

# PCI Reviews & Suggested Edits:

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

**a nice work in progress project with future upside**

---

Dear authors,

This is my first contribution handled through PCI, please apology the long handling time while figuring out how the system and the ideas behind PCI fully flesh out.

At first I was quite surprised to see that this is not a preprint per se but rather an outline of what eventually will become a nice study. I got thrilled on it as I really like the idea to embark an external board of reviewer that accompany a project already at the early stages. I think this is tremendously helpful, especially if one is trying to investigate new questions or using unfamiliar tools. It means more work for us though as I am now expecting an iterative process where the same team of reviewer will revisit your work to see how you have been doing. So it's far from a simple review what I was initially expecting but is rather a journey we take together - which I actually like the more I think about it.

With all that being said, you will find enclosed two reviews of two highly trusted peers of mine and I think they provide really good feedback what will help your envisioned research. Even though there might be some "make it or break it" issues I think they could be of help to navigate around them at that early stage of your work. E.g. considering some simulation work to frame null models against which your findings could be compared (like in anything related to phylogenetic relationships/distances).

I hope you will be able to answer the excellent points provided by the reviewers.

All the best

Jan

by *Jan Oliver Engler*, 2021-03-04 16:38

Manuscript: <https://doi.org/10.17605/OSF.IO/2UFJ5>

Thank you Jan for reviewing this preregistration. We found the insight from yourself and two reviewers extremely helpful! We provide our responses to each comment in blue (in the same way our text appears here), as well as a 'track changed' document showing the revisions and finally, a clean document to show the current version.

We also wish to apologize for our delay in submitting this revision and rebuttal. The lead author is an undergraduate and was busy with the spring semester coursework, in addition to strains caused by the COVID-19 pandemic on family life.

In response to your feedback regarding the consideration of simulation work to frame the null hypotheses: 1) We have reframed the hypotheses to better identify what we think is a true null (i.e., the acoustic adaptation hypothesis), and edited our approach to include accounting for phylogenetic history and to integrate the ensemble model explicitly with vocal signal characters. 2) Instead of reproducing simulations, we introduce a new *Background* section to refer the reader to Hudson and Creanza (2021) who modeled the evolution of birdsong as a functional trait in oscine birds. Although Hudson and Creanza (2021) focus on birdsong, we believe the simulation of vocal skill with respect to meaningful fitness differences (including survival) to also apply to alarm calls. The results of these simulations will help us interpret patterns in tone, range, pitch and vocalization complexity from our research.

We thank you sincerely,  
The Authors

### Review by **Guillermo Fandos**, 2021-03-01 12:59

Bird calls may provide crucial information on the effects of geographic isolation between subspecies or populations and may reflect local adaptation to diverging ecological conditions. Studies of variation in learned vocalisations are a promising complementary tool to traditional ones for identifying potential dispersal barriers. To this end, the authors provide a well-designed and hypotheses-driven approach that could yield exciting insights into the mechanisms that govern range expansion in *Q. mexicanus* subspecies. I find it very interesting. Nevertheless, I have some recommendations and concerns about the methods and statistical analysis to fully use the exciting outcomes.

Thank you for your enthusiastic and supportive review! Please find our responses to each of your comments in blue (in the same way our text appears here), as well as a 'track changed' document showing the revisions and finally, a clean document to show the current version.

First, in oscine Passerines, learning influence vocal development. Thus, vocal variation may be assigned to cultural differences that occur among different populations. The wide range of vocalisation in *Q. mexicanus* is likely the evolutionary result of the intense courtship and sexual selection behaviours. However, alarm calls development could be driven by different selective pressures that determinate a homogenisation between subspecies. For instance, the similarity in alarm call structure among species and subspecies might facilitate predator detection. I find that the manuscript will benefit from further explaining the implications of analysing alarm calls and not other

social calls that could drive a better picture of how social communication drives differences in the range expansion.

Thank you for this very thought-provoking comment. We initially wanted to only look at non-breeding vocalizations because of the variation in latent variables that could affect vocalizations during the breeding season. These include things such as the number of females per male, the stage of reproduction (i.e., whether or not females are fertile and may be mate-guarded), or whether there are neighboring males. During the COVID-19 pandemic, grackles were more difficult to locate to record their non-breeding vocalizations and so we elected to extend our sampling to a second year and include vocalizations during the breeding season. This necessitates accounting for 'year' and adding exemplar breeding season song files for conducting the playback experiment during the breeding season. We updated our 'Data collection procedures' to accommodate this change.

To help us distinguish expected outcomes of the alternate hypotheses and implications to range expansion we re-organized our hypotheses and predictions and now provide a new *Background* section. We point the reader to Hudson and Creanza (2021) who modeled the evolution of birdsong as a functional trait in oscine birds. We hope that these two revisions paint a better picture of how social communication could drive differences in range expansion among *Quicalus mexicanus* subspecies.

