Ecology Letters, (2021) **24**: 1282–1284 doi: 10.1111/ele.13732

TECHNICAL COMMENT

Macrogenetic studies must not ignore limitations of genetic markers and scale

Ivan Paz-Vinas,[1,2*†] Evelyn L. Jensen,[3*†] Laura D. Bertola,[4†‡] Martin F. Breed,[5†‡] Brian K. Hand,[6†] Margaret E. Hunter,[7†‡] Francine Kershaw,[8†] Deborah M. Leigh,[9†] Gordon Luikart,[6†‡] Joachim Mergeay,[10,11†‡] Joshua M. Miller,[12†] Charles B. Van Rees,[6§] Gernot Segelbacher[13†‡] and Sean Hoban[14†‡]

Abstract

Millette *et al.* (Ecology Letters, 2020, 23:55–67) reported no consistent worldwide anthropogenic effects on animal genetic diversity using repurposed mitochondrial DNA sequences. We reexamine data from this study, describe genetic marker and scale limitations which might lead to misinterpretations with conservation implications, and provide advice to improve future macrogenetic studies.

Keywords

anthropogenic impacts, COI, conservation, Genbank, genetic data archiving, genetic diversity patterns, macroecology, macrogenetics, mitochondrial sequences, population genetics.

Ecology Letters (2021) **24**: 1282–1284

INTRODUCTION

Macroecology and conservation biology now include 'macrogenetic' studies that repurpose genetic data from public databases to explore patterns and drivers of intraspecific genetic diversity (IGD) across large taxonomic, spatial and/or temporal scales (Blanchet *et al.*, 2017). Millette *et al.*, (2020) conducted a macrogenetic study to elucidate relationships between human impacts and animal IGD, but technical limitations of their approach may prevent the detection of anthropogenic effects on IGD. Although the authors acknowledged several constraints, and presented their results with more nuance than previous macrogenetic studies (e.g. Miraldo *et al.*, 2016), issues remain that cannot be resolved or adequately addressed by tempering the interpretation.

ARE COI SEQUENCES THE MOST APPROPRIATE DATA?

Millette *et al.* used 175 247 mitochondrial cytochrome c oxidase subunit 1 (COI) sequences from 17 082 vertebrate species (average 10 sequences *per* species) deposited in BOLD and GenBank. COI is a popular marker for species molecular barcoding due to its low within-species and high between-species variation. However, these characteristics, coupled with others (Table S1), make COI inappropriate for measuring IGD, as Millette *et al.* acknowledge. Despite these well-known issues,

[1]*Laboratoire Evolution & Diversité Biologique, Centre National pour la Recherche Scientifique, Institut de Recherche pour le Développement, Université de Toulouse, UPS, CNRS, IRD, UMR 5174, 118 route de Narbonne, Toulouse 31062, France*
[2]*Laboratoire Ecologie Fonctionnelle et Environnement, Université de Toulouse, UPS, CNRS, INP, UMR 5245, 118 route de Narbonne, Toulouse 31062, France*
[3]*Department of Ecology and Evolutionary Biology, Yale University, 21 Sachem St, New Haven, CT 06520, USA*
[4]*City College of New York, 160 Convent Ave, New York, NY 10031, USA*
[5]*College of Science and Engineering, Flinders University, Bedford Park, SA 5042, Australia*
[6]*Flathead Lake Biological Station, 32125 Bio Station Ln, Polson, MT 59860, USA*
[7]*U.S. Geological Survey, Wetland and Aquatic Research Center, 7920 NW 71st St, Gainesville, FL 32653, USA*
[8]*Natural Resources Defense Council, 40 West 20th Street, New York, NY 10011, USA*
[9]*WSL Swiss Federal Research Institute, Zürcherstrasse 111, Birmensdorf 8903, Switzerland*
[10]*Research Institute for Nature and Forest, Gaverstraat 4, Geraardsbergen 9500, Belgium*
[11]*Aquatic Ecology, Evolution and Conservation, KULeuven, Charles Deberiotstraat 32, box 2439, Leuven 3000, Belgium*
[12]*Department of Biological Sciences, University of Alberta, Edmonton, AB, Canada*
[13]*Chair of Wildlife Ecology and Management, University Freiburg, Tennenbacher Str. 4, Freiburg D-79106, Germany*
[14]*Center for Tree Science, The Morton Arboretum, 4100 Illinois Rt 53, Lisle 60532, USA*
**Correspondence: E-mail: ivanpaz23@gmail.com*
**These authors contributed equally to the work.*
†Group on Earth Observation Biodiversity Observation Network (GEO BON) Genetic Composition Working Group.
‡Conservation Genetics Specialist Group, International Union for Conservation of Nature (IUCN), 1196 Gland, Switzerland.
§Group on Earth Observation Biodiversity Observation Network (GEO BON) Species Populations and Freshwater BONs.

