Phylogenetic structure of European forest vegetation

Josep Padullés Cubino, Zdeňka Lososová, Gianmaria Bonari, Emiliano Agrillo, Fabio Attorre, Erwin Bergmeier, Idoia Biurrun, Juan Campos, Andraž Čarni, Mirjana Ćuk, et al.

Phylogenetic structure of European forest vegetation

Running title: Phylogenetic structure of European forests

Josep Padullés Cubino[1*], Zdeňka Lososová[1], Gianmaria Bonari[1,2], Emiliano Agrillo[3], Fabio Attorre[4], Erwin Bergmeier[5], Idoia Biurrun[6], Juan Antonio Campos[6], Andraž Čarni[7,8], Mirjana Ćuk[9], Michele De Sanctis[4], Adrian Indreica[10], Borja Jiménez-Alfaro[11], Larisa Khanina[12], Ilona Knollová[1], Jonathan Lenoir[13], Remigiusz Pielech[14], Valerijus Rašomavičius[15], Željko Škvorc[16], Jens-Christian Svenning[17,18], Kiril Vassilev[19], Wolfgang Willner[20], Milan Chytrý[1]

[1]Department of Botany and Zoology, Faculty of Science, Masaryk University, Brno, Czech Republic

[2]Faculty of Science and Technology, Free University of Bozen-Bolzano, Bolzano, Italy

[3]Institute for Environmental Protection and Research, Rome, Italy

[4]Department of Environmental Biology, Sapienza University of Rome, Rome, Italy

[5]Department of Vegetation and Plant Diversity Analysis, University of Göttingen, Göttingen, Germany

[6]Department of Plant Biology and Ecology, University of the Basque Country UPV/EHU, Bilbao, Spain

[7]Research Centre of the Slovenian Academy of Sciences and Arts, Institute of Biology, Ljubljana, Slovenia

[8]University of Nova Gorica, Nova Gorica, Slovenia

[9]Department of Biology and Ecology, Faculty of Science, University of Novi Sad, Novi Sad, Serbia

[10]Department of Silviculture, Transilvania University of Brasov, Brasov, Romania

[11]Research Unit of Biodiversity (CSIC-UO-PA), University of Oviedo, Mieres, Principado de Asturias, Spain

[12]Institute of Mathematical Problems of Biology of RAS - branch of the Keldysh Institute of Applied Mathematics of Russian Academy of Sciences, Pushchino, Russia

[13]Ecologie et Dynamique des Systèmes Anthropisés (EDYSAN, UMR 7058 CNRS), Université de Picardie Jules Verne, Amiens Cedex 1, France

[14]Department of Forest Biodiversity, University of Agriculture in Kraków, Kraków, Poland

[15]Nature Research Centre, Institute of Botany, Vilnius, Lithuania

[16]Faculty of Forestry, University of Zagreb, Zagreb, Croatia

[17]Center for Biodiversity Dynamics in a Changing World (BIOCHANGE), Department of Biology, Aarhus University, Aarhus C, Denmark

[18]Section for Ecoinformatics and Biodiversity, Department of Biology, Aarhus University, Aarhus C, Denmark

[19]Department of Plant and Fungal Diversity and Resources, Institute of Biodiversity and Ecosystem Research, Bulgarian Academy of Sciences, Sofia, Bulgaria

[20]Department of Botany and Biodiversity Research, University of Vienna, Rennweg 14, 1030 Vienna, Austria

*Corresponding author: padullesj@gmail.com

ORCID numbers:

Josep Padullés Cubino: 0000-0002-2283-5004

Zdeňka Lososová: 0000-0001-9152-7462

Gianmaria Bonari: 0000-0002-5574-6067

Emiliano Agrillo: 0000-0003-2346-8346

Fabio Attorre: 0000-0002-7744-2195

Erwin Bergmeier: 0000-0002-6118-4611

Idoia Biurrun: 0000-0002-1454-0433

Juan Antonio Campos: 0000-0001-5992-2753

Andraž Čarni: 0000-0002-8909-4298

Mirjana Ćuk: 0000-0002-8261-414X

Michele De Sanctis: 0000-0002-7280-6199

Adrian Indreica: -

Borja Jiménez-Alfaro: 0000-0001-6601-9597

Larisa Khanina: -

Ilona Knollová: 0000-0003-4074-789X

Jonathan Lenoir: 0000-0003-0638-9582

Remigiusz Pielech: 0000-0001-8879-3305

Valerijus Rašomavičius: -

Željko Škvorc: 0000-0002-2848-1454

Jens-Christian Svenning: 0000-0002-3415-0862

Kiril Vassilev: 0000-0003-4376-5575

Wolfgang Willner: 0000-0003-1591-8386

Milan Chytrý: 0000-0002-8122-3075

Acknowledgements

We thank all botanists and foresters who sampled the vegetation plots used in our analyses, and data contributors from the different vegetation-plot databases. We thank Stephan Hennekens for managing the EVA database, Jan Divíšek and Martin Večeřa for helpful discussion, and Irena Axmanová for providing a code to clean up the data. The study was supported by the Czech Science Foundation (19-28491X). IB and JAC were funded by the Basque Government (IT936-16). JCS considers this work a contribution to his VILLUM Investigator project "Biodiversity Dynamics in a Changing World" funded by VILLUM FONDEN (grant 16549).

Abstract

Aims: (1) To determine the contribution of current macro-environmental factors in explaining the phylogenetic structure of European forest vegetation, (2) to map and describe spatial patterns in their phylogenetic structure, and (3) to examine which lineages are the most important contributors to phylogenetic clustering and whether their contribution varies across forest types and regions.

Location: Europe.

Taxon: Angiosperms.

Methods: We analyzed the phylogenetic structure of 61,816 georeferenced forest vegetation plots across Europe considering alternative metrics either sensitive to basal (ancient evolutionary dynamics) or terminal (recent dynamics) branching in the phylogeny. We used boosted regression trees to model metrics of the phylogenetic structure as a function of current macro-environmental factors. We also identified clades encompassing significantly more taxa than under random expectation in phylogenetically clustered plots.

Results: Phylogenetic clustering was driven by climatic stress and instability and was strong in the areas glaciated during the Pleistocene, likely reflecting limited postglacial migration, and to a lower extent in areas of northern-central Europe and in summery-dry Mediterranean regions. Phylogenetic overdispersion was frequent in the hemiboreal zone in Russia, in some areas around the Mediterranean Basin, and along the Atlantic seaboard of the Iberian Peninsula. The

families Ericaceae, Poaceae and Fagaceae were overrepresented in clustered plots in different regions of Europe.

Main conclusions: We provide the first maps and analyses on the phylogenetic structure of European forest vegetation at the plot level. Our results highlight the role of environmental filtering, postglacial dispersal limitation, and spatial transitions between major biomes in determining the distribution of plant lineages in Europe.

Keywords: Community assembly; community phylogenetic structure; environmental filtering; European forests; glacial refugia; phylogenetic relatedness; postglacial recolonization; vegetation-plot data.

Introduction

Species composition of local communities is determined by the interplay between ecological and evolutionary processes (Götzenberger et al., 2012; Ricklefs, 2004). A given community can only contain species whose environmental tolerances allow them to maintain a population within the range of ecological conditions in that location (Kraft & Ackerly, 2014). The ability of a species to persist within that range of ecological conditions is in turn constrained by its evolutionary history. As a result, co-occurring species in different communities may present different levels of evolutionary relatedness, which is usually referred to as 'phylogenetic community structure' (Webb, Ackerly, McPeek, & Donoghue, 2002).

Previous studies examining patterns in the phylogenetic structure of plant communities have generally found that communities in more stressful (i.e., resource-limited; e.g., colder, drier, more seasonal) conditions exhibit higher internal phylogenetic relatedness (clustering) among their species because strong environmental filtering combined with phylogenetic niche conservatism would select for a subset of lineages adapted to these extreme environments (Lososová et al., 2015, 2021; Lu et al., 2018; Ma et al., 2016; Qian et al., 2019; Qian & Sandel, 2017; Thornhill et al., 2016). However, other processes operating at different spatial and temporal scales can also be highly influential, such as competitive exclusion (Graham, Parra, Rahbek, & McGuire, 2009), speciation (Guevara et al., 2016), climate-driven non-random extinctions (Eiserhardt, Borchsenius, Plum, Ordonez, & Svenning, 2015), geographic isolation (Weigelt et al., 2015), selective postglacial recolonization (Feng et al., 2014), or the species exchange between floras in transitional biogeographical regions (Kissling et al., 2012). Climatic oscillations on geological time scales and, more recently, human activities, have also shaped the distribution and composition of the European flora (Svenning, Normand, & Skov, 2009), which has left significant imprints in the phylogenetic structure of current forest vegetation (Eiserhardt

et al., 2015; Lososová et al., 2021). The effects of the aforementioned factors on the phylogenetic community structure are often interrelated, and their relative importance and direction may vary across biogeographical regions, groups of organisms, and community types, making it difficult to formulate general predictions at broad spatial scales.

The phylogenetic structure and diversity of vascular floras have been explored and mapped in a number of studies across multiple spatial scales (e.g., Lu et al., 2018; Massante et al., 2019; Mastrogianni, Kallimanis, Chytrý, & Tsiripidis, 2019). To date, most of such studies have either focused on specific clades (e.g., Kissling et al., 2012; Mishler et al., 2014) or life forms (mainly trees; e.g., Hawkins, Rueda, Rangel, Field, & Diniz-Filho, 2014; Qian & Sandel, 2017), and used either floras or regional checklists (e.g., Qian et al., 2019; Thornhill et al., 2016), thus omitting the potential effect of fine-scale processes, such as species interactions, at the plant community level. The recent explosion of global and continental vegetation-plot databases (e.g., Bruelheide et al., 2019; Chytrý et al., 2016), along with environmental data and novel analytical methods and tools, has created unprecedented opportunities for exploring fine-scale patterns of phylogenetic structure at large spatial scales and understanding their determinants.

Forests are a major component of the terrestrial ecosystems of Europe, currently representing over 40% of the land surface (European Environment Agency, 2016). Despite considerable knowledge on the distribution of species richness in European forests (e.g., Jiménez-Alfaro et al., 2018; Večeřa et al., 2019), we still lack information on the spatial variability of their phylogenetic structure. Here, we conducted a study on the phylogenetic structure of 61,816 forest vegetation plots sampled across Europe and classified into three broad forest types: coniferous, broadleaved deciduous, and broadleaved evergreen. These forest types represent the main biogeographical groups of European forests, and their distribution is determined by different ecological and historical processes (e.g., Tzedakis, Emerson, & Hewitt, 2013; Večeřa et al., 2019). Therefore, we also expected differences in the drivers of their

phylogenetic structure. In addition, we examined the relationships between metrics of phylogenetic structure and macroclimatic and other environmental variables and identified which clades, with their distinct evolutionary histories and ecological roles, were overrepresented (i.e., contained more species than random samplings) in phylogenetically clustered plots for the three forest types. We combined this information to understand the phylogenetic patterns of forest vegetation in Europe.

To explore the phylogenetic structure of European forest vegetation, we considered alternative metrics either sensitive to basal (ancient evolutionary dynamics) or terminal (recent dynamics) branching in the phylogeny (Mazel et al., 2016). Simultaneous basal and terminal phylogenetic clustering can be a consequence of environmental filtering and niche conservatism operating at both large and fine evolutionary time scales (Fig. 1; scenario A). Alternatively, basal overdispersion (i.e., the co-occurrence of distantly-related species) and terminal clustering reflects the legacy of ancient evolutionary radiation, and the effect of environmental filtering of clade-specific habitat preferences acquired over time (Fig. 1; scenario B) (Kooyman, Rossetto, Cornwell, & Westoby, 2011). Communities showing this pattern could represent both a refuge for old lineages ('museum'), and a center of speciation ('cradle') at finer evolutionary time scales (Jablonski, Roy, & Valentine, 2006; Koenen, Clarkson, Pennington, & Chatrou, 2015). Basal clustering and terminal overdispersion can arise when environmental filtering operates on long-term conserved niche requirements, while niche differentiation develops between close relatives (Fig. 1; scenario C) (Hardy & Senterre, 2007). Finally, simultaneous basal and terminal phylogenetic overdispersion mainly reflects the combined effect of long-term environmental stability, biogeographic processes, and recent niche differentiation on community assembly (Fig. 1; scenario D) (Kooyman et al., 2011).

We expected basal and terminal clustering of European forest vegetation to occur in high-latitude areas glaciated in the Pleistocene by the continental ice-sheet because of strong

environmental filtering in deglaciated areas and limited postglacial recolonization and speciation. In contrast, we expected basal and terminal overdispersion to occur in areas with relatively high historical climate stability due to either lack of severe drought, such as in some coastal and mountain areas, or lack of severe cold, such as in the southern European peninsulas (Iberian, Italian, and Balkan). These areas are considered to have been refugia for many species and lineages during the Pleistocene glaciations (Bennett, Tzedakis, & Willis, 1991; Tzedakis et al., 2013). To get insights into these expectations, we (1) determined the contribution of current macro-environmental factors in explaining the phylogenetic structure of European forest vegetation, (2) mapped and described spatial patterns in their phylogenetic structure, and (3) examined which lineages are the most important contributors to phylogenetic clustering and whether their contribution varies across forest types and regions.

Methods

Vegetation data

We obtained georeferenced vegetation-plot records from the European Vegetation Archive (EVA; Chytrý et al., 2016; accessed on June 28th 2019), which contains more than 1.5 million vegetation plots sampled across Europe (see Appendix S1: Table S1.1 in Supporting Information for an overview of contributing datasets). Forestry plantations, defined as forests with the dominance of non-native trees in the target area, were excluded from this selection, and we only selected plots with areas ranging between 100 and 1,000 m^2, which were the most common plot sizes in the database. We also removed plots sampled before 1950 in order to avoid potential bias related to long-term vegetation change, and plots where the potential spatial difference between the actual sampling point and the geographical coordinates provided in the database (i.e.,

'location uncertainty') were larger than 10 km. However, we retained plots with no information on location uncertainty because some regions were represented only by this sort of data (in most of such cases location uncertainty is lower than 10 km in the EVA database). To identify forest vegetation plots in the EVA database, we used a classification expert system of European EUNIS habitats (EUNIS-ESy v. 2020-06-08; Chytrý et al., 2021) run in the JUICE program (Tichý, 2002). We selected those plots which were classified by the expert system into (i) coniferous forests, (ii) broadleaved deciduous forests, and (iii) broadleaved evergreen forests (see Appendix S1: Table S1.2 for an overview of plots assigned to particular forest habitats). These three forest habitats correspond to the main forest categories proposed by the EUNIS habitat classification (Davies, Moss, & Hill, 2004). The classification of forest plots into these three types allowed us to limit the potential confounding effects of different ecological and evolutionary histories in their phylogenetic structure, and make our results comparable to previous studies examining the geographic patterns and the biogeographical drivers of alpha-diversity in forest types across Europe (Večeřa et al., 2019).

After the selection of plots and their classification into forest types, we obtained three datasets: 'coniferous forests' containing 31,769 plots, 'broadleaved deciduous forests' containing 100,404 plots, and 'broadleaved evergreen forests' containing 6,248 plots. To avoid the oversampling of particular areas, we performed a stratified resampling in each dataset. First, we assigned each vegetation plot to a geographical grid of 0.2° × 0.2°. Second, we determined the minimum number of plots required from a given cell as a truncated linear function of the β-diversity of that cell (Wiser & De Cáceres, 2013), which was defined as the overall compositional variation of the plots within the cell, measured with the turnover component of Sørensen's dissimilarity index (Legendre, Borcard, & Peres-Neto, 2005). Third, for those grid cells in which the number of recorded plots exceeded the minimum number required, we conducted a heterogeneity-constrained random (HCR) resampling (Lengyel, Chytrý, & Tichý,

2011) with R package 'vegclust' (De Cáceres, Font, & Oliva, 2010). Otherwise, all plots were retained. The resampled datasets contained (i) 16,382 plots for coniferous forests, (ii) 41,198 plots for broadleaved deciduous forests, and (iii) 4,236 plots for broadleaved evergreen forests. For each plot, we obtained data on species occurrences and removed non-vascular plants. We also standardized the species names in our datasets according to The Plant List using the R package 'Taxonstand' (Cayuela, Stein, & Oksanen, 2017).

Phylogenetic data

We used the R package 'V.PhyloMaker' (Jin & Qian, 2019) to link the species names in our datasets with those in the megaphylogeny implemented in the same package. This megaphylogeny was derived from two mega-trees (Smith & Brown, 2018; Zanne et al., 2014), and included 74,533 species and all families of extant vascular plants. We used the 'scenario 3' approach implemented in the same R package to add missing species (~43%) to the phylogeny (for further details see Jin & Qian, 2019; Qian & Jin, 2016). We pruned our complete phylogeny to create three separate trees which included the species recorded in coniferous (5,364 species), broadleaved deciduous (5,122 species) and broadleaved evergreen (2,936 species) forests.

Calculating and mapping the phylogenetic structure

We estimated the phylogenetic structure of each plot by calculating the phylogenetic Mean Pairwise Distance (MPD) and Mean Nearest Taxon Distance (MNTD), using the R package 'PhyloMeasures' (Tsirogiannis & Sandel, 2017). This package offers efficient algorithms that can process large sample sizes and phylogenetic data for such measures. MPD was calculated as the mean branch length distance between all pairs of species occurring in a vegetation plot, and

MNTD as the mean branch length distance between each species and its phylogenetically nearest neighbor in the vegetation plot (Webb et al., 2002). Thus, MPD reflects relatedness between the basal branches of the phylogeny and MNTD reflects relatedness between the terminal branches of the phylogeny. In addition, we calculated Faith's PD (i.e., the sum of all phylogenetic branch lengths that connect species in an assemblage; Faith, 1992), which was significantly correlated with MNTD (Pearson's $r > 0.95$; Appendix S2: Table S2.1), and therefore subsequently discarded from analysis (but see results in Appendix S3: Fig. S3.1).

We measured the standardized effect size of MPD and MNTD (ses.MPD and ses.MNTD) to produce metrics of phylogenetic structure that were independent of species richness (Pavoine & Bonsall, 2011). Specifically, 'PhyloMeasures' calculates ses.MPD and ses.MNTD not based on a resampling approximation of the mean and variance, but rather based on exact solutions given a particular tree and species richness. To define these statistical moments, a null model and a defined species pool are required. Our null model considered all possible combinations of *S* species from the species pool (where *S* is the richness of a sample to be standardized) to have the same (uniform) probability of occurrence (Tsirogiannis & Sandel, 2016). We then considered three different species pools, one for each forest type (i.e., coniferous, broadleaved deciduous, and broadleaved evergreen), assuming that each forest type had a particular pool of species capable of growing only or predominantly in that particular type. This approach allowed us to meaningfully compare patterns in the phylogenetic structure of communities in the three forest types with the independence of the effect of filtering processes acting on a larger pool of forest species (i.e., all sampled species). We used the *mpd.pval* and *mntd.pval* functions in package 'PhyloMeasures' to compute *P*-values associated with ses.MPD and ses.MNTD, respectively (Tsirogiannis & Sandel, 2016).

We performed analysis only for angiosperms, thus excluding gymnosperms and pteridophytes (~3% of the species), to avoid inflated overdispersion due to a few taxa connected

to deep phylogenetic nodes (see Appendix S3: Fig. S3.2 for results with all vascular plants). Gymnosperms and pteridophytes represented, on average, 15%, 7%, and 6% of the total species counts in vegetation plots of coniferous, broadleaved deciduous, and broadleaved evergreen forests, respectively. We maintained original names of coniferous forest types despite removing their dominant species because we retained the majority of species characteristic of these habitats.

To map ses.MPD and ses.MNTD of forest vegetation plots in a comprehensive manner, we created a grid cell of 1° × 1° and reclassified all forest vegetation plots inside each grid cell as follows: -1 if the plot was significantly clustered, 0 if the plot had a random phylogenetic structure, and 1 if the plot was significantly overdispersed. Then, we calculated the overall value of each grid cell as the average of all plots included in the cell. We mapped only grid cells with a minimum of five plots. We repeated the analysis with different grid-cell resolutions and a minimum number of plots to test for their effect on the results, but they all showed similar patterns (Appendix S4: Figs. S4.1-S4.2). We reported the average ses.MPD and ses.MNTD values per grid cell in Appendix S3: Fig. S3.3.

Environmental data

We selected a set of present macroclimatic, topographic and soil substrate variables to test the effect of environmental conditions on the phylogenetic structure of forest vegetation plots across Europe (Ma et al., 2016; Qian et al., 2019; Sandel et al., 2011). Present macroclimatic variables included 'mean annual temperature' (MAT; ºC), 'annual precipitation' (AP; mm), 'temperature seasonality' (Tses; the standard deviation of mean monthly temperatures; ºC), and 'precipitation seasonality' (Pses; coefficient of variation in monthly precipitation). We also calculated the 'velocity of temperature change' (VTC; meters/year) since the Last Glacial Maximum (LGM; ~

21,000 years BP) following Sandel et al. (2011). VTC was used to capture the magnitude of glacial-interglacial fluctuation. It represents a rate of local displacement of temperature, which can be interpreted as the rate at which a species would have to migrate in space to maintain the same climate conditions over time (Sandel et al., 2011). In addition, we calculated the 'extent of analogous temperature' (EAT), which is a measure of climate commonness, based on the climate rarity measure of Ohlemüller et al. (2008). We calculated EAT as the proportion of surrounding cells within a 100 km radius of each plot where MAT did not differ by more than 1° from the temperature in the location of the plot. Modern macroclimatic variables were downloaded from the WorldClim v. 2.1 database (Fick & Hijmans, 2017; freely available at https://worldclim.org) at a resolution of 30 arc seconds (~1 km^2). However, temperature values for LGM were only available at 2.5 arc minutes (~4.6 km^2) from the WorldClim database. Thus, we also obtained modern temperature at 2.5 arc minutes to calculate VTC.

Topographic variables included 'elevation' (m a.s.l.) and the 'Terrain Ruggedness Index' (TRI; m). Elevation and TRI were derived from the global digital elevation model developed by the U.S. Geological Survey's Center for Earth Resources Observation and Science with a spatial resolution of 30 arc seconds (freely available at earthexplorer.usgs.gov). TRI is the mean of the absolute differences between the value of a cell where a plot is located and the value of its surrounding cells. It was calculated with the R package 'raster' (Hijmans, 2019). We additionally obtained data for soil pH (at 15 cm depth) at a resolution of ~250 m (rescaled to 30 arc seconds) from the SoilGrids database (https://soilgrids.org/; Hengl et al., 2017), the geographic latitude of each plot, and the maximum extent of the LGM glacier following Peltier, Argus, & Drummond (2015).

We extracted environmental values for each plot following two approaches. First, we extracted values for the specific locations of plots. Second, we delimited a circular buffer with a fixed radius of 2.5 km around each plot and calculated the mean value of all cells contained in

the buffer zone. The buffer approach smoothed potentially extreme values of environmental variables from point extractions and reduced biases derived from the inaccuracy in the coordinates. These two approaches yielded virtually identical results, and we only report results from the first approach. Correlations among environmental variables can be found in Appendix S2: Tables S2.2 to S2.4.

Statistical analysis

We used boosted regression trees (BRTs; Elith, Leathwick, & Hastie, 2008) implemented in R package 'dismo' (Hijmans, Phillips, Leathwick, & Elith, 2017) to test for the relative influence of current environmental variables on ses.MPD and ses.MNTD of particular vegetation plots in each forest type. We used BRTs with Gaussian distribution rather than conventional regression because of the ability of BRTs to handle complex non-linear relationships and account for collinearity (De'ath, 2007; Elith et al., 2008). The optimum number of regression trees was obtained using a 10-fold cross-validation procedure (see Appendix S5 and Fig. S5.1 therein for details on model fitting evaluation). For each environmental variable, we obtained its contribution score (in %) as a measure of its relative importance in BRTs models. We used partial dependence plots to visualize the shape of the relationships between response and predictor variables (Appendix S5: Figs. S5.2-S5.7). We tested for spatial autocorrelation in the residuals of each model using Moran's *I* statistics for distance classes defined using Sturges' rule. Because we were interested in spatial autocorrelation at short spatial distances, we randomly selected 500 plots and calculated the spatial correlogram for all plots located within 2º (~225 km) of the target plot. Resulting correlograms were then summarized using local polynomial regression fitting with a span of 0.75 (Appendix S5: Fig. S5.8).

To identify nodes on the phylogenetic tree encompassing significantly more taxa present in vegetation plots than under random expectation, we used the *nodesig* function originally implemented in PHYLOCOM and adapted for R by Abellán, Carrete, Anadón, Cardador, & Tella (2016). This function determines the position of phylogenetic clustering by testing each node of the phylogeny for an overabundance of terminal taxa. Observed patterns were compared to the null model prediction that species are a random sample from the entire phylogeny in each forest type. We used 999 randomizations. As a result, the function returns for each plot the randomization rank and the associated *P*-value of each node in the tree. We carried out this analysis only for vegetation plots showing significant basal and terminal phylogenetic clustering (Fig. 1; scenario A) to reduce the computational time (885 plots for coniferous forests, 640 for broadleaved deciduous forests, and 55 for broadleaved evergreen forests).

We further grouped significantly clustered plots within each forest type based on the relative contribution of each node to its phylogenetic structure. For this, we first created a plot × node matrix (including all nodes of the respective tree) and assigned to each node its respective randomization rank as obtained from *nodesig*. We used the plot × node matrix and hierarchical clustering with Ward distance to group forest vegetation plots. To determine the optimal number of groups for each forest type, first, we repeated hierarchical clustering with all possible numbers of groups from 2 to 10. Second, we performed distance-based partial Redundancy Analysis (db-RDA) on the plot × node matrix and tested the effect of each possible number of groups, and the intercept (i.e., no groups), on node composition. We used Sørensen's dissimilarity index in the R package 'vegan' (Oksanen et al., 2019). Third, we ranked the db-RDA models by their AIC and selected the model with the lowest AIC as that containing the optimal number of groups. As a result, we grouped plots of both coniferous and broadleaved deciduous forests in two groups (Appendix S6: Table S6.1). Conversely, because broadleaved evergreen forests were represented by a relatively lower number of plots and a more restricted geographic range than coniferous and

broadleaved deciduous forests, we did not perform clustering for this forest type, and we kept it as a single group. Finally, we ranked nodes in each group by summing up their randomization rank.

We established significance at $\alpha < 0.05$ and performed all the analyses in R v. 3.5.3 (R Core Team, 2019).

Results

Environmental drivers of phylogenetic structure

The BRT models explained between 15% and 47% of the variation in ses.MPD and ses.MNTD, with the highest proportion of explained variation achieved for broadleaved evergreen forests and the lowest for broadleaved deciduous forests (Fig. 2). In coniferous and broadleaved deciduous forests, ses.MPD and ses.MNTD were best predicted by latitude, with a strong negative effect north of approximately 58º N (Appendix S5: Figs. S5.2-S5.5). Temperature seasonality was also among the top three predictors of ses.MPD and ses.MNTD in coniferous and broadleaved deciduous forests, with generally a multimodal response. Basal (ses.MPD) and terminal (ses.MNTD) phylogenetic clustering consistently increased with elevation and precipitation seasonality in coniferous and broadleaved deciduous forests, respectively, whereas annual precipitation consistently promoted both basal and terminal phylogenetic overdispersion in broadleaved deciduous forests. In broadleaved evergreen forests, variation in ses.MPD was best explained by latitude, terrain ruggedness, and temperature seasonality, whereas variation in ses.MNTD was best explained by temperature seasonality, terrain ruggedness, and elevation. Basal and terminal phylogenetic clustering in broadleaved evergreen forests consistently increased with temperature and precipitation seasonality, elevation, and the extent of analogous

temperature, whereas basal and terminal phylogenetic overdispersion increased with mean annual temperature. We found negligible spatial autocorrelation in residuals from all models (Appendix S5: Fig. S5.8).

Patterns of phylogenetic structure

We found similar patterns for both ses.MPD and ses.MNTD in forest vegetation across Europe (Fig. 3; Appendix S3: Fig. S3.4). In particular, both coniferous and broadleaved deciduous forests tended to be phylogenetically clustered for both metrics (Fig. 1; scenario A) in Fennoscandia (Fig. 3; see also Appendix S3: Fig. S3.5 for aggregated results across all forest types). To a lesser extent, phylogenetic clustering for both metrics was also observed in coniferous forests in Belarus, northern Poland, and Scotland. In southern Europe, phylogenetic clustering of forest vegetation was sporadically observed in Greece, south and western Anatolia, and interior regions of the Iberian Peninsula.

