

# 1 **Robust, data-driven bioregionalizations emerge from diversity concordance**

2

3 Cristian S. Montalvo-Mancheno <sup>a,\*</sup>, Jessie C. Buettel <sup>a,b</sup>, Stefania Ondei <sup>a</sup>, Barry W. Brook <sup>a,b</sup>

4

5 <sup>a</sup> School of Natural Sciences, University of Tasmania, Sandy Bay 7005 TAS Australia

6 <sup>b</sup> ARC Centre of Excellence for Australian Biodiversity and Heritage (CABAH), University of

7 Tasmania

8 \* Correspondence to: [cristian.montalvomancheno@utas.edu.au](mailto:cristian.montalvomancheno@utas.edu.au)

9

10

## 11 **Abstract**

12 **Aim:** Despite the increasing interest in developing new bioregionalizations and assessing the  
13 most widely accepted biogeographic frameworks, no study to date has sought to systematically  
14 define a system of small bioregions nested within larger ones that better reflect the distribution  
15 and patterns of biodiversity. Here, we examine how an algorithmic, data-driven model of  
16 diversity patterns can lead to an ecologically interpretable hierarchy of bioregions.

17 **Location:** Australia.

18 **Time period:** Present.

19 **Major taxa studied:** Terrestrial vertebrates and vascular plants.

20 **Methods:** We compiled information on the biophysical characteristics and species occupancy of  
21 Australia's geographic conservation units (bioregions). Then, using cluster analysis to identify  
22 groupings of bioregions representing optimal discrete-species areas, we evaluated what a  
23 hierarchical bioregionalization system would look like when based empirically on the within-

24 and between-site diversity patterns across taxa. Within an information-analytical framework, we  
25 then assessed the degree to which the World Wildlife Fund's (WWF) biomes and ecoregions and  
26 our suite of discrete-species areas are spatially associated and compared those results among  
27 bioregionalization scenarios.

28 **Results:** Information on biodiversity patterns captured was moderate for WWF's biomes (50–  
29 58% for birds' beta, and plants' alpha and beta diversity, of optimal discrete areas, respectively)  
30 and ecoregions (additional 4–25%). Our plants and vertebrate optimal areas retained more  
31 information on alpha and beta diversity across taxa, with the two algorithmically derived  
32 biogeographic scenarios sharing 86.5% of their within- and between-site diversity information.  
33 Notably, discrete-species areas for beta diversity were parsimonious with respect to those for  
34 alpha diversity.

35 **Main conclusions:** Nested systems of bioregions must systematically account for the variation  
36 of species diversity across taxa if biodiversity research and conservation action are to be most  
37 effective across multiple spatial or temporal planning scales. By demonstrating an algorithmic  
38 rather than subjective method for defining bioregionalizations using species-diversity  
39 concordances, which reliably reflects the distributional patterns of multiple taxa, this work offers  
40 a valuable new tool for systematic conservation planning.

41

#### 42 **Keywords**

43 Biogeography, Biodiversity conservation, Regionalization, Ecoregions, Multivariate methods,  
44 Species-area relationship, Alpha diversity, Beta diversity, Vertebrate taxa, Vascular plants

45

46 **1. Introduction**

47 The division of the Earth's surface into regions of unique biotic communities or similar  
48 ecological processes is a cornerstone of biogeographic and macroecological research (Ebach &  
49 Parenti, 2015; Mackey, Berry, & Brown, 2008). The identification of alternative macrounits of  
50 biodiversity reflects differences in scale of analysis, methods of classification, and type of data  
51 governing the existing biogeographic frameworks (Mackey et al., 2008). Many  
52 bioregionalizations have delineated precise geographic units based on differences in species  
53 composition (Kreft & Jetz, 2010; Wallace, 1876) and/or discontinuities in the abiotic  
54 environment (Olson et al., 2001; Omernik, 2004). Recently, biogeographic frameworks have  
55 focused instead on either using phylogenetic data (Daru, Elliott, Park, & Davies, 2017; Holt et  
56 al., 2013; Maestri & Duarte, 2020) to define those hard boundaries, or have taken a 'softer'  
57 approach to their geographic delineation by identifying transition zones (Edler, Guedes, Zizka,  
58 Rosvall, & Antonelli, 2017; Vilhena & Antonelli, 2015). Another key difference is that while  
59 some biogeographers have sought to define systems of small units nested within larger ones—  
60 such is the case of World Wildlife Fund (WWF) Terrestrial Ecoregions (Dinerstein et al., 2017;  
61 Olson et al., 2001)—others have rejected this as a desirable outcome of the study system (Ebach  
62 & Parenti, 2015). At finer resolution, the spatial delineation of homogeneous areas characterized  
63 by broad, landscape-scale natural features and environmental processes (called land systems,  
64 bioregions, or equivalentents) has also been crucial in the development of many bioregionalizations  
65 at sub-continental scales (Mackey et al., 2008). Yet, despite the great contribution that this  
66 diversity of frameworks has made to our understanding of biodiversity patterns, creating an  
67 objective, repeatable and transferable hierarchical system of geographic operational units to

68 meaningfully aggregate biodiversity from-regional-to-global scale, across multiple taxonomic  
69 groups, remains elusive (Antonelli, 2017; Morrone, 2018).

70

71         Alongside Wallace’s zoogeographic regions (Wallace, 1876), WWF’s hierarchical  
72 framework (Olson et al., 2001) of discrete areas of natural communities (i.e., ecoregions),  
73 spatially nested within larger distinct areas that reflect relations between climate, flora and fauna  
74 (i.e., biomes), is the most widely accepted global bioregionalization that has been the foundation  
75 of scientific research, environmental policy, resource management and conservation for almost  
76 two decades (Kier et al., 2005; Mackey et al., 2008; Smith et al., 2018). However, few  
77 continental-level bioregionalizations that correspond to a spatial subdivision of WWF’s  
78 geographic operational units have been developed and adopted over this period (Omernik, 2004;  
79 Thackway & Cresswell, 1995). Notably, only the bioregionalization for Australia—known as the  
80 Interim Biogeographic Regionalization for Australia (IBRA) framework—has been explicitly  
81 defined as a more detailed geographic division of WWF’s ecoregions (Department of  
82 Agriculture, 2012a).

83

84         Since its inception, the IBRA framework has been framed as a tool to guide the  
85 systematic conservation planning of Australia’s biodiversity. The initial release divided the  
86 continent in 80 biogeographic regions, called bioregions (Thackway & Cresswell, 1995). Over  
87 the subsequent 25 years, new bioregions have been identified, and boundaries updated to  
88 coincide better with faunal and floral patterns and environmental processes that influence the  
89 functioning of entire ecosystems. In an effort to aid regional conservation, IBRA bioregions have  
90 been further divided into subregions, using finer-scale differences in biophysical attributes—

91 such as geology and vegetation—to spatially define major regional ecosystems. Like WWF’s  
92 global ecoregions and other continental-scale bioregionalization templates (Mackey et al., 2008;  
93 Omernik, 2004), the IBRA framework reflects a hierarchical structure delineated bio-topically  
94 where the spatial aggregation of subregions makes up bioregions (Department of Agriculture,  
95 2012a). The creation and update of WWF’s and Australia’s biogeographic frameworks have also  
96 been similar, in that the tacit knowledge of an expert panel was used to compile a suite of  
97 disparate spatial information to define regions within which geographic phenomena associated  
98 with differences in ecosystems’ characteristics (i.e., health, quality, and integrity) coincide  
99 (Dinerstein et al., 2017; Olson et al., 2001; Omernik, 2004; Thackway & Cresswell, 1995). The  
100 subjective, expert-based derivation of these two bioregionalizations has prompted criticism.  
101 Nonetheless, only WWF’s biomes and ecoregions have been scrutinized quantitatively, with  
102 their capacity to discriminate species diversity shown to perform better than a random allocation  
103 of boundaries (Smith et al., 2018), but worse in comparison to remotely sensed productivity  
104 clusters (Coops, Kearney, Bolton, & Radeloff, 2018).