Second, although authors already state that disturbed habitats could influence the range expansion in some of their hypotheses, it would be necessary to provide further background and discussion about the role of urbanisation (as specific disturbed habitat) in these range expansion differences heterogenous intraspecific social communication. Some studies point out the importance of urbanisation in explaining the northward expansion differences between subspecies (Christensen, 2000; Wehtje, 2003). In that sense, urbanisation can influence range expansion, and at the same time, drive differences in vocal communication (high degree of behavioural flexibility) since they are found at much higher densities in urban areas than anywhere else, less predation and higher availability of food resources (Christensen, 2000). I suggest to the authors to include urbanisation variables to control differences in the alarm calls and in the ensemble distribution modelling approach and discuss further the urbanisation role in their hypothesis.

Thank you for this comment. We agree and understand the predicted influences of urban noise on vocalizations. We now include a Global Human Modification of Terrestrial Systems data layer in the ensemble model approach to account for metrics of urbanization (in *Large scale habitat data* section) as well as a priori explanations of how urbanization could affect vocalizations (in the *Background* section). We do not, however,

have predictions for how differences in vocalizations driven by urban noise could affect range expansion. This is because selection could act to promote the most effective signals (in either alarm or song) within any soundscape (i.e., following the acoustic adaptation hypothesis - our null hypothesis). We will plan to interpret our findings following our specific predictions and offer posthoc potential explanations if interesting results are found at the intersection of urbanization and vocalizations among *Q. mexicanus* subspecies.

Third, I find a lack of detail in the ensemble distribution modelling approach. Probably, authors will need another step and metrics for niche comparison between subspecies (e.g. niche overlap; Broennimann et al. 2012). In that sense, how they are planning to include the background information in the modelling approach, what algorithms they are planning to use?. For instance, a specific background for each subspecies (this could result in an overfitting model for each subspecies). Besides, niche models and differences between subspecies niches could be influenced by different sampling effort, and observation bias between subspecies ranges that could contain eBird data. Western states are better sampled and contain more complete lists than Central USA. Following similar reasoning, differences between generalist and specialist could be drawn by limitations and sampling biases (different realised niche between subspecies but the same fundamental niche).

Thank you for pointing this out. We agree that the point about eBird data having known drawbacks should be discussed and added text to explain how we plan to minimize the effects of this bias in our results. To accommodate the potential problems that you identified here, we plan to randomly select the same number of eBird lists for each subspecies with the intention of applying the same proportion/probability of model overfitting to each subspecies. An alternative dataset to eBird could be georeferencing locality data from museum databases, but we don't think this is the best approach since those data have large spatial error and could introduce unnecessary noise in our predictive distribution modelling approach. We also plan to remove all eBird records for *Q. mexicanus* from our background points layer in our ensemble model to additionally minimize model overfitting.

We added more detail to our explanation of the ensemble distribution modelling approach. We clarify that the regional models will be set so that each subspecies represents a region allowing differences between regions to be identified.

The point concerning our hypothesis about specialists compared to generalists is a good one. We have elected to remove this line of reasoning because of the potential for a circular argument where we provide no other lines of evidence besides our data and eBird data that could be biased in the ways you describe.

Finally, I found a lack of explanation of the integration between the alarm call experiment and niche modelling approach. How will be tested if alarm call heterogeneity is correlated with niche differences or heterogeneous landscape? Maybe further integration with other tools (e.g. genetic).

Thank you for bringing this to our attention. Initially we were using these two methods to test separate hypotheses explaining the differences in range expansion (i.e., original H1: social communication versus Alternative 1a: habitat availability). This comment has helped us to clarify that both approaches can test the 'acoustic adaptation hypothesis'. We have included additional text and references in the Abstract to clarify this, in addition to reframing our hypotheses to reflect how both approaches test each hypothesis. We also elected to account for the role of evolutionary history by turning two of our analyses into phylogenetic analyses (i.e., PPCA, and PGLS).

#### Specific comments

- While I am no expert about call analysis, authors should include information about the microphone distance to the target individual since it could influence some of the experiment's metrics.

Thanks for bringing this up. We added the distance from the bird in 'Data collection procedures' to emphasize that the tone, range, and pitch of recorded vocalizations are being collected from within a distance where sound quality is not expected to decay (i.e., 5 m away from the vocalizing individual).

- Could the close geographical distance of the proposed research areas from *Q. m. nelsoni* and *Q. m. monsoni* influence the playback experiment?

