RESPONSE LETTER TO THE COMMENTS OF THE RECOMMENDER AND REVIEWERS

Resubmission MS Title:

Unraveling genetic load dynamics during biological invasion: insights from two invasive insect species

Available at: https://doi.org/10.1101/2024.09.02.610743

Authors:

Eric Lombaert, Aurélie Blin, Barbara Porro, Thomas Guillemaud, Julio S. Bernal, Gary Chang, Natalia Kirichenko, Thomas W. Sappington, Stefan Toepfer, Emeline Deleury

Submitted to: PCI Evolutionary Biology

Dear Recommender,

We are deeply sorry for the long delay in response. We found the reviewer's comments and criticisms extremely helpful and cogent. We have incorporated most of their suggestions and feel our manuscript is much improved. We hope that you will find the MS appropriate for recommendation in PCI Evolutionary Biology.

Please find below our detailed responses (in italic characters) to your comments as well as to those of the two reviewers (in bold characters).

Best wishes,

Eric Lombaert on behalf of all authors

Recommender's comments:

Decision for round #1: ***Revision needed***

Dear authors,

The manuscript entitled "Unraveling genetic load dynamics during biological invasion: insights from two invasive insect species" by Lombaert et al has been evaluated by two referees. They agree that this study is of great interest, both from a theoretical and applied standpoint. However, the two reviewers also raised some concerns that I am also sharing. In particular I suggest the reviewers follow the suggestions of weighting each SNP as a function of the probability of the ancestral state, to reduce uncertainty around polarization, to better assess or present the robustness of the RXY according to the inferred demographic history and to provide results for synonymous SNP. Related to demographic history, more details would be welcome: in the supplementary ABC-RF analyses, the different scenarios do not allow for any gene flow, which is surprising. More importantly, no estimates of the robustness of the RF classifier is provided, nor any estimate of the classification error, whereas the latter is readily available in the ABC-RF package. These details are much needed. Second, estimates of the levels of inbreeding (e.g. FROH) would be useful, as suggested by reviewer 2. In addition, the reviewer #2 also asks for founder population size estimates, I agree that it would be great if such estimates are available.

Most importantly, reviewer #2 asks to better present the hypothesis/goal being tested (in the abstract/introduction) and whether this can help or not assess the load's role during biological invasion.

Based on reviewer #2 comments extensive work on the distinction between purging by drift versus purging by inbreeding is needed. A careful reading of the manuscript to ensure proper terminology is used would be welcome.

Finally, I was surprised to see a mention of the expansion load concept in the discussion but to see nothing in the introduction: is it relevant to the concept of biological invasion? If yes, how ? how does the expansion load hypothesis compare with (contradicts?) the purging hypothesis ? At least, either provide more details or remove references to this concept.

The statement of a "reduced efficacy of selection" (Line 370), based on the sole observation of HeN/HeS ratio seems a bit hasty. In addition, if the selection efficacy is reduced, shouldn't we expect a stronger load, at least at the early stage of the invasion process, before any purge could take place?

In summary I strongly suggest the authors to carefully consider all the reviewers' recommendations.

I would be happy to reconsider this manuscript for a recommendation in PCI Evol Biol if the authors can address point by point the reviewer's comments.

We thank you for your constructive and detailed feedback on our manuscript. We have carefully addressed your suggestions and those of the two reviewers, implementing a substantial revision to improve the manuscript. Below, we provide responses to your questions and remarks, some of which are then elaborated in greater detail in our responses to the reviewers.

1. *SNP weighting and polarization:*

   We have clarified our polarization methodology and added supplementary analyses to evaluate the robustness of our results under different probability thresholds (Figure S2 and Figure S3). While acknowledging the limitations of using arbitrary thresholds, we demonstrated that their impact on the retained SNPs was marginal. We also explained why we did not use a weighting approach, as it requires independent methodological validation. Regarding synonymous SNPs, our dataset is derived from exome capture and uses synonymous mutations as neutral references in the Rxy calculation, which prevents the possibility of separately computing Rxy for synonymous mutations. Please see our response to reviewer #1 for more detailed explanations.

2. *Gene flow in ABC analyses:*

   Our simulations do not account for gene flow due to a technical limitation of the tool employed, a point we now highlight in the supplementary information. However, this limitation does not undermine the validity of our findings. First, we are analyzing large-scale invasions involving spatially distant populations, where gene flow is unlikely. Second, in the case of the North-East American population of Harmonia axyridis, we tested scenarios with and without admixture between source populations under the assumption of no gene flow. If gene flow were present but not accounted for, one might expect the admixture scenario to be favored even in the absence of true admixture. The fact that our results robustly reject the admixture scenario despite this limitation strengthens our confidence in the conclusion that gene flow did not occur. Moreover, while accounting for gene flow might refine our analyses, this is not the central focus of our study, and transitioning to a different simulation framework would be logistically too demanding in the context of this study. DIYABC-RF offers several advantages over other tools that make it particularly suitable for our work, including the ability to optimally simulate Pool-Seq data and the use of a large set of summary statistics encompassing intra-population metrics as well as pairwise and higher-order statistics.

3. *RF analysis quality metrics:*

   We enhanced the presentation of the model choice ABC-RF results by highlighting key quality metrics, such as prior error rate (i.e. mean classification error), and robustness evaluation (see Appendix S1).

4. *Estimates of inbreeding levels:*

While we agree that inbreeding metrics such as ROH would be useful, these measures are not accessible with Pool-Seq data, which lacks individual-level genotypes. To address this limitation, we have included a detailed explanation in the revised manuscript, highlighting the methodological constraints and emphasizing that our use of allele frequency-based metrics serves as a proxy for exploring genetic load. Please see our response to reviewer #2 for a more detailed answer.

