Dear Sebastian Ernesto Ramos-Onsins,

Your article, entitled **Detection of Domestication Signals through the Analysis of the Full Distribution of Fitness Effects using Simulations**

The referees' comments and the recommender's decision are shown below. As you can see,the recommender found your article very interesting but suggests certain revisions. We shall, in principle, be happy to recommend your article as soon as it has been revised in response to the points raised by the referees. When revising your article, we remind you that your article must contain the following sections (see our Guide for Authors in the Help section of the PCIEvolBiol website):

1) **Data, script and code availability (if applicable)**
- Data, statistical scripts, command lines and simulation code must be made available to readers. They should either be included in the article or deposited **in** an open repository such as Zenodo **with a DOI**. A perennial URL can be provided if no DOI is available; please note that GitHub URL are not perennial.
- If deposited in an open repository, a reference to **Data, statistical scripts, command lines and simulation code**, with a DOI or a perennial URL, must be provided in the reference list and in the "Data, script and code availability" section
- The "Data, script and code availability" section must clearly indicate **where and how** data can be accessed.
- Wherever possible, data, scripts and code should be provided in machine-readable formats. Avoid PDFs other than for textual supplementary information.
- Metadata should accompany the data, to make the data understandable and reusable by the reader.

2) **Supplementary information (if applicable)**
- Supplementary information (text, tables, figures, videos, etc.) can be referred to in the article. It must be available in an open repository (such as Zenodo, Dryad, OSF, Figshare, Morphobank, Morphosource, Github, MorphoMuseuM, Phenome10k, etc. or any institutional repository, etc...) with a DOI. A perennial URL can be provided if no DOI is available.
- A reference to the supplementary information, with a DOI or a perennial URL, must be provided in the reference list and in the "Supplementary information" section.
- List all documents attached to the manuscript as Supplementary Information in the "Supplementary Information" section.

3) **Funding (mandatory)**
- All sources of funding must be listed in a separate "Funding section". The absence of funding must be clearly indicated in this section.

4) **Conflict of interest disclosure (mandatory)**
- Authors should declare any potential non-financial conflict of interest (financial conflicts of interest are forbidden, see the PCI code of conduct).
- In the absence of competing interests, the authors should add the following sentence to the "Conflict of interest disclosure" section: "The authors declare they have no conflict of interest relating to the content of this article." If appropriate, this disclosure may be completed by a sentence indicating that some of the authors are PCIrecommenders: "XXX is a recommender for PCI XX."

5) **Materials and methods (mandatory)**
- Details of experimental procedures and quantitative analyses must be made **fully available** to readers, in the text, as appendices, or as Supplementary Information deposited in an open repository, such as Zenodo, Dryad or institutional repositories with a DOI.
- For specimen-based studies, **complete repository information** should be provided and institutional abbreviations should be listed in a dedicated subsection (if applicable). Specimens on which conclusions are based **must be deposited in an accessible and permanent repository**.

When your revised article is ready, please:

1) Upload the new version of your manuscript onto your favorite open archive and **wait until it appears online**;

2) Follow this link https://evolbiol.peercommunityin.org/user/my_articles or logging onto the PCIEvolBiol website and go to 'For Contributors -> Your submitted preprints' in the top menu and **click on the blue ‘VIEW/EDIT' button at the right end of the line** referring to the preprint in question.

3) Click on the black ‘EDIT YOUR ARTICLE DATA’ button (mandatory step). You can then edit the title, authors, DOI, abstract, keywords, disciplines, and DOI/URL of data, scripts and code. Do not forget to save your modifications by clicking on the green button.

4) Click on the blue ‘EDIT YOUR REPLY TO THE RECOMMENDER’ button (mandatory step). You could then write or paste your text, upload your reply as a PDF file, and upload a document with the modifications marked in TrackChange mode. If you are submitting the final formatted version ready to be recommended, you should only add a sentence indicating that you posted the final version on the preprint server. Do not forget to **save your modifications by clicking on the green button**.

