

Biodiverse, a tool for the spatial analysis of biological and related diversity

Author:

Laffan, Shawn; Lubarsky, Eugene; Rosauer, Dan

Publication details:

Ecography

v. 33

Chapter No. 4

pp. 643-647

1600-0587 (ISSN)

Publication Date:

2010

Publisher DOI:

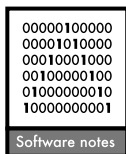
<http://dx.doi.org/10.1111/j.1600-0587.2010.06237.x>

License:

<https://creativecommons.org/licenses/by-nc-nd/3.0/au/>

Link to license to see what you are allowed to do with this resource.

Downloaded from <http://hdl.handle.net/1959.4/45539> in <https://unsworks.unsw.edu.au> on 2024-04-29



Biodiverse, a tool for the spatial analysis of biological and related diversity

Shawn W. Laffan, Eugene Lubarsky and Dan F. Rosauer

S. W. Laffan (Shawn.Laffan@unsw.edu.au) and E. Lubarsky, School of Biological, Earth and Environmental Sciences, Univ. of New South Wales, Australia, 2052. – D. F. Rosauer, School of Biological, Earth and Environmental Sciences, Univ. of New South Wales, Australia, 2052, and Centre for Plant Biodiversity Research, CSIRO Plant Industry, Australia, 2601, and Visiting Fellow, Centre for Macroevolution and Macroecology, Research School of Biology, Australian National Univ., Australia, 0200.

Biodiverse is a tool for the spatial analysis of diversity using indices based on taxonomic, phylogenetic and matrix (e.g. genetic dissimilarity) relationships. The explosion in georeferenced biological specimen and survey data means there is an increasing need for such tools. Biodiverse supports four processes: 1) linked visualisation of data distributions in geographic, taxonomic, phylogenetic and matrix spaces; 2) spatial moving window analyses including richness, endemism, phylogenetic diversity and beta diversity; 3) spatially constrained agglomerative cluster analyses; and 4) randomisations for hypothesis testing. Biodiverse is open-source and supports user developed extensions. It can be used both through a graphical user interface and scripts. Biodiverse can be downloaded from <www.purl.org/biodiverse>.

The spatial analysis of biological diversity is central to biogeography and ecology. The increased availability of georeferenced specimen and survey data, along with phylogenetic trees and matrix-based relatedness data, means it is now possible for researchers to analyse the spatial patterns of a broad array of biodiversity indices. Examples include species richness, taxonomic range restriction and endemism (Crisp et al. 2001, Laffan and Crisp 2003), genetic diversity (Bickford et al. 2004), phylogenetic diversity (Faith 1992), phylogenetic endemism (Rosauer et al. 2009), and species turnover (Whittaker 1960), as well as correlations of these patterns with environmental and other related conditions (Bickford and Laffan 2006, Ferrier et al. 2007). Biodiverse is a software tool developed to perform these and related analyses.

Biodiverse implements commonly used diversity, endemism and dissimilarity measures as well as new measures not previously available in published software, such as phylogenetic endemism (Rosauer et al. 2009). It also provides a simple platform for data exploration, such as viewing species distributions and patterns of species turnover, and exploring phylogeographic relationships across geographic space by clicking branches on a phylogenetic tree. The range of features in Biodiverse complement or build on previously released systems used in analysis of biodiversity pattern, for example Worldmap (Williams 1996), PATN (Belbin 2004), Diva GIS (Hijmans et al. 2005) and SAM (Rangel et al. 2006).

Biodiverse is a flexible system that can be used either through a graphical user interface (GUI) or user developed

scripts. The functionality can be extended by users, thereby enabling the rapid incorporation and development of new methods and indices, for example for the analysis of functional or phenotypic diversity. Biodiverse is open source software, released under the GNU Lesser General Public License. It is written in the Perl programming language, with the graphical user interface written using GTK (<www.gtk.org>). It will work on any computer platform for which these have been implemented, including Windows, Linux/UNIX and Macintosh.

While Biodiverse has been principally developed to analyse the spatial diversity of biological data, the basic principles of diversity analysis are generic and can be applied to any appropriate subject, for example lexical diversity (Jones and Laffan 2008). Similarly, spatial analyses can be applied using any data space, including geographic and environmental such as with rainfall and temperature. For this reason, while sample data in Biodiverse often comprise species in cells mapped across a study area, the units of analysis are referred to more generically as labels and groups.

