

1 **Positive associations among rare species and their persistence in ecological**
2 **assemblages**

3 Joaquín Calatayud^{1,2*}, Enrique Andivia^{3,4}, Adrián Escudero⁵, Carlos J. Melián⁶, Rubén
4 Bernardo-Madrid⁷, Markus Stoffel^{8,9,10}, Cristina Aponte¹¹, Nagore G. Medina^{12,13},
5 Rafael Molina-Venegas¹⁴, Xavier Arnan¹⁵, Martin Rosvall¹, Magnus Neuman¹, Jorge
6 Ari Noriega², Fernanda Alves-Martins², Isabel Draper¹³, Arantzazu Luzuriaga⁵, Juan
7 Antonio Ballesteros-Cánovas^{8,10}, César Morales-Molino^{16,17}, Pablo Ferrandis¹⁸, Asier
8 Herrero^{3,19}, Luciano Pataro¹³, Leandro Juen²⁰, Alex Cea²¹, Jaime Madrigal-González¹⁰.

9 ¹ Integrated Science Lab, Department of Physics, Umeå University, 901 87 Umeå,
10 Sweden.

11 ²Departamento de Biogeografía y Cambio Global, Museo Nacional de Ciencias
12 Naturales (MNCN-CSIC), C/ José Gutiérrez Abascal 2, 28006, Madrid, Spain.

13 ³Departamento de Ciencias de la Vida, Edificio de Ciencias, Universidad de Alcalá,
14 ctra. Madrid-Barcelona km 33.6, 28805 Alcalá de Henares, Madrid, Spain.

15 ⁴Departamento de Biodiversidad, Ecología y Evolución, Universidad Complutense de
16 Madrid, C/ José Antonio Novais 12, 28040, Madrid, Spain.

17 ⁵Departamento de Biología, Geología, Física y Química inorgánica. ESCET.,
18 Universidad Rey Juan Carlos, C/ Tulipán s/n, Móstoles, C.P. 28933, Madrid, Spain.

19 ⁶Department of Fish Ecology and Evolution, Eawag, Seestrasse 79, 6047
20 Kastanienbaum, Switzerland.

21 ⁷Departamento de Biología de la Conservación, Estación Biológica de Doñana CSIC, C/
22 Americo Vespucio, 26, 41092, Isla de la Cartuja, Sevilla, Spain.

23 ⁸Dendrolab, Department of Earth Sciences, University of Geneva, rue des Maraîchers
24 13, CH-1205 Geneva, Switzerland.

25 ⁹Department F.-A. Forel for Environmental and Aquatic Sciences, University of
26 Geneva, Boulevard Carl-Vogt 66, 1205, Geneva, Switzerland.

27 ¹⁰Climate Change Impacts and Risks in the Anthropocene (C-CIA), Institute for
28 Environmental Sciences (ISE), University of Geneva, 66 Boulevard Carl Vogt, 1205,
29 Geneva, Switzerland.

30 ¹¹Plant Science, The University of Melbourne, Burnley, 500 Yarra Boulevard,
31 Richmond, Victoria, 3121, Australia.

32 ¹²Department of Botany, Faculty of Sciences, University of South Bohemia, České
33 Budějovice, Czech Republic.

34 ¹³Departamento de Biología (Botánica), Facultad de Ciencias, Universidad Autónoma
35 de Madrid, calle Darwin 2, Madrid, Spain.

36 ¹⁴Plant Ecology, Institute of Plant Sciences (IPS), University of Bern, Altenbergrain 21,
37 3013 Bern, Switzerland.

38 ¹⁵CREAF (Centre de Recerca Ecològica i Aplicacions Forestals) Campus de Bellaterra
39 (UAB), Edifici C, 08193, Cerdanyola del Vallès Catalunya, Spain.

40 ¹⁶UMR CNRS 5805 EPOC – OASU, Université de Bordeaux Site de Talence -
41 Bâtiment B18N Allée Geoffroy Saint-Hilaire, CS 50023, 33615 Pessac Cedex, France.

42 ¹⁷Institute of Plant Sciences and Oeschger Centre for Climate Change Research,
43 University of Bern, Altenbergrain 21, 3013, Bern, Switzerland.

44 ¹⁸Higher Technical School of Agricultural and Forestry Engineering and Botanical
45 Institute, University of Castilla-La Mancha, 02071 Albacete, Spain.

46 ¹⁹Department of Plant Biology and Ecology, Pharmacy Faculty, University of Basque
47 Country, 01006 Vitoria-Gasteiz, Basque Country.

48 ²⁰Laboratório do Ecologia e Zoologia de Invertebrados, Instituto de Ciências Biológicas,
49 Universidade Federal do Pará, Rua Augusto Correia, 1, Barrio Guama, CEP 66075-110,
50 Belém, Pará, Brazil.

51 ²¹Departamento de Biología, Facultad de Ciencias, Universidad de La Serena, Avenida
52 Raúl Bitrán Nachary, 1305, Casilla 554, La Serena, Chile.

53 * Correspondence to: j.calatayud.ortega@gmail.com

54

55 **According to the competitive exclusion principle, species with low competitive**
56 **abilities should be excluded by more efficient competitors, and yet they generally**
57 **remain as rare species. Here, we describe the positive and negative spatial**
58 **association networks of 326 disparate assemblages, showing a general organization**
59 **pattern that simultaneously supports the primacy of competition and the**
60 **persistence of rare species. Abundant species monopolize negative associations in**
61 **about 90% of the assemblages. Contrarily, rare species are mostly involved in**
62 **positive associations, forming small network modules. Simulations suggest that**
63 **positive interactions among rare species and microhabitat preferences are the most**
64 **likely mechanisms underpinning this pattern and rare species persistence. The**
65 **consistent results across taxa and geography suggest a general explanation for the**
66 **maintenance of biodiversity in competitive environments.**

67 Rare species, in terms of low abundance, are the main component of the diversity of
68 ecological assemblages¹. However, despite decades of intense investigation, general
69 mechanisms behind the persistence of these species remain unclear. In theory, the
70 widely assumed effects of competition between pairs of species should preclude the

71 persistence of weak competitors and the high diversity observed in natural
72 assemblages^{2,3}. Explanations for this diversity paradox include the differential roles of
73 niche partition, intraspecific competition, facilitation, indirect and neutral interactions⁴⁻⁹,
74 among others. Yet, there is no consensus to explain rare species persistence across taxa
75 and environmental conditions so far.

