biomapME: biodiversity mapping and metrics, an introduction using TERN Ausplots vegetation data

Greg Guerin

2022-05-30

Introduction to biodiversity mapping

The spatial distribution of biodiversity is a topic of immense importance to understanding macroecological processes and conservation efforts based on limited resources. Biodiversity is generally concentrated into centres or hotspots characterised by high species richness. While hotspots of species richness have long been recognised, biodiverse areas can defined on the basis of more sophisticated metrics such as uniqueness or presence of threatened, range-restricted or complementary species and ecological communities.

Range rarity (i.e. rarity in terms of a restricted geographic distribution rather than population numbers per se) is a major consideration for conservation planning because it informs spatial management priorities and can be estimated from readily available incidence data (Crisp et al. 2001). Measures of area of occupancy (AOO), extent of occurrence (EOO), climatic range, or landscape-scale frequency are basic conservation metrics because species with more restricted ranges are considered at greater risk of extinction and contribute to biological uniqueness.

Added to strictly species-based metrics are the new wave of phylogeny-based ones. Phylogenetic diversity is a measure of the evolutionary history represented by a set of species in a community or samples of species in plots or map grid cells, while phylogenetic endemism weights PD by the relative geographic range restriction of parts of that history - phylogenetic branch lengths.

Biodiversity mapping as an endeavour is an empirical exercise seeking to represent patterns in species inventory data. However, this leads to immediate questions of how sampling influences the resulting patterns, especially given that large-scale biodiversity mapping studies typically use data from existing sources such as herbarium or museum records collected non-systematically and for a different purpose. Biases in the underlying data may influence the results.

Sampling issues include 1) uneven and under-sampling, resulting in biased richness estimates; and 2) the correlation of many endemism or phylogenetic metrics with species richness. These factors can make it difficult to disentangle metrics from sampling intensity and species richness. Plot-based data recording species according to systematic effort criteria make it possible to use the accumulation of species to estimate actual species richness from that observed, for example via rarefaction or non-parametric estimators. However, for the purposes of identifying relative patterns of biodiversity, it is not always necessary to correct metrics for sampling: it may be more relevant to ask whether metrics are higher than expected, given the context of sampling.

Introduction to biomapME

The **biomapME** R package contains a set of functions for manipulating, analysing and rasterising species inventory data for use in macroecological analysis and mapping of biodiversity metrics. The functions are primarily geared towards generating (and analysing) gridded biodiversity maps based on various metrics from species incidence (and environmental) data.

This vignette demonstrates the main functionality of the package using a step-by-step workflow starting with accessing data, installing the package, and generating a set of R data objects and subsequently maps and metrics of several biodiversity metrics. The package is designed (to some degree, depending on the application) to take either plot-based species records (in the form of a species x sites matrix and associated site coordinates matrix) or individual species records (in the form of a row per species observation and columns for species name and spatial coordinates).

The user can provide pre-existing raster maps if desired. If only species inventory data are available, rasters are automatically generated at the desired extent and resolution. Some site-based metrics are also calculated.

This demonstration makes use of TERN Ausplots a standardised continental survey of the species composition of Australian vegetation, which approaches 800 sites in extent www.tern.org.au. For comparison, incidence data are also sourced from Australia's Virtual Herbarium (URL).

Installation and getting started

To install the latest version of **biomapME**, use the devtools package, which must be installed first (I also recommend you install dependencies first), e.g.:

```
devtools::install_github("GregGuerin/biomapME", build_vignettes = TRUE, dependencies = FALSE)
```

For this vignette, we will use the tools of **biomapME** in conjunction with ecosystem monitoring data from TERN Ausplots, which is accessed through an API in R:

```
install.packages('ausplotsR')
```

First we load these packages:

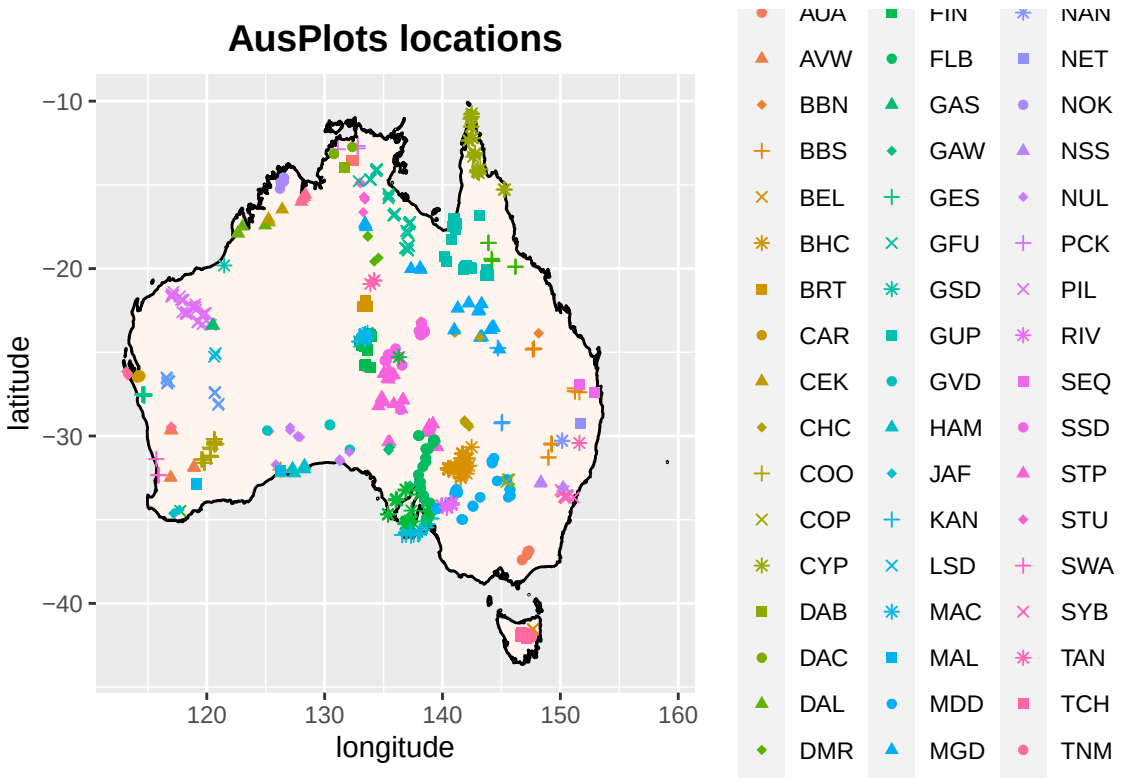
```r
library(ausplotsR)
library(biomapME)
```

