**Understanding the Processes Underpinning Patterns of Phylogenetic Regionalization**

**(Supplementary Material)**

Barnabas H. Daru1,2,\*, Tammy L. Elliott3, Daniel S. Park1, T. Jonathan Davies4

**Affiliations:**

1Department of Organismic and Evolutionary Biology, Harvard University, Cambridge, MA 02138, USA

2Department of Plant Sciences, University of Pretoria, Private Bag X20, Hatfield 0028, Pretoria, South Africa

3Department of Biological Sciences, University of Cape Town, Private Bag X3, Rondebosch, 7701, Cape Town, South Africa

4Department of Biology, McGill University, Montreal, QC, H3A 0G4 Canada

\*Correspondence: barnabas\_daru@fas.harvard.edu (B.H. Daru)

**SUPPLEMENTARY MATERIAL**

**Appendix S1 for Box 2:** Empirical example of phylogenetic beta diversity of Madagascan mammals.

*Mammal Distribution Data*

We used extent-of-occurrence maps from the IUCN (http://www.iucnredlist.org/technical-documents/spatial-data) for all 205 mammal species native to Madagascar. We overlaid the maps onto 3,263 equally sized grids of 0.125 × 0.125° (Berhmann equal-area projection) to generate a presence/absence (1/0) matrix with names of species as columns and rows as grid cells. We excluded coastal grid cells with < 50% land area to minimize the influence of unequal sampling area.

***Phylogenetic data***

The phylogenies of the mammals used were adapted from 600 phylogenies for all extant and late Quaternary extinct mammal species of the world (Faurby & Svenning 2015).

***Clustering phyloregions***

We used Simpson's beta dissimilarity (βsim) metric and generated pairwise distance matrices of phylogenetic β-diversity (pβsim) and also species-level β-diversity (for comparison) between all pairs of grid cells using R (R Core Team 2015). Cluster analyses of βsim and pβsim, selection of the best algorithm and determination of the optimal number of phyloregions were quantified as described in the main text (Box 1).

For each phyloregion, we computed five facets of diversity: species richness (SR), phylogenetic endemism (PE), species endemism (SE), evolutionary distinctiveness (ED), and 'evolutionary distinctiveness and global endangerment' (EDGE) (Box 2). SR is a count of species in an area; PE quantifies the degree to which phylogenetic diversity is restricted to a landscape (Rosauer *et al.* 2009), EDGE combines evolutionary distinctiveness (ED; *i.e*. phylogenetic isolation of a species) with IUCN global endangerment (GE) to identify species that are phylogenetically isolated and threatened with extinction (Isaac *et al.* 2007).

All phylogenetic analyses including phylogenetic beta diversity, phylogenetic endemism, and 'evolutionary distinctiveness and global endangerment' were estimated using 100 trees randomly selected from a posterior distribution of 600 trees from Faurby & Svenning (2015). The spatial variations in these phylogenetic metrics for the 100 subsets were combined by taking the mean values across all grid cell comparisons.

Last, we identified the phylogenetic positions of the nodes that are significantly contributing to phylogenetic turnover in each phyloregion using the *nodesigl* function in PHYLOCOM (Webb *et al.* 2008).

**RESULTS**

The spatial distribution of regional species richness for Madagascan mammals shows a longitudinal gradient, with higher richness in the eastern Evergreen rainforest and the western evergreen forest than in the interior part of the island (Figure S1.1). The UPGMA algorithm outperformed the other algorithms tested for regionalization (cophenetic coefficients = 0.60 and 0.51, and Gower distance = 129787.7 and 39186.02 for βsim and pβsim, respectively) (Table S1.1). Using an UPGMA approach, and applying the elbow method to select the optimal number of clusters, we found nine clusters, *i.e.*, ‘phyloregions’, for pβsim (Figure IA in Box 2 and Figure S1.2) and 11 regions for βsim (Figure IB in Box 2 and Figure S1.3), but there was a strong positive correlation between matrices of pβsim and βsim (r = 0.85, p = 0.001, Mantel test pβsim and βsim and 999 permutations).

