

1 **Running head:** Counting niches in a Neotropical forest

2 **Title:** Counting niches: Abundance-by-trait patterns reveal niche partitioning in a Neotropical  
3 forest

4 **Authors:** Rafael D’Andrea<sup>1,2,7</sup>, John Guittar<sup>1,3,8</sup>, James P. O’Dwyer<sup>2</sup>, Hector Figueroa<sup>1</sup>, S. J.  
5 Wright<sup>4,9</sup>, Richard Condit<sup>5,6</sup> and Annette Ostling<sup>1</sup>

6  
7 <sup>1</sup>Dept. Ecology & Evolutionary Biology, University of Michigan, Ann Arbor, MI, USA

8 <sup>2</sup>Dept. Plant Biology, University of Illinois, Urbana-Champaign, IL, USA

9 <sup>3</sup>Dept. Plant Biology, Michigan State University, Hickory Corners, MI, USA

10 <sup>4</sup>Smithsonian Tropical Research Institute, Apartado 0843–03092, Balboa, Republic of Panama

11 <sup>5</sup>Field Museum of Natural History, 1400 S. Lake Shore Dr., Chicago, IL 60605, USA

12 <sup>6</sup>Morton Arboretum, 4100 Illinois Rte. 53, Lisle, IL 60532, USA

13

14 <sup>7</sup> Corresponding author. E-mail: [rdandrea@illinois.edu](mailto:rdandrea@illinois.edu); ORCID 0000-0001-9687-6178

15 <sup>8</sup> ORCID 0000-0002-8897-8841

16 <sup>9</sup> ORCID 0000-0003-4260-5676

17 **Abstract**

18 Tropical forests challenge us to understand biodiversity, as numerous seemingly similar species  
19 persist on only a handful of shared resources. Recent ecological theory posits that biodiversity is  
20 sustained by a combination of species differences reducing interspecific competition and species  
21 similarities increasing time to competitive exclusion. Together, these mechanisms  
22 counterintuitively predict that competing species should cluster by traits, in contrast with  
23 traditional expectations of trait overdispersion. Here, we show for the first time that trees in a  
24 tropical forest exhibit a clustering pattern. In a 50 ha. plot on Barro Colorado Island in Panama,  
25 species abundances exhibit clusters in two traits connected to light capture strategy, suggesting  
26 that competition for light structures community composition. Notably, we find four clusters by  
27 maximum height, quantitatively supporting the classical grouping of Neotropical woody plants  
28 into shrubs, understory, midstory, and canopy layers.

29

This is the author manuscript accepted for publication and has undergone full peer review but has not been through the copyediting, typesetting, pagination and proofreading process, which may lead to differences between this version and the [Version of Record](#). Please cite this article as [doi: 10.1002/ECY.3019](https://doi.org/10.1002/ECY.3019)

This article is protected by copyright. All rights reserved

30 **Keywords**

31 Niche differentiation, Trait-based clustering, Self-organized similarity, Emergent neutrality,  
32 Competition, Tropical forests, Barro Colorado Island, Community structure.

33

34 **Introduction**

35 A basic principle of community ecology is that coexistence requires niche differentiation, i.e.  
36 species differences that stabilize communities by reducing interspecific competition. However,  
37 the high diversity and seemingly continuous phenotypic variation of trees in tropical forests  
38 raises questions about the importance of niche differentiation as a contributor to species co-  
39 occurrence and patterns of diversity, relative to other community assembly processes such as  
40 dispersal and ecological drift (Hubbell 2001, Levine and Murrell 2003, Mouquet and Loreau  
41 2003, Chase and Myers 2011). Indeed, neutral models reflecting chance events provide a good fit  
42 to the distribution of species abundances in tropical forests (Hubbell 2001, Volkov et al. 2003),  
43 although other studies have found evidence of niche differentiation (Kraft and Ackerly 2010,  
44 Götzenberger et al. 2012). One likely reason for these mixed results is a lack of model-validated  
45 quantitative predictions for how niche-based competition shapes community trait structure.

46 Classical theory posits that only species whose trait differences are large enough to  
47 sufficiently lower interspecific competition will stably coexist. Hence, attempts to find evidence  
48 of niche differentiation in nature typically look for greater-than-chance differences in trait values  
49 among species (Chase and Leibold 2003, Kraft et al. 2008, Lake and Ostling 2009, Kraft and  
50 Ackerly 2010, Götzenberger et al. 2012, D'Andrea and Ostling 2016). However, competition can  
51 produce counterintuitive patterns. If two resident species are not distant enough from each other  
52 in trait space to allow a third species with an intermediate trait value to invade, the potential  
53 invader has higher fitness if it is similar to one of the residents than if it is maximally different  
54 from both (MacArthur and Levins 1967). This is because the increase in competition with the  
55 close resident is more than compensated by the decrease in competition with the other resident.  
56 When this phenomenon is extrapolated to multispecies communities, species are competitively  
57 excluded at rates inversely related to trait distance from the nearest trait optimum (D'Andrea and  
58 Ostling 2017). As competition among species proceeds, a transient pattern arises characterized  
59 by distinct clusters of species with similar traits, centered around locally optimal ecological  
60 strategies (Scheffer and van Nes 2006). While competition among species in each cluster

61 eventually leads to exclusion of all but the most competitive species, the similar strategies among  
62 the species renders their dynamics near-neutral; indeed, clusters can persist indefinitely under  
63 restorative forces such as immigration (D’Andrea et al. 2019), environmental fluctuations  
64 (Sakavara et al. 2018), and specialist enemies (Scheffer and van Nes 2006, Barabás et al. 2013).

65 In other words, while classical literature predicts that niche partitioning will lead to trait  
66 overdispersion, more recent literature suggests trait clusters rather than overdispersion as  
67 evidence of niche partitioning. We note that the same models that predict overdispersion in  
68 stable equilibrium will predict clusters as transients (Scheffer and van Nes 2006, D’Andrea et al.  
69 2019). Furthermore, it has been shown that several ecological forces can maintain the clusters  
70 indefinitely as a stationary state. Those forces include immigration (D’Andrea et al. 2019),  
71 periodic environments (Sakavara et al. 2018), and “hidden niches”, i.e. independent niche axes  
72 providing extra intraspecific regulation such as specialist enemies (Scheffer and van Nes 2006,  
73 Barabás et al. 2013). Given the widespread influence of these forces in nature, we expect clusters  
74 to be a more general signature of niche differentiation than the stricter circumstances under  
75 which overdispersion might arise.