Groups are typically rectangular cells to which the set of labels (e.g. specimen data) that occur within them are aggregated (binned). However, groups can also be points or based on nominal classes (for example bioregions, administrative units or source database name). They can also be a mixture of each type, with no limit on the number of axes used to define them. In this way one can represent data using, for example, two axes in degrees of latitude and longitude, a third axis being rainfall, and a fourth being bioregion name. This allows exploration of variation in

geographic and environmental spaces, with additional stratification by bioregion.

Labels typically represent species, either as a single name or as a combination of their taxonomic ranks. However, they can be anything identified as occurring within a group, for example operational taxonomic units, individual biological specimens, items in a word list, environmental variables, or times associated with events at a location (e.g. fires and cyclones). This approach to data representation makes the system extremely flexible when the data are analysed.

An important step when dealing with any data is to explore their distributions (Fig. 1). Biodiverse enables the simultaneous visualisation of data distributions in geographic, tree and matrix spaces. A typical use of this might be to visualise clade–area relationships. For example, a user may click on a species name (label) to see its location on the associated map (group plot) and phylogeny (tree), or click a clade on the phylogeny to see the distribution of its constituent taxa. Each group on the map and node on the tree can also be interrogated to list what is found within it. Users can also plot a geographic information systems (GIS) layer (currently polygon or polyline shapefiles) on the map to aid in interpretation.

Moving window analyses are a common form of spatial analysis. In this approach a predefined window is applied around each group in the data set, with some appropriate statistic calculated for each group based on the neighbours that fall within the moving window. By varying the size of the windows one can gain an understanding of the effects of the scale on the statistic being calculated (Laffan and Crisp 2003, Bickford et al. 2004).

The moving window analysis results are displayed as a map (Fig. 2). Hovering the mouse over a group will highlight the set of groups used in the neighbourhood(s) for that group. The user can also click on the groups to interrogate the analysis results, for example to see the set of species and the weights used in an assessment of endemism for a particular group.

The definition of neighbourhoods used in the moving window analysis is fundamental to spatial analysis and can have important influences on the results (Laffan 2002). Biodiverse supports a suite of standard geometric shapes, applied to as many axes as are used to define the groups. Following the principles of map algebra (Tomlin 1990), these include overlapping focal windows (e.g. circles, squares and ellipses), non-overlapping blocks, and zones (where one or more of the group axes is text based and

