Significant competitive dominance in mid-latitude European plant communities

José A. Capitán[1,2], Sara Cuenda[3], Alejandro Ordóñez[4], and David Alonso[2]

[1]Department of Applied Mathematics. Technical University of Madrid. Av. Juan de Herrera, 6. 28040 Madrid, Spain.
[2]Theoretical and Computational Ecology Group. Center for Advanced Studies (CEAB-CSIC). C. Accés Cala St. Francesc 14. 17300 Blanes, Catalonia, Spain.
[3]Universidad Autónoma de Madrid. Facultad de Ciencias Económicas y Empresariales. Depto. Análisis Económico: Economía Cuantitativa. C. Francisco Tomás y Valiente 5, 28049 Madrid, Spain
[4]Department of Bioscience. Aarhus University, Aarhus. Ny Munkegade 114, DK-8000, Aarhus C, Denmark.

May 31, 2016

Corresponding author	José A. Capitán / David Alonso
Address	Department of Applied Mathematics Av. Juan de Herrera, 6 28040 Madrid, Spain
Phone / Fax	+34-913 36 7594 / +34-913 36 7806
E-mail José A. Capitán	`ja.capitan@upm.es`
E-mail Sara Cuenda	`sara.cuenda@uam.es`
E-mail Alejandro Ordóñez	`alejandro.ordonez@bios.au.dk`
E-mail David Alonso	`dalonso@ceab.csic.es`
Type of Article	Letter
Running title	Significant dominance in plant communities
Words Abstract	150
Words Main Text	≈3100
Figures / Tables	3 / 0
References	32
Text Boxes	0

Authorship: JAC and DA conceived and designed the research; JAC and SC performed the statistical analysis and model simulations; JAC, SC, AO and DA analyzed the data and results; AO contributed materials; JAC and DA wrote the paper.

Abstract

Understanding the main determinants of species coexistence across space and time is a central question in ecology. However, ecologists still know little about the scales and conditions at which biotic interactions matter and their interplay with the environment to structure species assemblages. Here we develop ecological theory to analyze plant distribution and trait data across Europe and find that plant height clustering is related to evapotranspiration and gross primary productivity. Our analysis suggests competitive dominance as a plausible mechanism underlying community assembly patterns over continental scales. In particular, we find a clear signal of plant-to-plant competition in mid-latitude ecoregions, where conditions for growth (reflected in actual evapotranspiration rates and gross primary productivities) are optimal. Under severe conditions, either climate is too harsh and overrides the effect of competition or other interactions play a relevant role. Our approach bridges the gap between modern coexistence theory and large-scale species distribution data analysis.

Keywords: Competition, Macroecology, Community Assembly, Stochastic Community Models, Plant Biogeography, Biodiversity.

Introduction

Classical coexistence theory (Gause, 1934, MacArthur & Levins, 1967) assumes that the more similar two species are in their niche requirements, the more strongly they will compete over shared resources, an idea that can be traced back to Darwin (1859). Ever since, competition-similarity hypotheses have been at the front line of theoretical explanations for species coexistence (Abrams, 1983, MacArthur & Levins, 1967), exploring community assembly based on phylogenetic or functional similarity. This framework predicts that large species differences should be selected during community assembly to reduce competition. Therefore, trait and/or phylogenetic overdispersion have often been regarded as signatures of competitive interactions. However, progress in our understanding of how species differences influence the outcome of competitive interactions (Chesson, 2000, Mayfield & Levine, 2010) shows that this theoretical framework is too simplistic because it disregards the balance between stabilizing and equalizing species differences (Chesson, 2000). Stabilizing mechanisms are based on trait differences that cause species to be limited more by their own con-specifics than by their competitors, favoring species when they drop to low densities, which, in turn, promotes species coexistence. Equalizing mechanisms, by contrast, promote species dominance over potential competitors. In the absence of stabilizing species differences, superior competitors would drive other species to extinction through competitive exclusion. In communities controlled by equalizing mechanisms, species with similar trait values should be selected through competitive dominance, resulting in high levels of trait clustering even in the absence of environmental filtering (Mayfield & Levine, 2010). Significant trait clustering at local sites in this theoretical framework may be a fingerprint of biotic (competitive) interactions controlling the composition of ecological communities.

Accurately separating the effect of environmental filters and biotic interactions as structuring agents of community assembly is not trivial. Despite the undeniable success of species distribution models (Peterson *et al.*, 2011), there is an increasing recognition of the need for simple, process-based models to make robust predictions that help understand species responses to environmental change (Wisz *et al.*, 2013). However, we still lack clear evidence for the role of biotic interactions in shaping species assemblages at large spatial scales. Studies based on species randomization models have attempted to separate the outcomes of competitive exclusion and environmental filtering by assuming the competition-similarity hypothesis as a given (Webb *et al.*, 2002). Empirical studies, while they may be able to independently assess environmental stress and species competitive abilities, are often limited to small community sizes (Violle *et al.*, 2011) or restricted to single habitats (Kunstler *et al.*, 2012). Very few studies have explored the idea of competition as a driver of community assembly across biogeographic re-

gions (Kunstler *et al.*, 2016, Swenson *et al.*, 2012). Here we report the results of the first macro-ecological study, based on theoretical predictions from modern coexistence theory (Chesson, 2000, Mayfield & Levine, 2010), aimed at separating the combined effects of biotic (competition) and abiotic (environmental) factors shaping plant community assemblages at large geographical scales.

