

1 **Title:** Functional determinants of forest recruitment over broad scales.

2

3 Oscar Godoy^{1*}, Marta Rueda² and Bradford A. Hawkins²

4

5 ¹ Department of Ecology, Evolution & Marine Biology, University of California, Santa Barbara,
6 CA 93106, USA.

7

8 ² Department of Ecology and Evolutionary Biology, University of California, Irvine,
9 CA 92612, USA.

10

11 * Present address: Instituto de Recursos Naturales y Agrobiología de Sevilla (IRNAS), CSIC, PO
12 Box 1052, Sevilla E-41080, Spain, Phone +34 954 624 711, Fax +34 954 624 002

13

14 E-mails: ogodoy@irnas.csic.es, mrueda.eco@gmail.com, bhawkins@uci.edu

15

16 **Key words:** Conspecific negative density dependence, density independence, frost phenology,
17 functional traits, leaf N, North America, positive interactions, species abundances, wood density.

18 **Running title:** Forest recruitment and species traits

19

20 Words: 218 (abstract), 4947 (main body), 53 (references),

21 Tables: 2, Figures: 3, Appendices: 6

22

23 **Abstract**

24

25 **Aim** The drivers of tree recruitment over large spatial scales remain unexplored. Here, we ask
26 whether the species potential for recruitment and the strength of density dependent processes,
27 both inferred from species relative abundances, show emerging patterns that can be explained
28 upon the basis of climatic and functional trait information.

29 **Location** Eastern Forests of the USA.

30 **Methods** We document the geographical distributions and magnitudes of seedling recruitment
31 and the strength of density dependence and conspecific density dependence for the forests of the
32 eastern USA spanning >1.2 million km² across 88,854 local communities comprising 164 tree
33 species. We also compiled climatic variables and 16 traits representing several important
34 ecological axes of tree functional strategies to assess which factors were most strongly associated
35 with the emerging broad-scale spatial patterns.

36 **Results** Strong geographical variation in the potential of seedling recruitment and a latitudinal
37 change from negative to positive density dependence moving northward were associated with
38 adaptation to seasonal freezing temperatures and seed size. Wood density and leaf nitrogen, in
39 contrast, were related to the magnitude of the negative density dependence and conspecific
40 density dependence respectively, which were prevalent over most of the region.

41 **Main conclusions** Our results provide strong evidence that tree recruitment and the strength of
42 density dependent processes have broad-scale patterns explainable by a few key species
43 functional traits.

44 **Introduction**

45 “*The patterns must be understood as emerging from the collective behaviors of large ensembles*
46 *of smaller scale units*”(Levin, 1992).

47

48 Ecologists and foresters have studied tree recruitment for decades given its importance to
49 understanding forest dynamics (Clark *et al.*, 1998; Rees *et al.*, 2001; Clark *et al.*, 2010; Comita *et*
50 *al.*, 2010), the maintenance of forest diversity (Bolker & Pacala, 1999; Harms *et al.*, 2000;
51 Johnson *et al.*, 2012) and, more recently, species responses to climatic change (Peñuelas &
52 Boada, 2003; Zhu *et al.*, 2012). Much effort has focused on local scales, but studies of
53 biogeographical patterns emerging from local assemblages are rare (Johnson *et al.*, 2012; Coll *et*
54 *al.*, 2013; Carnicer *et al.*, 2014; Ruiz-Benito *et al.*, 2014). This severely limits our understanding
55 of the climatic and functional factors driving patterns of tree species recruitment at scales
56 relevant to global change.

57

58 Variation in the potential for recruitment among tree species can arise from endogenous factors
59 (i.e. density independent processes), such as genetic-based differences in fecundity, masting,
60 seedling tolerance to drought and cold, or age at maturity (Rejmánek & Richardson, 1996; Clark
61 *et al.*, 1998; Herrera *et al.*, 1998; Mueller *et al.*, 2005). This species potential for recruitment can
62 also be strongly affected by density dependent processes, either from interactions with neighbors
63 for limiting resources such as light and soil nutrients, or more specifically from conspecific
64 neighbors due to Janzen-Connell effects of pathogens and seed and seedling predators (Janzen,
65 1970; Connell, 1971; Canham *et al.*, 2006; Comita *et al.*, 2010; Rees, 2013). Importantly,
66 species with high relative recruitment often dominate local communities. This can occur from

67 producing copious offspring, which is insensitive to neighbors (Clark *et al.*, 1998), and from a
68 positive relationship between offspring survival and the number of conspecific neighbors
69 (McIntire & Fajardo, 2011). In contrast, a more diverse forest community is maintained when,
70 *inter alia*, species limit the establishment of their own seedlings more than the species with
71 which they interact (Chesson, 2000; Comita *et al.*, 2010; Johnson *et al.*, 2012).

72

73 Although both the species potential for recruitment and the strength of density dependent
74 processes strongly interact with abiotic and biotic drivers (e.g. light and moisture availability or
75 pathogen/herbivore populations) (Borchert *et al.*, 1989; Ribbens *et al.*, 1994), which may result
76 in high recruitment variability at local scales (Clark *et al.*, 2010), we hypothesize that they
77 should also have broad-scale patterns related to climatic variation. These emerging geographical
78 patterns may occur from the single effect of limiting conditions such as drought or freezing as
79 well as from the correlated effect that climate has on shaping the geographical structure of
80 important species traits influencing fecundity, germination and seedling mortality. For instance,
81 we hypothesize that the seed size-seed number trade-off (Smith & Fretwell, 1974) recently
82 demonstrated for trees (Adler *et al.*, 2014) will translate to a geographical pattern of higher
83 fecundity and thus higher recruitment at northern latitudes and higher altitudes where
84 communities are composed of tree species producing smaller seeds (Moles *et al.*, 2007; Hawkins
85 *et al.*, 2014).

86

87 We also hypothesize that an increase in the density of neighbors will reduce recruitment
88 (Lambers *et al.*, 2002; Comita *et al.*, 2010), but interactive effects may occur with climatic
89 conditions shifting from negative to positive with increasing environmental stress (Callaway *et*

90 *al.*, 2002; McIntire & Fajardo, 2011). For example, prior work suggests that wood density may
91 determine the geographical pattern of the strength of density dependence (Enquist *et al.*, 1999;
92 Swenson & Enquist, 2007; Chave *et al.*, 2009; Adler *et al.*, 2014). Species with dense wood,
93 which in North American forests are distributed in southern latitudes and lower elevations, tend
94 to be larger, have longer life spans and survive better in the face of drought or insect/pathogen
95 attack, which may produce stronger negative effects of neighbors for longer. Other functional
96 traits related to resource acquisition such as height (for light) or rooting depth (for water) can
97 also determine the strength of density dependence when resources are limiting (Sterck *et al.*,
98 2011).

99

100 Recent work has shown that the strength of conspecific density dependence in North American
101 forest decreases with latitude (Johnson *et al.*, 2012), perhaps because drivers limiting
102 recruitment, such as host-specific herbivores and pathogens, are themselves limited by climate
103 (Janzen, 1970; Connell, 1971; Harms *et al.*, 2000), although the only empirical test of this
104 hypothesis found no support (Lambers *et al.*, 2002). This suggests that the geographical pattern
105 of conspecific density dependence should be similar to overall density dependence. Less clear is
106 identifying the functional determinants of the strength of conspecific density dependence, as it
107 tends to vary with shade tolerance (Kobe & Vriesendorp, 2011). Shade tolerance is a complex
108 attribute that can be mechanistically driven by an array of functional traits including height, wood
109 density, seed size, leaf area, and leaf N (Kitajima, 1994; Niinemets, 1997; Hewitt, 1998). Such
110 complex interactions among suites of traits make it difficult to predict the distributions of
111 phenotypes across space.

