

History and environment shape species pools and community diversity in European beech forests

Borja Jiménez-Alfaro (ORCID ID: 0000-0001-6601-9597)^{1,2,3*}, Marco Girardello (0000-0003-0699-8628)^{4,5}, Milan Chytrý (0000-0002-8122-3075)³, Jens-Christian Svenning (0000-0002-3415-0862)^{4,6}, Wolfgang Willner^{7,8}, Jean-Claude Gégout⁹, Emiliano Agrillo (0000-0003-2346-8346)¹⁰, Juan Antonio Campos (0000-0001-5992-2753)¹¹, Ute Jandt (0000-0002-3177-3669)^{1,2}, Zygmunt Kącki (0000-0002-2241-1631)¹², Urban Šilc (0000-0002-3052-699X)¹³, Michal Slezák (0000-0002-6926-7139)^{14,15}, Lubomír Tichý³, Ioannis Tsiripidis (0000-0001-9373-676X)¹⁶, Pavel Dan Turtureanu¹⁷, Mariana Ujházyová (0000-0002-5546-1547)¹⁸ and Thomas Wohlgemuth (0000-0002-4623-0894)¹⁹.

¹ German Centre for Integrative Biodiversity Research (iDiv) Halle-Jena-Leipzig, Deutscher Platz 5e, 04103 Leipzig, Germany

² Martin Luther University Halle-Wittenberg, Institute of Biology/Geobotany and Botanical Garden, Am Kirchtor 1, 06108 Halle, Germany

³ Department of Botany and Zoology, Masaryk University, 61137 Brno, Czech Republic

⁴ Section for Ecoinformatics and Biodiversity, Department of Bioscience, Aarhus University, 8000 Aarhus C, Denmark

⁵ cE3c – Centre for Ecology, Evolution and Environmental Changes, Azorean Biodiversity Group, 9700-042 Angra do Heroísmo, Portugal

⁶ Center for Biodiversity Dynamics in a Changing World (BIOCHANGE), Aarhus University, 8000 Aarhus C, Denmark

⁷ VINCA - Vienna Institute for Nature Conservation and Analyses, Wien, Austria.

⁸ Department of Botany and Biodiversity Research, University of Vienna, 1030 Vienna, Austria

⁹ LERFoB, INRA, AgroParisTech, 54000 Nancy, France

¹⁰ Department of Environmental Biology, Sapienza University of Rome, 500185 Roma, Italy

¹¹ Department of Plant Biology and Ecology, University of Basque Country UPV/EHU, 48080 Bilbao, Spain

¹² Department of Vegetation Ecology, Botanical Garden, University of Wrocław, 50335 Wrocław, Poland

¹³ ZRC SAZU, Institute of Biology, 1001 Ljubljana, Slovenia

¹⁴ Institute of Forest Ecology, Slovak Academy of Sciences, 960 53 Zvolen, Slovakia

¹⁵ Plant Science and Biodiversity Center, Institute of Botany, Slovak Academy of Sciences, 845 23 Bratislava, Slovakia

¹⁶ Department of Botany, School of Biology, Aristotle University of Thessaloniki, 54124 Thessaloniki, Greece

¹⁷ Alexandru Borza Botanical Garden, Babeş-Bolyai University, 400015 Cluj-Napoca, Romania

¹⁸ Department of Applied Ecology, Faculty of Ecology and Environmental Sciences, Technical University in Zvolen, 96053 Zvolen, Slovakia

¹⁹ WSL Swiss Federal Research Institute, 8903 Birmensdorf, Switzerland

SUMMARY PARAGRAPH

A central hypothesis of ecology states that regional diversity influences local diversity through species-pool effects. Species pools are supposedly shaped by large-scale factors and then filtered into ecological communities, but understanding these processes requires the analysis of large datasets across several regions. Here, we use a framework of community assembly at a continental scale to test the relative influence of historical and environmental drivers, in combination with regional or local species pools, on community species richness and community completeness. Using 42,173 vegetation plots sampled across European beech forests, we found that large-scale factors largely accounted for species-pool sizes. At the regional scale, main predictors reflected historical contingencies related to postglacial dispersal routes, while at the local scale the influence of environmental filters was predominant. Proximity to Quaternary refugia and high precipitation were the main factors supporting community species richness, especially among beech-forest specialist plants. Models for community completeness indicate the influence of large-scale factors, further suggesting community saturation as a result of dispersal limitation or biotic interactions. Our results empirically demonstrate how historical factors complement environmental gradients to provide a better understanding of biodiversity patterns across multiple regions.

MAIN TEXT

A major challenge in biodiversity research is developing empirical studies for addressing large-scale patterns of ecological communities^{1,2}. Macroecological studies rely on the variation of diversity in coarse sampling units across large-scale ecological drivers^{3,4}, while community-level studies typically focus on local assembly processes or the relationships between species diversity and ecosystem functions^{5,6}. However, understanding the drivers of large-scale spatial variation in community diversity (e.g. species richness or species composition) requires an integrative approach to consider the interplay of different factors across multiple spatial and temporal scales^{2,7}.

In the framework of community assembly⁸, the integration of different scales has been traditionally addressed by the species-pool hypothesis, stating that regional diversity is further refined into local diversity^{9,10}. The species pool consists of all the species in a region that can disperse and establish in a specific site, and these pools are generally investigated for the species that can persist in a given habitat, or *habitat species pool*¹¹. However, simple correlations between regional and local diversity provide little information about the drivers of the observed relationships^{7,12}, resulting in operational limitations of the species-pool hypothesis for empirical research¹³. New perspectives to advance the understanding of species-pool effects include the development of new approaches for understanding the assembly of regional pools¹⁴, the use of appropriate model systems with known biogeographic histories¹³ or the integration of other diversity metrics such as dark diversity or community completeness¹⁵.

While a major focus has been on the relationships between species pools and local community diversity, less attention has been paid to the influence of biogeographic history on the formation of regional pools^{14,16,17}. It has been suggested that the historical abundance of habitats has a direct effect on habitat species pools^{18,19} but we still need to know to what extent regional diversity informs us about historical and environmental drivers of species assemblages at different scales¹³. The species pool is generally conceived at the regional scale (from tens to

hundreds of km), assuming that community assembly is mainly influenced by dispersal and environmental filtering at the metacommunity level^{11,20}. However, focusing on a single region ignores biogeographic variation among regions, which is essential for understanding the historical formation of species pools and, in turn, local diversity⁷. For more complete understanding of community assembly processes, it is essential to extend the species-pool framework to large scales (e.g. continents, realms) accounting for different regions^{21,22}. We thus propose a multi-regional framework in which community diversity results from the interplay of historical contingencies, environmental variation and dispersal events that filter continental floras across different scale-dependent processes (Fig. 1). Under this framework, we investigate the extent to which large-scale factors shape species pools and community diversity across European beech forests as a model system.

Our first aim is to test whether historical and environmental factors explain patterns of community species richness in conjunction with regional species pools. If large-scale predictors are able to explain species-pool effects, at least partially, they will provide useful information about the drivers of community diversity across regions. We also test whether the influence of those predictors is consistent when explaining community completeness. This metric reflects how much of the species pool is realized within local communities, facilitating comparisons across different regions to assess the relative influence of regional and local processes²³. We use 42,173 community plot records sampled across multiple regions and evaluate the influence of continental drivers related to biogeographic history and environmental gradients. We calculate regional and local species-pool sizes (Fig. 1) to assess the influence of large-scale factors under two spatial scales, and investigate the relationships between large-scale and local environmental factors. The analyses were carried out for whole plant communities and for a subset of plant specialists of beech forests, assuming that different species groups within the habitat species pool can respond differently to ecological drivers²⁴ and eventually also to historical drivers.

Results

Diversity patterns. We found the highest values of community species richness and species-pool sizes around mountain areas of southern and central Europe, namely the Alps, Carpathians, Dinarides and Pyrenees (Fig. 2a). Species-pool sizes calculated for large biogeographical regions (regional pools) and for 1 km² grid cells using a radius of 15 km (local pools) showed similar patterns, but also exhibited some differences. While regional pools showed a clear trend of larger sizes in the mountains from western to central Europe, local pools were highlighted in the Alps and to a lower extent in the Carpathians and the north-western areas of central Europe. Patterns of community species richness and species-pool sizes were similar when using subsets restricted to plant specialists of beech forests (Fig. 2b), except for larger regional pools detected for the specialists in the Apennines and large local pools in north-western Germany and the Netherlands. Species-pool sizes showed positive correlations with community species richness at regional (Pearson's $r = 0.21$) and local ($r = 0.27$) scales, but negative (regional scale; $r = -0.06$) and weak (local scale; $r = 0.01$) correlations with community completeness. For plant specialists, the correlations were stronger for regional pools than for local pools ($r = 0.38$ and 0.21 for community richness; $r = 0.10$ and 0.09 for completeness).

Drivers of community diversity. Using Boosted Regression Trees to model community species richness, we found that large-scale factors explained 24% and 30% of variance in two data sets defined for comparing regional and local pools (Table 1). Regional soil pH, summer precipitation and proximity to refugia were the most important predictors and showed a positive influence (Supplementary Fig. 1-4), while sampled area (plot size) had next to the lowest relative influence. Including species-pool sizes in the models reduced the importance of both historical and environmental factors, while the explained variance increased very slightly (up to 25% and 31%). Regional pools were among the best predictors (19% of total explained variance), but local pools were not as strong (10%). We found very similar results in models for

plant specialists (Supplementary Table 1), but the relative contribution of proximity to refugia and summer precipitation increased considerably. As a result, the influence of regional soil pH was reduced; and the climatic suitability in the Last Glacial Maximum (LGM) showed a stronger positive influence. Computing the same models for community completeness as the response variable, we found a lower explanatory power for both all species and specialists, in most cases between 15 and 20% of total explained variance, but with similar relative contribution of the large-scale predictors (Table 1). The influence of species-pool sizes on community completeness had a negative rather than a positive effect on community species richness, reflecting the complementary influence of local processes on community diversity.

Causal relationships. Structural equation models (SEMs) revealed the magnitude of multiple relationships between community species richness, species pools, large-scale predictors and local habitat conditions estimated from ecological indicator values (Methods). We found the strongest coefficients between regional species-pool sizes and proximity to refugia (0.53) (Fig. 3b); and between local species-pool sizes and annual precipitation (0.43) (Fig. 3c). Other predictors related to both historical and environmental factors had similar effects in the two data sets used for regional and local pools. We also found strong correlations between large-scale environmental variables and local habitat conditions (Fig. 3a), especially between regional soil pH and the ecological indicator values (EIVs) for soil reaction (path coefficient = 0.33); and between summer precipitation with EIVs for moisture (0.25) and EIVs for nutrients (0.24). Soil reaction also showed a strong relationship with community species richness. The effect of species-pool sizes on community species richness was always minor (0.06), providing higher AICs (thus lower fitting) than the same models computed without species pools (Fig. 3).

We found similar coefficients when computing the SEMs with the number of beech-forest plant specialists (Supplementary Fig. 5). A relevant difference was the stronger effect of summer (0.51) and annual (-0.24) precipitation on regional pools (versus -0.02 and -0.06 in the models with all species), and the stronger effect of EIVs for nutrients (0.16 versus 0.02) and

regional pools (0.11 versus 0.01) on local community species richness. Again, relatively lower AIC values (indicating better model fit) were provided by the models without species pools. The model computed with local species pools and beech forest specialists showed a stronger effect of nutrients (0.19) on community species richness with respect to the effect shown for total richness (0.02) and a lower coefficient of topographic heterogeneity (0.05 versus 0.14).

When computing the SEMs for community completeness instead of community species richness, the coefficients between large-scale historical or environmental factors and species-pool sizes were very similar, as well as the relationships between large-scale environmental factors and local habitat conditions (Supplementary Table 2). However, the coefficients between habitat conditions and community completeness were strongly reduced, especially the effect of EIVs for light (from an average of 0.15 to 0.07), EIVs for moisture (from 0.12 to 0.04) and EIVs for nitrogen in the specialists (0.17 to 0.08). This indicates a weaker influence of local habitat conditions on community species richness when the effect of species-pool size is removed.

Discussion

Our models with large-scale factors were able to explain about one third of the variation in community species richness, which is relatively high if we consider that other factors influencing forest diversity were not considered here, such as landscape configuration or forest management^{25,26}. Overall, our results indicate a limited effect of species pools on community species richness, beyond what can be explained solely by large-scale drivers. Proximity to refugia explained much of the variation in species-pool sizes of large biogeographic regions, while precipitation was more relevant as predictor of the sizes of local species pools. Such differences are certainly related to the scales at which species-pool sizes were calculated, supporting the idea that species-pools are scale-dependent¹¹. Nevertheless, both historical and environmental factors were complementarily relevant for explaining patterns of diversity in whole communities and in plant specialists. The effect of large-scale factors was also similar for

community species richness and community completeness, further supporting that the observed relationships are independent of the variation in species pools across regions. We recognize that part of our data corresponds to indirect estimates of biodiversity (e.g. extrapolation of species pools) or environmental factors (e.g. LGM climatic suitability), limiting the interpretation of causal relationships. Nevertheless, our results with different data sets were consistent in detecting major effects of large-scale and local factors on the study system.