We can now access data on plant species diversity (in this case, presence/absence) in a continental network of vegetation survey plots:

```r
# download site information and vegetation vouchers for all available Ausplots
my.ausplots <- try(get_ausplots(veg.PI = FALSE))
```

The returned `my.ausplots` object is a list that includes as elements the data.frames `$site.info` and `$veg.vouch`. We can use the former to map the sites and the latter create a species x sites matrix for further analysis:

```r
# automated map generation showing IBRA bioregions:
map_ausplots(my.ausplots)
```




```
# generate species x sites table
my.sppBYsites <- species_table(my.ausplots$veg.vouch, m_kind = "PA", species_name = "SN",
    strip_bryophytes = TRUE)
```

Species presence/absence in map grid cells

The above object can now be used in conjunction with spatial coordinates in the site.info table to compute and map biodiversity metrics. To start with, we can convert these presences and absences in plots to presences and absences in map grid cells of a raster layer. As we don't have a raster to base this on, we will ask the function to create one for us and return that raster plus presence/absence scores for each species per grid cell.

```
# species against grid cells (identified by their ID numbers in the
# automatically generated and returned raster):
cell.PA <- map.pa.matrix(my.sppBYsites, records = "site", site.coords = my.ausplots$site.info[my.ausplo
    rownames(my.sppBYsites), c("longitude", "latitude")], deg.resolution = c(2, 2))
#> Generating frame raster at  2 2  resolution and extent defined by:  113 153 -42 -10
#> Some point locations lie outside the frame raster -- trimming these records
#> Generating the gridded occurrence matrix
#> Occurrence matrix generated with dimensions:  100 5142
#> Collating occurrence matrix outputs
```

Convert plot/site data to individual species records

Next, we can use the data to generate a set of metrics relating to the size of species geographic ranges within the dataset. However, the relevant function to do that requires long format species records, so we will first need to convert our plot-based records:

```
mySPPrecords <- convert.site.data(my.sppBYsites, site.coords = my.ausplots$site.info[my.ausplots$site.i
    rownames(my.sppBYsites), c("longitude", "latitude")])
```

Species range sizes

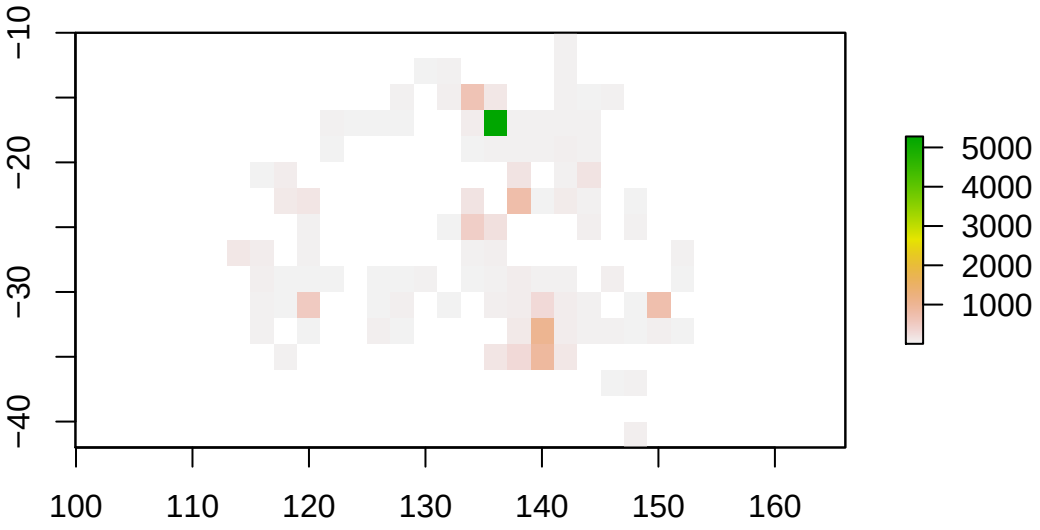
Now, we can calculate range sizes, in this case in terms of their recorded latitudinal range:

```
my.ranges <- spp.range.metrics(mySPPrecords, weight.type = "geo", geo.calc = "LAT",
    plot.out = FALSE, verbose = FALSE)

print(my.ranges$weights[101:110])
#>      Acacia.gonoclada      Acacia.hakeoides Acacia.hamersleyensis
#>              1.000000             21.692207              1.000000
#>      Acacia.hammondii    Acacia.harpophylla  Acacia.havilandiorum
#>              1.078713             23.168739             20.913277
#>     Acacia.hemignosta      Acacia.hemiteles       Acacia.hemsleyi
#>             10.843639             10.699656             19.740203
#>       Acacia.hilliana
#>              3.813333
```