112

113
114
115
116
117
118
119
120
121
122
123
124
125
126
127
128
129

In this paper we estimate the species potential for recruitment arising from density independent processes (hereafter ‘recruitment potential’, RP), and the strength of density dependence with all neighbors (hereafter ‘density dependence’, DD) and with conspecific neighbors only (hereafter ‘conspecific density dependence’, CDD) to document their community geographical patterns for the forests of the eastern United States. We considered in total 164 tree species occurring across 88,854 sites (approx. 1.27 million km²). These demographic variables were then analyzed using a machine learning method for regression (Random Forest) with respect to four climatic variables accounting for variation in temperature and precipitation, and with 16 functional traits (Table 1), representing several important ecological axes of woody plant strategies in terms of leaf economics, growth allocation, resource acquisition and regeneration (Westoby *et al.*, 2002; Westoby & Wright, 2006; Kraft *et al.*, 2008). We focus on three questions: (1) Do local communities have geographical structure of potential for recruitment and strength of density dependent processes based on species distribution? (2), If so, can these patterns be explained by climate and species traits? (3) What is the relative importance of climatic versus functional factors?

130 **Methods**

131 *Extracting information from the FIA database*

132 Our study centered on the forests of the USA delimited by the Northern and Southern Region
133 units of the US Forest Service’s Forest Inventory and Analysis (FIA). These forests extend from
134 Florida to Maine and from the Atlantic Coast to the Kansas-Colorado line. We downloaded from
135 the FIA (<http://www.fia.fs.fed.us>; accessed in February, 2012) data collected between 2005-

136 2010, corresponding to the most recently updated five-year cyclical inventory. The FIA protocol
137 is designed to record, per site and per species, the number of trees larger than 12.7 cm diameter
138 at breast height (dbh) in four subplots (168.3 m²) 36.6 m apart, and the number of seedlings in
139 contiguous nested microplots (13.5 m²). FIA defines as seedlings those individuals smaller than
140 2.54 cm dbh and taller than 30.5 cm for angiosperms and 12.2 cm for gymnosperms. We used
141 Access queries per state to select only those sites corresponding to natural stands (i.e. excluding
142 tree plantations and orchards), without human disturbance, and without fire occurrence in the
143 five years prior to being surveyed. We then used the FIPS code (the species code used by FIA)
144 to remove those entries corresponding to broad tree classifications (e.g. evergreen, deciduous)
145 and genera, and to merge species under currently recognized names. We used The Plant List
146 (<http://www.theplantlist.org>; accessed in March 2012) as the source of information on synonyms.
147 These selection criteria rendered a total of 88854 sites containing 164 species.

148

149 *Estimation of recruitment potential and density dependent effects*

150 We followed an analytical approach quasi-similar to Johnson *et al.* (2012). First, we used each
151 of the four subplot-microplot pairs per site as our unit of analysis (replicates). Second, according
152 to the scatterplots of the number of seedlings per focal species (S_i) as a function of the total
153 number of adult neighbors (T), we fitted an exponential function using maximum likelihood
154 methods (R version 2.13.2, function “optim,” method “L-BFGS-B”) with the specific form:

$$S_i = a_i e^{b_i T}$$

155

(1)

156 Here, a_i is the potential number of seedlings recruited per species in the absence of interaction
157 with adult neighbors (RP) (i.e. recruitment from density independent effects). b_i is the inflection

158 curve parameter, which indicates the direction and the magnitude of the effect of adult neighbors
159 density on RP. Negative values of b_i indicate that the number of seedlings decreases
160 exponentially as the number of adults increases (i.e. negative, or direct, density dependence)
161 whereas positive values indicate the reverse (i.e. positive, or inverse, density dependence)
162 (Appendix S1 in Supporting Information). For the fitting process, however, we selected a
163 negative binomial error structure giving the greatest heterogeneity observed in the increase of the
164 variance of the data with respect to the mean across species (overdispersion parameter k ranging
165 from 0.318 to 7835, median = 1.399, smaller values of k indicate greater heterogeneity) (Bolker,
166 2008). We followed this procedure to assess the average species response to the overall density-
167 dependent effect of conspecific and heterospecific neighbors (DD) as well as to the separate
168 effect of conspecific neighbors (CDD). For CDD, we selected sites only containing seedlings
169 and adults of a single focal species without the presence of other species. The main advantage of
170 this function is to obtain a joint fit for a_i (bounded to be >1), and b_i (set to be free). This also
171 makes values of the inflection curve (b_i) sensitive to those fitted to the intercept (a_i). Therefore,
172 we confirmed that differences in the strength of average DD versus average CDD were not due
173 to differences in average RP (the correlation between y-intercepts of DD and CDD fits was $r =$
174 0.809 , $p < 0.001$). For both DD and CDD, we set the threshold of the number of replicates per
175 species to ten subplot-microplot pairs to perform the fitting, and finally, we did not divide the
176 data into regional subunits because i) we needed to match the resolution of trait information,
177 which is available as a single value per species (see methods section “*climatic and species trait*
178 *information*”), and ii) we needed a single value per species for phylogenetic comparative
179 analyses to test whether obtained parameters can be considered a species characteristic in order
180 to be mapped (see next methods section).

181
182 *Phylogenetic analyses testing whether recruitment and density dependent effects are species*
183 *specific.*
184 We asked whether RP, DD, and CDD values can be considered a species characteristic or they
185 are evolutionarily labile. This was a crucial step to identify the type of information necessary to
186 explore emerging geographical patterns of recruitment potential and density dependent effects.
187 The rationale is that if they were labile or weakly conserved, they would likely be a product of
188 the local environment; that is, two closely related species could have either similar or very
189 different values. Thus, environmental data would be needed to map geographical patterns of RP,
190 DD, and CDD. In contrast, if they were conserved, we can assume that they are a species
191 characteristic and closely related species will tend to have similar values while values between
192 disparate lineages will tend on average to diverge. In such case, we would consider species
193 distributions instead of environment a better approach to explore geographical patterns.
194
195 We performed two types of phylogenetic comparative analyses: phylogenetic signal
196 representation curves (PSR) (Diniz-Filho *et al.*, 2012), complemented with the fitting by
197 phylogenetic Generalized Least Squares (pGLS) of three models of evolution given our species
198 phylogenetic relatedness (Appendix S2). Both approaches indicated that species recruitment and
199 density dependent parameters displayed large deviations from a Brownian model of evolution
200 and are consistent with an interpretation of trait evolution under an Ornstein-Uhlenbeck (OU)
201 process (Appendix S2). In addition, Blomberg's K values across all species indicated also a
202 significant phylogenetic signal according to an OU process ($K_{RP} = 0.249$, $p = 0.006$; $K_{DD} = 0.087$,
203 $p = 0.038$; $K_{CDD} = 0.147$, $p = 0.051$) (lower Blomberg's K values under a OU model indicate

204 stronger phylogenetic signal). This phylogenetic signal was partially driven by differences
205 between angiosperm and gymnosperm species. The RP of gymnosperms was significantly
206 higher ($F_{1,162} = 4.23$, $p = 0.048$) and their DD marginally weaker ($F_{1,162} = 2.67$, $p = 0.108$)
207 although both taxonomic groups showed similar CDD ($F_{1,162} = 1.31$, $p = 0.247$). In summary, all
208 methods of analysis indicated that species average RP, DD, and CDD are phylogenetically
209 conserved, so they were considered a species characteristic and they were subsequently mapped
210 using species distributions to test whether community geographical patterns can be predicted by
211 climatic variables and species functional traits. Note that in any case we aim to test the relative
212 contribution of phylogeny to the geographical patterns of RP, DD, and CDD and to their
213 functional predictors.