We found larger pool sizes and richer communities in mountain regions which are known as refugia for beech forests²⁷, suggesting the importance of historical contingencies related to long-time species persistence, or priority effects²⁸. The relevance of historical factors was stronger when only plant specialists of beech forests were considered, likely reflecting the legacies of post-glacial colonization in the distribution of these species²⁹. In the context of the species-pool hypothesis, these findings support the importance of historical abundance of habitats to shape both species pools and community diversity^{17,30}. Moreover, topographic heterogeneity was a better predictor for species richness of the whole community rather than for specialists, indicating that rough terrain increases the chance of integrating a wider spectrum of species from different habitats to a single community, in accordance with studies suggesting high species turnover in temperate mountain forests³¹. The positive influence of topographic heterogeneity on community species richness is well known at different spatial scales³², but in our system it should be also interpreted at the temporal scale, here contributing to the refugia of deciduous forests in southern European mountains³³⁻³⁵.

Our results with large-scale predictors also confirm the role of soil pH on species-pool sizes of temperate forests³⁶, and the effect of summer precipitation as a limiting ecological factor in southern Europe³⁷. These factors ultimately act at the local scale, but the majority of data currently stored in vegetation databases do not provide soil measurements, thus biodiversity models can only use regional environmental information as a surrogate of local habitat conditions³⁸. Our study supports this approach, revealing close relationships between large-

scale environmental factors and local habitat conditions expressed with species indicator values. The effects of soil reaction, nutrients and moisture on community richness also agree with regional studies investigating forest diversity with soil data³⁹, while the sensitivity of beech forest specialists to summer precipitation and related local factors has been interpreted as ecological adaptations to moist and nutrient-rich conditions^{29,35}. These are indeed two ecological pre-conditions for a region to be a Quaternary refugium⁴⁰, explaining the combined influence of historical and environmental drivers on species pools and richness of specialists.

With our results, one might think that large-scale drivers are effective enough to explain biogeographical patterns of community diversity, avoiding the empirical estimation of species-pool sizes and their related uncertainties^{11,41}. However, we were able to identify predictors for a well-known model system, and this may not be the case in other contexts. Calculations of species pools were also necessary to address community completeness, allowing us to confirm the effect of continental drivers in community diversity when controlling for species-pool sizes. Indeed, we found informative differences between the two diversity metrics; whereas community species richness was positively influenced by species-pool sizes, community completeness showed a negative relationship, indicating that the larger the species pool, the lower the degree of completeness. Despite the relatively weaker correlations between species pools and community completeness, our findings suggest the interplay of local processes in shaping community diversity, with dispersal limitation having a stronger effect on larger regions, and negative biotic interactions producing community saturation¹⁵. Although it is difficult to disentangle such processes with our data, the limited influence of local habitat conditions on community completeness might reflect a predominant effect of dispersal limitation or biotic interactions, rather than environmental filtering.

In conclusion, this study demonstrates how large-scale factors and related processes shape both species pools and community diversity, using a multi-regional framework for disentangling historical and ecological processes. This framework offers new possibilities to

investigate large-scale biodiversity patterns, potentially addressing taxonomical, functional or phylogenetic diversity⁴² under a biogeographical perspective of community assembly. We highlight that the use of large-scale predictors related to historical factors, when selected properly, may contribute to the species-pool hypothesis to a large extent, which is an assumption that had not been tested previously in empirical studies^{13,16}. We also note that large-scale environmental factors can be used as proxies for local habitat conditions, thus integrating different scales when assessing environmental filtering from species pools to local communities¹¹. Given the relatively little progress in the theoretical development of the species-pool hypothesis over the past 20 years⁷, and the on-going development of large data sets⁴³, we also advocate for the implementation of multi-regional studies to explain drivers of community diversity across spatio-temporal scales.

Methods

Study system. European beech forests are a suitable model system for testing the influence of environmental and historical drivers on community diversity. Firstly, European beech (*Fagus sylvatica*) is one of the most competitive trees in Europe, often forming monospecific, dense canopies, which makes beech-dominated understory ecologically homogeneous. There is also a good knowledge of the biogeographic history of beech forests and related species, with strong contractions during the Last Glacial Maximum (~21,000 years ago) followed by a postglacial expansion^{27,44}. These processes have influenced the geographic distribution of species associated with beech-dominated habitats through dispersal limitation and environmental filtering²⁹. European beech forests have also been profusely studied in terms of vegetation diversity⁴⁵ and major environmental drivers at regional scales³⁵. Given the wide distribution of European beech forests and the long-standing interest of field ecologists, they are probably among the best-sampled vegetation types worldwide.

Plant community data. We collected vegetation-plot data representing complete plant species composition in forest stands, comprising only those regions where *Fagus sylvatica* is native. We also included south-eastern European sites with *F. s. subsp. moesiaca* or *Fagus orientalis*, although we refer to all of them as European beech. A total of 42,173 georeferenced plots with the presence of beech were compiled from the European Vegetation Archive⁴⁶. To ensure a certain ecological homogeneity of the studied communities, we filtered the data to keep plots where beech is the dominant species, i.e. having an estimated cover > 50% and higher than any other tree species. We only used plots on the size between 100 and 400 m², which are the most frequent plot sizes used in the study system. Within this range, the correlation between plot size and community species richness was low (Pearson's $r = 0.02$), but we kept this variable in our analyses with different subsets. The plots were assigned to grid cells of 1 km² as the geographic operational units, which is the best grid resolution to fit with the geographical precision of the data. Most grid cells (67%) contained a single plot. When grid cells were sampled two (17%) or more times (16%), we randomly selected one plot to reduce the effect of pseudo-replication and spatial autocorrelation. Different random selections of grid cells did not affect the patterns observed in the data. The filtered data set contained 5,147 plots represented in 5,147 grid cells.

We stored the community data in TURBOVEG⁴⁷ version 3, which includes an automatic procedure for unifying taxonomic concepts from national species lists using the Euro+Med PlantBase (<http://www.emplantbase.org/>)⁴⁶. The conversion to standardized names was checked manually with JUICE software⁴⁸ to obtain a homogenized list, and lichens and bryophytes were removed. For each plot we calculated the total number of vascular plant species and the number of beech forest specialists, i.e. those species closely associated to European beech forests according to a previous review²⁹. We calculated community completeness as $\log(\text{plot richness}/(\text{species pool} - \text{plot richness}))$ to reflect the variation in species richness when species pools have been taken into account¹⁵. Patterns of community

richness were mapped along the distribution range of European beech using spatial interpolation by ordinary kriging. Spatial data were analysed with ArcGIS 10.4 (Redlands, CA, USA).

Large-scale predictors. We used the WorldClim bioclimatic variables at 1 km grid resolution (www.worldclim.org) to represent contemporary climatic gradients. Annual mean temperature was included as a key defining factor for the climatic niche of European beech and associated species, whose distribution range is limited in the driest (Mediterranean) and coldest (boreal and arctic) regions of Europe⁴⁹. In the study area, this variable is highly correlated with the minimum temperature of the coldest month (Pearson's $r = 0.91$) and the mean temperature of summer ($r = 0.94$). Annual precipitation was included to reflect the total amount of water per year, and summer precipitation to account for the limiting factor of water availability in the warmest and driest season. In addition, we used a variable estimating regional soil pH at a 1 km resolution provided by the ISRIC World Soil Information (<http://www.isric.org/>). This predictor is a good surrogate of dominant bedrock, clearly differentiating the regions with predominating base-rich (calcareous) and base-poor (acid) substrates.

Historical factors were characterized by proxies of glacial refugia and postglacial dispersal routes. First, we computed a climatic suitability model for the European beech in the Last Glacial Maximum (LGM), when the species found refugia in a mosaic of stands scattered in multiple regions²⁷ (Supplementary Fig. 6). The model was computed with six non-correlated bioclimatic variables using Maxent⁵⁰ with default parameters and a random selection of background data in all of Europe, and then projected onto the estimated climatic conditions of the LGM under CCSM and MIROC scenarios extracted from www.worldclim.org at a c. 4.2 km² resolution. Since there is no evidence of one of these scenarios being better than the other, the projections were averaged to a single model. The resulting layer was able to predict the occurrence of the fossil records reported for European beech in the LGM (Supplementary Fig. 6). However, several authors^{27,51} suggest that the European beech also found small refugia in

local sites with suitable climatic conditions not detected by the spatially coarse palaeoclimatic models. Therefore, we calculated (i) the proximity to the closest refugia using the mean Euclidean distance from each plot to the refugium localities reported by palaeobotanists; and (ii) topographical heterogeneity as a factor that favours habitat refugia under past or contemporary changes (Supplementary Fig. 7). The latter variable was firstly calculated at a resolution of 250 m using Riley's terrain ruggedness index in the Geomorphometric & Gradient Metrics toolbox, (<http://evansmurphy.wix.com/evansspatial>). All layers were finally resampled to a 1 km² grid size when necessary.

Estimation of species-pool sizes. Estimates of species-pool sizes were based on the initial data set of 42,173 plots, thus using all the species that co-occur with beech without any restriction of tree cover. Such estimates are largely independent of the number of plots we used to define community species richness ($n = 5,147$), and they represent the *habitat species pool*, i.e. the list of all species that are known to occur in the focal habitat¹³. We calculated regional species pools in large discrete regions defined by known biogeographic units of European beech-forests as a result of historical processes and long-distance dispersal²⁹ (Supplementary Fig. 8). These regions largely agree with traditional biogeographical units defined for European flora, providing a set of geographic areas with a common history and similar bioclimatic conditions. We estimated species-pool sizes using sample-based rarefaction curves⁵² for a unique sample size of 300 plots, a number that was found to properly reflect the upper bound of species richness in the regions with the highest sampling effort (with sample sizes between 400 and 1,500 plots). For the regions represented by 100 to 299 plots, the pools were estimated using non-parametric extrapolations to a total of 300 samples in EstimateS⁵². Regions with less than 100 plots were not used because extrapolations to more than three times the sample size could be unrealistic⁵². A total of 40 regions were used for calculating regional pools.

We also calculated local species pools for focal grid cells of 1 km² by counting all the species occurring in the surrounding cells, thus reflecting relatively smaller units mainly

influenced by metacommunity dynamics and related processes. This procedure is analogous to the principle of assemblage dispersion fields²² to create a grid-specific species-pool map, in this case with a minimum of one species (beech) overlapping. We chose a radius of 15 km to reflect maximum dispersal events between 1 km² grid cells within relatively short- or middle-time periods. The resulting patterns were very similar when using radiuses between 5 and 30 km. As done before, local species pools were based on 300 plots when they were available in the pre-defined radius, and extrapolated to this number for sample sizes between 100 and 299. Since many grid cells did not reach the minimum sample size for extrapolation, local species pools were finally calculated for 2,373 grid cells.

Local habitat conditions. We used Ellenberg Indicator Values (EIVs) as a surrogate of environmental conditions at the community level. EIVs are expert-based ordinal estimators (in most cases ranging from 1 to 9) of species' ecological preferences in central Europe⁵³. EIVs are based on field observations in plant communities, reflecting species realized niches along environmental gradients. The mean EIV calculated for a community at a given site reflects habitat conditions that implicitly consider biotic interactions. We assume that mean EIVs are not fully independent of community species diversity, causing risk of overestimating their effects on species composition, although to a lesser degree on species richness^{54,55}. Here, our main aim is to address correlations between factors operating at large scales and local habitat conditions estimated by EIVs (see statistical analyses). EIVs were obtained from the original Ellenberg values⁵³ and missing species from southern Europe were added from other sources using EIVs expressed on compatible scales^{56,57}. We used five EIVs describing local habitat conditions: light (L), temperature (T), moisture (M), soil reaction (R) and nutrients (N). The proportion of species with assigned indicator values per plot (mean percentage \pm SD; $n = 9314$) was 70.1 ± 11.5 for L, 39.3 ± 13.1 for T, 64.2 ± 13.2 for M, 57.3 ± 12.2 for R and 68.5 ± 12.2 for N.

Statistical analyses. We carried out two sets of complementary analyses. First, we used Boosted Regression Trees (BRTs) to test the relative influence of large-scale factors alone, or in

combination with regional and local species pools, on community species richness and community completeness. We decided to use BRTs rather than conventional regression methods because of their ability to handle complex non-linear relationships in heterogeneous datasets. The optimal number of BRT trees was selected using a 10-fold cross-validation procedure. The trees were gradually added to the model in groups of 100 and with a small learning rate (0.001) that represents contribution of each tree to the growing model⁵⁸. We repeated the same analyses for whole communities and for plant specialists of beech forests, and interpreted the effect plots for each factor to assess whether they reflect positive, negative or more complex (e.g. unimodal) responses.

Secondly, we used Structural Equation models (SEMs)⁵⁹ to quantify the causal relationships between large-scale factors, species pools and community diversity in combination with habitat conditions estimated from EIVs. We fitted the SEMs using different measures of community species richness (all species versus beech-forest specialists) and species pool size (regional versus local species pools). We used the Akaike Information Criterion (AIC) to compare each community model to an alternative model without species pool. All SEMs were fitted using a piecewise approach. This procedure overcomes the limitations of traditional variance-covariance SEMs, including distributional assumptions and data independence assumptions⁶⁰. Models with community species richness or community completeness as a response variable were assumed to follow the Poisson distribution, whereas all the other models were assumed to follow the Gaussian distribution. Standardized coefficients for SEM paths were used to compare the relative importance of predictor variables. We used a bootstrapping procedure to provide an estimate of the uncertainty of the estimated path coefficients, fitting the SEMs to 1,000 datasets resampled with replacement. We calculated 95% confidence intervals for the standardized path coefficients using the bootstrapped estimates. All correlations are presented in Supplementary Fig. 9.

