

Interactive maps for visualizing geographic distributions and phenotypes

Neil Rosser^{1,2} and James Mallet³

1. Department of Biology, University of Miami, Coral Gables, Florida 33146, USA. 2. Museum of Comparative Zoology, Harvard University, Cambridge, Massachusetts 02138, USA. 3. Department of Organismic and Evolutionary Biology, Harvard University, Cambridge, Massachusetts 02138, USA. *Author for correspondence; neil.rosser@miami.edu

Date of issue online: XXX

Electronic copies (ISSN 2575-9256) in PDF format at: <https://journals.flvc.org/troplep>; <https://zenodo.org>; archived by the Institutional Repository at the University of Florida (IR@UF), <http://ufdc.ufl.edu/ufir>; DOI: XXX

© The author(s). This is an open access article distributed under the Creative Commons license CC BY-NC 4.0 (<https://creativecommons.org/licenses/by-nc/4.0/>).

Abstract: Museum databasing initiatives have resulted in the release of large amounts of geographic locality data for a wide range of plants and animals. Species lists with images of type specimens are also increasingly available online. For specialists working with a particular group, it would be helpful to connect these phenotypic and geographic data within a single interactive interface that can be easily updated following taxonomic changes or the discovery of new taxa. Here, we present interactive web-maps for *Heliconius* butterflies and allies, together with a portal for downloading the underlying specimen and locality data. The maps and portal can be viewed here: <https://heliconius-maps.github.io>. The code and software underlying these maps are all open source, and the website is hosted for free by the GitHub servers. It should be straightforward for researchers to adapt our methods to their own taxa of interest.

Keywords: biodiversity informatics; butterflies; *Heliconius*; web-maps

INTRODUCTION

Since the turn of the century, a number of technological advances have revolutionised biodiversity informatics (Graham *et al.*, 2004). Large databasing initiatives have seen museums digitise metadata from vast numbers of specimens, and many of these datasets are freely available online. At the same time, the use of geographical information systems has become widespread, so that users can download species data and perform spatial analyses that address a broad range of questions in ecology, evolution and conservation. Many websites dedicated to taxonomy and identification have also been developed. For example, <http://butterfliesofamerica.com> currently comprises 160,500 images of over 8,300 species, including many type specimens. For researchers working on taxonomy and systematics, it would be helpful to visualise these phenotypic and distribution data simultaneously. However, we are not aware of any free and accessible platforms for creating such a resource.

Heliconius butterflies and allies (Nymphalidae: Heliconiini) comprise ~70-80 species (Lamas & Jiggins, 2017). In the past 30 years, research on *Heliconius* has increased enormously and they are the subjects of dozens of papers each year, the majority of which focus on evolutionary biology (Figure 1A). A striking feature of many species is their phenotypic variability. Many species exhibit locally adapted color patterns that mimic other co-occurring species (Merrill *et al.*, 2015). This variation is extreme - some species have as many as 30 geographic subspecies, and 450+ subspecies are currently recognised within the Heliconiini (Lamas & Jiggins, 2017).

While some excellent taxonomic resources exist for

Heliconius, they are nonetheless difficult to identify to species due to their mimicry. New molecular data also frequently lead to revisions and the description of new taxa (Rosser *et al.*, 2019; Thawornwattana *et al.*, 2023). Consequently, online databases of taxonomic and geographic distribution information such as GBIF are often riddled with errors, either due to incorrect museum specimen label data, errors during digitization, or because they have not been updated following nomenclatural or taxonomic changes. A recent paper on *Heliconius* biogeography explicitly avoided using GBIF data, “to ensure the use of data that have been curated by specialists both in terms of georeference and taxonomy” (Rueda-M *et al.*, 2021). To address these issues, we developed open-source, interactive web-maps that couple community-validated taxonomy and distribution data with representative images of species and subspecies of *Heliconius* and related genera, together with a portal for downloading the underlying data. These maps are accessed daily by users around the world (Figure 1B). Here we outline the underlying code and website architecture, should others wish to adopt our approach for different taxa.

WEB-MAPS FOR *HELICONIUS*

Maps using museum data

The web-maps comprise point localities corresponding to individual specimens. When a point locality is clicked, a pop-up bubble appears which displays key specimen information, together with an image of a representative phenotype, frequently the type specimen (Figure 1C). The data to be mapped are stored in tabular format in a comma-delimited text file. Fields include, but are not limited to, a unique identifier, genus,

PROOF

species and subspecies identity, collecting locality and decimal coordinates, and the collection source (for example, the Natural History Museum, London). The phenotype is displayed via a field containing a hyperlink to images hosted on either <https://butterfliesofamerica.com> or <https://cliniquevetodax.com/Heliconius>. Our museum database was compiled originally by Rosser *et al.* (2012) and we have subsequently incorporated a

number of new datasets, for example, from Rosser *et al.* (2021) and the collections at the Museum of Comparative Zoology, Harvard University.