214

215 *Climatic and species trait information*

216 We selected 20 variables to examine how climate and species functional traits are associated
217 with RP, DD and CDD. We selected four climatic variables representing the two most important
218 climatic gradients occurring at meso- and macro-scales in the eastern USA (temperature and
219 precipitation). Two measures of temperature (BIO5, maximum temperature of the warmest
220 month; BIO6, minimum temperature of the coldest month) and two measures of precipitation
221 (BIO12, annual precipitation; BIO18, precipitation in the warmest month) were extracted from the
222 30 arc-second WorldClim database (<http://www.worldclim.org>). Further, sixteen variables
223 representing a range of important axes of ecological strategies of tree species (Westoby *et al.*,
224 2002; Westoby & Wright, 2006; Moles *et al.*, 2007; Kraft *et al.*, 2008) were obtained from
225 various sources. This trait information is available at the species level (Table 1, Appendix S3).
226 These traits show geographical variation at local scales as well as large scales, and none of them

227 were strongly correlated ($r < 0.8$). We did not include specific leaf area (SLA), a widely used
228 plant trait, in the analysis due to limited information (values for fewer than 50% of the species
229 were available).

230

231 *Forest community data and geographical analyses with Random Forest models.*

232 To generate community data at the site level from species-level data we first created a presence-
233 absence matrix for the 164 species across FIA sites. Then, we calculated averages of the three
234 dependent variables (RP, DD, CDD) and the sixteen functional trait predictors at sites including
235 at least two species.

236

237 We then generated Random Forest models, each based on 200 regression trees with 1000
238 permutations per regression tree, in the R package RandomForest (Liaw & Wiener, 2002), first,
239 to statistically account for the spatial patterns of average community data across sites, second, to
240 statistically account for the variation across species without the spatial component. Comparing
241 both analyses served to test whether the drivers of recruitment match between community and
242 species level. Briefly, Random Forest modeling is a powerful machine-learning technique that
243 combines the predictions of multiple independent regression trees into a robust composite model.
244 The relative importance of the predictors is assessed by the decrease in explained variance
245 resulting from permutations of the focal variable. We selected Random Forest modeling over
246 more traditional linear modeling approaches because do not assume stationary of relationships.
247 This statistical technique is able to disentangle interacting effects and can identify nonlinear and
248 scale-dependent relationships among multiple, correlated predictors (Cutler *et al.*, 2007) that
249 often occur at the scale of this analysis.

250 For the community level analysis we generated nine Random Forest models, proceeding as
251 follows: for each dependent variable (RP, DD and CDD), we generated three models considering
252 the average site values of the predictors based on the subplot species composition; one including
253 only climatic variables, one including only species traits, and one including all predictors
254 (Appendix S4). This allows us to evaluate the extent to which the climate vs. trait predictors
255 with higher importance values have independent relationships with geographical structure of the
256 response variables or if their relative importance is largely due to correlations with the other
257 class of predictors. For each model, we recorded the percentage of the explained variance
258 (pseudo- R^2 ranging from 0 to 100%), and we ranked the relative importance of each variable
259 ranging from 100 (the strongest predictor) to 0 (having no predictive power) according to their
260 node purity values. The sign of the relationship between the dependent variable and the
261 predictors were further assessed with Pearson correlation. We found that models including all
262 predictors explained large amounts of variance, which might be due to including twenty
263 predictors. We therefore regenerated the Random Forest models including only those predictors
264 with high relative importance. To take a conservative approach the threshold was set to
265 importance value > 75 . These simplified models explained 12-20% less variance than the full
266 models (Appendix S4), but in all cases remained powerful in terms of explained variance
267 considering the small size of the FIA plots, the large sample size, and the geographical scope of
268 the data.

269

270 For the species level analysis, we generated three Random Forest models, one for each
271 dependent variable, considering trait values for each species as the predictors rather than site
272 averages. Here, we also recorded the percentage of explained variance of the model, and we

273 estimated the relative importance of each predictor and the sign of the relationship between the
274 dependent variable and the predictors as indicated above.

275

276 In the final step of the analysis we generated spatial autocorrelograms of Moran's I (the most
277 widely used autocorrelation metric) to evaluate the strengths of the geographical gradients of RP,
278 DD and CDD, followed by correlograms of the residuals extracted from the Random Forest
279 model combining all predictors. Moran's I values, ranging from ca. -1 to 1, indicate whether
280 communities connected at a given distance are more similar (positive correlation) or less similar
281 (negative correlation) than expected for randomly associated pairs of plots. This allows us to
282 evaluate the extent to which the climatic variables and traits are able to account for the spatial
283 patterns across the full range of scales, from local to subcontinental (Hawkins et al. 2014).

284

285 **Results**

286 *Recruitment potential (density independent effects)*

287 Average RP of local communities has a strong geographical pattern very similar to the
288 distribution of the minimum number of frost-free days (Min F-F days, Table 1) required by each
289 species to complete its seasonal phenological development and to seed size to a lesser extent
290 (Figs. 1a, 2a, 3a, b). Local communities comprising species able to live under longer seasonal
291 freezing conditions and with smaller seeds had the potential to recruit three times more
292 individuals (Table 2). Geographically, the potential of recruitment of tree communities in the
293 forests of the eastern USA has a latitudinal gradient with higher recruitment at northern latitudes,
294 an elevational gradient with higher recruitment in Appalachian forests compared to surrounding

295 lowland sites (Fig. 1a), and a visible longitudinal gradient with lowland forests nearer the east
296 coast having higher recruitment than forests at the forest—prairie interface.

297

298 *Density dependence*

299 Most species suffer reduced recruitment in the presence of neighbors (i.e. a negative DD)
300 (Appendix S5); thus negative density dependence in tree communities is expected to occur
301 across most, but not all, of the region (Fig. 1b inset). The geographical structure of DD was
302 relatively weaker than for RP, with which it was also weakly correlated ($r = 0.17$) (see Fig. 1b
303 for map and spatial correlogram; DD showed a flatter curve of the Moran's I values of the raw
304 data compared to RP). The spatial structure of DD at intermediate scales resulted, for instance,
305 in stronger density dependent effects at forest—prairie interface sites near the Texas-Louisiana-
306 Arkansas borders associated with oak-pine forest (Fig. 1b). Nevertheless, we found two clear
307 trait predictors: communities comprising species with higher wood densities and trunk diameters
308 tend to interact more negatively (Table 2 and Figs. 2b, 3d). In addition, tree communities also
309 show a sharp shift from negative to positive DD associated with longer, more severe winters
310 (Figs. 1b, 3e).

311

312 *Conspecific density dependence*

313 Average CDD shows similarities with DD. First, their magnitudes were comparable. Second,
314 negative CDD is also prevalent. Third, the geographical pattern of CDD is uncorrelated with RP
315 ($r = -0.02$), at least partially because species with stronger negative CDD interactions occur in
316 numerous scattered locations at smaller geographical scales, such as the lower basin of the
317 Mississippi River, parts of Florida, lowland forest along the east coast, Appalachian forest, and

318 Great Lakes forest. Fourth, there is again a shift from negative to more positive conspecific
319 interactions moving northward and westward related to Min F-F days (Figs. 1c, 3a), although the
320 spatial pattern is weak, albeit still significant ($P < 0.001$), based on the correlogram. A major
321 difference is that leaf N instead of wood density was the primary trait predictor of CDD (Table
322 2), with stands with lower leaf N showing the potential to interact more negatively between
323 conspecifics (positive correlation, Fig. 2c). Another difference is that CDD was predicted by a
324 larger number of traits than DD, including seed mass, maximum trunk diameter, wood density,
325 and anaerobic tolerance (Table 2 and Fig. 2c).