We tested for potential confounding effects of spatial autocorrelation (SAC) in our SEMs by computing spatial correlograms. Most models did not display a significant amount of SAC (Supplementary Fig. 10). We re-fitted those models displaying SAC using simultaneous autoregressive models (SAR)⁶¹. The SAR error model supplements Ordinary Least Squares regression with a spatial weight matrix (W) that accounts for SAC in model residuals. Spatial weight matrices were defined by successively fitting a SAR model with k -nearest neighbours of each site starting with $k=2$, gradually increasing the number of neighbours up to 7. The SAR models removed most residual SAC and produced standardized coefficients that were similar to non-spatial models (Supplementary Fig. 10). Thus, throughout the manuscript we focus on the non-spatial path coefficients from the SEMs.

All statistical analyses were performed using the R statistical programming language (v. 3.2.1, <http://www.R-project.org>). Standardized path coefficients and AIC values were calculated using the *piecewiseSEM* package⁶⁰ and BRTs models were fitted using the *dismo* package⁵⁸. Spatial autoregressive models were fitted using the *spdep* package⁶².

Data availability

The data is available upon request to the European Vegetation Archive (EVA). See the EVA Rules for details at <http://euroveg.org/download/eva-rules.pdf>

The R code used in this study is available here: <https://github.com/marcog77/fagus>

References

1. Ricklefs, R. E. Disintegration of the ecological community. *Am. Nat.* **172**, 741–50 (2008).
2. Pärtel, M., Bennett, J. A. & Zobel, M. Macroecology of biodiversity: disentangling local and regional effects. *New Phytol.* **211**, 404–410 (2016).
3. Kier, G. et al. Global patterns of plant diversity and floristic knowledge. *J. Biogeogr.* **32**, 1107–1116 (2005).
4. Šímová, I. et al. Global species-energy relationship in forest plots: Role of abundance,

- temperature and species climatic tolerances. *Glob. Ecol. Biogeogr.* **20**, 842–856 (2011).
5. Götzenberger, L. et al. Ecological assembly rules in plant communities: approaches, patterns and prospects. *Biol. Rev. Camb. Philos. Soc.* **87**, 111–27 (2012).
 6. Reich, P. B., Frelich, L. E., Voldseth, R. A., Bakken, P. & Adair, E. C. Understorey diversity in southern boreal forests is regulated by productivity and its indirect impacts on resource availability and heterogeneity. *J. Ecol.* **100**, 539–545 (2012).
 7. Lessard, J.-P., Belmaker, J., Myers, J., Chase, J. M. & Rahbek, C. Inferring local ecological processes amid species pool influences. *Trends Ecol. Evol.* **27**, 600–607 (2012).
 8. Diamond, J. M. in *Ecology and Evolution of Communities* (eds. Cody, M. L. C. & Diamond, J. M.) 342–444 (Harvard University Press, 1975).
 9. Eriksson, O. The species-pool hypothesis and plant community diversity. *Oikos* **68**, 71–374 (1993).
 10. Zobel, M. The relative role of species pools in determining plant species richness. An alternative explanation of species coexistence? *Trends Ecol. Evol.* **2**, 266–269 (1997).
 11. Cornell, H. V & Harrison, S. P. What are species pools and when are they important? *Annu. Rev. Ecol. Evol. Syst.* **45**, 45–67 (2014).
 12. Herben, T. Correlation between richness per unit area and the species pool cannot be used to demonstrate the species pool effect. *J. Veg. Sci.* **11**, 123–126 (2000).
 13. Zobel, M. The species pool concept as a framework for studying patterns of plant diversity. *J. Veg. Sci.* **27**, 8–18 (2016).
 14. Mittelbach, G. G. & Schemske, D. W. Ecological and evolutionary perspectives on community assembly. *Trends Ecol. Evol.* **30**, 241–247 (2015).
 15. Pärtel, M., Szava-Kovats, R. & Zobel, M. Community completeness: linking local and dark diversity within the species pool concept. *Folia Geobot.* **48**, 307–317 (2013).
 16. Ricklefs, R. E. Community diversity: relative roles of local and regional processes. *Science* **235**, 167–171 (1987).
 17. Pither, J. & Aarssen, L. W. The evolutionary species pool hypothesis and patterns of freshwater diatom diversity along a pH gradient. *J. Biogeogr.* **32**, 503–513 (2005).
 18. Pärtel, M. Local plant diversity patterns and evolutionary history at the regional scale. *Ecology* **83**, 2361–2366 (2002).
 19. Zobel, M. et al. The formation of species pools: historical habitat abundance affects current local diversity. *Glob. Ecol. Biogeogr.* **20**, 251–259 (2011).
 20. Scheiner, S. M. et al. The underpinnings of the relationship of species richness with space and time. *Ecol. Monogr.* **81**, 195–213 (2011).
 21. Brown, J. H. & Nicoletto, P. F. Spatial scaling of species composition: body masses of north american land mammals. *Am. Nat.* **138**, 1478–1512 (1991).

22. Graves, G. R. & Rahbek, C. Source pool geometry and the assembly of continental avifaunas. *Proc. Natl. Acad. Sci. U. S. A.* **102**, 7871–7876 (2005).
23. Fraser, L. H. et al. Response to comment on ‘worldwide evidence of a unimodal relationship between productivity and plant species richness’. *Science* **350**, 1177–1178 (2016).
24. Jiménez-Alfaro, B., Marcenò, C., Bueno, Á., Gavilán, R. & Obeso, J. R. Biogeographic deconstruction of alpine plant communities along altitudinal and topographic gradients. *J. Veg. Sci.* **25**, 160–171 (2014).
25. Horvat, V., Biurrun, I. & García-Mijangos, I. Herb layer in silver fir – beech forests in the western Pyrenees: Does management affect species diversity? *For. Ecol. Manage.* **385**, 87–96 (2017).
26. Paillet, Y. et al. Biodiversity differences between managed and unmanaged forests: Meta-analysis of species richness in Europe. *Conserv. Biol.* **24**, 101–112 (2010).
27. Magri, D. Patterns of post-glacial spread and the extent of glacial refugia of European beech (*Fagus sylvatica*). *J. Biogeogr.* **35**, 450–463 (2008).
28. Fukami, T. Historical contingency in community assembly: integrating niches, species pools, and priority effects. *Annu. Rev. Ecol. Evol. Syst.* **46**, 1–23 (2015).
29. Willner, W., Di Pietro, R. & Bergmeier, E. Phytogeographical evidence for post-glacial dispersal limitation of European beech forest species. *Ecography* **32**, 1011–1018 (2009).
30. Zobel, M. et al. The formation of species pools: historical habitat abundance affects current local diversity. *Glob. Ecol. Biogeogr.* **20**, 251–259 (2011).
31. Sabatini, F. M., Jiménez-Alfaro, B., Burrascano, S., Lora, A. & Chytrý, M. Beta-diversity of Central European forests decreases along an elevational gradient due to the variation in local community assembly processes. *Ecography* (2017). doi:10.1111/ecog.02809
32. Stein, A., Gerstner, K. & Kreft, H. Environmental heterogeneity as a universal driver of species richness across taxa, biomes and spatial scales. *Ecol. Lett.* **17**, 866–80 (2014).
33. Gavin, D. G. et al. Climate refugia: joint inference from fossil records, species distribution models and phylogeography. *New Phytol.* **204**, 37–54 (2014).
34. Wohlgemuth, T., Moser, B., Brändli, U.-B., Kull, P. & Schütz, M. Diversity of forest plant species at the community and landscape scales in Switzerland. *Plant Biosyst. - An Int. J. Deal. with all Asp. Plant Biol.* **142**, 604–613 (2008).
35. Ujházyová, M. et al. Diversity of beech forest vegetation in the Eastern Alps, Bohemian Massif and the Western Carpathians. *Preslia* **88**, 435–457 (2016).
36. Ewald, J. The calcareous riddle: Why are there so many calciphilous species in the Central European flora? *Folia Geobot.* **38**, 357–366 (2003).
37. Biurrun, I., Campos, J. A., García-Mijangos, I., Herrera, M. & Loidi, J. Floodplain forests of the Iberian Peninsula: Vegetation classification and climatic features. *Appl. Veg. Sci.* **19**, 336–354 (2016).

38. Franklin, J., Serra-Diaz, J. M., Syphard, A. D. & Regan, H. M. Big data for forecasting the impacts of global change on plant communities. *Glob. Ecol. Biogeogr.* **26**, 6–17 (2017).
39. Härdtle, W., von Oheimb, G. & Westphal, C. The effects of light and soil conditions on the species richness of the ground vegetation of deciduous forests in northern Germany (Schleswig-Holstein). *For. Ecol. Manage.* **182**, 327–338 (2003).
40. Tzedakis, P. C., Lawson, I. T., Frogley, M. R., Hewitt, G. M. & Preece, R. C. Buffered Tree Population Changes in a Quaternary Refugium: Evolutionary Implications. *Science* **297**, 2044–2047 (2002).
41. Karger, D. N. et al. Delineating probabilistic species pools in ecology and biogeography. *Glob. Ecol. Biogeogr.* **25**, 489–501 (2016).
42. Gerhold, P., Cahill, J. F., Winter, M., Bartish, I. V. & Prinzing, A. Phylogenetic patterns are not proxies of community assembly mechanisms (they are far better). *Funct. Ecol.* **29**, 600–614 (2015).
43. Wisser, S. K. Achievements and challenges in the integration, reuse and synthesis of vegetation plot data. *J. Veg. Sci.* **27**, 868–879 (2016).
44. Lehsten, D. et al. Modelling the Holocene migrational dynamics of *Fagus sylvatica* L. and *Picea abies* (L.) H. Karst. *Glob. Ecol. Biogeogr.* **23**, 658–668 (2014).
45. Willner, W. et al. Classification of European beech forests: a Gordian Knot? *Appl. Veg. Sci.* **20**, 494–512 (2017).
46. Chytrý, M. et al. European Vegetation Archive (EVA): An integrated database of European vegetation plots. *Appl. Veg. Sci.* **19**, 173–180 (2016).
47. Hennekens, S. M. & Schaminée, J. H. J. TURBOVEG, a comprehensive data base management system for vegetation data. *J. Veg. Sci.* **12**, 589–591 (2001).
48. Tichý, L. JUICE, software for vegetation classification. *J. Veg. Sci.* **13**, 451–453 (2002).
49. Houston Durrant, T., de Rigo, D. & Caudullo, G. in *European Atlas of Forest Tree Species* (eds. San-Miguel-Ayanz, J., de Rigo, D., Caudullo, G., Houston Durrant, T. & Mauri, A.) 94–95 (Publ. Off. EU, Luxembourg, pp. e012b90, 2016).
50. Elith, J. et al. A statistical explanation of MaxEnt for ecologists. *Divers. Distrib.* **17**, 43–57 (2011).
51. Magyari, E. K. et al. Late Pleniglacial vegetation in eastern-central Europe: Are there modern analogues in Siberia? *Quat. Sci. Rev.* **95**, 60–79 (2014).
52. Colwell, R., Mao, C. & Chang, J. Interpolating, extrapolating, and comparing incidence-based species accumulation curves. *Ecology* **85**, 2717–2727 (2004).
53. Ellenberg, H. et al. Zeigerwerte von Pflanzen in Mitteleuropa. *Scr. Geobot.* **18**, 1–248 (1991).
54. Zelený, D. & Schaffers, A. P. Too good to be true: pitfalls of using mean Ellenberg indicator values in vegetation analyses. *J. Veg. Sci.* **23**, 419–431 (2012).

55. Wildi, O. Why mean indicator values are not biased. *J. Veg. Sci.* **27**, 40–49 (2016).
56. Mayor, M. *Ecología de la Flora y Vegetación del Principado de Asturias* (Real Instituto de Estudios Asturianos, 1999).
57. Pignatti, S. Valori di bioindicazione delle piante vascolari della flora d'Italia. *Braun-Blanquetia* **39**, 1–97 (2005).
58. Elith, J., Leathwick, J. R. & Hastie, T. A working guide to boosted regression trees. *J. Anim. Ecol.* **77**, 802–813 (2008).
59. Grace, J. B. *Structural Equation Modeling and Natural Systems* (Cambridge University Press, 2006).
60. Lefcheck, J. S. piecewiseSEM: piecewise structural equation modelling in r for ecology, evolution, and systematics. *Methods Ecol. Evol.* **7**, 573–579 (2016).
61. Kissling, W. D. & Carl, G. Spatial autocorrelation and the selection of simultaneous autoregressive models. *Glob. Ecol. Biogeogr.* **17**, 59–71 (2008).
62. Bivand, R., Hauke, J. & Kossowski, T. Computing the Jacobian in Gaussian spatial autoregressive models: an illustrated comparison of available methods. *Geogr. Anal.* **45**, 150–179 (2013).