the large availability of COI sequences has, nevertheless, resulted in its continued use to represent IGD in macrogenetic studies (e.g. Miraldo *et al.,* 2016; Millette *et al.,* 2020; Theodoridis *et al.,* 2020; Manel *et al.,* 2020).

Even if COI could provide a useful IGD measure, we identified a subtle, yet serious, constraint of repurposing publicly available mitochondrial data due to inconsistent archiving practices. Specifically, it is common for only unique or newly discovered haplotypes to be deposited in repositories, and not the study's full dataset. We screened 18 issues from a leading population genetics journal (*Molecular Ecology*; Table S2). Of the 40 papers that deposited mitochondrial sequences in GenBank, 22 deposited all sequences generated, while 18 deposited only novel haplotypes (sequences detected for the first time) or exemplars of each haplotype. Therefore, deposited data may more accurately represent haplotype accumulation curves across space and time, rather than comparable snapshots of genetic diversity. This bias compromises attempts to quantify temporal trends in IGD using COI sequences from GenBank, as in Millette *et al.,* (2020), and is a potential issue in many spatial macrogenetic studies. Macrogenetic studies should extract metadata regarding sample sizes and complete haplotype (or allele) frequencies from the original manuscripts (e.g. Lawrence *et al.,* 2019) to avoid bias from inconsistently archived data (Table S1).

ARE THE SPATIAL AND TEMPORAL SCALES BIOLOGICALLY MEANINGFUL?

Millette *et al.* examined IGD temporal trends across 909 animal species where COI sequences were available for ≥ 4 years. Sequences were grouped across ≤ 1000 km to avoid 'conflating spatial and temporal effects'. This scale far exceeds the dispersal capabilities of many included species, the scale of habitat change affecting them and thus the scale at which population genetic processes influencing IGD operate. Additionally, the clustering algorithm used can 'daisy chain' locations together so that sites > 1000 km apart are grouped (Appendix A1). Grouping sequences into biologically implausible 'populations' likely obscures anthropogenic effects on IGD, especially when combined with the small sample sizes (< 10 sequences/year for 77% of time series overall) and the large number of locations sampled yearly (mean of three locations sampled *per* year for fish; Appendix A1). We reexamined 104 Inland and Coastal Bony Fish time series from Millette *et al.,* (2020), and found that the sequences included in most time series (96/104) were from multiple genetically and demographically independent locations (inland water bodies from a median of three disconnected drainage basins; see Figure 1 and Appendix A1) and incorrectly pooled into 'populations'. By pooling sequences from independent locations, changes in IGD attributable to anthropogenic pressures would be lost in the noise, with uneven sampling across space and time compounding the issue (Table S1).

Additionally, the median span of time series is only 7 years overall, and represents an average of just 2.2 generations for fish (Appendix A1). This example suggests that, for many taxa, the data cover an insufficient time span for most measurable changes in IGD.

CONCLUSIONS

We support the goals of Millette *et al.,* (2020) and recognize that some of the flaws outlined are not unique to their study, although their temporal focus presents novel issues. Combined, these constraints (see Table S1 for summary and recommendations for future macrogenetic studies) increase the risk of overinterpretation of macrogenetic studies' conclusions (e.g. no or no consistent anthropogenically driven IGD changes), which could misinform important conservation decisions. Macrogeneticists must not merely acknowledge such limitations and carry on with their studies regardless, especially when meta-analyses using appropriate molecular markers consistently show anthropogenically driven changes in IGD (e.g. due to habitat loss and fragmentation; Schlaepfer et al., 2018; González *et al.,* 2020). Macrogeneticists must accurately study the variables of interest using the most appropriate data (e.g. nuclear markers for overall IGD assessments; Schmidt *et al.,* 2020; functional genes for revealing adaptive IGD patterns; Yiming *et al.,* 2021; population-level data from temporal genetic assessments for exploring IGD changes over time; Leigh *et al.,* 2019) rather than the most abundant data.

