

# Infomap Bioregions: Interactive mapping of biogeographical regions from species distributions

Daniel Edler,<sup>1,2,\*</sup> Thaís Guedes,<sup>2,3,4</sup> Alexander Zizka,<sup>2</sup> Martin Rosvall,<sup>1</sup> and Alexandre Antonelli<sup>2,5</sup>

<sup>1</sup>*Integrated Science Lab, Department of Physics,  
Umeå University,  
SE-901 87 Umeå, Sweden*

<sup>2</sup>*University of Gothenburg,  
Department of Biological and Environmental Sciences,  
Box 461,  
SE-405 30 Göteborg, Sweden*

<sup>3</sup>*Federal University of São Paulo,  
09972-270, Diadema,  
Brazil*

<sup>4</sup>*Museum of Zoology of University of São Paulo,  
04263-000, São Paulo, Brazil*

<sup>5</sup>*Gothenburg Botanical Garden,  
Carl Skottsbergs gata 22A,  
SE-413 19 Gothenburg, Sweden*

Biogeographical regions reveal how species are spatially grouped and therefore are important units for conservation, historical biogeography, ecology and evolution. Several methods have been developed to identify bioregions based on species distribution data rather than expert opinion. One approach successfully applies network theory to simplify and highlight the underlying structure in species distributions data. However, there are no tools that make this methodology simple and efficient to use. Here we present Infomap Bioregions, an interactive web application that inputs species distribution data and generates bioregion maps. Species distributions may be provided as georeferenced point occurrences or range maps, and can be of local, regional or global scale. The application uses a novel adaptive resolution method to make best use of often incomplete species distribution data. The results can be downloaded as vector graphics, shapefiles or in table format. We validate the tool by processing large datasets of publicly available species distribution data of the world's amphibians using species ranges, and mammals using point occurrences. Potential applications include ancestral range reconstructions in historical biogeography and identification of indicator species for targeted conservation.

## Introduction

The Earth's biodiversity is unevenly distributed, and different regions contain different groups of species. Depending on the scale, taxonomic group and tradition, these broadly used biogeographical regions have received dozens of names, such as realms, biomes and ecozones, but here we simply refer to them as bioregions. In many disciplines, working with bioregions rather than single species is more effective. Conservation biology is a prime example, since protecting bioregions with high levels of biodiversity or uniqueness protects many species from extinction. In historical biogeography, bioregions may be used as operation areas for ancestral range reconstructions, in order to estimate how lineages in a phylogeny have evolved their geographical occupancy over time (Goldberg *et al.*, 2011; Matzke, 2014; Ree and Smith, 2008). Since different taxa exhibit different patterns of diversity, distribution and evolutionary history, the system under study and research question at hand will determine the best set of bioregions. Accordingly, in absence of one-size-fits-all bioregions, researchers depend on simple, effective and flexible tools for mapping bioregions.

While bioinformatic tools now can provide rapid and accurate coding of species into predefined areas (Töpel *et al.*,

2014; Zizka and Antonelli, 2015), choosing the areas in the first place has been a subjective procedure without quantitative support. Researchers have therefore developed a suite of algorithms for mapping grid cell areas into biologically relevant regions (Kozak and Wiens, 2006; Kreft and Jetz, 2010; Oliveira *et al.*, 2015), but often they involve multiple and overly technical steps. As a consequence, most biogeographical studies still use arbitrarily defined areas.

To make mapping of bioregions simple and effective, we present Infomap Bioregions, a web-based interactive mapping tool that identifies bioregions from species distributions. The underlying method clusters bipartite networks that contain both species and grid cells. This method was recently shown to outperform approaches that abstract away the species into species similarities between grid cells in unipartite networks (Vilhena and Antonelli, 2015). Moreover, the bipartite networks are clustered with the information-theoretic clustering algorithm known as Infomap (Rosvall and Bergstrom, 2008), which has been acclaimed as the best network clustering algorithm in several comparative studies (Aldecoa and Marín, 2013; Lancichinetti and Fortunato, 2009). Thanks to its simple and effective design, Infomap Bioregions has wide applications in biodiversity, conservation and related studies.