76 Clustering by organismal size has been reported in animal (Scheffer and van Nes 2006)  
77 and phytoplankton communities (Sakavara et al. 2018), but is not known to occur in tropical  
78 forests – a focal point for the development and tests of coexistence theory. Furthermore, previous  
79 efforts to detect species clusters have typically ignored species abundances and/or relied on  
80 arbitrary binning of trait space (Scheffer and van Nes 2006, Yan et al. 2012) or temporal data to  
81 distinguish between occasional and permanent species (Vergnon et al. 2009). However, recent  
82 work has proposed a parameter-free method which uses trait and abundance data to identify and  
83 count the clusters, which was validated using population dynamic models of competition with  
84 immigration (D’Andrea et al. 2019).

85 Here we use this approach to test whether high abundance values tend to be clustered  
86 with respect to trait values in the 50-ha Forest Dynamics Plot on Barro Colorado Island, Panama  
87 (Condit 1998, Hubbell et al. 1999, Condit et al. 2012a) (BCI). We use four morphological traits:  
88 seed mass, maximum tree height, wood density, and leaf mass per area (LMA). Seed mass is  
89 thought to reflect strategies along a tolerance-fecundity tradeoff axis (Muller-Landau 2010),  
90 whereby large-seeded species specialize in surviving stressful environmental conditions at the  
91 cost of low fecundity and vice-versa. Maximum tree height, wood density, and LMA are

92 associated with competition for light (Kunstler et al. 2016), and as such may reflect plant  
93 strategies related to partitioning light availability gradients both vertically (i.e. among forest  
94 canopy layers) and horizontally (i.e. across forest patches at different stages of succession). Such  
95 partitioning of light access is central to the forest architecture hypothesis (Kohyama 1993) and  
96 the related successional niche hypothesis (Grime 1979, Tilman 1988).

97 Trait pattern indicative of a niche mechanism will likely only become evident at a spatial  
98 scale above that of any spatial heterogeneity involved in the mechanisms (Kraft and Ackerly  
99 2010, D'Andrea and Ostling 2016). Therefore, we look for signatures of niche partitioning at the  
100 whole plot (50 ha) scale, large enough to encompass the spatial heterogeneity in patch-age since  
101 disturbance involved in these niche mechanisms (Hubbell et al. 1999). Because immigration can  
102 reinforce local clustering if surrounding forests have species clustered around the same trait-  
103 based niche optima, or mask local clustering if surrounding forests have species clustered around  
104 different trait-based niche optima, we compare results at the plot-scale with results from a  
105 combined pool of 40 census plots within 30km of BCI (Condit et al. 2013, 2016) (Fig. 1).

106

## 107 **Methods**

108 The Barro Colorado Island Forest Dynamics Plot is a 1,000-by-500 meter tract of tropical forest  
109 on Barro Colorado Island in Panama (Hubbell et al. 1999, Condit et al. 2012b) (Fig. 1). The plot  
110 has been censused for stems above 1 cm diameter at breast height (DBH) in 1982, 1985 and  
111 every five years through 2015. The plot contains approximately 210,000 stems with  $DBH \geq 1$   
112 cm, and about 300 tree species. Results shown in the main text pertain to the 2000 census;  
113 however, results are consistent across all available censuses (Appendix S1: Table S1, Figs. S3-  
114 S6).

115 We used four traits related to plant architecture, leaf structure, and seed size. Trait data,  
116 drawn from Wright et al. (2010), were based on the following protocols: 1) Maximum height  
117 was estimated as the mean height of the six largest (by DBH) individuals, measured using a  
118 rangefinder. 2) Wood density ( $g/cm^3$ ) was measured after drying at  $60^\circ C$ . 3) Leaf mass per area  
119 ( $g/m^2$ ) was measured on leaf laminae, excluding petioles and petiolules, averaged across six  
120 individuals. For species that reach the canopy we used leaves exposed to direct sunlight --- i.e.  
121 sun leaves. For all others we used shade leaves collected from the crowns of the tallest  
122 individuals in the 50 ha plot. 4) Seed mass (g) refers to endosperm plus embryo only and was

123 measured after drying at 60°C. Traits with skewed distributions (LMA and seed mass) were log-  
124 transformed to reduce the skew. Palms were excluded from wood and leaf trait analyses, due to  
125 their substantial differences in physiology from dicots.

126 We restricted our analysis to reproductive individuals (adults), since population dynamic  
127 models producing emergent clusters do not consider population structure. We consider  
128 individuals to be reproductive when their DBH exceeds half the maximum DBH observed for its  
129 species, a relationship that explains over 80% of interspecific variation in reproductive size  
130 thresholds (see Visser et al. 2016: Fig. S9). Maximum DBH was estimated as the average of the  
131 six largest individuals in the BCI plot in 2005, and an additional 150 ha of mapped tree plots  
132 located within 30 km and mostly within 10 km of BCI (Visser et al. 2016).

133 The regional pool consists of 40 1-ha sampling sites within 30km of BCI (Condit et al.  
134 2013, 2016). In 38 of these sites, individuals with maximum DBH > 1cm were censused in 40m-  
135 by-40m internal subplots, while individuals with DBH > 10cm were censused in the entire plot.  
136 In the remaining two plots, all individuals with maximum DBH > 10cm were censused in the  
137 entire plot. Hence, we standardized the counts by the respective sampling area to estimate  
138 densities in these sites. Of the circa 850 species contained in these sites, trait data were available  
139 for those also found on BCI (see Table 1, caption to Fig. 1). While lack of pattern among this  
140 subset of species in the regional pool does not rule out clustering among all species, it shows that  
141 the pattern on BCI is not simply mirroring regional-scale pattern inherited via immigration.

142 Our clustering metric has two foundations: the k-means clustering algorithm (MacQueen  
143 1967), an optimization procedure which assigns species to clusters by minimizing the average  
144 trait distance between individuals in the same cluster; and the gap method (Tibshirani et al.  
145 2001), which compares the observed data against null distributions, and selects the number of  
146 clusters that maximizes the difference between within-cluster trait dispersion in the observed and  
147 null communities. Specifically, given a candidate number of clusters,  $k$ , the k-means algorithm  
148 finds the arrangement of species into  $k$  clusters that minimizes within-cluster trait dispersion:  $D_k$

149  $= \frac{1}{k} \sum_C \sum_{i,j \in C} n_i n_j d_{ij}^2$ , where  $C$  refers to a cluster ( $1 \leq C \leq k$ ),  $n_i$  is the abundance of species  $i$ ,

150 and  $d_{ij}$  is the trait distance between species  $i$  and  $j$ . The algorithm starts with randomly chosen  
151 trait values in the local community as possible cluster centers, then puts species into the cluster  
152 whose center is the closest to them, then recalculates cluster centers, and so on until the  
153 algorithm converges or changes in  $D_k$  fall below a specified threshold. Because the result can

154 depend on the starting set of cluster centers, the procedure is repeated with different starting sets,  
155 and the final cluster arrangement with the lowest  $D_k$  across all sets is selected (note that the  
156 clusters need not be equally spaced). We use enough starting sets that larger numbers do not  
157 lower within-cluster dispersion further (typically between 1,000 and 10,000).

