Investigating diversity using *Biodiverse*

1 Introduction and Questions

The aims of this prac are to:

- Explore multi-species spatial patterns of diversity and endemism and consider the processes which may be responsible for these patterns
- Learn the meaning and applications of species and phylogenetic measures of diversity and endemism
- Understand how data on species distributions are derived, how diversity measures are generated, and the issues / uncertainties involved in these analyses.

Biodiverse is a free, open source computer program for spatial analysis of biodiversity.

It is described in: 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.

2 Installing Biodiverse on your computer

Biodiverse can be easily installed on Windows, Mac and Linux computers, following the **Installation Instructions**

2.1 Installing on Windows

- Go to the Biodiverse site here: http://shawnlaffan.github.io/biodiverse/ and click Download
- Under the heading **Stable release**, click on the download link for Windows.
- Create a new folder called /biodiverse, and unzip the files there.
- To launch *Biodiverse*, run the file BiodiverseGUI.exe which is found in the newly unzipped folder.

2.2 Installing on Mac

- Go to the Biodiverse site here: http://shawnlaffan.github.io/biodiverse/ and click Download
- Under the heading **Stable release**, click on the download link for MacOS.
- Unzip the files and drag and drop the biodiverse icon into the applications folder. See the Mac installation instructions on the Biodiverse site if needed.

2.3 Installing on Linux

- Go to the Biodiverse site here: http://shawnlaffan.github.io/biodiverse/ and click Download
- Under the heading **Stable release**, click on Linux.

There is also a helpful Quick Start Guide, although you shouldn't need it for this prac

Please install Biodiverse in advance of the session. If you have trouble, please contact Dan Rosauer for help on email dan.rosauer@anu.edu.au.

3 Sources of distribution information

For this exercise, all of the distribution data are provided.

However, for your own projects, it may be useful to consider the following sources of species location information

- Atlas of Living Australia
  www.ala.org.au
- Global Biodiversity Information Facility
  gbif.org
- IUCN Red List
  maps.iucnredlist.org
- Map of Life
  mol.org

4 Analysing species richness and endemism

4.1 Distribution data for Hylidae (Tree Frogs)

A dataset of frog species in the family Hylidae occurring in Australia is provided for your use. It is derived from:

- A species distribution model (SDM) for each species, which interpolates between locations where the species has been found, to predict the overall species range.
- The models were generated using a grid of 0.1 degree squares (roughly 11kmx11km), and thus predict the presence of each species for each 0.1 degree square across Australia
- Unzip the file: ModelledHylids8Sep08.zip. Have a brief look at the contents.

4.2 Start *Biodiverse* and import the species distributions

- To start Biodiverse open the /biodiverse folder and run BiodiverseGUI.exe

To import the Hylidae species ranges:

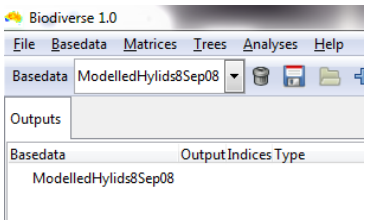
- From the Basedata menu, select Import
- Import: select the hylid distribution file ModelledHylids8Sep08.csv and click next
  Note: when opening a file, the **+** icon under the list of folders. Click this to bookmark a folder for quick access next time.
- Import Options: just click next, to accept the default values
- Choose columns: for each column in the file, you can choose how the column is used. Assign the columns as:
  - Species = Label (the names of the entities to map)
  - Longitude = Group
  - Latitude = Group (group the occurrence records by latitude and longitude)

- Set the Cell size to 0.25 for both latitude and longitude (define the grid cell size)
- Leave the other settings unchanged, click OK
- Re-order Columns. The groups order should be longitude then latitude (x then y).
- Click OK to start the import, which may take 1-2 minutes