Our phyloregions broadly match the existing vegetation map of Madagascar (*e.g.* Yoder & Nowak 2006), including Sambirano Forest, Grassland/woodland/bushland mosaic, and Dry deciduous forest. However, we also identified three distinct phyloregions (X, Y, and Z) that were previously overlooked by previous regionalization schemes. Phyloregion X occurs at the southern tip of Madagascar and encompasses parts of the Arid Spiny bush (Figure I in Box 2), phyloregion Y occurs along the southwestern coast of Mahajanga and phyloregion Z occurs in the northeast directly north of Helodrano Antongila Bay (Figure I in Box 2). The orders of differentiation among the phyloregions is also reflected in the NMDS ordination such that closely related phyloregions occur close together whereas distinct phyloregions fall separately away (Figure I in Box 2).

We indexed these phyloregions using common phylogenetic metrics (Figure II in Box 2). The previously undescribed phyloregion X in the south emerged as the most evolutionarily distinct (mean pβsim = 0.224), whereas the Evergreen rainforest I in the east captured both high species and unique phylogenetic diversity (phylogenetic endemism; mean PE = 470.54). The Dry deciduous Forest II phyloregion in the northern tip of the island captured the highest number of endemic species. Last, the second new phyloregion Y in the west encapsulated a high proportion of evolutionarily distinct and globally endangered species.

To identify clades contributing to the various patterns of phylogenetic turnover, we identified the nodes that are significantly overrepresented in each phyloregion (Borregaard *et al.* 2014). The most evolutionary distinct phyloregion X is dominated by Tenrecs (Tenrecidae) and Bats than primates *e.g.* Lemurs. The hot nodes contributing to the high phylogenetic endemism in the Evergreen rainforest I were mainly found within lineages of Mouse lemurs (Cheirogaleidae) and Tenrecs (Tenrecidae).

**DISCUSSION**

Our Madagascan mammal richness map matches the classical vertebrate diversity gradient of the island, which potentially mirrors the differences in relative diversification and extinction rates across the island: over-dispersion coupled with low extinction rates in the eastern wet regions *vs.* dry western habitats mediated by historical climate change (Razafindratsima et al. 2013).

Several mechanistic processes can underlie the patterns of phylogenetic turnover of the contemporary distribution of mammals in Madagascar. First, historical extinction during the Holocene to as recent as 500 years ago might have led to a large scale extinction of Madagascar's fauna (Turvey and Fritz 2011). This could underpin the pattern of phylogenetic turnover we found in this study where mammal assemblages formed distinct phylogenetic clusters.

**References**

Borregaard, M.K. *et al.* (2014) Node-based analysis of species distributions. *Methods Ecol. Evol.* 5, 1225–1235

Faurby, S. and Svenning, J.C. (2015) A species-level phylogeny of all extant and late Quaternary extinct mammals using a novel heuristic-hierarchical Bayesian approach. *Mol. Phylogenet. Evol.* 84, 14–26

Isaac, N.J. *et al.* (2007) Mammals on the EDGE: conservation priorities based on threat and phylogeny*. PLoS ONE*, 2, e296

R Core Team (2015) *R: A language and environment for statistical computing.* R Foundation for Statistical Computing, Vienna, Austria. URL http://www.R-project.org/.

Razafindratsima, O.H. *et al.* (2013) Extinctions, traits and phylogenetic community structure: insights from primate assemblages in Madagascar. *Ecography* 36, 47–56

Rosauer, D. *et al.* (2009) Phylogenetic endemism: a new approach for identifying geographical concentrations of evolutionary history. *Mol. Ecol.* 18, 4061–4072

Turvey, S. T. and Fritz, S.A. (2011) The ghosts of mammals past: biological and geographical patterns of global mammalian extinction across the Holocene. *Phil. Trans. R. Soc. B* 366, 2564–2576

Webb, C.O. *et al.* (2008) Phylocom: software for the analysis of phylogenetic community structure and trait evolution. *Bioinformatics* 24, 2098–2100

Yoder, A.D. and Nowak, M. (2006) Has vicariance or dispersal been the predominant biogeographic force in Madagascar? *Annu. Rev. Ecol. Evol. Syst*. 37, 405-431