76 The spatial arrangement of individuals plays a crucial role for unveiling the
77 mechanisms underpinning species assembly and coexistence¹⁰⁻²¹. Because individuals
78 within assemblages are not homogeneously distributed, their spatial organization may
79 both reflect important assembly processes^{10,11} and induce species coexistence *per se*¹².
80 For example, the patchy distribution of a dominant species might prevent the
81 monopolization of resources and allow the existence of its subordinate species^{12,13}.
82 Hence, considering spatial aspects of coexistence appears to be an important step in
83 elucidating assembly mechanisms¹². The spatial sorting of species can be the outcome
84 of divergent habitat preferences, dispersal abilities, and biotic interactions, although the
85 role of interactions is thought to prevail under rather homogeneous environmental
86 conditions, and especially at very fine spatial scales^{11,14,15}. The organization of species
87 within assemblages can be translated into association networks of species that are
88 spatially aggregated (positive networks) or segregated (negative networks). Association
89 networks of disparate biological assemblages can provide valuable empirical evidence
90 of the main forces driving the assembly of species¹⁶⁻²⁰, helping to reveal general
91 mechanisms underlying species coexistence.

92 Here we describe a general pattern of positive and negative species associations
93 that is consistent with the competitive exclusion paradigm but, at the same time, can
94 explain the persistence of rare species in natural assemblages. We base our results on a
95 dataset of 326 assemblages²¹ that meet the following criteria: (i) each assemblage

96 comprises taxa from only one trophic guild, thereby excluding the possibility that
97 species associations result from direct predation or parasitism; (ii) each assemblage
98 shows reduced spatial extent and low environmental heterogeneity, to increase the
99 likelihood that species associations are mainly due to biotic interactions; (iii) the
100 abundance of at least ten species is recorded in a minimum of ten samples, to improve
101 statistical power (Appendix S1); (iv) the dataset represents a wide variety of biomes
102 (*e.g.*, tropical forests, deserts, temperate steppes and polar climates), thus avoiding
103 biome-specific results; and (v) it encompasses a diversity of taxa (such as bryophytes,
104 vascular plants, and insects among others), to ensure the generalization of our results
105 across taxonomic and functional groups. We generated positive and negative association
106 networks for each assemblage by comparing the observed spatial association patterns
107 among species to a null model²². Species pairs that significantly deviate from random
108 expectations receive positive or negative links in their respective association networks
109 (Fig. 1).

110 We first analysed whether the structure of positive and negative association
111 networks can reflect predictions from the competitive exclusion principle. Given that
112 competition is heavily emphasized in the literature², one would expect species to be
113 more segregated than aggregated in natural systems. If so, negative networks should be
114 more densely connected (*i.e.*, more links per species) than their positive counterparts. In
115 accordance, negative networks were significantly more connected than their positive
116 pairs in a notable 93.2% of all assemblages ($t = 17.01$, $P < 0.001$, Fig. 2a). Differences
117 in connectivity remained similar after accounting for differences in network size ($t = -$
118 16.81 , $P < 0.001$, for 78.8% of the assemblages) or when calculating differences in the
119 average number of links (*i.e.* average species degree; $t = -14.69$, $P < 0.001$, for 69.0% of
120 the assemblages). Furthermore, if abundance is considered to be an expression of the

121 species' competitive abilities²³, the number of segregations should be monopolized by
122 the most abundant species. Accordingly, results indicated a strong positive correlation
123 between abundance and species degree in negative networks (mean Spearman's $\rho =$
124 0.65, SD = 0.23), but a weak or even negative correlation in positive networks (mean ρ
125 = 0.02, SD = 0.38), with differences between networks being statistically significant ($t =$
126 -23.88, $P < 0.001$, Fig. 2b). Moreover, we found evidence showing that a particular
127 species is more often involved in negative associations when it becomes abundant
128 (Appendix S2). Both the greater density of links and the relationship between species
129 degree and abundance in negative networks support current knowledge about the
130 prevailing role of competitive interactions in sustaining the dominance of abundant
131 species.

132 Yet, if the competitive exclusion principle is supported across several
133 assemblages, how can rare species persist? To search for potential mechanisms
134 answering this question we looked at the role played by rare species in association
135 networks. Curiously, we found that rare species are mostly involved in positive
136 associations in 91.7% of the assemblages studied, where positive networks showed a
137 higher incidence of less abundant species than their negative pairs ($t = 22.42$, $P < 0.001$,
138 Fig. 2c). Such spatial aggregations, however, do not occur among every rare species in
139 the assemblage. In fact, we found that 91.1% of positive networks were more modular
140 than their negative counterparts ($t = 39.68$, $P < 0.001$, Fig. 2d). This result remained
141 similar after accounting for network size and connectivity ($t = 11.31$, $P < 0.001$, for
142 67.3% of assemblages). Moreover, while 60.7% of positive networks were significantly
143 modular, only 13.8% of negative ones showed this pattern. Taken together, these
144 findings show that rare species tend to generate modular networks of positive spatial
145 associations.