158 We then apply this k-means algorithm to 1,000 null communities, and define the gap  
159 index for  $k$  clusters as  $G_k = \frac{1}{1000} \sum_{\text{nulls}} \log(D_k^*) - \log(D_k)$ , where  $D_k^*$  is the cluster dispersion in  
160 a given null community. This index measures the degree to which trait dispersion  $D_k$  in the  
161 observed community differs from its null expectation for  $k$  clusters (see Tibshirani et al. 2001,  
162 D'Andrea et al. 2019 for details). Notice that tighter than expected clustering in the data for  $k$   
163 clusters will lead to a positive gap  $G_k$ , and vice-versa. Finally, we find the number of clusters that  
164 maximizes the gap index. We call this maximal value -- corresponding to the peak in the gap  
165 curves in Figs. 2, 3 -- the gap statistic,  $G = \max(G_k)$ , and the value of  $k$  at which it occurs is the  
166 estimated number of clusters  $K$ . We note that Tibshirani et al. 2001 refer to the gap index  $G_k$  as  
167 the gap statistic, and estimate the number of clusters using a slightly different approach. Later  
168 studies found deficiencies with that approach (Dudoit and Fridlyand 2002, Sugar and James  
169 2003), and modifications have since been proposed (Yan and Ye 2007). Our method, designed to  
170 find whether high abundances tend to cluster with respect to trait values, was benchmarked using  
171 simulation tests where the number of clusters is known a priori (see D'Andrea et al. 2019).

172 We obtain a null distribution of gap statistics by performing the above routine on each of  
173 the null communities, and from this distribution we extract significance (p-value) and  
174 standardized effect size (z-score). The z-score is  $Z = (G - \mu)/\sigma$ , where  $\mu = \frac{1}{1000} \sum_{\text{nulls}} G_{\text{null}}$  is the  
175 mean of the null gap statistics and  $\sigma^2 = \frac{1}{1000} \sum_{\text{nulls}} (G_{\text{null}} - \mu)^2$  is the variance. The p-value is the  
176 proportion of null communities with a higher gap statistic than the observed community,  $P =$   
177  $\frac{1}{1000} \sum_{\text{nulls}} I(G_{\text{null}} > G)$ , where the indicator function  $I$  is 1 if its argument is true, and zero  
178 otherwise.

179 Our null communities contain the same set of observed trait values and abundances, with  
180 abundances randomized across the traits. We therefore test for a nonrandom association between  
181 traits and abundances, as opposed to a nonrandom set of traits or abundances. While a  
182 nonrandom trait distribution has interest in its own merit, it could be arguably caused by forces

183 beyond local competitive interactions. Local abundances reflect the outcome of local competitive  
184 interactions and environmental filtering to a higher degree than a presence-absence list of locally  
185 occurring species, as the latter is strongly influenced by dispersal from the regional species pool.  
186 Therefore, trait clustering among local species without regard for abundance could simply reflect  
187 the clustered trait distribution of the regional pool, leaving it unclear whether local competition is  
188 the cause rather than evolutionary history.

## 189 190 **Results**

191 We found significant clustering of species abundances by trait values on BCI based on species  
192 maximum height (Fig. 2A; Fig. 2E), with the gap statistic exceeding null expectations by 2.3  
193 standard deviations (Table 1). There were four height-based clusters, falling at approximately  
194 eight-meter intervals. Intriguingly, the pattern aligns well with the historical division of humid  
195 Neotropical forests into four strata of about ten meters each: shrubs, understory, midstory, and  
196 canopy (Paul W Richards 1952, Terborgh 1985, Condit et al. 1995) (Fig. 3, Appendix S1: Fig  
197 S1). We also found a significant pattern of five abundance-by-trait clusters based on wood  
198 density on BCI (Fig. 2F), exceeding null expectations by 2.1 standard deviations (Table 1).  
199 Notably, all species on BCI were also significantly clustered around a single wood density  
200 optimum. This occurred because species with intermediate wood densities were more common  
201 than those with extremely low or high wood densities (Fig. 2B), potentially indicating  
202 physiological costs or environmental filters against extreme trait values. We found no evidence  
203 for clustering based on LMA or seed mass.

204 The similar number of clusters by maximum height and wood density led us to ask if the  
205 two trait-based patterns reflected the same underlying niche axis. Even though maximum height  
206 and wood density did not correlate significantly (Pearson's  $\rho = -0.10$ , p-value = 0.06),  
207 maximum height clusters correlated with wood density clusters ( $\chi^2$ -test p-value = 0.02,  
208 Cramér's  $V = 0.20$ ), supporting this tentative hypothesis.

209 We found significant abundance-by-trait clustering by maximum height at the regional  
210 scale, with the highest gap index occurring at  $K = 1$  cluster, and a secondary but still significant  
211 peak at  $K = 4$  clusters (Appendix S1: Fig. S2A, B). The existence of four height-based clusters at  
212 both local and regional scales suggests a consistent organization of species into the same four  
213 niches across many tropical forest communities. However, a key difference between local and

214 regional height-based patterns is the presence of a significant single cluster at the regional scale.  
215 This single regional peak reflects a unimodal abundance trend along the maximum height axis,  
216 with trees in the second height cluster being consistently the most abundant in their communities.  
217 This abundance trend may result from an interplay of higher light incidence at larger heights and  
218 the potential for a larger number of smaller-sized trees in a given area.

219 As for lack of clustering by wood density in the regional pool, we hypothesize that  
220 differences in environmental or ecological conditions across the region may have created  
221 different trait optima (cluster centroid values) in different communities, leading to the erosion of  
222 cluster structure when the communities are combined.

223

## 224 **Discussion**

225 Our results underscore competitive partitioning of light as a major driver of tropical  
226 forest community structure. Specifically, we found cluster-based evidence for niche partitioning  
227 by maximum height, a trait associated with light capture, and wood density, a trait associated  
228 with shade tolerance (Wright et al. 2010). Conventional descriptions of tropical forest structure  
229 have long included references to different strata (sub-canopy, canopy, etc.), but to our knowledge  
230 there has never been any quantitative empirical support for their existence.