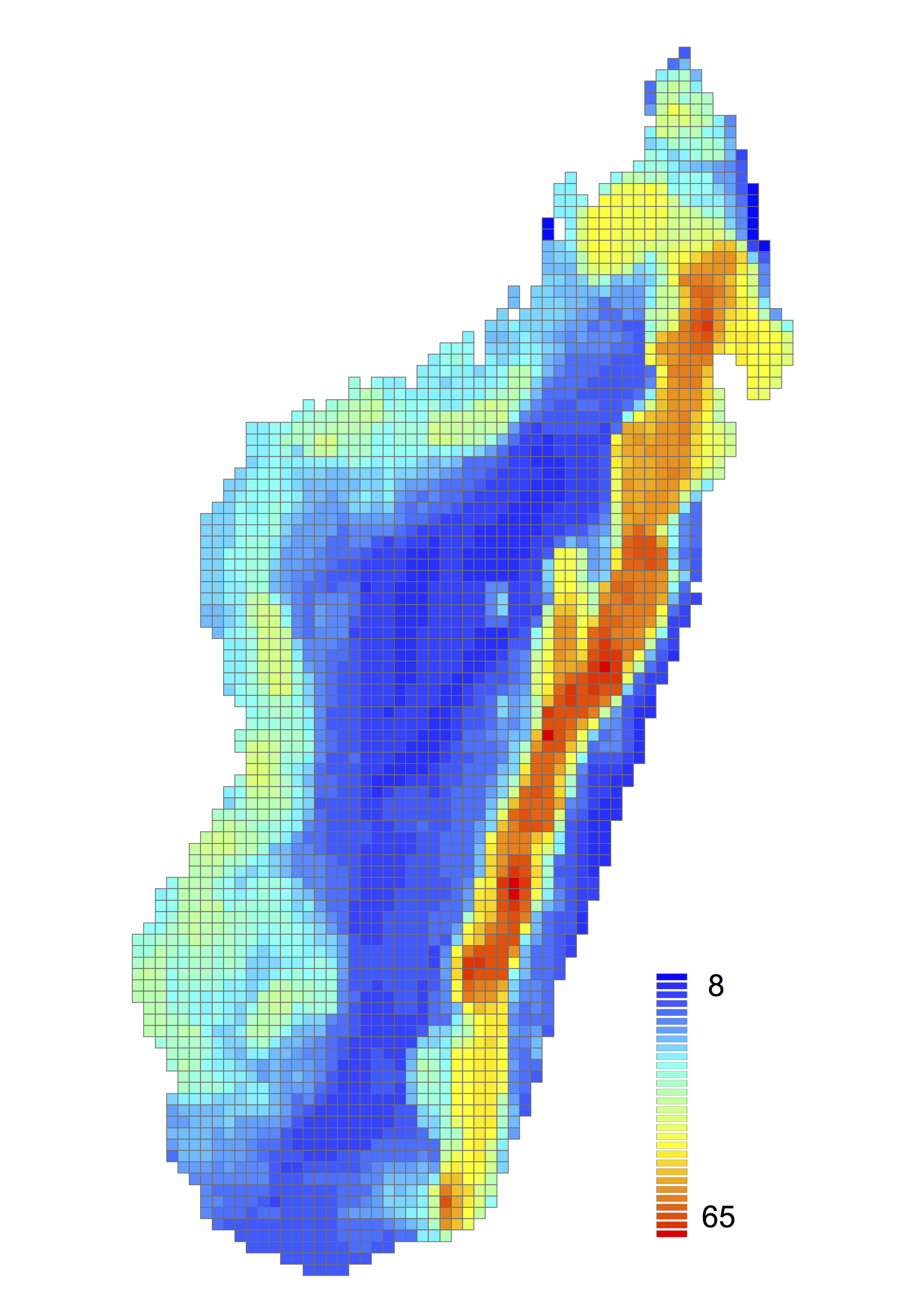


Figure S1.1. Species richness map of the native distribution of Malagasy mammals within 0.125 × 0.125 degree grids.