Hotspots of phylogenetic overdispersion for both ses.MPD and ses.MNTD (Fig. 1; scenario D) were found in interior regions of Russia for both coniferous and broadleaved deciduous forests (Fig. 3). Phylogenetic overdispersion for both metrics extended towards the Baltic countries and western areas of Belarus and Ukraine for broadleaved deciduous forests. Furthermore, all three forest types showed phylogenetic overdispersion of basal branches around the Adriatic Sea (Fig. 3; Appendix S3: Fig. S3.4). Broadleaved evergreen forests also showed a tendency towards phylogenetic overdispersion for both metrics in regions surrounding the Tyrrhenian Sea and coastal areas in the northern Iberian Peninsula. Basal and terminal phylogenetic overdispersion was also observed in the south-western Iberian Peninsula for broadleaved deciduous forests.

We found no consistent patterns of basal overdispersion and terminal clustering (i.e., positive ses.MPD and negative ses.MNTD; Fig. 1; scenario B) in coniferous and broadleaved deciduous forests throughout Europe. However, for broadleaved evergreen forests, we observed this pattern, albeit relatively weak, in some areas in Greece and sporadically in coastal areas of the Mediterranean (Appendix S3: Fig. S3.4). We found no consistent patterns of basal clustering and terminal overdispersion (Fig. 1; scenario C) throughout Europe for any forest type.

Among coniferous forests, phylogenetic clustering occurred especially in *Pinus* bog and *Picea* mire forests, whereas phylogenetic overdispersion characterized *Picea* and *Larix* taiga forests (Appendix S7: Figs. S7.1-S7.2). The contrasting patterns in the phylogenetic structure of coniferous forests in the border between Belarus and Russia (Fig. 3A) can be explained by the differences in sampled habitats in both regions: phylogenetically clustered *Pinus* forests in Belarus and overdispersed *Picea* and *Larix* taiga forests in adjacent Russia (Appendix S7: Fig. S7.2). Among broadleaved deciduous forests, phylogenetic clustering was intense in boreal and nemoral mountain *Betula* forests on mineral soils, whereas phylogenetic overdispersion was strong in temperate and boreal hardwood riparian forests and ravine forests (Appendix S7: Figs. S7.3-7.4). In broadleaved evergreen forests, we found the strongest phylogenetic overdispersion in mainland laurophyllous forests (Appendix S7: Figs. S7.5-7.6).

Overrepresented clades

Coniferous forests were divided into two main groups of vegetation plots along a latitudinal gradient based on the relative contribution of each node to its phylogenetic structure (Fig. 4A). The first group (C1) was mainly located at latitudes higher than 50º N and strongly overrepresented by Ericaceae, particularly *Vaccinium* (Fig. 4D; Appendix S8: Fig. S8.1). The second group (C2) was mainly located in the Mediterranean area. In this group, phylogenetic

clustering was mainly due to the overrepresentation of Poaceae and, to a lesser extent, Asteraceae. Nonetheless, the genus with the highest randomization rank was *Vaccinium* (Ericaceae). The two groups of plots tended to converge in the Alps.

Broadleaved deciduous forests were also divided into two groups following a north-west to south-east gradient (Fig. 4B). The first group (D1) was characterized by an overrepresentation of Fagales, particularly Betulaceae and Fagaceae (Fig. 4D; Appendix S8: Fig. S8.2). In the second group (D2), phylogenetic clustering was mostly driven by Poaceae, and the genus *Poa* therein. Broadleaved evergreen forests were strongly clustered for *Quercus* and, more broadly, Fagaceae (Fig. 4C-D; Appendix S8: Fig. S8.3).

Discussion

Environmental drivers of phylogenetic structure

Our results revealed that both basal and terminal phylogenetic clustering increased with latitude in coniferous and broadleaved deciduous forests. This effect was mainly driven by the influence of harsh environmental conditions and limited postglacial recolonization in areas north of approximately 58° N (see also Fig. 3 and associated discussion in the next section). Phylogenetic clustering was also consistently associated with higher elevation and precipitation seasonality in coniferous and broadleaved deciduous forests, respectively, and with higher elevation and climate seasonality, and lower mean annual temperature in broadleaved evergreen forests. These findings thus further support the role of environmental filtering in promoting phylogenetic clustering (e.g., Lu et al., 2018; Ma et al., 2016; Qian & Sandel, 2017; Thornhill et al., 2016).

The explanatory power of our models was highest for the broadleaved evergreen forests (narrowest environmental niche and geographic range) and lowest for the broadleaved deciduous

forests (broadest environmental niche and geographic range), suggesting that environmental drivers may cancel each other out, or be influential only in interaction with other factors, with an increasing spatial extent and environmental heterogeneity. The proportion of unexplained variation in our models could be attributed to the effect of fine-scale environmental conditions (partially derived from contrasting canopy structures), human management, and other biogeographical historical factors not accounted for by our predictors. However, the influence of biogeographical history on the phylogenetic structure of European forests can partially be interpreted from our maps, as we describe in the following sections. The potential scale mismatch between environmental and vegetation data may have also lowered the potential explanatory power of our models. Moreover, the relatively weak associations between environmental variables and plant phylogenetic structure in our vegetation plots contrast with results from previous studies that examined similar patterns in China (Lu et al., 2018; Qian et al., 2019) and North America (Ma et al., 2016; Qian, Zhang, Zhang, & Wang, 2013; Qian, Zhang, Sandel, & Jin, 2019). These differences can partially be explained by the fact that we examined phylogenetic patterns among all co-occurring species in particular vegetation plots, therefore accounting for fine-scale processes such as competition or facilitation, whereas previous works have mostly focused on regional floras, life forms, or particular lineages. The inclusion of all vegetation layers in our analyses was crucial to our goals, given that most of the plant diversity in temperate and boreal forest ecosystems is found in the herb layer (Gilliam, 2007).

Patterns of phylogenetic clustering and overrepresented clades

Following our expectations, we found a tendency towards basal and terminal phylogenetic clustering (Fig. 1; scenario A) in high-latitude areas glaciated in the Pleistocene by the continental ice-sheet. Phylogenetic clustering in these areas may reflect reduced postglacial

recolonization of some poorly dispersing and stress-intolerant lineages (e.g., Svenning, Normand, & Skov, 2008; Willner, Di Pietro, & Bergmeier, 2009), as well as recolonization by taxa from southern latitudes that underwent non-random extinctions of certain lineages (Eiserhardt et al., 2015). Accordingly, Ericaceae, which largely disperse by anemochory or endozoochory, and possess adaptations to nutrient-poor (oligotrophic) habitats, including ericoid mycorrhiza and tough evergreen leaves (Cairney & Meharg, 2003), were overrepresented in these areas and further south in northern continental Europe. Other clades, including wind-pollinated, cold-hardy and fast-growing species, such as Betulaceae, are also overrepresented in broadleaved deciduous forests of these same areas.

In the Mediterranean region, centers of basal and terminal phylogenetic clustering were located in inland areas of the Iberian Peninsula, and sporadically in the Aegean region. Like any semi-arid environment, the vegetation in inland areas of the Iberian Peninsula is severely constrained by scarce and unpredictable rainfall, but also by low temperatures at high altitudes (Whitford, 2002). Here, clustering was mainly driven by dominant long-living evergreen trees of the Fagaceae family, particularly *Quercus*. This clade is adapted to semi-arid environmental conditions (Leuschner et al., 2001), thanks to a wide range of adaptive traits such as a xeromorphic leaf structure or a deep-penetrating root system that allow them to cope with limited nutrient and water availability (Kubiskem & Abrams, 1993; Kuster, Arend, Günthardt-Goerg, & Schulin, 2013). In the eastern Mediterranean, and particularly in broadleaved deciduous forests, Poaceae accompanied Fagaceae as the main contributor to phylogenetic clustering. Poaceae are supposed to have originated in tropical forest-edge habitats, and their radiation correlates with an acquired ability to tolerate grazing pressure, fire, and drought (Bouchenak-Khelladi, Verboom, Savolainen, & Hodkinson, 2010; Clayton, 1981). Grasses are also prominent drivers of phylogenetic clustering in the forest-steppe zone of eastern Europe, where increasing continentality favors open forest structure (Erdős et al., 2018). In contrast, the influence of the

Atlantic wet climate in western Europe promotes more closed forests; here, the dominant woody species of Fagales are important contributors to phylogenetical clustering. The structure and composition of European forests have also been influenced by human activities since the last ice age, particularly by forest fragmentation, logging, forest grazing, and recurrent fires (Bradshaw, 2004). However, the magnitude of the effects of human management on the phylogenetic structure of forest vegetation remains a topic of further research.

Patterns of phylogenetic overdispersion

The southern European peninsulas (Iberian, Italian and Balkan) are considered to have been refugia for many lineages during the Pleistocene glaciations (Bennett et al., 1991; Tzedakis et al., 2013). Previous studies have also confirmed more localized refugia, and areas with high species richness, in the south-eastern and western Alps, the northern Dinaric Mountains, the southern Apennines, or northern Spain (Magri et al., 2006; Médail & Diadema, 2009; Večeřa et al., 2019; Willner et al., 2009). Confirming our initial expectations, we found partial overlap between these areas and hotspots of phylogenetic overdispersion (Fig. 1; scenario D) around the Adriatic and Tyrrhenian Seas, and Atlantic seaboard of the Iberian Peninsula, although with differences among forest types. It is possible that the persistence of long-term favorable environmental conditions in these refuge areas could result in limited extinction rates, and entailed stronger niche differentiation due to competitive interactions, resulting in the accumulation of taxa with different evolutionary histories (Jablonski et al., 2006). In addition, for broadleaved evergreen forests, we found evidence for basal (ancient) phylogenetic overdispersion but terminal (recent) clustering (Fig. 1; scenario B) in some areas in Greece and sporadically in coastal areas of the Mediterranean. Forests in these areas could represent both a former refuge for forest taxa ('museum') and a center of speciation ('cradle') at fine evolutionary time scales (Bartish et al.,

2016; Koenen et al., 2015; Stebbins, 1974). These findings highlight the need for simultaneous consideration of complementary metrics of phylogenetic structure sensitive to either basal or terminal branching to interpret distinct evolutionary patterns and their causes.

Our maps also revealed important hotspots of basal and terminal phylogenetic overdispersion (Fig. 1; scenario D) in the hemiboreal zone in Russia. For broadleaved deciduous forests, these areas extended towards the Baltic countries, Belarus, and Ukraine. The hemiboreal zone represents a transition between the boreal and temperate biogeographic regions (Tuhkanen, 1984), and hosts a mixture of boreal and temperate species, the latter occurring at the northern limit of their ranges (Diekmann, 1994). This biome-transition effect could explain why distant lineages with distinct evolutionary histories co-occur in the same communities. Ecological transition zones in Europe have also been reported to harbor communities with great genetic diversity due to the local co-occurrence of species adapted to different ecological zones (Petit et al., 2003). In our study, vegetation plots in these transitional areas were mainly represented by *Picea* taiga, and *Carpinus* and *Quercus* mesic deciduous forests. Nonetheless, further studies of taxonomic, phylogenetic, and functional turnover in the hemiboreal zone are required to explicitly test for the effect of ecological transition on forest diversity patterns, as well as in other vegetation types, such as grasslands. The patterns shown in our maps could also be enhanced in the future with the addition of more vegetation-plot data in the hemiboreal zone. We also suggest that hemiboreal areas should receive more attention from both the scientific and nature conservation perspectives in the pan-European context.

Conclusions

Our study stresses the role of postglacial dispersal limitation, environmental filtering, and species exchange in biome-transition regions as drivers of the distribution of plant lineages in European

forests. Phylogenetic clustering was important in areas glaciated during the Pleistocene, reflecting limited postglacial recolonization and speciation, and to a smaller extent, in the summer-dry Mediterranean regions, reflecting the effect of strong environmental filtering. Phylogenetic clustering occurred due to contributions of different clades in different regions, highlighting the importance of identifying overrepresented clades in relation to their shared conserved niches and biogeographical history. The hotspots of phylogenetic overdispersion, which we found scattered around the Mediterranean Basin (including the Adriatic Sea), along the Atlantic seaboard of the Iberian Peninsula, and especially in the hemiboreal zone in Russia, might have acted as refugia for the European forest flora, and deserve special focus in future biogeographic studies and conservation planning.

Figures

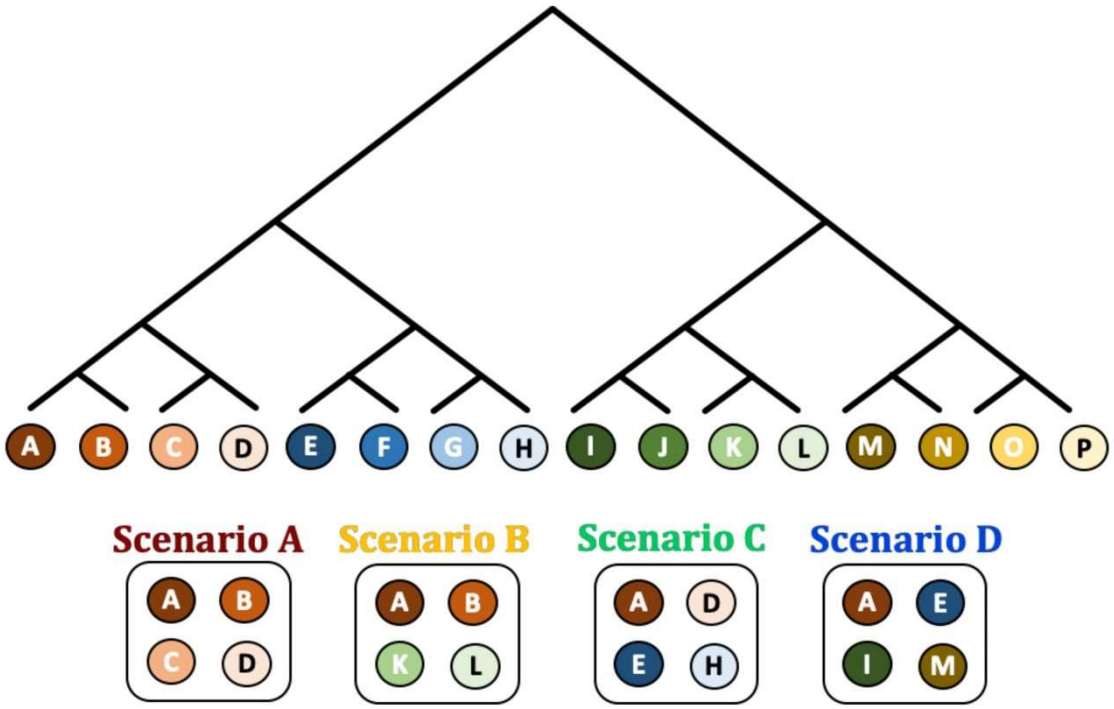


Fig. 1: Schematic representation of the four scenarios of phylogenetic structure. Each scenario is represented by a community with four species taken from the phylogeny. The community in scenario A is clustered at both basal and terminal levels. In contrast, the community in scenario D is overdispersed at both basal and terminal levels. The community in scenario B is overdispersed at the basal level but clustered at the terminal level (i.e., overdispersion of clustering; Mazel et al., 2016). Finally, the community in scenario C is clustered at the basal level but overdispersed at the terminal level (i.e., clustering of overdispersion).

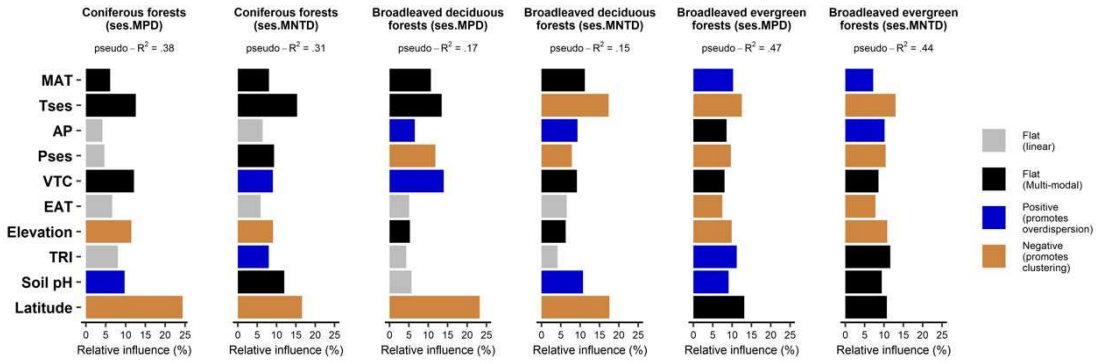


Fig. 2: Relative influence of environmental variables on ses.MPD and ses.MNTD of European coniferous, broadleaved deciduous, and broadleaved evergreen forests. The shapes of the effect of predictors on response variables were estimated from partial dependence plots (Appendix S5: Figs. S5.2-S5.7), and classified as follows: "Flat (linear)" if we did not observe any relationship; "Flat (multimodal)" if we observed a multimodal response with no consistent increase or decrease along the environmental gradient; "Positive" if we observed a linear, uni- or multimodal response with an overall increase along the environmental gradient (i.e., promoting overdispersion); and "Negative" if we observed a linear, uni- or multimodal response with an overall decrease along the environmental gradient (i.e., promoting clustering). We calculated pseudo-R^2 based on the square of the correlation coefficient between observed and predicted species richness. MAT = mean annual temperature; Tses = temperature seasonality, AP = annual precipitation, Pses = precipitation seasonality, VTC = velocity of temperature change since LGM; EAT = extent of analogous temperature; TRI = Terrain Ruggedness Index.

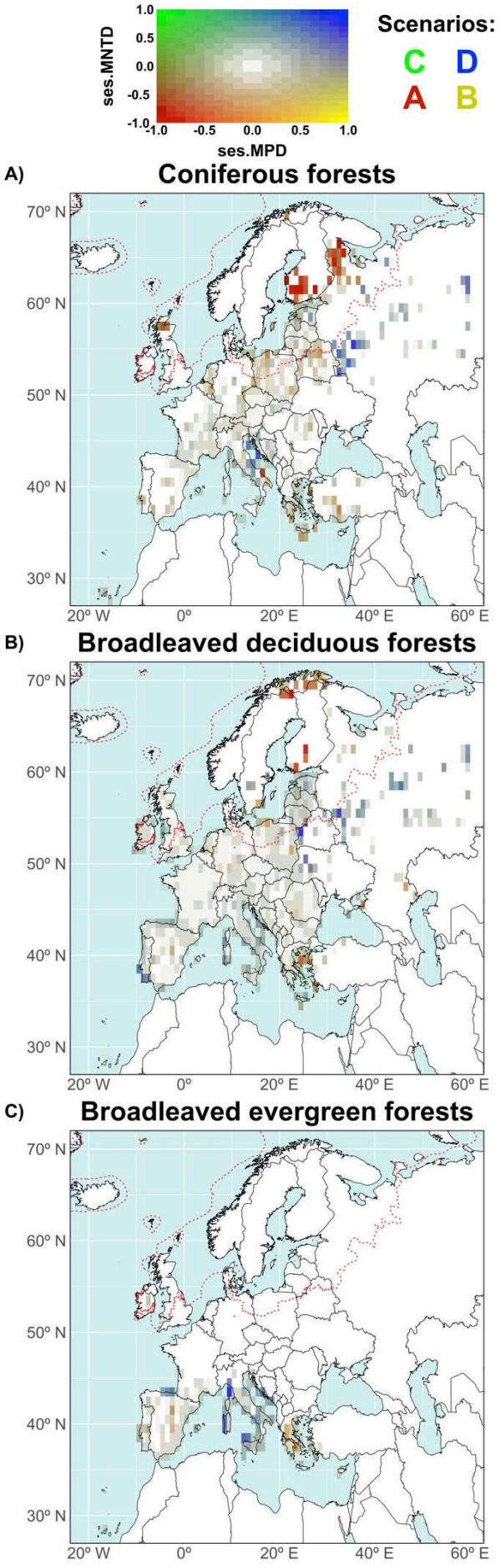

ses.MNTD
ses.MPD
1.0
0.5
0.0
-0.5
-1.0
Scenarios:
C D
A B
A)
Coniferous forests
B)
Broadleaved deciduous forests
C)
Broadleaved evergreen forests
70° N
60° N
50° N
40° N
30° N
20° W
0°
20° E
40° E
60° E

Fig. 3: Phylogenetic structure of European forests. The maps show standardized effect sizes of MPD and MNTD (i.e., ses.MPD and ses.MNTD) for coniferous (A), broadleaved deciduous (B), and broadleaved evergreen (C) forests including only angiosperms. Values of -1 and 1 indicate that all vegetation plots in the given cell were either significantly clustered or overdispersed, respectively, for each metric. Values of 0 indicate that plots in the given cell did not differ from the random expectation for each metric. Colors relate the four scenarios presented in Fig. 1. Disaggregated maps for ses.MPD and ses.MNTD can be found in Appendix S3: Fig. S3.4. The dashed red line indicates the extent of the ice-sheet over 50º N during the LGM.

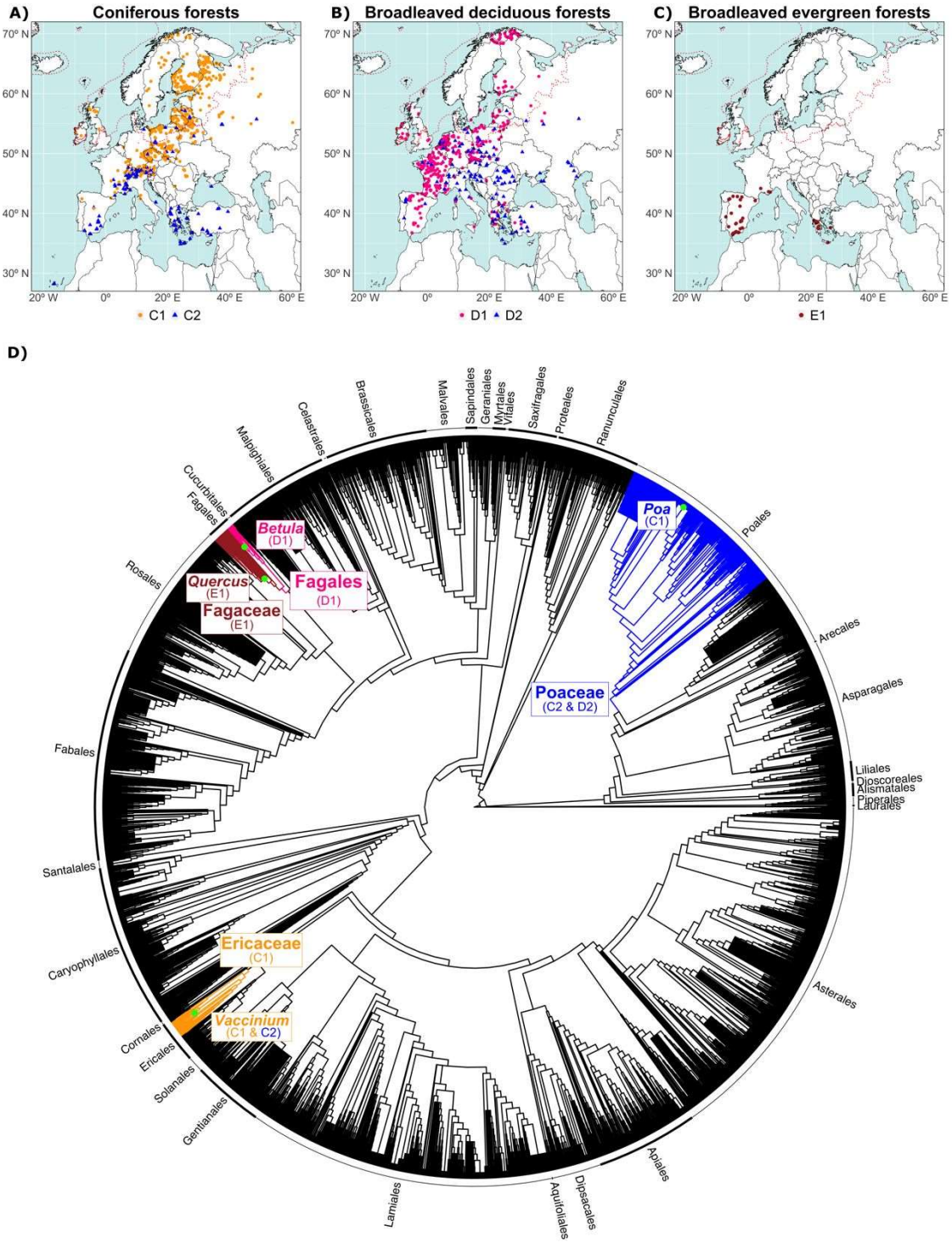


Fig. 4: Phylogenetic clustering of European forest vegetation. Maps show significantly clustered plots for both ses.MPD and ses.MNTD in coniferous (A), broadleaved deciduous (B), and broadleaved evergreen (C) forests. Vegetation plots were grouped based on the relative contribution of each node to the phylogenetic structure, and are represented by different colors

and shapes. For each group, we colored on the phylogeny (D) the clades with the highest randomization rank that also corresponded to families or orders (see Appendix S8: Figs. S8.1-S8.3 for greater detail). Green points indicate the nodes with the highest randomization rank that corresponded to a genus. Letters under taxonomic names indicate the group for which that clade or node had the highest randomization rank. The dashed red line in A), B) and C) indicates the extent of the ice-sheet over 50º N during the LGM.

Data Accessibility Statement

The data are available upon request from the European Vegetation Archive (EVA) (project number 82). See the EVA rules for details at http://euroveg.org/download/eva-rules.pdf.

References

Abellán, P., Carrete, M., Anadón, J. D., Cardador, L., & Tella, J. L. (2016). Non-random patterns and temporal trends (1912-2012) in the transport, introduction and establishment of exotic birds in Spain and Portugal. *Diversity and Distributions*, *22*(3), 263–273. https://doi.org/10.1111/ddi.12403

Bartish, I. V., Ozinga, W. A., Bartish, M. I., Wamelink, G. W. W., Hennekens, S. M., & Prinzing, A. (2016). Different habitats within a region contain evolutionary heritage from different epochs depending on the abiotic environment: Habitats contain evolutionary heritage from different epochs. *Global Ecology and Biogeography*, *25*(3), 274–285. https://doi.org/10.1111/geb.12408

Bennett, K. D., Tzedakis, P. C., & Willis, K. J. (1991). Quaternary Refugia of North European Trees. *Journal of Biogeography*, *18*(1), 103–115. https://doi.org/10.2307/2845248

Bouchenak-Khelladi, Y., Verboom, G. A., Savolainen, V., & Hodkinson, T. R. (2010). Biogeography of the grasses (Poaceae): a phylogenetic approach to reveal evolutionary history in geographical space and geological time. *Botanical Journal of the Linnean Society*, *162*(4), 543–557. https://doi.org/10.1111/j.1095-8339.2010.01041.x

Bradshaw, R. H. W. (2004). Past anthropogenic influence on European forests and some possible genetic consequences. *Forest Ecology and Management*, *197*(1–3), 203–212. https://doi.org/10.1016/j.foreco.2004.05.025

Bruelheide, H., Dengler, J., Jiménez-Alfaro, B., Purschke, O., Hennekens, S. M., Chytrý, M., … Zverev, A. (2019). sPlot – A new tool for global vegetation analyses. *Journal of Vegetation Science*, *30*(2), 161–186. https://doi.org/10.1111/jvs.12710

Cairney, J. W. G., & Meharg, A. A. (2003). Ericoid mycorrhiza: a partnership that exploits harsh edaphic conditions. *European Journal of Soil Science*, *54*(4), 735–740. https://doi.org/10.1046/j.1351-0754.2003.0555.x

Cayuela, L., Stein, A., & Oksanen, J. (2017). Taxonstand: Taxonomic Standardization of Plant Species Names. *R Package Version 2.0*. https://CRAN.R-project.org/package=Taxonstand

Chytrý, M., Hennekens, S. M., Jiménez-Alfaro, B., Knollová, I., Dengler, J., Jansen, F., … Yamalov, S. (2016). European Vegetation Archive (EVA): an integrated database of European vegetation plots. *Applied Vegetation Science*, *19*(1), 173–180. https://doi.org/10.1111/avsc.12191

Chytrý, M., Tichý, L., Hennekens, S. M., Knollová, I., Janssen, J. A. M., Rodwell, J. S., … Schaminée, J. H. J. (2021). EUNIS Habitat Classification: Expert system, characteristic species combinations and distribution maps of European habitats. *Applied Vegetation Science*, *24*. https://doi.org/10.1111/avsc.12519

Clayton, W. D. (1981). Evolution and Distribution of Grasses. *Annals of the Missouri Botanical Garden*, *68*(1), 5–14. https://doi.org/10.2307/2398808

Davies, C. E., Moss, D., & Hill, M. O. (2004). *EUNIS habitat classification*. Paris: European Topic Centre on Nature Protection and Biodiversity.

