Understanding spatial patterns of biodiversity: How sensitive is phylogenetic endemism to the randomisation model?

S.W. Laffan [1], A.H. Thornhill[2], J.T. Miller[3], N. Knerr[4], C.E. Gonzales-Orozco[5], B.D. Mishler[2]

[1]Centre for Ecosystem Science, School of Biological, Earth and Environmental Science, UNSW, Sydney, Australia, 2052
Email: shawn.laffan@unsw.edu.au
[2]University and Jepson Herbaria, and Dept. of Integrative Biology, University of California, Berkeley, CA 94720-2465, USA.
Email: {andrew.thornhill, bmishler}@berkeley.edu
[3]Division of Environmental Biology, National Science Foundation, Arlington, Virginia, USA
Email: joe@acaciamulga.net
[4]National Research Collections Australia, CSIRO National Facilities and Collections, GPO Box 1600, Canberra ACT 2601, Australia
Email: Nunzio.Knerr@csiro.au
[5]Corporación Colombiana de Investigación Agropecuaria, Corpoica, km 17 Vía Puerto López, Meta, Colombia
Email: cegonzalez@corpoica.org.co

Abstract

Mapping spatial patterns of phylogenetic diversity helps identify regions of unique evolutionary history warranting conservation. Randomisations form an integral component of this process. Here we test the sensitivity of a method used to identify unusual concentrations of old and new evolutionary history to the underlying randomisation. The results indicate low sensitivity to models of complete spatial randomness and spatial structure (proximal allocation and random walks).

1. Introduction

Knowledge of the spatial distribution of biodiversity is essential for the allocation of scarce conservation resources and for understanding the evolutionary histories of a region's biota. Biodiversity is many-faceted, and can be measured using components such as species, phylogenetic, and trait diversity (Laffan 2014).

Biodiversity indices are typically aggregate measures of the taxon assemblage found in a location, with geographic surfaces of indices commonly generated. The most commonly used index is species richness, calculated as the number of unique species in a sample. However, closely related species represent less unique diversity than do distantly related species (see Fig 1), e.g., a sample comprising a human, a gorilla, and an orangutan has less unique diversity than one comprising a snake, a cow, and a squid. If one is interested in the conservation and analysis of biodiversity at an evolutionary level then one needs to use phylodiversity indices (Laity et al. 2015).

Phylogenetic Diversity (PD; Faith 1992) is the simplest phylodiversity measure and is calculated as the sum of a tree's branch lengths in a sample (Figure 1). Phylogenetic Endemism (PE; Rosauer et al. 2009) is calculated in the same way as PD, but the branches are weighted by the fraction of their geographic ranges found in a location, such that wide-ranged branches contribute less than narrow-ranged branches of the same length. PE is used to identify regions containing lineages that are found in few other places.

A more recent development of PE is the CANAPE method (Categorical Analysis of Neo- and Palaeo-Endemism; Mishler et al. 2014). CANAPE uses PE with a randomisation test to identify regions of geographically restricted long or short branches. Regions of palaeo-endemism can be considered as museums of evolutionary history currently found in few other places, while regions of neo-endemism can be considered as cradles of new diversity. CANAPE classifies the remaining cells into three other classes, two that contain some

mixture of palaeo- and neo-endemism at differing levels (mixed and super), and those that are not significant given the randomisation test.

The randomisation test is integral to CANAPE as it is used to allocate cells to the different classes. The randomisation test used in studies to date follows a model of complete spatial randomness (CSR). In each random realisation, each species is allocated to cells randomly across the data set, with the dual constraints that each species is found in exactly the same number of cells as in the observed data set, and that each cell has exactly the same number of species as it does in the observed data set (thus range size and richness are held constant). A swapping process is used to ensure all species occurrences are allocated to satisfy the range constraint.

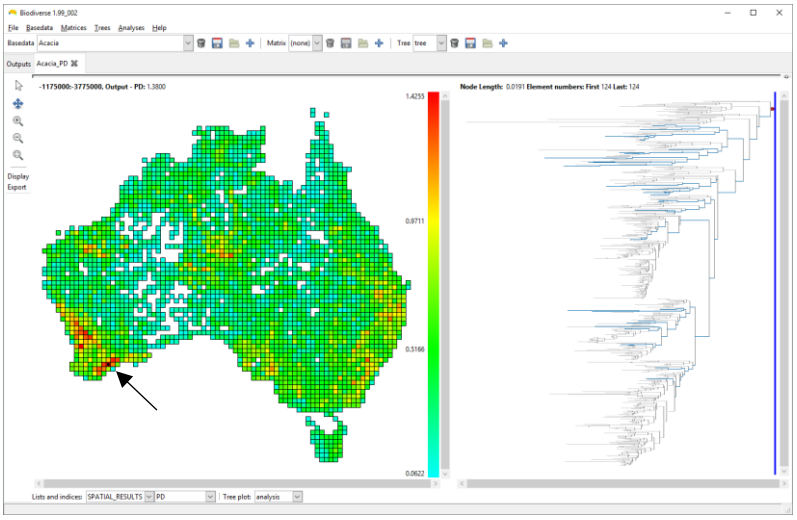

Figure 1. Phylogenetic diversity (PD) is measured as the sum of branch lengths on a phylogenetic tree that are found in a location. The example shows PD using a phylogenetic tree containing 506 *Acacia* species across Australia. Branches highlighted in blue occur in a cell in south-west Western Australia (arrow) and the sum of their lengths is the PD score for that cell.

