Comparison of the global distribution of functional and phylogenetic diversity in plant communities.

A study to highlight commonalities and differences in the distribution of vascular plants.

Masterarbeit

Angefertigt an der Naturwissenschaftlichen Fakultät I (Biowissenschaften),
Institut für Biologie, Institutsbereich für Geobotanik und Botanischer Garten,
Martin-Luther-Universität Halle-Wittenberg

Zur Erlangung des akademischen Grades

Master of Science (Biologie)

(M.Sc. Biologie, 120 Leistungspunkte)

vorgelegt von

Georg Hähn

geboren am 21. April 1998 in Leipzig

eingereicht am: 18.08.2022

Betreuer: Prof. Helge Bruelheide

Contents

List of figures

1. Introduction

The niche of a species is defined as the different environmental conditions under which a specific species can grow, survive and reproduce (Hutchinson 1978). Since resources (such as light and water) are finite, co-occurring plant species in a community partition the available niche space. Investigating how the niche space is occupied by co-existing species in plant communities is key for understanding the underlying mechanisms behind plant community assembly. However, it has proven conceptually and methodologically difficult to analyse the community niche space (Silvertown 2004, Pearman et al. 2008).

While the analysis of a species' fundamental niche, that is the range of abiotic conditions a species requires and tolerates, needs intensive experimental work (Duarte et al. 2010, Oliveira et al. 2018), the analysis of a species' realized niche, that is the range of conditions under which a species actually occurs, can be analysed from species occurrence and co-occurrence data (Rutherford et al. 1995, Gallego-Tévar et al. 2018). For the latter, two approaches have emerged that make use of a species' relationship to the other residents in the community. One focuses on the distance between species in terms of functional traits (McGill et al. 2006) and another in terms of their phylogeny (Webb et al. 2002).

Functional traits are characteristics related to an organism's survival, growth and reproduction measured at the individual level (Violle et al. 2007). Different functional traits represent different axes of a species' niche, that is its relationship with different environmental conditions. For example, specific leaf area (SLA) can be understood as a proxy of the leaf economics spectrum (LES), where low SLA values imply that a species tolerates environmental stress, while species with high SLA values make the most possible use of favourable environmental conditions (Reich et al. 2003, Wright et al. 2004 & 2005). Plant height indicates the aboveground performance, e.g. the ability to capture light, of a species in the realized niche (Bruelheide et al. 2018). Belowground performance, such as nutrient and water uptake, can be captured by a plant species specific root length (SRL) (García-Palacios et al. 2012). Focusing on multiple traits might help increase the understanding of a species' niche space, depending on how the traits covary (Statzner et al. 2001, Violle and Jiang 2009, Bruelheide et al. 2018).

While plant species do have a unique combination of functional traits, they also represent the endpoint of unique evolutionary histories (Reich et al. 2003). This evolutionary history cannot be captured by considering functional traits alone, because different lineages might have evolved similar trait values or states (trait convergence; Cavender-Bares et al. 2004, Ackerly 2009). In contrast, there are plant characteristics that are strongly linked to the species' evolution, in particular those related to biotic interactions. For example, metabolic characteristics that confer herbivore and pathogen

resistance are often shared and conserved within a phylogenetic clade (Núñez-Farfán et al. 2007, Castagneyrol et al. 2014). Considering the phylogenetic relatedness of the resident species in a community might help understanding the niche differentiation in this community.

The distances on the phylogeny of species present in a community can be understood as phylogenetic diversity and can be used as a proxy for unmeasured trait diversity (Faith 1992). However, the relationship between phylogenetic and functional diversity in community assemblage on a global scale remains unclear. As phylogenetic diversity is a biodiversity measure with the ability to capture homologous traits (Faith 1992), a strong relationship between phylogenetic and functional diversity would be expected. It has been shown that this is true when multiple traits are considered, so that phylogenetic diversity can be considered a proxy for high dimensional trait diversity (Tucker et al. 2018).

From the possible ways any given functional trait might map on the phylogenetic tree, the two extremes are trait conservatism or convergence, reflected in clustering or overdispersion of trait values on the phylogeny (Tofts and Silvertown 2000, Webb 2000, Cavender-Bares et al. 2004). Trait convergence indicates that species with high phylogenetic distance have more similar functional traits compared to phylogenetically more related species (Tofts and Silvertown 2000, Webb 2000). These are two extreme situations, however, and it is likely that different trait assemblages might be correlated with recent climate conditions. In addition, whether these functional traits are conserved or show convergence in regions with different climatic histories remains unclear (Reich et al. 2003).

The evolution of species and their assemblage is expected to be shaped by different long-term events such as continental drift or climate change after ice ages, resulting in different phylogenetic species' relatedness at specific regions (Gugerli and Holderegger 2001, Pennington et al. 2004). The velocity of climatic change after the last glacial maximum (LGM), for instance, has been shown to relate to patterns of phylogenetic diversity (Prentice et al. 1991, Wang et al. 2017), resulting in phylogenetic turnovers in regions with higher climatic change after the LGM (Cubino et al. 2021). Regions with rapid climate changes after the LGM went from annual mean temperatures below zero and a closed ice shield to climatic conditions where plants were able to establish, grow and reproduce (Gugerli and Holderegger 2001). So far, geographical patterns of phylogenetic diversity and the underlying evolutionary processes of plant species have only been shown for specific regions or specific taxa (Wright et al. 2000, Saladin et al. 2020, Qian et al. 2020).

This master thesis aims to fill the knowledge gaps related to the relationship between phylogenetic and functional diversity at the global scale, when a large number of species is considered, and to understand how the patterns of phylogenetic and functional diversity are related to past and present

climatic conditions. Based on "sPlot" – the global vegetation-plot database, the following hypotheses were tested:

H1:   Functional and phylogenetic diversity are related at the global scale.

H2:   Spatial patterns of functional and phylogenetic diversity differ from each other.

H3:   Distribution pattern of functional diversity depends on current climatic conditions.

H4:   Spatial pattern of phylogenetic diversity depends on past climatic events, i.e. climatic conditions after the last glacial maximum.

2. Methods

2.1 Species community data

The vegetation-plot database "sPlot" (www.idiv.de/splot) used in this thesis is a harmonised collection of 161 national- or regional-scale vegetation-plot datasets. The records provide geo-referenced information on presence and abundance of all vascular plants co-occurring in a sampling area, i.e. a vegetation-plot. The "sPlot" database version 3.0 contains 1,977,637 vegetation-plots collected between 1873 and 2019, including 76,912 species of vascular plant taxa (for version 2.1, see Bruelheide et al. 2019). The database provides information on vegetation type (i.e., forest, vs. non-forest) and biome (e.g. alpine, dry mid latitudes, tropics with summer rain) of each plot. In addition, the global, open-access dataset of vegetation-plots "sPlotOpen", which is an environmentally balanced subset of the "sPlot" vegetation database, was used to evaluate the method described below (Sabatini et al. 2021). Plots were only considered in this thesis if the cumulative coverage of those species for which both traits and phylogenetic information was available accounted for at least 50% of the relative cover of that plot. Despite the database is unbalance in covered areas, sPlot presents plots sampled in six continents across most biomes.

2.2 Phylogenetic Tree

All analyses were performed in R 4.1.3 (R Core Team 2022), all scripts are available at https://git.idiv.de/gh87wuma/master-thesis-pd-fd-splot. For all species in sPlot, a phylogenetic tree was built. The phylogenetic tree was calculated using the function *phylo.maker* from the R package *V.PhyloMaker* (Jin and Qian 2019). The phylogenetic backbone of the package is the combination of GenBank taxa with a backbone provided by the Open Tree of Life at version 9.1 (GBOTB) for seed plants (Smith and Brown 2018) and the clade in the phylogeny for pteridophytes (Zanne et al. 2014). Missing genera were inserted to the half point of the family tree. This approach was evaluated by Qian and Jin (2016), who showed that phylogenetic indices based on the calculated tree were highly correlated with indices based on the "PhytoPhylo megaphylogeny" (updated phylogenetic tree from Zanne et al. 2014). Species that could not be inserted by the *phylo.maker* were bound to the half of the terminal level of a sister species, if only one species was available in this genus, or to the most recent ancestor (MRCA), if the genus included more than one species. This additional binding was done with the *bind.node* function from the *phytools* package (Revell 2012).

2.3 Phylogenetic indices

For each vegetation-plot two phylogenetic indices were calculated using the *picante* package (Kembel et al. 2010). The mean pairwise distance (MPD, Webb et al. 2002) and Rao`s Quadratic Entropy (RQE, Rao 1982) were calculated based on the cophenetic distance of all the *n* species in the phylogenetic tree, pruned to contain only the species in that plot:

$$MPD = -1 \times (\overline{X_{obs}} - \frac{\overline{X_n}}{SD\ X_n})$$

$$RQE = \sum_{i=1}^{n} \sum_{j=i+1}^{n} d_{ij} p_i p_j$$

Where $\overline{X_{obs}}$ is the mean distance among all observed taxa in the community, $\overline{X_n}$ is the mean of all possible pairs of n taxa randomly distributed on the phylogeny and $SD\ X_n$ the standard deviation of all possible pairs of n taxa. For RQE d_{ij} is the nonnegative number representing the difference between species (i.e. cophenetic distance) *i* and *j*, p_i and p_j are the relative abundances of the species in the community, which in most cases is expressed as relative cover (i.e. cover of species *i* or *j*, divided by the sum of cover of all species in the community). All indices as well as relative cover were calculated based on the species present in the phylogeny and with known traits.

2.4 Functional traits
Three functional traits were selected as it was shown that functional diversity is more likely to be correlated to phylogenetic diversity with higher number of measured traits (Tucker et al. 2018). The selected functional traits, i.e., specific leaf area (SLA), plant height and specific root length (SLR), were available from the "TRY" database (Kattge et al. 2020) after gap-filling (Shan et al. 2012, Fazayeli et al. 2014, Schrodt et al. 2015). These three traits were selected to represent three different axes of a species' niche (Bruelheide et al. 2018). SLA is a proxy of the leaf economics spectrum, plant height shows the ability of a plant to capture light and belowground performance, such as nutrient uptake, is represented by SRL.

2.5 Functional indices
The two functional diversity indices, i.e. functional dispersion (FDis) and Rao's quadratic entropy (RQEF), were calculated for all vegetation-plot records based on the three selected plant traits. These two indices were selected because they can be considered functional analogues to the phylogenetic diversity indices described above. FDis, when calculated on presence-absence data, is the average distance of the species of a community to the community centroid (Anderson et al. 2006), and therefore, comparable to the calculated mean of pairwise species distances on the phylogenetic tree (MPD). Rao's quadratic entropy is based on a distance matrix which can be computed both on phylogenetic and functional trait dissimilarities. FDis was calculated using the *dbfd* function from the *FD* package (Laliberté and Legendre 2010 & Laliberté et al. 2014). The trait-based distance matrix to calculate Rao's quadratic entropy was computed using the *compute_dist_matrix* function from the *funrar* package (Grenié et al. 2017) based on the Gower distance metric (Gower 1971).

$$FDis = \frac{\sum_{j=1}^{n} z_j}{n}$$

Where z_j is the distance of species *j* present in the community to the centroid of the community and *n* the number of species.

2.6. Standardized effect size

Functional and phylogenetic diversity indices are known to depend on species richness (Petchey and Gaston 2002, Cadotte et al. 2009). Especially, for functional diversity, a higher number of species in a community is more likely to diverge in trait values than communities with fewer species (Petchey and Gaston 2002). Therefore, the standardised effect size of each diversity index was calculated for each vegetation-plot (Botta-Dukát 2018), where the number of species remained fixed and random species were allowed to be drawn among the species present in the whole dataset. Indices and standard deviation were calculated from 499 draws of randomly selected trait values for the species present in a given vegetation-plot. As monocultures were excluded from functional or phylogenetic diversity calculations, species richness of vegetation records ranged from two species to 412 species. The mean index of the randomized data was subtracted from the observed index and divided by the standard deviation of the index of the randomised data. For instance, in the case of MPD, this was done as follows:

$$SES.MPD = \frac{MPD_{obs} - \overline{MPD_{rand}}}{SD_{rand}}$$

2.7 Explanatory variables

Worldwide stable climate conditions after the last glacial maximum were derived from the open-access "StableClim" v.1.1 dataset, containing estimates of climate variability from 21,000 years ago at a 2.5° spatial resolution (Brown et al. 2020). Climate variability represents rapid global warming during the last deglaciation during the Bølling-Allerød transition (Renssen and Isarin 2001) on land and sea. A 90% threshold was used to identify rapid climatic changes. This threshold was calculated based on the ensembled average bootstrapped pre-industrial cumulative distribution function of trends of the global mean temperature to identify rates of centennial changes corresponding with rapid climate change (Brown et al. 2020). As this master thesis focuses on the effect of stable climate conditions, the conditions of the stable climate were defined as the inverse climate variability, mean temperature change per year ($\frac{1}{°C\ p.a.}$).

Current climate conditions (1981-2010) were represented by the 19 bioclimatic variables from "CHELSA" v.2.1 (Karger et al. 2017, 2018). A principal component analyses (PCA) was performed using all 19 bioclimatic variables to detect the predictors with the highest loading. Furthermore, a correlation matrix was calculated to select the climate variables with the lowest collinearity. According

to their explanatory power in the PCA and the weakest collinearity, five climate variables were selected. The first selection was based on the PCA and their first two axes (*Figure S 2*), the second selection was done based on the correlation matrix (*Figure S 4*). In addition, the third and fourth axes of the PCA (*Figure S 3*) was checked to avoid collinearity in this space of the PCA. Bioclimatic variables tending to capture seasonality effects were preferred.

All climate variables were extracted for each plot with the *extract* function from the *raster* package (Hijmans et al. 2022).

Four additional predictor variables were extracted from the vegetation-plot database "sPlot": 1) plot size, 2) which plants were recorded in the plot (e.g., complete vegetation, tree only, shrubs and trees), 3) vegetation type (forest vs. non forest) and 4) biome.

2.8 Statistical modelling

A generalised additive model (GAM) was used to analyse the relationship between functional and phylogenetic diversity. A GAM is a generalised linear model in which the linear response depends on unknown smooth functions of the explanatory variables. To account for the spatial structure of the data, the spatial coordinates were included as smooth spherical splines. All GAMs included a basis penalty smoother spline on the sphere (bs = "sos"), applied to the geographic coordinates of every site, thus taking spatial autocorrelation into account. The model was performed using the *gam* function from the *mgcv* package (Wood, 2017), defined as following:

$$gam(\quad SES.RQEF \sim SES.RQEP + s(Longitude, Latitude, bs = "sos"), family = "gaussian",$$
$$method = "REML")$$

SES.RQEF is the standardized effect size of Rao`s quadratic entropy based on the three selected functional plant traits and SES.RQEP is the standardized effect size of Rao`s quadratic entropy based on the phylogenetic distances of species present in the community.

Boosted regression trees (BRTs) were used to model the relationship between the functional and phylogenetic indices, and the explanatory variables, as well as to extract the relative influence of each explanatory variable. BRTs have few prior assumptions and are robust against overfitting and collinearity. Furthermore, they are known to uncover nonlinear relationships as well as interactions among predictors. The parameters of the BRT were set as follows: a tree complexity of five and a bag fraction of 0.5. The learning rate was set to 0.01 with a maximum number of 20,000 trees. To reduce the computational time needed to calculate BRTs, resampling was performed. A global-grid with a cell size of approximately 209,900 km^2 was created and 1,000 plots were randomly selected in each grid cell. This selection and the calculation of the BRTs were repeated 100 times, each time 181,151 plots

were selected. In addition, "sPlotOpen" was used as a benchmark for the random subset strategy used here. The BRTs were calculated using the *gbm.step* routine from the *dismo* package (Hijmans et al. 2021). An explanatory variable was considered relevant in the model if its mean relative influence over the 100 runs was greater than 10%, which is the expected influence of a variable if all the 10 predictors had an equal relative importance.