5. *Bottleneck intensity estimation:*

   We incorporated estimates of bottleneck intensities for all invasive populations, based on a consistent framework using ABC-RF. These results are detailed in Appendix S1 and referenced in the Methods, Results, and Discussion sections. See below for our more detailed response to reviewer #2.

6. *Clarification of objectives and terminology:*

   We have revised the manuscript to address the need for a clearer presentation of the hypothesis and the related concepts. Please, see below our more detailed responses replies to reviewer #2.

7. *Concept of "expansion load":*

   The interactions between the concepts of "expansion load" and purging of genetic load during biological invasions, which are associated with different stages of the invasion process (introduction/establishment versus expansion), indeed represent a fascinating area of research. However, to our knowledge, these interactions have not yet been thoroughly explored in the literature. Given that our data do not provide relevant insights into these questions and that the "expansion load" hypothesis falls outside the primary scope of our study, we have decided to remove any references to this concept from the manuscript.

8. *Concept of "reduced efficacy of selection":*

   We appreciate the comment regarding the use of "reduced efficacy of selection". We have revised the phrasing to better reflect the nuance of the data. While a higher HeN/HeS ratio in invasive populations aligns with expectations for populations experiencing a reduction in the efficacy of selection, we recognize that this alone does not provide conclusive evidence for such a process. We have, therefore, added caution to our interpretation, addressing the concerns while preserving the integrity of the analysis. These changes can be found in the following sections of the manuscript:

   Lines 319-322: "At the intra-population level, we used in-house R scripts to compute the synonymous expected heterozygosity HeS, as a measure of diversity, and the ratio of non-synonymous to synonymous expected heterozygosity HeN/HeS, which provides an indirect measure of the efficacy of selection."

Lines 391-395: "Overall, the differences between native and invasive populations were more pronounced in DVV, with invasive ranges showing lower diversities and higher HeN/HeS ratios. This could suggest a reduced efficacy of selection, although other factors may also contribute to the observed patterns."

In conclusion, we have addressed each comment from the reviewers (see below), clarified the limitations of our methods, and integrated additional analyses to strengthen the robustness of our findings. We hope these revisions meet your expectations and demonstrate the quality and relevance of our manuscript.

Anonymous reviewer #1's comments

In this article, the authors use a population genomic approach to study the effect of invasion history on the mutation load in two insect species, the western corn rootworm (Diabrotica virgifera virgifera, DVV hereafter and in the manuscript) and the harlequin ladybird (Harmonia axyridis, HA hereafter and in the manuscript). In particular, the aim is to test whether purging of the load has occurred during invasion because of episodes of bottlenecks. To do so, they sampled native and invasive populations of each species plus two/three outgroups to polarize mutations and use exome capture and pool sequencing. They annotate the deleteriousness of non-synonymous SNPs and compare patterns of genetic diversity and the load between native and invasive populations. The two species show contrasted patterns: they found a clear reduction in genetic diversity in DDV but no clear evidence of purging whereas in HA the reduction in diversity is weak and there are signs of increased load.

The manuscript addresses an important question, for both a theoretical point of view in evolutionary biology and a more applied one to understand the dynamics of invasive species. The experimental design is well thought, in particular using two or three outgroups allow polarizing alleles confidently, which is a key part of the analysis. Results are cautiously discussed and not over-interpreted. However, I have a few reserves about the analyses. I don't think it should affect the main patterns but I think some points should be clarified and I also suggest some re-analyses.

<u>Polarization of SNPs</u>

- Using two or three outgroups and the est-SFS tool is a robust approach to polarized SNPs. However, I don't understand why the authors repeated the polarization procedure for each native population separately instead of doing a single polarization. This constrained them to combine the results making some arbitrary choices about the thresholds.

We chose to repeat the polarization procedure for each native population (i.e. two per species) to enhance the robustness of our analyses. Using multiple native populations and replicating the analysis, we aim to minimize biases that could arise if a single population—potentially atypical—were used for polarization. While this approach is more conservative and excludes some SNPs, we believe this trade-off is essential to ensure the reliability of the polarized SNPs. Indeed, only one allele can be considered ancestral for each SNP; therefore, it needs to be confirmed in more than one population.

Regarding the combination of results, we applied clear and systematic criteria, as described in the manuscript, to harmonize the polarization across populations and further reduce the risk of

incorrect polarization. The dual constraint for a SNP – being polarized twice in the same direction and achieving an est-sfs probability above 0.75 – provides, in our view, a robust way to minimize errors in polarization.

We hope this clarification addresses the reviewer's concerns and would be happy to provide additional details if required.

- Even with a single polarization, est-SFS gives a probability of being ancestral. However, the threshold of 0.75 is not very stringent. In addition, using a threshold can sometimes biased the results. One solution would be to test different thresholds to assess the robustness of the results. Another, and I think better, approach would be to weight the SNPs as a function of the probability of ancestral state. For example, a mutation in frequency x with a probability p to be derived should count as p for frequency x and (1 – p) for frequency 1 – x. Or equivalently, the ancestral state could be randomly sampled with probability p / 1 – p.

As for the choice of the 0.75 probability threshold, we agree that it is somewhat arbitrary. However, there is no consensus in the literature on this matter, and all thresholds are inherently arbitrary. For instance, James et al. (2023, Mol. Biol. Evol., 40:1-16) used 0.6, while Chen et al. (2024, Sci. Adv., 10:26-31) applied 0.9. Importantly, we found that the threshold value had only a marginal impact on the number of polarized SNPs retained in our study. To support this, we computed the number of retained polarized SNPs using various probability thresholds, and the variation was limited (see Figure A below).