Light and water availability impose significant limitations on gross primary productivity which is reflected in actual evapotranspiration rates (Garbulsky *et al.*, 2010). These two resources vary at regional scales, placing strong, but opposing constraints on how tall a plant can grow within the limits of structural stability. In addition, plants have to compete strongly for these resources. The resulting plant height is a trait that reflects the ability of the individual to optimize its own growth within these regional environmental and biotic constraints (see Holmgren *et al.* (1997) and references therein). We analyzed presence-absence matrices of floral taxa across different European ecoregions to determine if competitive ability (reflected in maximum stem height) could help explain assemblage patterns across gradients of relevant environmentally-driven factors such as evapotranspiration. We examined how macro-ecological plant data match up to theoretical predictions generated by a synthetic, stochastic framework of community assembly (Capitán *et al.*, 2015, Haegeman & Loreau, 2011). Competition between hetero-specifics was measured by signed height differences, so that clustering could emerge by competitive exclusion of sub-dominant species. We find large fractions of local communities where clustering in maximum stem height is significant at intermediate latitudes, coinciding with a mid-latitude peak in evapotranspiration rates. Across Europe, actual evapotranspiration is lower at more southern latitudes (due to reduced precipitation levels) as well as at more northern latitudes (due to colder temperatures and low levels of sunlight). Species trait clustering is significant only in a latitudinal band where environmental constraints to plant growth are weaker, suggesting that it is only in these mid-latitude ecoregions that a clear signature of competitive dominance can be found in species assemblage patterns.

Materials and methods

European plant ecoregions

Plant community data were drawn from Atlas Florae Europaeae (Jalas & Suominen, 1964–1999). The distribution of flora is geographically described using equally-sized grid cells ($\sim 50 \times 50$ km) based on the Universal Transverse Mercator projection and the Military Grid Reference System, see Fig. 1. Each cell was assigned to a dominant habitat type based on the WWF Biomes of the World classification (Olson *et al.*, 2001), which defines different ecoregions, i.e., geographically distinct assemblages of species

subject to similar environmental conditions. We consider each cell in an ecoregion to represent a species aggregation (which we name 'local community'). We also refer to ecoregions as 'metacommunities'. A total of 3233 plant taxa were extracted from data sources.

Each species in an ecoregion was characterized by its maximum stem height H, an eco-morphological trait that relates to several critical functional strategies among plants. It represents an optimal trade-off between the gains of accessing light, water transport from soil, and sensitivity to biomass loss from mechanical disturbances, and is therefore a good proxy for competition under light- or water-limited conditions (Craine & Dybzinski, 2013). Height values were compiled from the LEDA database (Kleyer *et al.*, 2008) for as many species were available in that source. Missing values were taken from Ordonez *et al.* (2010). According to plant growth forms, 2610 species were grouped as herbaceous (aquatic, herbs, or graminoid) and 623 as woody (shrub or tree). Here, we report results for herbaceous plants, although similar conclusions were obtained for the entire data set (see Figs. S5, S6, and S7 in Supporting Information).

Species height values spanned several orders of magnitude, so we used a log-transformed variable ($h = \log H$) to measure species differences (using non-transformed heights yielded similar results, see Supporting Information, Figs. S5, S6, and S7). The values of h were standardized within ecoregions as $t = (h - h_{\min})/(h_{\max} - h_{\min})$ so that $0 \leq t \leq 1$. Species competitive dominance was measured by standardized, signed height differences, $\rho_{ij} = \hat{\rho}(t_i - t_j)$. This choice assumes the subsequent selection for low-trait valued species. Alternatively, the opposite choice $[\rho_{ij} = \hat{\rho}(t_j - t_i)]$ implies that large traits will be over-represented in local communities (Supporting Information, Sec. S2.3 and S3.2). The scale factor $\hat{\rho}$ measures the ratio between inter- *vs.* intraspecific competition. For all the species reported in an ecoregion, we formed a competition matrix with these ρ_{ij} values. The advantage of having these values represent trait differences between pairs of species is that any trend in competition strengths can be immediately translated into patterns of functional clustering or overdispersion.

For a metacommunity with S species, we calculated the average competitive strength as $\langle \rho \rangle = \frac{2}{S(S-1)} \sum_{i=1}^{S} \sum_{j=i+1}^{S} |\rho_{ij}|$. Of the S potential species, a number $s_k \leq S$ of species will form a local community assemblage at cell k. The coexistence probability was defined as the average fraction $p_{\mathrm{c}} = \langle s \rangle / S = \frac{1}{S N_{\mathrm{L}}} \sum_{k=1}^{N_{\mathrm{L}}} s_k$, with N_{L} representing the number of cells in the ecoregion. These quantities were used to compare model predictions with real data.