The variables that were considered as relevant from the BRTs were used in a GAM as explanatory variables of functional and phylogenetic diversity. The same syntax as for the relationship between functional and phylogenetic diversity was used for the GAMs. This means for each response variable (SES.RQEP, SES.MPD, SES.RQEF, SES.FDis) a GAM was performed with the explanatory variables that turned out to be relevant in the BRTs. The spatial coordinates were included as smooth spherical splines in the model as explained above. The prediction for each explanatory variable was performed using the *prediction* function from the R package *marginaleffects* (Arel-Bundock 2022) by predicting the response variable based on the sequence between the minimum and maximum of the selected explanatory variable and the GAM model. The residuals were extracted in the same way based on the prediction of the original data. The spatial smooth on the spherical splines was extracted as the predicted response value.

Functional and phylogenetic variables were plotted as mean for each grid cell with a size of approximately 865 km^2. The spatial smooth obtained from the GAM was plotted in the same way and the same resolution.

3. Results

3.1 Phylogenetic tree

The computed phylogenetic tree contained 68,052 of 76,912 species (88%) from the "sPlot" vegetation database. The additional binding resolved additional 3,802 species, resulting in a total of 71,854 species (93%) in the phylogenetic tree. Of these, 3,348 species were bound at the node of the most recent ancestor (MRCA) of already present sister species in the phylogenetic tree, while 454 species were bound at the half of the terminal level on the family node. The final computed phylogenetic tree contained 32,395 nodes. A total of 31,727 species in the phylogeny also had traits in the "TRY" database. These species were used to calculate the functional and phylogenetic indices. After selecting the plots where the species with both traits and phylogenetic information accounted for at least 50% of the plot relative cover, 1,782,777 out of 1,977,637 plots remained.

3.2 Patterns of functional and phylogenetic diversity

No clear geographical pattern was evident in the coupled distribution of functional and phylogenetic diversity, all combinations of high-high, low-low, high-low and low-high occurred in most areas. However at a closer look some patterns emerge, i.e. vegetation plots with both high functional and high phylogenetic diversity were found in Europe, Japan and New Zealand. Patterns of functional and phylogenetic poor plots were encountered in South-America, Africa, Australia and partly in Europe (*Figure 1*). High functional diversity (above the 95% quantile) expressed by standardized effect size of Rao's quadratic entropy (SES.RQEF) with low phylogenetic diversity (below the 5% quantile) expressed by standardized effect size of Rao's quadratic entropy (SES.RQEP) was found in Central Europe, in the West of North-America and partly in Japan and New-Zealand. High phylogenetic entropy was coupled to low functional entropy in Northern Europe, in the North-East of Russia, in parts of Japan and New Zealand and partly in North-America.

Patterns of functional dispersion expressed by standardized effect size of functional (SES.FDis) and phylogenetic dispersion expressed by the standardized effect size of the mean pairwise distances (SES.MPD) of species present in the community were similiar to the patterns described by SES.RQEF and SES.RQEP (*Figure 2*).

The relationship between functional diversity and phylogenetic diversity was modelled with a generalised additive model (GAM) accounting for the spatial structure. For functional entropy expressed by SES.RQEF a negative correlation with phylogenetic entropy (SES.RQEP) was predicted by the GAM (*Figure 3A*). Including the spatial smoothing term, SES.RQEP explained (*Figure 3C*) 6.78% of the deviance of SES.RQEF, with p-values below 0.001. Similarly, SES.FDis showed a negative relation with SES.MPD (*Figure 3B*) and the explained deviance of the model was 8.15%. For SES.FDis, the GAM

predicted really low values in the West of North-America (*Figure 3D*), while this pattern did not emerge for SES.RQEF. For both functional diversity indices high values were predicted in New Zealand and partly in Europe.
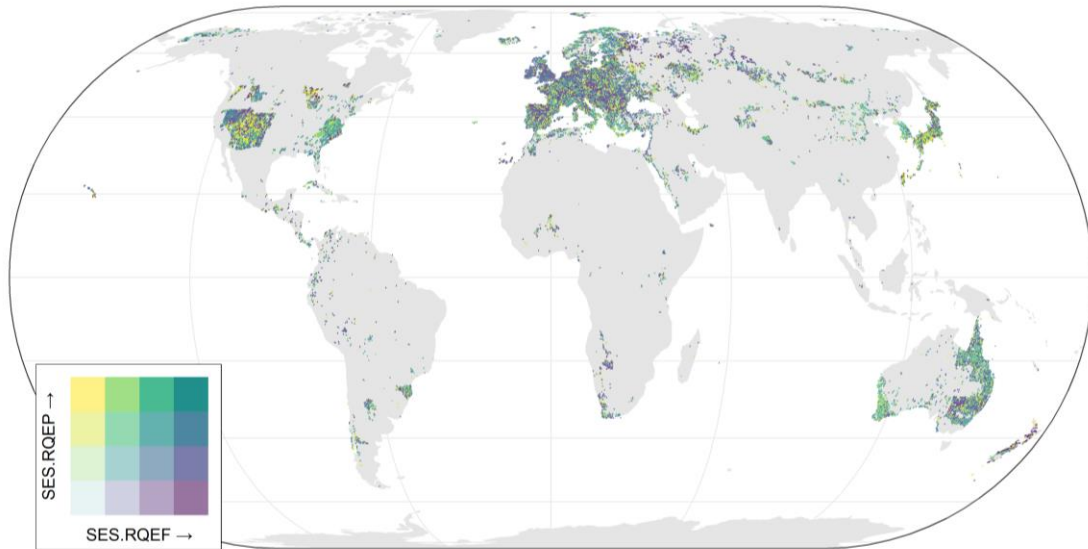
Figure 1: Mean standardized effect sizes of functional and phylogenetic diversity per raster cells (approx. 865 km². The spatial distribution pattern of phylogenetic entropy, indicated by the standard effect size of Rao's quadratic entropy (SES.RQEP), and functional entropy as the standard effect size of Rao's quadratic entropy (SES.RQEF). Functional entropy was based on three functional traits: specific leaf area, plant height and specific root length.
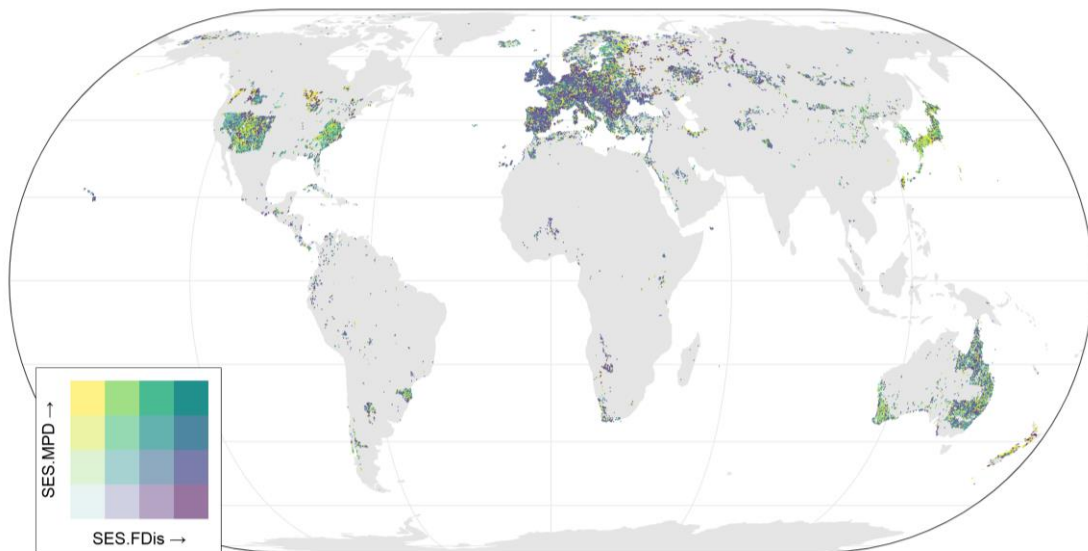
Figure 2: Mean standardized effect sizes of functional and phylogenetic dispersion per raster cells (approx. 865 km²).The spatial distribution pattern of phylogenetic dispersion, indicated by the standard effect size of mean pairwise distances of the species present in the community (SES.MPD), and of functional dispersion expressed as standard effect size of functional dispersion (SES.FDis). Functional dispersion was based on three functional traits: specific leaf area, plant height and specific root length.
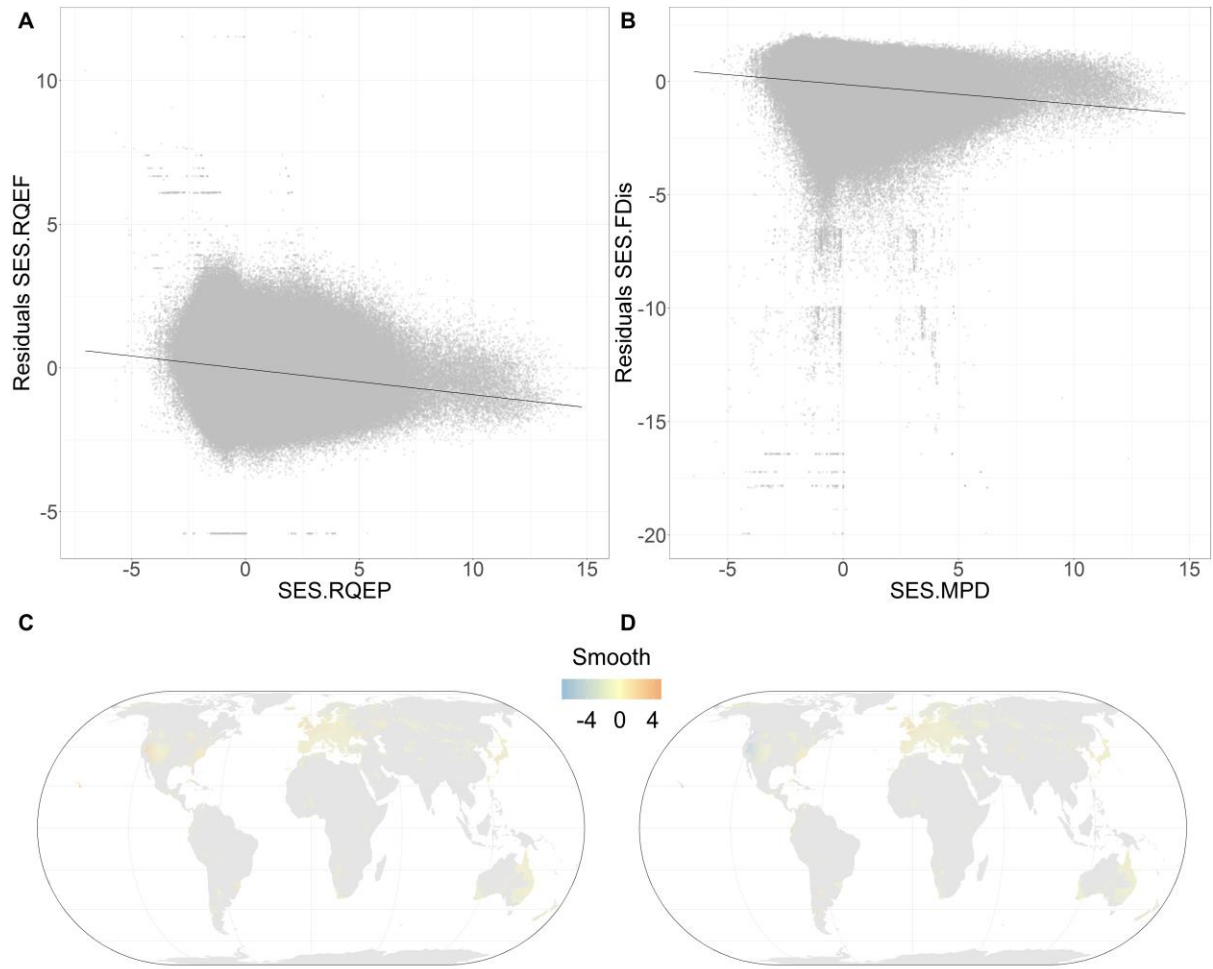
*Figure 3: The relationship of functional and phylogenetic diversity and their spatial distribution. **A** Functional entropy (SES.RQEF) as a function of phylogenetic entropy (SES.RQEP). **B** Standardized effect size of functional dispersion (SES.FDis) as a function of phylogenetic dispersion (SES.MPD). The solid line shows the regression obtained from the GAM. The grey points show the residuals of the GAM after accounting for spatial autocorrelation. Smooth model term showing the expected functional diversity based on space, when modelling **C** SES.RQEF as a function of SES.RQEP and **D** SES.FDis as a function of SES.MPD. Blue and red colours indicate lower and higher than average functional diversity related to space, respectively.*

