



## Research



**Cite this article:** Gross CP, Wright AM, Daru BH. 2025 A global biogeographic regionalization for butterflies. *Phil. Trans. R. Soc. B* **380**: 20230211. <https://doi.org/10.1098/rstb.2023.0211>

Received: 1 February 2024

Accepted: 17 October 2024

One contribution of 19 to a discussion meeting issue 'Bending the curve towards nature recovery: building on Georgina Mace's legacy for a biodiverse future'.

### Subject Areas:

ecology, ecosystem, evolution

### Keywords:

biogeographic regionalization, butterflies, phylogenetic beta diversity, lepidoptera

### Author for correspondence:

Barnabas H. Daru

e-mail: [bdaru@stanford.edu](mailto:bdaru@stanford.edu)

Electronic supplementary material is available online at <https://doi.org/10.6084/m9.figshare.c.7552614>.

# A global biogeographic regionalization for butterflies

Collin P. Gross<sup>1</sup>, April M. Wright<sup>2</sup> and Barnabas H. Daru<sup>1</sup>

<sup>1</sup>Department of Biology, Stanford University, Stanford, CA 94305, USA

<sup>2</sup>Department of Biological Sciences, Southeastern Louisiana University, Hammond, LA 70402, USA

CPG, 0000-0002-0896-8476; AMW, 0000-0003-4692-3225; BHD, 0000-0002-2115-0257

The partitioning of global biodiversity into biogeographic regions is critical for understanding the impacts of global-scale ecological and evolutionary processes on species assemblages as well as prioritizing areas for conservation. However, the lack of globally comprehensive data on species distributions precludes fine-scale estimation of biogeographical regionalization for numerous taxa of ecological, economic and conservation interest. Using a recently published phylogeny and novel curated native range maps for over 10 000 species of butterflies around the world, we delineated biogeographic regions for the world's butterflies using phylogenetic dissimilarity. We uncovered 19 distinct phylogenetically delimited regions (phyloregions) nested within 6 realms. Regional boundaries were predicted by spatial turnover in modern-day temperature and precipitation seasonality, but historical climate change also left a pronounced fingerprint on deeper- (realm-) level boundaries. We use a culturally and ecologically important group of insects to expand our understanding of how historical and contemporary factors drive the distribution of organismal lineages on the Earth. As insects and global biodiversity more generally face unprecedented challenges from anthropogenic factors, our research provides the groundwork for prioritizing regions and taxa for conservation, especially with the goal of preserving the legacies of our biosphere's evolutionary history.

This article is part of the discussion meeting issue 'Bending the curve towards nature recovery: building on Georgina Mace's legacy for a biodiverse future'.

## 1. Introduction

Biogeographic regions describe the spatial distribution of life on the Earth and form the basis of many studies in ecology, conservation and evolutionary biology [1–3]. Biogeographic regionalization involves dividing the Earth's surface into distinct regions based on shared ecological and evolutionary characteristics. These regions are characterized by unique assemblages of species, ecological communities and ecosystems. Understanding biogeographic regions is crucial across the fields of ecology, evolutionary biology and conservation because it allows the investigation of factors driving biodiversity patterns and ecosystem dynamics across ecological scales [4–6], and the possible changes in these factors in the past and into the future. Thus, delineating biogeographic regions can enable the identification of the underlying processes that shape species composition, community structure and ecosystem functioning [7] as well as enable the prediction of how contemporary factors such as habitat loss and climate change may alter these regions, and delineate areas and species for conservation priority [8–11].

Combining biogeographic regionalizations with phylogenetic estimation allows researchers to account for the role of evolutionary history of regional biotas in shaping species distributions. For instance, areas with extensive local

sympatric speciation or endemic radiations are likely to be more evolutionarily distinct from surrounding regions, potentially forming a separate phylogeographic cluster. On the other hand, in cases where speciation occurs in allopatry, especially due to geographical isolation, the separation of biogeographic clusters is less clear because sister species can occur on either side of a regional boundary [7]. Such knowledge can serve as a crucial step in addressing spatial evolution of biodiversity and identifying important units for conservation.

In terrestrial ecosystems, global biogeographic regionalizations have focused on vertebrates [1,12] or plants [13]. For instance, Holt and colleagues [1] updated Wallace's [3] zoogeographic regions using geographical distributions and phylogenetic relationships of the world's mammal, bird and amphibian species and divided global land areas into 20 regions nested into 11 larger realms, which are still widely used today [12,14,15]. However, these regionalizations rely heavily on globally comprehensive species geographical information, limiting the number and types of taxa that can be analysed. This information is not readily available for ecologically and economically invaluable invertebrate taxa, such as butterflies, which lack distribution data at a global scale [16–18]. Additionally, examining the patterns and drivers of regionalization schemes across disparate terrestrial taxa aids us in understanding the historical and modern-day processes that govern the broad-scale distribution of life on our planet.

Butterflies (Lepidoptera: Papilionoidea) are thought to have originated in the middle Cretaceous (102.5–100 Ma) in North or Central America [19,20]. Since then, they have diversified into approximately 19 000 extant species [21], play key ecosystem roles as pollinators (as adults) and herbivores (in their larval form) and are involved in a wide array of ecological and coevolutionary interactions ranging from myrmecophily to phytochemical sequestration [22–25]. Despite their ecological, cultural and evolutionary importance, no attempt has been made to incorporate phylogenetic structure into our understanding of butterfly regionalization, although recent work has illuminated the group's global phylogeographic history [19]. Comprehensive regionalization is hindered in part by a lack of comprehensive global distribution data at the species level. The country-level occurrence database compiled by Pinkert *et al.* [21] remained the only available dataset on global butterfly distributions until now [26]. Previous studies of butterflies have primarily focused on regional or continental assessments or have concentrated on specific clades [4,27–30], but these have largely overlooked their evolutionary history and how this history underlies modern distributions.

Here, using an unprecedented dataset including native range maps for 10 372 butterfly species, representing about 53% of global butterfly diversity along with their phylogenetic relationships, we provide the first global phylogenetic biogeographic regionalization (“phyloregions”) of butterfly assemblages. Specifically, we ask (i) what are the global phylogenetically delimited biogeographic regions for butterflies? (ii) what are the determinants of biogeographical boundaries for butterflies? (iii) how do biogeographic boundaries for butterflies compare with those observed in other terrestrial taxonomic groups? We hypothesized that butterfly phyloregions would be relatively congruent with those previously published for terrestrial plants and tetrapods [1,12,13]. In particular, because of butterflies' coevolutionary relationships with flowering plants [23–25], we expected butterfly and angiosperm phyloregions to be more congruent than those of butterflies and tetrapods. As terrestrial ectotherms, we also expected butterfly regionalization to largely be driven by global temperature patterns [31].

## 2. Methods

### (a) Source data

To begin our phylogenetic regionalization scheme for butterflies, we used a new set of native range polygons for butterflies created by Daru [26]. Briefly, these polygons are modelled distributions based on point data available from the Global Biodiversity Information Facility (GBIF [32]), clade-specific constraints on dispersal [33] and species distribution models using maximum entropy. To reduce the effects of uneven or biased sampling, we spatially thinned point records for species with at least five unique localities and ensured that no two points were within 30 km of one another to generate alpha hull polygons [34–36]. To facilitate scaling across a large number of species with very few occurrences, we converted alpha hull polygons into pseudo-occurrence records by first converting to a raster, converting raster grids to points and thinning points to a maximum of 500 records per species to reduce spatial autocorrelation [13,37,38]. Each species range map was stacked at the resolution of 100 km × 100 km grid cells to generate a presence and absence matrix of 10 372 species in 13 316 grid cells using the *polys2comm* function in the R package *phyloregion* [39].

### (b) Phylogeny

To examine how butterfly lineages are regionalized across the globe, we used a time-calibrated tree of 2254 species from Kawahara *et al.* [19] representing all families, subfamilies and tribes, and 92% of described genera. To complete the tree, we grafted the remaining species that were not in the original tree onto the backbone topology using the function *phylo.maker* in the package *V.Phylomaker2* under scenario 2 [40]. In this scenario, new tips are bound to randomly selected nodes at and below the genus-level (if congeners exist in the backbone;  $n = 8\,293$  species added in this way) or family-level (if congeners do not exist in the backbone;  $n = 401$  species added in this way) basal node. We created 100 random trees based on this method to build a consensus phylogenetic regionalization scheme for butterflies.