De Cáceres, M., Font, X., & Oliva, F. (2010). The management of vegetation classifications with fuzzy clustering: Fuzzy clustering in vegetation classifications. *Journal of Vegetation Science*, *21*(6), 1138–1151. https://doi.org/10.1111/j.1654-1103.2010.01211.x

De'ath, G. (2007). Boosted trees for ecological modeling and prediction. *Ecology*, *88*(1), 243–251. https://doi.org/10.1890/0012-9658(2007)88[243:BTFEMA]2.0.CO;2

Diekmann, M. (1994). Deciduous forest vegetation in Boreo-nemoral Scandinavia. *Acta Phytogeographica Suecica*, *80,* 1–116.

Eiserhardt, W. L., Borchsenius, F., Plum, C. M., Ordonez, A., & Svenning, J.-C. (2015). Climate-driven extinctions shape the phylogenetic structure of temperate tree floras. *Ecology Letters*, *18*(3), 263–272. https://doi.org/10.1111/ele.12409

Elith, J., Leathwick, J. R., & Hastie, T. (2008). A working guide to boosted regression trees. *Journal of Animal Ecology*, *77*(4), 802–813. https://doi.org/10.1111/j.1365-2656.2008.01390.x

Erdős, L., Ambarlı, D., Anenkhonov, O. A., Bátori, Z., Cserhalmi, D., Kiss, M., … Török, P. (2018). The edge of two worlds: A new review and synthesis on Eurasian forest-steppes. *Applied Vegetation Science*, *21*(3), 345–362. https://doi.org/10.1111/avsc.12382

European Environment Agency. (2016). *European forest ecosystems: state and trends.* Luxembourg: Publications Office of the European Union. http://bookshop.europa.eu/uri?target=EUB:NOTICE:THAL16005:EN:HTML

Faith, D. P. (1992). Conservation evaluation and phylogenetic diversity. *Biological Conservation*, *61*(1), 1–10. https://doi.org/10.1016/0006-3207(92)91201-3

Feng, G., Mi, X. C., Bøcher, P. K., Mao, L. F., Sandel, B., Cao, M., … Svenning, J.-C. (2014). Relative roles of local disturbance, current climate and paleoclimate in determining phylogenetic and functional diversity in Chinese forests. *Biogeosciences*, *11*(5), 1361–1370. https://doi.org/10.5194/bg-11-1361-2014

Fick, S. E., & Hijmans, R. J. (2017). WorldClim 2: new 1-km spatial resolution climate surfaces for global land areas. *International Journal of Climatology*, *37*(12), 4302–4315. https://doi.org/10.1002/joc.5086

Gilliam, F. S. (2007). The ecological significance of the herbaceous layer in temperate forest ecosystems. *BioScience*, *57*(10), 845–858. https://doi.org/10.1641/B571007

Götzenberger, L., de Bello, F., Bråthen, K. A., Davison, J., Dubuis, A., Guisan, A., … Zobel, M. (2012). Ecological assembly rules in plant communities–approaches, patterns and

prospects. *Biological Reviews*, *87*(1), 111–127. https://doi.org/10.1111/j.1469-185X.2011.00187.x

Graham, C. H., Parra, J. L., Rahbek, C., & McGuire, J. A. (2009). Phylogenetic structure in tropical hummingbird communities. *Proceedings of the National Academy of Sciences*, *106*, 19673–19678. https://doi.org/10.1073/pnas.0901649106

Guevara, J. E., Damasco, G., Baraloto, C., Fine, P. V. A., Peñuela, M. C., Castilho, C., … ter Steege, H. (2016). Low phylogenetic beta diversity and geographic neo-endemism in Amazonian white-sand forests. *Biotropica*, *48*(1), 34–46. https://doi.org/10.1111/btp.12298

Hardy, O. J., & Senterre, B. (2007). Characterizing the phylogenetic structure of communities by an additive partitioning of phylogenetic diversity. *Journal of Ecology*, *95*(3), 493–506. https://doi.org/10.1111/j.1365-2745.2007.01222.x

Hawkins, B. A., Rueda, M., Rangel, T. F., Field, R., & Diniz-Filho, J. A. F. (2014). Community phylogenetics at the biogeographical scale: cold tolerance, niche conservatism and the structure of North American forests. *Journal of Biogeography*, *41*(1), 23–38. https://doi.org/10.1111/jbi.12171

Hengl, T., Mendes de Jesus, J., Heuvelink, G. B., Ruiperez Gonzalez, M., Kilibarda, M., Blagotić, A., … Kempen, B. (2017). SoilGrids250m: Global gridded soil information based on machine learning. *PLOS ONE*, *12*(2), e0169748.

Hijmans, R. (2019). *raster: Geographic Data Analysis and Modeling*. R package version 2.9-5. https://CRAN.R-project.org/package=raster

Hijmans, R. J., Phillips, S., Leathwick, J., & Elith, J. (2017). *dismo: Species Distribution Modeling*. R package version 1.1-4. https://CRAN.R-project.org/package=dismo

Jablonski, D., Roy, K., & Valentine, J. W. (2006). Out of the tropics: Evolutionary dynamics of the latitudinal diversity gradient. *Science*, *314*(5796), 102–106. https://doi.org/10.1126/science.1130880

Jiménez-Alfaro, B., Girardello, M., Chytrý, M., Svenning, J.-C., Willner, W., Gégout, J.-C., … Wohlgemuth, T. (2018). History and environment shape species pools and community diversity in European beech forests. *Nature Ecology & Evolution*, *2*(3), 483–490. https://doi.org/10.1038/s41559-017-0462-6

Jin, Y., & Qian, H. (2019). V.PhyloMaker: an R package that can generate very large phylogenies for vascular plants. *Ecography*, *42*(8), 1353–1359. https://doi.org/10.1111/ecog.04434

Kissling, W. D., Eiserhardt, W. L., Baker, W. J., Borchsenius, F., Couvreur, T. L. P., Balslev, H., & Svenning, J.-C. (2012). Cenozoic imprints on the phylogenetic structure of palm species assemblages worldwide. *Proceedings of the National Academy of Sciences*, *109*(19), 7379–7384. https://doi.org/10.1073/pnas.1120467109

Koenen, E. J. M., Clarkson, J. J., Pennington, T. D., & Chatrou, L. W. (2015). Recently evolved diversity and convergent radiations of rainforest mahoganies (Meliaceae) shed new light on the origins of rainforest hyperdiversity. *New Phytologist*, *207*(2), 327–339. https://doi.org/10.1111/nph.13490

Kooyman, R., Rossetto, M., Cornwell, W., & Westoby, M. (2011). Phylogenetic tests of community assembly across regional to continental scales in tropical and subtropical rain forests. *Global Ecology and Biogeography*, *20*(5), 707–716. https://doi.org/10.1111/j.1466-8238.2010.00641.x

Kraft, N. J. B., & Ackerly, D. D. (2014). Assembly of plant communities. In R. K. Monson (Ed.), *The plant sciences: Ecology and the environment* (Vol. 8, pp. 67–88). Springer Nature.

Kubiskem, M. E., & Abrams, M. D. (1993). Stomatal and nonstomatal limitations of photosynthesis in 19 temperate tree species on contrasting sites during wet and dry years. *Plant, Cell and Environment*, *16*(9), 1123–1129. https://doi.org/10.1111/j.1365-3040.1996.tb02070.x

Kuster, T. M., Arend, M., Günthardt-Goerg, M. S., & Schulin, R. (2013). Root growth of different oak provenances in two soils under drought stress and air warming conditions. *Plant and Soil*, *369*(1–2), 61–71. https://doi.org/10.1007/s11104-012-1541-8

Legendre, P., Borcard, D., & Peres-Neto, P. R. (2005). Analyzing beta diversity: partitioning the spatial variation of community composition data. *Ecological Monographs*, *75*(4), 435–450. https://doi.org/10.1890/05-0549

Lengyel, A., Chytrý, M., & Tichý, L. (2011). Heterogeneity-constrained random resampling of phytosociological databases. *Journal of Vegetation Science*, *22*(1), 175–183. https://doi.org/10.1111/j.1654-1103.2010.01225.x

Leuschner, C., Backes, K., Hertel, D., Schipka, F., Schmitt, U., Terborg, O., & Runge, M. (2001). Drought responses at leaf, stem and fine root levels of competitive *Fagus sylvatica* L. and *Quercus petraea* (Matt.) Liebl. trees in dry and wet years. *Forest Ecology and Management*, *149*(1/3), 33–46.

Lososová, Z., Divíšek, J., Chytrý, M., Götzenberger, L., Těšitel, J., & Mucina, L. (2021). Macroevolutionary patterns in European vegetation. *Journal of Vegetation Science*, *32*. https://doi.org/10.1111/jvs.12942

Lososová, Z., Šmarda, P., Chytrý, M., Purschke, O., Pyšek, P., Sádlo, J., … Winter, M. (2015). Phylogenetic structure of plant species pools reflects habitat age on the geological time scale. *Journal of Vegetation Science*, *26*(6), 1080–1089. https://doi.org/10.1111/jvs.12308

Lu, L.-M., Mao, L.-F., Yang, T., Ye, J.-F., Liu, B., Li, H.-L., … Chen, Z.-D. (2018). Evolutionary history of the angiosperm flora of China. *Nature*, *554*(7691), 234–238. https://doi.org/10.1038/nature25485

Ma, Z., Sandel, B., & Svenning, J.-C. (2016). Phylogenetic assemblage structure of North American trees is more strongly shaped by glacial-interglacial climate variability in gymnosperms than in angiosperms. *Ecology and Evolution*, *6*(10), 3092–3106. https://doi.org/10.1002/ece3.2100

Magri, D., Vendramin, G. G., Comps, B., Dupanloup, I., Geburek, T., Gomory, D., … de Beaulieu, J.-L. (2006). A new scenario for the Quaternary history of European beech populations: palaeobotanical evidence and genetic consequences. *New Phytologist*, *171*(1), 199–221. https://doi.org/10.1111/j.1469-8137.2006.01740.x

Massante, J. C., Götzenberger, L., Takkis, K., Hallikma, T., Kaasik, A., Laanisto, L., … Gerhold, P. (2019). Contrasting latitudinal patterns in phylogenetic diversity between woody and herbaceous communities. *Scientific Reports*, *9*(1), 6443. https://doi.org/10.1038/s41598-019-42827-1

Mastrogianni, A., Kallimanis, A. S., Chytrý, M., & Tsiripidis, I. (2019). Phylogenetic diversity patterns in forests of a putative refugial area in Greece: A community level analysis. *Forest Ecology and Management*, *446*, 226–237. https://doi.org/10.1016/j.foreco.2019.05.044

Mazel, F., Davies, T. J., Gallien, L., Renaud, J., Groussin, M., Münkemüller, T., & Thuiller, W. (2016). Influence of tree shape and evolutionary time-scale on phylogenetic diversity metrics. *Ecography*, *39*(10), 913–920. https://doi.org/10.1111/ecog.01694

Médail, F., & Diadema, K. (2009). Glacial refugia influence plant diversity patterns in the Mediterranean Basin. *Journal of Biogeography*, *36*(7), 1333–1345. https://doi.org/10.1111/j.1365-2699.2008.02051.x

Mishler, B. D., Knerr, N., González-Orozco, C. E., Thornhill, A. H., Laffan, S. W., & Miller, J. T. (2014). Phylogenetic measures of biodiversity and neo- and paleo-endemism in Australian *Acacia*. *Nature Communications*, *5*(1), 4473. https://doi.org/10.1038/ncomms5473

Ohlemüller, R., Anderson, B. J., Araújo, M. B., Butchart, S. H. M., Kudrna, O., Ridgely, R. S., & Thomas, C. D. (2008). The coincidence of climatic and species rarity: high risk to small-range species from climate change. *Biology Letters*, *4*(5), 568–572. https://doi.org/10.1098/rsbl.2008.0097

Oksanen, J., Blanchet, F. G., Friendly, M., Kindt, R., Legendre, P., McGlinn, D., … Wagner, H. (2019). vegan: Community Ecology Package. *R Package Version 2.5-6*. https://CRAN.R-project.org/package=vegan

Pavoine, S., & Bonsall, M. B. (2011). Measuring biodiversity to explain community assembly: a unified approach. *Biological Reviews*, *86*(4), 792–812. https://doi.org/10.1111/j.1469-185X.2010.00171.x

Peltier, W. R., Argus, D. F., & Drummond, R. (2015). Space geodesy constrains ice age terminal deglaciation: The global ICE-6G_C (VM5a) model. *Journal of Geophysical Research: Solid Earth*, *120*(1), 450–487. https://doi.org/10.1002/2014JB011176

Petit, R. J., Aguinagalde, I., de Beaulieu, J.-L., Bittkau, C., Brewer, S., Cheddadi, R., … Vendramin, G. G. (2003). Glacial refugia: Hotspots but not melting pots of genetic diversity. *Science*, *300*(5625), 1563–1565. https://doi.org/10.1126/science.1083264

Qian, H., Deng, T., Jin, Y., Mao, L., Zhao, D., & Ricklefs, R. E. (2019). Phylogenetic dispersion and diversity in regional assemblages of seed plants in China. *Proceedings of the National Academy of Sciences*, *116*(46), 23192–23201. https://doi.org/10.1073/pnas.1822153116

Qian, H., & Jin, Y. (2016). An updated megaphylogeny of plants, a tool for generating plant phylogenies and an analysis of phylogenetic community structure. *Journal of Plant Ecology*, *9*(2), 233–239. https://doi.org/10.1093/jpe/rtv047

Qian, H., & Sandel, B. (2017). Phylogenetic structure of regional angiosperm assemblages across latitudinal and climatic gradients in North America. *Global Ecology and Biogeography*, *26*(11), 1258–1269. https://doi.org/10.1111/geb.12634

Qian, H., Zhang, J., Sandel, B., & Jin, Y. (2019). Phylogenetic structure of angiosperm trees in local forest communities along latitudinal and elevational gradients in eastern North America. *Ecography*, *43*(3), 419–430. https://doi.org/10.1111/ecog.04873

Qian, H., Zhang, Y., Zhang, J., & Wang, X. (2013). Latitudinal gradients in phylogenetic relatedness of angiosperm trees in North America. *Global Ecology and Biogeography*, *22*(11), 1183–1191. https://doi.org/10.1111/geb.12069

R Core Team. (2019). *R: A language and environment for statistical computing.* R Foundation for Statistical Computing. https://www.R-project.org/

Ricklefs, R. E. (2004). A comprehensive framework for global patterns in biodiversity. *Ecology Letters*, *7*(1), 1–15. https://doi.org/10.1046/j.1461-0248.2003.00554.x

Sandel, B., Arge, L., Dalsgaard, B., Davies, R. G., Gaston, K. J., Sutherland, W. J., & Svenning, J.-C. (2011). The influence of Late Quaternary climate-change velocity on species endemism. *Science*, *334*(6056), 660–664. https://doi.org/10.1126/science.1210173

Smith, S. A., & Brown, J. W. (2018). Constructing a broadly inclusive seed plant phylogeny. *American Journal of Botany, 105*(3), 302–314. https://doi.org/10.1002/ajb2.1019

Stebbins, G. L. (1974). *Flowering plants: Evolution above the species level*. The Belknap Press of Harvard Univ. Press. https:// doi.org/10.4159/harvard.9780674864856

Svenning, J.-C., Normand, S., & Skov, F. (2008). Postglacial dispersal limitation of widespread forest plant species in nemoral Europe. *Ecography*, *31*(3), 316–326. https://doi.org/10.1111/j.0906-7590.2008.05206.x

Svenning, J.-C., Normand, S., & Skov, F. (2009). Plio-Pleistocene climate change and geographic heterogeneity in plant diversity-environment relationships. *Ecography*, *32*(1), 13–21. https://doi.org/10.1111/j.1600-0587.2008.05732.x

Thornhill, A. H., Mishler, B. D., Knerr, N. J., González-Orozco, C. E., Costion, C. M., Crayn, D. M., … Miller, J. T. (2016). Continental-scale spatial phylogenetics of Australian angiosperms provides insights into ecology, evolution and conservation. *Journal of Biogeography*, *43*(11), 2085–2098. https://doi.org/10.1111/jbi.12797

Tichý, L. (2002). JUICE, software for vegetation classification. *Journal of Vegetation Science*, *13*(3), 451–453. https://doi.org/10.1111/j.1654-1103.2002.tb02069.x

Tsirogiannis, C., & Sandel, B. (2016). PhyloMeasures: a package for computing phylogenetic biodiversity measures and their statistical moments. *Ecography*, *39*(7), 709–714. https://doi.org/10.1111/ecog.01814

Tsirogiannis, C., & Sandel, B. (2017). *PhyloMeasures: Fast and Exact Algorithms for Computing Phylogenetic Biodiversity Measures*. R package version 2.1. https://CRAN.R-project.org/package=PhyloMeasures

Tuhkanen, S. (1984). A circumboreal system of climatic-phytogeographical regions. *Acta Botanica Fennica*, *127*, 1–50.

Tzedakis, P. C., Emerson, B. C., & Hewitt, G. M. (2013). Cryptic or mystic? Glacial tree refugia in northern Europe. *Trends in Ecology & Evolution*, *28*(12), 696–704. https://doi.org/10.1016/j.tree.2013.09.001

Večeřa, M., Divíšek, J., Lenoir, J., Jiménez-Alfaro, B., Biurrun, I., Knollová, I., … Chytrý, M. (2019). Alpha diversity of vascular plants in European forests. *Journal of Biogeography*, *46*(9), 1919–1935. https://doi.org/10.1111/jbi.13624

Webb, C. O., Ackerly, D. D., McPeek, M. A., & Donoghue, M. J. (2002). Phylogenies and community ecology. *Annual Review of Ecology and Systematics*, *3*, 475–505.

Weigelt, P., Kissling, W. D., Kisel, Y., Fritz, S. A., Karger, D. N., Kessler, M., …, Kreft, H. (2015). Global patterns and drivers of phylogenetic structure in island floras. *Scientific Reports*, *5*(1), 12213. https://doi.org/10.1038/srep12213

Whitford, W. G. (2002). *Ecology of desert systems.* San Diego, CA: Academic Press, an Elsevier Science Imprint.

Willner, W., Di Pietro, R., & Bergmeier, E. (2009). Phytogeographical evidence for post-glacial dispersal limitation of European beech forest species. *Ecography*, *32*(6), 1011–1018. https://doi.org/10.1111/j.1600-0587.2009.05957.x

Wiser, S. K., & De Cáceres, M. (2013). Updating vegetation classifications: an example with New Zealand's woody vegetation. *Journal of Vegetation Science*, *24*(1), 80–93. https://doi.org/10.1111/j.1654-1103.2012.01450.x

Zanne, A. E., Tank, D. C., Cornwell, W. K., Eastman, J. M., Smith, S. A., FitzJohn, R. G., … Beaulieu, J. M. (2014). Three keys to the radiation of angiosperms into freezing environments. *Nature*, *506*(7486), 89–92. https://doi.org/10.1038/nature12872

Biosketch

Josep Padullés Cubino is a postdoctoral research fellow interested in understanding the ecological and evolutionary determinants of forest plant diversity and distribution. He uses analysis of evolutionary relationships among plant species and plant functional traits to understand processes that shape ecosystems and the effects of human actions on these processes. Details about his research interests can be found at www.researchgate.net/profile/Josep_Padulles.

XSUPPORTING INFORMATION

Title: Phylogenetic structure of European forest vegetation

Journal: Journal of Biogeography

Table of contents

Appendix S1: Overview of the datasets included in this study

Table S1.1: Overview of the datasets included in our study. For each dataset, we provide the number and proportion of plots over the total after resampling procedure (N = 61,816).

Database	Database custodian	Number of plots	% of plots
French National Forest Inventory	-	15,087	24.4
European Boreal Forest Vegetation Database	Anni Kanerva	4,208	6.8
Polish Vegetation Database	Zygmunt Kącki	3,474	5.6
Vegetation Plot Database - Sapienza University of Rome	Emiliano Agrillo	3,188	5.2
Czech National Phytosociological Database	Milan Chytrý	3,147	5.1
Romanian Forest Database	Adrian Indreica	2,578	4.2
Hellenic Natura 2000 Vegetation Database (HelNatVeg)	Panayotis Dimopoulos	2,019	3.3
Austrian Vegetation Database	Wolfgang Willner	1,998	3.2
Swiss Forest Vegetation Database	Thomas Wohlgemuth	1,956	3.2
Iberian and Macaronesian Vegetation Information System (SIVIM) – Sclerophyllous forests	Federico Fernández-González	1,846	3.0
Slovak Vegetation Database	Milan Valachovič	1,588	2.6
German Vegetation Reference Database (GVRD)	Ute Jandt	1,346	2.2
Irish Vegetation Database	Úna FitzPatrick	1,341	2.2
EcoPlant	Jean-Claude Gégout	1,183	1.9
Iberian and Macaronesian Vegetation Information System (SIVIM) – Floodplain forests	Idoia Biurrun	1,156	1.9
SOPHY	Henry Brisse	1,095	1.8
Balkan Vegetation Database	Kiril Vassilev	1,038	1.7
Croatian Vegetation Database	Željko Škvorc	983	1.6
Dutch National Vegetation Database	Stephan Hennekens	959	1.6
CircumMed Pine Forest Database	Gianmaria Bonari	958	1.5
Iberian and Macaronesian Vegetation Information System (SIVIM) – Deciduous Forests	Juan Antonio Campos	927	1.5
Lithuanian vegetation Database	Valerius Rašomavičius	917	1.5
Vegetation database of Habitats in the Italian Alps - HabItAlp	Laura Casella	781	1.3

Vegetation-Plot Database of the University of the Basque Country (BIOVEG)	Idoia Biurrun	670	1.1
Vegetation Database of Slovenia	Urban Šilc	667	1.1
SE Europe forest database	Andraž Čarni	597	1.0
UK National Vegetation Classification Database	John S. Rodwell	583	0.9
Forest Database of Southern Poland	Remigiusz Pielech	554	0.9
The Nordic Vegetation Database	Jonathan Lenoir	551	0.9
VegetWeb Germany	Florian Jansen	507	0.8
INBOVEG	Els De Bie	433	0.7
Temperate Forests of European Russia	Larisa Khanina	352	0.6
Iberian and Macaronesian Vegetation Information System (SIVIM)	Xavier Font	345	0.6
Vegetation Database of Tatarstan	Vadim Prokhorov	324	0.5
VegItaly	Roberto Venanzoni	320	0.5
VegMV	Florian Jansen	289	0.5
Vegetation Database of Ukraine and Adjacent Parts of Russia	Viktor Onyshchenko	267	0.4
European Mire Vegetation Database	Tomáš Peterka	217	0.4
KRITI	Erwin Bergmeier	201	0.3
Database of Forest Vegetation in Republic of Serbia + Vegetation Database of Northern Part of Serbia (AP Vojvodina)	Mirjana Krstivojević Ćuk	148	0.2
CoenoDat Hungarian Phytosociological Database	János Csiky	127	0.2
Vegetation Database of Oak Communities in Turkey	Emin Uğurlu	127	0.2
Masaryk University Database	Milan Chytrý	113	0.2
Iberian and Macaronesian Vegetation Information System (SIVIM) - Macaronesia	Borja Jiménez-Alfaro	105	0.2
Vegetation Database Forest of Southern Ural	Pavel Shirokikh	99	0.2
Hellenic Woodland Database + Hellenic Beech Forests Database (Hell-Beech-DB)	Ioannis Tsiripidis	92	0.1
Dutch Military Ranges Vegetation Database (DUMIRA)	Iris de Ronde	85	0.1
VegetWeb Germany	Friedemann Goral	56	0.1
Semi-natural Grassland Vegetation Database of Latvia	Solvita Rūsiņa	56	0.1
Lower Volga Valley Phytosociological Database	Valentin Golub	56	0.1
Vegetation Database of Albania	Michele De Sanctis	44	0.1
Iberian and Macaronesian Vegetation Information System (SIVIM) - Catalonia	Xavier Font	43	0.1

Database Schleswig-Holstein (Northern Germany)	Joachim Schrautzer	11	0.0
VegetWeb Germany	Friedemann Goral	4	0.0

Table S1.2: Number of plots included in our final dataset (N = 61,816) belonging to particular habitat types of the revised EUNIS classification (Chytrý et al., 2020).

Code	Name	Number of plots
T1	Broadleaved deciduous forests	19,397
T11	Temperate and boreal *Salix* and *Populus* riparian forest	739
T12	Riparian *Alnus* forest	2,199
T13	Temperate and boreal hardwood riparian forest	450
T14	Mediterranean and Macaronesian riparian forest	712
T15	Broadleaved swamp forest on non-acid peat	880
T16	Broadleaved swamp forest on acid peat	555
T17	*Fagus* forest on non-acid soils	4,577
T18	*Fagus* forest on acid soils	974
T19	Temperate and submediterranean thermophilous deciduous forest	2,430
T1A	Mediterranean thermophilous deciduous forest	65
T1B	Acidophilous *Quercus* forest	2,458
T1C	Boreal-nemoral mountain *Betula* forest on mineral soils	95
T1D	Mediterranean and sub-Mediterranean mountain *Betula* and *Populus tremula* forests on mineral soils	91
T1E	*Carpinus* and *Quercus* mesic deciduous forest	3,490
T1F	Ravine forest	2,064
T1G	Non-riverine *Alnus* forest	22
T2	Broadleaved evergreen forests	508
T21	Mediterranean evergreen *Quercus* forest	3,483
T22	Mainland laurophyllous forest	65
T23	Macaronesian laurophyllous forest	41
T24	*Olea europaea -Ceratonia siliqua* forest	34
T25	South-Aegean *Phoenix* groves	8
T26	Canarian *Phoenix* groves	1
T27	Anatolian *Phoenix theophrasti* groves	5
T28	*Ilex aquifolium* forest	91
T3	Coniferous forests	4,911
T31	Temperate mountain *Picea* forest	1,612
T32	Temperate mountain *Abies* forest	1,600
T33	Mediterranean mountain *Abies* forest	171
T34	Temperate subalpine *Larix, Pinus cembra* and *Pinus uncinata* forests	460
T35	Temperate continental *Pinus sylvestris* forest	1,657
T36	Temperate and submediterranean montane *Pinus sylvestris-nigra* forest	820
T37	Mediterranean montane *Pinus sylvestris-nigra* forest	460

T38	Mediterranean and Balkan subalpine *Pinus heldreichii-peuce* forest	43
T39	Mediterranean lowland to submontane *Pinus* forest	1,541
T3A	*Pinus canariensis* forest	38
T3B	Mediterranean montane *Cedrus* forest	5
T3C	*Taxus baccata* forest	29
T3D	Mediterranean Cupressaceae forest	114
T3E	Macaronesian *Juniperus* forest	9
T3F	*Picea* taiga forest	914
T3G	*Pinus sylvestris* taiga forest	1,081
T3H	*Larix* taiga forest	9
T3J	*Pinus* bog forest	604
T3K	*Picea* mire forest	304

References

Chytrý, M., Tichý, L., Hennekens, S. M., Knollová, I., Janssen, J. A. M., Rodwell, J. S., … Schaminée, J. H. J. (2020). EUNIS Habitat Classification: Expert system, characteristic species combinations and distribution maps of European habitats. *Applied Vegetation Science*, avsc.12519. https://doi.org/10.1111/avsc.12519

Appendix S2: Correlations between variables

Table S2.1: Pearson's correlation matrix between ses.PD, ses.MPD, and ses.MNTD for each forest type (i.e., coniferous, and broadleaved deciduous and evergreen forests). All correlations were significant at $P < 0.001$, but P-values were greatly influenced by our large sample sizes.