Acknowledgements

We are grateful to data contributors of the European Vegetation Archive, and also Stephan Hennekens and Ilona Knollová for their support with data management. BJA was supported by the project “Employment of Best Young Scientists for International Cooperation Empowerment” (CZ.1.07/2.3.00/30.0037) co-financed from European Social Fund and the state budget of the Czech Republic, and by the German Centre for Integrative Biodiversity Research funded by the German Research Foundation. MC and LT were supported by the Czech Science Foundation (Centre of Excellence PLADIAS, 14-36079G). JCS considers this work a contribution to his VILLUM Investigator project funded by VILLUM FONDEN (grant 16549).

Author contributions

BJA designed the study with support of MC and JCS. BJA prepared the data and wrote the manuscript. MG performed the statistical analyses. The other co-authors contributed with data, interpreted the results and commented on the final version. The first six authors are ordered by relative contribution, the others are ordered alphabetically.

Additional information

Supplementary information is available for this paper.

Competing interests

The authors declare no competing financial interests.

FIGURE CAPTIONS

Figure 1. A multi-regional framework for community assembly. Multiple species pools calculated at either regional or local (landscape) scales are influenced by assembly processes interacting at different spatio-temporal scales to shape local community diversity.

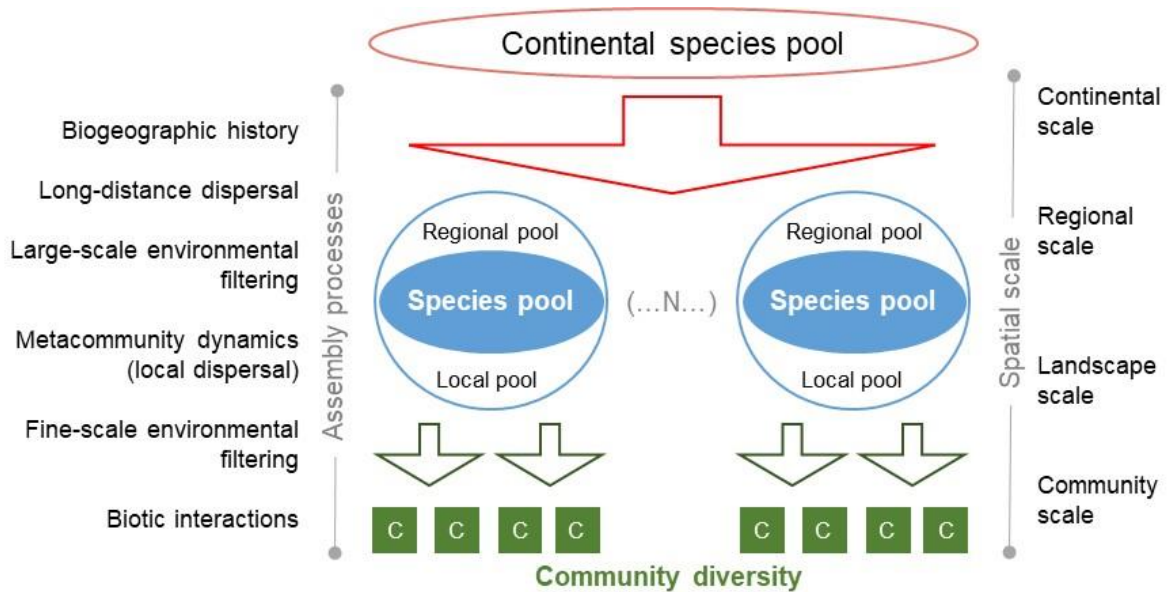


Figure 2. Number of plant species in European beech forests. **a**, species number for whole communities. **b**, species number for beech-forest plant specialists only. Community species richness was interpolated by kriging using plots where beech is the dominant species. Regional species pools were calculated for biogeographical regions; local species pools for grid cells of 1 km² and then interpolated by kriging. All maps are masked to the distribution range of beech in Europe as provided by www.euforgen.org.

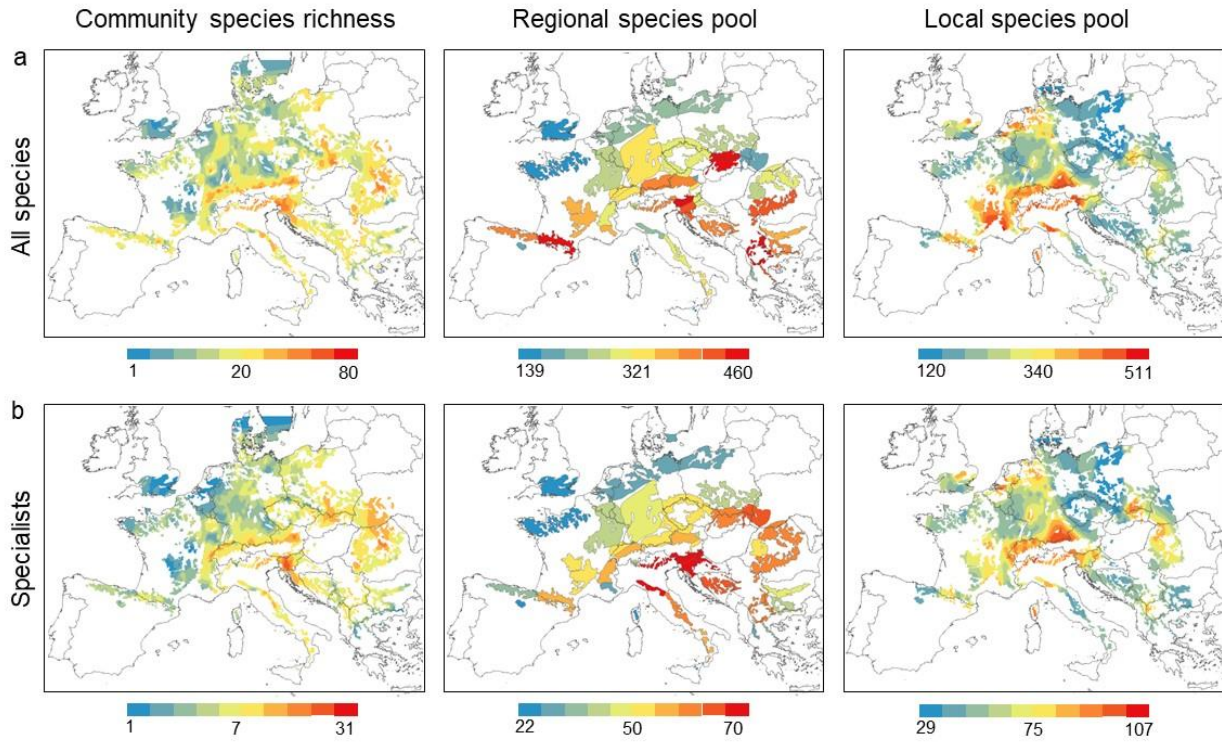


Figure 3. Path coefficients of structural equation models indicating relationships between large-scale environmental and historical factors, local habitat conditions, species pools and community species richness across European beech forests. a, general model for community species richness. **b,** model including regional pools estimated for large biogeographical regions. **c,** model including local pools estimated for grid cells of 1 km². The width of the lines represents the coefficients of relationships. Dashed lines represent negative signs. In **a**, coefficients were averaged from the analyses of two datasets providing almost identical coefficients and the same Akaike Information Criterion value. See Supplementary Fig. 6 for the plant specialists.

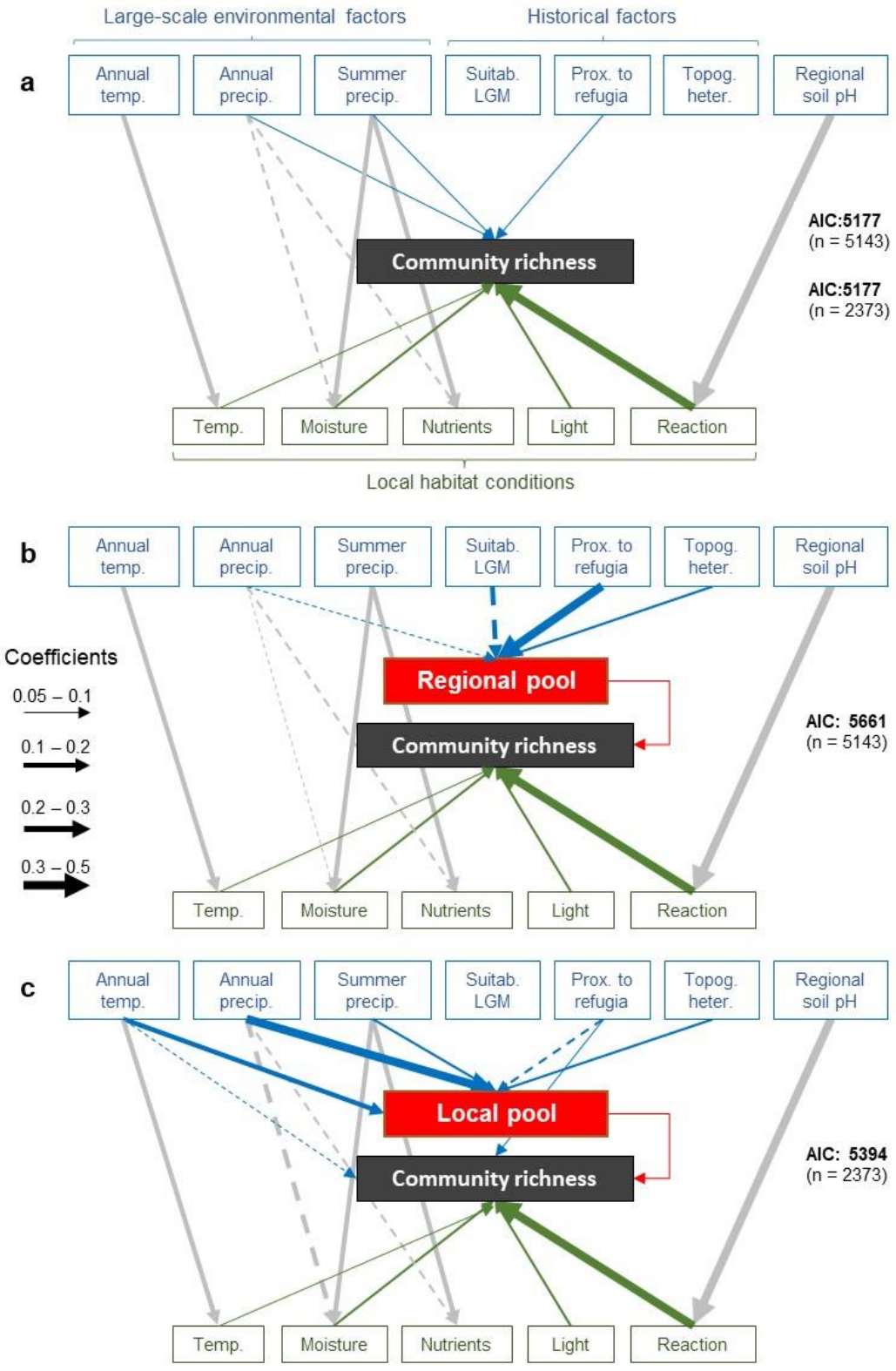


Table 1. Relative influence (in %) of environmental factors (E), historical factors (H), species-pool sizes (SP) and plot size on community species richness and community completeness across European beech forests, as estimated by Boosted Regression Trees. a, results from 5,143 plots for which species pools were estimated for large biogeographic regions. **b,** results for 2,373 plots for which species pools were estimated for 1 km². The effect of each predictor was estimated from effect plots of the models where they showed positive (+), negative (–) or unimodal/additive (^) responses (Supplementary Fig. 1-4). See Supplementary Table 1 for the plant specialists.

	Community species richness				Community completeness			
	~ E+H		~ E+H+SP		~ E+H		~ E+H+SP	
	<i>a</i>	<i>b</i>	<i>a</i>	<i>b</i>	<i>a</i>	<i>b</i>	<i>a</i>	<i>b</i>
Climate and soil (E)								
Regional soil pH	28 (+)	26 (+)	23 (^)	22 (+)	23 (+)	16 (+)	20 (+)	15 (+)
Summer precipitation	20 (+)	12 (+)	16 (+)	11 (+)	18 (^)	10 (^)	11 (^)	6 (^)
Annual precipitation	14 (+)	14 (+)	13 (+)	12 (+)	16 (+)	12 (^)	13 (+)	9 (^)
Mean annual temperature	5 (^)	13 (+)	3 (^)	15 (+)	3 (+)	17 (^)	2 (+)	12 (+)
History and refugia (H)								
Proximity to refugia	14 (+)	17 (+)	10 (^)	14 (+)	17 (–)	14 (+)	13 (–)	11(+)
Topographic heterogeneity	14 (+)	10 (+)	8 (+)	9 (+)	8 (+)	7 (^)	9 (+)	7 (+)
Suitability at the LGM	6 (^)	6 (^)	5 (^)	5 (+)	9 (^)	13 (^)	5 (^)	6 (^)
Species-pool sizes (SP)								
Regional species pool	.	.	19 (+)	.	.	.	20 (–)	.
Local species pool	.	.	.	10 (+)	.	.	.	22 (–)
Plot size								
	3 (+)	3 (+)	6 (+)	2 (+)	6 (+)	2 (^)	6 (+)	12 (^)
Total explained variance (D²)	24%	30%	25%	31%	15%	20%	19%	26%

SUPPLEMENTARY INFORMATION

History and environment shape species pools and community diversity in European beech forests

Borja Jiménez-Alfaro^{1,2,3*}, Marco Girardello^{4,5}, Milan Chytrý³, Jens-Christian Svenning^{5,6}, Wolfgang Willner^{7,8}, Jean-Claude Gégout⁹, Emiliano Agrillo¹⁰, Juan Antonio Campos¹¹, Ute Jandt^{1,2}, Zygmunt Kącki¹², Urban Šilc¹³, Michal Slezák^{14,15}, Lubomír Tichý³, Ioannis Tsiripidis¹⁶, Pavel Dan Turtureanu¹⁷, Mariana Ujházyová¹⁸ and Thomas Wohlgemuth¹⁹.