105

106 A plethora of studies seeking to develop new bioregionalization scenarios, and to assess  
107 the most widely accepted biogeographic templates, have emerged over the last decade (Ebach &  
108 Parenti, 2015; Kreft & Jetz, 2010). This revived interest is due to a number of recent  
109 developments. There is an increasing accessibility to ecological datasets that provide systematic  
110 information on species distributions, as well as on other facets of biodiversity (e.g., phylogenetic,  
111 and functional diversity) over large extents and at an increasingly finer resolution for many taxa  
112 (Daru et al., 2017; Ficetola, Mazel, & Thuiller, 2017). Alongside this, ecologists and  
113 biogeographers are increasingly using remotely sensed data to understand biodiversity patterns

114 and process across multiple spatial and/or temporal scales (Coops et al., 2018). Nonetheless,  
115 many—if not all—of the aggregative frameworks to emerge during the past 20 years would not  
116 have been possible without the use of high-performance computing infrastructure. Together with  
117 high volumes of processing power, spatially explicit aggregative and comparative techniques and  
118 new approaches to disentangle fundamental properties of ecological systems have also been  
119 introduced. These advances have opened the possibility to develop new algorithm-driven  
120 bioregionalizations that are objective, reproduceable, and tractable. However, whether the  
121 operational units of quantitative bioregionalizations—like in bioregionalizations defined by an  
122 expert-panel—can capture multiple facets of biodiversity remains highly contested and of much  
123 research interest among biogeographers (Ebach & Parenti, 2015; Mackey et al., 2008; Morrone,  
124 2018).

125  
126 Motivated by these problems with definition and implementation, we developed an  
127 integrative, data-driven approach to bioregionalization that leverages the information on species  
128 richness and composition within the bioregions of the IBRA framework. More specifically, we  
129 asked: 1) What would an IBRA framework look like when based empirically on the  
130 accumulation of species with increase in area (species-area relationship), and the within- and  
131 between-site species diversity (alpha and beta) for multiple taxonomic groups (hereafter referred  
132 as optimal discrete-species clusters)? 2) How well do WWF’ biomes and ecoregions match  
133 ‘optimal’ discrete-species clusters? 3) Is the spatial configuration of optimal-discrete-plant and -  
134 vertebrate clusters better associated with each other than with other discrete species clusters at  
135 lower taxonomic ranks, and when compared to the spatial association between WWF’s  
136 operational units (biomes and ecoregions) and those of optimal-discrete species clusters? By

137 answering those questions, we reveal a hierarchical system of spatial partitions that is  
138 ecologically interpretable, and thereby best suited to inform biodiversity policy, research, and  
139 conservation.

140

## 141 **2. Materials and Methods**

### 142 **2.1 Data collection and processing**

143 We collected spatial information on climate (Hallgren et al., 2016), elevation (Earth Resources  
144 Observation and Science (EROS) Center, NA), vegetation (Department of Agriculture, 2018),  
145 soil (Australian Soil Resource Information System, 2013), lithology (Raymond et al., 2012), and  
146 occurrences of terrestrial species native to Australia (Atlas of Living Australia, NA), which after  
147 pre-processing to minimize errors and biases included 25,995 native species: 23,248 vascular  
148 plants, 233 amphibians, 1,201 birds, 349 mammals, and 964 reptiles (see Appendix S1 in  
149 Supporting Information for details).

150

151 We downloaded version seven of IBRA subregions' names and borders (Department of  
152 Agriculture, 2012b) to derive a spatially coherent four-tier hierarchical system of geographic  
153 operational units for Australia (see Appendix S1 for details)—where 410 IBRA subregions are  
154 nested within 85 IBRA bioregions. Based on the most recent version of WWF's  
155 bioregionalization (Dinerstein et al., 2017) those bioregions are nested within 37 ecoregions, and  
156 those macrounits are themselves embedded in 7 broader-scale (and spatially coherent) biomes.  
157 We chose to use geographic operational units, because analyses based on lists of species within  
158 such units, as opposed to a grid-based approach, can highlight gradual changes in species  
159 diversity and are less likely to distort areal relationships due to heterogeneity in the sizes of

160 species ranges (Kreft & Jetz, 2010; Kreft, Sommer, & Barthlott, 2006; Morrone & Escalante,  
161 2002). We deem this a desirable feature since our goal was to reveal a bioregionalization's  
162 hierarchical system of discrete spatial clusters that is more directly relevant to biodiversity.

163

164 We characterized Australia's biophysical space by calculating the mean value of elevation  
165 and of nineteen climatic variables within IBRA subregions and bioregions. As the geographic-  
166 based measures for lithology, soil, and vegetation, we computed the percentage cover for each of  
167 these factors' categories relative to the size of subregions and bioregions (see Appendix S1 for  
168 details on excluded categories for these three discrete variables). Meanwhile, for the  
169 characterization of Australia's biotic space, we derived presence-absence matrices for vascular  
170 plants and four vertebrate classes (amphibians, birds, mammals, and reptiles) by intersecting  
171 IBRA bioregions and subregions with both species occurrences with less than 20 records post-  
172 equalization (i.e., a procedure to even out the difference in number of species occurrences among  
173 IBRA subregions by minimizing the variance within subregions' size-classes, while maximizing  
174 the variance between size classes) and our set of empirical extent-occurrence maps for those  
175 species with at least 20 records (see Appendix S1 for details and rationality). We joined  
176 amphibian and reptile presence-absence matrices into a single group (herpetofauna) to ensure  
177 that there were at least 10 species per IBRA operational units across taxonomic groups; we also  
178 created a presence-absence matrix for all vertebrate species. In terms of species, amphibians and  
179 reptiles follow different biogeographic patterns (Powney, Grenyer, Orme, Owens, & Meiri,  
180 2010), yet as a broad taxonomic group (i.e., herpetofauna), exothermic species represent a huge  
181 array of evolutionary adaptations that allow them to cover a wide range of potential niches. We  
182 used ArcGIS v. 10.5.1 (2017) to harmonize spatial data to a common format and coordinate



183 reference system (Australian Albers Equal Area; EPSG: 3577). All spatial calculations and  
184 feature engineering were done in R v. 3.6.3 (R Core Team, 2020) using several packages (see  
185 Appendix S1 for complete list).

186

## 187 **2.2 Metric to discriminate spatial clusters in biophysical space**

188 To assess the ecological significance of the ordination of geographic operational units of the  
189 IBRA framework based on biophysical factors, we used the species-area relationship (SAR) as a  
190 metric because SAR is one of the well-studied properties of ecological systems and has been  
191 applied in identifying priority areas for biodiversity conservation at large scales (Guilhaumon,  
192 Gimenez, Gaston, & Mouillot, 2008; Triantis, Guilhaumon, & Whittaker, 2012). Among nine  
193 alternative mathematical functions, we selected the logarithmic form of the power function to fit  
194 SAR for vascular plants and selected vertebrate species (bird, mammal, and herpetofauna) based  
195 on Akaike's Information Criteria (AIC) (Akaike, 1974) (see Appendix S2 for details and results).

196

## 197 **2.3 Selection of biophysical factors and IBRA unit of analysis**

198 Our selected geographic and environmental covariates represent a complex dataset ( $n = 71$   
199 variables) that describes IBRA subregions and bioregions by sets of variables structured into  
200 groups. We defined the distance between distinct IBRA spatial clusters at a hierarchical level to  
201 be based on an equal contribution of these five groups of continuous variables. To balance the  
202 influences of each group of variables in the description of distinct spatial clusters (Bécue-Bertaut  
203 & Pagès, 2008), we used multiple factor analysis (MFA) to assess the contribution of groups of  
204 variables to the characterization of IBRA operational units, and to identify the number of  
205 principal components needed to retain at least 90% of the variance, using their eigenvalues to

206 model the dissimilarity of IBRA operational units in the biophysical space instead of using  
207 geographic-based measures.

208

209         Additionally, we assessed the relevance of MFA results to discriminate biophysical  
210 factors, if necessary, and to identify the most appropriate hierarchical level of the IBRA  
211 framework for revealing the nature of the IBRA framework differences in terms of species  
212 diversity. We did this by visually exploring the spatial coherence and ecological interpretability  
213 of the ordination of subregions and bioregions into seven clusters—matching the number of  
214 WWF biomes—based on principal components and the ‘static’ technique to cut dendrograms  
215 (see next section for details). This assessment identified IBRA bioregions as the most appropriate  
216 geographic unit of analysis compared with subregions and reduced the biophysical dataset’s  
217 structure to include the eigenvalues of the 27 principal components based on climate, lithology,  
218 soil, and vegetation (see Appendix S3 for details) when constructing the ordination structure of  
219 IBRA bioregions in biophysical space. We used ‘FactoMineR’ v. 1.42 (Lê, Josse, & Husson,  
220 2008) package in the program R to perform MFA, with variables standardized, and the name of  
221 IBRA subregions and bioregions set as non-active variables.