	Coniferous forests			Broadleaved deciduous forests			Broadleaved evergreen forests		
	ses. PD	**ses. MPD**	**ses. MNTD**	**ses. PD**	**ses. MPD**	**ses. MNTD**	**ses. PD**	**ses. MPD**	**ses. MNTD**
ses.MPD	0.81			0.72			0.83		
ses.MNTD	0.95	0.66		0.95	0.58		0.96	0.72	

Table S2.2: Pearson's correlation matrix between predictor variables in coniferous forests. All correlations were significant at $P < 0.001$, but *P*-values were greatly influenced by our large sample size (n=16,382). Predictor variables are 'mean annual temperature (MAT)', 'temperature seasonality (Tses)', 'annual precipitation (AP)', 'precipitation seasonality (Pses)', 'velocity of temperature change since LGM (VTC)', 'extent of analogous temperature (EAT)', 'elevation', 'Terrain Ruggedness Index (TRI)', 'Soil pH', and 'Latitude'.

Predictors	**Tses**	**AP**	**Pses**	**VTC**	**EAT**	**Elevation**	**TRI**	**Soil pH**	**Latitude**
MAT	-0.52	-0.14	0.20	-0.37	-0.12	-0.21	-0.03	0.69	-0.63
Tses		-0.46	0.22	0.63	0.60	-0.45	-0.39	-0.41	0.71
AP			-0.42	-0.44	-0.54	0.54	0.46	-0.04	-0.29
Pses				0.04	0.02	-0.08	0.08	0.11	-0.14
VTC					0.79	-0.67	-0.60	-0.48	0.86
EAT						-0.81	-0.68	-0.36	0.73
Elevation							0.74	0.17	-0.58
TRI								0.22	-0.52
Soil pH									-0.67

Table S2.3: Pearson's correlation matrix between predictor variables in broadleaved deciduous forests. All correlations were significant at $P < 0.001$, but P-values were greatly influenced by our large sample size (n=41,198). Predictor variables are 'mean annual temperature (MAT)', 'temperature seasonality (Tses)', 'annual precipitation (AP)', 'precipitation seasonality (Pses)', 'velocity of temperature change since LGM (VTC)', 'extent of analogous temperature (EAT)', 'elevation', 'Terrain Ruggedness Index (TRI)', 'Soil pH', and 'Latitude'.

Predictors	**Tses**	**AP**	**Pses**	**VTC**	**EAT**	**Elevation**	**TRI**	**Soil pH**	**Latitude**
MAT	-0.51	-0.06	0.00	-0.51	0.02	-0.27	-0.10	0.57	-0.66
Tses		-0.40	0.32	0.39	-0.07	0.02	0.00	-0.08	0.24
AP			-0.24	-0.28	-0.30	0.33	0.34	-0.24	-0.12
Pses				0.12	-0.24	0.13	0.17	-0.02	-0.16
VTC					0.48	-0.43	-0.41	-0.39	0.81
EAT						-0.79	-0.67	-0.13	0.58
Elevation							0.73	0.01	-0.47
TRI								0.05	-0.43
Soil pH									-0.53

Table S2.4: Pearson's correlation matrix between predictor variables in broadleaved evergreen forests. All correlations were significant at $P < 0.001$, but P-values were greatly influenced by our large sample size (n=4,236). Predictor variables are 'mean annual temperature (MAT)', 'temperature seasonality (Tses)', 'annual precipitation (AP)', 'precipitation seasonality (Pses)', 'velocity of temperature change since LGM (VTC)', 'extent of analogous temperature (EAT)', 'elevation', 'Terrain Ruggedness Index (TRI)', 'Soil pH', and 'Latitude'.

Predictors	**Tses**	**AP**	**Pses**	**VTC**	**EAT**	**Elevation**	**TRI**	**Soil pH**	**Latitude**
MAT	-0.14	-0.34	0.55	-0.02	0.32	-0.65	-0.14	0.27	-0.43
Tses		-0.27	-0.10	0.03	-0.36	0.26	0.11	0.50	-0.05
AP			-0.09	0.06	-0.07	0.01	0.08	-0.61	0.37
Pses				-0.40	-0.05	0.06	0.36	-0.04	-0.80
VTC					0.32	-0.40	-0.51	0.12	0.62
EAT						-0.51	-0.38	-0.11	0.23
Elevation							0.49	-0.12	-0.33
TRI								-0.14	-0.43
Soil pH									-0.12

Appendix S3: Additional results and maps of phylogenetic diversity and structure

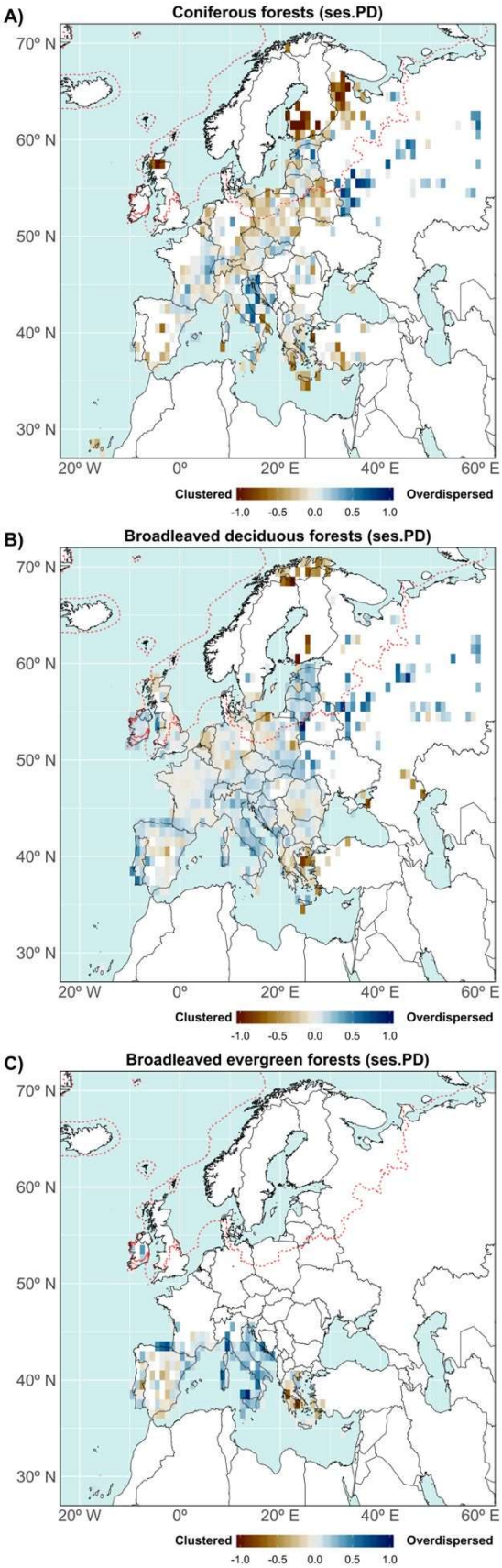


Fig. S3.1: Standardized effect sizes of Faith's PD (ses.PD) for coniferous (A), broadleaved deciduous (B), and broadleaved evergreen (C) forests including only angiosperms. Values of -1 and 1 indicate that all plots in the given cell were either significantly clustered or overdispersed, respectively. Values of 0 indicate that plots in the given cell did not differ from the random expectation for each metric. The dashed red line indicates the extent of the ice-sheet over 50º N during the LGM. We calculated ses.PD for each community following the same approach as described for ses.MPD and ses.MNTD in the Methods section.

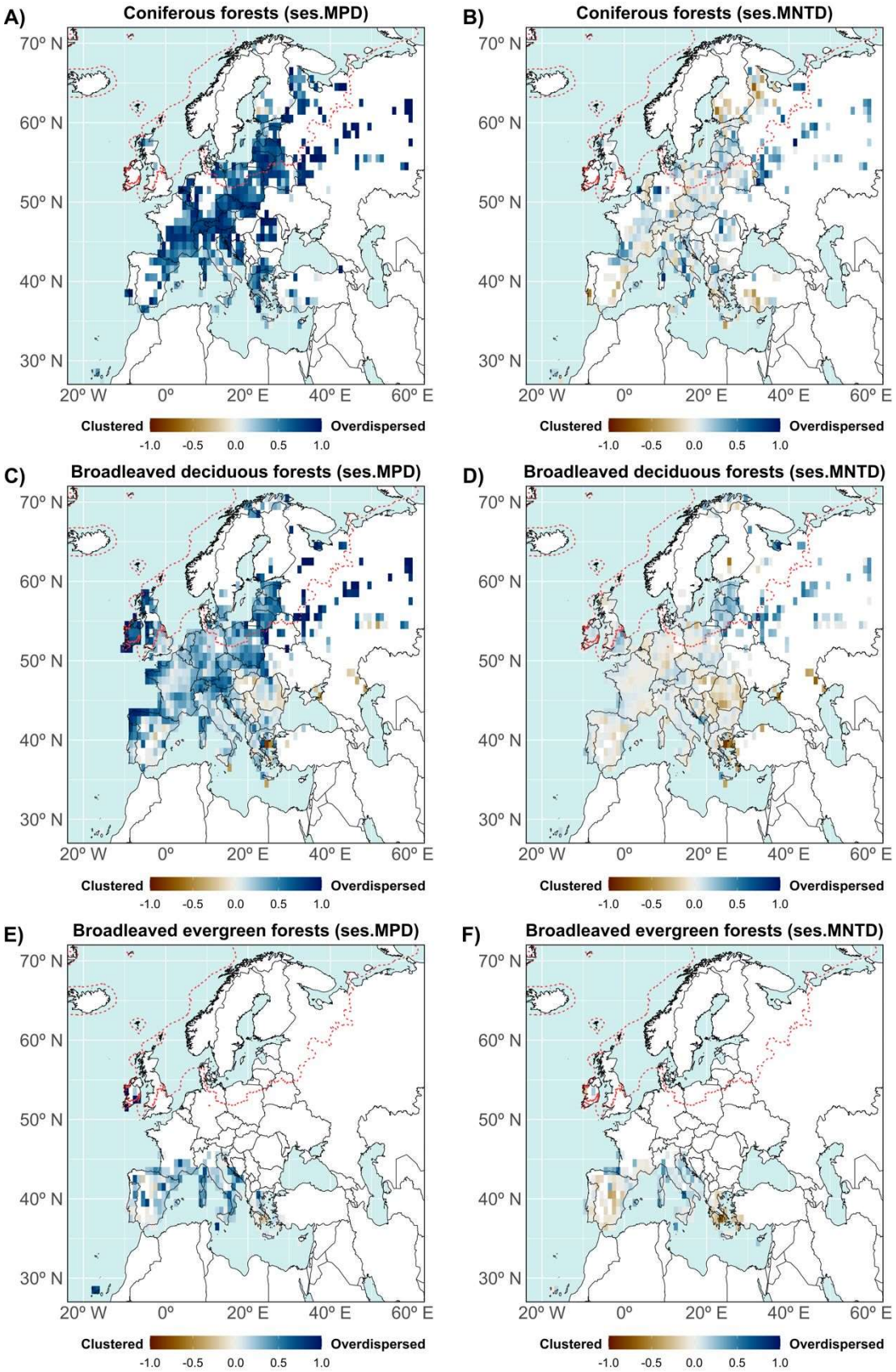

A)
Coniferous forests (ses.MPD)
B)
Coniferous forests (ses.MNTD)
C)
Broadleaved deciduous forests (ses.MPD)
D)
Broadleaved deciduous forests (ses.MNTD)
E)
Broadleaved evergreen forests (ses.MPD)
F)
Broadleaved evergreen forests (ses.MNTD)
70° N
60° N
50° N
40° N
30° N
20° W
0°
20° E
40° E
60° E
Clustered
Overdispersed
-1.0
-0.5
0.0
0.5
1.0

Fig. S3.2: Standardized effect sizes of MPD and MNTD (i.e., ses.MPD and ses.MNTD) for coniferous (A, B), broadleaved deciduous (C, D), and broadleaved evergreen (E, F) forests including all vascular taxa. Values of -1 and 1 indicate that all plots in that cell were either significantly clustered or overdispersed for both metrics, respectively. Values of 0 indicate that plots in the given cell did not differ from the random expectation for each metric. The dashed red line indicates the extent of the ice-sheet over 50º N during the LGM.

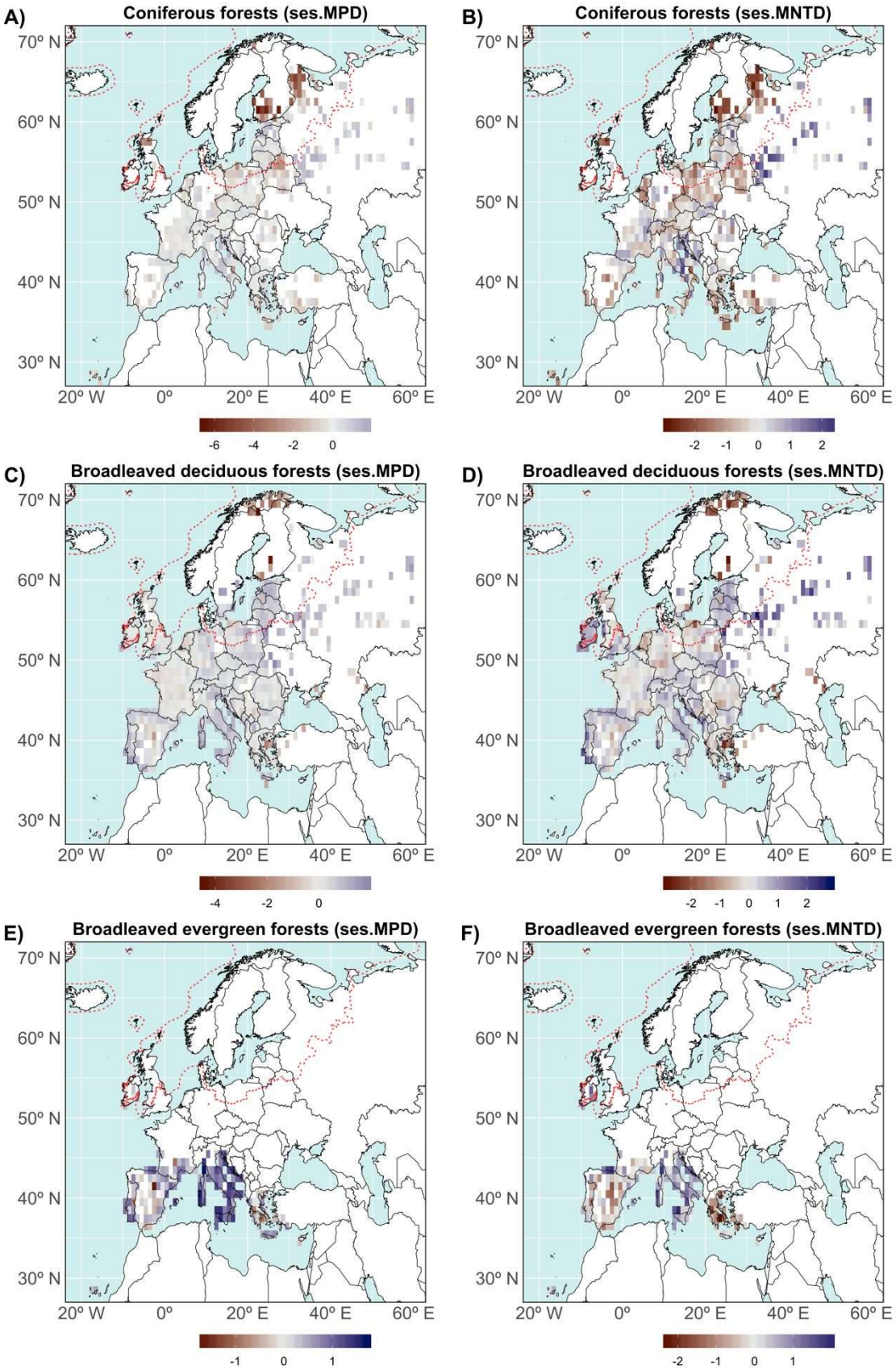


Fig. S3.3: Average ses.MPD and ses.MNTD values of all plots in each grid cell for coniferous (A, B), broadleaved deciduous (C, D), and broadleaved evergreen (E, F) forests including only

angiosperms. Negative values indicate phylogenetic clustering, and positive values indicate phylogenetic overdispersion. Values close to 0 indicate random phylogenetic structure. The dashed red line indicates the extent of the ice-sheet over 50º N during the LGM.

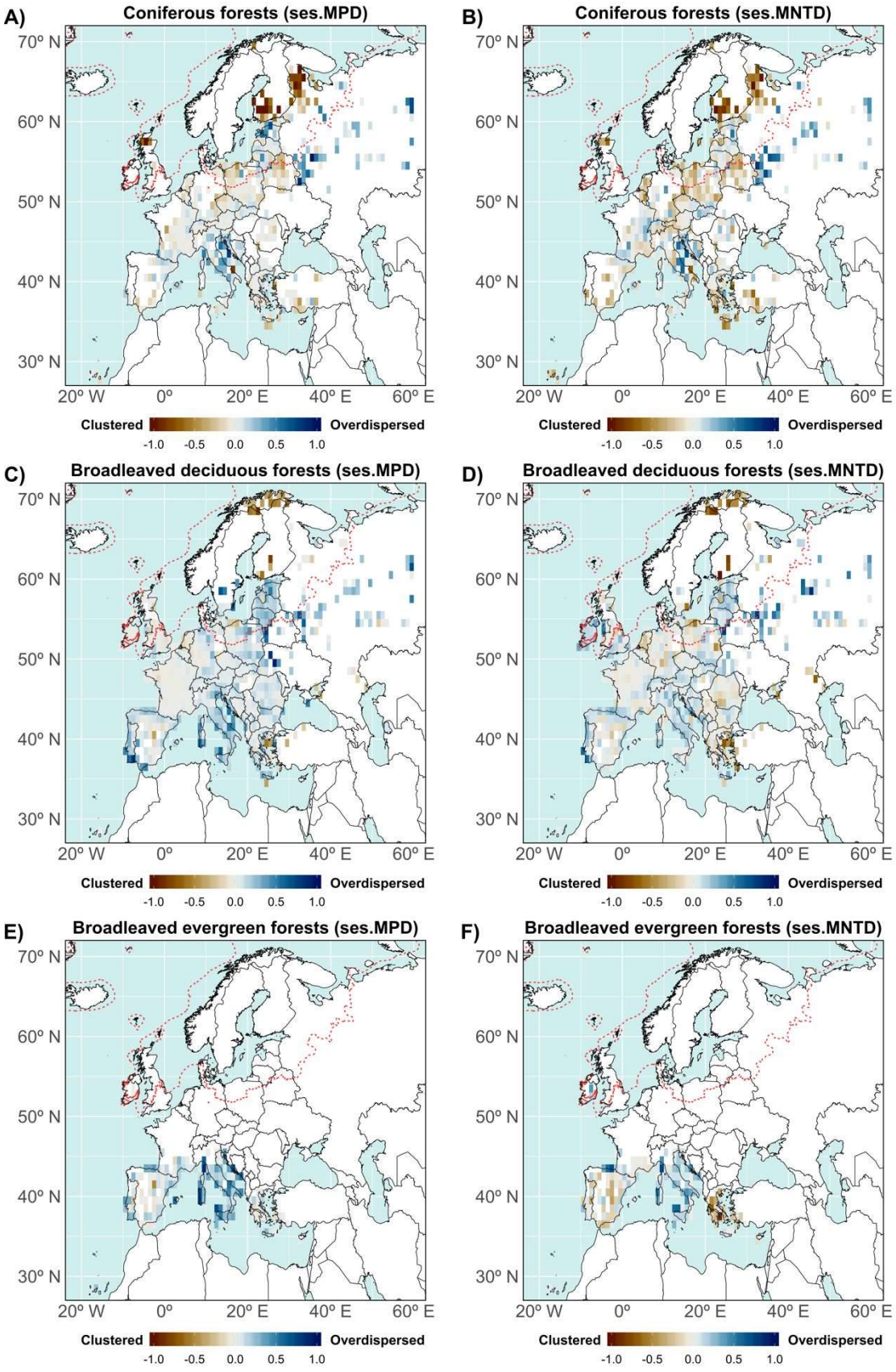


Fig. S3.4: Standardized effect sizes of MPD and MNTD (i.e., ses.MPD and ses.MNTD) for coniferous (A-B), broadleaved deciduous (C-D), and broadleaved evergreen (E-F) forests

including only angiosperms. Values of -1 and 1 indicate that all vegetation plots in the given cell were either significantly clustered or overdispersed, respectively. Values of 0 indicate that plots in the given cell did not differ from the random expectation for each metric. Grid cell resolution is 1º × 1º and each cell contains a minimum of 5 plots. The dashed red line indicates the extent of the ice-sheet over 50º N during the LGM.

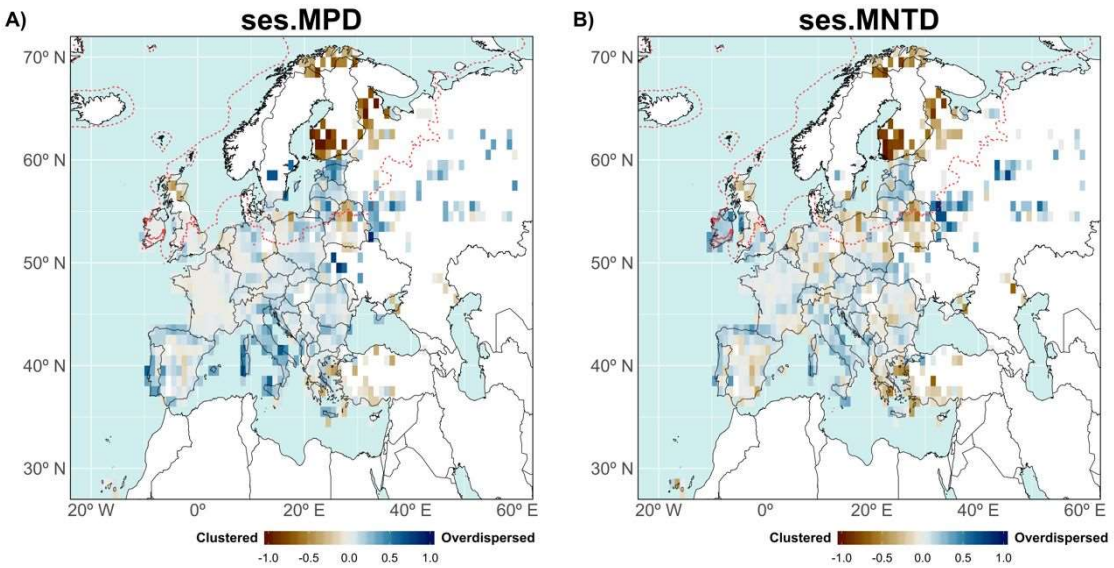


Fig. S3.5: ses.MPD (A) and ses.MNTD (B) for forest angiosperm species (i.e., combining vegetation plots of coniferous, and broadleaved deciduous and evergreen forests). Values of -1 and 1 indicate that all plots in the given cell were significantly clustered or overdispersed, respectively. Values of 0 indicate that all plots in the given cell did not differ from random. The dashed red line indicates the extent of the ice-sheet over 50º N during the LGM.

Appendix S4: Results of sensitivity analyses with different grid cell sizes

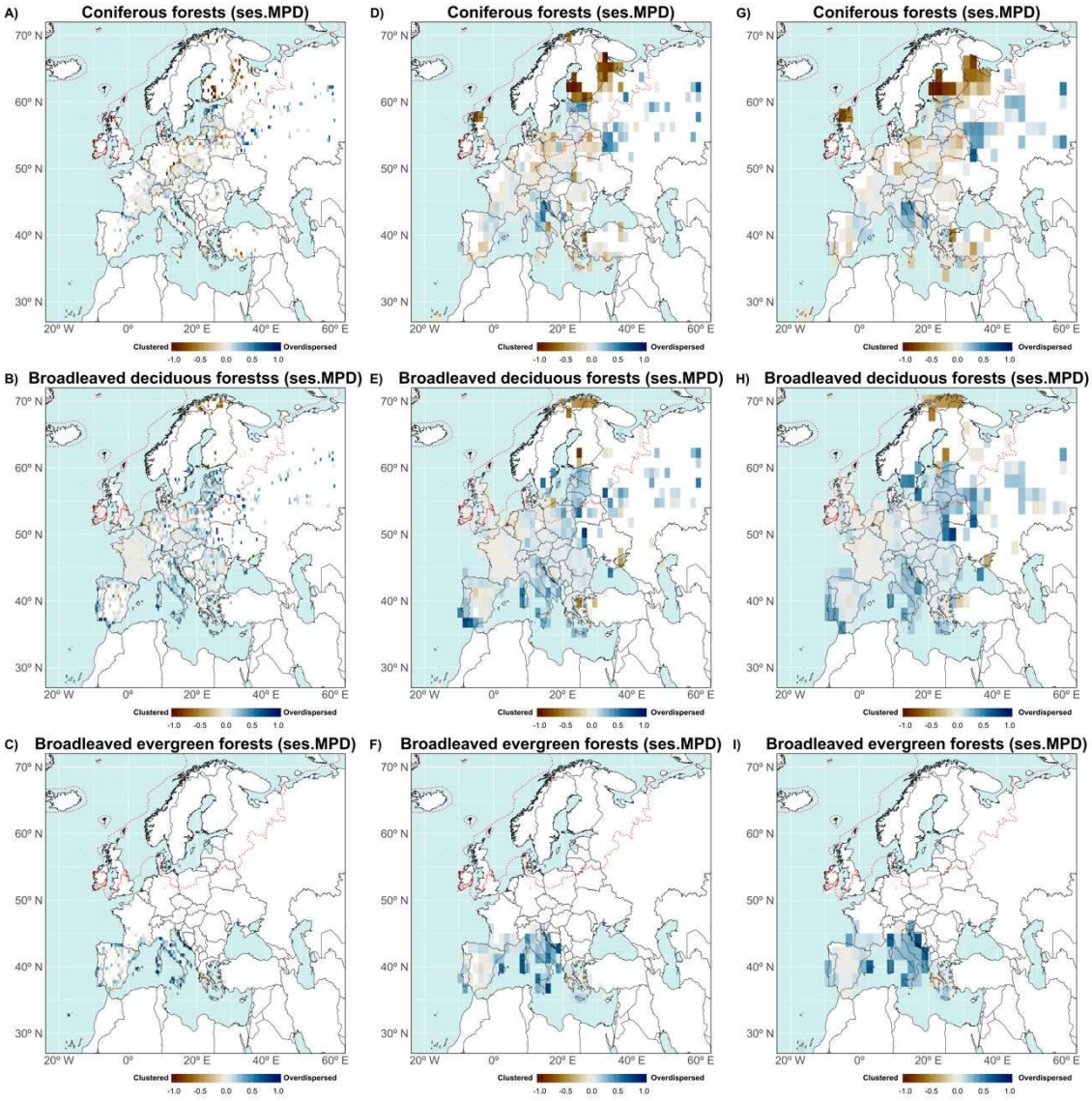


Fig. S4.1: Standardized effect sizes of MPD (ses.MPD) for coniferous (A, D, G), broadleaved deciduous (B, E, H), and broadleaved evergreen (C, F, I) forests including only angiosperms at different spatial resolutions. Values of -1 and 1 indicate that all plots in the given cell were either significantly clustered or overdispersed, respectively. Values of 0 indicate that plots in that cell did not differ from the random expectation for at least one of the metrics. In plots A-C, grid cell resolution is 0.5º × 0.5º and contain a minimum of 3 plots. In plots D-F, grid cell resolution is 1.5º × 1.5º and contain a minimum of 7 plots. In plots G-I, grid cell resolution is 2º × 2º and contain a minimum of 10 plots. See Appendix S3: Fig. S3.4 for a comparison with the actual

resolution used in this study. The dashed red line indicates the extent of the ice-sheet over 50º N during the LGM.