Range rarity richness

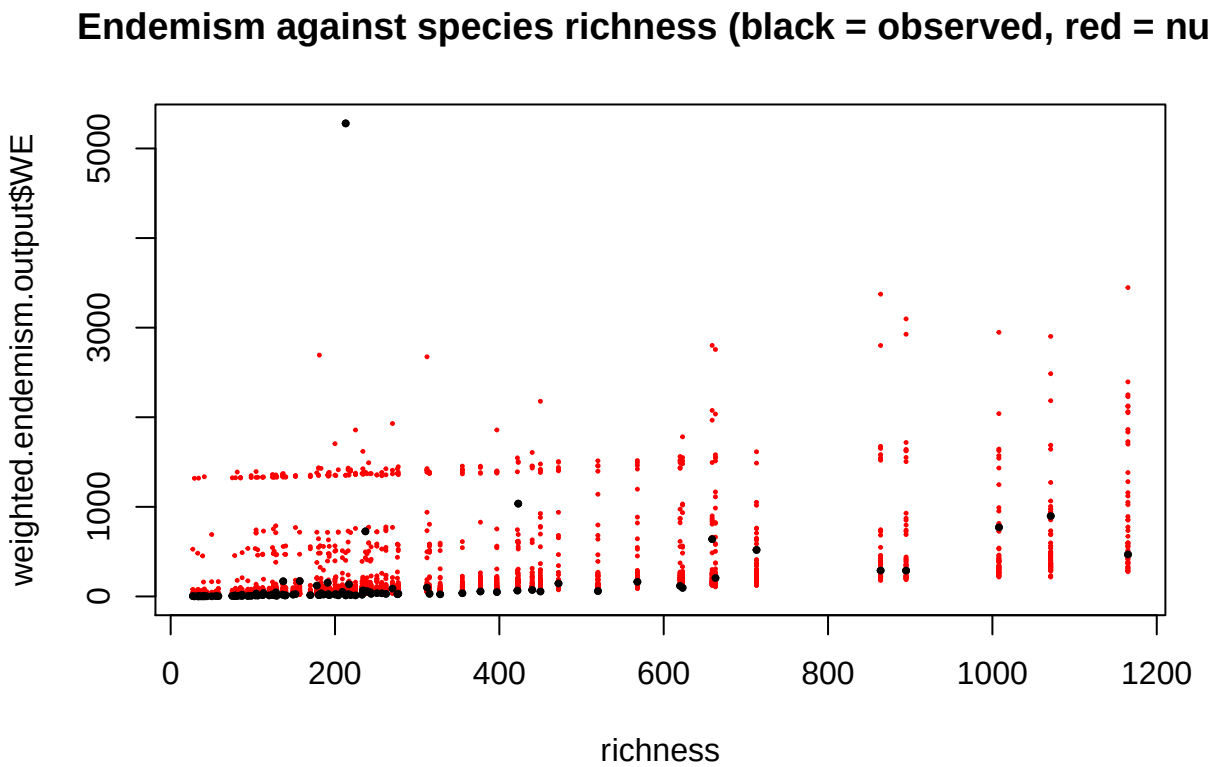
Given our range calculations above, we can now calculate and map species richness weighted by latitudinal range, using the `weighted.endemism` function, which by default calculates range rarity richness, or species richness weighted by the frequency of species in grid cells:

```
# this call will map richness of grid cells inversely weighted by species
# latitudinal range:
richness_lat_range_weight <- weighted.endemism(mySPPrecords, frame.raster = cell.PA$pa.raster,
    own.weights = my.ranges$weights[names(my.ranges$weights) %in% names(cell.PA$grid.matrix)],
    own.grid.matrix = cell.PA$grid.matrix, plot.raster = FALSE)
#> Some point locations lie outside the frame raster -- trimming these records
#> Reading user-defined gridded occurrence matrix
#> Calculating user supplied weights
#> Calculating weighted endemism
plot(richness_lat_range_weight$WE_raster)
```