326

327 *Species traits and climate: independent and correlated predictors*

328 It is worth noting that the combined Random Forest models including both climate and species
329 traits as predictors explained large amounts of the observed geographical patterns of RP, DD,
330 and CDD (Fig. 2) and captured virtually all spatial pattern; the Moran's I values of the model
331 residuals for each variable retained no detectable spatial autocorrelation at any scale (Fig. 1). In
332 addition, Random Forest models at the community level explained more variation of RP, DD and
333 CDD than models at the species level, perhaps because of the significant differences in sample
334 size (cf. Table 2 and Fig. 2). Surprisingly, climatic variables *per se* are not strong predictors; the
335 pseudo- R^2 did not increase appreciably between the Random Forest community model including
336 only the species traits and the model including both climate and traits (Fig. 2; Appendix S4).
337 However, the temperature variables, maximum temperature during summer (BIO5) (for RP) and
338 minimum temperature during winter (BIO6) (for CDD), were related to spatial patterns via their
339 relationships with specific species traits (BIO5 and BIO6 are correlated with Min F-F days ($r =$
340 0.56 and $r = 0.62$ respectively, Fig. 3c, e). Moreover, the geographical patterns for density

341 independent (RP) and density dependent (DD and CDD) processes were not identical, the former
342 being much more strongly structured than the latter. This suggests that the geographical
343 structure of the magnitude and sign of community density dependent processes do not depend on
344 the recruitment potential although they are modestly correlated at the species level ($r_{RP-DD} = -$
345 0.40 , $r_{RP-CDD} = -0.22$).

346

347 **Discussion**

348 Our central finding is that the recruitment from density independent and dependent processes of
349 trees species of eastern USA forest communities exhibit broad-scales patterns. We also found
350 clear functional predictors. Among them, variation in the magnitude of recruitment potential has
351 the clearest geographical structure, explained by the ability to tolerate seasonal freezing
352 conditions (Min F-F days), seed mass, and maximum temperature during summer (Table 2, and
353 Figs. 1a, 2). Several explanations for why forest communities at northern latitudes, at higher
354 elevations, and along the east coast can recruit up to three times more individuals are plausible
355 and not mutually exclusive. First, species have evolved smaller seeds in their adaptation to cold
356 climates in order to complete rapid seed development to avoid freeze-induced mortality in the
357 fall (Moles *et al.*, 2007; Hawkins *et al.*, 2014). Because species with smaller seeds tend to
358 produce more offspring per individual than species with larger seeds, we expect that being more
359 fecund may generate higher community recruitment since seed predation rate is generally not
360 related to seed mass (Moles *et al.*, 2003). Second, the activity of several drivers of seed/seedling
361 mortality such as pathogens, insects and mammals increases with temperature, mainly under
362 moist conditions (Bale *et al.*, 2002; Harvell *et al.*, 2002). Thus, being adapted to tolerate

363 freezing conditions for longer when active growth is not possible (i.e. low Min F-F days) reduces
364 this biotic risk of mortality. Finally, a longer growing season with a hot summer increases
365 seasonal drought and seedling mortality rates by water stress (Ruiz-Benito *et al.*, 2013).

366

367 We also found that density-dependent processes, either in the interactions with the total number
368 of adults (DD) or with conspecific adults only (CDD), reduced recruitment at almost all spatial
369 scales (Figs. 1b, c). Two distinctive functional traits arise when explaining the magnitude of
370 negative interactions: wood density for DD and leaf nitrogen for CDD (Table 2 and Figs. 2b, c).

371 A simple explanation for DD partially supported by our analyses may be that tree species with
372 higher wood density tend to be larger (Table 2, Max Trunk diameter) (Baker *et al.*, 2004), and
373 bigger individuals produce stronger negative competitive effects on both conspecific and
374 heterospecific neighbors (Rees, 2013). Why communities with lower leaf N tend to interact
375 more negatively only between conspecifics can be related to shade tolerance (Kobe &
376 Vriesendorp, 2011), although our categorical measure of shade tolerance was unable to account
377 for the geographical pattern of CDD. Shade-tolerant species show also enhanced survival
378 against attack by host-specific enemies because they invest in conservative functional strategies
379 such as low leaf N, large seeds and dense wood (traits with high relative importance in our
380 analyses, Table 2) (Kitajima, 1994; Coley & Barone, 1996; Hewitt, 1998). Also, lower leaf N
381 can reflect species living on nutrient-poor soils (Fig. 3f, lower Leaf N values correspond to
382 communities living in ultisols, which are acid forest soils with relatively low fertility).

383 Competition for nutrients in these environments is probably high compared to more fertile soils,
384 especially with conspecifics as they exploit similarly soil resources.

385

386 Tree communities shift from negative to positive DD and CDD from south to north, which
387 relates to the latitudinal pattern of Min F-F days (Figs. 1b, c, 3a). This change in the sign of
388 species interactions (at approx. 100 Min F-F days) with increasing environmental stress (colder
389 conditions at northern latitudes) is consistent with studies at the global scale supporting the
390 stress-gradient hypothesis (Callaway *et al.*, 2002). Seedlings are the stage most sensitive to cold
391 injury because the layer of cold air close to the ground can freeze their meristems; an effect that
392 is worsened in forest openings by radiation frost (Howe *et al.*, 2003). Thus, the presence of adult
393 trees likely favors seedling survival at northern latitudes by increasing soil surface temperature.
394 Nevertheless, the fact that the change in the sign of DD and CDD is abrupt and the other main
395 predictors are functional traits raises doubts about a simple climate-driven explanation. It may
396 be possible that negative to positive interactions depend partially on evolutionary histories of
397 plants. This is a plausible explanation given that we found DD marginally weaker for
398 gymnosperm species, and changes in composition from angiosperm-dominated forests (Oak-
399 Hickory, Aspen-Birch) to gymnosperm-dominated forests (Spruce-Fir) (Appendix S6) appear to
400 match the line separating negative from positive interactions (most evident again for DD).

401

402 One surprising result from both the trait and the combined (climate + trait) Random Forest
403 community models is that Min F-F days is consistently the predictor with the highest relative
404 importance (Fig. 2, Appendix S4). It is commonly thought that the ability to respond to
405 seasonally freezing temperatures defines range limits. At northern latitudes freezing
406 temperatures limit fruit maturation, and at southern latitudes budburst may not occur due to the
407 lack of chilling temperatures. Indeed, the physiological relationship of freezing with species
408 occurrences has been extensively explored using process-based species distribution models to

409 project shifts in the species ranges due to climate change (reviewed in Chuine, 2010). Our results
410 suggest an aspect related to demography not widely considered. Specifically, within species
411 ranges the degree of adaptation to the seasonal breadth of freezing (and thus the length of the
412 growing season) influences the magnitude of potential recruitment and the strength and sign of
413 density dependence of local communities. We believe that the combination of previous work
414 with a consideration of the role of phenology can stimulate research to predict not only future
415 forest composition with environmental change (Cleland *et al.*, 2007) but also the relative
416 abundances of tree species.

417

418 In sum, by combining demographic, functional, and geographical aspects of local forest
419 communities, we have found strong broad-scale patterns in three key components of forest
420 recruitment. We have also found that a phenological trait related to the tolerance to seasonal
421 freezing conditions combined with three functional traits (seed mass, wood density, and leaf N)
422 account most of the observed spatial structure. Importantly, we have quantified the magnitude
423 by which these traits explain reduction in recruitment or the shift from negative to positive
424 interactions. Nevertheless, we estimated an average value of RP, DD, and CDD per species, and
425 the observed geographical variation was based on differences in species occurrences among local
426 communities. Future research needs to estimate these three parameters within species at smaller
427 spatial scales to address the question of which drivers are operating locally that can change the
428 species potential for recruitment and the strength of density dependence, and hence, local
429 community patterns. This may account for the 12, 25, and 40% of variation in RP, DD, and
430 CDD unexplained by our models (there is no residual spatial autocorrelation after fitting the
431 predictors (Fig. 1), indicating that no additional spatially structured variables are needed to

432 explain the patterns statistically). Coupled with this, we also need more data on intraspecific
433 trait variation, at least across sites (information logistically difficult to obtain, but it could
434 initially be focused on seed mass, wood density and leaf N based on the results found here).
435 Finally, the subplots sampled by the FIA are small (0.017 h), which means that much of the
436 unexplained variance could be due to sampling error with respect to the trees actually present in
437 local communities, a potential problem that is difficult to access at the subcontinental scale.
438 These methodological issues aside, our results show that forest recruitment exhibit broad-scale
439 patterns, and they can be explained by a few key functional traits.