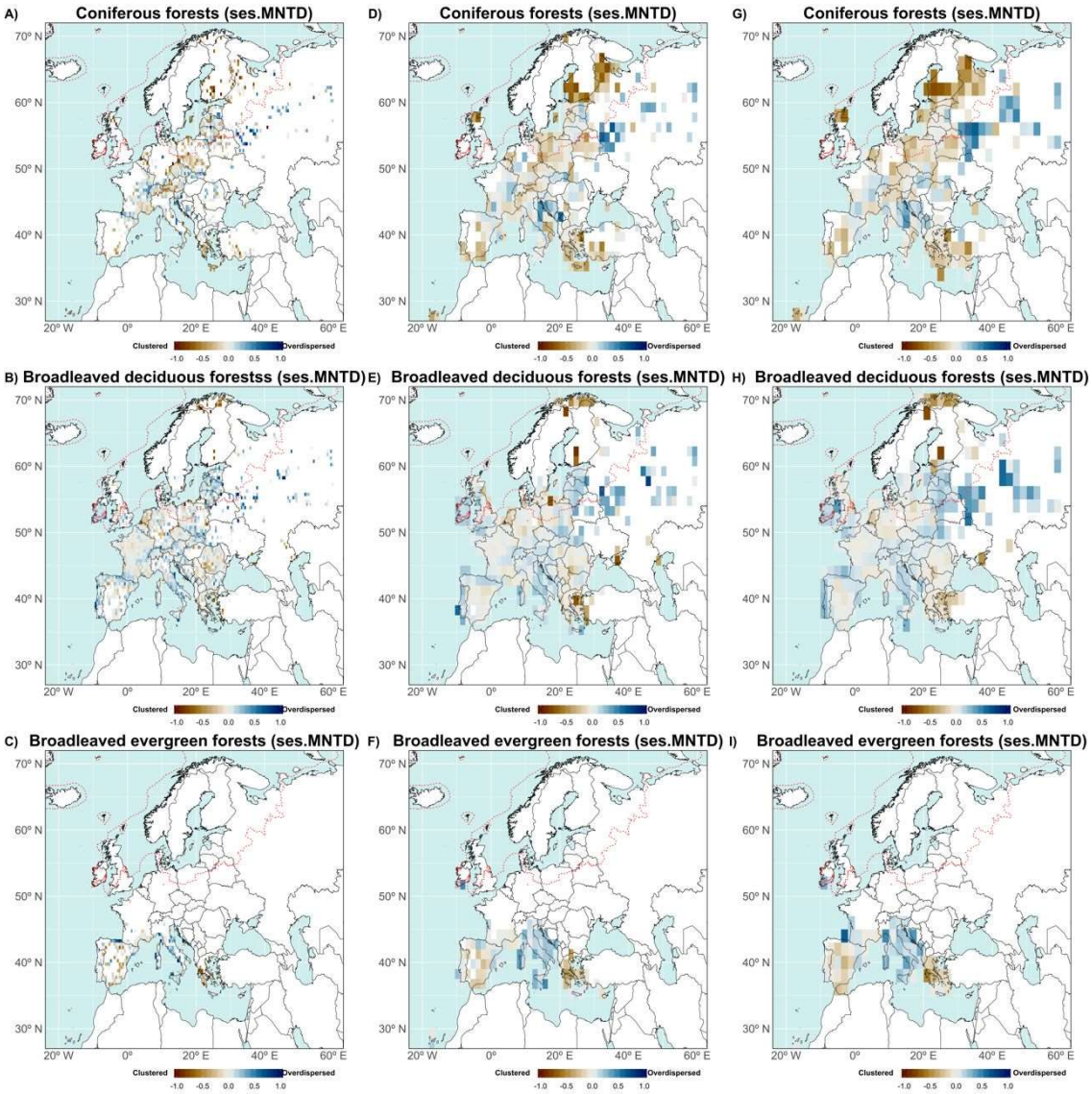


Fig. S4.2: Standardized effect sizes of MNTD (ses.MNTD) for coniferous (A, D, G), broadleaved deciduous (B, E, H), and broadleaved evergreen (C, F, I) forests including only angiosperms at different spatial resolutions. Values of -1 and 1 indicate that all plots in the given cell were either significantly clustered or overdispersed, respectively. Values of 0 indicate that plots in that cell did not differ from the random expectation for at least one of the metrics. In plots A-C, grid cell resolution is 0.5° × 0.5° and contain a minimum of 3 plots. In plots D-F, grid cell resolution is 1.5° × 1.5° and contain a minimum of 7 plots. In plots G-I, grid cell resolution is 2° × 2° and contain a minimum of 10 plots. See Appendix S3: Fig. S3.4 for a comparison with

the actual resolution used in this study. The dashed red line indicates the extent of the ice-sheet over 50º N during the LGM.

Appendix S5: Details on fitting and evaluation of boosted regression trees

We selected the optimum number of boosted regression trees (BRTs) for each model using a 10-fold cross-validation procedure (Fig. S5.1). Each tree was fitted on 50% of the data randomly sampled (without replacement) from the full training set. Introducing such randomness into a boosted model improves accuracy and speed and reduces overfitting (Friedman, 2002). The trees were gradually added to the model in groups of 100 and with a small learning rate of 0.001 that represents the contribution of each tree to the growing model (Elith et al., 2008). By setting the tree complexity parameter to 5, we allowed up to five-way interactions.

Variable selection in BRT is achieved because the model largely ignores non-informative predictors when fitting trees and variables have a minimal effect on prediction. The relative influence of predictor variables in each BRT model was estimated based on the number of times a variable was selected for splitting, weighted by the squared improvement to the model as a result of each split, and averaged over all trees. The relative influence (or contribution) of each variable is scaled so that the sum adds to 100, with higher numbers indicating a stronger influence on the response (Elith et al., 2008). Partial dependence plots were used to visualize the shape of the relationship between response and predictor variables within the model (Hastie et al., 2009). We interpreted the partial dependence plot for each factor to assess whether they reflect positive, negative or more complex responses.

We tested for spatial autocorrelation in the residuals of each model using Moran's I statistics for distance classes defined using Sturges' rule. Because we were interested in spatial autocorrelation at short spatial distances, we randomly selected 500 plots and calculated the spatial correlogram for all plots located within 2º (~225 km) of the focus plot. Resulting correlograms were then summarized using local polynomial regression fitting with a span of 0.75.

References

Elith, J., Leathwick, J. R., & Hastie, T. (2008). A working guide to boosted regression trees. *Journal of Animal Ecology*, *77*(4), 802–813. https://doi.org/10.1111/j.1365-2656.2008.01390.x

Friedman, J. H. (2002). Stochastic gradient boosting. *Computational Statistics & Data Analysis*, *38*(4), 367–378. https://doi.org/10.1016/S0167-9473(01)00065-2

Hastie, T., Tibshirani, R., & Friedman, J. (2009). *The elements of statistical learning: data mining, inference, and prediction* (2nd ed.). New York: Springer-Verlag.

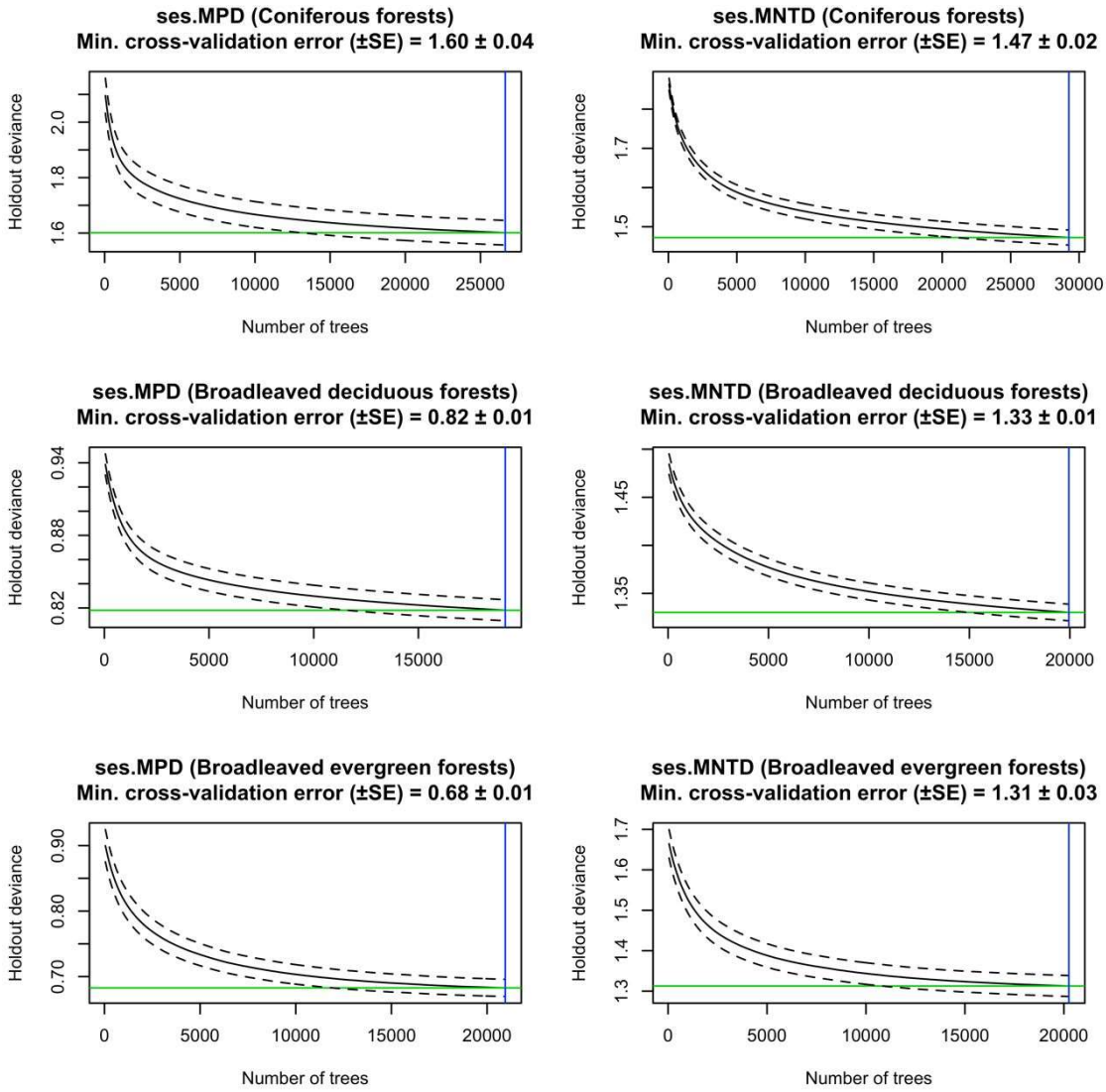


Fig. S5.1: Selection of the optimum number of regression trees in the BRT models for ses.MPD and ses.MNTD in coniferous, broadleaved deciduous, and broadleaved evergreen forests. Tree complexity parameter was set to 5 and learning rate to 0.001. Curves show a decrease of cross-validation error (± SE) as a function of an increasing number of regression trees. Green lines show the minimum cross-validation error reached in each case. Vertical blue lines show the optimum number of regression trees (a higher number of trees would cause an increase of cross-validation error and model overfitting).

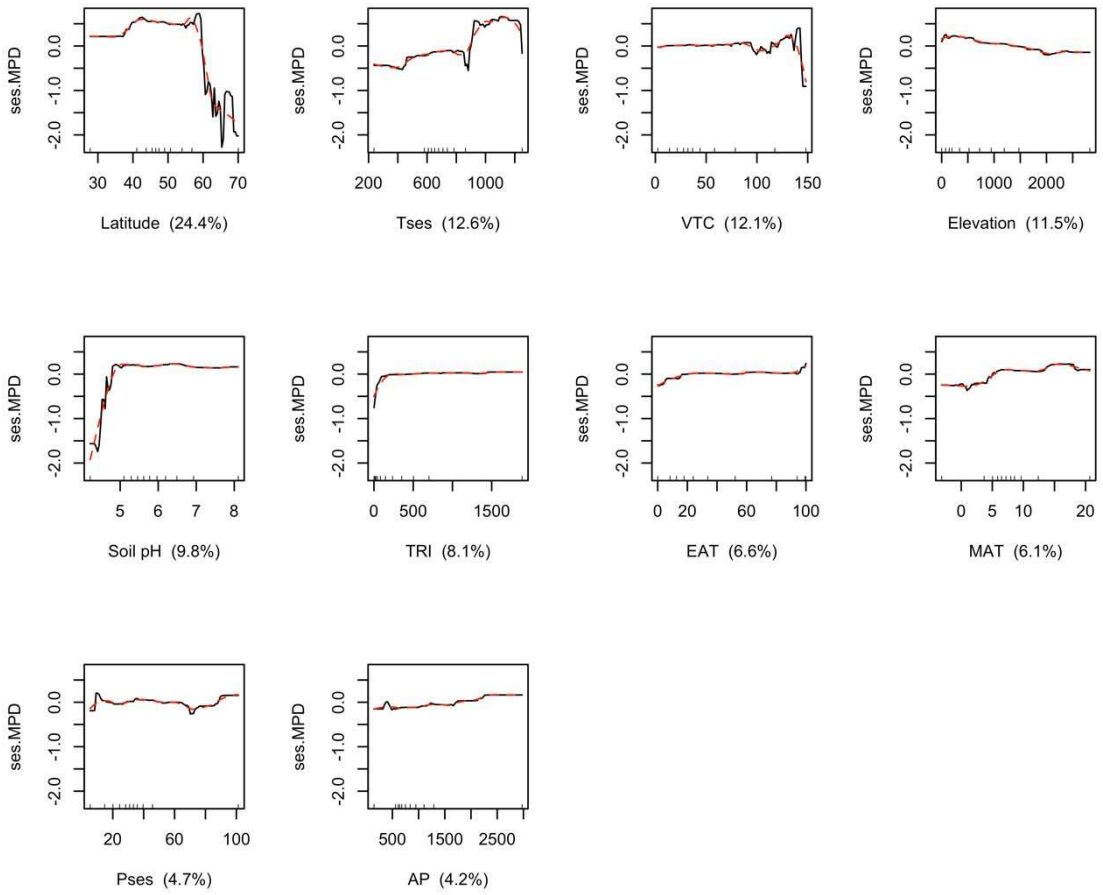


Fig. S5.2: Partial dependence plots with the effects of environmental variables on ses.MPD in coniferous forests (N = 16,382). Plots are ordered according to the importance (contribution, in %) of the variables in the BRT models. Smoothed versions of the fitted functions (red dashed curves) were calculated using local polynomial regression. Vertical ticks on the x-axis indicate deciles of the response variable. Predictor variables are ‘mean annual temperature (MAT)’, ‘temperature seasonality (Tses)’, ‘annual precipitation (AP)’, ‘precipitation seasonality (Pses)’, ‘velocity of temperature change since LGM (VTC)’, ‘extent of analogous temperature (EAT)’, ‘elevation’, ‘Terrain Ruggedness Index (TRI)’, ‘Soil pH’, and ‘Latitude’.

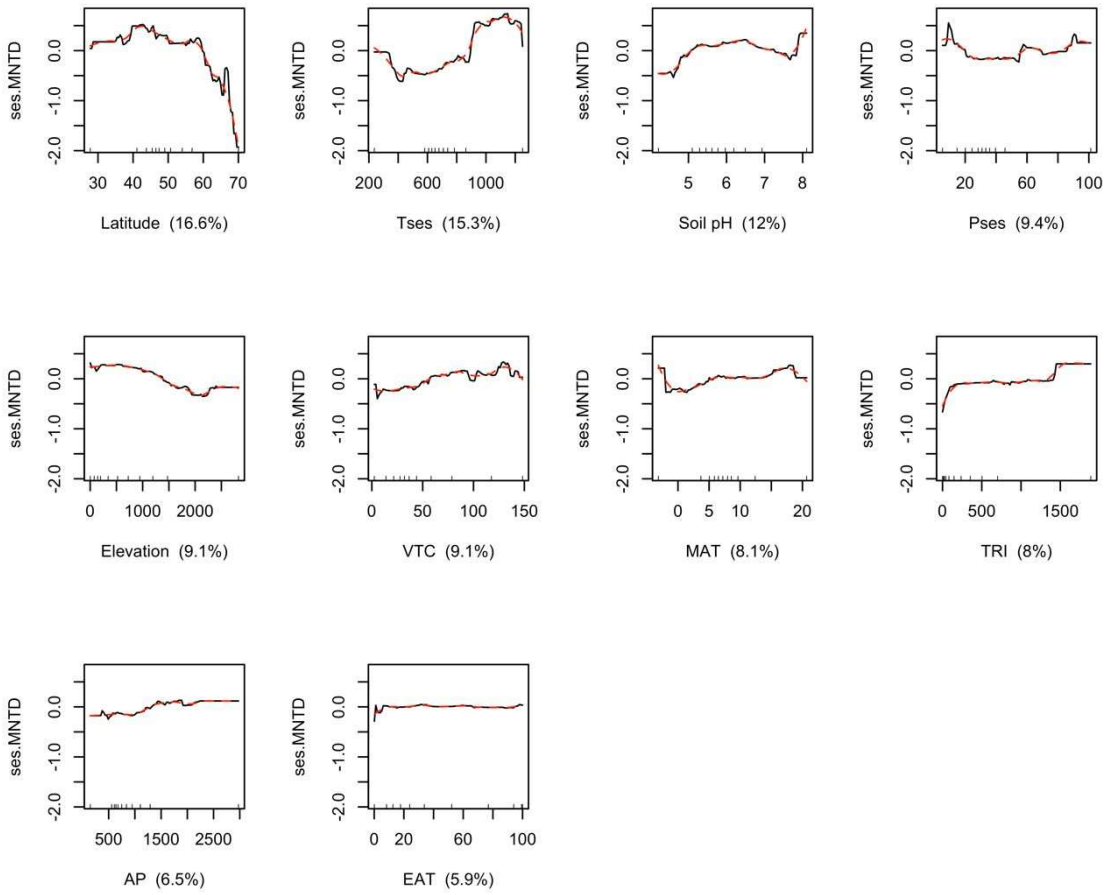


Fig. S5.3: Partial dependence plots with the effects of environmental variables on ses.MNTD in coniferous forests (N = 16,382). Plots are ordered according to the importance (contribution, in %) of the variables in the BRT models. Smoothed versions of the fitted functions (red dashed curves) were calculated using local polynomial regression. Vertical ticks on the x-axis indicate deciles of the response variable. Predictor variables are 'mean annual temperature (MAT)', 'temperature seasonality (Tses)', 'annual precipitation (AP)', 'precipitation seasonality (Pses)', 'velocity of temperature change since LGM (VTC)', 'extent of analogous temperature (EAT)', 'elevation', 'Terrain Ruggedness Index (TRI)', 'Soil pH', and 'Latitude'.

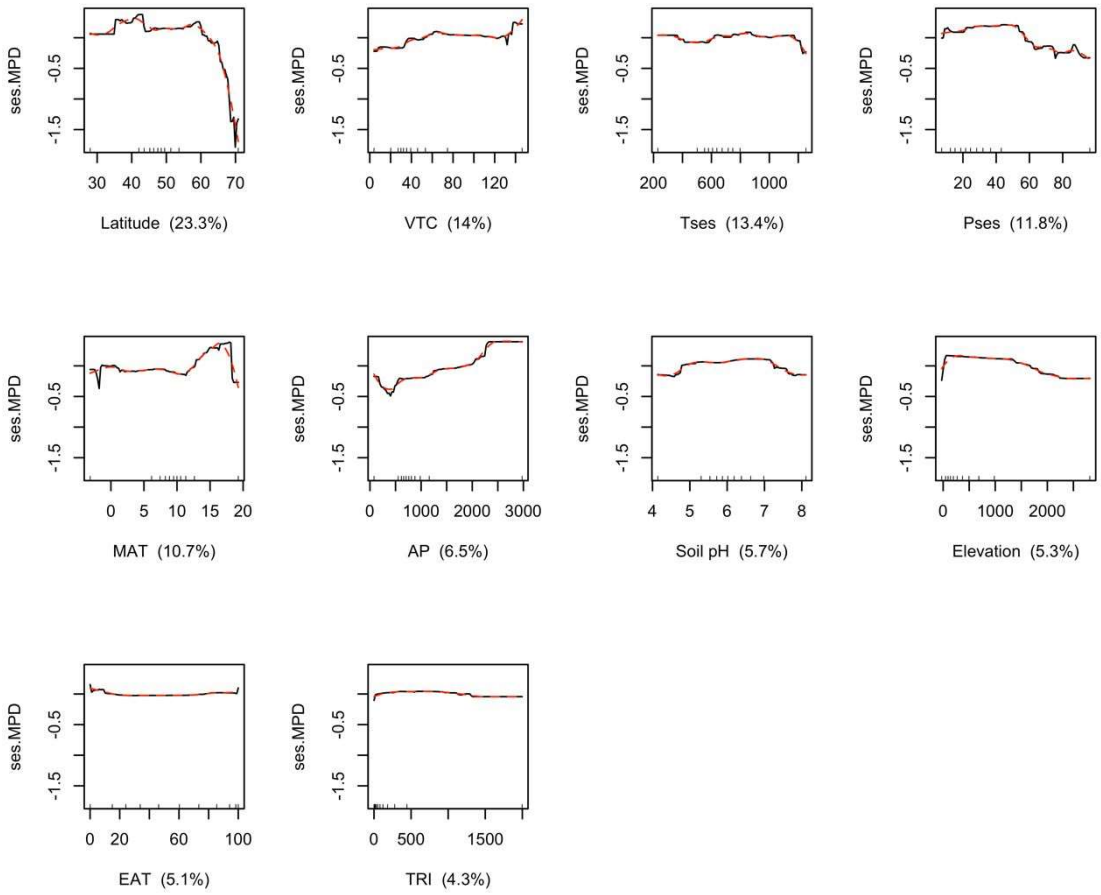


Fig. S5.4: Partial dependence plots with the effects of environmental variables on ses.MPD in broadleaved deciduous forests (N = 41,198). Plots are ordered according to the importance (contribution, in %) of the variables in the BRT models. Smoothed versions of the fitted functions (red dashed curves) were calculated using local polynomial regression. Vertical ticks on the x-axis indicate deciles of the response variable. Predictor variables are 'mean annual temperature (MAT)', 'temperature seasonality (Tses)', 'annual precipitation (AP)', 'precipitation seasonality (Pses)', 'velocity of temperature change since LGM (VTC)', 'extent of analogous temperature (EAT)', 'elevation', 'Terrain Ruggedness Index (TRI)', 'Soil pH', and 'Latitude'.

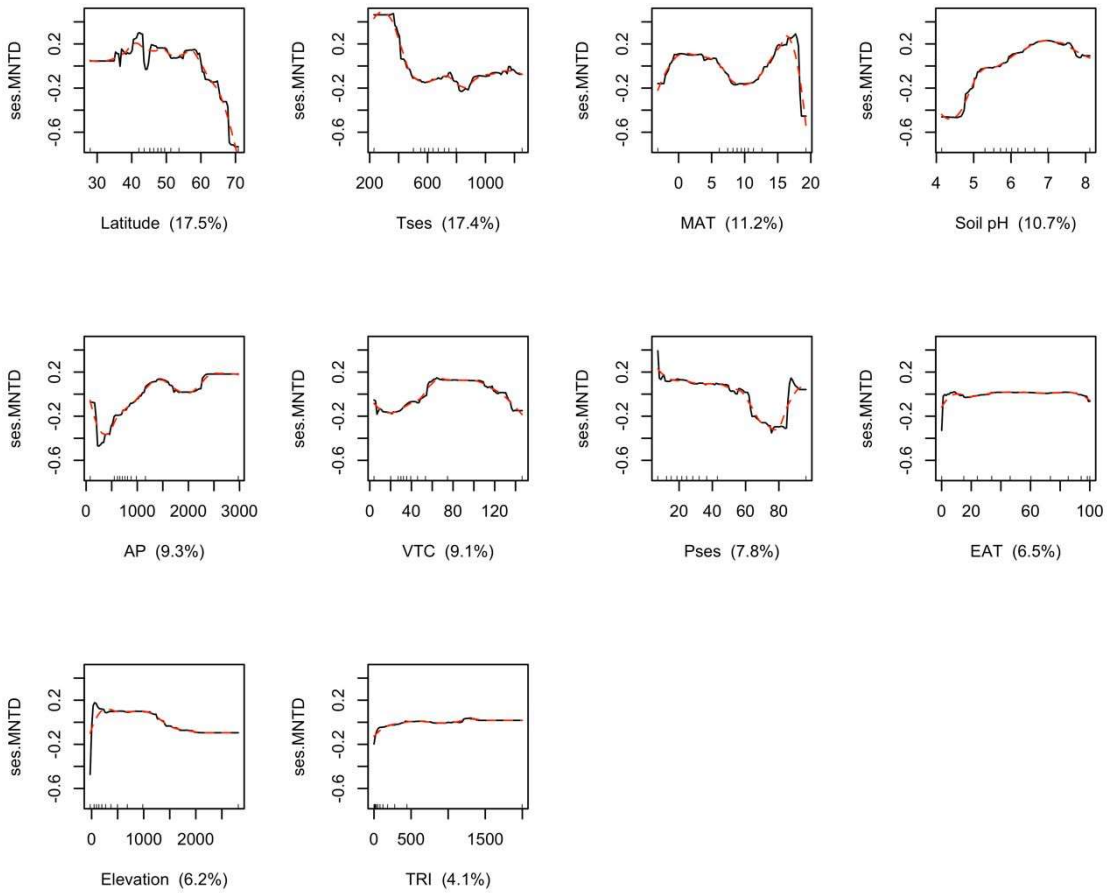


Fig. S5.5: Partial dependence plots with the effects of environmental variables on ses.MNTD in broadleaved deciduous forests (N = 41,198). Plots are ordered according to the importance (contribution, in %) of the variables in the BRT models. Smoothed versions of the fitted functions (red dashed curves) were calculated using local polynomial regression. Vertical ticks on the x-axis indicate deciles of the response variable. Predictor variables are 'mean annual temperature (MAT)', 'temperature seasonality (Tses)', 'annual precipitation (AP)', 'precipitation seasonality (Pses)', 'velocity of temperature change since LGM (VTC)', 'extent of analogous temperature (EAT)', 'elevation', 'Terrain Ruggedness Index (TRI)', 'Soil pH', and 'Latitude'.

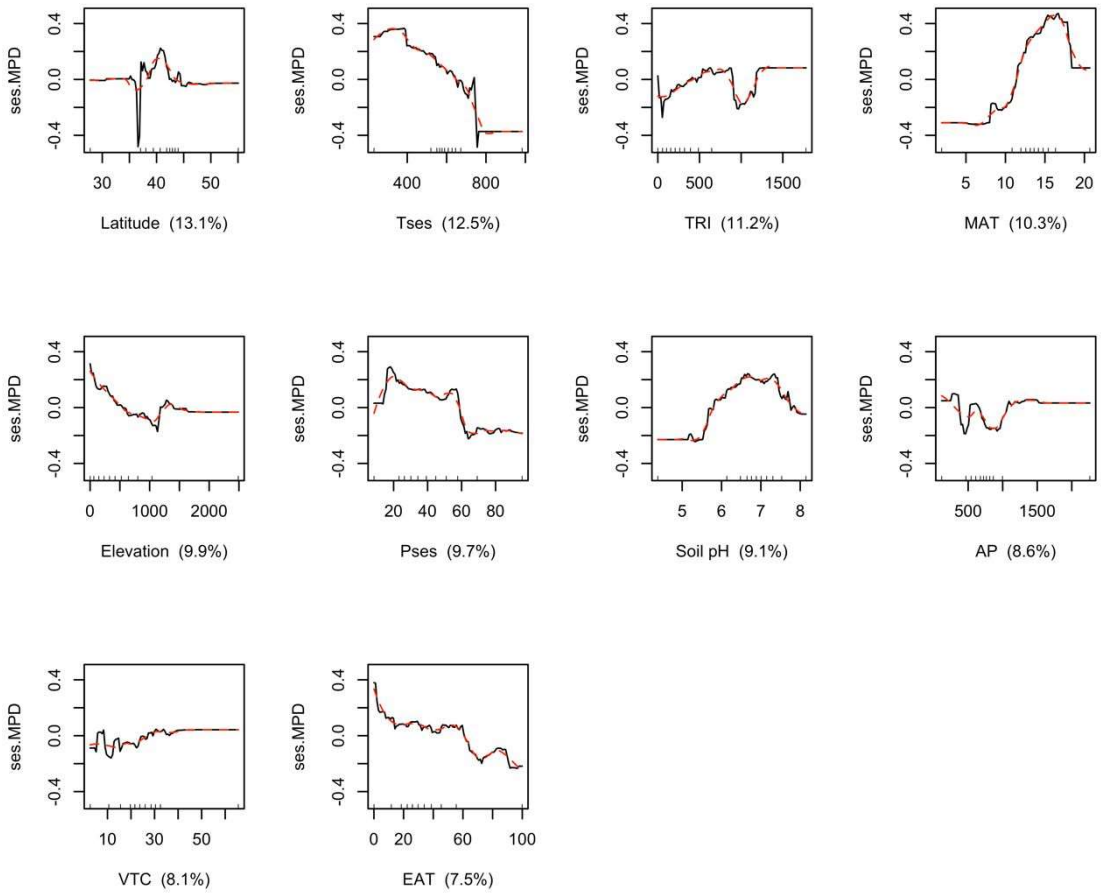


Fig. S5.6: Partial dependence plots with the effects of environmental variables on ses.MPD in broadleaved evergreen forests (N = 4,236). Plots are ordered according to the importance (contribution, in %) of the variables in the BRT models. Smoothed versions of the fitted functions (red dashed curves) were calculated using local polynomial regression. Vertical ticks on the x-axis indicate deciles of the response variable. Predictor variables are 'mean annual temperature (MAT)', 'temperature seasonality (Tses)', 'annual precipitation (AP)', 'precipitation seasonality (Pses)', 'velocity of temperature change since LGM (VTC)', 'extent of analogous temperature (EAT)', 'elevation', 'Terrain Ruggedness Index (TRI)', 'Soil pH', and 'Latitude'.