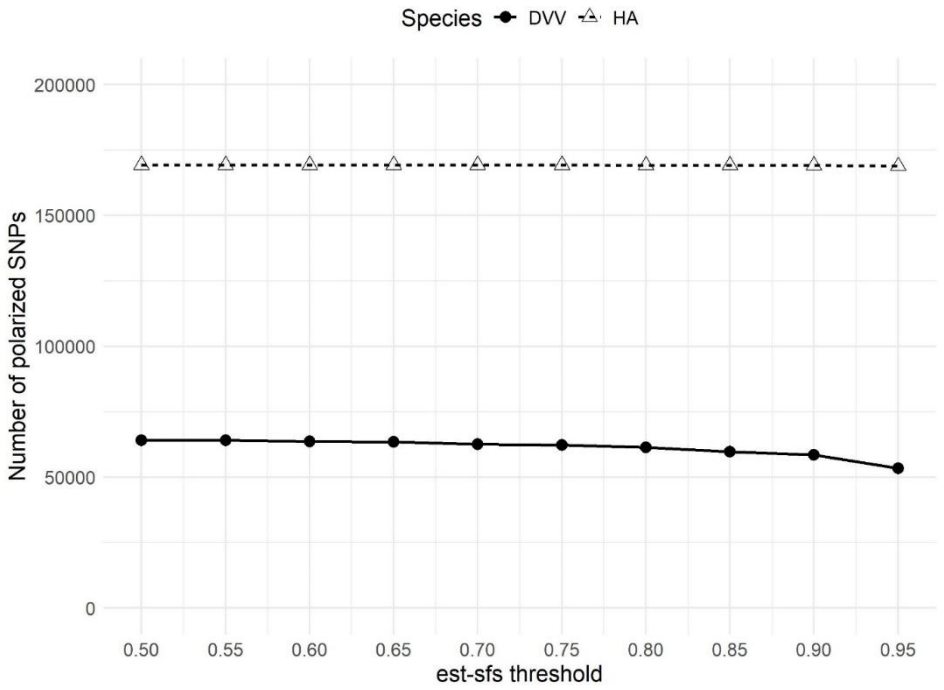

Figure A: Number of polarized SNPs as a function of the chosen est-sfs probability threshold.

What we observe in Figure A aligns with the distribution of est-sfs probabilies for both species, and across all target populations, as shown in Figure B. Notably, the vast majority of probabilities are close to 1, indicating that our polarization procedure is already robust and further confirming that that the choice of threshold has a minimal impact on the results.

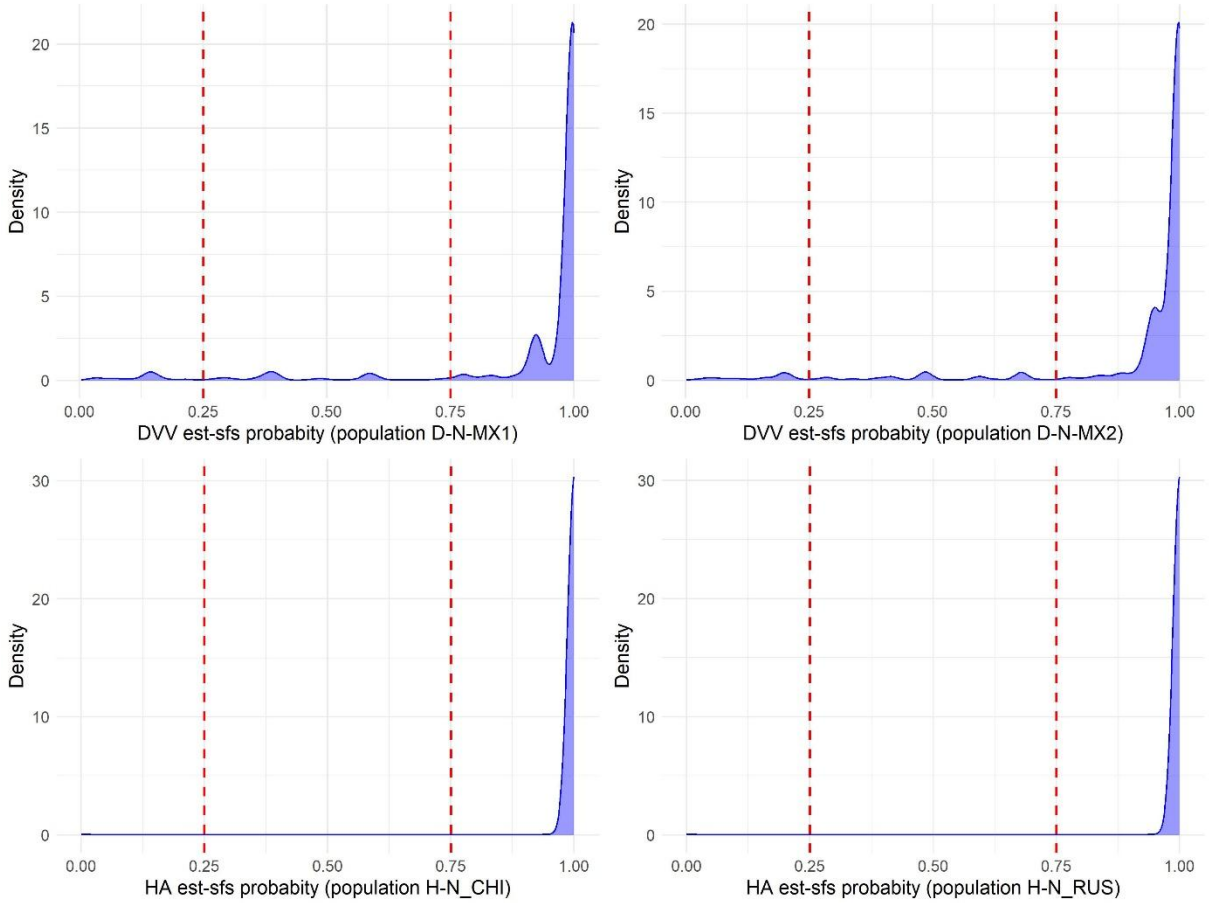

Figure B: est-sfs probability distribution for each species and each target population. The vertical dashed red lines illustrate the threshold chosen in the present study.