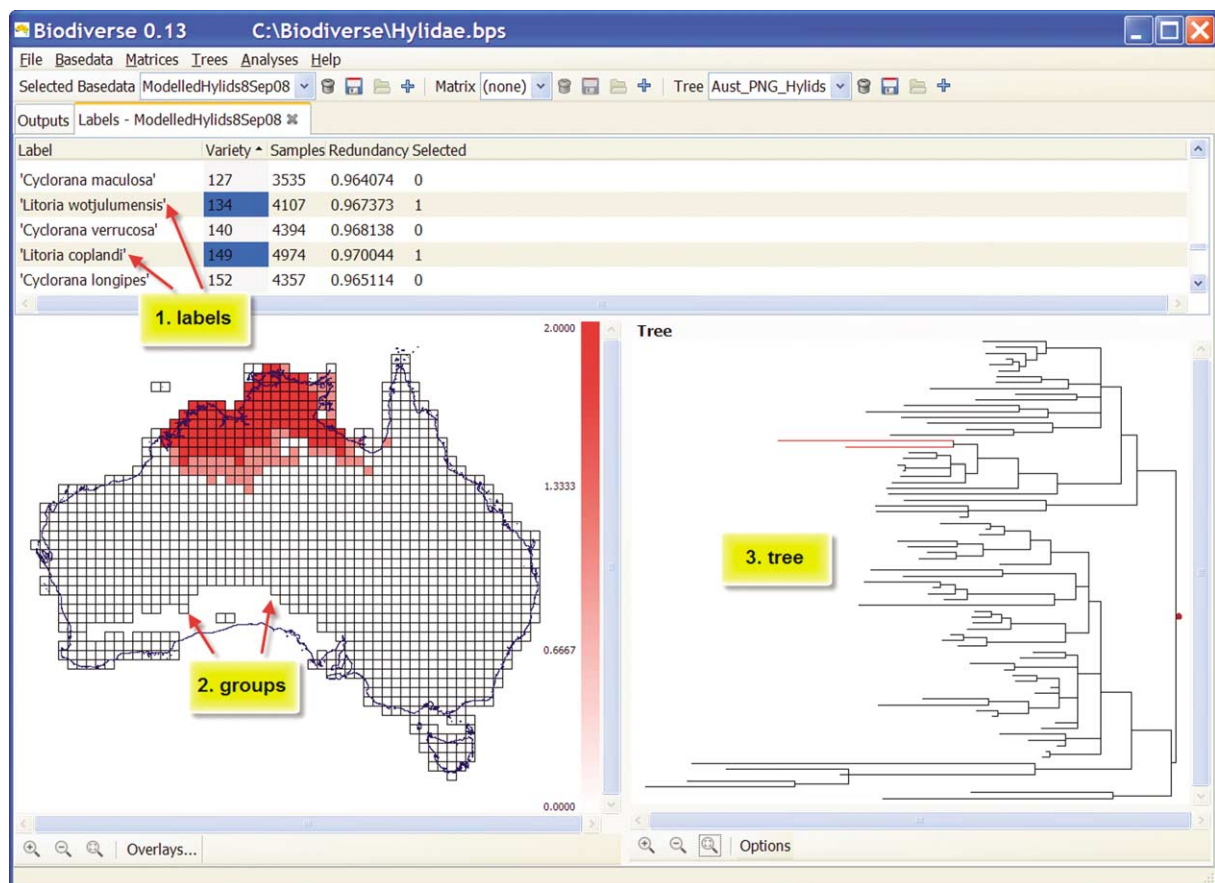


Figure 1. The Biodiverse data visualisation window displaying three linked views of a dataset. In each pane, the same two taxa are highlighted: 1) the labels, where the species or other entities are listed by name; 2) a map showing groups for which data were loaded. In this case the groups are 0.75 degree grid cells. The red shading indicates the relative number of the selected species (labels) recorded in each cell (group); 3) a tree showing a phylogeny for these taxa, with a clade of two species selected. Selecting an element in any pane highlights the corresponding elements in the other panes. The data shown are for Australian hylid frogs, as described in Rosauer et al. (2009). A GIS layer of a coastline is displayed for context.