---

\*Electronic address: [daniel.edler@umu.se](mailto:daniel.edler@umu.se)

## Description

Infomap Bioregions is an interactive web application that identifies taxon-specific bioregions from species distribution data. We first present the application’s workflow, and then describe each step in detail.

Given user-provided species distribution data, the application first bins the data into geographical grid cells with adaptive spatial resolution. When the data are sparse, the grid size is large; and when the data are dense, the grid size is small. This novel adaptive resolution offers a considerable advantage over conventional uniform binning when dealing with biodiversity data, which is notoriously unevenly distributed (Maldonado *et al.*, 2015).

The binning generates a bipartite network between species and grid cells, which is then clustered with the Infomap algorithm into bioregions (Edler and Rosvall, 2015). The application also identifies the most common and the most indicative species in each grid cell and bioregion, and shows the results as an interactive map together with supporting tables with information about the bioregions. All results can be exported in various formats.

### Input data

As input, Infomap Bioregions supports both point occurrence data and species range maps. Point occurrences are specified in a text file with either comma-separated (CSV) or tab-separated (TSV) values. The application requires a header with the column names, and the user must identify which columns that should be parsed as name, latitude and longitude, respectively. Range maps are specified in the shapefile format, which includes multiple files: a .shp file for species range polygons, a .dbf file for the attributes of each range polygon and, optionally, a .prj file for projection information. As for point occurrence data, the user must identify which attribute to parse as the name of the species.

### Adaptive resolution and bipartite network

Infomap Bioregions bins the species records into quadratic grid cells. To allow for adaptive spatial resolution, each grid cell can be recursively subdivided into four cells. The adaptive binning generates a so-called quadtree with subdivided grid cells that satisfy the following user-specified criteria, with decreasing priority from 1 to 3:

1. Given in degrees, no grid cell is larger than the specified *maximum cell size* or smaller than the specified *minimum cell size*.
2. Given as a natural number, no grid cell contains fewer records than the specified *minimum cell capacity*.
3. Given as a natural number, no grid cell contains more records than the specified *maximum cell capacity*.

For point occurrence data, these criteria make the adaptive binning straightforward. For range maps, the application first adds a species record to each cell of minimum size that intersects with the corresponding species range polygon, and then proceeds with the adaptive binning to satisfy the user-specified criteria.

Given the bins, the application then generates a bipartite network of species and grid cells. Each species is connected by an unweighted link to each grid cell in which it is present. We purposefully avoid weighting the links by the number of records, because that would make the results sensitive to biased sampling. Instead, we use the density of species records to increase the spatial resolution as described above. In this way, dense data give large networks with high resolution.

### Bioregions and indicator species

Infomap Bioregions clusters the bipartite network with Infomap for bipartite networks (Kheirkhah *et al.*, 2015). The resulting clusters contain both grid cells and species and form the bioregions. The map now displays the bioregions with different colours and a table for each bioregion provides summary statistics and lists selected species.

The application lists for grid cells and bioregions both the most common species and the most indicative species with the highest relative abundance. That is, for species  $s$  in region  $r$ , grid cell or bioregion, the indicative score  $I_{s|r}$  is defined as the ratio between the frequency  $f_{s|r}$  of the species in the region and the frequency  $f_s$  of the species in all regions,  $I_{s|r} = f_{s|r}/f_s$ . Thus an indicative score of 2 means that a species is twice as frequent in the region than in the entire dataset.