440

441 **Acknowledgements**

442 We thank Lorena Gómez-Aparicio, Irena Šímová, Pedro Villar, Jeff Diez and Elena Granda for
443 their comments. O.G and M.R acknowledge postdoctoral financial support provided by the
444 Spanish Ministry of Education and Science and Fulbright commission (FU-2009-0039 and BVA-
445 2010-0596, respectively).

446

447 **References**

- 448 Adler, P.B., Salguero-Gómez, R., Compagnoni, A., Hsu, J.S., Ray-Mukherjee, J., Mbeau-Ache,
449 C. & Franco, M. (2014) Functional traits explain variation in plant life history strategies.
450 *Proceedings of the National Academy of Sciences, USA*, **111**, 740-745.
- 451 Baker, T.R., Phillips, O.L., Malhi, Y., Almeida, S., Arroyo, L., Di Fiore, A., Erwin, T., Killeen,
452 T.J., Laurance, S.G. & Laurance, W.F. (2004) Variation in wood density determines
453 spatial patterns in Amazonian forest biomass. *Global Change Biology*, **10**, 545-562.

454 Bale, J.S., Masters, G.J., Hodkinson, I.D., Awmack, C., Bezemer, T.M., Brown, V.K.,
455 Butterfield, J., Buse, A., Coulson, J.C. & Farrar, J. (2002) Herbivory in global climate
456 change research: direct effects of rising temperature on insect herbivores. *Global Change*
457 *Biology*, **8**, 1-16.

458 Bolker, B.M. (2008) *Ecological models and data in R*. Princeton University Press, Princeton, NJ.

459 Bolker, B.M. & Pacala, S.W. (1999) Spatial moment equations for plant competition:
460 understanding spatial strategies and the advantages of short dispersal. *The American*
461 *Naturalist*, **153**, 575-602.

462 Borchert, M.I., Davis, F.W., Michaelsen, J. & Oyler, L.D. (1989) Interactions of factors affecting
463 seedling recruitment of blue oak (*Quercus douglasii*) in California. *Ecology*, **70**, 389-404.

464 Callaway, R.M., Brooker, R.W., Choler, P., Kikvidze, Z., Lortie, C.J., Michalet, R., Paolini, L.,
465 Pugnaire, F.I., Newingham, B., Aschehoug, E.T., Armas, C., Kikodze, D. & Cook, B.J.
466 (2002) Positive interactions among alpine plants increase with stress. *Nature*, **417**, 844-
467 848.

468 Carnicer, J., Coll, M., Pons, X., Ninyerola, M., Vayreda, J. & Peñuelas, J. (2014) Large-scale
469 recruitment limitation in Mediterranean pines: the role of *Quercus ilex* and forest
470 successional advance as key regional drivers. *Global Ecology and Biogeography*, **23**,
471 371-384.

472 Chave, J., Coomes, D., Jansen, S., Lewis, S.L., Swenson, N.G. & Zanne, A.E. (2009) Towards a
473 worldwide wood economics spectrum. *Ecology Letters*, **12**, 351-366.

474 Chesson, P. (2000) Mechanisms of maintenance of species diversity. *Annual review of Ecology*
475 *and Systematics*, **31**, 343-366.

- 476 Chuine, I. (2010) Why does phenology drive species distribution? *Philosophical Transactions of*
477 *the Royal Society B: Biological Sciences*, **365**, 3149-3160.
- 478 Clark, J.S., Macklin, E. & Wood, L. (1998) Stages and spatial scales of recruitment limitation in
479 southern Appalachian forests. *Ecological Monographs*, **68**, 213-235.
- 480 Clark, J.S., Bell, D., Chu, C., Courbaud, B., Dietze, M., Hersh, M., HilleRisLambers, J., Ibáñez,
481 I., LaDeau, S. & McMahon, S. (2010) High-dimensional coexistence based on individual
482 variation: a synthesis of evidence. *Ecological Monographs*, **80**, 569-608.
- 483 Cleland, E.E., Chuine, I., Menzel, A., Mooney, H.A. & Schwartz, M.D. (2007) Shifting plant
484 phenology in response to global change. *Trends in Ecology & Evolution*, **22**, 357-365.
- 485 Coley, P.D. & Barone, J. (1996) Herbivory and plant defenses in tropical forests. *Annual Review*
486 *of Ecology and Systematics*, **27**, 305-335.
- 487 Coll, M., Peñuelas, J., Ninyerola, M., Pons, X. & Carnicer, J. (2013) Multivariate effect
488 gradients driving forest demographic responses in the Iberian Peninsula. *Forest Ecology*
489 *and Management*, **303**, 195-209.
- 490 Comita, L.S., Muller-Landau, H.C., Aguilar, S. & Hubbell, S.P. (2010) Asymmetric density
491 dependence shapes species abundances in a tropical tree community. *Science*, **329**, 330-
492 332.
- 493 Connell, J.H. (1971) On the role of natural enemies in preventing competitive exclusion in some
494 marine animals and in rain forest trees. *Dynamics of populations* (ed. by J. Den Boer and
495 G. Gradwell), pp. 298-312. PUDOC, Wageningen.
- 496 Cutler, D.R., Edwards Jr, T.C., Beard, K.H., Cutler, A., Hess, K.T., Gibson, J. & Lawler, J.J.
497 (2007) Random forests for classification in ecology. *Ecology*, **88**, 2783-2792.

498 Diniz-Filho, F., Alexandre, J., Rangel, T.F., Santos, T. & Mauricio Bini, L. (2012) Exploring
499 patterns of interspecific variation in quantitative traits using sequential phylogenetic
500 eigenvector regressions. *Evolution*, **66**, 1079-1090.

501 Enquist, B.J., West, G.B., Charnov, E.L. & Brown, J.H. (1999) Allometric scaling of production
502 and life-history variation in vascular plants. *Nature*, **401**, 907-911.

503 Harms, K.E., Wright, S.J., Calderon, O., Hernandez, A. & Herre, E.A. (2000) Pervasive density-
504 dependent recruitment enhances seedling diversity in a tropical forest. *Nature*, **404**, 493-
505 495.

506 Harvell, C.D., Mitchell, C.E., Ward, J.R., Altizer, S., Dobson, A.P., Ostfeld, R.S. & Samuel,
507 M.D. (2002) Climate warming and disease risks for terrestrial and marine biota. *Science*,
508 **296**, 2158-2162.

509 Hawkins, B.A., Rueda, M., Rangel, T.F., Field, R. & Diniz-Filho, J.A.F. (2014) Community
510 phylogenetics at the biogeographical scale: cold tolerance, niche conservatism and the
511 structure of North American forests. *Journal of Biogeography*, **41**, 23-38.

512 Herrera, C.M., Jordano, P., Guitián, J. & Traveset, A. (1998) Annual variability in seed
513 production by woody plants and the masting concept: reassessment of principles and
514 relationship to pollination and seed dispersal. *The American Naturalist*, **152**, 576-594.

515 Hewitt, N. (1998) Seed size and shade-tolerance: a comparative analysis of North American
516 temperate trees. *Oecologia*, **114**, 432-440.