3.3 Functional and phylogenetic diversity and their relation with climatic conditions

The results from the boosted regression trees (BRTs) suggested that phylogenetic diversity was best explained by stable climate conditions, mean daily air temperature of the wettest quarter and vegetation type (*Figure 4*). Furthermore, the relative influences of all explanatory variables were similar between the two phylogenetic response variables SES.RQEP and SES.MPD. In contrast, no such strong commonality between the response variables was encountered for functional diversity. Here, the most relevant drivers for SES.FDis were plot size, mean monthly precipitation of the warmest quarter and mean daily air temperature of the wettest quarter. In addition, for SES.RQEF the mean daily air temperature of the warmest quarter and stable climate condition were relevant predictors. In particular, the differences of the relative influence of mean monthly precipitation of the warmest quarter and the mean daily air temperature of the warmest quarter between these two indices of functional diversity were higher than those of all other explanatory variables.
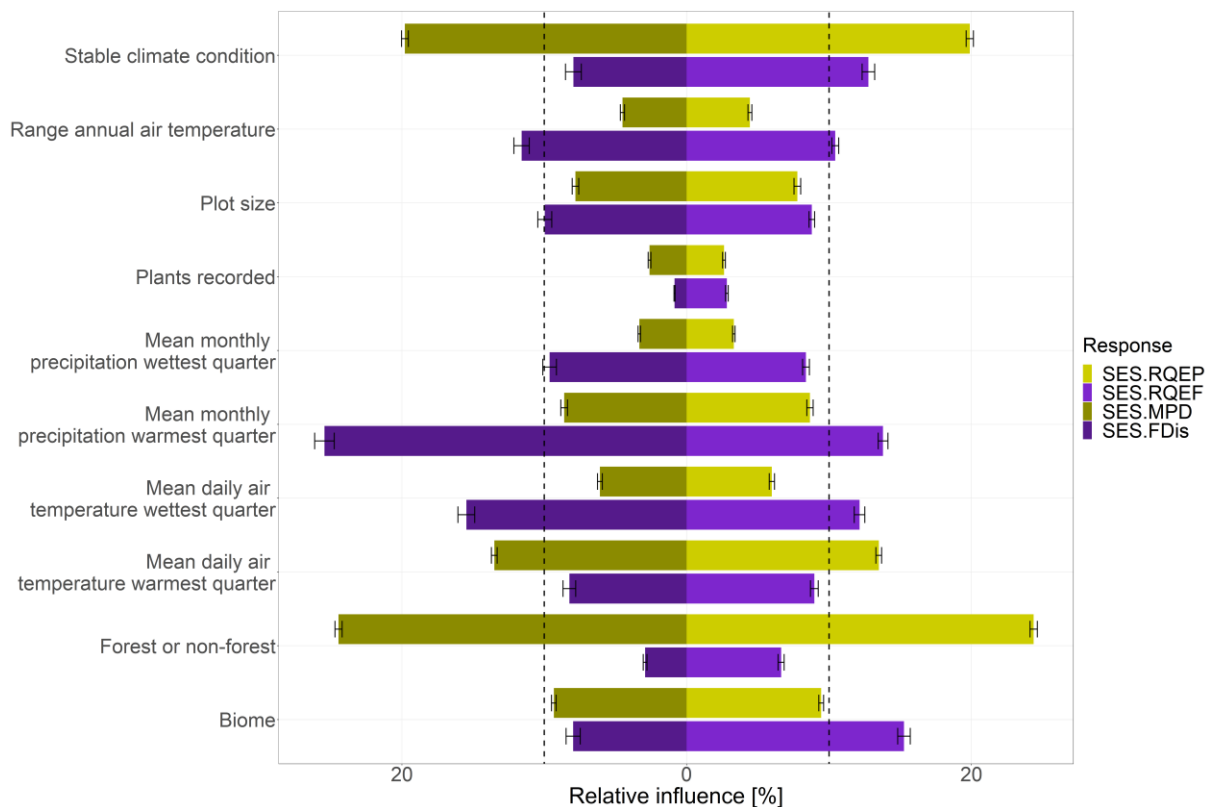
Figure 4: Result of the 100 sample runs of the boosted regression trees (BRTs), based on the five selected explanatory bioclimate variables with the highest loading in the principal component analysis (PCA, Fig. S2 and Fig. S3) and with the weakest correlation (Fig. S4), as well as the additional four explanatory variables available from the vegetation database "sPlot" and stable climate condition. An explanatory variable was considered relevant in the model when its mean relative influence over the 100 runs was greater than 10%, which is the expected influence of a variable if all ten predictors had an equal relative importance. This threshold is marked with the vertical dotted lines. The violet colours indicate the two functional diversity indices (SES.RQEF = standardized effect sizes of Rao's quadratic entropy, SES.FDis = functional dispersion expressed as standardized effect sizes of FDis), while the yellow colours show the phylogenetic indices (SES.RQEP = standardized effect size of Rao's quadratic entropy of the phylogenetic distance of the species present in the community and SES.MPD = standardized effect sizes of mean pairwise phylogenetic distances of the species present in the community).

The results from the BRTs based on the environmental balanced "sPlotOpen" subset (*Figure S5*) showed that for phylogenetic diversity, forest or non-forest was the most relevant explanatory variable followed by biome. All other explanatory variables were not considered relevant. For functional diversity, the results were more similar between the two indices than shown by the 100 random subset runs. Here, mean monthly precipitation of the warmest and mean daily air temperature of the wettest quarter were the most relevant variables, followed by mean daily air temperature of the warmest quarter. Mean monthly precipitation emerged as relevant variable for SES.FDis but not for SES.RQEF. In addition, the range of the annual air temperature and stable climate conditions had a relative influence above 10% for SES.RQEF.

The fitted smooth functions from the BRTs were extracted from the first random subset run and plotted between the 2.5th and the 97.5th quantile of the original data to remove the very extreme points. For functional entropy expressed by SES.RQEF, a negative correlation was predicted for mean monthly precipitation of the warmest quarter (*Figure 5A*) with a rapid decrease at low amounts of precipitation, a minimum at approximately 1,800 mm month^{-1} and a stable negative SES.RQEF at -0.1 for precipitation above 2500 mm month^{-1}. SES.RQEF was predicted to be positive for mean daily air temperature of the wettest quarter at approx. -5°C, while negative estimates were predicted below and above this value, with lowest values above 25°C (*Figure 5B*). For the range of annual air temperature (*Figure 5C*) the smooth term showed an almost linear negative correlation, with positive SES.RQEF between approximately 2°C and 8°C. The smoothing functions of the three mentioned explanatory variables for SES.RQEF were similar to the ones for SES.FDis (*Figure S7*). In addition, for SES.RQEF stable climate condition and the biome were found to be relevant predictors. The smooth term for stable climate condition suggested high SES.RQEF for unstable climate after the last glacial maximum (below $2 \frac{1}{°C \, p.a}$) and for stable climate conditions (above $8 \frac{1}{°C \, p.a}$), while SES.RQEF was low between these values.

For phylogenetic entropy expressed by SES.RQEP stable climate conditions (*Figure 6A*) and mean daily air temperature of the warmest quarter (*Figure 6B*) were identified as relevant variables by the BRTs. The smooth term extracted from the BRTs for SES.RQEP as function of stable climate condition showed high values for unstable climate (below $2 \frac{1}{°C \, p.a}$) and stable climate (above $8 \frac{1}{°C \, p.a}$). Low values were found between 1.5 and $3 \frac{1}{°C \, p.a}$, as well as between 6 and $8 \frac{1}{°C \, p.a}$. The smooth function for mean daily air temperature of the warmest quarter shows a hump-shaped curve with a maximum at approximately 12°C and a minimum at approximately 27°C. In addition, the BRT suggested higher phylogenetic entropy in forest than in non-forest vegetation-plots. The results found for phylogenetic

dispersion expressed by standardized effect size of mean pairwise distances (SES.MPD) of the species present in the community were similar to the ones for SES.RQEP (*Figure S8*).

*Figure 5: Smooth terms of functional entropy expressed as standardized effect size of Rao's quadratic entropy (SES.RQEF) as function of the relevant explanatory variables from the boosted regression trees, based on the first run of the 100 random subsets. **A** SES.RQEF as function of the bioclimatic variable mean monthly precipitation of the warmest quarter, **B** mean daily air temperature of the wettest quarter, **C** range of annual air temperature and **D** stable climate condition. Additional to the four shown explanatory variables the ten-levels factor biome was identified as relevant predictor. For functional dispersion expressed by the standardized effect size of FDis the same patterns were found (Fig. S6), except for stable climate conditions and biome which were not suggested as relevant in the BRTs.*
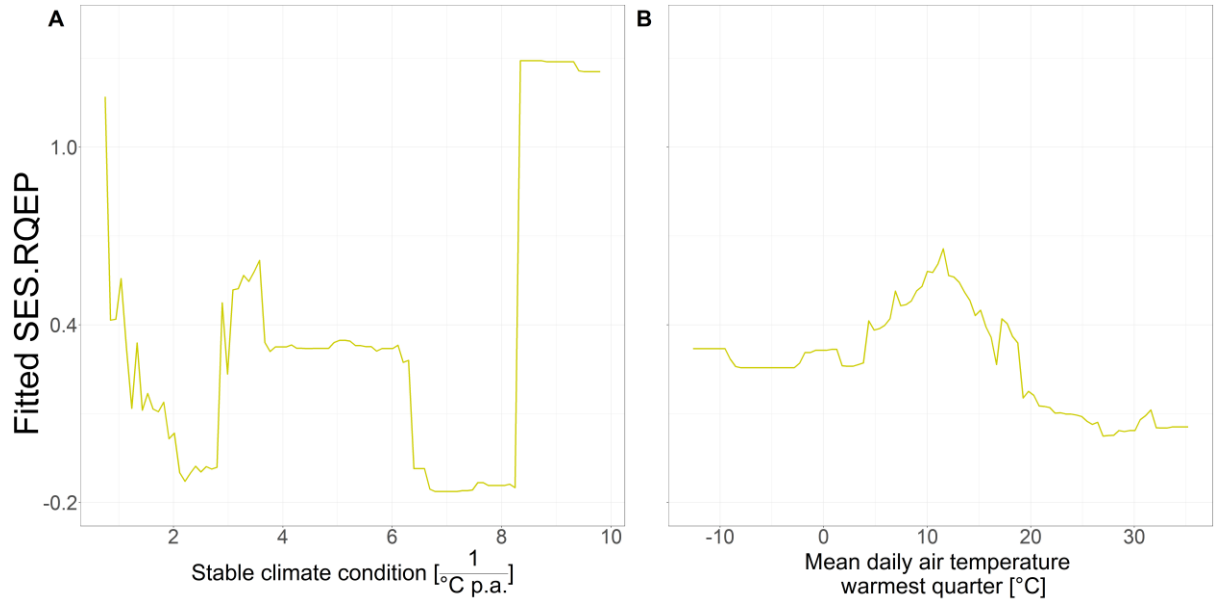
*Figure 6: Smooth terms of phylogenetic entropy expressed as standardized effect size of Rao's quadratic entropy (SES.RQEP) based on the distances of the species present in the community as function of the relevant explanatory variables from the boosted regression trees, based on the first run of the 100 random subsets. **A** SES.RQEP as function of stable climate conditions and **B** as function of mean daily air temperature of the warmest quarter. For phylogenetic dispersion expressed as standardized effect size of mean pair wise distances of the species present in the community the same patterns were found (Fig. S7).*

15

According to the general additive model (GAM) functional entropy (SES.RQEF) increased with increasing range of annual air temperature (*Figure 7A*), mean daily air temperature of the wettest quarter (*Figure 7B*) and stable climate conditions (*Figure 7D*). A negative relationship was found with the mean monthly precipitation of the warmest quarter (*Figure 7C*). Together with the smooth term of longitude and latitude (*Figure 7F*), the GAM explained 3.98% of the deviance of SES. For the second functional diversity index, functional dispersion (SES.FDis), the explanatory variables showed the same relationship to the response variable as for SES.RQEF (*Figure S9*). The GAM explained 5.73% of the deviance of SES.FDis.

