Relationships of local abundance of vascular plants with range-wide niche characteristics, and the role of functional traits

Dissertation

zur Erlangung des
Doktorgrades der Naturwissenschaften (Dr. rer. nat.)

der

Naturwissenschaftlichen Fakultät I – Biowissenschaften –

der Martin-Luther-Universität
Halle-Wittenberg,

vorgelegt

von Frau Maria Sporbert

geb. am 30.07.1987 in Halle (Saale)

GutachterInnen:
Prof. Dr. Helge Bruelheide
Prof. Dr. Isabell Hensen
Prof. Dr. Holger Kreft

Datum der Verteidigung: 06.04.2021

“Who can explain why one species ranges widely and is very numerous,
and why another allied species has a narrow range and is rare?”

Charles Darwin, On the Origin of Species (1859)

Contents

Summary

In plant ecology abundance is a measure of spatial density of individuals and species performance. At the interspecific level, there is an observed tendency of widespread species to also occur in higher densities compared to species restricted in their geographic distribution. However, at the intraspecific level, abundance patterns found at fine spatial scales are less clearly related to occurrence patterns at broad spatial scales, and quite often species are found at low abundance at the local scale yet are widespread at the global scale. Understanding spatial patterns of species abundance is of interest for applied issues such as the identification of suitable habitat conditions. Therefore, it remains crucial to understand whether species local abundances follow the same pattern as the species broad-scale distribution.

Plant species functional traits are used as proxies for species dispersal abilities, tolerance of environmental conditions and competitiveness and have been linked to species commonness and rarity on both, local and broad geographic scale. Biodiversity databases, which are a fundamental resource for studying and understanding macroecological patterns in species distribution, often suffer from sampling gaps or biases that can limit their applicability in ecological research. Therefore, a representative sampling coverage in both geographic and climatic space is crucial for reliable analyses of distribution patterns.

In this thesis, broad-scale range attributes were calculated based on Eurasian range maps for more than 500 species and combined with fine-scale abundance information from more than 800,000 vegetation plots and 20 functional traits that are expected to capture the essence of plant life forms and functions.

To support my analyses, I developed the Dynamic Match Coefficient, a measure to quantify the sampling coverage of species ranges by vegetation plots. I detected a positive correlation between the observed sampling coverage of species ranges by vegetation plots and the expected sampling coverage, based on null models, for both the geographic space and the climatic space. In my analyses I found no clear support for three macroecological theories that relate species local abundance with (1) their broad-scale distributions in geographic and climatic space, (2) the location of sampling plots within a species' range or (3) the predicted climatic suitability for species at the plot location. This indicated that, across entire species ranges, the distribution of abundance is highly heterogenous, because local drivers strongly influence plant species growth performance across their global range.

Finally, species local abundances were much more strongly related to traits than to geographic distribution attributes. Although no clear relationship could be found in species patterns of local abundance with broad-scale distribution, traits related to the leaf economics spectrum were found to be important for species abundance and occurrence at both spatial scales. This finding emphasizes the general importance of resource acquisition strategies for the abundance and distribution of vascular plant species.

Zusammenfassung

In der Pflanzenökologie ist die Abundanz ein Maß für die räumliche Vorkommensdichte von Individuen und die individuelle Leistung einer Art. Auf interspezifischer Ebene zeigen weitverbreitete Arten eine Tendenz zu höherer Vorkommensdichte im Vergleich zu jenen Arten mit begrenzter geographischer Verbreitung. Auf intraspezifischer Ebene hingegen, scheinen lokale Abundanzmuster weniger deutlich mit großräumigen Verbreitungsmustern in Beziehung zu stehen. So weisen Arten auf lokaler Skala oftmals eine geringe Abundanz auf, können auf großräumiger Skala hingegen weit verbreitet sein. In der praktischen Anwendung sind Abundanzmuster beispielsweise für die Identifikation von zukünftig geeigneten Schutzgebieten von Interesse. Daher ist es wichtig zu verstehen, ob und inwieweit die lokale Abundanz einer Art dem Muster ihrer großräumigen Verbreitung folgt.

Funktionelle Merkmale von Pflanzen, die als ein Maß für deren Ausbreitungsfähigkeit, ihrer Toleranz gegenüber Umweltbedingungen und Konkurrenzfähigkeit angesehen werden, wurden in Verbindung mit der Häufigkeit und Seltenheit einer Art, sowohl auf lokaler als auch auf großräumiger Skala, gebracht. Biodiversitätsdatenbanken, die eine fundamentale Grundlage für Untersuchungen und das Verständnis makroökologischer Muster in der Verbreitung von Arten sind, weisen oftmals Datenlücken oder Inhomogenitäten in der räumlichen Verteilung von Stichproben auf, die ihre Anwendbarkeit in der ökologischen Forschung einschränken können. Aus diesem Grund ist eine repräsentative Abdeckung von Stichproben im geographischen und klimatischen Raum wichtig, um Verbreitungsmuster verlässlich analysieren zu können.

In dieser Arbeit nutzte ich Verbreitungskarten um das großräumige Vorkommen von mehr als 500 Gefäßpflanzenarten im eurasischen Raum zu bestimmen. Diese Informationen verknüpfte ich mit lokalen Abundanzdaten aus mehr als 800,000 Vegetationsaufnahmen und mit Informationen zu 20 funktionellen Merkmalen, von denen angenommen wird, dass sie die Lebensformen und Funktionen von Pflanzen widerspiegeln.

Um meine Analysen zu stützen, habe ich den Dynamic Match Coefficient entwickelt, ein Maß, um die Abdeckung der Artareale durch Vegetationsaufnahmen zu quantifizieren. Ich konnte einen positiven Zusammenhang zwischen der realen Abdeckung der Artareale durch Vegetationsaufnahmen und der erwarteten

Abdeckung, die mittels Null-Modellen berechnet wurde, sowohl im geographischen als auch im klimatischen Raum finden. Jedoch fand ich keine einheitliche Bestätigung zu drei getesteten makroökologischen Theorien, die die lokale Abundanz in Beziehung setzen mit (1) der großräumigen Verbreitung einer Art im geographischen und klimatischen Raum, (2) der geographischen Lage einer Vegetationsaufnahme innerhalb des Areals, und (3) der vorhergesagten klimatischen Eignung am Ort der Vegetationsaufnahme. Diese Ergebnisse deuten darauf hin, dass die lokale Abundanz einer Art über ihr gesamtes Verbreitungsgebiet von sehr komplexen Einflüssen bestimmt wird.

Des Weiteren hat sich gezeigt, dass die lokale Abundanz einer Art stärker als ihre großräumige Verbreitung in Zusammenhang mit funktionellen Merkmalen steht. Obwohl kein eindeutiger Zusammenhang zwischen Mustern der lokalen Abundanz und der großräumigen Verbreitung gefunden wurde, zeigte sich, dass jene funktionellen Merkmale, die dem ‚leaf economics spectrum' zugeordnet werden, eine signifikante Bedeutung für die Abundanz und die großräumige Verbreitung der Arten auf beiden räumlichen Skalen zeigen. Dieses Ergebnis verdeutlicht, dass Kombinationen funktioneller Merkmale, die die Lebensstrategien von Arten charakterisieren, eine generelle Bedeutung für die Verbreitung von Gefäßpflanzenarten haben.

Chapter I

General Introduction

Biogeography aims at discovering, describing, and understanding geological, environmental and historic factors that determine the geographic distribution of organisms and biodiversity (Lomolino, Riddle, & Whittaker, 2007). In macroecological frameworks, one key goal is to identify the processes that underlie the broad-scale patterns by applying non-experimental and statistical investigations (Brown, 1995). Plant species abundance is understood as the result of species' growth and demographical performance and is influenced by biotic and abiotic factors acting at the local scale (Peterson et al., 2011). Analyses of species abundance patterns have been used to develop hypotheses about applied issues such as potential responses to climate change or the identification of suitable future locations for natural reserves (McGill & Collins, 2003). At the interspecific level, there is an observed tendency of widespread species to also occur in higher densities compared to species that are restricted in their geographic distribution (Brown, 1984). However, at the intraspecific level, abundance patterns found at fine spatial scales are less clearly related to occurrence patterns at broad spatial scales, and quite often species are found at low abundance at the local scale yet are widespread at the global scale (Rabinowitz, 1981; Murray & Lepschi, 2004). Most conservation strategies are implemented at the local scale but follow guidelines based on predictions that were made at the global or regional scale (Guerrero, McAllister, Corcoran, & Wilson, 2013). Therefore, it remains crucial to understand whether species local abundance follows the same pattern as the species broad-scale distribution and, following from this assumption, is mainly driven by the same predictor variables such as climatic conditions.

Plant functional traits that are related to productivity, competitive ability, dispersal, regeneration and persistence have been linked to species commonness and rarity (Gaston & Kunin, 1997). It remains a central question in macroecology to understand to what degree functional traits affect species local abundance and broad-scale distribution (McGill, Enquist, Weiher, & Westoby, 2006). It has been stated that studies carried out at a broader scale tend to be more representative of the global range extent of species (Sagarin, Gaines, & Gaylord, 2006; Pironon et al., 2017). Therefore, it would be crucial to investigate all the above-mentioned relationships based on the species' complete distribution range, and at the same time, achieve representativeness from local site measurements.

For this dissertation I had the opportunity to combine and analyse data on vascular plant species distribution throughout their Eurasian range with information on local abundance in vegetation plots all over Europe with measurements on functional traits that were collected from globally distributed sites and studies. All this information, nowadays stored in biodiversity databases and digitally available, is the long-term effort of researchers that went to the field, carried out vegetation surveys, collected plant material, analysed samples in the laboratory and conducted intense literature research. I acknowledge this effort which enabled me to conduct a cross-scale study on abundance distribution patterns and the effect of functional traits on the local abundance and broad-scale distribution of vascular plant species.

Broad-scale distribution metrics: geographic ranges and climatic niches

Over broad geographical extents, expert-drawn range maps are commonly the primary source of species distribution data (Hurlbert & Jetz, 2007). The availability of extensive data sets on species occurrences allows the investigation of macroecological patterns, such as range sizes (Gaston & Fuller, 2009). A measure of species' geographic range size is the area of occupancy (AOO), defined as the area of a species' actual occurrence within the species extent of occurrence (EOO, i.e. the area which lies within the outermost geographic limits to the occurrence of a species) (Gaston, 1991; Gaston & Fuller, 2009).

Following Hutchinson (1957), the fundamental niche is the sum of all environmental factors that constrain a species' performance and survival, conceived in a multidimensional hypervolume, comprising also unpopulated but suitable environmental conditions. As a subset of the fundamental niche, a species' realized niche is the range of abiotic and biotic conditions under which a species actually occurs, thereby taking into account the effects of species interactions and dispersal limitation (Soberón, Jiménez, Golubov, & Koleff, 2007; Colwell & Rangel, 2009).

The niche is a central conception in ecology to describe a species' occurrence in environmental space (Guisan & Thuiller, 2005; Soberón et al., 2007; Colwell & Rangel, 2009). The species range is conceived as a map, of which each occurrence is characterized by geographical coordinates. The environmental conditions at these coordinates define the corresponding niche space and allow for reciprocal projections between the geographic distribution of a species and its niche. However, the rules that define this duality, proposed by Hutchinson (1978), are not proportionally reciprocal. Whereas each point in the geographic space corresponds to exactly one point in niche space, a single point in niche space may correspond to many in the geographic space. With this, the duality between a species' realized climatic niche and geographic range is a suitable way to analyse biogeographical distributions in relation to spatial climatic patterns (Colwell & Rangel, 2009).

Patterns of local abundance throughout species distribution ranges

Species local abundances result of species populations growth and demographical performance and depend on factors like habitat suitability, the combination of environmental conditions and biotic interactions such as competition, operating at the local scale of species assemblages (Peterson et al., 2011; Staniczenko, Sivasubramaniam, Suttle, & Pearson, 2017). Within a species' geographic range, local abundance is often highly variable, and in most species, abundance is high at only a few sites and low at most remaining sites (Murphy, VanDerWal, & Lovett-Doust, 2006). These 'somewhere-abundant' species exhibit a right-skewed abundance frequency pattern, a common pattern in plant community ecology (McNellie, Dorrough, & Oliver, 2019).

Species can be rare or common within a local plant community. Similarly, some species have restricted distribution ranges while other species are widely distributed (Rabinowitz, 1981; Gurevitch, Scheiner, & Fox, 2002). Several macroecological theories relate species local abundance to their broad-scale distributions in geographic and climatic space and three of them, which are highly debated, have been investigated in this thesis.

The *abundance-range size relationship* states that locally rare species tend to be more narrowly distributed, whereas locally abundant species tend also to be more widespread (Brown, 1984; Gaston & Blackburn, 2008). Under the assumption that local abundance is driven by the same climatic factors as the species' geographic distribution, one would expect that species that tolerate a higher variability of climatic conditions are both locally abundant and widespread throughout the geographic range space. Although this relationship has been widely postulated, it has received only mixed empirical support (Gaston, Blackburn, & Lawton, 1997; Guisan & Thuiller, 2005).

The *abundance-range centre relationship* states that species performance (i.e. abundance) decreases from the centre to the edge of the geographic range (Brown, 1984; Hengeveld & Heck, 1982). Based on the assumption that species ranges are the spatial representation of the distribution of their environmental requirements (i.e. the niche), this theory predicts habitat suitability to decrease from the centre to the edge of a species' range. Subsequently, species abundance is expected to show the same distributional pattern. However, climatic conditions do not systematically follow geographic gradients, and in consequence, abundance might not systematically follow a geographical or climatic gradient from the centre to the edge of a species' range (Hargreaves, Samis & Eckert, 2013; Pironon et al., 2017).

Finally, the more recently proposed *abundance-suitability relationship* assumes species to show higher local abundance in environmentally more suitable areas, whereas they should achieve lower abundance in less suitable areas (Weber, Stevens, Diniz-Filho, & Grelle, 2017; Santini, Pironon, Maiorano, & Thuiller, 2019). Often, environmental suitability of a site is estimated by climate-based predictions from species distribution models (SDMs). The above-described *abundance-range centre relationship* serves as a basis for this ecological hypothesis (Van Couwenberghe, Collet, Pierrat, Verheyen, & Gégout, 2013). As species niches might show highest climatic suitability somewhere towards the edges and not in the very centre, the *abundance-suitability relationship* recognizes the non-reciprocal relationship between geographic and climatic gradients.

The effect of plant functional traits on local abundance and broad-scale distribution

Plant functional traits are valuable attributes that describe species morphological, physiological and life-history processes and their fitness and performance (Violle et al., 2007). The plant functional space is reflected by trade-offs between trait constellations that separate slow-growing and fast-growing species (Diaz et al., 2004). The leaf economics spectrum spans from species with the potential for fast return on investments of nutrition and leaf dry mass on the one end (e.g. high nutrient concentration, low leaf dry mass investment per leaf area) to species with slow potential rates of return on the other end of the spectrum (Wright et al., 2004). The global spectrum of plant form and function reflects a species' life-cycle, contrasting species with small stature, small diaspores and short lifespan (i.e. non-woody) with long-lived and tall woody plant species (Diaz et al., 2016).

Species functional traits are assumed to reflect the species' ability to maximise fitness in response to abiotic and biotic conditions on both local and broad geographic scale (Suding et al., 2008). Traits are used as proxies for species dispersal abilities, tolerance of environmental conditions and competitiveness (Thompson, Moles, Auld, & Kingsford, 2011; Bohner & Diez, 2020; Kunstler et al., 2016) and have been linked to species commonness and rarity on both local and broad geographic scale. On both local and broad geographic scales, species with taller stature, as a proxy for a species' competitive ability, were found to be more common and widespread than shorter species (Lavergne, Thompson, Garnier, & Debussche, 2004). Locally more abundant species were found to produce lighter and heavier seeds, while the contrasting pattern was encountered in widespread species that tended to produce heavier seeds in higher numbers (Hedge & Ellstrand, 1999; Kolb, Barsch, & Diekmann, 2006). Regarding species persistence, locally more abundant species have been found to be perennials rather than annuals and to be associated with clonal growth. In contrast, on a broad geographic scale, rare species are associated with clonal growth and longer life cycles, being shrubs rather than herbs (Eriksson & Jakobsson, 1998; Kelly & Woodward, 1996; Oakwood, Jurado, Leishman, & Westoby, 1993).

Many studies investigated single traits instead of trait combinations or trait syndromes and found no unequivocal support for a relationship between functional traits with local abundance and broad-scale distribution (Diaz et al, 2016; Guo et al., 2018). However, no single trait can completely describe a species' ecological strategy (Winnemiller, Fitzgerald, Bower, & Pianka, 2015) and species patterns in local abundance and broad-scale distribution might rather be affected by different sets of traits (Marino et al., 2020).

Advantages and limitations of biodiversity databases

Biodiversity databases, e.g. the Global Biodiversity Information Facility (GBIF; Edwards, Lane, & Nielsen, 2000) or the Botany Information and Ecology Network (BIEN, Enquist, et al., 2009), are a valuable and fundamental resource for studying and understanding macroecological patterns in species distribution (Meyer, Weigelt, & Kreft, 2016). Detailed information on the geographic occurrences of species are crucial for an effective conservation management and advances in ecological theory. However, biodiversity databases often suffer from sampling gaps or biases that can limit their applicability in ecological research. Uneven collection effort that may result in sampling bias includes the under-sampling of specific taxa (e.g. of rare species) or of specific geographical regions (e.g. due to limited physical accessibility of a site caused from terrain conditions or distance from a road (Zizka, Antonelli, & Silvestro, 2020). Furthermore, site-selection bias, i.e. a preference for sites that are either densely populated, must be considered especially when working with species abundance data (Mentges, Blowes, Hodapp, Hillebrand, & Chase, 2020).

At broad spatial scale, climate is one of the most important factors that shape species distribution (Woodward, 1987). Therefore, a representative sampling coverage in both geographic and climatic space is crucial for reliable distribution pattern analyses based on SDMs, which are highly sensitive to poor climatic sampling coverage (Fourcade, Engler, Rödder, & Secondi, 2014). However, until now only few studies have investigated how occurrence data are distributed in geographic and in climatic space (Bruelheide et al., 2019).

For broad-scale abundance analyses over the entire range of a species, sampling coverage of plots from vegetation plot databases, such as sPlot (Bruelheide et al., 2019) or the European Vegetation Archive (EVA, Chytrý et al., 2016), is important. A representative sampling coverage mainly depends on two factors: a sufficient sampling size and the even coverage of geographic and climatic gradients. Sampling coverage depends on the spatial resolution of the study side. Just by chance, sampling coverage will be higher towards a coarser spatial resolution, and therefore, the influence of sampling bias is related to spatial grain. Several studies have already investigated and developed tools to measure sampling biases in biodiversity databases (Meyer et al., 2016; Zizka et al., 2020; Mentges et al., 2020). Other studies have assessed the completeness of biodiversity datasets at different spatial resolution in geographic space (Lobo et al., 2018; Marsh, Barwell, Gavish, & Kunin, 2018). However, studies and tools that consider quantity and quality of species-specific sampling coverage in geographical and climatic space are missing.

Thesis Objectives

So far, the studies that investigated the relationships between species broad-scale distribution and local abundance with functional traits have focused on popular taxa, functional species groups and single traits. In contrast, in this study, I tested the relationships on a large group of herbaceous, dwarf shrub and shrub species over their full Eurasian distribution ranges and included trait syndromes in addition to single traits.

For this purpose, I extracted species distribution maps from the Chorological Database Halle (CDH, E. Welk et al., unpublished data) to derive information on the geographic range size for 564 plant species. CDH stores information on species distribution ranges for more than 17,000 vascular plant species and expert-drawn range maps were compiled for 5,583 taxa based on national and floristic databases and maps from floristic literature (Tralau, 1969-1981; Lundquist & Nordenstam, 1988; Lundquist, 1992; Lundquist & Jäger, 1995-2007). These data are published as distribution range maps (Meusel, Jäger, & Weinert, 1965; Meusel, Jäger, Rauschert, & Weinert,1978; Meusel & Jäger, 1992).

I used the number of 2.5-min grid cells occupied by a species (the area of occupancy, AOO) as a measure of its geographic range size. Species climatic niche size was measured as the number of 2.5-min grid cells occupied in bioclimatic space, based on 19 bioclimatic variables from the WorldClim2.0 database (Fick & Hijmans, 2017).

A test selection of 808,794 vegetation plots was provided by the European Vegetation Archive (EVA). EVA stores information on ~1.5 million vegetation plots from 57 countries on more than 10,000 vascular plant species (Chytrý et al., 2016). As a measure of species local abundance, I used the arithmetic mean of the percentage cover value from all vegetation plots within a 2.5-min grid cell in which a species was present. Finally, I compiled a complete species-trait-matrix of 20 plant functional traits from three trait databases that store information on species leaf and seed traits (TRY, Kattge et al., 2020), species life form and life span (BiolFlor, Kühn, Durka, & Klotz, 2004) and clonality (CLO-PLA, Klimešová, Danihelka, Chrtek, de Bello, & Herben, 2017).

The specific objectives of this thesis can be grouped in three research topics:

In **chapter II,** I propose and test a box-counting method to assess the sampling coverage of species distribution in biodiversity databases in geographic and climatic space across different spatial resolution (Figure I.1). I assessed the sampling coverage of 808,794 vegetation plots provided by EVA on the geographic ranges and climatic niches of 564 species derived from CDH. I applied null models to compare the observed sampling coverage with expectations from a random distribution of vegetation plots across a species' geographic range and climatic niche.

In **chapter III,** I test three macroecological hypotheses that link species local abundance to the following broad-scale distribution properties: (a) the interspecific abundance-range size relationship, (b) the intraspecific abundance-range centre relationship and (c) the intraspecific abundance-suitability relationship throughout the species whole Eurasian distribution ranges (Figure I.1). In this third chapter, I included broad-scale distribution data of 517 from the above mentioned 564 herbaceous, dwarf shrub and shrub species. The estimated local species abundances were based on 744,513 vegetation plots from EVA.

In **chapter IV,** I investigate whether species local abundance and broad-scale distribution can be predicted by single functional traits and sets of traits (trait syndromes) (Figure I.1). In this chapter, I included distribution information of 456 from the above mentioned 564 herbaceous, dwarf shrub and shrub species. The estimated local species abundances were based on 740,113 vegetation plots from EVA and used to calculate the species-specific skewness of cover values. The selected 20 functional traits are expected to capture the essence of plant life forms and functions (Wright et al., 2004, Diaz et al., 2016).

A synthesis in **chapter V** highlights the relationships between the partial results and creates the frame for the general discussion.

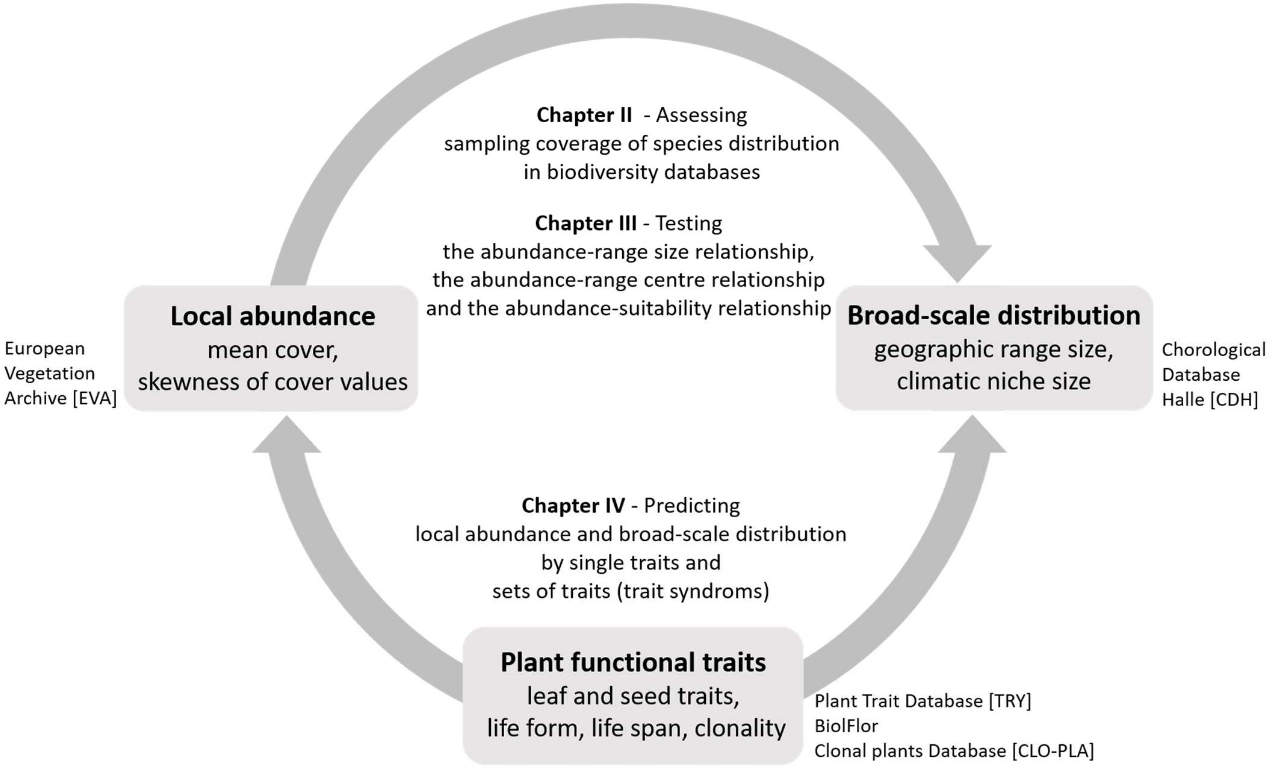

Figure I.1 Overview of the three objectives investigated in chapters II to IV in this thesis. The used database data on species local abundance, broad-scale distribution and functional traits are highlighted in grey boxes.

Chapter II

Assessing sampling coverage of species distribution in biodiversity databases

This chapter is published in *Journal of Vegetation Science* as

Sporbert, M., Bruelheide, H., Seidler, G., Keil, P., Jandt, U., Austrheim, G., Biurrun, I., Campos, J.A., Čarni, A., Chytrý, M., Csiky, J., De Bie, E., Dengler, J., Golub, V., Grytnes, J.-A., Indreica, A., Jansen, F., Jiroušek, M., Lenoir, J., Luoto, M., Marcenò, C., Moeslund, J. E., Pérez-Haase, A., Rūsiņa, S., Vandvik, V., Vassilev, K., & Welk, E. (2019). Assessing sampling coverage of species distribution in biodiversity databases. *Journal of Vegetation Science*, 30(4), 620–632

Abstract

Aim: Biodiversity databases are valuable resources for understanding plant species distributions and dynamics, but they may insufficiently represent the actual geographic distribution and climatic niches of species. Here we propose and test a method to assess sampling coverage of species distribution in biodiversity databases in geographic and climatic space.

Location: Europe.

Methods: Using a test selection of 808,794 vegetation plots from the European Vegetation Archive (EVA), we assessed the sampling coverage of 564 European vascular plant species across both their geographic ranges and realized climatic niches. Range maps from the Chorological Database Halle (CDH) were used as background reference data to capture species geographic ranges and to derive species climatic niches. To quantify sampling coverage, we developed a box-counting method, the Dynamic Match Coefficient (DMC), which quantifies how much a set of occurrences of a given species matches with its geographic range or climatic niche. DMC is the area under the curve measuring the match between occurrence data and background reference (geographic range or climatic niche) across grids with variable resolution. High DMC values indicate good sampling coverage. We applied null models to compare observed DMC values with expectations from random distributions across species ranges and niches.

Results: Comparisons with null models showed that, for most species, actual distributions within EVA are deviating from null model expectations and are more clumped than expected in both geographic and climatic space. Despite high interspecific variation, we found a positive relationship in DMC values between geographic and climatic space, but sampling coverage was in general more random across geographic space.

Conclusion: Because DMC values are species-specific and most biodiversity databases are clearly biased in terms of sampling coverage of species occurrences, we recommend using DMC values as covariates in macroecological models that use species as the observation unit.

Keywords: Chorological Database Halle (CDH), climatic niche, Dynamic Match Coefficient (DMC), European Vegetation Archive (EVA), macroecology, multi-scale, realized niche, sampling bias, spatial scale, species range, vascular plant, vegetation-plot databases.

Introduction

Broad-scale biodiversity databases (e.g. Global Biodiversity Information Facility (GBIF), Edwards, Lane, & Nielsen, 2000; Botany Information and Ecology Network (BIEN), Enquist, Condit, Peet, Schildhauer, & Thiers, 2009; sPlot, Bruelheide et al., 2019) are valuable resources for understanding species distributions and dynamics. Possible applications include broad-scale analyses across species or community types (e.g. Bruelheide et al., 2018; Jiménez-Alfaro et al., 2018), species distribution models (SDM) (Gomes et al., 2018; Wasof et al., 2015); and monitoring biodiversity changes over time (Bertrand et al., 2011; Jandt, von Wehrden, & Bruelheide, 2011). For broad-scale analyses covering the entire range of species, the quality of the sampling coverage across a given species range or throughout its realized niche is crucial. Hence, consistent data distribution is highly desirable across both the geographic and environmental space (Broennimann & Guisan, 2008; Pearman, Guisan, Broenniman, & Randin, 2008; Troia & McManamay, 2016). However, biodiversity databases often suffer from sampling gaps and biases limiting their application potential. Because of the uneven collection effort (Daru et al., 2018; Soria-Auza & Kessler, 2007; Speed et al., 2018) often caused by difficult access to some areas (Sousa-Baena, Garcia, & Peterson, 2014), broad regions of the world remain poorly sampled. Even comprehensive databases of species occurrences in well-surveyed regions are prone to geographic (Yang, Ma, & Kreft, 2013) and taxonomic biases (Pyke & Ehrlich, 2010; Soberón, Jiménez, Golubov, & Koleff, 2007). In an in-depth evaluation, Meyer, Weigelt, & Kreft (2016) found severe geographical bias in the GBIF database (Edwards et al., 2000), concluding that data limitations are rather the rule than the exception for most species and regions.

Species distribution models (SDM) are commonly used for macroecological niche analyses. They represent the estimation of species occurrence probabilities based on observed geographic distributions. Thereby, SDMs are sensitive to poor sampling coverage, especially if spatial bias results in climatically biased sampling (Fourcade, Engler, Rödder, & Secondi, 2014). In such situations, SDMs tend to misestimate species climatic niches (Titeux et al., 2017). Thus, for reliable analyses of biodiversity distribution patterns, sampling coverage needs to be representative for both the climatic and geographic space (Hortal, Jiménez-Valverde, Gómez, Lobo, & Baselga, 2008; Troia & McManamay, 2016). Unbiased sampling is typically obtained by meeting two interrelated requirements: sufficient sample size and even coverage of geographical and environmental gradients. Towards coarser spatial resolution, good coverage is easier to achieve and, as a consequence, sampling bias typically decreases. Consequently, the negative impact of sampling bias is clearly related to spatial grain. Several studies have analysed the importance of spatial scaling in niche studies (e.g. Pearman et al., 2008; Soberón et al., 2007; Hortal, Borges, & Gaspar, 2006). Recently, procedures have been developed to assess the completeness of a spatial dataset at different spatial resolutions in geographic space (*KnowBR,* Lobo et al., 2018; *downscale*, Marsh, Barwell, Gavish, & Kunin, 2018). At large spatial extent, climate is among the most important factors determining species distributions (Woodward, 1986). However, although including climate seems straightforward, until now, few studies have accounted for how evenly occurrence data cover species ranges in climatic space (e.g. Bruelheide et al., 2018). To our knowledge, no study has explicitly tested the degree to which the spatial distribution of occurrences represents the geographical range as well as the climatic niche of the sampled species.

Here we test the spatial and climatic coverage of plant occurrence data using an example dataset of the European Vegetation Archive (EVA). EVA is a key macroecological resource that incorporates information from 57 countries on approximately 1.5 million vegetation plots containing more than 10,000 vascular plant species (Chytrý et al., 2016). EVA data are used for various research objectives, yet the degree of unevenness in sampling effort across Europe's geographic and environmental space is unclear. A species' distribution database covering EVA's spatial extent, but otherwise independent from EVA, is the Chorological Database Halle (CDH) (Welk et al., unpubl.). CDH stores georeferenced information (range polygons and point occurrences) on the distribution range of more than 1,200 European vascular plant species. Species distribution data from CDH have already been used in several biodiversity studies (e.g. Csergő et al., 2017; San-Miguel-Ayanz, de Rigo, Caudullo, Houston Durrant, & Mauri, 2016; Schleuning et al., 2016) and as basis for biogeographical experiments on plant range limits (Bütof et al., 2012; Hofmann, Bütof, Welk, & Bruelheide, 2013; Welk, Welk, & Bruelheide, 2014). Here, we made use of expert-based range maps stored in CDH to extract information on both the species' geographic ranges and climatic niches and assess the sampling coverage of species occurrences stored in EVA across each of these two backgrounds (geographic and climatic).

To quantify sampling coverage, we developed the Dynamic Match Coefficient (DMC), a measure based on the area-under-the-curve (AUC) derived from threshold-independent box-counting statistics across variable spatial grains. We compared the observed DMC values with the values of plots randomly distributed across the species range and niche. Thereby, we produced an expected null reference distribution (Nunes & Pearson, 2017) within both the geographic and climatic space for a given sampling effort (sample size) and corresponding to the observed species frequency in the database. This enabled us to evaluate the observed plot distribution in geographic space (DMC_{GEO}) and climatic space (DMC_{CLIM}) in comparison to expectations of randomly distributed plots across the species range and realized climatic niche. We tested four hypotheses on sampling coverage of species occurrences across both the geographic and climatic space:

(1) Sampling coverage within the climatic space depends strongly on good sampling coverage across the geographic space because climatic conditions are spatially autocorrelated. We expect a positive correlation between sampling coverage in the geographic and climatic space.

(2) Sampling coverage is less representative in the climatic space than in the geographic space. The reason is the asymmetric transferability between points in the climatic and geographic space: a single point within the climatic space might translate to several geographic locations, while a single geographic location can only translate to one point in the climatic space. An increase in sampling coverage within the geographic space might thus be without positive effect on sampling coverage within the climatic space.

(3) Given the general sampling issues of biodiversity databases mentioned above and the heterogeneous nature of their source data, we expect that sampling coverage of the realized niches of plant species by such data is largely imperfect because of an underdispersed (clumped) distribution of species observations within the geographic space and supposedly also within the climatic space.

(4) Finally, for a given range size and macroclimatic niche size, we expect sampling coverage to increase with increasing sample size.

Material and Methods

We assessed the sampling coverage of European vascular plant species ranges (using species range data from the Chorological Database Halle, CDH) by a test selection of species occurrence data taken from vegetation plots from the European Vegetation Archive (EVA, Chytrý et al., 2016). We did this both in the geographic space (distribution range data from CDH) and in the climatic space (realized climatic niche space derived from CDH geographical distributions). We focused on species presence data (i.e. locations of vegetation plots in which the focal species was recorded) and examined the relationship between the geographic and climatic sampling coverage, as well as interspecific variability. The study area comprised all European countries plus Turkey, Georgia, Armenia and Azerbaijan (Figure II.1a).