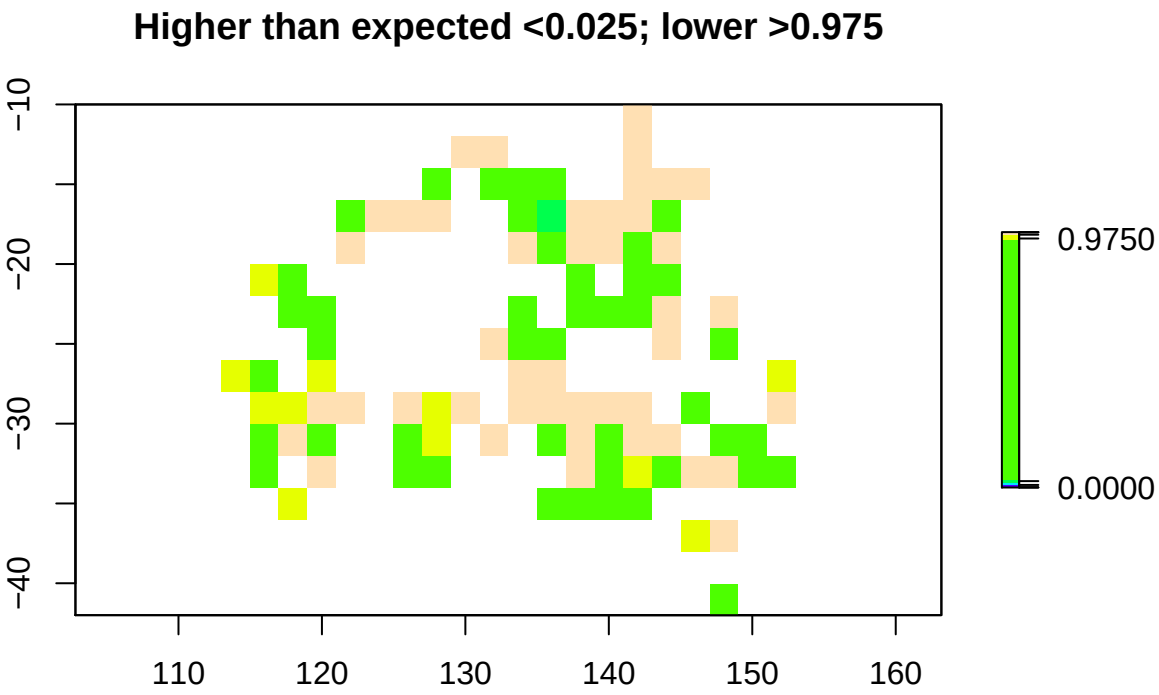


But hang on, this is just a map of species richness with a few tweaks isn't it? How do we map cells that have higher than expected concentrations of 'latitudinally restricted' species? Let compare the results to a null model by running the returned object in the `endemism.null.test` function:

```
my.null <- endemism.null.test(richness_lat_range_weight, nrep = 50, plot.all = FALSE)
#> Calculating species random sample probabilities and species richness
```



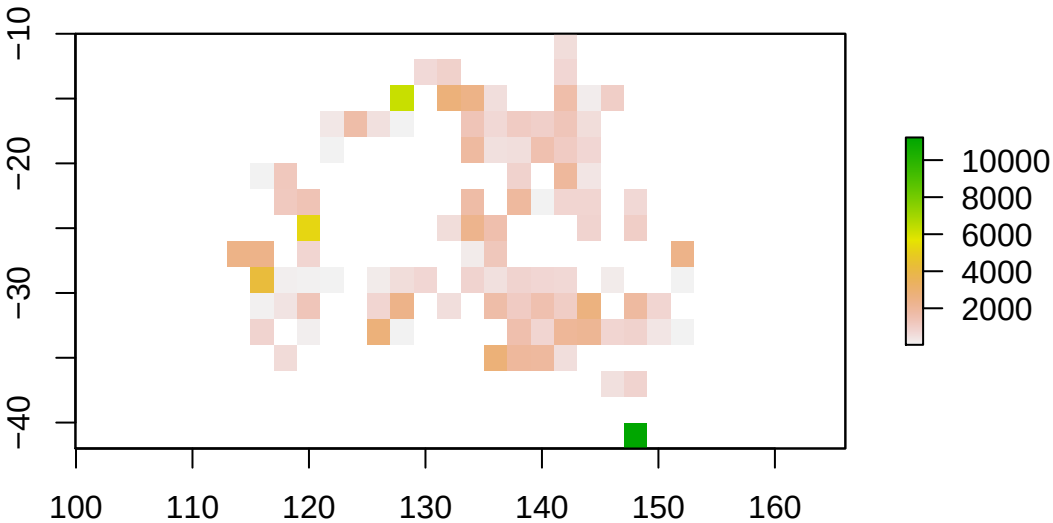

```
#> Generating null distributions for range of observed species richness:
#> Calculating outliers
#> Mapping outlier results
#> Calculating significance
#> Mapping significance results
#> Collating outputs

p.breaks <- c(0, 1e-04, 0.001, 0.01, 0.025, 0.975, 0.99, 0.999, 0.9999, 1)
plot(my.null$P.above.raster, breaks = p.breaks, col = topo.colors(9), main = "Higher than expected <0.0
```



Getting fancy - multi-plot metrics within grid cells or user-defined groups

Next, we use the added information content in these plot-based samples (as compared to presence-only records) to value-add to our biodiversity metrics using the function `map.multi.metrics`. This function groups field plots by the map grid cells they fall in and calculates several things - here we focus on non-parametric estimation of species richness or alpha diversity. That is, the species richness observed in each cell is probably (almost certainly!) under-estimated due to incomplete sampling, but we can use existing tools to estimate the real richness based on the number of occurrences of species in plots:

```
# the function can also do multi-site beta diversity and estimate phylogenetic
# diversity based on completeness of species sampling
estimated_richness <- map.multi.metrics(my.sppBYsites, site.coords = my.ausplots$site.info[my.ausplots$s
    rownames(my.sppBYsites), c("longitude", "latitude")], beta = FALSE, deg.resolution = c(2,
    2))
#> Generating frame raster at  2 2  resolution and extent defined by:  113 153 -42 -10
```




```
print(head(estimated_richness$alpha_result))
#>    Species      chao    chao.se    jack1 jack1.se    jack2     boot  boot.se n
#> 15     213  551.6988   78.24699 359.5714 66.55978 459.6905 274.7180 29.58080 7
#> 29      50  650.2500  624.75000  74.5000 24.50000  74.5000  62.2500 12.49750 2
#> 30     225  905.6108  178.19196 386.8750 74.41029 511.4821 291.1311 31.12806 8
#> 35     204  715.7175  130.80966 353.6250 68.73625 465.8036 265.4039 30.24931 8
#> 48     136 6256.0000 1053.47953 226.6667 66.26546 272.0000 176.2963 27.94017 3
#> 50     104 2756.2500 2703.75000 155.5000 55.97097 155.5000 129.7500 30.26859 2
```

But let's say we want to calculate the overall beta diversity across plots as grouped in a compositional classification.