517 Howe, G.T., Aitken, S.N., Neale, D.B., Jermstad, K.D., Wheeler, N.C. & Chen, T.H. (2003)
518 From genotype to phenotype: unraveling the complexities of cold adaptation in forest
519 trees. *Canadian Journal of Botany*, **81**, 1247-1266.

520 Janzen, D.H. (1970) Herbivores and the number of tree species in tropical forests. *The American*
521 *Naturalist*, **104**, 501-528.

522 Johnson, D.J., Beaulieu, W.T., Bever, J.D. & Clay, K. (2012) Conspecific negative density
523 dependence and forest diversity. *Science*, **336**, 904-907.

524 Kitajima, K. (1994) Relative importance of photosynthetic traits and allocation patterns as
525 correlates of seedling shade tolerance of 13 tropical trees. *Oecologia*, **98**, 419-428.

526 Kobe, R.K. & Vriesendorp, C.F. (2011) Conspecific density dependence in seedlings varies with
527 species shade tolerance in a wet tropical forest. *Ecology letters*, **14**, 503-510.

528 Kraft, N.J., Valencia, R. & Ackerly, D.D. (2008) Functional traits and niche-based tree
529 community assembly in an Amazonian forest. *Science*, **322**, 580-582.

530 Lambers, J.H.R., Clark, J.S. & Beckage, B. (2002) Density-dependent mortality and the
531 latitudinal gradient in species diversity. *Nature*, **417**, 732-735.

532 Levin, S.A. (1992) The problem of pattern and scale in ecology: the Robert H. MacArthur award
533 lecture. *Ecology*, **73**, 1943-1967.

534 Liaw, A. & Wiener, M. (2002) Classification and regression by randomforest. *R news*, **2**, 18-22.

535 McIntire, E.J. & Fajardo, A. (2011) Facilitation within species: a possible origin of group-
536 selected superorganisms. *The American Naturalist*, **178**, 88-97.

537 Moles, A.T., Warton, D.I. & Westoby, M. (2003) Do small-seeded species have higher survival
538 through seed predation than large-seeded species? *Ecology*, **84**, 3148-3161.

539 Moles, A.T., Ackerly, D.D., Tweddle, J.C., Dickie, J.B., Smith, R., Leishman, M.R., Mayfield,
540 M.M., Pitman, A., Wood, J.T. & Westoby, M. (2007) Global patterns in seed size. *Global*
541 *Ecology and Biogeography*, **16**, 109-116.

542 Mueller, R.C., Scudder, C.M., Porter, M.E., Talbot Trotter, R., Gehring, C.A. & Whitham, T.G.
543 (2005) Differential tree mortality in response to severe drought: evidence for long-term
544 vegetation shifts. *Journal of Ecology*, **93**, 1085-1093.

545 Niinemets, Ü. (1997) Role of foliar nitrogen in light harvesting and shade tolerance of four
546 temperate deciduous woody species. *Functional Ecology*, **11**, 518-531.

547 Peñuelas, J. & Boada, M. (2003) A global change-induced biome shift in the Montseny
548 mountains (NE Spain). *Global Change Biology*, **9**, 131-140.

549 Rees, M. (2013) Competition on productivity gradients—what do we expect? *Ecology letters*, **16**,
550 291-298.

551 Rees, M., Condit, R., Crawley, M., Pacala, S. & Tilman, D. (2001) Long-term studies of
552 vegetation dynamics. *Science*, **293**, 650-655.

553 Rejmánek, M. & Richardson, D.M. (1996) What attributes make some plant species more
554 invasive? *Ecology*, **77**, 1655-1661.

555 Ribbens, E., Silander Jr, J.A. & Pacala, S.W. (1994) Seedling recruitment in forests: calibrating
556 models to predict patterns of tree seedling dispersion. *Ecology*, **75**, 1794-1806.

557 Ruiz-Benito, P., Lines, E.R., Gómez-Aparicio, L., Zavala, M.A. & Coomes, D.A. (2013)
558 Patterns and drivers of tree mortality in Iberian forests: climatic effects are modified by
559 competition. *PloS One*, **8**, e56843.

560 Ruiz-Benito, P., Gómez-Aparicio, L., Paquette, A., Messier, C., Kattge, J. & Zavala, M.A.
561 (2014) Diversity increases carbon storage and tree productivity in Spanish forests. *Global*
562 *Ecology and Biogeography*, **23**, 311-322.

563 Smith, C.C. & Fretwell, S.D. (1974) The optimal balance between size and number of offspring.
564 *The American Naturalist*, vol?, 499-506.

- 565 Sterck, F., Markesteijn, L., Schieving, F. & Poorter, L. (2011) Functional traits determine trade-
566 offs and niches in a tropical forest community. *Proceedings of the National Academy of*
567 *Sciences, USA*, **108**, 20627-20632.
- 568 Swenson, N.G. & Enquist, B.J. (2007) Ecological and evolutionary determinants of a key plant
569 functional trait: wood density and its community-wide variation across latitude and
570 elevation. *American Journal of Botany*, **94**, 451-459.
- 571 Westoby, M. & Wright, I.J. (2006) Land-plant ecology on the basis of functional traits. *Trends in*
572 *Ecology & Evolution*, **21**, 261-268.
- 573 Westoby, M., Falster, D.S., Moles, A.T., Vesk, P.A. & Wright, I.J. (2002) Plant ecological
574 strategies: some leading dimensions of variation between species. *Annual Review of*
575 *Ecology and Systematics*, **33**, 125-159.
- 576 Zhu, K., Woodall, C.W. & Clark, J.S. (2012) Failure to migrate: lack of tree range expansion in
577 response to climate change. *Global Change Biology*, **18**, 1042-1052.

578

579 Additional references to the data sources may be found at the end of Appendices S2 and S3 at
580 [URL].

581

582 **Supporting information**

583 Additional supporting information may be found in the online version of this article at the
584 publisher's website.

585 **Appendix S1** FIA plot design and examples of the model.

586 **Appendix S2** Phylogenetic comparative analyses

587 **Appendix S3** Sources selected for trait information

588 **Appendix S4** Additional Random Forest models

589 **Appendix S5** Species values of recruitment potential and density dependence

590 **Appendix S6** Geographical pattern of angiosperm and gymnosperm forest communities

591

592 **Biosketch**

593 Oscar Godoy is a post-doctoral fellow interested in the functional and evolutionary determinants
594 of species interactions, mainly plant competition, and their consequences for community and
595 ecosystem dynamics.

596

597 Marta Rueda is a post-doctoral fellow interested in biogeography and global ecological and
598 evolutionary patterns. Her current research interest also includes fragmentation theory and
599 community phylogenetics at broad scales.

600

601 Bradford A. Hawkins is interested in ecological and phylogenetic patterns across a range of
602 spatial scales, with a focus on linking local and biogeographical processes.

603

604 **Author contributions**

605 O.G, M.R and B.A.H concieved the ideas, prepared the data, performed the analyses and wrote
606 the manuscript.

607

608

Tables

Table 1: Environmental and functional traits predictors included in the analyses. *N* represents the number of species with information for each trait. The total number of species in the database is 164. Source references are included in Appendix S3.