5) Click on the green ‘SEND RESUBMISSION’ button. This will result in your submission being sent to the recommender. Once the recommender has read the revised version, they may decide to recommend it directly, in which case the editorial correspondence (reviews, recommender’s decisions, authors’ replies) and a recommendation text will be published by PCIEvolBiol under the license CC-BY.

Alternatively, other rounds of reviews may be needed before the recommender reaches a favorable conclusion. They may also reject your article, in which case the reviews and decision will be sent to you, but they will not be published or publicly released by PCIEvolBiol. They will be safely stored in our database, to which only the Managing Board has access. You will be notified by e-mail at each stage in the procedure.

We thank you in advance for submitting your revised version.

Yours sincerely,

The Managing Board of PCIEvolBiol

Revision round #1

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

Revisions needed

Dear authors,

Thank you for your submission. Your study has been evaluated by two experts in the field. Both of them find it of relevance and importance, as your results provide a good contribution to the field. The focus on domestication has also been appreciated.

However, reviewers point to several concerns with your submission. Both indicate a lack of details in the description of the methods, especially with the new approach to infer demography and selection. The code provided on github is not sufficiently documented.

In this new version, we have described in higher detail the methodology used (see details in the review's response) and we have provided the necessary documentation of the code. We also have uploaded the code into a permanent repository at Zenodo (https://zenodo.org/records/14277802, see details in the review's response).

There are also some concerns over the interpretation of some results, where the reviewers appear to be less optimistic than you on the performance of your inferences. I also find Figure 3 to be too cluttered.

We have modified the interpretation of a number of results, as suggested by the reviewers. We have modified the Figures to give a clearer and more understandable message (see details in the review's response). In the new version we have included the Figures at the end of the manuscript to preserve the quality resolution (we noted that the resolution of the figures is modified when is included in MsWord).

While the study has merit, I encourage the authors to carefully consider all the points raised by the reviewers to clarify your text accordingly.

Matteo
by ***Matteo Fumagalli****, 25 Jun 2024 17:28*
Manuscript: 1204587247 **https://doi.org/10.1101/2022.08.24.505198**
version: 3
Review by anonymous reviewer 1, 15 Jun 2024 17:03

Title and abstract
Does the title clearly reflect the content of the article? Yes
Does the abstract present the main findings of the study? Yes

Introduction
Are the research questions/hypotheses/predictions clearly presented? Yes

Does the introduction build on relevant research in the field?Yes, but it lacks some citations. I have added those to my review.

Materials and methods
Are the methods and analyses sufficiently detailed to allow replication by other researchers? No (please explain) The methods are not presented in detail. I have specified that in my review.
Are the methods and statistical analyses appropriate and well described? No (please explain)

Results
In the case of negative results, is there a statistical power analysis (or an adequate Bayesian analysis or equivalence testing)? I don't know
Are the results described and interpreted correctly? Yes

Discussion
Have the authors appropriately emphasized the strengths and limitations of their study/theory/methods/argument? Yes
Are the conclusions adequately supported by the results (without overstating the implications of the findings)? Yes

In this manuscript, the authors model a very realistic population undergoing domestication. They model an ancestral wild population, a strong bottleneck, and changes in selection pressure that are likely to be experienced by the domesticated population. The authors ask if in such a model, when simulated realistically using forward simulations, it is possible to infer the population history and the full DFE. While population history is not recovered accurately, the authors suggest that the deleterious DFE is estimated accurately. The authors have also extended the method dadi to be able to infer parameters of a change in selective pressure experienced by a population. Overall, it is a good contribution and has the potential to yield interesting results in empirical applications. I have a few concerns listed below that I think should be addressed before it is acceptable for publication:

Major points:

1) If I understand correctly, the authors have extended dadi to infer new parameters of the DFE. For instance, you say in your abstract: "Third, using a novel joint DFE model, we are able to quantify the fraction of mutations that have experienced a change in their selection coefficient (*pc*) during domestication". But I don't see a description of the underlying methods. What were the diffusion equations that were solved to estimate p_c? How do you infer the beneficial DFE? These details must be laid out in the main textnclearly.

