

Distribution modelling of an introduced species: do adaptive genetic markers affect potential range?

Neftalí Sillero, Raymond B. Huey, George Gilchrist, Leslie Rissler and Marta Pascual

Article citation details

Proc. R. Soc. B **287**: 20201791.

<http://dx.doi.org/10.1098/rspb.2020.1791>

Review timeline

Original submission: 19 March 2020

1st revised submission: 25 July 2020

2nd revised submission: 20 August 2020

Final acceptance: 21 August 2020

Note: Reports are unedited and appear as submitted by the referee. The review history appears in chronological order.

Review History

RSPB-2020-0632.R0 (Original submission)

Review form: Reviewer 1

Recommendation

Accept with minor revision (please list in comments)

Scientific importance: Is the manuscript an original and important contribution to its field?

Good

General interest: Is the paper of sufficient general interest?

Good

Quality of the paper: Is the overall quality of the paper suitable?

Good

Is the length of the paper justified?

Yes

Should the paper be seen by a specialist statistical reviewer?

No

Do you have any concerns about statistical analyses in this paper? If so, please specify them explicitly in your report.

No

It is a condition of publication that authors make their supporting data, code and materials available - either as supplementary material or hosted in an external repository. Please rate, if applicable, the supporting data on the following criteria.

Is it accessible?

Yes

Is it clear?

Yes

Is it adequate?

Yes

Do you have any ethical concerns with this paper?

No

Comments to the Author

The manuscript "Distribution modelling of an introduced species: do adaptive genetic markers affect potential range?" investigates the ability of invasive species to occupy new environmental niches. By combining allele frequency data of several chromosomal inversions in *Drosophila subobscura* with climatic variables data, the authors perform niche modelling guided by genetic variation. They tested if *D. subobscura* (1) has the same environmental preference in the Americas compared to its ancestral habitats in Europe and found that the introduced populations share a large percentage of environmental space with the native populations. Based on genetic information of the source and invasive populations, they further studied (2) if habitat occupancy is depending on the geographic origin of the founders, i.e. if the source population came from warm or cold adapted regions. Finally, they aimed at (3) locating the probable geographic origin of the American invasion by projecting the environmental niche in America back to Europe and showed that invading flies most likely originated from warm adapted native populations around the Mediterranean. The authors suggest that these models can be used to understand initial colonization and spread of invasive species.

I enjoyed reading this manuscript, which is well written and without doubt of interest for a broader audience. This being said, I am missing a broader scope of the manuscript that reaches beyond *D. subobscura*. There are other, similar systems, such as, for example, *D. suzukii*, which expanded its native range to a world-wide scale only recently. It would be very useful to further discuss the prospects and limitations of applying your approach to the aforementioned and also to other systems.

Moreover, I think that important information on the genetic data and environmental variables are either missing or not optimally presented: For example, while arrangements are qualitatively classified as warm and cold-adapted, I am missing information about their identity in the main text. Moreover, it would be helpful to provide a table with statistical support for climatic adaptation, such as correlations with environmental variables.

It is unclear how the six bio variables are chosen. Did you do all possible pairwise correlations and then chose these six based on lowest overall collinearity? Why not performing a PCA on all variables and then using the first PC axes as predictors? Moreover, it is not clear which version of the Worldclim data was used? Worldclim1 has been replaced by more up-to-date data in 2017 (Fick & Hijmans 2017).

I am a bit concerned that arrangement gradients in the Americas are considered adaptive. To which extent could neutral mechanisms such as allele surfing, etc. have contributed to the allele frequency patterns in the Americas and how would that influence your conclusions?

This is a minor point and may be my personal preference, but I think it would enhance the comprehensibility if you choose another word for the INTRODUCED model.

Please find my list of minor comments below:

Line 47-50: This sentence is a bit long and hard to read.

Line 57: I do not understand this statement. Widespread species may show substantial amounts of differentiation. Moreover, a large proportion of markers may be involved in local adaptation, particularly in regions with high linkage, such as inversions.

Line 76 to 78: Here it remains unclear if the secondary invasion is also from European founders or from the first American invasion into other sites?

Line 85: provide a reference for rapid evolution of the arrangements.

Line 110: which errors? how were these validated?

Line 114: add a comma after the parentheses

Line 289: delete "Consequently," and start with "Ecological niche..."

Line 290: perhaps add more examples

Line 304: change -- to -

Line 315 to 317: Long and complicated sentence. Better split into two sentences.

Figure 3: perhaps restructure this plot into a grid with the four subplots arranged in two columns and two rows. This may facilitate to compare the subplots.

References

Fick SE, Hijmans RJ. 2017. WorldClim 2: new 1-km spatial resolution climate surfaces for global land areas. *International Journal of Climatology*. 37:4302–4315. doi: 10.1002/joc.5086.

Review form: Reviewer 2

Recommendation

Major revision is needed (please make suggestions in comments)

Scientific importance: Is the manuscript an original and important contribution to its field?

Good

General interest: Is the paper of sufficient general interest?

Excellent

Quality of the paper: Is the overall quality of the paper suitable?