We implement two approaches for mapping species distributions and provide example code and geographic data for replicating each at <https://github.com/heliconius-maps/web-map.examples>. Firstly, we produced an HTML web map for each

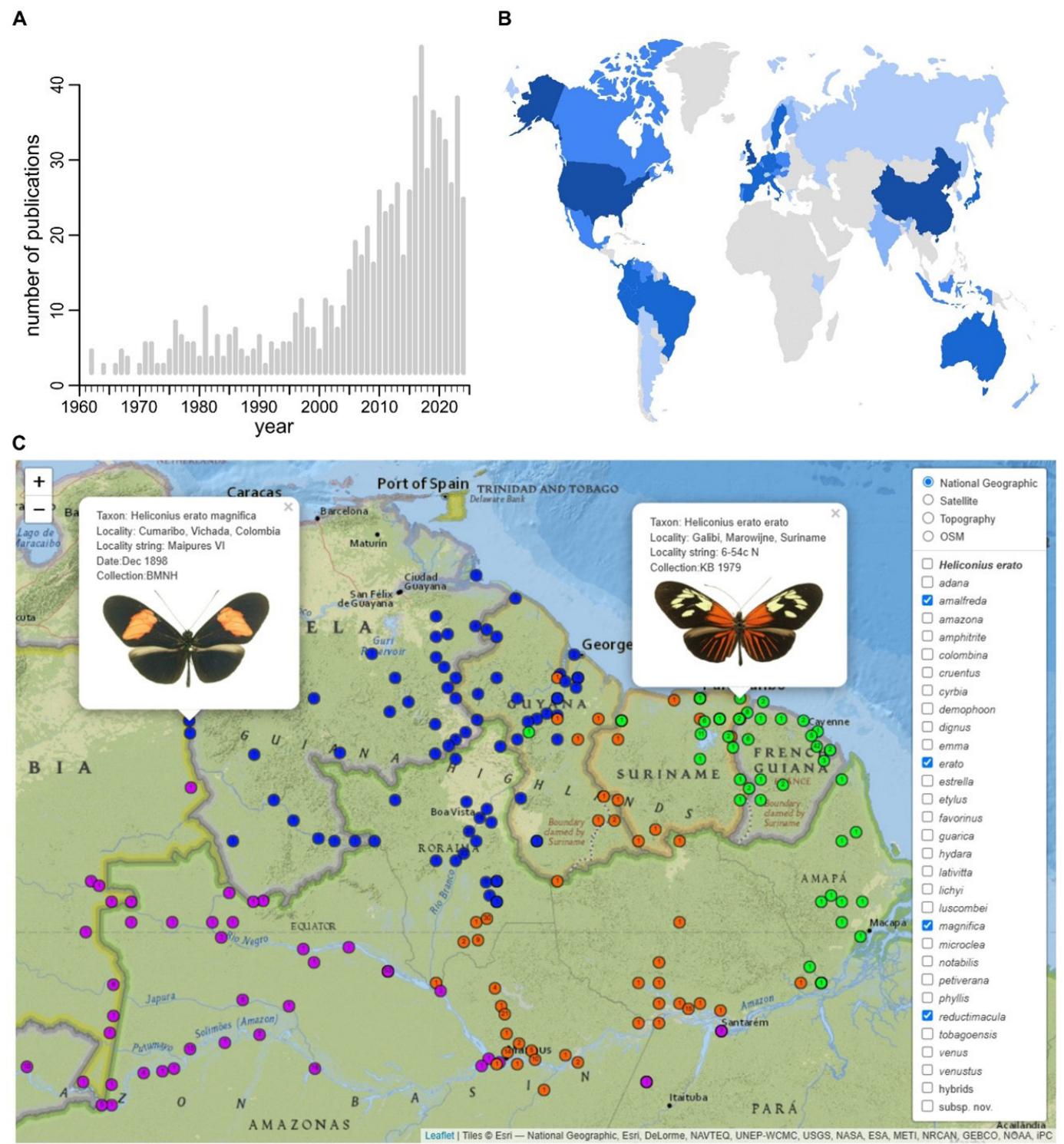


Figure 1. **a)** number of papers published per year which included *Heliconius* in the title, abstract or keywords (Web of Science, accessed 15th September 2024). **b)** Users from 47 countries have accessed the <https://heliconius-maps.github.io> homepage since it was put online in October 2021. Countries are shaded by number of users; USA has contributed the most users to data (n=660), followed by China (n=334). **c)** Example interactive web map for *Heliconius erato* subspecies.