Phylogenetic entropy expressed as SES.RQEP showed a negative correlation with stable climate conditions (*Figure 8A*) as well as with mean daily air temperature of the wettest quarter (*Figure 8B*). Forests showed a more positive effect on SES.RQEP compared to non-forest plant-communities (*Figure 8C*). The spatial smooth term suggested in general, low to mid phylogenetic entropy across the globe, but high phylogenetic diversity in New Zealand (*Figure 8D*). The complete model for SES.RQEP explained 39.4% of the deviance. The results for SES.MPD were similar, with an explained deviance of 39.5% (*Figure S10*).
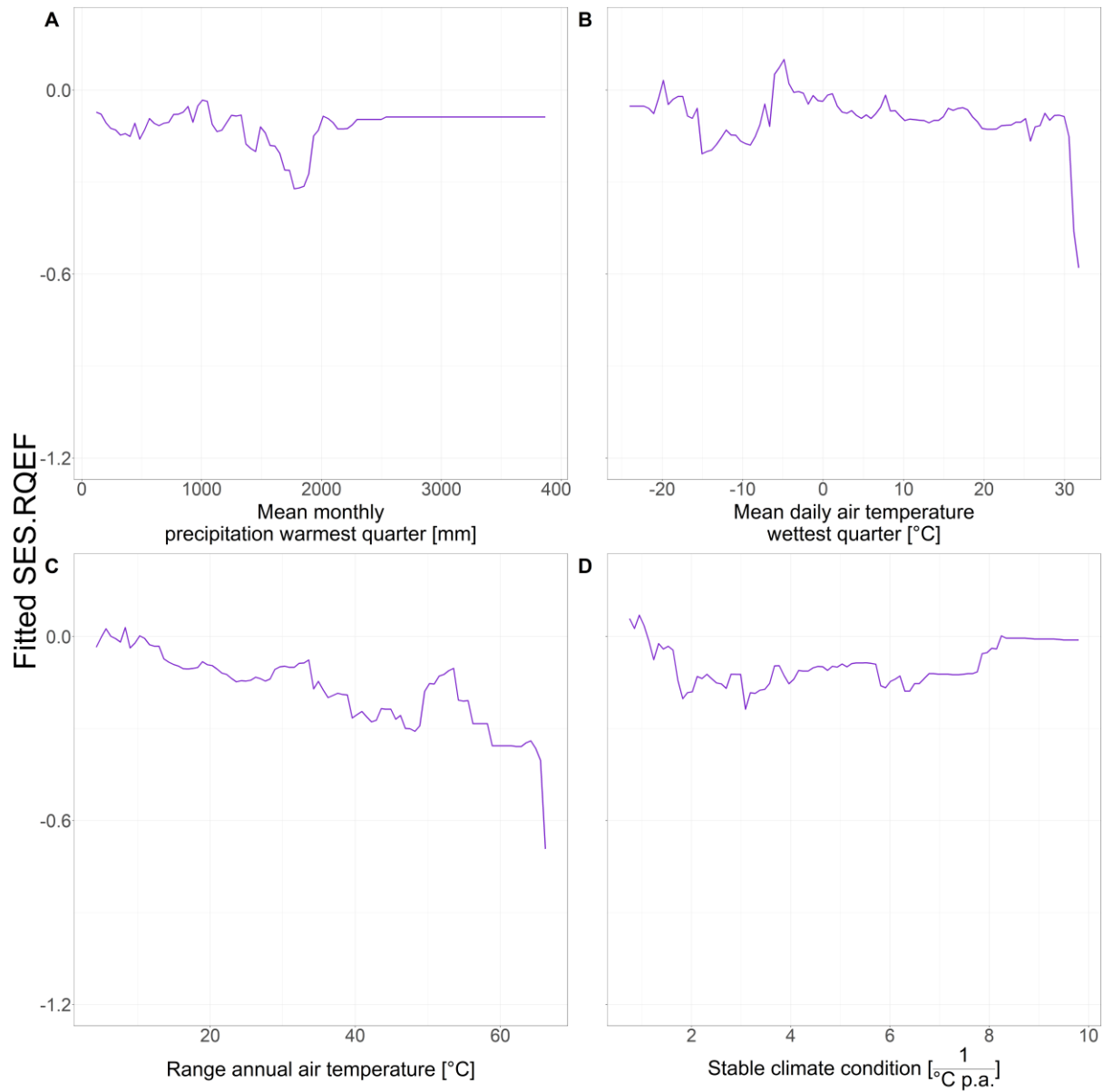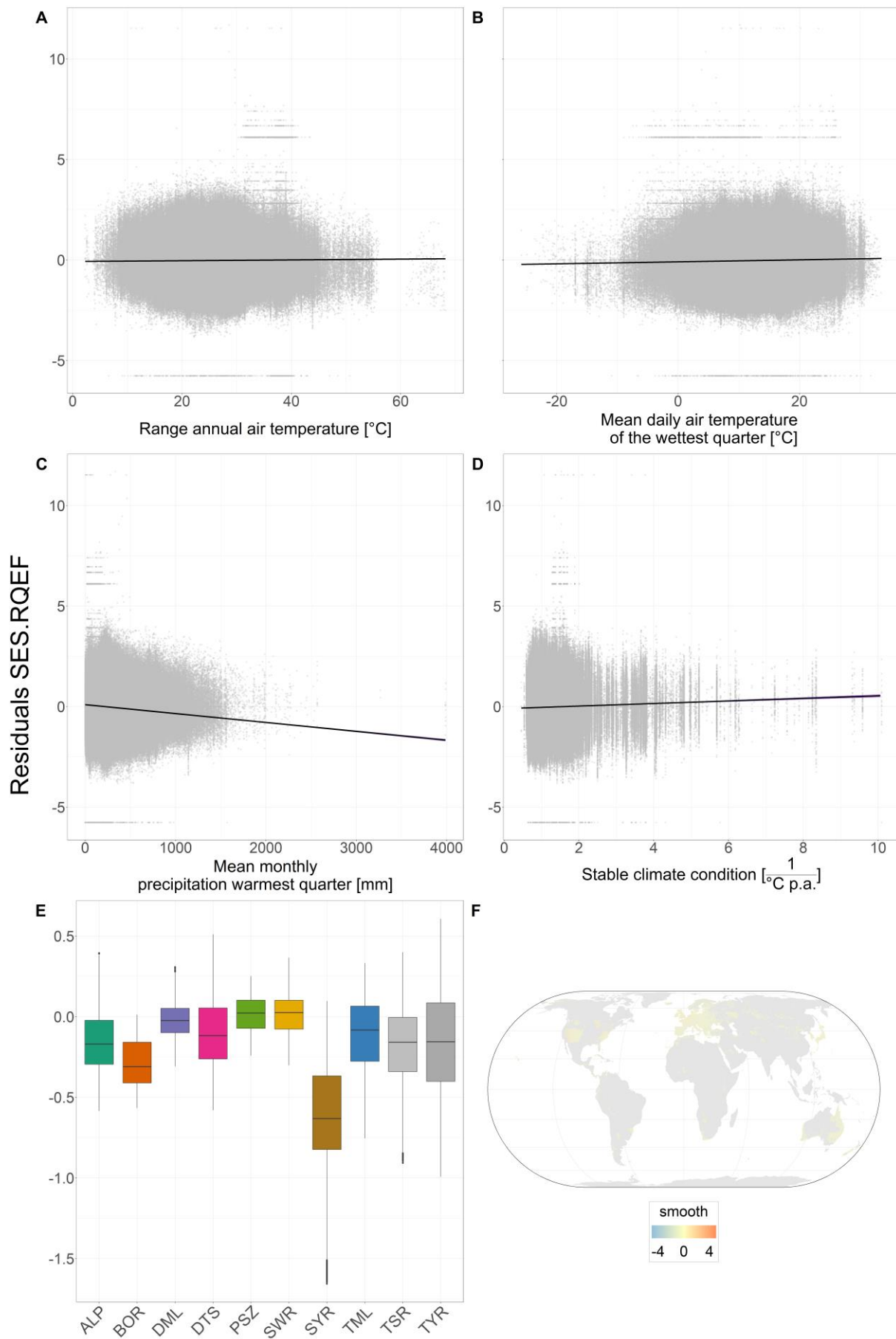
*Figure 7: Functional entropy expressed as standardized effect size of Rao's quadratic entropy (SES.RQEF) as a function of **A** the range in annual air temperature, **B** mean daily air temperature of the wettest quarter, **C** mean monthly precipitation of the warmest quarter, **D** stable climate condition, **E** biome. The general additive model (GAM) explained 3.98% of the deviance. The solid line shows the regression obtained from the GAM. The grey points show the residuals of the GAM after accounting for spatial autocorrelation and the other explanatory variables. **E** SES.RQEF as function of the ten categories of biomes (ALP = Alpine, BOR = Boreal zone, DML = Dry mid-latitudes, DTS = Dry tropics and subtropics, PSZ = Polar and subpolar zone, SWR = Subtropics with winter rain, SYR = Subtropics with year-round rain, TML = Temperate mid-latitudes, TSR = Tropics with summer rain, TYR = Tropics with year-round rain). **F** The smooth term of SES.RQEF in the space indicates lower and higher than expected functional diversity with blue and red colour, respectively.*
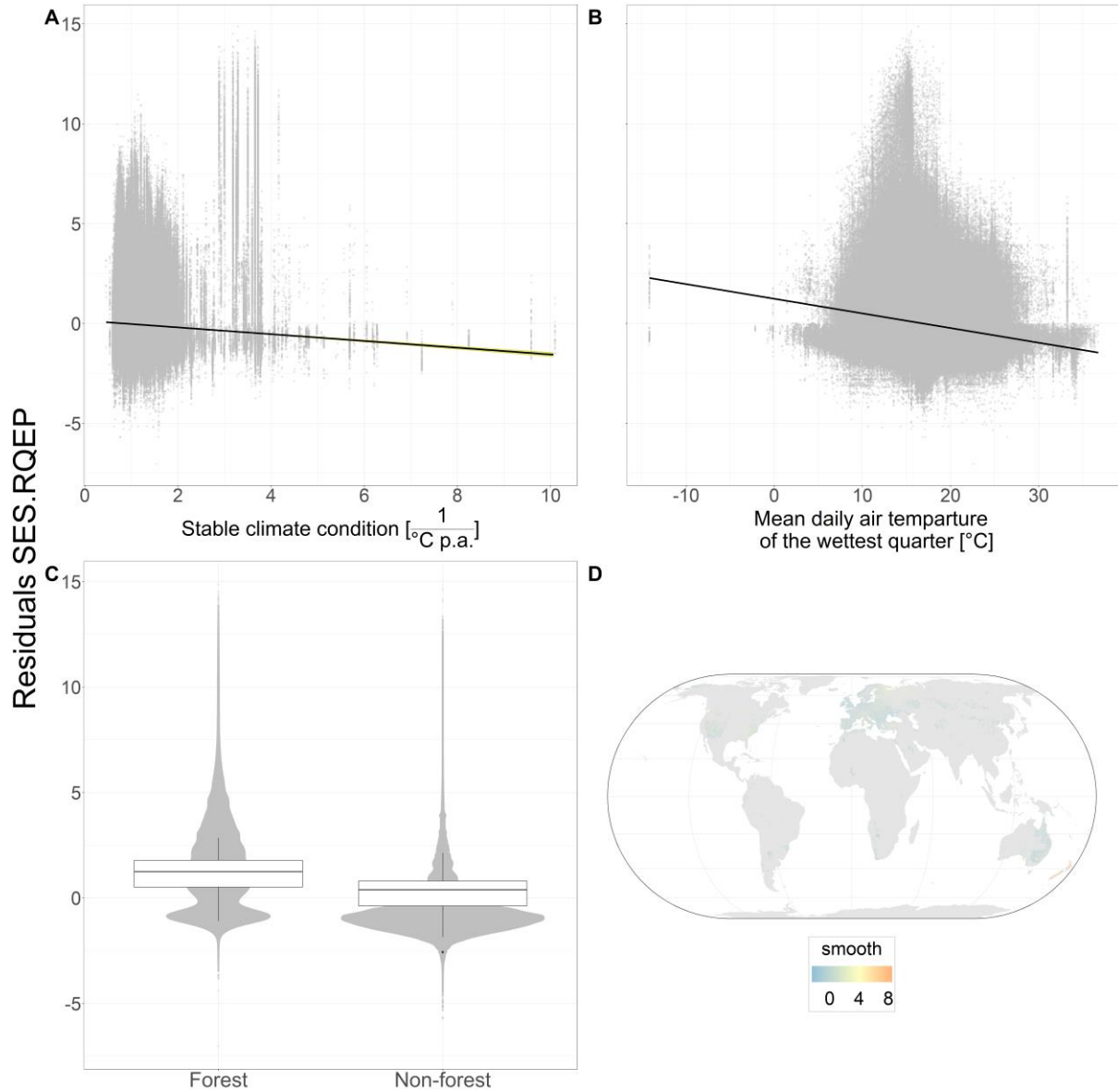
*Figure 8: Phylogenetic entropy expressed as standardized effect size of Rao's quadratic entropy (SES.RQEP) based on the species distances present in the community as a function of the explanatory variables selected from the BRT results: **A** stable climate condition, **B** mean daily air temperature of the wettest quarter, **C** vegetation type (forest vs. non-forest) and **D** the smooth term of longitude and latitude. The solid line shows the regression obtained from the general additive model (GAM). The grey points show the residuals of the GAM after accounting for spatial autocorrelation and the other explanatory variables. The explanatory variables explained 39.4% of the deviance in the model.*

18

4. Discussion

This thesis showed that functional and phylogenetic diversity differ in their distribution patterns and depend differently on environmental conditions and past climatic events. Following the initial hypotheses, the thesis demonstrated that (H1) functional and phylogenetic diversity are (negatively) correlated on the global scale, and that (H2) patterns of high and low functional and phylogenetic diversity are not restricted to single specific regions. Similarly, it could be confirmed that patterns of functional diversity are better explained by recent climatic conditions (H3), while patterns of phylogenetic diversity are better explained by past climatic events, i.e. climate change after the last glacial maximum (H4).

4.1 The relationship between functional and phylogenetic diversity

Functional diversity was negatively correlated with phylogenetic diversity. This could be shown for both indices of entropy and for those of dispersion. In contrast, theoretical studies showed the opposing pattern of a positive correlation between functional and phylogenetic diversity. In particular, it was demonstrated that this relationship becomes stronger with the number of traits used to calculate functional diversity (Tucker et al. 2018). If phylogenetic diversity is considered a proxy that captures all possible traits of a species (Reich et al. 2003), this positive relationship would have to be expected. However, studies carried out with smaller spatial extent showed varying trait patterns on the phylogeny, with either trait clustering or overdispersion (Cavender-Bares et al. 2004, Ackerly 2009). The negative correlation at the global scale, found in this thesis, indicate the tendency of a trait overdispersion in vegetation-plot records. This means that phylogenetically related species differ more in their trait values, while phylogenetic distant species share more of the same trait values. However, a proper statistical test for trait overdispersion is difficult, as it is not clear from which species pool trait values would have to be assembled at the global scale. It is also possible that the negative relationship was brought about by the low number of just three traits used in this thesis. It might be that an increased number of traits could turn the encountered negative relationship into a non-negative one (Tucker et al. 2018).

In addition, both functional and phylogenetic diversity did not display a clear spatial pattern. Vegetation-plots with all combinations from high to low functional and phylogenetic diversity were found at geographical close plant communities. This means that patterns of functional and phylogenetic diversity could not be explained by the geographical space itself. This result was not in line with other studies. The geographical patterns of functional diversity are often explained by climatic conditions, such as rain gradients (Zuo et al. 2021). For phylogenetic diversity it was shown that it changes along a latitudinal gradient (Massante et al. 2019). In addition, for specific countries, such as China, it was shown that phylogenetic diversity decreases towards the north (Cai et al. 2020).

However, these studies focused on specific species or species' groups and often do not cover big vegetation-plot data like this study. Moreover, it seems that specific regional factors are anticipated to drive a plant community high functional or phylogenetic diversity, instead of geographical space itself.

4.2 Drivers of functional and phylogenetic diversity

While it was expected that functional diversity could be explained by recent climatic conditions, the deviance explained by the general additive model (GAM) was relatively low, in particular in comparison to the models for phylogenetic diversity. The deviation of functional diversity was better explained by phylogenetic diversity, than by recent climate conditions. This means that functional diversity does not tend to be highly correlated with climatic conditions. Although it has been shown that functional diversity can be linked with climatic conditions at small spatial extents (Del Toro et al. 2015, Andrew et al. 2021), this thesis suggests that it is not the case on a global scale. The addition of more climatic variables to the model, which should capture more seasonally climatic conditions, could lead to a better understanding (Tonkin et al. 2017), but overfitting of correlated climatic variables might become a problem (Zangiabadi et al. 2021). Another reason for the low explanation of functional diversity could be that the functional composition of local communities strongly depends on local factors, such as land use (Flynn et al. 2009), soil properties and microclimatic conditions (Pauw et al. 2021). Unfortunately, such local factors are not represented in global datasets and the explained deviance of functional diversity remained low.

Phylogenetic diversity was well explained by GAM with the following variables: stable climate condition, mean daily air temperature of the wettest quarter, vegetation-plot classified as forest or non-forest, and the spatial space. Forest or non-forest classification had the most relative influence in the BRTs and shows significant influence in the GAM. Hence, the different structural layers of a forest lead to an increased phylogenetic diversity, but not in general to high functional diversity. This pattern could be interpreted as the ability of forests to buffer climatic events during the year, resulting in a more stable microclimate (Zellweger et al. 2020, Beugnon et al. 2021). In addition, trees differ from the understory in their evolutionary history and belong to families which are exclusively woody. This results in high phylogenetic diversity in forests (Qian et al. 2014, Coronado et al. 2015, Mastrogianni et al. 2019), while functional leaf and root traits of the plant species in different layers might be similar.

Functional diversity showed a positive relationship with the range of annual air temperature, meaning that functional diversity is high in regions with high interannual temperature shifts, i.e. unstable recent climate condition. It seems that resource competition in a specific niche is higher in such unstable environments and therefore plants need to adapt individually to outcompete other species in the community (Wright 2002, Brassard et al. 2009). In contrast, in a recent stable environment, i.e. a

forest, functional diversity is comparably low and plants might differ more in their phylogenetic distances. This could show that phylogenetic diversity captured traits that are more relevant to niche partitioning in a forest, than the three traits used in this study (SLA, SLR and plant height), such as metabolic characteristics that confer herbivore and pathogen resistance (Núñez-Farfán et al. 2007, Castagneyrol et al. 2013).

4.3 Validation of the subsampling method

The findings of the 100 subsamples method for functional diversity were mostly in line with the results of the "sPlotOpen" benchmark run. The minor difference between the two methods could be explained by the general difference of vegetation-plots present in "sPlotOpen". While "sPlotOpen" selected the most different floristic species compositions in each climate grid cell, the subsampling method resulted in a selection of the most abundant vegetation-plots in each spatial grid cell. The difference in the mean of each response variable per grid cell for "sPlotOpen" and the 100 subsample method was lower than the standard deviation of each response variable in the 100 subsamples runs (*Figure S6*).

For phylogenetic diversity, the difference between the two methods was greater than for functional diversity. In particular, stable climate was not suggested as relevant anymore by the boosted regression trees (BRTs) based on the "sPlotOpen" vegetation-database (*Figure S6*). One reason for this could be that "sPlotOpen" was subsampled from the "sPlot" database to cover the most different environmental conditions with the most different species. The selection in "sPlotOpen" might had increased the chance of having plant communities with very high functional and phylogenetic diversity. But this is not in line with the comparison between the two methods (*Figure S6*). In contrary, the results of the two methods showed a linear 1:1 correlation. But the maximum phylogenetic response values were lower for "sPlotOpen" than for the 100 subsample runs. However, a clear interpretation of the difference is difficult, as minor difference might lead to a bigger change in the calculated model based on the related response variables per grid cell.

5. Summary

In the past, the understanding of species community assemblage was driven by the methodological approaches of functional trait measurements. Due to the increased availability of species' sequence information, which could be understood as the sum of all functional traits of a species, phylogenetic diversity can be used to explain plant species community assemblage and even productivity. This thesis aimed to find the relation between functional and phylogenetic diversity, as well as their distribution patterns on a global scale, based on the vegetation-plot recordings of "sPlot".

From the possible ways any given functional trait might map on the phylogenetic tree, the two extremes are trait conservatism or convergence, resulting in clustering or overdispersion of trait values on the phylogeny. The results of the general additive model (GAM), including a geographical smooth, suggest that functional diversity is negatively correlated with phylogenetic diversity at the global scale, which means that plants in communities tend to be functionally clustered on the phylogeny. Functional and phylogenetic diversity showed no clear pattern in their spatial distribution. Furthermore, vegetation-plots with all combinations from high to low functional and phylogenetic diversity were found at geographical close plant communities.