## (c) Data analysis

### (i) Building regionalization schemes

To quantify spatial turnover in butterfly assemblages, we first calculated species-level beta diversity using the matrix of Simpson pairwise dissimilarities across all 100 km × 100 km grid cells using the *beta\_diss* function in the R package *phyloregion* [39]. We then calculated phylogenetic beta diversity among grid cells using the Simpson-derived pairwise phylogenetic dissimilarity matrix using the function *phylobeta* for each of our 100 trees and used the average dissimilarity matrix for downstream analyses.

To map biogeographic regions, we used the function *optimal\_phyloregion* to determine the optimal number of biogeographic regions for both species-level and phylogenetic beta diversity matrices based on the elbow method, which calculates the optimal number of clusters based on explained variance and internal consistency of clusters [41]. We then determined the best hierarchical clustering algorithm to use in building our biogeographic regionalization schemes by using the function *select\_linkage*, which compares eight commonly used clustering algorithms based on their cophenetic correlation coefficient, a measure of how well a dendrogram preserves Simpson distances from the original distance matrix. For both our phyloregions and species bioregions, this coefficient was highest for the unweighted pair-group method with arithmetic averages (UPGMA) method. Next, we examined the deeper relationships between regions by cutting the dendrogram at a depth that preserved at least 60% of the variation in phylogenetic beta dissimilarity to produce a set of phylogenetically delimited biogeographic realms with the phyloregions nested in them. This cut-off threshold consistently produced realms that were similar in shape and number to those previously established for other taxa [1,12,13,42].

To better characterize each region in terms of species and phylogenetic diversity, we computed the total species richness and phylogenetic diversity (PD, the sum of tree branch lengths in a community phylogeny; [43]) for each region. We also calculated the mean regional weighted species (species richness, with each species inversely weighted by the number of grid cells it occupies; [44]) and phylogenetic endemism (phylogenetic diversity, with branch lengths inversely weighted by the range size of their descendent clades; [45]) by averaging endemism values across all grid cells within the region, allowing us to determine which regions had larger numbers of species with small ranges rather than the number of species endemic to a whole region. We also computed the mean value of phylogenetic beta diversity between a focal phyloregion and all other phyloregions in the study area as a measurement of the evolutionary distinctiveness of each phyloregion.

To confirm the robustness of our regionalization schemes, we repeated these analyses eight times: once using the UPGMA algorithm with the subset of butterfly species shared between our dataset and the original Kawahara *et al.* tree ([19];  $n = 1691$ ), again using the UPGMA algorithm with all of our butterflies at the genus level ( $n = 1600$ ) and once on the full species dataset, each with six additional clustering algorithms: (1) centroid linkage clustering (unweighted pair-group method with centroids; UPGMC), (2) median linkage clustering (weighted pair-group method with centroids; WPGMC), (3) weighted pair-group method with arithmetic averages (WPGMA), Ward clustering with (4) and without (5) Ward's clustering criterion and (6) maximum (complete) lineage clustering.

### (ii) Comparing butterfly phyloregions with other taxa

To assess the congruency between butterfly phyloregions and those of other terrestrial clades, we used the V-measure (a measure of how homogeneous and complete clusters are) approach [46] on six previously published phylogenetic regionalization schemes for plants [13], reptiles [12], amphibians, birds, mammals, and combined amphibians, birds and mammals at the region level [1]. This approach compares two independent regionalization schemes and calculates a score (V-measure) between 0 (no association between compared regionalizations) and 1 (perfect association). For example, the V-measure score between butterflies and another taxon would decrease if single butterfly regions overlapped with multiple separate regions of the other taxon and *vice versa*. To test whether V-measures were significantly different from a random expectation, we compared the observed measures with a random distribution of measures created from 999 random regionalization schemes with the same number of regions as the scheme of comparison. We created random schemes by first randomly sampling a number of terrestrial cells corresponding to the number of random bioregions needed, then creating Voronoi polygons from those points using the *voronoi* function in the R package *terra* [47] and finally cutting the Voronoi polygons using the outline of global landmasses. We used the same approach to quantify the congruence between our phyloregions and species bioregions, genus-level phyloregions, phyloregions derived from the Kawahara *et al.* [19] subset and phyloregions and realms defined by the six additional clustering algorithms.

To assess the pairwise correlations of boundaries across different phylogenetic regionalization schemes, we first calculated the distance of each 100 km × 100 km grid cell from a boundary in each scheme. We then used a modified version of a *t*-test that is appropriate to test for correlations among spatial processes [48] to test whether the distance of each cell from a boundary is similar among the schemes. If two phylogenetic regionalization schemes are similar, boundaries cross the same areas of the planet, and any given cell shows a similar distance from the boundary in both schemes. Prior to testing, we log-transformed distance values in km.

### (iii) Examining predictors of phyloregion boundaries

We examined the predictors of boundaries between butterfly phyloregions using a simultaneous autoregressive (SAR) spatial modelling approach, which accounts for spatial autocorrelation when modelling the relationships between boundaries and

predictors [49,50]. Four predictor variables (derived from the WorldClim database; [51]) represented spatial climate heterogeneity, defined as the coefficient of variation between each cell and eight adjacent cells: temperature seasonality, precipitation seasonality, mean annual temperature and mean annual precipitation. Two additional variables included climate change velocity since the last glacial maximum (LGM) and terrain ruggedness index (TRI; mean absolute difference in elevation between each cell and eight adjacent cells). We centred all predictor variables on a mean of zero and scaled to a standard deviation of one. All predictors showed variance inflation factors of less than two, indicating minimal collinearity.

For our SAR modelling approach, we built a hierarchical generalized linear model (HGLM), in which the response variable was a binomially distributed binary variable determining whether a given cell is in contact with a phyloregion boundary or not. This model accounted for spatial autocorrelation among cells' predictor values using neighbourhood matrices in the error term. We used a neighbourhood size of 282.84 km, which was the maximum distance required for all cells to be connected to at least one other, and limited our dataset to only the cells within 200 km of either side of a boundary to avoid excessively zero-inflated data. We repeated this analysis three times: once for all boundaries, once for only the deep boundaries between realms and once for only the shallow boundaries within realms. We report the effect size of each variable as Fisher's  $z$ , a measure independent of sample size that enables robust comparisons of effects estimated across different models.

#### (iv) Identifying clades driving realm-level divisions

Finally, we were interested in the butterfly clades that were key drivers of faunal turnover across realm boundaries. To identify such 'indicator clades', we used a node-based approach [52]. First, we identified nodes in the global butterfly phylogeny whose daughter lineages had minimal overlap in their geographical distribution, defined as a Geographic Node Divergence score (GND) greater than 0.65. For each of these nodes, we identified the grid cells in which daughter lineages occurred and calculated a specific overrepresentation score (SOS) for each cell, with more positive scores indicating overrepresentation of one daughter lineage, more negative scores indicating overrepresentation of the other daughter lineage and scores closer to zero indicating equal representation of both lineages. We then modelled SOS values as a function of cell realm membership for each of 15 realm pairs to identify which of the high-GND clades contributed to differences in the butterfly faunas of paired realms, focusing only on comparisons between positive and negative mean SOS values. The contribution of clades to realm division was measured by the  $R^2$  value of each model.

### 3. Results

#### (a) Biogeographic regions for butterflies

Using the 'elbow' method, we identified 19 clusters based on shared phylogenetic affinity (phyloregions) and 13 regions based on species beta diversity (figure 1a, electronic supplementary material, figure S1) that explained 80% and 68% of the variation in beta diversity, respectively. The boundaries between species and phylogenetic regionalization schemes were markedly similar ( $V$ -measure = 0.84,  $p < 0.001$ ). Hierarchical clustering and nonmetric multidimensional scaling (NMDS) analyses revealed that deep divisions between six clusters (realms) accounted for 65% of the variation in phylogenetic beta diversity across grid cells (figure 1b): Holarctic (North American, Eurasian, Tibetan, Japanese), Indo-Malayan (Hengduan-Himalayan, Chinese, Indian, Novozelandic), African (Afrotropical, Maghrebi, Madagascan), Nusantara (Malesian, Papuasian), Australian and Neotropical (Mesoamerican, Caribbean, Amazonian, Pampeo-Andean, Valdivian).