Background data on species geographic range and climatic niche

The Chorological Database Halle (CDH) stores information on distribution ranges of about 17,000 vascular plant taxa. For 5,583 taxa, maps were compiled based on published distribution range maps (Meusel, Jäger, & Weinert, 1965; Meusel, Jäger, Rauschert, & Weinert, 1978; Meusel & Jäger, 1992), national and floristic databases and further maps from floristic literature (see bibliographic details in Index Holmiensis, Tralau, 1969-1981; Lundqvist & Nordenstam, 1988; Lundqvist, 1992; Lundqvist & Jäger, 1995-2007). CDH data can be requested for research objectives via http://chorologie.biologie.uni-halle.de/choro/. We retrieved from CDH the available geographical information for the distribution ranges of 1,200 European vascular plant species in electronic format (range polygons and point occurrences) in October 2015. The species range information was processed as raster layers of 2.5-min cell resolution, which is about 15 km² in Central Europe (Figure II.1a). The multi-dimensional climatic space (climatic niche) was determined by principal components analysis (PCA) of 19 bioclimatic variables from Worldclim with 2.5-min cell resolution (Hijmans, Cameron, Parra, Jones, & Jarvis, 2005) (for detailed information see Appendix S II.1 in the Supporting Information).

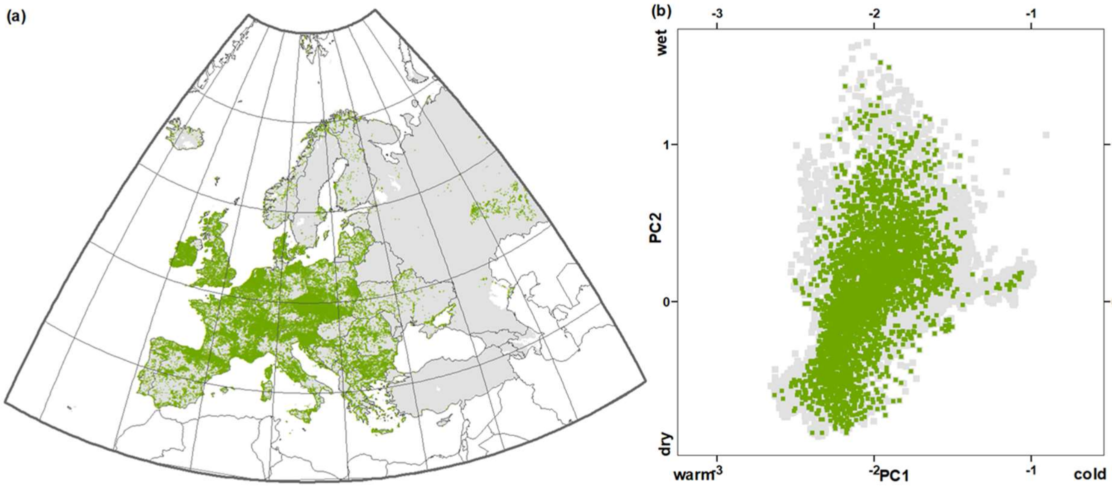

Figure II.1 Distribution of the 808,794 vegetation plots (green dots) extracted from EVA (European Vegetation Archive). Only plots with at least one of the 564 study species are shown. The study species merged distributions based on CDH are represented by grey cells. White areas (large water bodies, glaciers, and deserts) represent regions where none of the studied species occurs. (a) Distribution of vegetation plots in the geographic space. (b) Distribution of vegetation plots in climatic space represented by its first two PCA axes (74.1% and 13.9% variance explained by PC1 and PC2, respectively), where PC1 and PC2 were negatively and positively related to temperature and precipitation, respectively.

Vegetation plots

A test selection of vegetation plots was provided by the European Vegetation Archive in October 2015, containing information on 10,082 species from 933,228 vegetation plots. This selection included all the plots that were available in EVA at that time. Data for intraspecific taxa such as subspecies were merged at the species level. Further, we matched species names and checked for synonyms according to (i) the taxonomic reference list for Germany (German SL version 1.2, Jansen & Dengler, 2008) and (ii) all taxonomic reference lists available via the R package 'taxize' (Chamberlain & Szöcs, 2013; Chamberlain et al., 2018). We excluded trees, bryophytes, lichens, fungi, algae and species exotic to Europe. We also excluded 67,200 vegetation plots with location uncertainty larger than 10 km and 417 species that occurred in less than 10 plots.

After matching EVA and CDH species, 808,794 vegetation plots contained at least one of the 564 vascular plant species (herbs, dwarf shrubs and shrubs) with available digitized geographic distribution data in CDH. A list of these species is available online in the Supporting Information S II.2 (https://doi.org/10.1111/jvs.12763) and all the databases that provided vegetation plot data can be found in Appendix S II.3. The 808,794 vegetation plots from EVA were heterogeneously distributed across the study area in the geographic space. While some geographic regions were represented very well and with high density (e.g. the Czech Republic, the Netherlands), other regions were represented sparsely (e.g. Norway, Sweden, Finland, Belarus, parts of Russia; Figure II.1a). In contrast to geographic space, the study area was well represented by EVA vegetation plots in climatic space, except some marginal parts of the climatic background space (Figure II.1b). The maximum density of species was 396 species per 2.5 min raster cell in geographic space (Figure II.2a) and 528 species per cell in climatic space (Figure II.2b). Stacked CDH ranges of the 564 study species covered 98.5% of the study area in geographic space (154,455 raster cells of 2.5-min in total; Figure II.2a) and 100% in climatic space (9,931 cells in total; Figure II.2b).

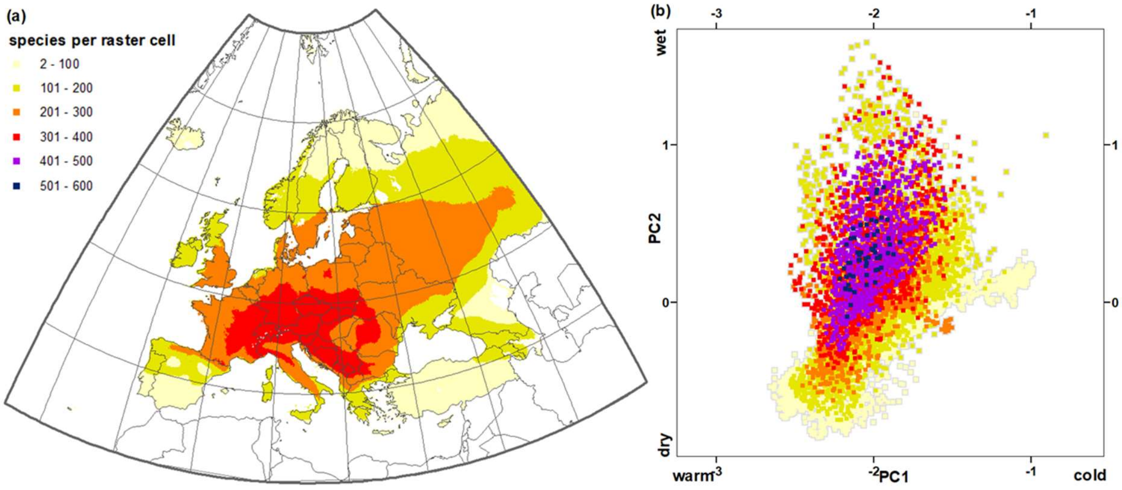

Figure II.2 Study species data density in the geographic and climatic space. (a) Data density on species geographic ranges of 564 vascular plant species included in this study in 2.5-min resolution raster. White areas (large water bodies, glaciers, and deserts) represent regions where none of the studied species occurs. (b) Data density on climatic niches of 564 species in the respective common climatic space represented by its first two PCA axes (74.1% and 13.9% variance explained by PC1 and PC2, respectively), where PC1 and PC2 were negatively and positively related to temperature and precipitation, respectively.

Dynamic Match Coefficient (DMC) - a measure of plot sampling coverage across spatial scales

Sampling bias is mainly a result of two interrelated issues: insufficient number of samples and inadequate sample distribution. The impact of sampling bias is related to spatial scale (spatial extent and grain size) and should decrease with increasing grain size. The spatial arrangement of sampling locations could be evaluated by classical methods of point pattern analysis (Boots & Getis, 1988; Wiegand & Moloney, 2013). However, there are two main issues related to the spatial pattern in the ecological domain of the data of interest. First, because of the generally irregular, often non-contiguous geometry of plant distribution ranges, traditional Euclidean geometry often fails to estimate characteristics of point patterns correctly (Pentland, 1984). Second, species ranges and niches cannot be regarded as merely geometric phenomena. Spatio-temporal population processes often result in complex range structures of genetic diversity, demographic performance and abundance (Peterson et al., 2011; Ricklefs, 2004).

To measure how well, i.e. how uniform vs. clustered and simultaneously how dense or scarce vegetation plots containing the focal species are located across the species' range or niche, we developed a measure inspired by fractal dimension analysis (Hall & Wood, 1993), which we call the Dynamic Match Coefficient (DMC). The DMC represents a measure of cell matches between a point pattern and spatial layers that are iterated across different raster cell resolutions (grain sizes), from fine to coarse (Figure II.3). Here, 20 iterative scaling steps were used, which resulted in a maximum achievable DMC of 2000 (20 × 100% match). The obtained values were standardized to 0-1. For all species, the starting grain size in geographic space was 1/20th of the respective species maximum North-South and East-West range extent. Hence, the initial grain size was smaller for small-range species (e.g. 50 km × 20 km for *Centaurea deustiformis*) than for large-range species (e.g. 211 km × 273 km for *Plantago major*) (see Appendices S II.2 and S II.4.1 in the Supporting Information for distribution of initial grain sizes in DMC calculations). Among the chosen starting grain sizes for the geographic space, even the finest grid cells (50 km × 20 km) are at a spatial resolution where climate conditions are considered the most important (Pearson & Dawson, 2003). The scaling procedure used in the climatic space was similar to that in the geographic space. Here the initial grain size was derived as the 1/20th fraction of the respective species maximum niche extent along the first two PCA axes. High DMC values indicate high sampling coverage, i.e. a more regular distribution and density of EVA vegetation plots across a species distribution range or within its realized climatic niche. In contrast, low DMC values indicate underdispersed sampling coverage, i.e. clumped distribution and/or inappropriately low density of EVA vegetation plots across a species distribution range or within its realized climatic niche (Figure II.3).

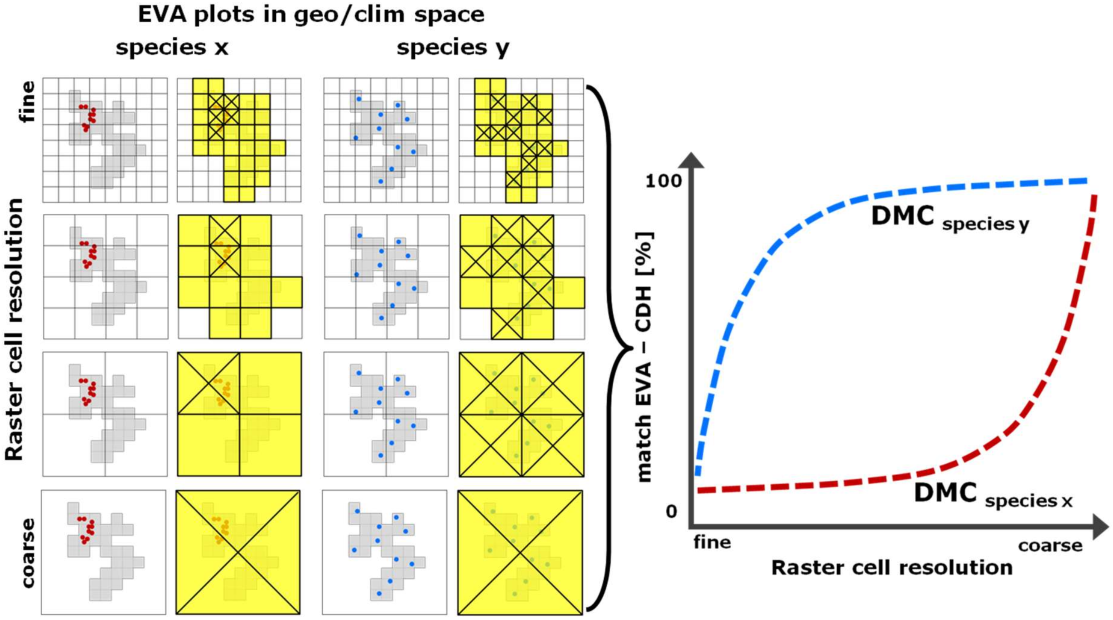

Figure II.3 Dynamic Match Coefficient (DMC) calculated for two example species X and Y with different plot distributions but similar ranges and climatic niches. DMC measures sampling coverage from fine resolution to coarse resolution as the area under the curve (AUC). Scaling for species X, with clumped plots (10 red dots) in the species range or climatic niche (grey background), results in a low DMC value. Scaling for species Y, with more regularly distributed plots (10 blue dots) in the species range or climatic niche (grey background), results in a high DMC value.

Figure II.4 shows how the DMC approach works for the geographic and climatic space and for two contrasting species: *Hieracium murorum*, a species with clumped distribution in EVA plots, and *Calluna vulgaris*, a species with a more regular distribution in EVA plots, both in the species range and in the realized climatic niche (Figure II.4a). Range size and the number of vegetation plots are similar in both species. The cell match ratio between species range and EVA vegetation plots was calculated in 20 iterations from fine to coarse raster cell resolution for both species in the geographic and climatic space (Figure II.4b). The cell match ratio at the 20 single raster steps was summed up, and this sum is what we term the final DMC value of a species in the geographic space (DMC_{GEO}) and climatic space (DMC_{CLIM}). For *Hieracium murorum*, DMC values reached 0.42 and 0.58 for the geographic (DMC_{GEO}) and climatic (DMC_{CLIM}) space, respectively. For *Calluna vulgaris*, DMC values reached 0.74 for both the geographic (DMC_{GEO}) and climatic (DMC_{CLIM}) space.

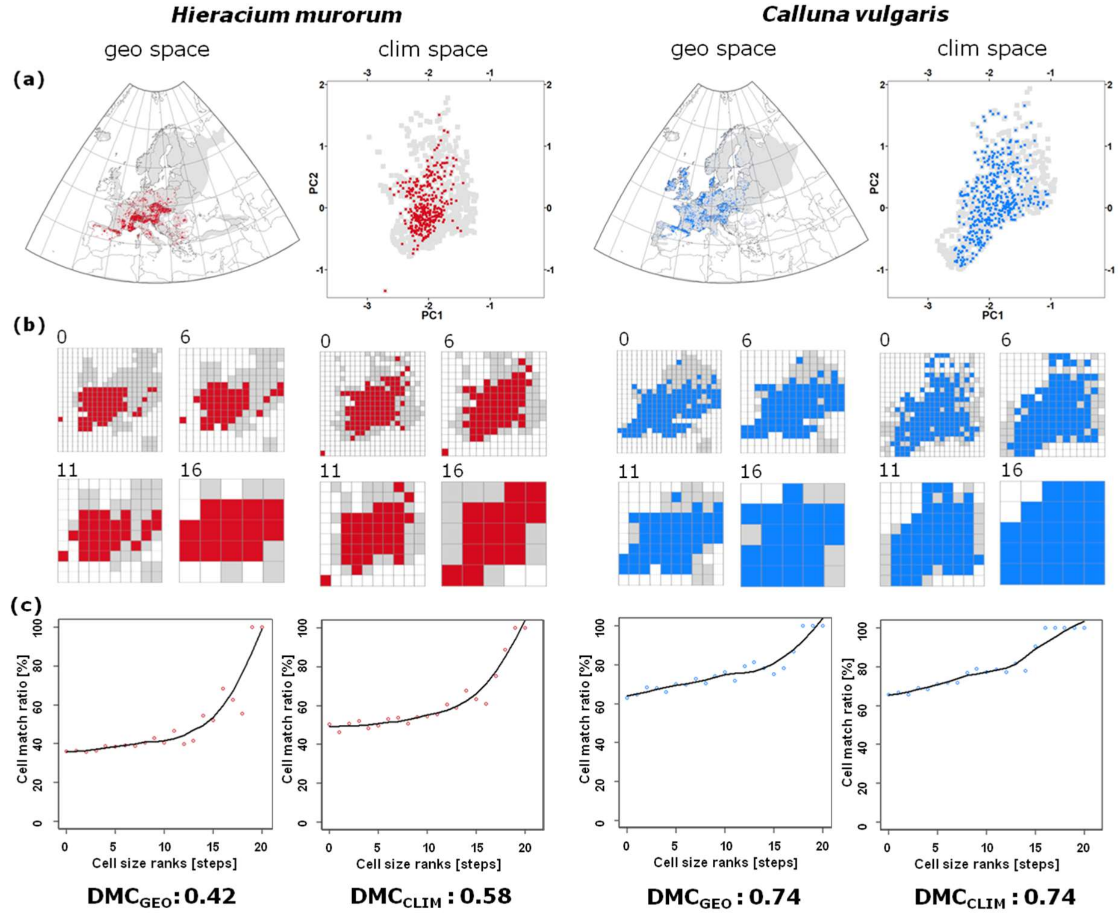

Figure II.4 The DMC scaling approach applied to the distribution of EVA vegetation plots inside species ranges in geographic space and inside species niches in climatic space (grey cells). (a) The distribution of EVA plots containing *Hieracium murorum* (left, red) and *Calluna vulgaris* (right, blue). (b) Four selected scaling steps from fine to coarse raster-cell resolution in geographic space (left-hand four panels in each set) and climatic space (right-hand four panels in each set). (c) The resulting DMC curves along 20 scaling steps, where the cell match ratio is the percentage of grey raster cells (species range or climatic niche) matched by a vegetation plot containing the species.

In all cases, the maximum achievable DMC is 1 (100% cell match in all scaling steps). DMC values reached 0.42 and 0.58 for the geographic ($\mathrm{DMC_{GEO}}$) and climatic ($\mathrm{DMC_{CLIM}}$) space for *Hieracium murorum* and 0.74 for both the geographic ($\mathrm{DMC_{GEO}}$) and climatic ($\mathrm{DMC_{CLIM}}$) space for *Calluna vulgaris*.

Observed vs. expected distributions

In order to quantify how far the observed DMC deviates from an expected random distribution, we applied a null model simulation (Nunes & Pearson, 2017) for each species. We randomly distributed a number of species occurrences for each species (n = number of plots containing the species) across its geographic range and climatic niche. We calculated the $\mathrm{DMC_{GEO}}$ and $\mathrm{DMC_{CLIM}}$ values for 100 such random distributions in the geographic and climatic space, respectively, and compared the simulated DMC distribution with the observed value. To quantify the deviation of the observed DMC value from the median of the simulated ideal random distribution ($\mathrm{DMC_{NULL}}$) we calculated a DMC ratio as:

$$DMC\ ratio\ = \frac{(DMC\ NULL - DMC\ observed)}{DMC\ observed}$$

A high DMC ratio corresponds to an underdispersed distribution of the EVA plots containing the species, while a low DMC ratio corresponds to a more random distribution. A negative ratio corresponds to an overdispersed distribution.

Effect of sample size on the DMC value

We analysed the effect of sample size (number of EVA plots containing a given species) on DMC values while accounting for range size (or niche size) by applying linear models with $\mathrm{DMC_{GEO}}$ (or $\mathrm{DMC_{CLIM}}$) values as the response variable, sample size as the main explanatory variable and range size (resp. niche size) as a covariate to correct for potential confounding effects of range size or niche size. In a first step, for each species, the percentage match of the species range (derived from CDH) by the respective EVA vegetation plots where the species occurred was calculated at 2.5-min raster cell resolution. Multiple occurrences per raster cell were reduced to presence-absence data per species and 2.5-min raster cell. In the second step, species ranges and the respective vegetation plots were projected into the climatic space. The study area in the climatic space is well represented by its first two PCA axes, which explain 88.0% of the data variance (for details see Appendix S1 in Supporting Information). Finally, the percentage of a species climatic niche matched by vegetation plots where the species occurred was calculated as the ratio of PCA cells of the respective EVA vegetation plots where the species occurred to all raster cells matched by the species range in the PCA space (species percentage match of its range and niche by EVA vegetation plots is provided in Appendix S II.2 in the Supporting Information).

Results

Overall, sampling coverage of European vascular plant species ranges by EVA vegetation plots was more complete within the geographic space than within the climatic space (Figure II.5), i.e. consistently higher DMC values were within the geographic space (DMC_{GEO}). The mean of DMC_{GEO} was slightly higher than that of DMC_{CLIM}, with values of 0.56 and 0.49, respectively. Species DMC_{GEO} values ranged from 0.08 to 0.94. For half of the species the DMC_{GEO} was between 0.48 and 0.65 (25th and 75th percentile). DMC_{CLIM} values ranged from 0.08 to 0.82 and for half of the species the DMC_{CLIM} was between 0.40 and 0.60 (25th and 75th percentile). We found a highly significant positive correlation (Spearman´s rho = 0.768; $p < 0.001$) between species geographic DMC values (DMC_{GEO}) and their climatic DMC values (DMC_{CLIM}) (Figure II.5). DMC_{CLIM} values were higher than DMC_{GEO} values for only 119 species (21.1%), while 445 species (78.9%) had higher DMC_{GEO} values than DMC_{CLIM} values. Furthermore, some species showed a high deviation in DMC values between the geographic and climatic space. For instance, *Arabis alpina* was more randomly sampled within the climatic space (DMC_{CLIM}: 0.55) than within the geographic space (DMC_{GEO}: 0.24), while this was the opposite for *Vinca major* (DMC_{GEO}: 0.63, DMC_{CLIM}: 0.29). In general, a positive relationship between species range size and niche size could be observed (Spearman´s rho = 0.805; $p <0.001$; Appendix S II.4.2 in Supporting Information).

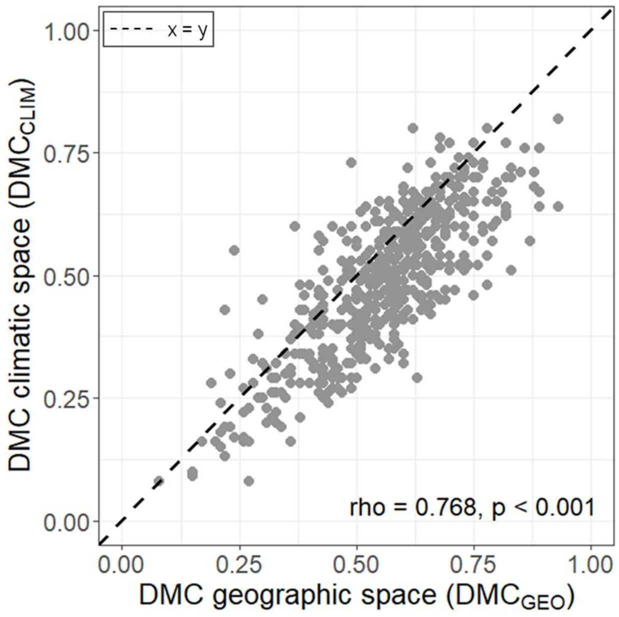

Figure II.5 Scatterplot and Spearman correlation coefficients (rho) of the relationship between DMC values in geographic space (DMC$_{GEO}$) and DMC values in climatic space (DMC$_{CLIM}$) for 564 plant species. Low DMC values indicate an underdispersed (more clumped) distribution of species occurrences in EVA vegetation plots, while high DMC values indicate a homogenous distribution in EVA vegetation plots, in the geographic range or realized climatic niche of a species.

Deviation of the observed DMC from the expected random distribution

We found a positive correlation between the observed DMC values and the expected DMC values, based on our null model, for both the geographic space (weaker, Spearman´s rho = 0.389; *p* <0.001) and the climatic space (stronger, Spearman´s rho = 0.824; *p* <0.001) (Figures II.6a and b). Importantly, a large majority (92.0%) of the observed species distributions in EVA were significantly underdispersed in both the geographic and climatic space. This is indicated by the position of most of the points above the 1:1 line, especially in the climatic space. Exceptionally, for a small number of species in the geographic space (43 species, 7.6%) (Figure II.6a) and for two species in the climatic space (Figure II.6b), the observed DMC values were higher than the null random expectation, indicating overdispersion.

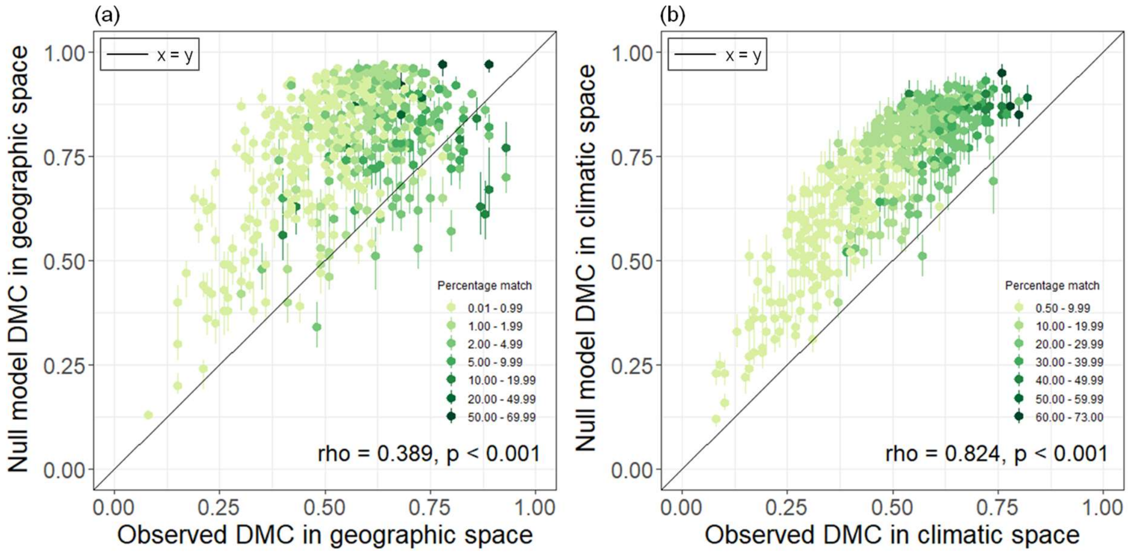

Figure II.6 Scatterplots and Spearman correlation coefficients (rho) of the relationships between the observed DMC and expected DMC derived by null models for (a) geographic space and (b) climatic space. Dots are medians; lines are inter-quartile ranges of the simulations from the null model. Colour gradient represents the percentage match of a species range by EVA vegetation plots in the geographic space (match at 2.5-min raster cell resolution) or climate space (ratio of PCA cells matched by EVA plots to all species-specific raster cells matched by the geographic range data in the PCA space).

For each species, we calculated the deviation of the observed DMC values from the null model DMC values in geographic and climatic space. While a low deviation of the observed DMC values from the null expectation indicates a more regular distribution of occurrences for a given species across its reference range or realized climatic niche, a high deviation indicates an underdispersed (more clumped) distribution. We found a positive correlation for the deviation of observed DMC values from the null model DMC values between geographic and climatic space (Spearman´s rho = 0.615; p <0.001). Despite a higher variability, DMC deviation from the null model was on average slightly lower in geographic space ($\text{min}_{\text{DEV\_GEO}}$: -0.31, $\text{max}_{\text{DEV\_GEO}}$: 2.47, $\text{median}_{\text{DEV\_GEO}}$: 0.46) than in climatic space ($\text{min}_{\text{DEV\_CLIM}}$: -0.10, $\text{max}_{\text{DEV\_CLIM}}$: 2.09, $\text{median}_{\text{DEV\_CLIM}}$: 0.47, see Figure II.7).

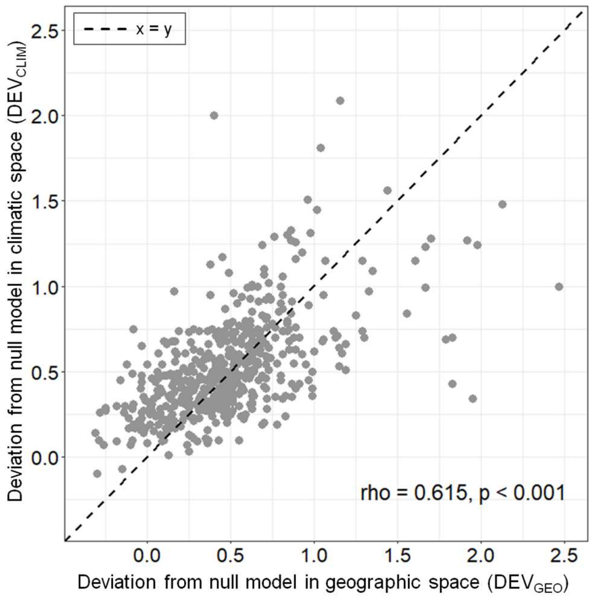

Figure II.7 Scatterplot and Spearman correlation coefficients (rho) of the relationship between the deviation of the observed DMC values from null model DMC values in the geographic space (DEV$_{GEO}$) and in climatic space (DEV$_{CLIM}$). Low deviation of the observed DMC values from the null expectation indicates a more regular distribution of occurrences for a given species across its reference range or realized climatic niche, a high deviation indicates an underdispersed (more clumped) distribution.

Effect of sample size on DMC values

In geographic space, the percentage match of species ranges by EVA vegetation plots containing the same species (measured as the percentage of the range containing the EVA plots at 2.5-min raster cell resolution) ranged from 0.01% to 67.6%. For half of the species, the percentage match was between 0.5% and 2.3% (25th and 75th percentile), with a mean of 1.1% in the geographic space. In the climatic space, the percentage match of species niches by EVA vegetation plots ranged from 0.5% to 72.7% and for half of the species the percentage match was between 7.6% and 22.1% (25th and 75th percentile), with a mean of 14.1%. The applied linear models revealed a positive effect of sample size (vegetation plots) on DMC values while accounting for range size or niche size in both the geographic space (multiple R^2: 0.212) and climatic space (multiple R^2: 0.571). We found a significantly positive correlation between the percentage match of the species range by EVA plots in both the geographic space (Spearman´s rho= 0.726; *p* <0.001) and climatic space (Spearman´s rho= 0.901;

p <0.001) (Figure II.8a and b). Furthermore, we encountered a significantly negative relationship between percentage match of species ranges by EVA vegetation plots and deviation from the null model in the geographic space (Spearman´s rho= -0.601; p <0.001) and climatic space (Spearman´s rho= -0.651; p <0.001) (Figure II.8c and d). Apart from this, a significantly positive correlation between the percentage match of the species range by EVA plots in the geographic space and climatic space could be found (Spearman´s rho= 0.865; p <0.001; Appendix S II.4.3 in Supporting Information).

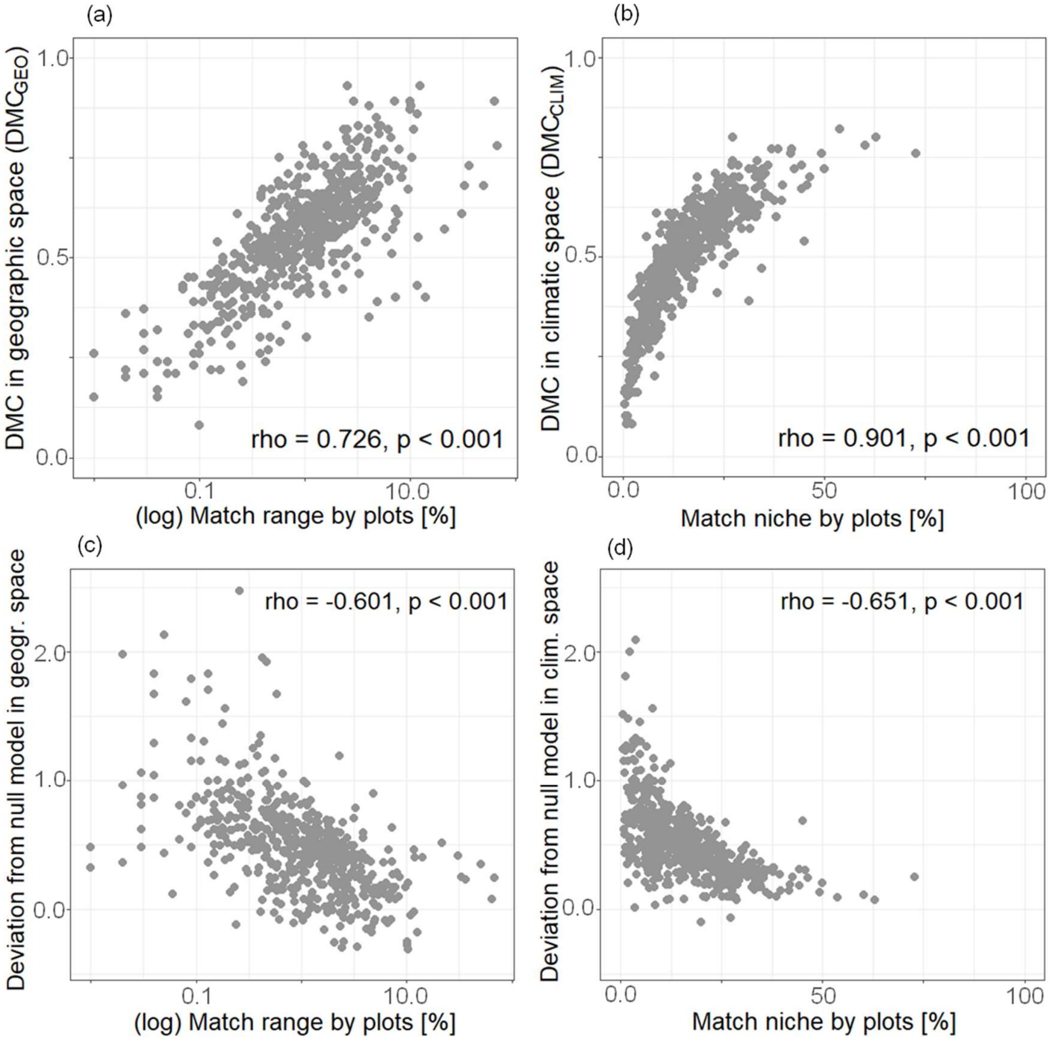

Figure II.8 Scatterplots and Spearman correlation coefficients (rho) of the relationships between percentage match of species ranges by EVA vegetation plots and (a) observed DMC in geographic space (DMC$_{GEO}$); (b) observed DMC in climatic space (DMC$_{CLIM}$); (c) deviation of observed DMC values from null model DMC values in geographic space (DEV$_{GEO}$); (d) deviation of observed DMC values from null model DMC values in climatic space (DEV$_{CLIM}$).

Discussion

Plot sampling coverage across spatial scales

In line with the general positive relationship between range size and niche size (see Appendix S4.2 in Supporting Information), we assumed that (1) a species will be well sampled throughout its multidimensional climatic niche (reaching high DMC_{CLIM} values) only if it is well sampled throughout its geographic range (high DMC_{GEO} values). The demonstrated positive correlation between DMC_{CLIM} and DMC_{GEO} confirms the first hypothesis. However, the relationship was far from perfect, since there are also species that are well sampled within the geographic space (reaching high DMC_{GEO} values) but less well sampled in the climatic space (reaching low DMC_{CLIM} values), and vice versa. Exceptions from the suggested positive relationship can arise especially due to high spatial heterogeneity in climatic conditions, e.g. in mountain regions (Hirst, Griffin, Sexton, & Hoffmann, 2017; Köckemann, Buschmann, & Leuschner, 2009).