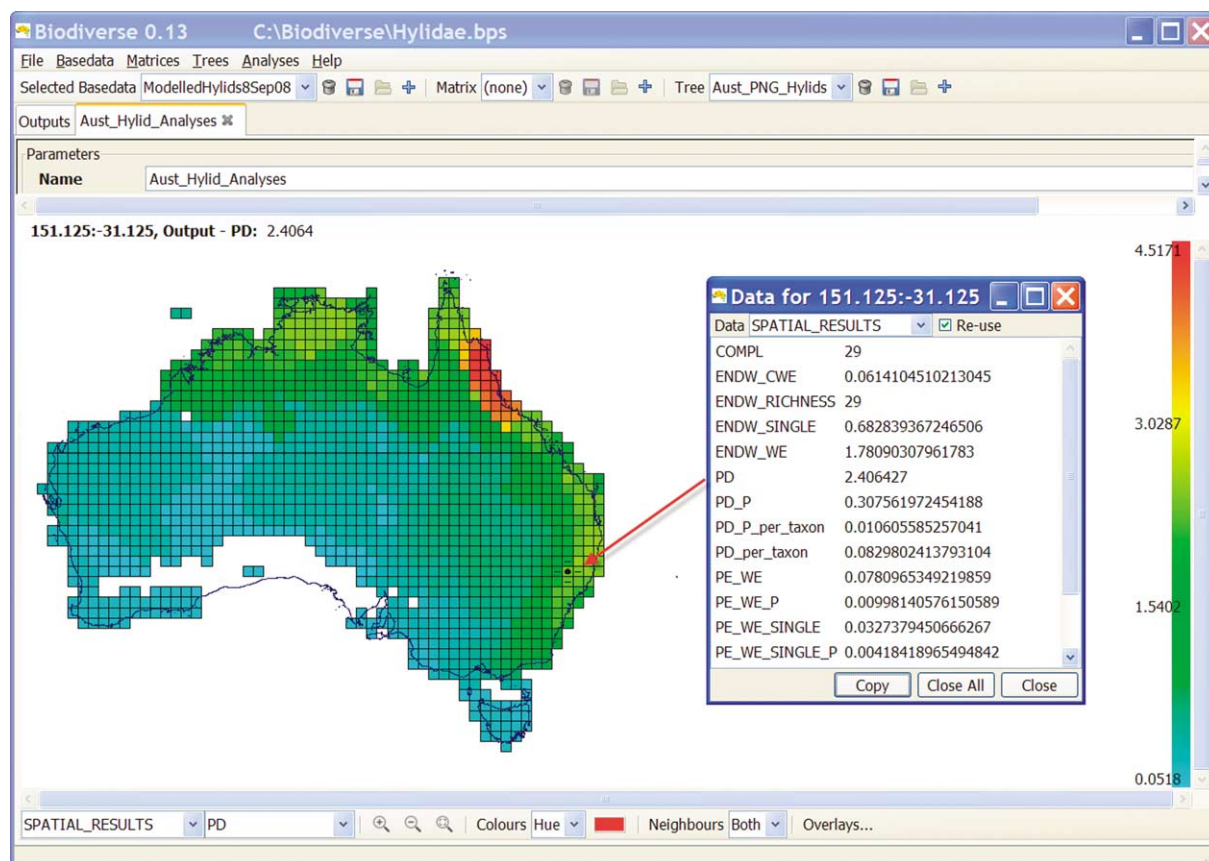


Figure 2. The Biodiverse analysis results window, displaying phylogenetic diversity scores. The spatial window used for this analysis is illustrated by the dot surrounded by four dashes. A popup window shows results of various analyses for a neighbourhood highlighted on the map.

describes, for example, bioregions or political units). There is also a notation to allow users to specify arbitrarily complex or adaptive windows, for example to vary window sizes by latitude. The effect of spatial scale can be explored by varying the sizes of the windows. Coarser and finer resolutions can be assessed by importing data at a different spatial resolution, or can be modelled using non-overlapping block windows.

Biodiverse supports the use of either one or two window definitions within an analysis. Two definitions are needed for analyses such as species turnover, where one compares the labels in one neighbourhood with those in another to assess the degree of change. Such window definitions do not need to overlap within an analysis, and any groups contained in the first window are excluded from the second to avoid double counting.

The set of indices that can be calculated using Biodiverse is large (see for example Legendre and Legendre 2000) and user-extendable. For species and related data, Biodiverse currently supports analyses including richness, beta diversity, endemism, rarity, phylogenetic diversity, phylogenetic endemism and genetic diversity (using matrices of genetic dissimilarity). For data sets where the labels represent numeric values, for example quantitative environmental data, the indices include summary statistics, while those numeric labels with temporal meaning can be assessed using inter-event interval statistics. Indices can be based on presence only data, sample counts or even local ranges

within the neighbourhood (as used for corrected weighted endemism in Laffan and Crisp 2003).

A simple block-based spatial index is used to reduce processing time by reducing the set of groups to be assessed for each spatial neighbourhood. Indexing delivers a significant performance gain for data sets with large numbers of groups. The exact gain depends on the nature of the data set, the index resolution, and the size of the specified neighbourhood.

Agglomerative clustering is commonly used in biodiversity assessments to identify a hierarchy of regions containing similar species complements or similar environments (Mackey et al. 2008). Clustering in Biodiverse can be based on a range of dissimilarity measures such as Jaccard, Sorenson, and Bray-Curtis. Several cluster linkage methods are supported to define the linkages between merged groups or clusters and the remainder of the data set, including average, minimum, maximum and recalculated (Bickford et al. 2004).

The clustering algorithm in Biodiverse is based on spatial principles, using the same underlying system as for the moving window analyses. The default is to use all pair-wise comparisons when building the dissimilarity matrix, clustering the most similar groups regardless of their geographic location. However, the user can specify that only groups within some neighbourhood around each group be considered. This allows the clustering to act as a region growing tool and is also useful when one wishes to exclude pairs

from clustering when they are separated by long distances. A hierarchy of spatial constraints can be specified to ensure the solution is spatially complete.

The groups used in the clustering are the same as those used in the moving window analyses. This means that the set of groups under each cluster node can be treated as a non-geometric window. All indices available for the moving window analyses that require only one neighbourhood can therefore be calculated for each of the cluster nodes. In this way one can calculate, for example, the degree of endemism or species richness for the set of groups that occur under a cluster node. This approach can be considered as a way to define neighbourhoods based on the properties of the data rather than some predefined window applied consistently across the data set.

Clustering results are displayed as a map linked to a tree in which nodes represent the clusters of similar groups (Fig. 3). This allows the user to observe the effect of different levels of aggregation, both across the tree and under a selected cluster node, both for individual clusters (up to 13, Brewer 2005) and for indices calculated for each node. As with the moving window analyses, the user can click on a group to interrogate the set of results for that group. They can also click on the tree nodes to access the results for each cluster. Hovering the mouse over a group highlights the path on the tree from the group to the root node, while hovering on a node in the tree highlights the set of groups that occur under that node.

