Mapping and understanding the digital biodiversity knowledge about vertebrates in the Atlantic Rainforest

Gracielle T. Higino1, 2 \*, Pablo Henrique da Silva3, Marcos V. C. Vital4

1 Department of Zoology and Biodiversity Research Centre, University of British Columbia, Vancouver, BC, Canada.

2 Programme de formation en sciences et services de la biodiversité computationnelle, Canada.

3 Instituto de Matemática e Estatística, Universidade Federal de Goiás, Goiânia, Brazil.

4 Instituto de Ciências Biológicas e da Saúde, Universidade Federal de Alagoas, Maceió, AL, Brazil

\*Corresponding author

\*[**graciellehigino@gmail.com**](mailto:graciellehigino@gmail.com)

# Abstract

Biases and gaps in biodiversity data can have serious consequences on ecological and conservation research and actions, as it can lead to incorrect conclusions. Although there is still digital knowledge to be found and organized, recent efforts on gathering global biodiversity data, such as data papers, have revealed long hidden information. Nevertheless, it is of major importance to map and describe the biases on the data we have available. Here we assessed terrestrial vertebrates’ digital inventory incompleteness at the Atlantic Rainforest and investigated if environmental variables are correlated to biodiversity knowledge. At a resolution of 0.5 degrees and considering the final slope of a rarefaction curve as a measure of inventory completeness, none of the grids can be considered well sampled units. However, the completeness of each group seems to have a non-uniform correlation to different environmental variables. By highlighting undersampled areas in the Atlantic Rainforest, these results provide important information for the biodiversity assessment of this highly impacted ecoregion. Combining them with the information about the magnitude of these impacts can help shape the agenda and priorities for conservation, but, also, the results alone can help us rethink what we understand about geographically structured biodiversity distribution.

Keywords: inventory completeness, rarefaction curves, data bias, Atlantic Rainforest.

# Introduction

Information about life diversity and distribution is a fundamental tool for understanding evolutionary and ecological processes (Graham *et al.* 2004; Rocchini *et al.* 2011; Jetz *et al.* 2012; Ladle & Hortal 2013; Meyer *et al.* 2015). After a long history of global biodiversity information collection by naturalists, taxonomists and, more recently, citizen scientists (Von Humboldt *et al.* 1850; Hawkins 2001; Willig *et al.* 2003; Chase 2012; Callaghan *et al.* 2020), researchers have been storing these data in electronic catalogues at slow pace since the 1970’s, connecting them, more recently, through web-based initiatives (Graham *et al.* 2004). As a result, we now have accessible and extensive information about biodiversity on big online databases such as the Global Biodiversity Information Facility (GBIF; <http://www.gbif.org/>) and the Map of Life (<https://mol.org/>), which compile museum, survey and observation data (Graham *et al.* 2004; Jetz *et al.* 2012; Beck *et al.* 2013). However, despite these recent efforts, our knowledge on species diversity and distribution is still biased and full of gaps due to the complex nature of this information (Brown and Lomolino 1998; Whittaker et al. 2005). These shortfalls have been recently revised (Hortal et al. 2015) and there is growing evidence that they can compromise ecological, evolutionary and conservation analyses (IUCN 2012; Ladle and Hortal 2013; Ficetola et al. 2014; Hortal et al. 2015). The wallacean shortfall (the lack of information about species’ real distribution) is present in every spatial and temporal scales (Whittaker et al. 2005; Hortal 2008; Hortal et al. 2015) and is a consequence of a myriad of biological, environmental and social factors. Characteristics of the species (such as crypsis, its natural history and behaviour), political borders and topography, for example, can lead to biases in biodiversity surveys and form gaps in information. On the other hand, clustered information also can lead to biased surveys, since researchers may prefer to assess places knowingly species-rich or that are undergoing a process of ecological change (Boakes et al. 2010; Ahrends et al. 2011; Rocchini et al. 2011; Yang et al. 2014). Information gaps may also be a consequence of data quality decay in space (e.g., when we extrapolate the distribution of a species based on polygons or species distribution models) and time (due to taxonomic reviews, climate change, land use, habitat loss, extinction and migration) (Ladle and Hortal 2013). Therefore, the measurement of geographical variation of biodiversity on the planet (represented by distribution maps) has an error associated that must be assessed (Hortal 2008; Rocchini et al. 2011; Ladle and Hortal 2013; Yang et al. 2013). The acknowledgement of error in biodiversity information is of major importance. The underestimation of species distribution can have consequences in conservation planning, such as the classification of species in risk of extinction following the range restriction criterion (IUCN 2012; Ladle and Hortal 2013; Ficetola et al. 2014; Hortal et al. 2015). Furthermore, bias can influence and even reverse ecogeographical patterns, leading us to associate certain factors to species richness when they are only proxies for sampling quality (Ficetola et al. 2014). Therefore, it has been recommended to include maps of ignorance in the results or to map data quality and use only well sampled locations on analyses (Hortal 2008; Ladle and Hortal 2013; Ficetola et al. 2014; Yang et al. 2014). Once researchers are aware of the error in their data sets, they can try to correct or diminish it and better analyse the results. This practice, in addition to guide future research, produces more reliable results, since the exact measure of uncertainty clarifies how explicative an inference can be.