Type	Name	<i>N</i>	Definition	Meaning
Climatic	Annual precipitation (mm)			
	Summer precipitation (mm)			
	Max temperature (°C)			
	Min temperature (°C)			
Species traits	Anaerobic tolerance	141	Relative tolerance to anaerobic soil conditions.	Stress tolerance
	CaCO ₃ tolerance	141	Relative tolerance to calcareous soils.	Stress tolerance
	Dispersal mode	159	Unassisted, animal, or wind.	Disturbance and dispersal
	Drought tolerance	141	Relative species tolerance to drought conditions compared to other species with the same growth habit from the same.	Stress tolerance
	Fire tolerance	139	Relative ability to resprout, regrow, or reestablish from residual seed after a fire.	Regeneration, disturbance and resource use
	Growth rate	137	Growth rate after successful establishment relative to other species with the same growth habit.	Resource use
	Max Height (m)	164		Light acquisition
	Leaf Nitrogen (N _{mass} %)	106		Leaf economics-resource
	Leaf Phosphorus (P _{mass} %)	82		Leaf economics-resource
	Max trunk diameter (m)	144		Light acquisition
	Min frost-free days	140	The minimum average number of frost-free days within the species' geographical range.	Stress tolerance
	Min root depth (cm)	141	The minimum depth of soil required for good growth.	Stress tolerance
	Leaf Phenology	164	Evergreen versus deciduous.	Resource use
	Seed mass (mg)	155		Regeneration
Shade tolerance	143	Relative tolerance to shade conditions.	Light acquisition	
Wood density (mg/cm ³)	113		Resource use, resistance to pathogens	

Table 2: Random Forest models at the species level including all trait predictors (environmental predictors are excluded). Shown are variable importance values (VI) of the species traits with VI >75 for recruitment potential, density dependence, and conspecific density dependence. 100 indicate the species trait with the highest relative importance for the Random Forest model. The sign of Pearson’s correlation are given to indicate the sign of the relationship. The percentage of explained variance (Pseudo- R^2) gives an overall fit for the model.

	Recruitment Potential (RP)	Density Dependence (DD)	Conspecific Density Dependence (CDD)
Pseudo- R^2	0.31	0.28	0.22
<i>Species trait</i>			
Anaerobic tolerance			80.7 (-)
Leaf N mass			100 (-)
Max trunk diameter		95.4 (-)	75.1 (-)
Min frost-free days	87.6 (-)	81.0 (-)	85.3 (-)
Seed mass	100 (-)		75.7 (-)
Wood density		100 (-)	78.3 (-)

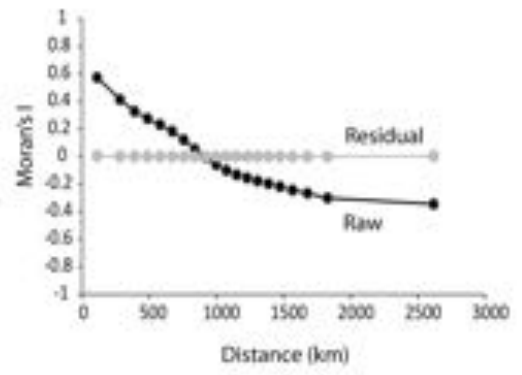
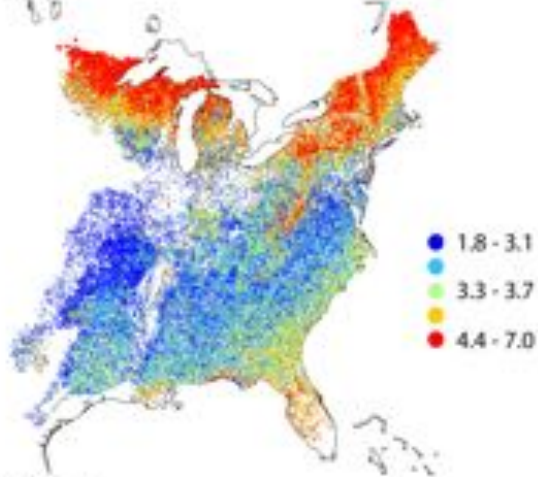
Figure Legends

Figure 1: Figure 1: Left column: (a) geographical pattern of mean recruitment potential (RP), (b) mean strength of density dependence (DD), and (c) mean strength of conspecific density dependence (CDD) for the eastern forest of the USA. Inset in (b) shows change from negative to positive interactions moving northwards. Right column: spatial correlograms of RP, DD, and CDD using random samples of 15,000 sites. Raw data and residuals correspond to Random Forest models provided in Fig. 2. All Moran's I values for the residuals are between -0.02 and 0.03.

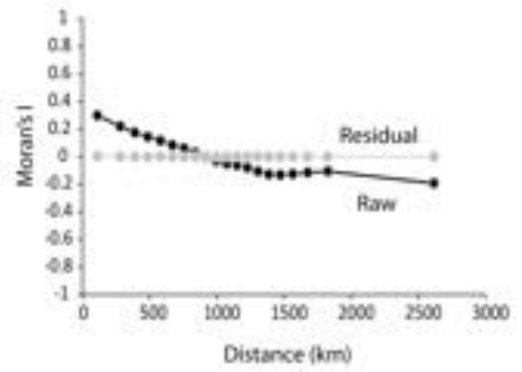
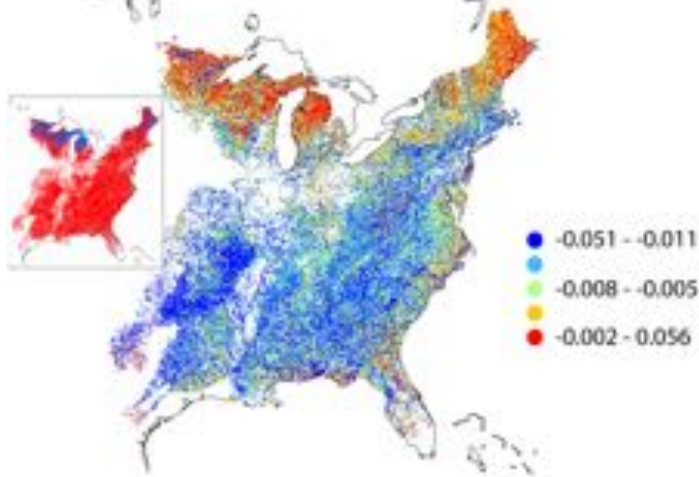
Figure 2: Variable importance values from a Random Forest model (based on 200 regression trees) of (a) mean of recruitment potential (RP), of (b) mean of the strength of density dependence (DD), and of (c) mean of the strength of conspecific density dependence (CDD), including all trait and environmental predictors. Broad-scale environmental variables are in white, and traits are in gray. Pearson's correlations are given for those important predictors to provide the sign of the relationship. Non-abbreviated value names are given in Table 1. Similar models performed only with the most important predictors are presented in Appendix S4.

Figure 3: Geographical pattern of the mean of the most important predictors resulting from the Random Forest analyses. For a detailed description see Table 1.

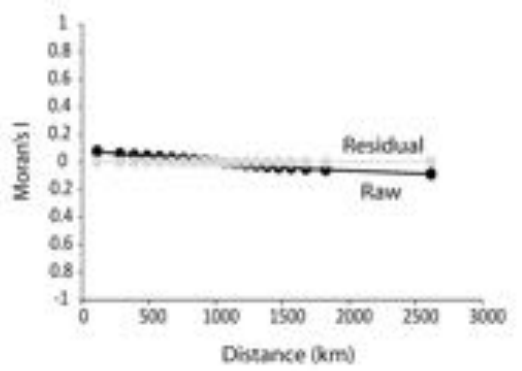
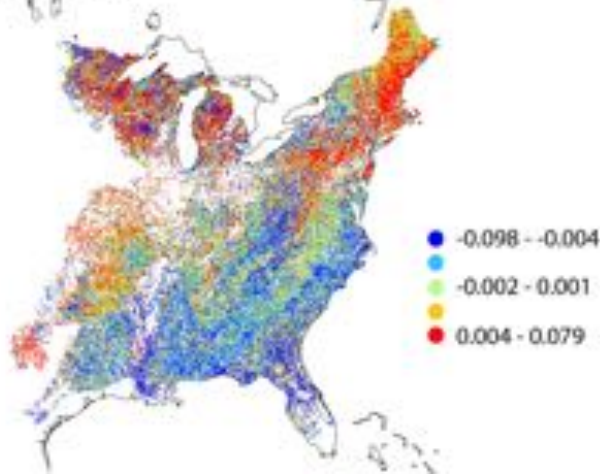
(a) RP