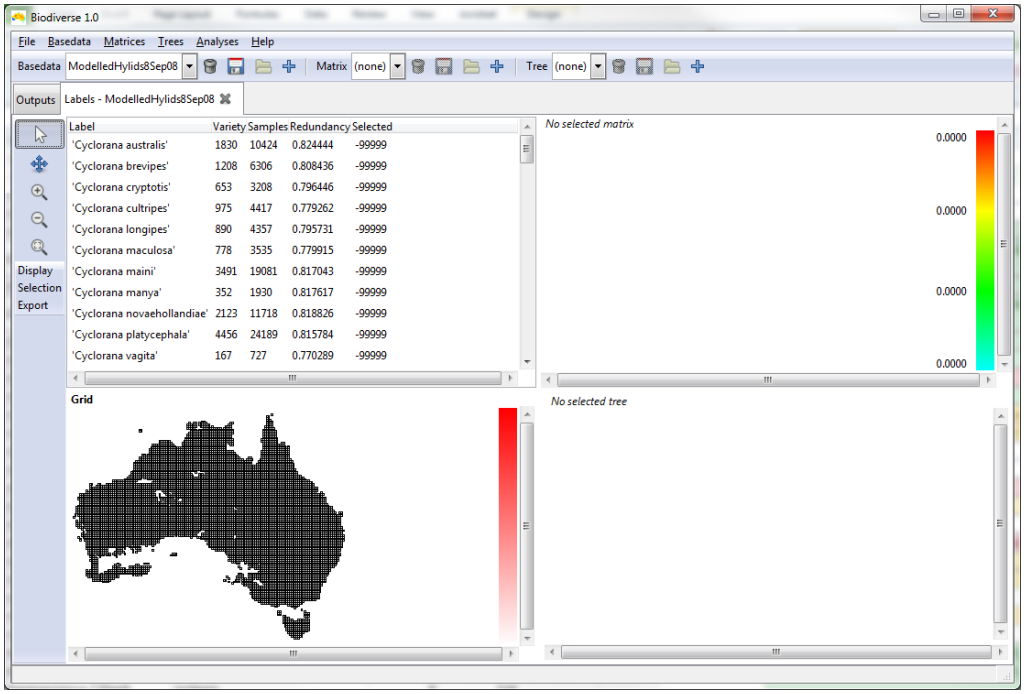
You have just summarised the occurrence data for Hylid frogs to a grid of 0.25 degree (roughly 27km) squares.

4.3 View the species distributions

- Select the name of the gridded spatial data set (basedata) that you just imported, and click Show.

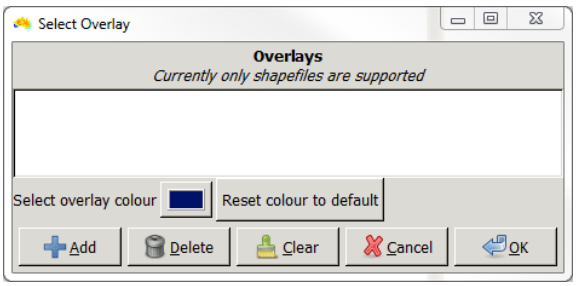


- You should now see a map with grid cells wherever species occur in this dataset. Like this:



- You can now:
  - Maximise the Biodiverse window for more space
  - Drag the boundaries between panes in the Biodiverse window to make more space for the map, and even drag until panes you are not using are completely hidden.

- Add a coastline overlay to the display:
    - Click the Display button to the left of the map, and choose Overlays.
    - A window like this will appear:



    - Click Add, then navigate to the folder with the data for the prac and select *aus5m_coast*.
    - This map layer should now appear in the Overlays window. Select it and click OK.
    - You should now have a coastline shown on the map.
      The coastline doesn't affect the analyses – just makes the map easier to interpret

- Explore the data:
    - Click or scroll through the species names to explore their ranges.
    - Click a grid cell, to see the number of the species in that cell which are found in each of the other cells.
    - Ctrl + click on a cell to get a species list for that cell
- If you're wondering…
    - Variety is the number of cells in which a species is recorded
    - Samples and Redundancy describe the number and density of known species locations, but aren't meaningful for this dataset, as it is based on modelled species ranges

4.4 Analyse richness and weighted endemism for species

Now calculate:

- **Species richness** – the number of species per grid cell
- **Weighted endemism** – like richness, but with each species weighted by the inverse of its range size. You can think of this as taking a value of one for each species, and dividing it equally across the cells in which the species occurs. A wide-ranging species will be spread thinly, while a narrow endemic has its value concentrated in a few cells.
  See Crisp, M., S. Laffan, et al. (2001). "Endemism in the Australian flora." Journal of Biogeography 28(2): 183-198.

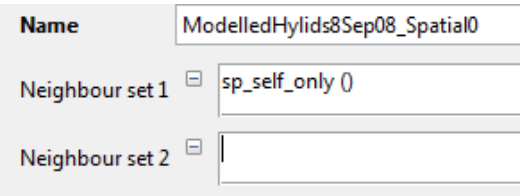
Steps:

- From the Analyses menu, select Spatial

Before running the analysis, there are 3 things you can set.

