Opinion

Understanding the Processes Underpinning Patterns of Phylogenetic Regionalization

Barnabas H. Daru,[1,2,*] Tammy L. Elliott,[3,4] Daniel S. Park,[1] and T. Jonathan Davies[5,6]

A key step in understanding the distribution of biodiversity is the grouping of regions based on their shared elements. Historically, regionalization schemes have been largely species centric. Recently, there has been interest in incorporating phylogenetic information into regionalization schemes. Phylogenetic regionalization can provide novel insights into the mechanisms that generate, distribute, and maintain biodiversity. We argue that four processes (dispersal limitation, extinction, speciation, and niche conservatism) underlie the formation of species assemblages into phylogenetically distinct biogeographic units. We outline how it can be possible to distinguish among these processes, and identify centers of evolutionary radiation, museums of diversity, and extinction hotspots. We suggest that phylogenetic regionalization provides a rigorous and objective classification of regional diversity and enhances our knowledge of biodiversity patterns.

Biogeographical Regionalization in a Phylogenetic Era

Biogeographical boundaries delineate the basic macrounits of diversity in biogeography, conservation, and macroecology. Their location and composition of species on either side of these boundaries can reflect the historical processes that have shaped the present-day distribution of biodiversity [1]. During the early 19th century, de Candolle [2] created one of the first global geographic regionalization schemes for plant diversity based on both ecological and historical information. This was followed by Sclater [3], who defined six zoological regions based on the global distribution of birds. In 1876, Wallace updated Sclater's zoological regions of the world [3] by incorporating taxonomic relationships among mammals, reptiles, and insects into his method. The Wallace Line, which delineates the faunas of Australia and southeast Asia, represents perhaps one of the most distinct biogeographic divisions observed [4], where species west of the line have a predominately Asiatic origin and those to the east show a mixture of Asian and Australian origins. Subsequently, ecologists and biogeographers have considered additional landscape features, such as structural differences in vegetation patterns (i.e., physiognomy) to classify biogeographic regions [5–7]. More recently, advances in molecular phylogenetics and sequencing have revealed many of the evolutionary relationships that connect the biota of the world, and a growing number of studies have integrated this phylogenetic information into **biogeographic regionalization** (see Glossary) schemes [8–13].

Such **phylogenetic regionalization** can help reveal the spatial and evolutionary structure of biodiversity [8,10–13]. However, the mechanisms underlying these patterns remain poorly understood. Here, we discuss the key processes that likely drive different phylogenetic

[1]Department of Organismic and Evolutionary Biology, Harvard University, Cambridge, MA 02138, USA
[2]Department of Plant Sciences, University of Pretoria, Private Bag X20, Hatfield 0028, Pretoria, South Africa
[3]Department of Biological Sciences, University of Cape Town, Private Bag X3, Rondebosch 7701, Cape Town, South Africa
[4]Institut de Recherche en Biologie Végétale, Département de Sciences Biologiques, Université de Montréal, 4101 Sherbrooke East, Montréal, QC, H1X 2B2, Canada
[5]Department of Biology, McGill University, Montreal, QC, H3A 0G4 Canada
[6]African Centre for DNA Barcoding, University of Johannesburg, APK Campus, PO Box 524, Auckland Park 2006, Johannesburg, South Africa

*Correspondence:
barnabas_daru@fas.harvard.edu
(B.H. Daru).

regionalization patterns, and link them to how species diversity is generated, distributed, and maintained across landscapes and seascapes. We then illustrate how these different processes influence the phylogenetic distance separating geographic regions. Finally, we discuss some of the challenges, conservation implications, and perspectives that might shape future research in this emerging research area.

Incorporating Phylogenetic Information in Biogeographic Regionalization

Phylogenetic regionalization involves delineating landscapes or seascapes into geographic units that capture information on the evolutionary relationships of the species assemblages within [8,12–14]. This process requires two types of data: spatial distribution data and knowledge of the phylogenetic relationships of the species present (Box 1). Distribution data can include point occurrence records from online repositories such as the Global Biodiversity Information Facility (GBIF [15]) and iDigBio[i], or range maps derived from herbarium specimens, published monographs, and field guides that have been validated by experts (e.g., IUCN range maps[ii]). These data can be represented as a matrix, often with species as columns and rows as spatial cells or communities (Box 1). Phylogenies are usually reconstructed from DNA sequence data or using supertree methods [16]. Here, we focus on one commonly used approach for delineating phylogenetic regions, outlined in Box 1; however, various other methods, such as using network theory [14] or parsimony analysis of endemism [17], can be used depending on the research question(s), taxa, and scale of the study. All these methods enable the association of species into distinct phylogenetically delimited biogeographic units, which we refer to as '**phyloregions**' [12], as illustrated for the mammals of Madagascar in Box 2.

Processes that Drive Phylogenetic Regionalization

The distribution of biodiversity across the surface of the globe is shaped by multiple processes, and differentiating among them has been a central challenge to the fields of biogeography and macroecology [18]. Here, we discuss the four major mechanisms that determine the broad spatial structure and composition of phylodiversity: **dispersal limitation**, **niche conservatism**, extinction, and **speciation** (Figure 1).

Dispersal

Dispersal involves the movement and establishment of a species in a new environment, and can be viewed as the ecological equivalent of gene flow in population genetics [19]. If dispersal distance is limited such that lineages are restricted geographically, co-occurring species are likely to represent clusters of close relatives and we would expect increased evolutionary separation between geographic regions. Thus, dispersal limitation predicts decreased similarity between phyloregions that are geographically distant (e.g., [20]). The effects of dispersal might be independent of species richness or composition, and can be evaluated by comparing the fit of phyloregions to the environmental dissimilarity between sites versus their fit to expectations under models of distance decay (e.g., [21]). We predict a strong negative correlation between phylogenetic similarity and geographic distance when communities are comprised of species with poor dispersal abilities, whereas the phylogenetic composition of communities including species with superior dispersal abilities is expected to be more homogenous across geographic space (Figure 1A). For instance, consider a phylogenetic regionalization for four taxa distributed discretely across a landscape, as depicted in Figure 2: if all else is equal (i.e., evolutionary rates are constant across the landscape, and geographical distances between phyloregions are more or less equivalent), phyloregions that cluster together will be composed of assemblages connected by high dispersal rates, whereas those that fall separately will represent assemblages characterized by low dispersal rates, which is reflected in their high **phylogenetic beta diversity** (PBD); that is, the boundaries separating phyloregions might become blurred if species disperse widely because there is a reduction in the evolutionary

distance between phyloregions and, ultimately, this can lead to homogeneity and loss of distinction between phyloregions (Figure 2A).

The rate of dispersal of a taxon and its colonization of new environments can depend on both species attributes and features of the landscape or seascape, such as the presence of deserts [22] and mountains [23], aridification [24], glaciation [25], and paleoclimatic changes [26]. Steep climatic gradients along heterogenous mountains have been invoked as barriers that can promote geographic isolation, thereby generating patterns of high PBD across relatively short distances [27]. The ecological barriers limiting exchange of taxa between regions differ in marine versus terrestrial systems [28]. As a result, marine organisms tend to be less constrained than terrestrial species, and can disperse across larger distances, often facilitated by ocean currents [28], whereas freshwater species tend to be more constrained by catchment boundaries. Seagrass phyloregions, for example, provide compelling evidence for the homogenizing effect of long-distance dispersal with geographically disjunct temperate phyloregions harboring closely related species assemblages [13]. In such cases, environmental filtering, rather than geographic distance, might be more important in shaping the distribution of phylogenetic diversity. Indeed, seagrass phyloregions show significant correlation with environmental variables, such as sea surface temperature [13], suggesting that environment selects and is, in part, responsible for the spatial variation of marine biological diversity (see following section on niche conservatism). Dispersal can also depend on the competitive ability and fecundity of a taxon. In plants, for example, factors such as the number of seeds produced, and seed-to-adult survival probability after displacement, as well as the availability of dispersal vectors (e.g., water, birds, insects, bats, large mammals, and wind) affect their overall dispersal ability [29]. Dispersal ability might additionally be linked to persistence in habitats (e.g., species longevity and woodiness) and rapid colonization (e.g., herbaceous annuals) [30,31]. Thus, phyloregions can not only help us better understand the relative importance of dispersal in shaping the distribution of biodiversity, but also provide insight into the ecological properties of the species that comprise them.

Box 1. Examples of Tools and Methods to Perform Phylogenetic Regionalization

Phylogenetic Beta Diversity

The first step in phylogenetic regionalization quantifies the turnover of shared branch lengths among pairs of communities (Figure IC) through metrics of phylogenetic beta diversity (PBD [72]). PBD can be quantified using a variety of approaches (reviewed in [75,76]), including PhyloSor, UniFrac, Rao's quadratic entropy, and Simpson's phylogenetic index ($p\beta_{sim}$ [8]). The $p\beta_{sim}$ index is commonly used in biogeographical regionalization because it is insensitive to differences in species richness [1,8,77].