Because of the one-to-n relationship between climatic and geographic data points we expected (2) a sparser species sample coverage (lower DMC values) in the climatic space. Accordingly, we found that the sampling coverage (DMC value) of species distribution in EVA was more random in the geographic space (DMC_{GEO}) than in the climatic space (DMC_{CLIM}) for 77.9% of the studied species. This more random sampling coverage in geographic space is explainable by the niche–biotope duality (Hutchinson, 1978). The same combination of climate factors can occur in only one location in geographic space, but will more likely occur in several localities with increasing spatial extent (Colwell & Rangel, 2009; Soberón & Nakamura, 2009). However, the rules that define the niche–biotope duality are not reciprocal (Colwell & Rangel, 2009; Soberón & Nakamura, 2009), and the climatic niche of a species might be fully captured even if only a part of its geographic distribution was sampled (Guisan, Petitpierre, Broennimann, Daehler, & Kueffer, 2014). This seems to be the case for 22.9% of the studied species that occupy ranges with highly heterogeneous climatic conditions (e.g. in mountain regions as mentioned above). For those species, the sampling coverage was higher in the climatic space (DMC_{CLIM}) than in geographic space (DMC_{GEO}).

Broad-scale biodiversity databases consist of heterogeneous, non-systematically sampled datasets with underdispersed observations within the geographic space and supposedly also within the climatic space. We therefore expected (3) the sampling coverage of species geographic ranges and climatic niches to be largely imperfect due to sampling biases. Accordingly, we found limited sampling coverage for most of the studied species. In almost all cases, the observed species distributions in EVA significantly underrepresented both the species geographic range and climatic niche space. It is achievable to identify species which are poorly represented in biodiversity databases relative to their geographic ranges or realized climatic niches (Boakes et al., 2010; Hoffmann et al., 2014). Since the observed and expected DMC values were highly positively correlated, the applied null model approach supports the usefulness of the presented DMC metric to assess sampling bias in the distribution of species occurrences in biodiversity databases.

We assumed that (4) on condition that range size and climatic niche size are correlated, sampling coverage increases with increasing sample size. The applied linear models revealed a positive effect of sample size on DMC values while accounting for range size and niche size, which supports our fourth hypothesis. Nevertheless, especially for the geographical space, high percentage cover of species range by the EVA plots cannot directly indicate high DMC values. In general, the correlation of percentage match of a species range by the EVA plots at 2.5-min raster cell resolution with DMC values was highly positive in geographic space. Nevertheless, there were species with higher percentage match that only reached lower DMC values while there were also species with lower percentage match that reached higher DMC values. Our results show that the number and thereby the density of observations across a species distribution range remains crucial. On the one hand, too small number of plots representing a species distribution range may be a sample of insufficient size even if the plots are distributed randomly (as suggested by the null model calculations). On the other hand, even a large number of vegetation plots may underrepresent a species range if their spatial distribution is underdispersed. Consequently, both clumping and density of occurrence observations have to be considered, computed and estimated simultaneously to evaluate the representativeness of biodiversity databases.

Possible applications of the DMC

Occurrence data and distribution maps for species of various taxa are increasingly being made available from biodiversity databases (e.g. Map Of Life, Jetz, McPherson, & Guralnick (2012); The IUCN Red List, IUCN (2019); Euro+Med Plantbase, Euro+Med (2019); The PLANTS Database, USDA, NRCS (2019)).

(I) Our DMC approach enables evaluation and comparison of the coverage of occurrence data across irregular and even non-contiguous background spaces. Thus, it helps identifying species with a suitable representation of their range / niche by existing point samples. In species distribution modelling, uneven or inconsistent representation of environmental gradients by occurrence records can strongly influence the model accuracy (Tessarolo, Rangel, Araújo, & Hortal, 2014), which can result in limited applicability for climate change predictions (Araújo & Guisan, 2006; Titeux et al., 2017).

(II) The DMC value calculation is applicable in both the climatic and geographic space and can help evaluate the coverage of species samples for species distribution modelling. Using such information derived from the DMC metric inside the modelling framework of SDM is likely to improve SDM predictive performance. Nevertheless, independent information on species geographic distribution is needed to correctly evaluate point sampling coverage for SDM studies. It is not recommended to generate range models based on sampling data of unknown coverage. While $DMC_{(GEO)}$ values generated this way might be used to gather information on species geographic point sampling quality, $DMC_{(CLIM)}$ values might be highly biased. Without independently generated distribution information, $DMC_{(CLIM)}$ values are not applicable for SDM evaluation. Since observed and expected DMC values (see the applied null model approach) were highly positively correlated, the deviation from the expected DMC is a suitable measure for the representativeness of species occurrence data. A high deviation corresponds to an underdispersed distribution of plots, while a low deviation corresponds to a more random distribution of plots and a negative deviation corresponds to an overdispersed distribution of plots.

(III) Data limitations (i.e. lack of fine-resolution data of species occurrences over large spatial extents) will remain the norm for most species and regions, and best-possible use should be made of limited information (Hoffmann et al., 2014; Meyer et al., 2016). Here, based on the curves resulting from the DMC calculations it would be possible to determine the raster cell resolution where results of the analyses are least vulnerable to errors due to the existing sampling gaps by calculating the inflection point of the DMC curve. Nevertheless, one must be aware that the achievable raster cell resolution always depends on the spatial extent of the study (e.g. regional, continental or global scale) (Hartley & Kunin, 2003; Pearson & Dawson, 2003; Willis & Whittaker, 2002).

(IV) The efficacy of database platforms strongly depends on the completeness of species inventories and the survey coverage across space and the environment (Hortal et al., 2008; Troia & McManamay, 2016), therefore it is necessary to continue surveys in undersampled areas (Beck et al., 2012; Engemann et al., 2015). Here, results of the DMC analyses can be used to identify these undersampled areas and help focus search efforts for data information in relevant literature or further databases. This would be possible by selecting undersampled parts of the niche and translate them back to the geographical space. Furthermore, the results of DMC analyses can be used to guide future botanical explorations and practical fieldwork, to make new sampling in geographical and climate spaces cost-efficient.

(V) Including both the DMC metrics as covariates in any model with species as the observational unit may help to account for potential confounding effects due to the varying sampling coverage of the sampled species distribution within both the climatic and geographic space. Since DMC values are species-specific, they can be included as weights in macroecological analyses and models, where well-represented species might be weighted higher than less-well represented species. Nevertheless, it might be necessary to apply re-sampling methods (e.g. Lengyel, Chytrý, & Tichý, 2011) to prevent spatial autocorrelation in model residuals.

Appendices

Appendix S II.1

Climatic resampling procedure and background PCA niche space of the study area

Multivariate approaches such as PCA or clustering algorithms are sensitive to the frequency distribution of the input variables values, and more average conditions might be lumped in the presence of extreme values. To minimize the spatial autocorrelation between species occurrences in terms of climatic data, it is desirable to sample climatic conditions equally.

Spatial climatic pre-stratification procedure

We developed a climatic pre-stratification to enable spatially unbiased resampling as follows:

1 – We used global layers with monthly mean values of temperature and precipitation at 2.5-min raster cell resolution (hemisphere-adjusted). All precipitation values were log-transformed to take into account the decreasing ecological importance of differences with increasing precipitation. Monthly mean values of temperature and (log) precipitation were separately standardized (0-1).

2 – After standardization (0-1), 10 classes (class width 0.1) per variable (monthly temperature and (log) monthly precipitation) were derived and labelled “A” to “J” (Figure S II.1.1a).

3 – The cells of a unique climate class are defined by an identical string of class labels (= climate class ID) containing 12 “A” to “J” combinations, one for each month.

All 2.5-min raster cells of one climatically homogeneous region are labelled by an identical climate class ID. By this, 29,550 unique climate class IDs were assigned to the raster cells of the analysed species ranges. One climatically homogeneous region was represented by one to many geographical patches of different size (Figure S II.1.1b). The smallest climatically homogeneous region consisted of only one 2.5-min raster cell while the largest climatically homogeneous region consisted of 38,577 2.5-min raster cells.

Based on this spatial pre-stratification, climatic data enclosed by species range polygons and extracted at species’ occurrence locations can be subsampled evenly from 29,550 differently sized, yet climatically homogeneous regions.

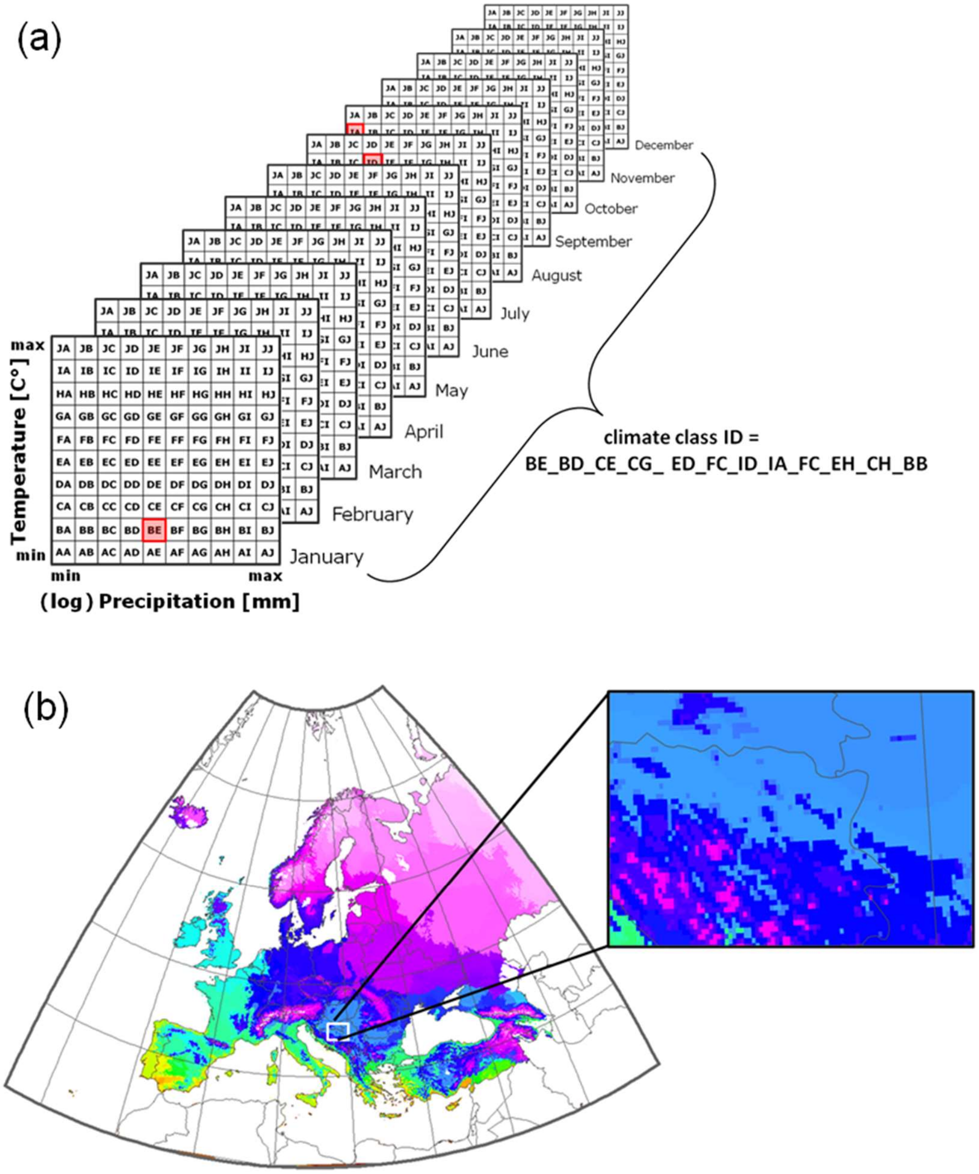

Figure S II.1.1 Illustration of the spatial climatic pre-stratification procedure. (a) Monthly mean values of temperature and (log) precipitation were standardized separately (0-1). Ten classes (class width 0.1) per variable were derived and labelled “A” to “J”. Cells of a unique climate class are labelled by a unique climate class ID. (b) All cells of one climatically homogeneous region are represented by identical colour. Black lines represent the country borders on the continent.

Background PCA niche space and determination of species niche size

To incorporate not only the number of populated climate classes but also between-class climatic similarity, the multivariate niche space of the species is determined by a PCA-approach:

1 – To generate data for a background niche, for each homogeneous climatic region (cells with identical climate class ID from climatic pre-stratification), we determined mean cell values for each of the 19 bioclimatic variables from WorldClim2.0 at 2.5-arc-min resolution (Fick and Hijmans, 2017).

2 – With these 29,550 class mean values, a multi-dimensional climatic space (or climatic niche) was determined by principal components analysis (PCA). The worldwide terrestrial climatic space is well represented by the first two PCA axes which explain 70.75% of the data variance (Figure S II.1.2). Pearson correlations of the 19 bioclimatic variables (BIO 01 – BIO 19) with the first two axes of the principal component analysis (PC1 and PC2) are given in Table S II.1.3. Accordingly, the PCA space is spread out across 29,550 data points, each representing a unique climate class.

3 – Since the class-point density is uneven, the whole PCA area was gridded into a 100 x 100 PCA-cell raster to enable a generalized, more robust species comparison.

4 – The multivariate niche space of a single species is finally determined as the occupied PCA area by counting all PCA-cells matched by a minimum convex hull of the occupied PCA space locations.

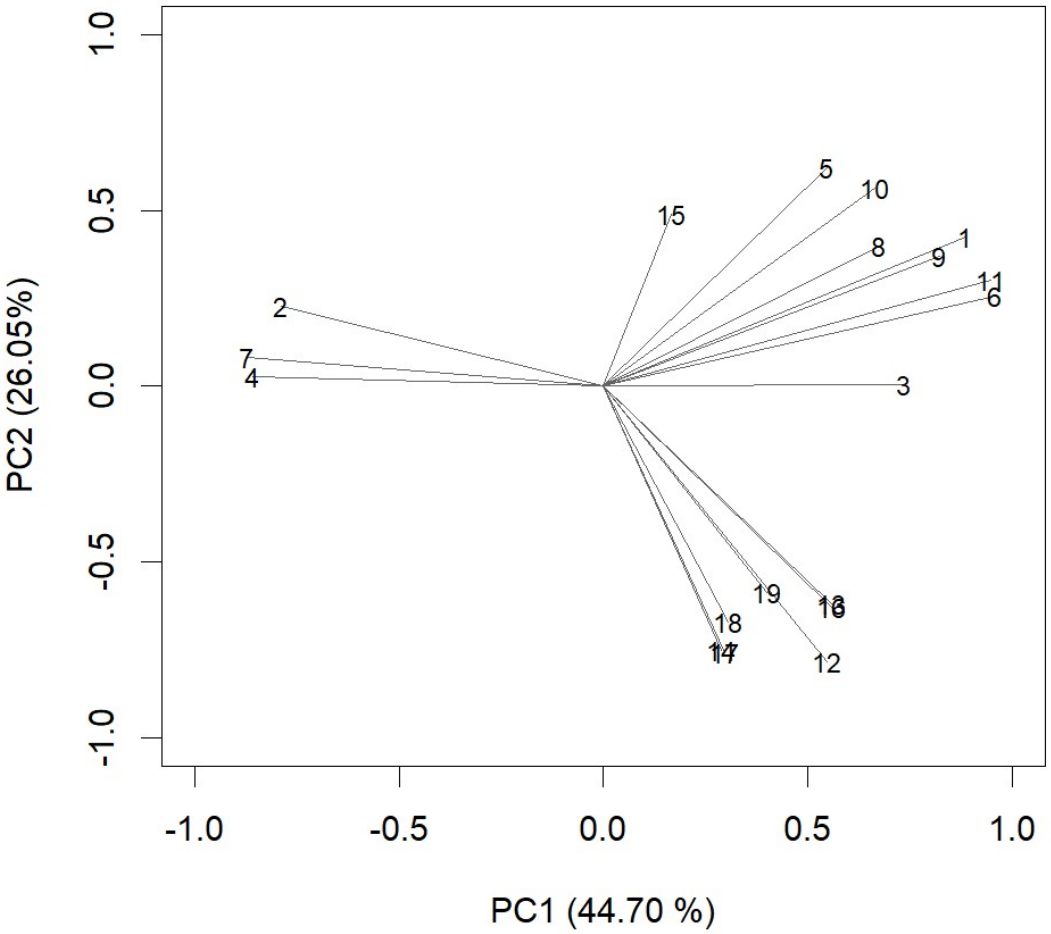

Figure S II.1.2 Principal component analysis (PCA) for bioclimatic variables in the global terrestrial regions excluding Antarctica. The two principal components (PC1 and PC2) explained 70.75% of the total variation in bioclimatic data. PC1 was positively related to temperature and negatively to temperature seasonality, while PC2 was negatively related to precipitation and positively to precipitation seasonality.

Table S II.1.3 Results of Pearson correlation between the 19 bioclimatic variables (BIO 01 – BIO 19) and the first two axes of the principal component analysis (PC1 and PC2).

Bioclim variable	**PC1**	**PC2**
BIO1 = Annual Mean Temperature	0.885	0.425
BIO2 = Mean Diurnal Range (Mean of monthly (max - min temp))	-0.789	0.227
BIO3 = Isothermality (BIO2/BIO7) (* 100)	0.735	0.005
BIO4 = Temperature Seasonality (standard deviation *100)	-0.857	0.028
BIO5 = Max Temperature of Warmest Month	0.545	0.621
BIO6 = Min Temperature of Coldest Month	0.958	0.257
BIO7 = Temperature Annual Range (BIO5-BIO6)	-0.873	0.081
BIO8 = Mean Temperature of Wettest Quarter	0.673	0.397
BIO9 = Mean Temperature of Driest Quarter	0.820	0.371
BIO10 = Mean Temperature of Warmest Quarter	0.664	0.562
BIO11 = Mean Temperature of Coldest Quarter	0.946	0.301
BIO12 = Annual Precipitation	0.548	-0.783
BIO13 = Precipitation of Wettest Month	0.558	-0.617
BIO14 = Precipitation of Driest Month	0.289	-0.751
BIO15 = Precipitation Seasonality (Coefficient of Variation)	0.165	0.489
BIO16 = Precipitation of Wettest Quarter	0.557	-0.628
BIO17 = Precipitation of Driest Quarter	0.296	-0.757
BIO18 = Precipitation of Warmest Quarter	0.304	-0.669
BIO19 = Precipitation of Coldest Quarter	0.400	-0.585

Appendix S II.2 Information on the 564 species included in this study. This Appendix can be found online, https://doi.org/10.1111/jvs.12763

Appendix S II.3 Information on the 59 databases that provided vegetation plots analysed in this thesis.

Official database name, database code in the Global Index of Vegetation-Plot Database (GIVD), name of the database custodian, total number of vegetation-plot samples included in the dataset [data access: October 2015], number of vegetation plot samples included in this study, proportional contribution of the datasets plot samples to this study [proportion = (Count Dataset Sporbert et al. *100) / 808794].

Database name	GIVD Code	Custodian name	Count Dataset total (10/2015)	Count Dataset Sporbert et al.	Proportion [%]
Vegetation Database of Eurasian Tundra	00–00–004	Risto Virtanen	1 132	294	0.036
Vegetation Database Forest of Southern Ural	00-RU-001	Pavel Shirokikh	997	222	0.027
Database Meadows and Steppes of Southern Ural + Database of South Ural Order Galietalia veri + Database of South Ural Order Arrhenatheretalia	00-RU-003 + 00-RU-004 + 00-RU-005	Sergey Yamalov	2 034	1 093	0.135
Forest Vegetation Database of Turkey – FVDT	00-TR-001	Ali Kavgacı	144	127	0.016
Vegetation Database of the Grassland Communities in Anatolia	AS-TR-001	Deniz Işık Gürsoy	20	6	0.001
Vegetation Database of Oak Communities in Turkey	AS-TR-002	Emin Uğurlu	68	61	0.008
Nordic-Baltic Grassland Vegetation Database	EU-00–002	Jürgen Dengler	6 062	6 056	0.749
Iberian and Macaronesian Vegetation Information System (SIVIM)	EU-00–004	Xavier Font	3 496	3 091	0.382
Iberian and Macaronesian Vegetation Information System (SIVIM) - Catalonia	EU-00–004	Xavier Font	3 875	3 512	0.434
Iberian and Macaronesian Vegetation Information System (SIVIM) – Grasslands	EU-00–004	Maria Pilar Rodríguez-Rojo	7 331	7 199	0.89
Iberian and Macaronesian Vegetation Information System (SIVIM) – Sclerophyllous forests	EU-00–004	Federico Fernández-González	3 799	3 170	0.392
Iberian and Macaronesian Vegetation Information System (SIVIM) - Shrublands	EU-00–004	Xavier Font	3 007	2 386	0.295
Vegetation-Plot Database of the University of the Basque Country (BIOVEG)	EU-00–011	Idoia Biurrun	18 429	16 405	2.028

Balkan Dry Grasslands Database	EU-00–013	Kiril Vassilev	8 152	4 769	0.59
Mediterranean Ammophiletea database	EU-00–016	Corrado Marcenò	6 835	4 843	0.599
European Coastal Vegetation Database	EU-00–017	John Janssen	4 311	2 251	0.278
The Nordic Vegetation Database	EU-00–018	Jonathan Lenoir	7 718	7 144	0.883
Balkan Vegetation Database	EU-00–019	Kiril Vassilev	9 579	7 092	0.877
WetVegEurope	EU-00–020	Flavia Landucci	1 994	6	0.001
SE Europe Forest Database	EU-00–021	Andraž Čarni	3 659	3 656	0.452
European Mire Vegetation Database	EU-00–022	Tomáš Peterka	10 099	9 047	1.119
Iberian and Macaronesian Vegetation Information System (SIVIM) - Deciduous Forests	EU-00–023	Juan Antonio Campos	6 630	6 286	0.777
Vegetation Database of Albania	EU-AL-001	Michele De Sanctis	290	193	0.024
Austrian Vegetation Database	EU-AT-001	Wolfgang Willner	30 659	23 941	2.960
INBOVEG	EU-BE-002	Els De Bie	13 541	8 204	1.014
Bulgarian Vegetation Database	EU-BG-001	Iva Apostolova	5 235	1 935	0.239
Swiss Forest Vegetation Database	EU-CH-005	Thomas Wohlgemuth	14 193	14 182	1.753
Czech National Phytosociological Database	EU-CZ-001	Milan Chytrý	110 534	97 650	12.074
VegMV	EU-DE-001	Florian Jansen	49 631	44 410	5.491
VegetWeb	EU-DE-013	Jörg Ewald	22 363	21 525	2.661
GVRD Vegetation Reference Database Halle	EU-DE-014	Ute Jandt	29 797	28 418	3.514
NATURDATA.DK	EU-DK-002	Jesper Erenskjold Moeslund	24 264	23 994	2.967
Iberian and Macaronesian Vegetation Information System (SIVIM) - Wetlands	EU-ES–001	Aaron Pérez-Haase	6 539	4 507	0.557
SOPHY	EU-FR-003	Henry Brisse	155 275	143 323	17.721

UK National Vegetation Classification Database	EU-GB-001	John S. Rodwell	25 485	24 104	2.980
Hellenic Natura 2000 Database (HelNatVeg)	EU-GR-005	Panayotis Dimopoulos	4 857	4 295	0.531
Hellenic Woodland Database + Hellenic Beech Forests Database (Hell-Beech-DB)	EU-GR-006 + EU-GR-007	Ioannis Tsiripidis	3 199	636	0.079
Croatian Vegetation Database	EU-HR-002	Željko Škvorc	8 517	8 249	1.020
CoenoDat Hungarian Phytosociological Database	EU-HU-003	János Csiky	5 104	812	0.1
Irish Vegetation Database	EU-IE-001	Úna FitzPatrick	26 687	25 010	3.092
VegItaly	EU-IT-001	Roberto Venanzoni	15 332	8 957	1.107
Vegetation database of Habitats in the Italian Alps (HabItAlp)	EU-IT-010	Laura Casella	3 562	3 496	0.432
Georeferenced Vegetation Database - Sapienza University of Roma	EU-IT-011	Emiliano Agrillo	12 665	10 981	1.358
Lithuanian Vegetation Database	EU-LT-001	Valerijus Rašomavičius	2 206	1 842	0.228
Semi-natural Grassland Vegetation Database of Latvia	EU-LV-001	Solvita Rūsiņa	5 594	5 581	0.69
Vegetation Database of the Republic of Macedonia	EU-MK-001	Renata Ćušterevska	1 269	370	0.046
Dutch National Vegetation Database	EU-NL-001	Joop H.J. Schaminée	93 812	83 968	10.382
Polish Vegetation Database	EU-PL-001	Zygmunt Kącki	56 989	53 381	6.600
Romanian Forest Database	EU-RO-007	Adrian Indreica	6 017	6 006	0.743
Romanian Grassland Database	EU-RO-008	Eszter Ruprecht	4 962	4 718	0.583
Vegetation Database Grassland Vegetation of Serbia	EU-RS-002	Svetlana Aćić	5 587	5 364	0.663
Database of Forest Vegetation in Republic of Serbia + Vegetation Database of Northern Part of Serbia (AP Vojvodina)	EU-RS-003 + EU-RS-004	Mirjana Krstivojević Ćuk	1 131	1 131	0.14
Lower Volga Valley Phytosociological Database	EU-RU-002	Valentin Golub	11 846	5 320	0.658
Vegetation Database of Tatarstan	EU-RU-011	Vadim Prokhorov	7 426	2 301	0.284

Vegetation Database of Slovenia	EU-SI-001	Urban Šilc	10 986	10 750	1.329
Slovak Vegetation Database	EU-SK-001	Milan Valachovič	36 266	33 320	4.120
Ukrainian Grassland Database	EU-UA-001	Anna Kuzemko	4 043	3 954	0.489
Halophytic and Coastal Vegetation Database of Ukraine	EU-UA-005	Tetiana Dziuba	4 399	13	0.002
Vegetation Database of Ukraine and Adjacent Parts of Russia	EU-UA-006	Viktor Onyshchenko	3 325	3 192	0.395

Appendix S II.4 Distribution of initial grain size (in km²) in DMC calculations (Figure S II.4.1). Scatterplots and Spearman correlation coefficients (rho) of the relationship between species range sizes and niche sizes (Figure S II.4.2). Scatterplots and Spearman correlation coefficients (rho) of the relationship between percentage match of species ranges by EVA vegetation plots and percentage match of species niches by EVA vegetation plots (Figure S II.4.3).

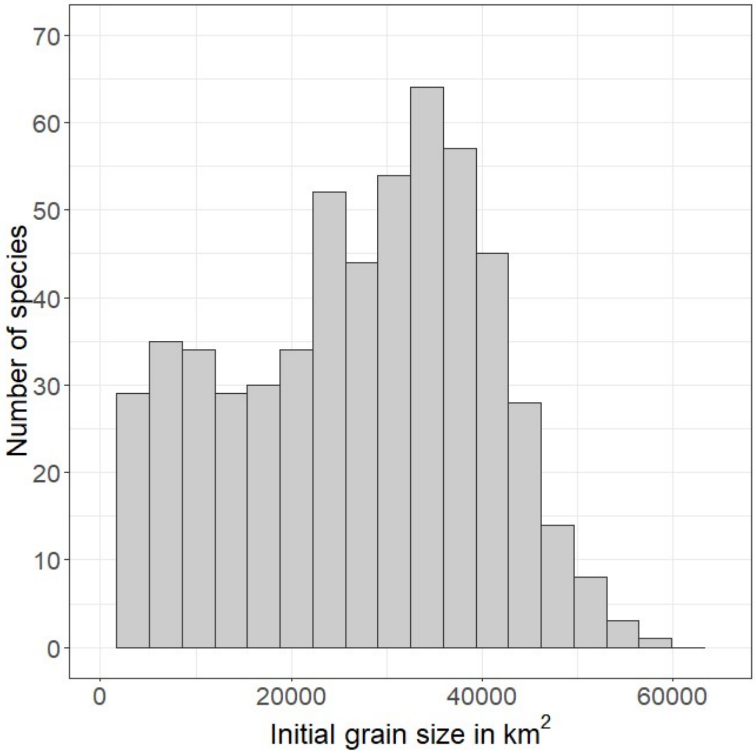

Figure S II.4.1 Distribution of initial grain size (in km²) in DMC calculations. Bandwidth of scaling steps were calculated species specific according to the species range sizes at 2.5-min raster cell resolution in geographic space.

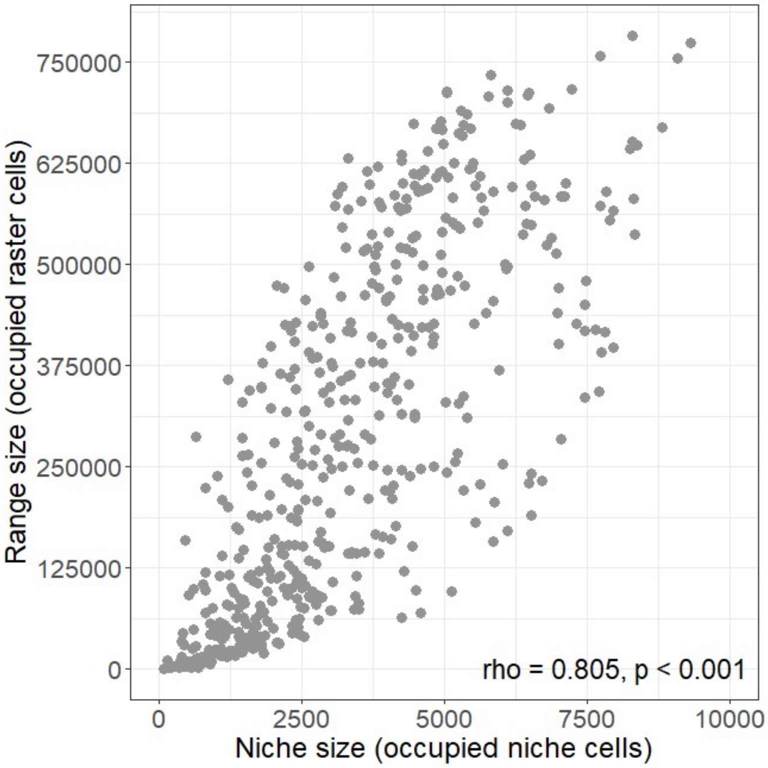

Figure S II.4.2 Scatterplots and Spearman correlation coefficients (rho) of the relationship between species range sizes (occupied raster cells at 2.5-min raster cell resolution) in geographic space and niche sizes (occupied niche cells) in climatic space.

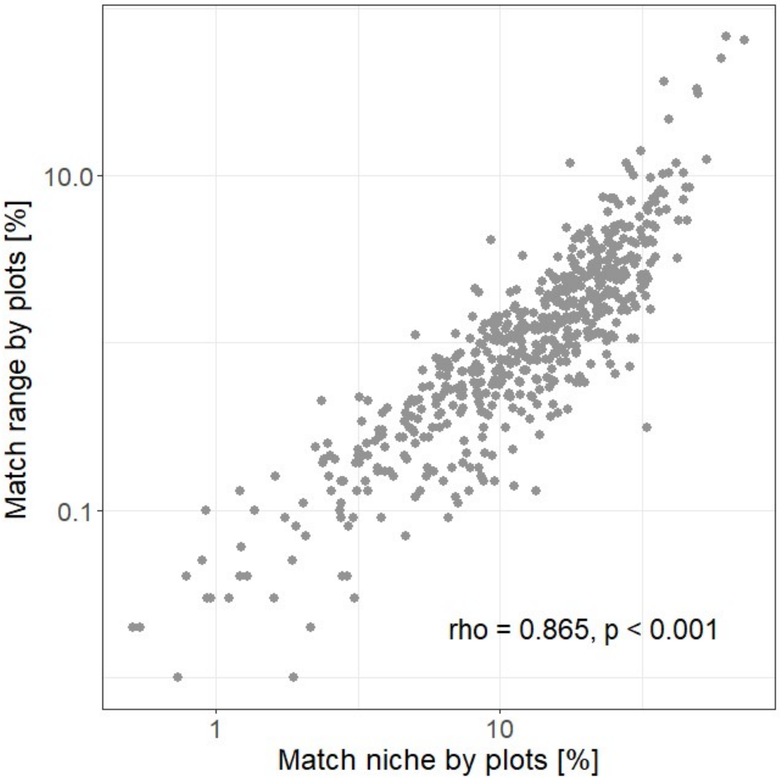

Figure S II.4.3 Scatterplots and Spearman correlation coefficients (rho) of the relationship between percentage match of species ranges by EVA vegetation plots and percentage match of species niches by EVA vegetation plots. X axis and y axis are log-transformed.

Chapter III

Testing macroecological abundance patterns: the relationship between local abundance and range size, range position and climatic suitability among European vascular plants

This chapter is published in *Journal of Biogeography* as

Sporbert, M., Keil, P., Seidler, G., Bruelheide, H., Jandt, U., Aćić, S., Biurrun, I., Campos, J.A., Čarni, A., Chytrý, M., Ćušterevska, R., Dengler, J., Golub, V., Jansen, F., Kuzemko, A., Lenoir, J., Marcenò, C., Moeslund, J.E., Pérez-Haase, A., Rūsiņa, S., Šilc, U., Tsiripidris, I., Vandvik, V., Vasilev, K., Virtanen, R., & Welk, E. (2020). Testing macroecological abundance patterns: the relationship between local abundance and range size, range position and climatic suitability among European vascular plants. *Journal of Biogeography,* 47(10), 2210–2222

Abstract

Aim: A fundamental question in macroecology centres around understanding the relationship between species' local abundance and their distribution in geographic and climatic space (i.e. the multi-dimensional climatic space or climatic niche). Here, we tested three macroecological hypotheses that link local abundance to the following range properties: (1) the abundance-range size relationship, (2) the abundance-range centre relationship, and (3) the abundance-suitability relationship.

Location: Europe

Taxon: Vascular plants

Methods: Distribution range maps were extracted from the Chorological Database Halle to derive information on the range and niche sizes of 517 European vascular plant species. To estimate local abundance, we assessed samples from 744,513 vegetation plots in the European Vegetation Archive, where local species' abundance is available as plant cover per plot. We then calculated the 'centrality', i.e. the distance between the location of the abundance observation and each species' range centre in geographic and climatic space. The climatic suitability of plot locations was estimated using coarse-grain species distribution models (SDMs). The relationships between centrality or climatic suitability with abundance were tested using linear models and quantile regression. We summarized the overall trend across species' regression slopes from linear models and quantile regression using a meta-analytical approach.

Results: We did not detect any positive relationships between a species' mean local abundance and the size of its geographic range or climatic niche. Contrasting yet significant correlations were detected between abundance and centrality or climatic suitability among species.

Main conclusions: Our results do not provide unequivocal support for any of the relationships tested, demonstrating that determining properties of species' distributions at large grains and extents might be of limited use for predicting local abundance, including current SDM approaches. We conclude that environmental factors influencing individual performance and local abundance are likely to differ from those factors driving plant species' distribution at coarse resolution and broad geographic extents.

Keywords: abundance, climatic suitability, commonness and rarity, range size, realized climatic niche, resolution, species distribution models, vegetation-plot data

Introduction

A fundamental question in macroecology centres around what drives spatial variation in species' abundance, with species' abundance across the geographic range having recently been proposed as one of the so-called essential biodiversity variables (Jetz et al., 2019). Much effort has gone into identifying the drivers of local abundance, and there is a substantial body of literature establishing that it depends on local factors such as environmental suitability and local biotic interactions (e.g. Andrewartha & Birch, 1954; MacArthur, Diamond, & Karr, 1972; Peterson et al., 2011; Staniczenko, Sivasubramaniam, Suttle, & Pearson, 2017). Similarly, it has been established that the limit of species' distribution is a function of the interplay between environmental conditions and barriers affecting dispersal and evolution (Baselga, Lobo, Svenning, & Araujo, 2012). Based on metapopulation and metacommunity concepts, local and regional processes are hypothesised to be mechanistically linked (Leibold et al., 2004), and relationships between species' local abundance and geographical distribution may be expected. However, supporting empirical data, especially for herbaceous plants, is lacking. In this paper, we explore whether local abundance is associated with geographic distribution in 517 European vascular plant species at broad geographic extents. Specifically, we focus on three macroecological relationships: (1) the abundance-range size relationship, (2) the abundance-range centre relationship, and (3) the abundance-suitability relationship (Table III.1). To test the relationships, we interrogated existing data on the species' geographic range characteristics and local abundances, deriving range information from the Chorological Database Halle (CDH; E. Welk et al., unpublished data) and local abundance data from the European Vegetation Archive (EVA; Chytrý et al., 2016). The EVA data are recorded as percentage cover per plot, which we extrapolated to determine local abundance following Preston (1948).