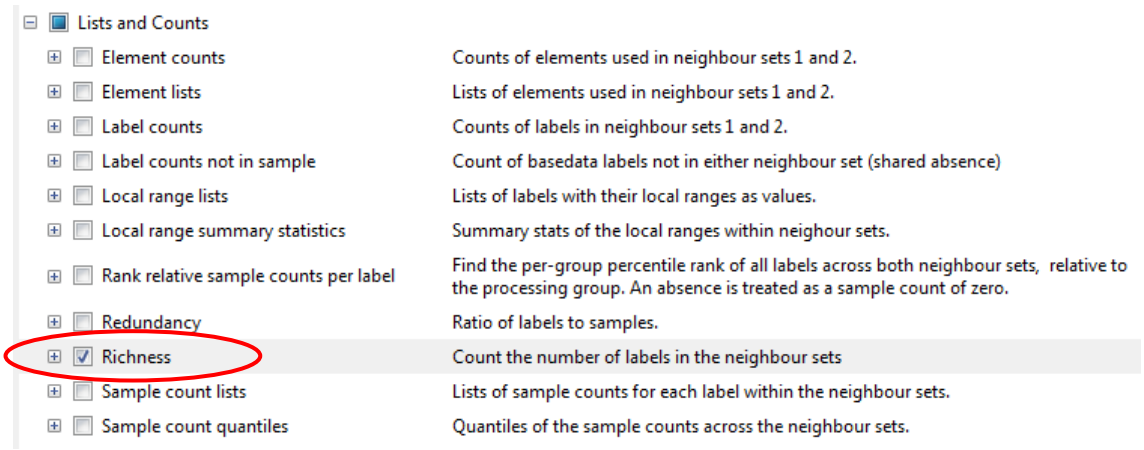
1. The analysis name – useful to change to describe your analysis, if you are running many. But you can also leave it as is.

2. The neighbour sets, define the neighbourhood of adjacent cells for which the measures are calculated, and allow the spatial scale of measures to vary.
   In this case, we will do the simplest neighbourhood, which is one cell at a time.
   - Leave neighbour set 1 unchanged, but clear the text from Neighbour set 2.

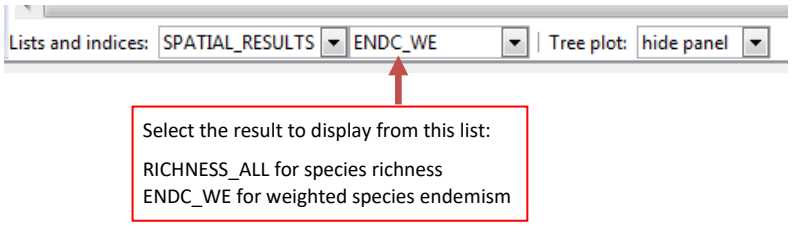


Biodiverse has a huge range of options for calculations to perform. They are detailed in the online help here: github.com/shawnlaffan/biodiverse/wiki/Indices. Don't be fazed by the long list, just select those you need.

3. Select two analyses from the list, using the triangles at the left to expand the categories.
   - From Lists and Counts select *Richness*
   - From Endemism select *Endemism central*



- All the settings are complete. Click *Go!* To run the analysis.
- Select *Yes* to display results, and you should get a colourful result map.



NOTE: the result list may default to ENDC_CWE. This is *corrected weighted endemism*, which is not used in this exercise. You want ENDC_WE, which is similar.

- Use the display bar to the left of the map to zoom in, pan or adjust colour options for your result map.

5 Load a phylogeny and link to the spatial data

The aim of this section is to link the spatial data to a phylogeny for these species, and see how the phylogenetic and spatial patterns are interrelated.

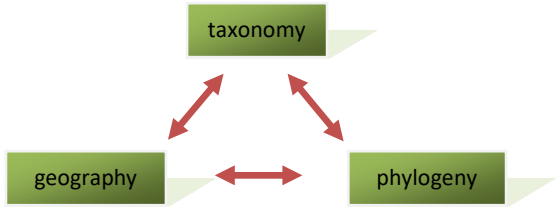
A phylogeny for 79 Australian Hylidae species is provided, adapted from:

> Rosauer, D., S. W. Laffan, et al. (2009). "Phylogenetic endemism: a new approach for identifying geographical concentrations of evolutionary history." Molecular Ecology 18(19): 4061-4072.