Selection of Cluster Algorithms

The second step is to choose a clustering algorithm that best describes the site × site PBD matrix [1,12]; alternatives include complete linkage, unweighted pair-group method using arithmetic averages, unweighted pair-group method using centroids, weighted pair-group method using arithmetic averages, weighted pair-group method using centroids, Ward's minimum variance, and the divisive hierarchical method of divisive analysis (DIANA) (Figure ID). The performances of these algorithms can be evaluated using statistics such as the cophenetic correlation coefficient [78] and Gower's distance [79]. The cophenetic correlation ranges from 0 to 1, and should be maximized. Gower distance should be minimized [80].

Determining Optimal Number of Clusters

After selecting the appropriate clustering algorithm, one must determine the optimal number of clusters. The latter has attracted particular scrutiny because the number of meaningful biogeographic regions may appear arbitrary [68]. We suggest that more objective approaches that can differentiate distinct phyloregions from transition zones be favored. The 'elbow' method [81] can be used for dividing the hierarchical dendrogram into meaningful clusters, based on the 'elbow' or 'knee' of an evaluation graph that corresponds to the point of optimal curvature (Figure IE). This can be easily achieved using the 'elbow' function in the R package GMD [82] or the Calinsky criterion in the R package vegan [83]. In addition, the hierarchical dendrogram could be cut at varying lengths to evaluate particular hypotheses, and evaluate the ability of different processes (e.g., tectonic movements and environmental shifts [84]) to explain the boundaries between phyloregions.

Describing Relationships among Phyloregions

The final step is to describe relationships among phyloregions, which can be done using hierarchical dendrograms of dissimilarity and ordination techniques, such as non-metric multidimensional scaling (NMDS) or principal coordinate analysis. Unlike a hierarchical cluster analysis, which distorts the original distance matrix (see [85]), ordination methods represent the original distance matrix with minimal loss of information [86]. Results can be presented as geographic maps, dendrograms, or ordination plots.
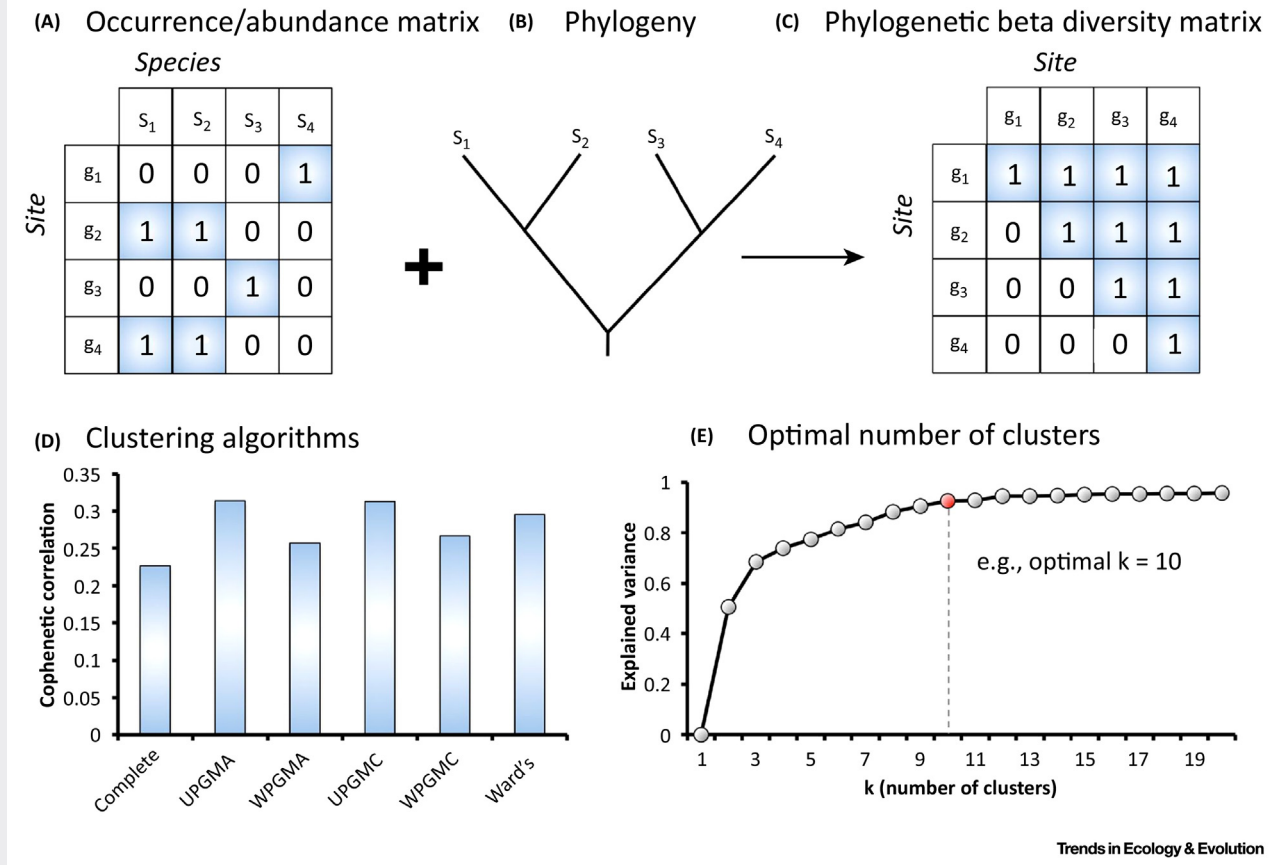

Figure I. Schematic Overview of Phylogenetic Regionalization. (A) Species presence/absence data are combined with (B) branch-length information from a phylogenetic tree to calculate (C) the phylogenetic beta diversity matrix. The performance of (D) clustering algorithms is evaluated using cophenetic correlation and (E) the threshold of explained variances to identify the optimal number of clusters. The 'elbow' (optimal cluster, k) of the graph in (E) is indicated by the red circle. Abbreviations: UPGMA, unweighted pair-group method using arithmetic averages; UPGMC, unweighted pair-group method using centroids; WPGMA, weighted pair-group method using arithmetic averages; WPGMC, weighted pair-group method using centroids.

Phylogenetic Niche Conservatism

The tendency of lineages to be restricted to their ancestral niche, potentially encouraging *in situ* radiation, is referred to as 'niche conservatism' [32,33]. Such phylogenetic niche conservatism can drive patterns of evolutionary divergence and the geographic isolation of lineages [34]. Here, we differentiate between niche conservatism as a pattern versus a process (*c.f.* [35]), placing emphasis on the latter. Several factors can prevent lineages from adapting to new niches, including gene flow, stabilizing selection, patterns of niche filling, and lack of genetic variation [36]. If niches evolve slowly, species will tend to retain their ancestral niche for longer and, thus, remain in their area of origin, regardless of dispersal ability. The radiation of particular clades or lineages in similar climate space can give rise to geographically distinct phyloregions (Figure 2B), which can be disjunct if climate space is patchy and species are good dispersers,

as in the aforementioned seagrasses. This is consistent with the climate-richness relationship hypothesis of Buckley *et al.* [37], which predicts a high tropical diversity for mammalian lineages with tropical origins versus low richness for lineages with temperate origins [38,39]. By contrast, if niches tend to be evolutionarily labile, lineages can more easily transition between different environments, and phyloregions might become more homogeneous (Figure 2B). Under this scenario, in our example in Figure 2, the two phyloregions in the non-metric multidimensional scaling (NMDS) ordination plot might represent species within clades characterized by high niche conservatism, whereas lineages that coalesce into the single phyloregion (lower panel in Figure 2B) might have relatively rapid rates of niche evolution (i.e., low niche conservatism), allowing for a greater exchange of taxa between them.

Extinction

Extinction can influence biogeographic patterns by effecting both taxon richness and composition. Past extinctions have likely shaped the current structure of **phylodiversity** observed today; for instance, the phylogenetic structure of present-day mammal assemblages strongly reflects the imprint of Pleistocene extinctions [40]. However, different patterns of extinctions lead to different phylodiversity patterns. If local assemblages are comprised of lineages that represent survivors of once more diverse clades (**paleoendemics**), then we might expect highly distinct patterns of phylodiversity (Figure 3A). The temperate seagrass phyloregions are one such example [13]. The closure of the Isthmus of Panama that coincided with a major glaciation event in the northern hemisphere approximately 3.1 million years ago (Ma) caused the extinction or southward migration of lineages to adjoining oceans [41]; thus, the remaining species assemblage is species-poor, but **evolutionarily distinct** [13]. However, if basal branches contributing to high phylogenetic turnover are lost through extinction of paleoendemics, we would then predict that the phylogenetic distances separating phyloregions would decrease. There is some evidence suggesting that more evolutionarily distinct species are more endangered within mangrove ecosystems [42], and in rockfish [43], but it is not clear whether this pattern is widely generalizable.