There is a growing interest in biodiversity data biases in the literature (see Boakes et al. 2010; Yang et al. 2013; Sousa-Baena et al. 2014). Nevertheless, studies mapping South American under-sampled sites are relatively few. This is worrying especially for the Atlantic Forest because this ecoregion is an important biodiversity and socio-climatic hotspot (Scarano and Ceotto 2015). Human activities and the growth of urban centres have reduced its original area to only 8% (Galindo-Leal and Câmara 2003; Scarano and Ceotto 2015), resulting in substantial loss of habitat. Despite that, it still hosts 1-8% of the world’s total species (Silva and Casteleti 2003). The Atlantic Rainforest is also a good model for ecological and evolutionary research because of its large latitudinal and altitudinal range, high endemicity, variation in temperature and precipitation, and historical connexion with other biomes (Silva and Casteleti 2003; Ribeiro et al. 2009; Batalha-Filho et al. 2013). Given that the Atlantic Rainforest is a biodiversity hotspot, with a high rate of deforestation, it becomes urgent to describe and map the information about species occurrence in this ecoregion that is available online.

# References

Beck, J., Ballesteros-Mejia, L., Nagel, P. & Kitching, I.J. (2013). Online solutions and the ’Wallacean shortfall’: What does GBIF contribute to our knowledge of species’ ranges? *Diversity and Distributions*, 19, 1043–1050.

Callaghan, C.T., Roberts, J.D., Poore, A.G.B., Alford, R.A., Cogger, H. & Rowley, J.J.L. (2020). Citizen science data accurately predicts expert-derived species richness at a continental scale when sampling thresholds are met. *Biodiversity and Conservation*.

Chase, J. (2012). Historical and Contemporary Factors Govern Global Biodiversity Patterns. *PLoS Biol*, 10, e1001294.

Graham, C.H., Ferrier, S., Huettman, F., Moritz, C. & Peterson, a.T. (2004). New developments in museum-based informatics and applications in biodiversity analysis. *Trends in Ecology and Evolution*, 19, 497–503.

Hawkins, B.A. (2001). Ecology’s oldest pattern? *Trends in Ecology & Evolution*, 16, 470.

Jetz, W., McPherson, J.M. & Guralnick, R.P. (2012). Integrating biodiversity distribution knowledge: Toward a global map of life. *Trends in Ecology and Evolution*, 27, 151–159.

Ladle, R. & Hortal, J. (2013). Mapping species distributions: Living with uncertainty. *Frontiers of Biogeography*, 5, 4–6.

Meyer, C., Kreft, H., Guralnick, R. & Jetz, W. (2015). Global priorities for an effective information basis of biodiversity distributions. *Nature Communications*, 6, 8221.

Rocchini, D., Hortal, J., Lengyel, S., Lobo, J.M., Jimenez-Valverde, A., Ricotta, C., *et al.* (2011). Accounting for uncertainty when mapping species distributions: The need for maps of ignorance. *Progress in Physical Geography*, 35, 211–226.

Von Humboldt, A., Otté, E.C. & Bohn, H.G. (1850). *Views of Nature: Or Contemplations on the Sublime Phenomena of Creation*. Cambridge University Press.

Willig, M.R., Kaufman, D.M. & Stevens, R.D. (2003). Latitudinal Gradients of Biodiversity: Pattern, Process, Scale, and Synthesis. *Annual Review of Ecology, Evolution, and Systematics*, 34, 273–309.