222

## 223 **2.4 Discontinuities of species diversity**

### 224 **2.4.1 Hierarchical clustering**

225 We ordered IBRA bioregions using the Ward’s method as the clustering algorithm, and  
226 Euclidian distance and species turnover—measured with the Beta-Simpson index (Lennon,  
227 Koleff, Greenwood, & Gaston, 2001; Simpson, 1943)—as the dissimilarity measures of  
228 bioregions in terms of principal components for our suite of biophysical factors (hereafter

229 referred as PC-biophysical) and species composition, respectively. Since no existing method is  
230 capable of maximizing both clustering criteria simultaneously (i.e., the amount of information  
231 retained in the dendrogram, and the clusters' internal coherence), we chose to use Ward's  
232 algorithm, because it has proven to perform best in the second criterion (Castro-Insua, Gómez-  
233 Rodríguez, & Baselga, 2018; Kreft & Jetz, 2010). In addition, the identification of a hierarchical  
234 system of distinct spatial clusters that minimizes within-cluster and maximizes between-cluster  
235 dissimilarity (i.e., clusters' internal coherence) in terms of biodiversity is a highly desirable  
236 outcome for any bioregionalization (Ebach & Parenti, 2015; Kreft & Jetz, 2010), and aligns  
237 clearly with our study's overarching goal. We used the 'stats' v. 3.6.3 (R Core Team, 2020), and  
238 the 'betapart' v. 1.5.1 (Baselga & Orme, 2012) packages to compute dissimilarity matrices.  
239 While bioregions' cluster analysis based on species compositional dissimilarity was done using  
240 'stats' too, we used the 'FactoMineR' (Lê et al., 2008) package to conduct that analysis in the  
241 PC-biophysical dissimilarity space.

242

#### 243 **2.4.2 Optimal discrete-species clusters**

244 We cut the dendrograms resulting from the hierarchical cluster analyses of IBRA bioregions in  
245 the PC-biophysical space and the dissimilarity in species composition for five taxonomic groups  
246 (birds, mammals, herpetofauna, vertebrate, and vascular plants) using different techniques to  
247 identify sets of prominent, spatially coherent biodiversity clusters. This included: defining  
248 continuous dendrogram branches based on a desired number of clusters ('static' technique),  
249 pruning branches based on their structure in the dendrogram ('dynamic' technique; specifically a  
250 bottom-up algorithm, called Dynamic Hybrid Cut) (Langfelder, Zhang, & Horvath, 2008), and/or  
251 identifying the intersection point between two straight lines that best fit an evaluation curve by

252 minimizing the total root mean square error ('L' technique) (Salvador & Chan, 2004). We used  
253 these pruning techniques together with goodness-of-fit and parsimony metrics (e.g., R-squared,  
254 and AIC and/or Bayesian Information Criteria – BIC) to discriminate prominent clusters in terms  
255 of species accumulation as the area sampled is increased (SAR models for prominent clusters in  
256 the PC-biophysical space), and the within- and across-variance of species richness and  
257 composition (alpha and beta diversity models, respectively). We selected the most parsimonious  
258 yet ecologically coherent model of the species-area relationship, and pairs of the best models of  
259 the variance in alpha and beta diversity, respectively, as the optimal discrete-species clusters (see  
260 Appendix S4 for methodological details).

261

## 262 **2.5 An ecologically meaningful algorithmic IBRA framework**

263 To reveal a hierarchical system of IBRA bioregions that meaningfully aggregates species  
264 diversity from-regional-to-global scale, we examined the ecological interpretability of our suite  
265 of eleven optimal discrete-species clusters (i.e., ten optimal partitions of species diversity, along  
266 with the optimal partition of SAR). We did this by calculating the slope and the standard error of  
267 the regression line of the species-area relationship across five taxonomic groups for (i) the spatial  
268 configuration of the optimal SAR model and (ii) the bespoke WWF's classification of IBRA  
269 bioregions into seven biomes. We then plotted the distribution, mean, median, and standard  
270 deviation of the merging height of nodes in the dendrograms of species composition, to visualize  
271 differences in the dissimilarity distance within and across the five target taxonomic groups.

272

273 We assessed the spatial coherence of optimal discrete-species clusters visually, and then  
274 quantitatively by estimating how well WWF's classification of IBRA bioregions into biomes and

275 ecoregions captured information on species richness and turnover stored in our set of maps for  
276 the optimal discrete-species clusters. We computed an overall global measure of association—  
277 called the ‘V-measure’ and implemented in the ‘sabre’ v. 0.3.2 (Nowosad & Stepinski, 2018)  
278 package—to quantify the degree of spatial association between these maps. The 0-1 range of the  
279 V-measure is grounded in information theory and interpretable in terms of analysis of variance,  
280 where 0 indicates absence of spatial association between two maps and 1 when the spatial  
281 association is perfect—respectively meaning that the amount of mutual information of a pairwise-  
282 map comparison totally differs or is identical. We computed this global measure of association to  
283 determine whether the amount of information on species richness and composition stored in the  
284 optimal discrete-plant clusters and the optimal discrete-vertebrate clusters is higher with each  
285 other than with other discrete-species clusters at lower taxonomic ranks, and then compared  
286 these results with those for the WWF ecoregions.

287

### 288 **3. Results**

#### 289 **3.1 Optimal discrete-species clusters**

290 To identify an optimal partition of IBRA bioregions in the PC-biophysical space, we fitted 150  
291 SAR models across five taxonomic groups to five sets of prominent clusters that were defined  
292 using three different dendrogram cutting techniques (see methods and Appendix S4 for details).  
293 Based on AIC scores, the best-selected SAR models across this set of prominent clusters  
294 frequently included the same grouping of bioregions as the best distinct spatial cluster among  
295 five taxonomic groups, except in three cases: SAR of plants where the seven groups were  
296 defined using the ‘dynamic’ technique, and SAR of mammals in prominent clusters with seven  
297 and nine groups based on ‘static’ and the ‘L’ techniques, respectively (Table 1).

298 *< Insert table 1 around here >*

299

300         These three special cases also had the highest variance in species richness explained by  
301 size of bioregions (Table 1). In the prominent clusters to which two of them belong (i.e.,  
302 prominent clusters with seven and nine groups based on ‘static’ and the ‘L’ techniques), the  
303 Tropical and Subtropical Moist Broadleaf Forests biome was reconstructed (Fig. S5.5 c and e in  
304 Supporting Information). Yet, the rate of increase in the number of species per standard area  
305 differed greatly among taxonomic groups for this biome (Biome 7 in Fig. 1b), with the slope of  
306 SAR models for vertebrates and vascular plants being 1.5 and 5.5 the slope of the bird SAR  
307 model. Considering this as a whole, the partition of bioregions into seven groups using the  
308 ‘dynamic’ tree-cutting technique—which based on its formulation is a technique that improves  
309 the detection of outlying group members in a prominent cluster (Langfelder et al., 2008)—was  
310 finally selected as the optimal partition for changes in species richness per standard area (i.e.,  
311 optimal partition for SAR).

312

313         Based on our approach to discriminate prominent clusters in terms of the variance of  
314 alpha or beta diversity across groups of bioregions in sets of prominent clusters (see Appendix  
315 S4), the clearest break in log-likelihood of the within- and across-variance of species richness  
316 (i.e., ANOVA models for birds, mammals, herpetofauna, vertebrate, and vascular plants) was  
317 evident when dendrograms of compositional dissimilarity were cut using the ‘dynamic’  
318 technique (Fig. S6.6). Notably, the optimal discrete-species cluster for vertebrates’ alpha  
319 diversity was defined using only the prominent cluster with the lowest BIC score rather than the  
320 agreement between AIC and BIC scores, as done for the other taxonomic groups. Likewise, the

321 highest ratio between the sum of multiple-site measures of compositional dissimilarity across  
322 groups of bioregions in sets of prominent clusters and the multiple-site measure of compositional  
323 dissimilarity across Australia’s bioregions appeared stable across taxa—particularly for birds,  
324 mammals, and vascular plants—when prominent clusters were defined by ‘dynamic’ tree-cutting  
325 technique (Fig. S7.7). As in the optimal partition for SAR, optimal partitions of alpha and beta  
326 diversity among all five taxonomic groups were also more appropriately identified when  
327 dendrograms of species compositional dissimilarity were cut using the ‘dynamic’ technique, and  
328 thereby this suite of eleven optimal discrete-species clusters were used in subsequent analysis.