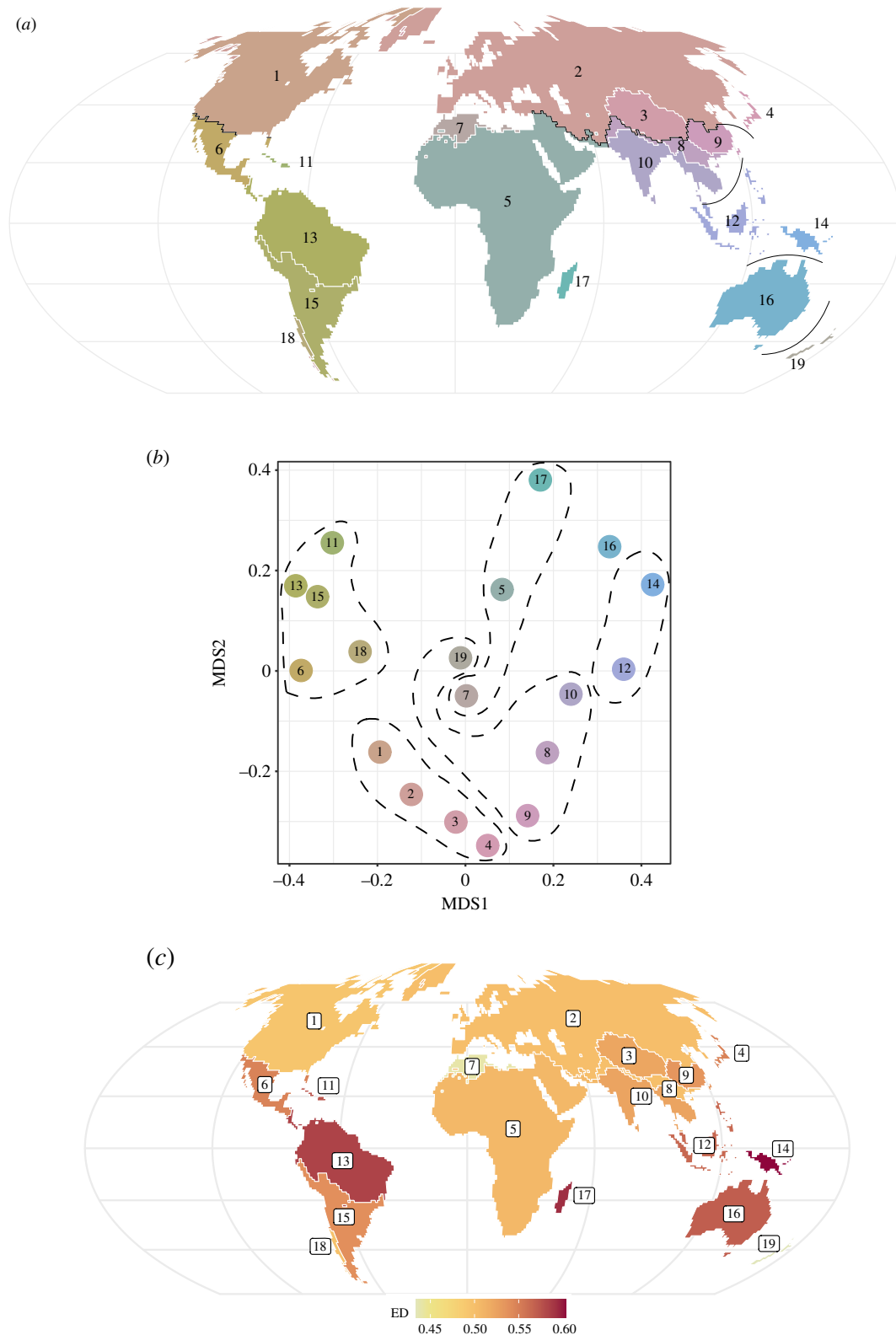
Given that 83% of the species in our dataset were grafted onto the backbone phylogeny, it is worth noting how the species in the original Kawahara *et al.* [19] tree are distributed among the regions. We found that although no region had more than 50% of its species present in the backbone, the Japanese region was most represented—with 49% of its species in the backbone—followed by the Caribbean and Valdivian regions, for which 37% and 34% of species were represented, respectively (electronic supplementary material, figure S2). In contrast, the Novozelandic, Afrotropical, Madagascan and Papuasian faunas were most underrepresented, with 13%, 13%, 15% and 16% of species in the original backbone, respectively (electronic supplementary material, figure S2). This is particularly striking in the case of the Afrotropical realm, where just 290 of its 2301 species were present in the original tree.

We found that the Papuasian, Madagascan and Amazonian phyloregions were the most evolutionarily distinct, each with mean phylogenetic beta diversity of 0.60, 0.59 and 0.58, respectively (figure 1c). The most species-rich and phylogenetically diverse regions were the Amazonian, Pampeo-Andean and Afrotropical (figure 2a,b), and the regions with the highest average endemism were the Caribbean, Papuasian and Malesian (figure 2c,d).

Phylogenetic regionalization using only the species in our dataset that are found in the Kawahara *et al.* [15] tree produced 13 regions that together explained 81% of the variation in phylogenetic beta diversity across cells and 5 realms that explained 64.3% of the variation (electronic supplementary material, figure S3). Our full phylogenetic regionalization scheme was robustly supported by this limited dataset, with high congruence between the schemes ( $V$ -measure = 0.78,  $p < 0.001$ ). The genus-level scheme produced 12 regions that explained 81% of the variation in phylogenetic beta diversity and with a lower  $V$ -measure score (0.75,  $p < 0.001$ ; electronic supplementary material, figure S4).

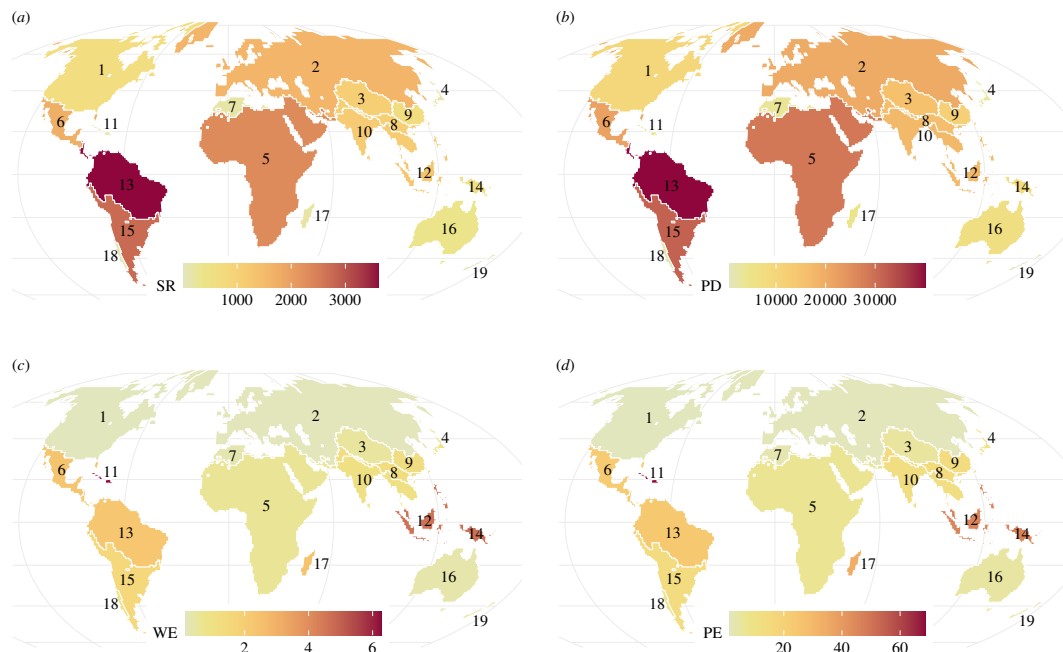
Across the clustering algorithms we used to create phyloregions, the Maghrebi-Afrotropical boundary and the northern border of the Hengduan-Himalayan region were consistent features, and distinct Madagascan and Papuasian regions were features of 6/7 algorithms (electronic supplementary material, figure S3 and S5). The Tibetan region appeared across 5/7 algorithms, and the Andean-Pampeo-Amazonian boundary appeared in 4/7 algorithms. Some form of boundary was generated