231 Testing for trait-based clustering is an appealing approach for the detection of niche  
232 partitioning because it conceptually unites the notions of stable species coexistence via  
233 differences and transient species coexistence via similarity (Holt 2006). Perhaps  
234 counterintuitively, it is precisely because species with similar strategies compete more strongly,  
235 and experience similar competitive effects from the rest of the community, that clusters emerge  
236 (D'Andrea et al. 2018). In other words, clusters reflect the simultaneous advantages of differing  
237 from others while being similar to those favored strategies.

238 We note there are special cases where competition could allow for coexistence of species  
239 with any trait value (Leimar et al. 2013). Specifically, the function specifying the relationship  
240 between competition and species traits (often referred to as the competition kernel) must have a  
241 strictly positive Fourier transform, which is not a general property of competition based on trait  
242 differences. For example, in the family of exponential kernels, only those with exponent  $\leq 2$  will  
243 satisfy the condition, and small perturbations of these kernels will break the property and lead to  
244 cluster formation. However, in the absence of empirical evidence that the competition kernels on



245 BCI possess those special properties, a more general expectation is for cluster formation rather  
246 than a continuum of traits.

247 The mechanistic origin of clustering may involve environmental selection favoring  
248 certain traits over others. For example, according to Terborgh's sunfleck model (Terborgh 1985),  
249 light availability below the canopy may peak at various vertical strata due to intersection of light  
250 cones originating from gaps in the canopy. Latitudinal parameters which affect the angle of  
251 incidence, combined with the shape of tree canopies in tropical forests, predict three such strata  
252 below the canopy, for a total of four optimal tree heights based on light availability.

253 Alternatively, the height-based clusters we identified may arise through tradeoffs between the  
254 light-related advantages of being tall and shading others, and the budgetary constraints that  
255 investment in height impinges on reproductive growth, such as through losses in recruitment or  
256 shade tolerance (Kohyama 1993, Kohyama and Takada 2009). These tradeoffs may result in an  
257 effective relationship between interspecific competition and height similarity. In such a case, the  
258 number of clusters would relate to how quickly competition drops with increasing separation in  
259 maximum height. Regardless of the mechanism, the height-based clusters reflect the existence of  
260 multiple light-related niches, as well as stabilization via light-related niche partitioning.

261 The existence of four clusters by height suggests that only four species can coexist on  
262 BCI based on differences in height alone, but it does not mean that only four species can stably  
263 coexist on BCI. Additional niche axes may support higher coexistence, and in fact may be  
264 responsible for the maintenance of the observed clusters. Barabás et al. (2013) pointed out that  
265 additional niche axes, such as susceptibility to different herbivores and pathogens, are required  
266 for the permanence of clusters in a closed community subject to no temporal variation in the  
267 environment. Though recent empirical studies report a good deal of overlap in enemy host use,  
268 models suggest that small differences in susceptibility to enemies between tree species may still  
269 contribute significantly to diversity maintenance (Sedio and Ostling 2013). Hence, they likely  
270 also contribute to the permanence of clusters along the trait axes we report in our study. Another  
271 example additional niche axis is the successional niche (Tilman 1988): trees sharing the same  
272 height niche may niche-differentiate by shade tolerance. This could be reflected in the  
273 distribution of wood density values and/or other traits related to successional strategy.

274 Wood density in Barro Colorado Island tree species has been found to be a good  
275 predictor of species niche strategies along a tradeoff axis between survival under stressful

276 conditions and rapid growth under optimal conditions (Wright et al. 2010). Therefore, the  
277 clustering we found in wood density suggests niche partitioning along a growth-mortality  
278 tradeoff axis, i.e. the successional niche (Tilman 1988). The precise positions of the clusters  
279 could have many interacting causes: the available pool of traits, external filtering from the  
280 environment, resource availability, and the shape of the competition kernel. In the case of wood  
281 density, given its connection to the successional niche, the distribution of light gaps in the forest  
282 may possibly determine the positions of the niches. The number of clusters may also be  
283 influenced by the rate at which competition drops between species of disparate wood density: the  
284 faster the drop, the more niches are allowed.

285         The wood density results are not fully independent from our maximum height results, as  
286 indicated by the correlation between height and wood density clusters. To fully delineate the  
287 niche space driving pattern on BCI, a multivariate trait treatment would be required. However,  
288 such an approach raises questions of how to best convert trait values along different trait axis  
289 into a good predictor of competition. Moreover, traits could combine in different ways to form  
290 multidimensional niche space (D'Andrea et al. 2018); for example, Euclidean distance may be a  
291 poor predictor of competition, and therefore not be useful for detecting clustering. Determining  
292 how competition maps to multivariate trait space requires mechanistic niche models going  
293 beyond the existing theoretical literature, which typically focuses on contributions of individual  
294 traits.

295         The absence of clustering in LMA and seed mass does not preclude a role for these traits  
296 in niche partitioning on BCI. D'Andrea et al. (2019) showed that a loose connection between  
297 trait value and niche strategy can mask clustering along the relevant niche axis. Seed mass is  
298 linked to dispersal and recruitment ability (Muller-Landau et al. 2008), and is therefore a  
299 candidate for clustering along a niche axis characterizing a competition-colonization tradeoff –  
300 hence its inclusion in our study. However, seed mass alone may be a poor predictor of a species'  
301 strategy along this niche axis, due for example to the effects of other traits such as drought  
302 tolerance and dispersal mode. LMA is theoretically connected to niche strategies (Wright et al.  
303 2010), but there is substantial ontogenetic variation in its predictive ability of growth, with the  
304 relationship being stronger in the seedling stage. Eventually, other aspects of physiology become  
305 limiting as the plant grows. Hence, LMA may be a noisy predictor of the niche strategy (Wright

306 et al. 2010). We conclude that the lack of clustering in these traits may be due to their loose  
307 connection to niche strategies rather than no involvement in niche partitioning.

308 Our study provides a new line of evidence that niche differentiation shapes tropical  
309 forests, notwithstanding neutral theory's success at fitting species abundance distributions  
310 (Rosindell et al. 2012). Yet, our results go further by identifying specific numbers of niches and  
311 their associated trait optima. Future extensions of clustering analysis could look at other traits  
312 related to defense chemistry or abiotic preferences, further delineating the niches contributing to  
313 coexistence in tropical forests.