The results of the boosted regression trees (BRTs) suggested that the stable climate condition after the last glacial maximum and if the plant community was a forest or a non-forest to be the most relevant explanatory variables for phylogenetic diversity. For functional diversity, recent climate variables showed the most relative influence. The variables selected by the BRTs were used in a GAM to include the term of geographical space. Phylogenetic diversity was suggested to be negatively related to stable climate condition after the last glacial maximum and to be higher if the plant community was classified as a forest compared to a non-forest community. While the GAM explained well the deviation of phylogenetic diversity, the explained variation was much lower for the models for functional diversity. Recent climatic conditions explained approximately only 4% of the deviation of functional diversity. This was unexpected, as on local scale climatic conditions have been shown to be a major driver of functional diversity. One reason for this could be that functional composition of local communities strongly depends on local factors which are not represented in global datasets.

In conclusion, this thesis gives a first idea of how functional and phylogenetic diversity are related across the world. However, the used methods could be improved by increasing the number of used functional traits, as well as the inclusion of more vegetation-plots of the global South. Furthermore, additional local scale factors might increase the understanding of the distribution of functional and phylogenetic diversity.

6. Zusammenfassung

In der Vergangenheit wurde das Verständnis von Artenzusammensetzungen in einer Gemeinschaft durch die Messung von funktionellen Eigenschaften der Pflanzen geprägt. Durch die angestiegene Verfügbarkeit von genetischen Sequenzen, welche als Summe aller funktionellen Eigenschaften verstanden werden können, kann die phylogenetische Diversität genutzt werden, um die Zusammensetzung und sogar die Produktivität von Pflanzengemeinschaften zu erklären. Diese Masterarbeit hatte das Ziel mit Hilfe der Vegetationsdatenbank "sPlot" die Beziehung zwischen funktioneller und phylogenetischer Diversität aufzudecken, sowie ihre globale Verteilung zu zeigen.

Unter allen Möglichkeiten der Positionierung von funktionellen Eigenschaften in einem phylogenetischen Baum, existieren zwei extreme, gegensätzliche Situationen: Konservatismus und Zerstreuung der Eigenschaften. Das bedeutet, dass entweder nahe Verwandte ähnliche Eigenschaften haben oder weit entfernte Arten. Das Ergebnis des generalisierten additiven Models (GAM) mit einer geographischen Glättung zeigte, dass funktionelle Diversität negativ mit phylogenetischer Diversität korreliert ist. Das bedeutet, dass Pflanzengesellschaften dazu neigen, funktionell geclustert in der Phylogenie zu sein. Allerdings zeigten funktionelle und phylogenetische Diversität keine klaren Muster einer globalen Verteilung. Des Weiteren, zeigten die Vegetationsplots alle Kombinationen von niedriger bis hoher funktioneller und phylogenetischer Diversität. Das Ergebnis der Regressionsbäume (boosted regression trees; BRT) zeigte, dass das stabile Klima nach der letzten Eiszeit und die Plotklassifizierung Wald bzw. nicht-Wald, den größten Einfluss auf die phylogenetische Diversität hatten. Für funktionelle Diversität zeigten die jüngsten klimatischen Eigenschaften den größten relativen Einfluss. Die ausgewählten Variablen wurden in einem GAM genutzt mit einem zusätzlichen Faktor für die geographische Glättung. Phylogenetische Diversität war negativ korreliert mit einem stabilen Klima nach der letzten Eiszeit und war höher in Vegetationsplots klassifiziert als Wald. Während das GAM sehr gut die Abweichung in der phylogenetischen Diversität erklärte, war das nicht der Fall für die funktionelle Diversität. Dieses Ergebnis war unerwartet, da auf lokaler Ebene bereits gezeigt wurde, dass klimatische Eigenschaften mit funktioneller Diversität korreliert sind. Allerdings wurde auch gezeigt, dass andere lokale Faktoren wie die Landnutzung einen großen Einfluss haben. Diese lokalen Faktoren sind jedoch kein Bestandteil globaler Datenbanken. Zusammenfassend lässt sich sagen, dass diese Arbeit eine erste Idee für die globalen Zusammenhänge von funktioneller und phylogenetischer Diversität gibt. Allerdings könnten die genutzten Methoden verbessert werden, zum Beispiel durch die Verwendung von mehreren funktionellen Eigenschaften oder einer besseren Verfügbarkeit von Vegetationsplots auf der Südhalbkugel. Des Weiteren könnten lokale Faktoren das Verständnis der Verteilungsmuster verbessern.

References

Ackerly, D. (2009), 'Conservatism and diversification of plant functional traits: Evolutionary rates versus phylogenetic signal', *Proceedings of the National Academy of Sciences*, 106 Suppl 2: 19699–19706.

Anderson, M. J., Ellingsen, K. E., and McArdle, B. H. (2006), 'Multivariate dispersion as a measure of beta diversity', *Ecology Letters*, 9/6: 683–693 <https://onlinelibrary.wiley.com/doi/full/10.1111/j.1461-0248.2006.00926.x?casa_token=TRNmHT24BwgAAAAA%3A82ctL876Lf0DfHfsRXqJaVdHHj_COuohj1-yTq8zDPRcNPTB8WKwe8WZVaki8YUk5DEObL-F6A6YOwiv>.

Andrew, S. C., Mokany, K., Falster, D. S. et al. (2021), 'Functional diversity of the Australian flora: Strong links to species richness and climate', *Journal of Vegetation Science*, 32/2: e13018.

Beugnon, R., Ladouceur, E., Sünnemann, M. et al. (2022), 'Diverse forests are cool: Promoting diverse forests to mitigate carbon emissions and climate change', *J of Sust Agri & Env*, 1/1: 5–8 <https://onlinelibrary.wiley.com/doi/full/10.1002/sae2.12005>.

Botta-Dukát, Z. (2018), 'Cautionary note on calculating standardized effect size (SES) in randomization test', *Community Ecology*, 19/1: 77–83 <https://akjournals.com/view/journals/168/19/1/article-p77.xml>.

Brassard, B. W., Chen, H. Y. H., and Bergeron, Y. (2009), 'Influence of Environmental Variability on Root Dynamics in Northern Forests', *Critical Reviews in Plant Sciences*, 28/3: 179–197.

Brown, S. C., Wigley, T. M. L., Otto-Bliesner, B. L. et al. (2020), 'StableClim, continuous projections of climate stability from 21000 BP to 2100 CE at multiple spatial scales', *Sci Data*, 7/1: 335 <https://www.nature.com/articles/s41597-020-00663-3#Sec20>.

Bruelheide, H., Dengler, J., Purschke, O. et al. (2018), 'Global trait-environment relationships of plant communities', *Nat Ecol Evol*, 2/12: 1906–1917 <https://www.nature.com/articles/s41559-018-0699-8.epdf?shared_access_token=h2peG44rtLnu-jbqsSa3FdRgN0jAjWel9jnR3ZoTv0MEpXxLjfwajqrCZkhFIwrTKmL1WaFmknhkGzaJzuP9sQ5GasMaR6Ayfa00sFleJqfjY676tcJt6f-gO9AdCYYDxwU4qgy9XiAGiIqFYqik5W0z9uT96SJsBIPBRYP0ph0%3D>.

Bruelheide, H., Dengler, J., Jiménez-Alfaro, B. et al. (2019), 'sPlot – A new tool for global vegetation analyses', *J Veg Sci*, 30/2: 161–186 <https://onlinelibrary.wiley.com/doi/full/10.1111/jvs.12710>.

Cadotte, M. W., Davies, J. T. and Regetz, J. et al. (2010), 'Phylogenetic diversity metrics for ecological communities: integrating species richness, abundance and evolutionary history', Ecology Letters, 13/1: 96–105.

Cai, H., Lyu, L., Shrestha, N. et al. (2021), 'Geographical patterns in phylogenetic diversity of Chinese woody plants and its application for conservation planning', Divers Distrib, 27/1: 179–194 <https://onlinelibrary.wiley.com/doi/full/10.1111/ddi.13180>.

Castagneyrol, B., Jactel, H., Vacher, C. et al. (2014), 'Effects of plant phylogenetic diversity on herbivory depend on herbivore specialization', *J Appl Ecol*, 51/1: 134–141 <https://besjournals.onlinelibrary.wiley.com/doi/full/10.1111/1365-2664.12175>.

Cavender-Bares, J., Ackerly, D. D., Baum, D. A. et al. (2004), 'Phylogenetic overdispersion in Floridian oak communities', *The American naturalist*, 163/6: 823–843.

Coronado, H. E. N., Dexter, K. G., Pennington, R. T. et al. (2015), 'Phylogenetic diversity of Amazonian tree communities', *Diversity and Distributions*, 21/11: 1295–1307 <https://onlinelibrary.wiley.com/doi/full/10.1111/ddi.12357>.

Da Duarte, L. S., Hofmann, G. S., Dos Santos, M. M. G. et al. (2010), 'Testing for the influence of niche and neutral factors on sapling community assembly beneath isolated woody plants in grasslands', *Journal of Vegetation Science*, 21/3: 462–471 <https://onlinelibrary.wiley.com/doi/full/10.1111/j.1654-1103.2009.01153.x>.

Del Toro, I., Silva, R. R., and Ellison, A. M. (2015), 'Predicted impacts of climatic change on ant functional diversity and distributions in eastern North American forests', *Diversity Distrib.*, 21/7: 781–791 <https://onlinelibrary.wiley.com/doi/full/10.1111/ddi.12331>.

Faith, D. P. (1992), 'Conservation evaluation and phylogenetic diversity', *Biological Conservation*, 61/1: 1–10.

Fazayeli, F., Banerjee, A., Kattge, J. et al. (2014), 'Uncertainty Quantified Matrix Completion Using Bayesian Hierarchical Matrix Factorization', *Proceedings - 2014 13th International Conference on Machine Learning and Applications, ICMLA 2014*, 2014: 312–317 <https://experts.umn.edu/en/publications/uncertainty-quantified-matrix-completion-using-bayesian-hierarchi-2>.

Flynn, D. F. B., Gogol-Prokurat, M., Nogeire, T. et al. (2009), 'Loss of functional diversity under land use intensification across multiple taxa', *Ecology Letters*, 12/1: 22–33 <https://onlinelibrary.wiley.com/doi/full/10.1111/j.1461-0248.2008.01255.x>.

Gallego-Tévar, B., Curado, G., Grewell, B. J. et al. (2018), 'Realized niche and spatial pattern of native and exotic halophyte hybrids', *Oecologia*, 188/3: 849–862 <https://link.springer.com/article/10.1007/s00442-018-4251-y>.

Grenié, M., Denelle, P., Tucker, C. M. et al. (2017), 'funrar: An R package to characterize functional rarity', *Diversity Distrib.*, 23/12: 1365–1371 <https://onlinelibrary.wiley.com/doi/10.1111/ddi.12629>.

Gugerli, F., and **Holderegger**, R. (2001), 'Nunatak survival, tabula rasa and the influence of the Pleistocene ice-ages on plant evolution in mountain areas', *Trends in Plant Science*, 6/9: 397–398 <https://www.sciencedirect.com/science/article/pii/S1360138501020532>.

Gower, J. C. (1971), 'A General Coefficient of Similarity and Some of Its Properties', *Biometrics*, 27/4: 857–871.

Hijmans, R. J. , van Etten , J., Sumner , M. et al. (2022), 'raster: Geographic Data Analysis and Modeling', 23 Jan <https://rdrr.io/cran/raster/>, accessed 13 May 2022.

Hijmans, R. J., Phillips, S. , Leathwick, J. and Elith, J. (2021), 'dismo: Species Distribution Modeling', 10 Nov <https://rdrr.io/cran/dismo/>, accessed 19 May 2022.

Hutchinson, G. E. (1978), *Introduction to population ecology* (Yale University Press) <https://agris.fao.org/agris-search/search.do?recordid=us201300539753>.

Jin, Y., and **Qian**, H. (2019), 'V.PhyloMaker: an R package that can generate very large phylogenies for vascular plants', *Ecography*, 42/8: 1353–1359 <https://onlinelibrary.wiley.com/doi/full/10.1111/ecog.04434>.

Karger, D. N., Conrad, O., Böhner, J. et al. (2017), 'Climatologies at high resolution for the earth's land surface areas', *Sci Data*, 4: 170122.

Karger, D. N., Conrad, O., Böhner, J. et al. (2018), *Data from: Climatologies at high resolution for the earth's land surface areas*.

Kattge, J., Reich, P. et al. (2012), 'Gap Filling in the Plant Kingdom - Trait Prediction Using Hierarchical Probabilistic Matrix Factorization', *undefined*, 2012 <https://www.semanticscholar.org/paper/Gap-Filling-in-the-Plant-Kingdom-Trait-Prediction-Shan-Kattge/86830b6b3525b5724c8e03f518f870670a0944e5>.

Kattge, J., Bönisch, G., Díaz, S. et al. (2020), 'TRY plant trait database - enhanced coverage and open access', *Global Change Biology*, 26/1: 119–188 <https://onlinelibrary.wiley.com/doi/full/10.1111/gcb.14904>.

Kembel, S. W., Cowan, P. D., Helmus, M. R. et al. (2010), 'Picante: R tools for integrating phylogenies and ecology', *Bioinformatics*, 26/11: 1463–1464 <https://www.researchgate.net/publication/43159817_Picante_R_tools_for_integrating_phylogenies_and_ecology>.

Laliberté, E., and Legendre, P. (2010), 'A distance-based framework for measuring functional diversity from multiple traits', *Ecology*, 91/1: 299–305.

Laliberté, E., Legendre, P. and Shipley, B. (2014), *FD: Measuring functional diversity from multiple traits, and other tools for functional ecology* (1)

<https://www.researchgate.net/publication/312463190_FD_Measuring_functional_diversity_fro m_multiple_traits_and_other_tools_for_functional_ecology>.

Massante, J. C., Götzenberger, L., Takkis, K. et al. (2019), 'Contrasting latitudinal patterns in phylogenetic diversity between woody and herbaceous communities', *Sci Rep*, 9/1: 6443 <https://www.nature.com/articles/s41598-019-42827-1>.

Mastrogianni, A., Kallimanis, A. S., Chytrý, M. et al. (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://www.sciencedirect.com/science/article/pii/S0378112719301161>.

McGill, B. J., Enquist, B. J., Weiher, E. et al. (2006), 'Rebuilding community ecology from functional traits', *Trends in Ecology & Evolution*, 21/4: 178–185 <https://www.sciencedirect.com/science/article/pii/S0169534706000334>.

Núñez-Farfán, J., Fornoni, J., and Valverde, P. L. (2007), 'The Evolution of Resistance and Tolerance to Herbivores', *Annual Review of Ecology, Evolution, and Systematics*, 38: 541–566 <http://www.jstor.org/stable/30033871>.

Oliveira, R. S., Costa, F. R. C., van Baalen, E. et al. (2019), 'Embolism resistance drives the distribution of Amazonian rainforest tree species along hydro-topographic gradients', *New Phytologist*, 221/3: 1457–1465 <https://nph.onlinelibrary.wiley.com/doi/full/10.1111/nph.15463>.

Pauw, K. de, Meeussen, C., Govaert, S. et al. (2021), 'Taxonomic, phylogenetic and functional diversity of understorey plants respond differently to environmental conditions in European forest edges', *J Ecology*, 109/7: 2629–2648 <https://besjournals.onlinelibrary.wiley.com/doi/full/10.1111/1365-2745.13671>.

Pearman, P. B., Guisan, A., Broennimann, O. et al. (2008), 'Niche dynamics in space and time', *Trends in Ecology & Evolution*, 23/3: 149–158 <https://www.sciencedirect.com/science/article/pii/S0169534708000372>.