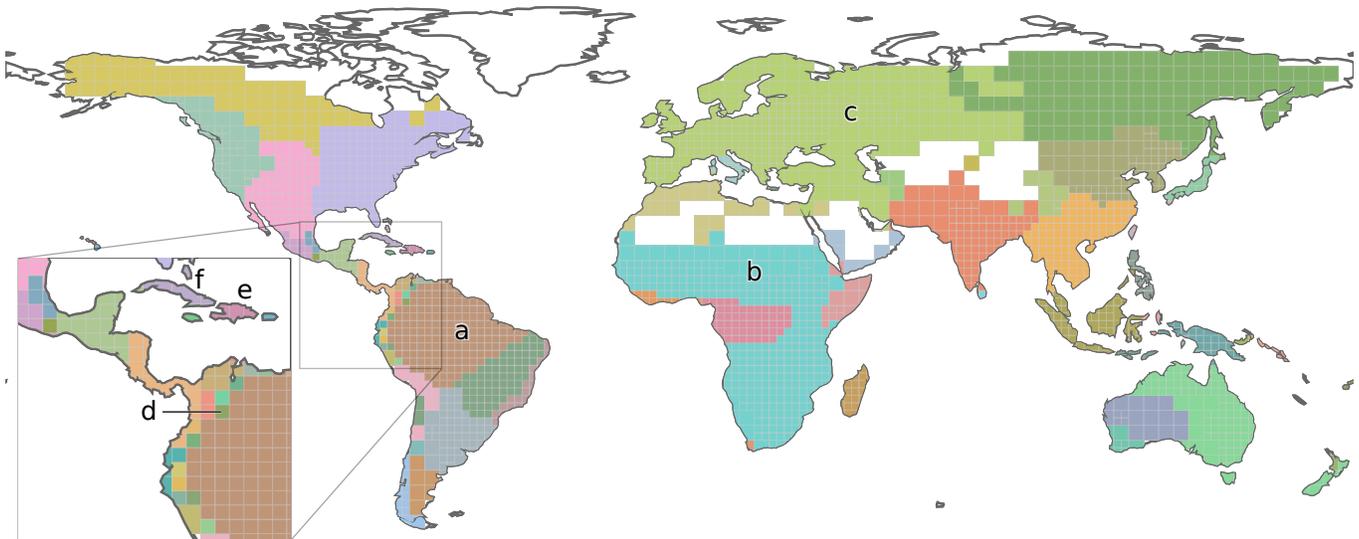
## Results and discussion

To validate Infomap Bioregions, we applied it to range maps of amphibians and point occurrences of terrestrial mammals. For amphibians, we downloaded global distribution data as range polygons for 6,069 species from the International Union for Conservation of Nature (IUCN, <http://www.iucn.org>). For terrestrial mammals, we compiled the global distribution of 4,362 species from a collection of georeferenced observation records obtained through the Global Biodiversity Information Facility (GBIF, <http://www.gbif.org>). We cleaned the mammal dataset in the R package `speciesgeocodeR` (Zizka and Antonelli, 2015), checking for obvious errors such as empty coordinates, sea species, terrestrial species reported in the sea, coordinates not belonging to the country they are reported from, etc.

We set the minimum cell size to 1 and the maximum cell size to 4, and allowed grid cells to range between 1° and 4° to reflect spatial differences in data density. Further, we set the minimum cell capacity to 5 and the maximum cell capacity to 100. That is, we allowed cells to include between 5 and 100 records. Below we show the bioregion maps of the amphibians and mammals, and highlight a few bioregions.

### Amphibians

We identified 87 bioregions of amphibians as illustrated in Fig. 1 (see Supplementary Materials for detailed results). In Table 1, we detail the three most species rich bioregions and three small bioregions with relatively many species. Most of the species belong to relatively large bioregions, but we also identified smaller bioregions such as in the Caribbean where island endemics are common and in the tropical Andes where species turnover is high and many species are located in just a few cells.



**Figure 1** | Bioregion map of the world's amphibians generated with Infomap Bioregions, using the IUCN species range maps. White areas have insufficient data and were excluded from the analysis. The inset shows a zoom-in of Central America, the West Indies and northwestern South America, depicting many small bioregions that reflect high species turnover and narrow range distributions characteristic for the region. Table 1 shows information about labelled bioregions.