Regarding the alternative approach suggested by the reviewer, we appreciate the suggestion of weighting SNPs by their ancestral state probabilities or randomly sampling the ancestral state based on probability ratios. To our knowledge, however, this method has not been described or validated in the literature, and its application would likely require a dedicated study to fully assess its effectiveness. Such a study, that could first be based on simulations, could significantly advance the methodological framework for allele polarization in this type of analysis. Nonetheless, we believe this question deserves its own focused investigation, as it lies beyond the scope of the present study. Additionally, in the context of our dataset, the distribution of est-sfs probabilities (Figure B) indicates that the proposed alternative method would have a limited effect on the results, given that the vast majority of probabilities are close to 1. Finally, we believe that our current method, based on clear thresholds and

replication across multiple populations, already provides a robust and conservative framework for polarization for this study.

The results presented to the reviewer in this response have been included as Supporting Information to provide readers with additional context and to help them better understand the rationale behind our decisions (Lines 295-296, Figure S2, and Figure S3).

Rxy analysis

- For DDV, the Rxy results differ among populations (even if only one case is significant). However, the reference population varies. This is based on the history of invasion but what is the robustness of the scenario?

The invasion scenarios have been thoroughly tested and validated across multiple studies using different datasets and statistical methods for DVV (Miller et al. 2005, Science, 310:992; Ciosi et al. 2008, Mol. Ecol., 17:3614-3627; Lombaert et al. 2018, Biol. Inv., 20:665-677) and HA (Lombaert et al. 2010, PLoS One, 5:e9743; Lombaert et al. 2011, Mol. Ecol., 20:4654-4670; Lombaert et al. 2014, Mol. Ecol., 23:5979-5997). Robustness assessments in these studies include high posterior probabilities and strong goodness-of-fit measures from posterior model checking in ABC analyses, among other metrics, providing substantial support for the reliability of the proposed invasion histories. As these studies provide a strong foundation, we believe that a detailed restatement of their findings would be redundant in the current manuscript. However, it should be noted that, in this manuscript, we have explicitly addressed the case of the eastern North American population of HA (Appendix S1), which has been identified in the literature as the population with the most uncertain origin (e.g., Lombaert et al. 2011, Mol. Ecol., 20:4654-4670).

We have included a general clarification in the Methods section (Line 138-142): "Both species are highly successful invaders with extensive invasive ranges (Gray et al. 2009; Roy et al. 2016). Additionally, their invasion routes are well-documented and supported by robust analyses using diverse datasets and methodologies (Miller et al. 2005; Ciosi et al. 2008; Lombaert et al. 2010, 2011, 2014, 2018)."

- As a control it could be useful to also give the results for synonymous SNPs

In some studies, neutral sites in the genome used for Rxy computation is represented by SNPs located in intergenic regions, which is feasible, for example, when data are derived from whole-genome sequencing. In these cases, Rxy values for synonymous mutations are often calculated as an informative control, since such mutations are expected to behave nearly neutrally and thus yield Rxy values close to 1. However, our data is derived from an exome capture protocol, which predominantly includes SNPs

from coding regions, specifically synonymous and non-synonymous mutations. Consequently, in our study, synonymous mutations are used in the computation of Rxy to represent neutral sites (referred to as S in Appendix S2) to account for genetic drift, consistent with approaches used in other studies (e.g. Do et al. 2015, Nat. Gen., 47:126-131; Narasimhan et al. 2016, Science, 352:474-477; Ochoa & Gibbs 2021, Mol. Eol., 30:5454-5469). Thus, it is not possible to calculate a separate Rxy specifically for synonymous mutations within this dataset. For further details on the Rxy calculation, we refer the reviewer to Lines 335-345 and Appendix S2.

Discussion

- The interpretations of mean allelic frequencies can be partly misleading. The load and inbreeding depression are not linear function of allelic frequencies, so they also depend on the variance (see for example, Bataillon, T., and M. Kirkpatrick. 2000. Inbreeding depression due to mildly deleterious mutations in finite populations: size does matter. Genetics Research 75:75–81). For example, the Hungarian population of DDV shows a signature of purging (Rxy < 1) but on figure 2, it seems that it has more fixed deleterious mutations and fixed mutations generate a load disproportionally larger than those maintained in low frequency (s, selection coefficient, versus u, mutation rate).

We agree with the general concern raised about the complexity of interpreting genetic load from mean allelic frequencies alone. This issue highlights a limitation of our study, which is based on pool-seq data. Pool-seq precludes access to genotypic frequencies and limits our ability to precisely estimate inbreeding depression and explore the non-linear effects of genetic load. That said, we tried to mitigate this limitation by presenting and exploring genetic load using multiple complementary statistics, including the proportion of derived alleles classified by frequencies (Figure 2), the mean derived allele frequency (Figure 3), and the Rxy statistic (Figure 4). We believe that no single measure can fully capture the dynamics of genetic load and its evolution across populations, and therefore, using a variety of metrics provides a more robust understanding.

Regarding the Hungarian DDV population, we would like to clarify that while the LoF (loss-of-function) mutations in this population show a signature of purging (Rxy < 1), the number of fixed deleterious LoF mutations in the Hungarian population is not higher than in other populations (indeed, only 3 out of 902 LoF SNPs were fixed for the derived allele in this population; Figure 2). Although missense mutations are slightly more fixed in this population, they do not show a purging signal in the Rxy values (Rxy not significantly different from 1 for the missense category; see Figure 4).