Pennington, R. T., Lavin, M., Prado, D. E. et al. (2004), 'Historical climate change and speciation: neotropical seasonally dry forest plants show patterns of both tertiary and quaternary diversification', *Philosophical transactions of the Royal Society of London. Series B, Biological sciences*, 359/1443: 515–537.

Petchey, O. L., and **Gaston**, K. J. (2002), 'Extinction and the loss of functional diversity', *Proceedings. Biological sciences*, 269/1501: 1721–1727.

Prentice, I. C., Bartlein, P. J., and Webb, T. (1991), 'Vegetation and Climate Change in Eastern North America Since the Last Glacial Maximum', *Ecology*, 72/6: 2038–2056.

Qian, H., Hao, Z., and Zhang, J. (2014), 'Phylogenetic structure and phylogenetic diversity of angiosperm assemblages in forests along an elevational gradient in Changbaishan, China', *J Plant Ecol*, 7/2: 154–165 <https://academic.oup.com/jpe/article/7/2/154/932721?login=true>.

Qian, H., and **Jin**, Y. (2016), 'An updated megaphylogeny of plants, a tool for generating plant phylogenies and an analysis of phylogenetic community structure', *J Plant Ecol*, 9/2: 233–239 <https://academic.oup.com/jpe/article/9/2/233/2928108>.

Qian, H., Jin, Y. and Leprieur, F. et al. (2020), 'Geographic patterns and environmental correlates of taxonomic and phylogenetic beta diversity for large-scale angiosperm assemblages in China', *Ecography*, 43/11: 1706–1716.

Rao, C. R. (1982), 'Diversity: Its measurement, decomposition, apportionment and analysis', *The Indian Journal of Statistics*, 41: 1–23 <https://www.jstor.org/stable/25050293>.

Reich, P. B., Wright, I. J., Cavender-Bares, J. et al. (2003), 'The Evolution of Plant Functional Variation: Traits, Spectra, and Strategies', *International Journal of Plant Sciences*, 164/S3: S143-S164.

Renssen, H., and **Isarin**, R. (2001), 'The two major warming phases of the last deglaciation at ~14.7 and ~11.5 ka cal BP in Europe: climate reconstructions and AGCM experiments', *Global and Planetary Change*, 30/1-2: 117–153 <https://www.researchgate.net/publication/222219032_The_two_major_warming_phases_of_the_last_deglaciation_at_147_and_115_ka_cal_BP_in_Europe_Climate_reconstructions_and_AGCM_experiments>.

Revell (2012), 'phytools: An R package for phylogenetic comparative biology (and other things). Methods in Ecology and', *Evolution*, 3: 217.

Rutherford, T. F. (1995), 'Extension of GAMS for complementarity problems arising in applied economic analysis', *Journal of Economic Dynamics and Control*, 19/8: 1299–1324 <https://www.sciencedirect.com/science/article/pii/0165188994008312>.

Sabatini, F. M., Lenoir, J. and Hattab, T. et al. (2021), 'sPlotOpen – An environmentally balanced, open-access, global dataset of vegetation plots', *Global Ecol Biogeogr*, 30/9: 1740–1764.

Saladin, B., Pellissier, L. and Graham, C. H. et al. (2020), 'Rapid climate change results in long-lasting spatial homogenization of phylogenetic diversity', *Nat Commun*, 11/1: 4663 <https://www.nature.com/articles/s41467-020-18343-6>.

Schrodt, F., Kattge, J. and Shan, H. et al. (2015), 'BHPMF - a hierarchical Bayesian approach to gap-filling and trait prediction for macroecology and functional biogeography', *Global Ecol Biogeogr*, 24/12: 1510–1521 <https://onlinelibrary.wiley.com/doi/full/10.1111/geb.12335>.

Shan, H., Kattge, J. and Reich, P. et al. (2012), 'Gap Filling in the Plant Kingdom - Trait Prediction Using Hierarchical Probabilistic Matrix Factorization', *undefined*, 2012

<https://www.semanticscholar.org/paper/Gap-Filling-in-the-Plant-Kingdom-Trait-Prediction-Shan-Kattge/86830b6b3525b5724c8e03f518f870670a0944e5>.

Silvertown, J. (2004), 'Plant coexistence and the niche', *Trends in Ecology & Evolution*, 19/11: 605–611 <https://www.sciencedirect.com/science/article/pii/S0169534704002630>.

Smith, S. A. and **Brown**, J. W. (2018), 'Constructing a broadly inclusive seed plant phylogeny.', *American Journal of Botany*, 105(3): 302–314.

Statzner, B., Hildrew, A. G. and Resh, V. H. (2001), 'Species traits and environmental constraints: entomological research and the history of ecological theory', *Annual Review of Entomology*, 46/1: 291–316 <https://www.researchgate.net/profile/alan-hildrew-2/publication/12216321_species_traits_and_environmental_constraints_entomological_research _and_the_history_of_ecological_theory>.

Tofts, R. and **Silvertown**, J. (2000), 'A phylogenetic approach to community assembly from a local species pool', *Proceedings. Biological sciences*, 267/1441: 363–369.

Tonkin, J. D., Bogan, M. T. and Bonada, N. et al. (2017), 'Seasonality and predictability shape temporal species diversity', *Ecology*, 98/5: 1201–1216 <https://esajournals.onlinelibrary.wiley.com/doi/full/10.1002/ecy.1761>.

Tucker, C. M., Davies, T. J. and Cadotte, M. W. et al. (2018), 'On the relationship between phylogenetic diversity and trait diversity', Ecology, 99/6: 1473–1479 <https://esajournals.onlinelibrary.wiley.com/doi/full/10.1002/ecy.2349?casa_token=ujZVwD1q-UAAAAA%3AN8GjICFEo8ohMet2kpFEwL9RLzffn7WLTTR1cQH4YR33ElN8zQMUHQyZGc6eJ9vA2ZqE DvN44tT7CoHg>.

Arel-Bundock, V. (2022), *Marginal Effects, Marginal Means, Predictions, and Contrasts [R package marginaleffects version 0.6.0]* (Comprehensive R Archive Network (CRAN)) <https://cran.r-project.org/web/packages/marginaleffects/index.html>.

Violle, C., and **Jiang**, L. (2009), 'Towards a trait-based quantification of species niche', *J Plant Ecol*, 2/2: 87–93.

Violle, C., Navas, M.-L. and Vile, D. et al. (2007), 'Let the concept of trait be functional!', *Oikos*, 116/5: 882–892 <https://onlinelibrary.wiley.com/doi/full/10.1111/j.0030-1299.2007.15559.x?casa_token=xZhooiv2zy8AAAAA%3AHnShAit3zvNn3pLayTKu9tTVl6gH1ibJNnW 4qUAepMi7aqCQHTRC7FjTtBCUkTP2izAC9j_G3p3uDZiv>.

Wang, S., Xu, X. and Shrestha, N. et al. (2017), 'Response of spatial vegetation distribution in China to climate changes since the Last Glacial Maximum (LGM)', *PLoS ONE*, 12/4: e0175742 <https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0175742>.

Webb, C. O. (2000), 'Exploring the Phylogenetic Structure of Ecological Communities: An Example for Rain Forest Trees', The American Naturalist, 156/2: 145–155.

Webb, C. O., Ackerly, D. D. and McPeek, M. A. et al. (2002), 'Phylogenies and Community Ecology', *Annual Review of Ecology and Systematics*, 33/1: 475–505.

Wood, S. N. (2017), *Generalized additive models*: *An introduction with R* (Chapman & Hall / CRC texts in statistical science; Second edition, Boca Raton: CRC Press) <https://www.taylorfrancis.com/books/mono/10.1201/9781315370279/generalized-additive-models-simon-wood>.

Wright, J. S. (2002), 'Plant diversity in tropical forests: a review of mechanisms of species coexistence', *Oecologia*, 130/1: 1–14 <https://link.springer.com/article/10.1007/s004420100809>.

Wright, I. J., Reich, P. B. and Westoby, M. et al. (2004), 'The worldwide leaf economics spectrum', *Nature*, 428/6985: 821–827 <https://www.nature.com/articles/nature02403>.

Wright, I. J., Reich, P. B. and Cornelissen, J. H. C. et al. (2005), 'Assessing the generality of global leaf trait relationships', *The New phytologist*, 166/2: 485–496 <https://nph.onlinelibrary.wiley.com/doi/10.1111/j.1469-8137.2005.01349.x>.

Wright, S. D., Yong, C. G. and Dawson, J. W. et al. (2000), 'Riding the Ice Age El Nino? Pacific biogeography and evolution of Metrosideros subg. Metrosideros (Myrtaceae) inferred from nuclear ribosomal DNA', *Proceedings of the National Academy of Sciences of the United States of America*, 97/8: 4118–4123.

Zangiabadi, S., Zaremaivan, H. and Brotons, L. et al. (2021), 'Using climatic variables alone overestimate climate change impacts on predicting distribution of an endemic species', *PLoS ONE*, 16/9: e0256918 <https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0256918>.

Zanne, A. E., Tank, D. C. and Cornwell, W. K. et al. (2014), 'Three keys to the radiation of angiosperms into freezing environments', *Nature*, 506/7486: 89–92 <https://www.nature.com/articles/nature12872>.

Zellweger, F., de Frenne, P. and Lenoir, J. et al. (2020), 'Forest microclimate dynamics drive plant responses to warming', *Science (New York, N.Y.)*, 368/6492: 772–775.

Zuo, X., Zhao, S. and Cheng, H. et al. (2021), 'Functional diversity response to geographic and experimental precipitation gradients varies with plant community type', *Funct Ecol*, 35/9: 2119–2132 <https://besjournals.onlinelibrary.wiley.com/doi/full/10.1111/1365-2435.13875>.
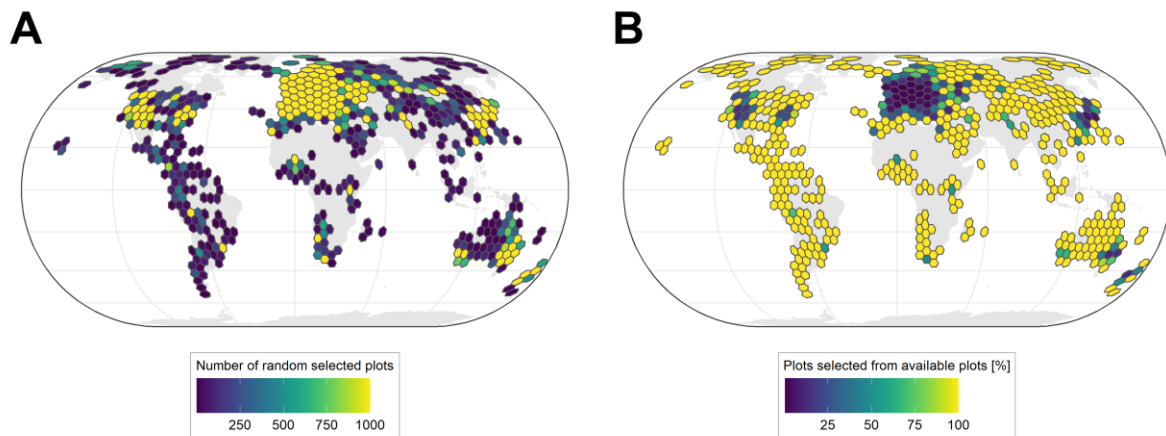
Supporting information