146 The patterns of negative and positive associations networks remain largely
147 invariant regardless of different probability thresholds to detect significant associations,
148 the use of quantitative links and assumptions of disparate null models (Appendix S1).
149 This robust and conspicuous spatial organization suggests that the underlying
150 mechanisms can also be responsible for the persistence of rare species. On the one hand,
151 dissimilar habitat preferences between dominant species and groups of weak
152 competitors¹⁴ may generate this pattern, enhancing also rare species persistence. Indeed,
153 numerical simulations show that this possibility increased the probabilities of
154 reproducing realized association network patterns, regardless of different interaction
155 networks reflecting hypothesized assembly mechanisms (Fig. 3a and Appendix S3).
156 This, however, mainly occurs when habitat preferences are strong, a situation that
157 should arise under marked environmental gradients most likely far from the reality of
158 the fine-scaled assemblages studied here. Complementarily, positive interactions within
159 groups of rare species may also contribute and/or generate these modular positive
160 networks. This may moreover increase the persistence of weak competitors since, just as
161 in harsh abiotic environments²⁴, the biotic harshness produced by superior competitors
162 could be counterbalanced by positive interactions among rare species. Accordingly,
163 simulations show that the inclusion of positive interactions within groups of weak
164 competitors increases the chance of species persistence by 58.2% compared to
165 assemblages ruled by competition alone. Our simulations also reveal that this
166 hypothesis most likely reproduces the observed patterns in association networks
167 compared to other stabilizing mechanisms such as neutral colonization-extinction
168 dynamics⁹, intransitive competition^{25,26}, differential density-dependent effects^{27,28} or
169 facilitation by nurse species⁷ (Figs. 3b). Interestingly, the combination of habitats and
170 positive interactions yields the highest probability of reproducing the observed network

171 patterns (Fig. 3a). This further suggests that, even under strong differences in habitat
172 preferences, stabilizing forces, such as facilitation or complementarity, would enhance
173 the coexistence of groups of rare species in reduced microhabitats^{15,17}. Besides habitat
174 selection, it seems that modular positive interactions among rare species can contribute
175 to the pattern we found and the persistence of these species, which agrees with recent
176 experimental evidence²⁹.

177 Overall, our results show that ecological assemblages are consistently organized
178 in positive and negative association networks across the main biological groups (*i.e.*,
179 animals and plants) and geography (Fig. 4 and Extended Table 1). This ubiquity sheds
180 light on the long-standing diversity paradox as the potential mechanisms leading to this
181 organizational pattern can also enhance the persistence of rare species. Modular positive
182 interactions among weak competitors emerged as a plausible mechanism even when
183 assessed in conjunction with different microhabitat preferences. Questions remain about
184 the relative contribution and feedbacks of these positive interactions and microhabitats.
185 Nevertheless, the generality of the findings presented here bring us closer to
186 understanding the assemblage of the vast biodiversity on Earth.

187 **Materials and Methods**

188 Data acquisition

189 Assemblage data were collected from published studies in peer-reviewed journals
190 and our own surveys²¹ (Appendix S4 and Supplementary Table 1). Each assemblage
191 consists of at least ten samples where the abundance of at least ten species of the same
192 trophic guild was recorded. In order to minimize the effects of environmental
193 heterogeneity and dispersion on spatial patterns, we only included datasets that showed
194 (i) low environmental variability across samples (excluding surveys where any kind of
195 environmental gradient was reported or no clear information about it was provided), (ii)

196 a very reduced spatial extent (median = 0.1 ha; ranging from 0.01 to 25.6 ha), (iii) a
 197 very small grain size to increase the probability of physical and/or chemical contact
 198 among all species in the samples (median = 100 m²; ranging from 0.002 to 400 m²,
 199 respectively), and (iv) standardization among samples to avoid sampling biases (e.g.
 200 effects of area). Following these criteria, we gathered a total of 385 datasets distributed
 201 worldwide and representing a wide taxonomic spectrum, including bryophytes (n= 71),
 202 tracheophytes (n=279), anthozoans (n=7), and insects (n=28). Abundance was estimated
 203 as the number of individuals per sample in most of the assemblages, but a small number
 204 of assemblages included abundance data estimated as the percentage cover of the
 205 sampled surface (especially in bryophytes and plants). Since some null models only
 206 accept integer data (see below), we rounded percentages when necessary. Finally, we
 207 only used those assemblages where both positive and negative networks showed at least
 208 two links (n=326).

209

210 Generation of association networks

211 For each assemblage, we calculated similarity in abundance distribution across
 212 samples for each species pair i and j using the Schoener's index³²,

$$213 \quad S(i, j) = 1 - \sum_{k=1}^N |p_{ik} - p_{jk}| / 2$$

214 where N is the number of samples and p_{ik} is the proportion of the total abundance of
 215 species i present in sample k ($p_{ik} = x_{ik} / \sum_{k=1}^N x_{ik}$). We compared observed similarities to
 216 999 null values obtained through randomization of species abundances using a fixed-
 217 fixed algorithm (i.e., row and column totals are kept constant). For each observed
 218 similarity value, two one-tailed p-values were calculated as the proportion of null values
 219 (plus the observation) that were higher than or equal to and lower than or equal to the

220 observed value for positive and negative associations, respectively. We considered an
221 aggregation or segregation significant in those cases where associated p-values in any of
222 the two tests were lower than or equal to 0.05. Alternative probability thresholds and
223 null models provided quantitatively and qualitatively similar results (Appendix S1;
224 Supplementary Figures 1 and 2). Significantly aggregated and segregated species pairs
225 were used to generate unweighted links in the positive and negative association
226 networks of each assemblage, respectively. It is important to note that the frequency of
227 spurious associations (i.e., Type I errors) may be thought to be relatively high in species
228 rich assemblages due to multiple comparisons (but see ref³³). However, species pairwise
229 similarities were compared against null values generated using a fixed-fixed
230 assemblage-wise null model (i.e., a strict null model making null hypotheses among
231 comparisons to be different but intrinsically interdependent). This partially alleviates the
232 detection of false positives while preventing the use of powerful false discovery rate
233 methods³⁴. Nevertheless, we used the same nominal error (i.e., $\alpha = 0.05$) to detect both
234 positive and negative associations, making the rate of false discoveries equal in both
235 types of networks, and allowing unbiased comparisons of their structures. Indeed,
236 results remained largely constant when using different nominal errors (Appendix S1 and
237 Supplementary Figure 1).

