Refining analyses of existing datasets is valuable for macrogenetics: a response to Paz-Vinas et al. (2021)

Katie Millette[1], Vincent Fugère[2], Chloé Debyser[1], Ariel Greiner[3], Frédéric Chain[4], and Andrew Gonzalez[1]

[1]McGill University
[2]Université du Québec à Trois-Rivières
[3]University of Toronto
[4]University of Massachusetts at Lowell

February 11, 2021

Abstract

Paz-Vinas et al. (2021) comment on methodological and data-related limits of our paper (Millette et al. 2020), which affect a small proportion of our datasets and analyses. These points do not refute our conclusions. We address their comments and support the call for the development of best practices for future macrogenetics research.

Title: Refining analyses of existing datasets is valuable for macrogenetics: a response to Paz-Vinas et al. (2021)

Authors : Katie L. Millette[1*], Vincent Fugère[2], Chloé Debyser[1], Ariel Greiner[3], Frédéric J. J. Chain[4] and Andrew Gonzalez[1,5]

[1]McGill University, Department of Biology

[2]Université du Québec à Trois-Rivières, Département des Sciences de l'Environnement

[3]University of Toronto, Department of Ecology & Evolution

[4]University of Massachusetts at Lowell, Department of Biological Sciences

[5]Quebec Centre for Biodiversity Science

***corresponding author** : Katie L. Millette, Department of Biology, McGill University, 1205 Docteur Penfield Ave, Montreal QC, Canada, H4V 2N2

Tel: 514-398-1622; email: millettek@gmail.com

Author emails:

Katie L. Millette: millettek@gmail.com

Vincent Fugère: Vincent.Fugere@uqtr.ca

Chloé Debyser: chloedebyser@gmail.com

Ariel Greiner: ariel.greiner@mail.utoronto.ca

Frédéric J. J. Chain: frederic_chain@uml.edu

Andrew Gonzalez: andrew.gonzalez@mcgill.ca

Statement of authorship: All authors contributed to the ideas and writing of the manuscript.

Data accessibility statement: No new data is used.

Key words: macrogenetics, population genetics, intraspecific genetic diversity, intraspecific genetic variation, COI

Type of article: Technical comment (response to: ELE-01078-2020.R1 – Macrogenetic studies must not ignore limitations of genetic makers and scale)

Number of words in abstract: 52

Number of words in main text: 810

Number of references: 3

Number of figures, tables, and text boxes: 0

Abstract

Paz-Vinas et al. (2021) comment on methodological and data-related limits of our paper (Millette et al. 2020), which affect a small proportion of our datasets and analyses. These points do not refute our conclusions. We address their comments and support the call for the development of best practices for future macrogenetics research.

Main text

We thank Paz-Vinas et al. for developing aspects of our macrogenetics paper that we were not able to cover in detail. We agree with some of their messages and support the call for careful use of compiled data, particularly when repurposing it for new questions.

Macrogenetics (Blanchet et al. 2017) investigates large-scale patterns in intraspecific genetic diversity (IGD) across many taxa, as they relate to ecological processes and anthropogenic impacts. Macrogenetics is a fast-emerging field, so we hope this exchange with Paz-Vinas contributes to its development.

Our paper presented a global analysis of the effects of human population density and land use on patterns of IGD; the first (and most substantive) part of our paper focused on a spatial analysis and the second (using a much smaller dataset) presented a time series analysis. Paz-Vinas et al. comment primarily on the latter.

Paz-Vinas et al.'s comment raises some points we addressed in Millette et al. (2020) and present two issues that question the main finding of our paper captured by our title: "No consistent effects of humans on animal genetic diversity worldwide". Yet, nothing in their comment undermines our conclusions. We address the two major points of their comment below.

Are COI sequences the most appropriate data?

It is certainly imperfect to use COI sequences for population estimates of IGD. We gave several reasons in our paper (page 65), including the inability of mtDNA to reflect neutral evolutionary patterns, and the potential sensitivity of patterns to the minimum number of sequences per population (end of page 57).

We are aware of the inconsistent archiving practices among data-depositors (i.e., uploading only unique sequences to repositories) and we were careful not to overextend our conclusions because of this. Paz-Vinas et al.'s recommendation to use haplotype diversity/richness rather than pairwise nucleotide diversity is valid given the unknown sampling effort behind archived sequences.

As a community, we must decide whether these limitations preclude the use of COI in macrogenetics. No other marker offers such broad taxonomic and spatial scope. Studies using mtDNA reveal large-scale patterns in IGD with respect to latitude and human impacts (e.g., Miraldo et al. 2016); refining analyses of these existing macrogenetic datasets to confirm the existence of patterns is necessary to advance the field. There

is an opportunity to gain further insight into IGD patterns by complementing large-scale analyses with fine-scale population genetic estimates for data-rich taxa.

Are the spatial and temporal scales biologically meaningful?

Paz-Vinas et al. focus their critique on our time series analyses. Our analysis does conflate space and time to a certain extent, although we used the smallest possible spatial scale (1000 km) with sufficient data to test our hypothesis. We explain this (page 58) and make careful statements not to over-emphasize these results (i.e., no consistent effect of humans on animal IGD).

Whether 1000 km exceeds species' dispersal capabilities is species-specific. Geographic and anthropogenic barriers can further fragment species' ranges. Because there is no 'right' resolution, we introduced a multi-scale approach to assess IGD patterns at the global extent. Ideally, species traits, ranges and geological features should be considered in future studies.

Paz-Vinas et al. selected fish in their comment, which in general have unique dispersal abilities, and are the most 'noisy' taxa in our study as a result of watershed constraints. We did not account for watershed structure which is a shortcoming of our analysis, although fish populations constitute only 11.4% of our temporal dataset. The purely spatial analysis, that is the focus of our paper, assesses IGD patterns at smaller spatial scales (sequences < 10 km apart) where the risk of pooling sequences from isolated populations with divergent land use is reduced. As we have discussed in our Appendix, there is little to no spatial variation among sequences in the majority of populations: "*the mean great circle distance among sequences was 0 for ˜50% of species in our dataset, and usually < 100 km for other populations (Fig. S1)* ". Thus, the "daisy-chaining" issue is irrelevant for the majority of our analyses. We would have liked Paz-Vinas et al. to acknowledge that the issue they raise only applies to our time series analysis. Future macrogenetic studies should be aware of the consequences of aggregating sequences into populations.

Regardless of the marker(s) chosen, or method of forming biologically meaningful populations, Paz-Vinas et al.'s comment does not indicate either way (positive or negative) how we purportedly over-interpret and misinform readers on the effects of humans on IGD. Theory predicts positive and negative effects of humans on IGD, so inconsistent trends are not surprising.

We agree that macrogenetics will progress faster when researchers align behind a set of best practices for data, analyses, and the communication of uncertainties in our understanding of the patterns arising from data limitations. We look forward to seeing future analyses as the field tackles the problem of assessing human impacts on genetic diversity.

References

Blanchet, S., Prunier, J.G. & De Kort, H. (2017). Time to go bigger: emerging patterns in macrogenetics. *Trends Genet* ., 33, 579–580.

Millette, K.L., Fugère, V., Debyser, C., Greiner, A., Chain, F.J.J. & Gonzalez A. (2020). No consistent effects of humans on animal genetic diversity worldwide. *Ecol. Lett* ., 23, 55–67.

Miraldo, A., Li, S., Borregaard, M.K., Flórez-Rodríguez, A., Gopalakrishnan, S., Rizvanovic, M., et al. (2016). An Anthropocene map of genetic diversity. *Science* , 353, 1532–1535.