The method developed is based on Huang et al. (2021). Here we extended that model to consider joint DFEs that include mutations that are beneficial in one or both populations. All the new implementations are now described in the Materials and Methods section, specifically in the subsection "dadi: 2D-SFS and 2D-DFE" line 274.

2) It could be nice to add a scenario where there are no beneficial mutations and there is only BGS. This would help understand if mildly beneficial mutations can be

causing the misinference of the timing of bottlenecks (lines 440-442).

We added six more scenarios: with and without migration and different proportions of sites changing selection coefficients but without beneficial mutations (see Table 1). We confirm that mildly beneficial mutations do not cause the misinference of demographic parameters. Therefore, we interpret that the misinference must be due to background selection (see Results section lines 458-471).

3) You say that mean s_d values are estimated quite accurately using both polyDFE and dadi (lines 579-582). However, I do not conclude that from Figure 3. I see quite a bit of misinference. Either the scale of misinference is unclear in the Figure or our interpretation is different. But I would not conclude what the authors are concluding here.

We have included into the main text the former Supplementary Figure 5 (now Figure 3) to mainly focus on the inference of the DFE distribution. We have split the former Figure 3 into two new supplementary figures for visual clarity: The new Supplementary Figure 4 shows the shape parameter and the new Supplementary Figure 5 shows the mean s_d values. Now we have rewritten this sentence as: "Regarding the mean of the deleterious DFE (s) in Figure 4, we observe that the inferred mean values across bootstrap replicates vary by up to 20% higher or lower, depending on the population, scenario, and inference tool. The largest misinference occurs when positive selection is strong and *dadi* is used and in the Domesticated population and polyDFE is used."

4) Lastly, it was a bit difficult to follow the authors' arguments in some sections (e.g., lines 522-523). It would be great if the authors could proofread the manuscript to improve clarity, sentence structure, etc.

Thanks. We improved the writing in all the manuscript. Specifically, for the sentence described, now it says (lines 535-538): "Second, when we contrast models with and without considering the positive DFE (that is, testing the nested models M1 vs M2 and M10 vs M20), it yields statistically significant results in all scenarios (see Supplementary Table 2, fourth and fifth columns)."

Minor issues:

1) Figure 2 – it's difficult to see the simulated population history, especially for the wild population. Would be good to change the colors of the simulated line to something that is clearly visible. Also is that relative Ne? Please specify what the y-axis is.

The Figure 2 has been modified to better contrast the patterns of the different simulated models. We have detailed the meaning of the x and y-axis in the Figure legend. Specifically, the x-axis indicates the number of generations in relation to the ancestral population size Na, while the y-axis show the population size at each time in relation to Na (that is, Ne/Na, where 1 means that Ne=Na).

2) Line 137: "Several studies have been focusing on the significant distortions in the DFE shape caused by the effect of linked selection (e.g., Gilbert et al. 2021, Johri et

al 2021, Shrider et al. 2016).” -> Neither Johri et al nor Schrider et al evaluated DFE inference. Citations need to be corrected or maybe the way you are phrasing it needs to be modified.

The Introduction section has been modified to better focus on the background of the topic here studied. This paragraph has been replaced and these works have been later discussed in the Results and Discussion section.

3) “Shrider” should be “Schrider”

Thanks, the paragraph and the citation has been erased.

4) Although Gilbert et al showed that the population-scaled DFE is misinferred with high rates of inbreeding, Daigle & Johri (2024) show that selection coefficients are misestimated only when HRI effects are very severe. The authors should discuss this.

We modified the text, and we have included a paragraph in the Results and Discussion section discussing these works in relation to the results obtained here (lines 555-565).