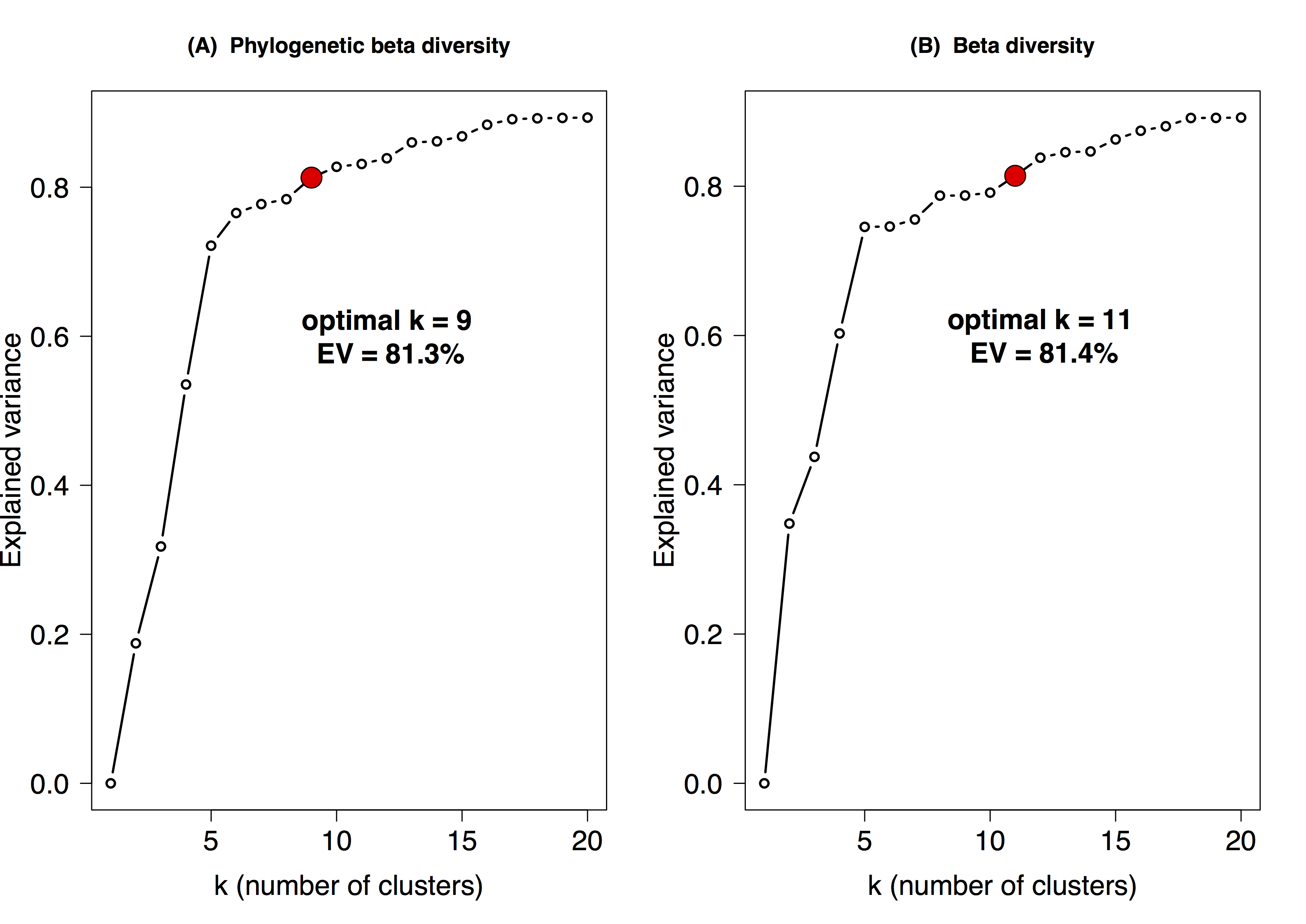


Figure S1.2. Identifying optimal number of clusters (k) using the 'elbow' method for k = 2-20 groups for phylogenetic beta diversity (A), and beta diversity (B). The optimal cluster of the graph is indicated by the red circle.