¹German Centre for Integrative Biodiversity Research (iDiv) Halle-Jena-Leipzig, Deutscher Platz 5e, 04103 Leipzig, Germany. ²Martin Luther University Halle-Wittenberg, Institute of Biology/Geobotany and Botanical Garden, Am Kirchtor 1, 06108 Halle, Germany. ³Department of Botany and Zoology, Masaryk University, 61137 Brno, Czech Republic. ⁴cE3c – Centre for Ecology, Evolution and Environmental Changes, Azorean Biodiversity Group, 9700-042 Angra do Heroísmo, Portugal. ⁵Section for Ecoinformatics and Biodiversity, Department of Bioscience, Aarhus University, 8000 Aarhus C, Denmark. ⁶Center for Biodiversity Dynamics in a Changing World (BIOCHANGE), Aarhus University, 8000 Aarhus C, Denmark. ⁷VINCA - Vienna Institute for Nature Conservation and Analyses, Wien, Austria. ⁸Department of Botany and Biodiversity Research, University of Vienna, 1030 Vienna, Austria. ⁹LERFoB, INRA, AgroParisTech, 54000 Nancy, France. ¹⁰Department of Environmental Biology, Sapienza University of Rome, 500185 Roma, Italy. ¹¹Department of Plant Biology and Ecology, University of Basque Country UPV/EHU, 48080 Bilbao, Spain. ¹²Department of Vegetation Ecology, Botanical Garden, University of Wrocław, 50335 Wrocław, Poland. ¹³ZRC SAZU, Institute of Biology, 1001 Ljubljana, Slovenia. ¹⁴Institute of Forest Ecology, Slovak Academy of Sciences, 960 53 Zvolen, Slovakia. ¹⁵Plant Science and Biodiversity Center, Institute of Botany, Slovak Academy of Sciences, 845 23 Bratislava, Slovakia. ¹⁶Department of Botany, School of Biology, Aristotle University of Thessaloniki, 54124 Thessaloniki, Greece. ¹⁷Alexandru Borza Botanical Garden, Babeş-Bolyai University, 400015 Cluj-Napoca, Romania. ¹⁸Department of Applied Ecology, Faculty of Ecology and Environmental Sciences, Technical University in Zvolen, 96053 Zvolen, Slovakia. ¹⁹WSL Swiss Federal Research Institute, 8903 Birmensdorf, Switzerland.

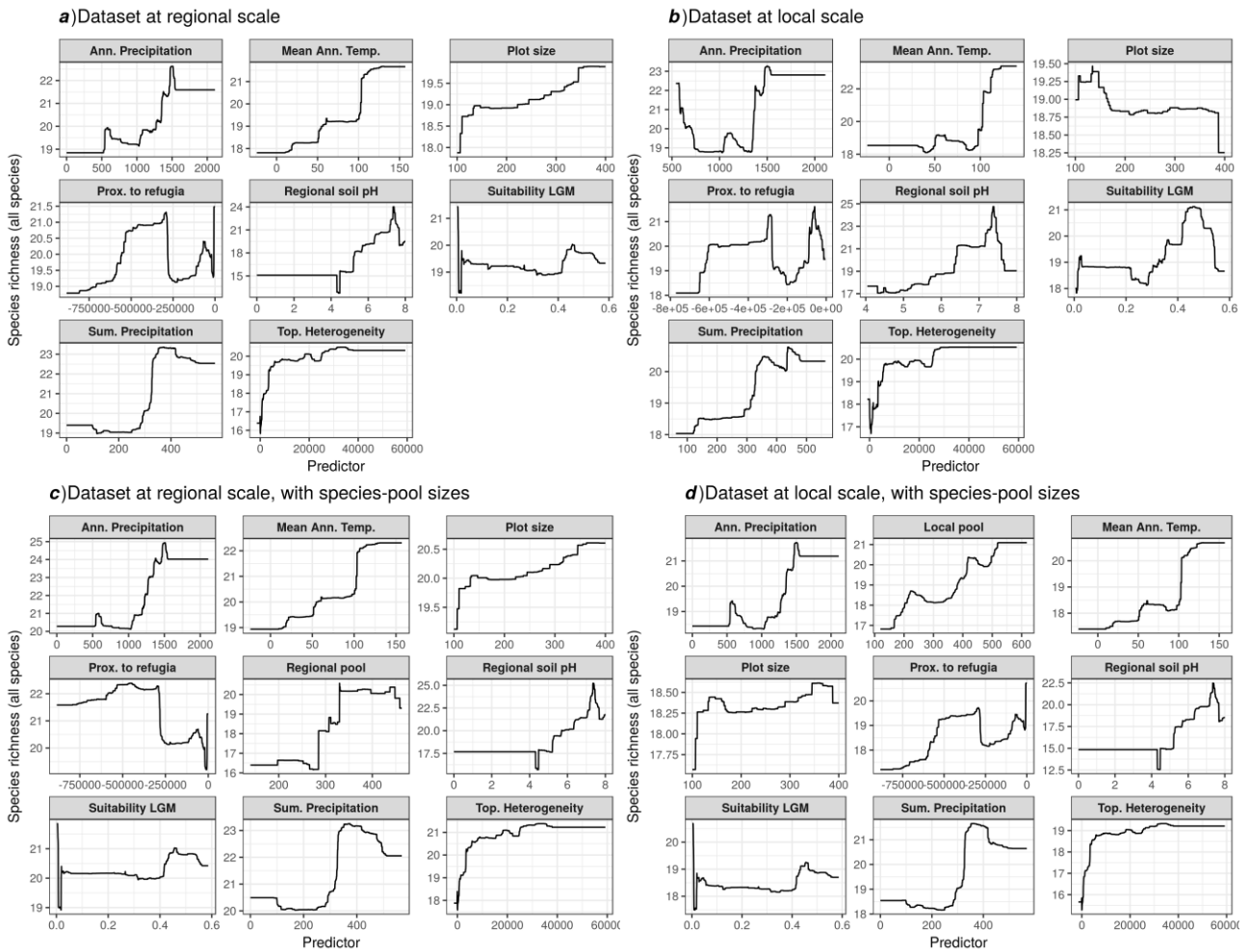
*e-mail: jimenezalfaro.borja@gmail.com

Supplementary Table 1. Relative influence (in %) of environmental factors (E), historical factors (H), species-pool sizes (SP) and plot size on community species richness and community completeness of European beech-forests plant specialists, as estimated by Boosted Regression Trees. a, Results from 5,143 plots for which species pools were estimated for large biogeographic regions. b, Results for 2,373 plots for which species pools were estimated for 1 km². The effect of each predictor was estimated from effect plots of the models where they showed positive (+), negative (–) or unimodal/additive (^) responses (Supplementary Fig. 1-4).

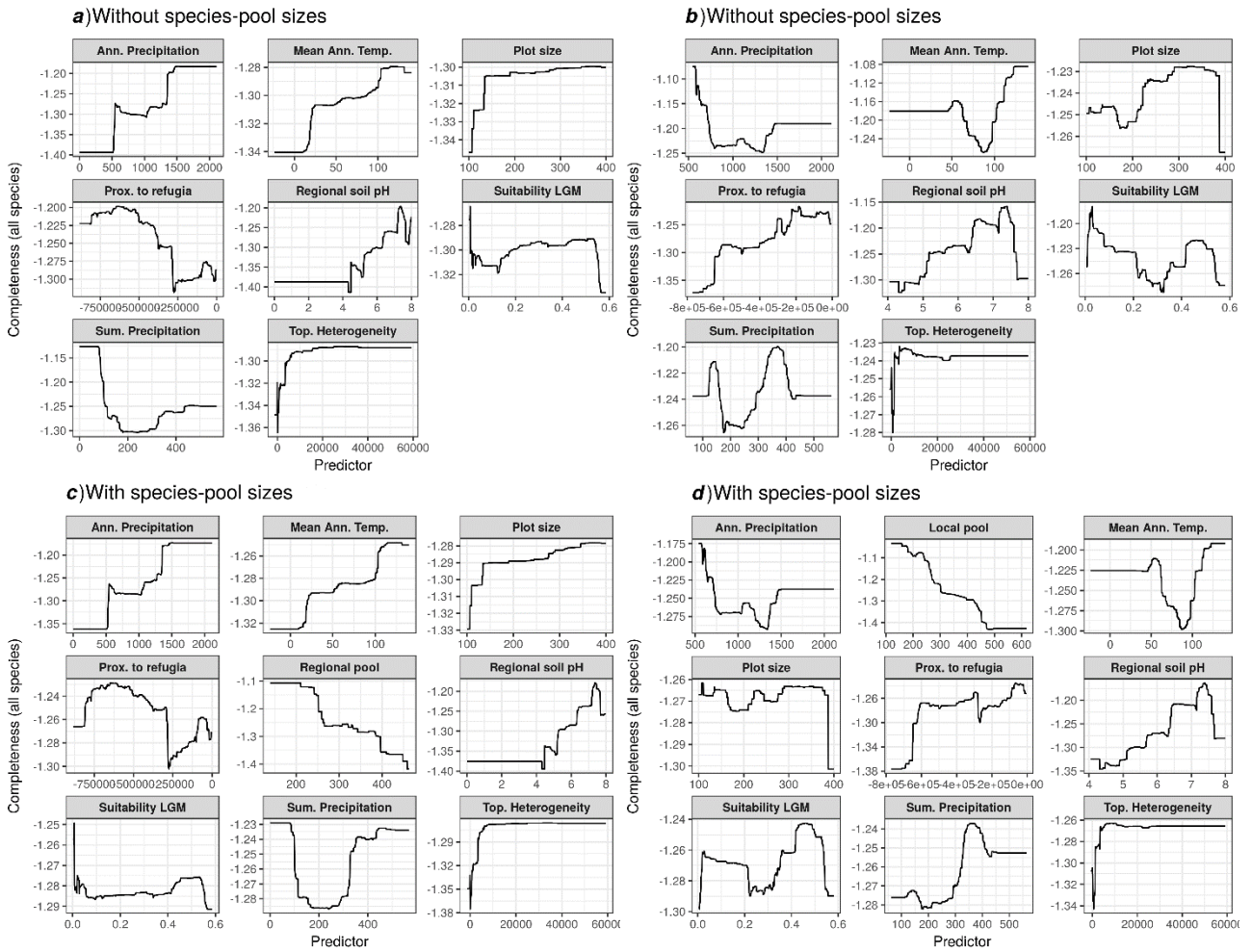
	Community species richness				Community completeness			
	~ E+H		~ E+H+SP		~ E+H		~ E+H+SP	
	a	b	a	b	a	b	a	b
Climate and soil (E)								
Summer precipitation	40 (^)	30 (^)	16 (^)	28 (^)	24 (+)	17 (+)	22 (+)	11 (+)
Regional soil pH	15 (+)	16 (+)	23 (+)	14 (+)	17 (+)	13 (+)	15 (+)	9 (+)
Mean annual temperature	6 (^)	9 (+)	3 (^)	9 (+)	10 (^)	12 (^)	9 (^)	9 (^)
Annual Precipitation	5 (+)	9 (+)	13 (^)	8 (^)	8 (^)	11 (–)	7 (+)	10 (^)
History and refugia (H)								
Proximity to refugia	16 (–)	20 (–)	10 (–)	17 (–)	18 (^)	20 (+)	16 (^)	13 (+)
Suitability at the LGM	7 (+)	10 (+)	5 (+)	9 (+)	10 (–)	14 (–)	9 (–)	10 (–)
Topographic heterogeneity	5 (+)	5 (+)	8 (+)	5 (+)	6 (^)	6 (^)	5 (+)	5 (+)
Species pool sizes (SP)								
Regional species pool	.	.	19 (+)	.	.	.	9 (^)	.
Local species pool	.	.	.	8 (^)	.	.	.	28 (–)
Plot size								
	6 (^)	2 (^)	6 (+)	2 (^)	7 (+)	7 (+)	7 (+)	5 (+)
Total explained variance (D²)	25%	29%	25%	29%	14%	20%	14%	26%

Supplementary Table 2. Path coefficients for all the relationships tested by structural equation modeling to detect causal relationships between community diversity (community species richness or completeness) and large-scale factors⁽¹⁾, habitat conditions estimated from Ellenberg indicator values⁽²⁾ and species pool sizes. Coefficients refer to eight different models computed for the whole communities (all species) and for plant specialists of European beech forests, and using data sets with species pools calculated at either regional or local scales. Colors reflect the gradient from the most positive (dark green) to the most negative (dark red) relationships. LGM: last glacial maximum.