1.generate cluster tree 2. cut and get groups 3. redo map multi with beta and map result

Adding phylogenetic information to mapped biodiverity metrics

By adding phylogenetic information, we can increase our repertoire to a wider range of metrics that get at different aspects of biodiversity.

For the purpose of demonstration, we will generate a random phylogenetic tree:

```
ausplots.Rtree <- ape::rtree(n = ncol(my.sppBYsites), tip.label = colnames(my.sppBYsites))
```

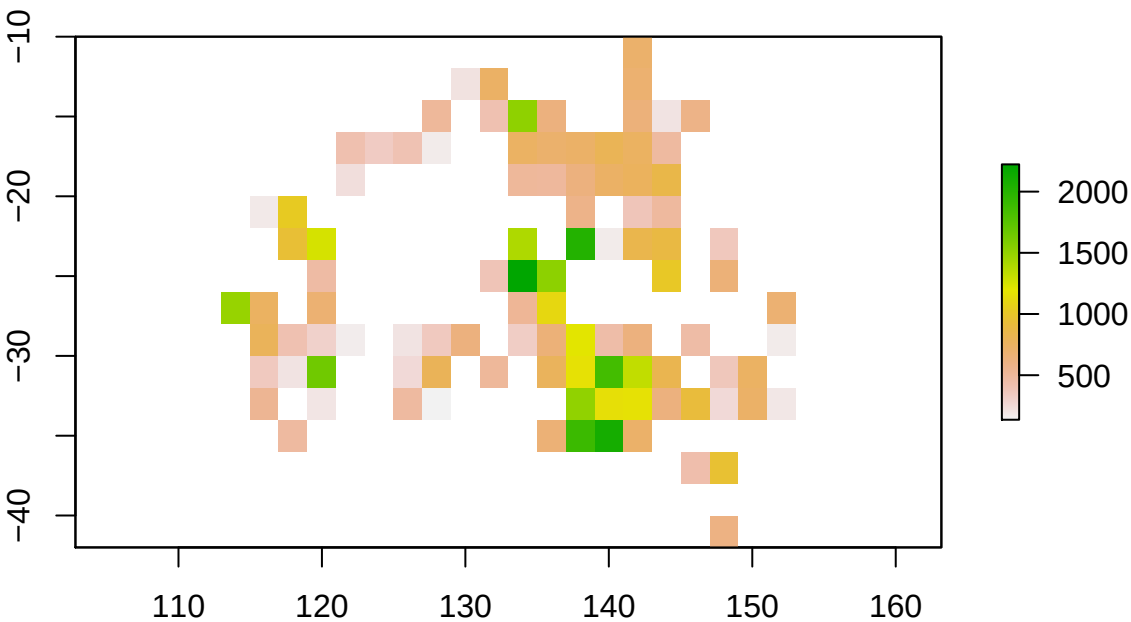
As TERN Ausplots is plot-based, we can generate a community level vector of phylogenetic diversity scores.

```
# Calculate for the first 20 sites:
phylogenetic.diversity.sites(my.sppBYsites[1:20, ], ausplots.Rtree)
#> Checking that occurrence matrix and phylo match.
#> Calculating phylogenetic diversity
#> Collating outputs
#> $PD
#> NSABBS0001-58580 NSABBS0002-58555 NSABBS0003-58581 NSABBS0004-58556 NSABBS0005-58582
#>        155.10291        183.69626        124.97623         93.57363        128.47387
#> NSABBS0006-58557 NSABHC0001-53596 NSABHC0002-53597 NSABHC0003-53598 NSABHC0004-53599
#>        125.21254        294.31947        305.06279        296.80761        411.53594
#> NSABHC0005-53600 NSABHC0006-53601 NSABHC0007-53602 NSABHC0008-53603 NSABHC0009-53604
#>        276.03854        300.45374        417.45769        348.55334        410.52723
#> NSABHC0009-58026 NSABHC0010-53605 NSABHC0011-53606 NSABHC0012-53607 NSABHC0012-58022
```

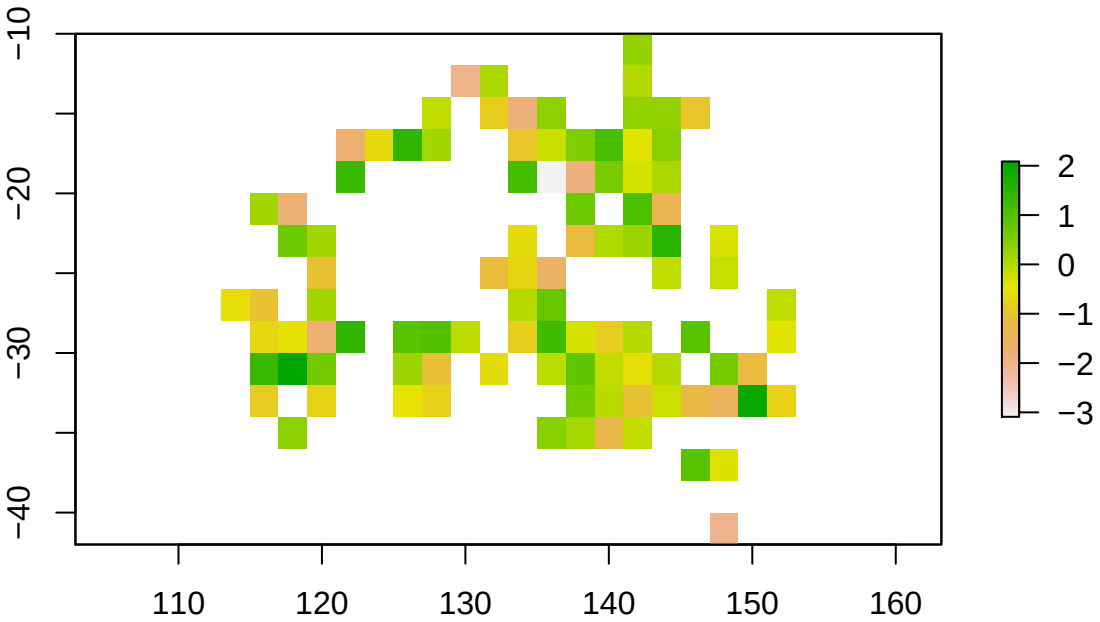
```
#>          347.64739          292.18243          284.20382          378.02675          341.36451
```