For the first relationship, we checked for associations between species' local abundance and the size of the respective geographic range or climatic niche and ask whether locally abundant species are generally more widely distributed, i.e. whether they have larger geographic ranges than species with relatively lower local abundance values (Table III.1). Such a relationship has been widely and empirically documented (Hanski, 1982; Brown, 1984; Gaston et al., 1997; Reif et al. 2006; Gaston & Blackburn, 2008), and several mechanisms have been proposed as drivers (Gaston et al., 1997). For instance, when a species can tolerate a large variability of climatic conditions across its range, it should be able to cope with local temporal climatic variability and thus perform better than local competitors with narrower climatic niches and limited plasticity, resulting in it showing higher local abundance. This mechanism relies on the assertion that a species' geographical distribution is driven by the same climatic tolerances that influence its local abundance, which may not be the case (Guisan & Thuiller, 2005; Mertes & Jetz, 2018). Gaston, Blackburn, & Lawton (1997) reviewed eight other mechanisms that could lead to this relationship, two of which were artefactual, while the others considered niche breadth, resource availability, habitat selection, dispersal limitation, metapopulation dynamics or position within the distributional range. In conclusion, the authors noted that none of the proposed mechanisms has received unequivocal support. As such, the abundance-range size relationship has seen mixed empirical support (Gaston et al., 1997; Köckemann, Buschmann, & Leuschner, 2009), and there have been concerns that species' abundance and range-size were mostly analysed at completely different, or inappropriate, scales (Thompson, Hodgson & Gaston, 1998; Conlisk, Conlisk, Kassim, Billick, & Harte, 2012; Kambach et al., 2019). For our study, on a large group of herbaceous plant and shrub species predominantly over their full distributional ranges, we expected locally less abundant species to be more narrowly distributed, and locally abundant species to be more widespread in geographic and climatic space (Figure III.1a).

The second intraspecific abundance-range centre relationship tested here links local abundance values to the distance of the locality from the centre of the species' geographic range or climatic niche (Table III.1). The hypothesis stems from the assumption that the environment tends to be more suitable at the centre of the range than near its edges (Grinnell, 1922; Hengeveld & Haeck, 1982; Brown, 1984). However, it has become clear that species' geographic ranges can be shaped by barriers such as mountains or coastlines (Hargreaves, Samis, & Eckert, 2013), and ecological conditions do not systematically follow geographic gradients (Pironon et al., 2017). Thus, the geographic distribution of suitable environment can be unrelated to the geometry of a species' range (Manthey et al., 2015). Empirical support for the positive form of this relationship in geographic space is weak (e.g. Gaston et al., 1997; Sagarin, Gaines, & Gaylord, 2006; Dallas, Decker, & Hastings, 2017; Pironon et al., 2017). While supported in theory, empirical support for the hypothesis in climatic space is mixed, with some studies finding stronger support for this relationship in climatic than in geographic space (Martínez-Meyer, Díaz-Porras, Peterson, & Yáñez-Arenas, 2013; Van Couwenberghe, Collet, Pierrat, Verheyen, & Gégout, 2013; Osorio-Olvera, Yáñez-Arenas, Martínez-Meyer, & Peterson, 2020), while others finding weak or no support (Abeli, Gentili, Mondoni, Orsenigo, & Rossi, 2014; Pironon, Villellas, Morris, Doak, & García, 2015; Dallas, Pironon, & Santini, 2020). For our study, we therefore expected the relationship between local abundance and distance from the range centre to be weak or absent in geographic space, while to be positive in climatic space (Figure III.1b).

The third abundance-suitability relationship assumes that the more suitable the climate for a species within a larger area, the higher the abundance of the species should be within local plots in that area (Table III.1). By equating occurrence probability with climatic suitability, it is expected that favourable climatic conditions support higher species' abundance as a result of improved population and/or individual performance (Araújo, Williams, & Fuller, 2002). While some previous studies have found strong support for this relationship (VanDerWal, Shoo, Johnson, & Williams, 2009; Weber et al., 2017), others failed to detect any effects (Gomes et al., 2018; Santini et al., 2019). As statistically significant relationships have only been recorded for a few species, a low generality of climatic suitability as a predictor of spatial patterns of abundance may be assumed (VanDerWal et al., 2009). In addition, species' responses to differing climatic conditions can be non-Gaussian (i.e. skewed, bi-modal, or truncated) (Austin, 1987). Therefore, species' climatic niche may not show highest climatic suitability in the very centre, but somewhere closer to the edges. For this study, we expected local abundance to be positively related to the coarse-grain climatic suitability predicted for the 15 km^2 grid cells containing the respective vegetation plots (Figure III.1c).

It is noted that there is a suite of mechanisms that can weaken or limit the above-described relationships by affecting species' local abundance but without affecting their overall geographic distribution. Examples include: environmental and demographic stochasticity (Lande, Engen, & Saether, 2003), particularly when they are temporally synchronous over large geographic extents; biotic interactions (Dallas et al., 2017); soil and disturbance parameters (VanDerWal et al., 2009), or isolation by dispersal barriers

(Reif et al., 2006). Since correlative models of climatic suitability do not integrate these factors, species might be absent or show low abundance at sites with predicted high climatic suitability (VanDerWal et al., 2009). Consequently, relationships could be absent or triangular (Figure III.1b and c), which are common forms of relationship between macroecological variables (Brown & Maurer, 1987), indicating more limiting factors at play than just a linear response of central tendency (as in Figure III.1a).

Table III.1 Description and references to the three tested relationships that link species' local abundance to distribution in coarse-grain geographic or climatic space.

Relationship	Description	References
(Interspecific) abundance-range size relationship	Locally rare species tend to be narrowly distributed, while locally abundant species tend to be more widespread in geographic or climatic space.	Brown, 1984; Gaston & Blackburn, 2008; Gaston et al., 1997; Hanski, 1982; Reif et al., 2006; Thompson et al., 1998
(Intraspecific) abundance-range centre relationship	Sites with low abundance are located towards range or niche margins, while sites with high abundance are clustered at the centre of a species' range or niche.	Brown, 1984; Hengeveld & Haeck, 1982; Osorio-Olvera et al., 2020; Pironon et al., 2017; Sagarin & Gaines, 2002; Sagarin et al*.,* 2006; Santini et al., 2019
(Intraspecific) abundance-suitability relationship	Species show lower local abundance in climatically less suitable areas but achieve higher local abundance in climatically more suitable areas.	Dallas & Hastings, 2018; Gomes et al., 2018; Santini et al., 2019; VanDerWal et al., 2009; Weber et al., 2017

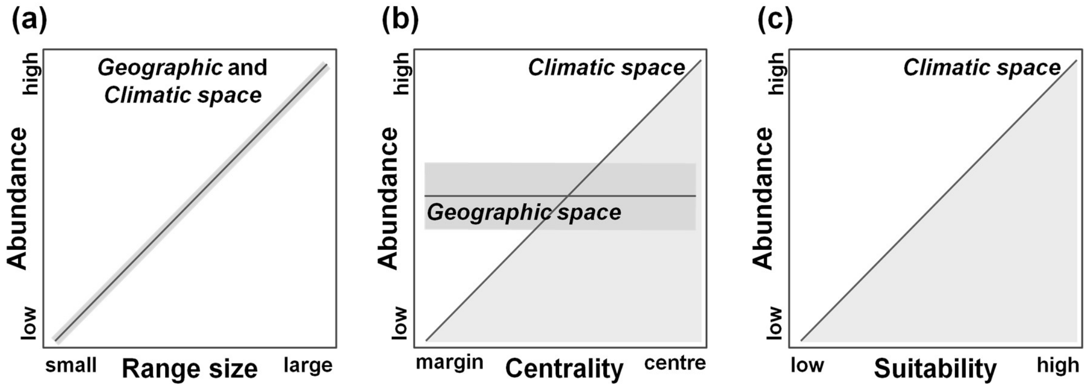

Figure III.1 Hypothesized relationships between local abundance and: (a) the size of the geographic range or climatic niche derived at the coarse grain; (b) the vegetation plot position (i.e. distance to centre) within the species' range in either geographic or climatic space; and (c) the climatic suitability of a grid cell within which a local plot is situated in the species' range in climatic space.

Material and Methods

Geographic ranges

For this study, we used existing data on the geographic ranges of European vascular plant species from the CDH (E. Welk et al., unpublished data). The study area comprised all geographical European countries as well as Turkey, Georgia, Armenia and Azerbaijan (see Figure II.1 and II.2 in chapter II). We only included species for which digitized GIS-data were available (i.e. range polygons and point occurrences). We excluded trees, bryophytes, lichens, fungi and algae from the vegetation-plot records to obtain a more homogeneous dataset of herbaceous species, dwarf shrubs and shrubs. Data on a total of 517 species were consequently amassed for the study, which represents approximately 10% of all Central European vascular plant species (Meusel & Jäger, 1992). Species' range information was processed to coarse-grain raster layers of 2.5-min resolution, which corresponded to grid cells covering approximately 15 km² each across Central Europe (see e.g. the range of *Inula conyzae* in Figure III.2a). The measure of range size for each species then corresponded to the number of grid cells it occupied (Area of Occupancy (AOO)) (IUCN, 2019).

Climatic niches

The multi-dimensional climatic space (or climatic niche) of each geographic range was determined using principal components analysis (PCA) of 19 bioclimatic variables from WorldClim (Hijmans, Cameron, Parra, Jones, & Jarvis, 2005) at 2.5-min cell resolution (Figure III.2b; climatic niche of *Inula conyzae*). The common European climatic space is represented by the first two PCA axes, which explain 88.0% of the data variance. Unique PCA space locations were considered and counted as grid cells in climatic space. Species' niche size was calculated as the number of occupied grid cells in climatic space (see Appendix S II.1 in chapter II).

Local abundance in vegetation plots

Local abundance values for a total of 744,513 vegetation plots were obtained from EVA (Chytrý et al., 2016) for the 517 study species in October 2015. For half of the plots, the recorded area was between 9 and 100 m² (25th and 75th percentile) with a median of 25 m². Data for subspecies were merged at the species level, and we matched synonymous species names according to: (i) a taxonomic reference list for Germany (GermanSL version 1.2, Jansen & Dengler, 2008); and (ii) for all taxonomic reference lists available via the R package 'taxize' (Chamberlain & Szöcs, 2013; R Core Team, 2018). We only included species that occurred in at least 100 vegetation plots in the EVA dataset, and plots with geographic location uncertainty of < 10 km. The number of vegetation plots for each species ranged from 101 plots for *Malva pusilla* to 23,464 plots for *Plantago lanceolata*. For half of the studied species, the number of plots ranged between 631 and 4,531 plots (25th and 75th percentile), with a median of 1,863 plots (see Appendix S III.1.1).

Cover-abundance values compiled in EVA that were based on different scales (e.g. Domin, 1928; Braun-Blanquet, 1951) were transformed to a common percentage scale (van der Maarel, 1979).

When more than one plot per species was present in a 2.5-min raster cell, we calculated mean values of abundance (%) to reduce effects of spatial autocorrelation and pseudo-replication. The percentage of grid cells with less than four plots ranged from 39.1% to 97.1% per species. For half of the studied species, the percentage of grid cells with less than four plots was between 66.7% and 77.4% (25th and 75th percentile) with a median of 71.5%. The percentage of grid cells with more than 10 plots ranged from 0.0% to 28.6%. For half of the studied species, the percentage of grid cells with more than 10 plots was between 2.6% and 6.1% (25th and 75th percentile), with a median of 4.3% (see Appendix S III.1.2). Information on source databases that provided vegetation plot data can be found in Appendix S II.3 in chapter II.

Distance from centre of the geographic range or climatic niche

To determine the centroids of each species' geographic range and climatic niche, all grid cells in which a species was indicated as present in the CDH database were considered. Geographic range centroids were calculated as the arithmetic mean of spatial central coordinates of grid cells over the species' CDH geographic range. To determine species' niche centroids, the multivariate climatic space was translated into two-dimensional space (using PCA), and species' geographic occurrences were projected into this climatic niche space. Niche centroids were determined as the arithmetic mean of PCA-coordinates of the respective species' raster cell values. Geographic distance (in kilometres) from each respective EVA vegetation plot to the respective species' CDH range centre was determined using Haversine great circle geographic distance (Figure III.2c and d). We calculated Mahalanobis distance to the climatic niche centroid as a measure in climatic space (Figure III.2e and f). Mahalanobis distance is considered as a good proxy for marginality since it takes into account the covariance structure of the data (Osorio-Olvera, Soberón, & Falconi, 2019; Osorio-Olvera et al., 2020). For each species' vegetation plot position, the distance to range or niche centroid was divided by the species-specific maximum distance to the range or niche centroid (distance/$\text{distance}_{\text{max}}$). This standardization enabled and simplified comparison among the species in our study.

Coarse-grain climatic suitability

We used species distribution modelling (SDM) to obtain spatial estimates of climatic suitability within each species' geographic range. SDMs estimate occurrence probabilities based on the relationship between species occurrence and environmental (climatic) characteristics. We used occurrence data from CDH and bioclimatic variables provided by the CHELSA project (Karger et al., 2017) at 2.5 min-resolution as explanatory variables to build SDMs. The distribution range data of CDH are point-polygon maps covering the complete distribution of the recorded species. Apart from isolated or fragmented single occurrences (points), the areas outside the range polygons are proven to be ‘absence areas’ by accumulated regional expert knowledge, as documented in national to regional floristic atlases, floras or floristic inventories. Thus, pseudo-absences were sampled from bordering regions in geographical and climatic space. The general sampling approach is based on a climatic pre-stratification of the species’ geographic range into differing climatic regions. A fixed number of presence and pseudo-absence samples were drawn randomly from each climatic region, irrespective of the size of the respective climatic region. As recommended by Barbet-Massin, Jiguet, Albert, & Thuiller (2012), we kept the number of selected pseudo-absence samples equal to the number of presence samples for all models respectively (see Appendix S III.2.1 and 2.2).

SDMs estimate spatial predictions of environmental suitability from 0 (not suitable) to 1 (most suitable) (Figure III2g and h). We applied four different suitability modelling methods from the three main groups of modelling approaches (i.e. machine learning methods, statistical modelling and similarity methods). The methods we applied are ‘bioclim’ (similarity method), ‘multivariate adaptive regression splines’ (mars) (statistical modelling), ‘random forest’ (rf) and ‘support vector machine’ (svm) (machine learning methods). We used the area under the receiver operating characteristic (ROC) curve (AUC; Bradley, 1997) to calculate model accuracy, within which high AUC values (i.e. those closer to 1) indicate a strong capacity for model discrimination (Bedia, Herrera, & Gutiérrez, 2013). Species distribution modelling was performed using the R package 'sdm' (version 1.0-67) (Naimi & Araújo, 2016). We fitted and evaluated the four models using 10 runs of subsampling replications withholding 30 percent samples as test data.

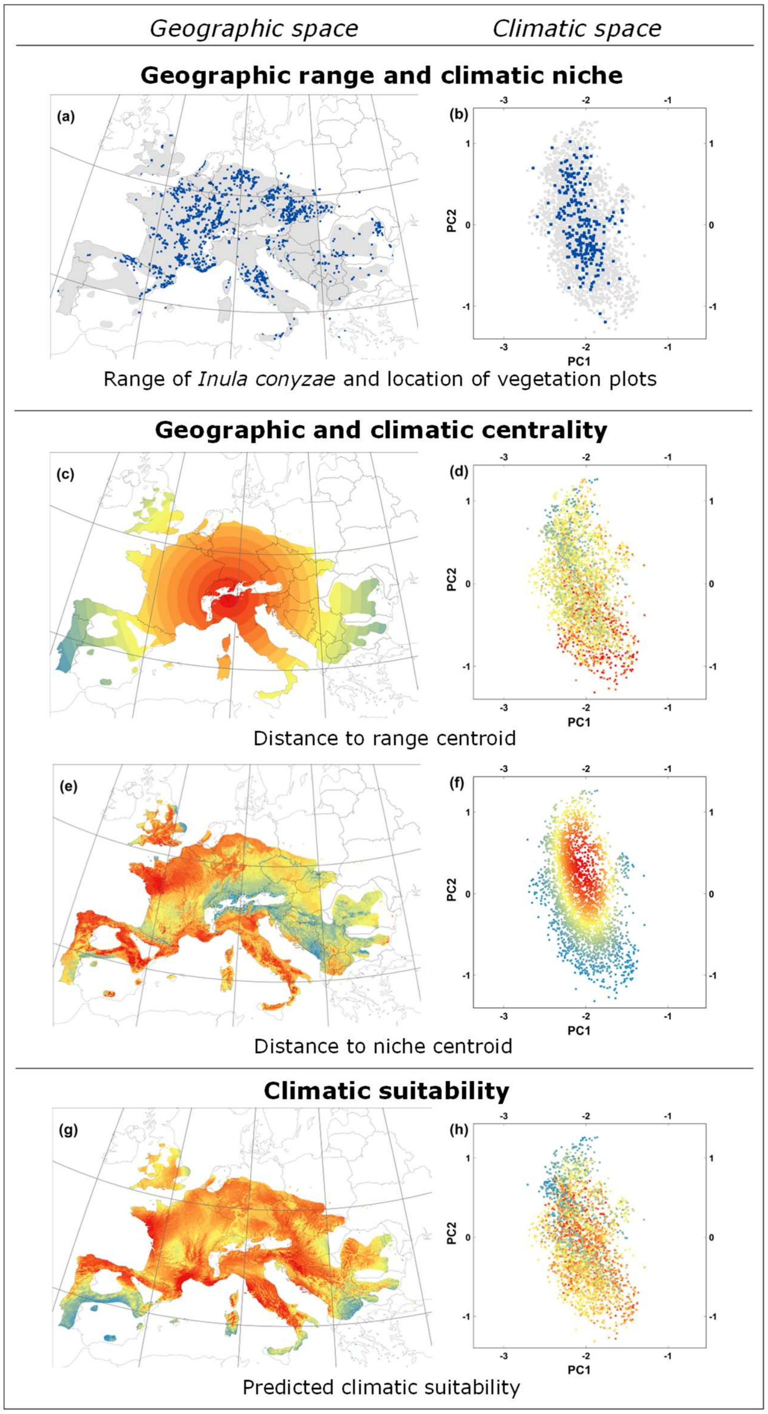

Figure III.2 Range of *Inula conyzae* (grey) and locations of vegetation plots (blue dots) from EVA in (a) geographic and (b) climatic space. Centrality and predicted climatic suitability for *Inula conyzae* illustrated in blue (low centrality/suitability) to red (high centrality/suitability). The distance to range centroid in (c) geographic and (d) climatic space refers to the abundance-range centre relationship. The distance to niche centroid in (e) geographic and (f) climatic space refers to the abundance-range centre relationship. The predicted climatic suitability from the model ‘random forest’ (rf) in (g) geographic and (h) climatic space refers to the abundance-suitability relationship.

Abundance vs range size, centrality and coarse-grain climatic suitability

We applied ‘ordinary least squares’ (OLS) linear regression models to examine the relationship of local abundance with range size, centrality and coarse-grain climatic suitability. We used ‘linear quantile regression’ to examine the relationship between centrality and coarse-grain climatic suitability to upper limits (90th quantile) of abundance values to test whether the relationships would be better described as “triangular”, as illustrated in Figures III.1b and c. To estimate the overall trend across species, a meta-analysis was performed on species’ slopes for each of the applied centrality and climatic suitability measures (Borenstein, 2009). We conducted a mixed-effect meta-analysis on the slopes and the associated variance of each of the centrality and suitability measures to calculate a summary effect size including species as a random factor. Quantile regressions were performed using the R package 'quantreg', (version 5.38) (Koenker, 2018) and the mixed-effect meta-analysis using the function 'rma' from package 'metafor' (Viechtbauer, 2010). All the analyses were rerun on a subset of the dataset, for which only grid cells with a minimum number of four occurrences of each species were included.

Results

Abundance vs range size and niche size

Species’ range size (no. of occupied grid cells) ranged from 1,202 in *Juniperus sabina* to 782,025 in *Stellaria media*. For half of the species, range size was between 74,867 and 476,865 (25^{th} and 75^{th} percentile), with a median of 254,579 grid cells. Species’ niche size (no. of occupied grid cells) ranged from 162 in *Scabiosa canescens* to 9,318 in *Plantago major*. For half of the species, the niche size was between 1,657 and 4,614 (25^{th} and 75^{th} percentile), with a median of 3,002 grid cells (see Supporting Information S III.3, available online https://doi.org/10.1111/jbi.13926). There was a strong significant positive relationship between species range and niche size ($R^2 = 0.616$, *p*-value < 0.001, see Appendix S III.4.1). Species’ local abundance (mean plot cover) was significantly, yet weakly, related to range size ($R^2 = 0.011$, *p*-value = 0.020) but not to niche size (R^2 = 0.001, *p*-value = 0.398) (Figure III.3a and b). For both relationships we found high intraspecific variation in species' abundance values, as indicated by the width of the vertical error bars in Figure III.3.

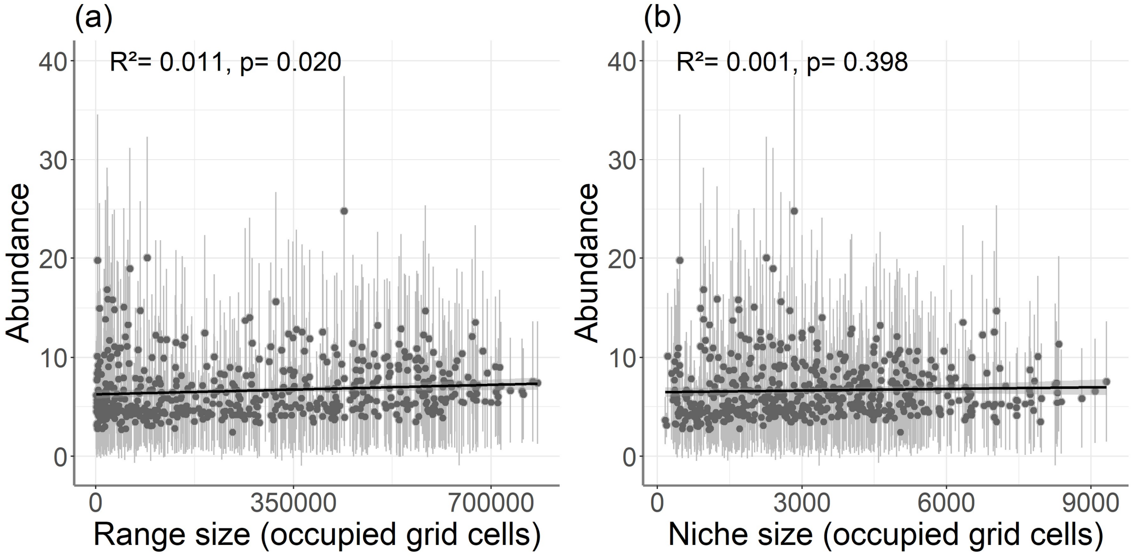

Figure III.3 Absence of a relationship between species' mean abundance with (a) range size and (b) niche size at coarse (2.5-min) raster grain. Points indicate species' mean abundance values; grey error bars are standard deviations, and black solid lines are linear regressions.

Abundance vs centrality in geographic space

Across the 517 species we found large variation in slopes of linear regressions relating abundance to the distance from the centre of geographic ranges, with species showing positive (e.g. *Brachypodium phoenicoides*), negative (e.g. *Luzula pilosa*), or no relationship (e.g. *Potentilla argentea*; Figure III.4), ranging from -0.31 to 0.38, with a median of -0.03 (Figure III.5a). Similarly diverse results were obtained using quantile regressions (90th quantile used to determine the upper limit of a triangular abundance-range centre relationship), with positive but also negative relationships ranging from -1.0 to 1.0, with a median of -0.08 (Figure III.5b). For 13.5% of the species, quantile regression revealed slopes not significantly different from 0 (Table III.2). When we summarized the slopes of both linear and quantile regressions with a mixed-effect meta-analysis, we found that the overall mean slope across 517 species was slightly negative and significantly different from 0 in both OLS linear models and linear quantile regressions, with values averaging -0.04 and -0.10, respectively (Table III.3). This indicates that the distance of the plot location to the centre of the species' geographic range is a weak and variable predictor of local abundance.

Abundance vs centrality in the climatic niche

Similar to the above-described analysis involving geographic space, we found large variation in regression slopes for the relationships between abundance and distance from the centre of climatic niches (Figure III.5c and d). Slopes from linear models ranged from -0.30 to 0.29 with a median of -0.02, while slopes from the quantile regression ranged from -1.00 to 0.89 and had a median of -0.01. For 24.1% of the species, quantile regression slopes were not significantly different from 0 (Table III.2). Summarized in a mixed-effect meta-analysis, the overall mean slope was slightly negative and significantly different from 0 in both linear models and quantile regressions, with means of -0.03 and -0.08 respectively (Table III.3). Again, this points to a weak and considerably varying relationship.

Abundance vs coarse-grain climatic suitability

The four SDM techniques ('bioclim', 'mars', 'rf', 'svm') predicted similar maps of coarse-grain climatic suitability (Figure III.2g and h). According to AUC scores, all techniques had high success rates, with AUC values averaged over 517 species being 0.947, 0.930, 0.916 and 0.718 for 'bioclim', 'mars', 'rf', and 'svm', respectively. Importantly, when predicted climatic suitability was used as a predictor of abundance, we found no general relationship, irrespective of the SDM technique used. Specifically, both OLS linear regression and linear quantile regression showed a number of relationships with both positive and negative slopes, as well as no relationship. For some of the species, linear quantile regression revealed no relationship (slope=0) between abundance and climatic suitability (Table III.2, Figure III.5e and f and Appendix S III.4.2).

Summarized in a mixed-effect meta-analysis, the overall mean slope was slightly negative for all applied SDM methods in OLS linear regression models and linear quantile regressions. The overall mean slope was significantly different from 0 for 'bioclim', 'svm' and 'rf' based predictions in OLS linear regressions and for all models in quantile regression, indicating a slightly negative trend in local plot abundance with increasing coarse-grain climatic suitability, which again points to a weak and heterogeneous relationship. The analyses were rerun on a subset of the dataset that included only grid cells with a minimum number of four occurrences for each species, which revealed similar results (see Appendix S III.4.3 and 4.4).

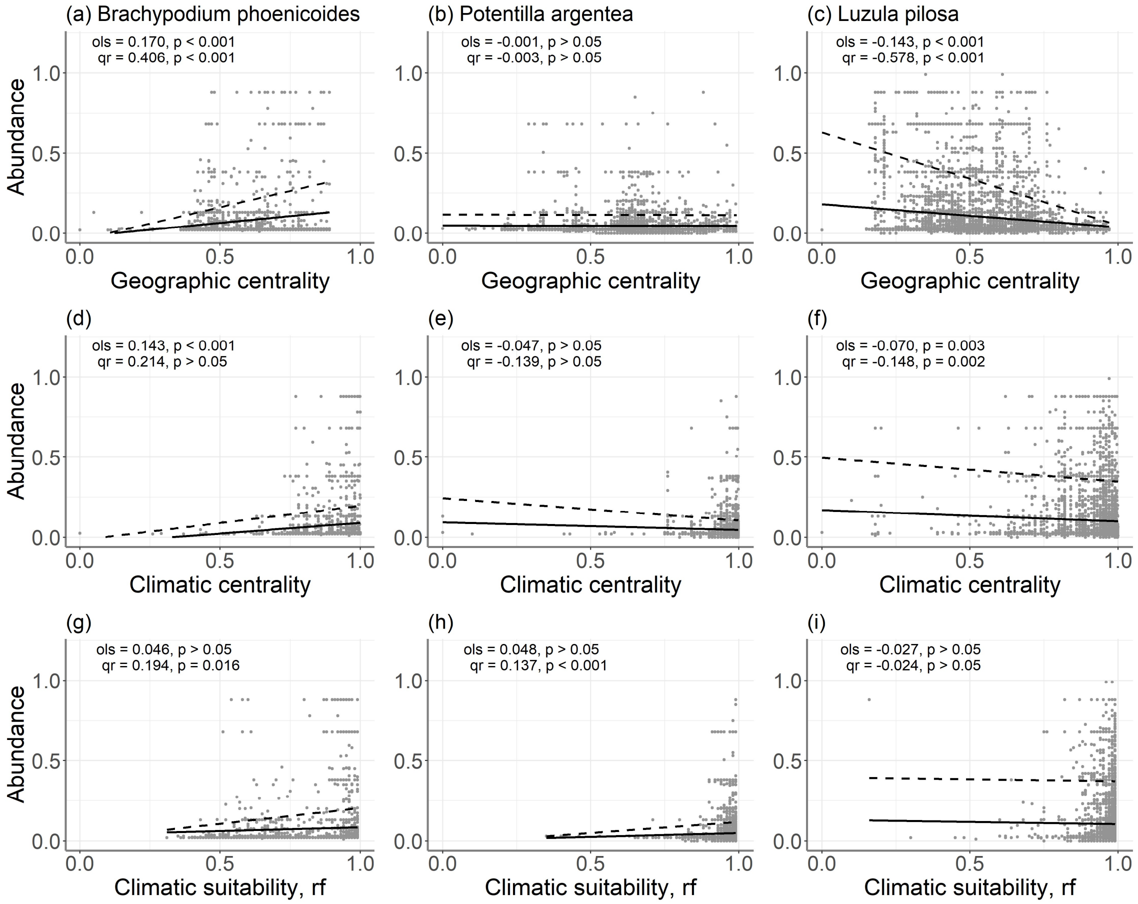

Figure III.4 Examples of herbaceous species having positive (*Brachypodium phoenicoides,* left column), nearly absent (*Potentilla argentea,* middle column), and negative (*Luzula pilosa,* right column) relationships between local abundance and properties derived from distributions in the coarse-grain geographic and climatic space. Scatterplot and regression coefficients from OLS linear regression models (ols) and linear quantile regressions (qr, 90th quantile) between abundance and centrality in geographic space (a-c), centrality in climatic space (d-f), and predicted climatic suitability from model 'random forest' (rf) (g-i). Bold lines represent the 50th quantile regression (regression from linear model); the dashed line represents the 90th quantile regression.

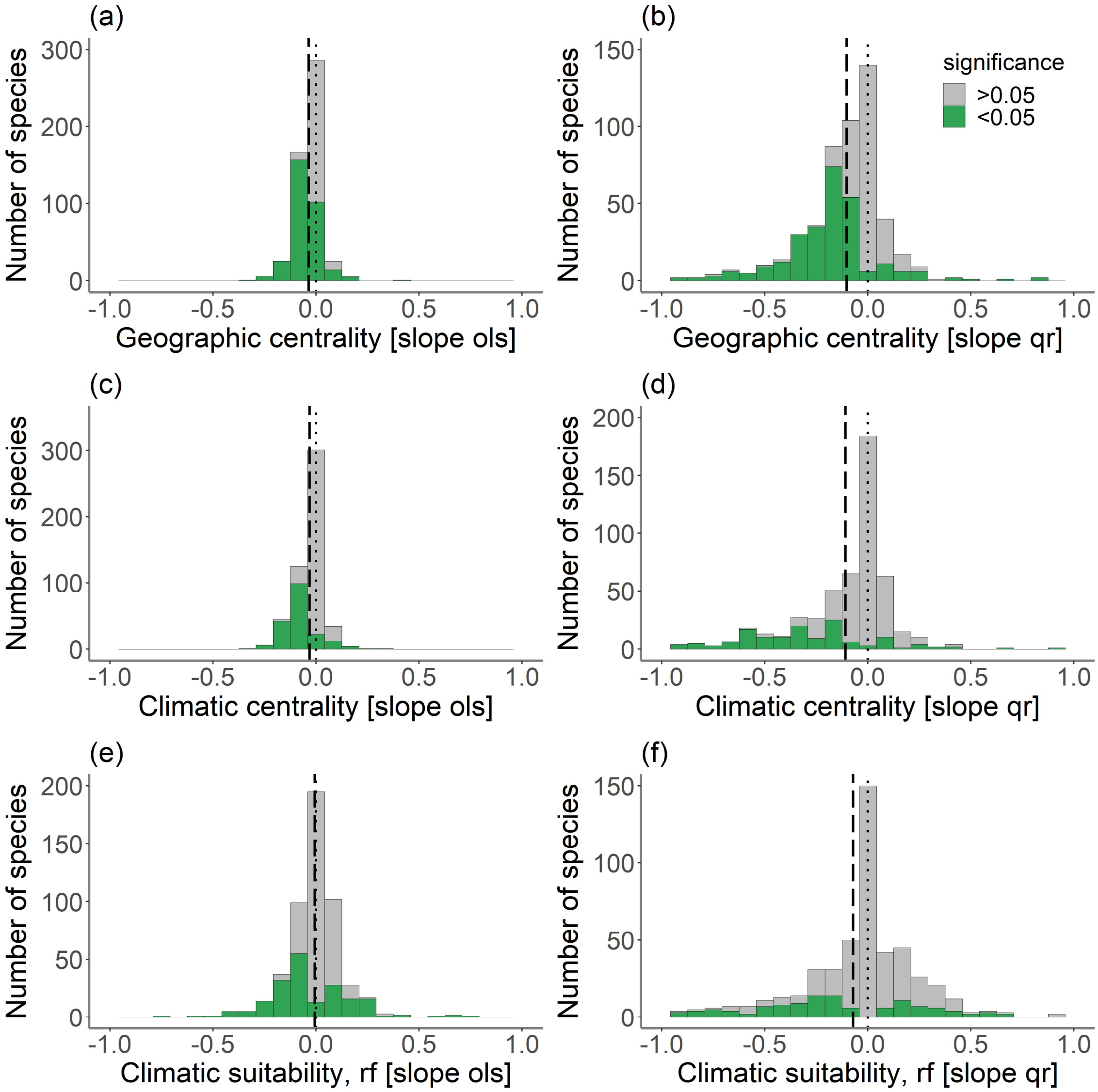

Figure III.5 Frequency distributions for 517 vascular plant species of slopes from OLS linear regressions (ols) (left) and linear quantile regressions (qr; 90th percentile) (right) between mean species local abundance and distance to range centroid (a,b); distance to niche centroid (c,d), and climatic suitability predicted from model 'random forest' (rf) (e,f). Green bars represent significant slope values. The dotted line represents slope=0; the dashed line represents the average slope.

Table III.2 Regression slopes derived from OLS linear regression models and quantile regressions (90th quantile) between species' local abundance and the six measures for distance and suitability applied (distance to range centroid, distance to niche centroid, predicted climatic suitability from 'bioclim', 'mars', 'rf' and 'svm') for 517 vascular plant species. The percentage of all 517 species included in this study showing significant ($p < 0.05$) positive, negative or no slope.