ACKNOWLEDGEMENTS

We acknowledge the support of the GEO BON Genetic Composition Working Group in the development of this manuscript. We also thank Arne Mooers and the reviewers for their comments. I.P-V works in a laboratory supported by the 'Laboratoire d'Excellence' (LABEX) entitled TULIP (ANR-10-LABX-41).

COMPETING INTERESTS

The authors declare no competing interests.

DISCLAIMER

Any use of trade, firm or product names is for descriptive purposes only and does not imply endorsement by the U.S. Government.

AUTHORSHIP

All authors contributed to the brainstorming of ideas and to the development of the manuscript. IPV and ELJ wrote the drafts of the manuscript, and all authors contributed substantially to revisions. IPV made the figures and GIS analyses, and ELJ conducted the literature survey. SH initiated the project, and IPV and SH supervised the project.

PEER REVIEW

The peer review history for this article is available at https://publons.com/publon/10.1111/ele.13732.

DATA AVAILABILITY STATEMENT

No new data have been used in this article. Mitochondrial COI sequence data for fish time series were made available online by Millette *et al.,* (2020) and can be found here: https://doi.org/10.5061/dryad.29rt7n0.

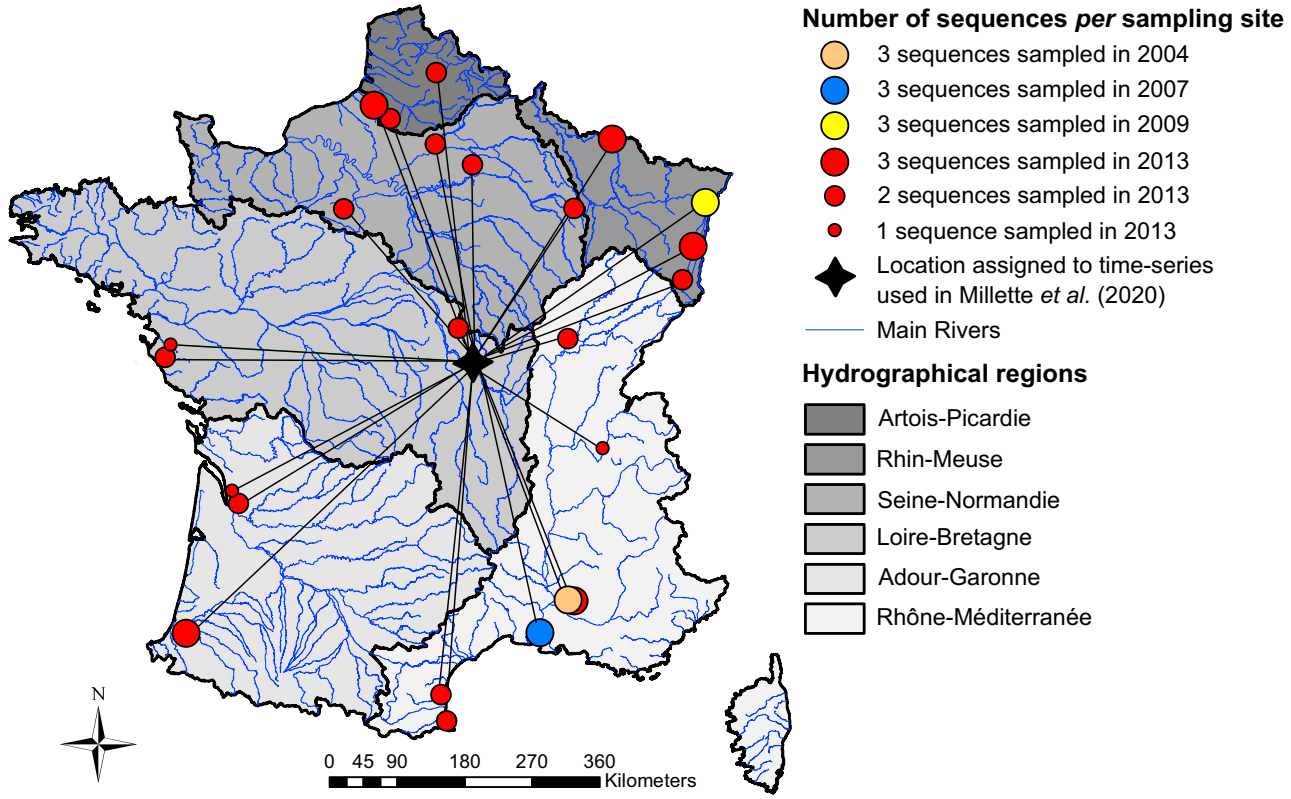

Figure 1 Map showing the grouping of sequences from the fish species *Gasterosteus gymnurus* (Cuvier, 1829; a junior synonym of *Gasterosteus aculeatus*, Linnaeus, 1758; Denys et al. 2015) into a single 'population' to measure change in IGD. This is one of the 909 time series datasets in Millette et al. (2020) and consists of 53 mitochondrial COI sequences collected at 24 different sampling sites (coloured dots). The sampling sites are all within the 1,000 km distance threshold set by Millette et al. for aggregating sequences into a 'population', despite being located in nine watersheds from six major disconnected hydrographical regions. Sample sizes are highly uneven across the time series, with just three sequences from a single site each in 2004, 2007 and 2009, and then 44 sequences from 21 sites in 2013. Millette et al. (2020) analysed the trend in nucleotide diversity across these temporal points, despite the 2013 sample consisting of sequences pooled across many different regions, while the other years had a single site, in different regions.

