1 Title:

2 Community structure informs species geographic distributions

3

4 Authors: Alicia Montesinos-Navarro *[1,2], Alba Estrada[3,], Xavier Font[4], Migel G. Matias[1, 5],
5 Catarina Meireles[1], Manuel Mendoza[1,6], Joao P. Honrado[7,8], Hari D. Prasad[9], Joana R.
6 Vicente [7,8], Regan Early[10].

7

8 [1] InBIO/CIBIO - Centro de Investigação em Biodiversidade e Recursos Genéticos,
9 Universidade de Évora. Casa Cordovil 2º andar, Rua Dr. Joaquim Henrique da Fonseca,
10 7000-890. Évora, Portugal
11 [2] Spanish Scientific Council (CSIC). Centro de Investigaciones sobre Desertificación (CIDE,
12 CSIC-UV-GV), Carretera de Moncada-Náquera Km 4.5 46113 Moncada, Valencia, Spain.
13 [3] Research Unit of Biodiversity (UMIB, UO-CSIC-PA), Oviedo University – Campus
14 Mieres, Spain
15 [4] Departament de Biologia Vegetal. Facultat de Biologia, Universitat de Barcelona, Av.
16 Diagonal 643, 08028 Barcelona, España
17 [5] Imperial College London. Buckhurst Road. Ascot, Berks. United Kingdom
18 [6] Spanish Scientific Council (CSIC). National Museum of Natural History (MNCN).
19 Department of Biogeography and Global Change. C/ Serrano 115bis, 28006 Madrid, Spain
20 [7] InBIO / CIBIO - Centro de Investigação em Biodiversidade e Recursos Genéticos, Campus
21 Agrário de Vairão, Universidade do Porto, 4485-601 Vairão, Portugal
22 [88] Faculdade de Ciências da Universidade do Porto, Edifício FC4 (Biologia), Rua do Campo
23 Alegre, s/n, 4169-007 Porto, Portugal
24 [9] Physical Sciences and Engineering Division, King Abdullah University of Science and
25 Technology, Thuwal, Saudi Arabia
26 [10] Centre for Ecology and Conservation, College of Life and Environmental Sciences,
27 University of Exeter, Cornwall Campus TR10 9FE, Cornwall, UK
28

29 *Corresponding author:

30 Email: ali.montesinos@gmail.com
31 ORCID-ID: 0000-0003-4656-0321

32 Running headline: Community structure and species distributions

33

Abstract

Understanding what determines species' geographic distributions is crucial for assessing global change threats to biodiversity. Measuring limits on distributions is usually, and necessarily, done with data at large geographic extents and coarse spatial resolution. However, survival of individuals is determined by processes that happen at small spatial scales. The relative abundance of coexisting species (i.e. 'community structure') reflects assembly processes occurring at small scales, and are often available for relatively extensive areas, so could be useful for explaining species distributions. We demonstrate that Bayesian Network Inference (BNI) can overcome several challenges to including community structure into studies of species distributions, despite having been little used to date. We hypothesized that the relative abundance of coexisting species can improve predictions of species distributions. In 1570 assemblages of 68 Mediterranean woody plant species we used BNI to incorporate community structure into Species Distribution Models (SDMs), alongside environmental information. Information on species associations improved SDM predictions of community structure and species distributions moderately, though for some habitat specialists the deviance explained increased by up to 15%. We demonstrate that most species associations (95%) were positive and occurred between species with ecologically similar traits. This suggests that SDM improvement could be because species co-occurrences are a proxy for local ecological processes. Our study shows that Bayesian Networks, when interpreted carefully, can be used to include local conditions into measurements of species' large-scale distributions, and this information can improve the predictions of species distributions.

Introduction

Current topics in ecology such as biological invasions or species responses to global change rely on a better understanding of the drivers governing species distributions [1,2]. Although at large geographical scales climatic conditions are the main factor determining species distributions (but see [3]), several studies have shown that non-climatic biotic and abiotic factors (e.g. landscape dynamics, disturbance regimes, micro-topography, biotic interactions between species such as competition or predation) are important at finer spatial resolutions [4–9]. Therefore, information reflecting local ecological processes would be valuable for improving forecasts of responses to environmental change by species distribution models (SDMs). Nevertheless this information is rarely included (but see [10]).

A potential reason why local factors are not usually included in SDMs is the lack of suitable fine scale data over large areas. Although data on micro-environmental and biotic interactions are usually not available at a large scale, for many taxa, in particular plant species, the relative abundance of coexisting species in a community is well documented across large geographic areas (e.g. in vegetation databases such as SIVIM (http://www.sivim.info/sivi/), BDN (http://www.magrama.gob.es/es/biodiversidad/servicios/banco-datos-naturaleza/), BIEN (http://bien.nceas.ucsb.edu/bien/)). An additional challenge specific to biotic interactions is finding statistical techniques to deal with the large amount of potential interactions. There have been previous attempts to include biotic information into SDMs [11], one approach is to focus on a small number of pair-wise species dependencies (< 25 species)

80   [12–15] and another to use surrogates for biotic interactions, such as species richness [16].

81   However, these approaches are either unable to assess all potential species interactions (there

82   are $N^2 - N/2$ possible pair-wise interactions in a community that contains N species), or they

83   rely on extremely detailed ecological knowledge. Finally, the statistical challenge is made

84   much more complicated when considering that species live in complex interaction networks,

85   where co-occurrence patterns are affected by not only pair-wise but also indirect interactions

86   influenced by the presence of a third species [17,18].

87   Bayesian network inference (BNI) can be a useful tool to overcome these major challenges.

88   These analyses are used to study the conditional dependencies (represented by directed

89   edges) among a set of either abiotic (i.e. climatic, edaphic or land-use-related) and/or biotic

90   (i.e. species abundances) variables (represented by nodes). BNI has been widely used to study

91   interaction patterns in molecular biology, medical informatics, economics and social science

92   research [19–23]. However, BNI has only been recently applied to ecological research

93   questions: to microbial community ecology, to the study of assembly rules in invertebrate and

94   bird species, to inform management decisions, and to disentangle direct and indirect

95   associations between environmental variables and species distribution patterns [24–31]. BNI

96   estimates the effect of specific interactions on a focal species considering all the potential

97   direct and indirect relationships among the rest of species in the community. To calculate the

98   effect of every direct and indirect interaction requires the estimation of a very high number of

99   parameters (i.e. assigning a probability to each potential combination of states of every

100 species). This is unfeasible using regression techniques, but is possible with BNI due to its

101 heuristic nature. BNI uses a heuristic search of graphs proposed by different algorithms,

102 which are sequentially compared to the dataset through goodness-of-fit statistics. The graph

103 that best matches the relationships between variables in the data is kept [23]. In addition, BNI

104 decomposes the global probability distribution of the abundance of a focal node (species),

4

105   into a local probability distribution, only affected by a set of conditioning variables [23].

106   Thus, BNI can combine abiotic and biotic information, and consider the potential effects of

107   the composition and relative abundance of every species in a community (hereafter

108   'community structure') on a focal species [23,32]. Based on this information, BNI

109   summarizes the entire community structure by calculating the strength of the effect of 'parent

110   nodes' on 'child nodes'[23,32], and each species can be a parent or child to any other species.

111   Larsen et al. (2012) [29] were the first to show that BNI can be combined with regression

112   techniques to improve predictions of species' relative abundances in a community. They

113   suggested that BNI can be used to identify the most influential parent and child nodes for a

114   target species. Each of these nodes (species) can be entered into SDMs, which are used to

115   predict the target species' distribution and resulting community structure.

116   Although BNI can identify patterns in species associations, it cannot disentangle the two

117   major underlying processes shaping the relative abundance of species in a community, biotic

118   interactions and environmental filtering [33,34]. Biotic interactions can prevent a species

119   from occupying all areas that are environmentally suitable for it (e.g. competition, predation),

120   but at the same time extend the distribution of a given species into areas that would be

121   environmentally unsuitable in the absence of the biotic interaction (e.g. facilitative

122   interactions) [35,36]. Environmental filtering restricts species distributions to sites where

123   environmental conditions are suitable for a given species. This includes environmental

124   conditions that vary at large spatial scales (e.g. climate or lithology), and micro-

125   environmental factors that vary at local scales (e.g. pH, soil humidity or shade). At local

126   scales, the presence of species with certain requirements could indicate the availability of

127   suitable micro-environmental conditions for other species that share similar environmental

128   requirements. Thus, the same pattern of species co-occurrence could be caused by both biotic

129 interactions and micro-environmental filtering. As the use of co-occurrences to study biotic

130 interactions becomes more widespread, it is important to consider how these two processes

131 could be disentangled. A solution to this problem might lie in addressing the ecological

132 requirements of the species involved, as indicated by species traits, and we explore how this

133 could be done.