species using the R package leaflet (Cheng, *et al.*, 2022). This package allows users to create and customize interactive maps via the Leaflet JavaScript library and the htmlwidgets R library. When viewing maps at broader scales, specimens from the same locality are clustered and represented by a single marker, which displays the number of underlying points. The scale at which markers are clustered is determined using a parameter (`maxClusterRadius`). A legend allows users to select which taxa and base map to display. The point localities for each taxon are assigned a unique color automatically and different taxa can be plotted simultaneously for comparison. A variety of different base maps showing satellite imagery, topography etc. can be easily added from sources such as ESRI and OpenStreetMap. Once a suitable template has been decided on, a map for each species in the dataset can be generated automatically using the batch script provided. The resulting maps can be explored here <https://heliconius-maps.github.io>.

Secondly, we created a Shiny interactive web application (Chang *et al.*, 2021). In brief, Shiny applications comprise a user interface object (UI) and a server function. A shinyApp function creates shinyApp objects from an explicit UI/server pair. The shinyApp also uses Leaflet to map species and subspecies distributions (the server function). However, the user selects species and subspecies by using a drop-down dashboard (the UI). Thus, there is a single HTML webpage for all species. For diverse taxa with hundreds of species, this might be a more tractable approach than mapping each species separately. We also provide options to filter the data by country and department, allowing species lists for a region to be quickly generated. The dashboard also provides a convenient portal for the user to download datasets for their own purposes, and we include a button to allow this. The shinyApp can be accessed and used here: <https://heliconius-maps.github.io/Download>.

Maps using citizen science data

Our core dataset was derived from museum databasing initiatives and comprises specimen metadata. Specimens were identified by the authors, or by the other specialists listed in the acknowledgments. We do not have photographs of each specimen available to us, but in any case, *Heliconius* phenotypes are highly variable, due to polymorphism and/or hybridisation. We therefore provide an image of the type or a specimen we consider canonical for the species/subspecies. Given our primary objective of providing a practical guide to facilitate identification, we feel this is more useful than a specimen by specimen documentation of variation.

Nonetheless, it would be straightforward for the pop-ups to display a photograph of the specimen corresponding to each geographical record. To illustrate this, we added an additional section of the website, which includes 187,414 observations downloaded from the citizen science initiative iNaturalist (last accessed: 15th September 2024, filtered to verifiable, research grade, wild observations). When these observations are clicked on, the pop-ups display the default image from the iNaturalist observation (<https://heliconius-maps.github.io/iNaturalist>). In addition, the user can quickly access the iNaturalist observation by clicking on a link. As well as providing a vast amount of additional data, these maps also provide a convenient way to

check iNaturalist records for errors, for example, by looking for clear geographic outliers.

DISCUSSION

Research on *Heliconiini* has contributed significantly to our understanding of biodiversity, in fields ranging from evolutionary developmental biology to neotropical biogeography (Merrill *et al.*, 2015). Prior to widespread digitization of museum collections, the most comprehensive geographical reference for *Heliconius* butterflies was Keith S. Brown's 1979 "Livro Docência" (habilitation degree) for the University of Campinas in São Paulo, Brazil. Brown's 296-page monograph was a monumental achievement and the definitive reference manual for *Heliconius* field workers for decades. In 2012, we compiled a database of 58,000 records for species in the *Heliconiini* tribe (Rosser *et al.*, 2012). We made species maps and data available 1) as static images in an appendix to Rosser *et al.*, (2012) and 2) through Google Fusion Tables, which was a free web service provided by Google for data visualisation and sharing. Despite some limitations, Fusion Tables provided a reasonable solution for making data available for mapping and downloading. However, the service was discontinued in December 2019, leaving few alternatives. Those that did exist were commercial ventures and prohibitively expensive.