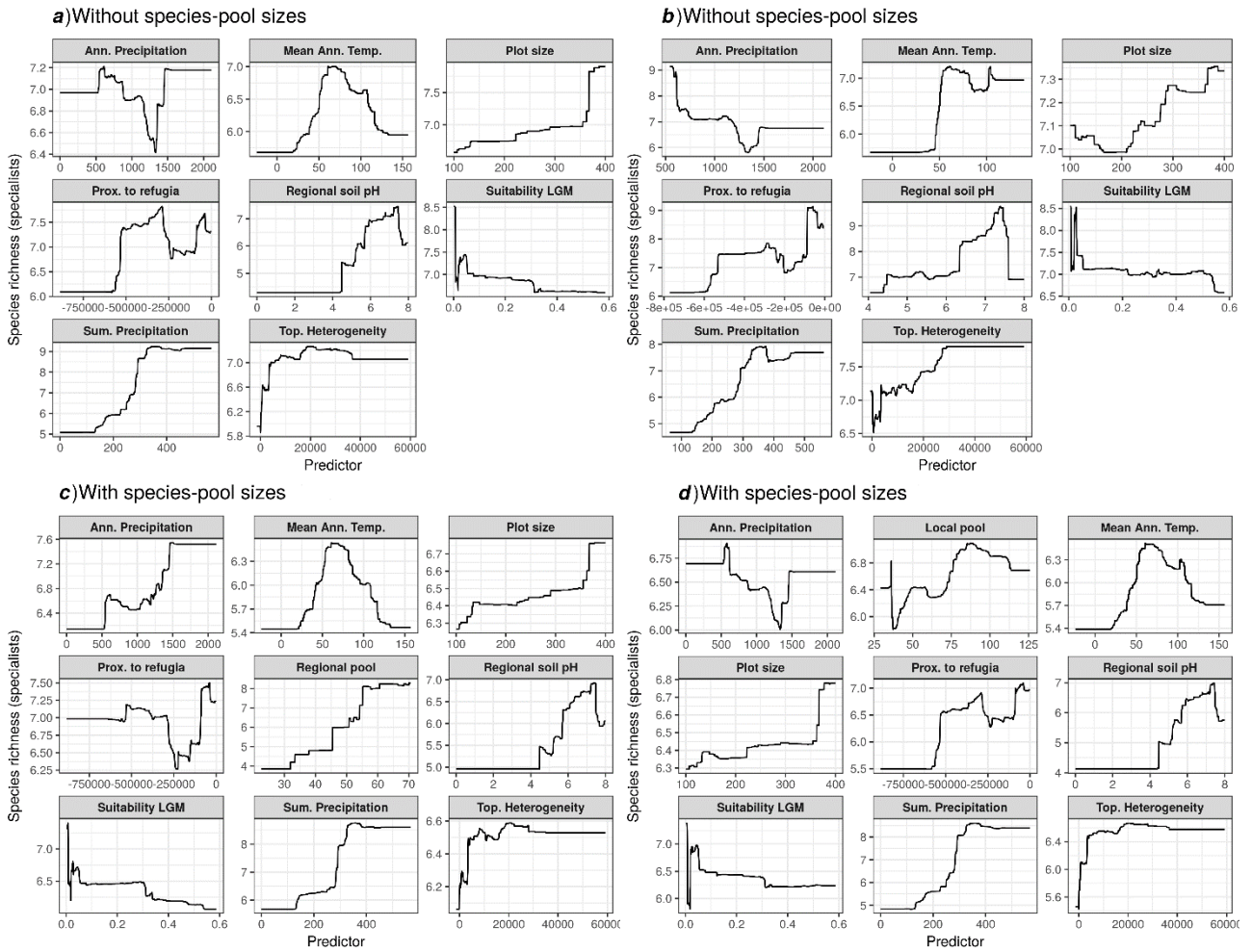
Response variable	Predictor	Community species richness				Community completeness			
		All species		Specialists		All species		Specialists	
		Regional pool	Local pool	Regional pool	Local pool	Regional pool	Local pool	Regional pool	Local pool
Species pool size	Distance to refugia ⁽¹⁾	0.53	-0.24	0.40	-0.23	0.50	-0.20	0.37	-0.20
Community diversity	Reaction ⁽²⁾	0.34	0.35	0.48	0.49	0.13	0.11	0.19	0.19
Reaction ⁽²⁾	Regional soil pH ⁽¹⁾	0.33	0.37	0.33	0.37	0.32	0.36	0.32	0.36
Moisture ⁽²⁾	Summer precipitation ⁽¹⁾	0.25	0.30	0.25	0.30	0.27	0.38	0.27	0.38
Temperature ⁽²⁾	Annual temperature ⁽¹⁾	0.25	0.21	0.25	0.21	0.30	0.34	0.30	0.34
Nitrogen ⁽²⁾	Summer precipitation ⁽¹⁾	0.24	0.22	0.24	0.22	0.26	0.24	0.26	0.24
Community diversity	Light ⁽²⁾	0.15	0.16	-0.15	-0.13	0.07	0.08	-0.08	-0.06
Species pool size	Topographic heterogeneity	0.14	0.14	0.08	0.05	0.13	0.14	0.08	0.05
Community diversity	Moisture ⁽²⁾	0.14	0.12	0.12	0.10	0.05	0.05	0.03	0.04
Community diversity	Summer precipitation ⁽¹⁾	0.06	0.04	0.07	0.06	0.01	0.02	0.02	-0.01
Community diversity	Distance to refugia ⁽¹⁾	0.05	0.08	0.05	0.13	-0.02	-0.01	0.00	0.01
Community diversity	Annual precipitation ⁽¹⁾	0.05	0.00	0.07	0.01	0.04	0.02	0.04	0.06
Community diversity	Temperature ⁽²⁾	0.05	0.06	0.02	0.03	0.03	0.02	0.02	0.02
Community diversity	Plot size	0.03	-0.01	0.05	0.00	0.02	-0.01	0.03	0.00
Community diversity	Regional soil pH ⁽¹⁾	0.02	0.03	0.02	0.04	0.01	0.00	0.00	0.01
Community diversity	Topographic heterogeneity	0.02	0.01	0.00	0.00	0.01	0.00	0.01	0.00
Community diversity	Species pool size	0.01	0.06	0.11	0.04	-0.07	0.02	-0.04	0.01
Species pool size	Annual temperature ⁽¹⁾	0.00	0.20	0.02	0.27	-0.02	0.18	0.03	0.27
Community diversity	Nutrients ⁽²⁾	-0.02	0.02	0.16	0.19	0.00	0.00	0.09	0.08
Species pool size	Summer precipitation ⁽¹⁾	-0.02	0.13	0.51	0.19	0.01	0.14	0.52	0.21
Community diversity	Annual temperature ⁽¹⁾	-0.02	-0.09	-0.05	-0.11	0.00	-0.02	-0.01	-0.03
Community diversity	Suitability at LGM ⁽¹⁾	-0.04	-0.04	-0.05	-0.06	-0.02	-0.01	-0.03	-0.02
Species pool size	Annual precipitation ⁽¹⁾	-0.06	0.43	-0.24	0.45	-0.08	0.41	-0.23	0.44
Moisture ⁽²⁾	Annual precipitation ⁽¹⁾	-0.10	-0.20	-0.10	-0.20	-0.10	-0.26	-0.10	-0.26
Nitrogen ⁽²⁾	Annual precipitation ⁽¹⁾	-0.18	-0.09	-0.18	-0.09	-0.22	-0.20	-0.22	-0.20
Species pool size	Suitability at LGM ⁽¹⁾	-0.22	0.21	-0.14	0.12	-0.20	0.22	-0.15	0.13



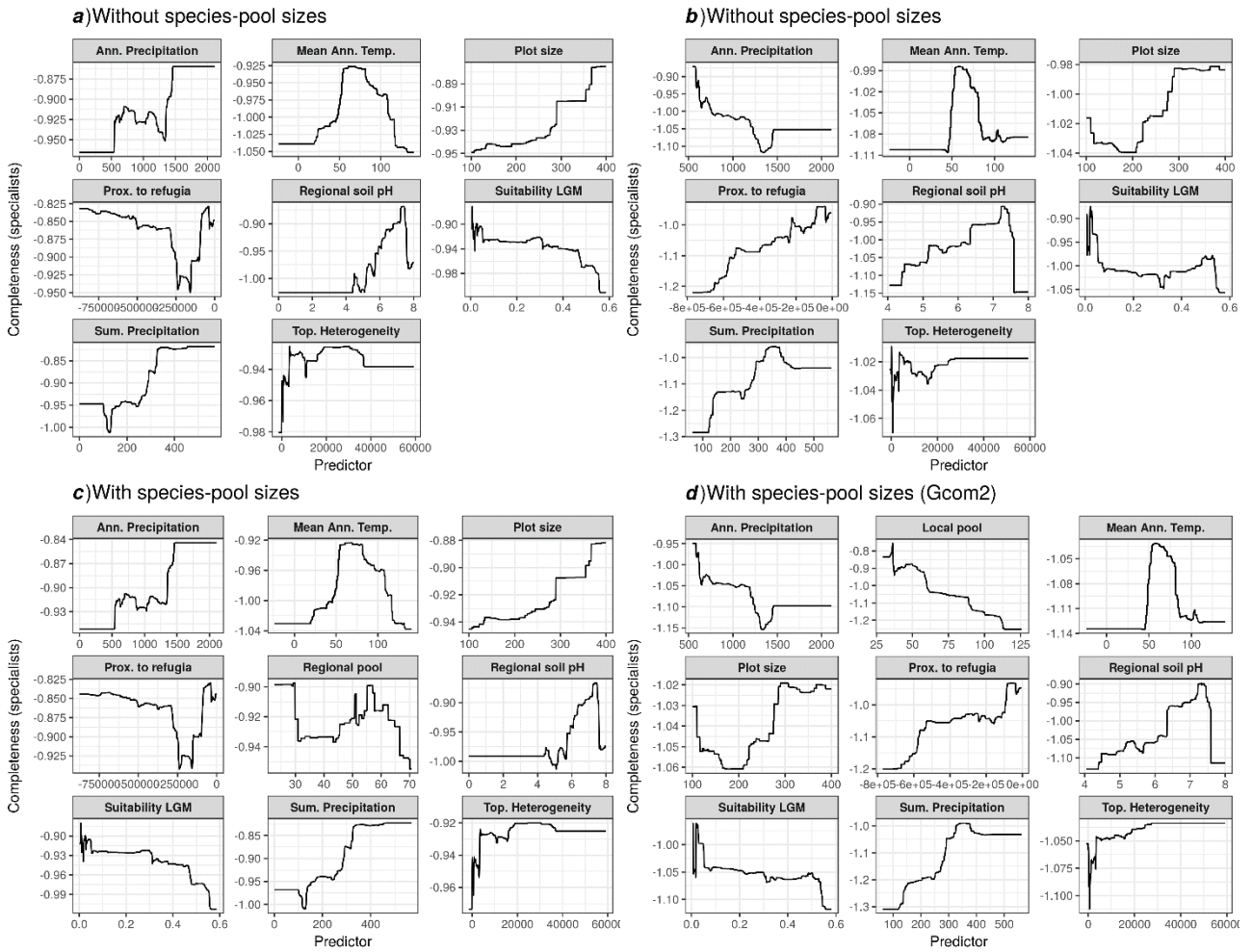
Supplementary Figure 1. Partial dependence plots showing the relationships of Boosted Regression Trees models fitted to the community species richness dataset including all the species sampled in vegetation plots across European beech forests. The graphs depict the effect of each predictor (x axis) on species richness (y axis) after accounting for the average effects of all other variables in the model. **a, c**, models computed with a dataset defined to analyze the influence of regional species pools ($N = 5,143$). **b, d**, models computed with a second database to analyze the influence of local species pools ($N = 2,373$).