Good

Is the length of the paper justified?

Yes

Should the paper be seen by a specialist statistical reviewer?

No

Do you have any concerns about statistical analyses in this paper? If so, please specify them explicitly in your report.

Yes

It is a condition of publication that authors make their supporting data, code and materials available - either as supplementary material or hosted in an external repository. Please rate, if applicable, the supporting data on the following criteria.

Is it accessible?

Yes

Is it clear?

Yes

Is it adequate?

Yes

Do you have any ethical concerns with this paper?

No

Comments to the Author

This manuscript invoked adaptive genetic markers into ecological niche modeling and then estimate the most probable geographic source of colonizers, evaluate whether the genetic bottleneck experienced by founders affects their potential distribution, and test whether this species has spread to all its potential suitable habitats worldwide.

It is the first manuscript I ever read which tried to invoke genetic information into the ecological niche model. It is interesting and valuable. Be honest, I am not familiar with Genomics. So I have to read some papers on this topic to make an objective evaluation for this manuscript.

In the manuscript, authors grouped the presence datapoints in the native range into three datasets according to their chromosomal inversion frequency: NATIVE, COLD, and WARM. To build these datasets they identified locations in the native area with a mean frequency of cold-adapted arrangements present in both American continents ≥ 0.5 . Here I am very curious about the distribution of frequency of cold-adapted arrangements. Why authors pick 0.5 as a threshold? Why didn't they separate occurrences into three or more groups? It might be better if they use 0-0.4, 0.4-0.6, 0.6-1 to split the occurrences into three groups. And the group in the middle (0.4-0.6) is the most interesting one because the individuals in this group have both warm-adapted and cold-adapted abilities.

The ENM part looks ok, except for some minor defects.

First, why did they use AUC to evaluate the models? AUC has been proved to have some potential problems and misleadings on evaluating models. See the papers below.

Lobo, J. M., et al. (2008). "AUC: a misleading measure of the performance of predictive distribution models." *Global Ecology and Biogeography* 17(2): 145-151.

Peterson, A. T., et al. (2008). "Rethinking receiver operating characteristic analysis applications in ecological niche modeling." *Ecological Modelling* 213(1): 63-72.

Line 167, remove 'which requires presence-only records,'. It is an ambiguous statement. In some categories, Maxent was put into a type if 'presence-background', rather than presence-only.

Lines 168, 175, replace 'realised' with 'ecological'. It is also an ambiguous statement. Is SDM/ENM modelling fundamental niche, potential niche or realized niche? We haven't get a consensus. To void the misleading, I suggest to use 'ecological niche'.

Line 185, change to "we involved maximum TSS (true skill statistic) as the threshold".

TSS= sensitivity+ specificity-1. The term 'TSS' is clearer than what author used in the manuscript. Allouche, O., et al. (2006). "Assessing the accuracy of species distribution models: prevalence, kappa and the true skill statistic (TSS)." *Journal of Applied Ecology* 43(6): 1223-1232.

Line 301, change "Drosophila subobscura" to "D. subobscura"

Decision letter (RSPB-2020-0632.R0)

01-Jun-2020

Dear Dr Pascual:

I am writing to inform you that your manuscript RSPB-2020-0632 entitled "Distribution modelling of an introduced species: do adaptive genetic markers affect potential range?" has, in its current form, been rejected for publication in *Proceedings B*.

This action has been taken on the advice of referees, who have recommended that substantial revisions are necessary. With this in mind we would be happy to consider a resubmission, provided the comments of the referees are fully addressed. However please note that this is not a provisional acceptance.

The resubmission will be treated as a new manuscript. However, we will approach the same reviewers if they are available and it is deemed appropriate to do so by the Editor. Please note that resubmissions must be submitted within six months of the date of this email. In exceptional circumstances, extensions may be possible if agreed with the Editorial Office. Manuscripts submitted after this date will be automatically rejected.

Please find below the comments made by the referees, not including confidential reports to the Editor, which I hope you will find useful. If you do choose to resubmit your manuscript, please upload the following:

- 1) A 'response to referees' document including details of how you have responded to the comments, and the adjustments you have made.
- 2) A clean copy of the manuscript and one with 'tracked changes' indicating your 'response to referees' comments document.
- 3) Line numbers in your main document.

To upload a resubmitted manuscript, log into <http://mc.manuscriptcentral.com/prsb> and enter your Author Centre, where you will find your manuscript title listed under "Manuscripts with Decisions." Under "Actions," click on "Create a Resubmission." Please be sure to indicate in your cover letter that it is a resubmission, and supply the previous reference number.

Sincerely,
 Professor Gary Carvalho
 mailto: proceedingsb@royalsociety.org

Associate Editor

Board Member: 1

Comments to Author:

The manuscript "Distribution modelling of an introduced species: do adaptive genetic markers affect potential range?" by Sillero and colleagues combines ecological niche modelling with genetic information to estimate the most probable geographic source of invasive *Drosophila subobscura*. In particular, the authors genetically differentiate between cold- and warm-adapted strains, and it is this information that allows them to pinpoint the source of invaders, and to show that accounting for this differentiation substantially improves invasion forecasts. There are a couple other manuscripts in the literature (references 45-47) that also use genetically informed ecological niche models, so the approach in and of itself is not entirely novel. Nonetheless, I think that this manuscript presents an interesting and elegant application to an important invasive species, with intuitive results that to my knowledge have not been discussed previously. I find the manuscript well written and easy to follow, including for people that know ecological niche models but don't have a genetic background like myself.