134 In this study, we hypothesized that the relative abundance of coexisting species can improve

135 the predictions of species geographic distributions made by SDMs. For 1570 assemblages of

136 68 Mediterranean woody plant species, we applied (BNI) to incorporate community structure

137 into SDMs. We assessed the accuracy of predictions of species abundance and community

138 structure based on SDMs with and without information on coexisting species. We used

139 species trait data to interpret the ecological processes potentially underlying the species

140 associations inferred by BNI.

Materials and methods

Overview of the methodology

143 Following [29], we used BNI to infer a) the "overall network" (i.e. considering all species

144 and environmental variables) and select the parent nodes of each focal species and the sign of

145 the inter-specific association; and b) another network for each species, in which only the

146 focal species and environmental factors were included. Then, for each species we fitted two

147 SDMs in which the predictor variables were the parent nodes of the focal species in each of

148 the two networks (hereafter called "Env+Bio" and "Env" predictors respectively). Next, we

149 compared the ability of SDMs with the two predictor types to predict the abundance of each

150 species, and the community structure of each site.

In order to explore the ecological processes underlying the inferred species associations, we classified 68 Mediterranean plant species into two groups, each consisting of similar combinations of life-history traits and ecological requirements (see *Species syndromes* below). We used a chi-square test to assess whether species with positive or negative abundance co-variance tend to be more similar (belong to the same group) or dissimilar (belong to different groups) than expected by chance.

Study site and community structure database

Within the Iberian Peninsula (mainland Portugal and Spain) (S1 Fig), we aimed to select a pool of plant species that do not have extremely different environmental requirements, for which differences in their distributions are entirely driven by the local conditions (for example avoiding the mix of plants from alpine and saltmarsh vegetation). In order to detect effects of the local environment or biotic interactions, the study species needed to differ in subtler aspects of their niche (for example shade or soil moisture requirements). The goal was to obtain assemblages that contain many of the same species, but that have different community structure (i.e. relative abundances). In order to obtain this species pool, we selected a species with restricted habitat requirements but which is broadly distributed throughout the Iberian Peninsula, the cork-oak (*Quercus suber*), and the pool of plant species associated with it. To determine the species associated with *Q. suber*, we used data from the SIVIM database (Sistema de Información de la Vegetación Ibérica y Macaronésica; http://www.sivim.info/sivi/). SIVIM compiles plant community information from phytosociological relevés (hereafter 'plots') consisting of directly submitted data, publications, and unpublished documents (e.g. theses or reports) [37]. For each plot the species composition and relative abundance (percentage of cover) of each species was reported (more details in Methods appendix). We extracted all SIVIM plots in the Iberian

175 Peninsula in which *Q. suber* was present, and the relative abundances of co-occurring species

176 in those plots. This resulted in 1570 plots occupied by 68 plant species (S1 Table).

Environmental variables

178 Each plot was characterized based on the following environmental variables: climate,

179 geology, land use (agriculture or forest-shrub), orientation, and dominant growth-form of the

180 vegetation (trees or shrubs). Climatic variables were obtained from a dynamical downscaling

181 method using the Weather Research and Forecasting model [38] (more details in Methods

182 appendix). Geological information was obtained from the digital geological map data

183 provided by OneGeology-Europe (http://www.onegeology.org/), and each plot was assigned

184 to the dominant geological type, i.e. that which covered ≥70% of the 10 km grid cell in which

185 each plot was located. If no single type fulfilled this requirement, the plot was assigned to a

186 type called "mix" (more details about geological types in Methods appendix). Land use

187 information was extracted from the European Environment Agency website (Corine Land

188 Cover 2006; http://www.eea.europa.eu/). We classified each 10 km UTM (Universal

189 Transverse Mercator coordinates system) grid cell into just one of the two main land uses,

190 agriculture and forest-shrub, based on the dominant land use type, or into a third category

191 (mix) when neither of the land uses covered 70% of the surface. Orientation determines the

192 solar irradiance a site receives, affecting the microclimatic conditions, and resulting in larger

193 hydric stress in south oriented aspects. Plot orientation (North (N), South (S), East (E), West

194 (W), North-East (NE), North-West (NW), South-East (SE), South-West (SW)) was extracted

195 from the information included in each entry of the SIVIM database. The dominant growth

196 form (trees or shrubs) was considered "tree" if the percent of tree cover reported for that plot

197 was more than 50%, and "shrub" if tree cover was less than 25%. If the percentage of tree

198 cover was between 25-50% the plot was considered a 'mix'. In order to account for trends in

199   the data across large geographical distances, the longitude and latitude of the grid cell in

200   which each plot was located was also used as an environmental variable.

Network inference

202   We used BNI to infer relationships between the relative abundance of the 68 plant species

203   across the 1570 plots. BNI can identify which variables (i.e. the relative species abundance or

204   environmental conditions in each plot) significantly condition the probability of finding a

205   given abundance of a given species [39]. The nodes of these networks represent the variables,

206   while the directed edges (links) show the dependency between the two variables involved.

207   Directed edges point from parent to child nodes. As species abundance was recorded as

208   ranges of percent cover, we used multinomial Bayesian networks, in which all the variables

209   are categorical (see details about the criteria to define categories and selection of the

210   algorithms to infer the network in the Methods appendix).

211   Milns et al (2010) pointed out that directionality in a BN is hard to assess as there are

212   multiple configurations of the network that can equally maximize the match with the

213   observed relationships among variables. In order to overcome this issue, we used a two-step

214   process following Sachs et al. [40]: (i) Candidate associations among random variables were

215   identified using the 50% cut-off. The network structure is learned 500 times and the links and

216   directions that consistently (i.e. in > 50% of the runs) show a given direction across the 500

217   runs are selected). The number of runs in which a link showed the same direction was used to

218   quantify the robustness of the direction. (ii) Significant associations were identified based on

219   the threshold approach proposed by Scutari et al. (2013) [41]. For all significant links, we

220   calculated the sign of the interaction using a Jonckheere trend test for ordered factors [42]

221   (see Network inference section in Methods appendix for more details about the order of the

222 categorical variables). We partially constrained the inference by not allowing the species

223 abundance to influence environmental variables and by not allowing any environmental

224 variable to influence the following variables: the temperature in the warmest quarter of the

225 year, mean annual precipitation, geological type and orientation. All analyses were performed

226 using the package "bnlearn" implemented in the software R version 3.1.2 [32].

Similarity in species life-history traits and ecological requirements: species syndromes

229 The plant species that currently co-exist in the Mediterranean basin are a mixture of species

230 that originated at different times and under different environments [43]. The dry, hot

231 summers of the Mediterranean climate originated in the late Pliocene [44]. At that time, most

232 of the plants in the Mediterranean that required summer rain became extinct and

233 predominantly those species with traits that confer tolerance to summer drought persisted

234 until today [45–48]. However, other plant lineages that also currently inhabit Mediterranean

235 areas originated more recently and have evolved under Mediterranean climate [44].

236 Differences in the selective pressures experienced by these two groups of Mediterranean

237 plant lineages has resulted in different morphological-functional trait combinations and

238 regeneration niche requirements, which we term "syndromes" [43,46]. The recent lineages

239 (with a Quaternary syndrome) are characterized by non-sclerophyllous leaves, facultative

240 summer deciduousness, hermaphroditic, large, colored flowers, small seeds and pollination

241 by large insects. Ancient lineages (with a Tertiary syndrome) are evergreen plants with

242 sclerophyllous leaves, reduced-greenish-unisexual flowers, medium to large seeds, fleshy

243 fruits dispersed by vertebrates, and pollination by wind or small insects [49].

Most of the plant species considered in this study (60 out of 68) belong to genera that have been previously assigned to one of these two syndromes according to the outcome of a principal component analysis based on their ecological traits and regeneration niche requirements [43,46] (33 as Tertiary (T) and 27 as Quaternary (Q); S1 Table). We therefore restricted this part of the analysis to those 60 species. We used a χ^2 test to assess whether positive abundance covariance between species that have similar (the same syndrome) or dissimilar (different syndromes) ecological requirements occur more frequently than expected by chance.