314

### 315 **Acknowledgments**

316 RD and AO conceptualized the study; SJW and RC provided the data; RD, JG, HF, and AO  
317 analyzed the data; RD and JG wrote the manuscript; RD, JG, and AO contributed to revisions of  
318 the manuscript. RD and JG contributed equally to the execution of this manuscript. We thank C.  
319 Klausmeier for helpful discussions. BCI data were from the forest dynamics research project,  
320 founded by S.P. Hubbell and R.B. Foster and now managed by R. Condit, S. Lao, and R. Perez  
321 under CTFS and the Smithsonian Tropical Research in Panama. Hundreds of field workers also  
322 contributed. Many organizations provided funding, principally the U.S. National Science  
323 Foundation (NSF). J.O.D. acknowledges the Simons Foundation Grant #376199 and McDonnell  
324 Foundation Grant #220020439. A.O. acknowledges the NSF Grant #1038678, Niche versus  
325 neutral structure in populations and communities funded by Advancing Theory in Biology.

326

### 327 **Literature cited**

328 Barabás, G., R. D'Andrea, R. Rael, G. Meszéna, and A. M. Ostling. 2013. Emergent Neutrality  
329 or Hidden Niches? *Oikos* 122:1565–1572.

330 Chase, J. M., and M. Leibold. 2003. *Ecological Niches: Linking Classical and Contemporary*  
331 *Approaches*. The University of Chicago Press.

332 Chase, J. M., and J. a. Myers. 2011. Disentangling the importance of ecological niches from  
333 stochastic processes across scales. *Philosophical Transactions of the Royal Society B:*  
334 *Biological Sciences* 366:2351–2363.

335 Condit, R. 1998. *Tropical forest census plots: methods and results from Barro Colorado Island,*  
336 *Panama and a comparison with other plots*. Springer Science & Business Media.

- 337 Condit, R., R. A. Chisholm, and M. O. Hunter. 2012a. Thirty Years of Forest Census at Barro  
338 Colorado and the Importance of Immigration in Maintaining Diversity. *PLoS ONE* 7:1–6.
- 339 Condit, R., B. M. J. Engelbrecht, D. Pino, R. Pérez, and B. L. Turner. 2013. Species distributions  
340 in response to individual soil nutrients and seasonal drought across a community of tropical  
341 trees. *Proceedings of the National Academy of Sciences of the United States of America*  
342 110:5064–8.
- 343 Condit, R., S. Lao, R. Perez, S. B. Dolins, R. Foster, and S. Hubbell. 2012b. [Dataset:] Barro  
344 Colorado Forest Census Plot Data (Version 2012).
- 345 Condit, R. S., S. P. Hubbell, and R. B. Foster. 1995. Mortality rates of 205 Neotropical tree and  
346 shrub species and the impact of a severe drought. *Ecological Monographs* 65:419–439.
- 347 Condit, R. S., R. Pérez, S. Aguilar, and S. Lao. 2016. Data from Tree Censuses and Inventories  
348 in Panama. [repository.si.edu/handle/10088/28866](https://repository.si.edu/handle/10088/28866).
- 349 D’Andrea, R., and A. Ostling. 2017. Biodiversity maintenance may be lower under partial niche  
350 differentiation than under neutrality. *Ecology* 98:3211–3218.
- 351 D’Andrea, R., and A. M. Ostling. 2016. Challenges in linking trait patterns to niche  
352 differentiation. *Oikos* 125:1369–1385.
- 353 D’Andrea, R., A. Ostling, and J. P. O’Dwyer. 2018. Translucent windows: How uncertainty in  
354 competitive interactions impacts detection of community pattern. *Ecology Letters*:826–835.
- 355 D’Andrea, R., M. Riolo, and A. M. Ostling. 2019. Generalizing clusters of similar species as a  
356 signature of coexistence under competition. *PLoS Computational Biology* 15:1–19.
- 357 Dudoit, S., and J. Fridlyand. 2002. A prediction-based resampling method for estimating the  
358 number of clusters in a dataset. *Genome biology* 3:1–21.
- 359 Götzenberger, L., F. de Bello, K. A. Bråthen, J. Davison, A. Dubuis, A. Guisan, J. Lepš, R.  
360 Lindborg, M. Moora, M. Pärtel, L. Pellissier, J. Pottier, P. Vittoz, K. Zobel, and M. Zobel.  
361 2012. Ecological assembly rules in plant communities—approaches, patterns and prospects.  
362 *Biological Reviews* 87:111–127.
- 363 Grime, J. P. 1979. Plant strategies and vegetation processes. Page *Plant strategies and vegetation*  
364 *processes*. John Wiley and Sons.
- 365 Holt, R. D. 2006. Emergent neutrality. *Trends in Ecology & Evolution* 21:531–3.
- 366 Hubbell, S. P. 2001. The Unified Neutral Theory of Biodiversity and Biogeography. Page  
367 *Monographs in Population Biology*. Princeton University Press, Princeton, NJ.

- 368 Hubbell, S. P., R. B. Foster, S. T. O'Brien, K. E. Harms, R. Condit, B. Wechsler, S. J. Wright,  
369 and S. Loo de Lao. 1999. Light-gap disturbances, recruitment limitation, and tree diversity  
370 in a Neotropical forest. *Science* 283:554–557.
- 371 Kohyama, T. 1993. Size-Structured Tree Populations in Gap-Dynamic Forest--The Forest  
372 Architecture Hypothesis for the Stable Coexistence of Species. *Journal of Ecology* 81:131–  
373 143.
- 374 Kohyama, T., and T. Takada. 2009. The stratification theory for plant coexistence promoted by  
375 one-sided competition. *Journal of Ecology* 97:463–471.
- 376 Kraft, N. J. B., and D. D. Ackerly. 2010. Functional trait and phylogenetic tests of community  
377 assembly across spatial scales in an Amazonian forest. *Ecological Monographs* 80:401–422.
- 378 Kraft, N. J. B., R. Valencia, and D. D. Ackerly. 2008. Functional traits and niche-based tree  
379 community assembly in an Amazonian forest. *Science (New York, N.Y.)* 322:580–2.
- 380 Kunstler, G., D. Falster, D. A. Coomes, F. Hui, M. Kooyman, Robert, D. C. Laughlin, L. Poorter,  
381 M. Vanderwel, G. Vieilledent, S. J. Wright, M. Aiba, C. Baraloto, J. Caspersen, J. H. C.  
382 Cornelissen, S. Gourlet-Fleury, M. Hanewinkel, B. Herault, J. Kattge, H. Kurokawa, Y.  
383 Onoda, J. Penuelas, H. Poorter, M. Uriarte, S. Richardson, and P. Ruiz-Benito. 2016. Plant  
384 functional traits have globally consistent effects on competition. *Nature* 529:1–15.
- 385 Lake, J. K., and A. M. Ostling. 2009. Comment on “Functional traits and niche-based tree  
386 community assembly in an Amazonian forest”. *Science (New York, N.Y.)* 324:1015; author  
387 reply 1015.
- 388 Leimar, O., A. Sasaki, M. Doebeli, and U. Dieckmann. 2013. Limiting similarity, species  
389 packing, and the shape of competition kernels. *Journal of Theoretical Biology* 339:3–13.
- 390 Levine, J. M., and D. J. Murrell. 2003. The Community-Level Consequences of Seed Dispersal  
391 Patterns. *Annual Review of Ecology, Evolution, and Systematics* 34:549–574.
- 392 MacArthur, R. H., and R. Levins. 1967. The Limiting Similarity, Convergence, and Divergence  
393 of Coexisting Species. *The American Naturalist* 101:377–385.
- 394 MacQueen, J. B. 1967. Some Methods for classification and Analysis of Multivariate  
395 Observations. 5th Berkeley Symposium on Mathematical Statistics and Probability 1967  
396 1:281–297.
- 397 Mouquet, N., and M. Loreau. 2003. Community Patterns in Source-Sink Metacommunities. *The*  
398 *American Naturalist* 162:544–557.