**Figure 1.** A global phylogenetic regionalization of butterfly species ( $n = 10\,372$  species). (a) Map showing the 19 major butterfly regions around the world, defined by phylogenetic similarity. White lines represent divisions between regions (80% of variation in phylogenetic dissimilarity explained), while black lines represent deeper divisions between realms (65% of variation explained). (b) Non-metric multidimensional scaling (NMDS) plot showing relationships between regions (numbered circles) and realms (dotted outlines). (c) Regions coloured by evolutionary distinctiveness (ED). A region is darker if the mean value of phylogenetic beta diversity between it and all other regions is greater. Numbered regions: 1) North American, 2) Eurasian, 3) Tibetan, 4) Japanese, 5) Afrotropical, 6) Mesoamerican, 7) Maghrebi, 8) Hengduan–Himalayan, 9) Chinese, 10) Indian, 11) Caribbean, 12) Malesian, 13) Amazonian, 14) Papuanian, 15) Pampeo-Andean, 16) Australian, 17) Madagascan, 18) Valdivian and 19) Novozelandic. Maps are Mollweide projections.

by 4/7, 4/7 and 5/7 algorithms between eastern and western North America, between the wet and dry tropics of southern Africa, and between northern and southern Australia, respectively. V-measure scores between the 7 clustering algorithms were greater than 0.6 for both regions and realms (electronic supplementary material, tables S1 and S2), and spatial correlation coefficients were high (greater than 0.8) among clustering algorithms for borders between regions and realms (electronic supplementary material, tables S3 and S4), indicating strong support overall for our regionalization scheme.



**Figure 2.** Diversity and endemism in the 19 major butterfly regions around the world. Species richness (a) and phylogenetic diversity (Faith's PD; (b) are calculated across all species in each region. Weighted species (c) and phylogenetic endemism (d) are averaged across 100 km  $\times$  100 km grid cells within a region.

## (b) Similarity of phyloregions across taxonomic groups

We contrasted the boundaries of our butterfly phyloregions with those of other terrestrial taxonomic groups using the V-measure method [46], which examines the intersections among regions of butterflies and other taxa. We found high similarity between butterfly phyloregions and those of other taxonomic groups, with all pairwise comparisons (V-measure scores) significantly higher than 0.7 (table 1). However, butterfly phyloregions showed the highest congruence with tetrapod phyloregions (V-measure = 0.77;  $p < 0.001$ ). Within tetrapods, regions defined by reptile clades were most similar to those defined by butterflies (V-measure = 0.76,  $p < 0.001$ ). Plant-defined phyloregions were least congruent (V-measure = 0.71,  $p = 0.002$ ), largely driven by differences in the temperate Southern Hemisphere, the Mediterranean and Central Asia (electronic supplementary material, figure S6).

When we compared the relative positions of butterfly phyloregion boundaries with those of other taxa, we again found significant correlations between the positions of these boundaries (table 1), but the strongest correlation was between butterfly and plant regional boundaries (correlation = 0.66,  $p < 0.001$ ), while butterfly and reptile regional boundaries showed the weakest correlation (correlation = 0.45,  $p < 0.001$ ).

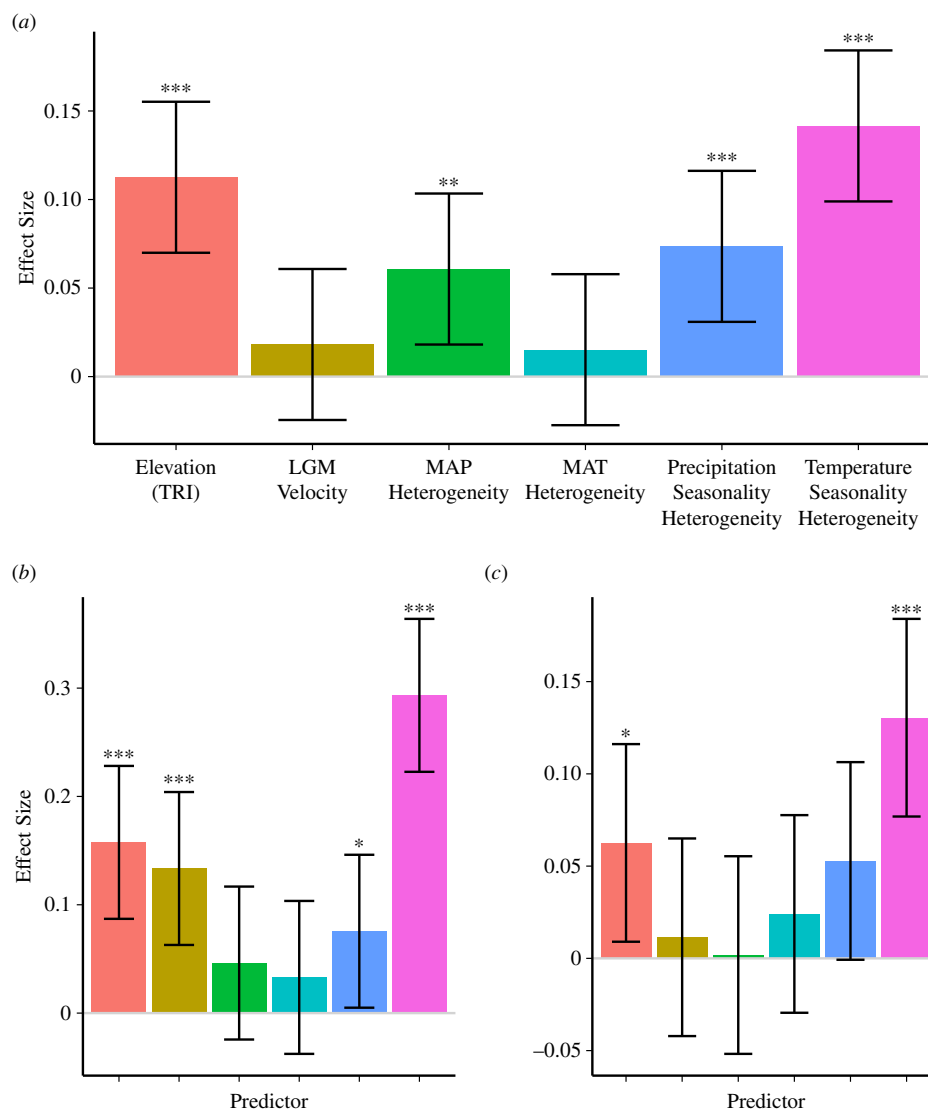
## (c) Determinants of butterfly biogeographic boundaries

Our spatially explicit HGLM assessing the factors related to the presence of all biogeographic boundaries for butterflies showed that boundaries were more likely to occur within regions of high TRI (mean absolute difference in elevation between each cell and eight adjacent cells) and turnover in temperature seasonality and precipitation (figure 3a). The strongest of these effects was turnover in temperature seasonality ( $z = 0.14$ ,  $p < 0.001$ ). Climate change velocity since the LGM and mean annual temperature heterogeneity were not significant predictors across all boundaries.

Closer inspection of deep (between-realm) boundaries revealed that TRI, turnover in temperature seasonality and climate change velocity since the LGM were all strongly predictive of boundaries (figure 3b). Analysis of only shallow (within-realm) boundaries again showed that TRI and turnover in temperature seasonality were the two strongest predictors of boundaries (figure 3c).