238

239 Network structure comparison

240 To explore whether positive and negative association networks reflected
241 competitive processes we compared their connectivity and their relationships between
242 abundance (calculated as the sum of the abundances across samples) and species degree
243 (i.e., species' number of links) for each pair of network types. Connectivity is defined as
244 the number of realised links relative to the number of potential links. This measure of

245 connectivity may be negatively correlated with network size. Hence, we also used
246 residual connectivity obtained from the residuals of a linear regression between the
247 number of observed and potential links, both log-transformed³⁵. On the other hand, the
248 relationship between abundance and species degree was assessed using the Spearman's
249 ρ correlation coefficient. Finally, to search for differences between network types we
250 used a paired Student's t-test, where the alternative hypothesis was that negative
251 networks present higher means than their positive pairs.

252 To determine if rare species have a larger participation in positive association
253 networks, we compared the average relative abundance, weighted by the number of
254 links of each species in the network, between the species involved in positive and
255 negative networks. We also explored if positive networks were more modular than their
256 negative pairs calculating modularity with the index proposed by Newman³⁶ (Q) along
257 with the optimisation algorithm of Louvain³⁷. The algorithm was run 100 times, and we
258 selected the partition that showed the highest modularity value. Since modularity can be
259 related to network size and connectivity, we compared observed and null modularity
260 values from random networks generated using a null model that maintains the number
261 of links and nodes, as well as the degree sequence (implemented in the RandNetGen
262 software³⁸). Then, we computed relative modularity values as $Q_r = -2(P - 0.5)$,
263 where P represents the proportion of null cases showing modularity higher than or equal
264 to the observation. A paired Student's t-test was used to explore the differences between
265 network types in all cases.

266 Finally, we explored whether the probability of finding the above-explained
267 differences of positive and negative networks was related to the number of samples per
268 assemblage (as indicative of sampling effort), an approximation of null model degrees
269 of freedom (Appendix S1), latitude, longitude, taxonomic group (i.e., animals or plants)

270 and species richness. To do so, we firstly generated four binomial dependent variables,
271 based on whether i) the negative networks of each assemblage were more densely
272 connected than their positive pairs; ii) the negative networks present higher positive
273 abundance-degree relationships; iii) the positive networks tend to be composed of less
274 abundant species; and iv) the positive networks were more modular. Then, we fitted
275 logistic models with a logit link function.

276

277 Numerical simulations

278 We ran simulations to explore whether different interaction matrices and/or habitat
279 preferences can generate the patterns observed in association networks. We designed a
280 simulation model composed of 20 samples and ten species, whose individuals were
281 randomly distributed at the outset. Individuals reproduce, colonise a randomly chosen
282 sample or die, with probabilities dependent on the density of individuals and the sample
283 carrying capacity ($K = 100$). We subsequently incorporated the effects of both
284 competition and positive interactions by modifying these probabilities depending on the
285 species identities of co-occurring individuals (Appendix S3). That is, individuals of
286 dominant species reduce the probability of reproduction and colonisation, while
287 increasing mortality probability, of co-occurring individuals of subordinate species.
288 Benefactor individuals have the opposite effects on beneficiary individuals'
289 probabilities (Supplementary Figure 3a and b).

290 We further incorporated the effects of dissimilarities in habitat preferences by
291 setting four habitats preferred by different groups of species (Supplementary Figure 3c).
292 Specifically, the probabilities of reproduction, survival (i.e., one minus the mortality
293 probability) and colonisation in non-preferred habitats were multiplied by a habitat-
294 tolerance coefficient, β , ranging between 0 (null tolerance) and 1 (total tolerance;

295 Appendix S3). Hence, when $\beta = 0$, individuals are highly specialist and only allowed to
296 reproduce, survive or colonize in preferred habitat, whereas $\beta = 1$ corresponds to a
297 neutral habitat scenario.

298 We ran simulations following six hypotheses explaining species assembly and
299 coexistence (Fig. 3 and Appendix S3). i) A neutral interaction model, where all species
300 were ecologically equivalent⁹. ii) A hierarchical competition model with one strong
301 competitor. iii) An intraspecific density-dependent model, where superior competitors
302 suffer more from intraspecific competition⁵. iv) A model incorporating intransitive
303 competition²⁶, where the superior competitor is outcompeted by three species, which, in
304 turn, outcompete all species except specific pairs (i.e., theoretically promoting the
305 generation of empirical association patterns; see Appendix S3). v) A nurse model⁷ with
306 four superior competitors, three of which facilitate different pairs of subordinate
307 species. vi) A model reflecting positive interactions within three groups of three rare
308 species. Fourteen additional matrices with different settings to these six general models
309 were also explored (see Appendix S3; Supplementary Figure 4).

310 Simulations were run using a wide range of combinations ($n=216$) where
311 demographic rates (i.e., reproduction, mortality and dispersal) had different relative
312 importance (Appendix 3.4). In addition, we also used five values of the habitat-
313 tolerance parameter (β). For each interaction matrix and parameter combination, we ran
314 25 replicates of 5,000 iterations each. We quantified the probability of simulated
315 association networks showing empirical patterns (i.e., differences between positive and
316 negative networks in connectivity, abundance-degree relationship, abundance and
317 modularity), as well as the probability of persistence of all species (i.e., non-extinction),
318 as the proportion of all our replicates showing these patterns. Finally, these probabilities
319 were averaged across the parameter space defined by demographic rates where the

320 *Competition* model, under neutral habitat preferences, fulfilled expectations from the
321 competitive exclusion principle (i.e., weak competitors went extinct; P (non-extinction)
322 = 0; see Appendix S3 and Supplementary Figure 5).