Species distribution models

We fitted SDMs to each of the 68 species. Following Larsen et al. (2012) [29] , we used the network structure learned using BNI to identify the parent nodes of each species and used those nodes as explanatory variables. We used the mean percent of cover of each species in each plot as the dependent variable to construct a generalized additive model (GAM) with a binomial error distribution, including the longitude and latitude interaction of the 10 km grid cell as a smoothing term [50–52]. Cross-validation was used to estimate the optimal amount of smoothing (λ). During cross-validation, the optimal λ, and the effective degrees of freedom was obtained by choosing different values of λ and then minimizing the sum of squares of the linear regression penalized by the smoothing splines. This was performed using "mgcv" package implemented in the software R version 3.1.2 (Wood 2011). We fitted the GAMs using all the parent nodes of each focal species identified by BNI (usually 1-4 variables per species, S2 Table). If the species did not have any parent node, the GAM was fitted using the intercept as the only explanatory variable (indicated as ~ 1, in S2 Table). For longitude, latitude, mean temperature of warmest quarter and annual mean precipitation we used continuous data in the GAMs. As we aim to compare predictions made with the best available

268   information on the drivers of each species distribution in the presence and absence of species

269   co-occurrence data, the environmental predictors may differ between Env and Env+Bio

270   SDMs for a given species (S2 Table). Finally, we also asked whether the models used

271   following this procedure predicted the observed abundances better than the models based on

272   randomly selected variables (Methods in appendix).

273

Comparing SDMs with "Env+Bio" and "Env" variables

275   Following Larsen et al. (2012) [29], for each species we fitted two models using "Env+Bio"

276   and "Env" predictor variables separately. To identify "Env+Bio" variables we inferred a

277   single BN considering all species relative abundances and environmental variables, so that

278   either species or environmental variables could be parent nodes of the focal species. For

279   "Env" variables, we inferred network structure for each species, which contained the focal

280   species' relative abundance and all the environmental variables. In this way, the parent nodes

281   of each species could only be environmental variables.

282   The two sets of predictor variables represent different knowledge situations. 'Env' asks

283   which environmental variables we would think are important if we knew nothing about co-

284   occurring species. Env+Bio asks which environmental variables and species co-occurrences

285   are important when we have knowledge of both of these factors.

286   In order to evaluate the explanatory power of the SDMs with and without biotic data, we

287   randomly selected two thirds of the plots in which each species was present to construct

288   GAMs with the two sets of relevant explanatory variables ('calibration plots'). The same

289   plots were used to evaluate SDMs with and without biotic data. In order to account for

290   variation in the number of explanatory variables used in "Env+Bio" and "Env" models, we

291 calculated the Akaike Information Criterion (AIC) of each model, which penalizes against the

292 addition of explanatory variables. We compared AICs between models using a paired t-test.

293 We also calculated the percentage of variance explained by the two GAMs as a proxy for the

294 absolute quality of the models. The analyses were performed using the R package "MASS"

295 and "mgcv" implemented in the software R version 3.1.2 [53,54].

296 In order to evaluate the predictive power of SDMs, we used the GAMs constructed with

297 calibration plots to predict the community structure (species composition and abundance) in

298 the remaining one third of the plots ('validation plots'). We calculated the Spearman

299 correlation coefficient (rho) between the observed species abundance and the abundances

300 predicted by the "Env+Bio" and "Env" predictors. A paired t-test on the rho values was used

301 to test whether the predictions by the GAMs using "Env+Bio" or "Env" predictors correlate

302 better with the observed abundances. Finally, we used the Bray-Curtis (BC) dissimilarity

303 index to estimate the similarity between the predicted and observed community structure in

304 each of the validation plots. Hereafter we will use the similarity index 1-BC (where 1 is the

305 most similar, implying better predictions and 0 the most dissimilar and implying worse

306 predictions) and refer it as "BC similarity index". A paired t-test was used to test whether the

307 BC similarity index was higher when the "Env+Bio" or "Env" predictors were used. These

308 analyses were performed using the R package "vegan" implemented in the software R

309 version 3.1.2 [55].

310 # Results

311 ## Overall BNI network

The overall network, including all species and environmental variables, contained a total of 138 significant links (Fig 1), 104 of which were positive (75%) and 20 (15%) negative. For 14 links the Jonckheere trend was not strong enough to assign a sign. Of the 138 significant links, 75 occurred between species. Most species-species links (95%) were positive, indicating that the probability of finding a higher abundance of one species increases when the other species is also abundant. Only four links between species were negative (Table S3). On average, each species had 1.94 ± 0.08 (mean $\pm$ SE) parent nodes and 1.29 ± 0.18 children nodes.
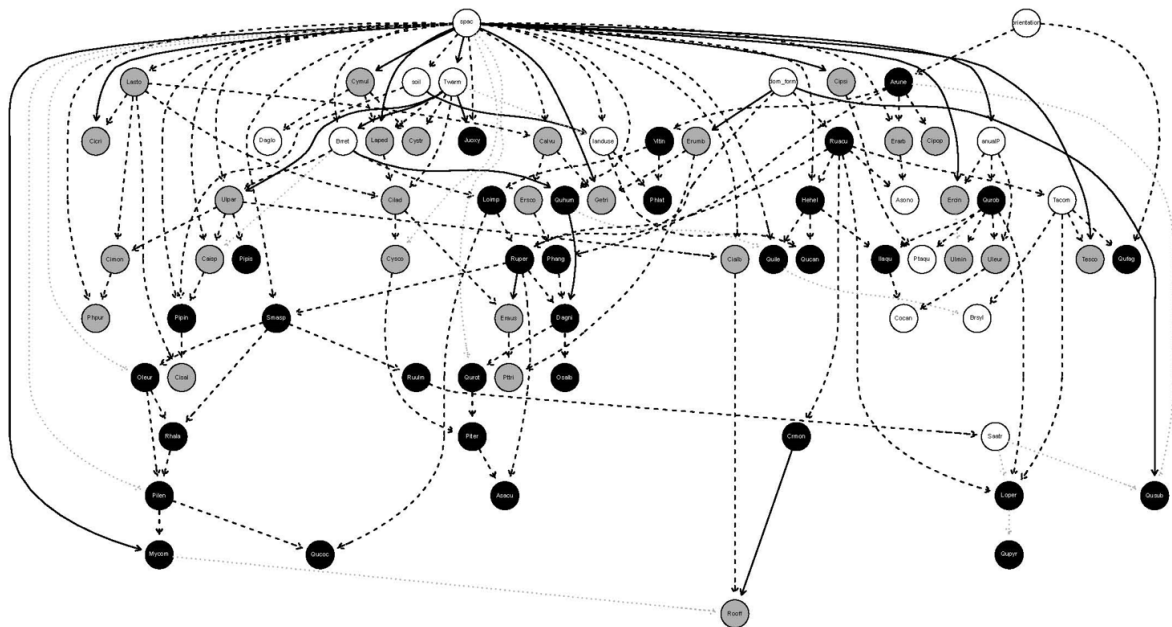
Figure 1. Network structure learned using Bayesian network inference. Only significant links are presented, and grey lines indicating links with no sign was detected. Grey and black circles represent species with a Quaternary and Tertiary syndrome respectively. White circles are either environmental variables (mean temperature in the warmest quarter of the year (Twarm), annual precipitation (anualP), soil types (soil), land use (landuse), orientation (orientation), dominant form (dom_form) and spatial location (spac)) or species with no syndrome associated. Continuous and dashed lines represent negative and positive

14

328   associations respectively. Complete names for species are provided in the appendix and

329   environmental variable categories in the methods section.

The accuracy of SDMs when informed by community structure

331   Across all species, the "Env+Bio" predictors resulted in models of species abundance that

332   have greater explanatory power than did the "Env" predictors (mean (±SE) decrement in AIC

333   = -146 ± 100; t_{paired} = -3.97, df = 67, p-value < 0.0001) (S1 Table). Across all species, the

334   models of species abundance using "Env+Bio" predictors explains a slight but significantly

335   higher percentage of deviance than the models using "Env" predictors, (mean increment in

336   the percentage of deviance explained (±SE) = 1.5 % ± 0.42; t_{paired} = 3.54, df = 67, p-value<

337   0.001), but there was considerable variation across species, ranging from species for which

338   the model using "Env+Bio" predictors decreased the deviance explained by 6%

339   (*Pterospartum tridentatum*) to species in which the model using "Env+Bio" predictors

340   increased the deviance explained by 15% (*Salix atrocinerea*). The models using "Env+Bio"

341   predictors also predicted the observed abundances better than the models based on randomly

342   selected variables; on average, Env+Bio predictors explained a higher percentage of deviance

343   (3.18 % ± 1.29; t_{paired} = 2.36, df = 67, p-value= 0.01) (more details in Methods in appendix).