You will also find that the two reviewers were positive about your work, but do raise several concerns with respect to the analyses and presentation, including the use of the AUC-metric, the presentation of genetic data and choices of environmental variables, and a lacking broader scope that reaches beyond *D. subobscura*. In particular, for a broader audience, I see the main value of the manuscript in its combination of genetic with environmental information, and the potential for such approaches to improve predictive ability substantially. As such, I think it would be particularly helpful to revise the introduction and discussion to include a broader discussion on genetically informed ENMs, and emphasize how the current work fits into this literature. I hope you find these comments helpful for revising your manuscript.

Reviewer(s)' Comments to Author:

Referee: 1

Comments to the Author(s)

The manuscript "Distribution modelling of an introduced species: do adaptive genetic markers affect potential range?" investigates the ability of invasive species to occupy new environmental niches. By combining allele frequency data of several chromosomal inversions in *Drosophila subobscura* with climatic variables data, the authors perform niche modelling guided by genetic variation. They tested if *D. subobscura* (1) has the same environmental preference in the Americas compared to its ancestral habitats in Europe and found that the introduced populations share a large percentage of environmental space with the native populations. Based on genetic information of the source and invasive populations, they further studied (2) if habitat occupancy is depending on the geographic origin of the founders, i.e. if the source population came from warm or cold adapted regions. Finally, they aimed at (3) locating the probable geographic origin of the American invasion by projecting the environmental niche in America back to Europe and showed that invading flies most likely originated from warm adapted native populations around the Mediterranean. The authors suggest that these models can be used to understand initial colonization and spread of invasive species.

I enjoyed reading this manuscript, which is well written and without doubt of interest for a broader audience. This being said, I am missing a broader scope of the manuscript that reaches beyond *D. subobscura*. There are other, similar systems, such as, for example, *D. sukukii*, which expanded its native range to a world-wide scale only recently. It would be very useful to further discuss the prospects and limitations of applying your approach to the aforementioned and also to other systems.

Moreover, I think that important information on the genetic data and environmental variables are either missing or not optimally presented: For example, while arrangements are qualitatively classified as warm and cold-adapted, I am missing information about their identity in the main

text. Moreover, it would be helpful to provide a table with statistical support for climatic adaptation, such as correlations with environmental variables.

It is unclear how the six bio variables are chosen. Did you do all possible pairwise correlations and then chose these six based on lowest overall collinearity? Why not performing a PCA on all variables and then using the first PC axes as predictors? Moreover, it is not clear which version of the Worldclim data was used? Worldclim1 has been replaced by more up-to-date data in 2017 (Fick & Hijmans 2017).

I am a bit concerned that arrangement gradients in the Americas are considered adaptive. To which extent could neutral mechanisms such as allele surfing, etc. have contributed to the allele frequency patterns in the Americas and how would that influence your conclusions?

This is a minor point and may be my personal preference, but I think it would enhance the comprehensibility if you choose another word for the INTRODUCED model.

Please find my list of minor comments below:

Line 47-50: This sentence is a bit long and hard to read.

Line 57: I do not understand this statement. Widespread species may show substantial amounts of differentiation. Moreover, a large proportion of markers may be involved in local adaptation, particularly in regions with high linkage, such as inversions.

Line 76 to 78: Here it remains unclear if the secondary invasion is also from European founders or from the first American invasion into other sites?

Line 85: provide a reference for rapid evolution of the arrangements.

Line 110: which errors? how were these validated?

Line 114: add a comma after the parentheses

Line 289: delete "Consequently," and start with "Ecological niche..."

Line 290: perhaps add more examples

Line 304: change -- to -

Line 315 to 317: Long and complicated sentence. Better split into two sentences.

Figure 3: perhaps restructure this plot into a grid with the four subplots arranged in two columns and two rows. This may facilitate to compare the subplots.

References

Fick SE, Hijmans RJ. 2017. WorldClim 2: new 1-km spatial resolution climate surfaces for global land areas. *International Journal of Climatology*. 37:4302–4315. doi: 10.1002/joc.5086.

Referee: 2

Comments to the Author(s)

This manuscript invoked adaptive genetic markers into ecological niche modeling and then estimate the most probable geographic source of colonizers, evaluate whether the genetic bottleneck experienced by founders affects their potential distribution, and test whether this species has spread to all its potential suitable habitats worldwide.

It is the first manuscript I ever read which tried to invoke genetic information into the ecological niche model. It is interesting and valuable. Be honest, I am not familiar with Genomics. So I have to read some papers on this topic to make an objective evaluation for this manuscript.