To take into account your feedback, we have made adjustments in the discussion section (Lines 566-571), adding the following sentence to emphasize the limitation of our study: "One limitation lies in our reliance on allelic frequencies derived from pool-seq data, which precludes access to genotypic

frequencies and therefore limits our ability to precisely estimate inbreeding depression and explore the non-linear effects of genetic load (Bataillon and Kirkpatrick 2000). While we used complementary statistics to mitigate this issue, future studies based on individual-level sequencing would allow for more accurate assessments of genetic variation and its relationship to genetic load."

- The results suggest that HA may have accumulated instead of purged deleterious mutations, which is at variance with the results of Facon et al. (2011) based on direct fitness measures. However, this can also be explained by the effect mentioned above. Alternatively, very strongly deleterious mutations that can contribute to inbreeding depression can be maintained in too low frequencies to be detected in the samples. For example, assuming a mutation rate of 10^-6, a recessive lethal is expect to segregate in frequency sqrt[u] = 1/1000 in a large panmicitic population, which is almost impossible to capture with a sample size of 80.

The issue of very strongly deleterious mutations being maintained at extremely low frequencies is indeed a plausible explanation for this divergence between Facon et al. (2011) and the present study. We have incorporated this point into the discussion, Lines 541-544: "Moreover, very highly deleterious mutations, which are critical contributors to inbreeding depression, are expected to segregate at extremely low frequencies in large populations and may thus escape detection in our sampling due to their rarity."

This addition complements the other hypotheses we discuss, including the potential disproportionate effects of a few key genes on invasion success, the imperfect nature of our categorization and the limitations of our exome capture protocol. We hope this addresses the reviewer's concern and underscores the robustness of our revised interpretation.

<u>Minor comments</u>

- L77-78: This situation often results in a "mutational meltdown". This is a theoretical expectation but to my knowledge there is very few empirical evidence of mutational meltdown in natural conditions. Please some empirical examples or rephrase this sentence: for example, "may result" instead of "often results".

We agree; this was a misuse of language. We have corrected the text as proposed (lines 82-85).

- In the different figures, it would help the reader to more clearly mark the difference between native and invasive populations. There is only the letter N and I in the middle of other letters.

In each of the figures in the manuscript, the names of the native populations are now underlined for ease of reading (see Figure 1, 2, 3 and 4).

Anonymous reviewer #2's comments

Dear Editor and authors,

Thank you for the opportunity to review the manuscript entitled 'Unraveling genetic load dynamics during biological invasion: insights from two invasive insect species'. I had great pleasure reading and reviewing the manuscript and provide detailed comments below.

I hope my comments will be helpful and will contribute to the improvement of the manuscript.

Best Regards.

<u>Editorial questions:</u>

Introduction

Are the research questions/hypotheses/predictions clearly presented? [ ] Yes, [X] No (please explain), [ ] I don't know

>>> As stated in my comments, there should be a clear distinction between reduction in genetic load via drift or purging (i.e. inbreeding facilitating purifying selection). It is also not clear how the authors actually test for purging. Data on inbreeding should also be presented.

Does the introduction build on relevant research in the field? [] Yes, [X] No (please explain), [ ] I don't know

>>> Overall, yes. However, more information on the colonisation history, number of introductions, founder (effective) population sizes should be discussed in the introduction, if available.

Results

Are the results described and interpreted correctly? [] Yes, [X] No (please explain), [ ] I don't know

>>> As stated above and in my comments, the authors seem to think that any loss of deleterious variation is evidence for purging. However, drift can also induce a loss in deleterious variation. The interpretation of the results should thus be clarified.

Discussion

Have the authors appropriately emphasized the strengths and limitations of their study/theory/methods/argument? [ ] Yes, [X] No (please explain), [ ] I don't know

>>> shortcomings related to estimates of inbreeding (e.g. FROH) and a reduced set of genes should be discussed and highlight how to test better for evidence for purging of load. The authors should

also discuss the advantage of using forward simulations (e.g. Slim) to recapitulate and test the invasion history of those 2 species.

In this study, Lombaert et al. quantify genetic diversity and load in two invasive insect species using exome data. They find no strong evidence for purging in one species and an excess in deleterious alleles in the other one, highlighting the contrasted dynamics of load during the invasion process.
Overall, I think it is a very interesting study and it is great to read a paper on the dynamics of load during the invasion process. However, while the manuscript is well written and the data analyses are overall sound, there seems to be some degree of confusion regarding what purging actually is. Some clarification is thus needed on what purging is. Furthermore, I think it should be more clear whether the authors test for purging of load and how it can contribute to the invasion success. Reading the abstract, it seems that they are testing this hypothesis, but in the introduction, they stress that it is not what they are testing. So this should be clarified. In light of these clarifications, the discussion of the results should be updated, making clear whether the pattern observed is the result of purging, drift or a combination of both.
We thank the reviewer for his comments, remarks and suggestions. We respond to each of the "major comments" and "minor comments" below.

<u>Main comments:</u>

First of all, I think that the way the hypothesis to be tested is formulated is not completely correct. The authors refer to the 'so-called purge hypothesis' in the abstract. I would avoid using the 'so-called', especially if no reference to this hypothesis is given. Furthermore, I would avoid referring to 'intriguing' on l. 73 as this hypothesis is not new at all and has been tested before many times, even before the era of genomics.
While our intention was to highlight that this hypothesis is relatively recent in invasion biology compared to conservation and domestication biology, we agree that the phrasing was misleading. Therefore, we have removed both "so-called" (Line 42) and "intriguing" (Line 74) to avoid any confusion. Additionally, we have made modifications to the text to better reflect the existing literature and to present the hypothesis more precisely (see below for details).

Also, on l-73-86. The way purging is referred to is a bit unclear. As it reads now, it suggests that purging is a process that can occur immediately at the start of the invasion and it almost reads as if loss of deleterious variation is always caused by purging. The latter is incorrect and it will take several

generations for inbreeding to increase and facilitate the exposure of deleterious variation in homozygous state and thus to selection.