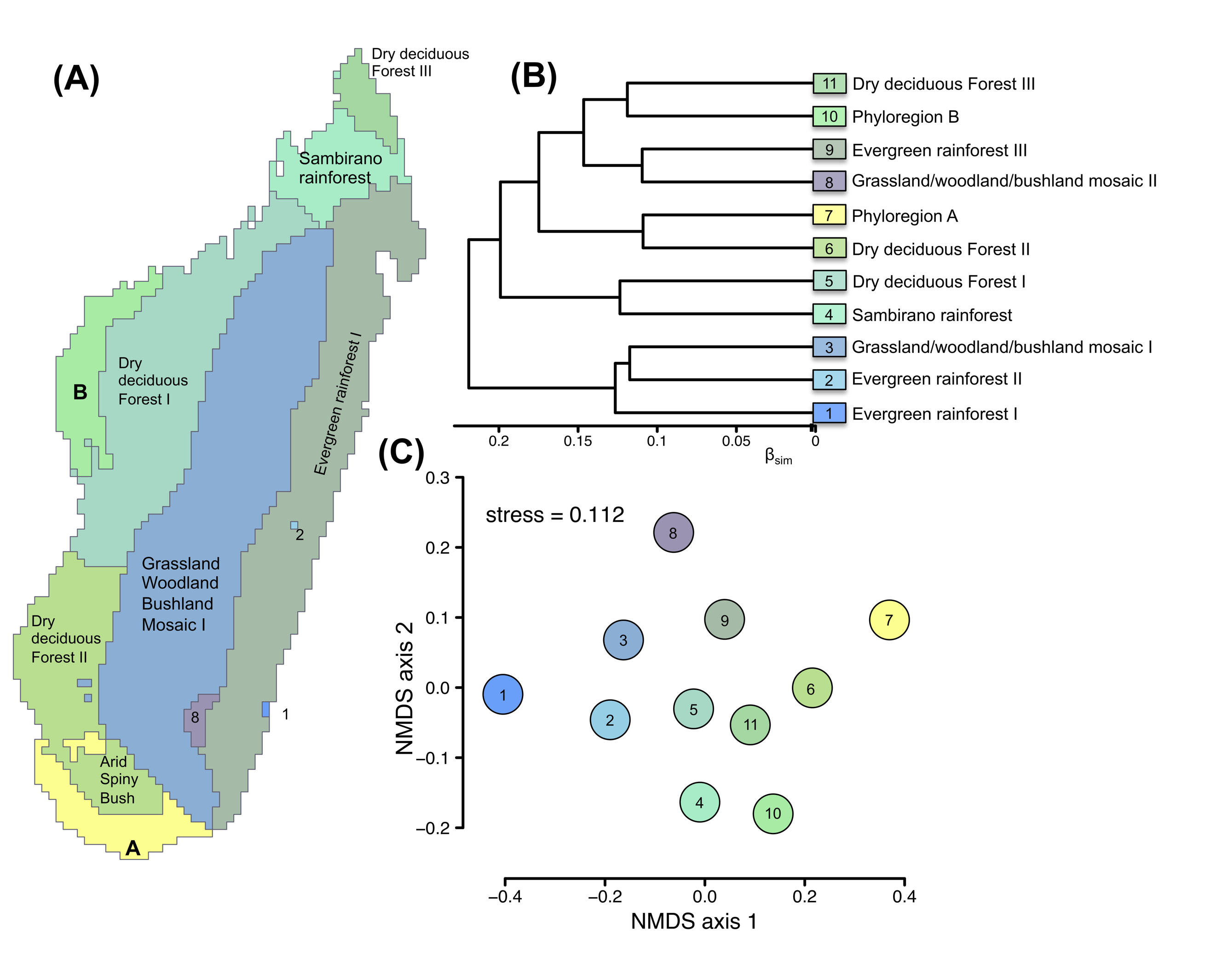


Figure S1.3 Relationships among biogeographic regions for the native mammals of Madagascar based on species-level beta diversity (βsim) values, represented (A) in geographic space, (B) as a dendrogram of compositional dissimilarity, and (C) as an NMDS ordination plot. Grid cells cluster into 11 regions.