In the manuscript, authors grouped the presence datapoints in the native range into three datasets according to their chromosomal inversion frequency: NATIVE, COLD, and WARM. To build these datasets they identified locations in the native area with a mean frequency of cold-adapted arrangements present in both American continents ≥ 0.5 . Here I am very curious about the distribution of frequency of cold-adapted arrangements. Why authors pick 0.5 as a threshold? Why didn't they separate occurrences into three or more groups? It might be better if they use 0-0.4, 0.4-0.6, 0.6-1 to split the occurrences into three groups. And the group in the middle (0.4-0.6) is the most interesting one because the individuals in this group have both warm-adapted and cold-adapted abilities.

The ENM part looks ok, except for some minor defects.

First, why did they use AUC to evaluate the models? AUC has been proved to have some potential problems and misleadings on evaluating models. See the papers below.

Lobo, J. M., et al. (2008). "AUC: a misleading measure of the performance of predictive distribution models." *Global Ecology and Biogeography* 17(2): 145-151.

Peterson, A. T., et al. (2008). "Rethinking receiver operating characteristic analysis applications in ecological niche modeling." *Ecological Modelling* 213(1): 63-72.

Line 167, remove 'which requires presence-only records,'. It is an ambiguous statement. In some categories, Maxent was put into a type if 'presence-background', rather than presence-only.

Lines 168, 175, replace 'realised' with 'ecological'. It is also an ambiguous statement. Is SDM/ENM modelling fundamental niche, potential niche or realized niche? We haven't get a consensus. To void the misleading, I suggest to use 'ecological niche'.

Line 185, change to "we involved maximum TSS (true skill statistic) as the threshold".

TSS= sensitivity+ specificity-1. The term 'TSS' is clearer than what author used in the manuscript.

Allouche, O., et al. (2006). "Assessing the accuracy of species distribution models: prevalence, kappa and the true skill statistic (TSS)." *Journal of Applied Ecology* 43(6): 1223-1232.

Line 301, change "Drosophila subobscura" to "D. subobscura"

Author's Response to Decision Letter for (RSPB-2020-0632.R0)

See Appendix A.

RSPB-2020-1791.R0

Review form: Reviewer 1

Recommendation

Accept as is

Scientific importance: Is the manuscript an original and important contribution to its field?

Good

General interest: Is the paper of sufficient general interest?

Good

Quality of the paper: Is the overall quality of the paper suitable?

Good

Is the length of the paper justified?

Yes

Should the paper be seen by a specialist statistical reviewer?

No

Do you have any concerns about statistical analyses in this paper? If so, please specify them explicitly in your report.

No

It is a condition of publication that authors make their supporting data, code and materials available - either as supplementary material or hosted in an external repository. Please rate, if applicable, the supporting data on the following criteria.

Is it accessible?

Yes

Is it clear?

Yes

Is it adequate?

Yes

Do you have any ethical concerns with this paper?

No

Comments to the Author

After carefully reading the revised manuscript and the response to the editor's and previous reviewers' comments, I feel that the authors substantially improved the manuscript and thoroughly addressed most of the editor's and reviewer's concerns either directly in the manuscript or in the response letters. I have no additional comments.

Decision letter (RSPB-2020-1791.R0)

13-Aug-2020

Dear Dr Pascual

I am pleased to inform you that your Review manuscript RSPB-2020-1791 entitled "Distribution modelling of an introduced species: do adaptive genetic markers affect potential range?" has been accepted for publication in Proceedings B.

The referee(s) do not recommend any further changes. Therefore, please proof-read your manuscript carefully and upload your final files for publication. Because the schedule for publication is very tight, it is a condition of publication that you submit the revised version of your manuscript within 7 days. If you do not think you will be able to meet this date please let me know immediately.

To upload your manuscript, log into <http://mc.manuscriptcentral.com/prsb> and enter your Author Centre, where you will find your manuscript title listed under "Manuscripts with Decisions." Under "Actions," click on "Create a Revision." Your manuscript number has been appended to denote a revision.

You will be unable to make your revisions on the originally submitted version of the manuscript. Instead, upload a new version through your Author Centre.

Before uploading your revised files please make sure that you have:

1) A text file of the manuscript (doc, txt, rtf or tex), including the references, tables (including captions) and figure captions. Please remove any tracked changes from the text before submission. PDF files are not an accepted format for the "Main Document".

2) A separate electronic file of each figure (tiff, EPS or print-quality PDF preferred). The format should be produced directly from original creation package, or original software format. Please note that PowerPoint files are not accepted.

3) Electronic supplementary material: this should be contained in a separate file from the main text and the file name should contain the author's name and journal name, e.g. `authorname_procb_ESM_figures.pdf`

All supplementary materials accompanying an accepted article will be treated as in their final form. They will be published alongside the paper on the journal website and posted on the online figshare repository. Files on figshare will be made available approximately one week before the accompanying article so that the supplementary material can be attributed a unique DOI. Please see: <https://royalsociety.org/journals/authors/author-guidelines/>

4) Data-Sharing and data citation

It is a condition of publication that data supporting your paper are made available. Data should be made available either in the electronic supplementary material or through an appropriate repository. Details of how to access data should be included in your paper. Please see <https://royalsociety.org/journals/ethics-policies/data-sharing-mining/> for more details.

