Round #2

Decision by Michelle DiLeo, 2019-12-05 13:44

Manuscript: https://doi.org/10.1101/557066 (https://doi.org/10.1101/557066) version 2 (October 3, 2019)

Dear Editor,

We have now revised our MS taking into account all comments. We are very thankful to the reviewers and editor for their interest and constructive suggestions to improve the flow of the MS.

Sincerely,

The authors

Dear Authors,

First, I would like to sincerely apologise for the delay in handling your manuscript. The revision came in while I was out of office, and finding a third reviewer to evaluate the methodological additions to the manuscript took longer than expected.

I have now received three reviews on the revised manuscript. I applaud the authors on the thoroughness of this revision. It is clear that a tremendous amount of work went into this, and I believe that the new analyses show that the results are relatively robust, and also introduce approaches that I have not seen used elsewhere in landscape-genetic studies. The reviewers for the most part share my enthusiasm, however, I agree with reviewer 3 that the extensive additions have now led to a decrease in readability. I unfortunately I cannot recommend the manuscript at this time, however I believe only minor changes are needed before the manuscript can be accepted. The suggested changes do not require new analyses, and can be addressed through additional clarification and by moving some of the new methods and results to the supplementary material.

I invite the authors to respond carefully to all reviewer comments. In particular, please work on increasing the readability. For example, both the RF methods produced similar results, so please only present one and move the methods and results of the other to the supplement. Similarly, the authors can consider omitting the model selection on the INLA results and instead just present the full model, with interpretation following from a comparison of coefficients.

=> We have followed the suggestions from the reviewers and editor to improve the readability of the MS: Details on one of the RF approach have been moved to the supplementary material and the model selection has been removed. We have also tried to simplify the presentation of the different approaches and rewritten the discussion in order to increase the readability.

Reviewed by anonymous reviewer, 2019-11-19 18:56

I have revised the new version of this work and I find that the authors have satisfactorily addressed my concerns, in particular adding the simulation study to assess the issue of sample size bring more robustness to the results found. I have only very few minor comments. Overall I think this is an interesting study, which contributes to understand the relationship between urbanization and population dynamics of invasive fauna.

Specific comments: Ln223: of a minimum of 20 captured...

=> Done

Ln270: groups

=> Done

Ln304: explain why these selection of radius (300, 600, 1000, 1500), does it have to do with is social organization, dispersal, etc. It is more interesting to know why this minimum and maximum value.

=> We actually defined the buffer size based on the dispersion of the individuals sampled within each of the 13 sampled sites as for the error circle in the MAPI analysis.

Ln514: R2 to R2 (some are in capital and others no)

=> "*r*" refers to the Pearson correlation coefficient

In the conclusion, maybe it can be added a sentence of what can be done next to bring this type of studies to more practical recommendations for rodent control, o to other research avenues the authors consider important to address in the near future.

=> We discuss about implications for rodent control in the last paragraph of the discussion and we also added the following sentence in the conclusion: "This study highlights the need to investigate for potential effects of past demographic events when conducting genetic studies on non-native invasive commensal rodents in order to assess gene flow to inform control actions in highly urbanized environment."

Reviewed by Tuomas Aivelo, 2019-10-25 15:34

Stragier et al. have revised their previous manuscript to include a population-based analysis approach, which shows results consistent with previous results. In addition, the authors did a simulation to assess whether MAPI can be biased due to small number of sampling sites and they did resampling comparison to show that the original results are present in even smaller number of sampling sites. Furthermore, the authors have provided additional analysis on land cover variables and their relationship to genetic estimates. All in all, this brings quite a lot new results to the manuscript, but also makes the case much stronger.

My main concern was about the relationship between historical versus contemporary aspects of urban environment and I am happy with the additional analysis, explanations and revisions that the authors did.

I think the discussion needs still a bit polishing (some suggestions below): the flow is sub-optimal and some paragraphs are not very clear as they contain a lot of different ideas. I would suggest using less names of variables and more emphasis on the conclusions. For example, "Result from the RF and CAR models also suggested that the land cover class "Spontaneous" had a negative impact on genetic differentiation" does not open a new paragraph in very accessible way. State first your claim and they argue for it.

=> We agree with the reviewer and we re-organized and simplified the discussion to be less technical and focus more on the interpretation of the results and conclusions.

Some minor comments: - In abstract, maybe outline shortly what it actually means that "current genetic structure reflects the interplay between the historical dynamics of urbanization and the variation of contemporary urban habitats" – what are your actual results, i.e., something similar to line 693-695.

=> We added the following sentence to be more explicit about our results:

"Results are consistent with an introduction of the house mouse at colonial time and current genetic structure exhibits a gradient-like pattern reflecting the historical process of spatially continuous expansion of the city from the first European Settlement. The genetic patterns further suggest that population dynamics of the house mouse is also driven by the spatial heterogeneity of the current cityscape, including socio-economics features, that translate in habitat quality."

-Lines 581-594: this paragraph reads much more technical than previous paragraphs – less about individual variables and more about broader implications might bring it more into line with other parts of discussion

=> We have now rewritten this paragraph

- Line 595: Maybe start this paragraph with what you are trying to say, now the beginning reads as a result. Also, this paragraph is too long for all the contents you are trying to put there.
=> This part has now been simplified and reorganized

- Line 666: I would not describe your sample size small, but rather what you say in parenthesis - suboptimal sampling sites for spatial coverage.
=> Done

Review#3

Overall impression

The main results and scope of the study are presented clearly in the abstract. The introduction is well written and logically structured. It gives the rationale for the study of the population genetic structure of urban rodents and shows how the historical context can make the house mouse in Dakar an interesting case study. The historical spatial data and their processing seem adequate for the scope and resolution of the genetic data. The permissions for sampling and handling of animals were duly reported in the manuscript. The multiplicity of methods and analyses used in the study is a challenge to the reader and nor does it always contribute to understanding of the data. The text of the results and discussion are at times hard to follow due to the amount of detail included, but the tables and figures help clarify the text.

