Predictors of intraspecific diversity in herptiles using machine learning

Maria João Paúl [1,2,3], Matthew Moreira [1,3], André Vicente Liz [1,2,3], Sydney K. Decker [4], Craig Moritz [5], Bryan C. Carstens [4], Sílvia B. Carvalho [1,3]

[1] CIBIO, Centro de Investigação em Biodiversidade e Recursos Genéticos, InBIO Laboratório Associado, Campus de Vairão, Universidade do Porto, Vairão, Portugal;
[2] Departamento de Biologia, Faculdade de Ciências, Universidade do Porto, Porto, Portugal;
[3] BIOPOLIS Program in Genomics, Biodiversity and Land Planning, CIBIO, Campus de Vairão, Vairão, Portugal;
[4] Department of Evolution, Ecology and Organismal Biology, The Ohio State University, Columbus, OH, USA;
[5] Research School of Biology, The Australian National University, Canberra, ACT, Australia.

I. Introduction

Given that an important element of variation is distributed within species, inferring **intraspecific genetic diversity** and identifying its **predictors** is of utmost importance so that we can develop effective conservation strategies.

Here, we focus on the genetic diversity of **herpetiles** across **six regions**: the Iberian Peninsula, Northwestern Africa, the Arabian Peninsula, Madagascar, Northern Australia, and the Brazilian Atlantic Forest (Figure 1).

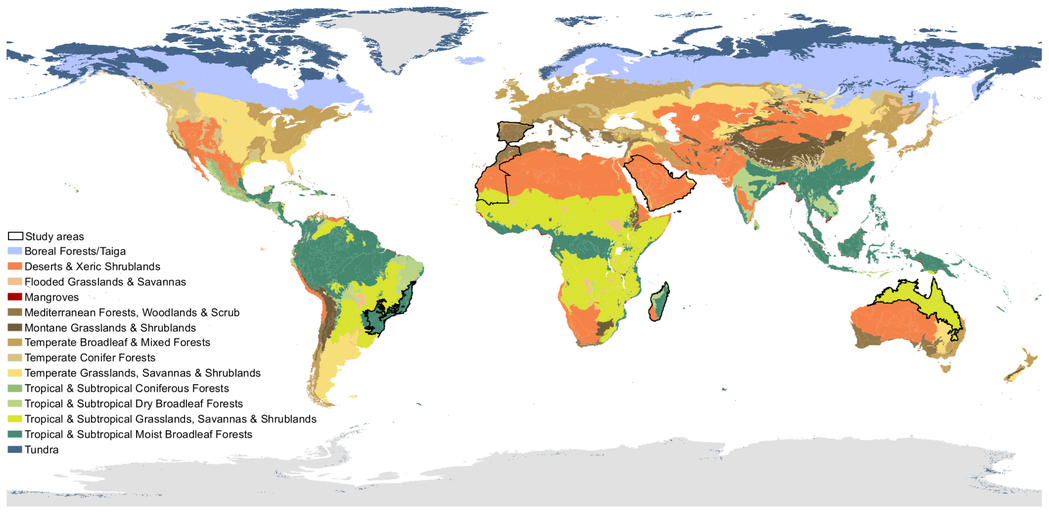


Figure 1. Biomes of the study regions of this study. Biomes adapted from Dinerstein et al. 2017.