If you wish to submit your data to Dryad (<http://datadryad.org/>) and have not already done so you can submit your data via this link

<http://datadryad.org/submit?journalID=RSPB&manu=RSPB-2020-1791> which will take you to your unique entry in the Dryad repository.

If you have already submitted your data to dryad you can make any necessary revisions to your dataset by following the above link.

5) For more information on our Licence to Publish, Open Access, Cover images and Media summaries, please visit <https://royalsociety.org/journals/authors/author-guidelines/>.

Once again, thank you for submitting your manuscript to Proceedings B and I look forward to receiving your final version. If you have any questions at all, please do not hesitate to get in touch.

Sincerely,
Professor Gary Carvalho
<mailto:proceedingsb@royalsociety.org>

Associate Editor

Comments to Author:

I agree with the reviewer that the authors addressed all comments adequately and that this substantially improved the manuscript. A few minor editorial corrections follow, but otherwise I have no further concerns.

1.69 & a couple other places: add scientific names of species

124: "revising" -& "reviewing"; "removed" -& "removing";

126: "species" -& "species' "

142: delete the second "only"?

153: "meet" -& "meeting"

192: it may not be intuitive to everyone what you mean by "background" -& define

207-208: This requires a clarifying sentence. Naively speaking, aren't you comparing four different models (NATIVE, COLD, WARM, BACKPROJECTION)?

236: incomplete sentence

346: the insertion is not clear, please rephrase.

352: insert "the" before "Western"

353: delete "to"

621: it seems there may be a word missing between "America" and "in the Native"

624: does the contour line also exclude ocean areas, or only Antarctica?

643: INTRODUCED etc were all capitalized in main text, so should be capitalized here, too.

Reviewer(s)' Comments to Author:

Referee: 1

Comments to the Author(s).

After carefully reading the revised manuscript and the response to the editor's and previous reviewers' comments, I feel that the authors substantially improved the manuscript and thoroughly addressed most of the editor's and reviewer's concerns either directly in the manuscript or in the response letters. I have no additional comments.

Sincerely,

Decision letter (RSPB-2020-1791.R1)

21-Aug-2020

Dear Dr Pascual

I am pleased to inform you that your manuscript entitled "Distribution modelling of an introduced species: do adaptive genetic markers affect potential range?" has been accepted for publication in Proceedings B.

You can expect to receive a proof of your article from our Production office in due course, please check your spam filter if you do not receive it. PLEASE NOTE: you will be given the exact page length of your paper which may be different from the estimation from Editorial and you may be asked to reduce your paper if it goes over the 10 page limit.

If you are likely to be away from e-mail contact please let us know. Due to rapid publication and an extremely tight schedule, if comments are not received, we may publish the paper as it stands.

If you have any queries regarding the production of your final article or the publication date please contact procb_proofs@royalsociety.org

Your article has been estimated as being 10 pages long. Our Production Office will be able to confirm the exact length at proof stage.

Open Access

You are invited to opt for Open Access, making your freely available to all as soon as it is ready for publication under a CC BY licence. Our article processing charge for Open Access is £1700.

Corresponding authors from member institutions

(<http://royalsocietypublishing.org/site/librarians/allmembers.xhtml>) receive a 25% discount to these charges. For more information please visit <http://royalsocietypublishing.org/open-access>.

Paper charges

An e-mail request for payment of any related charges will be sent out shortly. The preferred payment method is by credit card; however, other payment options are available.

Electronic supplementary material:

All supplementary materials accompanying an accepted article will be treated as in their final form. They will be published alongside the paper on the journal website and posted on the online figshare repository. Files on figshare will be made available approximately one week before the accompanying article so that the supplementary material can be attributed a unique DOI.

You are allowed to post any version of your manuscript on a personal website, repository or preprint server. However, the work remains under media embargo and you should not discuss it with the press until the date of publication. Please visit <https://royalsociety.org/journals/ethics-policies/media-embargo> for more information.

Thank you for your fine contribution. On behalf of the Editors of the Proceedings B, we look forward to your continued contributions to the Journal.

Sincerely,

Editor, Proceedings B

<mailto:proceedingsb@royalsociety.org>

Appendix A



UNIVERSITAT DE
BARCELONA

Departament de Genètica,
Microbiologia i Estadística
Facultat de Biologia
Av. Diagonal 643
08028 Barcelona
Tel. 934034850
Fax 934034420

We are pleased to submit the revision of our manuscript RSPB-2020-0632 entitled "Distribution modelling of an introduced species: do adaptive genetic markers affect potential range?" following all reviewers and editor comments as detailed in [blue](#) below.

We are including the clean copy of the manuscript and a copy with 'tracked changes' in blue to facilitate identifying the changes.

In the response to referees below, we indicate the line numbers in bold that match both the clean copy and the one with tracked changes (all changes have been marked in blue in the "tracked changes" manuscript to facilitate visualization).