5) “…, there are more rare synonymous polymorphisms compared to what we would expect if there was free recombination under a constant population size (Nielsen 2005).” -> Needs many more citations (e.g. Charlesworth et al 1995, Nicolaisen and Desai 2013, Ewing and Jensen 2016, Messer and Petrov 2013)

We have included more references (line 467-471): " This signal of a recent expansion in the Wild population is expected because when we consider how linked selection affects the SFS, there are more rare synonymous polymorphisms compared to what we would expect if there was free recombination under a constant population size (Charlesworth et al. 1993, 1995, Nielsen 2005, Zeng and Charlesswoth 2011, Messer and Petrov 2013, Nicolaisen and Desai 2013, Ewing and Jensen, 2016). "

6) I think Figure 1 C and D is quite confusing, although Table 2 helps. It is up to the authors eventually, but it would be great if Figure 1 C and D could be improved for clarity.

We have modified the Figure 1 to clarify the meaning. We have included headers with titles and short sentences explaining the meaning of the plots represented in the Figure. In summary, the left side of the Figure 1 shows the scenario for demographic and selective parameters used in simulations, and the right side of the Figure 1 shows the demographic and selective parameters used for inference analysis. Note that are not necessarily the same; we have included more demographic parameters for inference that were used in simulation. On the other hand and given the limitation of 2-D dadi analysis, the positive selective parameters are limited to one fixed positive selective value for each inference performed.

References:

Austin Daigle and Parul Johri. 2024. Hill-Robertson interference may bias the inference of fitness effects of new mutations in highly selfing species. bioRxiv 2024.02.06.579142.

Charlesworth, D, B Charlesworth, and M T Morgan, 1995 The pattern of neutral molecular variation under the background selection model. Genetics 141: 1619–1632.

Nicolaisen, L E, and M M Desai, 2013 Distortions in genealogies due to purifying selection and recombination. Genetics 195: 221–230.

Ewing, G B, and J D Jensen, 2016 The consequences of not accounting for background selection in demographic inference. Mol. Ecol. 25: 135–141.

Messer, P W, and D A Petrov, 2013 Frequent adaptation and the McDonald–Kreitman test. Proc. Natl. Acad. Sci. USA 110: 8615–8620.

Review by Miguel de Navascués, 14 Jun 2024 07:59

Title and abstract

Does the title clearly reflect the content of the article? No: Title can be more clear (discussed below)
Does the abstract present the main findings of the study? Yes

Introduction

Are the research questions/hypotheses/predictions clearly presented? No: Objectives of the study could be presented in a more clear way (discussed below)
Does the introduction build on relevant research in the field? Yes

Materials and methods

Are the methods and analyses sufficiently detailed to allow replication by other researchers? No: Code is provided to replicate simulations and analyses, but description is insufficient to be understood (discussed below)
Are the methods and statistical analyses appropriate and well described? No: Methods are appropriate as far as I can tell, but description is insufficient (discussed below)

Results

In the case of negative results, is there a statistical power analysis (or an adequate Bayesian analysis or equivalence testing)? I don’t know
Are the results described and interpreted correctly? Yes

Discussion

Have the authors appropriately emphasized the strengths and limitations of their study/theory/methods/argument? Yes
Are the conclusions adequately supported by the results (without overstating the implications of the findings)? Yes

In this work, the authors address the performance of statistical methods to infer demography and selection (distribution of fitness effects, DFE) under an evolutionary scenario representing a domestication process. Some of the methods evaluated are widely used in the community, and one is a newly developed approach that accounts for the change of selection regime in the domesticated population. While the results reflect previously published evaluations of the negative effects of linked selection on demographic inference and the challenges of disentangling selection and demographic effects, the work merits attention for its specific focus on domestication and the development of a new approach. Contrary to the opinion of previous reviewers, I argue that the domestication scenario is a societally important evolutionary process with broad interest. Moreover, a scenario of divergent populations with one undergoing a bottleneck and a change of selection regime may also be relevant in other contexts, such as the case of invasive species. Unfortunately, the text presenting the work remains unclear, particularly regarding the description of the new approach. Below, I provide a list of points (roughly ordered from "concerns" to "suggestions") that need to be addressed before the work can be recommended.