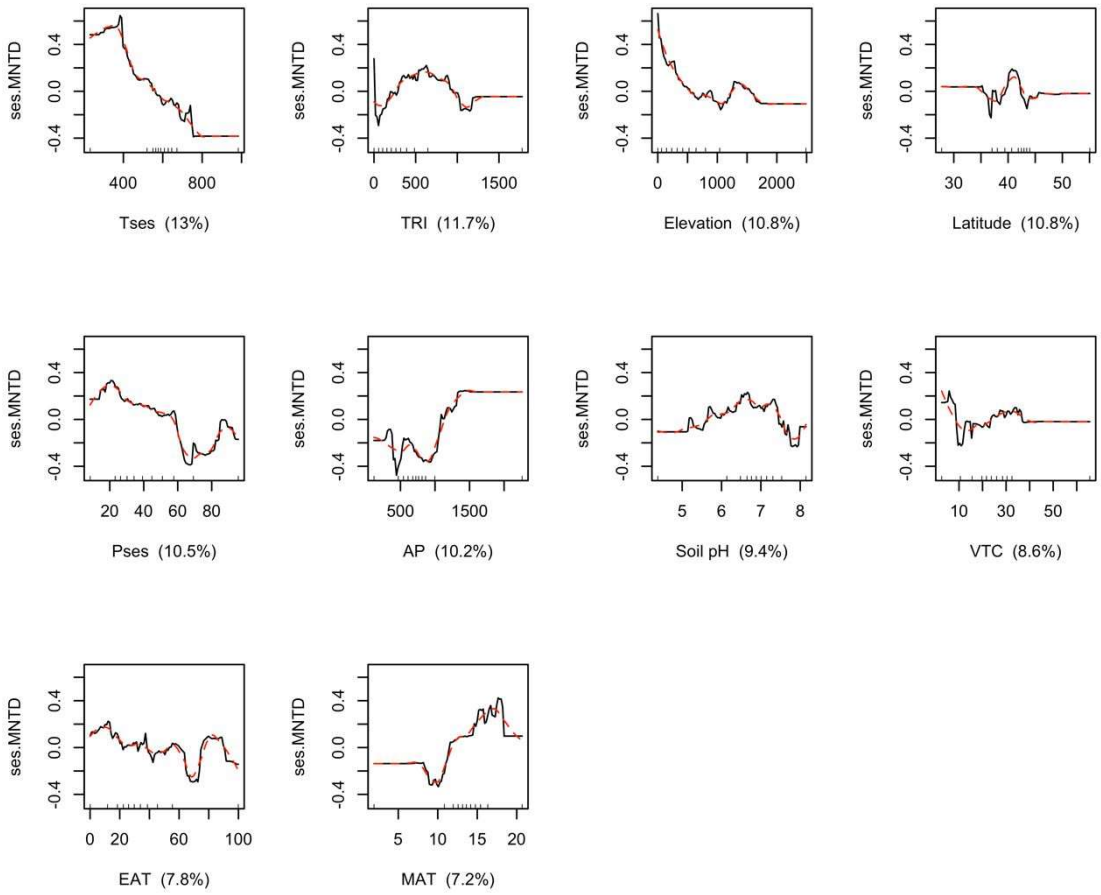


Fig. S5.7: Partial dependence plots with the effects of environmental variables on ses.MNTD in broadleaved evergreen forests (N = 4,236). Plots are ordered according to the importance (contribution, in %) of the variables in the BRT models. Smoothed versions of the fitted functions (red dashed curves) were calculated using local polynomial regression. Vertical ticks on the x-axis indicate deciles of the response variable. Predictor variables are 'mean annual temperature (MAT)', 'temperature seasonality (Tses)', 'annual precipitation (AP)', 'precipitation seasonality (Pses)', 'velocity of temperature change since LGM (VTC)', 'extent of analogous temperature (EAT)', 'elevation', 'Terrain Ruggedness Index (TRI)', 'Soil pH', and 'Latitude'.

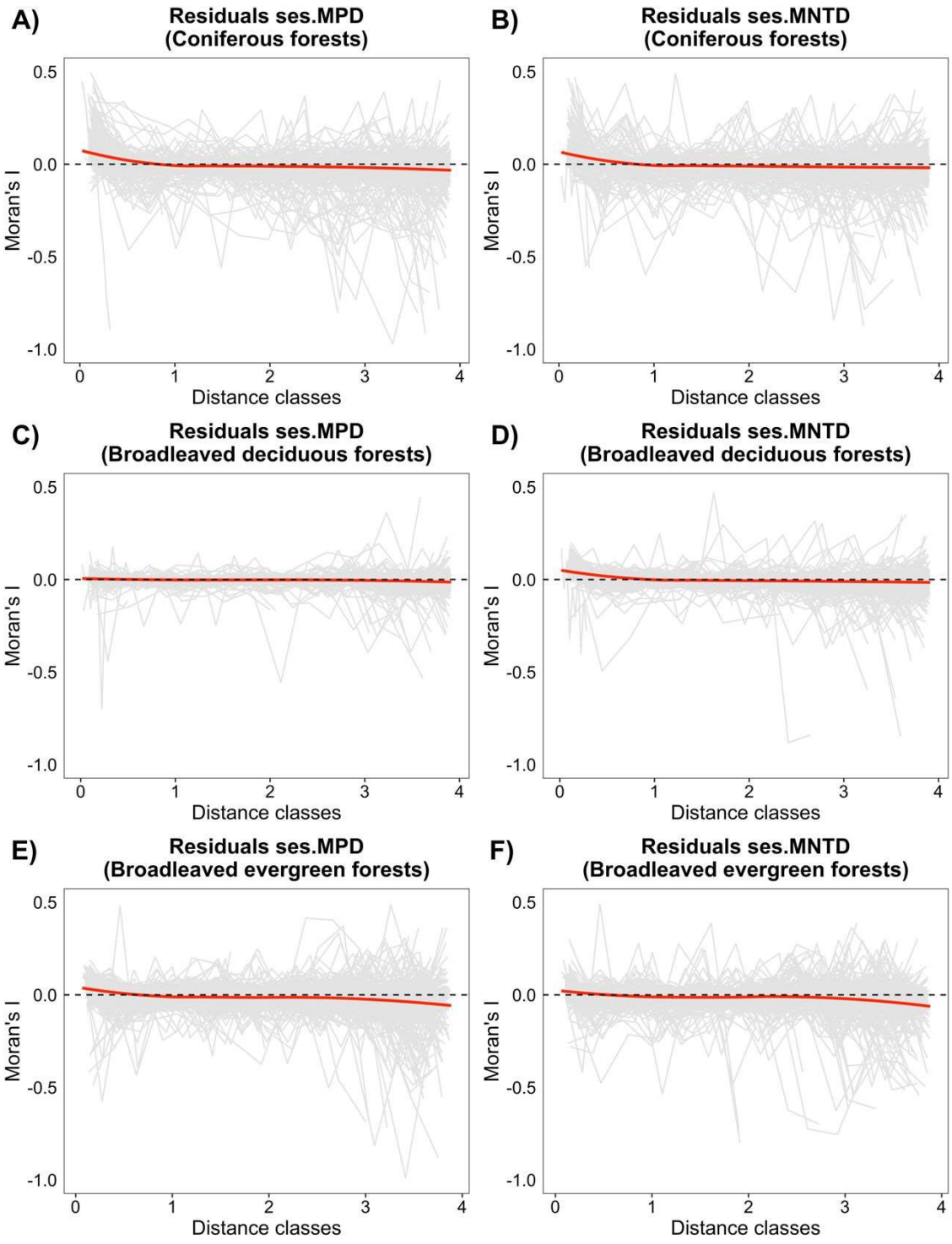


Fig. S5.8: Spatial correlograms of model residuals for each response variable in each forest type: coniferous (A-B), broadleaved deciduous (C-D), and broadleaved evergreen (E-F). We calculated Moran's *I* statistic within 500 circular windows with a 2º radius randomly distributed across the sampled regions.

Appendix S6: Results of distance-based partial Redundancy Analysis (db-RDA)

Table S6.1: Akaike Information Criterion (AIC) values from distance-based partial Redundancy Analysis (db-RDA) with all possible numbers of groups (from 2 to 10) as predictors. The lowest AIC for each forest type is shown in bold.

Number of groups	AIC
Coniferous forests	
1 (Intercept)	122.22
2	**-6.15**
3	80-09
4	82.74
5	81.50
6	96.81
7	91.77
8	97.24
9	90.93
10	88.18
Broadleaved deciduous forests	
1 (Intercept)	242.69
2	**152.05**
3	171.43
4	185.66
5	194.92
6	215.35
7	208.68
8	201.74
9	205.12
10	207.41

Appendix S7: Phylogenetic structure of vegetation plots sorted by EUNIS habitat type

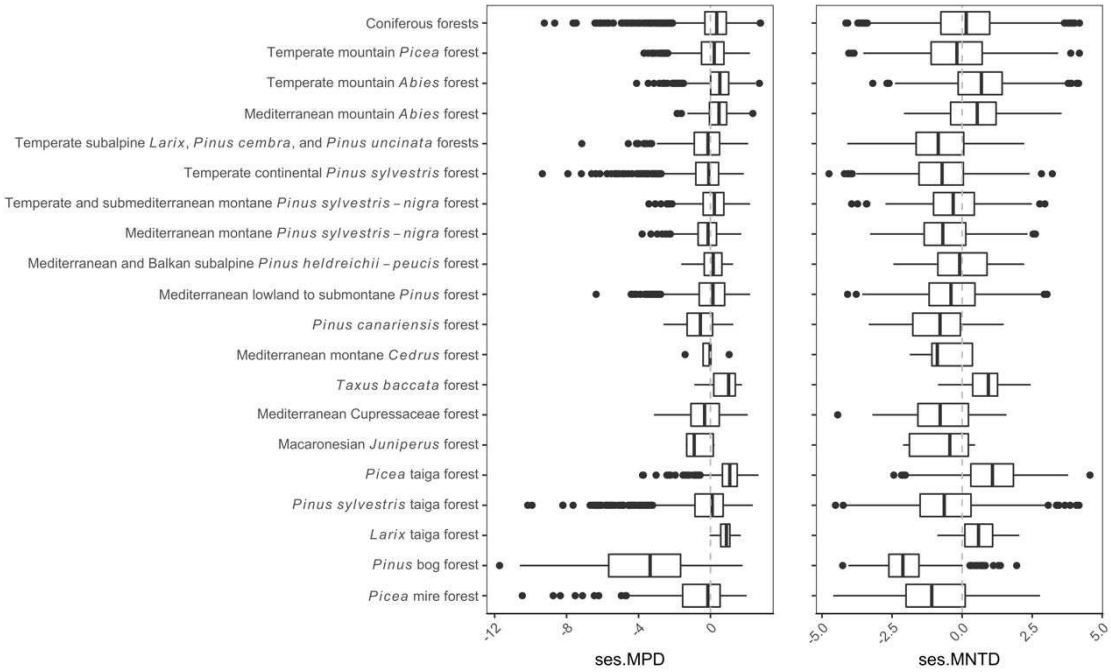
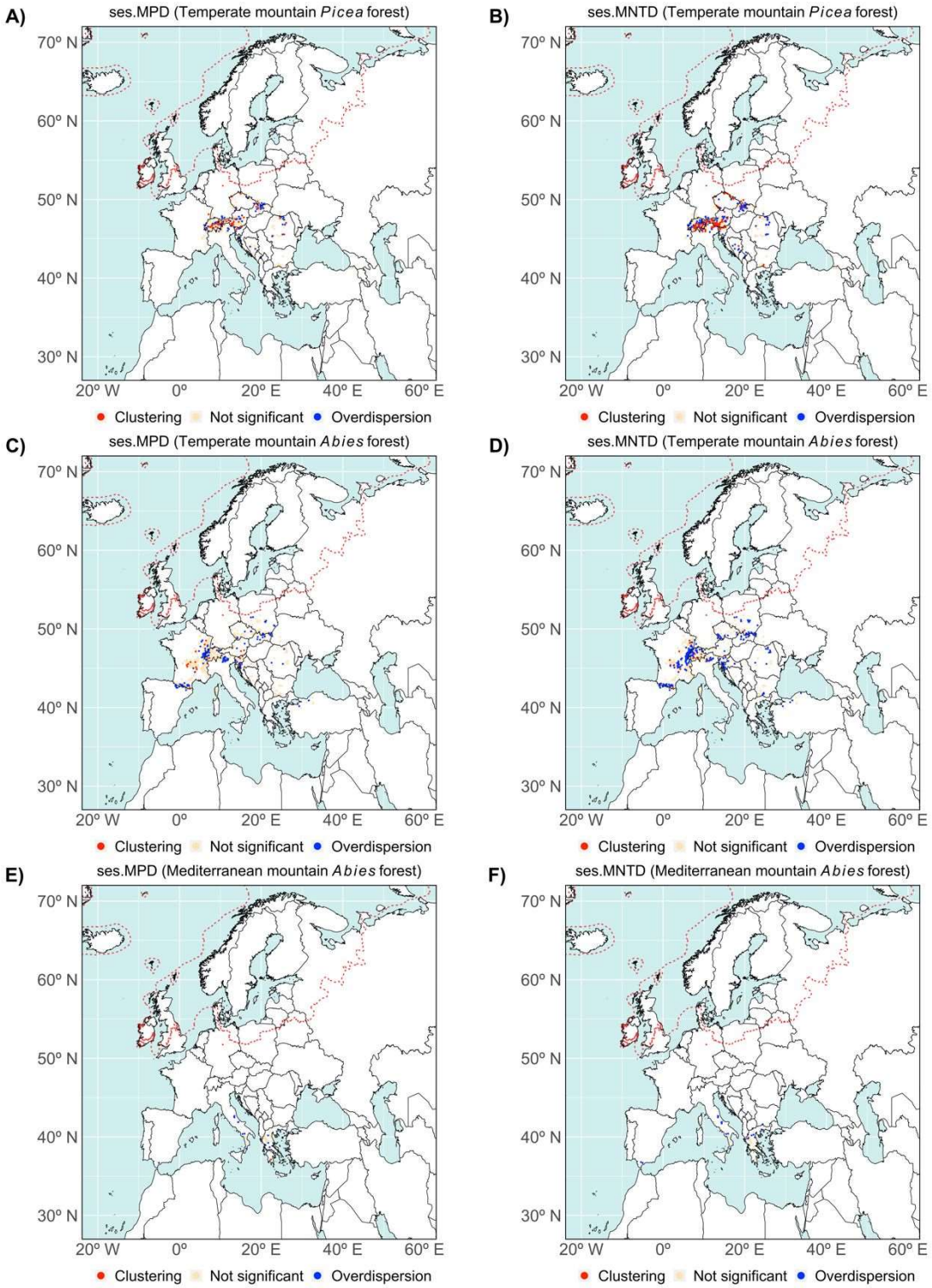
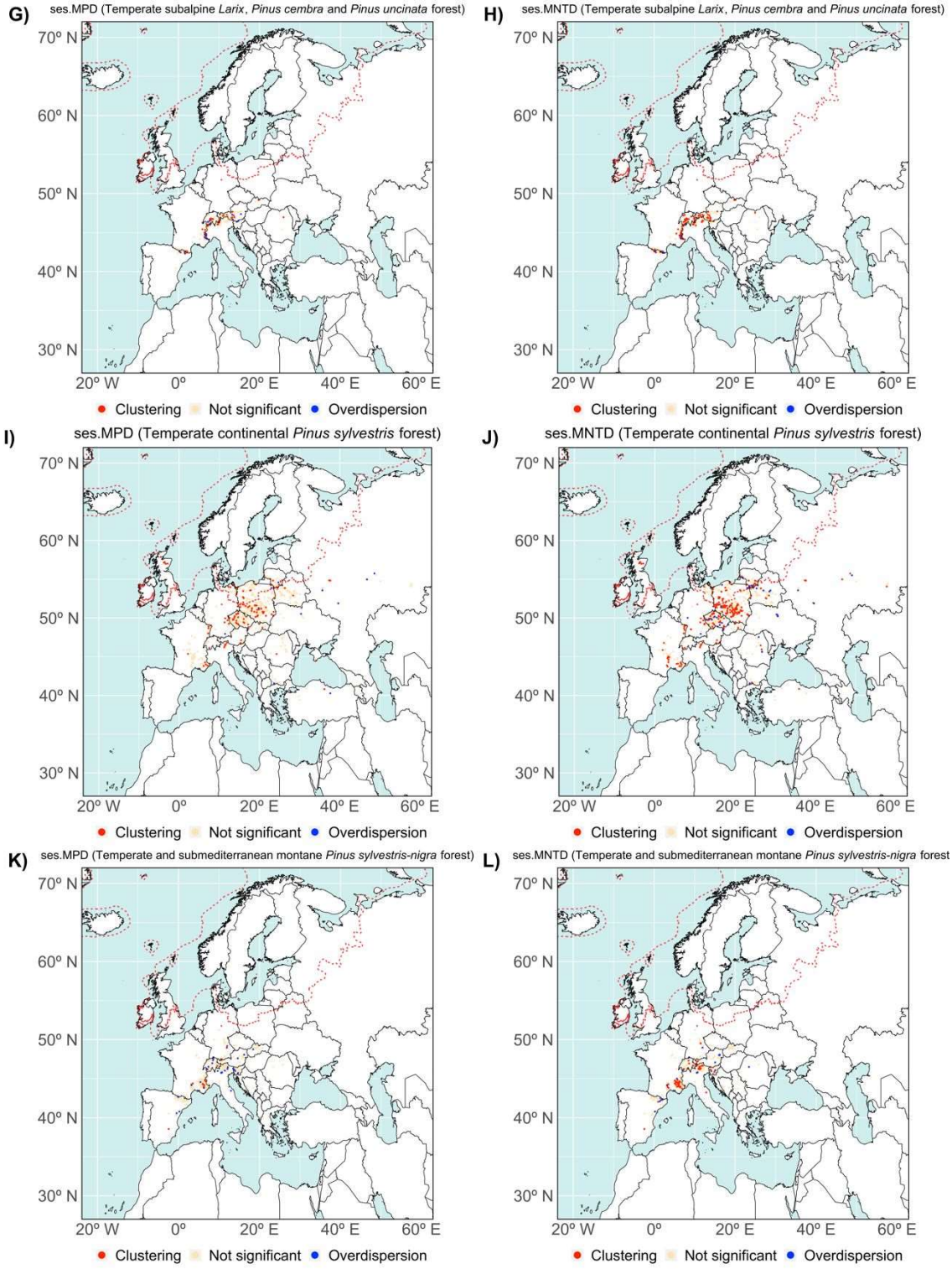
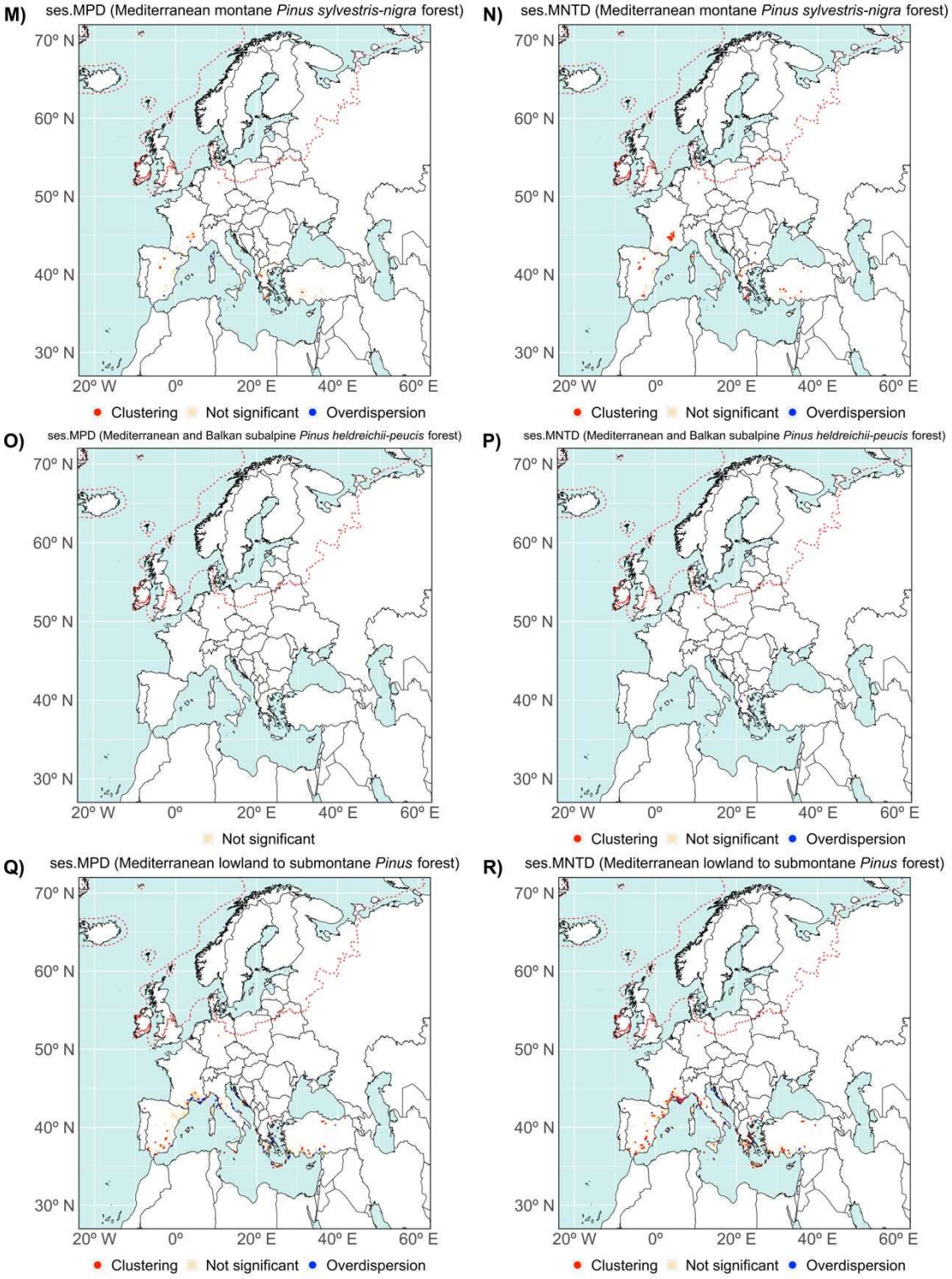
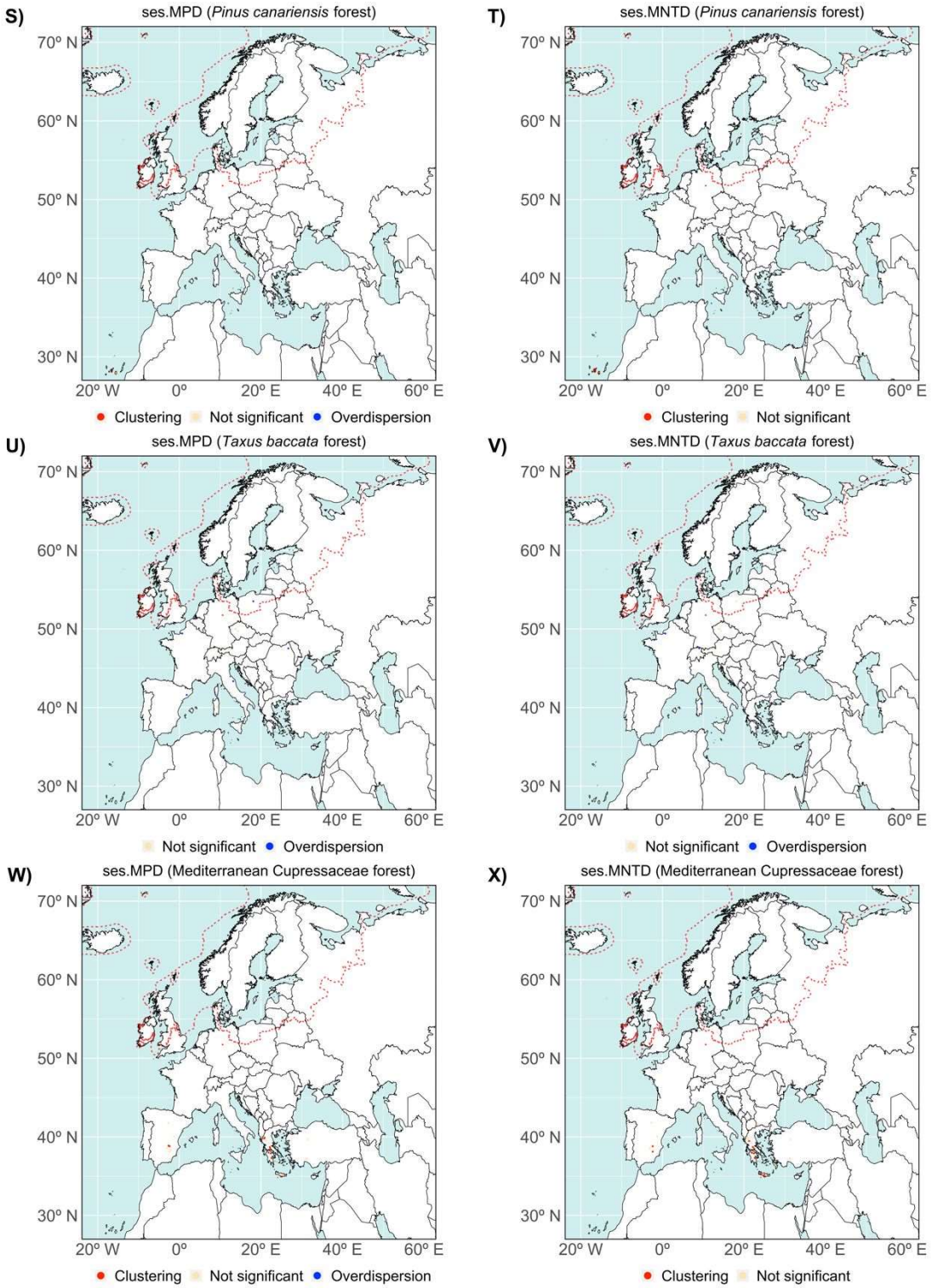


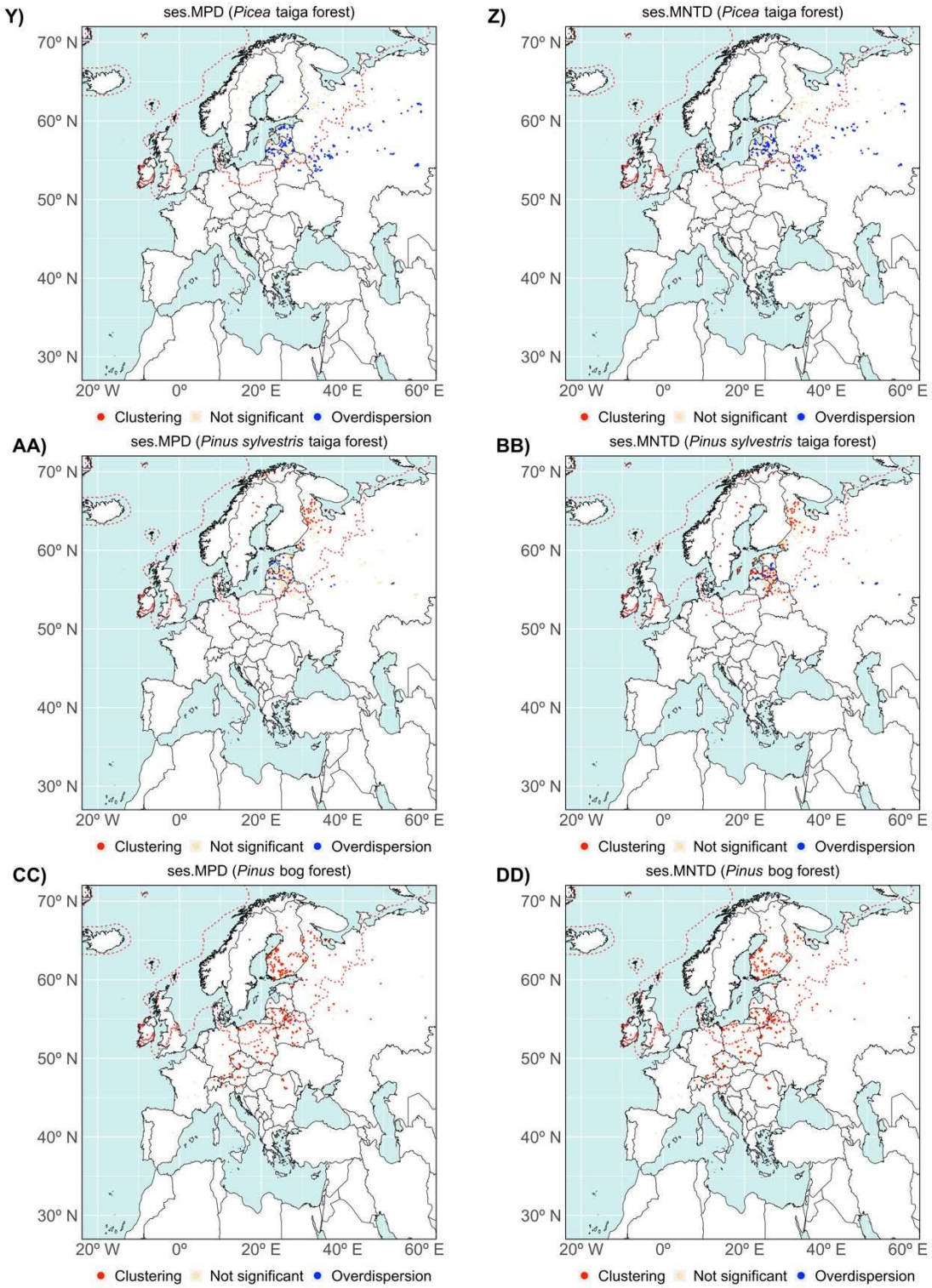
Fig. S7.1: Boxplots (median and quartiles; whiskers show the 5th and 95th percentile) for ses.MPD and ses.MNTD in coniferous vegetation plots sorted by EUNIS habitat types. The 'coniferous forests' category refer to plots that could not be subdivided into more precise EUNIS habitat types.


