review, but I understand why they are necessary. On a global note, the legends are clear but the figure themselves are not always easy to read because text, bars, etc appear very small/thin.

Reply: We have revised all figures for clarity.

The discussion about the role of selection pressures and their signature on G matrices (with a full overview of the effect of divergent, stabilizing selection and drift) is very clear and elegant. I have to admit that I had to read the paper a few times to fully grasp all of it.

I appreciated very much the honesty of the authors who mentioned the various scenarii that could explain their results (original lack of standing variation, sampling error, etc.) I agree that the indirect effects of environmental covariances, which could be interacting with this kind of study, is unlikely to underly unexpected genetic variance in this particular study.

Pleiotropic effects are indeed a possible underlying mechanism, but I’d tend to say that this is not a wild population and that selection was artificially directed at these traits. I’m not saying that it excludes the potential pleiotropy but it will be a lot more messy in the wild. I find that you could have included more information about the balancing selection hypothesis. I find fluctuating selection is very interesting because its impact can vary from loss to gain in genetic variance depending on the intensity and frequency of directional changes, which generates very

interesting processes. I'm not saying that you are in the presence of fluctuating selection. I'm merely saying that a bit more information about the features of the potential balancing selection at play here, in a context of stabilizing selection, would make this point more interesting and clear to me, and potentially others.

Reply: We have a few sentences about balancing selection, based on our published and unpublished work (third paragraph discussion). It is unclear if there is frequency-dependent selection, although we can speculate about it, especially in the context of sexual selection (first paragraph of the discussion).

I wrote this before reading the next paragraph: "About divergence, I wonder (since I'm not basing this idea on thorough knowledge) whether the interaction of random processes with a somewhat polygenic determinism may have unexpected consequences. I'm not saying that you should consider this idea in the text. I can't fully wrap my head around it. I'm just putting this here, in case you know something about this that I don't." It is therefore only logical that I say that I appreciated your discussion of the departure from the infinitesimal model.

Reply: Yes, very difficult to understand adaptive divergence in the context of a polygenic model. To our understanding there are very few, if any, analytical models particularly considering multiple traits (for single trait, see Matuszewski et al. 2015 Genetics). For single traits, the accepted intuition is that large effect loci will respond first to selection if they are at intermediate frequencies (which will not be the case with sudden "peak shifts") but that stabilizing selection and drift will result in the loss of the best genotypes. Hill-Robertson effects (selective interference) will further contribute to genetic differentiation and phenotypic divergence.

LI 591-593 These are indeed the most described mechanisms but many others are at play and were described in papers.

Reply: We have reworded this sentence and refer to your 2018 review, and Estes and Arnold AmNat 2007, where a discussion of other mechanisms is done (last paragraph discussion).

The conclusion, as well as the whole paper, is very clear. I find that the conclusion fails to present the fundamental implication of this brilliant study. It is necessary to clarify why and how the research presented here contributes to move the domain and the theory forward. In terms of natural history, it seems to me that the situation described here is potentially existing in many other wild populations that are described as being in stasis because they are at mutation-selection equilibrium post adaptation. I have my idea but I'd like you to state why it is important to show that genetic divergence can occur while there is a stasis.

Reply: We have added a few sentences about the importance of our study to short-term and long-term evolution in natural populations (last paragraph discussion).

Review by **Pedro Simões**

In this study the authors address the stability of G-matrix during a 100-generation window of phenotypic stasis. They found that both genetic drift and (stabilizing) selection shape the evolution of G matrices. The evolution of G-matrices is a quite relevant topic, as it allows insight on the processes shaping phenotypic evolution and can provide a framework to help predicting evolutionary responses of polygenic traits. My general assessment is that the study is interesting and worthy of publication although I think there are some important issues that need to be addressed particularly concerning the interpretation and discussion of some of the results obtained. I first highlight these below and then move on to some additional comments.

Reply: We appreciate your comments and suggestions.

- I have some concerns about the marked differences between the G-matrix of A6140 and those of the derived CA(1-3) populations at generations 50 and 100 (as seen in Figure 2b and c for example). These substantial changes look a bit counterintuitive considering that phenotypic stasis was already prevalent by the start of the focal stage and had in fact already been persistent for about 140 generations since the intercross of the 16 inbred founders (Figure 1). There are some (methodological) issues that could at least in part explain these differences namely i) the occurrence of (slight) temporal changes in the environmental settings of the experimental system or ii) sampling effects during the derivation of the several replicates from A6140, which could explain the substantial loss of genetic variation relative to the ancestral state. Can you rule out these possible explanations? Concerning this last point, additional information should be provided on the derivation of the 6 replicates under study (lines 100-101).

Reply: Thank you for pointing out this problem. It is indeed suspicious that at generation 50 and 100 we have much less genetic variance, though we note that this pattern is expected just by drift was shown in Figure 5 where all observed genetic variance is plotted along the selection dimensions.

We cannot rule out that methodological issues explain the decrease in genetic variance. Temporal changes in laboratory culture are possible, in particular because pre- and post-generation 140 culture was done at two different lab locations, even if all materials and equipment used were similar. We suspect, however, that changes in the phenotypic means of the outbred populations would have been noted, which was not the case.

Regarding founder effects, they are highly unlikely because each replicate population was founded from at least 10^3 individuals from the same, very large, pool of revived A6140 samples. 10^3 is the expected effective population size in our demographic and environmental conditions (see Chelo and Teotónio Evolution 2013). We now provide more details about how the replicate populations were derived from the A6140 population (second paragraph methods).