	Slope OLS linear regression		Slope linear quantile regression (90th quantile)		
Measure	[+] slope	[-] slope	[+] slope	[-] slope	no slope
Geographic centrality	6.2	54.0	6.2	45.8	13.5
Climatic centrality	3.9	32.5	4.3	23.4	24.1
Suitability bioclim	13.0	34.6	10.4	29.0	17.7
Suitability mars	17.2	22.8	11.8	18.8	14.7
Suitability rf	13.5	24.2	8.5	17.6	20.9
Suitability svm	10.8	28.0	7.4	22.6	14.6

Table III.3 Slopes of the mixed-effect meta-analysis for linear models and quantile regressions (90th quantile) between species' local abundance and the six measures for distance and suitability applied (distance to range centroid, distance to niche centroid, predicted climatic suitability from 'bioclim', 'mars', 'rf' and 'svm'); summary effect size (SE) is given in brackets; *p*-values: *= $p < 0.05$, ** = $p < 0.01$, *** = $p < 0.001$.

	Slope (SE) *p*-value	
Measure	OLS linear regression	linear quantile regression (90th quantile)
Geographic centrality	-0.036 (0.002) ***	-0.100 (0.008) ***
Climatic centrality	-0.032 (0.003) ***	-0.081 (0.009) ***
Suitability bioclim	-0.013 (0.002) ***	-0.036 (0.005) ***
Suitability mars	-0.005 (0.006)	-0.043 (0.015) **
Suitability rf	-0.010 (0.005) *	-0.031 (0.011) **
Suitability svm	-0.029 (0.004) ***	-0.092 (0.013) ***

Discussion

In our attempt to link species' local abundance to their distribution at coarse resolution and broad extents in geographic and climatic space we tested for three macroecological relationships: (1) the abundance-range size relationship, (2) the abundance-range centre relationship, and (3) the abundance-suitability relationship.

For (1), we found no empirical association between species' local abundance and the size of the species' geographic range or the size of its climatic niche estimated at a coarse spatial grain. For (2), contrary to some of our expectations, we found that, on average, species' local abundance was not related to distance with respect to the centre of its geographic range or climatic niche; however, a generally weak and slightly negative relationship indicates substantial variation, with many species having strong positive or strong negative forms of the relationship. For (3), we again unexpectedly found species' local abundance to be nearly unrelated to its predicted climatic suitability, notwithstanding the weak negative trend and, again, considerable variation. As such, despite the significant statistical associations discussed below, most of the relationships were weak and highly variable. While this finding contradicts some early macroecological propositions (Brown, 1984), it accords with some more systematic and data-intensive empirical evaluations that often showed similarly noisy and weak relationships (Gaston et al., 1997; Köckemann et al., 2009; Dallas & Hastings, 2018).

Whereas our study tested the relationships on a large group of herbaceous plant and shrub species predominantly over their full distributional ranges, many other studies focused on popular taxa and functional species groups (e.g. trees and amphibians, VanDerWal et al., 2009; mammals and trees, Dallas & Hastings, 2018; birds, Osorio-Olvera et al., 2020), or they were restricted to specific geographic regions (e.g. Australian rainforest, VanDerWal et al., 2009; China, Ren et al., 2013).

Overall, we offer two general and plausible explanations for our results. First, that processes driving species' local abundance can differ from those driving occupancy across their geographic or climatic space (Shmida & Wilson, 1985; Wiens, 1989; Bradley, 2016; Mertes & Jetz, 2018). Continental-scale presence or absence of species may be primarily shaped by broad-scale environmental conditions (e.g. climate), historical factors (Brändle & Brandl, 2001) and long-distance dispersal limitation. In contrast, species' abundance at the local scale may be more influenced by abiotic microhabitat characteristics such as soil or microclimatic conditions

(Köckemann et al., 2009; De Frenne et al., 2013), or the prevailing disturbance regimes or successional stages (Meurant, 2012; Morris, Ehrlén, Dahlgren, Loomis, & Louthan, 2020). Furthermore, biotic interactions may strongly influence local species' abundance (Moeslund et al., 2017; Dallas & Hastings, 2018).

The second explanation emphasizes the central role of temporal scale and short-term stochasticity (Lande et al., 2003), which can particularly disrupt patterns of local abundance, thereby disconnecting them from coarse-grain occurrence patterns, which are driven by long-term average conditions. This acknowledges that vegetation plots only reflect abundance at any one point in time (and thus vary within and across years, e.g. due to stochasticity), while species' coarse-grain occurrence patterns, and any derived assumptions on their climatic niches, represent long-term averages.

We consider both explanations to be plausible as that they can explain results associated with the three examined relationships, and because the grain of the vegetation plots is particularly small and thus potentially highly sensitive to both local spatial variation of microhabitats and short-term stochasticity. This is also supported by the observed high variation of abundance values around the determined relationships (see Figure III.4).

Our results have several important practical implications: The first concerns the interpretation of SDMs, which have become an essential tool in conservation planning and assessment (Peterson et al., 2011) and have been used to identify priority species and regions (Hoffmann et al., 2010). Species' coarse-grain occurrence and local abundance do not frequently display similar patterns, and they often do not even correlate well (Mi, Huettmann, Sun, & Guo, 2017). In a study by Johnston et al. (2015), locations that had been prioritized for conservation by SDMs based on species' occurrence only showed 10–58% overlap with locations prioritized by Species Abundance Models (SAMs) based on species' abundance. They conclude that SDMs do not typically identify locations of highest abundance that are crucial for the conservation of populations. Our results support this conclusion as we found no clear correlation between coarse-grain climatic suitability predicted by SDMs and local species' abundance. Since most conservation strategies are implemented at the local scale, we highlight the need to check predictions made at a global or regional scale and their transferability to the local scale, as previously proposed by Guerrero, McAllister, Corcoran, & Wilson (2013).

Another practical implication concerns rarity as a proxy for threat or conservation status. Our results indicate that for European plants, range size, especially when used as the only proxy for local abundance, seems to be a weak predictor of local rarity. Species with small ranges may be locally abundant, while those with large ranges may be locally rare. Thus, we argue that range size and local abundance should be independently considered when developing conservation measures for species or habitats.

In conclusion, we found conflicting evidence for some of the hypothesized links between species' coarse-grain distribution and local abundance, which may be due to the nature of dispersal barriers across Europe. In addition, we note that the size of the vegetation plots used can make species' abundance values more or less sensitive to local microhabitat variation and stochasticity. These results call into question any assumptions made on species' abundance at the local scale, particularly where predictive SDMs on coarse grain occurrence data were used. As such, we recommend that further detailed investigations of the processes driving species' local abundance in relation to their geographic range are required, particularly to better inform conservation measures.

Appendices

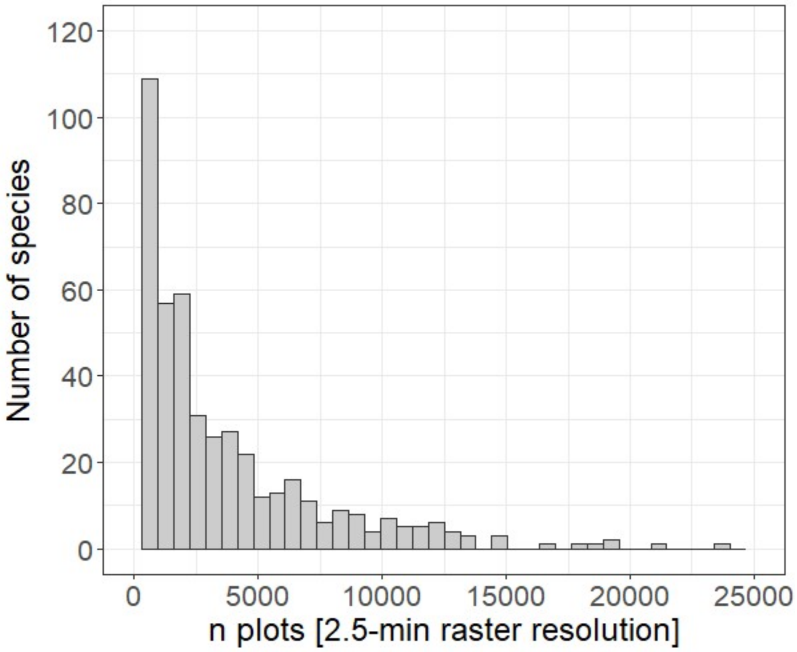

Figure S III.1.1 Distribution of number of vegetation plots per species (at 2.5-min raster resolution). Number of plots occupied per species ranged from 101 to 23,464 and for half of the studied species; number of plots was between 631 and 4,531 (25th and 75th percentile) with a median of 1,863 plots.

Moran's I was calculated to test for spatial autocorrelation between plot size and plot occurrence coordinates (latitude/longitude) in geographic space. Moran's I is ranging from -1 to 1, with perfect dispersion (negative spatial autocorrelation) at -1, complete random arrangement at 0, and perfect correlation (positive spatial autocorrelation) at +1. Here, Moran's I was 0.075, meaning plot size is randomly distributed in the studied geographic space.

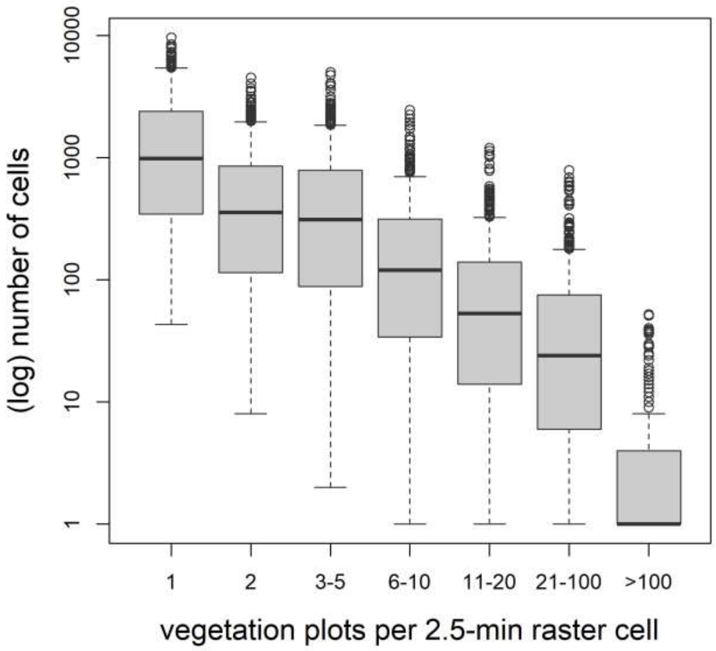

Figure S III.1.2 Boxplots on the species' specific distribution of number of vegetation plots per grid cell. Seven groups were classified by number of vegetation plots per 2.5-min raster cell.

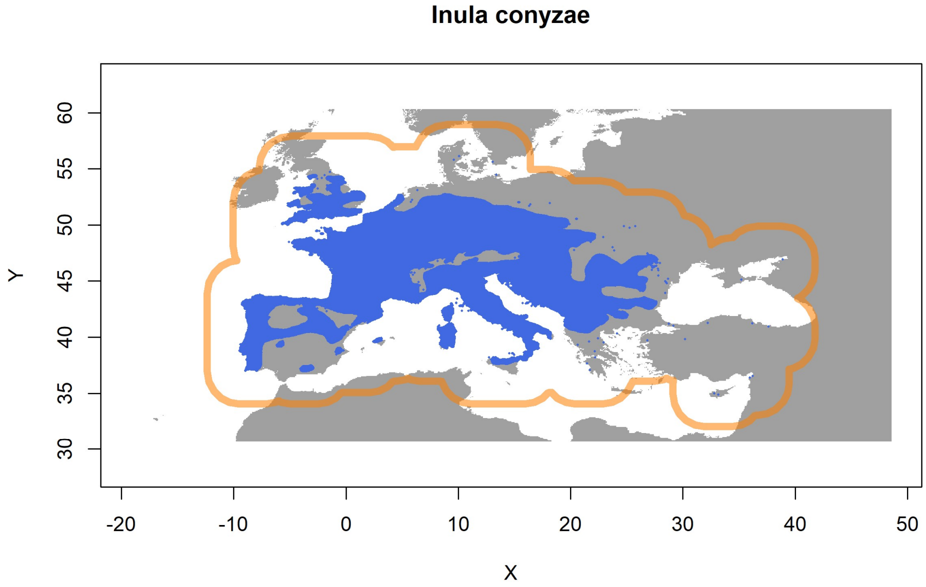

Figure S III.2.1 Geographic range (blue polygon) and range size related buffer (orange line) for *Inula conyzae.*

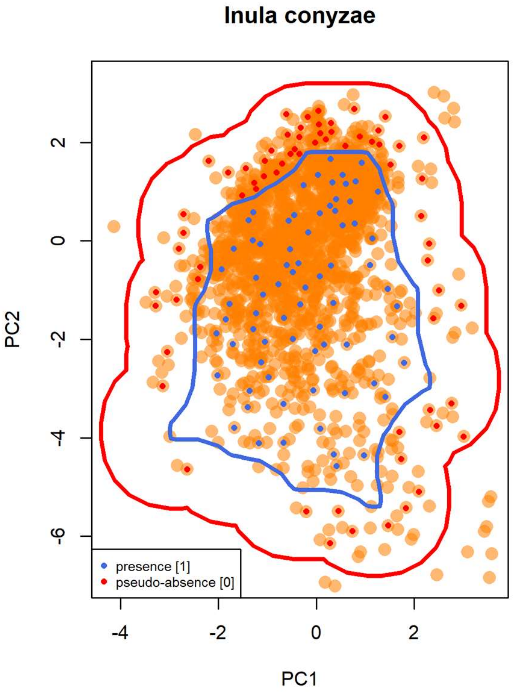

Figure S III.2.2 Available climatic combinations based on range size related buffer zone (orange dots); species realized niche based on geographic occurrences from CDH (blue line); niche size related buffer zone (red line); selected presences (blue dots); selected pseudo-absences (red dots) for *Inula conyzae.*

Appendix S III.3 Information on the 517 species included in this study. This Appendix can be found online, https://doi.org/10.1111/jbi.13926

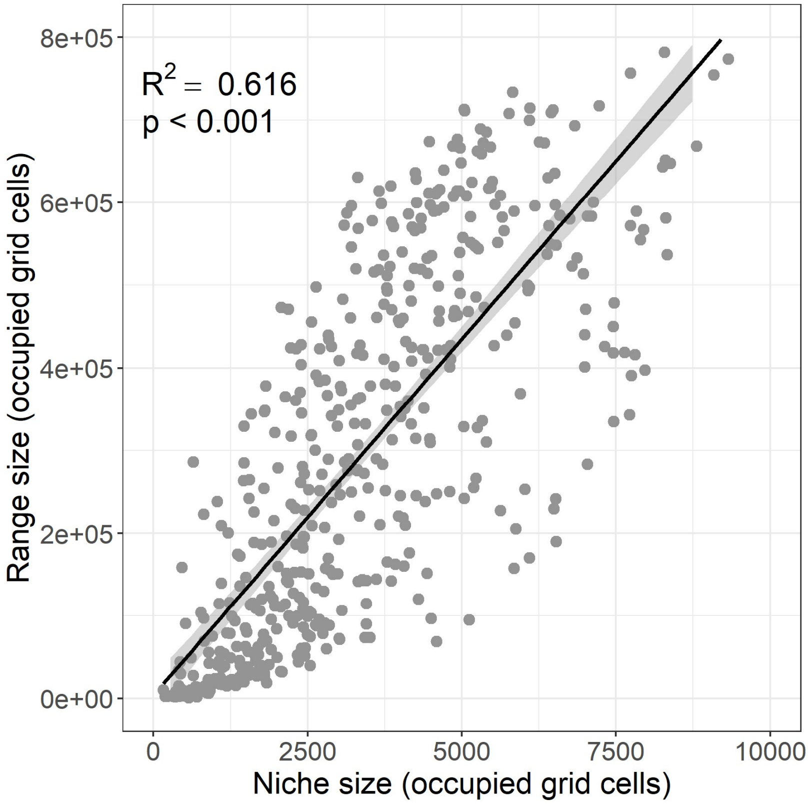

Figure S III.4.1 Relationship between niche size and range size for the 517 species included in this study

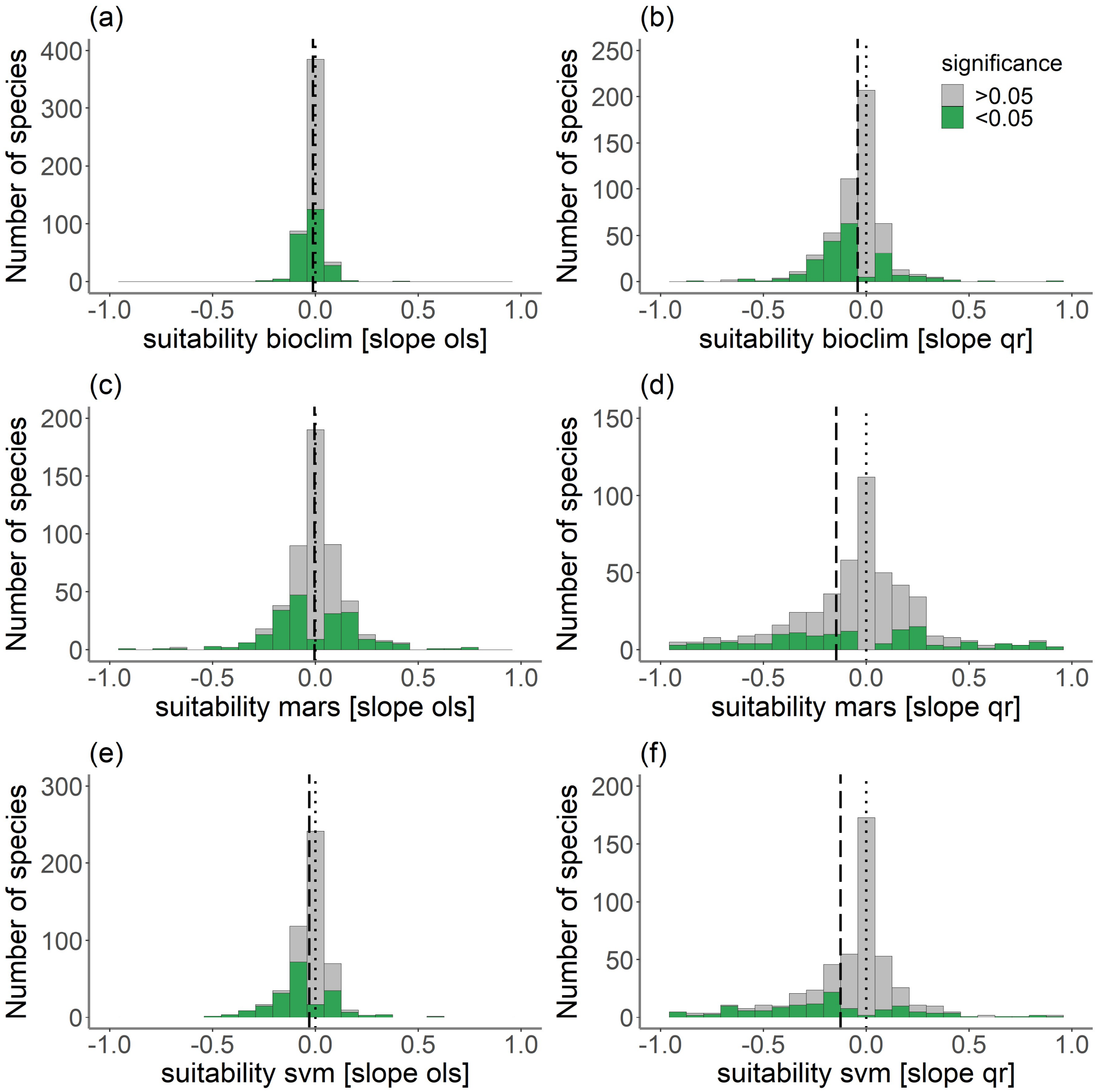

Figure S III.4.2 Frequency distributions for 517 vascular plant species of slopes from OLS linear regressions (ols) (left) and linear quantile regressions (qr, 90th percentile) (right) between mean species local abundance and climatic suitability predicted by 'bioclim' (a, b), 'mars' (c, d) and 'svm' (e, f). Green bars represent significant slope values. Dotted line represents slope=0, the dashed line represents average slope.

Table S III.4.3 Regression slopes derived from OLS linear regression models and quantile regressions (90th quantile) between species' local abundance and the six measures for distance and suitability applied (distance to range centroid, distance to niche centroid, predicted climatic suitability from 'bioclim', 'mars', 'rf' and 'svm') for 517 vascular plant species. The percentage of all species included in this study showing significant ($p < 0.05$) positive or negative or no slope. Results based on a subset dataset (including only cells with a minimum number of four plots per cell and species).

	Slope OLS linear regression		Slope linear quantile regression (90th quantile)		
Measure	[+] slope	[-] slope	[+] slope	[-] slope	no slope
Geographic centrality	9.0	22.7	6.2	17.2	2.7
Climatic centrality	7.8	27.5	8.2	20.3	1.4
Suitability bioclim	92.0	0.0	60.4	0.2	0.2
Suitability mars	na	na	na	na	na
Suitability rf	11.9	19.3	11.9	17.0	2.3
Suitability svm	7.6	32.6	6.8	23.8	1.2

Table S III.4.4 Slopes of the mixed-effect meta-analysis for linear models and quantile regressions (90th quantile) between species' local abundance and the six measures for distance and suitability applied (distance to range centroid, distance to niche centroid, predicted climatic suitability from 'bioclim', 'mars', 'rf' and 'svm'); summary effect size (SE) is given in brackets; *p*-values: *= $p < 0.05$; ** = $p < 0.01$; *** = $p < 0.001$. Results based on a subset dataset (including only cells with a minimum number of four plots per cell and species).

	Slope (SE) *p*-value	
Measure	OLS linear regression	linear quantile regression (90th quantile)
Geographic centrality	-0.038 (0.003) ***	-0.097 (0.006) ***
Climatic centrality	-0.034 (0.003) ***	-0.068 (0.008) ***
Suitability bioclim	-0.016 (0.002) ***	-0.037 (0.004) ***
Suitability mars	na	na
Suitability rf	-0.023 (0.005) ***	-0.049 (0.012) ***
Suitability svm	-0.036 (0.005) ***	-0.087 (0.012) ***

Chapter IV

Different sets of traits explain abundance and distribution patterns of European plants at different spatial scales

This chapter is under revision in *Journal of Vegetation Science* as

Sporbert, M., Welk, E., Seidler, G., Jandt, U., Aćić, S., Biurrun, I., Campos, J.A., Čarni, A., Cerabolini, B.E.L., Chytrý, M., Ćušterevska, R., Dengler, J., De Sanctis, M., Dziuba, T., Fagúndez, J., Field, R., Golub, V., He, T., Jansen, F., Lenoir, J., Marcenò, C., Martín-Forés, I., Moeslund, J.E., Moretti, M., Niinemets, Ü., Penuelas, J., Pérez-Haase, A., Vandvik, V., Vassilev, K., Vynokurov, D. & Bruelheide, H.. Different sets of traits explain abundance and distribution patterns of European plants at different spatial scales

Abstract

Aim: Plant functional traits summarize major axes describing plant form and function across taxa and biomes. We assess whether geographic range size, climatic niche size, and local abundance of plants relate to predictive sets of traits (trait syndromes) or are driven by single traits.

Location: Europe (vegetation plots), Eurasia (distribution ranges)

Methods: Plant species distribution maps were extracted from the Chorological Database Halle to derive information on the geographic range size and climatic niche size for 456 herbaceous, dwarf shrub and shrub species. We estimated local species abundances based on 740,113 vegetation plots from the European Vegetation Archive, where abundances were available as plant species cover per plot. We compiled a complete species-trait-matrix of 20 plant functional traits from trait databases (TRY, BiolFlor and CLO-PLA). The relationships of species geographic range size, climatic niche size and abundance with single traits and trait syndromes were tested with multiple linear regression models.

Results: All plant distribution and abundance variables were related to functional traits. Generally, local abundances were much more strongly related to traits than geographic distributions, yet the specific traits also differed. Geographic range size increased with plant height, whereas climatic niche size decreased with leaf C content. Both, geographic range size and climatic niche size increased with specific leaf area (SLA). Species local abundance increased with leaf area and SLA.

Conclusion: Functional traits matter for species' distributions at both local and large regional scales. Local abundances are shaped by a different combination of traits as compared to broad-scale distributions, pointing to filtering by different environmental and ecological factors acting at distinct spatial scales. However, traits related to the leaf economics spectrum were found to be important at both spatial scales, emphasizing the general importance of the acquisitive-conservative trait syndrome axis for the distribution of vascular plant species.

Keywords: Chorological Database Halle (CDH), climatic niche, commonness and rarity, European Vegetation Archive (EVA), functional traits, geographic range, macroecology, multi-scale, vegetation-plot data.

Introduction

A central aim in functional macroecology is to understand to what degree plant morphological and physiological fitness-determining traits (Violle et al., 2007) affect species' distributions at broad spatial scales (i.e. occurrences across the geographic and climatic space) as well as local abundances within communities (Brown, 1995; McGill et al., 2006). The limits of species' broad-scale distributions reflect the interplay between dispersal barriers and bioclimatic conditions that affect range dynamics and evolution (Baselga et al., 2012). In contrast, species' local abundance depends on factors operating at the local scale of species assemblages, such as habitat suitability, the local combination of environmental conditions, and biotic interactions (Peterson et al., 2011; Staniczenko et al., 2017). Under the assumption that species' functional traits reflect the mechanisms through which species respond to abiotic and biotic conditions to maximise their fitness, these traits are expected to predict both broad-scale distribution and local abundances (Suding et al., 2008; Heino and Tolonen, 2018).

Plant communities consist of different kinds of locally co-occurring species. Among these, some species are widely distributed, while others have restricted distributions. Similarly, some co-occurring species are locally abundant, while others are locally rare (Rabinowitz, 1981; Gurevitch et al., 2002; Enquist et al., 2019). It has been observed that widely distributed species with large geographic range sizes tend to have broad environmental tolerances (i.e. broad climatic niches), while geographically narrowly distributed species are also more likely to be narrowly distributed in climatic space (Slatyer et al., 2013). A positive relationship between climatic niche size and geographic range size across species thus seems to be a general macroecological pattern (Gaston, 2000; Slatyer et al., 2013; Cardillo et al., 2019). At the local scale, species abundance values are frequently used as descriptors of species performance and are an important characteristic of the composition of herbaceous plant communities (Kent and Coker, 1992; Chiarucci et al., 1999). Within the geographic distribution range of a species, local abundance is often highly variable. Typically, species exhibit low cover at most sites and high cover at only a few sites across their distribution range (Murphy et al., 2006). In contrast to 'everywhere-sparse' species, these `somewhere-abundant´ species are reflected in right-skewed species abundance patterns that are ubiquitous in plant community ecology (McNellie et al., 2019). This skewness in local abundance might be caused by the distribution of optimal habitats, and thus, might represent a species characteristic itself. It has to be assumed that mean abundance does not capture the full variability of skewed frequency distributions. Thus, both the mean and the skewness of a species' cover value distribution should be considered as proxies for rarity or commonness.

Functional traits have been used as proxies for species' dispersal abilities (Greene and Johnson, 1993; Thompson et al., 2011), environmental tolerances (Loehle 1998; Bohner and Diez, 2020) or competitiveness (Kunstler et al., 2016). Specific functional traits have been linked to commonness and rarity on both local and broad spatial scales. For example, studies have found plant height, used as a proxy for species competitive ability, to be positively correlated with range size, with taller species more widespread than shorter ones (Lavergne et al., 2004; Kolb et al., 2006). On the local scale, common species have again been associated with taller stature and with other traits that are proxies for species' productivity, i.e. larger specific leaf area (SLA) and higher leaf nitrogen (N) content (Grime et al., 1997; Hegde and Ellstrand, 1999; Lavergne et al., 2004; Mariotte, 2014). Nitrogen (N) and phosphorus (P) availability limit plant growth in most terrestrial ecosystems (Güsewell, 2004). Low nutrient availability (e.g. phosphorus limitation) may weaken the relationships between productivity-related traits and macroclimate (Bruelheide et al., 2018). As a consequence, there might be a negative correlation between species' N:P ratio and both their local abundance and broad-scale distributions. Regarding species persistence, locally more abundant species have been associated with perennial life span and clonal growth (Eriksson and Jakobsson, 1998; Kolb et al., 2006). In contrast, at broad spatial scales, rare species have been associated with clonal growth (Kelly and Woodward; 1996) and being shrubs and trees rather than herbs (Oakwood et al., 1993). Several studies investigated the relationship linking dispersal or regeneration-related traits with species' local abundance and broad-scale distribution. On the local scale, more abundant species were found to produce fewer and lighter seeds than rare species (Hedge and Ellstrand, 1999; Guo et al., 2000; Kolb et al., 2006). In contrast, across broad spatial extents, geographically widespread species have been found to produce significantly more and heavier seeds than restricted plant species (Lavergne et al., 2004; Kolb et al., 2006; Van der Veken et al., 2007).

While some studies have found relationships between functional traits and local abundance and/or broad-scale distribution, others failed to detect a clear correlation (see Table IV.1).

Table IV.1 Traits used in this study, their function in the community, and their reported correlation with local abundance and broad-scale distribution being unimodal (—), positive (↑) or negative (↓).

Trait	**Function**	**Reported correlation with**	
		local abundance	**broad-scale distribution**
Specific leaf area, Leaf C, Leaf N, Leaf P, Leaf dry matter content	Productivity, competitive ability	Specific leaf area↑[1], Leaf N↑[1]	Specific leaf area—[2], Leaf N—[2], leaf dry matter content—[2]
Leaf N:P ratio	Nutrient supply	↓[3,14]	↓[3,14]
Plant height, Leaf area	Competitive ability	Plant height↑[1,2,4,5], Leaf area↑[1]	Plant height—[6]↑[2,7]
Seed mass, Seed number per reproductive unit, Dispersal unit length	Dispersal, regeneration	Seed mass—[6,7,13]↓[10], Seed number per reproductive unit↓[4,7]	Seed mass—[2,9]↑[7]↓[10,11], Seed number per reproductive unit↑[2,7,8]
Life span, Life form, Clonality	Persistence	Perennials↑[6,7], Clonal growth↑[6,7]	Therophytes↑[11], Phanerophytes↓[11], Clonal growth—[6]↓[12]

[1] Mariotte, 2014; [2] Lavergne et al., 2004; [3] Bruelheide et al., 2018; [4] Hedge and Ellstrand, 1999; [5] Grime et al., 1997; [6] Eriksson and Jakobsson, 1998; [7] Kolb et al., 2006; [8] Van der Veken et al., 2007; [9] Thompson et al., 1999; [10] Guo et al., 2000; [11] Oakwood et al., 1993; [12] Kelly and Woodward, 1996; [13] Leishman and Murray, 2001; [14] Güsewell, 2004

So far, the majority of studies have focused on single traits rather than on trait combinations or trait syndromes (Diaz et al., 2016; Guo et al., 2018) as predictors of broad and local distribution patterns. However, no single trait can completely describe a species' ecological strategy (Winemiller et al., 2015; Marino et al., 2020). Rather, species' local abundance and broad-scale distribution patterns might be affected by different sets of traits (Marino et al., 2020). It has been suggested that locally rare and geographically restricted plant species differ systematically from more common species in functional traits that are related to species productivity, competitive ability, dispersal, regeneration and persistence (Murray et al., 2002). However, the different states and values of traits cannot be unconditionally combined. Diaz et al. (2004) highlighted that the functional space occupied by vascular plant species is strongly constrained by trade-offs between traits. On the one hand, the leaf economics spectrum describes a productivity-persistence trade-off and contrasts species with a set of successful trait combinations for quick returns on investments of nutrients and dry mass in leaves to species with a slower potential rate of return (Wright et al., 2004). On the other hand, the size spectrum reflects the species life cycle, with small stature species, smaller seeds and short lifespans versus long-lived woody plants (Díaz et al., 2016; Table IV.1).

In this study, we aim to unravel the relationship existing between single traits and trait syndromes and species distribution and dominance at broad and local scales. Specifically, we focused on 20 traits that are expected to respond to bioclimatic drivers and capture the essence of plant life forms and functions (Wright et al., 2004; Petchey and Gaston, 2006; Diaz et al., 2016; Bruelheide et al., 2018). We tested for these relationships across 456 European herbaceous, dwarf shrub and shrub species by inter-relating existing data on functional traits with the species' (i) geographic range size, (ii) climatic niche size, and (iii) species' local abundance, which was measured as (iii. a) mean cover from all the vegetation plots in which a species was present and (iii.b) skewness of cover values. We expect climatic niche size and geographic range size to be driven by the same underlying environmental factors and ecological processes (Colwell and Rangel, 2009), and therefore to be positively correlated, and to be predicted by many of the same single traits or trait syndromes (Table IV.1).

Material and Methods

Broad-scale distribution metrics: geographic ranges and climatic niches

We used available digitized species distribution data (i.e. range polygons and point occurrences) of the Chorological Database Halle (CDH) (E. Welk et al., unpublished data) to assess the Eurasian geographic ranges of 456 herbaceous, dwarf shrub and shrub species, including their neophytic occurrences. A list of these species can be found in Appendix S IV.1, available online, https://figshare.com/s/e61c08fea258e420f309. CDH stores information on distribution ranges of about 17,000 vascular plant taxa. For 5,583 taxa, maps were compiled based on published distribution range maps (Meusel and Jäger, 1992; Meusel et al., 1978; Meusel et al., 1965), national and floristic databases and further maps from floristic literature (see bibliographic details in Index Holmiensis: Lundqvist, 1992; Lundqvist and Jäger, 1995-2007; Lundqvist and Nordenstam, 1988; Tralau, 1969-1981). CDH data can be requested for research objectives via chorologie.biologie.uni-halle.de/choro/. We aggregated species' point and polygon distribution data using a raster grid layer of 2.5 arc-min resolution, which corresponds to grid cells covering approximately 15 km² each across Central Europe. As a measure of range size for each species, we counted the number of grid cells occupied (approximating the area of occupancy in the geographical space).

We determined the multi-dimensional climatic space (or climatic niche) of each geographic range based on principal components analysis (PCA) of 19 bioclimatic variables from the WorldClim2.0 database (Fick and Hijmans, 2017), also at 2.5 arc-min resolution. The resulting global background climatic space is well represented by the first two PC axes, which accounted for 70.75% of the total climatic variance. The two-dimensional PCA space was rasterized into 100 x 100 PCA-grid cells, considered as the background climatic niche, as explained in Appendix S II.1 in chapter II. The species' niche size was then calculated as the number of PCA grid cells occupied in the climatic space (i.e., the area of occupancy in the bioclimatic niche space) (for detailed information see S II.1 in chapter II).