Stochastic community model

Community assembly results from the interplay of four fundamental processes: speciation, ecological drift, dispersal, and selection acting across

space and time (Vellend, 2010). New species are added to local communities through speciation and/or immigration, species abundances are then shaped by stochastic births and deaths (ecological drift), ongoing dispersal and differential growth driven by species interactions and/or specific adaptations to local environment (selection). Our approach disregards speciation, does not account for environmental factors and assembles communities based only on dispersal, ecological drift and asymmetric competition processes. Community dynamics is mathematically described as a birth-death-immigration process in continuous time (Supporting Information, Sec. S2.2). Let $\mathbf{n} = (n_1, \ldots, n_S)$ be a vector of population numbers for all species in the metacommunity. At each time step, one of the following events can take place: (1) immigrants arrive from the metacommunity at rate μ, (2) local births and deaths occur at rates $\alpha^+ n_i$ and $\alpha^- n_i$, respectively, for species i, (3) two individuals of the same species i compete at a rate $\alpha n_i^2/K$, where $\alpha = |\alpha^+ - \alpha^-|$ and K is interpreted as a carrying capacity, (4) two individuals of distinct species i and j compete at a rate $\alpha \rho_{ij} n_i n_j / K$, resulting in an increase of population size n_i if $\rho_{ij} < 0$ and a decrease if $\rho_{ij} > 0$ (Supporting Information, Sec. S2.1). For the sake of simplicity, trait values t_i in the model were drawn from a Gaussian distribution with zero mean and variance σ^2. Interactions were chosen as $\rho_{ij} = \frac{\langle\rho\rangle\sqrt{\pi}}{2\sigma}(t_i - t_j)$ so that the expected value of $|\rho_{ij}|$ is precisely $\langle\rho\rangle$ (Supporting Information, Sec. S2.1).

The stochastic dynamics predicts the identities of species selected by competition in local communities, as well as the observed local diversity relative to the metacommunity richness (i.e., the coexistence probability). We used the variability in coexistence probability as a function of average competitive strength and the distribution of pairwise competitive strengths observed in real communities to test how closely they confirm the predictions of the stochastic community assembly model.

Randomization tests

Null models for community assembly (Webb *et al.*, 2002) compare the properties of actual communities against random samples of the same size extracted from the metacommunity. This approach assumes that local communities are built up through the independent arrival of equivalent species from the metacommunity (Alonso *et al.*, 2015, MacArthur & Wilson, 1967) regardless of species preferences for particular environments or species interactions (see Supporting Information, Sec. S2.5, for details). Local species richness is determined by inherent site-level differences in the ability to harbor different numbers of species (Chase *et al.*, 2011). Our randomization tests were based on the average competitive strength of a cell with s species, $\langle\rho\rangle_{\mathrm{L}} = \frac{2}{s(s-1)} \sum_{i=1}^{s} \sum_{j=i+1}^{s} |\rho_{ij}^{\mathrm{L}}|$, where (ρ_{ij}^{L}) is the submatrix of the metacommunity competition matrix restricted to the species present in the local community. Compared to metacommunity samples, the lower (higher) the

empirical local average $\langle\rho\rangle_{\mathrm{L}}^{\star}$ is, the higher (lower) is the degree of species clustering in the local community. For each local community we calculated the probability $p = \Pr(\langle\rho\rangle_{\mathrm{L}}^{\star} \leq \langle\rho\rangle_{\mathrm{L}})$ that the empirical local average is smaller than the randomly-sampled competition average. At a 5% significance level, if $p > 0.95$ the empirical competition average is significantly larger than the average measured for random metacommunity samples, signifying that average trait differences in local communities are larger than expected at random. On the other hand, if $p < 0.05$, observed trait differences are, on average, significantly smaller than those obtained for random metacommunity samples. In either case the null hypothesis (i.e., local communities are built as random assemblages from the metacommunity) can be rejected, which implies clear signals of "significant trait overdispersion" ($p > 0.95$) or "significant trait clustering" ($p < 0.05$).

Results

Two predictions from theory

A first theoretical prediction from the stochastic community model is that, as competition increases, coexistence probability shows a power-law decay whose exponent is controlled by the immigration rate μ (Supporting Information, Fig. S4). In particular, the curves for different metacommunity sizes collapse when represented as a function of the competitive strength scaled by the metacommunity richness S, $p_{\mathrm{c}} \sim (\langle\rho\rangle S)^{-\gamma}$ (see Fig. 2a and Supporting Information, Sec. S2.3, for a theoretical derivation of the curve collapse). This collapse eliminates the variability in S, so that empirical coexistence probabilities, which arise from different metacommunity sizes, can be fitted together (Fig. 2a).