Spatial data commonly violate the assumption of independence between data points, thus making many standard tests of significance inappropriate. Randomisation analyses alleviate this issue (Manly 2007) and also allow the specification of a variety of null hypotheses. Biodiverse currently supports null models based on complete spatial randomness at the group level, where whole groups and their labels are randomly rearranged, as well as models structured to replicate the observed richness patterns in groups within user specified tolerances (Laffan and Crisp 2003, Slatyer et al. 2007). A separate pseudo-random number generator (PRNG) stream is used for each randomisation, and the PRNG state is stored to enable continuation of the sequence in additional iterations. Biodiverse also stores the PRNG starting seed to enable replication of results if needed.

The Biodiverse data and analysis results are stored in a native format for re-use between sessions. They can also be exported as tables in delimited text file format for import into GIS, database or statistical packages. Cluster results and other tree based structures can be exported to the Newick and Nexus file formats.

In summary, Biodiverse is a tool that has been developed to harness the considerable increase in data for georeferenced specimens, matrices of genetic dissimilarity, phylogenetic trees and environmental data, with a focus on their spatial analysis. It supports visualisation and analysis of the spatial relationships in and between these data, thus enabling

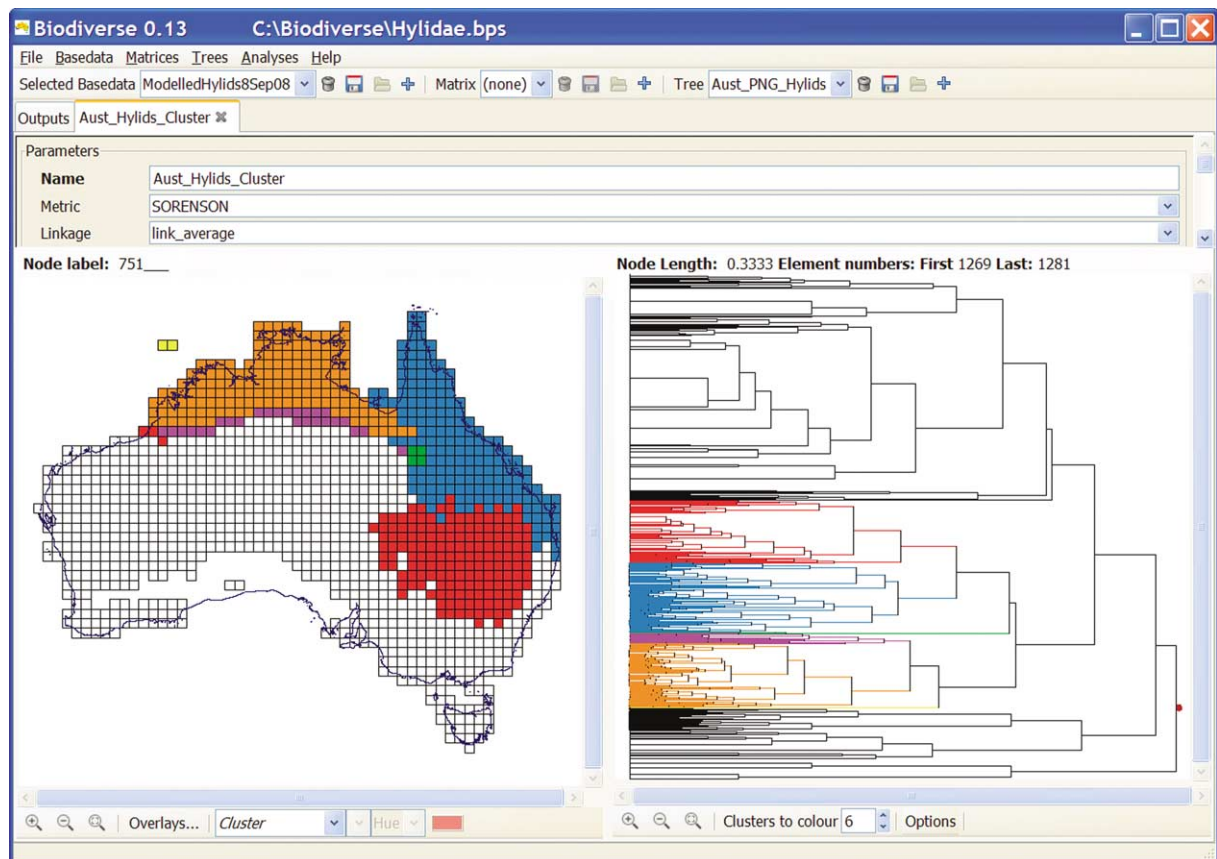


Figure 3. The cells (groups) on the map have been clustered based on a Sorenson distance. The tree on the right represents similarity between clusters of cells on the map (not between taxa). The user has clicked on a branch of the tree to highlight the branches beneath it, and the corresponding areas on the map. These are coloured after being divided into six sub-clusters based on their node lengths.

the exploration of questions related to a broad range of biodiversity applications in biogeography and ecology.

