Trends in Ecology & Evolution

CellPress REVIEWS

Spotlight

Which Latitudinal Gradients for Genetic Diversity?

Paolo Gratton,[1,*] Silvio Marta,[2] Gaëlle Bocksberger,[1] Marten Winter,[3] Petr Keil,[3] Emiliano Trucchi,[4] and Hjalmar Kühl[1]

A recent global analysis of GenBank DNA sequences from amphibians and mammals indicated consistent poleward decrease of intraspecific genetic diversity in both classes. We highlight that this result was biased by not accounting for distance decay of similarity and reanalyse the datasets, revealing distinct latitudinal gradients in mammals and amphibians.

Geographic patterns of intraspecific genetic diversity are informative about past range dynamics [1] and large-scale analyses of public DNA databanks can boost our understanding of the global distribution of biodiversity [2]. Miraldo *et al*. [3] recently presented an analysis of geographic patterns of intraspecific genetic diversity in terrestrial mammals and amphibians. Similar to [2], they retrieved and geocoded mtDNA sequences from GenBank and showed maps of knowledge and ignorance. However, they took an important step further, presenting a global survey of intraspecific mtDNA variation.

A major result by Miraldo *et al*. [3] was the observation of a poleward decrease in genetic diversity [4] in both mammals and amphibians. The key analyses supporting this result were 'band-wise' beta regressions, where sequences were binned within 10° latitudinal bands and genetic (nucleotide) diversity was estimated for each species in each band (e.g., sequences of *Mustela erminea* sampled between 40°N and 50°N, ranging through Canada, Europe, and Japan, were aligned to obtain a single measure of diversity). In these regressions, diversity averaged across species in each band was the response and the central latitude of the bands was the explanatory variable.

However, these analyses neglected the 'first law of geography': that 'near things are more related than distant things' [5]. Ignoring distance decay of similarity might have affected Miraldo *et al*.'s [3] results in two ways. First, as the similarity of two gene copies tends to decay with their geographic distance, the expected genetic diversity in a set of DNA sequences increases with the average geographic distance between pairs of conspecific sequences in it (hereafter, *D*). Because of Earth's sphericity and the arrangement of landmasses and species ranges, *D* is expected to vary widely across latitudinal bands, although, ultimately, empirical *D*s depend on the actual spatial arrangement of samples. Indeed, *D* varies across Miraldo *et al*.'s [3] bands. Furthermore, as expected from the larger land area in the northern hemisphere (Figure 1A,B), *D* strongly correlates with latitude in the mammals dataset ($R^2 = 0.75$, $P = 0.0003$) and, less markedly, in amphibians ($R^2 = 0.38$, $P = 0.0328$). A second, distinct facet of distance decay of similarity is that measures of genetic diversity are expected to be spatially autocorrelated. If extensive autocorrelation is present in the data, binning data from distant regions into the same band can compromise statistical analyses and contrasting, for example, a 10°S–0° band containing mostly South American data with a 40°N–50°N band with mostly European data (Figure 1D) can be misleading. Therefore, the regressions of genetic diversity versus latitude presented in [3], which do not account for the effect of *D* and ignore spatial autocorrelation, might misrepresent actual patterns.

To estimate global patterns of mtDNA diversity while accounting for distance decay of similarity, we downloaded 'cell-wise' estimates of genetic diversity (cytochrome b gene) provided by [3], whereby sequences were binned within the cells of an equal-area grid (data downloaded from http://macroecology.ku.dk/resources/imapgenes on 21 October 2016). To account for the effect of *D*, we computed its value for each grid cell using the coordinates of individual sequences (also from http://macroecology.ku.dk/resources/imapgenes) and fitted generalised additive models for amphibians and mammals (formula: diversity ~ latitude + latitude^2 + *D*; weights: number of sequences in cell; error structure: beta; link: logit; *R* package: mgcv [6]). We checked for spatial autocorrelation by plotting and testing correlograms of model residuals (Figure 1C,F).

For both amphibians and mammals, genetic diversity depended significantly on *D* (amphibians: $z = 29.2$, $P < 0.001$; mammals: $z = 24.5$, $P < 0.001$) and, quadratically, on latitude (amphibians: $z = -29.4$, $P < 0.001$; mammals: -10.49, $P < 0.001$). Model residuals were not significantly autocorrelated (Figure 1C,F), indicating that latitude accounts for most of the spatial autocorrelation in these datasets and that there is no need for an additional term (e.g., spline smoothing) for longitude.

