

Response letter to the Recommender

Manuscript PCI Ecology # 579

March 6, 2024

Dear Paul Savary,

Your article, entitled How does dispersal shape the genetic structure of animal populations in European cities? A simulation approach, has now been reviewed.

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.

We thank you in advance for submitting your revised version.

Yours sincerely,

The Managing Board of PCIEcology

Revision round # 1

Decision for round # 1: Revision needed

Revision needed for the preprint "How does dispersal shape the genetic structure of animal populations in European cities? A simulation approach"

Dear Dr Savary and collaborators,

After unexpected difficulties to secure reviews, I have finally received two reports on your preprint entitled "How does dispersal shape the genetic structure of animal populations in European cities? A simulation approach". I sincerely apologize for the unusual long time that was necessary before getting back to you about this preprint.

Both reviewers are enthusiastic about your manuscript. I agree with them that overall, it presents a very interesting and stimulating study and that it is nicely written. The two reviewers highlighted a few limitations in your simulations that need at least to be discussed in details. They also made a number of suggestions of improvements in the writing that I suggest you follow.

I will be happy to read the new version of your manuscript and assess it for recommendation.

Best,

Aurélie Coulon.

by Aurélie Coulon, 09 Jan 2024 17:23

Manuscript: <https://doi.org/10.32942/X2JS41>

version: 1

Response to the Recommender

Dear Dr. Coulon,

We thank you very much for this opportunity to revise our manuscript. We found the reviewers' comments to be very constructive and thorough, which helped us improving the overall clarity of the initial version. Following the reviewers' comments, we have modified the main text in order to explicitly address their most important concerns. Moreover, we have incorporated most of the minor suggestions into the revised manuscript. The responses to some very relevant comments could not be included within the manuscript in their entirety. However, we have included below a point-by-point response to all these comments. Our responses are illustrated by 8 figures (numbered R1 to R8) reporting the results of the additional simulations and analyses we have carried out to address the reviewers' concerns. We hope these detailed responses offer comprehensive explanations and our revised manuscript better stresses both the significance and limitations of our study.

Before providing a detailed response to every reviewers' comments, we deemed important to clarify two major points that repeatedly appeared in the reviewers' comments, i.e., the major goal and scope of our experimental simulation framework, and consequently, the status of the 325 European cities considered in this framework. We provide further clarifications below.

In order to increase the clarity of our responses, we have numbered each of them, which made it possible to refer to previous elements when addressing related points.

We would like to extend our appreciation for your time and consideration, and look forward to hearing from you.

Sincerely,

The Authors

General responses

As stated in the introduction, we wanted to answer the following question: how does dispersal limitation explain the variability of neutral genetic patterns in urban tolerant species? For that purpose, we assessed how variations in dispersal abilities across urban areas affect genetic patterns, independently from any other processes. Besides, we wanted to assess whether habitat type and local variations in the connectivity of these habitats affected the genetic patterns when subject to variations in dispersal. Indeed, connectivity patterns are commonly the main moderator of the dispersal process.

Considering that the two main processes shaping neutral genetic patterns are genetic drift and dispersal-driven gene flow, we needed to control for the intensity of genetic drift to isolate the effect of dispersal variations on genetic patterns. We thus designed an experiment in which variations in dispersal movement behaviors were the main effect, while maintaining drift intensity constant. These variations affected populations located in two types of habitat (forest, UGS) with distinct spatial locations, and more or less connected to each other ("interface" vs other), which was described through local (i.e., within-city) connectivity assessments. Carrying out such an experiment in a real system with sufficient replications would be costly. Moreover, the intensity of drift can hardly be controlled in real systems, thereby preventing from assessing the effect of dispersal alone. We thus adopted a simulation approach.

Our simulation results can be considered as predictions deriving from simulations rooted in a mechanistic understanding of the pattern-process relationships shaping neutral genetic variations. They shed light on the expected effect of dispersal variations on genetic patterns, all else being equal, and could inform the future predictions of empirical studies in urban evolutionary biology.

Due to the goal and experimental strategy described above, we had to strike a balance between

(i) the generality of our findings, (ii) the realism of urban settings, and (iii) the low variability among replicates. For that purpose, we systemically controlled variations in dispersal by using three distinct scenarios. Their ecological meaning was based on assumed differences among species in their ability to cross artificial areas (either sealed or built-up). The second "treatment" we systematically considered was the type of habitat patch (forest or UGS) occupied by the simulated populations. Using a land cover database, we could differentiate habitat types precisely and realistically take into account their differential location within each city. Assessing the relative local connectivity of each patch within each city using the same dispersal scenarios, and discretizing these variations was a way to refine the resolution of our "experimental treatment" to help the interpretation of the overall patterns. Finally, to limit any unwanted variations among replicates, we selected the study areas by controlling their degree of urbanization. Moreover, variations of connectivity were quantitatively assessed in each city to evidence the differences in connectivity among habitat types, as they are a common moderator of responses to dispersal variations in urban settings (see Figure 2).

Accordingly, each city was merely a replicate in which the same simulation was repeated according to three dispersal cost scenarios, all else being equal (i.e., population sizes, number of populations, dispersal rate, maximum cost-distance covered during dispersal). Hence, differences among cities were not our variable of interest. However, we took further precautions to make sure that these differences did not affect the results in unwanted ways:

- First, we included "city" as a random effect (random intercept) in our statistical analyses of the simulation outputs. This is common statistical practice for preventing Type-I error rate inflation due to pseudo-replication, e.g., when carrying out experiments under a block design. In our case, each city can actually be considered as a block. This also ensures that the main effect is properly estimated, irrespective of inter-block variations in the overall value of the response variable. The latter variations were precisely quantified and reported as ICC values in the main text.

- Second, we broke down our data into groups of cities (4 quartiles) depending either on study area radius or on total habitat area. We thereby made sure that these city-level properties were not confounding factors liable to affect the qualitative nature of our results (see the Supporting information).

In this context, addressing the reviewers' comments asking for more details about the variations among cities and their effects on genetic responses would have been out of the scope of our study. Similarly, to increase the generality of our results, we intentionally made very conservative assumptions when parameterising the simulations. For instance, not assuming any differences in population density among habitat types ensured that dispersal-driven gene flow would be responsible for all the variations in genetic responses among these habitats.

This being said, we totally understand why the reviewers were interested in the effect of city-level properties on the genetic patterns, apart from those of the dispersal cost scenarios and of local patch connectivity. This second question is currently under investigation by our research group. Note that a technical report describing the outcomes of the larger project from which this work originates can be found online: <https://hal.science/hal-03936492/>

This technical report is only available in French language but presents the preliminary findings we obtained when investigating this second question. We are working on another study regarding the role of urban forms on genetic patterns, extending some of these findings, and better embedding this approach into the relevant literature and theory. In this case, a single dispersal cost scenario is used and the focus switches from the effect of contrasted movement scenarios to the role of differences among cities. These differences are analysed by paying close attention to the influence of urban sprawl, habitat area, and habitat subdivision on habitat connectivity, and consequently on genetic diversity (contrasting these effects in forests and UGS).

We hope that the previous points better spell out the details of our study and contribute to the clarity of our subsequent responses to the reviewers' comments.

Review by anonymous reviewer 1, 19 Oct 2023 06:32

General comments:

The paper presents a very interesting and original simulation-based investigation of the effect of the permeability of urban environments (i.e. dispersal limitation) on neutral genetic patterns in populations of urban tolerant species, using the example of forest species occupying both forest and urban green spaces (UGS). Authors simulated genetic data across a set of 325 cities in Europe. Disentangling the influence of drift and gene flow (/dispersal) on genetic patterns is one of the biggest challenges in landscape genetics. Simulation studies are useful to produce predictions and investigate the effect of these processes on observed genetic patterns (genetic diversity and genetic structure). In this context, the present paper is a nice and original contribution in the specific context of urban ecology. I found the study well designed, relevant and timely. The paper is well written and easy to get through. I have though a few suggestions to improve it (see specific comments). Overall, I find the methods convincing and the reported findings interesting. My suggestion is to accept the paper with minor modifications.

We are very grateful to the encouraging appreciation of our study by the Reviewer 1, and would like to thank them for the close scrutiny of our manuscript. We provide below detailed responses to their main concerns and explain how we took into considerations their suggestions in the revised version of the manuscript.