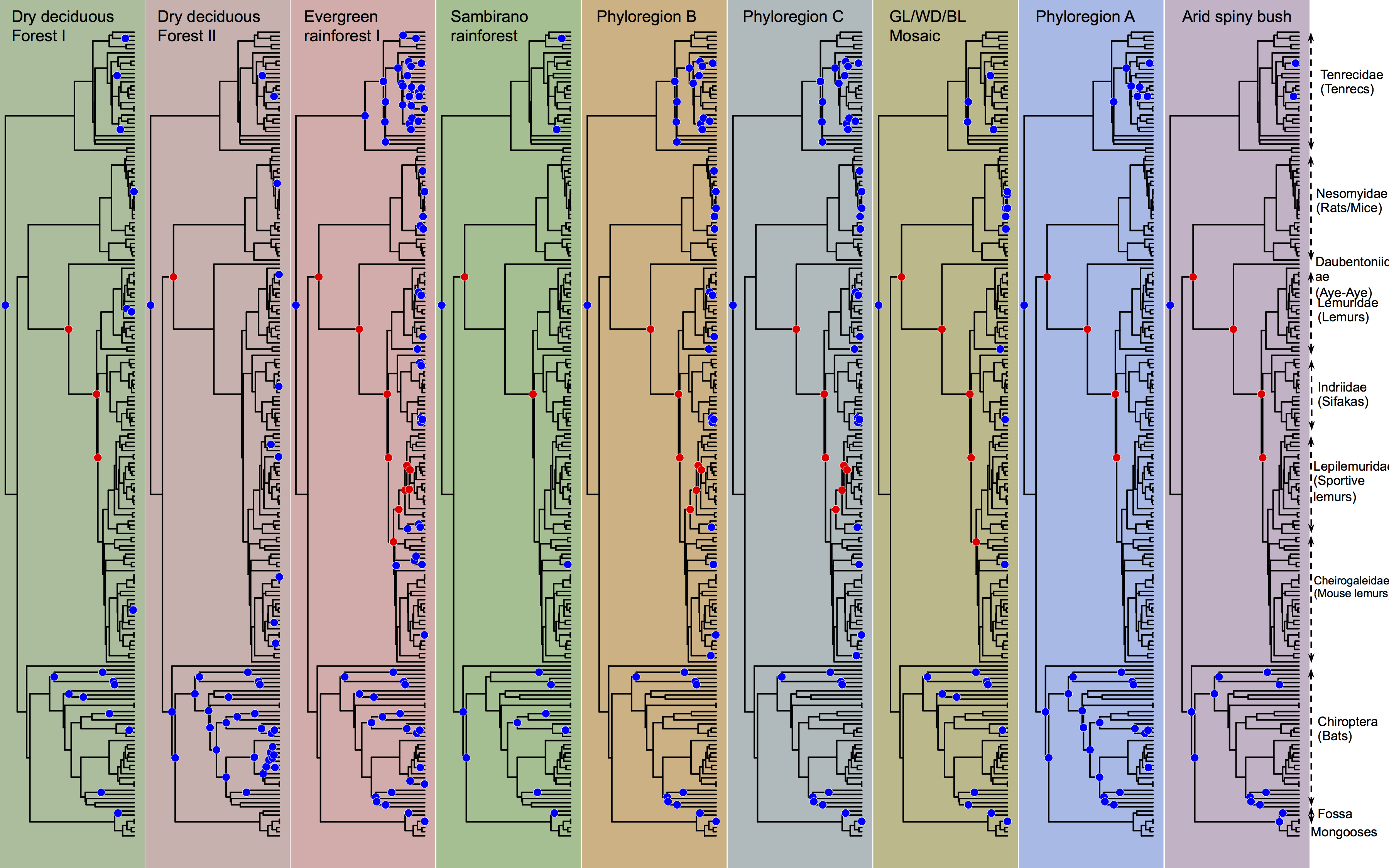


Figure S1.4. Clade representation within phyloregions based on a randomly selected subset of 100 trees from a posterior distribution of Bayesian 600 trees (from Borregaard *et al.* 2014). Blue dots are nodes that are significantly overrepresented in a phyloregion and red indicates significantly underrepresented nodes.

Table S1 Performance of clustering algorithms for phylogenetic beta diversity (pβsim) and beta diversity (βsim) for the flora of North America.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Cluster Method | **Phylogenetic beta diversity** | | **Beta diversity** | |
| Algorithm | Cophenetic correlation coefficient | Gower distance | Cophenetic correlation coefficient | Gower distance |
| Single linkage agglomerative clustering  method = "single" | 0.052 | 152616.8 | 0.0091 | 776589 |
| Complete linkage agglomerative clustering  method = "complete" | 0.47 | 847453.1 | 0.53 | 1356862 |
| UPGMA agglomerative clustering  method = "average" | 0.51 | 39186.02 | 0.60 | 129787.7 |
| WPGMA agglomerative clustering  method = "mcquitty" | 0.46 | 43785.94 | 0.53 | 159390.5 |
| UPGMC agglomerative clustering  method = "centroid" | 0.51 | 62156.49 | 0.57 | 280265.7 |
| WPGMC agglomerative clustering  method = "median" | 0.49 | 56008.29 | 0.51 | 243887.9 |
| Ward's minimum variance clustering  method = "ward" | 0.45 | 273239109 | 0.56 | 733355234 |
| DIANA | 0.48 | 913515.5 | 0.54 | 1746825 |

The cophenetic correlation coefficient measures the fit of each algorithm and ranges between 0 and 1. Cluster algorithms are tested by grouping county assemblages for the flora of North America.