Figure S 1: A global grid (of approx. 209,900 km²) was aligned and in each grid cell 1,000 plots were randomly selected and used for the further modelling. **A** shows the number of plots randomly selected in each grid cell, where yellow colour indicates that 1,000 plot records were selected, while violet colour means that fewer than 250 plot records were selected. A selection of fewer than 1,000 plots occurs when ƒewer records were available in that grid cell. **B** shows the percentage of randomly selected plots of available plots per grid cell.
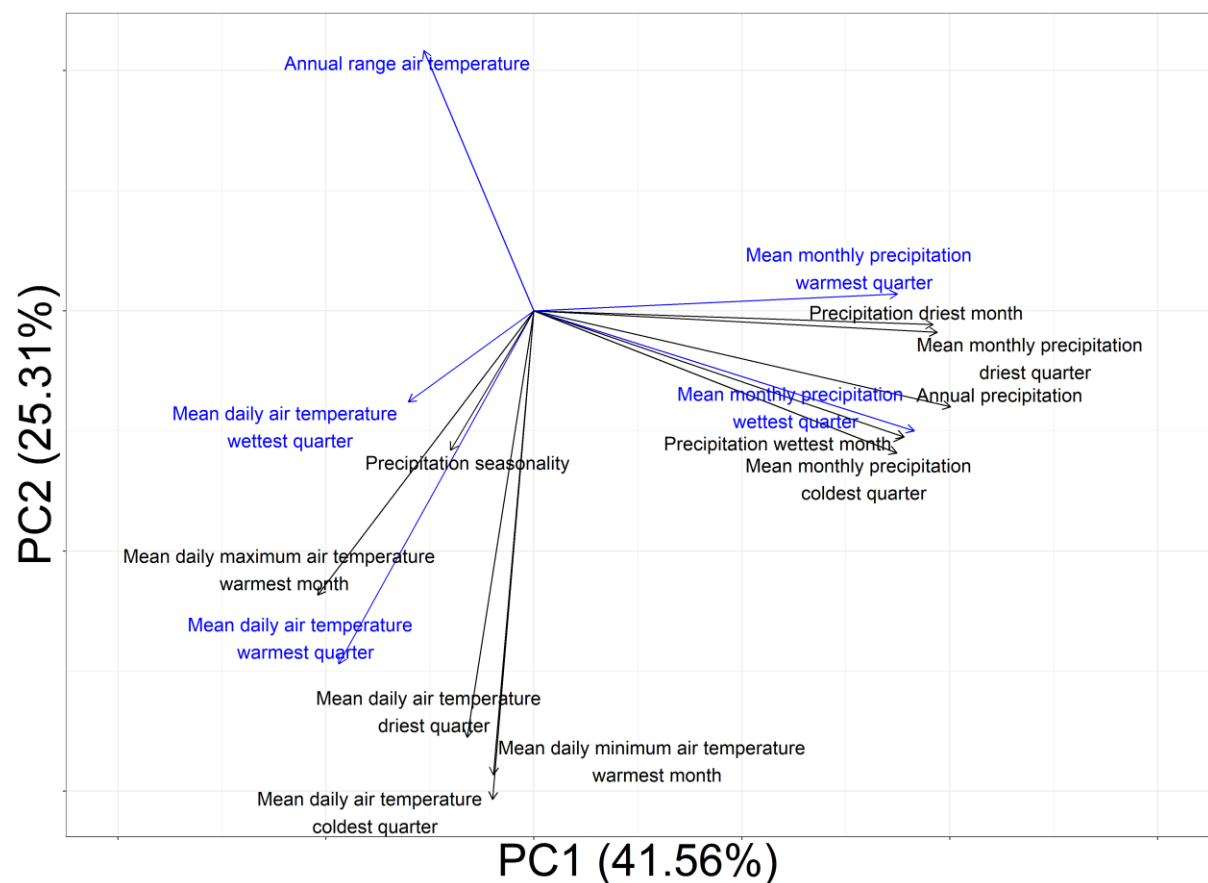
Figure S 2: Principial component analyses (PCA) of all 19 climate variables. Based on this analysis the explanatory variables for the further modelling were selected. The arrows indicate the strength and direction of the correlation between the climate variables and the first and second axes.
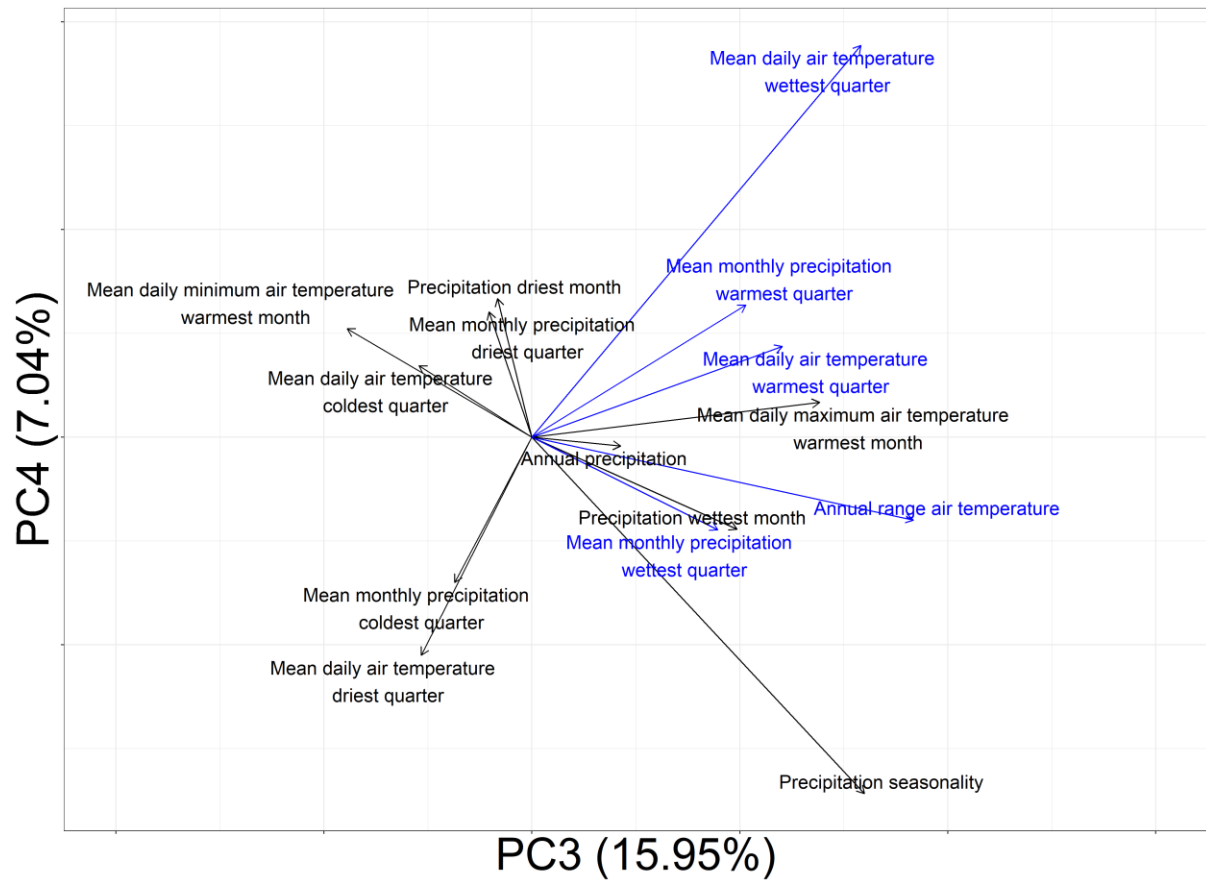
Figure S 3: Principial component analyses (PCA) of all 19 climate variables. Based on this analysis the explanatory variables for the further modelling were selected. The arrows indicate the strength and direction of the correlation between the climate variables and the third and fourth axes.
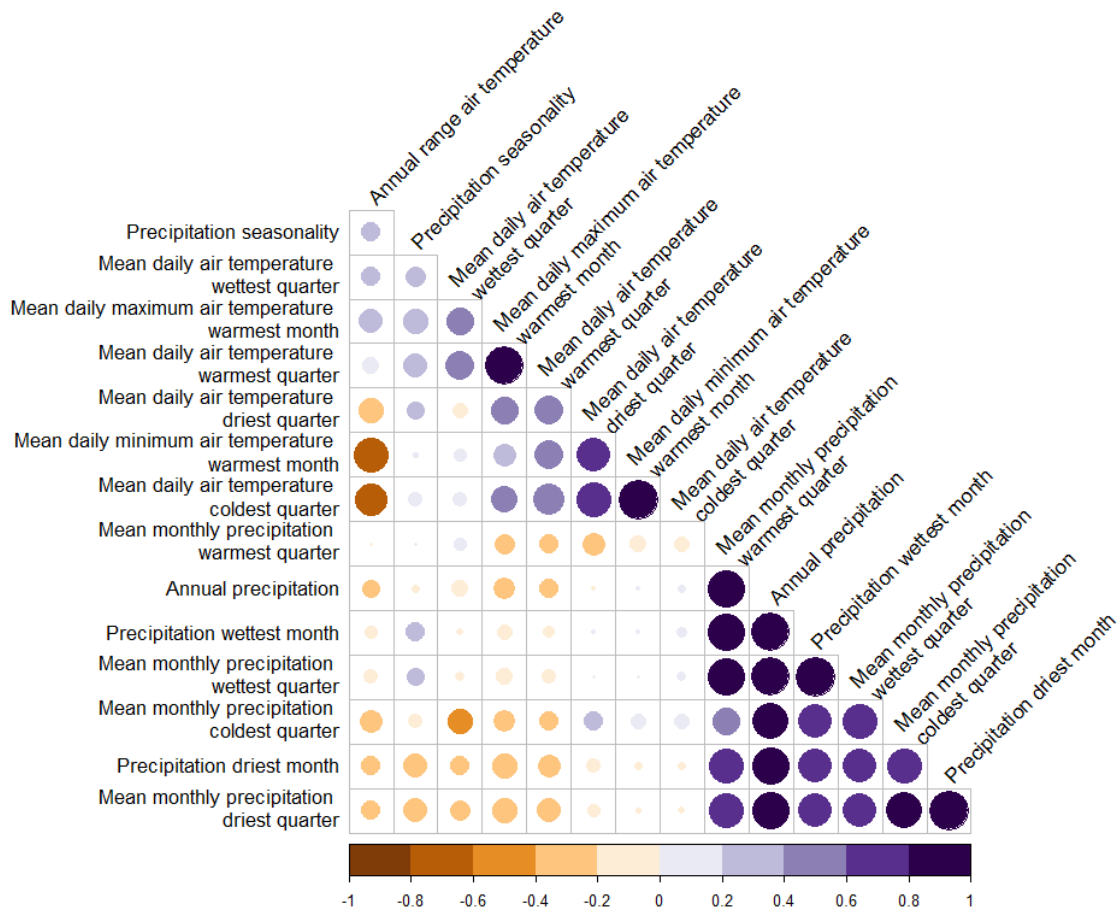
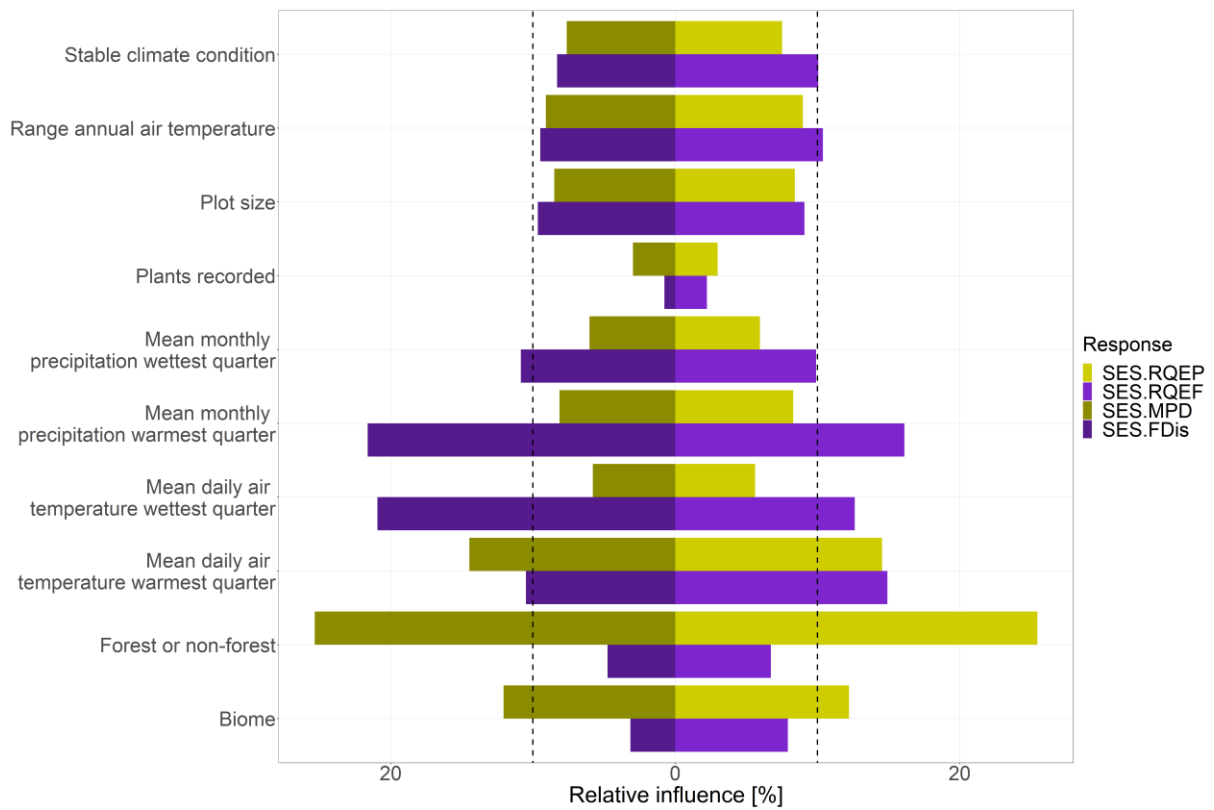*Figure S 4: Correlation matrix of the recent climate variables with Pearson correlation coefficients.*