See definition from Dussex et al. (2023):

Purging: reduction in genetic load by purifying selection operating against recessive deleterious variants exposed in a homozygous state due to inbreeding in small populations, through population fragmentation or under positive assortative mating.

Similarly, this statement is rather vague and incorrect: 'Conversely, bottlenecks may also purge deleterious alleles, thereby increasing the mean fitness of the introduced individuals.' It is not the bottleneck that will purge load, but natural selection, aided by increased inbreeding.

I would thus reframe this paragraph (and abstract) and instead mention how drift and/or purifying selection + inbreeding (i.e. purging) can contribute to the reduction in genetic load. While the two processes are linked and can produce the same effect, they are not the same.

We agree that the explanation of the purging process was unclear and that the distinction between the roles of genetic drift and purifying selection in reducing genetic load required further clarification. To address these issues, we have made substantial revisions to the manuscript. In particular, we have refined the description of purging, aligning it with the definition proposed by Dussex et al. (2023). We have emphasized more clearly how genetic drift and purging are distinct processes in the reduction of genetic load. These changes provide a clearer and more nuanced explanation of the mechanisms influencing genetic load during biological invasions. Specific revisions can be found in the abstract and the introduction section. Here are some examples:

- *Lines 42-44: "Here we investigated the purge hypothesis, which suggests that demographic bottlenecks may facilitate conditions (e.g., increased homozygosity and inbreeding) under which natural selection can purge deleterious mutations, thereby reducing genetic load."*
- *Lines 78-87: "While genetic drift may randomly eliminate some deleterious alleles, thereby reducing part of the genetic load, it also contributes to the transformation of the masked load (i.e., the load which may become express in future generations; Bertorelle et al. 2022) into a realized load (i.e., the load which reduces fitness in the current generation; Bertorelle et al. 2022). This shift may result in a "mutational meltdown" (Lynch et al. 1995; Simberloff 2009), where expression of deleterious mutations increases risk of extinction, ultimately leading to failure of the introduced population to establish. Conversely, bottlenecks may create conditions where increased homozygosity exposes recessive deleterious alleles to purifying selection (Dussex et al. 2023), potentially reducing genetic load and increasing mean fitness of the introduced individuals."*

- *Lines 91-93: "The potential for such purging in introduced populations is particularly interesting for its role in facilitating successful invasions by reducing inbreeding depression and enabling inbred individuals to maintain high fitness levels."*

Related to this point, if one really wants to assess whether purging occurred or not, estimates of inbreeding (FIS) or better based on Runs of Homozygosity (FROH) would be needed. While the latter may not be possible with transcriptome-based exome data, FIS should be estimated here based on synonymous variation. I would also show boxplots with the number of homozygous and heterozygous alleles for each category, after filtering the data for missing sites. This will provide a visually explicit representation of masked and realised load. In addition, it would be to estimate realised and masked load as described in Mathur&DeWoody (2021) - Genetic load has potential in large populations but is realized in small inbred populations.

Our study relies on Pool-Seq data, which inherently limits our ability to compute individual-based metrics such as FIS, FROH, or the numbers of homozygous and heterozygous individuals. Pool-Seq provides population-level allele frequency estimates, but not individual genotypes, precluding these types of analyses. Additionally, as you noted, our use of transcriptome-based exome capture further restricts our ability to investigate Runs of Homozygosity (ROH) because these analyses require genome-wide, high-density genotype data.

Despite these limitations, we addressed the distinction between masked and realized load using allele frequency distributions. Specifically, as described Lines 326-330, Figure 2 categorizes SNPs based on derived allele frequencies, distinguishing between rare alleles (which are more likely to remain heterozygous in the population, contributing to masked load) and common or fixed alleles (which are more likely to be homozygous, contributing to realized load). While this approach does not provide the granularity of individual-based metrics and remains imperfect, it serves as a population-level proxy to explore these dynamics within the constraints of Pool-Seq data.

To make these methodological limitations more explicit, we have emphasized in the manuscript that our measures of masked and realized load are proxies (e.g. Lines 419-425). We have also added text in the final paragraph of the manuscript to highlight that our approach is limited to population-level inferences due to the nature of Pool-Seq data. For example, Lines 566-571: "One limitation lies in our reliance on allelic frequencies derived from pool-seq data, which precludes access to genotypic frequencies and therefore limits our ability to precisely estimate inbreeding depression and explore the non-linear effects of genetic load (Bataillon and Kirkpatrick 2000). While we used complementary statistics to mitigate this issue, future studies based on individual-level sequencing would allow for more accurate assessments of genetic variation and its relationship to genetic load."

We hope this explanation provides clarity and demonstrates how we have worked within the constraints of our dataset to address the dynamics of genetic load in invasive populations.

Secondly, there is not information on the founder population sizes for these two species. Is there information available from previous studies? Do we know if these introductions were serial or a unique event? I think this would be really important to add because it is crucial to understand the dynamics of load.

The invasion routes of both Diabrotica virgifera virgifera *and* Harmonia axyridis *have been extensively studied and are well-documented in the literature using diverse datasets and robust analytical approaches (Miller et al. 2005, Science, 310:992; Ciosi et al. 2008, Mol. Ecol., 17:3614–3627; Lombaert et al. 2010, PLoS One, 5:e9743; Lombaert et al. 2011, Mol. Ecol., 20:4654–4670; Lombaert et al. 2014, Mol. Ecol., 23:5979–5997; Lombaert et al. 2018, Biol. Inv., 20:665–677). These studies collectively confirm that all invasive populations of both species originate from unique introduction events rather than multiple introductions (with the only exception being the eastern North American population of HA, which was specifically explored in Appendix S1). We have added a clarification on this point, Lines 189-193: "None of the selected invasive populations resulted from multiple introductions. The east North American population of HA was previously hypothesized to be an admixture between two populations (Lombaert et al. 2011), but our analysis using ABC (approximate Bayesian computation; Beaumont et al. 2002) with synonymous SNPs from the current dataset indicates that this is unlikely (see Appendix S1 for details)."*