To test the significance of competitive dominance in local communities, we generated a second model prediction applying randomization tests to model communities. Fig. 2b shows the distribution of p-values observed when simulated local communities are compared to metacommunity samples (Supporting Information, Sec. S2.5). The model predicts high levels of trait clustering for low immigration rates and high carrying capacity values (Fig. 2b), as well as the selection for low/large-trait valued species (depending on how ρ_{ij} is defined, either as $\rho_{ij} = \hat{\rho}(t_i - t_j)$ or $\rho_{ij} = \hat{\rho}(t_j - t_i)$, see Supporting Information, Sec. S2.1). When immigration increases, local communities show a broad distribution of p-values. At the highest immigration rates, model communities are essentially random samples of the metacommunity since immigration overrides competition in this regime (Etienne & Alonso, 2005). Fig. 2a provides a hint of the range in which model parameters best fit real data. For a realistic fit, the exponent of the empirical power law is obtained for $\mu/\alpha \sim 0.1$ individuals per generation. Since plant communities operate in a low-immigration regime, the non-dimensional im-

migration rate $\lambda = \mu/\alpha K$ must satisfy $\lambda = 0.1/K \ll 1$, hence the carrying capacity must be large. With these parameter values, the model predicts a significant degree of species clustering (Fig. 2b).

Testing model predictions against data

Regarding the first prediction, we found a significant correlation between coexistence probability and the scaled competitive overlap based on empirical data ($r^2 = 0.52$, $p < 10^{-3}$, Fig. 2a). Apparently, a model driven solely by dominant competitive interactions reliably predicts the average richness of local plant communities across different ecoregions. In addition, this theoretical prediction allowed an indirect estimation of the relative importance $\hat{\rho}$ of inter- *vs.* intraspecific average effects: the average ratio of inter- to intraspecific competition strength is about 4% (see Supporting Information, Sec. S2.1, for details on the estimation procedure).

On the other hand, we calculated p-values for randomization tests applied to all local communities in each ecoregion, which represent the empirical metacommunity distribution of p-values (Fig. 2c). At the parameter values that make plant data consistent with the first prediction, our model predicts significant trait clustering (Fig. 2b). When this prediction is compared with empirical data, we observe that some ecoregions fit best with our model, based on competitive dominance, while others clearly do not. In addition, no ecoregion is consistent with trait overdispersion (Fig. 2c).

Ecoregion clustering and actual evapotranspiration rates

In order to better quantify the propensity of an ecoregion to exhibit clustering in maximum stem height, we define a clustering index q for an ecoregion as the fraction of its local communities that lie within the 5% range of significant clustering ($p < 0.05$). An ecoregion for which significant clustering is found in most of its local communities will tend to score high in the q index. We examined how the clustering index varied across the continent (geographical location of ecoregion centroids) as well as with actual evapotranspiration (Fig. 3). Evapotranspiration maps were obtained from data estimated through remote sensing (Mu *et al.*, 2011). Water availability acts as a factor limiting plant growth at geographical scales, and correlates with gross primary productivity (Garbulsky *et al.*, 2010), see Fig. 3d. Therefore, for a given region, mean annual evapotranspiration is a distinct measure of environmental constraints on plant growth (Garbulsky *et al.*, 2010). Panels a and b of Fig. 3 show a clear latitudinal trend: there is an intermediate range of ecoregion latitudes where both clustering indices and evapotranspiration are large, indicating that evapotranspiration measures can robustly predict clustering indices (Fig. 3c). More importantly, since evapotranspiration is a powerful proxy of environmental constraints on plant growth, this clustering

in maximum stem height appears to be strongest at ecoregions less limited by environmental conditions. As environments become harsher and less optimal for plant growth, these clustering patterns disappear. This is particularly true for the severe climatic conditions characteristic in the Mediterranean (with erratic rainfall, limited water availability and drought), as well as of boreal zones (with low radiation incidence and cold temperatures). The harshness of these conditions likely override the effects of competition, and other processes such as species tolerances and facilitation (Valiente-Banuet & Verdú, 2007) may be critical community drivers at climatic extremes. According to model predictions, the overall clustering patterns found at middle-range latitudes are consistent with species competitive dominance driven by height differences (as well as with the over-representation of low-height values in local communities, see Supporting Information, Sec. S3.2 and Fig. S8).

Discussion

To the best of our knowledge, this is the first time that predictions from modern coexistence theory (Chesson, 2000) and the competition-similarity paradigm (Mayfield & Levine, 2010) have been tested with macro-ecological trait data at large spatial scales. While potential evapotranspiration decreases with latitude, actual evapotranspiration peaks at intermediate latitudes, and is strongly associated with higher levels of local trait clustering. Critically, actual evapotranspiration is positively correlated with gross primary productivity (GPP) across terrestrial ecosystems (see Fig. 3d and Garbulsky *et al.* (2010)), which also peaks at intermediate latitudes across Europe (Supporting Information, Fig. S3). The confrontation of model predictions against plant community data across Europe reveals a clear signature of competition in local communities in the environmentally conducive middle-range latitudes; as environmental conditions get increasingly extreme, they override competitive effects resulting in non-significant clustering in plant heights.