329 *< Insert figure 1 around here >*

330

### 331 **3.2 An ecologically meaningful IBRA framework**

332 Using the spatial configuration of our suite of optimal clusters (Fig. S8.8 b–l) to examine what an  
333 IBRA framework would look like when based empirically on patterns of species diversity, we  
334 found that while the most parsimonious, ecologically coherent model of Australia’s biophysical  
335 dissimilarity (i.e., optimal partition of SAR)—like WWF’s biome map—also aggregated  
336 bioregions into seven groups, our algorithmically driven bioregionalization differed from the  
337 bespoke WWF’s expert-derived classification (Fig. 1 a). When assessing the ecological  
338 significance of the spatial configuration of these two biogeographic scenarios—based on a log-  
339 log implementation of the power function to fit the species-area relationship—we found that the  
340 increase in species richness with area varies among distinct spatial regions and taxonomic groups  
341 within regions (Fig. 1 b and c). Notably, similar—if not the same—patterns of species richness  
342 per standard area across taxonomic groups were detected between two regions of the optimal  
343 partition of SAR and those of WWF’s expert-derived classification that overlapped (i.e., Group 1

344 vs. Biome 4 - Temperate Broadleaf & Mixed Forests; and Group 3 vs. Biome 6 - Tropical &  
345 Subtropical Grasslands, Savannas & Shrublands), despite geographical differences due to the  
346 number of bioregions within their boundaries.

347 *< Insert figure 2 around here >*

348

349 When hierarchical systems of IBRA bioregions were defined based on pairs of the best  
350 models of the variance in species richness and multiple-site compositional dissimilarity (i.e.,  
351 optimal partitions of alpha and beta diversity), we found that like for the optimal partition of  
352 SAR, the distinct spatial clusters for the alpha- and beta-diversity of birds, mammals,  
353 herpetofauna, vertebrates and vascular plants largely included adjacent bioregions (Fig. S8.8 c–  
354 l), even with no explicit spatial aggregation or distance penalty being imposed on the algorithm.  
355 For vertebrate and vascular plant species, the spatial configuration and the numbers of optimal  
356 discrete clusters were the same for their alpha and beta diversity (Fig. S8.8 i–l)—with distinct  
357 plant-species areas almost perfectly collapsing within those for vertebrates (Fig. 2). Further, the  
358 distribution of the height of nodes in the species composition dendrograms was right skewed  
359 across all taxa, with the variation of compositional dissimilarity being larger for herpetofauna  
360 and vascular plants (Fig. 3).

361 *< Insert figure 3 around here >*

362

363 Within an information-theoretical analytical framework (Nowosad & Stepinski, 2018),  
364 the degree of spatial association between WWF's biome map and those of optimal discrete-  
365 species clusters was moderate (Table 2), ranging from 50% of the information shared between  
366 the biome map and the spatial configuration of beta diversity in birds, and 58% for alpha and



367 beta diversity in vascular plants. At the ecoregional level, WWF's bioregionalization of Australia  
368 captured an additional 4 to 25% of information on patterns of species diversity embodied within  
369 our suite of optimal clusters, except when compared to the optimal discrete-bird map of beta  
370 diversity (i.e., - 4.3% spatial association between maps). The spatial concordance of optimal  
371 discrete-plant and -vertebrate clusters was slightly worse with each other than with other optimal  
372 discrete-species clusters at lower taxonomic ranks only for the pairwise-map comparisons  
373 between discrete-vertebrate and -mammal clusters (Table 2), for which the loss of information on  
374 alpha and beta diversity ranged from 0.8% to 4%, respectively. When contrasting the spatial  
375 association results of these two biogeographic scenarios, an algorithmically driven IBRA  
376 framework of discrete-plant and -vertebrate clusters retained more information on species alpha  
377 and beta diversity patterns within bioregions across multiple taxonomic groups than WWF's  
378 hierarchical system of biomes and ecoregions. Nonetheless, the bespoke expert partitioning of  
379 Australia (IBRA) performed slightly better at retaining differences in species richness between  
380 bioregions of varying sizes (Table 2).

381 *< Insert table 2 around here >*

382

### 383 **3. Discussion**

384 Just as endemism is a commonly used basis for bioregionalizations (Ebach & Parenti, 2015;  
385 Morrone, 2014), species' intrinsic traits are also responsible for defining ecologically meaningful  
386 clusters at large scale. We found strong spatial concordances across taxonomic groups and  
387 patterns of species richness and composition among the suite of optimal discrete-species clusters,  
388 signaling the effect that species' biology and evolutionary history have in the identification of  
389 distinct areas of co-occurring species. The variation of compositional dissimilarity across our  
390 five target taxonomic groups (Fig. 3) reflects the overall small distributions of species with

391 restricted range and/or low occupancy (Kreft & Jetz, 2010) and suggests that species' dispersal  
392 abilities are important determinants of the emergent biogeographic divisions. Furthermore,  
393 having discrete plant-species areas almost perfectly nested within larger vertebrate-species areas  
394 suggests that interspecific relationships are also involved in explaining most of the variation in  
395 species diversity. Consequently, basing a hierarchical system of bioregions on plants and  
396 vertebrate optimal clusters is ecologically intuitive, because plants are essential to all animals,  
397 and interspecific interactions are largely responsible for generating community structure (Wisz et  
398 al., 2013).

399

400 Underpinning hierarchical systems of bioregions solely in the analysis of geographic and  
401 environmental covariates is not necessarily so strong as to capture the distributional patterns of  
402 multiple taxa in unison. Variation in the z-values (log-slopes) of the species-area relationship  
403 among discrete macrounits of biodiversity, such as the case of biomes, has already been  
404 documented (Kier et al., 2005). Yet, when combined with both the variability of z-values across  
405 taxonomic groups, and the parsimony of distinct spatial clusters for beta diversity with respect to  
406 those for alpha diversity (Fig. S8.8 c-l), it suggests that although environmental gradients are not  
407 impenetrable barriers to the arrangement of distinct communities, they can be important  
408 determinants of species richness within the structure of those communities. This demonstrates  
409 that while environmental heterogeneity is a well-established driver of species richness (Stein,  
410 Gerstner, & Kreft, 2014), defining a hierarchical system of bioregions that meaningfully  
411 aggregates biodiversity should not rely on species richness alone, because it varies among higher  
412 taxonomic groups, and does not necessarily have a positive relationship with species endemism  
413 (Koleff & Gaston, 2002).

414

415           Considering that WWF's biomes were defined using associations of climate and  
416 dominant vegetation forms and structure to broadly classify terrestrial ecosystems (Kier et al.,  
417 2005; Mackey et al., 2008), the relatively moderate degree of the overall spatial association  
418 between WWF's biomes and our eleven optimal-cluster maps was expected. This finding aligns  
419 qualitatively with those reported in recent studies (Coops et al., 2018; Edler et al., 2017), and  
420 they together give strength to our conclusion that complex interactions between biophysical  
421 factors and species' intrinsic attributes are reflected in a nested hierarchy of bioregions (Sexton,  
422 McIntyre, Angert, & Rice, 2009). This means that hierarchical bioregionalization systems must  
423 account for variability in the distributional patterns of different taxa (Morrone, 2018) if they are  
424 to direct more efficient and appropriately targeted biodiversity research, policy, and conservation  
425 across multiple spatial or temporal planning scales.

426

427           Bias in biodiversity conservation is systemic (Butchart et al., 2015) and usually attributed  
428 to conservation efforts being implemented, for practical reasons, in areas of low agricultural  
429 productivity (Joppa & Pfaff, 2009) and/or targeting specific groups of species, such as  
430 charismatic species (Colléony, Clayton, Couvet, Saint Jalme, & Prévot, 2017). However,  
431 biogeographic frameworks might also contribute to this bias. When comparing WWF's  
432 hierarchical system (ecoregions nested within biomes) to an algorithmic, data-driven hierarchical  
433 system of IBRA bioregions based on plant- and vertebrate-oriented optimal areas, the difference  
434 in the amount of information on Australian biodiversity patterns that each tier of these two  
435 bioregionalization scenarios captured was greater for the (bespoke) WWF's expert-derived  
436 classification, with neither biomes nor ecoregions outperforming an integrative, data-driven

437 alternative at retaining information on alpha and beta diversity across taxa (Table 2). This  
438 finding suggests that—from inception—systematic conservation planning may have been  
439 undermined by how a particular biogeographic framework was defined, which in the case of  
440 WWF’s global bioregionalization included biotopic and biocoenotic classification approaches  
441 (Mackey et al., 2008). This is troublesome if we consider that, in addition to being one of the  
442 most widely used biogeographic templates for biodiversity conservation, WWF’s ecoregions  
443 have served as the basis of other well-established conservation strategies at global scale  
444 (Lamoreux et al., 2006). Given that spatial information on biodiversity patterns is essential for  
445 effective biodiversity conservation, the design of environmental policies, the establishment of  
446 protected area networks, and the implementation of more recent *in situ* interventions (e.g.,  
447 rewilding, species’ translocations), there is a substantial need for a hierarchical system of  
448 geographic operational units that is ecologically interpretable across broad spatial and taxonomic  
449 breaths.

