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.

- 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).

- 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 in 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.
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.

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??

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.

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

- Lines 275. “This decomposition of” … **G** ?

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

- 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?

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...

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)?

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

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

Figure 5. There is no Y legend.

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.