We can also calculate phylogenetic diversity aggregated at grid cell level and standardise the result either according to the completeness of species sampling in the grid cells, or according to expectation for a given level of species richness. Finally we can calculate phylogenetic range rarity for grid cells and standardise that by comparing the result to a null distribution.

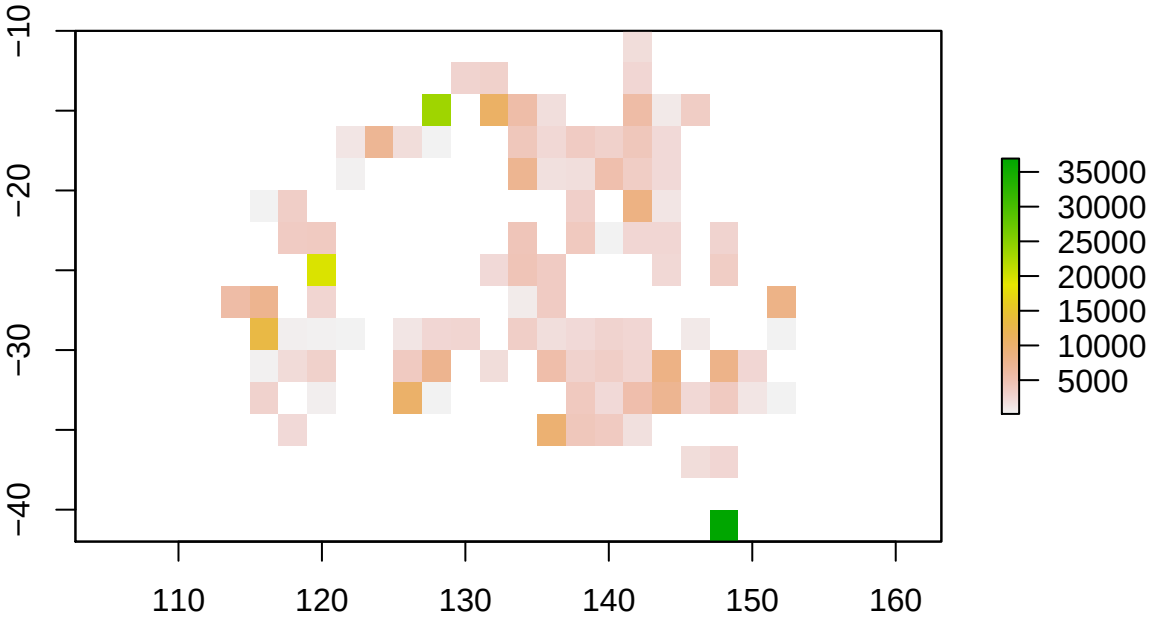
```
# unweighted PD for grid cells:
myPD <- phylogenetic.endemism(mySPPrecords, sep.comm.spp = "none", phylo.tree = ausplots.Rtree,
    sep.phylo.spp = "none", pe.type = "unweighted", plot.raster = FALSE, own.grid.matrix = cell.PA$grid
    frame.raster = cell.PA$pa.raster)
#> Some point locations lie outside the frame raster -- trimming these records
#> Reading user-defined gridded occurrence matrix
#> Occurrence matrix generated with dimensions:  100 5142
#> Checking that occurrence matrix and phylo match.
#> Calculating phylogenetic diversity
#> Collating outputs
plot(myPD$PD_raster)
```




```
# richness-standardised PD for cells:
myPD_rich <- phylogenetic.endemism(mySPPrecords, phylo.tree = ausplots.Rtree, frame.raster = cell.PA$pa
    pe.type = "unweighted", own.grid.matrix = cell.PA$grid.matrix, plot.raster = FALSE,
    pd.standard = TRUE, sep.comm.spp = "none", sep.phylo.spp = "none")
#> Some point locations lie outside the frame raster -- trimming these records
#> Reading user-defined gridded occurrence matrix
#> Occurrence matrix generated with dimensions:  100 5142
#> Checking that occurrence matrix and phylo match.
#> Calculating phylogenetic diversity
#> Standardising phylogenetic diversity by species richness
#> Collating outputs
plot(myPD_rich$PD_raster)
```