If, by contrast, extinction operates disproportionately upon more derived lineages of species (**neoendemics**), the loss of a few species might have only a small effect (Figure 3A). Such a scenario is not unrealistic. For instance, it has been shown that threatened plant species in the Cape of South Africa, a global biodiversity hotspot, cluster within short branches at the tips of the phylogeny of the flora [44]. However, because the intensity of anthropogenic drivers of extinction, such as habitat transformation, pollution, and climate change, are spatially

---

Box 2. Phylogenetic Beta Diversity of Madagascan Mammals

In this example, the $p\beta_{sim}$ metric was used to divide 3263 equally sized cells of $0.125° \times 0.125°$ into phyloregions (Figure S1.1 in the supplemental information online). The phylogenetic regionalization of $p\beta_{sim}$ yielded nine phyloregions (Figure IA) that closely matched regions delimited by species-level beta diversity (Figure IB) ($r = 0.85$, $P = 0.001$, Mantel test $p\beta_{sim}$ and β_{sim} and 999 permutations). There was also a broad correspondence between the phyloregions and the existing vegetation map of Madagascar (e.g., [87]), including regions matching closely to Sambirano forest, grassland/woodland/bushland mosaic, and dry deciduous forest. However, we additionally highlight three novel regions (labelled X, Y, and Z in Figure I). Region X occurs at the southern tip of Madagascar encompassing parts of the arid spiny bush (Figure IA), region Y occurs in the west coast along the southwestern coast of Mahajanga, and region Z occurs in the northeast directly north of Helodrano Antongila Bay (Figure IA). Critically, region Z is only identified using the metric of $p\beta_{sim}$.

Phyloregions can be indexed using common phylogenetic metrics (Figure II). Phyloregion X in the south emerged as the most evolutionarily distinct (mean $p\beta_{sim} = 0.224$), perhaps indicating that it is a center of speciation or nonrandom extinction, whereas the evergreen rainforest I in the east coast captured both high species richness and unique phylogenetic diversity (phylogenetic endemism; mean PE = 470.54). This evergreen rainforest habitat might represent a hotspot of neo- or paleoendemism [88]. The dry deciduous forest II phyloregion in the northernmost tip of the island

captured the highest number of endemic species. Last, phyloregion Y in the west coast encapsulated a high proportion of evolutionarily distinct and globally endangered (EDGE) species. This new phyloregion could be a target to focus conservation efforts on because it represents an area where unique evolutionary history is currently at risk of extinction.

To identify clades that are significantly contributing to the various patterns of phylogenetic turnover in each phyloregion, we used a node-based approach following [89]. We found that the most distinct phyloregion X is dominated by tenrecs (Tenrecidae) and bats. The nodes contributing to the high phylogenetic endemism in the evergreen rainforest I were mainly found within mouse lemurs (Cheirogaleidae) and tenrecs (Figure S1.4 in the supplemental material online).
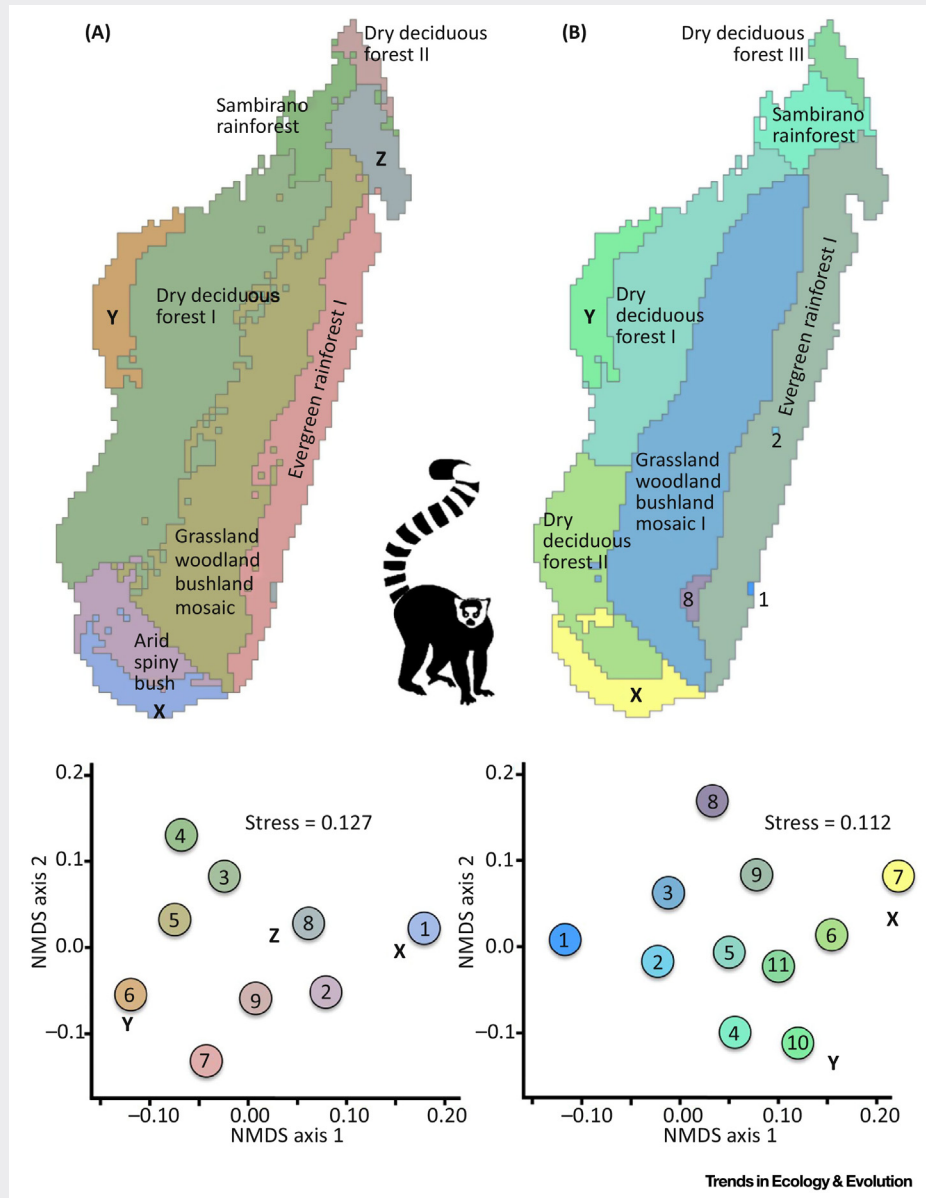