The interactive web-maps and data portal described here present an alternative and accessible solution for mapping species distributions. We stress that these are not intended to be competitors to GBIF and similar global data sources. Rather, they fill a different niche with three primary advantages. First, unlike GBIF or other global biodiversity databases, every specimen was examined and identified by specialists using the latest agreed taxonomy (*Heliconius* Genome Consortium, 2012; Lamas & Jiggins, 2017). Second, the underlying software is all free and open source, and the code and maps are easy to replicate and customize. Thirdly, the data and website can be quickly updated. In our particular implementation, the georeferenced specimen data are held in three differently formatted files, consisting of (1) the first author's personal database, (2) our community-validated museum database, and (3) the data downloaded from iNaturalist. These can be continually curated to a high standard by eliminating errors and employing the latest taxonomy, and new data can be added over time. To update the website, we simply run a custom R script to combine the three databases into a single comma-delimited file. We then run two further scripts that generate the Leaflet webmaps and the shinyApp. These can then be posted online within a matter of minutes. Researchers should be able to easily adapt our methods to any organisms with available geographic and phenotypic data. Our hope is that they find them as useful as we do.

ACKNOWLEDGMENTS

We thank all the researchers who contributed to the data presented here, including Keith Willmott, Blanca Huertas, Gerardo Lamas, Camilo Salazar, Andrew Brower, Keith Brown, Jean Francois Le Crom, Chris Jiggins, Dan Janzen, José Luis

PROOF

Salinas-Gutiérrez, Jason Hall, André Freitas, Leila Shirai and others. We are also indebted to Michel Cast and the Butterflies of America team for making phenotypic data for heliconiine butterflies freely available via their websites. We also thank Wendy Valencia-Montoya, Keith Willmott and André Freitas for providing helpful comments on the manuscript.

LITERATURE CITED

- Brown, K. S. Jr.** 1979. *Ecologia Geográfica e Evolução nas Florestas Neotropicais*. Tese apresentada à Universidade Estadual de Campinas como parte das exigências de um Concurso de Livre Docência, área de Ecologia, Campinas, Brazil.
- Chang, W., Cheng, J., Allaire, J., Sievert, C., Schloerke, B., Xie, Y., Allen, J., McPherson, J., Dipert, A., Borges, B.** 2021. shiny: Web Application Framework for R. R package version 1.7.1. <https://CRAN.R-project.org/package=shiny>.
- Cheng, J., Karambelkar, B., Xie, Y.** 2022. Leaflet: Create Interactive Web Maps with the JavaScript ‘Leaflet’ Library. R package version 2.1.1. <https://CRAN.R-project.org/package=leaflet>.
- 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 & Evolution* 19: 497-503.
- Heliconius Genome Consortium.** 2012. Butterfly genome reveals promiscuous exchange of mimicry adaptations among species. *Nature* 487: 94-98.
- Lamas, G., Jiggins, C. D.** 2017. *Taxonomic List*, pp. 214-244. In: Jiggins, C. D. (Ed.), *The Ecology and Evolution of Heliconius Butterflies*. New York, Oxford University Press.
- Merrill, R. M., Dasmahapatra, K. K., Davey, J. W., Dell’Aglia, D. D., Hanly, J. J., Huber, B., Jiggins, C. D., Joron, M., Kozak, K. M., Llaurens, V., Martin, S. H., Montgomery, S. H., Morris, J., Nadeau, N. J., Pinharanda, A. L., Rosser, N., Thompson, M. J., Vanjari, S., Wallbank, R. W. R., Yu, Q.** 2015. The diversification of *Heliconius* butterflies: what have we learned in 150 years? *Journal of Evolutionary Biology* 28: 1417-1438.
- Rosser, N., Freitas, A. V. L., Huertas, B., Joron, M., Lamas, G., Mérot, C., Simpson, F., Willmott, K. R., Mallet, J., Dasmahapatra, K. K.** 2019. Cryptic speciation associated with geographic and ecological divergence in two Amazonian *Heliconius* butterflies. *Zoological Journal of the Linnean Society* 186: 233-249.
- Rosser, N., Phillimore, A. B., Huertas, B., Willmott, K. R., Mallet, J.** 2012. Testing historical explanations for gradients in species richness in heliconiine butterflies of tropical America. *Biological Journal of the Linnean Society* 105: 479-497.
- Rosser, N., Shirai, L. T., Dasmahapatra, K. K., Mallet, J., Freitas, A. V. L.** 2021. The Amazon river is a suture zone for a polyphyletic group of co-mimetic heliconiine butterflies. *Ecography* 44: 177-187.
- Rueda-M, N., Salgado-Roa, F. C., Gantiva-Q, C. H., Pardo-Díaz, C., Salazar, C.** 2021. Environmental drivers of diversification and hybridization in neotropical butterflies. *Frontiers in Ecology and Evolution* 9.
- Thawornwattana, Y., Seixas, F. A., Yang, Z., Mallet, J.** 2023. Major patterns in the introgression history of *Heliconius* butterflies. *eLife* 12.