450

451 Over the past 200 years, the history of bioregionalization in Australia has been driven as  
452 much by changes in foci (i.e., from exploration, to the conservation of biodiversity) as by  
453 theoretical and methodological advances and data availability (Ebach, 2012). Multiple studies  
454 have found that different biogeographic templates, encompassing those both qualitative and  
455 quantitative perspectives, were mostly congruent with each other (Bloomfield, Knerr, & Encinas-  
456 Viso, 2018; González-Orozco, Laffan, Knerr, & Miller, 2013; González-Orozco, Thornhill,  
457 Knerr, Laffan, & Miller, 2014). Yet, even when applying similar biogeographic approaches to  
458 partition Australia’s landscape, phyto- and zoo-geographers disagreed on the boundaries of some  
459 areas of Australia, such as the arid region (Ebach & Murphy, 2020). In our algorithmic, data-

460 driven model, the partition of arid Australia is greater than in any of the bioregionalization  
461 scenarios of Bloomfield et al. (2018). Yet, despite methodological differences, a visual  
462 comparison of these two approaches' bioregionalization scenarios suggests that geographic  
463 operational units of optimal discrete-vertebrate and -plant clusters in arid Australia are spatially  
464 nested within the zones delineated in the study of Bloomfield et al. (2018), which the authors in  
465 turn argued to be spatially consistent with Eremaean biogeographic region (i.e., arid Australia  
466 based on its flora).

467  
468 As a basis for developing ecologically sensible bioregionalizations that are both  
469 methodologically robust and repeatable, we argue that this new approach is highly innovative  
470 and can be applied in many contexts. Nonetheless, there are some caveats. First, no consensus  
471 exists on how to decisively identify biogeographic divisions and to delineate their boundaries  
472 (Antonelli, 2017; Morrone, 2018). Yet, the plethora of aggregative approaches to emerge in  
473 biogeographic research has been instrumental in our understanding that any single  
474 bioregionalization cannot hope to consistently capture the distributional patterns of multiple taxa  
475 (Coops et al., 2018; Edler et al., 2017; Kreft & Jetz, 2010; Vilhena & Antonelli, 2015). While  
476 our study reinforces this conclusion, we do show that a hierarchical system more directly  
477 relevant to biodiversity can be derived systematically by leveraging information on species  
478 diversity and composition within a bioregionalization's geographic operational units. Second,  
479 biased, and inadequate knowledge on species' distribution can confound the identification of  
480 natural biogeographic areas (Pimm et al., 2014; Whittaker et al., 2005). However, a hierarchical  
481 bioregionalization system based on our algorithmic approach can be readily and transparently  
482 revised to reflect the increasing availability of distributional data, and the species' responses to

483 global and regional environmental changes. Third, because diversity is unevenly distributed  
484 between taxa, as well as in space and time (Whittaker et al., 2005) other biodiversity dimensions  
485 beyond species might be relevant. As functional and phylogenetic data continue to accumulate,  
486 and comparative approaches are advanced, a similar approach might be used to investigate  
487 whether a general bioregionalization is maintained under such patterns, and how the different  
488 dimensions of biodiversity interrelate within a particular biogeographic framework.

489  
490 This work demonstrates how a systematic, objective examination of the patterns of  
491 species diversity within the geographic operational units of a biogeographic template—in this  
492 study, 85 IBRA bioregions—can be used to develop a rigorous hierarchical system of discrete  
493 spatial partitions that is directly relevant for aggregating biodiversity. The use of existing,  
494 geographically restricted operational units, which might also be delineated using quantitative  
495 techniques, makes our approach not only generic (Mackey et al., 2008), but also sufficiently  
496 flexible to account for the increasing knowledge of biodiversity. As such, this robust, data-driven  
497 method can underpin the design and implementation of *in situ* conservation initiatives in other  
498 regions and even globally, inform policy for meaningful environmental and biodiversity  
499 outcomes across multiple planning scales, and bolster interests in the factors and the processes  
500 driving discontinuities of biodiversity. This is crucial, given the effect of human-mediated habitat  
501 transformations on the boundaries of biogeographical regions globally.

502

503 **REFERENCES**

- 504 Akaike, H. (1974). A new look at the statistical model identification. *IEEE Transactions on*  
505 *Automatic Control*, 19(6), 716–723. doi:10.1109/TAC.1974.1100705
- 506 Antonelli, A. (2017). Biogeography: drivers of bioregionalization. *Nature Ecology & Evolution*,  
507 1(4), 0114. doi:10.1038/s41559-017-0114
- 508 Atlas of Living Australia. (NA). Atlas of Living Australia. Retrieved from  
509 <https://www.ala.org.au/>
- 510 Australian Soil Resource Information System. (2013). Atlas of Australian soils. Retrieved from  
511 [https://www.asris.csiro.au/themes/Atlas.html#Atlas\\_Downloads](https://www.asris.csiro.au/themes/Atlas.html#Atlas_Downloads)
- 512 Baselga, A., & Orme, C. D. L. (2012). betapart: an R package for the study of beta diversity.  
513 *Methods in Ecology and Evolution*, 3(5), 808–812. doi:10.1111/j.2041-  
514 210X.2012.00224.x
- 515 Bécue-Bertaut, M., & Pagès, J. (2008). Multiple factor analysis and clustering of a mixture of  
516 quantitative, categorical and frequency data. *Computational Statistics & Data Analysis*,  
517 52(6), 3255–3268. doi:10.1016/j.csda.2007.09.023
- 518 Bloomfield, N. J., Knerr, N., & Encinas-Viso, F. (2018). A comparison of network and clustering  
519 methods to detect biogeographical regions. *Ecography*, 41(1), 1–10.  
520 doi:10.1111/ecog.02596
- 521 Butchart, S. H. M., Clarke, M., Smith, R. J., Sykes, R. E., Scharlemann, J. P. W., Harfoot,  
522 M., . . . Burgess, N. D. (2015). Shortfalls and solutions for meeting national and global  
523 conservation area targets. *Conservation Letters*, 8(5), 329–337. doi:10.1111/conl.12158

- 524 Castro-Insua, A., Gómez-Rodríguez, C., & Baselga, A. (2018). Dissimilarity measures affected  
525 by richness differences yield biased delimitations of biogeographic realms. *Nature*  
526 *Communications*, 9(1), 5084. doi:10.1038/s41467-018-06291-1
- 527 Colléony, A., Clayton, S., Couvet, D., Saint Jalme, M., & Prévot, A.-C. (2017). Human  
528 preferences for species conservation: animal charisma trumps endangered status.  
529 *Biological Conservation*, 206, 263–269. doi:10.1016/j.biocon.2016.11.035
- 530 Coops, N. C., Kearney, S. P., Bolton, D. K., & Radeloff, V. C. (2018). Remotely-sensed  
531 productivity clusters capture global biodiversity patterns. *Scientific Reports*, 8(1), 16261.  
532 doi:10.1038/s41598-018-34162-8
- 533 Daru, B. H., Elliott, T. L., Park, D. S., & Davies, T. J. (2017). Understanding the processes  
534 underpinning patterns of phylogenetic regionalization. *Trends in Ecology & Evolution*,  
535 32(11), 845–860. doi:10.1016/j.tree.2017.08.013
- 536 Department of Agriculture, Water and the Environment. (2012a). Australia's bioregions (IBRA).  
537 Retrieved from <http://www.environment.gov.au/land/nrs/science/ibra>
- 538 Department of Agriculture, Water and the Environment. (2012b). Interim Biogeographic  
539 Regionalisation for Australia (Subregions) v. 7 (IBRA) [ESRI shapefile]. Retrieved from  
540 <https://www.environment.gov.au/fed/catalog/search/resource/details.page?uuid=%7B8B9E3F42-9856-4487-AE9E-C76A322809A1%7D>
- 541
- 542 Department of Agriculture, Water and the Environment. (2018). National vegetation information  
543 system (NVIS) data products. Retrieved from  
544 [http://www.environment.gov.au/land/native-vegetation/national-vegetation-information-](http://www.environment.gov.au/land/native-vegetation/national-vegetation-information-system/data-products)  
545 [system/data-products](http://www.environment.gov.au/land/native-vegetation/national-vegetation-information-system/data-products)