However, while for amphibians the latitudinal trend from our cell-wise analysis was similar to the band-wise analysis in [3] (Figure 1E), the mammal regressions were strikingly different (Figure 1B). In particular, the band-wise regression from [3] predicts a more pronounced effect of latitude, higher diversity in the northern hemisphere, and much lower diversity in the southern hemisphere than our cell-wise regression. The regression curve

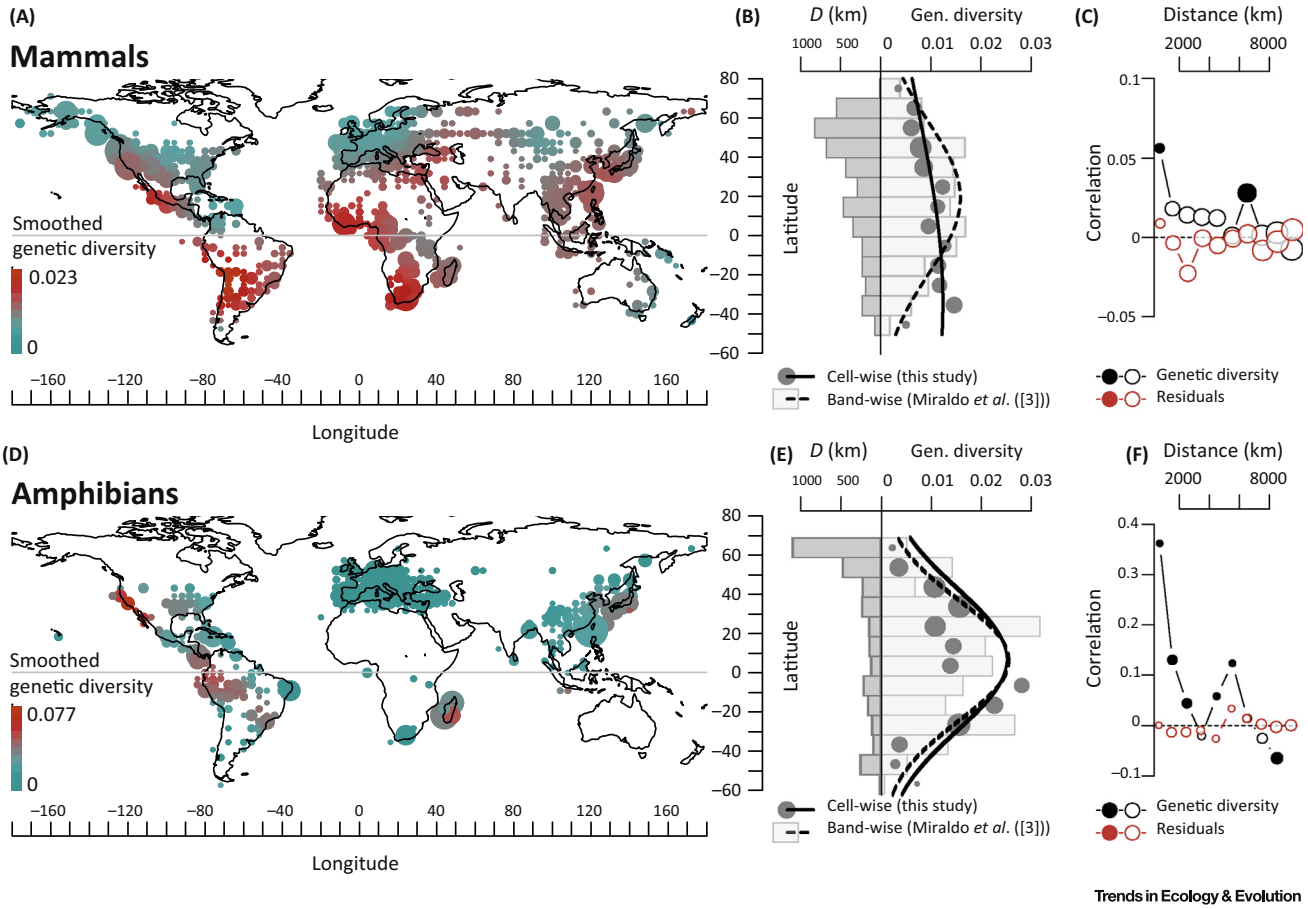

Figure 1. Global Distribution of Cytochrome b Diversity in Mammals and Amphibians. (A,D) Colour gradients show cytochrome b nucleotide diversity per cell of an equal-area grid predicted by generalised additive models (GAMs) with Gaussian smoothing splines for latitude and longitude and controlling for average distance between conspecific sequences (*D*). Raw data were downloaded from http://macroecology.ku.dk/resources/imapgenes. We plotted smoothed model predictions to obtain easily readable maps of genetic diversity that remove the confounding effect of average distance between conspecific samples (*D*) (predictions are shown for *D* = 100 km in all cells). Diameters of circles are proportional to the logarithm of the number of sequences in the cell. (B,E) Dark-grey bars to the left of the vertical axes show *D* for latitudinal bands analysed by Miraldo *et al*. [3]. Light-grey bars to the right of the vertical axes show band-wise estimates of genetic diversity from [3]. Dark-grey dots represent estimates of genetic diversity averaged across grid cells [same grid as in (A,D)] within 10° latitudinal bands. Size of dots is proportional to the average number of sequences across cells in each band. Lines show predictions from beta regressions based on band-wise estimates of genetic diversity as in [3] (broken line) and based on cell-wise estimates of genetic diversity and accounting for *D* (unbroken line). (C,F) Correlograms (1000-km increments) for raw genetic diversity (black) and for the residuals of a GAM of genetic diversity as a function of latitude and *D* (red). Statistical significance was computed through 1000 permutations. Significant correlation is indicated by solid dots. Size of dots is proportional to the number of cell pairs in each distance bin.

estimated in [3] (Figure 1B, broken line) also contrasts with patterns noted by the same authors, who, in commenting on their global map of genetic diversity, emphasised the high diversity of South America and southern Africa (see also Figure 1A). Underestimation of genetic diversity in the southern hemisphere by the band-wise approach is expected because of southern bands having lower *D* than northern bands (Figure 1B). Interestingly, Miraldo *et al*. [3] also performed a cell-wise regression of genetic diversity versus latitude (presented in their supplementary materials and not controlling for either *D* or autocorrelation of residuals). However, they could not appreciate the extent of the differences between the band-wise and cell-wise analyses because their cell-wise analysis conflated hemispheres by considering absolute latitude.