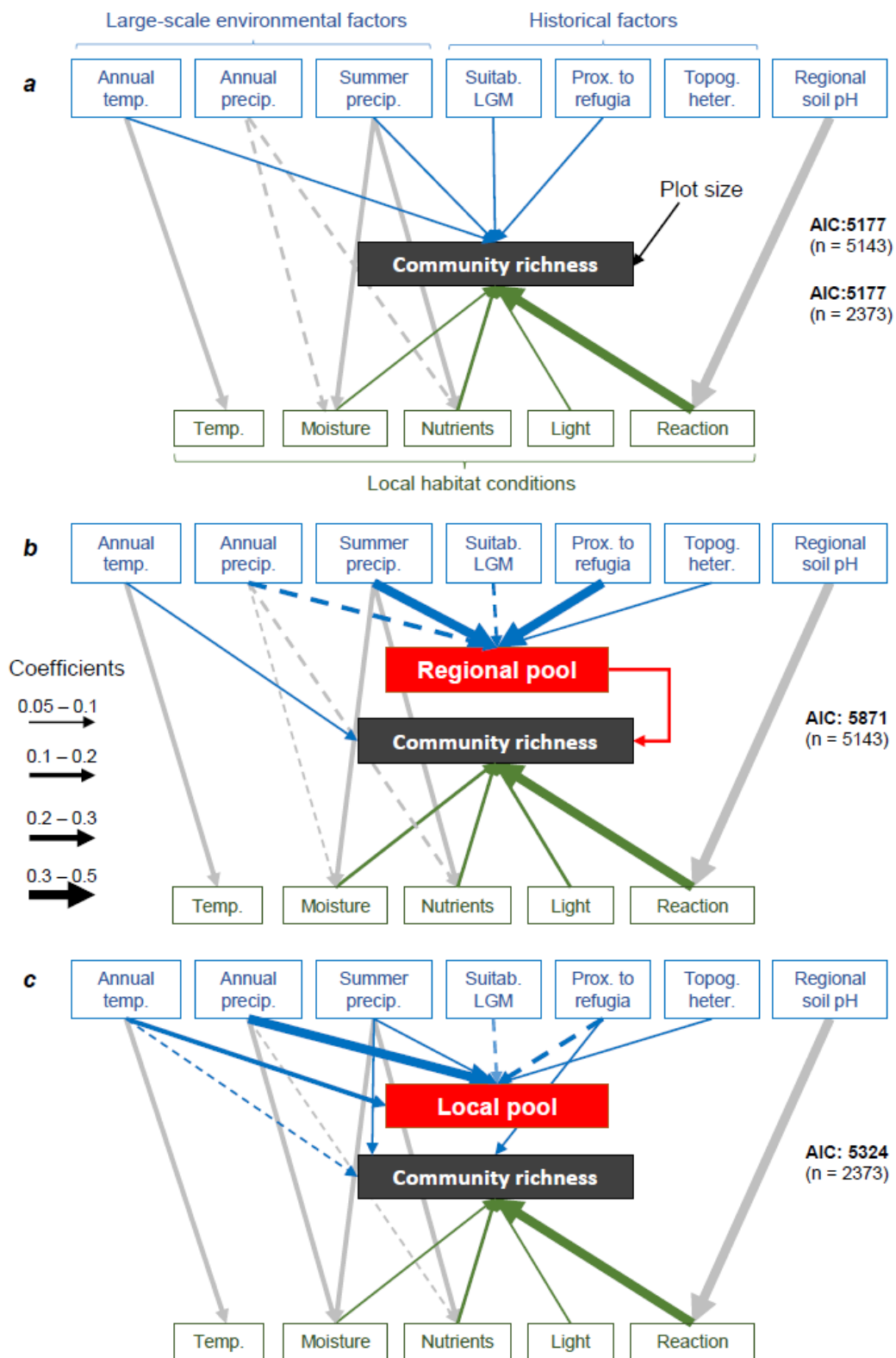
Supplementary Figure 2. Partial dependence plots showing the relationships of Boosted Regression Trees models fitted to the community completeness dataset including all the species sampled in vegetation plots across European beech forests. The graphs depict the effect of each predictor (x axis) on completeness (y axis) after accounting for the average effects of all other variables in the model. **a, c**, models computed with a dataset defined to analyze the influence of regional species pools ($N = 5,143$). **b, d**, models computed with a second database to analyze the influence of local species pools ($N = 2,373$).



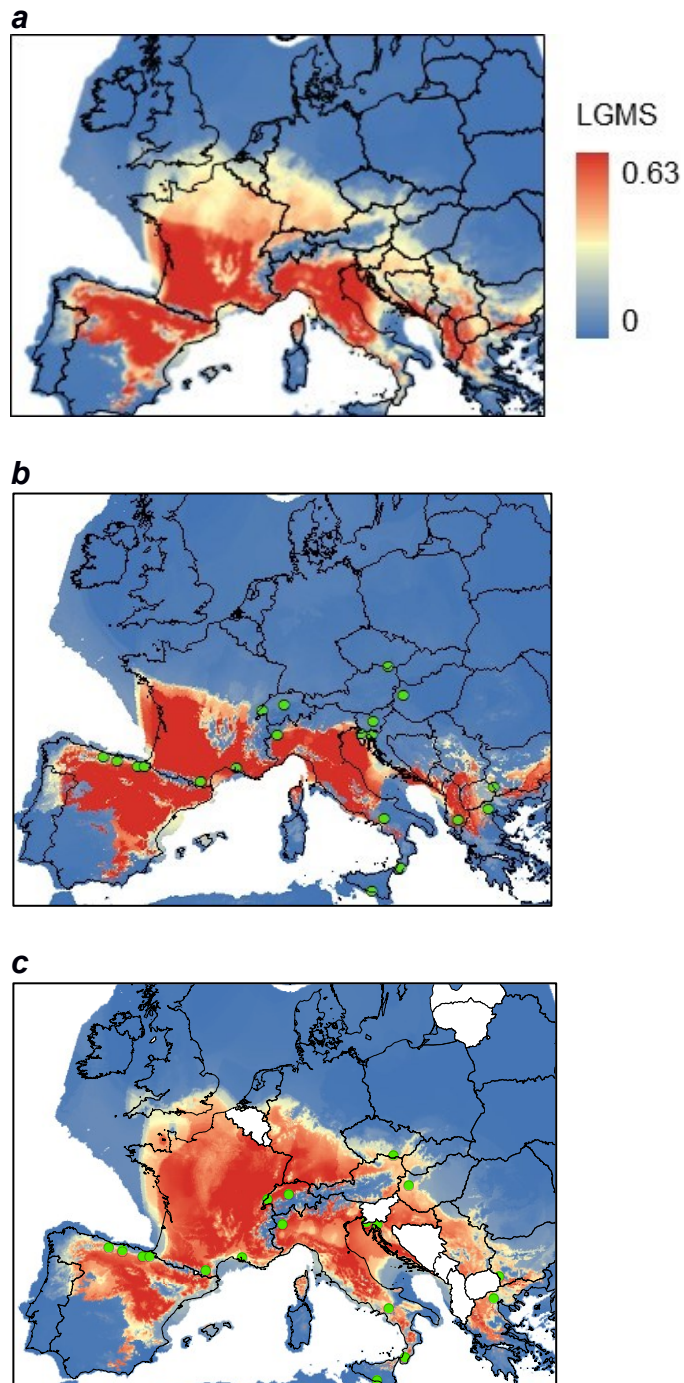
Supplementary Figure 3. Partial dependence plots showing the relationships of Boosted Regression Trees models fitted to the community species richness dataset including plant specialists sampled in vegetation plots across European beech forests. The graphs depict the effect of each predictor (x axis) on species richness (y axis) after accounting for the average effects of all other variables in the model. **a, c**, models computed with a dataset defined to analyze the influence of regional species pools ($N = 5,143$). **b, d**, models computed with a second database to analyze the influence of local species pools ($N = 2,373$).



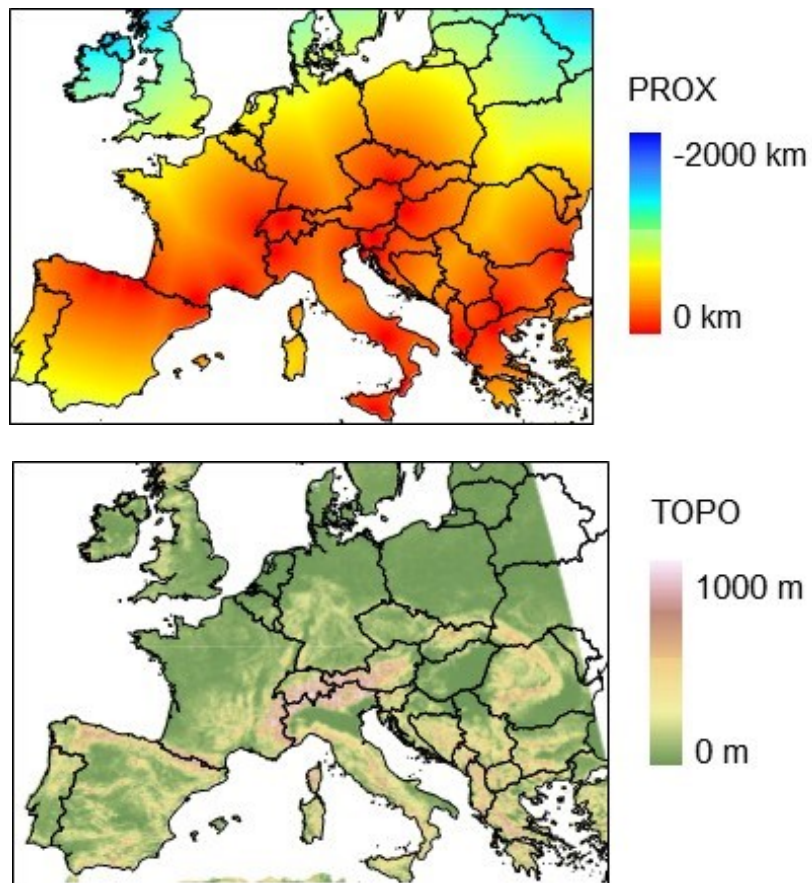
Supplementary Figure 4. Partial dependence plots showing the relationships of Boosted Regression Trees models fitted to the **community completeness** dataset including **plant specialists** sampled in vegetation plots across **European beech forests**. The graphs depict the effect of each predictor (*x* axis) on completeness (*y* axis) after accounting for the average effects of all other variables in the model. **a, c**, models computed with a dataset defined to analyze the influence of regional species pools ($N = 5,143$). **b, d**, models computed with a second database to analyze the influence of local species pools ($N = 2,373$).



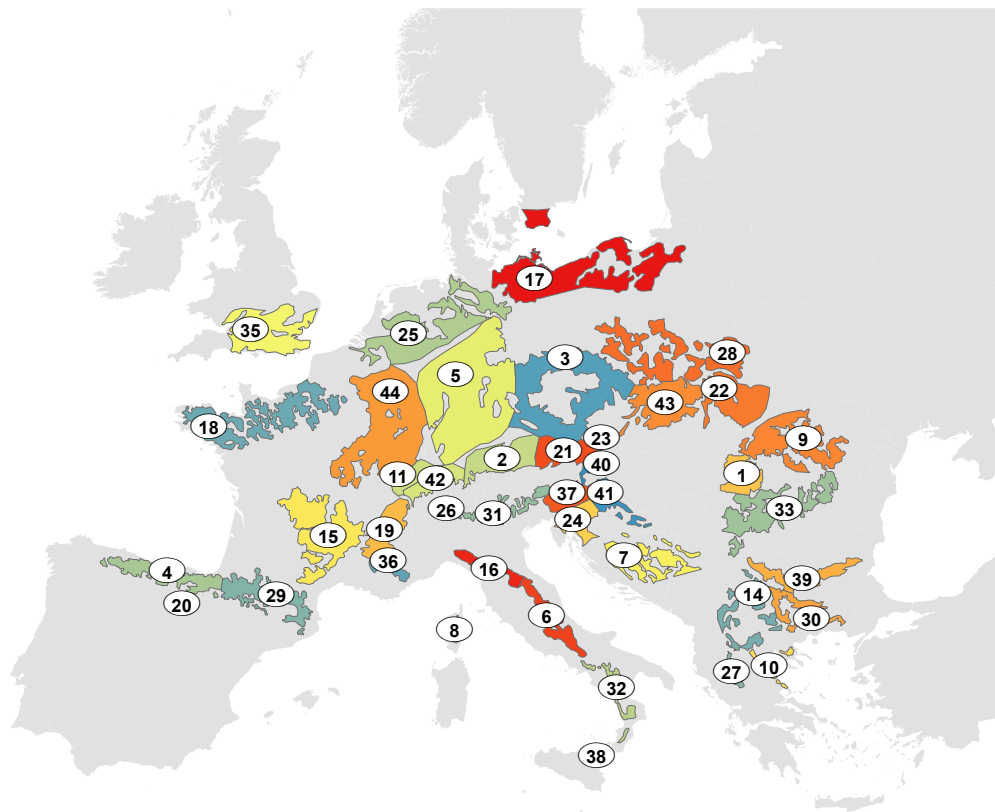
Supplementary Figure 5. Path coefficients of structural equation models for plant specialists of European beech forests, indicating relationships between large-scale environmental and historical factors, local habitat conditions, species pools and community species richness. a, model for community species richness. **b**, model including regional pools estimated for large biogeographical regions. **c**, model including local pools estimated for grid cells of 1 km². The width of the lines represent the coefficients of relationships. Dashed lines represent negative signs. In **a**, coefficients were averaged from the analyses of two datasets providing almost identical coefficients and the same Akaike Information Criterion value.



Supplementary Figure 6. Climatic suitability of European beech during the Last Glacial Maximum (LGMS). *a*, final Model averaged from projections modeled with Maxent for the MIROC and CCSM palaeoclimatic scenarios. *b*, MIROC model. *c*, CCSM model. Green dots show the distribution of fossil records according to Magri, D. Patterns of post-glacial spread and the extent of glacial refugia of European beech (*Fagus sylvatica*). *J. Biogeogr.* **35**, 450–463 (2008).