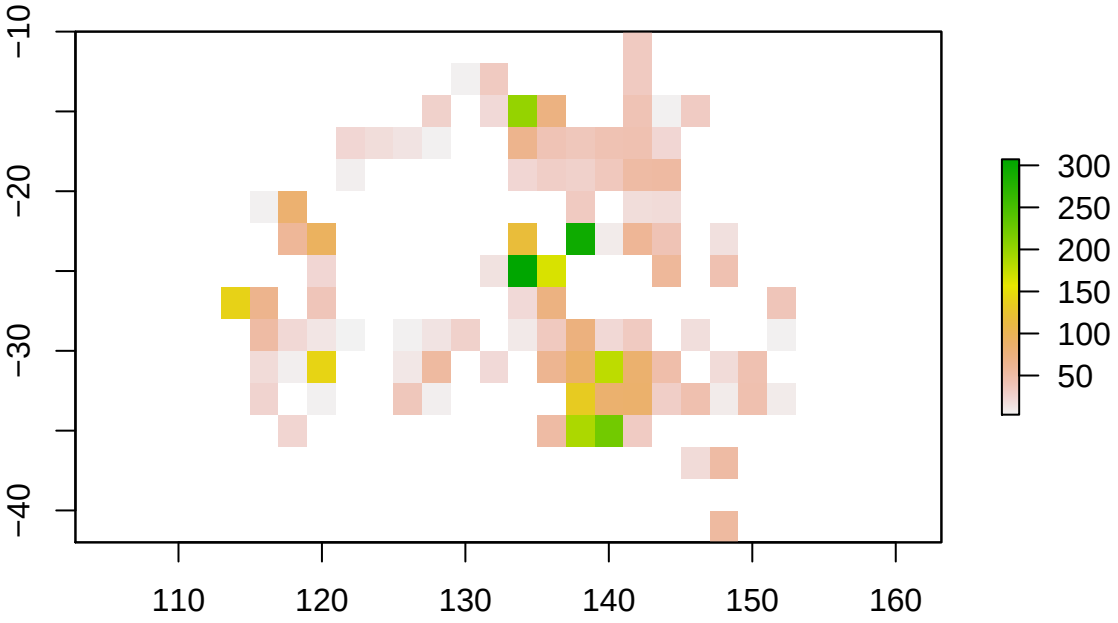



```
##################


# PD in grid cells corrected according the to completeness of species sampling:
PD_complete <- map.multi.metrics(my.sppBYsites, site.coords = my.ausplots$site.info[my.ausplots$site.in
    rownames(my.sppBYsites), c("longitude", "latitude")], beta = FALSE, phylogenetic = TRUE,
    phylo.tree = ausplots.Rtree, frame.raster = cell.PA$pa.raster, plot.raster = FALSE)
#> Checking that occurrence matrix and phylo match.
#> Calculating phylogenetic diversity

plot(PD_complete$phylogenetic_raster)
```

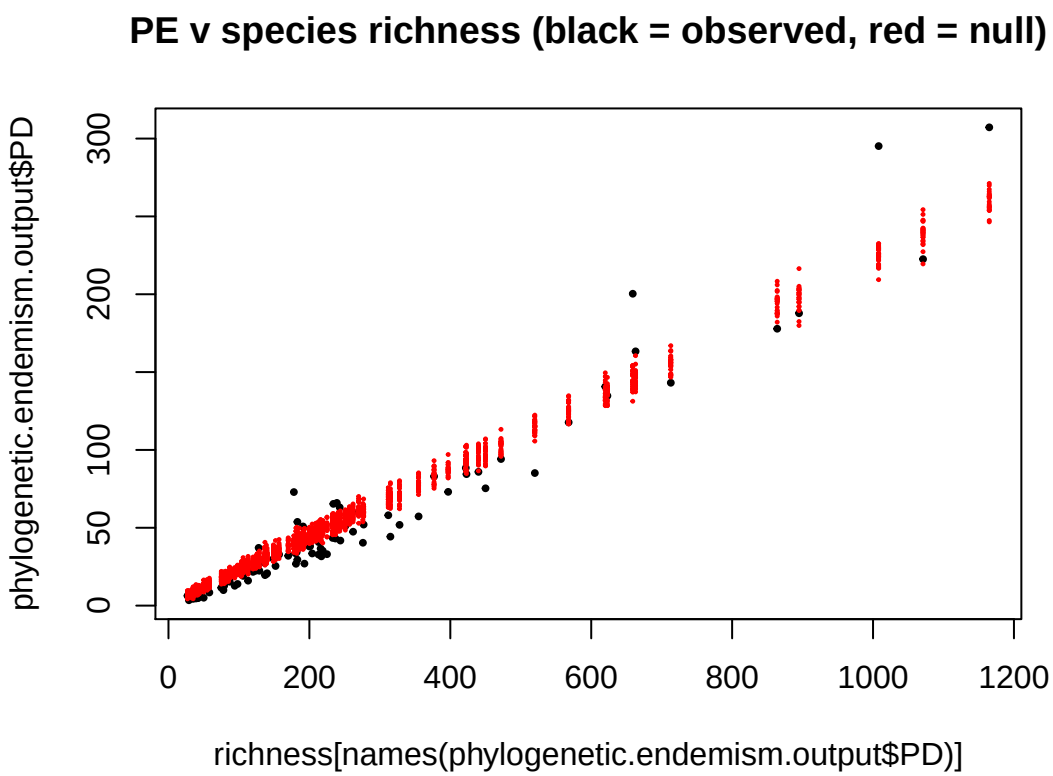



```
##### phylogenetic range rarity compared to a null distribution
```