We thank the editor and both reviewers for their useful comments. We hope that this revised version will be clearer to a broad audience.

Shortly before sending the manuscript one of the authors, Dr. George Gilchrist, passed away and we would like to dedicate this paper on his memory. Please tell us how to proceed in case this is possible.

We wish that our manuscript in this revised form fits all the requirements for publication in Proceedings B.

Sincerely,

Marta Pascual on behalf of all authors

Associate Editor: Board Member: 1

Comments to Author:

The manuscript "Distribution modelling of an introduced species: do adaptive genetic markers affect potential range?" by Sillero and colleagues combines ecological niche modelling with genetic information to estimate the most probable geographic source of invasive *Drosophila subobscura*. In particular, the authors genetically differentiate between

cold- and warm-adapted strains, and it is this information that allows them to pinpoint the source of invaders, and to show that accounting for this differentiation substantially improves invasion forecasts. There are a couple other manuscripts in the literature (references 45-47) that also use genetically informed ecological niche models, so the approach in and of itself is not entirely novel. Nonetheless, I think that this manuscript presents an interesting and elegant application to an important invasive species, with intuitive results that to my knowledge have not been discussed previously. I find the manuscript well written and easy to follow, including for people that know ecological niche models but don't have a genetic background like myself.

You will also find that the two reviewers were positive about your work, but do raise several concerns with respect to the analyses and presentation, including the use of the AUC-metric, the presentation of genetic data and choices of environmental variables, and a lacking broader scope that reaches beyond *D. subobscura*. In particular, for a broader audience, I see the main value of the manuscript in its combination of genetic with environmental information, and the potential for such approaches to improve predictive ability substantially. As such, I think it would be particularly helpful to revise the introduction and discussion to include a broader discussion on genetically informed ENMs, and emphasize how the current work fits into this literature. I hope you find these comments helpful for revising your manuscript.

We have followed the editor's comments and revised the manuscript accordingly.

We have broadened the introduction and the discussion sections to highlight the important role of chromosomal arrangements in adaptation and the potential of combining environmental and genetic adaptive information on modelling to improve predictive ability on species distributions.

We have clarified the choice of environmental variables and indicated why chromosomal arrangements are considered warm or cold adapted.

Regarding the AUC-metric or other related metrics such as TSS proposed by the reviewer are affected by the size of the study area. However, this is not a problem as long as models from different areas are not being compared, as in the present case. Thus AUC remains an effective validation metric. We have clarified this in the text.

We hope that this revised version will be clearer to a broad audience.

Referee: 1 Comments to the Author(s)

The manuscript "Distribution modelling of an introduced species: do adaptive genetic markers affect potential range?" investigates the ability of invasive species to occupy new environmental niches. By combining allele frequency data of several chromosomal inversions in *Drosophila subobscura* with climatic variables data, the authors perform niche modelling guided by genetic variation. They tested if *D. subobscura* (1) has the same environmental preference in the Americas compared to its ancestral habitats in Europe and

found that the introduced populations share a large percentage of environmental space with the native populations. Based on genetic information of the source and invasive populations, they further studied (2) if habitat occupancy is depending on the geographic origin of the founders, i.e. if the source population came from warm or cold adapted regions. Finally, they aimed at (3) locating the probable geographic origin of the American invasion by projecting the environmental niche in America back to Europe and showed that invading flies most likely originated from warm adapted native populations around the Mediterranean. The authors suggest that these models can be used to understand initial colonization and spread of invasive species.

I enjoyed reading this manuscript, which is well written and without doubt of interest for a broader audience. This being said, I am missing a broader scope of the manuscript that reaches beyond *D. subobscura*. There are other, similar systems, such as, for example, *D. suzukii*, which expanded its native range to a world-wide scale only recently. It would be very useful to further discuss the prospects and limitations of applying your approach to the aforementioned and also to other systems.

We have broadened the introduction and the discussion section to highlight the important role of chromosomal arrangements in adaptation and the potential of combining environmental and genetic adaptive information on modelling to improve predictive ability on species distribution. **Lines 61-63, Lines 317-320, Lines 330-336**

The case of *D. suzukii* is indeed very interesting and we thank the reviewer for pointing it out. This species has had multiple independent invasions from latitudinally diverse locations -- as suggested by genetic markers, recurrent introductions and subsequent admixture in invasive regions (Fraimout et al. 2017). Global distribution modelling has shown a shift in the ecological niche (Ørsted and Ørsted, 2019) that could be mediated by the mixed origin of colonizers. Under these scenarios, information on adaptive genetic markers in the native and introduced areas, specially linked to chromosomal arrangements due to their recombination reduction in heterokaryotypes, can improve the ability to predict species distribution. Most interestingly in these complex invasions, forward projections modelling from native areas partitioning presence data according to the genetic composition of invaded areas, as done in our study, can further improve distribution modelling and help understanding and predicting the future geographic distribution of these invasive species. Overall, we show that genetically informed environmental niche modelling -- using candidate loci for local adaptation -- can help predict habitat suitability and help us understand and monitor the distribution of invasive species, even those with complex invasive scenarios. This information has been provided in **Lines 323-330**

Similarly to our back projection simulations, this approach can be used in species with complex invasion scenarios, presenting multiple introductions and admixture in certain areas, as in *D. suzukii* (Fraimout et al. 2017). In such complex invasion reverse projections of groups of genetically independent colonized locations could provide an ecological based assessment of the potential origin of colonizers in each event. **Lines 373-376.**

Moreover, I think that important information on the genetic data and environmental variables are either missing or not optimally presented: For example, while arrangements are qualitatively classified as warm and cold-adapted, I am missing information about their identity in the main text. Moreover, it would be helpful to provide a table with statistical support for climatic adaptation, such as correlations with environmental variables.