Simulation studies in landscape genetics are generally based on simulated landscapes. Using real landscapes is an originality of the paper. The number of selected cities ($N = 325$) is large enough to allow authors to draw general conclusions. However, I wonder why the authors choose to base the study on real cities rather than simulating urban landscapes (and controlling their topology). I feel it might have been a more simple approach to address the research objective (in addition, the process to select cities is not totally clear to me, see specific comment below). Controlling parameters would have allowed to test some predictions. I do not recommend to re-run analyses with simulated urban landscapes but it will be interesting to explain and discuss this choice in the paper.

R1.1: While we agree with the Reviewer 1 regarding the usefulness of simulated landscapes, we decided to use existing land cover data to increase the realism of the urban settings considered. This also allowed us to assess connectivity contrasts among habitat types and within urban areas based on real data. These contrasts were important and consistent moderators of dispersal variations' effects on genetic patterns. Simulating such contrasts would have required the formulation of strong assumptions regarding their magnitudes and their consistency across cities, potentially casting doubt on the nature and generality of our findings. Using real landscapes therefore appeared as the optimal strategy, especially since harmonized land cover data covering European agglomerations are freely available.

In addition, note that linear landscape features play an important role for the connectivity of habitats in urban areas. Simulating them in a realistic manner is very complex. The lack of such features in simulated landscapes is actually a limitation of previous studies by our research group (see Savary et al., 2021), and we decided to circumvent this issue by using real landscapes in the present study.

In general, there was a lack of descriptive information of the data. For example, it would be interesting to present metrics describing raw differences between forested and UGS patches (location, average size, ...). I wonder if the results from genetic simulations do not simply reflect these differences

(plus the differences in habitat connectivity that were nicely summarized with the F and EC metrics). Actually, parameters of the genetic simulations are identical for forested and UGS patches (same dispersal cost value, no differences in "carrying capacity" or in dispersal rates). Genetic simulation results could have been analyzed without contrasting these two habitats. In fact, this distinction just summarizes differences in size, location and connectivity between the two habitat types. I wonder if genetic response modelling would have been more informative by directly using such continuous metrics in the genetic response modelling (at least location and connectivity such as the Flux connectivity metric).

R1.2: Raw differences between forest and UGS patches were assessed by local connectivity assessments, and summarized at the city-level. These connectivity variations are the direct moderators of the dispersal process and result from the interplay of patch size, type, and location. We thus decided to focus on this ultimate driver, rather than on its concurrent factors. We discretized connectivity variations to increase the generality of our qualitative findings. Because these variations were strong moderators of the responses, we included them in the statistical analyses. Interpreting the parameters of our models specifically associated with the dispersal cost scenarios amounts to considering their effects irrespective of the habitat types occupied by populations. These parameters are very informative and we therefore agree with the Reviewer 1 regarding the usefulness of analyses not considering habitat type variations.

As explained in the main response to the Recommender, we believe that having conservative assumptions (e.g., similar carrying capacities and dispersal rates in each habitat type, depending only on their area) made it possible to interpret our results through the lens of variations in dispersal scenarios, and of the connectivity patterns moderating dispersal, without further confounding factor.

This being said, we provide below two figures representing the proportions of forest or UGS areas at increasing distances from the city center of the studied cities (Figure R1) and a histogram of habitat patch sizes for these two types of habitats (Figure R2). The interplay of these variations is responsible for the city-level differences in EC metrics displayed on Figure 2, and for the local connectivity variations driving the patterns presented in the subsequent figures of the main text.

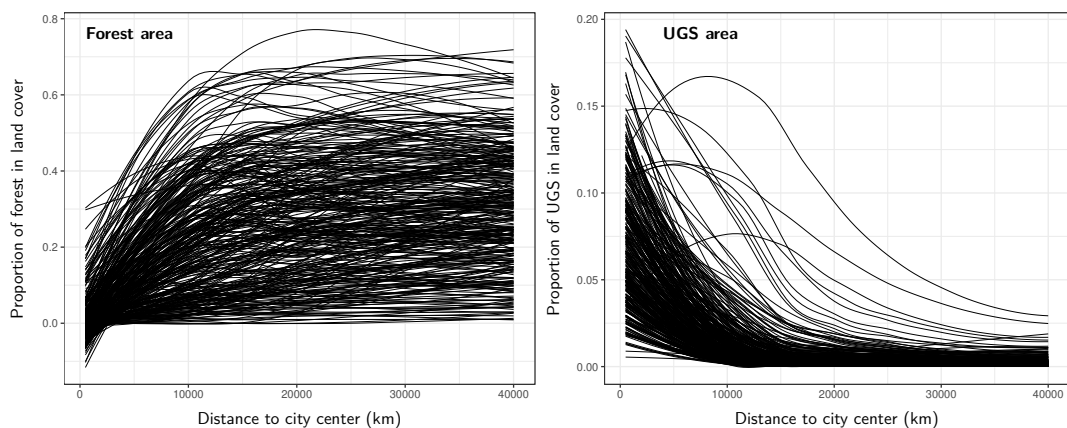


Figure R1: Proportions of forest (left panel) or UGS areas (right panel) at increasing distances from the city center of the studied cities.

Replicates of genetic simulations. Regarding genetic simulations, it seems that there was only one simulation per city and scenario. I wonder if it would be possible to investigate the sensitivity of results to initial conditions in each city x scenario simulation (in particular populations sizes and their location, given that not all habitats patches are occupied). Running e.g. 5 or 10 simulations per city would probably lead more consistent results. Although I don't really measure the amount of computing time this would take ($10 \times 3 \times 325 =$ almost 10,000 simulations). You might at least explore and quantify this variability on a few cities (eg with contrasted number of habitat patches).

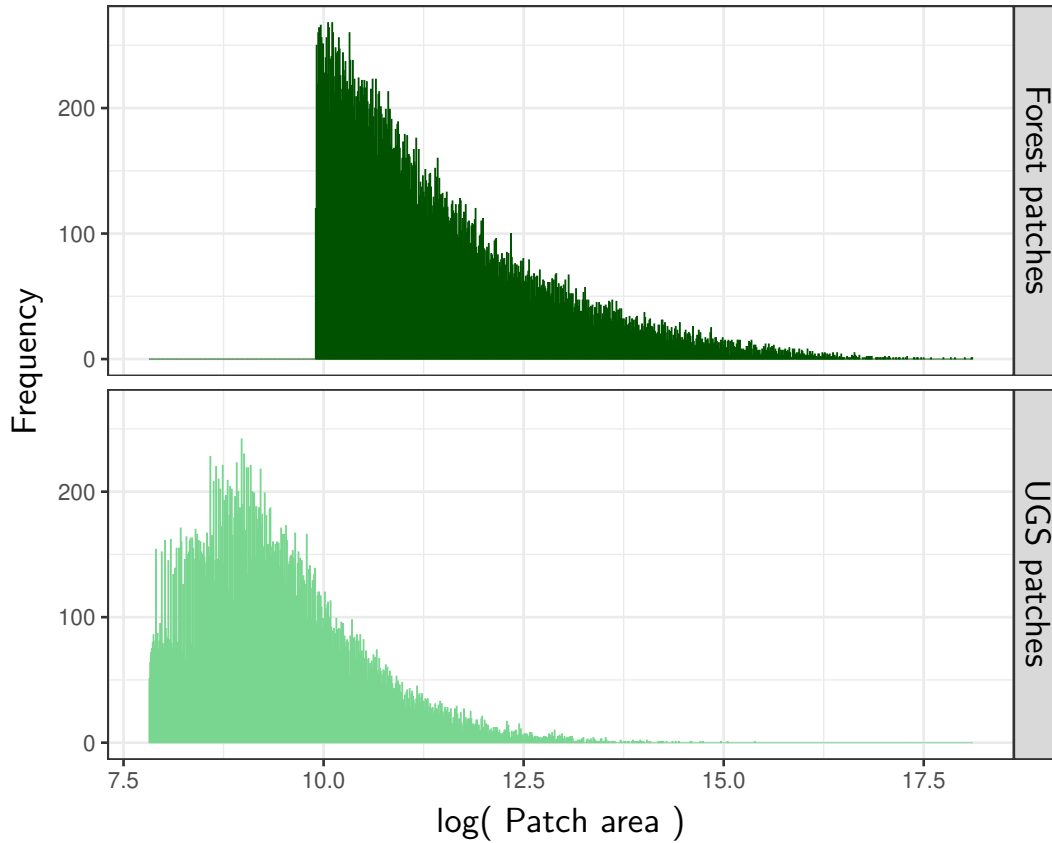


Figure R2: Histogram of habitat patch sizes for both forest (top panel, dark green) and UGS (bottom panel, light green). Patch areas are displayed in logarithmic scales. Note that the truncated distribution of forest patch areas is due to a Minimum Mapping Unit of 1 ha for rural land cover types in the Urban Atlas land cover database.