Harsh environments either interfere with competitive interactions or create conditions for other types of interactions to drive community assembly. In either case, plant trait clustering is not apparent. Although competition may play a role, particularly at small spatial scales, facilitation is likely to be more important under stressful conditions (Maestre *et al.*, 2009). By contrast, the mild environments characteristic of mid-latitude ecoregions impose less stringent limits to plant growth and the effect of species interactions through competitive dominance can be observed even at large biogeographic scales. At these intermediate latitudes, differences in competitive ability tend to be weaker, on average, and trait clustering is likely to arise from competitive interactions. In these ecoregions, competition, by

filtering out subdominant species, leaves a significant trace on local community assembly and, as empirical data confirm, species tend to cluster around smaller height values (Supporting Information, Fig. S8). Interestingly, the relevant role of competitive dominance driven by species trait hierarchies has been also reported at much smaller spatial scales for forest trees in the French Alps (Kunstler *et al.*, 2012). In addition, a recent study of the assembly of forest communities across East Asia shows that a phylogenetic-based species similarity index tends to be smaller the higher the minimum temperature of the coldest month is (Feng *et al.*, 2015). Together with our results, these studies suggest that trait clustering is generally likely to occur where conditions for plant growth are less restrictive. Our model indicates that the process underlying this pattern is competitive dominance, although it is likely that community assembly for other taxa may be driven by other biotic or environmental filters. For instance, phytoplankton communities appear to be driven by limiting similarity creating clumpy species coexistence in estuarine ecosystems (Segura *et al.*, 2012).

Theoretical approaches to understand the forces shaping ecological communities rely on mathematical models of species interactions in order to predict community assembly rules, species coexistence, and community stability. Empirical approaches rely, instead, on sampling existing ecological communities to collect as much information as possible on species composition, abundances, and associated environmental variables across temporal and spatial scales in order to infer regularities using statistical models. Finding a theoretically robust and ecologically meaningful rapprochement at relevant scales remains a challenge for ecology, and we trust that our work will inspire new contributions in this direction.

Conclusion

This analysis uncovers a new macro-ecological empirical pattern involving the relationship between trait clustering in maximum stem height of plant species and plant primary productivity (and actual evapotranspiration) over large spatial scales. The intensity of this clustering increases with primary productivity, this is, it is more significant in regions with better conditions for plant growth. Conversely, in harsh environments this clustering disappears. Our theoretical investigations point to the role of competitive dominance as a plausible explanation underlying this pattern. The relevance of this mechanism at driving community assembly, which tends to equalize differences in competitive ability, should increase with gross primary productivity. Our strategy provides a clear direction for quantitatively assessing the generality of this prediction in other regions, for different taxa, and for other traits.

Acknowledgements

The authors thank Mercedes Pascual, Jonathan Chase and Stefano Allesina for their comments, and are indebted to Rohan Arthur for his constructive criticism of the manuscript. This work was funded by the Spanish "Ministerio de Economía y Competitividad" under the project SITES (CGL2012-39964, DA, JAC) and the Ramón y Cajal Fellowship program (DA). JAC acknowledges partial financial support from the Department of Applied Mathematics, Technical University of Madrid.

References

1.
Abrams, P. A. (1983). The theory of limiting similarity. *Annu. Rev. Ecol. Syst.*, 14, 359–376.

2.
Alonso, D., Pinyol-Gallemí, A., Alcoverro, T. & Arthur, R. (2015). Fish community reassembly after coral mass mortality: higher trophic groups are subject to increased rates of extinction. *Ecol. Lett.*, 18, 451–461.

3.
Capitán, J. A., Cuenda, S. & Alonso, D. (2015). How similar can co-occurring species be in the presence of competition and ecological drift? *J. R. Soc. Interface*, 12, 20150604.

4.
Chase, J. M., Kraft, N. J. B., Smith, K. G., Vellend, M. & Inouye, B. D. (2011). Using null models to disentangle variation in community dissimilarity from variation in α-diversity. *Ecosphere*, 2, 24.

5.
Chesson, P. L. (2000). Mechanisms of maintenance of species diversity. *Ann. Rev. Ecol. Syst.*, 31, 343–366.

6.
Craine, J. R. & Dybzinski, R. (2013). Mechanisms of plant competition for nutrients, water and light. *Funct. Ecol.*, 27, 833–840.

7.
Darwin, C. (1859). *On the origin of species by means of natural selection.* John Murray, London, UK.

8.
Etienne, R. S. & Alonso, D. (2005). A dispersal-limited sampling theory for species and alleles. *Ecol. Lett.*, 8, 1147–1156.

9.
Feng, G., Mi, X., Eiserhardt, W. L., Jin, G., Sang, W., Lu, Z. *et al.* (2015). Assembly of forest communities across east-asia —insights from phylogenetic community structure and species pool scaling. *Scientic Reports*, 5, 9337.