- 399 Muller-Landau, H. C. 2010. The tolerance-fecundity trade-off and the maintenance of diversity  
400 in seed size. *Proceedings of the National Academy of Sciences of the United States of*  
401 *America* 107:4242–7.
- 402 Muller-Landau, H. C., S. J. Wright, O. Calderón, R. Condit, and M. O. Hunter. 2008.  
403 Interspecific variation in primary seed dispersal in a tropical forest. *Journal of Ecology*  
404 96:653–667.
- 405 Paul W Richards. 1952. *The Tropical Rain Forest*. The Univeristy Press, Cambridge.
- 406 Rosindell, J., S. P. Hubbell, F. He, L. J. Harmon, and R. S. Etienne. 2012. The case for  
407 ecological neutral theory. *Trends in Ecology & Evolution* 27:204–209.
- 408 Sakavara, A., G. Tsirtsis, D. L. Roelke, R. Mancy, and S. Spatharis. 2018. Lumpy species  
409 coexistence arises robustly in fluctuating resource environments. *Proceedings of the*  
410 *National Academy of Sciences*.
- 411 Scheffer, M., and E. H. van Nes. 2006. Self-organized similarity, the evolutionary emergence of  
412 groups of similar species. *Proceedings of the National Academy of Sciences of the United*  
413 *States of America* 103:6230–6235.
- 414 Sedio, B. E., and A. M. Ostling. 2013. How specialised must natural enemies be to facilitate  
415 coexistence among plants? *Ecology Letters* 16:995–1003.
- 416 Sugar, C. A., and G. M. James. 2003. Finding the Number of Clusters in a Dataset: An  
417 Information-Theoretic Approach. *Journal of the American Statistical Association* 98:750–  
418 763.
- 419 Terborgh, J. 1985. The Vertical Component of Plant Species Diversity in Temperate and  
420 Tropical Forests. *The American Naturalist* 126:760–776.
- 421 Tibshirani, R., G. Walther, and T. Hastie. 2001. Estimating the number of clusters in a data set  
422 via the gap statistic. *Journal of the Royal Statistical Society: Series B (Statistical*  
423 *Methodology)* 63:411–423.
- 424 Tilman, D. 1988. *Plant strategies and the dynamics and structure of plant communities*. Princeton  
425 University Press.
- 426 Vergnon, R., N. K. Dulvy, and R. P. Freckleton. 2009. Niches versus neutrality: uncovering the  
427 drivers of diversity in a species-rich community. *Ecology letters* 12:1079–90.
- 428 Visser, M. D., M. Bruijning, S. J. Wright, H. C. Muller-Landau, E. Jongejans, L. S. Comita, and  
429 H. de Kroon. 2016. Functional traits as predictors of vital rates across the life cycle of

430 tropical trees. *Functional Ecology* 30:168–180.

431 Volkov, I., J. R. Banavar, M. O. Hunter, and A. Maritan. 2003. Neutral theory and relative  
 432 species abundance in ecology. *Nature* 424:1035–7.

433 Wright, S. J., K. Kitajima, N. J. B. Kraft, P. B. Reich, I. J. Wright, D. E. Bunker, R. Condit, J. W.  
 434 Dalling, S. J. Davies, S. Díaz, B. M. J. Engelbrecht, K. E. Harms, M. O. Hunter, C. O.  
 435 Marks, M. C. Ruiz-Jaen, C. M. Salvador, and A. E. Zanne. 2010. Functional traits and the  
 436 growth-mortality trade-off in tropical trees. *Ecology* 91:3664–74.

437 Yan, B., J. Zhang, Y. Liu, Z. Li, X. Huang, W. Yang, and A. Prinzing. 2012. Trait assembly of  
 438 woody plants in communities across sub-alpine gradients: Identifying the role of limiting  
 439 similarity. *Journal of Vegetation Science* 23:698–708.

440 Yan, M., and K. Ye. 2007. Determining the number of clusters using the weighted gap statistic.  
 441 *Biometrics* 63:1031–1037.

442

443 **Tables**

444 **Table 1: Clustering results for the 50-ha plot and regional pool.** For each header, the left  
 445 column refers to the 50-ha BCI plot (“50 ha.”), and the right column refers to the regional pool  
 446 (“Region”). The total number of species for which trait data were available is listed under  
 447 “Species”, and the total number of individuals represented is listed under “Individuals.” The  
 448 number of clusters  $K$  is the value at which the gap index peaks. z-scores and p-values were  
 449 obtained by comparing the gap statistic against 1,000 null communities. In the case of maximum  
 450 height, in addition to the peak at  $K = 1$  listed below, the regional pool also had a significant  
 451 subpeak at  $K = 4$  ( $Z = 2.0$ ,  $P = 0.03$ ). Significant results ( $P < 0.05$ ) are marked with an  
 452 asterisk.

	Species		Individuals		Clusters (K)		z-score (Z)		p-value (P)	
	50 ha.	Regio n	50 ha.	Regio n	50 ha.	Regio n	50 ha.	Regio n	50 ha.	Regio n
<b>Maximum height</b>	259	242	20,62	13,743	4*	1*	2.3*	3.4*	0.01	0.02*
			0						*	
<b>Wood density</b>	229	219	17,13	9,273	5*	20	2.1*	1.6	0.02	0.06
			9						*	

<b>Leaf mass/area</b>	250	234	18,36	9,528	3	3	0.2	1.5	0.40	0.07
			8							
<b>Dry seed mass</b>	185	177	18,25	10,667	15	3	0.2	1.0	0.38	0.16
			9							

453

## 454 **Figure Legends**

455 **Figure 1. Field site.** Intact tropical forest covers 20% of Panama. Our regional pool is an  
 456 aggregate of 40 sampling sites within 30km of BCI. Combined, these sites comprise  
 457 approximately 850 species, of which trait data were available for 242 species (max height), 219  
 458 (wood density), 234 (leaf mass per area), 177 (seed mass), comprising between 9,528 and 13,743  
 459 adult trees. Barro Colorado Island, spanning roughly 16 km<sup>2</sup>, sits in an artificial lake in the  
 460 Panama Canal. The 50-ha Forest Dynamics Plot, located near its center, is a 1,000m by 500m  
 461 patch of forest containing 300 tree species and 21,000 adult trees.