1. The description of the new approach (lines 275-334) lacks sufficient detail for the reader to understand the new development and its implementation. First, little explanation is provided about the previous developments upon which the new approach is based. There is just a sentence citing four works, leaving the reader2. 3. approach is based. There is just a sentence citing four works, leaving the reader to decipher why these works are cited, what key elements they provide for the new method, and how they were used. I assume the work builds upon the developments of Huang et al. 2021, and that Jerison et al. 2014 and Ragsdale et al. 2016 are cited for the concept of joint DFE. Assuming this is correct, the text should briefly describe the approach in Huang et al. 2021. Jerison et al. 2014 might fit better in the introduction, and the relevance of Ragsdale et al. 2016 (which focuses on the joint DFE of triallelic sites) is unclear. If my assumption is incorrect, the text should be revised to clarify this for the reader.

The text describes the proportions of different types of sites under selection used in the new approach. However, it is unclear how these proportions affect the site frequency spectrum (SFS) and how they are applied in practice for the inference. The authors refer readers to the code implementing the model without providing a mathematical description of it. It is imperative that the article clearly states how these proportions of different sites under selection modify the equations/algorithms (presumably from Huang et al. 2021) that provide the expected SFS. Otherwise, understanding the model would require reverse engineering from the code, which is unreasonable. Describing the work upon which the new model is built will provide the necessary elements for explaining the new model.

As in the answer of the first review: the method developed is based on Huang et al. (2021). Here we extend that model to consider joint DFEs that include mutations that are beneficial in one or both populations. All the new implementations are now described in the Materials and Methods section, specifically in the subsection "dadi: 2D-SFS and 2D-DFE" (subsection starting at line 274).

2. The description of the mutation process simulation is unclear. A more detailed description is needed on how the location of different types of sites is “randomly precalculated” and the proportion of different types of sites (m1, m2, etc.) in the genome (authors only report *pc* for the simulations). It might also be useful for the authors to explicitly explain whether the different proportions of sites are predetermined as fixed proportions or if each site type is randomly assigned and how many realized mutations of each type occur in the simulation. Clarify whether *pc*(p_c) is a probability or a proportion and how that translates to the actual proportion of mutations involved in domestication.

We have modified the paragraph and included a new column in the Table 2 to clarify the calculations (lines 231-235): "...A mutation at m7 sites will change its selection coefficient from deleterious in the Wild to beneficial in the Domesticated population (see probabilities included in Table 2). For each of the 24x100 simulation runs, we randomly pre-calculate independently the location of each site type (except for the permanent disposition of two non-synonymous sites followed by a synonymous site within codons) and their selective effect using an ad hoc R script. "

The Table 2 now includes a probability column with the probabilities to have every different mutation type.

3. The notation throughout the paper requires thorough revision. While some remarks on notation might be considered arbitrary conventions, others are necessary to avoid confusion. Symbols that represent quantities or variables should be in italics while labels ("e" for effective, "d" for deleterious, etc.) should be in roman ("upright"): several indexes in the notation used should be in roman: $p\mathrm{c}$, $N\mathrm{e}$, $s\mathrm{d}$, etc. I suggest to avoid the use of letter "x" as symbol for multiplication: no symbol (multiplication between two symbols) or the multiplication symbol (×) is a less ambiguous. I suggest to avoid using operators ("+", "-") as labels when they can create confusion in equations (e.g. line 286), or, if kept, use a less confusing arrangement, for instance $p\mathrm{w}+$ instead of $p\mathrm{+w}$ (as in $p\mathrm{c}+$). I think that three letters (e.g. p, q and g or some other choice) might be even a simpler (thus better) option to $p\mathrm{c}$, $p\mathrm{c}+$ and $p\mathrm{w}+$. In line 266, I suggest to replace n for i or to clarify the relationship between n and i. I suggest to use "%" only for percentages, i.e. express selection coefficients as -0.01 instead of -1%, express probabilities as numbers between 0 and 1 (note, it is not the same that each locus has a probability 0.25 of changing their selection regime that 25% of loci change their selection regime). It would be good to be more consistent with the notation throughout the manuscript: in the methods $S\mathrm{b}$ and $S\mathrm{d}$ are defined respect to $N\mathrm{e}$ (line 222, $S\mathrm{d}=2N\mathrm{e}s\mathrm{d}$); however, in the simulation model and the inference model described in figure 1, $N\mathrm{e}$ is not defined (does it correspond to $N\mathrm{a}$?, to $N\mathrm{ew}$?); then, γ is used for the population ($N\mathrm{a}$) scaled selection coefficients instead of S which can lead to further confusion. The authors try to clarify some of this in lines 353-355 but it is still unclear the relationship between parameters because there is a lack of explicit definition of equivalences ($4N\mathrm{e}s=2N\mathrm{a}s$?). I would suggest to use a uniform notation throughout the paper and provide a complete description of the relationships between parameters. If there are not equivalences between models (e.g. simulation model and inference model) I think that author should state clearly so when comparing those parameters. Their lack of equivalence does not come from scaling respect to $N\mathrm{e}$ or $N\mathrm{a}$, because $N\mathrm{a}$ is defined in both models: I think authors should rescale, if necessary, to the same reference N to be consistent. Regarding the equivalence to SLiM notation, the explanation is only relevant for readers that will use the simulation code and (in my opinion) that explanation should be a comment within the code, not the main text. This will simplify the text for the general reader that does not need those technical details for understanding the analysis.