**Table 1** | Selected amphibian bioregions. For exact locations, the indices (a)–(f) are displayed on the bioregion map in Fig. 1. Bioregions (a)–(c) are the most species rich and (d)–(f) are hand-picked to illustrate how even small bioregions can contain relatively many species. Common names from Encyclopedia of Life at <http://eol.org>

Location	Records	Species	Cells	Most common species	Most indicative species
(a) South America	42,161	719	167	<i>Trachycephalus venulosus</i> (600) Veined tree frog	<i>Lithobates palmipes</i> (3.3) Amazon River frog
(b) Africa	27,267	553	333	<i>Kassina senegalensis</i> (970) Senegal running frog	<i>Hildebrandtia ornata</i> (2.1) African ornate frog
(c) Europe and Asia	13,083	103	313	<i>Rana arvalis</i> (1,547) Moor frog	<i>Triturus cristatus</i> (1.3) Northern crested newt
(d) Andes	121	75	1	<i>Pristimantis nericus</i> (13) –	<i>Pristimantis nericus</i> (157) –
(e) Caribbean Islands	181	65	4	<i>Hypsiboas heilprini</i> (28) Los Bracitos tree frog	<i>Osteopilus vastus</i> (73) Hispaniola tree frog
(f) Cuba	214	61	4	<i>Osteopilus septentrionalis</i> (28) Cuban tree frog	<i>Eleutherodactylus varleyi</i> (73) –

The identified bioregions largely coincide with those found by Vilhena and Antonelli (2015), except for some differences due to the adaptive resolution and its settings. These results also coincide with the findings by the widely used classification of biomes proposed by Olson *et al.* (2001), which was defined on non-explicit biotic and abiotic data. For the Neotropics, our clustering also reflects the regionalization proposed by Morrone (2006, 2014) for some sub-regions and provinces such as the Amazonian subregion, the Parana subregion and the Chacoan dominion.

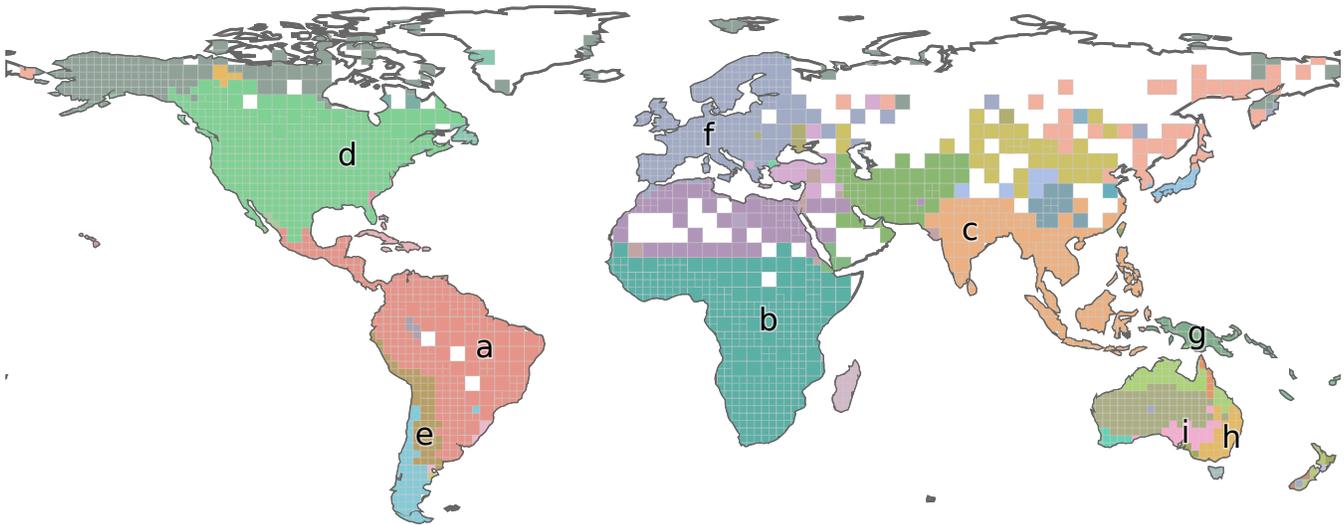
Infomap Bioregions also successfully identified small bioregions in, for example, the Caribbean region and the tropical Andes, which contain high diversity and species turnover, and therefore should be particularly considered for conserva-

tion. Other examples of small-scale bioregions include the Cape region in South Africa and the Dahomey gap in West Africa (Fig. 1). As a comparison, the method used by Holt *et al.* (2013) identified 15 bioregions compared with the in total 87 bioregions shown in Fig. 1.