462

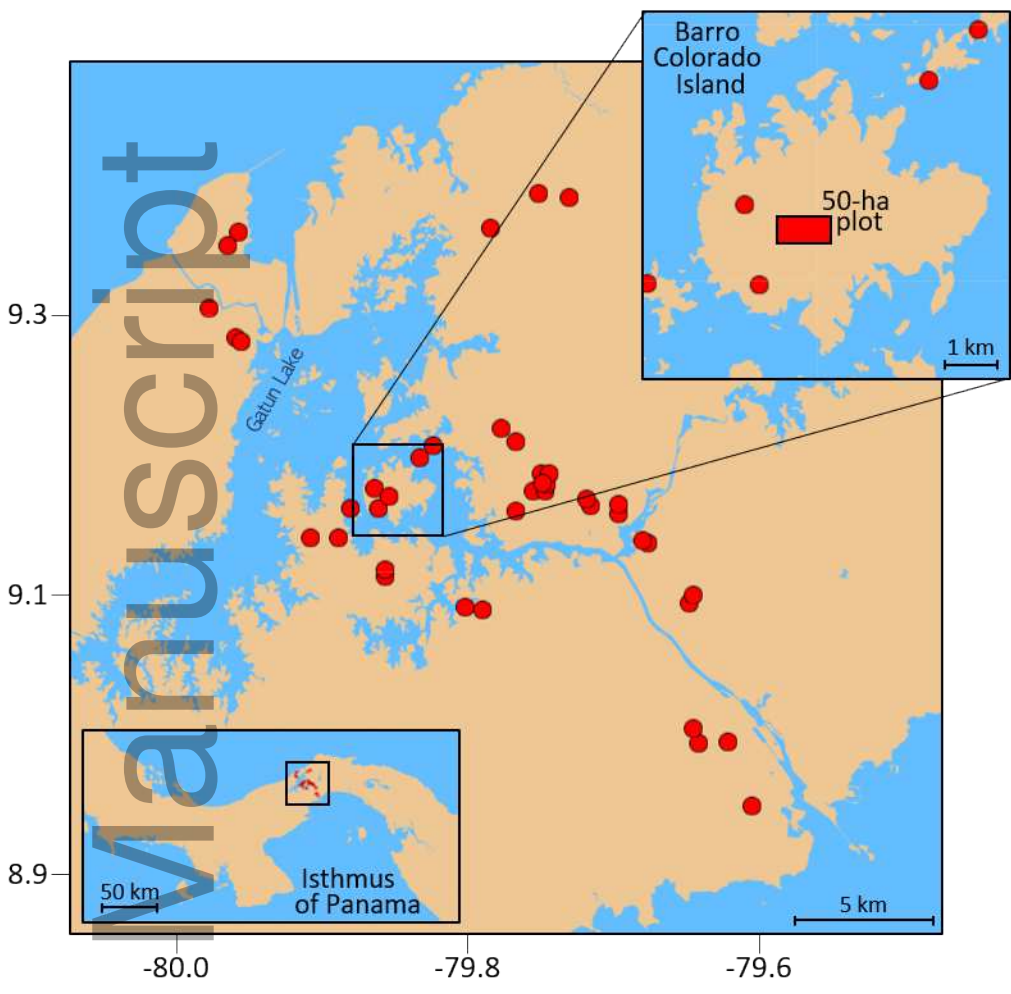
463 **Figure 2. Results.** Stem plots show trait distribution in the 50-ha plot on BCI for maximum  
 464 height (**A**), wood density (**B**), leaf mass per area (**C**), and seed mass (**D**). Each stem represents a  
 465 species, with its trait value plotted on the abscissa and abundance on the ordinate. Data shown  
 466 for the 2000 census. The corresponding gap curves (**E-H**) plot the gap index against the potential  
 467 number of clusters. The gap statistic is the highest point on the curve, with the red line indicating  
 468 the 95<sup>th</sup> quantile of the expected gap statistic under no pattern, obtained from 1,000 null sets. The  
 469 peaks above the red line in (**E**) and (**F**) reveal clustering by maximum height, with four clusters,  
 470 and wood density, with five. The alternating colors in the respective stem plots show cluster  
 471 membership of each species.