323

324 **Acknowledgments.** We are very grateful to Joaquín Hortal and Stefano Allesina for
325 their critical comments on an early version of the manuscript. The simulations were
326 performed on resources provided by the Swedish National Infrastructure for Computing
327 (SNIC) at HPC2N. J.C. is supported by the Carl Tryggers Foundation for Scientific
328 Research (CTS 16:384). E.A. is supported by a postdoctoral grant founded by the
329 Universidad Complutense de Madrid. C.J.M is supported by the Swiss National Science
330 Foundation (SNSF-31003A-144162). R.B.M. is supported by the Spanish Ministry of
331 Science and Innovation Predoctoral Fellowship BES-2013-065753. M.S., J.A.B.C. and
332 J.M.G acknowledge support from the University of Geneva. X.A. is supported by a
333 Ramón y Cajal research contract by the Spanish Ministry of Economy and
334 Competitiveness (RYC-2015-18448). M.R. is supported by the Swedish Research
335 Council grant 2016-00796. JAN was supported by a Colombian COLCIENCIAS PhD
336 scholarship. F.A.M. is grateful to CAPES for a PhD scholarship (120147/2016-01).
337 A.L., P.F. and J.M.G. were funded by AGORA Project (ref. CGL2016-77417-P;
338 MINECO, Spain). C.M.M. was supported by an IdEx Bordeaux Postdoctoral
339 Fellowship (VECLIMED project). A.H. was supported by the University of Alcalá Own
340 Research Programme's 2018 Postdoctoral Grant and Basque Country Government
341 funding support to FisioClimaCO2 (IT1022-16) research group. L.J. received
342 productivity grants from of CNPq (process: 307597/2016-4).

343 **Author contributions.** J.C. and J.M.G. conceived the ideas; J.C. and J.M.G. designed
344 the analyses with contributions from E.A., A.E., C.J.M. and R.B.M.; J.C., E.A., R.B.M.,

345 M.S., C.A., X.A., N.G.M., J.A.N., F.A.M., I.D., A.L., J.A.B.C., C.M.M., P.F., A.H.,
346 L.P., L.J., A.C., and J.M.G. collected the data; J.C. analyzed the data with assistance
347 from C.J.M., M.R. and M.N.; J.C., E.A. and J.M.G. lead the writing in close
348 collaboration with A.E., C.J.M., R.B.M., M.S., C.A., and R.M.V.; all authors
349 contributed to the development and writing of the paper.

350 **Code availability.** The R scripts used in this study are freely available at
351 <https://doi.org/10.6084/m9.figshare.9906092>

352 **Data availability.** The dataset used in this study is freely available at
353 <https://doi.org/10.6084/m9.figshare.9906092>