R1.3: We understand the concern of the Reviewer 1 regarding the number of runs. As explained in our main response to the Recommender, every city was a replicate in which we applied the three cost scenarios in a consistent way. Having 325 cities ensured that we covered a range of realistic urban settings when implementing our experimental simulation framework. This being said, we randomly selected 8 cities, equally spaced along the first axis of the PCA of land cover and urban form variables (Figure R3) and performed 10 runs of the simulations with each of the 3 scenarios within them. These simulations were run for 500 generations to address a subsequent point (see R1.27). We present below the equivalent of the Figures 3 and 4 of the main text considering multiple runs. We split the results among the 8 cities to display the variability of the response variables among runs, that is, the allelic richness and the mean genetic differentiation (DPS) among types of population pairs (Figures R4 and R5, respectively). Note that in both cases (i.e., allelic richness and genetic differentiation), the variability among simulation runs for a given city and scenario is very low, which confirms that one run per city was sufficient.

A larger number of runs per city would have required much higher computing time. We believe that by having 325 cities, we did not need to randomize further the spatial and size distribution of habitat patches within each of them to cover a wider range of variations in this parameter.

Statistical modelling. I liked the use of statistical modelling to summarize the simulations results for genetic metrics. Regarding the genetic differentiation metric, the response variable is a paired genetic distance. I do not think this is appropriate to run LM/GLM/GLMMM on this type of data as they are not independent (a single population is involved in multiple pairs). Permutation tests should be more appropriate in that case. However, since you do not interpret the P values I think this should not be considered as a too serious problem.

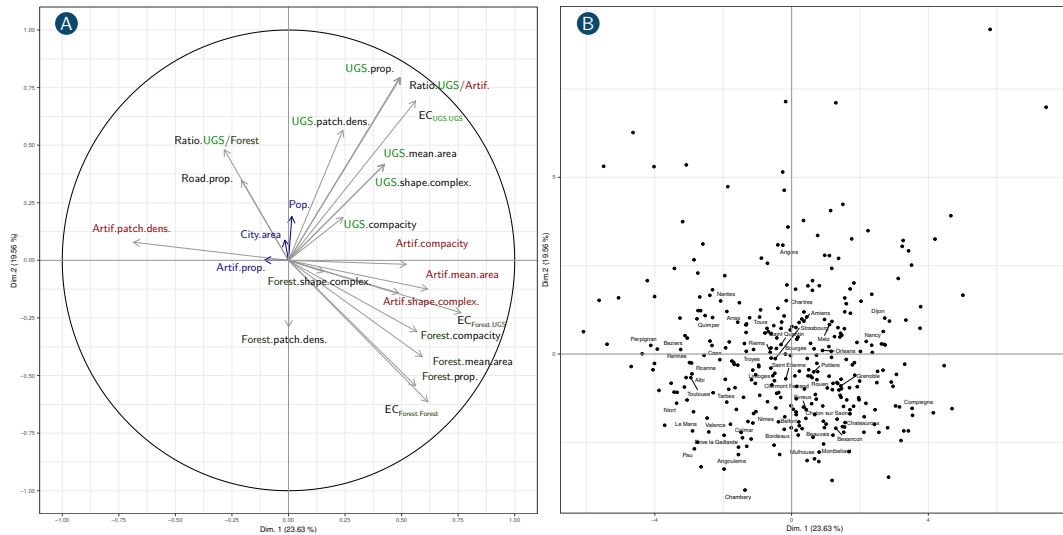


Figure R3: PCA of a set of land cover, urban form and habitat connectivity variables. A) Circle of variable correlations. B) Scatterplot of the 325 studied cities according to their coordinates on the two first principal components. The name of French cities is displayed to provide few examples. The 8 cities considered for the new runs of simulations were randomly sampled, and equally spaced along the first principal component.

R1.4: We thank the reviewer for the opportunity to clarify this technical point. We totally agree that in sets of pairwise measures, such as genetic differentiations between population pairs, independence assumptions are not met because each population is involved in several pairs. Population-level random effects models (e.g., MLPE models, Clarke et al., 2002) or permutation tests (e.g., Mantel tests) are usually employed to prevent the lack of independence from increasing Type-I error rates. However, in our case, pairwise values were averaged per type of population pairs. This averaging removes the dependence of the genetic differentiation indices on the specific populations involved in each given pair.

Accordingly, there were three types of values per city (forest-forest, UGS-UGS, and forest-UGS pairs) and this created a dependence among values measured in the same city. This dependence was one of the reasons why we used mixed-effects models and considered cities as a random effect when testing for the effect of pair type and cost scenarios on genetic differentiation.

The reasons why we did not consider the p-values associated with the parameters of the models are twofold: 1) Simulations studies such as ours have a full control over sample sizes because they mainly depend on the number of in-silico replications carried out (White et al., 2014). Reporting p-values when degrees of freedom can be easily increased is not very informative and can mislead the reader regarding the significance of the findings. 2) The degrees of freedom considered when assessing the significance of parameters from mixed-effects models are subject to debate. Hence, several authors do not recommend to report p-values when using these models (see Baayen et al., 2008 for a demonstration and alternative solutions). We therefore focused on effect sizes, while specifying that the p-values reported by the R software almost always took the smallest value reported in this software.

Specific comments:

Title : "Genetic pattern" might be better than "genetic structure" as you explored both genetic diversity and genetic structure.

R1.5: We modified the title according to Reviewer 1's suggestion.

L22-23: this is only because of simulation parameters that reflected size difference between forest and UGS patches. You should focus on the interaction with scenario.

R1.6: We thank the Reviewer for this relevant comment. We modified the abstract to mention the differences in size and location among these two habitat types, as main drivers of the resulting genetic patterns (L20-22). Given that the interaction effect was not very strong statistically speaking, we decided not to mention it in the abstract.

L58 : « neutral genetic markers are also affected by urbanization » I see what you mean, and agree, however, I found it a bit clumsy. Please reformulate. Eg « neutral genetic patterns...»

R1.7: Absolutely. We reformulated this sentence (L57-58).

L69 : This section seems to provide general background on processes acting on genetic patterns, while the following section focuses on urban tolerant species. Thus, I would not start with a sentence specifying "in these urban tolerant species" as everything following in the section is general and apply to all species (same for "urban populations" L75).

R1.8: We are grateful to the Reviewer 1 for this relevant comment. « in these urban tolerant species » was replaced by « in urban settings » (L69).

L85-87: I suggest to move this sentence (which is very important) to the previous section.

R1.9: We followed this suggestion as this very important sentence closes the previous paragraph adequately (L85-86).

L94-97: The idea is not clear, I suggest to remove the sentence (or reformulate).

R1.10: In hindsight, we acknowledge that the message conveyed by this sentence was both complex and convoluted. We followed the suggestion and removed it (L92).

The initial message could have been framed as follows: an increasing body of literature documents the respective roles of phenotypic plasticity and genetic adaptation for the survival and demographics of species in urban settings. In the light of our results, we deem crucial to consider the role that dispersal alone can have on species demographics and maintenance of genetic variations in these settings, especially when aiming to precisely disentangle the role of these other processes or mechanisms on such outcomes.

L108 : could you please be more specific on the « potential factors » ?

R1.11: We would like to thank the Reviewer 1 for this opportunity to clarify the status of connectivity analyses in our experimental simulation framework. We modified this sentence to explain that the connectivity assessments broken down by habitat type helped us interpreting variations in genetic patterns across scenarios and habitat types in the light of connectivity differences (L115-116).

L116-120: I suggest to move this section to the introduction. Would be nice at the end of the 4th section.

R1.12: We reiterate our thanks to Reviewer 1 for this very relevant suggestion. We slightly modified these two sentences to include them at the end of the penultimate introduction paragraph (L100-104).

L122 : change to «... forest species occupying both forest and UGS »

R1.13: We followed this suggestion (L125-126).

L135-136: It would be interesting to somehow quantify this differential distribution of forest and UGS patches in cities (e.g. kind of density plot with distance from city center for each type and / or of their relative proportion).