472

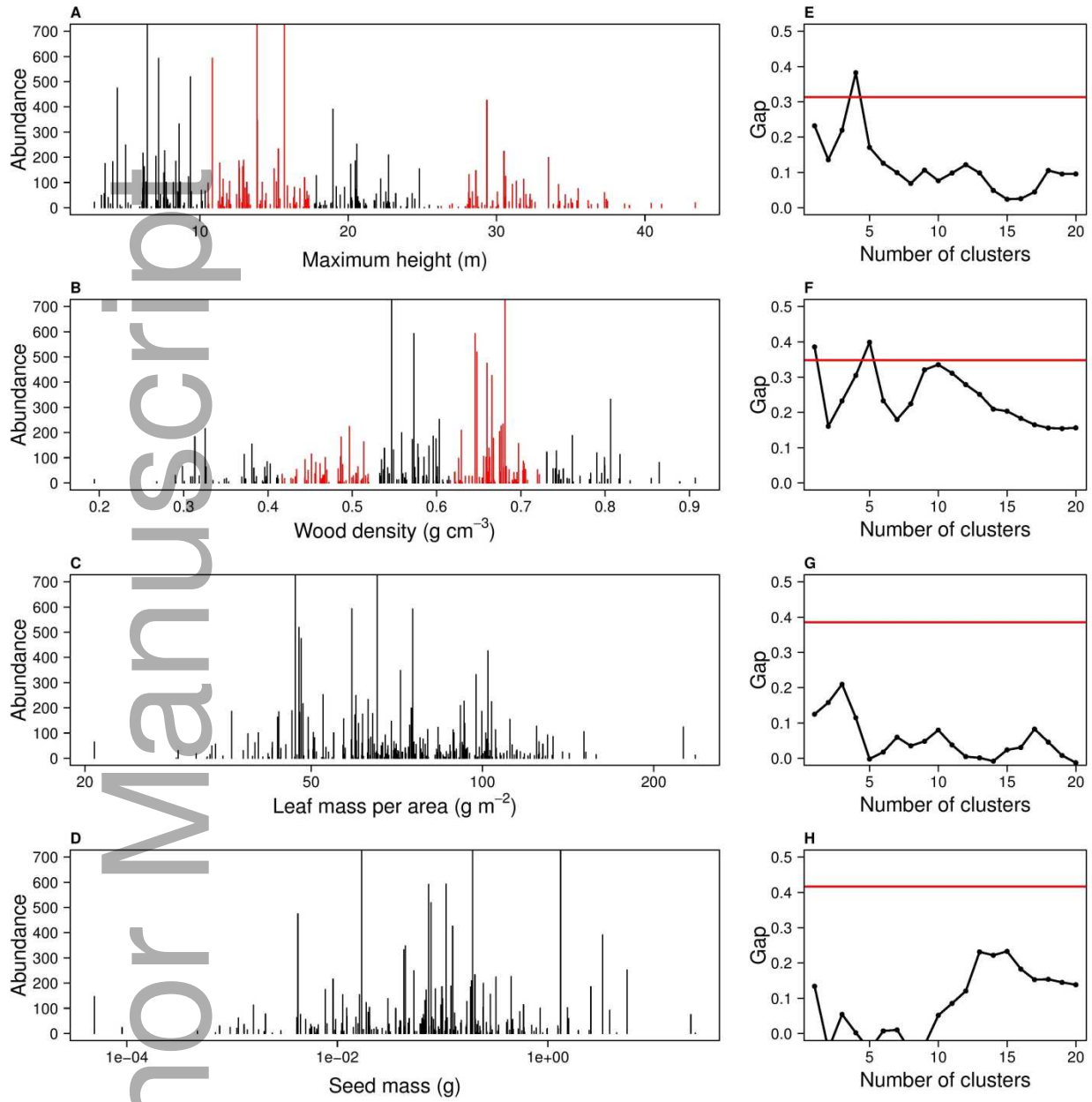
473 **Figure 3. Height layers.** Visual representation of the forest on Barro Colorado Island,  
 474 highlighting the four height-based groups identified by our metric: shrubs ( $61 \pm 2$  species;  $8,071$   
 475  $\pm 712$  adult individuals), understory treelets ( $71 \pm 3$ ;  $7,308 \pm 165$ ), midstory trees ( $61 \pm 1$ ;  
 476  $2,957 \pm 114$ ), and canopy trees ( $64 \pm 1$ ;  $2,836 \pm 82$ ), where the numbers represent mean and  
 477 standard deviation across the 7 censuses. Horizontal lines show maximum height of the dominant  
 478 (most abundant) species in each group.

479





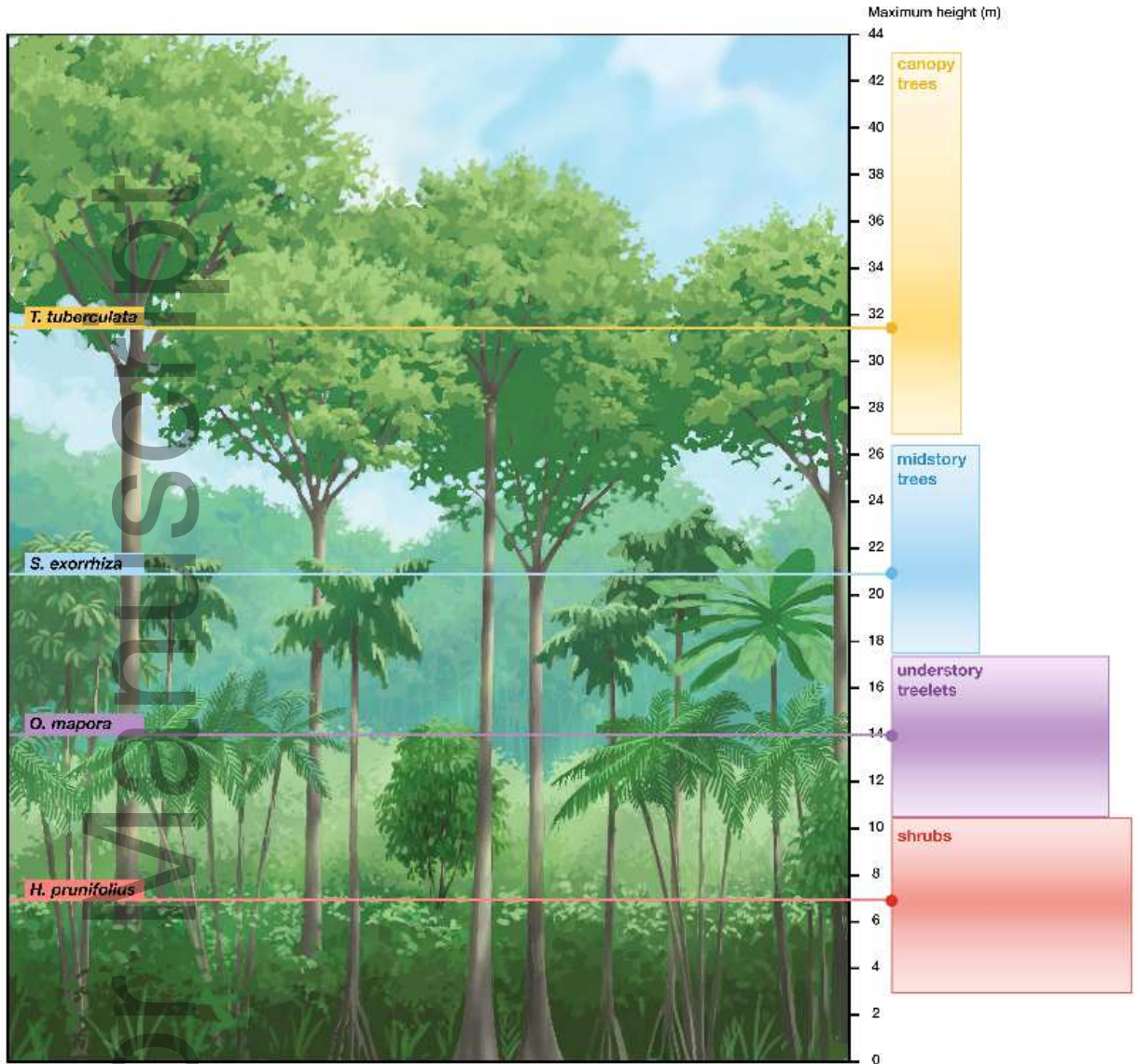
481  
482 **Figure 1.**



483

484

Figure 2.



485

486 Figure 3.