The inbred lines from the CA1-3 populations of generation 50 were phenotyped by a different experimenter than the inbred lines from the other generations. We reviewed and re-analyzed the original raw data and continue to find that at this generation there is a large reduction in genetic variance. The phenotypic variance is also greatly reduced, and is accompanied by a reduction in the mean body size (among the inbred lines). We suspect that the height of the camera during recording was different from that when generation 0 (A6140)

and generation 100 inbred lines were phenotyped. We could not find a way to properly scale all generations together and thus removed generation 50 from all analysis of divergence (new Figures 3 and 5). We have maintained, however, the analysis of generation 50 inbred lines, separately from the other generations, to argue for differentiation between replicate populations (supplementary Figure S10).

We now present analysis of divergence between generation 0 and generation 100 inbred lines phenotyped by the same experimenter at the same time. This new analysis includes data not previously presented, although it is less powerful because of a reduced number of lines. It confirms the reduction in genetic variance at generation 100 (Figure S9).

• I think the Discussion is a bit too lengthy and an excessive focus on some aspects that were not directly tested, or are somewhat questionable:

1) Directional Selection is mentioned in several instances in the discussion as playing an important role on the patterns obtained (particularly in lines 472-474 and lines 596-598). I do not think such references are warranted as there is no evidence of Directional Selection in the system (and trait used) aside from the initial stage of the founder lines (based on Figure 1). Furthermore, the occurrence of directional selection was not directly tested at any stage and is therefore beyond the scope of this paper.

Reply: We agree that evidence for directional selection is marginal and partly based on the phenotypic evolution from founders to the ancestor hybrid population. We did, however, explicitly test for directional selection, using the Lande-Arnold regression methods, but could not find much evidence for it (line 304 in old draft, line 353 in revised version).

We don't agree that the occurrence of directional selection is beyond the scope of our work. We would argue that its occurrence is crucial to infer effective stabilizing selection, where during experimental evolution populations meander by drift in a "plateau" and are bounce back when reaching its "edges" because of directional selection (first and last paragraph of discussion).

2) The pattern of increase variance of y5 and y6 axis of selection (Figure 5) looks of very modest magnitude and particularly questionable for y5. Reducing the extended discussion on the topic (lines 496-548) would help increase focus on the most relevant findings.

Reply: We agree. We have eliminated this discussion.

I think the focus of the manuscript should remain on its more robust findings namely 1) the occurrence of phenotypic stasis with the underlying evolution of G-matrices being governed by drift and, to a lesser extent, stabilizing selection; 2) divergence relative to the ancestral state and a transient differentiation among replicates during the following period. Additionally, a final consideration framing the importance of this study within the context of evolution under environmental change would be relevant. For instance, how would the observed occurrence of phenotypic stasis with underlying reduction of genetic variance impinge on the ability of populations to cope with sudden environmental shifts??

Reply: We re-worked the discussion and tried to focus it, as suggested. We have added a brief conclusion about the implications of our findings to natural populations and theory (last paragraph).

Additional comments:

The introduction does a good job in explaining the aim of the study and framing it within the existing literature. I miss some information on the relevance of locomotion behaviour in this system that could explain why its choice as a trait of interest.

Reply: We have added more information about the significance of genetic variance in, and selection on, locomotion behavior in our environmental conditions (last paragraph introduction).

- Line 15. "Repeatable" in what sense? do you mean maintained?

Reply: Thank you, we have replaced "repeatable" by "maintained".

- Lines 275. "This decomposition of" ... G ?

Reply: Corrected.

- Lines 323-324. Weird phrasing, do you mean "compare empirical posterior distributions with each other"?

Reply: Rephrased.

- Lines 396-398. I don't think you can say this, there appears to be an overlap of observed values with the null distribution by generation 100 (although not by generation 50). Also, looking at Figure 2B it appears that values of the null distribution are consistently lower in generation 50 than at generation 100. Wouldn't it be expectable to be the other way around (lower expected variance at a more advanced generation)? Do you have any explanation for this pattern?

Reply: We no longer show generation 50 results with the other generations, please see reply above. Regarding differences between estimated genetic variances with the null, in all cases the mode of the empirical distribution falls outside the null distribution of posterior means (please see methods section, "inference of effects").

Lines 451-453. not very clear, please quantify the number of replicates with non-significant overlap. to me it looks like there are few replicates that do not overlap with the null model particularly for y5...

Reply: We've eliminated this discussion.

Lines 572—576. How do you reconcile this reasoning with the observation of a “transient” differentiation between replicates (with apparently lower replicate differentiation by generation 100 than by generation 50)?

Reply: Thank you. We have added a sentence about the loss of genetic variance, which can explain why transient differentiation might not be of future consequence. We have no way, however, of testing this hypothesis without knowing the relevant QTL and further evolution experiments in new environments.

Figure 2C. “V” missing in the y axis label.

Reply: Corrected.

Figure 3A. “Replace “bars, 95%...” by “black bars, 95%...”

Reply: Corrected.

Figure 5. There is no Y legend.

Reply: Sorry about this, corrected.

Tables 1 & 2. The coding for the transition stages is missing here. This coding should be explained somewhere in the text to avoid redundant information in the different Tables and Figures.

Reply: We have added an explanation in the Table 1 legend. We have also added a new panel in Figure 1 defining transitions rates.