### Mammals

We identified 60 bioregions of mammals as illustrated in Fig. 2 (see Supplementary Materials for detailed results). A few large bioregions cover most of the recorded area while smaller bioregions cover areas known to have high biodiversity. For example, more than 10 bioregions cover the Australian continent with its many endemic species (see Table 2).

The identified bioregions for mammals largely coincide with the zoo-geographical regions recognized by Holt *et al.*



**Figure 2** | Bioregion map of world mammals generated with Infomap Bioregions, using species point occurrences from GBIF. White areas have insufficient data and were excluded from the analysis. Table 2 shows information about labelled bioregions.

**Table 2** | Highlighted mammalian bioregions, sorted on species richness. For exact locations, the indices (a)–(i) are displayed on the bioregion map in Fig. 2. Common names from Encyclopedia of Life at <http://eol.org>

Location	Records	Species	Cells	Most common species	Most indicative species
(a) South America	76,403	1,479	282	<i>Glossophaga soricina</i> (1,584) Pallas's long-tongued bat	<i>Tapirus terrestris</i> (36) Brazilian tapir
(b) Africa	38,226	1,102	322	<i>Mastomys natalensis</i> (991) Common African rat	<i>Mus musculoides</i> (57) Kasai mouse
(c) Greater India	13,757	907	205	<i>Rattus exulans</i> (400) Polynesian rat	<i>Eonycteris spelaea</i> (143) Blanford's fruit bat
(d) North America	277,816	807	423	<i>Peromyscus maniculatus</i> (17,600) Deer mouse	<i>Dipodomys ordii</i> (3.2) Ord's kangaroo rat
(e) Andes	5,605	429	51	<i>Phyllotis xanthopygus</i> (296) Yellow-rumped leaf-eared mouse	<i>Akodon albiventer</i> (191) White-bellied grass mouse
(f) Europe	635,177	393	268	<i>Meles meles</i> (46,299) Eurasian badger	<i>Talpa europaea</i> (1.2) European mole
(g) New Guinea	5,107	325	46	<i>Syconycteris australis</i> (260) Southern blossom bat	<i>Echymipera kalubu</i> (220) Common Echymipera
(h) SE Australia	298,374	269	39	<i>Phascolarctos cinereus</i> (26,029) Koala	<i>Petaurus australis</i> (2.2) Yellow-bellied glider
(i) SE Australia	44,874	148	24	<i>Macropus robustus</i> (9,180) Hill wallaroo	<i>Petrogale xanthopus</i> (6.2) Yellow-footed rock-wallaby

(2013), despite the fact that we used point occurrences instead of species range polygons. We must acknowledge that the automated cleaning steps described above are not sufficient to fully validate the dataset. As a consequence, our results may be affected by sampling biases, inaccurate georeferenc-

ing or incorrect identifications. These issues prevent us from discerning, for example, whether the scattered occurrence of small bioregions in e.g. Russia is a real biological result or more likely an artefact of the publicly available data.

## Conclusions

Designed to make mapping of bioregions simple and effective, we introduced the web application Infomap Bioregions and demonstrated its flexibility. A user can load species data from both point occurrences and range polygons, modify parameters directly in the web interface and export results to various formats for high-quality printing or further biogeographical analyses. Moreover, the web application uses adaptive spatial resolution, can process millions of records in a few minutes and applies bipartite network clustering that outperforms traditional methods based on similarity indices. We validated the application on two large datasets of amphibians and mammals and anticipate that Infomap Bioregions will become a standard tool in many studies in ecology, evolution, conservation biology and historical biogeography.