10.
Garbulsky, M. F., Peñuelas, J., Papale, D., Ardö, J., Goulden, M. L., Kiely, G. *et al.* (2010). Patterns and controls of the variability of radiation use efficiency and primary productivity across terrestrial ecosystems. *Global Ecology and Biogeography*, 19, 253–267.

11.
Gause, G. F. (1934). *The struggle for existence.* Williams & Wilkins, Baltimore, MD.

12.
Haegeman, B. & Loreau, M. (2011). A mathematical synthesis of niche and neutral theories in community ecology. *J. Theor. Biol.*, 4, 263–271.

13.
Holmgren, M., Scheffer, M. & Huston, M. A. (1997). The interplay of facilitation and competition in plant communities. *Ecology*, 78, 1966–1975.

14.
Jalas, J. & Suominen, J. (1964–1999). *Atlas Florae Europaeae. Distribution of vascular plants in Europe, Vol. 1–12.* Societas Biologica Fennica Vanamo, Helsinki.

15.
Kleyer, M., Bekker, R. M., Knevel, I. C., Bakker, J. P., Thompson, K., Sonnenschein, M. *et al.* (2008). The leda traitbase: a database of life-history traits of the northwest european flora. *Journal of Ecology*, 96, 1266–1274.

16.
Kunstler, G., Falster, D., Coomes, D. A., Hui, F., Kooyman, R. M., Laughlin, D. C. *et al.* (2016). Plant functional traits have globally consistent effects on competition. *Nature*, 529, 204–207.

17.
Kunstler, G., Lavergne, S., Courbaud, B., amd G. Vieilledent, W. T., Zimmermann, N. E., Kattge, J. *et al.* (2012). Competitive interactions between forest trees are driven by species' trait hierarchy, not phylogenetic or functional similarity: implications for forest community assembly. *Ecol. Lett.*, 15, 831–840.

18.
MacArthur, R. H. & Levins, R. (1967). The limiting similarity, convergence and divergence of coexisting species. *Am. Nat.*, 101, 377–385.

19.
MacArthur, R. H. & Wilson, E. O. (1967). *The theory of island biogeography.* Princeton University Press, Princeton.

20.
Maestre, F. T., Callaway, R. M., Valladares, F. & Lortie, C. J. (2009). Refining the stress-gradient hypothesis for competition and facilitation in plant communities. *Journal of Ecology*, 97, 199–205.

21.
Mayfield, M. M. & Levine, J. M. (2010). Opposing effects of competitive exclusion on the phylogenetic structure of communities. *Ecol. Lett.*, 13, 1085–1093.

22.
Mu, Q., Zhao, M. & Running, S. W. (2011). Improvements to a MODIS global terrestrial evapotranspiration algorithm. *Remote Sens. Environ.*, 115, 1781–1800.

23.
Olson, D. M., Dinerstein, E., Wikramanayake, E. D., Burgess, N. D., Powell, G. V. N., Underwood, E. C. *et al.* (2001). Terrestrial ecoregions of the world: A new map of life on Earth. *BioScience*, 51, 993–938.

24.
Ordonez, A., Wright, I. J. & Olff, H. (2010). Functional differences between native and alien species: A global-scale comparison. *Fun. Ecol.*, 24, 1353–1361.

25.
Peterson, A. T., Soberón, J., Pearson, R. G., Anderson, R. P., Martínez-Meyer, E., Nakamura, M. *et al.* (2011). *Ecological Niches and Geographic Distributions.* Princeton University Press.

26.
Segura, A. M., Kruk, C., Calliari, D., García-Rodríguez, F., Conde, D., Widdicombe, C. E. *et al.* (2012). Competition drives clumpy species coexistence in estuarine phytoplankton. *Scientific Reports*, 3, 1037.

27.
Swenson, N. G., Enquist, B. J., Pither, J., Kerkhoff, A. J., Boyle, B., Weiser, M. D. *et al.* (2012). The biogeography and filtering of woody plant functional diversity in North and South America. *Global Ecology and Biogeography*, 21, 798–808.

28.
Valiente-Banuet, A. & Verdú, M. (2007). Assembly through facilitation can increase the phylogenetic diversity of plant communities. *Ecol. Lett.*, 10, 1029–1036.

29.
Vellend, M. (2010). Conceptual synthesis in community ecology. *Q. Rev. Biol.*, 85, 183–206.

30.
Violle, C., Nemegut, D. R., Pu, Z. & Jiang, L. (2011). Phylogenetic limiting similarity and competitive exclusion. *Ecol. Lett.*, 14, 782–787.

31.
Webb, C. O., Ackerly, D. D., McPeek, M. A. & Donoghue, M. J. (2002). Phylogenies and community ecology. *Ann. Rev. Ecol. Syst.*, 33, 475–505.

32.
Wisz, M., Pottier, J., Kissling, W. D., Pellisier, L., Lenoir, J., Damgaard, C. *et al.* (2013). The role of biotic interactions in shaping distributions and realised assemblages of species: Implications for species distribution modelling. *Biol. Rev.*, 88, 15–30.