R1.14: We agree with the Reviewer 1 that this representation of the proportion of forest and UGS areas at increasing distances from the city center is informative. We used the Urban Atlas land cover database to represent these variations up to a distance of 40 km from the city centers on the Figure R1 included in the present document. Alike the figure R2 also included above and in the Supporting Information (Figure S1), this figure confirms that forest and UGS patches have different areas and locations. These differences underlie the connectivity variations summarized at the city level in the manuscript (Figure 2). Consequently, we prefer to focus on the figure presenting variations in EC metrics in the main text. Indeed, the overall connectivity of forest and UGS patches is a better proxy for the drivers of the genetic patterns' variations we describe and results from the interplay of patch size and location and the nature of the landscape matrix among them.

L139-140: I think a single reference could be enough here.

R1.15: As the recent study by Lemoy and Caruso (2020) focuses on the radial pattern of increase of built-up areas around city centers across European cities, we only kept this reference to follow this suggestion (L144).

L148-156: I found the method of cities and radius size selection complex and its description was unclear. Did you first set up a minimal/maximal radius? Did you increase radius step by step (size of steps?), until having a proportion of artificial area at 20 %? Besides, the sensitivity analysis on radius size in the supporting information is useful and convincing.

R1.16: We acknowledge that this methodological step was complex and its description needed some clarification. We reformulated these sentences (L152-156).

The figure R6, included in the present document, represents variations in built-up area proportions at increasing radius from city centers in more than 600 cities for which we had Urban Atlas land cover data. We computed these proportions at all radius from 1 km to 40 km (steps of 500 m), although we did not consider the smallest radius (≤ 5 km) in our final selection. During these preliminary analyses, we then searched for an optimal "target proportion" for which we could conserve the larger number of cities while minimizing the variations in the radius at which this proportion was reached. As can be seen on the figure R6, 20 % was an optimal proportion in that respect.

L158-159: I suggest to remove the first sentence.

R1.17: We thank the Reviewer 1 for this suggestion. The first sentence was removed and the following one completed with "genetic response" for the sake of clarity (L163).

L159: these responses -> 'genetic responses' or 'neutral genetic patterns'

R1.18: Please refer to R1.17 and L163.

L196: The Flux connectivity metric was only calculated to identify "Forest interfaces" and "UGS interfaces" used in the genetic response modelling. I suggest to move this section to the supporting information in order to reduce the length of M& M. It would also be nice to present the distribution of Flux values with a supplementary Figure.

R1.19: While we agree with the Reviewer 1 that the Flux connectivity metric was only used for distinguishing two types of patches from the others, we deem important to include details about the method used to identify these "interface" patches. Indeed, this is one of the first time this metric is used in this way (i.e., considering two types of habitat and their inter-connections). Besides, we show that the populations occupying "interface" patches identified with this F metric undergo particular demographic dynamics affecting their genetic responses. We therefore removed a sentence related to the F metric in this method section, while referring to a preprint describing in details the novelty of the multiple-habitat graph modeling framework (Savary et al., in prep.).

L231-232: first sentence is not necessary to me.

R1.20: We removed this sentence (L234).

L233: it is also unlikely that all habitat patches are occupied in real metapopulations.

R1.21: Absolutely. We added a sentence stressing that point (L235-236).

L238-239: This sentence is a bit misleading. Do you mean that you sampled patches to populate them with populations? Did you consider patch size? I guess it is more likely that larger patches are occupied by a population.

R1.22: We thank the Reviewer 1 for this reformulation. We now explain that we sampled patches, before populating them with individuals (L241-242). Although we agree with the assumption made by the Reviewer 1 regarding the dependence of patch occupancy on patch size, we chose not to make such an extra-assumption in our study, for being as conservative as possible and reaching a higher level of generality.

L240-242: Could you provide the average size of forested and UGS patches? (maybe better in the "land cover data" section?)

R1.23: The average areas of forest patches and UGS patches are respectively 37.0 ha and 2.5 ha, while their median areas are 6.5 ha and 1.0 ha. These differences are easily explained by looking at the histogram of their areas (displayed according to a logarithmic scale). To follow this suggestion by the Reviewer 1, we therefore decided to include the Figure R2 in the Supporting Information (Figure S1) and to refer to it in the revised version of this paragraph (L253).

L251: remove "each" (typo?), "on average" sounds better.

R1.24: We followed this suggestion (L255).

L256: did the simulation process allowed the colonization of empty patches?

R1.25: Unfortunately, the simulation process did not allow for the colonization of empty patches. This would have required the use of a different simulation program and we could not develop it during the time frame of this project.

L251-255: Could you indicate the proportion of habitat patches (per type, forested and UGS) that are occupied? And its variation across cities? Do you think this parameter could affect simulations results (regarding contrasts among scenarios)?

R1.26: The number of occupied habitat patches over the total number of habitat patches directly follows from the formula included in the main text (after L239). Given that differences in genetic responses were interpreted within each city (i.e., within-block variations in responses to treatment),

variations in this proportion, as other inter-city variations, do not influence the results. We refer to the General responses to the Recommender regarding that point.

L259: did you test the sensitivity of simulations to the number of generations? Did genetic simulations reach an equilibrium after 250 generations?

R1.27: As explained in R1.3, we reproduced the simulations for 8 cities for 500 generations instead of 250, to show that the qualitative nature of our results stayed similar when extending the simulations beyond 250 generations. Please find below the Figures R7 and R8 which compare the results obtained at generations 250 and 500 in terms of allelic richness and genetic differentiation, respectively. The results are qualitatively very consistent.

L265: not clear. To keep the total population or the local population constant?

R1.28: We thank the Reviewer 1 for this opportunity to clarify this detail. We maintained the local populations constant, and consequently, the total population constant. We specified that (L270), as the reverse is not necessarily true (i.e., local variations are possible when only the total population is kept constant).

L287-288: Sentence not necessary. I suggest to remove.

R1.29: We removed this sentence (L291).

L294-295: It might be more clear to indicate that it correspond to a ratio of the maximal DMC value possible for a given city, thus varying from 0 to 1.

R1.30: We rephrased this sentence to follow this relevant suggestion (L297-299).

L330: genetic structure -> genetic pattern (unless you think it only affected genetic structure?)

R1.31: We followed this suggestion (L334).

L339-340: I did not get the point. I agree that you controlled population sizes (according to patch size which are different between Forest and UGS) and dispersal cost values associated with scenarios, so you expect significant p-values related to these factors. However, models should be useful to test the interaction between scenarios and habitat type/type of pair.

R1.32: We believe there was a misunderstanding here, due to our use of the term "sample sizes", which leads to confusion in this context. We replaced it by "the number of replicates". We included a reference supporting this choice and refer to R1.4 for further explanations (L342-344).

L343-344: "When..." I first understood that habitat type/type of pair and cost scenario were included in all models (and so the interaction). Please explain. Also, why don't you show and discuss the estimate of the interaction and its p-value? It could give stronger support to your main conclusions.

R1.33: As explained L347-348, we could only consider "habitat type" or "type of pairs" when modelling allelic richness and genetic differentiation. For these response variables, we thus considered the interaction of this fixed effect with the cost scenarios, because we agree that it gives more insights into the processes at play, and a stronger support to our conclusions. The DMC and ARI indices can only be computed at the city-level, thus preventing the inclusion of any within-city variations as predictors of their variations.

L356: please provide at least the range or maybe more details on the distribution of radius sizes in

the dataset (in supplementary data). Is there a strong correlation between radius size and population (inhabitants) size?

R1.34: We provided the range, mean and median values of the radius sizes in the main text (L360-361). There is a strong correlation between the number of inhabitants of the radius size (Pearson's correlation coefficient = 0.82), which is expected. That is why we controlled for the potential effect of radius size on our qualitative findings (see section "Consistency of analysis results among cities having different sizes and habitat amount" and Supporting Information), and found none.

L377: very high -> higher

R1.35: We rephrased this sentence, that was indeed misleading (L382-384).

L378-379: I don't understand how did you get a separate estimate for forest and UGS interfaces. It is not clear in the methods and results sections. did you include interface as a fixed effect? Ran a separate model? The "genetic response modelling" section needs to be improved with more details.

R1.36: We thank the Reviewer 1 for this comment stressing a lack of clarity in that respect. The patch type predictor was a qualitative variable (and fixed effect) with four levels (Forest Interface, Forest(-other), UGS Interface, UGS(-other)). That is why we obtained a value for each combination of these levels and scenarios (considering the interaction). We extended the sentence L347-348 in the Methods section to clarify this point (see also R1.33).