Figure I. Empirical Example Comparing Patterns of Phylogenetic Regionalization Based on Phylogenetic Beta Diversity (A) versus Species-Level Regionalization Based on Beta Diversity (B) for Mammals of Madagascar. Relationships among clusters are represented in geographic space (maps) and in NMDS ordination (graphs). Colors differentiating regions in geographic space and non-metric multidimensional scaling (NMDS) are similar and depict the amount of phylogenetic (A) or species (B) turnover among regions.
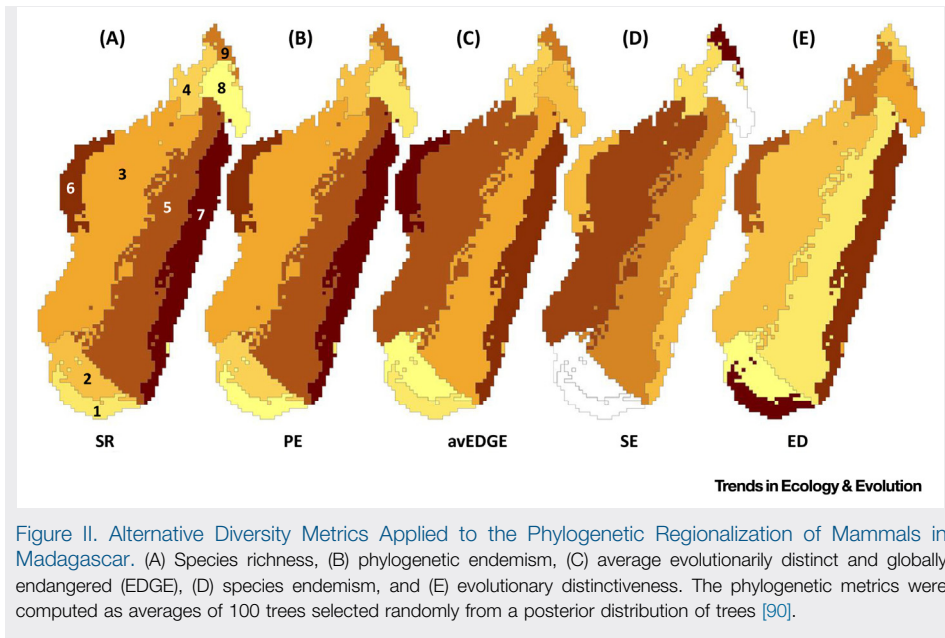
Figure II. Alternative Diversity Metrics Applied to the Phylogenetic Regionalization of Mammals in Madagascar. (A) Species richness, (B) phylogenetic endemism, (C) average evolutionarily distinct and globally endangered (EDGE), (D) species endemism, and (E) evolutionary distinctiveness. The phylogenetic metrics were computed as averages of 100 trees selected randomly from a posterior distribution of trees [90].
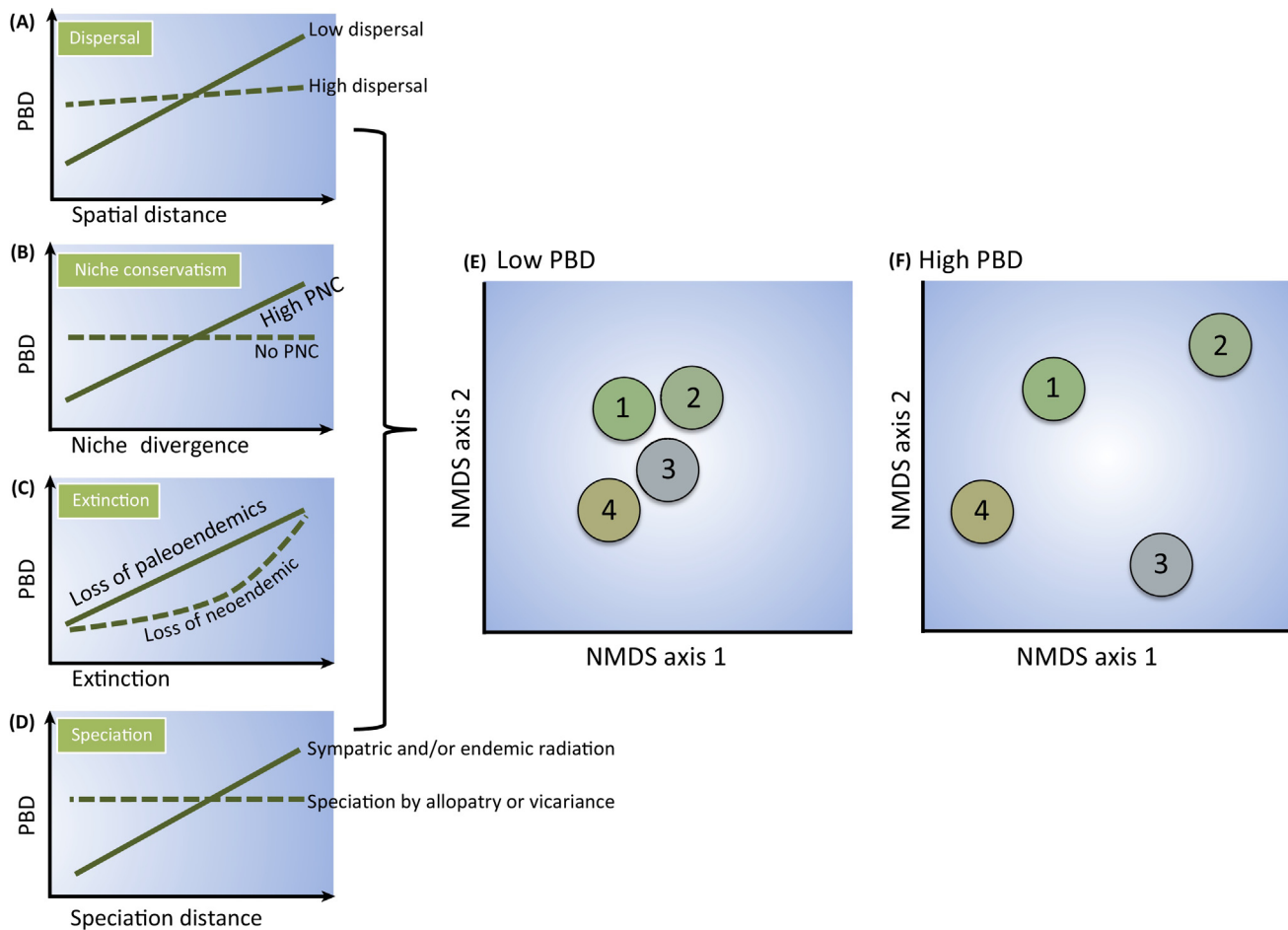
structured, we risk losing entire clades from some regions, and the future structure of phyloregions might then look very different.

Speciation

The geographical patterns of speciation are much debated. In plants, for example, speciation can be promoted by long-distance dispersal, flower–pollinator coevolution, polyploidization, and hybridization that might favor the formation of small and reproductively isolated populations [45]. Rates of new species formation might be higher when populations become geographically isolated from one another (**allopatric speciation**), although it is also possible for species to diverge without geographic isolation (**sympatric speciation** [46,47]).