Download instructions for the software along with a user manual can be found online at the Biodiverse website <www.purl.org/biodiverse>.

To cite Biodiverse or acknowledge its use, cite this Software note as follows, substituting the version of the application that you used for “Version 0.13”:

Laffan, S. W., Lubarsky, E. and Rosauer, D. F. 2010. Biodiverse, a tool for the spatial analysis of biological and related diversity. – *Ecography* 33: 643–647 (Version 0.13).

Acknowledgements – This research was supported by ARC linkage grant LP0562070 and UNSW Faculty Research Grants Program funding to Laffan. Sophie Bickford and David Phillips provided important feedback on the software, while Kristen Williams, Marcel Cardillo and two anonymous reviewers provided valuable comments on earlier drafts of this manuscript.

References

- Belbin, L. 2004. PATN. – Blatant Fabrications, <www.patn.com.au/>.
- Bickford, S. A. and Laffan, S. W. 2006. Multi-extent analysis of the relationship between Pteridophyte species richness and climate. – *Global Ecol. Biogeogr.* 15: 588–601.
- Bickford, S. A. et al. 2004. Spatial analysis of taxonomic and genetic patterns and their potential for understanding evolutionary histories. – *J. Biogeogr.* 31: 1715–1733.
- Brewer, C. A. 2005. Designing better maps: a guide for GIS users. – ESRI Press.
- Crisp, M. D. et al. 2001. Endemism in the Australian flora. – *J. Biogeogr.* 28: 183–198.
- Faith, D. P. 1992. Conservation evaluation and phylogenetic diversity. – *Biol. Conserv.* 61: 1–10.
- Ferrier, S. et al. 2007. Using generalized dissimilarity modelling to analyse and predict patterns of beta diversity in regional biodiversity assessment. – *Divers. Distrib.* 13: 252–264.
- Hijmans, R. J. et al. 2005. Diva GIS. – <www.diva-gis.org/>.
- Jones, C. and Laffan, S. W. 2008. Lexical similarity and endemism in historical wordlists of Australian Aboriginal languages of the greater Sydney region. – *Trans. Philol. Soc.* 106: 456–486.
- Laffan, S. W. 2002. Using process models to improve spatial analysis. – *Int. J. Geogr. Inform. Sci.* 16: 245–257.
- Laffan, S. W. and Crisp, M. D. 2003. Assessing endemism at multiple spatial scales, with an example from the Australian vascular flora. – *J. Biogeogr.* 30: 511–520.
- Legendre, P. and Legendre, L. 2000. Numerical ecology – Elsevier.
- Mackey, B. G. et al. 2008. Reconciling approaches to biogeographical regionalization: a systematic and generic framework examined with a case study of the Australian continent. – *J. Biogeogr.* 35: 213–229.
- Manly, B. F. J. 2007. Randomization, bootstrap, and Monte Carlo methods in biology. – Chapman and Hall/CRC.
- Rangel, T. F. L. V. B. et al. 2006. Towards an integrated computational tool for spatial analysis in macroecology and biogeography. – *Global Ecol. Biogeogr.* 15: 321–327.
- Rosauer, D. F. et al. 2009. Phylogenetic endemism: a new approach to identifying geographical concentrations of evolutionary history. – *Mol. Ecol.* 18: 4061–4072.
- Slatyer, C. et al. 2007. An assessment of endemism and species richness patterns in the Australian Anura. – *J. Biogeogr.* 34: 583–596.
- Tomlin, C. D. 1990. Geographic information systems and cartographic modeling. – Prentice Hall.
- Whittaker, R. H. 1960. Vegetation of the Siskiyou mountains, Oregon and California. – *Ecol. Monogr.* 30: 279–338.
- Williams, P. H. 1996. Worldmap 4 windows: software and help document. – <www.nhm.ac.uk/eb/worldmap.shtml>.