344   Including community structure in SDMs improved the accuracy of the species' observed

345   abundance predictions, as there was a slight but significant higher correlation between the

346   observed and the predicted abundances using "Env+Bio" predictors than using "Env"

347   predictors (mean increment in rho (±SE) = 0.02 ± 0.006; t_{paired} = -3.1, df = 67, p-value <

348   0.002) (Fig 2). However, there were six species for which the models using "Env+Bio"

349   predictors resulted in an increment of the Spearman correlation coefficient above 0.10,

350   indicating a considerably more accurate prediction of these species' abundances (S2 Table).

351   Models using "Env+Bio" predictors also improved the predictions of the whole community

352   structure in each plot. Overall, the Bray-Curtis similarity index was higher when using

353   "Env+Bio" predictors than when using "Env" predictors (mean increment in Bray-Curtis

354   similarity index (±SE) = 0.1 ± 0.004; N = 524; t_{paired} = 2.1861, df = 523, p-value < 0.0001)

355   (Fig 3).

Rho correlation coefficient
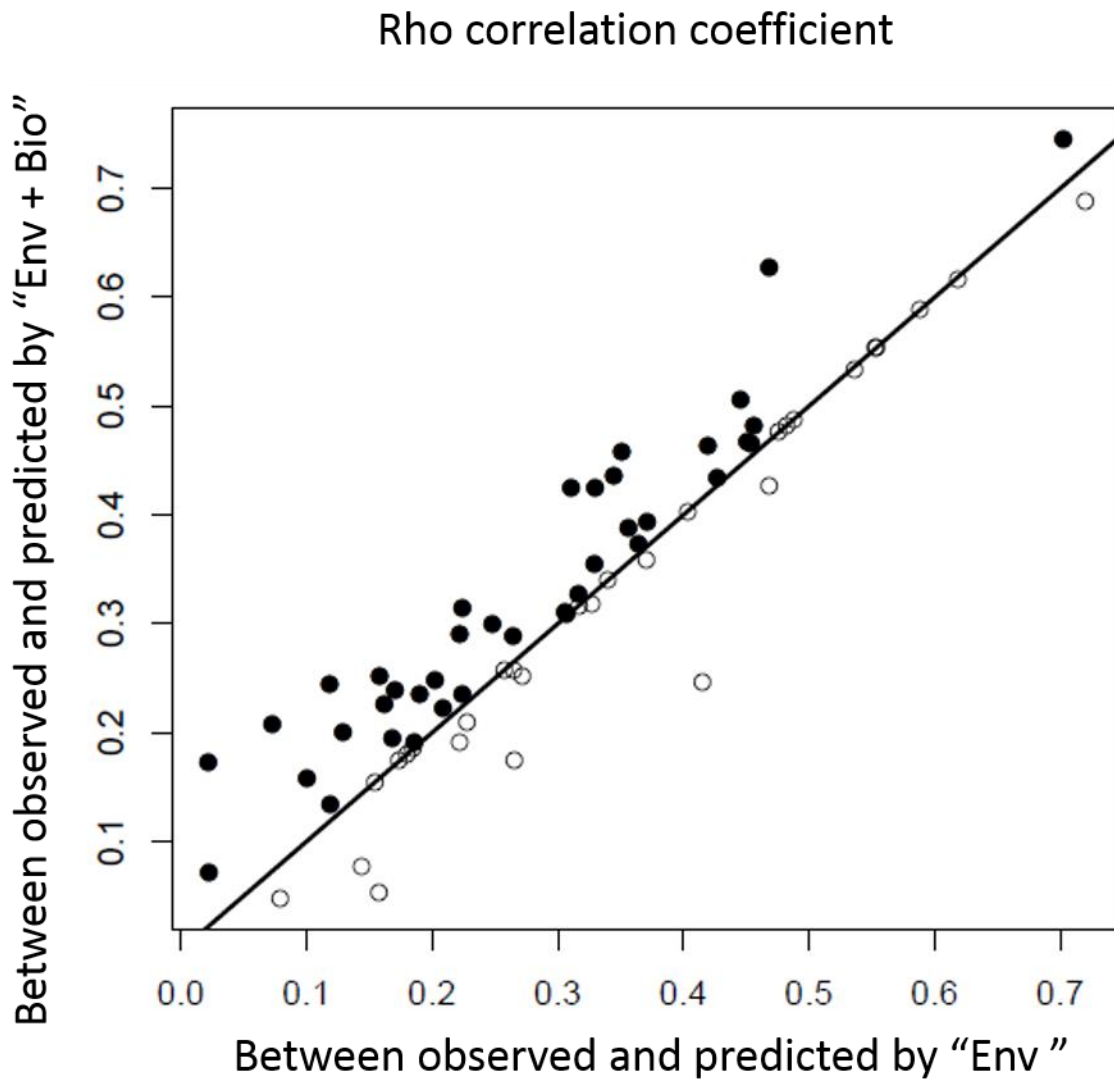
356

357   **Figure 2. Correlation between the prediction of "Env" and "Env + Bio" models.** Rho

358   coefficient of the correlation between the observed species abundances and abundances

359   predicted by "Env+Bio" vs. the correlation coefficient between abundances observed and

16

360 predicted by "Env" models, for the 68 species. Black points above the line represent species

361 with higher Spearman's rho correlation coefficients values using "Env+Bio" rather than

362 "Env" predictors. The opposite is true for white points below the line.

Bray-Curtis similarity index
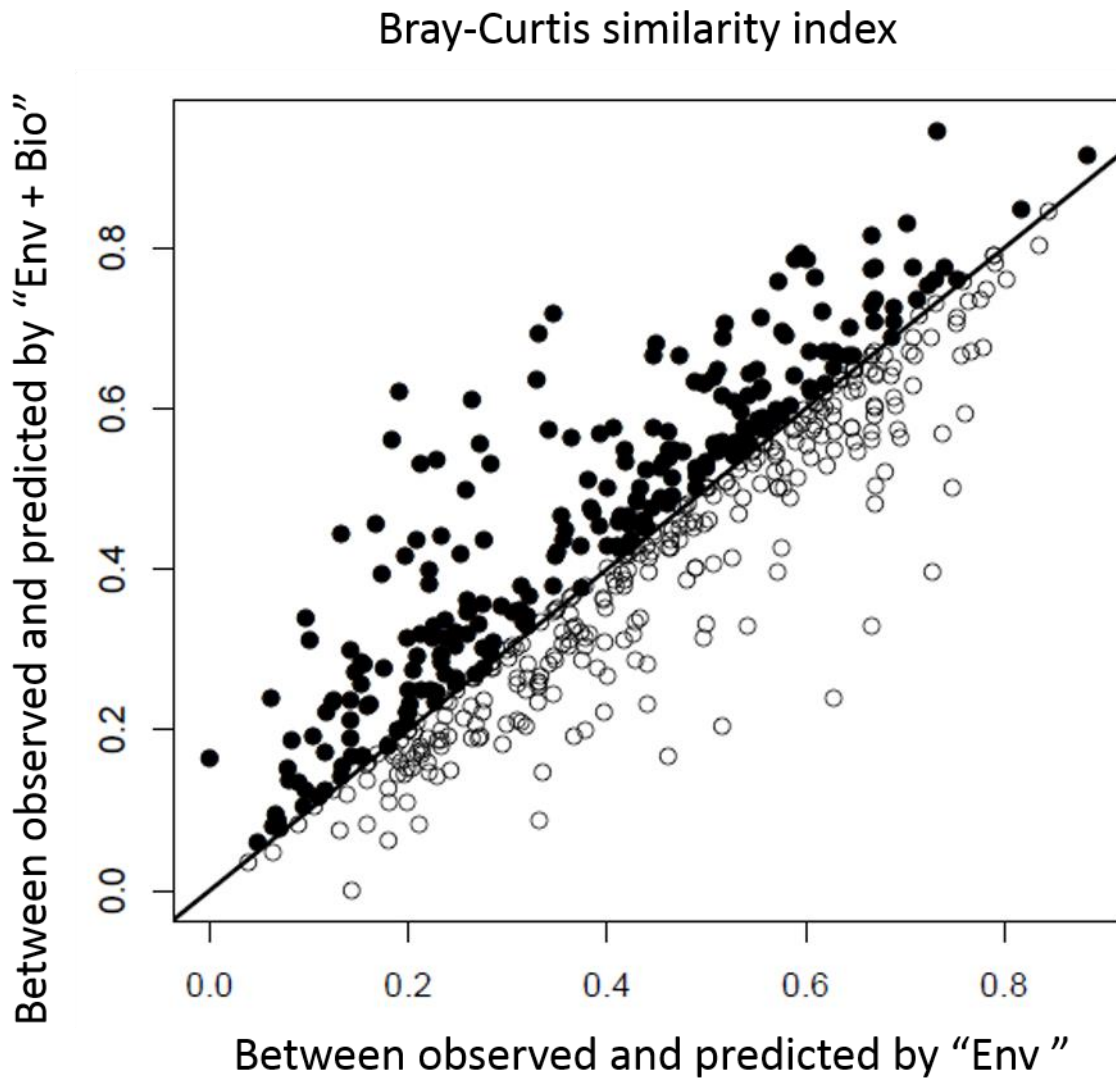
363

364 **Figure 3. Correlation between the Bray-Curtis similarity index using "Env" and "Env +**