While some previous studies have provided estimates of founder population sizes or related parameters, we chose to obtain results for all populations in this study within a unified framework. To this end, we simulated the final invasion scenarios for both species and estimated bottleneck intensities using Approximate Bayesian Computation Random Forest (ABC-RF; Raynal et al. 2019, Bioinformatics, 35:1720-1728). Bottleneck intensity was quantified as a composite parameter, the ratio of bottleneck duration (DB) to the effective population size of founders (NF). This measure captures the intrinsic link between DB and NF, which jointly influence genetic diversity, and thus provides a more integrative assessment of bottleneck than NF alone. We have added a dedicated section in Appendix S1 that details the ABC methodology and provides the estimated bottleneck intensities for all populations. To ensure clarity, we have revised the manuscript to explicitly reference these estimates and have included updated lines as follows:

- *Lines 323-325: "Additionally, we estimated bottleneck intensities for all invasive populations of both species using ABC analyses (See Appendix S1 for details)."*

- *Lines 383-385: "Overall, the loss of diversity was more pronounced in DVV invasive populations, consistent with our ABC results, which indicate that bottlenecks were generally more intense in DVV than in HA (see Appendix S1)."*
- *Lines 491-494: "DVV exhibits marked genetic differences between native and invasive populations. Consistent with previous studies (Ciosi et al. 2008; Lombaert et al. 2018), invasive populations, including the long-established population in Colorado, display sharp declines in genetic diversity, consistent with the strong bottlenecks inferred from ABC analyses."*
- *Lines 557-560: "For instance, the two species experienced markedly different bottleneck intensities, yet our data do not allow us to establish a direct link between these differences and the observed dynamics of genetic load."*

  We believe that these additions and clarifications address the reviewer's concerns.

<u>Minor comments:</u>

l. 62-63. 'The key factors determining the success of invasive species remain largely hypothetical'. Do they really? I would have thought that at least some of those factors would be well known (e.g. absence of natural competitor). Maybe this statement needs to be rephrased or toned down or a reference should be given to support it.

We have toned down the sentence (Lines 63-64): "The key factors determining the success of invasive species are not yet fully understood". While some hypotheses are better supported than others, none provides a comprehensive understanding of the invasion process, which is likely multifactorial. This issue is further discussed in the following sentence of the manuscript.

l. 65. Is 'simultaneously' needed here?

We have removed the word "simultaneously" (Line 66).

l. 78. I may be good to mention that 'mutational meltdown' also refers to an increase in the expression of deleterious variation which may increase the risk of extinction.

We have revised the sentence to incorporate the reviewer's suggestion (Lines 82-85): "This shift may result in a "mutational meltdown" (Lynch et al. 1995; Simberloff 2009), where expression of deleterious mutations increases risk of extinction, ultimately leading to failure of the introduced population to establish."

l. 94. '…did not experience the inbreeding depression suffered'. Please rephrase such as 'affected by inbreeding depression' or 'showing evidence of inbreeding depression'

As suggested by the reviewer, we have modified the sentence as follow (Lines 99-103): "One notable formal test of this hypothesis was conducted on the invasive Asian ladybird Harmonia axyridis, where measurement of life history traits revealed that invasive populations showed no evidence of the inbreeding depression observed in native ones, suggesting that deleterious alleles were purged during the invasion process (Facon et al. 2011)."

l. 96. I would use 'dynamics of load' instead of 'evolution of load'.

The text has been changed accordingly (Line 103).

l. 99. 'past' instead of 'last'

This has been corrected (Line 106).

l. 101. 'applied to' instead of 'expanded into the fields of'

This has been corrected (Line 108).

l. 105. 'reduction' instead of 'losses'

This has been corrected (Line 113).

l. 107 'High quality genomic resources' instead of 'A good knowledge of the genome'

This has been corrected (Line 114).

l. 107-112. I would suggest to maybe rephrase these few sentences along the lines of 'while WGS used to be prohibitively expensive, it has now become more affordable/routine and progress in bioinformatics analyses have also improved greatly', or something along those lines.

We have changed the text as follows (Lines 114-119): "High quality genomic resources are essential for such studies, which may explain the limited application of these methods to invasive species, which are mostly non-model organisms. However, advances in genome sequencing technologies and bioinformatics have significantly reduced costs and improved accessibility, making these methods increasingly routine and affordable (Bertorelle et al. 2022)."

l. 112. I would rephrase and say either 'test the hypothesis of purging during the process of invasion' or 'examined the dynamics of load in two invasive species…'

We agree that we are not directly testing the hypothesis of purging as a mechanism behind invasion success. Instead, we aim to assess whether purging of deleterious mutations occurred during the invasion process. We have reworded the sentence to better reflect the study's goals, as follow (lines

120-123): "In this study, we examined the dynamics of genetic load during the invasion of two successful insect species, namely the western corn rootworm, Diabrotica virgifera virgifera, and the harlequin ladybird, Harmonia axyridis, by directly measuring/estimating genetic load using genomic data from feral native and invasive populations."

l. 114. 'measuring/estimating load' instead of 'assessing'

The text has been changed accordingly (Line 122).

l. 115. I am not sure I understand this statement: 'Importantly, the purpose of this study did not include testing the instrumental role of purging in invasion success'. Based on the introduction and the previous statement, I would have thought that assessing the role of purging in the success or history of invasion was the main goal of the study. Surely, if load is purged rapidly, this would facilitate the invasive potential (along with other processes, e.g. absence of natural predator, etc…). So maybe it could be rephrased along the lines of 'assessing the potential contribution of purging to invasion success'. If one species shows evidence or purging and the other not, but the 2 are 'equally' successful, maybe purging is not necessarily needed? This will of course depend on the founder population size, but this may provide clues on the role of purging in the invasion success.