Figure captions

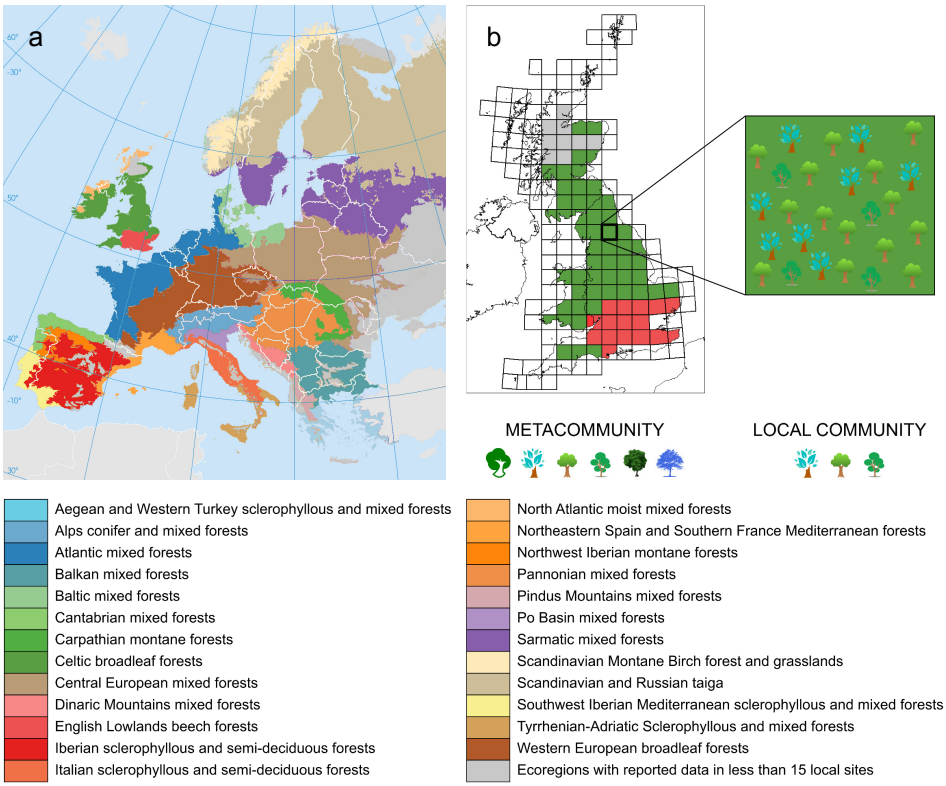

Figure 1: **Geographical description of plant data across European ecoregions.** **a**, 25 different habitats covering most of Europe are shown in the map and listed below. Ecoregions are regarded as metacommunities comprising all plant species observed in that region. **b**, The Military Grid Reference System divides ecoregions in grid cells, each one considered as a local community formed by a species sample of the metacommunity.