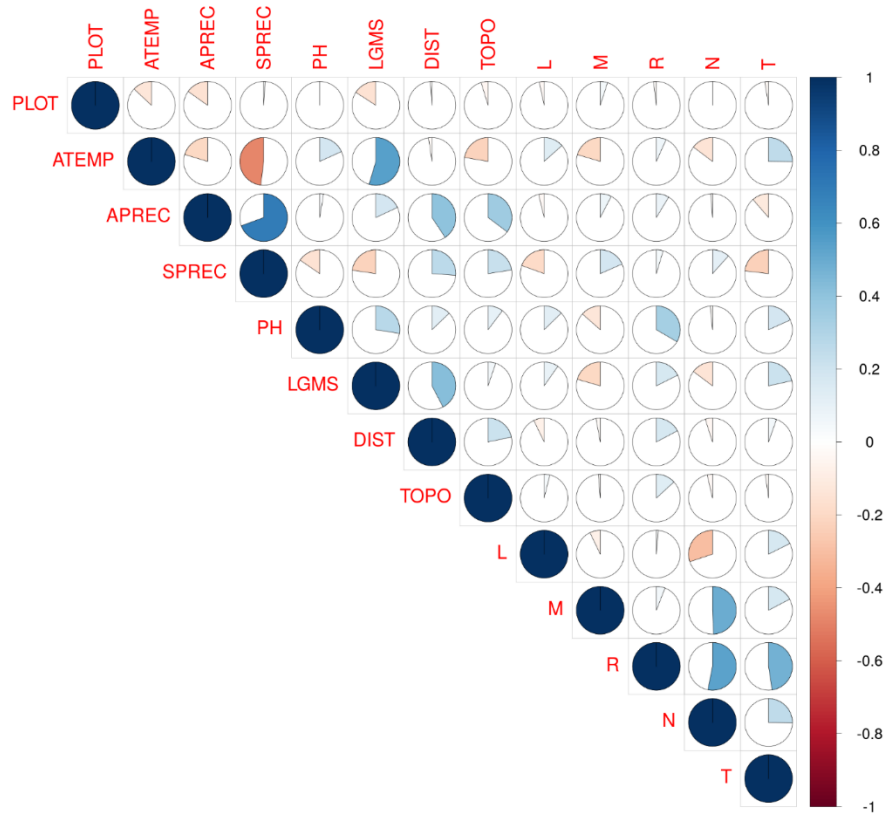
Supplementary Figure 7. Predictors of refugia. Proximity to glacial refugia sites of European beech in the last glacial maximum (PROX) according to fossil records reported by Magri, D. Patterns of post-glacial spread and the extent of glacial refugia of European beech (*Fagus sylvatica*). *J. Biogeogr.* **35**, 450–463 (2008); and Topographic heterogeneity (TOPO) based on Riley's terrain ruggedness index and scaled at 1 km grid cell.



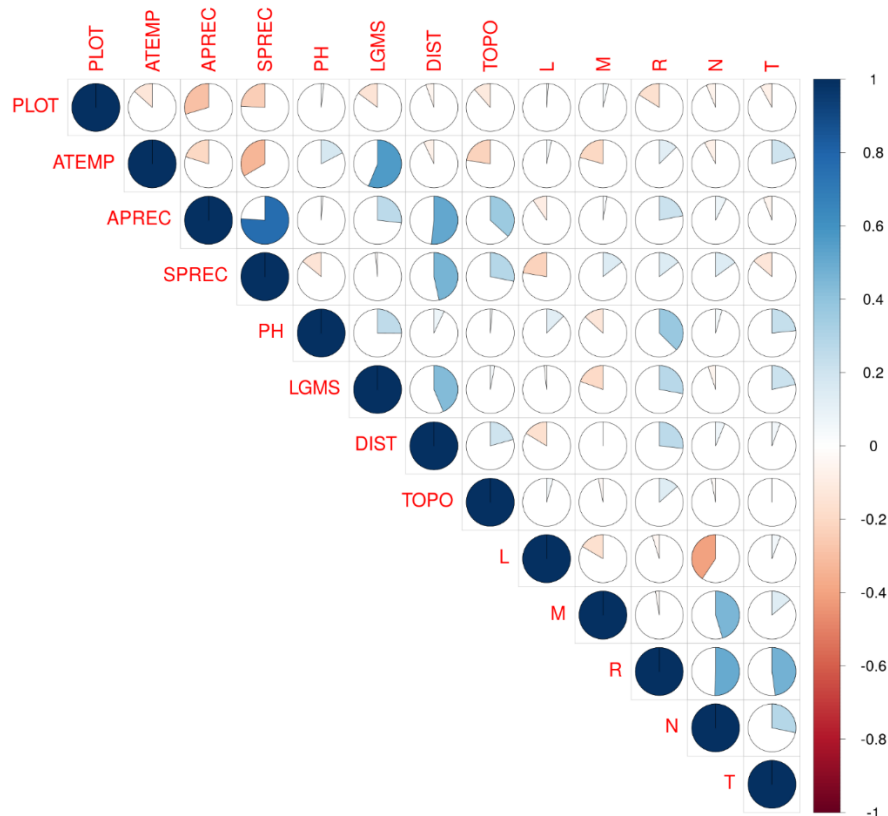
- | | | |
|------------------------------|------------------------------|--------------------------|
| 1, Apuseni Mountains | 17, N European Plain | 31, S (Italian) Alps |
| 2, Bavarian-Tyrolian Alps | 18, N France | 32, S Appennins |
| 3, Bohemian Massif | 19, N French Alps | 33, S Carpathians |
| 4, Cantabro-Atlantic | 20, N Iberian | 35, S England |
| 5, Central (German) Low Mts | 21, NE (Austrian) Alps | 36, S French Alps |
| 6, Central Appennins | 22, NE Carpathians | 37, SE (Slovenian) Alps |
| 7, Central Dinaric Mountains | 23, NW (Austrian) Alps | 38, Sicily |
| 8, Corsica | 24, NW Dinaric Montains | 39, Stara Planina |
| 9, E Carpathians | 25, NW European Plain | 40, Styrian Alps |
| 10, E Central Greece | 26, Piedmonte-Insubrian Alps | 41, SW Pannonia |
| 11, French and Swiss Jura | 27, Pindos and S Albania | 42, Swiss Alps & Molasse |
| 14, Macedonian/Serbian Mts | 28, Poland-Ukraine | 43, W Carpathians |
| 15, Massif Central | 29, Pyrenees | 44, W Lower Mountains |
| 16, N Appennins | 30, Rhodopi -Rila-Pirin | |

Supplementary Figure 8. Biogeographic regions used to calculate regional species pools. The regions are based on Willner, W. et al. Classification of European beech forests: a Gordian Knot? *Appl. Veg. Sci.* **20**, 494–512 (2017). Regions finally used for data analyses excluded numbers 1, 7, 34 and 40 because they had < 100 vegetation plots.

a
(Regional pools)
N = 5,143

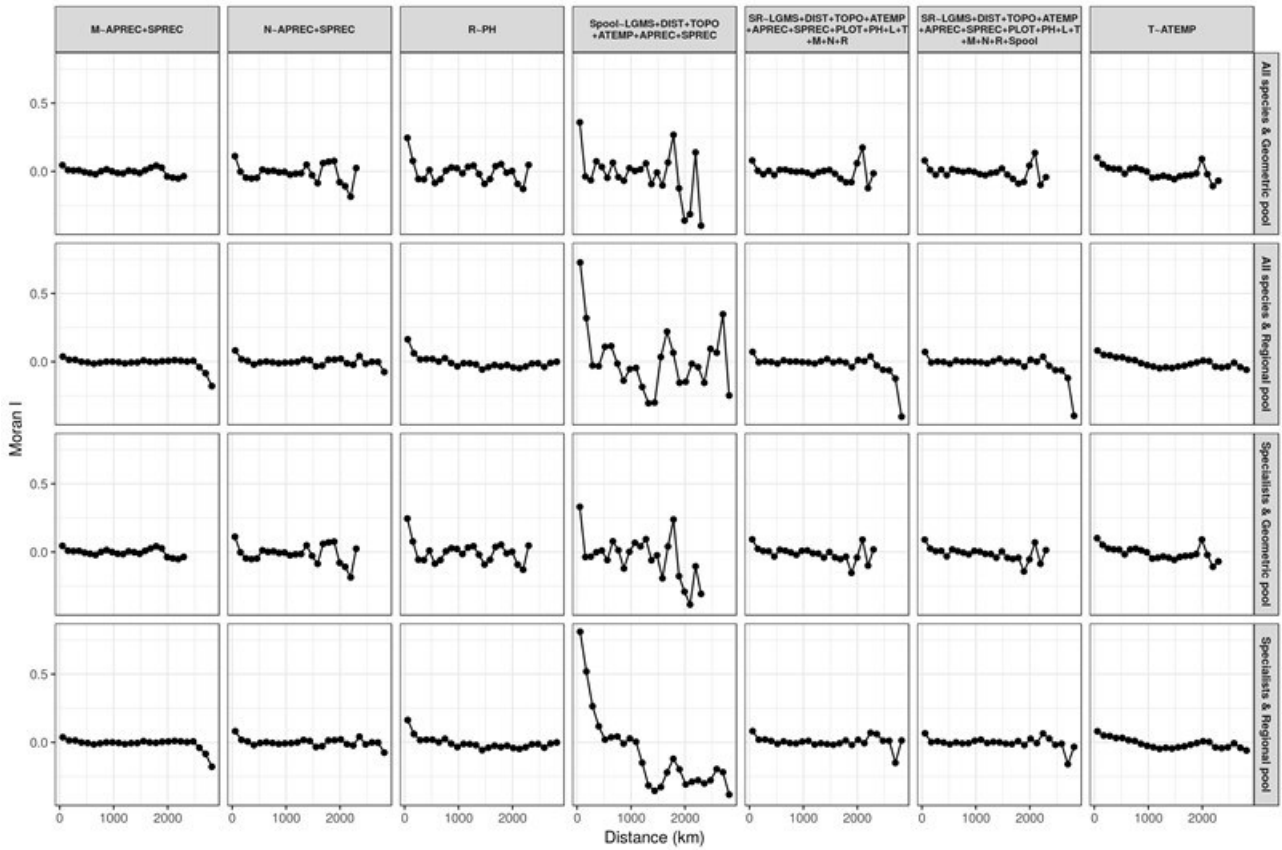


b
(Local pools)
N = 2,373

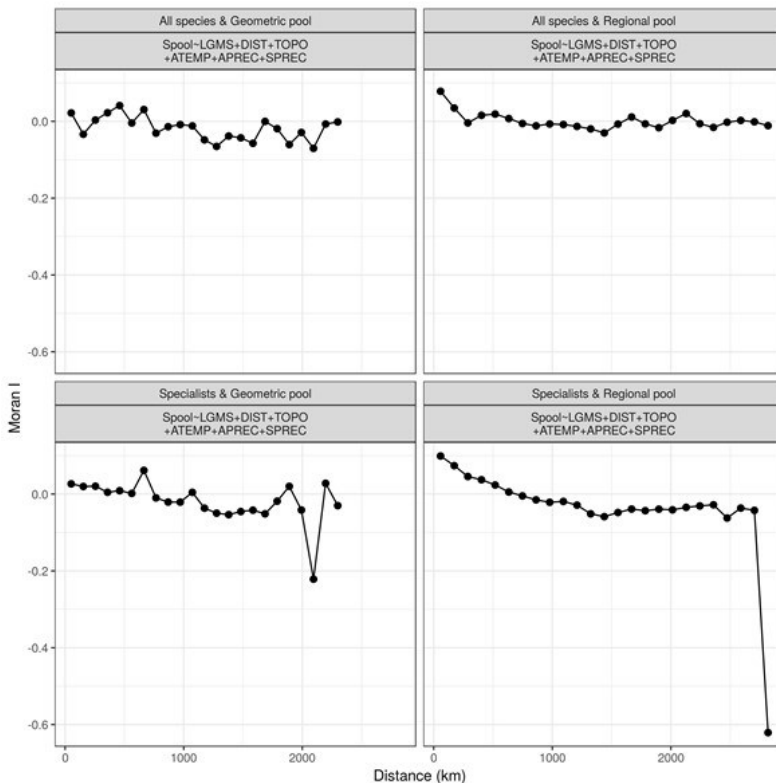


Supplementary Figure 9. Correlations between explanatory variables in the two data sets used for estimating regional and local species pool sizes of European beech forests. PLOT: plot size; ATEMP: Annual mean temperature; APREC: Annual precipitation; SPREC: summer precipitation; PH: regional soil pH; LGM: suitability at the last glacial maximum; DIST: proximity to refugia; TOPO: topographic heterogeneity; L, M, R, N, T: Ellenberg indicator values for light, moisture, soil reaction, nutrients and temperature, respectively.

a



b



Supplementary Figure 10. a, correlograms for residuals of piecewise SEM models. b, Correlograms fitted as spatial autoregressive error models (SAR). PLOT: plot size; ATEMP: Annual mean temperature; APREC: Annual precipitation; SPREC: summer precipitation; PH: regional soil pH; LGM: suitability at the last glacial maximum; DIST: proximity to refugia; TOPO: topographic heterogeneity; L, M, R, N, T: Ellemberg indicator values for light, moisture, soil reaction, nutrients and temperature, respectively. Geometric pool refers to local species pool.