Figure S 5: Results of the boosted regression trees (BRT) based on the environmental balanced "sPlotOpen" subset. This result was used as Benchmark of the 100 subsample method by grid cells for the "sPlot" vegetation-database.
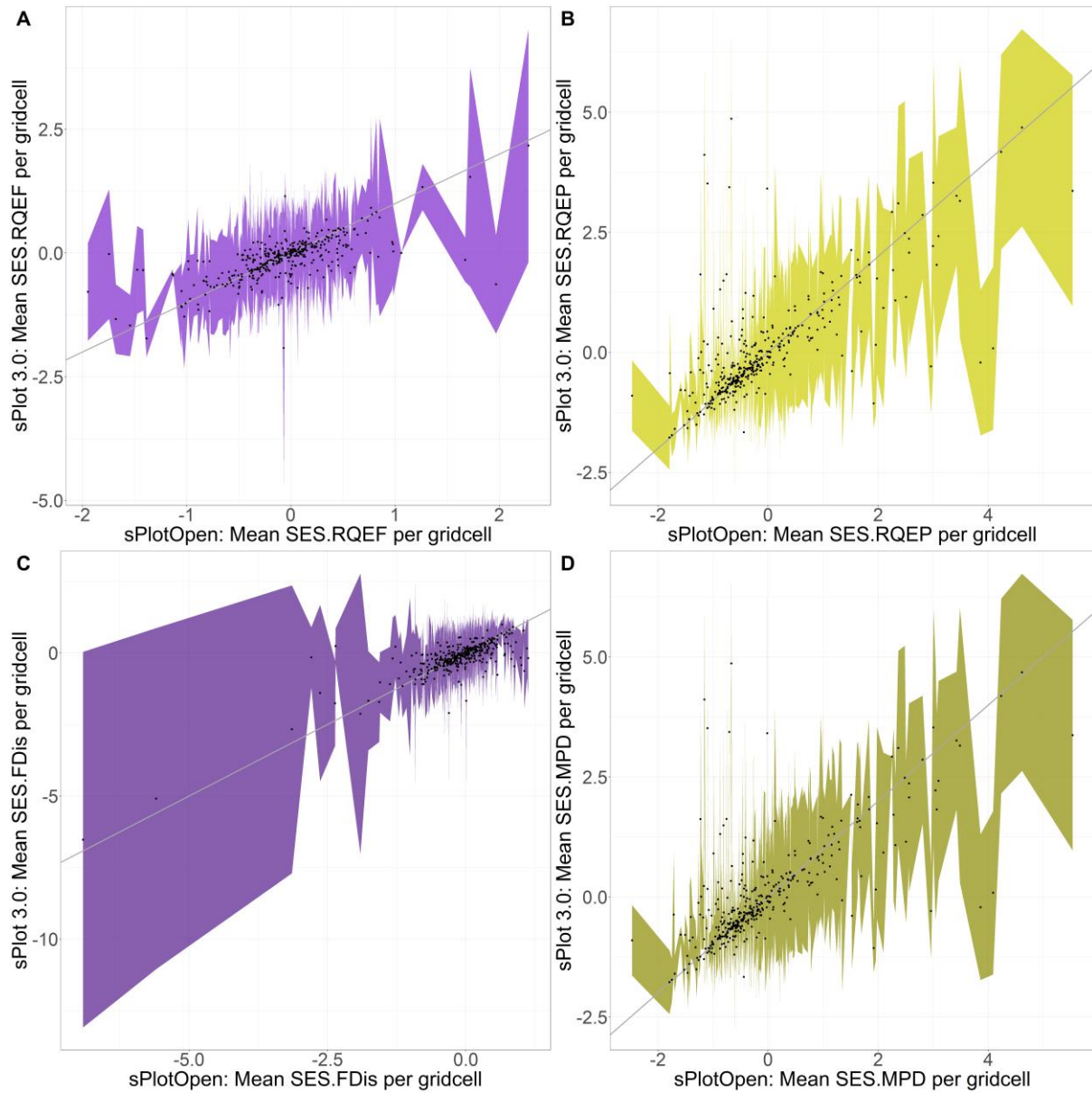
*Figure S 6: The mean per grid cell of the "sPlot" data against the mean per grid cell of the "sPlotOpen" data for: **A** Standardized effect size of Rao`s quadratic entropy based on the functional traits (SES.RQEF), **B** Standardized effect size of Rao`s quadratic entropy based on the phylogenetic distances of species present in the community (SES.RQEP), **C** Standardized effect size of functional dispersion (SES.FDis), **D** Standardized effect size of the mean pairwise distances of species present in the community (SES.MPD). The coloured area shows the standard deviation of the "sPlot" data. The grey line express y = x. For points above the line: values calculated based on the "sPlot" data are higher than the values calculated based on the "sPlotOpen" data.*
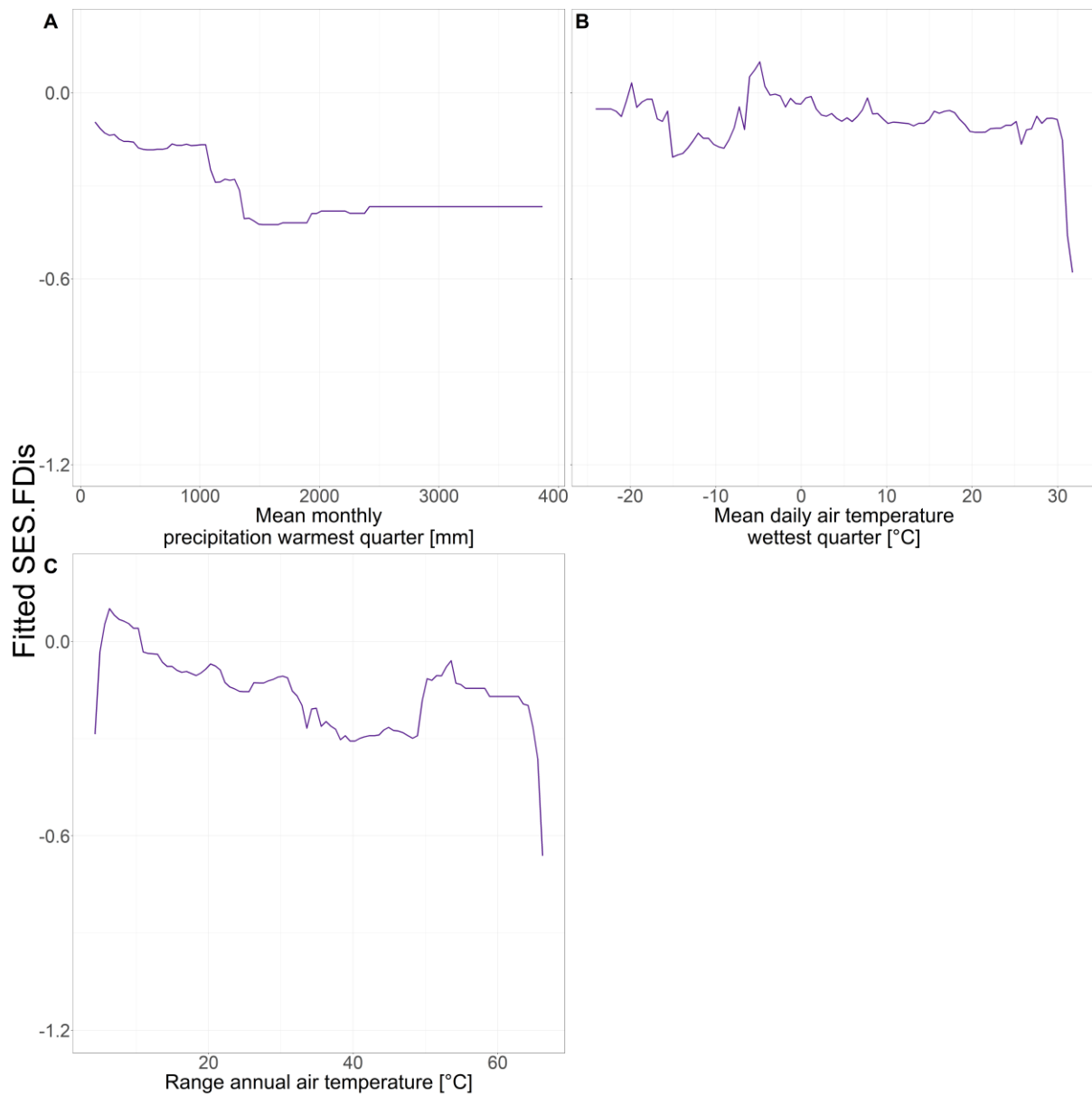
*Figure S 7: Smooth terms of functional dispersion expressed as standardized effect size of FDis (SES.FDis) as a function of the relevant explanatory variables from the boosted regression trees, based on the first run of the 100 random subsets. **A** SES.FDis as function of the bioclimatic variable mean monthly precipitation of the warmest quarter, **B** mean daily air temperature of the wettest quarter, **C** range of annual air temperature.*
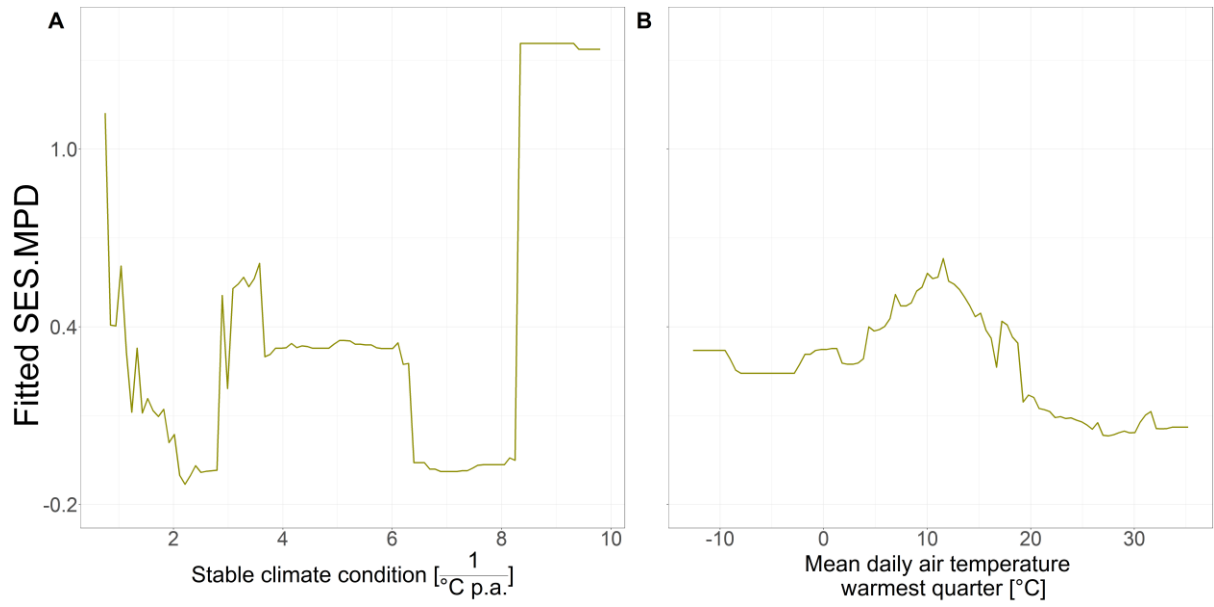
*Figure S 8: Smooth terms of phylogenetic dispersion expressed as standardized effect size of mean pairwise distances (SES.MPD) of the species present in the community as function of the relevant explanatory variables from the boosted regression trees, based on the first run of the 100 random subsets. **A** SES.MPD as function of stable climate condition and **B** as function of mean daily air temperature of the warmest quarter.*
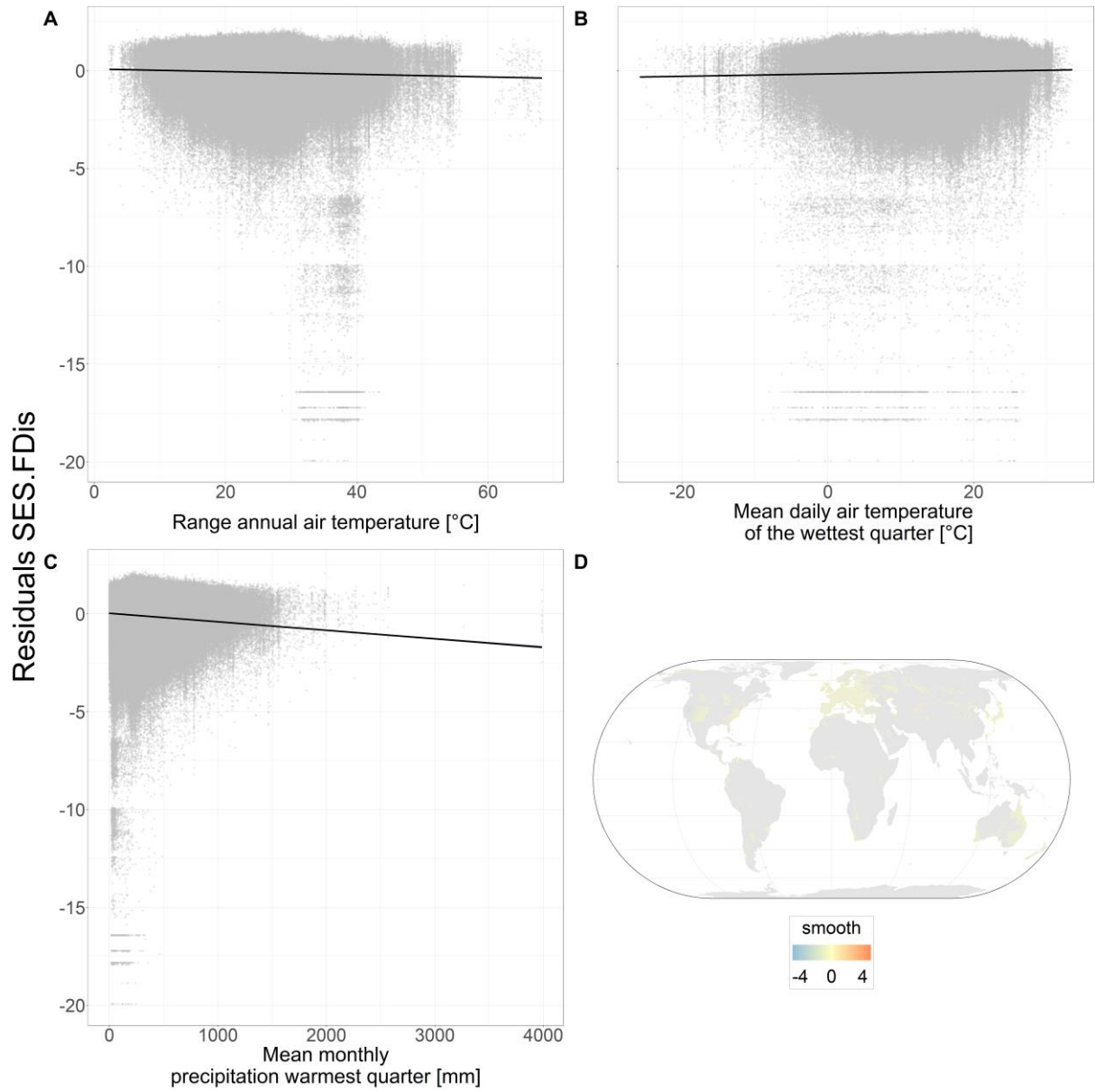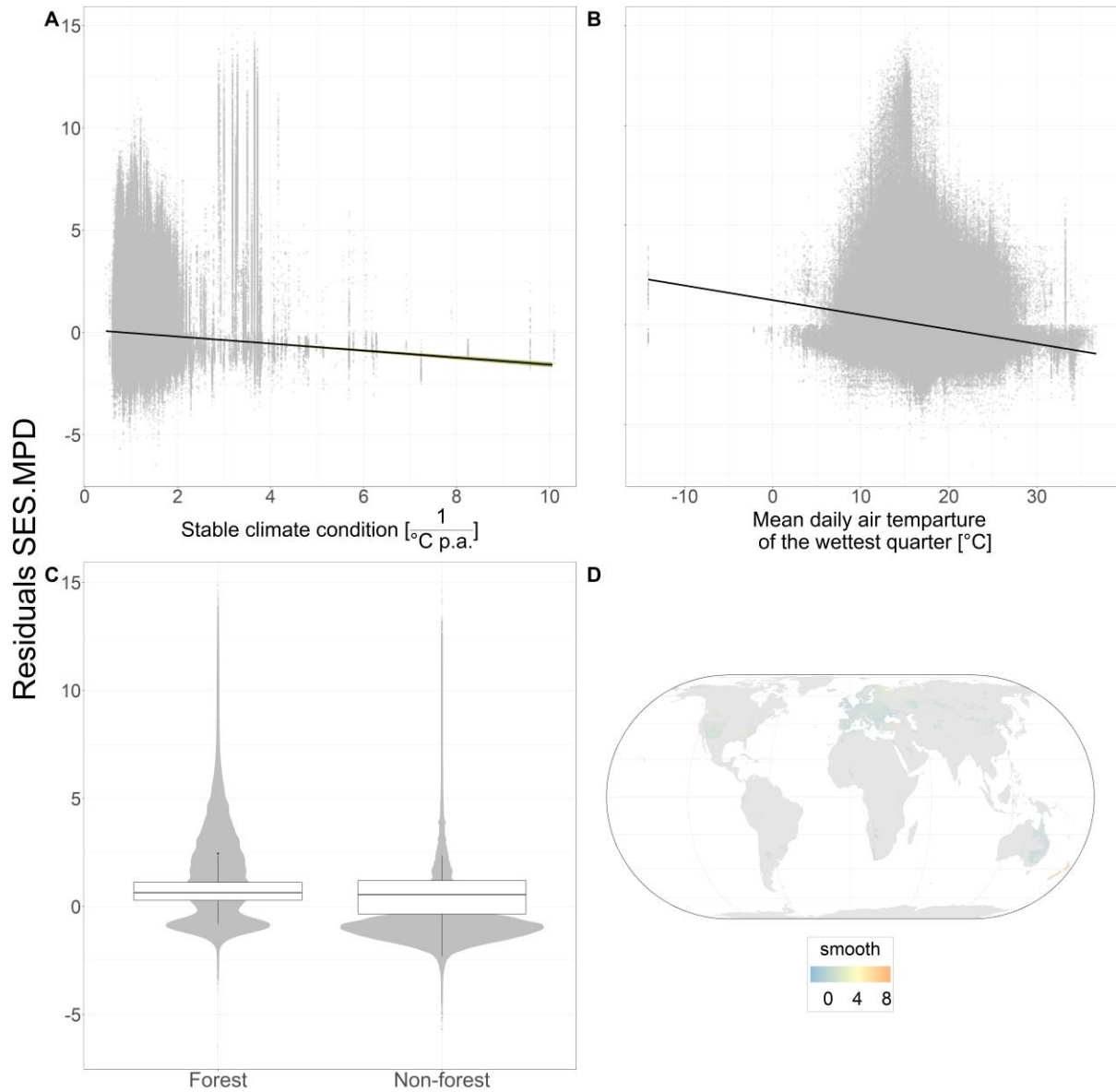
*Figure S 9: Functional dispersion expressed as standardized effect size of FDis (SES.FDis) as function of **A** range annual air temperature, **B** mean daily air temperature of the wettest quarter, **C** mean monthly precipitation of the warmest quarter, **D** stable climate condition and **D** the smoothing term of the spatial space. The solid line shows the regression obtained from the general additive model (GAM). The grey points show the residuals of the GAM after accounting for spatial autocorrelation and the other explanatory variables. The model explained 5.73% of the deviance.*

*Figure S 10: Phylogenetic dispersion expressed as standardized effect size of mean pairwise species distances (SES.MPD) present in the community as a function of the explanatory variables selected from the boosted regression trees (BRTs) results: **A** stable climate condition, **B** mean daily air temperature of the wettest quarter, **C** the two-levels factor variable forest or non-forest and **D** the smooth term of longitude and latitude. The solid line shows the regression obtained from the general additive model (GAM). The grey points show the residuals of the GAM after accounting for spatial autocorrelation and the other explanatory variables. The model explained 39.5% of the deviance.*

Acknowledgements

Statutory Declaration

Eidesstattliche Erklärung

Hiermit erkläre ich, dass ich die vorliegende Arbeit mit dem Titel „Comparison of the global distribution of functional and phylogenetic diversity in plant communities. A study to highlight commonalities and differences in the distribution of vascular plants.", selbstständig angefertigt habe und keine anderen als die von mir angegebenen Quellen und Hilfsmittel verwendet wurden. Diese Arbeit hat weder in gleicher noch einer ähnlichen Form einer Prüfungskommission vorgelegen.

Ort, Datum   Unterschrift