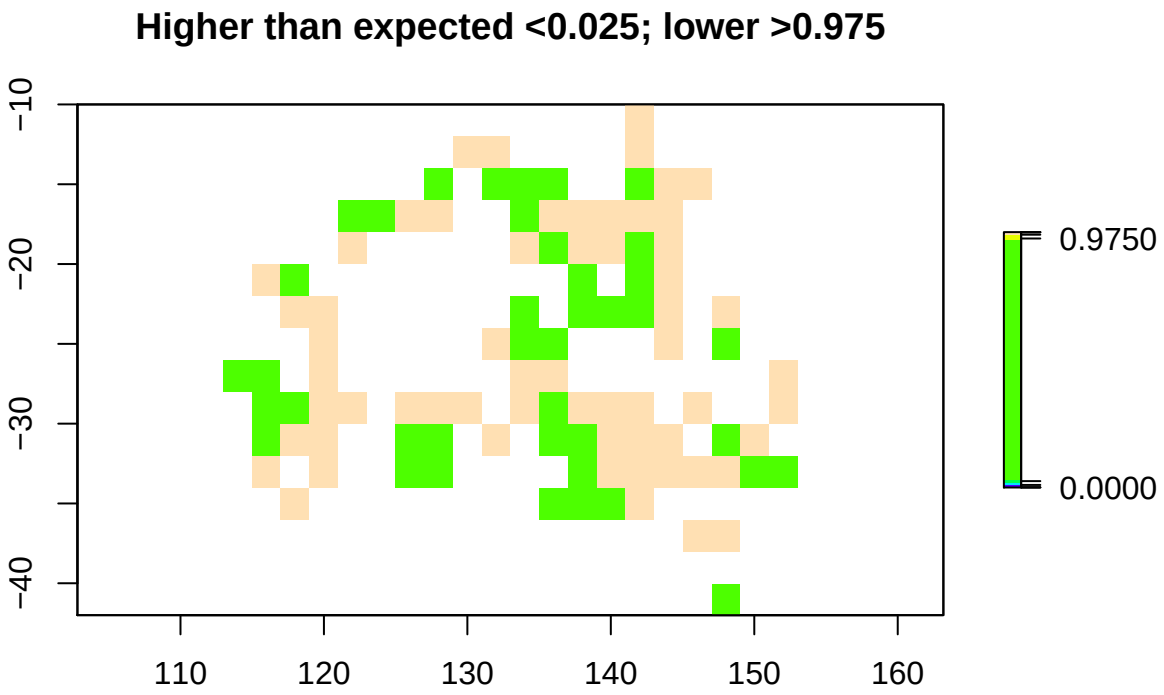
```r
myPE <- phylogenetic.endemism(mySPPrecords, phylo.tree = ausplots.Rtree, frame.raster = cell.PA$pa.raste
    own.grid.matrix = cell.PA$grid.matrix, plot.raster = FALSE, sep.comm.spp = "none",
    sep.phylo.spp = "none")
#> Some point locations lie outside the frame raster -- trimming these records
#> Reading user-defined gridded occurrence matrix
#> Occurrence matrix generated with dimensions:  100 5142
#> Checking that occurrence matrix and phylo match.
#> Converting phylo.tree to matrix
#> Re-ordering the matrices to match...
#> Phylo matrix generated with dimensions,  5142 10282 and range  0 1
#> Generating a community phylogenetic branch matrix.
#> Calculating range weights
#> Calculating phylogenetic endemism.

plot(myPE$PD_raster)
```




```r
# generate null distribution and detect outliers
myPE.test <- pe.null.test(myPE, phylo.tree = ausplots.Rtree, nrep = 25, plot.all = FALSE)
#> Calculating species random sample probabilities and species richness
```




```
#> Generating null distributions for range of observed species richness:
#> Calculating outliers
#> Mapping outlier results
#> Calculating significance
#> Mapping significance results
#> Collating outputs

p.breaks <- c(0, 1e-04, 0.001, 0.01, 0.025, 0.975, 0.99, 0.999, 0.9999, 1)
plot(myPE.test$P.above.raster, breaks = p.breaks, col = topo.colors(9), main = "Higher than expected <0
```



Conclusion

With these examples, we have seen how **biomapME** can be used to easily generated gridded biodiversity maps from simple species incidence data, using metrics such as species richness and endemism, multi-site beta, phylogenetic diversity and endemism, with several choices of approach for correcting estimates for undersampling and the correlation of most additive indices with simple species richness.

References

Crisp, M.D., Laffan, S., Linder, H.P. & Monro, A. (2001) Endemism in the Australian flora. *Journal of Biogeography* **28**, 183-198.

Guerin, G.R., Ruokolainen, L., Lowe, A.J. (2015) A georeferenced implementation of weighted endemism. *Methods in Ecology and Evolution* **6**, 845–852.

Guerin, G.R. & Lowe, A.J. (2015) Mapping phylogenetic endemism in R using georeferenced branch extents. *SoftwareX* **3-4**, 22–26.

Guerin, G.R. & Lowe, A.J. (2015) 'Sum of inverse range sizes' (SIR), a biodiversity metric with many names and interpretations. *Biodiversity and Conservation* **24**, 2877–2882.

Rosauer, D., Laffan, S.W., Crisp, M.D., Donnellan, S.C. & Cook, L.G. (2009) Phylogenetic endemism: a new approach for identifying geographical concentrations of evolutionary history. *Molecular Ecology* **18**, 4061-4072.

Sparrow, B., Foulkes, J., Wardle, G., Leitch, E., Caddy-Retalic, S., van Leeuwen, S., Tokmakoff, A., Thurgate, N., Guerin, G.R. and Lowe, A.J. (2020) A vegetation and soil survey method for surveillance monitoring of rangeland environments. *Frontiers in Ecology and Evolution*, 8:157.

Tsirogiannis, C. & B. Sandel. (2016) PhyloMeasures: A package for computing phylogenetic biodiversity measures and their statistical moments. *Ecography* **39**, 709-714,