We have broadened the explanation of what chromosomal arrangements are and their role on environmental adaptation in different taxa. Moreover, we have explained the relation of chromosomal arrangements in *D. subobscura* with latitude and how the same clinal patterns observed in the native area were independently reproduced in both colonized areas, showing their adaptive value. Moreover, those arrangements more frequent in low latitudes, considered warm adapted, have increased worldwide in the last decades along with temperature changes due to global warming. We have included this information in the manuscript and rewritten some parts for clarity. **Lines 64-80, Lines 94-100.**

It is unclear how the six bio variables are chosen. Did you do all possible pairwise correlations and then chose these six based on lowest overall collinearity? Why not performing a PCA on all variables and then using the first PC axes as predictors? Moreover, it is not clear which version of the Worldclim data was used? Worldclim1 has been replaced by more up-to-date data in 2017 (Fick & Hijmans 2017).

We calculated pairwise correlations between all 19 variables. When two or more variables presented Pearson correlation coefficients higher than 0.75 only one variable was retained to reduce co-linearity. We have added this explanation in the text (**Lines 163-165**).

We agree with the reviewer that using a PCA approach is a good method for eliminating the correlation between variables and has been used in some studies, although many other works choose to remove correlated variables above a threshold as in the present work. We prefer to use this conventional method because it is more straightforward to interpret the importance of the variables in the model.

We now included the version information of Wordclim. **Line 163.**

I am a bit concerned that arrangement gradients in the Americas are considered adaptive. To which extent could neutral mechanisms such as allele surfing, etc. have contributed to the allele frequency patterns in the Americas and how would that influence your conclusions?

The invasion of South and North America by *D. subobscura* allowed corroborating the adaptive value of chromosomal arrangements, as the same clinal patterns observed in the native area were independently reproduced in both colonized areas and in a relatively short period of time. Moreover, the analyses of the frequency of chromosomal arrangements across time in both native and introduced areas has shown the capacity of this species to adapt to global warming by the increase of warm adapted chromosomal arrangements at a global scale. We have provided this information and rewritten this part for clarity. **Lines 91-100.**

This is a minor point and may be my personal preference, but I think it would enhance the comprehensibility if you choose another word for the INTRODUCED model.

Different words have been used in the literature to refer to populations of alien species such as introduced, invasive, colonizing, or non-native. We think that using “introduced” for the range and the model of that range mirrors what is done in the case of the native range. We have indicated that presence data points in South and North America combined constituted the introduced range (INTRODUCED). **Lines 135-136.**

Line 47-50: This sentence is a bit long and hard to read.

We have split the sentence and rewritten it for clarity. **Lines 47-50.**

Line 57: I do not understand this statement. Widespread species may show substantial amounts of differentiation. Moreover, a large proportion of markers may be involved in local adaptation, particularly in regions with high linkage, such as inversions.

This sentence indicates that most genetic markers can be considered neutral and show no signs of genetic differentiation within species, and that only a few markers will be involved in local adaptation. We have rewritten the two sentences to avoid confusion. **Lines 57-61.**

Line 76 to 78: Here it remains unclear if the secondary invasion is also from European founders or from the first American invasion into other sites?

We have rewritten the sentence for clarity since there was only one invasion from Europe and the rest were secondary from Chile. **Lines 82-86.**

Line 85: provide a reference for rapid evolution of the arrangements.

This part has been rewritten for clarity. **Lines 91-94.**

Line 110: which errors? how were these validated?

The source literature reported in TAXODROS was revised for those localities where we did not know if *D. subobscura* was present. We detected a few cases where the species name was mentioned in the literature indicating that it was not present but inadvertently considered as present in the database. These error points were removed. We have explained that in the text. **Lines 124-126.**

Line 114: add a comma after the parentheses

Done

Line 289: delete “Consequently,” and start with “Ecological niche...”

Done. **Line 317.**

Line 290: perhaps add more examples

We have included other examples and rewritten the sentence. **Lines 317-320.**

Line 304: change -- to –

Done

Line 315 to 317: Long and complicated sentence. Better split into two sentences.

We have simplified the sentence. **Lines 351-352.**

Figure 3: perhaps restructure this plot into a grid with the four subplots arranged in two columns and two rows. This may facilitate to compare the subplots.

Done.