I found the "sensitivity" analysis very interesting and convincing. Did you consider running models with Radius area or total habitat area as covariates?

R1.37: We thank the Reviewer 1 for this positive appreciation. We followed the rationale explained in the main response to the Recommender and considered "cities" as random effects. Under this approach, models only quantify the variability among cities and consider the multiple properties of each city as random effects. Although we do not estimate the effects of these individual properties (no additional fixed effects), as it is out of the scope of the present study, we account for their potential confounding effects.

L447: "dispersal behavior" Not sure about the use of this term. It reflects both dispersal rate (set constant in the simulations) and dispersal distance. In simulations you only explored distance through the modulation of matrix resistance / permeability. You should be more specific.

R1.38: We agree with the Reviewer 1 that this term can be misleading. Dispersal behaviors, in our study, refer to the movement behavior of individuals when traversing the landscape matrix, while their emigration propensity is fixed at a given rate (i.e., dispersal rates) and their subsequent settlement success is assumed. We therefore rephrased to "variations in dispersal movement behaviors" (L452) to explicitly refer to the traversal stage of dispersal events.

L448-449: (we did not find any comments in the Reviewer's report)

L518-524: I suggest to remove this section.

R1.39: We removed a related section from the introduction following Reviewer's 1 suggestion. We agree that this was needed in order not to confuse the readers at an early stage of the manuscript. However, we believe that mentioning these points in the discussion provides insights into the actual implications of our somehow theoretical and abstract simulation study, mainly focused on dispersal and gene flow, and not on other eco-evolutionary processes.

L525: To assess the spatial distribution of urban habitat types it could be interesting to include distance from city center in genetic response modelling analyses (at least in models for genetic diversity).

R1.40: This is a very relevant point but as explained previously (response to the Recommender, R1.2, and R1.14), this would have prevented from focusing on the ultimate moderators of dispersal variations, i.e., connectivity patterns.

Fig 1: I unsuccessfully tried to figure out which city is shown in the example. Not sure this information is important to improve the quality of the MS but I still like to know the answer :)

R1.41: We are pleased by the curiosity of the Reviewer 1 regarding the city chosen as an example. This is the city of Besançon (France), where a large part of our research group is based. We decided to mention its name in the revised version. Note that we paid close attention to the land cover data in this particular city, because we know it and this gave us crucial information regarding the thematic resolution and precision of the Urban Atlas database.

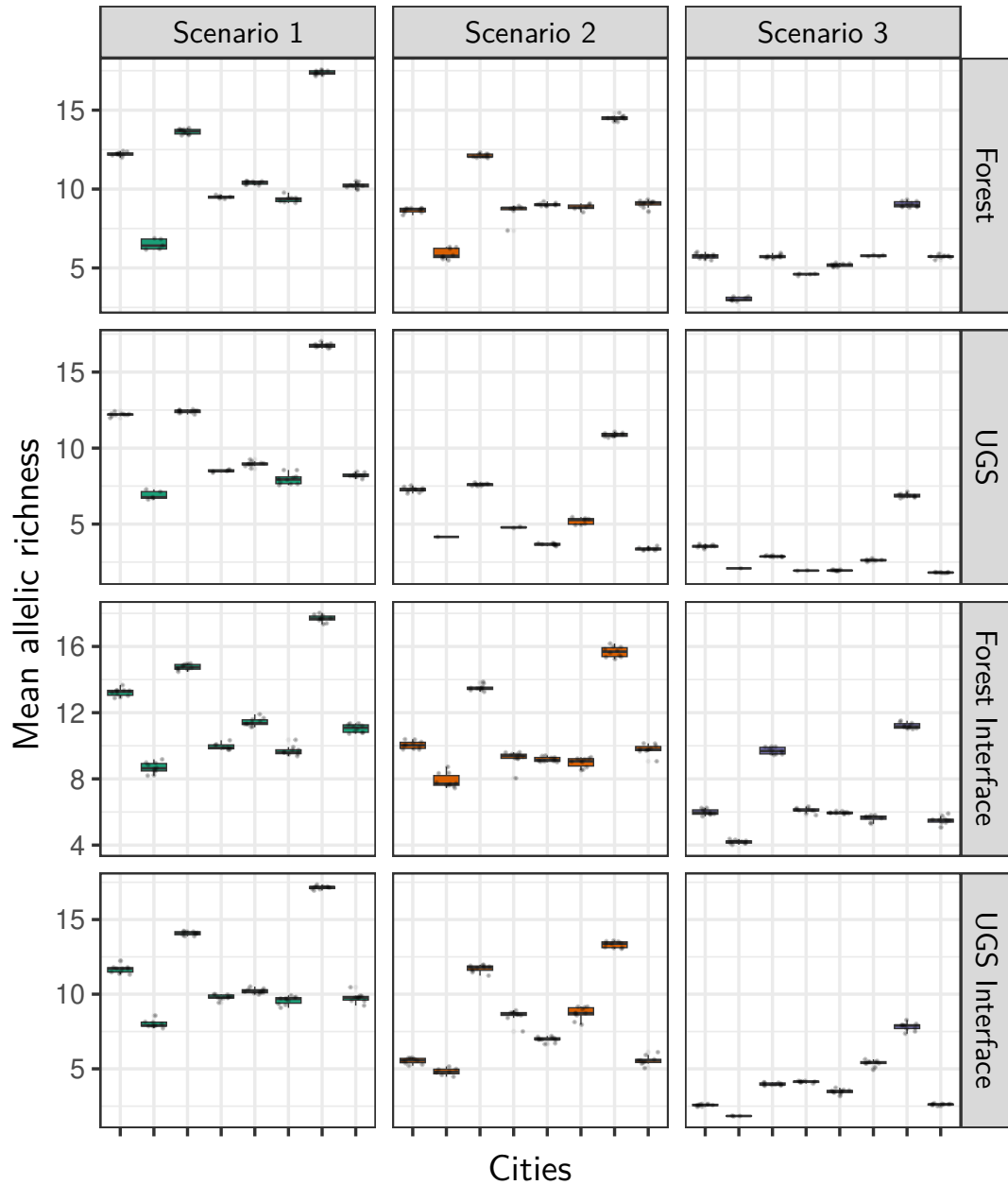


Figure R4: Distribution of the mean allelic richness of populations located in "Forest", "Forest Interface", "UGS" and "UGS Interface" patches in 8 cities for the three dispersal cost scenarios, after 250 generations. Each simulation was performed 10 times in each city for each scenario (n=10 values per box). The x axis distinguishes the 8 cities. The panels separate the scenarios vertically and the habitat types horizontally. Within-box variations represent the variability of the mean allelic richness in each habitat type from one run to another for a given city and scenario.