- Select *Open* from the *Trees* menu, and choose the file *Hylid phylogeny.bts.* A tree should appear to the right of the map.

5.1 Explore the phylogeny and geography together

You now have 3 ways of viewing the same taxa – by taxonomy, phylogeny and geography. Any selection on one is shown on the others.



- Move your mouse over the branches of the tree. As you hover over a branch, the species or clade represented by a branch, is blacked out on the map.
- Click on a grid cell, and the branches found there are highlighted.
- Select a species name, and its location is highlighted on the map and the tree.

6 Calculate phylogenetic diversity and endemism

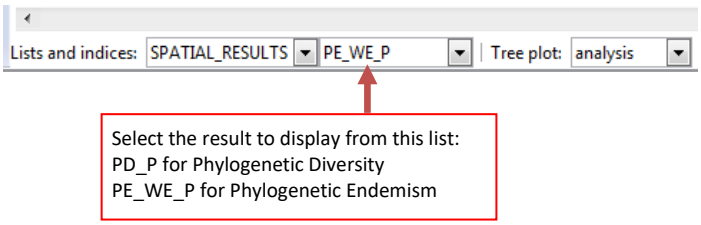
The next step is to calculate two phylogenetic measures of the distribution of evolutionary diversity:

- Phylogenetic diversity (PD), the sum of the branch lengths linking a set of species to the root of the tree.
  See Faith, D. P. (1992) Conservation evaluation and phylogenetic diversity. - Biological Conservation 61: 1-10.

- Phylogenetic Endemism (PE), endemism of PD. The degree to which units of evolutionary diversity, measured as branch lengths on the phylogeny, are restricted in a particular area. High PE is found where species which represent a lot of branch length on the phylogeny, have a small distribution range.
  See Rosauer, D., Laffan, S. W., Crisp, M. D., Donnellan, S. C. and Cook, L. G. 2009. Phylogenetic endemism: a new approach for identifying geographical concentrations of evolutionary history. - *Molecular Ecology* **18**: 4061-4072.

- Just as you did in section 4.4, from the Analyses menu, select Spatial to open a new analysis tab. Or go back to the one you were using before.
- Use the same neighbourhood settings as before.
- From the phylogenetic indices section of the calculations list, select *Phylogenetic Diversity* ;
- And from the phylogenetic endemism section select *Phylogenetic Endemism* (not any of its variants)
- All the settings are complete. Click *Go!* To run the analysis.

- Once the analysis is complete, you can explore the results as before:

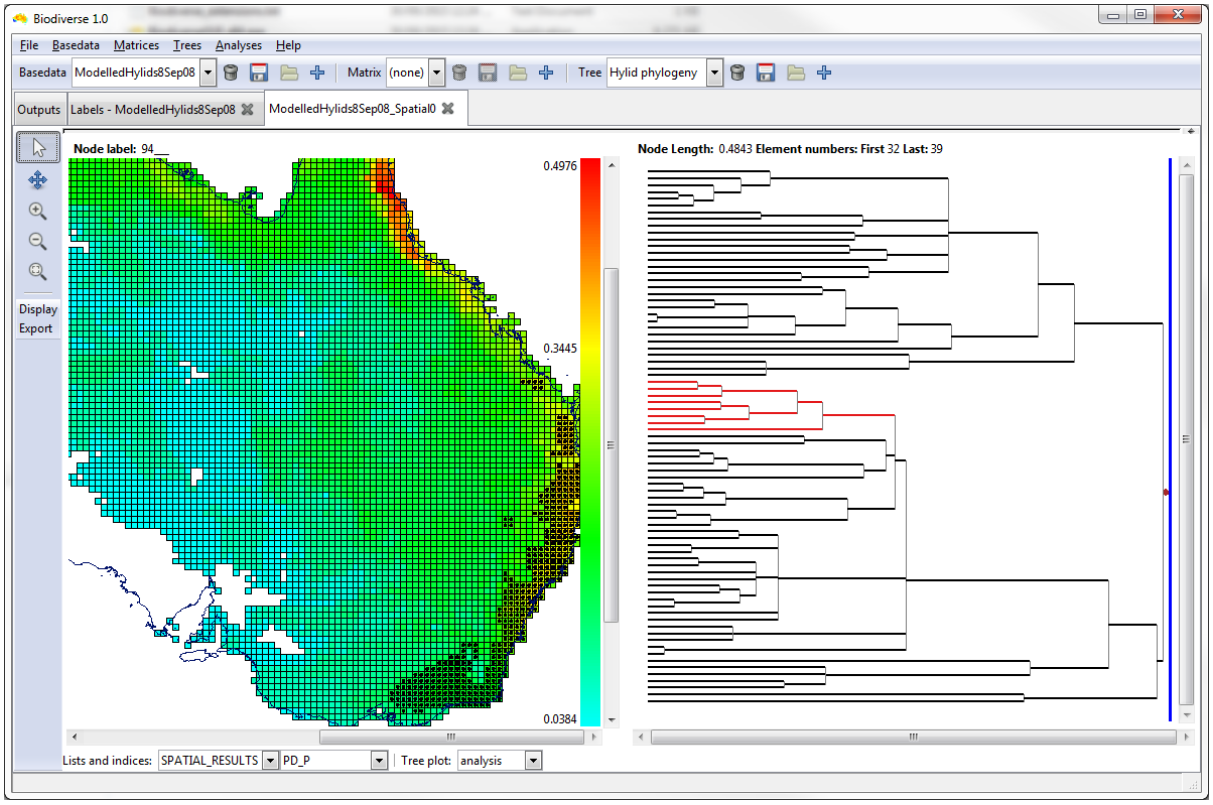


- The suffix '_P' on the indices to display stands for proportional. This means that these results are given as proportions of the total length of the tree. So for proportional measures:
    - PD scores are always between 0 and 1. A PD_P score of 0.33 means that the species in that cell represent 33% of the phylogenetic diversity on this tree.
    - The sum of all PE scores across all grid cells sum to 1.
- You will find the colour scaling options under the display menu helpful in exploring the wide range of values in the results

6.1 Delve deeper into the results

- Which area has the highest PE score?
- What species occur there?
- Look for this species on the phylogeny, and look it up on the web to find out more about the species
- So what gives this area its high score?

As you explore the results, hover and click on the tree or on the map, to help determine which species and clades of species are driving the scores on particular areas.



7 Defining biogeographic regions using beta diversity

In this step we identify biogeographic regions on the basis of their shared biota's or shared evolutionary history.

Pairwise beta diversity, also known as compositional dissimilarity, is a measure of the difference between two biotas, for example between the assemblages of species found in two areas. Typically, pairwise beta diversity varies from 0 for sites with identical species composition, to 1 for sites which share no taxa.

Here we use two measures of beta diversity:

- Sorenson species dissimilarity – to measure the degree to which the biotas of different areas share the same species
- Sorenson phylogenetic dissimilarity – to measure the degree to which the biotas of different areas share close related species

Using these metrics we will generate a site x site distance matrix comparing each grid cell to all others.

We can then use hierarchical clustering to identify compositionally similar areas.

This approach is similar to that used by:

Holt, B.G., Lessard, J.-P., Borregaard, M.K., Fritz, S.A., Araújo et al (2013) An Update of Wallace's Zoogeographic Regions of the World. *Science*, **339**, 74-78

and

Rosauer, D.F., Catullo, R.A., Vanderwal, J., Moussalli, A., Hoskin, C.J. & Moritz, C. (2015) Lineage range estimation method reveals fine-scale endemism linked to Pleistocene stability in Australian rainforest herpetofauna. *PLoS One*, **10**, e0126274

7.1 Reload the distribution data, gridded at a coarse 1 degree resolution

Pairwise comparison and clustering of all cells is computationally intensive, so for this session we will use larger grid cells to ensure a quick calculation.

- Load in the distribution data as you did in stage 4.2, but this time set the Cell size to 1 for latitude and longitude

7.2 Cluster the cells using the (species) Sørensen metric

- Select cluster from the analyses menu
- For metric select Sørensen
- Click Go! to run the analysis
  For 1 degree cells it should take 1 -2 minutes

7.3 Repeat using the phylo-Sørensen metric

- Again select cluster to create a new cluster tab in Biodiverse
- This time choose phylo-Sorensen from the list of metrics
- Before running edit the name of the analysis, to distinguish this table from species-Sorensen

7.4 Compare the results

How do the zones generated by species and phylogenetic beta diversity differ from each other. Why do they differ?