- 546 Dinerstein, E., Olson, D., Joshi, A., Vynne, C., Burgess, N. D., Wikramanayake, E., . . . Saleem,  
547 M. (2017). An ecoregion-based approach to protecting half the terrestrial realm.  
548 *Bioscience*, 67(6), 534–545. doi:10.1093/biosci/bix014
- 549 Earth Resources Observation and Science (EROS) Center. (NA). Shuttle radar topography  
550 mission (SRTM) 1 arc-second global. Retrieved from  
551 <http://dx.doi.org/10.5066/F7PR7TFT>
- 552 Ebach, M. C. (2012). A history of biogeographical regionalisation in Australia. *Zootaxa*, 3992,  
553 1–34. doi:10.11646/zootaxa.3392.1.1
- 554 Ebach, M. C., & Murphy, D. J. (2020). Carving up Australia’s arid zone: a review of the  
555 bioregionalisation of the Eremaean and Eyrean biogeographic regions. *Australian*  
556 *Journal of Botany*, 68(3), 229–244. doi:10.1071/BT19077
- 557 Ebach, M. C., & Parenti, L. R. (2015). The dichotomy of the modern bioregionalization revival.  
558 *Journal of Biogeography*, 42(10), 1801–1808. doi:10.1111/jbi.12558
- 559 Edler, D., Guedes, T., Zizka, A., Rosvall, M., & Antonelli, A. (2017). Infomap bioregions:  
560 interactive mapping of biogeographical regions from species distributions. *Systematic*  
561 *Biology*, 66(2), 197–204. doi:10.1093/sysbio/syw087
- 562 Environmental Systems Research Institute. (2017). ArcGIS Desktop: Release 10.5. Redlands,  
563 CA: Environmental Systems Research Institute.
- 564 Ficetola, G. F., Mazel, F., & Thuiller, W. (2017). Global determinants of zoogeographical  
565 boundaries. *Nature Ecology & Evolution*, 1(4), 1–7. doi:10.1038/s41559-017-0089
- 566 González-Orozco, C. E., Laffan, S. W., Knerr, N., & Miller, J. T. (2013). A biogeographical  
567 regionalization of Australian Acacia species. *Journal of Biogeography*, 40(11), 2156–  
568 2166. doi:10.1111/jbi.12153

- 569 González-Orozco, C. E., Thornhill, A. H., Knerr, N., Laffan, S., & Miller, J. T. (2014).  
570 Biogeographical regions and phytogeography of the eucalypts. *Diversity and*  
571 *Distributions*, 20(1), 46–58. doi:10.1111/ddi.12129
- 572 Guilhaumon, F., Gimenez, O., Gaston, K. J., & Mouillot, D. (2008). Taxonomic and regional  
573 uncertainty in species-area relationships and the identification of richness hotspots.  
574 *Proceedings of the National Academy of Sciences*, 105(40), 15458–15463.  
575 doi:10.1073/pnas.0803610105
- 576 Hallgren, W., Beaumont, L., Bowness, A., Chambers, L., Graham, E., Holewa, H., . . . Weis, G.  
577 (2016). The biodiversity and climate change virtual laboratory: where ecology meets big  
578 data. *Environmental Modelling & Software*, 76, 182–186.  
579 doi:10.1016/j.envsoft.2015.10.025
- 580 Holt, B. G., Lessard, J.-P., Borregaard, M. K., Fritz, S. A., Araújo, M. B., Dimitrov, D., . . .  
581 Rahbek, C. (2013). An update of Wallace’s zoogeographic regions of the world. *Science*,  
582 339(6115), 74–78. doi:10.1126/science.1228282
- 583 Joppa, L. N., & Pfaff, A. (2009). High and far: Biases in the location of protected areas. *PLoS*  
584 *ONE*, 4(12), e8273. doi:10.1371/journal.pone.0008273
- 585 Kier, G., Mutke, J., Dinerstein, E., Ricketts, T. H., Küper, W., Kreft, H., & Barthlott, W. (2005).  
586 Global patterns of plant diversity and floristic knowledge. *Journal of Biogeography*,  
587 32(7), 1107–1116. doi:10.1111/j.1365-2699.2005.01272.x
- 588 Koleff, P., & Gaston, K. J. (2002). The relationships between local and regional species richness  
589 and spatial turnover. *Global Ecology and Biogeography*, 11(5), 363–375.  
590 doi:10.1046/j.1466-822x.2002.00302.x

- 591 Kreft, H., & Jetz, W. (2010). A framework for delineating biogeographical regions based on  
592 species distributions. *Journal of Biogeography*, 37(11), 2029–2053. doi:10.1111/j.1365-  
593 2699.2010.02375.x
- 594 Kreft, H., Sommer, J. H., & Barthlott, W. (2006). The significance of geographic range size for  
595 spatial diversity patterns in Neotropical palms. *Ecography*, 29(1), 21–30.  
596 doi:10.1111/j.2005.0906-7590.04203.x
- 597 Lamoreux, J. F., Morrison, J. C., Ricketts, T. H., Olson, D. M., Dinerstein, E., McKnight, M. W.,  
598 & Shugart, H. H. (2006). Global tests of biodiversity concordance and the importance of  
599 endemism. *Nature*, 440(7081), 212–214. doi:10.1038/nature04291
- 600 Langfelder, P., Zhang, B., & Horvath, S. (2008). Defining clusters from a hierarchical cluster  
601 tree: the Dynamic Tree Cut package for R. *Bioinformatics*, 24(5), 719–720.  
602 doi:10.1093/bioinformatics/btm563
- 603 Lê, S., Josse, J., & Husson, F. (2008). FactoMineR: an R package for multivariate analysis.  
604 *Journal of Statistical Software*, 25(1), 1–18. doi:10.18637/jss.v025.i01
- 605 Lennon, J. J., Koleff, P., Greenwood, J. J. D., & Gaston, K. J. (2001). The geographical structure  
606 of British bird distributions: diversity, spatial turnover and scale. *Journal of Animal  
607 Ecology*, 70(6), 966–979. doi:10.1046/j.0021-8790.2001.00563.x
- 608 Mackey, B. G., Berry, S. L., & Brown, T. (2008). Reconciling approaches to biogeographical  
609 regionalization: a systematic and generic framework examined with a case study of the  
610 Australian continent. *Journal of Biogeography*, 35(2), 213–229. doi:10.1111/j.1365-  
611 2699.2007.01822.x

- 612 Maestri, R., & Duarte, L. (2020). Evoregions: mapping shifts in phylogenetic turnover across  
613 biogeographic regions. *Methods in Ecology and Evolution*, *11*(12), 1652–1662.  
614 doi:10.1111/2041-210X.13492
- 615 Morrone, J. J. (2014). Parsimony analysis of endemism (PAE) revisited. *Journal of*  
616 *Biogeography*, *41*(5), 842–854. doi:10.1111/jbi.12251
- 617 Morrone, J. J. (2018). The spectre of biogeographical regionalization. *Journal of Biogeography*,  
618 *45*(2), 282–288. doi:10.1111/jbi.13135
- 619 Morrone, J. J., & Escalante, T. (2002). Parsimony analysis of endemism (PAE) of Mexican  
620 terrestrial mammals at different area units: when size matters. *Journal of Biogeography*,  
621 *29*(8), 1095–1104. doi:10.1046/j.1365-2699.2002.00753.x
- 622 Nowosad, J., & Stepinski, T. F. (2018). Spatial association between regionalizations using the  
623 information-theoretical V-measure. *International Journal of Geographical Information*  
624 *Science*, *32*(12), 2386–2401. doi:10.1080/13658816.2018.1511794
- 625 Olson, D. M., Dinerstein, E., Wikramanayake, E. D., Burgess, N. D., Powell, G. V. N.,  
626 Underwood, E. C., . . . Kassen, K. R. (2001). Terrestrial ecoregions of the world: a new  
627 map of life on Earth. *Bioscience*, *51*(11), 933–938. doi:10.1641/0006-  
628 3568(2001)051[0933:TEOTWA]2.0.CO;2
- 629 Omernik, J. M. (2004). Perspectives on the nature and definition of ecological regions.  
630 *Environmental Management*, *34*(1), S27–S38. doi:10.1007/s00267-003-5197-2
- 631 Pimm, S. L., Jenkins, C. N., Abell, R., Brooks, T. M., Gittleman, J. L., Joppa, L. N., . . . Sexton,  
632 J. O. (2014). The biodiversity of species and their rates of extinction, distribution, and  
633 protection. *Science*, *344*(6187), 1246752. doi:10.1126/science.1246752