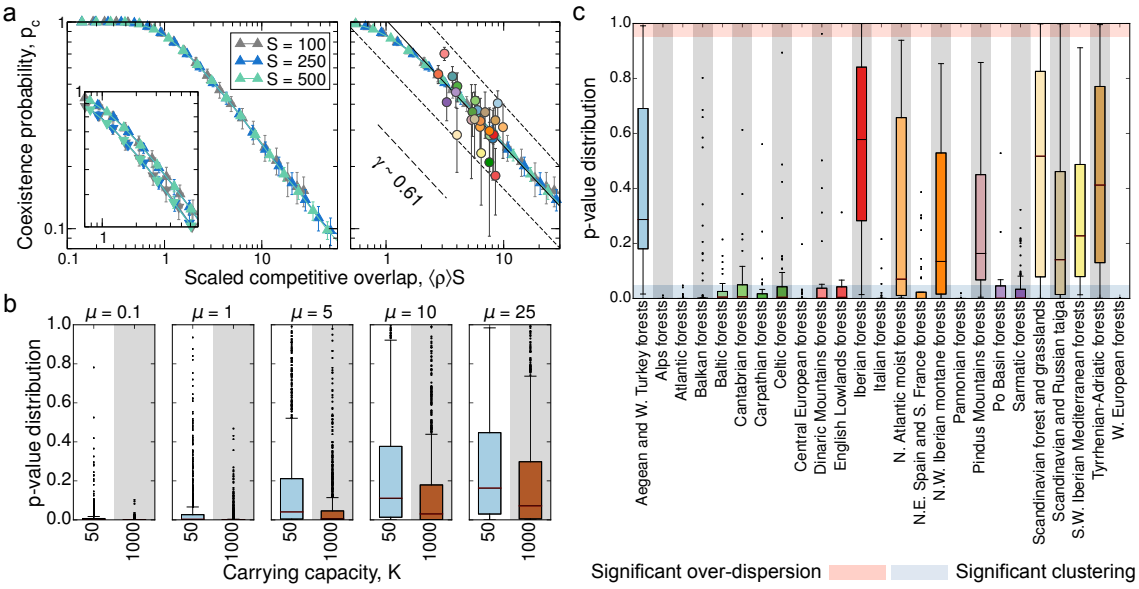

Figure 2: **Two theoretical predictions tested against data.** **a**, Left: Average fraction p_c of coexisting species in local communities as a function of the scaled competitive strength, $\langle\rho\rangle S$. Simulation parameters are $\alpha^+ = 50$, $\alpha^- = 0.1$, $\mu = 5$, $\sigma = 0.2$ and $K = 50$ (down triangles, inset) or $K = 1000$ (up triangles). Right: The model predicts a power-law decay $p_c \sim (\langle\rho\rangle S)^{-\gamma}$ regardless of the metacommunity size S, which permits fitting a power law to data ($r^2 = 0.52$, $p < 10^{-3}$, 95% prediction intervals shown). In order for model curves to match empirical data, we need to choose $\hat{\rho} = 0.039$ in the calculation of ρ_{ij} (Supporting Information). **b**, Model randomization tests for different immigration rates and carrying capacities; here $\langle\rho\rangle = 0.06$ and $S = 100$ (Supporting Information, Sec. S2.5). The closer the distribution is to 0, the larger is the fraction of cells where trait clustering is significant. For parameter values that fit data in the low immigration regime ($\mu \sim 5$, $\mu/\alpha K \ll 1$, panel **a**), the model indicates a clear signature of clustering. **c**, Empirical randomization tests. The majority of ecoregions are consistent with model predictions as the distributions (Tukey boxplots) lie in the 5% range of significant clustering. Color codes for data in panels **a** and **c** match codes in Fig. 1.

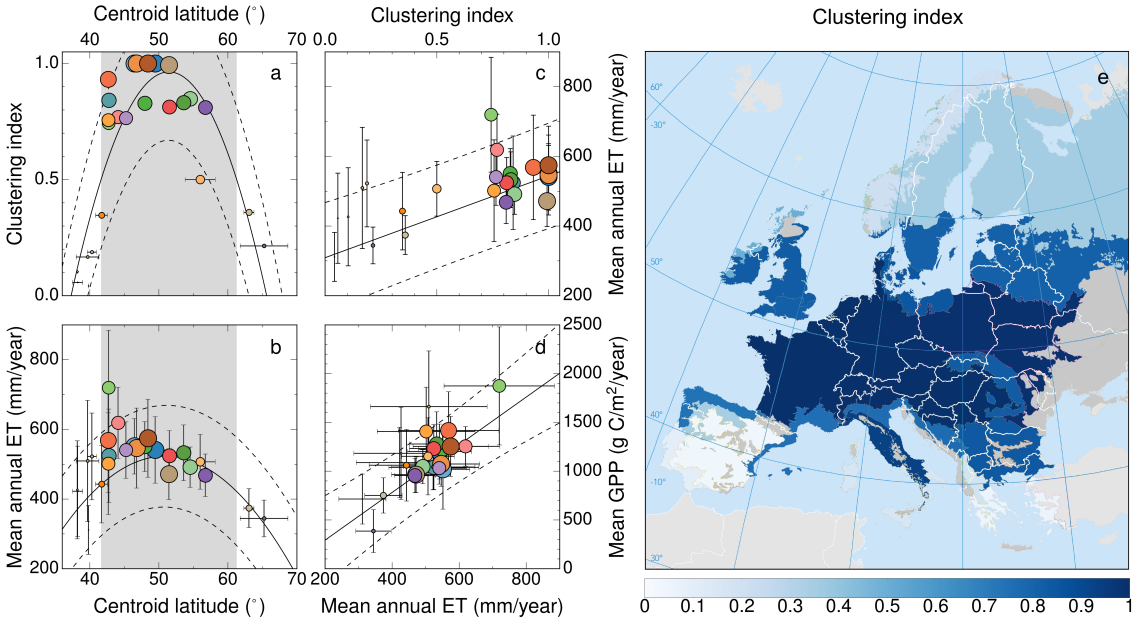

Figure 3: **Linking height clustering to geographical and environmental variables.** **a**, Variation in the clustering index (q) with latitude (φ). Quadratic fit: $r^2 = 0.77$, $p < 10^{-3}$. **b**, Latitudinal variation in mean annual evapotranspiration (ET) data. Quadratic weighted regression: $r^2 = 0.63$, $p < 10^{-3}$. The shaded areas in panels **a** and **b** represent the latitudinal range for which the adjusted dependence $q(\varphi) \geq 1/2$, where both height clustering and evapotranspiration are maximal. **c**, Linear weighted regression for ET as a function of the clustering index; $r^2 = 0.60$, $p < 10^{-3}$. **d**, Correlation between mean gross primary productivity (GPP) and mean annual ET; linear weighted fit: $r^2 = 0.73$, $p < 10^{-3}$. In the first four panels, the radius of each circle is proportional to the clustering index. Symbol colors refer to different ecoregions (Fig. 1). All the fits show the 95% prediction intervals. **e**, Geographical distribution of clustering indices for ecoregions across Europe.