We have changed the confusing notation and hope that it clarifies the methodology and comparative analysis carried out. We have included the italic types only for inferred parameters and we have modified the notation of the indices, eliminating the "+" and "-". We have eliminated the "x" used as a multiplication and included "·" instead. We have removed the percentages that refer to probabilities in the manuscript. The term Ne has been replaced by Na to clarify the study, and all population scaled coefficients have been transformed to the same value (2Nas) for ease of understanding.

4. In several parts of the text the description of reality, models and inferences is sometimes confusing, attributing model features to reality or describing inference

results as processes. For instance, the abstract states that “Artificial selection can alter the selection coefficients”. A change in the selection regime can be modelled as changes of selection coefficients, as done by the authors, but it can also be modelled as a new optimum for a quantitative trait determined by a set of loci whose effects on fitness is not determined by a selection coefficient but by their effect on the trait and the genetic background they are in. There is not such a things as selection coefficients for real alleles in a domesticated population, only models that try to describe the actual process in a useful way. In lines 468-469 authors state that strong selection “erase[s] the demographic history” and “recreate[s] large recent population expansions”. The presence of strong selection does not change demography (in the simulated model of this work), but it can challenge its inference: I suggest the authors to use words that convey that idea (e.g. “erases the signal of demography” and “generates genetic diversity patterns similar to recent population expansions”). In line 449, is it the “ demographic history” or “the inferred demographic history”? I suggest to revise the whole text to this regards.

Yes. There are a number of unfortunate and confusing sentences that have been rewritten to clarify the discussion of models and inference separately from the description of reality. The Abstract section has been extensively modified, we directly address the reviewer concerns in the introduction and clarify that with our simulated change in selection coefficients we try to model the actual process in a useful way. Several paragraphs in the Introduction and in the Results and Discussion section have been rewritten, as well as sentences in the Conclusions section to clarify the differences between models, inference, and reality.

5. Results and discussion section introduce analysis and methodological details that should have been described in the methods section. The description of genetic diversity presented in lines 652-683 is not described in the methods and it is not described how the different is calculated. In line 381 a key information about the codominance of simulated mutations if given instead of being stated in the methods. Also, methods do not clearly state from which data all the calculation are produced (samples of individuals of the last generation?)

We have included the missing information about the methodology used to perform the analysis described in the Fig S7 (lines 278-283). We have included the information about the co-dominance effect on mutations in the Material and Methods section, specifically in the description of the simulated models (lines 239-245).

6. Regarding the discussion of the results from the estimation of the marginal DFE the authors speculate about domesticated populations having higher number of advantageous mutations due to migration. I do not understand this reasoning because migration should also introduce deleterious mutations. I think the authors should provide a more clear explanation of their reasoning and they should backup their statement by looking at the simulated data the actual proportions of advantageous mutations of their models.