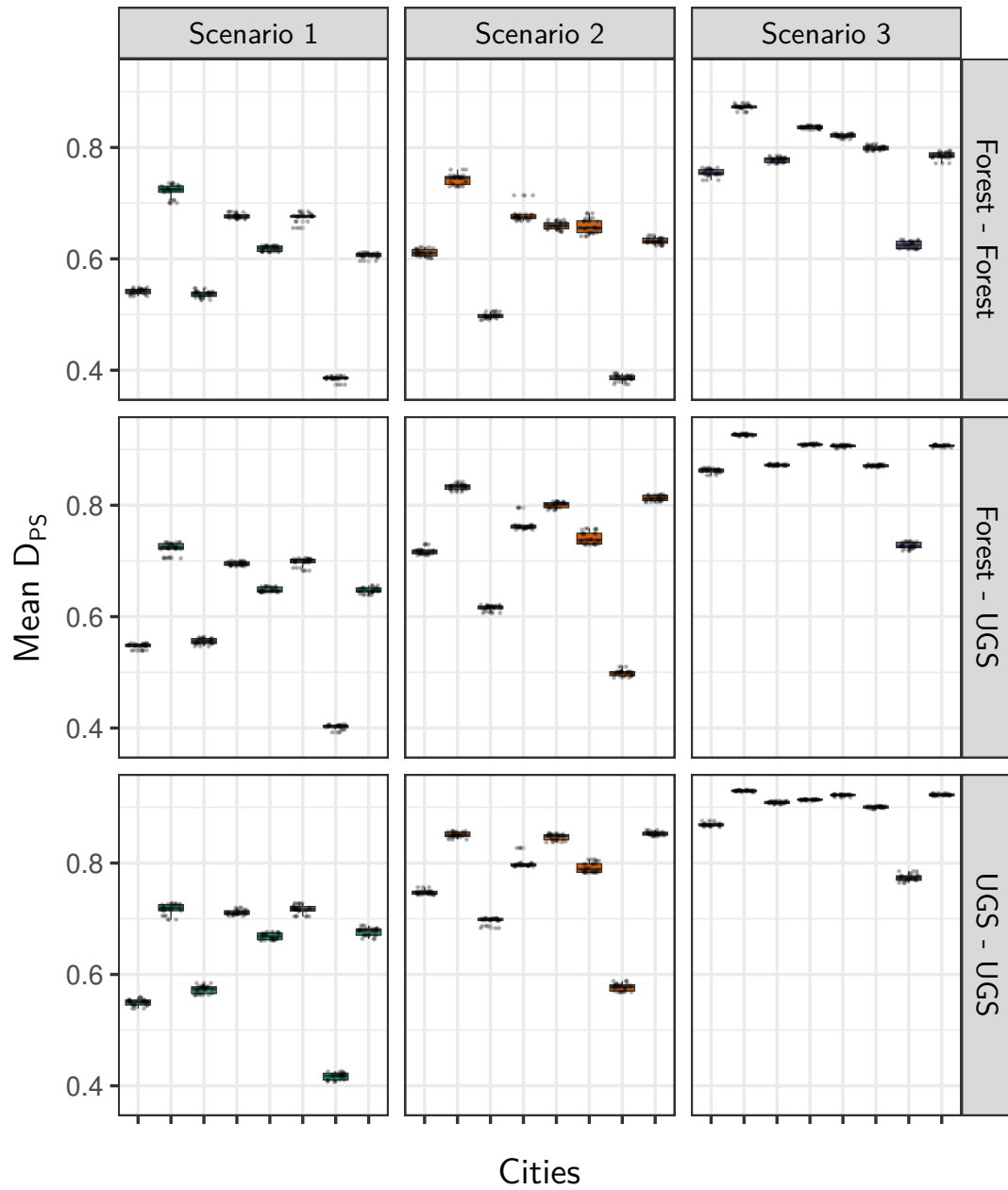


Figure R5: Distribution of the mean genetic differentiation (DPS index) among forest patches (Forest.Forest), forest and UGS patches (Forest.UGS) and UGS patches (UGS.UGS) in 8 cities for the three dispersal cost scenarios, after 250 generations. Each simulation was performed 10 times in each city for each scenario (n=10 values per box). The x axis distinguishes the 8 cities. The panels separate the scenarios vertically and the types of population pairs horizontally. Within box variations represent the variability of the mean genetic differentiation between each type of population pairs from one run to another for a given city and scenario.

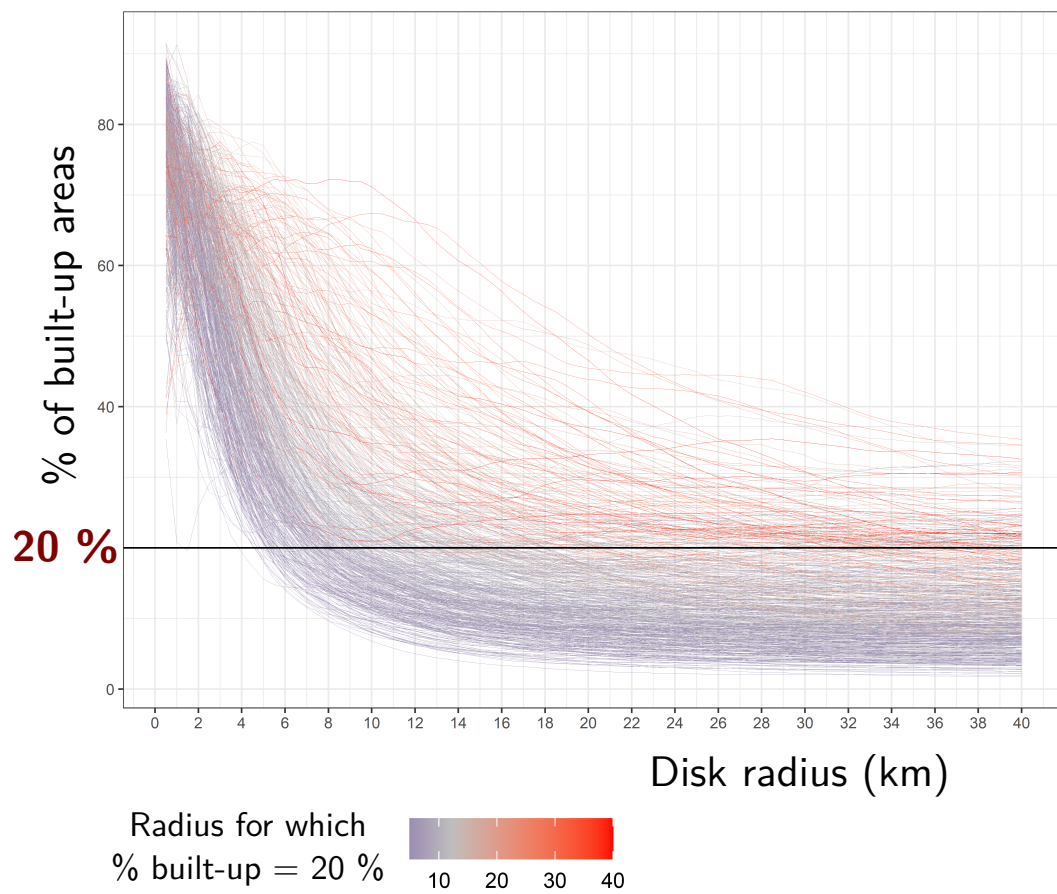


Figure R6: Proportion of built-up areas at an increasing distance from the city center of the studied cities. The horizontal line represents the 20 % built-up areas cut-off we used for delineating city areas.