The geographical distance between *Q. m. nelsoni* and *Q. m. monsoni* is small, however, the two subspecies are not thought to come into contact during the non-breeding season. This may potentially be due to the Cochise Filter Barrier (Provost et al. 2021), which may have played a role in the biogeography of these two subspecies. We now include discussion on this matter in a new 'Study taxa' section and add the DeCosta et al. (2008) citation showing the distant relatedness predicted for these two adjacent subspecies.

· What the authors mean with “near wetlands” in the statement of the Alternative 1b hypothesis?. Do you plan to draw a buffer around the wetlands? In that sense, it would be necessary to include some dispersal information.

Yes we will draw a buffer around the wetlands by including pixels within the typical territory size from wetland edges (i.e., 5 kilometers). We now include this information in the *Predictors and Covariates* subsection.

## References

- Broennimann, O., Fitzpatrick, M. C., Pearman, P. B., Petitpierre, B., Pellissier, L., Yoccoz, N. G., ... & Guisan, A. (2012). Measuring ecological niche overlap from occurrence and spatial environmental data. *Global ecology and biogeography*, 21(4), 481-497.
- Christensen, A. F. (2000). The fifteenth- and twentieth-century colonization of the basin of Mexico by the great-tailed grackle (*Quiscalus mexicanus*). *Global Ecology and Biogeography*, 9(5), 415-420.
- Wehtje, W. (2003) The range expansion of the great-tailed grackle (*Quiscalus mexicanus* Gmelin) in North America since 1880. *Journal of Biogeography*, 30(10), 1593-1607.

Thank you.  
Sincerely,  
The Authors

## Review by anonymous reviewer, 2021-02-05 11:27

Dear authors,

You deal with an interesting study topic. Range extensions certainly deserve more attention and I like your very integrative approach. Your study concept seems to be thoughtful and well organized. I mainly have some smaller concerns but I assume that you uploaded your concept to PCI Ecology in

order to look for an “advocatus diaboli”. Hence, I also want to stress some larger concerns dealing with study design in general. From my point of view this is not only a very interesting experimental study dealing with plenty of possible predictors and pretty rigorous hypothesis testing but also a comparative study.

Thank you for your “advocatus diaboli” review! Please find our responses to each of your comments in blue (in the same way our text appears here), as well as a ‘track changed’ document showing the revisions and finally, a clean document to show the current version.

1) Comparative studies deal conceptually with coalescence and are typically not feasible with such a small number of species/taxa. Differences among the three subspecies could be based on their common ancestry, i.e. two taxa are almost certainly more closely related to each other than to the third group. Any differences among them could be “simply” related to their past history. Typical comparative studies deal with rather taxon-rich phylogenies in order to figure out evolutionary driving forces. It seems hard to study this in a three-taxa system. I am fully aware that literature is full of comparisons among few (perhaps even between two) species (or subspecies or populations) and that far-ranging conclusions are inferred from this kind of studies but actually, great caution is warranted. I am not sure that you will be able to fully reject any null hypotheses if one keeps this in mind. This also relates to my next concern.

Thank you for pointing this out. As we are familiar with how to conduct comparative phylogenetic analyses, we have elected to include this as part of our approach to this study. We will now conduct a phylogenetic principal components analysis (PPCA) and phylogenetic generalized least squares (PGLS) to identify whether differences between vocalizations are more or less than expected given the relatedness of each taxon. These analyses replace our original PCA and ANOVA. We think that this revision will greatly enhance our final product and are grateful that you brought it up (even though we still only have 3 taxa in this project)!

2) You look for differences among three groups but these groups are spatially clustered – the three taxa occur (necessarily for subspecies) in three different geographical areas. Although you check for biological meaningful variables associated with different areas (habitat), this does not exclude any other random factor associated with a west-east gradient (e.g. longitude per se or even any other absurd factor could not be excluded due to inherent problems of study design). This argument is similar to the first concern. I do not know if you follow my argumentation. The problem is that there is no true solution for these issues (theoretically, you could check for relationships among subspecies) but you should at least be aware of the problem. I have some additional minor concerns. It should be more clearly stated if you interpret possible differences in bioacoustics as consequence of different social behavior or as plastic (or evolutionary) consequence of to the environment. As far as I understood, you tackle all these aspects but this needs to be distinguished right at the beginning of the outline. Note, that calls are generally believed to be genetically determined (I do not know if this

has been shown in your study species, too?). Note also, that any differences in call parameters can be simply due to chance (e.g. based on a Brownian motion model often assumed in character evolution and actually a kind of null model against one should test evolutionary changes, see main concern 1 above). I am pretty sure that you are aware of this and playback experiments are a great idea to tackle this issue. Also note that there are size differences among subspecies. Song frequency is necessarily heavily influenced by size and differences in pitch can be simply based on size (environmental or social factors might be actually irrelevant in this case).