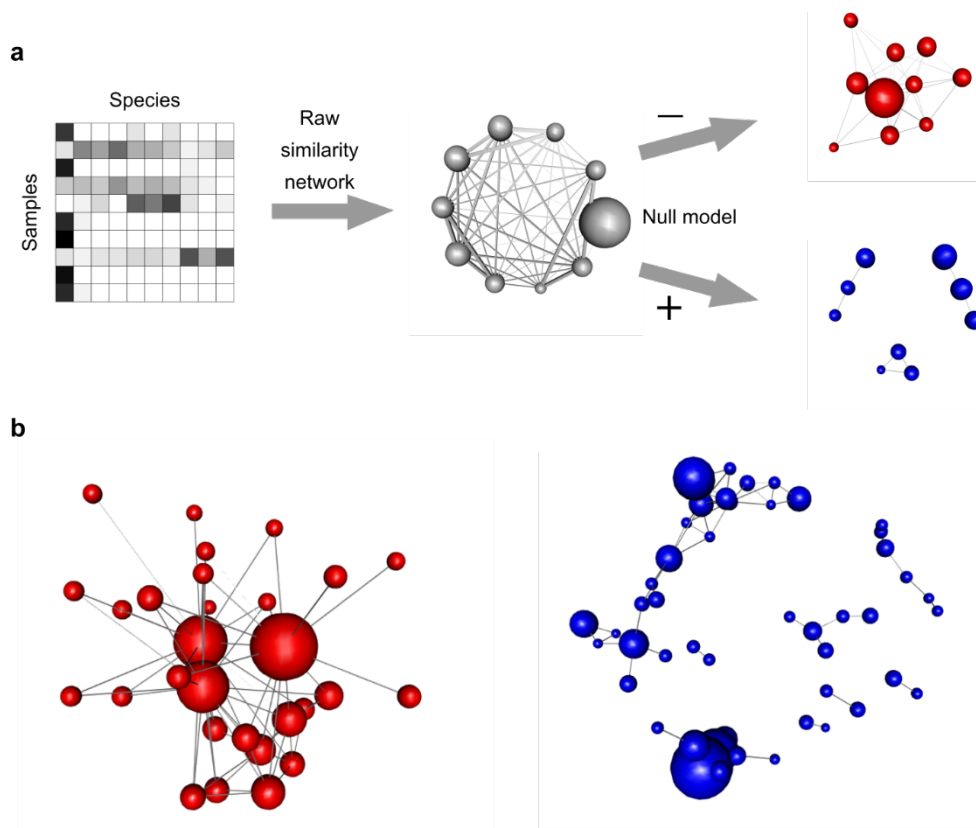
354 **References**

- 355 1. Gaston, K.J. *Rarity*. (Chapman & Hall, London, 1994).
- 356 2. Hardin, G. The competitive exclusion principle. *Science* **131**, 1292-1297(1960).
- 357 3. Hutchinson, G. E. The paradox of the plankton. *Am. Nat.* **95**, 137-145 (1961).
- 358 4. Schoener, T. W. Resource partitioning in ecological communities. *Science*, **185**,
359 27-39 (1974).
- 360 5. Yenni, G., Adler, P. B. & Ernest, S. M. Strong self-limitation promotes the
361 persistence of rare species. *Ecology* **93**, 456-461 (2012).
- 362 6. Chesson, P. Mechanisms of maintenance of species diversity. *Annu. Rev. Ecol.*
363 *Evol. Syst.* **31**, 343-366 (2000).
- 364 7. Soliveres, S. et al. A missing link between facilitation and plant species
365 coexistence: nurses benefit generally rare species more than common ones. *J.*
366 *Ecol.* **103**, 1183-1189 (2015).
- 367 8. Grilli, J. et al. Higher-order interactions stabilize dynamics in competitive network
368 models. *Nature* **548**, 210 (2017).

- 369 9. Hubbell, S. P. *The unified neutral theory of biodiversity and biogeography*, vol. 32
370 (Princeton University Press, 2001).
- 371 10. Durrett, R. & Levin, S. Spatial aspects of interspecific competition. *Theor. Popul.*
372 *Biol.* **53**, 30-43 (1998).
- 373 11. McIntire, E. J. & Fajardo, A. Beyond description: the active and effective way to
374 infer processes from spatial patterns. *Ecology* **90**, 46-56 (2009).
- 375 12. Arnan, X., Gaucherel C. & Andersen A.N. Dominance and species co-occurrence
376 in highly diverse ant communities: a test of the interstitial hypothesis and
377 discovery of a competition cascade. *Oecologia* **166**, 783-794 (2011).
- 378 13. Atkinson, W. D. & Shorrocks, B. Competition on a divided and ephemeral
379 resource: a simulation model. *J. Anim. Ecol.* **50**, 461-471 (1981).
- 380 14. Hart, S. P., Usinowicz, J. & Levine, J. M. The spatial scales of species
381 coexistence. *Nat. Ecol. Evol.* **1**, 1066 (2017).
- 382 15. Chacón-Labela, J. et al. Evidence for a stochastic geometry of biodiversity: the
383 effects of species abundance, richness and intraspecific clustering. *J. Ecol.* **105**,
384 382-390 (2017).
- 385 16. Saiz, H. et al. Evidence of structural balance in spatial ecological networks.
386 *Ecography* **40**, 733-741 (2017).
- 387 17. Freilich, M. A., Wieters, E., Broitman, B. R., Marquet, P. A. & Navarrete, S. A.
388 Species co-occurrence networks: can they reveal trophic and non-trophic
389 interactions in ecological communities? *Ecology* **99**, 690-699 (2018).
- 390 18. A. Faisal, et al. Inferring species interaction networks from species abundance
391 data: A comparative evaluation of various statistical and machine learning
392 methods. *Ecol. Inform.* **5**, 451-464 (2010).

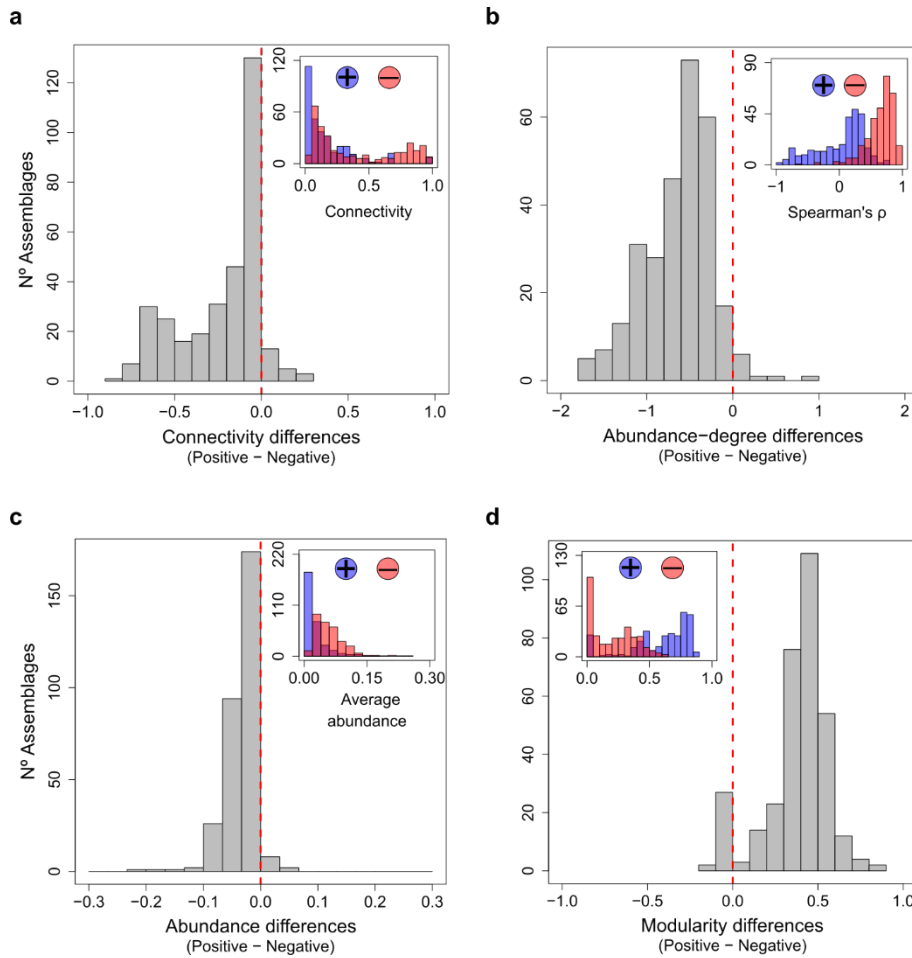
- 393 19. Barberán, A., Bates, S. T., Casamayor, E. O. & Fierer, N. (2012). Using network
394 analysis to explore co-occurrence patterns in soil microbial communities. *The*
395 *ISME journal* **6**, 343 (2012).
- 396 20. Borthagaray, A. I., Arim, M. & Marquet, P. A. Inferring species roles in
397 metacommunity structure from species co-occurrence networks. *Proc. R. Soc.*
398 *Lond. B*, **281**, 20141425 (2014)
- 399 21. Calatayud, et al. (2019) Data from: Positive associations among rare species and
400 their persistence in ecological assemblages. figshare. DOI:
401 10.6084/m9.figshare.9906092
- 402 22. Ulrich, W. & Gotelli, N. J. Null model analysis of species associations using
403 abundance data. *Ecology* **91**, 3384-3397 (2010).
- 404 23. Tilman, D. Resource competition between plankton algae: an experimental and
405 theoretical approach. *Ecology* **58**, 338-348 (1977).
- 406 24. Callaway, R.M. et al. Positive interactions among alpine plants increase with
407 stress. *Nature* **417**, 844-848 (2002).
- 408 25. Allesina, S. & Levine, J. M. A competitive network theory of species
409 diversity. *Proc. Nat. Acad. Sci.* **108**, 5638-5642(2011).
- 410 26. Gallien, L. et al. The effects of intransitive competition on coexistence. *Ecol.Lett.*
411 **20**, 791-800 (2017).
- 412 27. Comita, L. S. et al. Asymmetric density dependence shapes species abundances in
413 a tropical tree community. *Science* **329**, 330-332 (2010).
- 414 28. Cody, M. L. & Diamond, J. M. *Ecology and evolution of communities*. (Belknap
415 Press, Cambridge, 1975).