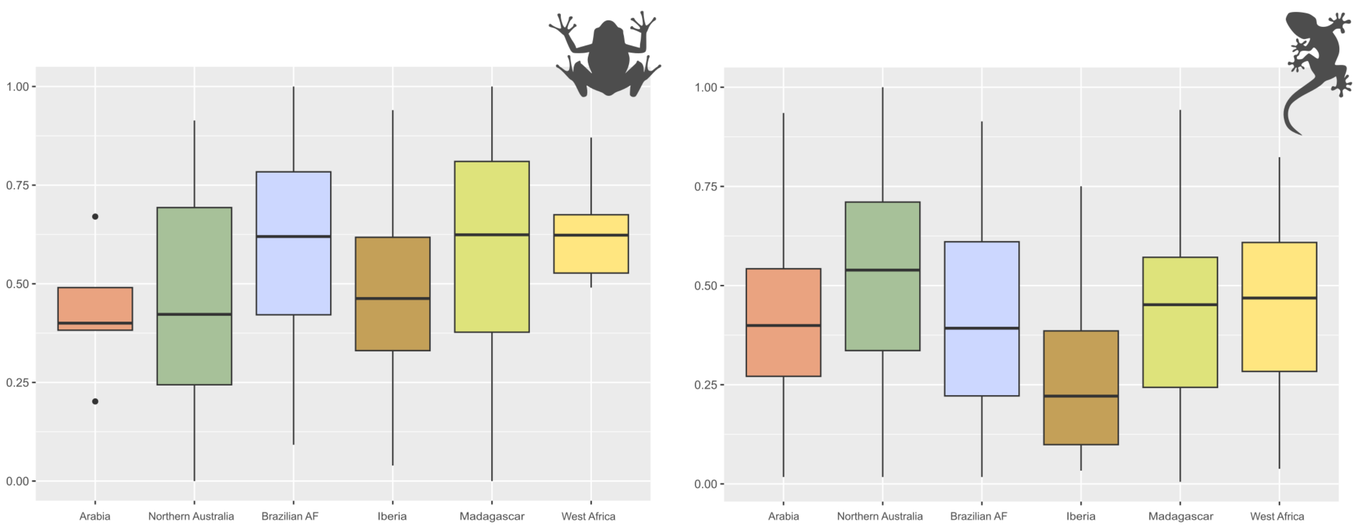


Figure 2. Boxplots of amphibian and reptile mean nucleotide diversity per study region. Each box-whisker plot indicates the median (bold lines) and the interquartile range (boxes).

II. Methods

1. Mitochondrial data compilation, alignment and georeferencing
2. Genetic diversity estimation and analysis
3. Predictor compilation:
   - Range characteristics
   - Land characteristics
   - Life history
   - Taxonomy
   - Miscellaneous
4. Identification of important predictors using machine learning

III. Results

Mean nucleotide diversity was significantly different in both amphibians (*p-value*: 0.0007) and reptiles (*p-value*: 0.0003) among study regions (Figure 2).

In amphibians, the top 3 most important predictors in our Random Forest Regression (RMSE = 0.2086) were body size, altitude sd and precipitation of the driest quarter (bio17) range (Figure 3).

In reptiles, the top 3 most important predictors in our Random Forest Regression (RMSE = 0.2054) were maximum body mass, minimum temperature of coldest month (bio06) range and mean offspring (Figure 3).