365 **Bio models".** Bray-Curtis similarity index between the observed community structure and the

366 community structure predicted by "Env+Bio" vs. the similarity index between the observed

367 community structure and that predicted by "Env" models, for the 524 validation plots. Values

368 of Bray-Curtis similarity index closer to 1 imply that community structure is predicted more

369 accurately and values closer to 0 indicate less accurate predictions. Black points above the

370 line represent plots with higher similarity between the observed values and those predicted

371 using the "Env+Bio" rather than the "Env" predictors. The opposite is true for white points

372 below the line.

Potential ecological processes underlying abundance covariance between species

375 The links between species inferred by BNI do not occur between random pairs of species.

376 Positive links between species with the same syndrome (Tertiary-Tertiary (TT) or

377 Quaternary-Quaternary (QQ)) are significantly more frequent than expected by chance ($\chi2 =$

378 26.68, df = 1, p-value < 0.0001). The links were significantly more frequent between species

379 with the same syndrome than between species with a different syndrome (Number of links:

380 QQ = 20, TT = 32, QT = 4, TQ = 7; $\chi2 = 63$, df = 3, p-value < 0.0001), and especially

381 between those sharing a Tertiary syndrome (Table S3). Only four of the significant links were

382 negative, which prevented us from performing any statistical inference for negative links.

Discussion

384 For 80% of the 68 species, including information on community structure in SDMs appears

385 to improve predictions of species distributions. The improvements in SDM performance are

386 of a similar magnitude to those recently found by [56], who used BNs to directly model biotic

387 interactions and shared habitat requirements' relationships among species in a community.

388 Positive associations between Mediterranean woody plants tend to occur between

389 ecologically similar (i.e. 'Tertiary') species. This association pattern suggests that positive

390 associations might be driven by a match between the requirements of similar species and the

18

391 presence of environmental conditions, in particular shade and moisture. The species

392 associations we observe appear to reflect the conditions that occur within vegetation plots,

393 and so at a much finer spatial resolution than is usually possible to study with most sources of

394 climate data. Moreover, we selected a study system in which the macro-climatic conditions

395 did not vary greatly. Thus, we propose that species distribution predictions might have been

396 improved because information about the community structure acts as a proxy for micro-

397 environmental conditions, for which direct data are not available.

Incorporating community structure in SDMs

399 SDM predictions of species distribution and community structure improved when

400 information on community structure was included. Several of the species for which

401 community structure information improved SDMs have specific habitat requirements.

402 *Corynephorus canescens* requires bare and sandy soils [57], *Salix atrocinerea* occupies river

403 banks and permanently wet soils [58], and *Quercus canariensis* occupies shaded and humid

404 canyons [59]. By contrast, species for which community structure information does not

405 improve SDMs often have wide distributions in the Mediterranean region (*Quercus ilex* [59])

406 or are highly generalist and exhibit invasive behaviour in non-native regions (*Brachypodium*

407 *sylvaticum, Hedera helix* [60–62]) (S2 Table). Therefore, the micro-environmental data

408 added by community structure might be especially informative for species with restrictive

409 ecological requirements, and less relevant for more generalist species.

410 Information about micro-climatic conditions is rarely available across large spatial extents

411 such as the Iberian Peninsula (though climate data can be downscaled [63]). However,

412 information about the community structure of coexisting plant species is often available

19

413 across large extents, and can act as a substitute for micro-climatic information that cannot be
414 otherwise included in SDMs.

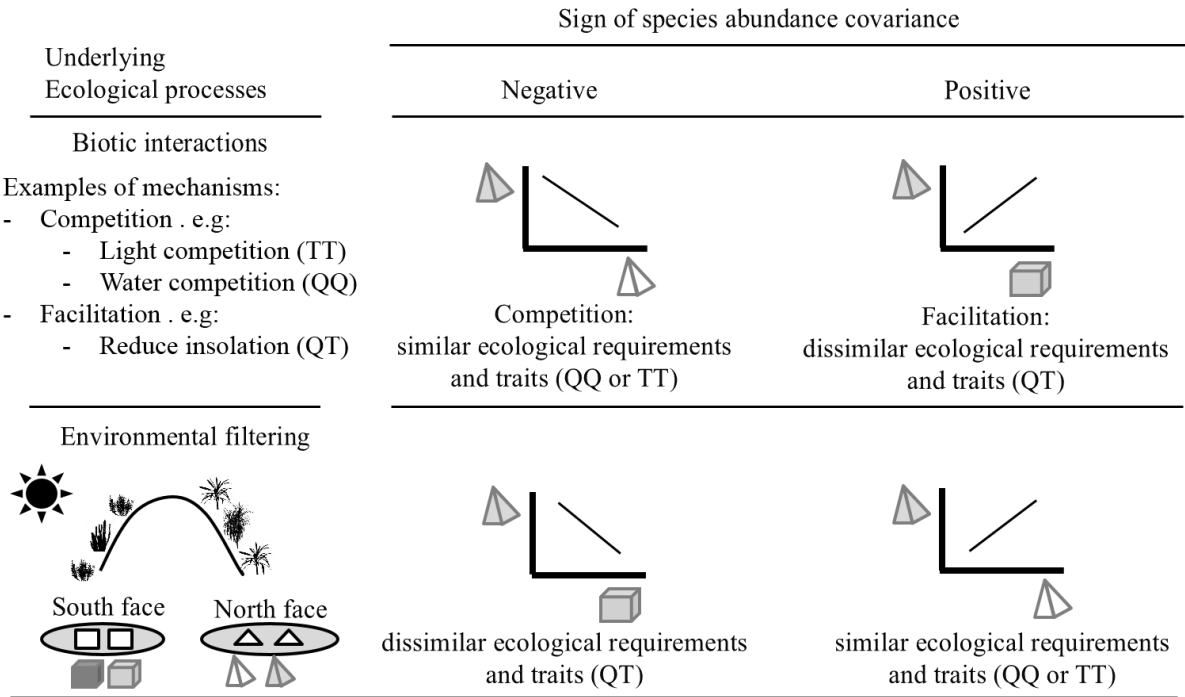## Using traits to explore ecological processes underlying abundance covariance between species

417 We caution against simply assuming that co-occurrence patterns reflect biotic interactions.
418 Instead, we suggest that asking whether associations occur between species with similar or
419 dissimilar ecological requirements can provide insight into the predominance of biotic
420 interactions and environmental filtering. Community assembly theory suggests that biotic
421 interactions and environmental filtering can affect the distribution of trait values within
422 communities (i.e. by permitting different sets of species to co-exist). Environmental filtering
423 leads to coexisting species having similar traits as a result of shared ecological tolerances
424 [64,65]. However, non-consumptive interactions like competition and facilitation can have
425 varying effects on traits, depending on the traits and details of the interaction. For example,
426 most studies focusing on competition have been based on the common assumption that
427 species with similar ecological strategies compete more intensely for resources than species
428 with different strategies [66] resulting in co-existing species having different traits. On the
429 other hand, competition can magnify the effects of environmental filtering by causing species
430 with similar traits to co-occur. For example, competition for light in shaded environments can
431 lead to species with the same light-adaptation traits outcompeting species with different traits
432 [67]. Positive biotic interactions such as facilitation (i.e. one species directly promotes the
433 presence of another [68]) can result in a positive association between ecologically dissimilar
434 species, because this ecological process is frequent between phylogenetically distant plant
435 species [35,69,70]. Alternatively, facilitative interactions driven by shared mutualists such as

436 pollinators, can result in a positive association between plants with similar floral traits, as

437 similar flowers enhance the attraction of shared pollinators [71]. The potential for different

438 trait co-occurrence patterns to arise from the same type of biotic interactions therefore adds

439 complexity to the interpretation of trait data to explain species co-occurrence. However, we