A)
ses.MPD (Temperate mountain Picea forest)
B)
ses.MNTD (Temperate mountain Picea forest)
C)
ses.MPD (Temperate mountain Abies forest)
D)
ses.MNTD (Temperate mountain Abies forest)
E)
ses.MPD (Mediterranean mountain Abies forest)
F)
ses.MNTD (Mediterranean mountain Abies forest)
70° N
60° N
50° N
40° N
30° N
20° W
0°
20° E
40° E
60° E
Clustering
Not significant
Overdispersion


G)
ses.MPD (Temperate subalpine Larix, Pinus cembra and Pinus uncinata forest)
H)
ses.MNTD (Temperate subalpine Larix, Pinus cembra and Pinus uncinata forest)
I)
ses.MPD (Temperate continental Pinus sylvestris forest)
J)
ses.MNTD (Temperate continental Pinus sylvestris forest)
K)
ses.MPD (Temperate and submediterranean montane Pinus sylvestris-nigra forest)
L)
ses.MNTD (Temperate and submediterranean montane Pinus sylvestris-nigra forest
70° N
60° N
50° N
40° N
30° N
20° W
0°
20° E
40° E
60° E
Clustering
Not significant
Overdispersion


M)
ses.MPD (Mediterranean montane Pinus sylvestris-nigra forest)
N)
ses.MNTD (Mediterranean montane Pinus sylvestris-nigra forest)
O)
ses.MPD (Mediterranean and Balkan subalpine Pinus heldreichii-peucis forest)
P)
ses.MNTD (Mediterranean and Balkan subalpine Pinus heldreichii-peucis forest)
Q)
ses.MPD (Mediterranean lowland to submontane Pinus forest)
R)
ses.MNTD (Mediterranean lowland to submontane Pinus forest)
70° N
60° N
50° N
40° N
30° N
20° W
0°
20° E
40° E
60° E
Clustering
Not significant
Overdispersion


S)
ses.MPD (Pinus canariensis forest)
T)
ses.MNTD (Pinus canariensis forest)
U)
ses.MPD (Taxus baccata forest)
V)
ses.MNTD (Taxus baccata forest)
W)
ses.MPD (Mediterranean Cupressaceae forest)
X)
ses.MNTD (Mediterranean Cupressaceae forest)
70° N
60° N
50° N
40° N
30° N
20° W
0°
20° E
40° E
60° E
Clustering
Not significant
Overdispersion


Y)
ses.MPD (*Picea* taiga forest)
Z)
ses.MNTD (*Picea* taiga forest)
AA)
ses.MPD (*Pinus sylvestris* taiga forest)
BB)
ses.MNTD (*Pinus sylvestris* taiga forest)
CC)
ses.MPD (*Pinus* bog forest)
DD)
ses.MNTD (*Pinus* bog forest)
70° N
60° N
50° N
40° N
30° N
20° W
0°
20° E
40° E
60° E
Clustering
Not significant
Overdispersion

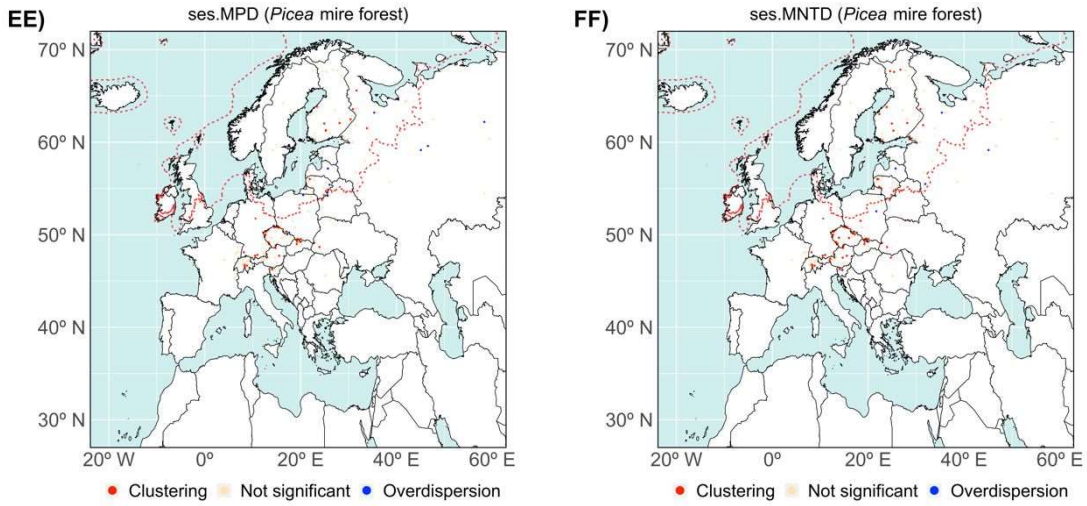


Fig. S7.2: Standardized effect sizes of MPD and MNTD (ses.MPD and ses.MNTD) for coniferous forests sorted by EUNIS habitat type (angiosperms only). Red and blue points indicate significantly clustered and overdispersed plots, respectively. Light-orange denotes plots with random phylogenetic structure. We only mapped habitat types with at least 10 plots (Appendix S1: Table S1.2). The dashed red line indicates the extent of the ice-sheet over 50º N during the LGM.

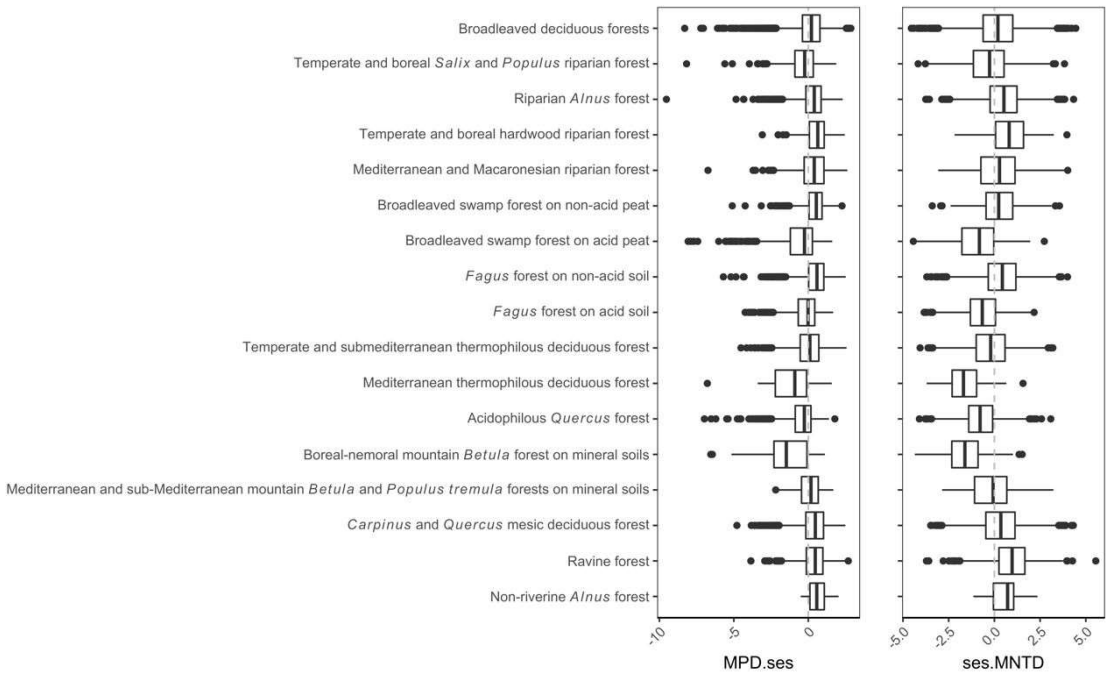
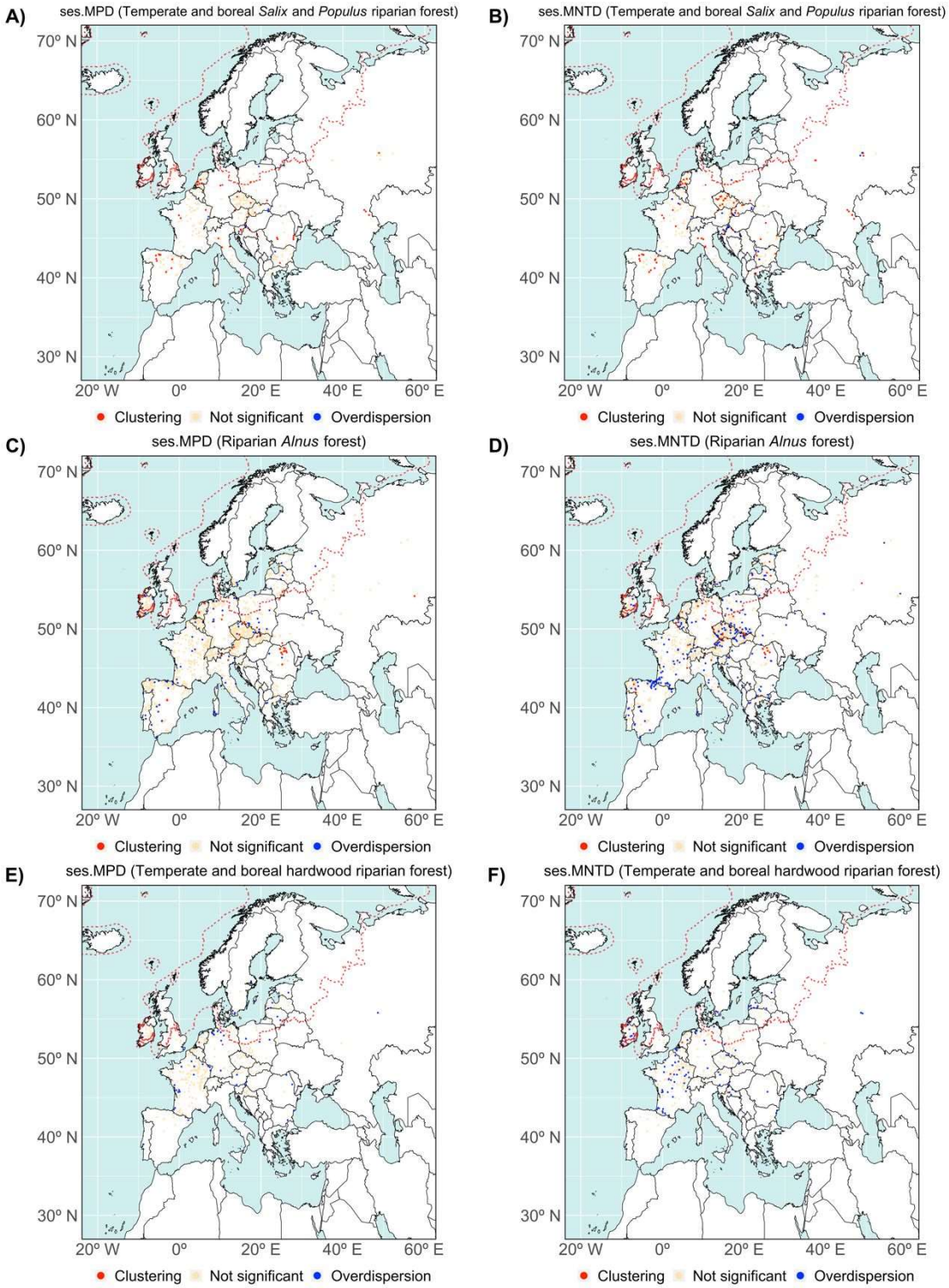
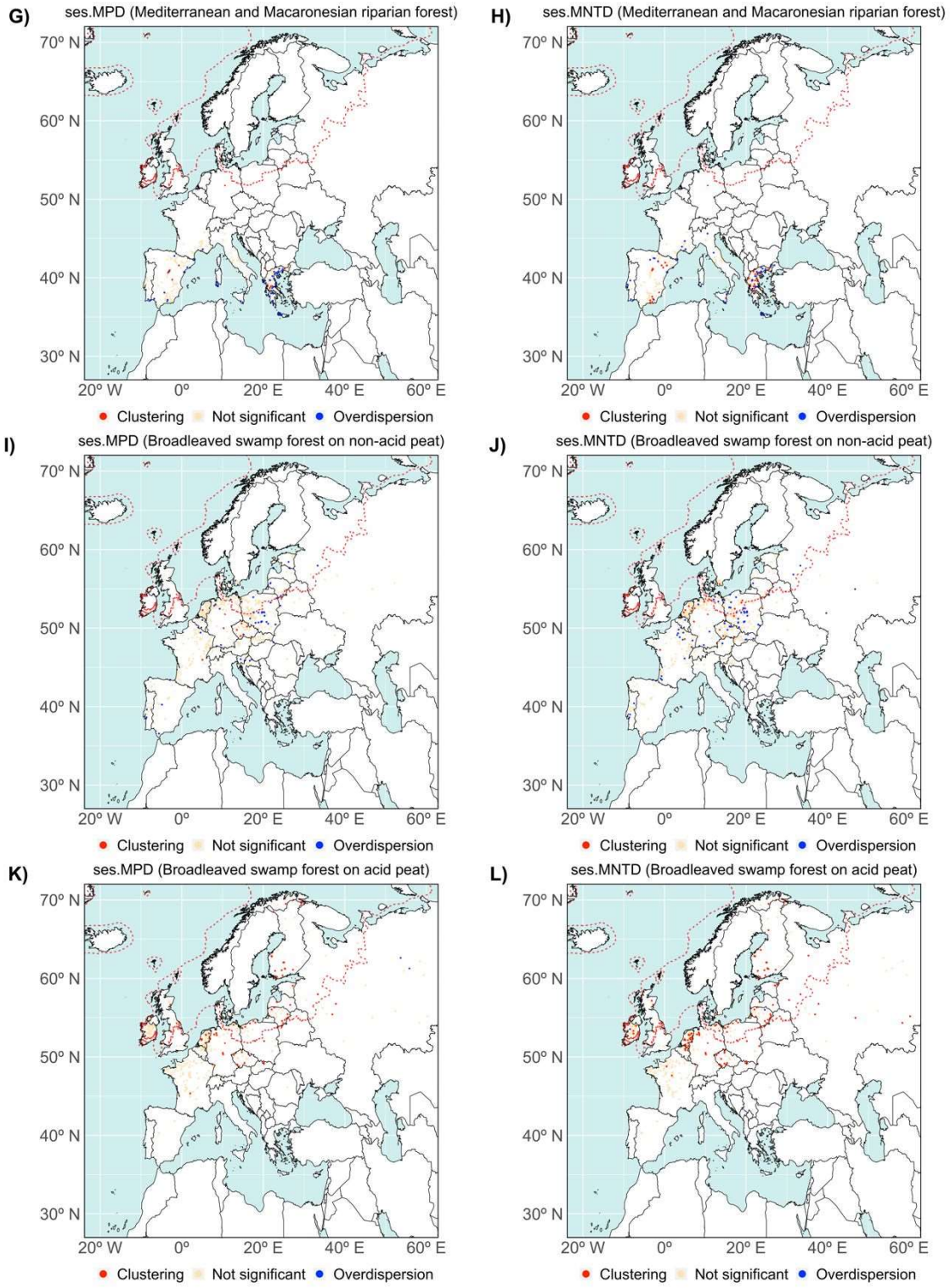
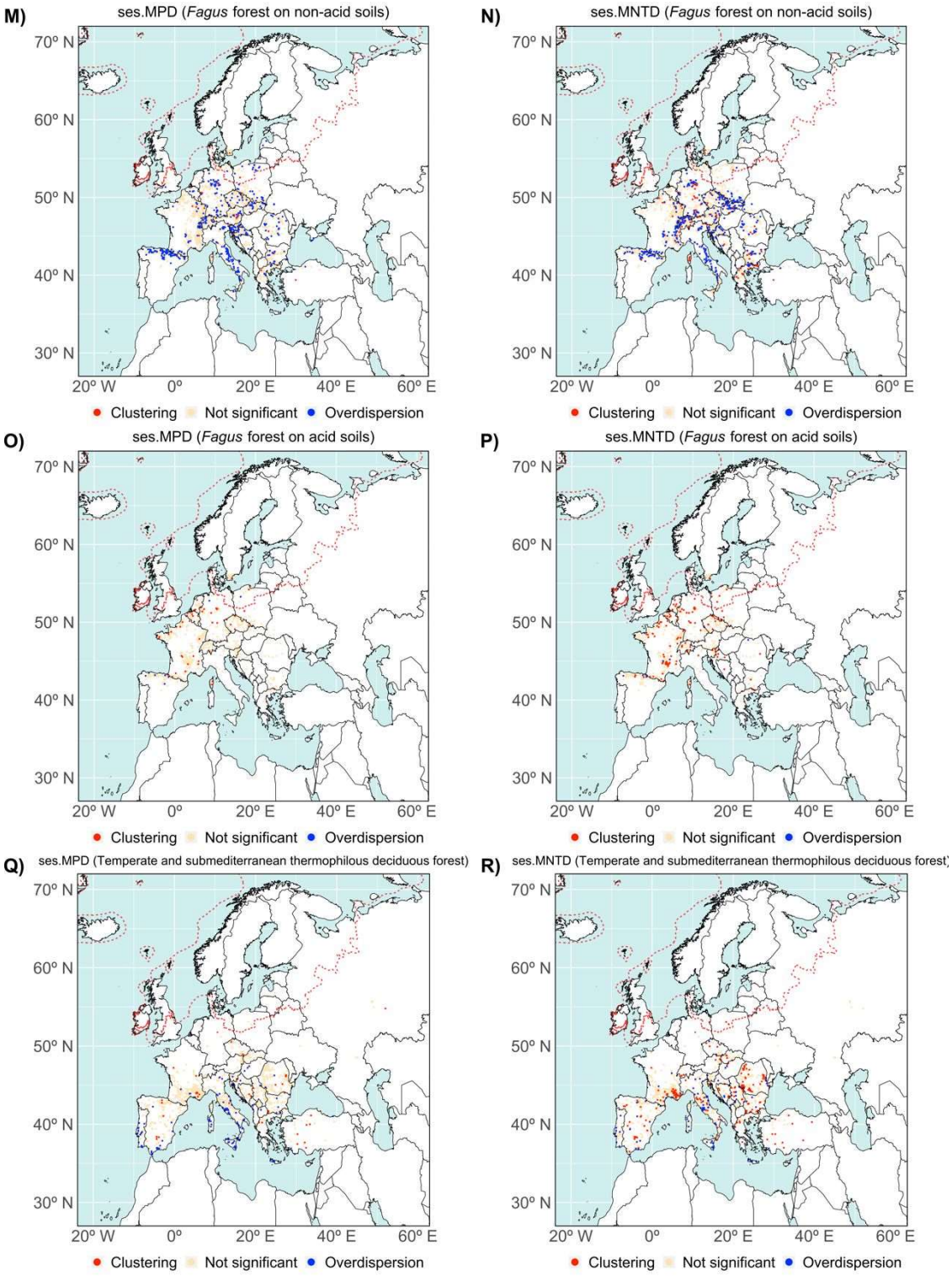
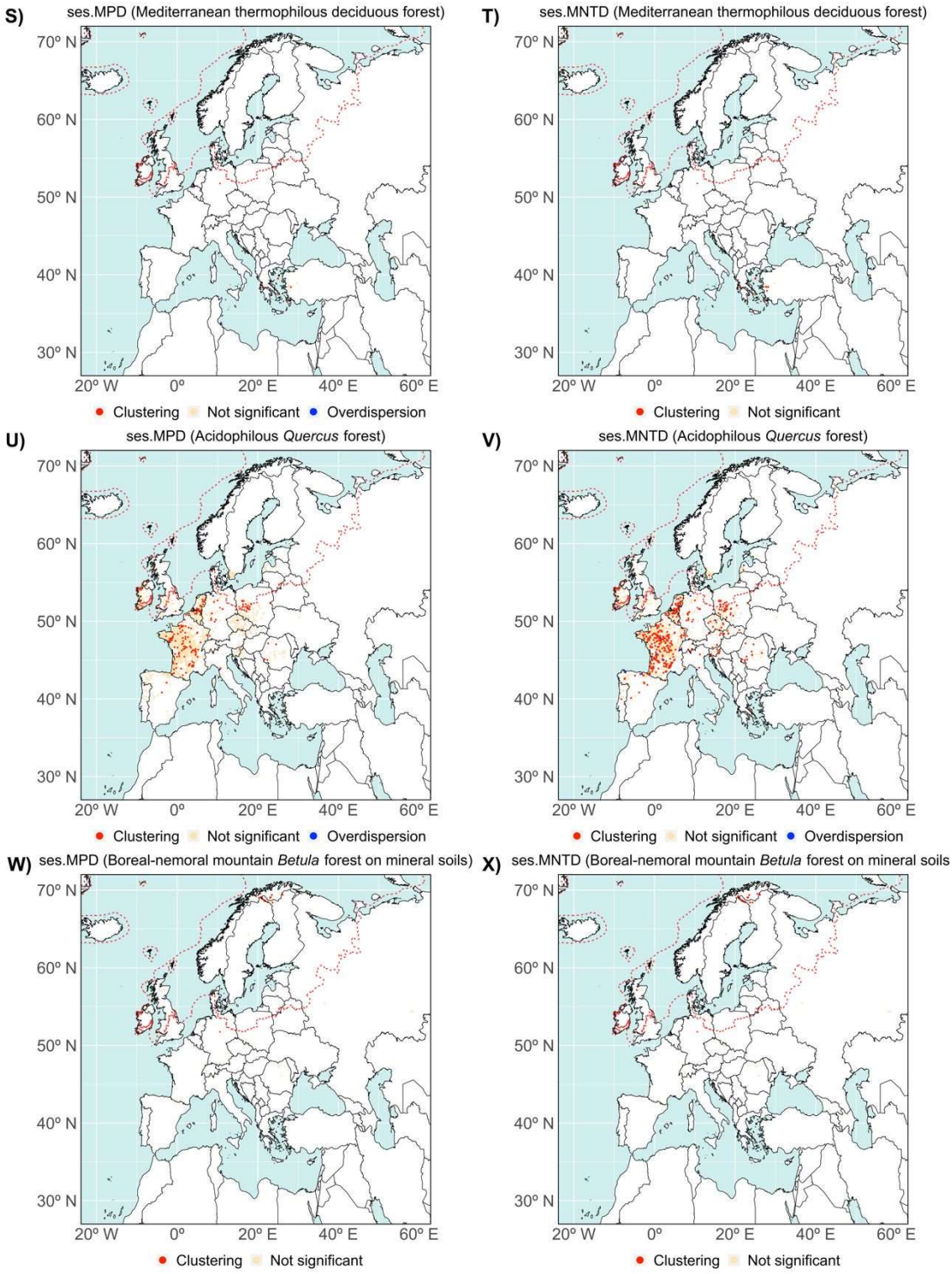


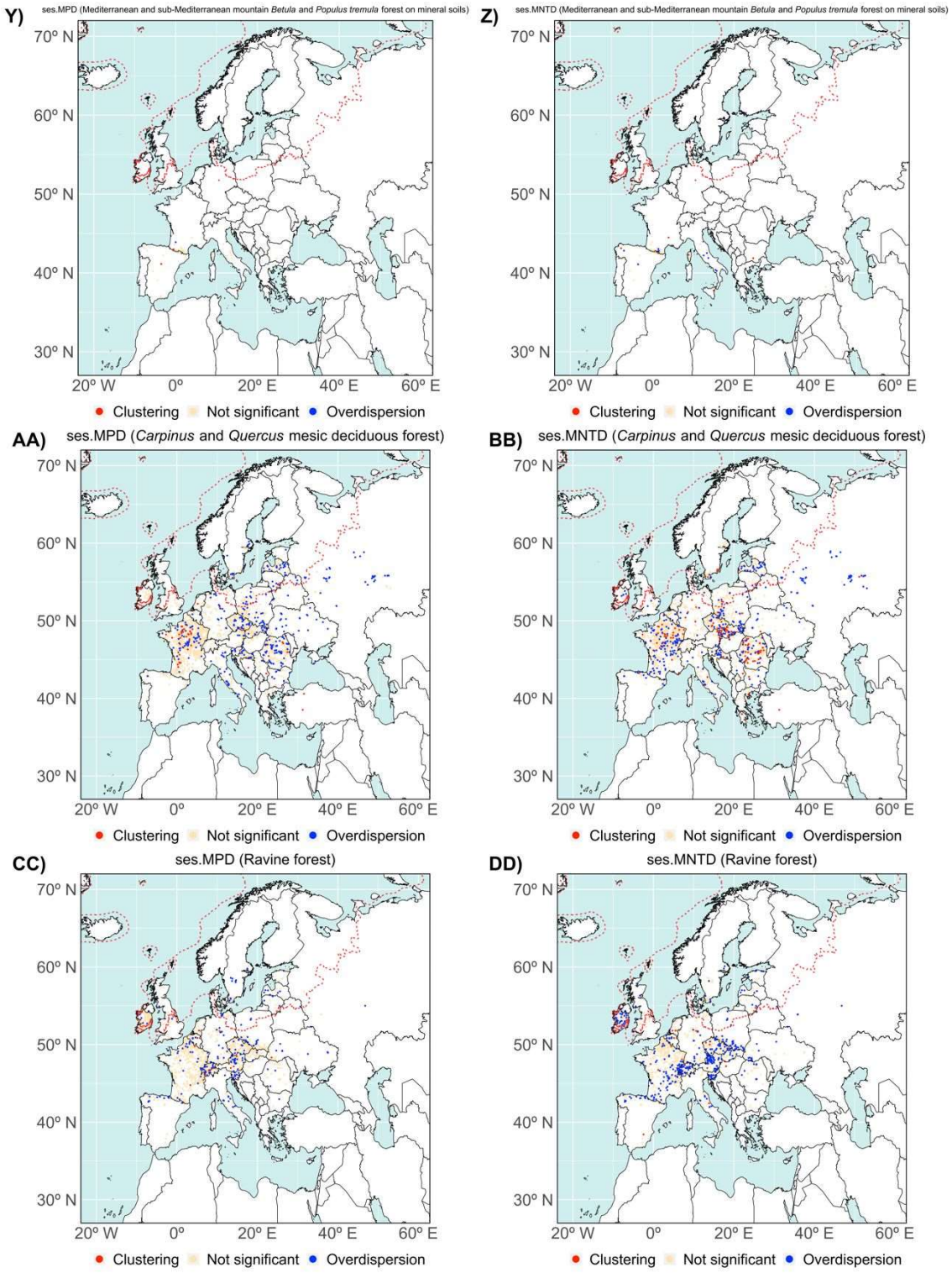
Fig. S7.3: Boxplots (median and quartiles; whiskers show the 5th and 95th percentile) for ses.MPD and ses.MNTD in broadleaved deciduous vegetation plots sorted by EUNIS habitat types. The 'broadleaved deciduous forests' category refer to plots that could not be subdivided into more precise EUNIS habitat types.