The limitations of the CSR model are well documented (see for example O'Sullivan and Unwin 2010), with spatially structured randomisations potentially offering a more rigorous test of the results (O'Sullivan and Perry 2013). Such effects have previously been explored for species level endemism indices (Laffan and Crisp 2003), but not for more complex phylogenetic indices such as CANAPE. An understanding of the effect of more spatially structured randomisations on the CANAPE analysis is therefore needed.

2. Spatial randomisations

Two additional models have been implemented in the Biodiverse software (Laffan et al. 2010), proximal allocation (PA) and a random walk (RW) (Figure 2). Both models use the swapping approach from the CSR model to ensure all occurrences are allocated.

In the PR model a species and seed cell are selected, and the species occurrence is allocated to that cell. Subsequent occurrences of that species are then allocated in order of increasing distance from the seed cell, with random selection in the event of ties. Cells are skipped if they have already reached their richness target. A spatial window centred on the seed cell controls how far species will be allocated from the seed cell. Once all allocatable cells in the window have been used, a new seed location is chosen and the process continues until all occurrences are allocated or there remain no cells to assign to. This approach is similar to the circular model used in Laffan and Crisp (2003), but with greater flexibility as

Biodiverse supports arbitrarily complex spatial windows. (The allocation order can also be random instead of proximal, but that is not used here).

The RW model is a long established approach (O'Sullivan and Perry 2013) and is simply a variation on the PA approach that uses a different allocation method. From the seed cell, the method selects a neighbouring cell to which it allocates the next species occurrence. It then allocates to one of that cell's neighbours, and the process repeats until all occurrences have been allocated or no more cells can be assigned to. If a cell has no assignable neighbours then the system backtracks to the most recently allocated cell with such neighbours, or it restarts at a new seed location. The RW model has the advantage that the random distributions will remain within a region if there are gaps that the spatial window does not span, for example islands.

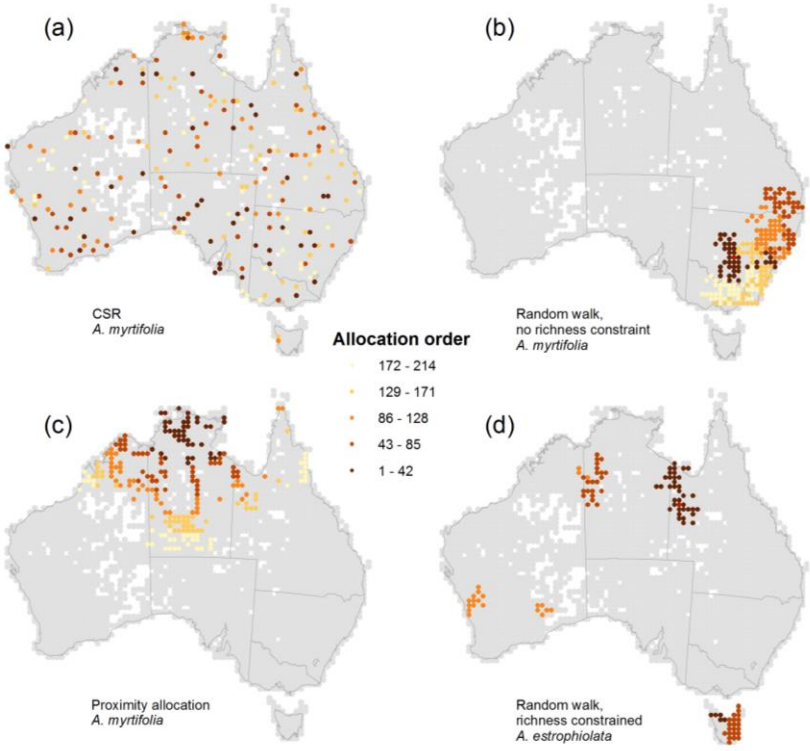

Figure 2. Example randomised distributions. (a) CSR, (b) RW without richness constraints, (c) PA, and (d) RW with richness constraint. Colours show allocation order.

3. Analyses and discussion

The CANAPE index was calculated for a data set of 506 *Acacia* species aggregated to 3037 cells at a 50 km resolution (Mishler et al. 2014). Four randomisations were used: CSR, RW with 100 km radius windows, PA with 100 km radius windows, and PA with no window constraint.