440 suggest that considering traits appropriate to the situation can be highly informative when

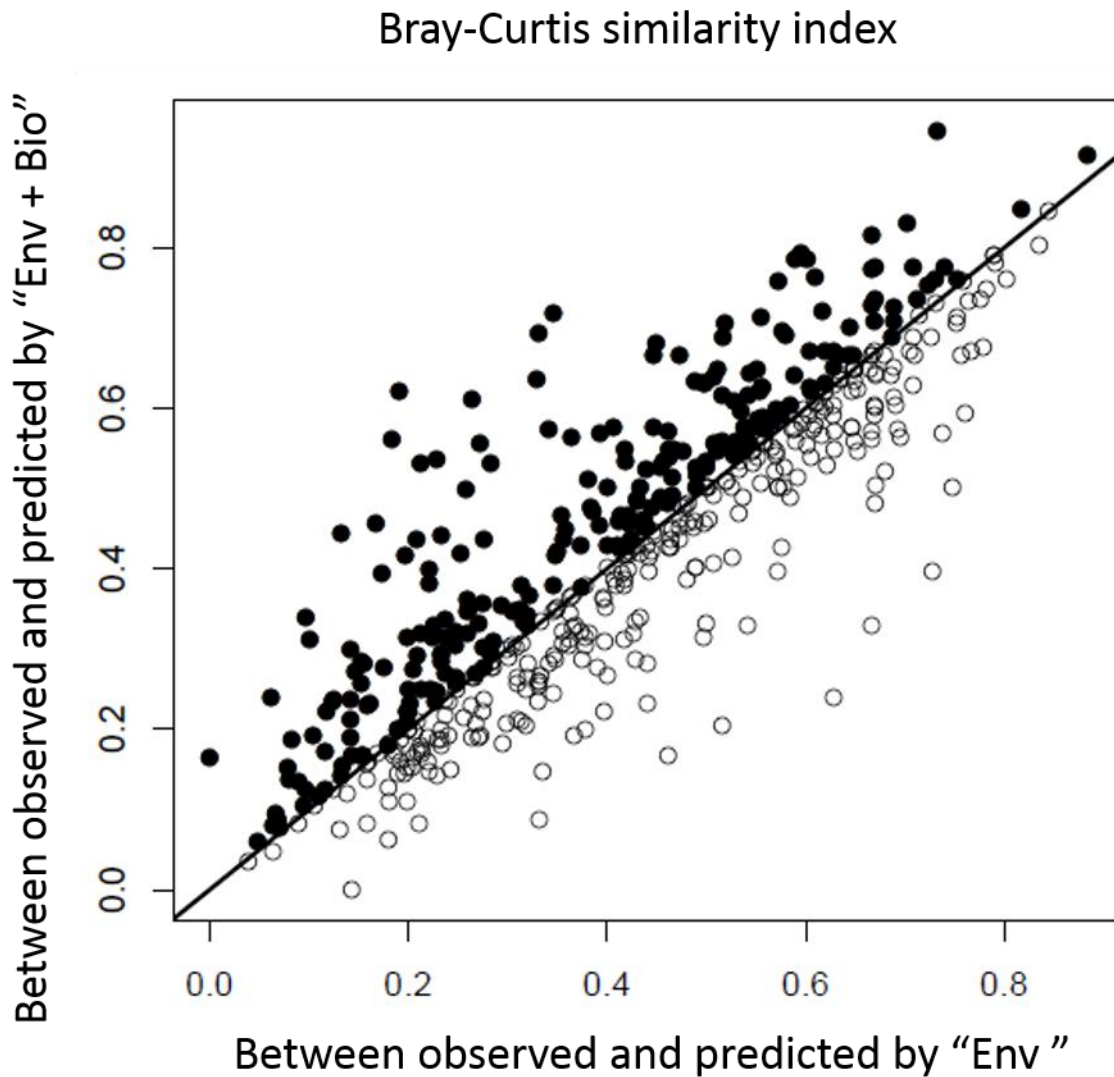441 interpreting causes of co-occurrence patterns.

442

443 The tendency for Tertiary species (which are associated with humid, shaded areas) to co-

444 occur, suggests that their presence can provide information about micro-environmental

445 conditions, specifically shade and soil moisture (Fig 4). An alternative explanation could be

446 that the species are facilitating each other's reproduction by attracting shared pollinators [71].

447 However, only two of the 14 morphological and functional traits used to define Quaternary

448 and Tertiary syndromes relate to the pollination syndromes [49]. In addition, the plants

449 studied showed neither entomophily or anemophilia, so there was little inter-specific

450 variation in floral morphology. Therefore, although we cannot completely rule out the

451 possibility that facilitation through enhanced attraction of shared pollinators underlies the co-

452 occurrence of ecologically similar plant species in our study, we consider it unlikely.

453

	Sign of species abundance covariance	
Underlying Ecological processes	Negative	Positive
Biotic interactions Examples of mechanisms: - Competition . e.g: - Light competition (TT) - Water competition (QQ) - Facilitation . e.g: - Reduce insolation (QT)	Competition: similar ecological requirements and traits (QQ or TT)	Facilitation: dissimilar ecological requirements and traits (QT)
Environmental filtering South face North face	dissimilar ecological requirements and traits (QT)	similar ecological requirements and traits (QQ or TT)

454

Bray-Curtis similarity index

Figure 4. Expected covariance between species involved in biotic interactions and environmental filtering. The combination of 3-d shapes and colors represent four different species. Species with similar requirements (syndromes) are represented by the same shape (pyramids: Tertiary (T), cubes: Quaternary(Q)), but distinct colors. Environmental filters are represented as grey ellipses in which only species with certain traits can survive (e.g. moist and shaded environments on north facing slopes where species with a tertiary syndrome can survive, or sunny environments on south facing slopes where quaternary species can survive: the 3-d shapes must match the shape of the ellipse). In the case of negative abundance covariance, competition is expected to be more intense between species with similar traits

23

465 and ecological requirements resulting in spatial segregation between species with similar

466 requirements and traits, while environmental filtering will result in spatial segregation

467 between species with dissimilar requirements and traits. In the case of positive abundance

468 covariance, facilitation promotes the co-occurrence between species with dissimilar

469 requirements and traits, while habitat filtering results in the co-occurrence of species with

470 similar requirements and traits.

471

472 Although environmental filtering appears to explain the co-occurrence patterns found,

473 environmental filtering would also be expected to result in negative links among species that

474 inhabit different habitat types, with the same frequency as positive links [72,73]. The

475 predominance of positive links in our network (Fig 1 and TS 3) might be because the study

476 system is defined by the presence of *Quercus suber* which has relatively restricted habitat

477 requirements, resulting in insufficient environmental variation to reveal strong segregation

478 between Quaternary and Tertiary species. The predominance of positive species associations

479 has been also reported in other studies of species associations [25,74–76].

480 Although our results suggest that environmental filtering drives species associations, plant-

481 plant facilitation (positive interactions) between species with Quaternary and Tertiary

482 syndromes is known to have played a crucial role in the persistence of the latter [46]. It may

483 be possible to detect facilitation at an even finer spatial resolution than we studied.

484 Quaternary-Tertiary facilitation may often take the form of improved seedling recruitment

485 under adult plants, which might be apparent if networks are created using plant abundance

486 data on the scale of a few meters. The ecological processes captured by network inference

487 may therefore depend on the spatial resolution of the analysis.

488 In conclusion, we show how BNIs can improve understanding of species distributions, and

489 how this could improve SDMs. The network structure provided by the BNI can be combined

490 with ecological trait data to explore potential processes underlying species associations.

491 However, these interpretations should be made cautiously, given that different mechanisms

492 could result in similar patterns. Taking this into account, we consider it likely that species

493 abundance in Mediterranean woody plant communities, at the resolution studied, arise from

494 micro-environmental associations that are rarely detectable using standard SDM approaches.