Of the modelling approaches the most concerning part is the analysis of the MAPI smoothing output with the spatial regression models (the pairwise random forest and INLA analyses). Analysing smoothed point estimates based on sparse data as if they were actual field observations is prone to lead to overconfident results. However, as the population level analyses and simulations gave similar results, this does not suggest that the overall results of the study are in some way unreliable.

I expect that with a more extensive sampling scheme for the genetic material a richer pattern would have emerged together with more solid understanding of the relative roles of colonization history vs. contemporary urban land use. Nonetheless, even with the limited sample size the study manages to make the case that both factors play a role in defining the population genetic structure of the house mouse in Dakar.

Comments, questions and suggestions

Lines 168–171 It is not clear to me what standardized time period refers to. Does it mean that the variables used in the analyses are the difference between the date of Built-up or Connected and 2016? Please clarify.

=> We computed the difference between 2016 (the year of sampling) and the year at which a cell is considered as built-up and connected. Values for these time periods, computed in years were rescaled between 0 and 1 and then considered as the variable "Built-up" and "Connection" in the different analyses. We modified the text to clarify this point (L169-L172) as follow: "We then computed the time period elapsed between the year during which most of the sites have been sampled (2016) and both, the year at which a spatial unit was considered as built-up and as connected to the first European Settlement. These time periods were rescaled between 0 and 1 for further analyses and thereafter referred to as the historical variables "Built-up" and "Connection", respectively."

Lines 303–310 For future studies I suggest looking into distance weighing (e.g. Aue et al. 2011 1) as a more interpretable alternative to estimating separate effects to each of the distance buffers. While this requires choosing or estimating the parameters of the weighing kernel there would be only one result per covariate to report and discuss.

=> We thank the reviewer for pointing out this approach.

Lines 372–377 It is not clear to me that this model comparison is necessary for evaluating whether adding the historical information is useful in explaining the population genetic structure. The covariate results in Table 3 already demonstrate alone that Built-up and Connection have an effect and that Connection is stronger in terms of its regression coefficient.

=> As suggested by the reviewer and editor we omitted the selection model procedure and now only present the full model.

Lines 375–375 waic in R-INLA refers to the widely applicable information criterion or Watanabe–Akaike information criterion not weighted AIC. The case of the acronym should thus also be WAIC and not wAIC.

=> Automatism… We now only present the full model so we no longer present WAIC and DIC.

Lines 624–662 This paragraph is rather long and is mostly speculation about potential specific factor contributing to the patterns observed in Dakar, with less general relevance. I would consider expressing the ideas in this section more directly and briefly.

=> The discussion was re-organized and this part has been reduced to be more on point.

Figure 2, panel C Why is IDG not included? The omission should be corrected or mentioned in the methods as is done for the other analyses that exclude IDG. See also comment regarding Table S3.1. Also the value is called local F ST everywhere else — not GESTE F ST . I would advocate for consistency.

=> The voluntary omission of IDG in both the GESTE and MAPI analyses was already mentioned in the method section untitled "*Spatial patterns of genetic variation and cityscape heterogeneity*" (L.297-L300) as follow: "The relationship between the spatial patterns of genetic diversity and differentiation and the spatio-temporal urbanization process within the Cap-Vert peninsula (i.e. Gorée Island, locality IDG, was excluded from the dataset) was analyzed using both a punctual (i.e. population-based genetic estimates) and network (i.e. pairwise genetic estimate) approach."

=> To be consistent with the text we replaced "GESTE Fst" by "Local-F_{ST}" and MAPI values (Fst) by "smoothed-F_{ST}" in Fig 2C and 2D, respectively.

Supplementary material

Table S3.1 Should local F ST for IDG not be included in the table as well?

=> We discarded IDG from the GESTE and MAPI analyses as we were interested in describing the effect of the spatially continuous process of urbanization within the peninsula. This is why we did not present local-Fst value for IDG in Table S3.1.

Table S3.2 It does not make sense to report the correlations among a r , H S and local F ST as if they had been measured in the 300m, 600m, 1000m and 1500m buffers.

=> We now mention in the table caption that we present correlations between genetic estimates computed within sites in the column corresponding to the 300m buffer size.

Figure S4.2, legend The legend should make clear which posterior probabilities the figure presents without the reader having to go back to the text, e.g. similary as indicated in Table 3.

=> We clarified the legend for the posterior probabilities in the same way as in Table 3

Spelling and notation

Line 121 extent not extend

=> Done

Lines 204–206 The sentence could be restuctured so there is no need to repeat the acronyms.

=> Done

Line 234 QIAGEN should be in parentheses

=> Done

Line 455 Use of "-" to mark the range of the confidence interval (CI) is inconsistent with other lines, use ";" throughout.

=> "-" was changed to ";"

Bibliography The style is inconsistent, especially notation for volumes, issues and page ranges.

=> The style of the references has been homogeneized.

Line 802 a generalization not ageneralization

=> Corrected

Figure 1, legend time series not ime series

=> Corrected

Figure 2, legend Last sentence "Small inserts representing [. . .] from Natural Earth (public domain)." probably belongs to the legend of Figure 1? Also possibly inset instead of insert.

=> Corrected

Supplementary material

Section 1 "In the 90s" not "In the 90 "

=> Corrected

Section 4 "a thirteenth location was" not "a thirteen location was "

=> Corrected