## (d) Clades driving realm-level divisions

Clades with high contribution to realm divisions ('indicator clades') varied among realms but appeared across the butterfly phylogeny (figure 4). The median age of clades with the greatest contribution to realm-level division was 34.6 Ma, but these nodes ranged from the earliest division between Papilionidae and all other butterflies (100 Ma) to divisions within Hesperiid and Nymphalid tribes (6.9–11.6 Ma; electronic supplementary material, figure S7). The Australian and Nusantara realms were delineated from all other realms by the oldest nodes, while the Holarctic and African realms were delineated by the youngest nodes (electronic supplementary material, figure S7). The basal node in our tree was especially important for distinguishing the Nusantara and Indo-Malayan realms from all other realms except the Holarctic (average  $R^2 = 0.605$ ). Within families, the greatest contributions to divisions were by nodes within Papilionidae, especially between the Nusantara and other realms (average  $R^2 = 0.776$ ). The Nusantara realm was also distinguished from the Neotropical by clades within Riodinidae

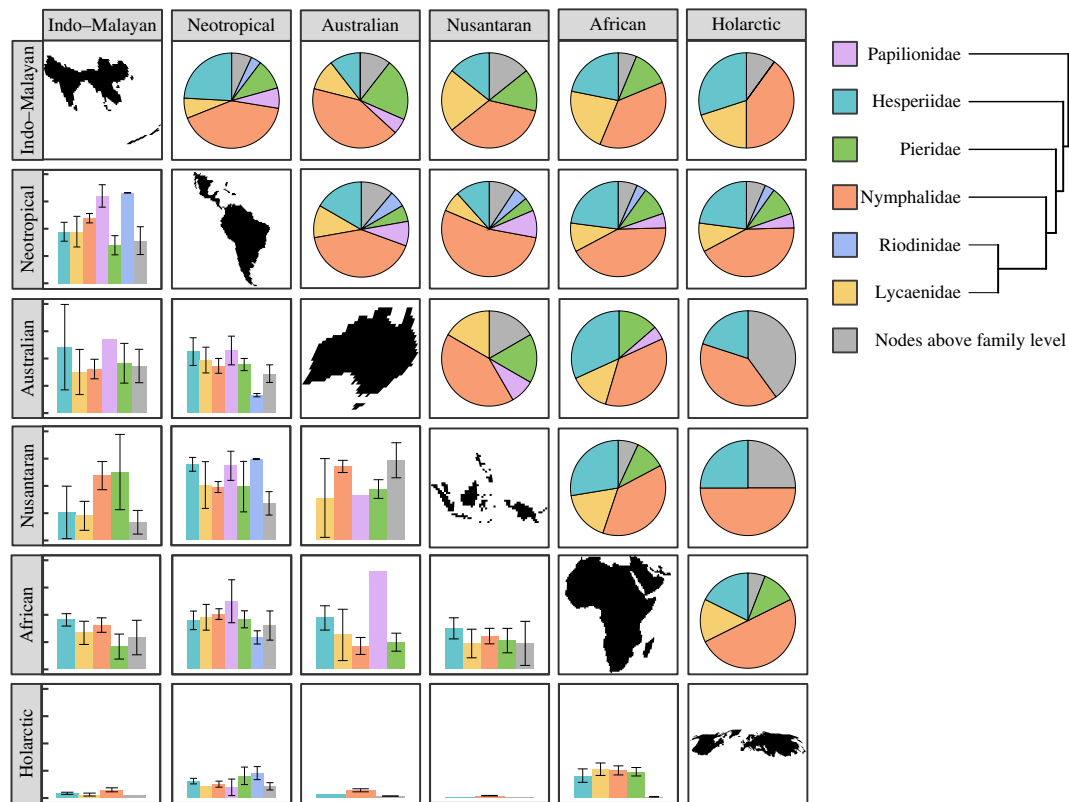


**Figure 3.** Effect sizes from spatially explicit hierarchical generalized linear models of whether a 100 km  $\times$  100 km grid cell was in contact with (a) any border between butterfly phyloregions, (b) deep boundaries between realms or (c) shallow regional boundaries within realms. Terrain ruggedness index (TRI) represents the mean absolute difference in elevation between each cell and eight adjacent cells; climate heterogeneity represents the coefficient of variation between each cell and eight adjacent cells. Bars indicate effect sizes (Fisher's  $z$ ); error bars represent 95% confidence intervals of  $z$ . Asterisks represent effect sizes significantly different from zero;  $0 < p < 0.001$ : \*\*\*;  $0.001 < p < 0.01$ : \*\*;  $0.01 < p < 0.05$ : \*.

**Table 1.** Comparing regions defined by other clades with those defined by butterflies using V-measure scores and correlation coefficients from spatially explicit  $t$ -tests. Higher V-measure scores indicate higher congruency between regional polygons for butterflies and the focal taxon; higher correlation coefficients indicate that the boundaries of regions defined by the two taxa are closer in space.

taxon	V-measure score	V-measure $p$ -value	spatial correlation	spatial correlation $p$ -value
all tetrapods	0.770	0	0.648	$4.91 \times 10^{-42}$
reptiles	0.763	0	0.453	$6.65 \times 10^{-23}$
birds	0.751	0	0.644	$1.16 \times 10^{-40}$
amphibians	0.741	0	0.526	$8.65 \times 10^{-34}$
mammals	0.736	0	0.641	$8.78 \times 10^{-44}$
plants	0.711	0.002	0.665	$2.12 \times 10^{-33}$

(Nemeobiini versus Euselasiini,  $R^2 = 0.745$ ; Nemeobiinae versus Riodininae,  $R^2 = 0.736$ ). The clade subtending the Riodinidae and Lycaenidae additionally had a relatively high contribution to the distinctiveness of the Neotropical realm (average  $R^2 = 0.414$ ).



**Figure 4.** Contributions of 83 potential 'indicator' butterfly clades to divisions between 6 biogeographic realms, totaling 411 pairwise comparisons. Colours represent nodes within families (including between and within subfamilies, tribes and genera) or older nodes above the level of family. Above the diagonal, each pie chart represents the relative contribution of each type of node to distinguishing between the pair of realms. Below the diagonal, bars indicate the mean  $R^2$  value in a model of specific overrepresentation score (SOS) response to realm identity across all the nodes within a group for each realm pair (electronic supplementary material, table S5). Y-axis tick marks correspond to 0.0, 0.25, 0.50, 0.75 and 1.00. Hedyliidae, sister to Hesperidae, is not shown because neither nodes within the family nor the node subtending Hesperidae and Hedyliidae contributed to divisions between realms.

## 4. Discussion

Overall, butterfly phyloregions shared many similarities with those published for other terrestrial taxa [1,12,13], including distinct North American, Eurasian, Australian, Madagascan, Novozelandic and Afrotropical regions. However, we found some notable differences. For instance, we found a separate realm including the islands of maritime Southeast Asia and New Guinea (the Nusantara realm) that was distinct from the rest of the Sunda Shelf and Australia, a distinction not present in any published phylogenetic regionalization scheme for other taxa. While our Nusantara realm was particularly distinct for its clades of swallowtail butterflies (Papilionidae, figure 4), Southeast Asia as a whole is believed to be the geographical origin of much of the world's butterfly diversity, particularly brush-footed butterflies (Nymphalidae; [19,53,54]), and the high concentration of islands and history of sea level change in this region likely contribute to the evolutionary distinctiveness and high endemism of the region ([55–57]; figures 1 and 2). These patterns may also help to specify the centre of diversity in the region believed to be the source of diversity in surrounding areas; other studies have attempted to understand the geographical origins and dispersal of butterfly biodiversity [19,53] but used an *a priori* Wallacean [3] regionalization scheme to constrain their analyses. Our phylogenetic regionalization scheme can thus be further used to understand the geographical sources and sinks of butterfly diversity worldwide.

Like other terrestrial regionalization schemes, we uncovered distinct Novozelandic and Saharan (Maghrebi) phyloregions, although these were among the least stable across algorithms and datasets (electronic supplementary material, figures S1, S5 and S4). The native butterflies of New Zealand are unusual among other taxa from the archipelago in that they do not represent recent endemic radiations, paleoendemic relics or close affinities with other Oceanian biotas [58–61]; of the only 16 Novozelandic species in our dataset, 11 were endemic and 4 (*Argyrophenga*, *Dodonidia*, *Erebolia* and *Percnodaimon*) represent endemic genera. The low species richness and low phylogenetic endemism in our dataset likely contributed to the relatively low distinction of the Novozelandic region, which grouped within the Indo-Malayan and Holarctic realms across datasets and algorithms. Also of note here is that only 2 of the 16 Novozelandic species in our full list were included in the original backbone phylogeny [19], making this area of the world particularly subject to variation among the random trees we generated (electronic supplementary material, figure S2). The Maghrebi region—which in our dataset contains 177 species, of which 10 are endemic—jumps affinities between Afrotropical, Eurasian and Turanian regions, likely owing to a high proportion of shared lineages between these regions.