REFERENCES

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

Denys, G.P.J., Geiger, M., Persat, H., Keith, P. & Dettai, A. (2015). Invalidity of *Gasterosteus gymnurus* (Cuvier, 1829) (Actinopterygii, Gasterosteidae) according to integrative taxonomy. *Cybium*, 39, 37–45.

González, A.V., Gómez-Silva, V., Ramírez, M.J. & Fontúrbel, F.E. (2020). Meta-analysis of the differential effects of habitat fragmentation and degradation on plant genetic diversity. *Cons Biol*., 34, 711–720.

Lawrence, E.R., Benavente, J.N., Matte, J.-M., Marin, K., Wells, Z.R.R., Bernos, T.A. *et al.* (2019). Geo-referenced population-specific microsatellite data across American continents, the MacroPopGen Database. *Sci. Data*, 6, 14.

Leigh, D.M., Hendry, A.P., Vázquez-Domínguez, E. & Friesen, V.L. (2019). Estimated six per cent loss of genetic variation in wild populations since the industrial revolution. *Evol. Appl*., 12, 1505–1512.

Manel, S., Guerin, P.-E., Mouillot, D., Blanchet, S., Velez, L., Albouy, C. *et al.* (2020). Global determinants of freshwater and marine fish genetic diversity. *Nat. Commun*., 11, 692.

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.

Schlaepfer, D.R., Braschler, B., Rusterholz, H.-P. & Baur, B. (2018). Genetic effects of anthropogenic habitat fragmentation on remnant animal and plant populations: a meta-analysis. *Ecosphere*, 9, e02488.

Schmidt, C., Domaratzki, M., Kinnunen, R.P., Bowman, J. & Garroway, C.J. (2020). Continent-wide effects of urbanization on bird and mammal genetic diversity. *Proc. R. Soc. B Biol. Sci*., 287, 20192497.

Theodoridis, S., Fordham, D.A., Brown, S.C., Li, S., Rahbek, C. & Nogues-Bravo, D. (2020). Evolutionary history and past climate change shape the distribution of genetic diversity in terrestrial mammals. *Nat. Commun*., 11, 2557.

Yiming, L., Siqi, W., Chaoyuan, C., Jiaqi, Z., Supen, W., Xianglei, H. *et al.* (2021). Latitudinal gradients in genetic diversity and natural selection at a highly adaptive gene in terrestrial mammals. *Ecography*, 44, 206–218.

SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section at the end of the article.

Editor, Arne Mooers
Manuscript received 18 September 2020
First decision made 14 November 2020
Manuscript accepted 6 December 2020