- 416 29. Bimler, M. D., Stouffer, D. B., Lai, H. R. & Mayfield, M. M. Accurate predictions
417 of coexistence in natural systems require the inclusion of facilitative interactions
418 and environmental dependency. *J. Ecol.* **106**, 1839-1852 (2018).
- 419 30. Fruchterman, T. M. & Reingold, E. M. Graph drawing by force-directed
420 placement. *Softw. Pract. Exp.* **21**, 1129-1164 (1991).
- 421 31. Bailey, R. G. Explanatory supplement to ecoregions map of the continents.
422 *Environ. Conser.* **16**, 307-309, (1989).
- 423 32. Schoener, T. W. The Anolis lizards of Bimini: resource partitioning in a complex
424 fauna. *Ecology* **49**, 704-726 (1968).
- 425 33. Rothman, K. J. No adjustments are needed for multiple comparisons.
426 *Epidemiology* **1**, 43-46 (1990).
- 427 34. Benjamini, Y. & Yekutieli, D. The control of the false discovery rate in multiple
428 testing under dependency. *Ann. Stat.* **29**, 1165-1188 (2001).
- 429 35. Araújo, W.S. et al. Contrasting Effects of Land Use Intensity and Exotic Host
430 Plants on the Specialization of Interactions in Plant-Herbivore Networks. *PLoS*
431 *ONE* **10**, e0115606 (2015).
- 432 36. Newman, M. E. Modularity and community structure in networks. *Proc. Nat.*
433 *Acad. Sci. U.S.A.* **103**, 8577-8582 (2006).
- 434 37. Blondel, V. D. et al. Fast unfolding of communities in large networks. *J. Stat.*
435 *Mech. Theory Exp.* **2008**, P10008 (2008).
- 436 38. Colomer-de-Simón, P., Serrano, M. A., Beiró, M. G., Alvarez-Hamelin, J. I. &
437 Boguná, M. Deciphering the global organization of clustering in real complex
438 networks. *Sci. Rep.* **3**, 2517 (2013).
- 439



440

441 **Fig. 1. Approaching assembly mechanisms through the lens of positive and negative**
 442 **association networks. a**, Species segregations and aggregations can inform on the main
 443 mechanisms underlying ecological assemblages. These spatial patterns are measured
 444 between species pairs using the similarity in the spatial distribution of their individuals.
 445 Observed similarities are compared with those obtained by a null model to distinguish
 446 actual associations from those generated by chance. Species pairs whose individuals are
 447 more aggregated in samples than expected by chance receive a positive link in association
 448 networks (blue nodes). Species pairs whose individuals are more segregated than
 449 randomly expected receive a negative link in association networks (red nodes). **b**, Positive
 450 (blue) and negative (red) networks of a tropical rainforest tree assemblage (see
 451 “Barra_Paraguacu” in Supplementary Table 1). The size of the nodes is proportional to
 452 the species’ abundances at the assemblage level. Networks were plotted using the
 453 Fruchterman-Reingold force-directed layout algorithm³⁰.



454

455

Fig. 2. The contrasting patterns of positive and negative association networks.

456

a, The higher connectivity of negative networks indicates that species segregation

457

dominates over species aggregation. **b,** These segregations are monopolized mostly by

458

dominant species, as shown by the strong relationship between abundance and species

459

degree (i.e., number of links of a species) in negative networks. **c,** In contrast, less

460

abundant species are more prone to generate positive associations, although, **d,** these

461

associations only occur among specific groups of rare species, as indicated by the higher