Local abundance metrics in vegetation plots: mean cover and skewness of cover values

As a measure of local abundance, cover percentage values were obtained for each of the 456 study species in 740,113 vegetation plots from the European Vegetation Archive (EVA) (Chytrý et al., 2016) in October 2015. EVA comprised vegetation plots from all European countries plus Turkey, Georgia, Armenia, Azerbaijan and parts of Russia. We included vegetation plots from all vegetation types available from EVA, except aquatic vegetation. We matched synonymous species names according to the taxonomic reference list for Germany (German SL version 1.2, Jansen and Dengler, 2008) and to four taxonomic reference lists available via the R package 'taxize'

(Chamberlain and Szöcs, 2013; R Core Team, 2018), i.e. Encyclopedia of Life (EOL), International Plant Names Index (IPNI), Integrated Taxonomic Information Service (ITIS) and Tropicos. In cases where no exact match was found, taxon names were resolved using Taxonomic Name Resolution Service (TNRS) and all names matched or converted from a synonym where considered accepted taxon name when probabilities were ≥ 95%. We merged the data for subspecies at the species level following the taxonomic hierarchy in TNRS. We only selected species that occurred within at least 100 vegetation plots in the selected EVA dataset. Vegetation plots with geographic location uncertainty of more than 10 km were removed prior to this selection. The median number of vegetation plots per species was 2,162 (interquartile range 846 to 5,137). Information on source databases that provided vegetation plot data can be found in Appendix S II.3 in chapter II. Cover or cover-abundance values that were based on ordinal scales (e.g. Domin, 1928; Braun-Blanquet, 1951) were transformed to percentage cover (van der Maarel, 1979). When more than one plot per species was present in a 2.5 arc-min raster cell, mean values of percentage cover were calculated.

For each species, we measured two aspects of its abundance across the vegetation plots. First, we calculated its ‘mean cover’: the arithmetic mean of the percentage cover values from all the vegetation plots at 2.5 arc-min raster cells in which the species was present in EVA. Second, we evaluated the frequency distribution of these percentage cover values (see Figure IV.1 for details on the procedure for three example species). To do this, we computed the shape of the distribution function of the percentage cover values. As those values are not normally distributed (Figure IV.1 a-c), we developed a non-parametric approach for measuring the shape of the distribution function. This was achieved by calculating the distribution quantiles in 5%-steps, resulting in 20 quantile values. We then fitted a non-linear model on the extracted 20 quantile values and obtained the estimate and the credible interval of the area under the histogram, AUH (Figure IV.1 d-f). We applied a Bayesian Markov chain Monte Carlo (MCMC) method following Feng et al. (2017), using an exponential distribution, 0.95 confidence level and 10,000 iterations. The resulting AUH value ranged from 0 to 1, with values lower or higher than 0.5 meaning that the distribution of cover values for a focal species is right- or left-skewed, respectively. The lower the AUH value, the higher was the rarity (i.e. the proportion of relatively low cover values). Thus, the AUH values are suitable as proxies for abundance structure across the vegetation plots. Hereafter, we refer to the AUH values as ‘skewness of cover values’ and use it as an alternative metric, additional to mean cover, to assess across-plot species abundance.

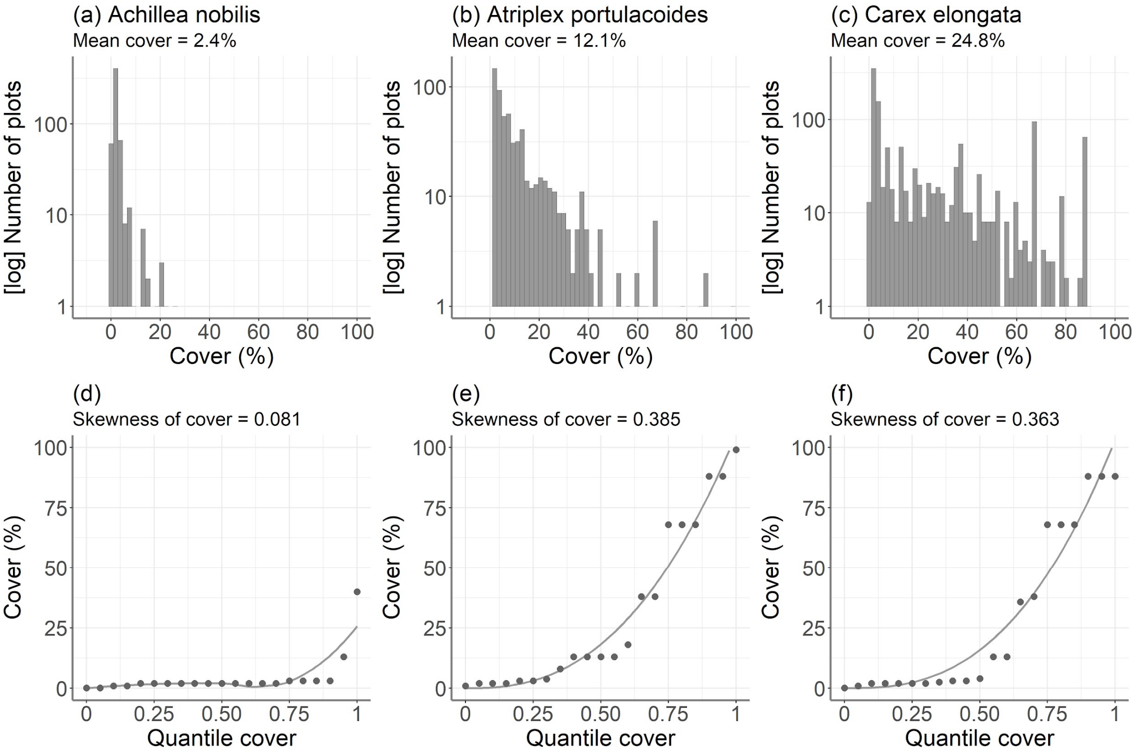

Figure IV.1 Examples of distribution of species’ cover values from vegetation plots and calculated mean cover value for (a) *Achillea nobilis,* the species with the lowest mean cover value, (b) *Atriplex portulacoides*, a species with intermediate mean cover value and (c) *Carex elongata,* the species with the highest mean cover value. Note the log scale for frequency. Distribution quantiles from species’ cover values were calculated and used to compute the shape of the frequency distribution function for each species, respectively (d-f). Non-linear models on the extracted quantile values were applied to calculate the area under the histograms of cover values (AUH), ranging from 0 to 1, with values close to 0 indicating for a strongly right-skewed distribution whereas values close to 1 point to a strongly left-skewed distribution of cover values.

Plant functional traits

We compiled a complete species trait matrix with 20 plant functional traits (see Table IV.2 and Appendix S IV.1). The trait matrix included nine binary variables: five for life form following Raunkiaer (1934), three for life span (derived from BiolFlor database;

Kühn et al., 2004) and the latter for clonality (derived from CLO-PLA database; Klimešová et al., 2017). We included information on 11 continuous trait variables from the global plant-trait database TRY (Kattge et al., 2020). All continuous traits were derived from Bruelheide et al. (2018) who applied a gap-filling approach with Bayesian Hierarchical Probabilistic Matrix Factorization (BHPMF; Schrodt et al., 2015) to fill gaps in the observed species-by-trait matrix data received from TRY. Continuous trait variables were ln-transformed prior to analysis. Functional traits values for each species can be found in Appendix S IV.1.

Table IV.2 Traits, abbreviation of trait names, units of measurement, mean and standard deviation of observed trait values and ln-transformed trait values.

Trait	Abbreviation	Unit	Mean	SD	Mean (ln)	SD (ln)
Leaf area	LeafArea	mm^2	2128.74	6346.69	6.36	1.64
Specific leaf area	SLA	m^2/kg	23.30	8.99	3.08	0.37
Leaf C content	LeafC	mg/g	451.36	24.93	6.11	0.06
Leaf N content	LeafN	mg/g	24.59	7.61	3.16	0.31
Leaf P content	LeafP	mg/g	2.14	0.95	0.67	0.45
Leaf dry matter content	LDMC	mg/g	0.22	0.08	-1.58	0.35
Leaf N:P ratio	LeafNPratio	g/g	12.13	6.35	2.41	0.40
Plant height	PlantHeight	m	0.41	0.44	-1.17	0.71
Seed mass	SeedMass	mg	2.36	4.59	-0.22	1.59
Seed number per reproductive unit	SeedNumRepUnit		42956.49	447429.20	6.15	2.38
Dispersal unit length	DispUnitLeng	mm	3.18	2.29	0.93	0.69
		Proportion of the 456 study species [%]				
Life span						
Annual	annual	16.2	-	-	-	-
Biennial	biennial	9.4	-	-	-	-
Perennial	perennial	81.8	-	-	-	-
Life form			-	-	-	-
Phanerophyte	Phaneroph	5.0	-	-	-	-
Chamaephyte	Chamaeph	6.1	-	-	-	-
Hemicryptophyte	Hemicrypt	78.2	-	-	-	-
Geophyte	Geoph	10.7	-	-	-	-
Therophyte	Theroph	14.3	-	-	-	-
Clonality			-	-	-	-
Clonal Growth	ClonalGrowth	88.8	-	-	-	-

A principal component analysis (PCA) of the 20 traits included in this study was generated using the package ‘factoextra’ (Kassambara and Mundt, 2017), allowing the visualization of the trait contributions (loadings) to the first and second axis of the PCA

(Figure IV.2). The first axis corresponded to traits of life form, life span and clonal growth and accounted for 18.8% of the total variation in trait values. The second axis, corresponding to leaf traits, accounted for 14.6% of the total variation in trait values. The third and fourth axis corresponded to dispersal traits and life form and accounted for 11.4% and 8.7% of the total variation in trait values, respectively. The mean and standard deviation of observed trait values and ln-transformed trait values are given in Table IV.2. A correlation matrix including the correlation coefficients of all trait combinations can be found in Appendix S IV.2.1. The values of trait contributions (loadings) to all PCA axes are given in Appendix S IV.2.2.

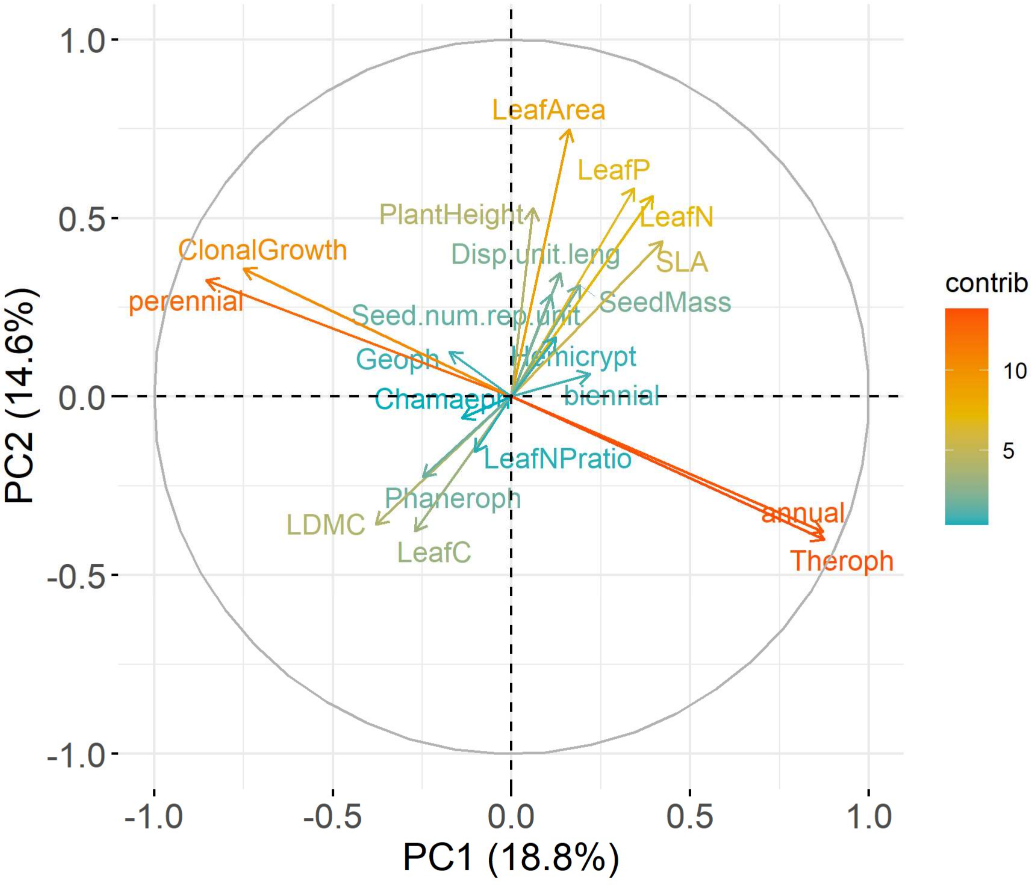

Figure IV.2 Principal component analysis of the 20 traits included in this study. Colour represents the trait contributions (loadings) to the PCA. The first and second axis accounted for 18.8% and 14.6% of the total variation in trait values, respectively. For abbreviation of the trait names see Table IV.2.

Relationships linking geographic range size, climatic niche size, and mean cover and skewness of cover values as response variables with functional traits as explanatory variables

We used the function ‘phylo.maker’ from the package ‘V.PhyloMaker’ (Jin and Qian, 2019) to create a phylogenetic tree of the studied species. The function ‘phylo4d’ from the package ‘phylobase’ (Hackathon et al., 2013) was applied to link trait data to the species’ phylogeny. We applied Pagel’s Lambda statistic (Pagel, 1999) and Fritz and Purvis’ *D* (2010) statistics to quantify the strength of phylogenetic signal in each of the 20 studied trait variables. Pagel’s Lambda statistic revealed a strong phylogenetic signal in all continuous trait variables. Fritz and Purvis’ *D* revealed a phylogenetic signal in all binary trait variables (see Appendix S4.3). Therefore, we ran phylogenetic generalized least squares models using the function ‘pgls’ from the package ‘caper’ (Orme et al., 2018), that take into account the phylogenetic non-independence between species trait values when analyzing the relationships linking plant functional traits and mean cover values, skewness of cover values, geographic range size, and climatic niche size. To allow fair comparisons of the magnitude of effect sizes between continuous and binary trait variables, all continuous and ln-transformed trait variables were standardized by subtracting the mean and dividing by two standard deviations (Gelman, 2008).

The lasso procedure in function ‘glmnet’ from the package ‘glmnet’ (Friedman et al., 2010) was applied to extract those of the 20 trait variables that contributed the most in the four linear models (i.e. mean cover values, skewness of cover values, geographic range size, and climatic niche size being the four studied response variables). The minimum lambda ratio value in each of the four linear models was 0.01, 0.0005, 0.0001 and 0.01 for mean cover values, skewness of cover values, geographic range size and climatic niche size, respectively (see Appendix S IV.2.4 for the predictor variables included in the models). To test the effect of trait combinations and interactions for each of the four response variables, we then tested all possible combinations of the predictor variables that contributed the most, including their two-fold interaction terms, applying the ‘dredge’ function from the ‘MuMIn’ package (Barton, 2019). We allowed for a maximum of three predictor terms to be included in a given candidate model (m.max=3). Finally, the Akaike Information Criterion (AIC), with delta AIC < 2 was used to identify the most parsimonious candidate model for each of the four studied response variables. We computed the Variance Inflation Factor (VIF) for each predictor term in the most parsimonious models to check for potential multicollinearity issues among the continuous predictor variables, using the function ‘vif’ from the package ‘car’ (Fox & Weisberg, 2019). To account for the fact that traits can be summarized by a few trait dimensions (Díaz et al. 2016), we repeated this analysis with using principal components instead of trait values. We used all the species scores on all PC axes as predictor terms in the models described above and tested for combinations and interactions of PC scores in the same way as described for traits. By applying the ‘dredge’ function we tested all possible combinations of the predictor variables that contributed the most, including two-fold interaction between PC scores, for each of the four response variables. AIC with delta AIC < 2 was used to identify the most parsimonious candidate model with a maximum of three predictor terms for each of the four studied response variables (see Appendix S IV.2.2 for the trait contributions (loadings) to all PC axes).

Results

Broad-scale distribution metrics: geographic range size and climatic niche size

Species’ range size (number of occupied grid cells in geographical space) ranged from 1,947 in *Dactylorhiza sambucina* to 782,025 in *Stellaria media*. The median range size was 310,070 cells (interquartile range 108,374 to 511,619). Species’ climatic niche size (number of occupied PCA grid cells within the climatic niche space) ranged from 162 in *Scabiosa canescens* to 9,318 in *Plantago major* (median 3,236; interquartile range 1,831 to 4,828. We found a positive relationship between species geographic range and climatic niche size ($R^2 = 0.605$, p-value < 0.001 in a phylogenetically corrected model; Figure IV.3a).

Local abundance metrics: mean cover and skewness of cover values

Species’ mean cover from all the vegetation plots in which a species was present ranged from 2.4% for *Achillea nobilis* to 24.8% for *Carex elongata* (Figure IV.1a and c). For the middle two quarters of the studied species, the mean cover ranged from 4.6% to 8.1% (interquartile range between the 25th and 75th percentile) with a median of 5.9%. Species’ skewness of cover values ranged from 0.081 (strongly right-skewed distribution of low cover values) in *Achillea nobilis* to 0.385 in *Atriplex portulacoides* (Figure IV.1d and e). The interquartile range was 0.158 to 0.226 and the median was

0.180. Species' mean cover was positively related to species' skewness of cover values in a phylogenetically corrected model ($R^2 = 0.763$, *p*-value < 0.001; Figure IV.3b).

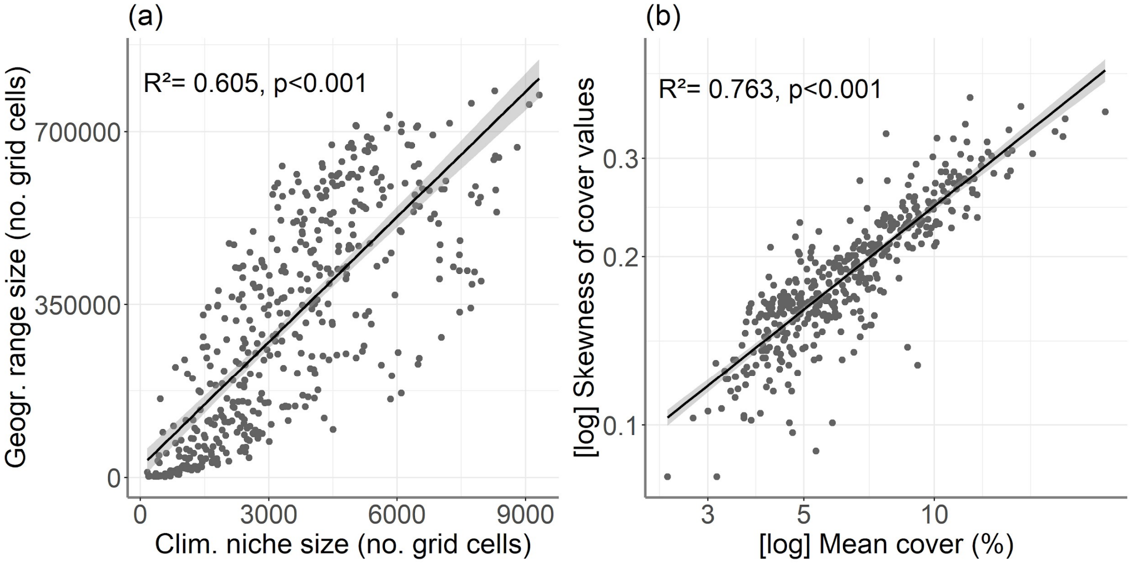

Figure IV.3 Scatterplot and regression coefficient from phylogenetic generalized least squares models between (a) species' climatic niche size and geographic range size and (b) abundance measured as mean cover and frequency distribution of cover values

Relationships between mean cover, distribution of cover values, geographic range size, and climatic niche size with functional traits

In general, species geographic range and climatic niche size were larger in species that had a higher SLA value. In addition, geographic range size was larger in species with taller stature and a lower leaf N:P ratio. In contrast, climatic niche size was larger in species that had a lower leaf C content. The mentioned functional traits were significantly related to species' geographic range size ($R^2 = 0.090$, *p*-value < 0.001) and climatic niche size ($R^2 = 0.069$, *p*-value < 0.001) in the phylogenetic generalized least squares models (Table IV.3, Figure IV.4a and b). Species mean cover and the skewness of cover values was higher in species with a higher SLA value and with higher leaf area values. The interaction of the variables SLA and leaf area was positive and the named functional traits were significantly related to species mean cover ($R^2 = 0.211$, *p*-value < 0.001) and the AUH measure of the skewness of cover values ($R^2 = 0.169$, *p*-value < 0.001; Table IV.3, Figure IV.5a and b).

The above-mentioned traits that were identified in the final models for the four predictor variables had also high loadings of the PC axes that were identified as important predictors in the PCA based models (see Appendix S IV.2.2 for the trait contributions (loadings) to the first (PC1) to 20th (PC20) PC axes). In addition, the axes represented some more traits with maximum absolute loadings that were not selected in the final models based on only single trait variables, such as leaf area for geographic range size, leaf P content for climatic niche size, clonal growth for mean cover and leaf area for skewness of cover values. However, for each of the four response variables, the three PCs in the final models explained less variation than in the trait-based models: geographic range size ($R^2 = 0.063$, *p*-value < 0.001; in the sequence of importance, the model included PC 12, 2 and 1), climatic niche size ($R^2 = 0.069$, *p*-value < 0.001; based on PC 1, 12 and 19), mean cover ($R^2 = 0.123$, *p*-value < 0.001; based on PC 13, 2 and 4) and skewness of cover values ($R^2 = 0.094$, *p*-value < 0.001; based on PC 6, 2 and 4).

Table IV.3 Effects of species traits on species geographic range size, climatic niche size, mean cover and skewness of cover values. Phylogenetic generalized least squares models were applied to test for the relationships. Following the lasso procedure in function 'glmnet' from the package 'glmnet' (Friedman et al., 2010) to extract those of the 20 trait variables that contributed the most in the four linear models, Akaike Information Criterion (AIC) was used to identify the most parsimonious model for each of the four response variables. We allowed for a maximum of three terms to be included in the respective models, including two-fold interactions. Interaction terms are indicated by '*'. Computed Variance inflation factor (VIF) for each predictor variable was low, indicating that the predictors were not correlated with each other.

Response variable	Multiple R^2; *p*-value	Predictor terms entered in model	Regression coefficient	*p*-value	VIF
Geographic range size	0.090; < 0.001				
		SLA (m^2/kg)	0.213	< 0.001	1.015
		Plant height (m)	0.140	< 0.01	1.014
		Leaf N:P ratio (g/g)	-0.140	< 0.01	1.020
Climatic niche size	0.069; < 0.001				
		SLA (m^2/kg)	0.181	< 0.001	1.063
		Leaf C (mg/g)	-0.126	< 0.01	1.043
		Therophyte	0.108	0.096	1.024
Mean cover	0.211; < 0.001				
		SLA (m^2/kg)	2.146	< 0.001	1.066
		Leaf area (mm^2) * SLA (m^2/kg)	1.722	< 0.001	1.041
		Leaf area (mm^2)	1.417	< 0.001	1.076
Skewness of cover values	0.169; < 0.001				
		SLA (m^2/kg)	0.033	< 0.001	1.066
		Leaf area (mm^2) * SLA (m^2/kg)	0.030	< 0.001	1.041
		Leaf area (mm^2)	0.025	< 0.001	1.076

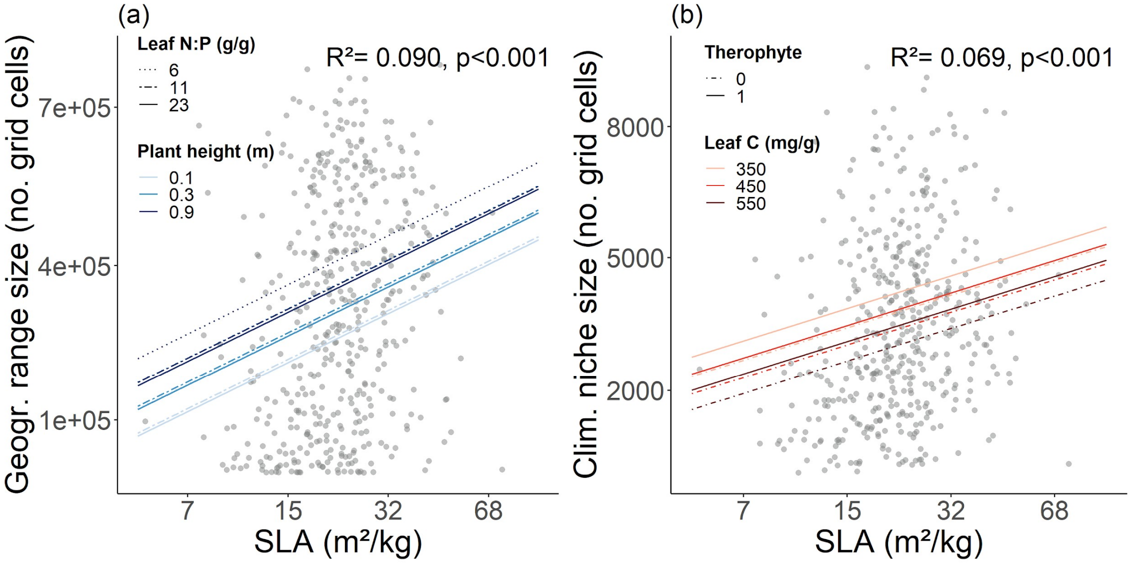

Figure IV.4 Scatter plot of observed values and regression lines from phylogenetic generalized least squares models, showing the effects of the three most predictive terms on species' (a) geographic range size and (b) climatic niche size. For geographic range size, coloured and dashed lines represent the 5th, 50th and 95th percentile in values for plant height and leaf N:P ratio, respectively; for climatic niche size, coloured lines, solid and dotted, represent the 5th, 50th and 95th percentile in leaf C content values for therophytic and non-therophytic species, respectively.

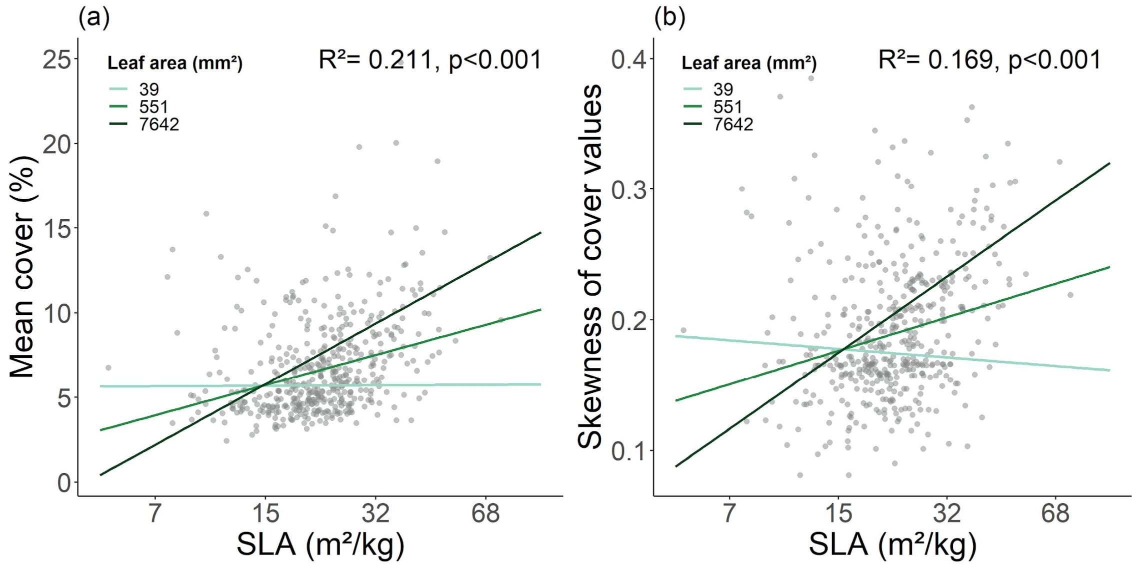

Figure IV.5

Scatter plot of observed values and regression lines from phylogenetic generalized least squares models, showing the effects of the relevant species traits on species' local abundance. Plots show the effect of the interaction between leaf area and SLA on (c) mean cover and (d) skewness of cover values. Coloured lines represent the 5th, 50th and 95th percentile in values for leaf area.

Discussion

Our main finding was that local species abundance was better predicted by plant traits than broad-scale distribution metrics. This indicates that plant traits to better capture local processes acting at the community level, such as biotic processes, than to capture broad-scale macroecological processes. Our results largely confirm trends previously reported about the existing association between species geographic range and climatic niche size, with widely distributed species also having broad climatic tolerances and geographically narrowly distributed species also narrowly distributed in climatic space (Gaston, 2000; Slatyer et al., 2013; Cardillo et al., 2019). We found an overall right-skewed distribution in cover values for most of the studied species, with species exhibiting low cover at most sites and high cover in only a few sites across their distribution range. The species mean cover values were positively related to the skewness of cover values. Therefore, for our species set, we consider the measure of skewness, calculated as area under the histogram (AUH), a robust tool to capture both, the mean and the variability of cover values across a species' whole distribution range.

Specific leaf area (SLA) had a significant positive effect and explained most of the observed variation in all four models predicting species local abundance and broad-scale distribution. In line with our findings, several studies state common species to be associated with higher SLA (Grime et al., 1997; Diaz et al., 2004; Mariotte, 2014). SLA is a productivity- and competitive ability- related trait, that reflects species strategies for rapid acquisition of resources, with higher SLA values allowing a species for e.g. better light capture (Wright et al., 2004; Diaz et al., 2004; Mariotte, 2014). Our results support these findings, as we found that species with higher SLA values are more common than species with low SLA values in the local community and on a broad spatial scale. Furthermore, at the local scale, leaf area showed a significantly positive effect on species commonness. On the one hand, this is the direct consequence of how cover was estimated, as large-leaved-plants of the same size will attain higher cover values than small-leaved ones. On the other hand, like SLA, leaf area is interpreted as a trait that is positively related to species productivity and competitive ability (Wright et al., 2004; Diaz et al., 2004). This result confirms the assumption that species with larger leaves, allowing better light capture, are found to be generally more common than species with smaller leaves (Mariotte, 2014). However, leaf area was particularly important in interaction with SLA values, as local abundance was higher in species with large leaves and high SLA values than expected by the single traits alone. In contrast, the ability to grow clonally was not selected in any of the final models based on only single trait variables. The ability of species for clonal growth plays an important role both in short-distance spread and in persistence within habitats (Benot et al., 2010) and previous studies found clonality to be positively associated with local abundance (Eriksson and Jakobsson, 1998; Kolb et al., 2006). Accordingly, this trait had high axes loadings in the PCA-based model with mean cover value as response variable. However, clonal growth was not included in our final models based on original trait values, which probably is explained by a significantly negative correlation of clonal growth with SLA in our set of species.

At broad spatial scale, species distribution in geographic space (i.e. geographic range size) was positively related to species plant height (i.e. with taller species being more widespread). High stature is known to have a competitive advantage and to be associated with common species. Greater plant height of widely distributed species suggests that these species may have higher competitive ability for space and light than narrowly distributed species. A similar positive pattern between plant height and geographic range size was found in studies on Mediterranean herbaceous species by Lavergne et al. (2004) and on herbaceous forest species in Germany by Kolb et al. (2006). We found species leaf N:P ratio to be negatively correlated with geographic range size. Nitrogen (N) and phosphorus (P) availability can limit plant growth in terrestrial ecosystems, and N:P ratios are on average higher in stress-tolerant species compared to ruderals (Güsewell, 2004). Ruderal species are characterized by rapid growth and they establish much quicker and thrive better in disturbed habitats than stress-tolerant and competitor species (Grime, 1979; Wright et al., 2004; Guo et al., 2018) and generally undergo long-distance dispersal (Baker, 1965). With this, we see a plausible explanation by which low N:P ratio, typically displayed by ruderal species, had a positive effect on species geographic range size.

In our study, the distribution range in climatic space was larger in species with lower leaf carbon content, even when accounting for SLA. In general, carbon content would be expected to be negatively related to SLA (Reich, 2014), but both traits seem to explain independent proportions in climatic niche size. This was brought about by species with a broad climatic niche size, for which SLA alone was a poor predictor, such as species with a tendency to succulence (e.g. *Plantago major*), which however have leaves with low SLA and a low leaf C content. Species with a therophytic life form (i.e. annual plants that overwinter as a seed) did show a marginally positive effect towards greater climatic tolerance (i.e. broader climatic niche size). A short generation time is a selective advantage to annuals over biennials and perennials (Pysek and Richardson, 2007), with annual species being capable of faster reproduction and spread by seeds than perennials, e.g. many weed species are annual ruderal species that generally undergo long-distance dispersal (Baker, 1965). Finally, traits related to dispersal, regeneration and persistence did not show significant effect on species local commonness or broad-scale distribution in our models. These results confirm previous studies that found no significant relationship between species commonness with seed mass (Leishman & Murray, 2001; Kolb et al., 2006; Lavergne et al., 2004) or life form (Oakwood et al., 1993). Our findings show that especially traits related to the leaf economic spectrum, especially SLA, affect both broad-scale distribution and local abundances.

Nevertheless, single plant traits and trait syndromes only weakly explained the total observed variation in species' broad-scale distribution metrics. Our study was carried out on the species' whole Eurasian distribution range, and with this comprises plenty of different habitat types and different bioclimatic zones. Species functional traits are expected to be related to those environmental conditions under which the species occurs (Lavorel & Garnier, 2002; McGill et al., 2006). Thus, a species can be common at one site while being rare in another, depending on the specific local environmental conditions (Wellstein et al. 2013; Mariotte, 2014). For example, Aerts & Chapin (2000) found fast-growing species from nutrient-rich habitats to have a combination of high SLA, high tissue N concentrations and high rates of C and N uptake, while the opposite direction for those traits characterized species from nutrient-poor habitats. In consequence, the association between plant species' functional traits and both local abundance and broad-scale distribution is highly context-dependent, differing between habitats, vegetation types and geographical regions (Murray et al., 2002; Kolb et al., 2006).

Several studies have provided evidence of significant intraspecific trait variation and global variability in leaf traits (Reich and Oleksyn, 2004; Albert et al., 2011; Moles et al., 2014; Niinemets, 2015; Wright et al., 2017). Therefore, we encourage future studies to include intraspecific trait variation in addition to mean values for species traits when investigating studies over broad geographic scales. Furthermore, to confirm the key role of context-dependency, future studies should incorporate habitat variability, by means of comparisons among single habitat types (e.g. by applying EUNIS habitat classification; Chytry et al., 2020), and test for consistency of the role of traits for patterns of species commonness at different spatial scales.

Appendices

Appendix S IV.1 Information on the 456 species included in this study. This Appendix is available online from
Figshare Digital Repository: https://figshare.com/s/e61c08fea258e420f309

Table S IV.2.1 Correlation matrix including the correlation coefficients of all trait combinations; significant correlations (p < 0.01) are highlighted in bold.