Acknowledgments

507 **DATA ACCESSIBILITY STATEMENT**

508 The datasets supporting this article have been uploaded as part of the supplementary material.

509 **REFERENCES**

510 1. Parmesan C. Ecological and evolutionary responses to recent climate change. Annual
511   Review of Ecology, Evolution, and Systematics. JSTOR; 2006;637–69.
512

513 2. Peterson AT. Predicting the geography of species' invasions via ecological niche
514   modeling. The quarterly review of biology. The University of Chicago Press;
515   2003;78(4):419–33.
516

517 3. Heikkinen RK, Luoto M, Virkkala R, Pearson RG, Körber J-H. Biotic interactions
518   improve prediction of boreal bird distributions at macro-scales. Global Ecology and
519   Biogeography. Wiley Online Library; 2007;16(6):754–63.
520

521 4. Keith DA, Akçakaya HR, Thuiller W, Midgley GF, Pearson RG, Phillips SJ, et al.
522   Predicting extinction risks under climate change: coupling stochastic population models
523   with dynamic bioclimatic habitat models. Biology Letters. The Royal Society;
524   2008;4(5):560–3.
525

526 5. Anderson BJ, Akçakaya HR, Araújo MB, Fordham DA, Martinez-Meyer E, Thuiller W,
527   et al. Dynamics of range margins for metapopulations under climate change. Proc Biol
528   Sci. 2009;276(1661):1415–20.
529

530 6. Valladares F, Matesanz S, Guilhaumon F, Araújo MB, Balaguer L, Benito-Garzón M, et
531   al. The effects of phenotypic plasticity and local adaptation on forecasts of species range
532   shifts under climate change. Ecology letters. Wiley Online Library; 2014;17(11):1351–
533   64.
534

535 7. Kearney M, Porter W. Mechanistic niche modelling: combining physiological and
536   spatial data to predict species' ranges. Ecol Lett. 2009;12(4):334–50.
537

538 8. Araújo MB, Luoto M. The importance of biotic interactions for modelling species
539   distributions under climate change. Global Ecology and Biogeography. Wiley Online
540   Library; 2007;16(6):743–53.
541

542 9. Scott J, Heglund P, Morrison M, Haufler J, Raphael M, Wall W, et al. Predicting
543   species occurrences: issues of scale and accuracy. Predicting species occurrences: Issues
544   of scale and accuracy. Island Press Washington D. C.; 2002.
545

546 10. Meineri E, Hylander K. Fine-grain, large-domain climate models based on climate
547   station and comprehensive topographic information improve microrefugia detection.
548   Ecography. Wiley Online Library; 2017;40(8):1003–13.
549

550 11. Kissling WD, Dormann CF, Groeneveld J, Hickler T, Kühn I, McInerny GJ, et al.
551   Towards novel approaches to modelling biotic interactions in multispecies assemblages
552   at large spatial extents. Journal of Biogeography. Wiley Online Library;
553   2012;39(12):2163–78.
554

555 12. Ovaskainen O, Hottola J, Siitonen J. Modeling species co-occurrence by multivariate
556   logistic regression generates new hypotheses on fungal interactions. Ecology. Eco Soc
557   America; 2010;91(9):2514–21.
558

559   13.   Sebastián-González E, Sánchez-Zapata JA, Botella F, Ovaskainen O. Testing the
560   heterospecific attraction hypothesis with time-series data on species co-occurrence.
561   Proceedings of the Royal Society B: Biological Sciences. The Royal Society;
562   2010;277(1696):2983–90.

564   14.   Meier ES, Kienast F, Pearman PB, Svenning J-C, Thuiller W, Araújo MB, et al. Biotic
565   and abiotic variables show little redundancy in explaining tree species distributions.
566   Ecography. Wiley Online Library; 2010;33(6):1038–48.

568   15.   Pollock LJ, Tingley R, Morris WK, Golding N, O'Hara RB, Parris KM, et al.
569   Understanding co-occurrence by modelling species simultaneously with a Joint Species
570   Distribution Model (JSDM). Methods in Ecology and Evolution. Wiley Online Library;
571   2014;5(5):397–406.

573   16.   Pellissier L, Pradervand J-N, Pottier J, Dubuis A, Maiorano L, Guisan A. Climate-based
574   empirical models show biased predictions of butterfly communities along
575   environmental gradients. Ecography. Wiley Online Library; 2012;35(8):684–92.

577   17.   Wisz MS, Pottier J, Kissling WD, Pellissier L, Lenoir J, Damgaard CF, et al. The role of
578   biotic interactions in shaping distributions and realised assemblages of species:
579   implications for species distribution modelling. Biological Reviews. Wiley Online
580   Library; 2013;88(1):15–30.

582   18.   Levine JM, Bascompte J, Adler PB, Allesina S. Beyond pairwise mechanisms of species
583   coexistence in complex communities. Nature. Nature Research; 2017;546(7656):56–64.

585   19.   Goulding R, Jayasuriya N, Horan E. A Bayesian network model to assess the public
586   health risk associated with wet weather sewer overflows discharging into waterways.
587   Water Research. Elsevier; 2012;46(16):4933–40.

589   20.   Chai LE, Loh SK, Low ST, Mohamad MS, Deris S, Zakaria Z. A review on the
590   computational approaches for gene regulatory network construction. Computers in
591   Biology and Medicine. Elsevier; 2014;48:55–65.

593   21.   Kupfer P, Huber R, Weber M, Vlaic S, Häupl T, Koczan D, et al. First-time application
594   of multi-stimuli network inference to synovial fibroblasts of rheumatoid arthritis
595   patients. BMC medical genomics. BioMed Central Ltd; 2014;7(1):40.

597   22.   Kuppuswamy U, Ananthasubramanian S, Wang Y, Balakrishnan N, Ganapathiraju M.
598   Predicting gene ontology annotations of orphan GWAS genes using protein-protein
599   interactions. Algorithms for Molecular Biology. 2014;9(1).

601   23.   Scutari M, Denis J-B. Bayesian Networks: With Examples in R. CRC Press; 2014.

603   24.   Mori T, Saitoh T. Flood disturbance and predator-prey effects on regional gradients in
604   species diversity. Ecology. Eco Soc America; 2014;95(1):132–41.
605

25. Milns I, Beale CM, Smith VA. Revealing ecological networks using Bayesian network inference algorithms. Ecology. Eco Soc America; 2010;91(7):1892–9.

26. Marcot BG, Holthausen RS, Raphael MG, Rowland MM, Wisdom MJ. Using Bayesian belief networks to evaluate fish and wildlife population viability under land management alternatives from an environmental impact statement. Forest ecology and management. Elsevier; 2001;153(1):29–42.

27. Wilson AJ, Ribeiro R, Boinas F. Use of a Bayesian network model to identify factors associated with the presence of the tick Ornithodoros erraticus on pig farms in southern Portugal. Preventive veterinary medicine. Elsevier; 2013;110(1):45–53.

28. Douglas SJ, Newton AC. Evaluation of Bayesian networks for modelling habitat suitability and management of a protected area. Journal for Nature Conservation. Elsevier; 2014;22(3):235–46.

29. Larsen PE, Field D, Gilbert JA. Predicting bacterial community assemblages using an artificial neural network approach. Nature methods. Nature Publishing Group; 2012;9(6):621–5.

30. Shafiei M, Dunn KA, Chipman H, Gu H, Bielawski JP. BiomeNet: A Bayesian Model for Inference of Metabolic Divergence among Microbial Communities. PLoS computational biology. Public Library of Science; 2014;10(11):e1003918.

31. Faust K, Raes J. Microbial interactions: from networks to models. Nature Reviews Microbiology. Nature Publishing Group; 2012;10(8):538–50.

32. Scutari M. Learning Bayesian Networks with the bnlearn R Package. Journal of Statistical Software. American Statistical Association; 2010;35(i03).

33. Weiher E, Keddy PA. The assembly of experimental wetland plant communities. Oikos. JSTOR; 1995;323–35.

34. Wilson JB. Assembly rules in plant communities. Ecological assembly rules: perspectives, advances, retreats. Cambridge University Press Cambridge; 1999.

35. Valiente-Banuet A, Verdú M. Plant Facilitation and Phylogenetics. Annual Review of Ecology, Evolution, and Systematics. Annual Reviews; 2013;44:347–66.

36. Bruno JF, Stachowicz JJ, Bertness MD. Inclusion of facilitation into ecological theory. Trends in Ecology & Evolution. Elsevier; 2003;18(3):119–25.

37. Font X, Rodriguez-Rojo MP, Acedo C, Biurrun I, Fernández-González F, Lence C, et al. SIVIM: an on-line database of Iberian and Macaronesian vegetation. Wald ökologie, Landschaftsforschung und Naturschutz. 2010;8:15–22.

38. Skamarock W, Klemp J, Dudhia J, Gill D, Barker D, Duda M, et al. A description of the advanced research WRF version 2 NCAR Tech. Note NCAR/TN-468+STR, 123 pp.

654   2005;

656 39. Needham CJ, Bradford JR, Bulpitt AJ, Westhead DR. Inference in Bayesian networks.
657   Nature biotechnology. New York, NY: Nature Pub. Co., 1996-; 2006;24(1):51–4.