- 634 Powney, G. D., Grenyer, R., Orme, C. D. L., Owens, I. P. F., & Meiri, S. (2010). Hot, dry and  
635 different: Australian lizard richness is unlike that of mammals, amphibians and birds.  
636 *Global Ecology and Biogeography*, *19*(3), 386–396. doi:10.1111/j.1466-  
637 8238.2009.00521.x
- 638 R Core Team. (2020). R: a language and environment for statistical computing. Vienna, Austria:  
639 R Foundation for Statistical Computing. Retrieved from <https://www.R-project.org/>
- 640 Raymond, O. L., Gallagher, R., Shaw, R., Yeates, A. N., Douth, H. F., Palfreyman, W. D., . . .  
641 Highet, L. (2012). Surface geology of Australia 1:2.5 million scale dataset 2012 edition.  
642 Retrieved from <http://dx.doi.org/10.26186/5c636e559cbe1>
- 643 Salvador, S., & Chan, P. (2004, 15–17 Nov. 2004). *Determining the number of clusters/segments*  
644 *in hierarchical clustering/segmentation algorithms*. Paper presented at the IEEE  
645 International Conference on Tools with Artificial Intelligence, Boca Raton, FL, USA.
- 646 Sexton, J. P., McIntyre, P. J., Angert, A. L., & Rice, K. J. (2009). Evolution and ecology of  
647 species range limits. *Annual Review of Ecology, Evolution, and Systematics*, *40*(1), 415–  
648 436. doi:10.1146/annurev.ecolsys.110308.120317
- 649 Simpson, G. G. (1943). Mammals and the nature of continents. *American Journal of Science*,  
650 *241*(1), 1–31. doi:10.2475/ajs.241.1.1
- 651 Smith, J. R., Letten, A. D., Ke, P.-J., Anderson, C. B., Hendershot, J. N., Dhami, M. K., . . .  
652 Daily, G. C. (2018). A global test of ecoregions. *Nature Ecology & Evolution*, *2*(12),  
653 1889–1896. doi:10.1038/s41559-018-0709-x
- 654 Stein, A., Gerstner, K., & Kreft, H. (2014). Environmental heterogeneity as a universal driver of  
655 species richness across taxa, biomes and spatial scales. *Ecology Letters*, *17*(7), 866–880.  
656 doi:10.1111/ele.12277

- 657 Thackway, R., & Cresswell, I. D. (1995). *An interim biogeographic regionalisation for*  
658 *Australia: a framework for setting priorities in the national reserves system cooperative*  
659 *program, Version 4.0*. Canberra, Australia: Australian Nature Conservation Agency.
- 660 Triantis, K. A., Guilhaumon, F., & Whittaker, R. J. (2012). The island species–area relationship:  
661 biology and statistics. *Journal of Biogeography*, *39*(2), 215–231. doi:10.1111/j.1365-  
662 2699.2011.02652.x
- 663 Vilhena, D. A., & Antonelli, A. (2015). A network approach for identifying and delimiting  
664 biogeographical regions. *Nature Communications*, *6*, 6848. doi:10.1038/ncomms7848
- 665 Wallace, A. (1876). *The geographical distribution of animals* (Vol. I & II). New York: Harper  
666 and Brothers.
- 667 Whittaker, R. J., Araújo, M. B., Jepson, P., Ladle, R. J., Watson, J. E. M., & Willis, K. J. (2005).  
668 Conservation biogeography: assessment and prospect. *Diversity and Distributions*, *11*(1),  
669 3–23. doi:10.1111/j.1366-9516.2005.00143.x
- 670 Wisz, M. S., Pottier, J., Kissling, W. D., Pellissier, L., Lenoir, J., Damgaard, C. F., . . . Svenning,  
671 J.-C. (2013). The role of biotic interactions in shaping distributions and realised  
672 assemblages of species: implications for species distribution modelling. *Biological*  
673 *Reviews*, *88*(1), 15–30. doi:10.1111/j.1469-185X.2012.00235.x
- 674

675 **Data availability statement**

676 The data on climate, elevation, vegetation, soil, lithology, and the subregions of the Interim  
677 Biogeographic Regionalization for Australia (IBRA) framework are freely available online. All  
678 the sources are cited in the manuscript. The custom computer code used for this study, as well as  
679 the processed spatial information for subregions and bioregions of the IBRA framework, the  
680 species occurrences post-equalization, the geographic-based measures of biophysical factors, the  
681 presence-absence matrices for vascular plants and four vertebrate classes (amphibians, birds,  
682 mammals, and reptiles), and the information on all fitted models for the relationship between  
683 species richness and size of IBRA geographic operational units are freely available at: [DOI will  
684 be provided should the manuscript be accepted for publication].

685

686 **Appendices**

687 Appendix S1. Additional information on data collection and processing

688 Appendix S2. Selection of species-area relationship (SAR) mathematical function

689 Appendix S3. Selection of biophysical factors and IBRA unit of analysis

690 Appendix S4. Selection of optimal discrete-species clusters

691 Figs. S5 to S8

692 List of references

693

694 **Table 1. Best models of the species-area relationship (SAR).**

Taxa	Clusters (9)			Clusters (7)						Clusters (5)								
	'L' technique			WWF's Biomes			'Static' technique			'Dynamic' technique			'Static' technique			'Dynamic' technique		
	n	z	R <sup>2</sup>	n	z	R <sup>2</sup>	n	z	R <sup>2</sup>	n	z	R <sup>2</sup>	n	z	R <sup>2</sup>	n	z	R <sup>2</sup>
Birds	14	0.10	0.43	23	0.14	0.47	14	0.10	0.43	14	0.10	0.43	14	0.10	0.43	14	0.10	0.43
Mammals	5	0.62	1.00	4	0.03	0.01	20	0.50	0.85	14	0.13	0.26	14	0.13	0.26	14	0.13	0.26
Herpetofauna	14	0.18	0.64	23	0.21	0.73	14	0.18	0.64	14	0.18	0.64	14	0.18	0.64	14	0.18	0.64
Vertebrate	14	0.13	0.52	23	0.17	0.56	14	0.13	0.52	14	0.13	0.52	14	0.13	0.52	14	0.13	0.52
Vascular plants	14	0.13	0.20	20	0.15	0.17	14	0.13	0.20	6	0.40	0.89	14	0.13	0.20	14	0.13	0.20

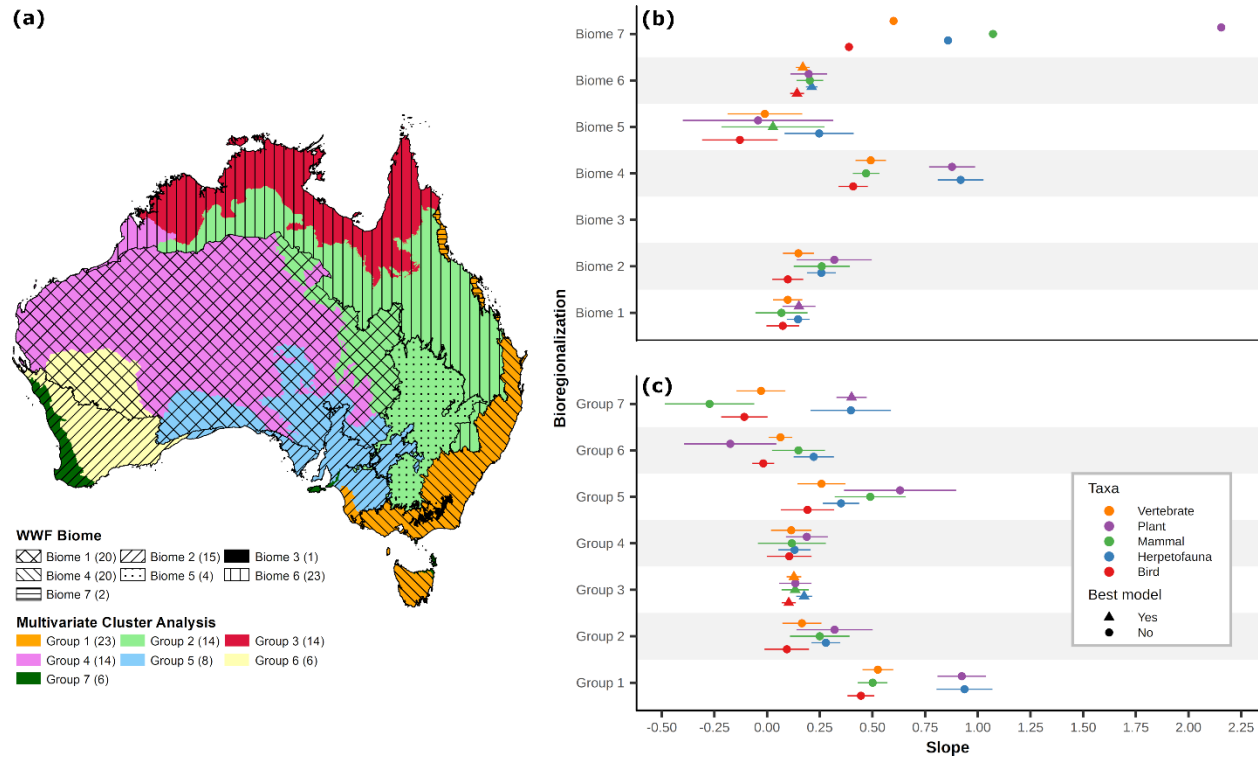
695 Note: SAR in a log-log implementation of the power function for the partition of bioregions of the Interim Biogeographic  
696 Regionalization for Australia (IBRA) framework based on the World Wildlife Fund's (WWF) expert classification of Australia into  
697 seven biomes (Dinerstein et al., 2017), and for the partition of the hierarchical cluster analysis of the principal components of  
698 Australia's biophysical dissimilarity space of IBRA bioregions into a pair of five and another of seven groups using the 'static' and  
699 'dynamic' tree-cutting techniques, and nine groups defined by the 'L' technique. Best model corresponds to the most parsimonious  
700 SAR model for birds, mammals, herpetofauna, vertebrates, and vascular plants based on the Akaike's Information Criteria (i.e., best-  
701 selected SAR models). The number of bioregions in groups/biomes (n), the slope of SAR (z), and the amount of variance explained by  
702 SAR model (R<sup>2</sup>) are reported. The color of cells match those of the discrete spatial clusters from which these values were estimated  
703 (see Fig. S3.4 a and Fig. S5.5 in Supporting Information for WWF biomes and the best-selected SAR models, respectively). The  
704 number of partitions of prominent discrete clusters in the PC-biophysical space are shown between parenthesis.