	LeafArea	SLA	LeafC	LeafN	LeafP	Plant Height	Seed Mass	LDMC	LeafNP ratio	Seed. num.rep. unit
LeafArea		**0.168**	**-0.152**	**0.286**	**0.349**	**0.526**	**0.291**	**-0.278**	**-0.129**	**0.261**
SLA	**0.168**		**-0.122**	**0.462**	**0.401**	0.054	0.037	**-0.456**	**-0.124**	0.024
LeafC	**-0.152**	**-0.122**		**-0.115**	**-0.174**	-0.114	0.029	**0.346**	0.059	**-0.217**
LeafN	**0.286**	**0.462**	**-0.115**		**0.477**	**0.242**	**0.315**	**-0.323**	**0.171**	0.121
LeafP	**0.349**	**0.401**	**-0.174**	**0.477**		**0.139**	0.062	**-0.335**	**-0.596**	0.150
PlantHeight	**0.526**	0.054	**-0.114**	**0.242**	**0.139**		**0.299**	0.043	0.108	**0.348**
SeedMass	**0.291**	0.037	0.029	**0.315**	0.062	**0.299**		0.062	**0.208**	**-0.359**
LDMC	**-0.278**	**-0.456**	**0.346**	**-0.323**	**-0.335**	0.043	0.062		0.108	**-0.230**
LeafNPratio	**-0.129**	**-0.124**	0.059	**0.171**	**-0.596**	0.108	**0.208**	0.108		-0.070
Seed.num. rep. unit	**0.261**	0.024	**-0.217**	0.121	**0.150**	**0.348**	**-0.359**	**-0.230**	-0.070	
Disp.unit. length	**0.358**	-0.081	-0.074	0.116	0.003	**0.342**	**0.704**	0.071	0.202	-0.162
annual	-0.112	**0.160**	-0.042	**0.128**	0.043	-0.085	0.078	**-0.150**	**0.002**	0.012
biennial	0.064	0.000	-0.086	0.057	0.045	0.103	-0.019	-0.110	-0.010	0.155
perennial	0.100	**-0.162**	0.049	**-0.148**	-0.098	0.066	-0.067	**0.162**	0.052	-0.047
Clonal Growth	0.092	**-0.124**	0.010	-0.053	-0.002	0.073	-0.065	0.087	0.019	0.000
Phaneroph	-0.054	**-0.158**	**0.130**	-0.096	**-0.147**	0.013	0.006	**0.193**	0.047	-0.107
Chamaeph	**-0.128**	0.006	-0.036	-0.038	0.013	-0.115	-0.029	-0.104	-0.071	-0.010
Hemicrypt	0.061	0.023	-0.041	0.018	0.053	0.050	-0.005	0.021	-0.012	0.091
Geoph	0.078	0.085	0.010	0.024	-0.004	0.016	-0.006	-0.058	0.026	-0.059
Theroph	-0.128	0.148	-0.026	0.119	0.049	-0.103	0.072	-0.137	-0.018	0.008

Table S IV.2.1 (continuing) Correlation matrix including the correlation coefficients of all trait combinations; significant correlations ($p < 0.01$) are highlighted in bold.

	Disp.unit.length	annual	biennial	perennial	Clonal Growth	Phaneroph.	Chamaeph.	Hemicrypt.	Geoph.	Theroph.
LeafArea	**0.358**	-0.112	0.064	0.100	0.092	-0.054	-0.128	0.061	0.078	-0.128
SLA	-0.081	0.160	0.000	-0.162	-0.124	-0.158	0.006	0.023	0.085	0.148
LeafC	-0.074	-0.042	-0.086	0.049	0.010	0.130	-0.036	-0.041	0.010	-0.026
LeafN	0.116	0.128	0.057	-0.148	-0.053	-0.096	-0.038	0.018	0.024	0.119
LeafP	0.003	0.043	0.045	-0.098	-0.002	-0.147	0.013	0.053	-0.004	0.049
PlantHeight	**0.342**	-0.085	0.103	0.066	0.073	0.013	-0.115	0.050	0.016	-0.103
SeedMass	**0.704**	0.078	-0.019	-0.067	-0.065	0.006	-0.029	-0.005	-0.006	0.072
LDMC	0.071	-0.150	-0.110	0.162	0.087	**0.193**	-0.104	0.021	-0.058	-0.137
LeafNPratio	**0.202**	0.002	-0.010	0.052	0.019	0.047	-0.071	-0.012	0.026	-0.018
Seed.num. rep. unit	-0.162	0.012	0.155	-0.047	0.000	-0.107	-0.010	0.091	-0.059	0.008
Disp.unit. length		0.022	0.009	-0.007	-0.028	-0.035	-0.076	0.060	-0.022	0.014
annual	0.022		0.084	**-0.843**	**-0.737**	-0.110	-0.084	-0.053	-0.151	0.959
biennial	0.009	0.084		**-0.249**	-0.100	-0.081	-0.081	**0.173**	-0.112	0.055
perennial	-0.007	**-0.843**	**-0.249**		**0.665**	0.115	0.115	0.032	0.159	**-0.866**
Clonal Growth	-0.028	**-0.737**	-0.100	**0.665**		0.001	0.089	-0.140	0.123	**-0.770**
Phaneroph	-0.035	-0.110	-0.081	0.115	0.001		0.016	**-0.467**	-0.087	-0.105
Chamaeph	-0.076	-0.084	-0.081	0.115	0.089	0.016		**-0.356**	-0.087	-0.105
Hemicrypt	0.060	-0.053	**0.173**	0.032	-0.140	**-0.467**	**-0.356**		**-0.409**	-0.071
Geoph	-0.022	-0.151	-0.112	0.159	0.123	-0.087	-0.087	**-0.409**		-0.145
Theroph	0.014	**0.959**	0.055	**-0.866**	**-0.770**	-0.105	-0.105	-0.071	-0.145	

Table S IV.2.2 Abbreviation of trait names and values of trait contributions (loadings) to the first (Dim1) to twentieth (Dim20) PCA axes.

Abbreviation	Dim1	Dim2	Dim3	Dim4	Dim5	Dim6	Dim7	Dim8	Dim9	Dim10	Dim11	Dim12	Dim13	Dim14	Dim15	Dim16	Dim17	Dim18	Dim19	Dim.20
LeafArea	0.69	19.21	0.61	0.17	4.00	2.17	0.17	2.66	0.52	0.02	14.23	3.23	0.04	7.35	43.91	0.03	0.30	0.00	0.69	0.00
SLA	4.72	6.48	2.58	4.76	3.17	2.65	2.00	11.04	0.00	0.15	6.80	24.31	6.92	21.30	2.49	0.07	0.02	0.00	0.52	0.02
LeafC	1.95	4.88	1.82	0.02	0.11	5.71	6.89	18.49	0.92	30.93	22.28	3.93	0.03	0.24	1.63	0.16	0.02	0.01	0.00	0.00
LeafN	4.14	10.81	0.43	2.86	0.05	3.81	0.03	23.18	0.01	2.37	15.34	3.79	5.54	2.04	1.13	0.24	0.03	0.53	23.66	0.00
LeafP	3.12	11.68	4.22	0.48	3.99	17.64	0.21	1.13	0.63	0.26	7.47	1.16	2.41	0.23	3.47	6.54	0.11	0.36	34.86	0.02
PlantHeight	0.10	9.55	6.37	0.70	19.53	1.38	0.02	0.00	1.38	1.68	0.05	24.20	6.09	10.80	16.06	1.39	0.00	0.03	0.63	0.04
SeedMass	0.97	3.36	27.48	1.38	5.18	0.44	0.58	0.19	1.09	0.02	0.05	0.10	0.15	0.01	1.29	54.24	0.94	0.01	2.52	0.00
LDMC	3.84	4.39	8.63	3.53	0.21	7.46	2.00	0.18	0.04	5.79	24.29	9.98	0.01	4.78	24.13	0.15	0.01	0.01	0.57	0.01
LeafNPratio	0.28	0.82	10.66	0.19	3.49	41.47	0.49	4.20	1.14	0.10	0.05	0.77	0.28	0.45	0.22	3.69	0.04	0.22	31.43	0.00
Seed.num.rep.unit	0.33	2.78	7.81	4.66	27.27	0.07	0.54	0.09	6.31	8.87	0.28	4.79	1.54	23.37	0.97	9.57	0.22	0.05	0.48	0.00
Disp.unit.length	0.49	4.12	25.14	0.01	1.17	0.37	1.54	6.79	0.04	0.00	1.51	4.04	0.14	26.55	2.09	22.22	0.45	0.00	3.33	0.00
annual	20.14	4.89	0.24	0.36	0.25	0.05	0.05	0.38	0.56	0.24	0.07	0.42	2.18	0.21	0.35	0.34	17.70	20.79	0.36	30.43
biennial	1.28	0.14	0.22	8.21	4.15	0.95	4.90	0.32	75.25	0.07	0.07	0.43	0.14	0.13	0.41	0.09	3.08	0.01	0.00	0.15
perennial	19.40	3.65	0.02	0.00	0.73	0.14	0.00	0.03	1.64	0.41	0.72	0.85	6.29	0.01	0.01	0.67	64.27	0.02	0.22	0.90
ClonalGrowth	14.93	4.40	0.44	0.22	0.47	0.41	0.09	0.16	1.15	0.20	4.00	9.75	42.38	0.07	0.00	0.07	0.00	18.80	0.15	2.30
Phaneroph.	1.63	1.74	1.41	6.86	13.13	10.84	2.84	14.39	0.04	24.52	0.85	0.01	4.77	1.44	0.01	0.06	1.15	13.44	0.06	0.81
Chamaeph.	0.51	0.12	1.61	11.99	1.63	0.00	40.86	2.40	0.09	22.47	0.55	5.00	3.26	0.48	1.06	0.10	1.19	6.09	0.03	0.55
Hemicrypt.	0.41	0.94	0.00	40.58	9.08	1.28	0.18	0.99	1.99	0.15	1.14	1.80	6.89	0.39	0.19	0.10	2.95	28.56	0.36	2.01
Geoph.	0.80	0.54	0.09	12.60	2.13	2.93	36.34	12.97	6.31	1.51	0.10	0.91	9.69	0.05	0.35	0.03	1.12	10.86	0.14	0.53
Theroph.	20.29	5.50	0.22	0.41	0.25	0.23	0.25	0.42	0.87	0.22	0.17	0.54	1.27	0.10	0.21	0.23	6.40	0.22	0.00	62.21

Table S IV.2.3 Test for phylogenetic signal in the trait variables. We applied Pagel's Lambda statistic (Pagel, 1999) and Fritz and Purvis's *D* (2010) statistics to quantify the strength of a phylogenetic signal in the 20 studied trait variables. Pagel's Lambda statistic ranges from 0 for traits being phylogenetically unrelated to 1 for traits following trait evolution according to a Brownian motion (BM) model; Pagel 1999; Freckleton, Harvey and Pagel, 2002). Fritz and Purvis's *D* is equal to 0 if the observed binary trait has a phylogenetically random distribution across the tips of the phylogeny; *D* is equal 1 if the observed trait is as clumped as if it had evolved by Brownian motion. Pagel's Lambda statistic revealed a strong phylogenetic signal in all continuous trait variables. Fritz and Purvis's *D* revealed a phylogenetic signal in all binary trait variables.

Trait	**Lambda**	***p*-value**
Leaf area	0.911	< 0.001
Specific leaf area	0.811	< 0.001
Leaf C content	0.897	< 0.001
Leaf N content	0.835	< 0.001
Leaf P content	0.729	< 0.001
Leaf dry matter content	0.890	< 0.001
Leaf N/P ratio	0.809	< 0.001
Plant height	0.906	< 0.001
Seed mass	0.984	< 0.001
Seed number per reproductive unit	0.773	< 0.001
Dispersal unit length	0.975	< 0.001

	Estimated D	***p*-value (D > 0)**	***p*-value (D < 1)**
Annual	0.746	< 0.001	< 0.001
Biennial	0.829	< 0.001	< 0.051
Perennial	0.804	< 0.001	< 0.001
Phanerophyte	0.448	< 0.001	< 0.05
Chamaephyte	0.816	< 0.001	< 0.05
Hemicryptophyte	0.638	< 0.001	< 0.001
Geophyte	0.893	n.s.	< 0.001
Therophyte	0.772	< 0.001	< 0.001
Clonal Growth	0.802	< 0.001	< 0.001

Table S IV.2.4 The lasso procedure in function ‘glmnet’ from the package ‘glmnet’ (Friedman et al., 2010) was applied to extract those of the 20 trait variables that contributed the most in the four linear models (i.e. mean cover values, skewness of cover values, geographic range size, and climatic niche size being response variable). The minimum lambda ratio (s) in the models was 0.01 (mean cover values), 0.0005 (skewness of cover values), 0.0001 (geographic range size) and 0.01 (climatic niche size).

Response variable	**Mean cover**		**Skewness of cover values**		**Geographic range size**		**Climatic niche size**	
Multiple R²; *p*-value	0.245; < 0.001		0.209; < 0.001		0.112; < 0.001		0.113; < 0.001	
Trait (Abbreviation)	Regression coefficient	*p*-value	Regression coefficient	*p*-value	Regression coefficient	*p*-value	Regression coefficient	*p*-value
LeafArea	0.531	0.160	0.011	0.096	-0.108	0.097	0.061	0.222
SLA	1.976	< 0.001	0.032	< 0.001	0.209	< 0.001	0.052	< 0.001
LeafC	0.284	0.336	0.008	0.147	-0.034	0.503	0.050	< 0.001
LeafN					0.107	0.209		
LeafP	0.205	0.614			-0.158	0.107	0.071	0.002
LDMC	-0.738	0.066	-0.012	0.099	-0.023	0.707		
LeafNPratio	-0.439	0.259	-0.011	0.031	-0.279	0.002	0.065	< 0.001
PlantHeight	1.013	0.003	0.012	0.042	0.214	< 0.001	0.055	0.002
SeedMass	-0.804	0.129	-0.017	0.092	-0.018	0.829		
SeedNumRepUnit	-0.277	0.388	-0.007	0.203	0.003	0.959		
DispUnitLength	1.010	0.047	0.027	0.004	-0.006	0.939		
annual	1.159	0.322			-0.065	0.770		
biennial	-0.227	0.618	-0.001	0.862	-0.076	0.382		
perennial	-0.252	0.715	0.001	0.937	-0.072	0.586	0.117	0.577
Phaneroph.	0.839	0.320	0.022	0.078	0.185	0.228	0.101	0.453
Chamaeph.	-0.736	0.295	-0.004	0.742	0.109	0.406		
Hemicrypt.	-0.640	0.296	-0.005	0.523	0.142	0.207		
Geoph.	-0.348	0.553			0.172	0.116	0.075	0.211
Theroph.	-1.812	0.231	-0.001	0.946	0.146	0.607	0.124	0.706
ClonalGrowth	0.392	0.617	0.017	0.188	0.131	0.361		

Chapter V

Synthesis

In this thesis, I investigated abundance distribution patterns of vascular plant species and the effect of functional traits on species local abundance and broad-scale distribution. I analysed the relationships for a large group of herbaceous, dwarf shrub and shrub species over their full Eurasian distribution ranges and included trait syndromes in addition to single traits. The following results are based on analyses that included information on the geographic range size for 564 plant species from the Chorological Database Halle (CDH, E. Welk et al., unpublished data), 808,794 vegetation plots provided by the European Vegetation Archive (EVA, Chytrý et al., 2016) and a complete species-trait-matrix of 20 plant functional traits that was compiled from three trait databases that store information on species leaf and seed traits (TRY, Kattge et al., 2020), species life form and life span (BiolFlor, Kühn et al., 2004) and clonality (CLO-PLA, Klimešová et al., 2017).

Summary of Results

In **chapter II,** I proposed and tested a box-counting method to assess **the sampling coverage of species distribution in biodiversity databases in geographic and climatic space across different spatial resolution**. First, I found a positive relationship between species geographic range size and climatic niche size. Second, the sampling coverage of European vascular plant species ranges by EVA vegetation plots was more complete within the geographic space than within the climatic space. Third, I detected a positive correlation between the observed sampling coverage and the expected sampling coverage, based on null models, for both the geographic space and the climatic space. Importantly, a large majority (92.0%) of the observed species distributions in EVA were significantly underdispersed in both the geographic and climatic space. Fourth, I found a positive effect of sample size (number of vegetation plots in which a species occurred) on sampling coverage while accounting for range size or niche size in both the geographic space and climatic space.

In **chapter III,** with respect to the **abundance-range size relationship**, I found that species local abundance was significantly, yet weakly, related to range size but not to niche size, with high intraspecific variation in species abundance values. Regarding both, the **abundance-range centre relationship** and **the abundance-suitability relationship,** I found large variation in regression models relating abundance to the distance from the centre of the species' geographic ranges and climatic niches and the predicted climatic suitability of the plot location, with species showing positive, negative or no relationship. Summarized in a mixed-effect meta-analysis, the overall mean slope was slightly negative for the two distance measures and for all applied SDM methods in the applied regression models. This indicated that both, the distance to the centre of the species' geographic range and the predicted climatic suitability of the plot location are poor predictors of local abundance.

In **chapter IV,** I investigated whether species **local abundance** and **broad-scale distribution** can be predicted by single **functional traits** and sets of traits (trait syndromes). First, all plant distribution and abundance variables were related to functional traits. Second, I found local abundances to be much more strongly related to traits than to geographic distributions. While geographic range size increased with plant height, climatic niche size decreased with leaf C content. The species' mean cover was positively related to the species' skewness of cover values and both increased with leaf area and specific leaf area (SLA). Both geographic range size and climatic niche size increased with SLA. For each of the four response variables, sets of traits (trait syndromes) explained less variation than single functional traits.

Discussion

Brown, Stevens, & Kaufman (1996) pointed out that distribution maps of species geographic ranges are simplifications of the complex spatial and temporal distribution patterns of species on earth. Furthermore, they highlight that range maps suffer from problems of lacking precision, accuracy and completeness. Polygon range maps potentially encompass a large number of localities where a species does not really occur. One might argue that this might also be the case for range maps used in this thesis and provided by the CDH. Although range maps stored in the CDH are based on expert-drawn range maps, they are constantly updated with information from national and floristic databases and maps from floristic literature and information on species occurrences from online databases like GBIF. Nevertheless, I am aware that the range maps from CDH also include geographical areas that might be unoccupied by the species. Due to advances in phylogenetic systematics through molecular studies, the variation at the molecular level has resulted in splitting of species into multiple new species based on distinctive molecular genetic characteristics in the past (Brown et al., 1996). CDH range maps are therefore constantly updated regarding species taxonomy. Besides the completeness of species distribution maps, for this study on broad-scale abundance patterns the quality of sampling coverage of plots from EVA over the entire range of species was crucial. Although the observed species distributions in EVA were significantly underdispersed in both the geographic and climatic space, the correlation between the observed sampling coverage and the expected sampling coverage, based on null models, was positive (see chapter II). Furthermore, both CDH and EVA data were processed to a coarse-grain raster layer of 2.5 min resolution, which corresponds to grid cells covering approximately 15km² each across Central Europe. This coarse-grain resolution did reduce the influence of possible geographic location uncertainty at fine-grain resolution.

The positive relationship that I found in chapter II between species geographic range size and climatic niche size confirms the assumption that the shapes of a species' range and the range boundaries reflect the influence of limiting climatic conditions (Brown et al., 1996). That these climatic factors also account for the pattern of variation in abundance of species among sites within their range is the theoretical explanation for the **abundance-range size relationship** (Brown et al., 1996). In addition, the **abundance-range centre relationship** and **the abundance-suitability relationship** assume that species reach highest abundances where the environmental conditions are assumed to be most suitable. As a consequence, species abundance should be related to range position, and in case of the **abundance-range centre relationship**, the favourable conditions should be met towards the centre of the species range. Nevertheless, in this thesis, I found no clear support for any of the three macroecological hypotheses (chapter III) and there are several plausible explanations for these results.

First, there are different abiotic and biotic factors that limit range expansion along different margins of species ranges in geographic space. While biotic interactions tend to limit species distribution and abundance at lower latitudes, abiotic factors are more likely to be limiting at higher latitudes (MacArthur, 1984; Normand et al., 2009; Greiser et al., 2020) or towards higher elevation (Bruelheide & Scheidel, 1999; Rumpf et al., 2018). This means that increasing physical stress limits species distributions in one direction along the gradient, while in the other direction they are limited by e.g. an increasing number of competitors and suffer from increasing impacts of competition (Louthan, Doak & Angert, 2015). Furthermore, geographic barriers such as coastlines or mountains can impede further range expansion in otherwise climatically suitable habitats, and in consequence, may lead to marginal range areas that are highly climatically suitable. Therefore, range boundaries and abundance towards the range edges can be determined by multiple abiotic and biotic factors that interact in complex ways and do not strictly follow solely climatic gradients (Sagarin et al., 2006).

Second, the environmental factors that drive species distribution at broad geographic extents may differ from those that influence species performance and abundance on the local scale, and consequently, the pattern of abundance relationships with range characteristics is blurred. In general, species broad-scale distribution, measured on the regional or the continental scale, is shaped by coarse-scale bioclimatic conditions and geographic barriers that affect range dynamics and long-distance dispersal (Wiens, 1989; Baselga, Lobo, Svenning, & Araújo, 2012). In contrast, species abundance, that is measured on the small local scale of a few square meters, depends on factors that are operating at the local scale (Bruelheide et al., 2018) such as habitat structure (e.g. soil or microclimatic conditions, De Frenne et al., 2013), prevailing disturbance regimes or successional stages and biotic interactions (e.g. competition, Moeslund et al., 2017).

Third, species range size on the broad geographic scale and abundance at the local scale are measured on very different temporal scales. Species abundance, at different sites throughout its geographic range, is only reflected at any one point in time by a vegetation record and thus, varies between and even within years. A species might be locally abundant at specific times but not others (Sagarin et al., 2006). In contrast, species broad-scale geographic ranges are driven by long-term average conditions. Although species range boundaries are dynamic, range shifts, expansions or contractions, are, in most cases, the result of long-term environmental changes. For example, in Europe and North America, northern range boundaries within the last 10,000 years tracked the retreat of the last continental ice sheets (Brown, 1995). Further range shifts were caused by human activities mainly during the last two centuries (Mack et al., 2000). Murray and Lepschi (2004) conclude that if distributions of species abundance are variable in time, it is unlikely that higher-order processes, such as gene flow between populations or responses of ranges to climate change can be predicted based on a single description of a distribution of abundance.

The results of this thesis, with no unequivocal support for any of the three macroecological hypotheses, are in line with previous empirical studies (e.g. Gaston et al., 1997; Pironon et al., 2017; Santini et al., 2019; Dallas, Pironon, & Santini, 2020). The tested general macroecological abundance patterns are probably only valid when local abundance is measured by species individuals' density but not when measured by species growth performance (i.e. cover values). This indicated that, across entire species ranges, the distribution of abundance is highly heterogenous, because local drivers strongly influence plant species growth performance across their global range. Therefore, I agree with the statement by Sagarin et al. (2006) that even in systems, where entire species ranges can be studied, the distribution of abundance involves a high complexity.

The *jack-of-all-trades-master-of-none hypothesis* states that habitat specialists can have higher abundance in their preferred habitat than habitat generalists, i.e. species able to occupy a wide range of environmental conditions (Devictor et al., 2010). This hypothesis reflects the assumption of Rabinowitz (1981), that there are habitat specialists with small geographic ranges, yet they dominate the community they are found in.

As plant species' **functional traits** are assumed to reflect the mechanisms through which species respond to abiotic and biotic conditions to maximise their fitness (Suding et al., 2008), I was interested to investigate whether and which traits or sets of traits can explain patterns in species local abundance and broad-scale distribution. In general, sets of traits (trait syndromes) explained less variation than single functional traits. I found local abundances to be much more strongly related to traits than to geographic distributions, indicating that functional traits better capture processes acting at the local community level than broad-scale processes. Variation in species broad-scale distribution was e.g. explained by leaf nitrogen (N) and phosphorus (P) concentration. N:P ratios are on average lower in ruderal species which also establish quicker and perform better in disturbed habitats than stress-tolerant or competitor species (Grime, 1979; Guo et al., 2018). Therefore, it seems plausible that low N:P ratios had a positive effect on species broad-scale distribution. Traits related to dispersal and regeneration (e.g. seed size and weight) and persistence (e.g. life span) did not show any significant effect on species local abundance or broad-scale distribution, which is in line with findings from previous studies (Lavergne et al., 2004; Oakwood et al., 1993). Confirming these results, Fried et al. (2020) found no direct relationship between traits related to dispersal and colonization ability with broad-scale (national) distribution in arable weeds.

Although I found no clear relationship of species patterns of local abundance with broad-scale distribution (chapter II), it is even more interesting that specific leaf area, a trait related to the leaf economics spectrum, did have a positive effect on both the species' local abundance and broad-scale distribution. SLA is interpreted as a trait that is positively related to species productivity and competitive ability, with higher SLA values allowing a better light capture (Wright et al., 2004). Several studies found more abundant species to be associated with higher SLA (Grime, 1997; Mariotte, 2014). Moreover, SLA tends to be higher in ruderal species that are adapted to frequent disturbance especially in productive habitats (Guo et al., 2018; Fried et al., 2020).

I see two plausible explanations for the partially contradicting or missing results found in previous studies (see Table IV.1) and in this thesis on the relationship between traits with local abundance and broad-scale distribution.

First, the association between functional traits and both local abundance and broad-scale distribution is highly context-dependent (Murray et al., 2002). Theory suggests that temperature and precipitation are major determinants of plant traits at a global scale (Moles et al., 2014). Wright et al. (2017) characterized a worldwide pattern in leaf size, with leaf size to be on average larger in equatorial regions and smaller toward the poles and higher elevations. Reich and Oleksyn (2004) found a similar pattern for plant leaf N and P contents, with leaf N and P to decline and the N:P ratio to increase toward the equator as average temperature and growing season length increase. As I stated above, a plausible explanation for a missing abundance pattern following broad-scale climatic gradients are other factors than macroclimate that operate at the local scale and influence species local abundance, such as microclimatic conditions or biotic interactions. Thus, a species can be common at one site within its geographic range, while being rare in another in close proximity, simply because of differing specific local environmental conditions (Mariotte, 2014). As a species' functional traits are expected to be related to the environmental conditions under which the species occurs, rare and common species might be characterized by the same traits in different habitats, vegetation types or geographic regions, depending on the local conditions (Aerts & Chapin, 2000).

Second, functional traits express not only species-specific characteristics, but also intraspecific variability (Albert, Grassein, Schurr, Vieilledent, & Violle, 2011; Niinemets, 2015). Intraspecific variation in functional traits reflects phenotypic plasticity, and this variation influences plant responses to abiotic conditions and biotic interactions (Fridley & Grime, 2010). Studies detected intraspecific trait variation e.g. in SLA values and N:P ratios along elevation gradients (Kichenin, Wardle, Peltzer, & Freschet, 2013) or in SLA values and height along flooding gradients in flood meadows (Jung, Violle, Mondy, Hoffmann, & Muller, 2010). In a global meta-analysis, Siefert et al. (2015) found intraspecific trait variation to account for 25% of the total trait variation within plant communities. This intraspecific trait variation may influence the interactions among and between species and their environment and, therefore, might influence species performance and plant species community assembly (Bolnick et al., 2003; Siefert et al., 2015).

Conclusion and Future perspectives

Although I found no unequivocal support for any of the three macroecological hypotheses tested, this is an important finding and highlights the complexity of factors that determine species abundance throughout their geographic range. This complexity can strongly influence predictions about habitat conservation (Hampe & Petit, 2005) and species responses to climate change (Helmuth, Kingsolver, & Carrington, 2005), which previously were based on assumptions from Species Distribution Models (SDMs). SDMs are applied to identify places that are suitable for the survival of populations of a species by identifying their environmental requirements (Soberon & Nakamura, 2009). Since the importance of local environmental heterogeneity that influences species local abundance has been highlighted in several studies (see e.g. De Frenne et al., 2013; Köckemann et al., 2009), I advocate for including information on microclimatic conditions that are available at fine spatial resolution (e.g. soil temperature, Lembrechts et al., 2020) in addition to commonly used bioclimatic variables to improve predictions of species local abundance based on broad-scale occurrence data.

Furthermore, since some studies have detected abundance patterns at regional scales (Fried et al., 2020) and pointed out the importance of habitat variability that influences species local abundance (Murray et al., 2002), it seems promising to include this in between regional scale, e.g. by applying habitat classifications (EUNIS, Chytrý et al., 2020), to investigate the existence of abundance patterns at different spatial scales. I found that especially functional traits related to the leaf economics spectrum affect species local abundance and broad-scale distribution. As evidence has been provided of global variability in leaf traits, it seems straightforward to include intraspecific trait variation in addition to mean trait values in future studies to gain more specific insights when investigating abundance patterns over broad geographic scales. However, this would require recording traits with abundance values on the same individuals. Finally, as a species' abundance is variable also temporally, it would be highly interesting to investigate whether patterns of temporal change in abundance can be found in species characterized by their functional traits by resurvey analyses from vegetation surveys.

Literature Cited

Abeli, T., Gentili, R., Mondoni, A., Orsenigo, S., & Rossi, G. (2014). Effects of marginality on plant population performance. *Journal of Biogeography*, 41, 239–249.

Aerts, R. & Chapin III, F.S. (2000). The mineral nutrition of wild plants revisited: a re-evaluation of processes and patterns. *Advances in Ecological Research*, 30, 1–67.

Albert, C. H., Grassein, F., Schurr, F. M., Vieilledent, G., & Violle, C. (2011). When and how should intraspecific variability be considered in trait-based plant ecology? *Perspectives in Plant Ecology, Evolution and Systematics*, 13(3), 217–225.

Andrewartha, H. G. & Birch, L. C. (1954). The distribution and abundance of animals. Chicago, IL, USA: University of Chicago Press.

Araújo, M. B. & Guisan, A. (2006). Five (or so) challenges for species distribution modelling. *Journal of Biogeography*, 33, 1677–1688.

Araújo, M. B., Williams, P. H., & Fuller, R. J. (2002). Dynamics of extinction and the selection of nature reserves. *Proceedings of the Royal Society of London. Series B: Biological Sciences*, 269, 1971–1980.

Austin, M. (1987). Models for the analysis of species' response to environmental gradients. *Vegetatio*, 69, 35–45.

Baker, H.G. (1965). Characteristics and modes of origin of weeds. In: Baker HG, Stebbins GL (eds) The genetics of colonizing species. Academic Press, New York, pp 147–172.

Barbet-Massin, M., Jiguet, F., Albert, C. H., & Thuiller, W. (2012). Selecting pseudo-absences for species distribution models: how, where and how many? *Methods in Ecology and Evolution*, 3, 327–338.

Barton, K. (2019). MuMIn: Multi-Model Inference. R package version 1.43.15. [cited November 20, 2020]. Available at: https://CRAN.R-project.org/package=MuMIn

Baselga, A., Lobo, J. M., Svenning, J. C., & Araújo, M. B. (2012). Global patterns in the shape of species geographical ranges reveal range determinants. *Journal of Biogeography*, 39, 760–771.

Beck, J., Ballesteros-Mejia, L., Buchmann, C. M., Dengler, J., Fritz, S. A., Gruber, B., ..., Dormann, C. F. (2012). What's on the horizon for macroecology? *Ecography*, 35, 673–683.

Bedia, J., Herrera, S., & Gutiérrez, J. M. (2013). Dangers of using global bioclimatic datasets for ecological niche modeling. Limitations for future climate projections. *Global and Planetary Change*, 107, 1–12.

Benot, M. L., Bittebiere, A. K., Ernoult, A., Clement, B., & Mony, C. (2013). Fine-scale spatial patterns in grassland communities depend on species clonal dispersal ability and interactions with neighbours. *Journal of Ecology*, 101(3), 626–636.

Bertrand, R., Lenoir, J., Piedallu, C., Riofrío-Dillon, G., de Ruffray, P., Vidal, C., …, Gégout, J.-C. (2011). Changes in plant community composition lag behind climate warming in lowland forests. *Nature*, 479, 517–520.

Boakes, E. H., McGowan, P. J. K., Fuller, R. A., Chang-qing, D., Clark, N. E., O'Connor, K., & Mace, G. M. (2010). Distorted views of biodiversity: spatial and temporal bias in species occurrence data. *PLOS Biology*, 8, e1000385.

Bohner, T. & Diez, J. (2020). Extensive mismatches between species distributions and performance and their relationship to functional traits. *Ecology Letters*, 23(1), 33–44.

Bolnick, D. I., Svanbäck, R., Fordyce, J. A., Yang, L. H., Davis, J. M., Hulsey, C. D., & Forister, M. L. (2003). The ecology of individuals: incidence and implications of individual specialization. *The American Naturalist*, 161(1), 1–28.

Boots, B. N. & Getis, A. (1988). Point pattern analysis (Vol. 8). Newbury Park, CA, US: Sage Publications Inc.

Borenstein, M. (2009). Effect sizes for continuous data. In Cooper, H., Hedges, L.V. & J.C. Valentine (Eds., The handbook of research synthesis and meta-analysis (pp. 279-293). New York, NY: The Russal Sage Foundation.

Bradley, A. P. (1997). The use of the area under the ROC curve in the evaluation of machine learning algorithms. *Pattern Recognition*, 30, 1145–1159.

Bradley, B. A. (2016). Predicting abundance with presence-only models. *Landscape Ecology*, 31, 19–30.

Brändle, M. & Brandl, R. (2001). Distribution, abundance and niche breadth of birds: scale matters. *Global Ecology and Biogeography*, 10, 173–177.

Braun-Blanquet, J. (1951). Pflanzensoziologie: Grundzüge der Vegetationskunde. 2nd ed. 631 pp. Wien, Austria: Springer.

Broennimann, O. & Guisan, A. (2008). Predicting current and future biological invasions: both native and invaded ranges matter. *Biology Letters*, 4, 585–589.

Brown, J. H. (1984). On the relationship between abundance and distribution of species. *The American Naturalist*, 124, 255–279.

Brown, J. H. (1995). Macroecology. Chicago: University of Chicago Press.

Brown, J. H. & Maurer, B. A. (1987). Evolution of species assemblages: effects of energetic constraints and species dynamics on the diversification of the North American avifauna. *The American Naturalist*, 130, 1–17.

Brown, J. H., Stevens, G. C., & Kaufman, D. M. (1996). The geographic range: size, shape, boundaries, and internal structure. *Annual Review of Ecology and Systematics*, 27(1), 597–623.

Bruelheide, H., Dengler, J., Jiménez-Alfaro, B., Purschke, O., Hennekens, S., Chytrý, M., …, Zverev, A. (2019). sPlot – a new tool for global vegetation analyses. *Journal of Vegetation Science*, 30, 161–186.

Bruelheide, H., Dengler, J., Purschke, O., Lenoir, J., Jiménez-Alfaro, B., Hennekens, S. M., …, Jandt, U. (2018). Global trait–environment relationships of plant communities. *Nature Ecology & Evolution*, 2, 1906–1917.

Bütof, A., von Riedmatten, L.R., Dormann, C.F., Scherer-Lorenzen, M., Welk, E., & Bruelheide, H. (2012). The responses of grassland plants to experimentally simulated climate change depend on land use and region. *Global Change Biology*, 18, 127–137.

Cardillo, M., Dinnage, R., & McAlister, W. (2019). The relationship between environmental niche breadth and geographic range size across plant species. *Journal of Biogeography*, 46(1), 97–109.

Chamberlain, S. A. and Szöcs, E. (2013). taxize – Taxonomic search and retrieval in R. F1000 Research, 2, 191. [cited November 20, 2020]. Available at: https://doi.org/10.12688 /f1000research.2-191.v1

Chamberlain, S. A., Szöcs, E., Foster, Z., Arendsee, Z., Boettiger, C., Ram, K., Baratomeus, I., …, O'Donnell, J. (2018). taxize: Taxonomic information from around the web. R package version 0.9.3.

Chiarucci, A., Wilson, J. B., Anderson, B. J., & De Dominicis, V. (1999). Cover versus biomass as an estimate of species abundance: does it make a difference to the conclusions?. *Journal of Vegetation Science*, 10(1), 35–42.

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, 173–180.

Chytrý, M., Tichý, L., Hennekens, S. M., Knollová, I., Janssen, J. A., Rodwell, J. S., …, Hájek, M. (2020). EUNIS Habitat Classification: Expert system, characteristic species combinations and distribution maps of European habitats. *Applied Vegetation Science*. [cited November 20, 2020]. Available at: https://doi.org/10.1111/avsc.12519.

Colwell, R. K. & Rangel, T. F. (2009). Hutchinson's duality: the once and future niche. *Proceedings of the National Academy of Sciences of the United States of America*, 106, 19651–19658.