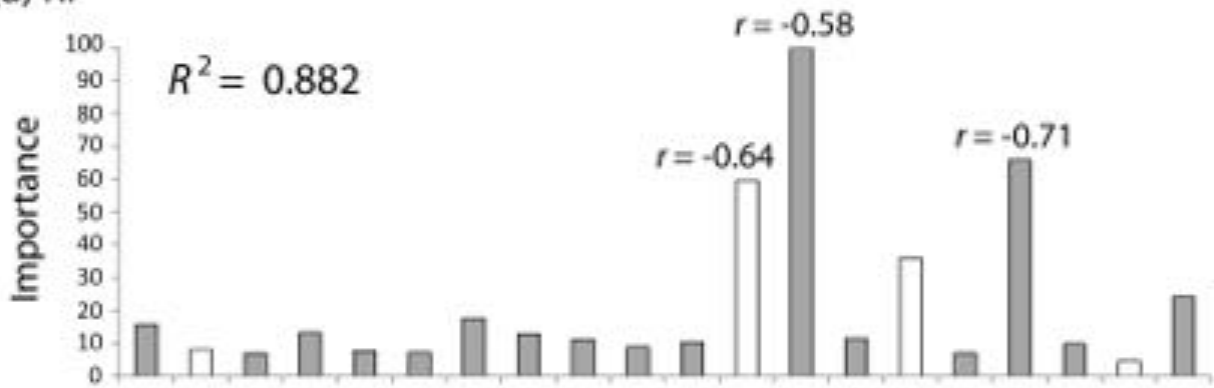
(b) DD



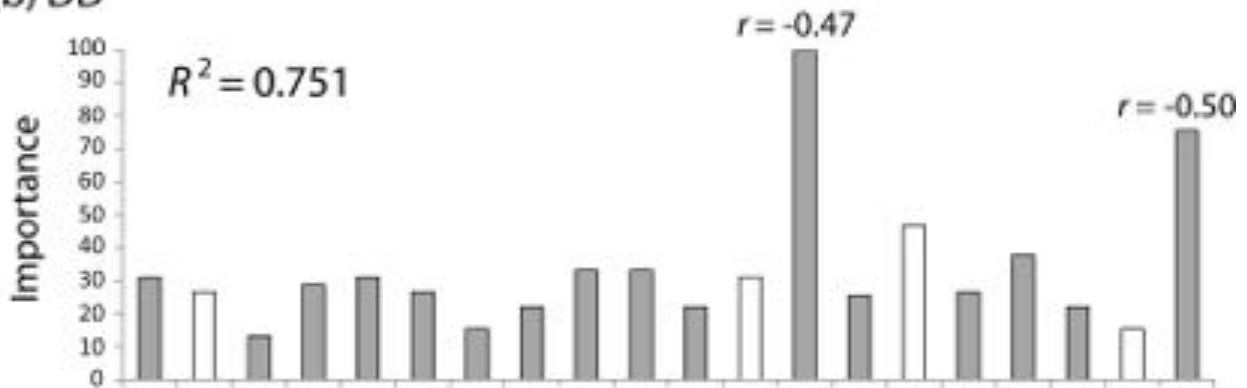
(c) CDD



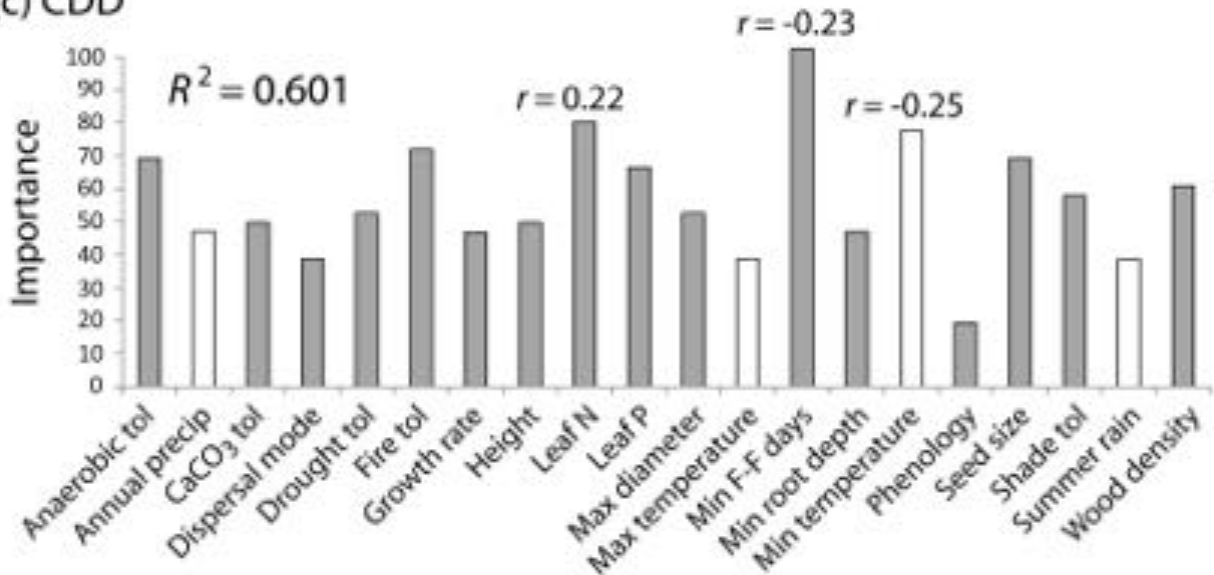
(a) RP



(b) DD



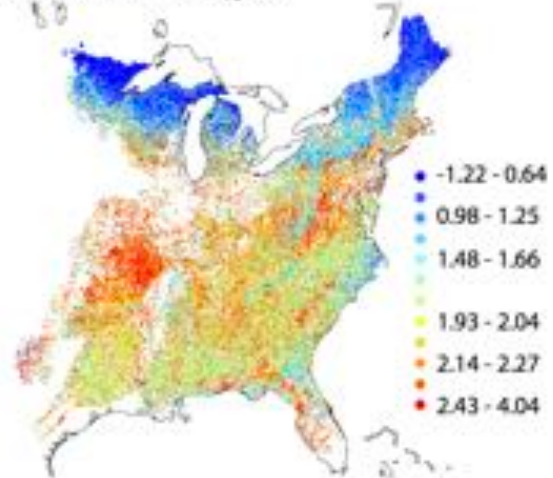
(c) CDD



(a) Min number of frost-free days



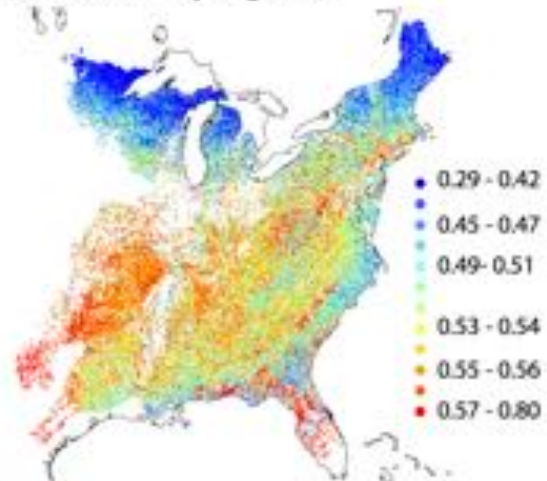
(b) Seed mass (\log_{10} g)



(c) Max temperature in warmest month ($^{\circ}\text{C}$)



(d) Wood density (mg/cm^3)



(e) Min temperature in coldest month ($^{\circ}\text{C}$)



(f) Leaf nitrogen (log percentage)