We appreciate the reviewer’s comment regarding migration introducing both deleterious and beneficial mutations from wild to domesticated populations. We agree that migration plays a dual role in shaping genetic variation. However, as selection coefficients can change in domesticated populations (specifically when pc>0), mutations that are effectively neutral or slightly deleterious in wild populations may become beneficial in domesticated populations. In the absence of migration, this scenario would manifest as positive selection from standing variation. With migration, however, these mutations that have become beneficial can be continuously reintroduced during the domestication bottleneck, potentially increasing the number of beneficial mutations segregating in domesticated populations compared to scenarios without migration.

While this observation is interesting, we recognize that it is tangential to the primary question addressed in our study. To maintain focus, we have decided to remove this point from the manuscript. We thank the reviewer for highlighting this aspect and prompting us to clarify its relevance.

7. I suggest to remove altogether the paragraph about polygenic adaptation (lines 83-96). I do not see the relevance. The discussion of considering the fitness as a trait is also misleading, the type of dynamics that selection of quantitative traits can have could be very different to what it is modelled in this work. An allele of a quantitative trait locus is not deleterious or beneficial per se, it would be positively selected or negatively selected depending on the diversity of the individual genomes in which it occurs and on the whole genomic composition of the population,and this can change through time even if the phenotypic optimum remains constant. In the same line, I suggest to replace “considers polygenic adaptation (considering fitness as a trait)” for “considers multilocus adaptation”

We have removed the paragraph since it is not relevant.

8. I suggest to revise the title, which is misleading. Current version seems to state that “detection of domestication signals” is done “using simulations”. I would suggest removing “using simulations” because keeping that in the title would lead to a more cumbersome explanation: e.g. “Evaluation of methods to detect … using simulations”.

Thanks. Done.

In the same line, I suggest to revise the first sentence of the materials and methods (line 170). A “simulation analysis of the domestication process” does not seems a proper description of what it was done. The simulation are used to evaluate the performance of several inference methods, there is no analysis of the domestication process.

Right. Done.

And again, in the conclusions (line 703), the simulations “provide valuables insights” on method performance, not on demography and adaptation in the context of domestication. By using a simulation model, the genetic dynamics of domestication are assumed to follow specific processes determined by the authors vision of domestication. As discussed above, there is some confusion between reality and model/inferences in the discourse of the authors.

Thanks, we have modified the sentence (lines 685-687): now it says "In summary, our use of forward-in-time simulations has provided valuable insights into the inference of complex genetic demographic history and distribution of fitness effects (DFE) for both new and standing amino acid mutations in the context of domestication."

9. The authors share the code used in their work through a GitHub repository. This is good but it must be noted that this does not completely comply with FAIR criteria. Research data repositories should assure the permanent availability of the shared data. A GitHub project could be removed by decision of the project owner or GitHub could cease to exists (GitHub has not vocation of being a permanent repository; therefore, there are not provisions to assure the persistence of research data contained in their projects). I strongly encourage the authors to deposit their code in a proper research data repository. I suggest, as fast and easy solution, to use Zenodo, as it has tools to link with GitHub projects.

Yes, you are right. We include all these information in a Zenodo repository (https://zenodo.org/records/14277802). Supplemntary Information subsection, line 707.

10. Line 586. It is unclear what is meant with "easier". I think the authors mean that the inference of deleterious DFE is robust to demographic model misspecification, but in any case the sentence would benefit from some revision.

We have modified the entire paragraph (lines 555-565). Now says: "Inferred parameters related to the deleterious DFE: Supplementary Figure 4 and 5 depicts the distribution of parameters related to the deleterious DFE that are estimated by performing bootstrap analysis using polyDFE and dadi. We observe that both tools have a tendency to marginally overestimate the shape parameter of the gamma distribution employed to model the deleterious DFE (Supplementary Figure 4). The overestimation is particularly significant in polyDFE, when positive selection is strong. In such scenarios, dadi's shape estimation is sometimes rather noisy. Regarding the mean of the deleterious DFE (s) (Supplementary Figure 5), we observe that the inferred mean values across bootstrap replicates vary by up to 20% higher or lower, depending on the population, scenario, and inference tool. The largest misinference occurs when positive selection is strong and *dadi* is used and in the Domesticated population when polyDFE is used."