462

modularity of positive networks. Main histograms show the differences in network

463

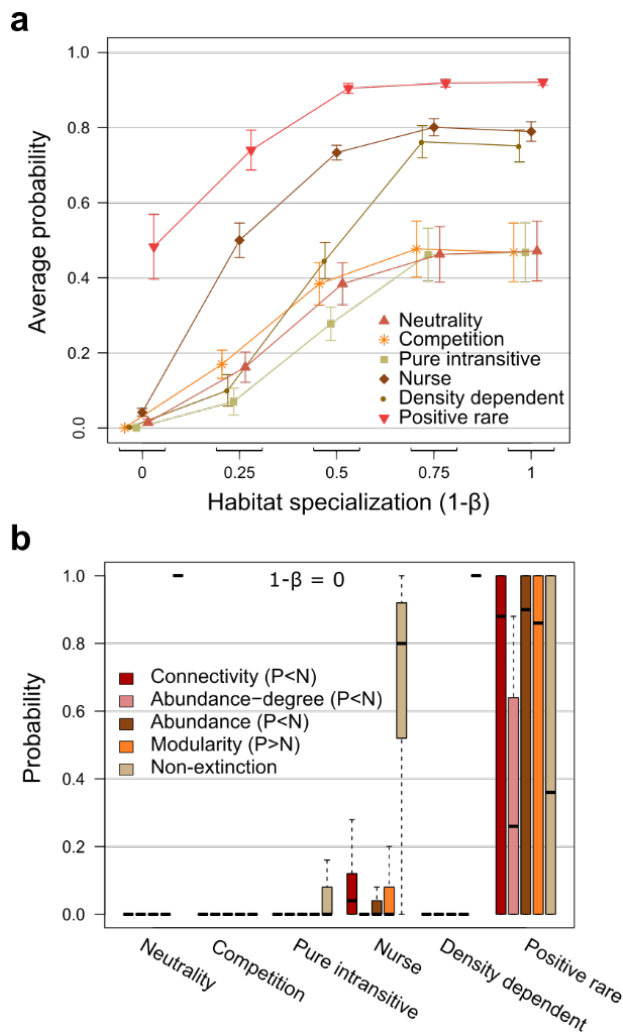
features between pairs of positive and negative networks. Insets show the raw values for

464

both types of networks, where purple colour represents the overlap between both

465

distributions.



466

467

Fig. 3. Positive interactions among weak competitors alone or together with

468

habitat preferences reproduce realized association patterns. **a**, Dissimilarities in

469

habitat preferences between dominants and groups of rare species may generate the

470

empirical patterns of association networks, regardless of different assembly mechanisms.

471

However, this only occurs when habitat specialization is strong. Moreover, the

472

combination of habitats and positive interactions among weak competitors (*Positive rare*)

473

yields the highest probabilities. The y-axis represents the average probabilities of finding

474

the four empirical patterns, and the x-axis depicts a gradient of habitat specialisation (see

475

Methods and Appendix S3). Error bars depict confidence intervals at $\alpha = 0.05$. **b**, All

476

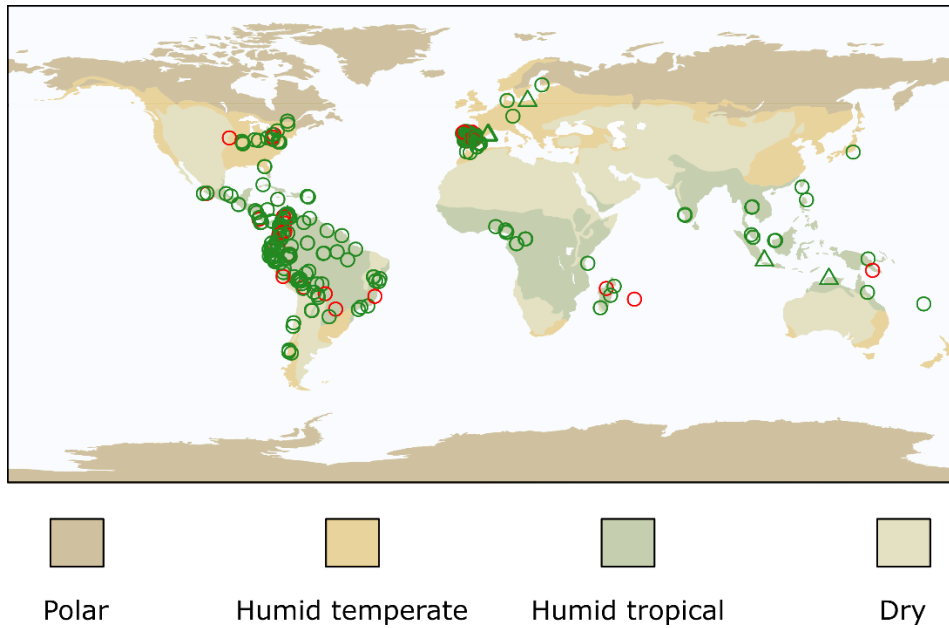
theoretical models explaining species coexistence increase the chance of species

477

persistence (Non-extinction) relative to simulated assemblages only driven by

478 hierarchical competition. However, positive interactions among groups of rare species is
479 the most likely model to generate simulated assemblages showing the same association
480 networks as empirical assemblages (Connectivity, Fig. 2a; Abundance-degree, Fig. 2b;
481 Abundance, Fig. 2c; and Modularity, Fig. 2d). The y-axis represents the probability of
482 simulated association networks showing empirical differences between positive and
483 negative networks across different combinations of reproduction, mortality and dispersal
484 rates where interactions are expressed (see Methods and Appendix S3). In boxplots the
485 centre line shows median probabilities, being the box between 25th and 75th percentiles
486 and the whiskers at minimum and maximum probability once outliers are discounted.
487 Outliers are not shown. P: positive networks. N: negative networks.

488



490 **Fig. 4. Organization of association networks remains invariant across the**
 491 **globe and regardless of taxa.** Circles and triangles represent plant and animal
 492 assemblages, respectively. Green colour depicts assemblages where positive networks
 493 were both composed of less abundant species and more modular than negative
 494 counterparts, whereas red colour shows assemblages where these patterns were not found.
 495 Map colours represent the Earth climatic zones proposed by Bailey³¹.

496