A)
ses.MPD (Temperate and boreal Salix and Populus riparian forest)
B)
ses.MNTD (Temperate and boreal Salix and Populus riparian forest)
C)
ses.MPD (Riparian Alnus forest)
D)
ses.MNTD (Riparian Alnus forest)
E)
ses.MPD (Temperate and boreal hardwood riparian forest)
F)
ses.MNTD (Temperate and boreal hardwood riparian forest)
70° N
60° N
50° N
40° N
30° N
20° W
0°
20° E
40° E
60° E
Clustering
Not significant
Overdispersion


G)
ses.MPD (Mediterranean and Macaronesian riparian forest)
H)
ses.MNTD (Mediterranean and Macaronesian riparian forest)
I)
ses.MPD (Broadleaved swamp forest on non-acid peat)
J)
ses.MNTD (Broadleaved swamp forest on non-acid peat)
K)
ses.MPD (Broadleaved swamp forest on acid peat)
L)
ses.MNTD (Broadleaved swamp forest on acid peat)
70° N
60° N
50° N
40° N
30° N
20° W
0°
20° E
40° E
60° E
Clustering
Not significant
Overdispersion


M)
ses.MPD (*Fagus* forest on non-acid soils)
N)
ses.MNTD (*Fagus* forest on non-acid soils)
O)
ses.MPD (*Fagus* forest on acid soils)
P)
ses.MNTD (*Fagus* forest on acid soils)
Q)
ses.MPD (Temperate and submediterranean thermophilous deciduous forest)
R)
ses.MNTD (Temperate and submediterranean thermophilous deciduous forest)
70° N
60° N
50° N
40° N
30° N
20° W
0°
20° E
40° E
60° E
Clustering
Not significant
Overdispersion


S)
ses.MPD (Mediterranean thermophilous deciduous forest)
T)
ses.MNTD (Mediterranean thermophilous deciduous forest)
U)
ses.MPD (Acidophilous Quercus forest)
V)
ses.MNTD (Acidophilous Quercus forest)
W)
ses.MPD (Boreal-nemoral mountain Betula forest on mineral soils)
X)
ses.MNTD (Boreal-nemoral mountain Betula forest on mineral soils
70° N
60° N
50° N
40° N
30° N
20° W
0°
20° E
40° E
60° E
Clustering
Not significant
Overdispersion


Y)
ses.MPD (Mediterranean and sub-Mediterranean mountain Betula and Populus tremula forest on mineral soils)
Z)
ses.MNTD (Mediterranean and sub-Mediterranean mountain Betula and Populus tremula forest on mineral soils)
AA)
ses.MPD (Carpinus and Quercus mesic deciduous forest)
BB)
ses.MNTD (Carpinus and Quercus mesic deciduous forest)
CC)
ses.MPD (Ravine forest)
DD)
ses.MNTD (Ravine forest)
70° N
60° N
50° N
40° N
30° N
20° W
0°
20° E
40° E
60° E
Clustering
Not significant
Overdispersion

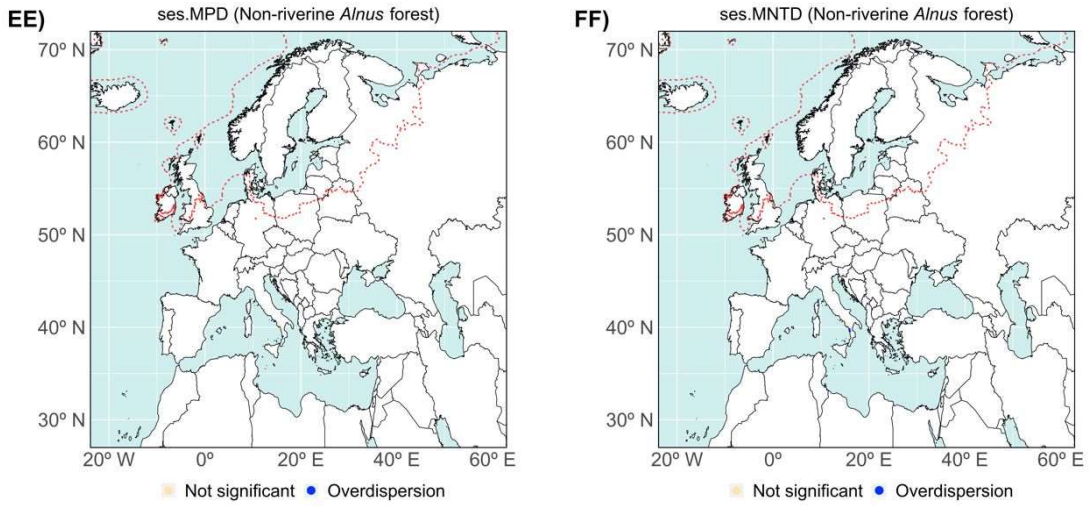


Fig. S7.4: Standardized effect sizes of MPD and MNTD (ses.MPD and ses.MNTD) for broadleaved deciduous forests sorted by EUNIS habitat type (angiosperms only). Red and blue points indicate significantly clustered and overdispersed plots, respectively. Light-orange denotes plots with random phylogenetic structure. We only mapped habitats with at least 10 plots (Appendix S1: Table S1.2). The dashed red line indicates the extent of the ice-sheet over 50º N during the LGM.

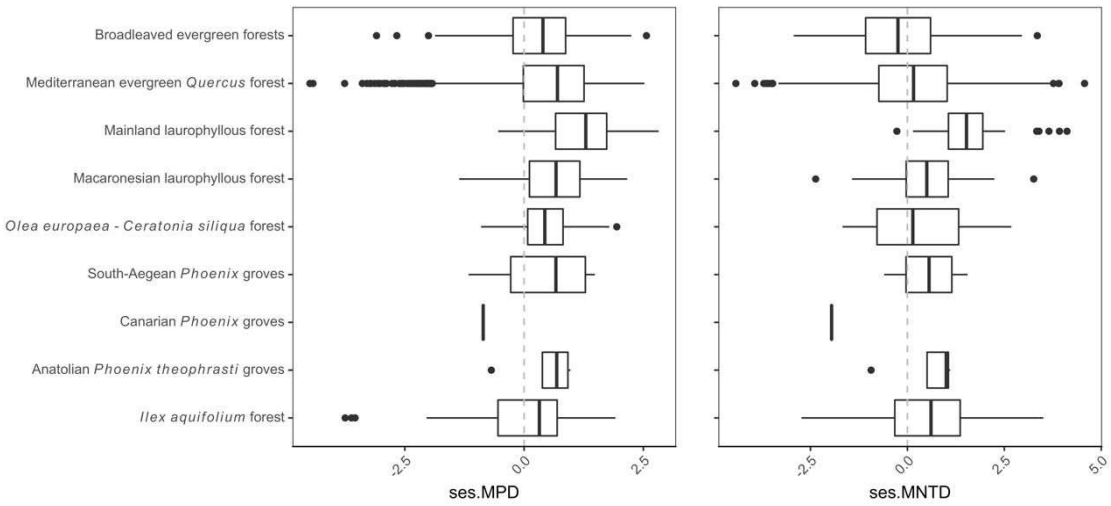
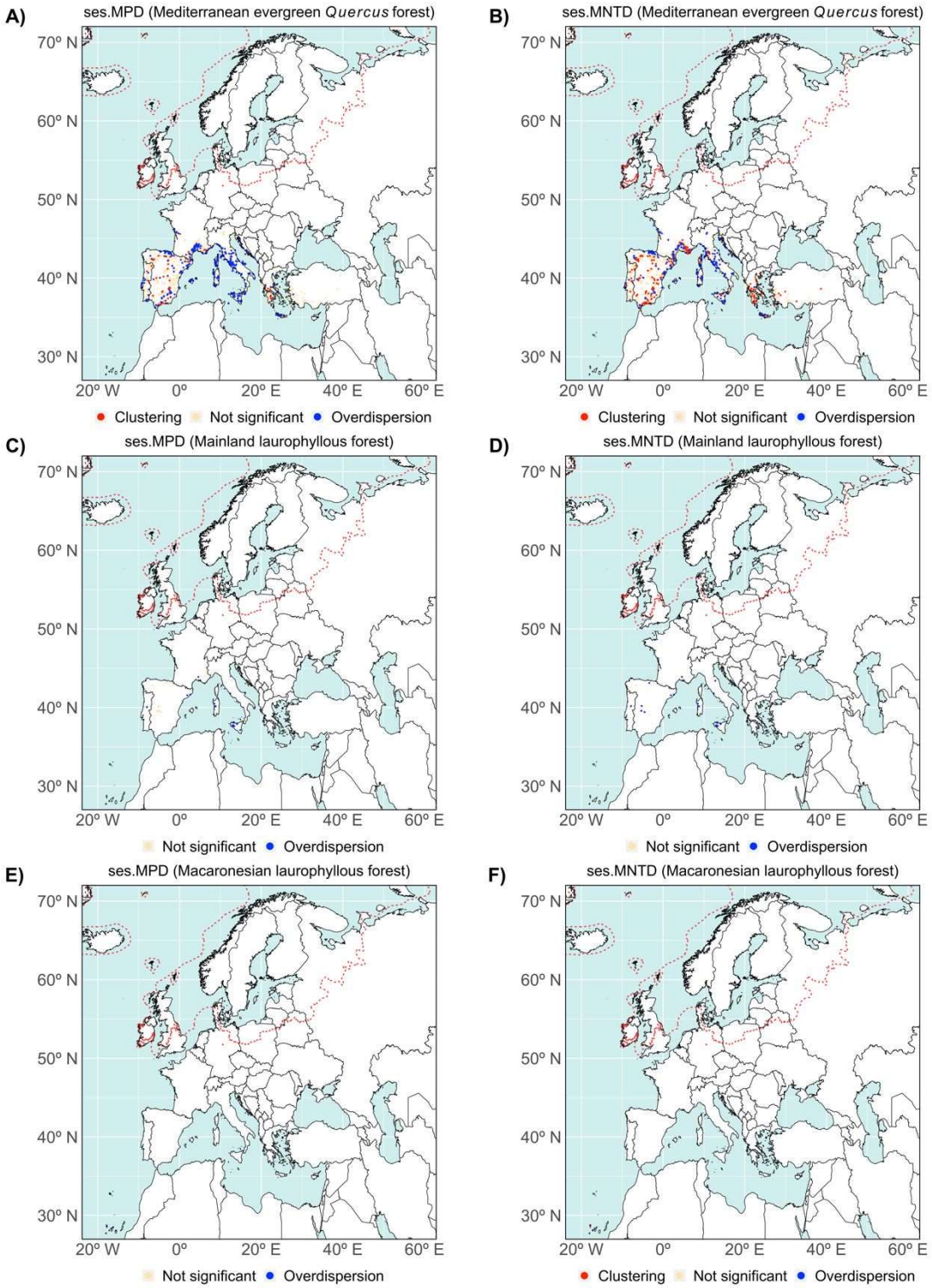


Fig. S7.5: Boxplots (median and quartiles; whiskers show the 5th and 95th percentile) for ses.MPD and ses.MNTD in broadleaved evergreen vegetation plots sorted by EUNIS habitat types. The 'broadleaved evergreen forests' category refer to plots that could not be subdivided into more precise EUNIS habitat types.


A)
ses.MPD (Mediterranean evergreen Quercus forest)
B)
ses.MNTD (Mediterranean evergreen Quercus forest)
C)
ses.MPD (Mainland laurophyllous forest)
D)
ses.MNTD (Mainland laurophyllous forest)
E)
ses.MPD (Macaronesian laurophyllous forest)
F)
ses.MNTD (Macaronesian laurophyllous forest)
70° N
60° N
50° N
40° N
30° N
20° W
0°
20° E
40° E
60° E
Clustering
Not significant
Overdispersion

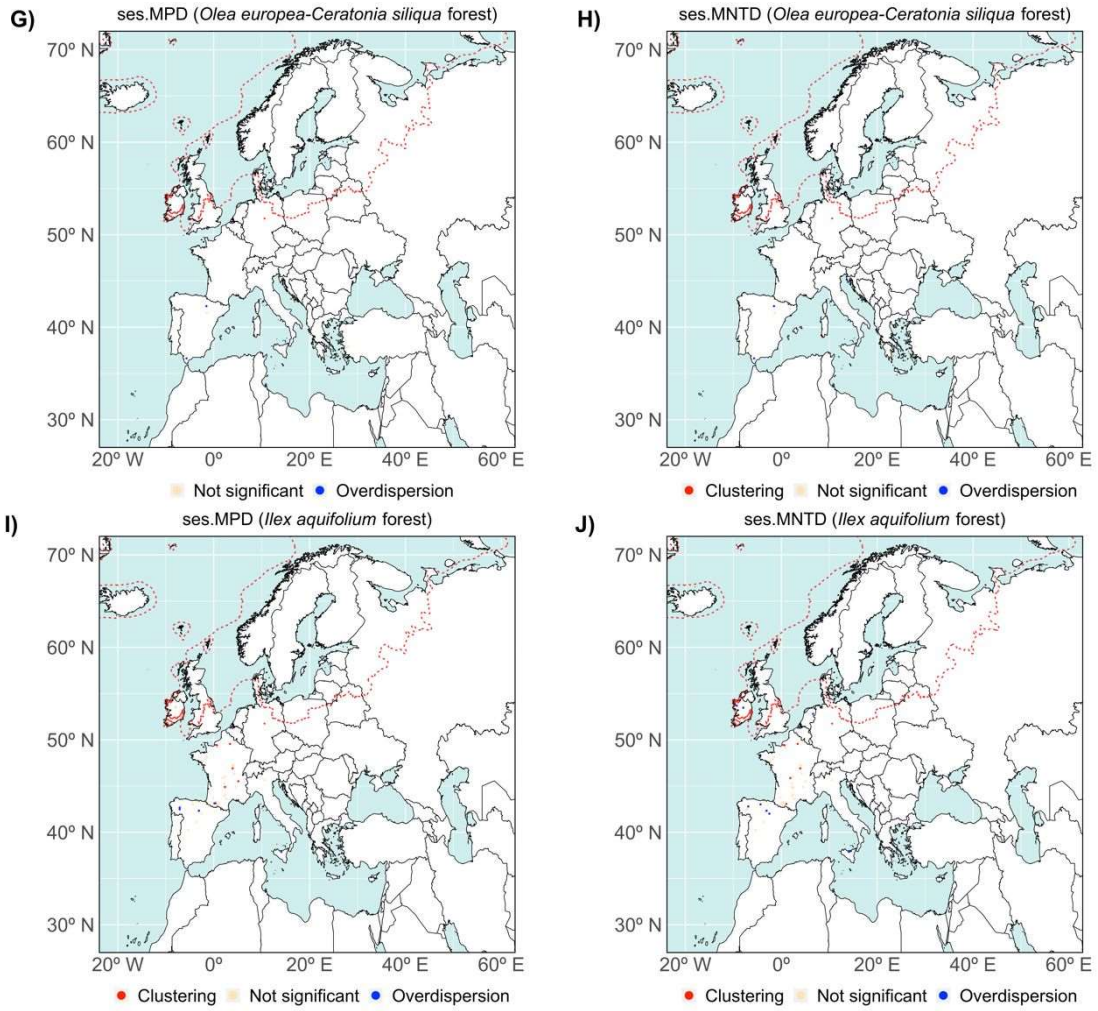


Fig. S7.6: Standardized effect sizes of MPD and MNTD (ses.MPD and ses.MNTD) for broadleaved evergreen forests sorted by EUNIS habitat type (angiosperms only). Red and blue points indicate significantly clustered and overdispersed plots, respectively. Light-orange denotes plots with random phylogenetic structure. We only mapped habitats with at least 10 plots (Appendix S1: Table S1.2). Dashed line indicates the extent of the ice-sheet over 50º N during the LGM.

Appendix S8: Additional results of the *nodesig* analysis

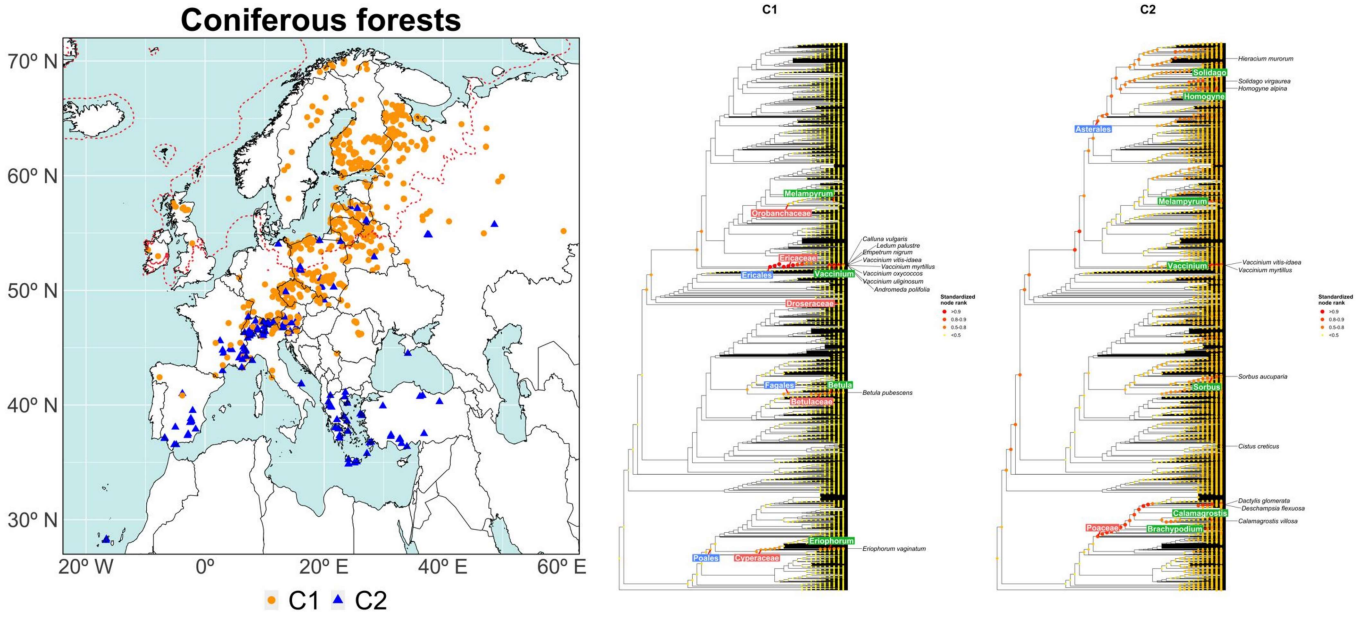


Fig. S8.1: Locations of significantly clustered plots of coniferous forests. Forest vegetation plots were grouped based on the relative contribution of each node to the phylogenetic structure, and are represented by different colors and shapes. For each group, we plotted on top of the phylogeny those nodes with standardized node rank > 0.5 (C1 and C2). Then, we selected the 20% of the nodes with the highest standardized node rank (i.e., the highest contribution to clustering; see Methods), and labelled those corresponding to the whole genera (green), families (red), or orders (blue). For each group, we also included the 15 most abundant species. The dashed red line indicates the extent of the ice-sheet over 50º N during the LGM.

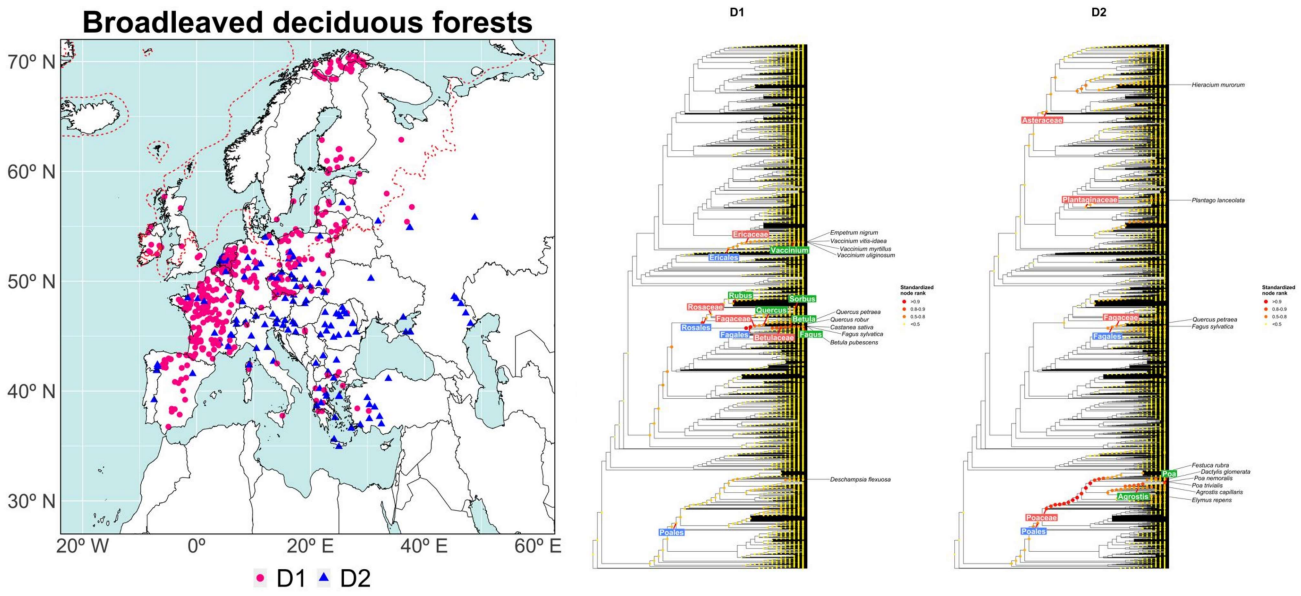
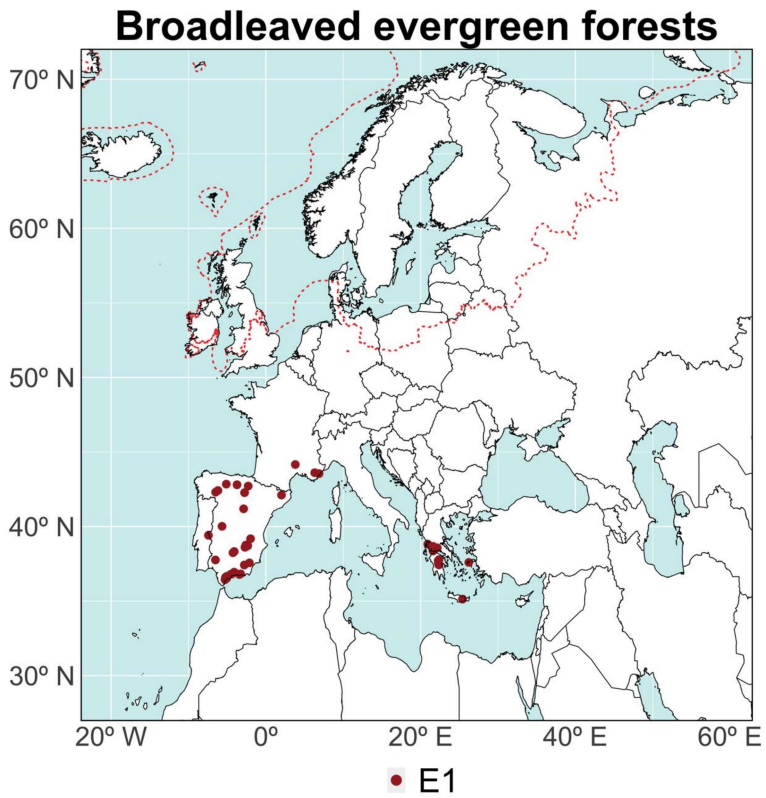


Fig. S8.2: Locations of significantly clustered plots of broadleaved deciduous forests. Forest vegetation plots were grouped based on the relative contribution of each node to the phylogenetic structure, and are represented by different colors and shapes. For each group, we plotted on top of the phylogeny those nodes with standardized node rank > 0.5 (D1 and D2). Then, we selected the 20% of the nodes with highest standardized node rank (i.e., the highest contribution to clustering; see Methods), and labelled those corresponding to whole genus

(green), families (red), or orders (blue). For each group, we have also included the 15 most abundant species. The dashed red line indicates the extent of the ice-sheet over 50º N during the LGM.


Broadleaved evergreen forests
70° N
60° N
50° N
40° N
30° N
20° W
0°
20° E
40° E
60° E
E1

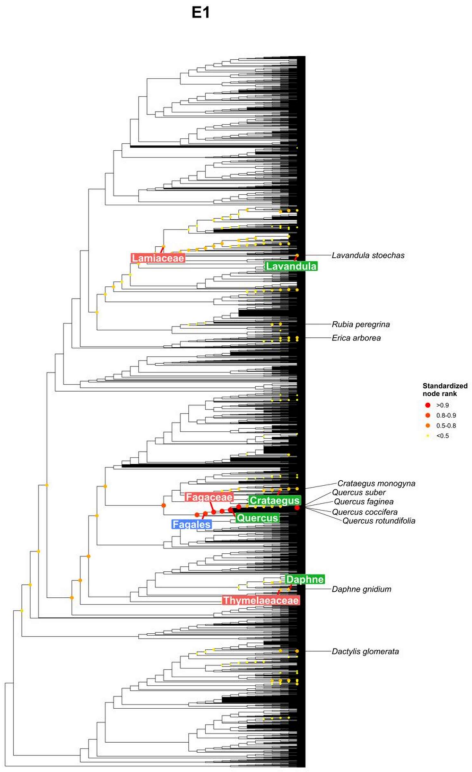

E1
Lamiaceae
Lavandula
Lavandula stoechas
Rubia peregrina
Erica arborea
Standardized node rank
>0.9
0.8-0.9
0.5-0.8
<0.5
Fagaceae
Castanea
Quercus
Fagales
Castanea monogyna
Quercus suber
Quercus faginea
Quercus coccifera
Quercus rotundifolia
Daphne
Daphne gnidium
Thymelaeaceae
Dactylis glomerata

Fig. S8.3: Locations of significantly clustered plots of broadleaved evergreen forests. We plotted on top of the phylogeny those nodes with standardized node rank > 0.5. Then, we selected the 20% of the nodes with highest standardized node rank (i.e., the highest contribution to clustering; see Methods), and labelled those corresponding to whole genus (green), families (red), or orders (blue). We have also included the 15 most abundant species. The dashed red line indicates the extent of the ice-sheet over 50º N during the LGM.