659 40. Sachs K, Perez O, Pe'er D, Lauffenburger DA, Nolan GP. Causal protein-signaling
660   networks derived from multiparameter single-cell data. Science. American Association
661   for the Advancement of Science; 2005;308(5721):523–9.

663 41. Scutari M, Nagarajan R. Identifying significant edges in graphical models of molecular
664   networks. Artificial Intelligence in Medicine. Elsevier; 2013;57(3):207–17.

666 42. Jonckheere AR. A distribution-free k-sample test against ordered alternatives.
667   Biometrika. JSTOR; 1954;41:133–45.

669 43. Herrera CM. Historical effects and sorting processes as explanations for contemporary
670   ecological patterns: character syndromes in Mediterranean woody plants. American
671   Naturalist. JSTOR; 1992;421–46.

673 44. Axelrod DI. History of the Mediterranean ecosystem in California. Mediterranean type
674   ecosystems. Springer; 1973.

676 45. Ackerly DD. Community assembly, niche conservatism, and adaptive evolution in
677   changing environments. International Journal of Plant Sciences. JSTOR;
678   2003;164(S3):S165–S184.

680 46. Valiente-Banuet A, Rumebe AV, Verdú M, Callaway RM. Modern Quaternary plant
681   lineages promote diversity through facilitation of ancient Tertiary lineages. Proceedings
682   of the National Academy of Sciences of the United States of America. National Acad
683   Sciences; 2006;103(45):16812–7.

685 47. Axelrod DI. Evolution and biogeography of Madrean-Tethyan sclerophyll vegetation.
686   Annals of the Missouri Botanical Garden. JSTOR; 1975;280–334.

688 48. Palamarev E. Paleobotanical evidences of the Tertiary history and origin of the
689   Mediterranean sclerophyll dendroflora. Plant Systematics and evolution. Springer;
690   1989;162(1-4):93–107.

692 49. Herrera CM. Tipos morfológicos y funcionales en plantas del matorral mediterráneo del
693   sur de España. Studia Oecologica. 1984;5:7–34.

695 50. Hastie TJ, Tibshirani RJ. Generalized additive models. CRC Press; 1990.

697 51. F Dormann C, M McPherson J, B Araújo M, Bivand R, Bolliger J, Carl G, et al.
698   Methods to account for spatial autocorrelation in the analysis of species distributional
699   data: a review. Ecography. Wiley Online Library; 2007;30(5):609–28.

701 52. Cressie N. Statistics for spatial data. Wiley ; 2015.
702

703   53.   Venable W, Ripley BD. Modern applied statistics with S-PLUS. Springer, New York.
704   2002;

706   54.   Wood SN. Fast stable restricted maximum likelihood and marginal likelihood
707   estimation of semiparametric generalized linear models. Journal of the Royal Statistical
708   Society: Series B (Statistical Methodology). Wiley Online Library; 2011;73(1):3–36.

710   55.   Oksanen J, Blanchet FG, Kindt R, Legendre P, Minchin PR, O'Hara R, et al. vegan:
711   Community Ecology Package. R package version 2.2-1. 2015. 2015;

713   56.   Staniczenko P, Sivasubramaniam P, Suttle KB, Pearson RG. Linking macroecology and
714   community ecology: refining predictions of species distributions using biotic interaction
715   networks. Ecology Letters. Wiley Online Library; 2017;

717   57.   Marshall JK. Corynephorus canescens (L.) P. Beauv. as a model for the Ammophila
718   problem. Journal of Ecology. JSTOR; 1965;447–63.

720   58.   Castroviejo S, Aedo C, Cirujano S, Laínz M, Montserrat P, Morales R, et al. Flora
721   Ibérica 3. Real Jardín Botánico, CSIC, Madrid; 1993.

723   59.   Do Amaral Franco J. Flora Iberica 16. Real Jardin Botánico, CSIC; 2014.

725   60.   Metcalfe DJ. Hedera helix L. Journal of Ecology. Wiley Online Library;
726   2005;93(3):632–48.

728   61.   Holmes SE, Roy BA, Reed JP, Johnson BR. Context-dependent pattern and process: the
729   distribution and competitive dynamics of an invasive grass, Brachypodium sylvaticum.
730   Biological Invasions. 2010;12(7):2303–18.

732   62.   Ramakrishnan AP, Musial T, Cruzan MB. Shifting dispersal modes at an expanding
733   species' range margin. Molecular Ecology. Wiley Online Library; 2010;19(6):1134–46.

735   63.   Lenoir J, Graae BJ, Aarrestad PA, Alsos IG, Armbruster WS, Austrheim G, et al. Local
736   temperatures inferred from plant communities suggest strong spatial buffering of
737   climate warming across Northern Europe. Glob Chang Biol. 2013;19(5):1470–81.

739   64.   Cornwell WK, Schwilk DW, Ackerly DD. A trait-based test for habitat filtering: convex
740   hull volume. Ecology. Eco Soc America; 2006;87(6):1465–71.

742   65.   Diamond J. Assembly of species communities. In: Ecology and evolution of
743   communities. Harvard Univ. Press, pp. 342-444; 1975.

745   66.   Darwin C. The Origin of Species by Means of Natural Election, Or the Preservation of
746   Favored Races in the Struggle for Life. AL Burt.; 1859.

748   67.   Mayfield MM, Levine JM. Opposing effects of competitive exclusion on the
749   phylogenetic structure of communities. Ecology letters. Wiley Online Library;
750   2010;13(9):1085–93.
751

752   68.   Callaway RM. Positive interactions and interdependence in plant communities.
753   Springer; 2007.
754
755   69.   Castillo JP, Verdú M, Valiente-Banuet A. Neighborhood phylodiversity affects plant
756   performance. Ecology. Eco Soc America; 2010;91(12):3656–63.
757
758   70.   Valiente-Banuet A, Verdú M. Facilitation can increase the phylogenetic diversity of
759   plant communities. Ecol Lett. 2007;10(11):1029–36.
760
761   71.   Sargent RD, Ackerly DD. Plant-pollinator interactions and the assembly of plant
762   communities. Trends in Ecology \& Evolution. Elsevier; 2008;23(3):123–30.
763
764   72.   Bernard-Verdier M, Navas M-L, Vellend M, Violle C, Fayolle A, Garnier E.
765   Community assembly along a soil depth gradient: contrasting patterns of plant trait
766   convergence and divergence in a Mediterranean rangeland. Journal of Ecology. Wiley
767   Online Library; 2012;100(6):1422–33.
768
769   73.   Price JN, Gazol A, Tamme R, Hiiesalu I, Pärtel M. The functional assembly of
770   experimental grasslands in relation to fertility and resource heterogeneity. Functional
771   Ecology. Wiley Online Library; 2014;28(2):509–19.
772
773   74.   Haemig PD. Symbiotic nesting of birds with formidable animals: a review with
774   applications to biodiversity conservation. Biodiversity and Conservation. Springer;
775   2001;10(4):527–40.
776
777   75.   Quinn JL, Prop J, Kokorev Y, Black JM. Predator protection or similar habitat selection
778   in red-breasted goose nesting associations: extremes along a continuum. Animal
779   Behaviour. Elsevier; 2003;65(2):297–307.
780
781   76.   Stamps J, Krishnan V. Nonintuitive cue use in habitat selection. Ecology. Eco Soc
782   America; 2005;86(11):2860–7.
783
784

785

786 SUPPORTING INFORMATION

787 METHODS APPENDIX. Word file. Further detailed information about plot characterization,

788 environmental variables, climatic variables, geological information, network inference,

789 variables selection, and "Env+Bio" and "Env" comparison.

790 SUPPLEMENATRY MATERIAL.

791 S1 Table. Sp. Syndrome. Names of the species and code used for each of them, syndrome

792 assigned and reference supporting the assignment to that syndrome

793 S2 Table. Env+Bio and Env Models. Summary of the SDMs constructed used for each

794 species. Spearman correlation between their predictions and the observed abundance for each

795 species, considering the validated plots ("validate") and those used in the analysis ("test"),

796 the deviance and deviance explained for each model, and the difference between the

797 correlation with the observed data obtained using the "Env+Bio" and "Env" model for each

798 species.

799 S3 Table. Links. Summary for all the significant links inferred between species. Species

800 involved (from: parent node, to: children node), strength and direction of the association

801 based on the number of times that the link appears in the resampled networks, sign and

802 significance of the sign based on the Jonckheere trend test and the syndrome code for the

803 interspecific association.

804 S1 Fig. Sampling area and location of the plots used in the study.

805