The separation of phyloregions can be driven by broad patterns of speciation at the (regional/ global geographic) scale on which phylogenetic regionalization is performed. Sympatric, or local (fine-scale) allopatric speciation can result in highly speciose, yet phylogenetically clustered phyloregions across the landscape. The Fynbos biome of southern Africa provides one such example. This biome comprises high species richness, but for the most part this diversity can be traced back to the recent radiation of just 33 clades *ca.* 18–8 Ma [48]. As a consequence, the Fynbos phyloregion is the most evolutionarily distinct floristic phyloregion in southern Africa [12]. If species arise through sympatric speciation, sibling species will tend to occupy the same broad environmental niche space (of course, we might still observe fine-scale niche partitioning, especially in the process of adaptive radiations or ecological speciation), elevating the phylogenetic distance between environments and distinctiveness of phyloregions (Figure 3B). By contrast, allopatric speciation across a landscape can give rise to several smaller, less distinct, phylogenetically overdispersed phyloregions. For instance, vicariance events can drive the formation of new species, but not the phylogenetic separation between regions, because sister species would occur on either side of the barrier (Figure 3B).
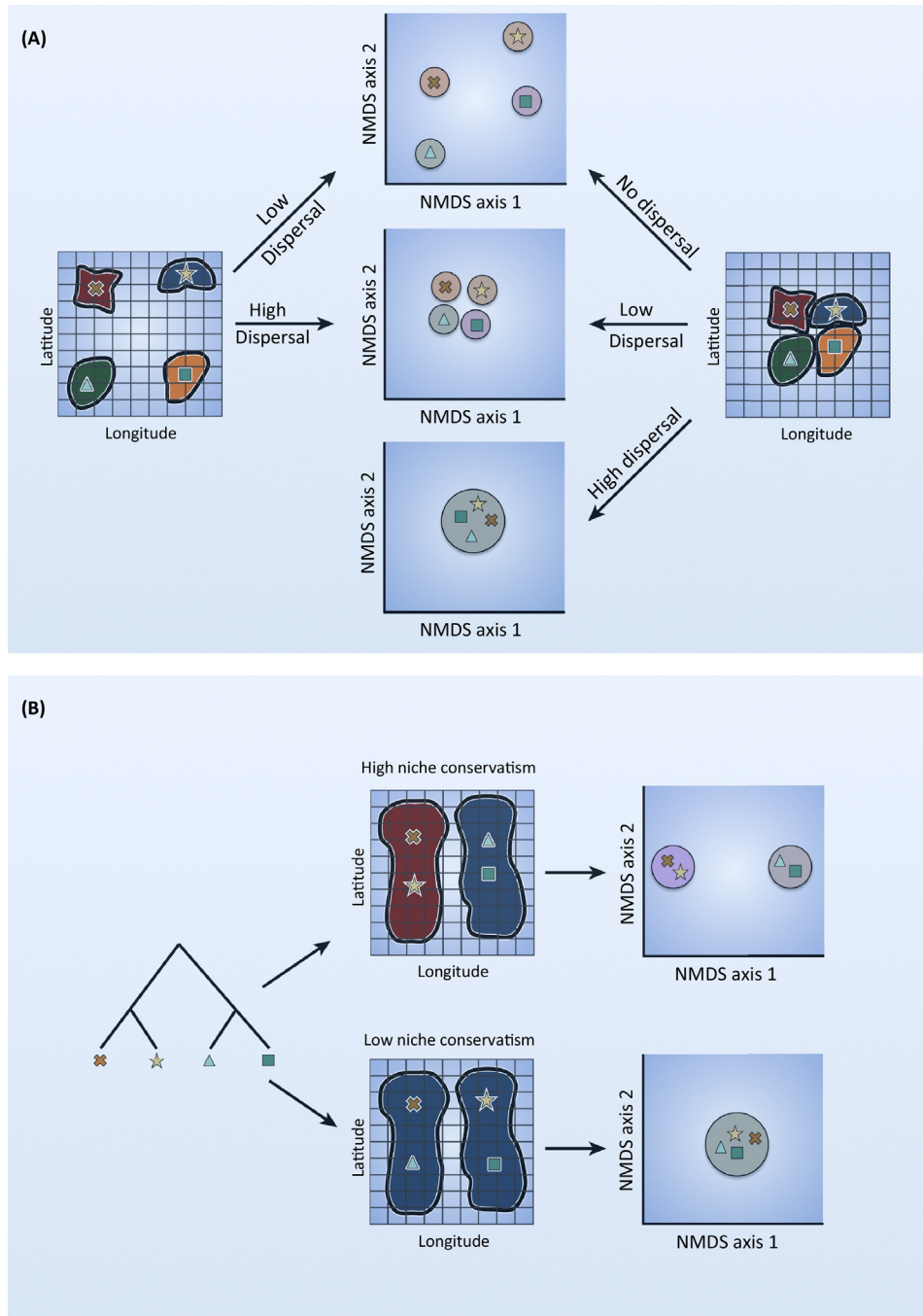
The examples we present here are greatly simplified scenarios; multiple mechanisms of speciation can act on the same communities and landscapes, generating complex phylogenetic patterns. There are also additional modes of speciation that we do not discuss, such as
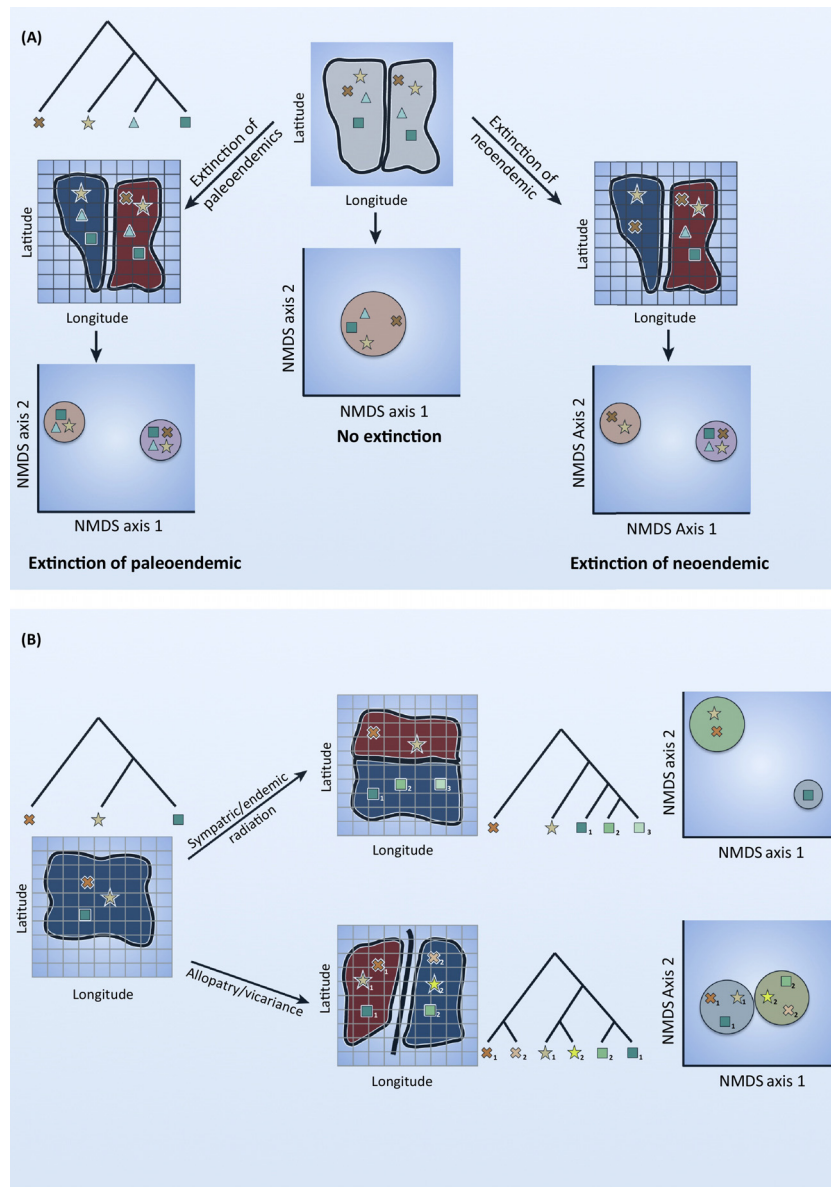
Figure 1. The Four Principal Processes Underlying the Formation of Phyloregions and their Relationship with Phylogenetic Beta Diversity (PBD). (A) Dispersal limitation: this will cause the aggregation of close relatives leading to high evolutionary separation among phyloregions. (B) Niche conservatism: lineages with conserved niches are predicted to remain in their ancestral environments and, thus, demonstrate high PBD, resulting in the clustering of lineages into distinct phyloregions. Lineages with evolutionarily labile niches can transition between different environments, leading to greater homogeneity and loss of distinction among phyloregions. (C) Extinction: losing members of an ancient beta lineage (paleoendemics) from a region can cause large changes in the phylogenetic diversity of the region, and alter greatly the phylogenetic distance separating regions. However, the loss of more phylogenetically derived species will have less impact on phylogenetic structure, at least initially. If extinction is phylogenetically clustered and progressive, eventually entire clades will be lost, elevating rates of phylogenetic change. (D) Speciation: when the formation of new species from a single ancestral species occurs without geographical isolation (e.g., via sympatric speciation), a high PBD leading to increased separation of phyloregions will result. Under a model of allopatric speciation with geographical isolation, PBD decreases leading to homogeneity among phyloregions. The distribution of phyloregions in non-metric multidimensional scaling (NMDS) space under a scenario of (E) low and (F) high PBD. Abbreviation: PNC, phylogenetic niche conservatism.

para- or peripatric speciation [49]. Hence, the effect of speciation on phylogenetic regionalization patterns is likely to be dependent on both the geographic and taxonomic scale of interest. We do not suggest that patterns of phylogenetic regionalization can be simply translated into process, and we acknowledge that multiple processes can lead to similar patterns (Figures 1–3). However, we propose that there is nonetheless a mapping of process to pattern that allows alternative hypotheses to be contrasted, and some processes to be excluded (e.g., separation of geographically close phyloregions in NMDS space is inconsistent with high dispersal between them).

Trends in Ecology & Evolution

Figure 2. Effects of Dispersal and Niche Conservatism on the Structure and Composition of Phyloregions.
(A) When dispersal is low, phylogenetic beta diversity (PBD) is predicted to increase with spatial distance, resulting in a greater separation of phyloregions in non-metric multidimensional scaling (NMDS) ordination space. Under high dispersal, we predict only a weak correlation between PBD and geographic distance because species would tend to distribute widely across the landscape, resulting in little separation among phyloregions in NMDS ordination space. (B) If niches are conserved, high PBD is expected, which can lead to the formation of evolutionarily distinct phyloregions (top row). If niche conservatism is low, beta diversity will also tend to be low, and the separation of phyloregions is reduced (bottom row). Different-colored symbols refer to species combinations (A) or individual species (B), and circles within the NMDS plots refer to the formation of the species assemblages into phyloregions.

Figure 3. Effects of Extinction and Speciation on the Structure of Phyloregions. (A) The local extinction of a paleoendemic lineage can lead to an increased separation between phyloregions. However, the extinction of a neoendemic might have only small effect on phylogenetic structure, at least initially. It is only when we start to lose entire clades that loss of neoendemics will result in a significant change in phylogenetic diversity (see also Figure 1C in the main text). (B) Speciation can have different impacts on phylogenetic structure and the separation of phyloregions depending on the predominant mode of species formation. Where speciation is sympatric or occurs within an endemic radiation, phylogenetic beta diversity and the separation of phyloregions will be increased. In cases where speciation occurs in allopatry, especially as a consequence of geographic separation, the separation of phyloregions is less pronounced because the species within each region are interdigitated on the phylogenetic tree. Different-colored symbols refer to different species, and circles within non-metric multidimensional scaling (NMDS) plots refer to the formation of the species assemblages into phyloregions.