## Availability and forthcoming extensions

Infomap Bioregions is distributed under the GNU AGPL v3+ license. It is written in JavaScript and builds on a set of open source libraries (see dependencies in `package.json`). Because it is a pure client-side application, all data stay and all calculations run on the user's computer. Moreover, all heavy calculations run in a background thread. This means improved privacy and performance.

Infomap Bioregions is available at <http://www.mapequation.org/bioregions> and the source code is freely available at <http://github.com/mapequation/bioregions>.

Possible forthcoming extensions include batch runs, additional methods to find indicator species and bioregions, hierarchical clustering of bioregions, significance clustering using the bootstrap and incorporation of phylogenetic information.

## Supplementary Materials

Supporting tables with detailed information about all bioregions are available at <https://github.com/mapequation/infomap-bioregions-supplementary-materials/raw/master/appendix.pdf>.

## Funding

This work was supported by the São Paulo Research Foundation (2013/04170-8 and 2015/18837-7 to T.B.G.); the Swedish

Research Council (B0569601 to A.A. and 2012-3729 to M.R.); the European Research Council under the European Unions Seventh Framework Programme (FP/2007-2013 and ERC Grant Agreement n. 331024 to A.A.); and a Wallenberg Academy Fellowship to A.A.

## References

- Aldecoa, R., and I. Marín, 2013, *Sci. Rep.* **3**, 2216.
- Eidler, D., and M. Rosvall, 2015, The infomap software package, <http://www.mapequation.org>.
- Goldberg, E. E., L. T. Lancaster, and R. H. Ree, 2011, *Syst. Biol.* **60**(4), 451.
- Holt, B., J. Lessard, M. Borregaard, S. Fritz, M. Araújo, D. Dimitrov, P. Fabre, C. Graham, G. Graves, K. Jönsson, *et al.*, 2013, *Science* **339**(6115), 74.
- Kheirikhah, M., A. Lancichinetti, and M. Rosvall, 2015, arXiv:1511.01540.
- Kozak, K. H., and J. Wiens, 2006, *Evolution* **60**(12), 2604.
- Kreft, H., and W. Jetz, 2010, *J. of Biogeogr.* **37**(11), 2029.
- Lancichinetti, A., and S. Fortunato, 2009, *Phys. Rev. E* **80**(5), 056117.
- Maldonado, C., C. I. Molina, A. Zizka, C. Persson, C. M. Taylor, J. Albn, E. Chilquillo, N. Rnsted, and A. Antonelli, 2015, *Global Ecol. Biogeogr.* **24**(8), 973, ISSN 1466-8238, URL <http://dx.doi.org/10.1111/geb.12326>.
- Matzke, N. J., 2014, *Syst. Biol.* , syu056.
- Morrone, J. J., 2006, *Annu. Rev. Entomol.* **51**, 467.
- Morrone, J. J., 2014, *Zootaxa* **3782**(1), 1.
- Oliveira, U., A. D. Brescovit, and A. J. Santos, 2015, *PLoS ONE* **10**(1), e0116673, URL <http://dx.doi.org/10.1371/journal.pone.0116673>.
- Olson, D. M., E. Dinerstein, E. D. Wikramanayake, N. D. Burgess, G. V. Powell, E. C. Underwood, J. A. D'amico, I. Itoua, H. E. Strand, J. C. Morrison, *et al.*, 2001, *BioScience* **51**(11), 933.
- Ree, R. H., and S. A. Smith, 2008, *Syst. Biol.* **57**(1), 4.
- Rosvall, M., and C. T. Bergstrom, 2008, *Proc. Natl. Acad. Sci. USA* **105**(4), 1118.
- Töpel, M., M. F. Calió, A. Zizka, R. Scharn, D. Silvestro, and A. Antonelli, 2014, *bioRxiv* , 009274.
- Vilhena, D. A., and A. Antonelli, 2015, *Nature Comm.* **6**.
- Zizka, A., and A. Antonelli, 2015, *BioRxiv*. doi: <http://dx.doi.org/10.1101/032755> .