Butterfly regions had significantly high congruence with all the regions defined by other taxa that we investigated (table 1). Butterflies and plants shared the lowest V-measure scores, which was surprising given the coevolutionary relationships between butterflies and larval host plants [23–25,62], but this V-measure was only 0.059 less than the highest V-measure (butterflies



and tetrapods). These values may all be relatively high because we have defined regions at the global scale at a relatively low phylogenetic beta diversity cut-off, which glosses over finer distinctions between biotas at sub-regional scales. Although there are noticeable differences in regionalization schemes among taxa, at the level at which we have defined regions, larger global-scale processes are the primary architects of discrete regions, which tend to be common across taxa [12,14]. Were we to focus on sub-regional clusters, we anticipate much higher variation in pairwise V-measures between butterflies and other taxa, as the processes occurring at these smaller spatial scales are more idiosyncratic and more difficult to generalize across taxonomic groups [12,63].

Spatial correlations between boundaries seem to be a better descriptor of similarity between regionalizations than V-measures in our analyses, showing that butterfly and plant boundaries were most highly correlated (table 1). This was unsurprising given the coevolutionary relationships between butterflies and their larval host plants [23–25,62], even though codiversification and constraints on host-switching within particular butterfly taxa tend to occur for nodes in the host and butterfly phylogenies that are deeper than those used to build these regionalizations (species; for example, subfamilies within the butterfly Pieridae feed on multiple plant orders, but host use is more conserved within subfamilies; [62,64]). Thus, while the codivergence of butterflies and their hostplants likely play a role in the correlation between their regional boundaries, shared abiotic filters such as climate and topography likely play an equivalent role in shaping how lineages in these taxa are distributed around the globe [42].

We found that the deep terrestrial boundaries between butterfly realms were consistent with those observed for terrestrial tetrapod taxa [1,12], including between the Neotropical and Holarctic realms and between the Holarctic and African and Indo-Malayan realms (figure 1), pointing to similar processes driving the divisions between realms across taxa. Rather than turnover in mean annual temperature or precipitation, butterfly realms appeared to be formed primarily by turnover in seasonality; that is, boundaries between realms occurred when regional climates shifted from relatively consistent year-round to those with more pronounced seasonal variation (figure 3). We additionally saw a strong significant effect of past climate change in driving realm boundaries. Greater historical and modern climate stability may reduce extinction rates among older lineages while simultaneously increasing speciation rates [65,66], potentially increasing the phylogenetic distinctiveness of realms and regions. On the other hand, greater seasonal variation may increase extinction rates and facilitate range expansions through adaptive migration [67,68], reducing phylogenetic beta diversity within regions. Additionally, species whose ranges abut or cross regional boundaries may have reduced adaptive capacity to overcome differences in seasonality regimes with strategies such as diapause or changes in voltinism [69–71].

Here we have presented one of the first known global biogeographic regionalization schemes for a terrestrial invertebrate taxon, including the phylogenetic and external abiotic drivers behind regional boundaries. Our results expand our understanding of how historical and contemporary factors drive the distribution of organismal lineages on the planet and provide the groundwork for future efforts to model historical dispersal, diversification and niche conservatism across taxa worldwide. Additionally, recent studies have revealed significant declines in insect abundance and diversity, but these declines are not evenly distributed across the globe or the insect tree of life [72,73]. Our results can contribute to butterfly conservation and biodiversity assessment in three key ways. First, highlighting particular global phyloregions as areas of high endemism, and of phylogenetic distinctiveness (e.g. Madagascar, the Caribbean and the Nusantara regions) can help prioritize regions for conservation efforts [8,74]. Second, identifying ‘indicator clades’ can be useful to focus conservation efforts that preserve distinctions between regions. Finally, identifying the species and environmental factors that drive regional and realm-level divisions can provide a framework for understanding shifts in biotic regionalization and biodiversity homogenization in a changing world [9,75]. Given their extremely high diversity and wide array of complicated relationships with other organisms, studies of global insect biogeography and regionalization are much needed. We hope that studying the regionalization of a charismatic group like butterflies is an important step towards understanding how future climate and land use change will affect the processes driving global insect biogeography.

**Ethics.** This work did not require ethical approval from a human subject or animal welfare committee.

**Data accessibility.** We used a new set of native range polygons for butterflies created by Daru [26]. Data are available at Dryad [76]. Supplementary material is available online [77].

**Declaration of AI use.** We have not used AI-assisted technologies in creating this article.

**Authors' contributions.** C.P.G.: formal analysis, methodology, writing—original draft, writing—review and editing; A.M.W.: funding acquisition, writing—review and editing; B.H.D.: conceptualization, data curation, funding acquisition, project administration, software, supervision, validation, writing—original draft, writing—review and editing.

All authors gave final approval for publication and agreed to be held accountable for the work performed therein.

**Conflict of interest declaration.** We declare we have no competing interests.

**Funding.** B.H.D. was supported by the US National Science Foundation (awards 2345994 and 2416314).

## References

- Holt BG *et al.* 2013 An update of Wallace's zoogeographic regions of the world. *Science* **339**, 74–78. (doi:10.1126/science.1228282)
- Whittaker RJ, Riddle BR, Hawkins BA, Ladle RJ. 2013 The geographical distribution of life and the problem of regionalization: 100 years after Alfred Russel Wallace. *J. Biogeogr.* **40**, 2209–2214. (doi:10.1111/jbi.12235)
- Wallace AR. 1876 *The geographical distribution of animals*. Cambridge, UK: Cambridge University Press.