11. Line 124. Does not the work by Huang et al. 2021 provide a method to detect differences in DFE between closely related populations?

We have clarified this point (lines 135-140): "Importantly, we introduce a novel methodology based on Huang et al. (2021) that incorporates an additional parameter critical for distinguishing populations experiencing rapid selective change: the selective effects of a fraction of existing variants can change (e.g., from beneficial to deleterious, or vice versa) in domesticated populations. This method jointly infers the full DFE parameters for both wild and domesticated populations, including shifts in the selective effects of shared variants."

12. In the abstract, the objectives of the study, as presented in the first four sentences, are vague. First, there is a mention to “genetic consequences”, which could mean anything. Then there is a confusing sentence mixing the domestication process with one model to describe it (see previous comment) and the next sentence starts with “To investigate this” but it seems unclear to me what “this” refers to. I suggest to revise this sentences.

We have modified extensively the Abstract to clarify the objectives, the work performed and the main findings (lines 13-29).

13. lines 118-120. Please develop or give some references, it seems unclear to me to what kind of study this sentence is referring.

We have modified the Introduction paragraphs. The new sentences say now (lines 77-84: “In short, genomic analyses of domestication have traditionally focused on identifying strong selection footprints, often driven by loci with large effects responsible for phenotypic differences (e.g. Groenen et al 2012; Carneiro et al 2019; Qanbari et al 2019; Li et al 2020). However, Leno-Colorado et al. (2022) found that domesticated and wild pig populations did not differ in the number and type of non-synonymous fixed mutations, contradicting the idea that most domesticated traits follow a Mendelian genetic architecture. Thus, the hard selective sweep model may be the exception rather than the rule in pigs domestication.”.

14. lines 25-26: I suggest to remove (or modify) “such as the shape and strength of the deleterious DFE”. First, a distribution has no “strength”, I think authors mean the “mean” of the distribution. Then, the type of assumed distribution is not mentioned, so mentioning the “shape” parameter does not make much sense here. Also, note that this is also a confusion between reality and the model (as discussed previously), DFE have a gamma distribution in (some) models.

We have modified the Abstract to clarify the objectives, the work performed and the main findings (lines 13-29).

15. line 23: replace “fluctuating” with “changing” (or equivalent). There is no fluctuating selection in the models considered in this work.

The abstract has been modified and the word has been replaced.

16. line 246: What is it meant with “disposition”? “arrangement”?

We have replaced the word by “location". (line 233)

17. line 143: I suggest to replace “SLiM simulations (ref)” for “forward-in-time simulations” (here the specific software used in not relevant and the reader cannot understand the type of simulation if they don't know about SLiM)

Agree. Changed (line 127).

18. figue 1: I suggest to remove “=1” from “*Na*=1” or explain what does it mean.

We have removed the “=1” in the Figure 1.

19. lines 111-113. This explanation might be better placed in the first mention of the terms 1D-SFS and 2D-SFS.

In the new paragraphs of the Introduction section this explanation is in the lines 97-99.” Previous studies estimating the DFE have primarily relied on contrasting the site frequency spectrum (SFS) of synonymous and non-synonymous mutations within a single population (1D-SFS).”.

20. lines 693-694: Maybe this can be clarified or developed more. I am unsure if the authors are referring to already existing methods (which ones) or methods that should be adapted or developed.

Thanks. We have clarified this paragraph (lines 663-664): “Nevertheless, under the assumptions used in this work, we believe that the discretized deleterious DFE is estimated with reasonable accuracy.”.

21. line 51: I might be missing some subtle meaning here but, isn’t “artificial selection” always “human-induced”?

we have erased this redundancy (line 46).

22. line 611: typo, it should be “out” not “our”

Done (line 597)

23. line 223-224: format of references should be revised

Done (line 209)

24. line 271: “R” should be removed

Done (line 273)

25. line 847: incomplete reference.

Thanks. We have completed (lines 827-829).