The results indicate little difference in the overall patterns (Figure 3). There is a small increase in the neo-endemic locations using the spatial randomisations, and the unconstrained PA has more non-significant cells, but the general patterns remain the same.

It is likely that the richness constraints have a large influence on initial allocations, with the swapping process further disrupting the spatial structure of the initial distributions. This is the likely reason the internal branch ranges remain larger than the observed data, and are not substantially different among the models (data not shown). This is also broadly consistent with the results of Laffan and Crisp (2003) for a single cell analysis.

Further testing will assess the effect of spatial scale on the results, considering the degree to which branches in a cell are restricted to regions surrounding them. More complex RW

models could include random back-tracking instead of sequential to generate shorter paths, and constraining the overall size and shape of a distribution to create more compact walks.

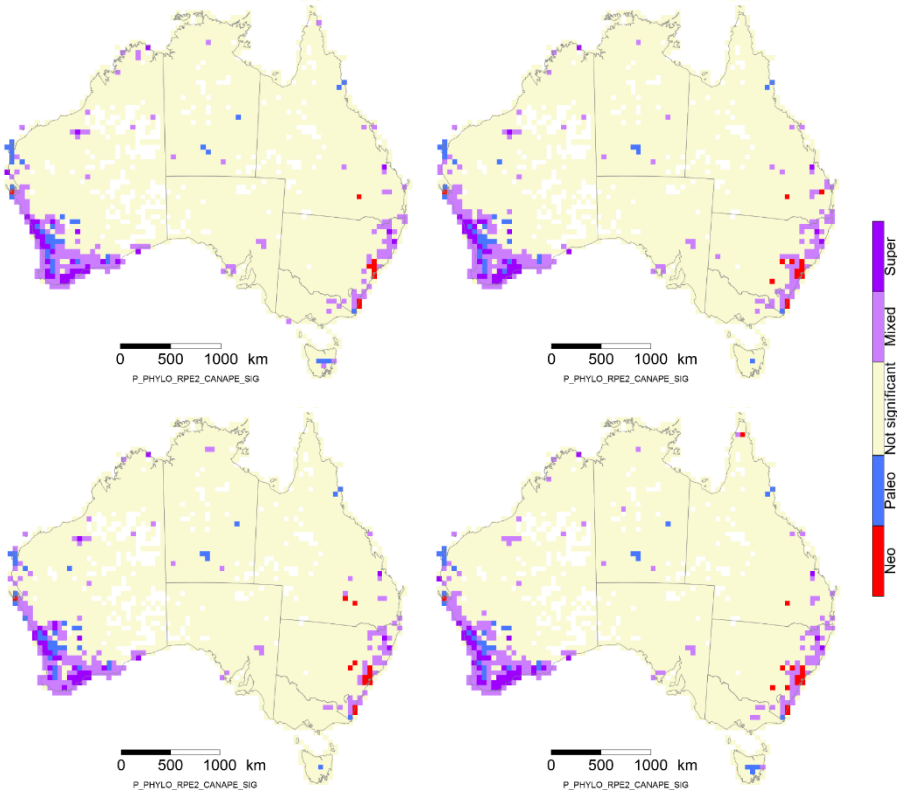

Figure 3. CSR model (upper left), random walk model (upper right), proximal allocation (100km radius; lower left), proximal allocation to any cells (lower right).

Acknowledgements

This manuscript includes work done by JTM while serving at the National Science Foundation. The views expressed in this article do not necessarily reflect those of the National Science Foundation or the United States Government. NSF grant DEB- 1354552 provided partial support for the UC Berkeley portion of this project.

References

Faith, D.P. 1992. Conservation evaluation and phylogenetic diversity. *Biological Conservation,* 61, 1-10.

Laffan, S.W., 2014. GeoComputation: Applications in Biology. *In:* Abrahart, R. J. and See, L. eds. *GeoComputation, Second Edition.* 2 ed. London: CRC Press, 125-142.

Laffan, S.W. and Crisp, M.D. 2003. Assessing endemism at multiple spatial scales, with an example from the Australian vascular flora. *Journal of Biogeography,* 30, 511-520.

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.

Laity, T.*, et al.* 2015. Phylodiversity to inform conservation policy: An Australian example. *Science of the Total Environment,* 534, 131-143.

Mishler, B.D.*, et al.* 2014. Phylogenetic measures of biodiversity and neo- and paleo-endemism in Australian Acacia. *Nature Communications,* 5, 4473.

O'Sullivan, D. and Perry, G.L.W., 2013. *Spatial Simulation: Exploring Pattern and Process.* Chichester, UK: John Wiley & Sons, Ltd.

O'Sullivan, D. and Unwin, D.J., 2010. Geographic Information Analysis, 2nd edition. *Geographic Information Analysis.* 2nd ed. Hoboken, USA: John Wiley & Sons, Inc., 406.

Rosauer, D.F.*, et al.* 2009. Phylogenetic endemism: a new approach to identifying geographical concentrations of evolutionary history. *Molecular Ecology,* 18, 4061-4072.