4. Andrew ME, Wulder MA, Coops NC. 2011 How do butterflies define ecosystems? A comparison of ecological regionalization schemes. *Biol. Conserv.* **144**, 1409–1418. (doi:10.1016/j.biocon.2011.01.010)
5. Elliott TL, Davies TJ. 2019 A comparison of phylogenetic and species beta diversity measures describing vegetation assemblages along an elevation gradient. *J. Veg. Sci.* **30**, 98–107. (doi:10.1111/jvs.12700)
6. Calatayud J, Hortal J, Medina NG, Turin H, Bernard R, Casale A, Ortuño VM, Penev L, Rodríguez MÁ. 2016 Glaciations, deciduous forests, water availability and current geographical patterns in the diversity of European *Carabus* species. *J. Biogeogr.* **43**, 2343–2353. (doi:10.1111/jbi.12811)
7. Daru BH, Elliott TL, Park DS, Davies TJ. 2017 Understanding the processes underpinning patterns of phylogenetic regionalization. *Trends Ecol. Evol.* **32**, 845–860. (doi:10.1016/j.tree.2017.08.013)
8. Daru BH, van der Bank M, Davies TJ. 2018 Unravelling the evolutionary origins of biogeographic assemblages. *Div. Distrib.* **24**, 313–324. (doi:10.1111/ddi.12679)
9. Daru BH, Rock BM. 2023 Reorganization of seagrass communities in a changing climate. *Nat. Plants* **9**, 1034–1043. (doi:10.1038/s41477-023-01445-6)
10. Chan KO, Grismer LL. 2021 Integrating spatial, phylogenetic, and threat assessment data from frogs and lizards to identify areas for conservation priorities in Peninsular Malaysia. *Glob. Ecol. Conserv.* **28**, e01650. (doi:10.1016/j.gecco.2021.e01650)
11. Huang M, Liu H, Tong Y, Li S, Hou Z. 2024 Diversity of endemic cold-water amphipods threatened by climate warming in northwestern China. *Divers. Distrib.* **30**, e13798. (doi:10.1111/ddi.13798)
12. Falaschi M, Marta S, Lo Parrino E, Roll U, Meiri S, Ficetola GF. 2023 Global bioregions of reptiles confirm the consistency of bioregionalization processes across vertebrate clades. *Glob. Ecol. Biogeogr.* **32**, 1272–1284. (doi:10.1111/geb.13694)
13. Carta A, Peruzzi L, Ramírez-Barahona S. 2022 A global phylogenetic regionalization of vascular plants reveals a deep split between Gondwanan and Laurasian biotas. *New Phytol.* **233**, 1494–1504. (doi:10.1111/nph.17844)
14. Ficetola GF, Mazel F, Falaschi M, Marta S, Thuiller W. 2021 Determinants of zoogeographical boundaries differ between vertebrate groups. *Glob. Ecol. Biogeogr.* **30**, 1796–1809. (doi:10.1111/geb.13345)
15. Weil SS, Gallien L, Nicolai MPJ, Lavergne S, Börger L, Allen WL. 2023 Body size and life history shape the historical biogeography of tetrapods. *Nat. Ecol. Evol.* **7**, 1467–1479. (doi:10.1038/s41559-023-02150-5)
16. Troudet J, Grandcolas P, Blin A, Vignes-Lebbe R, Legendre F. 2017 Taxonomic bias in biodiversity data and societal preferences. *Sci. Rep.* **7**, 9132. (doi:10.1038/s41598-017-09084-6)
17. Daru BH, Rodriguez J. 2023 Mass production of unvouchered records fails to represent global biodiversity patterns. *Nat. Ecol. Evol.* **7**, 816–831. (doi:10.1038/s41559-023-02047-3)
18. Stork NE. 2018 How many species of insects and other terrestrial arthropods are there on Earth? *Annu. Rev. Entomol.* **63**, 31–45. (doi:10.1146/annurev-ento-020117-043348)
19. Kawahara AY *et al.* 2023 A global phylogeny of butterflies reveals their evolutionary history, ancestral hosts and biogeographic origins. *Nat. Ecol. Evol.* **7**, 903–913. (doi:10.1038/s41559-023-02041-9)
20. Wahlberg N, Wheat CW, Peña C. 2013 Timing and patterns in the taxonomic diversification of Lepidoptera (butterflies and moths). *PLoS One* **8**, e80875. (doi:10.1371/journal.pone.0080875)
21. Pinkert S, Barve V, Guralnick R, Jetz W. 2022 Global geographical and latitudinal variation in butterfly species richness captured through a comprehensive country-level occurrence database. *Glob. Ecol. Biogeogr.* **31**, 830–839. (doi:10.1111/geb.13475)
22. Pierce NE, Braby MF, Heath A, Lohman DJ, Mathew J, Rand DB, Travassos MA. 2002 The ecology and evolution of ant association in the Lycaenidae (Lepidoptera). *Annu. Rev. Entomol.* **47**, 733–771. (doi:10.1146/annurev.ento.47.091201.145257)
23. Nishida R. 2002 Sequestration of defensive substances from plants by Lepidoptera. *Annu. Rev. Entomol.* **47**, 57–92. (doi:10.1146/annurev.ento.47.091201.145121)
24. van der Linden CFH, WallisDeVries MF, Simon S. 2021 Great chemistry between us: the link between plant chemical defenses and butterfly evolution. *Ecol. Evol.* **11**, 8595–8613. (doi:10.1002/ecs3.7673)
25. Janz N, Nylin S. 1998 Butterflies and plants: a phylogenetic study. *Evolution* **52**, 486–502. (doi:10.1111/j.1558-5646.1998.tb01648.x)
26. Daru BH. 2024 A global database of butterfly species native distribution. *Ecology* **e4462**. (doi:10.1002/ecy.4462)
27. Rueda M, Rodríguez MÁ, Hawkins BA. 2010 Towards a biogeographic regionalization of the European biota. *J. Biogeogr.* **37**, 2067–2076. (doi:10.1111/j.1365-2699.2010.02388.x)
28. Dapporto L, Fattorini S, Vodà R, Dincă V, Vila R. 2014 Biogeography of western Mediterranean butterflies: combining turnover and nestedness components of faunal dissimilarity. *J. Biogeogr.* **41**, 1639–1650. (doi:10.1111/jbi.12315)
29. Romo H, García-Barros E. 2010 Biogeographic regions of the Iberian Peninsula: butterflies as biogeographical indicators. *J. Zool.* **282**, 180–190. (doi:10.1111/j.1469-7998.2010.00730.x)
30. Dapporto L, Giolli G, Dennis RLH, Fox R, Shreeve TG. 2015 A new procedure for extrapolating turnover regionalization at mid-small spatial scales, tested on British butterflies. *Methods Ecol. Evol.* **6**, 1287–1297. (doi:10.1111/2041-210X.12415)
31. Sunday JM, Bates AE, Dulvy NK. 2011 Global analysis of thermal tolerance and latitude in ectotherms. *Proc. R. Soc. B* **278**, 1823–1830. (doi:10.1098/rspb.2010.1295)
32. GBIF. 2023 GBIF: Global Biogeographic Information Facility. See <https://www.gbif.org/>.
33. Louca S. 2021 Phylogeographic estimation and simulation of global diffusive dispersal. *Syst. Biol.* **70**, 340–359. (doi:10.1093/sysbio/syaa061)
34. Edelsbrunner H, Kirkpatrick D, Seidel R. 1983 On the shape of a set of points in the plane. *IEEE Trans. Inf. Theory* **29**, 551–559. (doi:10.1109/TIT.1983.1056714)
35. Burgman MA, Fox JC. 2003 Bias in species range estimates from minimum convex polygons: implications for conservation and options for improved planning. *Anim. Conserv.* **6**, 19–28. (doi:10.1017/S1367943003003044)
36. Meyer L, Diniz-Filho JAF, Lohmann LG. 2017 A comparison of hull methods for estimating species ranges and richness maps. *Plant Ecol. Divers.* **10**, 389–401. (doi:10.1080/17550874.2018.1425505)
37. Fourcade Y. 2016 Comparing species distributions modelled from occurrence data and from expert-based range maps. Implication for predicting range shifts with climate change. *Ecol. Inform.* **36**, 8–14. (doi:10.1016/j.ecoinf.2016.09.002)
38. Roll U *et al.* 2017 The global distribution of tetrapods reveals a need for targeted reptile conservation. *Nat. Ecol. Evol.* **1**, 1677–1682. (doi:10.1038/s41559-017-0332-2)
39. Daru BH, Karunarathne P, Schliep K. 2020 phyloregion: R package for biogeographical regionalization and macroecology. *Methods Ecol. Evol.* **11**, 1483–1491. (doi:10.1111/2041-210X.13478)
40. Jin Y, Qian H. 2022 VPhyloMaker2: an updated and enlarged R package that can generate very large phylogenies for vascular plants. *Plant Divers.* **44**, 335–339. (doi:10.1016/j.pld.2022.05.005)
41. Salvador S, Chan P. 2004 Determining the number of clusters/segments in hierarchical clustering/segmentation algorithms. In *16th IEEE International Conference on Tools with Artificial Intelligence, Boca Raton, FL, 15–17 November 2004*, pp. 576–584. (doi:10.1109/ICTAI.2004.50)
42. Liu Y *et al.* 2023 An updated floristic map of the world. *Nat. Commun.* **14**, 2990. (doi:10.1038/s41467-023-38375-y)