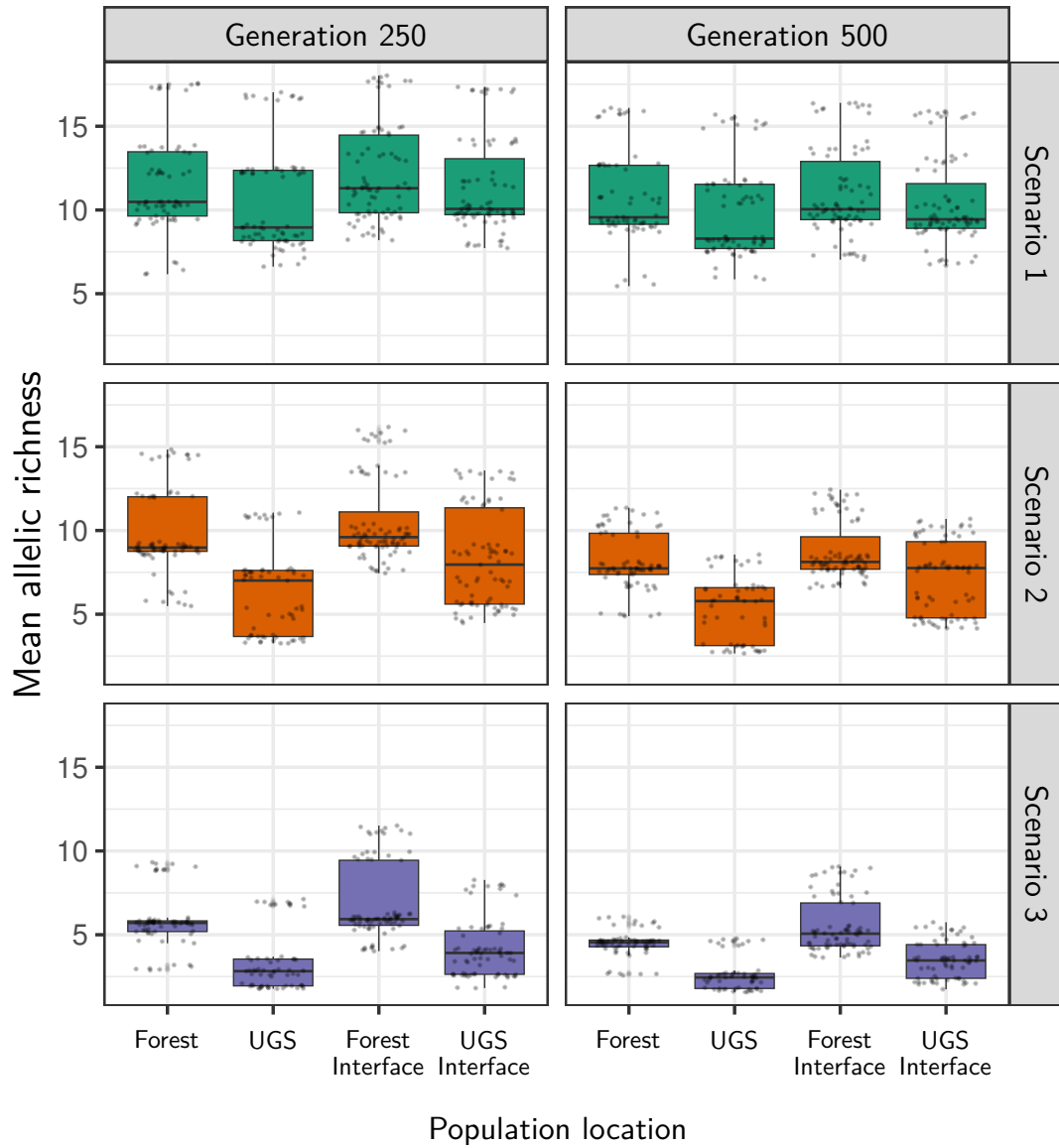


Figure R7: Variations in allelic richness among types of populations at generations 250 (left panel) or 500 (right panel). The results are shown for 8 cities and 10 runs per cost scenario.

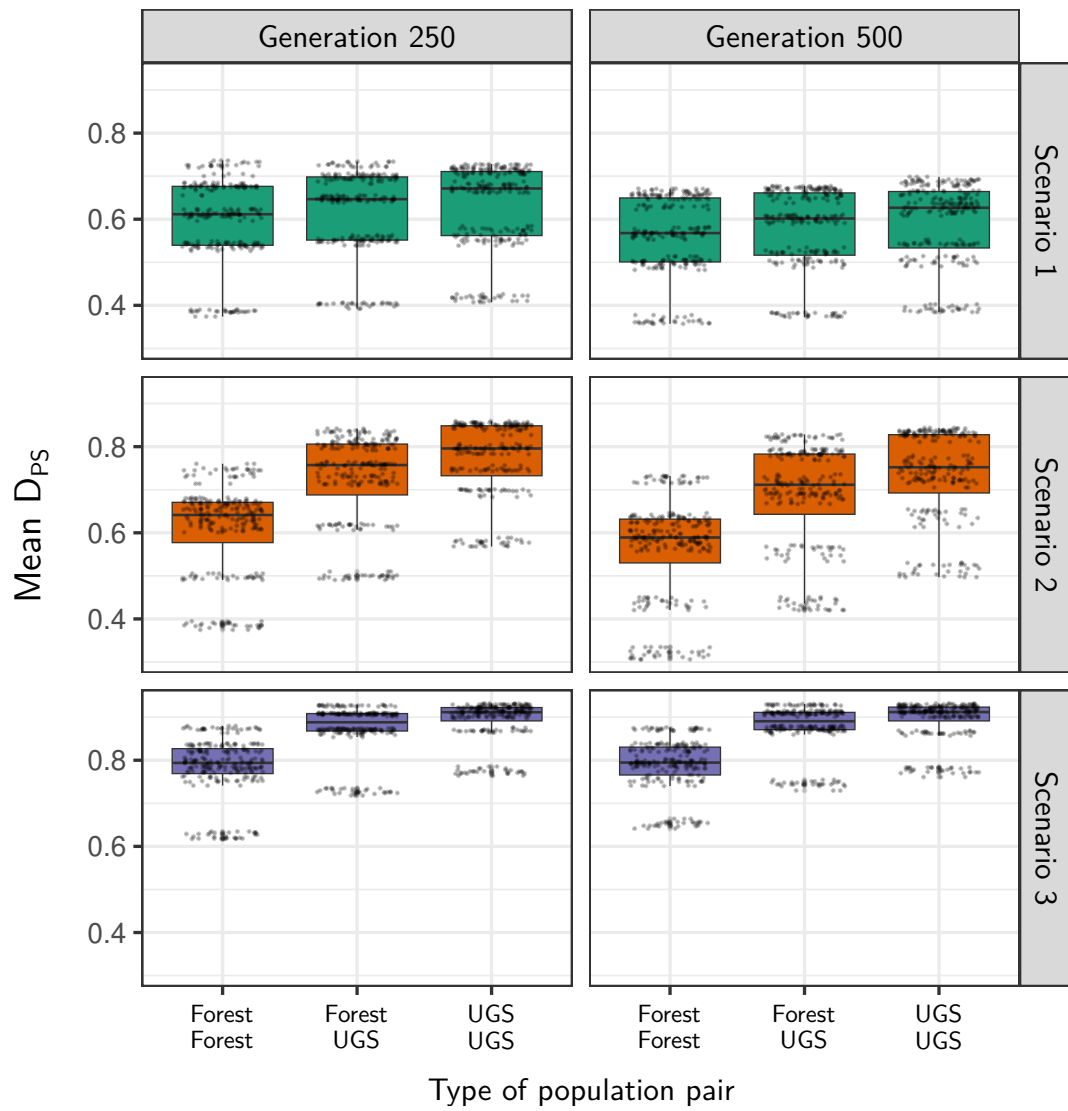


Figure R8: Variations in genetic differentiation among types of pairs of populations at generations 250 (left panel) or 500 (right panel). The results are shown for 8 cities and 10 runs per cost scenario.

Review by anonymous reviewer 2, 05 Jan 2024 22:37

The reviewed manuscript, "How does dispersal shape the genetic structure of animal populations in European cities? A simulation approach", seeks to understand the role of gene flow and broad-scale landscape effects on the genetic outcomes and population trajectories of urban tolerant wildlife species. The authors present an impressively large simulation study, which differs from many theoretical landscape genetic studies by using empirical spatial data representing 325 cities across Europe on which they seeded random genetic variation and simulated migration patterns across three cost scenarios. The analyses provide important insights into the relative role of gene flow vs. genetic drift and provide actionable advice for the conservation management of urban wildlife populations. The writing, methodology, and results are all of very high quality with few major issues.

We are very grateful to the positive evaluation of our study by the Reviewer 2 and thank them for the opportunity to improve our manuscript and discuss important points in our responses to the comments below.

Important issues:

Two important simplifications were used within the presented analyses, which may have biased analyses. First, landcover classes were grouped and simplified, treating all types of built urban areas as "artificial areas", including low and very low density urban fabric. While urban wildlife may prefer dispersing through continuous green spaces (forest or UGS), built urban areas likely provide supplementary dispersal pathways that are ignored in this analysis. Considering the wide-reaching interpretation about the permeability of urban landscapes, I would like to see results for cost scenarios in which the built environment is treated in a more nuanced way, perhaps with artificial areas broken up into moderate and high cost areas (e.g. with low-density areas given a moderate cost) rather than the entire built environment treated uniformly as high cost. While the "Limitations" section does mention the lack of detail here, citing the need for finer-grained resolution data ignores the presence of additional detail within current spatial data that was ignored through study design. Second, the focus on dispersal via the shortest or least cost path simplifies the simulated dispersal behavior and reduces the importance of patch pairs that may have multiple similar dispersal routes. Circuit-theory approaches would better account for the presence of these more diffuse dispersal networks. I don't necessarily think that these simplifications negate or discount the presented results, but I would like to see more discussion of these aspects in the discussion section.

R2.1: We agree with the Reviewer 2 that these two points (i.e., land cover thematic resolution and dispersal distance algorithms) are critical for the connectivity analyses underlying our simulation framework.

First, the land cover data we used (i.e., Urban Atlas database) make a distinction between what we classified as artificial areas and roads. The former includes different types of urban fabric, as relevantly pointed out by the Reviewer 2, whereas the latter includes roads, streets and other areas associated with transport infrastructures. Given this classification and the fine spatial resolution of the raw data and of the raster maps created from them (5 m), all the streets, parking lots and other sealed but not built-up areas creating interstices in the urban matrix were distinguished from the buildings themselves. Considering a slightly lower cost for the latter areas than for the artificial areas in all three cost scenarios (90 vs 100, 900 vs 1,000, 9,000 vs 10,000, as per Table 1) ensured that the most favorable paths within the urban matrix could be followed, as soon as the virtual species was able to venture into a dense urban fabric. We therefore agree with the comment of Reviewer 2 regarding the potential movement paths provided by more permeable areas within cities, and we believe that our scenarios included these potential paths in a realistic way.