Additional Insights from Phyloregions

Phylogenetic regionalization can provide a window into the past history of biomes [50]. In the absence of a detailed fossil record, phylogeny can allow us to investigate the origin and ages of present-day phyloregions by exploring the evolutionary turnover (PBD) separating these areas at different phylogenetic depths. For example, we can deconstruct the phylogenetic tree at successive time intervals and reconstruct phyloregions at each iteration to determine the phylogenetic depth at which the signal for their present-day structure emerges. This approach allows us to draw inference on past environments that might have differed from those occupied by present-day descendants [51]. For instance, Daru (B.H. Daru, PhD thesis, University of Johannesburg, 2015) showed that most terrestrial phytogeographic regions in southern Africa, including the Fynbos, Grassland, Nama Karoo, and the Savanna, likely originated during the Miocene, from approximately 20 Ma to the present day, which represented an era of warmer global climates and coincided with the appearance of grasslands [52].

In contrast to previous biogeographic regionalization schemes that treat all species as being equally related, phylogenetic regionalization allows even geographic regions that do not share any taxa in common to be included in analyses, and can highlight patterns that are not apparent from species-level analyses. For example, in a global study of major vertebrate groups, Holt *et al.* [8] found a lack of support for previously described Palearctic boundaries, but defined for the first time three novel associations (the Panamanian, Sino-Japanese, and Oceanian realms) that were not well recognized by previous species-level classifications. Similarly, a recent phylogenetic regionalization of southern Africa [12] revealed biogeographic groupings that matched broadly to traditional classification schemes, as well as two phylogenetically distinct regions (the 'Gariep Karoo' and 'Zambezian transition zone') that had previously been over-looked. In Figure I (Box 2), we identify a new floristic region in Madagascar, phyloregion Z, which is missed by the species-level approach. Additionally, using species taxonomic differences, the dry deciduous forest differentiated into two regions, but these are generally phylogenetically similar and so grouped together in phylogenetic regionalization (Figure IA, Box 2). We also lose the arid spiny bush when we consider only species-level beta diversity. Thus, phylogenetic regionalization allows us to capture important information that can be missed by species-level approaches.

However, interpreting patterns of phylogenetic regionalization is subject to many of the same constraints that make it difficult to infer process from patterns of phylogenetic clustering and overdispersion. These more standard **ecophylogenetic** metrics have nonetheless provided many new insights into the structure of biodiversity gradients and community composition (and the literature is still expanding rapidly in this field). We strongly believe that patterns of phylogenetic regionalization complement community-level analyses of phylogenetic clustering and overdispersion, because phylogenetic regionalization also evaluates the shared evolutionary relationships among phyloregions (Box 3). Critically, phylogenetic regionalization provides insights into the ecological and evolutionary processes (such as niche conservatism, dispersal, speciation, and extinction) that group species (or sites), whereas testing for phylogenetic under- or overdispersion requires us to predefine the sites and/or communities of interest. We believe that this is a fundamental distinction between these approaches.

Limitations to Detecting Patterns of Phylogenetic Regionalization

Although incorporating phylogenetic information into regionalization schemes can provide novel insights, uncertainties, gaps, and biases in a dataset can have important consequences in delineating biogeographic regions [53]. Possible impacts include biased ecological inferences [54–56] and inefficient conservation management [57]. While it might be impractical to sample every morphotype in an area, it is important to sample representatives of all major lineages present, such that most of the phylogenetic diversity in a region is represented [58].

Box 3. Comparison of Community Analysis of Phylogenetic Clustering and/or Overdispersion versus Phylogenetic Regionalization

Consider a community analysis of clustering and overdispersion of a landscape with six communities (g1–g6) that results in patterns of clustering (shaded), and overdispersion (unshaded) as depicted in Figure I. Assuming all else is equal (i.e., constant evolutionary rates across the landscape), community analysis of clustering and/or overdispersion reveal the phylogenetic structure of one community relative to another, but do not provide information on the shared relationships among and within communities (Figure IC). However, phylogenetic regionalization can reveal not only the shared relationships among all clusters (phyloregions) (Figure ID), but also the amount of evolutionary distinctiveness captured by each phyloregion.
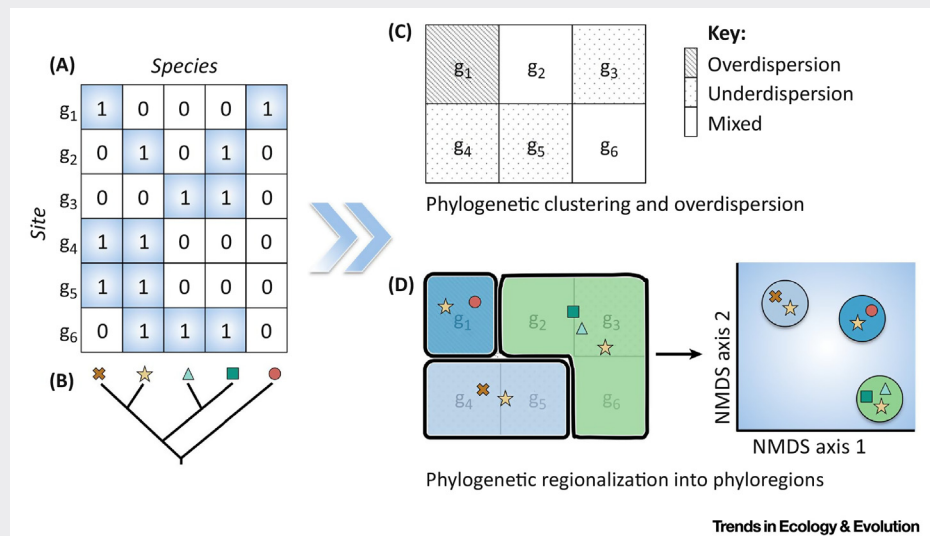

Figure I. Differences between Community Analysis of Phylogenetic Clustering and/or Overdispersion versus Phylogenetic Regionalization. Both analyses require the use of geographical distribution data of the species presented as a species × site matrix (A) and a phylogenetic tree (B). However, the resulting patterns in both cases are different: clustering and overdispersion in (C) versus phylogenetic regionalization into phyloregions in (D). The colors of phyloregions in geographic space and non-metric multidimensional scaling (NMDS) are similar and indicate levels of differentiation of species assemblages in different phyloregions. Phyloregions with similar colors have similar clades, whereas those with different colors differ in the clades they enclose.

Most macroecological studies are plagued by problems of missing taxa; however, there is an increasing number of easily accessible data sets (e.g., on species distributions and phylogeny) in public repositories, such as the IUCN[iii], GBIF [15], and GenBank [59], that make available data spanning large taxonomic and spatial breadths. With the growing efforts in digitization, global scale data sets are becoming available for major vertebrate groups, such as birds [60], amphibians, and mammals, and particular species-rich clades, such as corals, mangroves, and seagrasses[iv]. At the regional scale, additional data are available, for example, for many woody plant species [10,61,62]. Nonetheless, it remains a challenge to obtain reliable data at the appropriate scale for many taxa, and the data that are available are often biased and incomplete [53]. In the absence of a complete data set, a common practice in conservation planning is to use cross-taxon surrogates [63]. This approach assumes that groups for which distributions and evolutionary relationships are well known (typically birds or mammals) can effectively serve as surrogates for taxa that are lacking in information (e.g., plants). However, empirical studies using such surrogates have been mixed in their results (e.g., [64,65]), and the effectiveness of cross-taxon surrogacy might depend on the scale or context of the study (e.g., [63,66]). These issues are nontrivial when delineating phyloregions because regions with similar spatial patterns of species diversity can have different phylogenetic structures [67].

Even when taxonomic sampling is adequate, poor phylogenetic resolution can present additional problems for biogeographic regionalization [68]. The available phylogenetic trees for large-scale regionalization, which might encompass many thousands of species, are often poorly resolved and/or lacking branch length information (e.g., as is common for plants where often resolution below the family level is lacking; see [69,70]). As a consequence, spurious patterns might emerge.

Although our examples are simplistic and presented so as to emphasize the links between process and pattern for each mechanism separately, it is important to note that multiple processes can generate similar patterns, and there might be complicated interactions between processes that shape phylogenetic regionalization. It is challenging to make clear predictions in such multidimensional scenarios, and we instead suggest that simulations, such as those by Pigot and Etienne [71], might provide a powerful way to compare and contrast alternative scenarios.

Finally, geographic scale is also an important consideration, and we might expect more pronounced turnover of lineages at large regional and global scales than at local or community scales [72]. For example, a phylogenetic regionalization at a local scale in Canada did not match the environmental turnover across the landscape, suggesting that abiotic site preferences are not well captured by phylogenetic information at local scales (T.L. Elliott and T.J. Davies, unpublished data, 2017). The lack of phylogenetic regionalization pattern at small scales does not exclude the possibility that ecological processes that operate over large spatial scales, such as environmental filtering, are not relevant at these smaller scales, but suggests that other processes, such as competition, biotic interactions, and short-distance dispersal, are more important [73,74].

