

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.

Lines 303–310 *For future studies* I suggest looking into distance weighing (e.g. Aue *et al.* 2011¹) 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.

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.

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*.

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.

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.

Supplementary material

Table S3.1 Should local *F_{ST}* for IDG not be included in the table as well?

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.

¹doi: 10.1111/j.2041-210X.2011.00130.x

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. similarly as indicated in Table 3.

Spelling and notation

Line 121 *extent* not *extend*

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

Line 234 *QIAGEN* should be in parentheses

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

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

Line 802 *a generalization* not *ageneralization*

Figure 1, legend *time series* not *ime series*

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*.

Supplementary material

Section 1 “*In the 90s*” not “*In the 90*”

Section 4 “*a thirteenth location was*” not “*a thirteen location was*”