706 **Table 2. Mutual information on species richness and composition across pairwise-map**  
 707 **comparisons.**

Optimal clusters' maps	WWF's bioregionalization		Algorithm-driven IBRA framework	
	Biome's map	Ecoregion's map	Discrete-vertebrate's map	Discrete-plant's map
	V-measure (%)	V-measure (%)	V-measure (%)	V-measure (%)
Biophysical heterogeneity				
Species-area relationship	57.7	61.9	53.5	54.8
Within-site diversity (Alpha)				
Birds	53.3	76.8	83.0	81.9
Mammals	55.9	69.4	90.5	82.2
Herpetofauna	53.1	77.6	80.2	83.7
Vertebrates	56.4	70.4	100.0	86.5
Vascular Plants	58.0	77.8	86.5	100.0
Between-site diversity (Beta)				
Birds	50.0	45.7	68.5	60.9
Mammals	54.8	67.5	87.3	79.0
Herpetofauna	54.3	58.3	70.9	71.1
Vertebrates	56.4	70.4	–	86.5
Vascular Plants	58.0	77.8	86.5	–

708 Note: Within an information-theoretical analytical framework (Nowosad & Stepinski, 2018), the  
 709 spatial association between the World Wildlife Fund's (WWF) biomes and ecoregions of  
 710 Australia (Dinerstein et al., 2017) and our suite of eleven optimal discrete-species cluster maps  
 711 (Fig. S8.8) were computed to quantify the overall degree of similarity between these maps. This  
 712 global index, called the V-measure, changes from 0 for no spatial association between two maps  
 713 to 1 when perfect. '–' indicates that degree of spatial association was not computed because  
 714 pairwise map comparison would have been with the optimal discrete-species clusters itself.  
 715



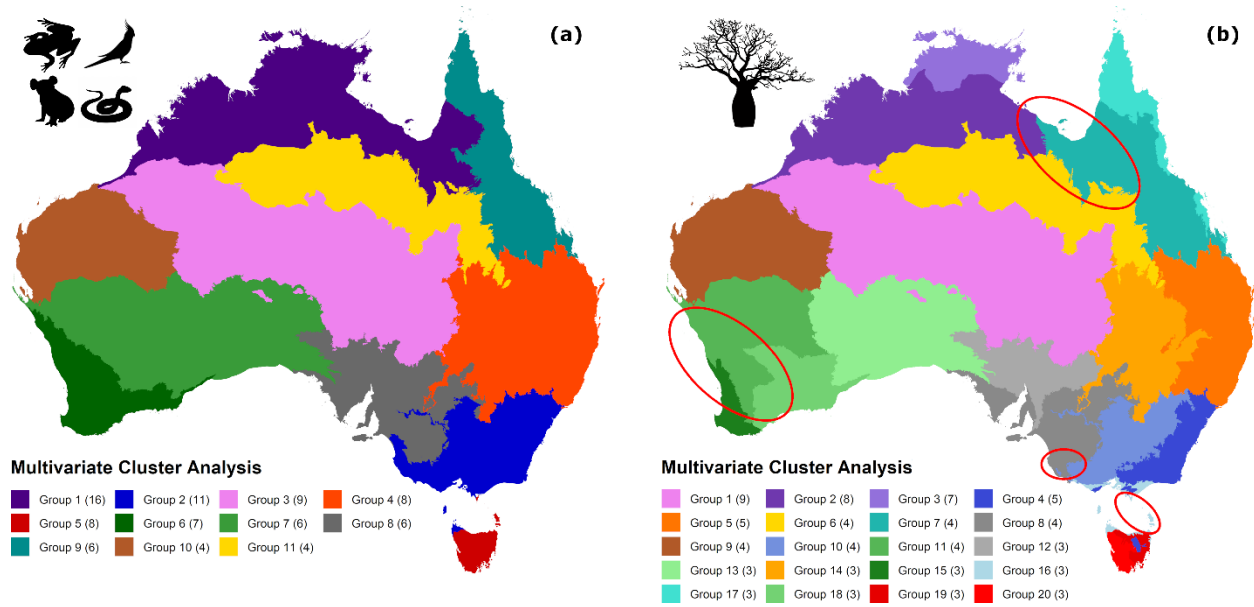
716

717 **FIGURE 1 Relationship between biophysical factors and species richness.** (a) Overlap of  
 718 distinct spatial clusters of Australia's bioregions based on the World Wildlife Fund's (WWF)  
 719 biomes (Dinerstein et al., 2017), and the optimum partition for changes in species richness per  
 720 standard area (i.e., optimal partition for SAR), with the number of bioregions in biomes/groups  
 721 found between parentheses. Species-area relationship (SAR) in a log-log implementation of the  
 722 power function for the aggregations of bioregions according to (b) WWF's biomes and (c)  
 723 groups of the optimal partition for SAR. Point represents the slope (z-value) of the shape of  
 724 SAR, and line shows the standard error of the regression line. Due to Biome 3 including only one  
 725 bioregion, neither slope nor the standard error could be computed, whereas only the slope could  
 726 be estimated for Biome 7. Biome 1 = Deserts & Xeric Shrublands; Biome 2 = Mediterranean  
 727 Forests, Woodlands, Scrub; Biome 3 = Montane Grasslands & Shrublands; Biome 4 =  
 728 Temperate Broadleaf & Mixed Forests; Biome 5 = Temperate Grasslands, Savannas &

729 Shrublands; Biome 6 = Tropical & Subtropical Grasslands, Savannas & Shrublands; and Biome

730 7 = Tropical & Subtropical Moist Broadleaf Forests.

731



732

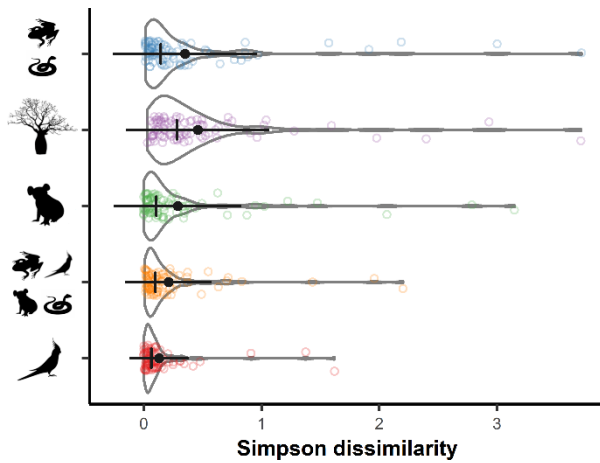
733 **FIGURE 2 Spatial configuration of the optimal clusters for species diversity. (a) Vertebrate**

734 and (b) vascular plant species. Ellipses highlight the areas where plant clusters do not collapse

735 within vertebrates. The number of bioregions in distinct spatial clusters are found between

736 parentheses.

737



738

739 **FIGURE 3 Height of nodes in five dendrograms.** Kernel density plot shows the distribution of  
740 the nodes' height for the spatial turnover of bird, mammal, herpetofauna, vertebrate, and vascular  
741 plant species. Color circle corresponds to the height of nodes as clusters are merged. Point and  
742 crossbar respectively represent the mean and median dissimilarity among nodes' height, and dark  
743 line shows the standard deviation.

744