Implications for Science, Conservation, and Policy-Making

Phyloregions represent evolutionarily coherent units, a product of species assembly processes and coevolutionary dynamics that have operated over many millions of years. Thus, phyloregions can function as an important unit for the conservation of evolutionary history and help identify unique floras or faunas overlooked by traditional biogeographic delimitations, such as the newly defined Gariep Karoo and Zambezian transition zone in southern Africa [12] or the Panamanian, Sino-Japanese, and Oceanian zoogeographic realms [8]. By enclosing lineages that tend to be conserved in their ancestral areas of origin, phylogenetic regionalization can identify species assemblages that are evolutionary distinct, such as the Fynbos of southern Africa or the Australian zoogeographic realm [50]. Phyloregions with high evolutionary distinctiveness might represent geographic regions that harbor species with rare features not found elsewhere (i.e., locally restricted radiations) and, therefore, should rank highly for conservation purposes [8,12,13]. While evolutionary distinct phyloregions do not necessarily capture the greatest taxonomic richness or unique phylogenetic diversity, it is easy to integrate more traditional conservation metrics, such as richness, into a phyloregions framework for conservation prioritization.

Concluding Remarks

Here, we have highlighted how the emerging field of phylogenetic ecology provides a new perspective on biogeographic patterns and has advanced our understanding on how biodiversity is generated, distributed, and maintained. We have illustrated how phyloregions provide insights into the processes underlying biodiversity gradients that are not captured by more traditional approaches based on species membership. We identify four key drivers that shape broad patterns of phylodiversity: dispersal limitation, extinction, speciation, and niche conservatism, and discuss how it is possible to differentiate among them. Future work can

Outstanding Questions

To what extent does a phylogenetically informed biogeographic regionalization for plants compare with the pre-existing zoogeographic realms?

How do environmental factors, such as soil types, productivity, temperature, habitat and so on, influence the formation of species assemblages into phyloregions?

Can PBD predict the turnover of functional diversity across a landscape and/or seascape?

What are the processes underlying the formation of functionally delimited biogeographic units?

draw from new analytical tools, mathematical and theoretical models, digitized data sets, and genomics to further advance and transform our understanding of phylodiversity. For example, a global-scale phylogenetic regionalization of plants is still lacking, but would represent a major advance because plant distributions are more closely linked to climate, and modern ecozones are more often defined by climate and floristic composition (e.g., WWF ecoregions [5]) with little regard for their evolutionary history (see Outstanding Questions). Finally, we suggest phyloregions represent important biotic units that deserve conservation attention.

Appendix A Resources

[i]www.idigbio.org

[ii]www.iucnredlist.org/technical-documents/spatial-data

[iii]www.iucnredlist.org

[iv]www.iucnredlist.org/technical-documents/spatial-data

Supplemental Information

Supplemental information associated with this article can be found, in the online version, at http://dx.doi.org/10.1016/j.tree.2017.08.013.

References

1. Kreft, H. and Jetz, W. (2010) A framework for delineating biogeographical regions based on species distributions. *J. Biogeogr.* 37, 2029–2053

2. De Candolle, A.P. (1820) Geógraphie botanique. In *Dictionnaire des Sciences Naturelles*, pp. 359–422, Paris, France, F. Levrault

3. Sclater, P.L. (1858) On the general geographical distribution of the members of the class Aves. *J. Proc. Linn. Soc. Lond. Zool.* 2, 130–136

4. Huxley, T.H. (1868) On the classification and distribution of the Alectoromorphae and Heteromorphae. *Proc. Zool. Soc. Lond.* 1868, 296–319

5. Olson, D.M. *et al.* (2001) Terrestrial ecoregions of the worlds: a new map of life on Earth. *Bioscience* 51, 933–938

6. Burgess, N. *et al.* (2004) *The Terrestrial Ecoregions of Africa and Madagascar,* Island

7. Spalding, M.D. *et al.* (2007) Marine ecoregions of the world: a bioregionalization of coastal and shelf areas. *Bioscience* 57, 573–583

8. Holt, B.G. *et al.* (2013) An update of Wallace's zoogeographic regions of the world. *Science* 339, 74–78

9. Qian, H. *et al.* (2013) Phylogenetic beta diversity of angiosperms in North America. *Glob. Ecol. Biogeogr.* 22, 1152–1161

10. Kubota, Y. *et al.* (2014) Beta diversity of woody plants in the Japanese archipelago: the roles of geohistorical and ecological processes. *J. Biogeogr.* 41, 1267–1276

11. Li, R. *et al.* (2015) A phylogenetically informed delineation of floristic regions within a biodiversity hotspot in Yunnan, China. *Sci. Rep.* 5, 9396

12. Daru, B.H. *et al.* (2016) A novel phylogenetic regionalization of phytogeographical zones of southern Africa reveals their hidden evolutionary affinities. *J. Biogeogr.* 43, 155–166

13. Daru, B.H. *et al.* (2017) Phylogenetic regionalization of marine plants reveals close evolutionary affinities among disjunct temperate assemblages. *Biol. Conserv.* 213, 351–356

14. Vilhena, D.A. and Antonelli, A. (2015) A network approach for identifying and delimiting biogeographical regions. *Nat. Commun.* 6, 6848

15. Edwards, J.L. *et al.* (2000) Interoperability of biodiversity databases: biodiversity information on every desktop. *Science* 289, 2312–2314

16. Webb, C.O. and Donoghue, M.J. (2005) Phylomatic: tree assembly for applied phylogenetics. *Mol. Ecol. Resour.* 5, 181–183

17. Rosen, B.R. and Smith, A.B. (1988) Tectonics from fossils? Analysis of reef-coral and sea-urchin distributions from late Cretaceous to recent, using a new method. In *Gondwana and Tethys* (Audley-Charles, M.G. and Hallam, A., eds), pp. 275–306, Oxford University Press

18. Crisp, M.D. *et al.* (2009) Phylogenetic biome conservatism on a global scale. *Nature* 458, 754–756

19. Vellend, M. (2010) Conceptual synthesis in community ecology. *Q. Rev. Biol.* 85, 183–206

20. Hubbell, S.P. (2001) *The Unified Neutral Theory of Biodiversity and Biogeography (MPB-32),* Princeton University Press

21. Gilbert, B. and Lechowicz, M.J. (2004) Neutrality, niches, and dispersal in a temperate forest understory. *Proc. Natl. Acad. Sci. U. S. A.* 101, 7651–7656

22. Schuster, M. *et al.* (2006) The age of the Sahara Desert. *Science* 311, 821

23. Janzen, D.H. (1967) Why mountain passes are higher in the tropics. *Am. Nat.* 101, 233–249

24. Jürgens, N. (1997) Floristic biodiversity and history of African arid regions. *Biodivers. Conserv.* 6, 495–514

25. Provan, J. and Bennett, K.D. (2008) Phylogeographic insights into cryptic glacial refugia. *Trends Ecol. Evol.* 23, 564–571

26. Gonçalves, D.V. *et al.* (2012) Phylogeny of North African Agama lizards (Reptilia: Agamidae) and the role of the Sahara desert in vertebrate speciation. *Mol. Phylogenet. Evol.* 64, 582–591

27. Kerr, J.T. and Packer, L. (1997) Habitat heterogeneity as a determinant of mammal species richness in high-energy regions. *Nature* 385, 252–254

28. Kinlan, B.P. and Gaines, S.D. (2003) Propagule dispersal in marine and terrestrial environments: a community perspective. *Ecology* 84, 2007–2020

29. Nathan, R. (2006) Long-distance dispersal of plants. *Science* 313, 786–788

30. Maurer, K. *et al.* (2003) Frequency of plant species in remnants of calcareous grassland and their dispersal and persistence characteristics. *Basic Appl. Ecol.* 4, 307–316

31. Saar, L. *et al.* (2012) Which plant traits predict species loss in calcareous grasslands with extinction debt? *Divers. Distrib.* 18, 808–817

32. Webb, C.O. *et al.* (2002) Phylogenies and community ecology. *Annu. Rev. Ecol. Evol. Syst.* 33, 475–505

33. Wiens, J.J. (2004) Speciation and ecology revisited: phylogenetic niche conservatism and the origin of species. *Evolution* 58, 193–197

34. Park, D.S. and Potter, D. (2015) Why close relatives make bad neighbours: phylogenetic conservatism in niche preferences and dispersal disproves Darwin's naturalization hypothesis in the thistle tribe. *Mol. Ecol.* 24, 3181–3193

35. Wiens, J.J. *et al.* (2010) Niche conservatism as an emerging principle in ecology and conservation biology. *Ecol. Lett.* 13, 1310–1324