43. Faith DP. 1992 Conservation evaluation and phylogenetic diversity. *Biol. Conserv.* **61**, 1–10. (doi:10.1016/0006-3207(92)91201-3)
44. Crisp MD, Laffan S, Linder HP, Monro A. 2001 Endemism in the Australian flora. *J. Biogeogr.* **28**, 183–198. (doi:10.1046/j.1365-2699.2001.00524.x)
45. Rosauer D, Laffan SW, Crisp MD, Donnellan SC, Cook LG. 2009 Phylogenetic endemism: a new approach for identifying geographical concentrations of evolutionary history. *Mol. Ecol.* **18**, 4061–4072. (doi:10.1111/j.1365-294X.2009.04311.x)
46. Nowosad J, Stepinski TF. 2018 Spatial association between regionalizations using the information-theoretical V-measure. *Int. J. Geogr. Inf. Sci.* **32**, 2386–2401. (doi:10.1080/13658816.2018.1511794)
47. Hijmans RJ. 2024 Terra: spatial data analysis. *R package v. 1.7-82*. See <https://rspatial.org/>.
48. Vallejos R, Osorio F, Bevilacqua M. 2020 *Spatial relationships between two georeferenced variables with applications in R*. Cham, Switzerland: Springer Nature Switzerland. (doi:10.1007/978-3-030-56681-4)
49. Warren DL, Cardillo M, Rosauer DF, Bolnick DI. 2014 Mistaking geography for biology: inferring processes from species distributions. *Trends Ecol. Evol.* **29**, 572–580. (doi:10.1016/j.tree.2014.08.003)
50. McIntire EJB, Fajardo A. 2009 Beyond description: the active and effective way to infer processes from spatial patterns. *Ecology* **90**, 46–56. (doi:10.1890/07-2096.1)
51. Fick SE, Hijmans RJ. 2017 WorldClim 2: new 1-km spatial resolution climate surfaces for global land areas. *Int. J. Climatol.* **37**, 4302–4315. (doi:10.1002/joc.5086)
52. Borregaard MK, Rahbek C, Fjeldsø J, Parra JL, Whittaker RJ, Graham CH. 2014 Node-based analysis of species distributions. *Methods Ecol. Evol.* **5**, 1225–1235. (doi:10.1111/2041-210X.12283)
53. Chazot N *et al.* 2021 Conserved ancestral tropical niche but different continental histories explain the latitudinal diversity gradient in brush-footed butterflies. *Nat. Commun.* **12**, 5717. (doi:10.1038/s41467-021-25906-8)
54. Tseng HY, Chiba H, Lohman DJ, Yen SH, Aduse-Poku K, Ohshima Y, Wu LW. 2022 Out of Asia: Intercontinental dispersals after the Eocene-Oligocene transition shaped the zoogeography of Limenitidinae butterflies (Lepidoptera: Nymphalidae). *Mol. Phylogenet. Evol.* **170**, 107444. (doi:10.1016/j.ympev.2022.107444)
55. Lohman DJ, de Bruyn M, Page T, von Rintelen K, Hall R, Ng PKL, Shih HT, Carvalho GR, von Rintelen T. 2011 Biogeography of the Indo-Australian Archipelago. *Annu. Rev. Ecol. Evol. Syst.* **42**, 205–226. (doi:10.1146/annurev-ecolsys-102710-145001)
56. Reilly SB *et al.* 2022 Diverge and Conquer: Phylogenomics of southern Wallacean forest skinks (Genus: *Sphenomorphus*) and their colonization of the Lesser Sunda Archipelago. *Evolution* **76**, 2281–2301. (doi:10.1111/evo.14592)
57. Letsch H, Balke M, Kusy D, McKenna DD, Pramesa Narakusumo R, Sagata K, Toussaint EFA, White LT, Riedel A. 2023 Beetle evolution illuminates the geological history of the World's most diverse tropical archipelago. *Ecography* **2023**, e06898. (doi:10.1111/ecog.06898)
58. Goldberg J, Trewick SA, Paterson AM. 2008 Evolution of New Zealand's terrestrial fauna: a review of molecular evidence. *Phil. Trans. R. Soc. B* **363**, 3319–3334. (doi:10.1098/rstb.2008.0114)
59. Winkworth RC, Wagstaff SJ, Glenny D, Lockhart PJ. 2005 Evolution of the New Zealand mountain flora: origins, diversification and dispersal. *Org. Divers. Evol.* **5**, 237–247. (doi:10.1016/j.ode.2004.12.001)
60. Koot EM, Morgan-Richards M, Trewick SA. 2020 An alpine grasshopper radiation older than the mountains, on Kā Tiritiri o te Moana (Southern Alps) of Aotearoa (New Zealand). *Mol. Phylogenet. Evol.* **147**, 106783. (doi:10.1016/j.ympev.2020.106783)
61. Grehan JR, Mielke CGC. 2018 Evolutionary biogeography and tectonic history of the ghost moth families Hepialidae, Mnesarchaeidae, and Palaeosetidae in the Southwest Pacific (Lepidoptera: Exoporia). *Zootaxa* **4415**, 243–275. (doi:10.11646/zootaxa.4415.2.2)
62. Braga MP, Janz N, Nylin S, Ronquist F, Landis MJ. 2021 Phylogenetic reconstruction of ancestral ecological networks through time for pierid butterflies and their host plants. *Ecol. Lett.* **24**, 2134–2145. (doi:10.1111/ele.13842)
63. Cornell HV, Harrison SP. 2013 Regional effects as important determinants of local diversity in both marine and terrestrial systems. *Oikos* **122**, 288–297. (doi:10.1111/j.1600-0706.2012.20691.x)
64. Braby MF, Trueman JWH. 2006 Evolution of larval host plant associations and adaptive radiation in pierid butterflies. *J. Evol. Biol.* **19**, 1677–1690. (doi:10.1111/j.1420-9101.2006.01109.x)
65. Fine PVA. 2015 Ecological and evolutionary drivers of geographic variation in species diversity. *Annu. Rev. Ecol. Evol. Syst.* **46**, 369–392. (doi:10.1146/annurev-ecolsys-112414-054102)
66. Archibald SB, Bossert WH, Greenwood DR, Farrell BD. 2010 Seasonality, the latitudinal gradient of diversity, and Eocene insects. *Paleobiology* **36**, 374–398. (doi:10.1666/09021.1)
67. Abraham JO, Upham NS, Damian-Serrano A, Jesmer BR. 2022 Evolutionary causes and consequences of ungulate migration. *Nat. Ecol. Evol.* **6**, 998–1006. (doi:10.1038/s41559-022-01749-4)
68. Chowdhury S, Fuller RA, Dingle H, Chapman JW, Zalucki MP. 2021 Migration in butterflies: a global overview. *Biol. Rev.* **96**, 1462–1483. (doi:10.1111/brv.12714)
69. Schmidt C, Muñoz G, Lancaster LT, Lessard JP, Marske KA, Marshall KE, Garroway CJ. 2022 Population demography maintains biogeographic boundaries. *Ecol. Lett.* **25**, 1905–1913. (doi:10.1111/ele.14058)
70. Batz ZA, Clemento AJ, Fritzenwanker J, Ring TJ, Garza JC, Armbruster PA. 2020 Rapid adaptive evolution of the diapause program during range expansion of an invasive mosquito. *Evolution* **74**, 1451–1465. (doi:10.1111/evo.14029)
71. Wepprich T, Henry E, Haddad N. In press. Voltinism shifts in response to climate warming generally benefit populations of multivoltine butterflies. *Ecol. Lett.*
72. Saunders ME, Janes JK, O'Hanlon JC. 2020 Moving on from the insect apocalypse narrative: engaging with evidence-based insect conservation. *Bioscience* **70**, 80–89. (doi:10.1093/biosci/biz143)
73. van Klink R, Bowler DE, Gongalsky KB, Shen M, Swengel SR, Chase JM. 2024 Disproportionate declines of formerly abundant species underlie insect loss. *Nature* **628**, 359–364. (doi:10.1038/s41586-023-06861-4)
74. Braby MF, Williams MR, Coppen RAM, Williams AAE, Franklin DC. 2020 Patterns of species richness and endemism of butterflies and day-flying moths in the monsoon tropics of northern Australia. *Biol. Conserv.* **241**, 108357. (doi:10.1016/j.biocon.2019.108357)
75. Bernardo-Madrid R, Calatayud J, González-Suárez M, Rosvall M, Lucas PM, Rueda M, Antonelli A, Revilla E. 2019 Human activity is altering the world's zoogeographical regions. *Ecol. Lett.* **22**, 1297–1305. (doi:10.1111/ele.13321)
76. Daru B. 2024 Data from: A global database of butterfly species native distributions [dataset]. Dryad Digital Repository. (doi:10.5061/dryad.bzkh189h9)
77. Gross C, Wright A, Daru B. 2024 Supplementary material from: A global biogeographic regionalization for butterflies. Figshare (doi:10.6084/m9.figshare.c.7552614)