We fully agree that purging could theoretically facilitate invasion success by reducing genetic load, and this hypothesis is central to our study. However, our data do not allow us to draw conclusions on whether purging played an instrumental role in invasion success. Our work focuses on measuring the evolution of genetic load during the invasion of the two species. The only outcomes that can emerge from our results are whether the genetic load has been purged or not. We cannot infer from these findings whether the observed evolution of genetic load contributed, positively or negatively, to the success of the invasion. Indeed, evidence of purging does not imply that purging was the reason for invasion success: it could simply be a concomitant process rather than a causal factor. Invasive success is influenced by a wide array of factors, including ecological conditions, phenotypic plasticity, and stochastic processes, which could obscure or outweigh the role of genetic purging.

To rigorously test whether purging plays an instrumental role, even as a contributing factor, in invasion success, a much broader study would be required. This would involve examining multiple species, populations, and introduction events, including both successful and failed invasions. Such data are challenging to obtain, especially for failed introductions, but they are essential to disentangle causality from correlation in this context.

In light of the reviewer's suggestion, we have rephrased the paragraph to clarify that our goal is to assess the occurrence of purging, rather than to establish its causal role in invasion success (lines 123-127): "Importantly, our study was not designed to assess the instrumental role of purging in

invasion success, but rather to investigate its occurrence during these invasions. A broader study of both successful and failed invasions across many species would be required to draw conclusions on purging's role in invasion success."

l. 119-121. Maybe state 'the fate of the genetic load during the invasion process' or similar.

The sentence has been modified as followed (lines 130-131): "Our results offer insights into the fate of the genetic load and provide valuable perspectives on the purge hypothesis in the context of biological invasions."

l. 128 and elsewhere: 'purging of genetic load' instead of 'genetic load purging'

We have modified the wording as suggested by the reviewer throughout the manuscript.

l. 287. Could you give more detail on what ''essential information from the vcf file' refer to?

The paragraph has been changed to include more detailed information (lines 300-308). The first sentence is as follows: "Key information was extracted from the vcf file using the SnpSift program v5.0 (Cingolani et al. 2012a), including SNP coordinates, reference and alternative alleles, allele depths, predicted effects, impact annotations, and protein-level changes."

l. 387 'In all populations studied and for each species'

The text has been changed accordingly (Line 407).

l. 463. Use 'dynamics' instead of 'fate'

The text has been changed accordingly (Line 486).

l. 470 maybe add 'and excess in LoF variants, albeit non-significant'? This may apply to l. 476-478 as well.

The paragraph has been changed as follows (lines 500-502): "In North America, despite a strong initial loss of genetic diversity, invasive populations showed no significant relative deficit or excess of deleterious mutations compared to their source, although a trend toward an excess of loss-of-function variants was observed."

l. 478 'changes' instead of 'evolution'

The sentence has been removed entirely following a comment made by reviewer #1.

l. 486-468. 'This suggests that the demographic and selective constraints in this population were effective at purging highly deleterious mutations but perhaps not moderately deleterious ones.' This goes back to main comment. This should be clarified and the distinction between drift and purging (selection) should be clear. If you cannot show that purging actually facilitated the reduction of load, you can state that this reduction was facilitated by a combination of drift and purifying selection. This is also why some measure of inbreeding is important to add.

We agree that drift and purifying selection work in tandem to shape genetic load in invasive populations, and our data do not allow us to disentangle their respective contributions. The sentence has been revised to reflect this distinction more explicitly (lines 506–509): "This suggests that the demographic and selective constraints in this population were effective at reducing highly deleterious mutations through a combination of genetic drift and purifying selection, while moderately deleterious mutations have been less affected".

While inbreeding measures, such as FIS or metrics based on runs of homozygosity (ROH), could provide additional insights, our study relies on Pool-Seq data, which precludes the calculation of individual-level metrics (see above). Instead, we used allele frequency distributions to examine the dynamics of genetic load, focusing on proxies for masked and realized load. This limitation is now acknowledged in the manuscript (lines 569-573), where we describe the methodological constraints and highlight the need for individual-level data to refine these analyses in future studies.

L. 490. Replace 'significant' with 'severe'

The text has been changed accordingly (Line 511).

l. 494. Again, it is not clear whether purging or drift facilitated this loss of deleterious variation.

The text has been changed as follows (lines 513-516): "Finally, the invasive northwestern Italian population, despite originating from an independent introduction (Miller et al. 2005), showed patterns similar to that of the Central-Southeastern European population. However, it did not exhibit significant signals of deficit or excess of deleterious alleles, regardless of the mutation severity considered."

l. 504-507. This would give more support for drift leading to fixation of deleterious alleles and should be mentioned here.

The end of the paragraph has been modified as follows (lines 525-529): "This observation aligns with a significant trend, although subtle, towards a relative excess of deleterious mutations in invasive populations compared to their native source, regardless of putative fitness impact. This tendency to fix alleles contributing to genetic load, likely driven by drift, is significant across all populations and severity

categories, except for the highly deleterious mutations in western North America (sampled in Washington).”

l. 535. ‘our study is one of the first in the context of biological invasions’ . Could the authors cite a few of those studies? What have they shown? Have they assessed the role of purging on the invasion success?

To our knowledge, our study is the first to apply population genomics methods to specifically investigate the evolution of genetic load in the context of biological invasions. We have revised the sentence as follows (lines 561-564): “While population genomics approaches to assess the evolution of genetic load have become increasingly popular in conservation and domestication biology (Moyers et al. 2018; Bertorelle et al. 2022), our study is, to our knowledge, the first to use these methods to specifically investigate the evolution of genetic load in the context of biological invasions.”