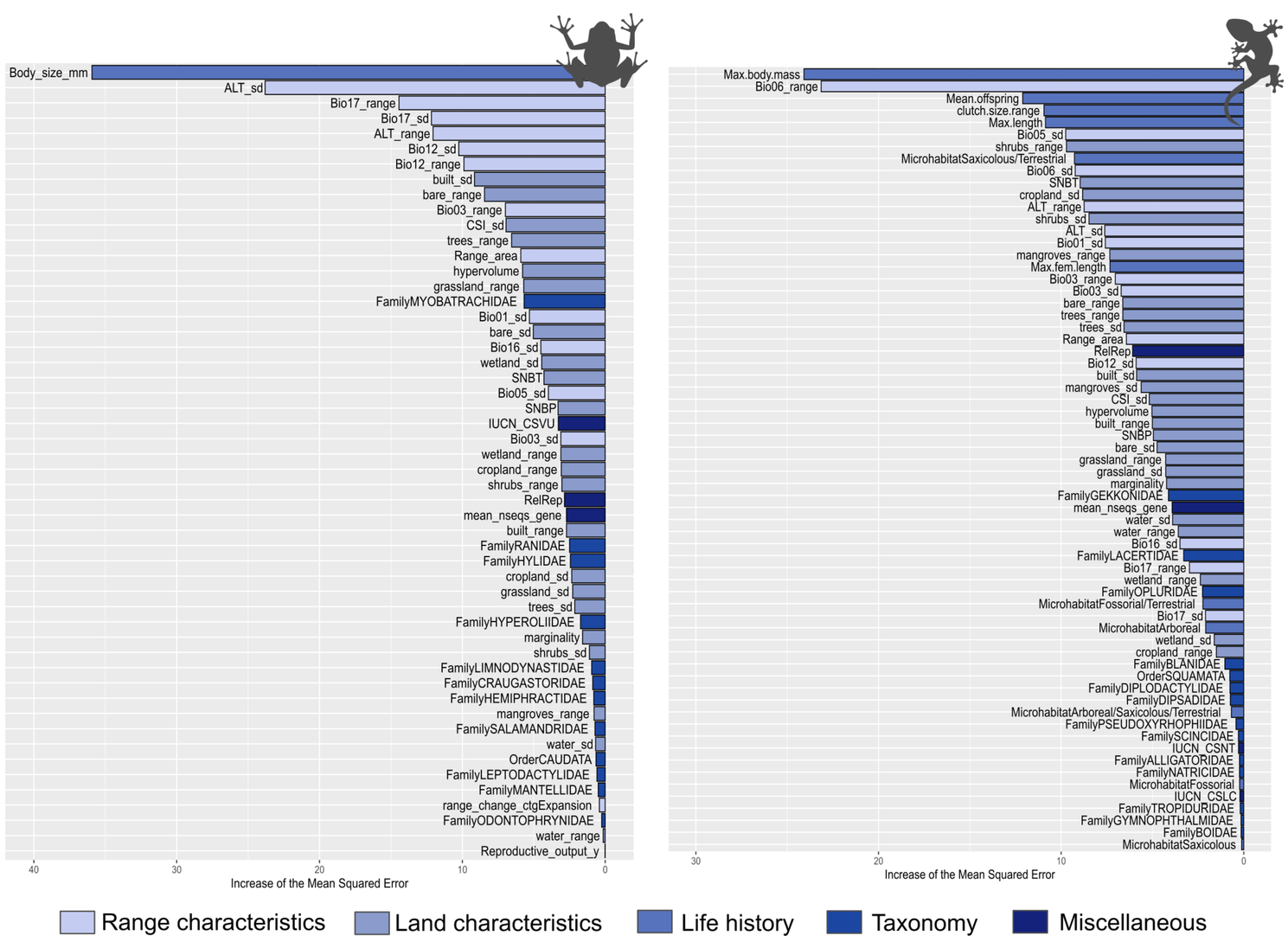


Figure 3. Most important predictor variables for the consensus Random Forest Regression.

IV. Final Remarks

There are **regional variations** in both amphibian and reptile genetic diversity.

Life history traits, **range** and **land characteristics** are key predictors of herptile genetic diversity in our six study regions.

Acknowledgements

This work was funded by Fundação para a Ciência e Tecnologia, I.P, via national funds, under the scope of project PLACES - "PLAnning for the Conservation of life on Earth in the light of evolution and Sustainability" (DOI 10.54499/2022.04598.CEECIND/CP1730/CT0006). MJP and SBC were supported by Fundação para a Ciência e Tecnologia through a PhD Grant (SFRH/BD/08664/2020), and an individual scientific employment program contract (CEECIND/CP1730/CT0006), respectively.

References

Dinerstein, Eric, David Olson, Anup Joshi, Carly Vynne, Neil D. Burgess, Eric Wikramanayake, Nathan Hahn, et al. 2017. 'An Ecoregion-Based Approach to Protecting Half the Terrestrial Realm'. BioScience 67 (6): 534–45.