Summarising, our analysis suggests that the band-wise quadratic regressions that constitute the main result of [3] incorrectly describe global patterns of genetic diversity (at least for mammals), mainly because of a systematic northward bias in *D*. Our analysis revealed differences between latitudinal gradients of genetic diversity in mammals and amphibians that were overlooked by [3] and might reflect differences in dispersal and thermoregulation between the two classes. Nonetheless, for latitudes above 30°N a poleward decrease of genetic diversity in both mammals and amphibians is broadly confirmed.

Another important result in [3] concerned the reduction of genetic diversity from more pristine to more anthropised habitats. As this analysis also did not account for distance decay of similarity, we suggest that a thorough recalculation is necessary to verify this result.

Both our analyses and Miraldo *et al*.'s [3] did not account for species identities, thus potentially conflating within-species gradients and differences among species occupying different regions. Disentangling these effects requires the analysis of large numbers of range-wide intraspecific datasets, which are still undeveloped [2]. It is therefore encouraging that Miraldo *et al*. [3] joined our earlier call [2] for a collective effort toward richer and accessible georeferenced inventories of global genetic diversity.

Acknowledgments

The authors thank Fabrizio Mafessoni, Roger Mundry, and Ludovic Orlando for valuable comments and support. P.G., G.B., and H.K. acknowledge funding by the Max Planck Society through the Pan African Programme (http://panafrican.eva.mpg.de). P.K. and M.W. acknowledge funding by the DFG for iDiv (FZT 118).

[1]Department of Primatology, Max Planck Institute for Evolutionary Anthropology, Deutscher Platz 6, 04103 Leipzig, Germany
[2]Dipartimento di Biologia, Università di Roma 'Tor Vergata', Via della Ricerca Scientifica 1, 00133 Roma, Italy
[3]iDiv German Center for Integrative Biodiversity Research, Deutscher Platz 5, 04103 Leipzig, Germany
[4]Department of Life Sciences and Biotechnology, University of Ferrara, Via Ludovico Ariosto 35, 44121 Ferrara, Italy

*Correspondence:
paolo_gratton@eva.mpg.de (P. Gratton).
http://dx.doi.org/10.1016/j.tree.2017.07.007

References

1. Hewitt, G.M. (2000) The genetic legacy of the Quaternary ice ages. *Nature* 405, 907–913
2. Gratton, P. *et al.* (2016) A world of sequences: can we use georeferenced nucleotide databases for a robust automated phylogeography? *J. Biogeogr.* 44, 475–486
3. Miraldo, A. *et al.* (2016) An Anthropocene map of genetic diversity. *Science* 353, 1532–1535
4. Pereira, H. (2016) A latitudinal gradient for genetic diversity. *Science* 353, 1494–1495
5. Tobler, W. (1970) A computer movie simulating urban growth in the Detroit region. *Econ. Geogr.* 46, 234–240
6. Wood, S.N. (2011) Fast stable restricted maximum likelihood and marginal likelihood estimation of semiparametric generalized linear models. *J. R. Stat. Soc. Series B* 73, 3–36