Second, choosing between least-cost path algorithms and algorithms based on the circuit theory is always a critical modeling choice for connectivity analyses. Both approaches have their pros and cons,

in terms of computation times, realism, and their respective empirical support has not yet been resolved. We provide below the reasons why we chose least-cost paths over circuits for the present study.

- Applying a cost scenario to compute dispersal distances among habitat patches amounts to modeling how animal movement behaviors incur a distorted perception of Euclidean distances associated with movements across heterogeneous landscapes. Depending on the objective of the modeling, the choice of an algorithm over another does not have the same consequences. On the one hand, if one has empirical data related with dispersal and wants to identify the scenario best reflecting species movement behavior, it is crucial to make the correct decision. If animal dispersal is somehow informed, or if the studied patterns depend strongly on the most frequent movements, using least-cost path algorithms is relevant. If species dispersal commonly mimics random walks, or is known to involve multiple sub-optimal paths, then Circuitscape is a better solution. This will ensure that the subsequent cost inferences from empirical data (or conservation decisions based on spatial models) rely upon correct movement assumptions.

On the other hand, when one wants to understand how dispersal movement assumptions themselves could affect biodiversity patterns, the choice of the algorithm becomes less significant. We are in this latter case. Given that the same algorithm was used for all scenarios across all cities, our treatment (three distinct cost scenarios affecting relative distances among sites) was applied in a consistent way among replicates and allowed us to understand the dependence of genetic patterns on dispersal movement assumptions.

- Strictly speaking, our conclusions would only be valid for species whose dispersal only happens along least cost paths. However, we have good reasons to believe that similar variations in resistance distances from Circuitscape would have affected genetic patterns in similar ways. Whether we would have needed other cost contrasts to reproduce the same patterns with Circuitscape is a relevant question, though out of our scope.

- Computing cost-distances among every pairs of habitat patches across 325 European cities at a spatial resolution of 5 m and for 3 cost scenarios was already computationally intensive and required the use of a server with multiple-core processors. The same computations with Circuitscape would have been even more time-consuming.

Given the use of true urban landscapes within the simulations, the authors are in a position to provide valuable resource to municipal and national institutions seeking to improve wildlife connectivity. If possible, the authors should seek to provide maps or quantitative descriptions of important habitat patches or dispersal routes/barriers at the individual city level. At a minimum, an offer should be made in the discussion to provide this information through private correspondence.

R2.2: We are very grateful to the Reviewer 2 for this very encouraging comment. Actually, the overall project from which the present study originates included a more applied work package which gave rise to fruitful exchanges with biodiversity managers from the Greater Paris region. We replicated the same analyses in 5 administrative entities of this French region (Saint-Germain-en-Laye, Aulnay-sous-Bois, Gennevilliers, Marne et Gondoire, Paris Vallée de la Marne), at a finer spatial resolution (2 m), and involved the professionals in both the design and interpretation of the outputs. The result maps provided insights into potential biodiversity conservation measures in these territories. Although not publicly available, several technical reports have been issued and could be transmitted upon agreement of the professionals involved. A webinar was organized at the end of this study, where several professionals were invited. It is available online: <https://youtu.be/PdSIEgtghZY>

Minor issues noted:

L20: the claim that simulations "reproduced empirically observed results" is not supported with

quantitative analysis, so this phrase is misleading. The discussion provides extremely brief claims that outcomes mirrored several cited publications but provides no visual or quantitative comparison for the audience. Either include more explicit comparisons in the text, or soften this language.

R2.3: We thank the Reviewer 2 for this suggestion. We nuanced our conclusions (L20-21).

L44: "timely needed" ...missing comma or remove one word.

R2.4: We removed one word (L44).

L73: populations must not be "well connected" to facilitate genetic exchange. Moderately or even poorly connected populations can allow gene flow, and only minor levels of gene flow are needed to prevent allelic dropout or intense differentiation.

R2.5: Absolutely. We removed "well connected" in this sentence (L72-73).

L77: considering the citation of specific gene flow/drift scenarios from Hutchison & Templeton in the discussion, a more in-depth description of the theoretical expectations noted here is warranted.

R2.6: We thank the Reviewer 2 for this opportunity to describe at an early stage of the manuscript the theoretical pillars of our predictions and subsequent interpretations. We added a section describing the archetypal genetic patterns introduced by Hutchison & Templeton (1999) (L79-86).

L82: double closed parentheses

R2.7: We apologize for this formatting error, due to the use of a LaTeX template.

L90: define UGS in this first use of the acronym in main text

R2.8: We defined the acronym (L96).

L122: "forest species"... should this be "forest-dwelling"

R2.9: We included this suggestion (L125).

L130: define OECD

R2.10: We defined OECD as Organisation for Economic Co-operation and Development (L133-134).

L136: given the importance of UGS vs Forest patches throughout the study, a more detailed description of UGS is warranted. Are there canopy cover cutoffs that would make an interior patch into a "forest". Are there more detailed descriptions of the habitat available?

R2.11: In hindsight, we acknowledge that the description of UGS was not sufficient. We included the following sentence in the main text to address this issue: "UGS included public green areas predominantly dedicated to recreational use (e.g., gardens, parks), as well as suburban natural areas that have become and are managed as urban parks." (L140-142).

L145: "led differences" edited to "led to differences"

R2.12: This sentence was indeed confusing and we rephrased it (L149).

L266: "for every 20 loci", does this indicate that there are a total of 20 loci. It is unclear how many loci are being simulated.

R2.13: We indeed simulated a total of 20 loci and clarified this point (L271).

L377: provide examples of what is meant by "high" and "very high" allelic richness, particularly in relation to the starting genetic diversity.

R2.14: This sentence was rephrased (see R1.35 and L382-384).

L472: the claim that the individuals "range" does not change seems inaccurate. IF the total endured cost stayed the same but the cost per cell increased, then the "geographic range" of the dispersers would decrease. Perhaps clarification on what is meant by "range" is warranted here or earlier in the text.

R2.15: We understood the ambiguity associated with the term "range" here and replaced it by "the effective distance they could cover" (L478).

L478-485: comparisons to empirical data are extremely simplified and provide little detail. I would like to see more explicit descriptions of the species involved, the empirical landscape being compared to, and the findings of those papers. The citation of only two empirical studies is underwhelming given claims made in the abstract.

R2.16: We understand the concern of the Reviewer 2. We tuned down the abstract as suggested (R2.3). Several references are cited in the introduction as well, and we here refer the readers to two review articles (Fusco et al., 2021 ; Miles et al., 2019) assessing variations in genetic patterns across urban areas.

References

Baayen, R. H., Davidson, D. J., & Bates, D. M. (2008). Mixed-effects modeling with crossed random effects for subjects and items. *Journal of memory and language*, 59(4), 390-412.

Clarke, R. T., Rothery, P., & Raybould, A. F. (2002). Confidence limits for regression relationships between distance matrices: estimating gene flow with distance. *Journal of Agricultural, Biological, and Environmental Statistics*, 7, 361-372.

Fusco, N. A., Carlen, E. J., and Munshi-South, J. (2021). Urban landscape genetics: Are biologists keeping up with the pace of urbanization? *Current Landscape Ecology Reports*, pages 1–11.

Miles, L. S., Rivkin, L. R., Johnson, M. T., Munshi-South, J., and Verrelli, B. C. (2019). Gene flow and genetic drift in urban environments. *Molecular Ecology*, 28(18):4138–4151.

Savary, P., Foltête, J.-C., Moal, H., Vuidel, G., & Garnier, S. (2021). Analysing landscape effects on dispersal networks and gene flow with genetic graphs. *Molecular Ecology Resources*, 21(4), 1167-1185. <https://doi.org/10.1111/1755-0998.13333>

Savary, P., Clauzel, C., Foltête, J.C., Vuidel, G., Girardet, X., Bourgeois, M., Martin, F.M., Ropars, L. & Garnier, S. Multiple habitat graphs: how connectivity brings forth landscape ecological processes. Submitted to *Ecological Applications*. Preprint available: <https://doi.org/10.32942/X2CP66>

White, J. W., A. Rassweiler, J. F. Samhuri, A. C. Stier, & C. White. (2014). Ecologists should not use statistical significance tests to interpret simulation model results. *Oikos*. 123, 385-388.