Thank you for your attention to detail here. We have now added an explanation in a new *Background* section that identifies that vocalizations have both genetic and learned aspects and so we conduct phylogenetically constrained analyses to account for evolutionary history. To account for size differences among subspecies, we will measure museum specimens of each subspecies to use a covariate in comparative phylogenetic analyses. We now state this in the Other Variables subsection under “7. Data collection procedures”.

Details Abstract Q. m. monsoni in italics

Thank you. We have now italicized this.

4 Hypotheses Alternative 1a Habitat availability is certainly a very crucial factor. Keep in mind that you have to consider the area reachable for the birds in a given time. The delimitation of the background is very important for any spatial modelling and these considerations could be added to your outline.

Thank you for this comment. We added more detail about our ensemble model approach including how we selected background points to avoid model overfitting. To improve our selection of background points, we also exclude pixels where the species has been seen from eBird.

7 Data collection procedures I highly recommend some training, keeping bird in your view (and record age+sex) + recording of background noise + actually recording and subsequently performing your experiment + covariates like weather etc. is not easy. I am pretty sure you already know that 😊 You use many covariates and this is certainly warranted. However, it seems reasonable to limit experiments and recordings to suitable weather conditions (no storm, no rain). For playback experiments always use uncompressed files (e.g. wav) as birds might not react to compressed mp3-files (I am sure you are aware of this, too but I once made this mistake myself when trying to map nightjars).

Thank you for this insightful point. We hadn't thought of this and now include a subsection called '*Training*' under the Data collection procedures section to address these points.

10 Stopping rule Your power analysis seems good to me but you also mention a few drawbacks. You are dealing with field data that only later might have to be excluded (e.g. because problems with the recording not immediately obvious). So, in practice, always add a few more recordings. Actually, the sample size is very ambitious, anyway! You should expect differences in sexes (e.g. due to size differences) and while you consider sex as a demographic variable you might have to make your analyses for males and females separately. Not that differences among sexes could be specific for each subspecies making comparisons particularly difficult.

Thank you for laying out this critical concern. We had already planned to record 5-10 vocalizations prior to conducting the playback experiment to limit problems with recording (e.g., background noise influencing recording quality), and now make that plainer in our new *Training* section. Part of our analysis in warbleR removes low signal-to-noise vocalizations from analyses which we now state in the '*Describing Vocalizations*' subsection. Because we started collecting data after submitting the initial preregistration, we were able to see that it is almost exclusively the males that alarm-call. Thus, we now only use males in our study.

12 Measured variables Note that AUC has been heavily criticized in the literature (e.g. Lobo et al. 2008, Global Ecol. Biogeogr.). It might make sense to add an additional value for model evaluation (e.g. Kappa). Variable contribution is a big task in spatial modelling. I am not familiar with the approach by Curry et al. 2018 but most approaches (like Maxent) are not built for this task Which scale do you chose for co-occurrence? Not a very big problem but you should select an appropriate pixel/grid cell size for co-occurrence, e.g. do birds that occur in the same 1x1, 5x5 or 10x10 km grid actually influence each other (the smaller the better)?

Thank you for bringing this to our attention. We were aware of the criticism of AUC but had not considered alternatives. We now include the root mean square error (RMSE) and AUC used by Curry et al. 2018 to evaluate model performance. All data sets (occurrence, co-occurrence, and geospatial environmental) will be set to the largest pixel size which is 30 m resolution. This is now stated at the beginning of the '*Measured variables*' section.

15 Blinding What about the third population? I am a bit nitpicking but do not name the subspecies according to US federal states. The birds are more widely distributed (e.g. in Mexico) and for outsiders this is more difficult to follow than scientific names.

Thank you for pointing out this inappropriate shorthand that we used. We have now edited this section to state the subspecies names instead of the sampling locations organized by state. For clarification here: the first author will record *Q. m. nelsoni* and *Q. m. monsoni* and the second author will analyze that data. The second author will record *Q. m. prosopidicola* and the first author will analyze that data.

18 Ensemble Distribution Models Did you consider methodological alternatives? As far as I understood Curry et al. 2018 is not a typical ensemble approach as they do not rely on different modelling algorithms like e.g. Biomod. I do not want to criticize Curry et al. 2018 but recommend selecting the approach most suitable for your data and Biomod or Maxent might be alternatives (which you might simply reject for a good reason). Given that modelling is very complex, this aspect of your outline could be a bit more detailed.

We appreciate this comment and now provide some more justification for choosing the Curry et al. 2018. Specifically, the ensemble model approach by Curry et al. 2018 enables comparing regions in the same model to identify differences across space. We will use this to identify whether there are regional (subspecies) differences in the importance of habitat variables and vocalizations. Other 'ensemble' approaches compare different types of SDMs to identify and use the output from the best model.

I look forward to see your first results being published.

Kind regards

Thank you.  
Sincerely,  
The Authors