Referee: 2 Comments to the Author(s)

This manuscript invoked adaptive genetic markers into ecological niche modeling and then estimate the most probable geographic source of colonizers, evaluate whether the genetic bottleneck experienced by founders affects their potential distribution, and test whether this species has spread to all its potential suitable habitats worldwide.

It is the first manuscript I ever read which tried to invoke genetic information into the ecological niche model. It is interesting and valuable. Be honest, I am not familiar with Genomics. So I have to read some papers on this topic to make an objective evaluation for this manuscript.

In the manuscript, authors grouped the presence datapoints in the native range into three datasets according to their chromosomal inversion frequency: NATIVE, COLD, and WARM. To build these datasets they identified locations in the native area with a mean frequency of cold-adapted arrangements present in both American continents ≥ 0.5 . Here I am very curious about the distribution of frequency of cold-adapted arrangements. Why authors pick 0.5 as a threshold? Why didn't they separate occurrences into three or more groups? It might be better if they use 0-0.4, 0.4-0.6, 0.6-1 to split the occurrences into three groups. And the group in the middle (0.4-0.6) is the most interesting one because the individuals in this group have both warm-adapted and cold-adapted abilities.

The chromosomal inversion polymorphism is very rich in *Drosophila subobscura* and has more than 80 identified arrangements in the native region. Due to the strong bottleneck experienced in the colonization of America, only 23% of chromosomal arrangements are present in the introduced area. We used the 0.5 threshold applied to arrangements present in the introduced area to build the different datasets in the native area as a compromise to get the extreme values and retain a fair number of localities with chromosomal inversion frequency to define the groups and avoid geographic overlapping. Overall, 32 localities with chromosomal arrangements information were used to define the warm area building a rectangle where all presence data points were used to build the climatic model. Similarly, 34 localities with information on chromosomal arrangements frequency were used to define the area for the cold rectangle. Considering only the frequency of warm adapted arrangements for building three or more groups could be misleading since arrangements decreasing with latitude but not present in America could unintentionally contribute to build putatively

interesting datasets. Moreover, the number of localities for shaping the rectangles would be reduced.

For instance, although in each chromosome there is only one cold arrangement present in high frequency both in the native and introduced areas, there are many arrangements especially in the southern localities in the native region. Thus having medium frequency of warm arrangements present in America does not indicate having medium frequency of cold arrangements which makes it difficult to select localities with medium frequencies for both types of chromosomal arrangements. For instance in Figure 1, where the mean frequency of warm-adapted chromosomal arrangements present in the Native and Introduced area is color coded, the low frequency of those arrangements in the eastern area in the native range is due not to the presence of cold-adapted arrangements but to the presence of other arrangements in high frequency not found in the introduced area. Thus, we have made the COLD and WARM partitions to focus on the areas with major frequencies of chromosomal arrangements present in America while restricting it to mostly cold or mostly warm arrangements.

We have explained the frequency distribution of arrangements and the reasons for the selection to clarify how we built the two reduced models. **Lines 139-158**

The ENM part looks ok, except for some minor defects.

First, why did they use AUC to evaluate the models? AUC has been proved to have some potential problems and misleading on evaluating models. See the papers below. Lobo, J. M., et al. (2008). "AUC: a misleading measure of the performance of predictive distribution models." *Global Ecology and Biogeography* 17(2): 145-151. Peterson, A. T., et al. (2008). "Rethinking receiver operating characteristic analysis applications in ecological niche modeling." *Ecological Modelling* 213(1): 63-72.

We agree with the reviewer that AUC value increases with the size of the study area, as it is easier to place absences (or pseudo-absences) in unsuitable habitats, but so do other validation metrics coming from a contingency matrix, such as the True Skills Statistic (TSS).

VanderWal et al (2009) show that AUC remains an effective validation metric when species from different study areas are not compared, which is the case in the present study and thus this metric remains a good estimate of model fit. We have acknowledged this problem and clarified its relevance to our work. **Lines 206-208.**

Line 167, remove 'which requires presence-only records,'. It is an ambiguous statement. In some categories, Maxent was put into a type if 'presence-background', rather than presence-only.

The Referee is right, thank you for indicating our mistake. We have changed the sentence to 'which requires presence and background'. We prefer to be accurate in naming Maxent as presence-background method. **Line 191-192.**

Lines 168, 175, replace 'realised' with 'ecological'. It is also an ambiguous statement. Is SDM/ENM modelling fundamental niche, potential niche or realized niche? We haven't get a consensus. To void the misleading, I suggest to use 'ecological niche'.

Done. Lines 192, 200, 210 and Fig 3.

Line 185, change to "we involved maximum TSS (true skill statistic) as the threshold". $TSS = sensitivity + specificity - 1$. The term 'TSS' is clearer than what author used in the manuscript. Allouche, O., et al. (2006). "Assessing the accuracy of species distribution models: prevalence, kappa and the true skill statistic (TSS)." *Journal of Applied Ecology* 43(6): 1223-1232.

The name used "maximum training sensitivity plus specificity logistic threshold" is the one provided by Maxent and we indicate that in the text. **Lines 212-213.**

Line 301, change "Drosophila subobscura" to "D. subobscura"

Done. Line 337.