36. Wiens, J.J. and Graham, C.H. (2005) Niche conservatism: integrating evolution, ecology, and conservation biology. *Annu. Rev. Ecol. Evol. Syst.* 36, 1–689

37. Buckley, L.B. *et al.* (2010) Phylogeny, niche conservatism and the latitudinal diversity gradient in mammals. *Proc. R. Soc. B* 277, 2131–2138

38. Hawkins, B.A. *et al.* (2006) Post-Eocene climate change, niche conservatism, and the latitudinal diversity gradient of New World birds. *J. Biogeogr.* 33, 770–780

39. Pyron, R.A. and Burbrink, F.T. (2009) Can the tropical conservatism hypothesis explain temperate species richness patterns? An inverse latitudinal biodiversity gradient in the New World snake tribe Lampropeltini. *Glob. Ecol. Biogeogr.* 18, 406–415

40. Davies, T.J. and Buckley, L.B. (2012) Exploring the phylogenetic history of mammal species richness. *Glob. Ecol. Biogeogr.* 21, 1096–1105

41. Barry, R.G. (1989) The present climate of the Arctic Ocean and possible past and future states. In *The Arctic Seas: Climatology, Oceanography, Geology, and Biology* (Herman, Y., ed.), pp. 1–46, Van Nostrand Reinhold

42. Daru, B.H. *et al.* (2013) A global trend towards the loss of evolutionarily unique species in mangrove ecosystems. *PLoS One* 8, e66686

43. Magnuson-Ford, K. *et al.* (2009) Rockfish (*Sebastes*) that are evolutionarily isolated are also large, morphologically distinctive and vulnerable to overfishing. *Biol. Conserv.* 142, 1787–1796

44. Davies, T.J. *et al.* (2011) Extinction risk and diversification are linked in a plant biodiversity hotspot. *PLoS Biol.* 9, e1000620

45. Rieseberg, L.H. and Willis, J.H. (2007) Plant speciation. *Science* 317, 910–914

46. Savolainen, V. *et al.* (2006) Sympatric speciation in palms on an oceanic island. *Nature* 441, 210–213

47. Via, S. (2001) Sympatric speciation in animals: the ugly duckling grows up. *Trends Ecol. Evol.* 16, 381–390

48. Linder, H.P. (2003) The radiation of the Cape flora, southern Africa. *Biol. Rev.* 78, 597–638

49. Coyne, J.A. and Orr, H.A. (2004) *Speciation,* Sinauer Associates

50. Pennington, R.T. *et al.* (2006) Insights into the historical construction of species-rich biomes from dated plant phylogenies, neutral ecological theory and phylogenetic community structure. *New Phytol.* 172, 605–616

51. Rosauer, D.F. *et al.* (2014) Phylogenetic generalised dissimilarity modelling: a new approach to analysing and predicting spatial turnover in the phylogenetic composition of communities. *Ecography* 37, 21–32

52. Behrensmeyer, A.K. *et al.* (1992) *Terrestrial Ecosystems through Time: Evolutionary Paleoecology of Terrestrial Plants and Animals,* University of Chicago Press

53. Meyer, C. *et al.* (2016) Multidimensional biases, gaps and uncertainties in global plant occurrence information. *Ecol. Lett.* 19, 992–1006

54. Prendergast, J.R. *et al.* (1993) Correcting for variation in recording effort in analyses of diversity hotspots. *Biodivers. Lett.* 1, 39–53

55. Hortal, J. *et al.* (2008) Historical bias in biodiversity inventories affects the observed environmental niche of the species. *Oikos* 117, 847–858

56. Syfert, M.M. *et al.* (2013) The effects of sampling bias and model complexity on the predictive performance of MaxEnt species distribution models. *PLoS One* 8, e55158

57. Grand, J. *et al.* (2007) Biased data reduce efficiency and effectiveness of conservation reserve networks. *Ecol. Lett.* 10, 364–374

58. Davies, T.J. *et al.* (2012) Incompletely resolved phylogenetic trees inflate estimates of phylogenetic conservatism. *Ecology* 93, 242–247

59. Clark, K. (2016) GenBank. *Nucleic Acids Res.* 44, D67–D72

60. BirdLife International (2012) *BirdLife International and NatureServe Bird Species Distribution Maps of the World.* (Ver. 2), BirdLife International

61. Maurin, O. *et al.* (2014) Savanna fire and the origins of the "underground forests" of Africa. *New Phytol.* 204, 201–214

62. González-Caro, S. *et al.* (2014) Phylogenetic alpha and beta diversity in tropical tree assemblages along regional-scale environmental gradients in northwest South America. *J Plant Ecol.* 7, 145–153

63. Rodrigues, A.S.L. and Brooks, T.M. (2007) Shortcuts for biodiversity conservation planning: the effectiveness of surrogates. *Annu. Rev. Ecol. Evol. Syst.* 38, 713–737

64. Grenyer, R. *et al.* (2006) Global distribution and conservation of rare and threatened vertebrates. *Nature* 444, 93–96

65. Zupan, L. *et al.* (2014) Spatial mismatch of phylogenetic diversity across three vertebrate groups and protected areas in Europe. *Divers. Distrib.* 20, 674–685

66. Ferrier, S. and Watson, G. (1997) *An Evaluation of the Effectiveness of Environmental Surrogates and Modelling Techniques in Predicting the Distribution of Biological Diversity,* Environment Australia

67. Hawkins, B.A. *et al.* (2012) Different evolutionary histories underlie congruent species richness gradients of birds and mammals. *J. Biogeogr.* 39, 825–841

68. Kreft, H. and Jetz, W. (2013) Comment on "An update of Wallace's zoogeographic regions of the world". *Science* 341, 343

69. Chave, J. *et al.* (2006) Regional and phylogenetic variation of wood density across 2456 neotropical tree species. *Ecol. Appl.* 16, 2356–2367

70. Kerkhoff, A.J. *et al.* (2014) The latitudinal species richness gradient in New World woody angiosperms is consistent with the tropical conservatism hypothesis. *Proc. Natl. Acad. Sci. U. S. A.* 111, 8125–8130

71. Pigot, A.L. and Etienne, R.S. (2015) A new dynamic null model for phylogenetic community structure. *Ecol. Lett.* 18, 153–163

72. Graham, C.H. and Fine, P.V.A. (2008) Phylogenetic beta diversity: linking ecological and evolutionary processes across space and time. *Ecol. Lett.* 11, 1265–1277

73. Vamosi, S.M. *et al.* (2009) Emerging patterns in the comparative analysis of phylogenetic community structure. *Mol. Ecol.* 18, 572–592

74. Cavender-Bares, J. *et al.* (2009) The merging of community ecology and phylogenetic biology. *Ecol. Lett.* 12, 693–715

75. Swenson, N.G. (2011) Phylogenetic beta diversity metrics, trait evolution and inferring the functional beta diversity of communities. *PLoS One* 6, e21264

76. Leprieur, F. *et al.* (2012) Quantifying phylogenetic beta diversity: distinguishing between 'true' turnover of lineages and phylogenetic diversity gradients. *PLoS One* 7, e42760

77. Koleff, P. *et al.* (2003) Measuring beta diversity for presence-absence data. *J. Anim. Ecol.* 72, 367–382

78. Sokal, R.R. and Rohlf, F.J. (1962) The comparison of dendrograms by objective methods. *Taxon* 11, 33–40

79. Gower, J.C. (1983) Comparing classifications. In *Numerical Taxonomy* (Felsenstein, J., ed.), pp. 137–155, Springer

80. Borcard, D. *et al.* (2011) *Numerical Ecology with R,* Springer

81. Salvador, S. and Chan, P. (2004) Determining the number of clusters/segments in hierarchical clustering/segmentation

algorithms. In *Proceedings of the Sixteenth IEEE International Conference on Tools with Artificial Intelligence*, pp. 576–584, Institute of Electrical and Electronics Engineers

82. Zhao, X. *et al*. (2011) Systematic clustering of transcription start site landscapes. *PLoS One* 6, e23409

83. Oksanen, J. *et al*. (2015) *Vegan: Community Ecology Package,* CRAN-R

84. Ficetola, G.F. *et al*. (2017) Global determinants of zoogeographical boundaries. *Nat. Ecol. Evol.* 1, 0089

85. Kaesler, R.L. (1970) The cophenetic correlation coefficient in paleoecology. *Geol. Soc. Am. Bull.* 81, 1261–1266

86. Minchin, P.R. (1987) An evaluation of relative robustness of techniques for ecological ordinations. *Vegetatio* 69, 89–107

87. 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

88. Mishler, B.D. *et al*. (2014) Phylogenetic measures of biodiversity and neo- and paleo-endemism in Australian Acacia. *Nat. Commun.* 5, 4473

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

90. 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