Conlisk, J., Conlisk, E., Kassim, A. R., Billick, I., & Harte, J. (2012). The shape of a species' spatial abundance distribution. *Global Ecology and Biogeography*, 21, 1167–1178.

Csergő, A. M., Salguero-Gómez, R., Broennimann, O., Coutts, S. R., Guisan, A., Angert, A. L., ..., Buckley, Y. M. (2017). Less favourable climates constrain demographic strategies in plants. *Ecology Letters*, 20, 969–980.

Dallas, T. A., & Hastings, A. (2018). Habitat suitability estimated by niche models is largely unrelated to species abundance. Global Ecology and Biogeography, 27, 1448–1456.

Dallas, T., Decker, R. R., & Hastings, A. (2017). Species are not most abundant in the centre of their geographic range or climatic niche. *Ecology Letters*, 20, 1526–1533.

Dallas, T., Pironon, S., & Santini, L. Weak support for the abundant niche-centre hypothesis in North American birds. BioRxiv [Preprint]. February 28, 2020 [cited November 20, 2020]. Available at: https://doi.org/10.1101/2020.02.27.968586

Damgaard, C. (2009). On the distribution of plant abundance data. *Ecological Informatics*, 4(2), 76–82.

Daru, B. H., Park, D. S., Primack, R. B., Willis, C. G., Barrington, D. S., Whitfeld, T. J. S., ..., Davis, C. C. (2018). Widespread sampling biases in herbaria revealed from large-scale digitization. *New Phytologist*, 217, 939–955.

De Frenne, P., Rodríguez-Sánchez, F., Coomes, D. A., Baeten, L., Verstraeten, G., Vellend, M., ... Decocq, G. M. (2013). Microclimate moderates plant responses to macroclimate warming. *Proceedings of the National Academy of Sciences*, 110, 18561–18565.

Devictor, V., Clavel, J., Julliard, R., Lavergne, S., Mouillot, D., Thuiller, W., ..., Mouquet, N. (2010). Defining and measuring ecological specialization. *Journal of Applied Ecology*, 47(1), 15–25.

Diaz, S., Hodgson, J. G., Thompson, K., Cabido, M., Cornelissen, J. H., Jalili, A., …, Band, S. R. (2004). The plant traits that drive ecosystems: evidence from three continents. *Journal of Vegetation Science*, 15(3), 295–304.

Díaz, S., Kattge, J., Cornelissen, J. H., Wright, I. J., Lavorel, S., Dray, S., …, Garnier, E. (2016). The global spectrum of plant form and function. *Nature*, 529(7585), 167.

Domin, K. (1928). The relations of the Tatra mountain vegetation to the edaphic factors of the habitat: A synecological study. *Acta Botanica Bohemica*, 6(7), 133–163.

Donoghue, M.J. (2008). A phylogenetic perspective on the distribution of plant diversity. *Proceedings of the National Academy of Sciences of the United States of America*, 105: 11549–11555.

Edwards, J. L., Lane, M. A., & Nielsen, E. S. (2000). Interoperability of biodiversity databases: Biodiversity information on every desktop. *Science*, 289, 2312–2314.

Engemann, K., Enquist, B. J., Sandel, B., Boyle, B., Jørgensen, P. M., Morueta–Holme, N., …, Svenning, J.-C. (2015). Limited sampling hampers "big data" estimation of species richness in a tropical biodiversity hotspot. *Ecology and Evolution*, 5, 807–820.

Enquist, B. J., Feng, X., Boyle, B., Maitner, B., Newman, E. A., Jørgensen, P. M., …, Couvreur, T.L. (2019). The commonness of rarity: Global and future distribution of rarity across land plants. *Science Advances*, 5, eaaz0414.

Enquist, B. J., R. Condit, B. Peet, M. Schildhauer, B. Thiers, and BIEN working group. (2009). The Botanical and Information Ecology Network (BIEN): Cyberinfrastructure for an integrated botanical information network to investigate the ecological impacts of global climate change on plant biodiversity. [cited November 20, 2020]. Available at: http:// www.iplantcollaborative.org/sites/default/files/ BIEN_White_Paper.pdf

Eriksson, O. & Jakobsson, A. (1998). Abundance, distribution and life histories of grassland plants: a comparative study of 81 species. *Journal of Ecology*, 86(6), 922–933.

Euro+Med. Euro+Med PlantBase – the information resource for Euro-Mediterranean plant diversity. [cited November 20, 2020]. Available at: http://ww2.bgbm.org/EuroPlusMed/.

Feng, D., Cortese, G., & Baumgartner, R. (2017). A comparison of confidence/credible interval methods for the area under the ROC curve for continuous diagnostic tests with small sample size. *Statistical Methods in Medical Research*, 26(6), 2603–2621.

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.

Fourcade, Y., Engler, J. O., Rödder, D., & Secondi, J. (2014). Mapping species distributions with MAXENT using a geographically biased sample of presence data: a performance assessment of methods for correcting sampling bias. *PLOS ONE*, 9, e97122.

Fox, J. & Weisberg, S. (2019). An {R} Companion to Applied Regression. Third Edition. Thousand Oaks CA: Sage. [cited November 20, 2020]. Available at: https://socialsciences.mcmaster.ca/jfox/Books/Companion/

Freckleton, R. P., Harvey, P. H., & Pagel, M. (2002). Phylogenetic analysis and comparative data: a test and review of evidence. *The American Naturalist*, 160(6), 712–726.

Fridley, J. D., & Grime, J. P. (2010). Community and ecosystem effects of intraspecific genetic diversity in grassland microcosms of varying species diversity. *Ecology*, 91(8), 2272–2283.

Fried, G., Armengot, L., Storkey, J., Bourgeois, B., Gaba, S., Violle, C., & Munoz, F. (2020). Do ecological specialization and functional traits explain the abundance–frequency relationship? Arable weeds as a case study. *Journal of Biogeography*. [cited November 20, 2020]. Available at: https://doi.org/10.1111/jbi.13980

Friedman, J., Hastie, T., & Tibshirani, R. (2010). Regularization Paths for Generalized Linear Models via Coordinate Descent. *Journal of Statistical Software*, 33(1), 1–22.

Fritz, S. A. & Purvis, A. (2010). Selectivity in mammalian extinction risk and threat types: a new measure of phylogenetic signal strength in binary traits. *Conservation Biology*, 24(4), 1042–1051.

Gaston, K. J. (1991). How large is a species' geographic range?. *Oikos*, 434–438.

Gaston, K. J. (2000). Global patterns in biodiversity. *Nature*, 405(6783), 220–227.

Gaston, K. J. & Blackburn, T. M. (2008). Pattern and Process in Macroecology. Hoboken, NJ, USA: John Wiley & Sons.

Gaston, K. J., Blackburn, T. M., & Lawton, J. H. (1997). Interspecific abundance-range size relationships: an appraisal of mechanisms. *Journal of Animal Ecology*, 66, 579–601.

Gaston, K. J. & Kunin, W. E. (1997). Rare–common differences: an overview. In: Kunin, W. E. & and Gaston, K. J. (Eds.), The Biology of Rarity: Causes and Consequences of Rare–Common Differences. London, UK: Chapman & Hall, pp. 12–29.

Gelman, A. (2008). Scaling regression inputs by dividing by two standard deviations. *Statistics in Medicine*, 27(15), 2865-2873.

Gomes, V. H., IJff, S. D., Raes, N., Amaral, I. L., Salomão, R. P., de Souza Coelho, L., ..., Guevara, J. E. (2018). Species Distribution Modelling: Contrasting presence-only models with plot abundance data. *Scientific Reports*, 8, 1003.

Green, D. F & Johnson, E. A. (1993). Seed mass and dispersal capacity in wind-dispersed diaspores. *Oikos*, 67, 69–74.

Greiser, C., Hylander, K., Meineri, E., Luoto, M., & Ehrlén, J. (2020). Climate limitation at the cold edge: contrasting perspectives from species distribution modelling and a transplant experiment. *Ecography*, 43(5), 637–647.

Grime, J. P. (1979) Plant strategies and vegetation processes. Chichester, UK: Wiley.

Grime, J. P., Thompson, K., Hunt, R., Hodgson, J. G., Cornelissen, J. H. C., Rorison, I. H., …, Booth, R.E. (1997) Integrated screening validates primary axes of specialisation in plants. *Oikos*, 79. 259–281.

Grinnell, J. (1922). The role of the ‘accidental’. *The Auk*, 39, 373–380.

Guerrero, A. M., McAllister, R. R., Corcoran, J., & Wilson, K. A. (2013). Scale mismatches, conservation planning, and the value of social-network analyses. *Conservation Biology*, 27, 35–44.

Guisan, A., & Thuiller, W. (2005). Predicting species distribution: offering more than simple habitat models. *Ecology Letters*, 8, 993–1009.

Guisan, A., Petitpierre, B., Broennimann, O., Daehler, C., & Kueffer, C. (2014). Unifying niche shift studies: insights from biological invasions. *Trends in Ecology & Evolution*, 29, 260–269.

Guo, Q., Brown, J. H., Valone, T. J., & Kachman, S. D. (2000). Constraints of seed size on plant distribution and abundance. *Ecology*, 81(8), 2149–2155.

Guo, W. Y., van Kleunen, M., Winter, M., Weigelt, P., Stein, A., Pierce, S., …, Kreft, H. (2018). The role of adaptive strategies in plant naturalization. *Ecology Letters*, 21(9), 1380–1389.

Gurevitch, J., Scheiner, S. & Fox, G. (2002). The ecology of plants. Sunderland, MA, USA: Sinauer Associates.

Güsewell, S. (2004). N: P ratios in terrestrial plants: variation and functional significance. *New Phytologist*, 164(2), 243–266.

Hackathon, R., Bolker, B., Butler, M., Cowan, P., De Vienne, D., Eddelbuettel, D., …, Michonneau, F. (2013). phylobase: Base package for phylogenetic structures and comparative data. R package version 0.8, 4. [cited November 20, 2020]. Available at: https://cran.r-project.org/web/packages/phylobase/index.html

Hall, P., & Wood, A. (1993). On the performance of box-counting estimators of fractal dimension. *Biometrika*, 80, 246–252.

Hampe, A., & Petit, R. J. (2005). Conserving biodiversity under climate change: the rear edge matters. *Ecology Letters*, 8(5), 461–467.

Hanski, I. (1982). Dynamics of regional distribution: the core and satellite species hypothesis. *Oikos*, 38, 210–221.

Hargreaves, A. L., Samis, K. E., & Eckert, C. G. (2013). Are species' range limits simply niche limits writ large? A review of transplant experiments beyond the range. *The American Naturalist*, 183, 157–173.

Hartley, S., & Kunin, W. E. (2003). Scale dependency of rarity, extinction risk, and conservation priority. Conservation Biology, 17, 1559–1570.

Hegde, S.G. & Ellstrand, N.C. (1999). Life history differences between rare and common flowering plant species of California and the British Isles. *International Journal of Plant Sciences*, 160, 1083–1091.

Heino, J. & Tolonen, K. T. (2018). Ecological niche features override biological traits and taxonomic relatedness as predictors of occupancy and abundance in lake littoral macroinvertebrates. *Ecography*, 41(12), 2092–2103.

Helmuth, B., Kingsolver, J. G., & Carrington, E. (2005). Biophysics, physiological ecology, and climate change: does mechanism matter?. *Annual Review of Physiology*, 67, 177–201.

Hengeveld, R. & Haeck, J. (1982). The distribution of abundance. I. Measurements. *Journal of Biogeography*, 9, 303–316.

Hijmans, R. J., Cameron, S. E., Parra, J. L., Jones, P. G., & Jarvis, A. (2005). Very high resolution interpolated climate surfaces for global land areas. *International Journal of Climatology*, 25, 1965–1978.

Hirst, M. J., Griffin, P. C., Sexton, J. P., & Hoffmann, A. A. (2017). Testing the niche-breadth–range-size hypothesis: habitat specialization vs. performance in Australian alpine daisies. *Ecology*, 98, 2708–2724.

Hoffmann, M., Hilton-Taylor, C., Angulo, A., Böhm, M., Brooks, T. M., Butchart, S. H., ..., Darwall, W. R. (2010). The impact of conservation on the status of the world's vertebrates. *Science*, 330, 1503–1509.

Hoffmann, A., Penner, J., Vohland, K., Cramer, W., Doubleday, R., Henle, K., …, Häuser, C. L. (2014). Improved access to integrated biodiversity data for science, practice, and policy - the European Biodiversity Observation Network (EU BON). *Nature Conservation*, 6, 49–65.

Hofmann, M., Bütof, A., Welk, E., & Bruelheide, H. (2013). Relationship between fundamental and realized niches in terms of frost and drought resistance. *Preslia*, 85, 1–17.

Hortal, J., Borges, P. A., & Gaspar, C. (2006). Evaluating the performance of species richness estimators: sensitivity to sample grain size. *Journal of Animal Ecology*, 75, 274–287.

Hortal, J., Jiménez-Valverde, A., Gómez, J. F., Lobo, J. M., & Baselga, A. (2008). Historical bias in biodiversity inventories affects the observed environmental niche of the species. *Oikos*, 117, 847–858.

Hurlbert, A. H. & Jetz, W. (2007). Species richness, hotspots, and the scale dependence of range maps in ecology and conservation. *Proceedings of the National Academy of Sciences*, 104(33), 13384–13389.

Hutchinson, G. E. (1957). Concluding Remarks. Population studies: Animal Ecology and Demography. Cold Spring Harbor Symposium on Quantitative Biology, 22, 415–442.

Hutchinson, G. E. (1978). An introduction to population ecology. New Haven, CT, US: Yale University Press.

IUCN 2019. Guidelines for using the IUCN Red List categories and criteria, ver. 14. – IUCN Standards and Petitions Committee. Gland, Switzerland and Cambridge, UK: IUCN.

Jandt, U., von Wehrden, H., & Bruelheide, H. (2011). Exploring large vegetation databases to detect temporal trends in species occurrences. *Journal of Vegetation Science*, 22, 957–972.

Jansen, F. & Dengler, J. (2008). GermanSL – Eine universelle taxonomische Referenzliste für Vegetationsdatenbanken in Deutschland. *Tuexenia*, 28, 239– 253.

Jetz, W., McGeoch, M. A., Guralnick, R., Ferrier, S., Beck, J., Costello, M. J., ..., Meyer, C. (2019). Essential biodiversity variables for mapping and monitoring species populations. *Nature Ecology & Evolution*, 3, 539.

Jetz, W., McPherson, J. M., & Guralnick, R. P. (2012). Integrating biodiversity distribution knowledge: toward a global map of life. *Trends in Ecology and Evolution*, 27, 151–159.

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, 483–490.

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

Johnston, A., Fink, D., Reynolds, M. D., Hochachka, W. M., Sullivan, B. L., Bruns, N. E., ..., Kelling, S. (2015). Abundance models improve spatial and temporal prioritization of conservation resources. *Ecological Applications*, 25, 1749–1756.

Jung, V., Violle, C., Mondy, C., Hoffmann, L., & Muller, S. (2010). Intraspecific variability and trait-based community assembly. *Journal of Ecology*, 98(5), 1134–1140.

Kambach, S., Lenoir, J., Decocq, G., Welk, E., Seidler, G., Dullinger, S., ..., Vittoz, P. (2019). Of niches and distributions: range size increases with niche breadth both globally and regionally but regional estimates poorly relate to global estimates. *Ecography*, 42, 467–477.

Karger, D. N., Conrad, O., Böhner, J., Kawohl, T., Kreft, H., Soria-Auza, R. W., ..., Kessler, M. (2017). Climatologies at high resolution for the earth's land surface areas. *Scientific Data*, 4, 170122.

Kassambara, A. & Mundt, F. (2017). Package 'factoextra'. The Comprehensive R Archive Network (CRAN). [cited November 20, 2020]. Available at: https://cran.r-project.org/web/packages/factoextra/index.html

Kattge, J., Bönisch, G., Díaz, S., Lavorel, S., Prentice, I. C., Leadley, P., …, Acosta, A. T. (2020). TRY plant trait database–enhanced coverage and open access. *Global Change Biology*, 26(1), 119–188.

Kelly, C. K. & Woodward, F. I. (1996). Ecological correlates of plant range size: taxonomies and phylogenies in the study of plant commonness and rarity in Great Britain. *Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences*, 351(1345), 1261–1269.

Kichenin, E., Wardle, D. A., Peltzer, D. A., Morse, C. W., & Freschet, G. T. (2013). Contrasting effects of plant inter-and intraspecific variation on community-level trait measures along an environmental gradient. *Functional Ecology*, 27(5), 1254–1261.

Klimešová, J., Danihelka, J., Chrtek, J., de Bello, F., & Herben, T. (2017). CLO-PLA: A database of clonal and bud-bank traits of the Central European flora. *Ecology*, 98(4), 1179–1179.

Köckemann, B., Buschmann, H., & Leuschner, C. (2009). The relationships between abundance, range size and niche breadth in Central European tree species. *Journal of Biogeography*, 36, 854–864.

Koenker, R. (2018). quantreg: Quantile Regression. R package version 5.38. [cited November 20, 2020]. Available at: https://CRAN.R-project.org/package=quantreg

Kolb, A., Barsch, F., & Diekmann, M. (2006). Determinants of local abundance and range size in forest vascular plants. *Global Ecology and Biogeography*, 15(3), 237–247.

Kühn, I., Durka, W., & Klotz, S. (2004). BiolFlor: a new plant-trait database as a tool for plant invasion ecology. *Diversity and Distributions*, 10(5/6), 363-365.

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

Küster, E. C., Kühn, I., Bruelheide, H., & Klotz, S. (2008). Trait interactions help explain plant invasion success in the German flora. *Journal of Ecology*, 96(5), 860–868.

Lande, R., Engen, S., & Saether, B. E. (2003). Stochastic population dynamics in ecology and conservation. Oxford, UK: Oxford University Press.

Lavergne, S., Thompson, J. D., Garnier, E., & Debussche, M. (2004). The biology and ecology of narrow endemic and widespread plants: a comparative study of trait variation in 20 congeneric pairs. *Oikos*, 107(3), 505–518.

Lavorel, S. & Garnier, E. (2002). Predicting changes in community composition and ecosystem functioning from plant traits: revisiting the Holy Grail. *Functional Ecology*, 16(5), 545–556.

Leibold, M. A., Holyoak, M., Mouquet, N., Amarasekare, P., Chase, J. M., Hoopes, M. F., ..., Loreau, M. (2004). The metacommunity concept: a framework for multi-scale community ecology. *Ecology Letters*, 7, 601–613.

Leishman, M. R. & Murray, B. R. (2001). The relationship between seed size and abundance in plant communities: model predictions and observed patterns. *Oikos*, 94, 151–161.

Lembrechts, J. J., Aalto, J., Ashcroft, M. B., De Frenne, P., Kopecký, M., Lenoir, J., ..., García, R. A. (2020). SoilTemp: a global database of near-surface temperature. *Global Change Biology*, [cited November 20, 2020]. Available at: https://doi.org/10.1111/gcb.15123.

Lengyel, A., Chytrý, M., & Tichý, L. (2011). Heterogeneity-constrained random resampling of phytosociological databases. Journal of Vegetation Science, 22, 175–183.

Lobo, J. M., Hortal, J., Yela, J. L., Millán, A., Sánchez-Fernández, D., García-Roselló, E., …, Guisande, C. (2018). KnowBR: An application to map the geographical variation of survey effort and identify well-surveyed areas from biodiversity databases. *Ecological Indicators*, 91, 241–248.

Loehle, C. (1998). Height growth rate tradeoffs determine northern and southern range limits for trees. *Journal of Biogeography*, 25, 735–742.

Lomolino, M. V., Riddle, B. R., & Whittaker, R. J. (2017). Biogeography, Fifth Edition. (5), 730. Sunderland, MA: Oxford University Press.

Louthan, A. M., Doak, D. F., & Angert, A. L. (2015). Where and when do species interactions set range limits?. *Trends in Ecology & Evolution*, 30(12), 780–792.

Lundqvist, J. (1992). Index Holmiensis, Vol. 7. Stockholm, Sweden: Swedish Museum of Natural History.

Lundqvist, J. & Jäger, E. J. (1995-2007). Index Holmiensis. Stockholm, Sweden: Swedish Museum of Natural History.

Lundqvist, J. & Nordenstam, B. (1988). Index Holmiensis, Vol. 6. Stockholm, Sweden: Swedish Museum of Natural History.

MacArthur, R. H. (1984). Geographical Ecology: Patterns in the Distribution of Species. Princeton University Press.

MacArthur, R. H., Diamond, J. M., & Karr, J. R. (1972). Density compensation in island faunas. *Ecology*, 53, 330–342.

Mack, R. N., Simberloff, D., Lonsdale, W. M., Evans, H., Clout, M., & Bazzaz, F. A. (2000). Biotic invasions: causes, epidemiology, global consequences, and control. *Ecological Applications*, 10(3), 689–710.

Manthey, J. D., Campbell, L. P., Saupe, E. E., Soberón, J., Hensz, C. M., Myers, C. E., ..., Lira-Noriega, A. (2015). A test of niche centrality as a determinant of population trends and conservation status in threatened and endangered North American birds. *Endangered Species Research*, 26, 201–208.

Marino, N. A., Céréghino, R., Gilbert, B., Petermann, J. S., Srivastava, D. S., de Omena, P. M., …, Barberis, I. M. (2020). Species niches, not traits, determine abundance and occupancy patterns: A multi-site synthesis. *Global Ecology and Biogeography*, 29(2), 295–308.

Mariotte, P. (2014). Do subordinate species punch above their weight? Evidence from above- and below-ground. New Phytologist, 203(1), 16–21.

Marsh, C. J., Barwell, L. J., Gavish, Y., & Kunin, W. E. (2018). downscale: an R package for downscaling species occupancy from coarse-grain data to predict occupancy at fine-grain sizes. *Journal of Statistical Software*, 86.

Martínez-Meyer, E., Díaz-Porras, D., Peterson, A. T., & Yáñez-Arenas, C. (2013). Ecological niche structure and rangewide abundance patterns of species. *Biology Letters*, 9, 20120637.

McGill, B. & Collins, C. (2003). A unified theory for macroecology based on spatial patterns of abundance. *Evolutionary Ecology Research*, 5(4), 469–492.

McGill, B. J., Enquist, B. J., Weiher, E., & Westoby, M. (2006). Rebuilding community ecology from functional traits. *Trends in Ecology & Evolution*, 21(4), 178–185.

McNellie, M. J., Dorrough, J., & Oliver, I. (2019). Species abundance distributions should underpin ordinal cover abundance transformations. *Applied Vegetation Science*, 22, 361–372.

Mentges, A., Blowes, S. A., Hodapp, D., Hillebrand, H., & Chase, J. M. (2020). Effects of site-selection bias on estimates of biodiversity change. *Conservation Biology*. [cited November 20, 2020]. Available at: https://doi.org/10.1111/cobi.13610

Mertes, K. & Jetz, W. (2018). Disentangling scale dependencies in species environmental niches and distributions. *Ecography*, 41, 1604–1615.

Meurant, G. (2012). The ecology of natural disturbance and patch dynamics. Cambridge, MA, USA: Academic Press.

Meusel, H. & Jäger, E. J. (1992). Vergleichende Chorologie der zentraleuropäischen Flora, Karten, Vol. III. Jena, Germany: Gustav Fischer Verlag.

Meusel, H., Jäger, E. J., & Weinert, E. (1965). Vergleichende Chorologie der zentraleuropäischen Flora, Karten, Vol. I. Jena, Germany: VEB Gustav Fischer Verlag.

Meusel, H., Jäger, E. J., Rauschert, S., & Weinert, E. (1978). Vergleichende Chorologie der zentraleuropäischen Flora, Karten, Vol. II. Jena, Germany: VEB Gustav Fischer Verlag.

Meyer, C., Weigelt, P., & Kreft, H. (2016). Multidimensional biases, gaps and uncertainties in global plant occurrence information. *Ecology Letters*, 19, 992–1006.

Mi, C., Huettmann, F., Sun, R., & Guo, Y. (2017). Combining occurrence and abundance distribution models for the conservation of the Great Bustard. *PeerJ*, 5, e4160.

Moeslund, J. E., Brunbjerg, A. K., Clausen, K. K., Dalby, L., Fløjgaard, C., Juel, A., & Lenoir, J. (2017). Using dark diversity and plant characteristics to guide conservation and restoration. *Journal of Applied Ecology*, 54, 1730–1741.

Moles, A. T., Perkins, S. E., Laffan, S. W., Flores-Moreno, H., Awasthy, M., Tindall, M. L., …, Anand, M. (2014). Which is a better predictor of plant traits: temperature or precipitation? *Journal of Vegetation Science*, 25(5), 1167–1180.

Morris, W. F., Ehrlén, J., Dahlgren, J. P., Loomis, A. K., & Louthan, A. M. (2020). Biotic and anthropogenic forces rival climatic/abiotic factors in determining global plant population growth and fitness. *Proceedings of the National Academy of Sciences*, 117, 1107–1112.

Murphy, H.T., VanDerWal, J., & Lovett-Doust, J. (2006). Distribution of abundance across the range in eastern North American trees. *Global Ecology and Biogeography*, 15(1),63–71.

Murray, B. R. & Lepschi, B. J. (2004). Are locally rare species abundant elsewhere in their geographical range?. *Austral Ecology*, 29(3), 287–293.

Murray, B. R., Thrall, P. H., Gill, A. M., & Nicotra, A. B. (2002). How plant life-history and ecological traits relate to species rarity and commonness at varying spatial scales. *Austral Ecology*, 27(3), 291–310.

Naimi, B. & Araújo, M. B. (2016). sdm: a reproducible and extensible R platform for species distribution modelling. *Ecography*, 39, 368–375.

Niinemets, Ü. (2015). Is there a species spectrum within the world-wide leaf economics spectrum? Major variations in leaf functional traits in the Mediterranean sclerophyll Quercus ilex. *New Phytologist*, 205(1), 79–96.

Normand, S., Treier, U. A., Randin, C., Vittoz, P., Guisan, A., & Svenning, J. C. (2009). Importance of abiotic stress as a range-limit determinant for European plants: insights from species responses to climatic gradients. *Global Ecology and Biogeography*, 18(4), 437–449.

Nunes, L. A. & Pearson, R. G. (2017). A null biogeographical test for assessing ecological niche evolution. *Journal of Biogeography*, 44, 1331–1343.

Oakwood, M., Jurado, E., Leishman, M., & Westoby, M. (1993). Geographic ranges of plant species in relation to dispersal morphology, growth form and diaspore weight. *Journal of Biogeography*, 20, 563–571.

Orme, D, Freckleton, R., Thomas, G., Petzoldt, T., Fritz, S., Isaac, N., & Pearse, W. (2018). caper: Comparative Analyses of Phylogenetics and Evolution in R. R package version 1.0.1. [cited November 20, 2020]. Available at: https://CRAN.R-project.org/package=caper

Osorio-Olvera, L., Soberón, J., & Falconi, M. (2019). On population abundance and niche structure. *Ecography*, 42, 1415–1425.

Osorio-Olvera, L., Yañez-Arenas, C., Martínez-Meyer, E., & Peterson, A.T. (2020). Relationships between population densities and niche-centroid distances in North American birds. *Ecology Letters*, 23, 555–564.

Pagel, M. (1999). Inferring the historical patterns of biological evolution. *Nature*, 401(6756), 877–884.

Pearman, P. B., Guisan, A., Broennimann, O., & Randin, C. F. (2008). Niche dynamics in space and time. *Trends in Ecology & Evolution*, 23, 149–158.

Pearson, R. G. & Dawson, T. P. (2003). Predicting the impacts of climate change on the distribution of species: are bioclimate envelope models useful? *Global Ecology and Biogeography*, 12, 361–371.

Pentland, A. P. (1984). Fractal-based description of natural scenes. *IEEE Transactions on Pattern Analysis & Machine Intelligence*, 6, 661–674.

Petchey, O. L. & Gaston, K. J. (2006). Functional diversity: back to basics and looking forward. *Ecology Letters*, 9(6), 741–758.

Peterson, A. T., Soberón, J., Pearson, R. G., Anderson, R. P., Martínez-Meyer, E., Nakamura, M., & Araújo, M. B. (2011). Ecological niches and geographic distributions (MPB-49) (Vol. 56). Princeton, NJ, USA: Princeton University Press.

Pironon, S., Papuga, G., Villellas, J., Angert, A. L., García, M. B., & Thompson, J. D. (2017). Geographic variation in genetic and demographic performance: new insights from an old biogeographical paradigm. *Biological Reviews*, 92, 1877–1909.

Pironon, S., Villellas, J., Morris, W. F., Doak, D. F., & García, M. B. (2015). Do geographic, climatic or historical ranges differentiate the performance of central versus peripheral populations? *Global Ecology and Biogeography*, 24, 611–620.

Preston, F. W. (1948). The commonness, and rarity, of species. *Ecology*, 29, 254–283.

Pyke, G. H. & Ehrlich, P. R. (2010). Biological collections and ecological/environmental research: a review, some observations and a look to the future. *Biological Reviews*, 85, 247–266.

Pysek, P. & Richardson, D.M. (2007). Traits associated with invasiveness in alien plants: where do we stand? In: Nentwig, W. (Ed.), Biological Invasions. Berlin and Heidelberg, Germany: Springer-Verlag, pp. 97–125.

R Core Team (2018). R: A language and environment for statistical computing. Vienna, AT: R Foundation for Statistical Computing.

Rabinowitz, D. (1981). Seven forms of rarity. In: Synge, H. and Chichester, J. (Eds.), The Biological Aspects of Rare Plant Conservation. New York, USA: John Wiley & Sons, pp. 207–217.

Raunkiaer, C. (1934). The life forms of plants and statistical plant geography, being the collected papers of C. Raunkiær. Oxford, UK: Clarendon Press.

Reich, P. B. (2014). The world-wide ‘fast–slow’ plant economics spectrum: a traits manifesto. *Journal of Ecology*, 102(2), 275–301.

Reich, P. B. & Oleksyn, J. (2004). Global patterns of plant leaf N and P in relation to temperature and latitude. *Proceedings of the National Academy of Sciences*, 101, 11001–11006.

Reif, J., Hořák, D., Sedláček, O., Riegert, J., Pešata, M., Hrázský, Z., ..., Storch, D. (2006). Unusual abundance–range size relationship in an Afromontane bird community: the effect of geographical isolation? *Journal of Biogeography*, 33, 1959–1968.

Ren, H., Condit, R., Chen, B., Mi, X., Cao, M., Ye, W., …, Ma, K. (2013). Geographical range and local abundance of tree species in China. *PLOS ONE*, 8, e76374.

Ricklefs, R. E. (2004). A comprehensive framework for global patterns in biodiversity. *Ecology Letters*, 7, 1–15.

Rumpf, S. B., Hülber, K., Klonner, G., Moser, D., Schütz, M., Wessely, J., ..., Dullinger, S. (2018). Range dynamics of mountain plants decrease with elevation. *Proceedings of the National Academy of Sciences*, 115(8), 1848–1853.

Sagarin, R. D. & Gaines, S. D. (2002). The ‘abundant centre’ distribution: to what extent is it a biogeographical rule? *Ecology Letters*, 5, 137–147.

Sagarin, R. D., Gaines, S. D., & Gaylord, B. (2006). Moving beyond assumptions to understand abundance distributions across the ranges of species. *Trends in Ecology & Evolution*, 21, 524–530.

San-Miguel-Ayanz, J., de Rigo, D., Caudullo, G., Houston Durrant, T., & Mauri, A. (Eds.) (2016). European atlas of forest tree species. Luxembourg, LU: Publication Office of the European Union. DOI: 10.2788/038466

Santini, L., Pironon, S., Maiorano, L., & Thuiller, W. (2019). Addressing common pitfalls does not provide more support to geographical and ecological abundant-centre hypotheses. *Ecography*, 42, 696–705.

Schleuning, M., Fründ, J., Schweiger, O., Welk, E., Albrecht, J., Albrecht, M., ..., Hof, C. (2016). Ecological networks are more sensitive to plant than to animal extinction under climate change. *Nature Communications*, 7, 13965.

Schrodt, F., Kattge, J., Shan, H., Fazayeli, F., Joswig, J., Banerjee, A., …, Gillison, A. (2015). BHPMF–a hierarchical Bayesian approach to gap-filling and trait prediction for macroecology and functional biogeography. *Global Ecology and Biogeography*, 24(12), 1510–1521.

Shmida, A. V. I. & Wilson, M. V. (1985). Biological determinants of species diversity. *Journal of Biogeography*, 12, 1–20.

Siefert, A., Violle, C., Chalmandrier, L., Albert, C. H., Taudiere, A., Fajardo, A., ..., de L. Dantas, V. (2015). A global meta-analysis of the relative extent of intraspecific trait variation in plant communities. *Ecology Letters*, 18(12), 1406–1419.

Slatyer, R. A., Hirst, M., & Sexton, J. P. (2013). Niche breadth predicts geographical range size: A general ecological pattern. *Ecology Letters*, 16, 1104–1114.

Soberón, J. & Nakamura, M. (2009). Niches and distributional areas: Concepts, methods, and assumptions. *Proceedings of the National Academy of Sciences of the United States of America*, 106, 19644–19650.

Soberón, J., Jiménez, R., Golubov, J., & Koleff, P. (2007). Assessing completeness of biodiversity databases at different spatial scales. *Ecography*, 30, 152–160.

Soria-Auza, R. W. & Kessler, M. (2007). The influence of sampling intensity on the perception of the spatial distribution of tropical diversity and endemism: a case study of ferns from Bolivia. *Diversity and Distributions*, 14, 123–130.

Sousa-Baena, M. S., Garcia, L. C., & Peterson, A. T. (2014). Completeness of digital accessible knowledge of the plants of Brazil and priorities for survey and inventory. *Diversity and Distributions*, 20, 369–381.

Speed, J. D. M., Bendiksby, M., Finstad, A. G., Hassel, K., Kolstad, A. L., & Prestø, T. (2018). Contrasting spatial, temporal and environmental patterns in observation and specimen based species occurrence data. *PLOS ONE*, 13, e0196417.

Sporbert, M., Keil, P., Seidler, G., Bruelheide, H., Jandt, U., Aćic, S., …, Welk, E. (2020). Testing macroecological abundance patterns: The relationship between local abundance and range size, range position and climatic suitability among European vascular plants. *Journal of Biogeography*, 47(10), 2210–2222.

Staniczenko, P. P., Sivasubramaniam, P., Suttle, K. B., & Pearson, R. G. (2017). Linking macroecology and community ecology: refining predictions of species distributions using biotic interaction networks. *Ecology Letters*, 20, 693–707.

Suding, K. N., Lavorel, S., Chapin III, F. S., Cornelissen, J. H., Diaz, S., Garnier, E., …, Navas, M. L. (2008). Scaling environmental change through the community-level: a trait-based response-and-effect framework for plants. *Global Change Biology*, 14(5), 1125–1140.

Tessarolo, G., Rangel, T., Araújo, M. B., & Hortal, J. (2014). Uncertainty associated with survey design in Species Distribution Models. *Diversity and Distributions*, 20, 1258–1269.

Thompson, K., Gaston, K. J., & Band, S. R. (1999) Range size, dispersal and niche breadth in the herbaceous flora of central England. *Journal of Ecology*, 87(1), 150–155.

Thompson, K., Hodgson, J. G., & Gaston, K. J. (1998). Abundance–range size relationships in the herbaceous flora of central England. *Journal of Ecology*, 86, 439–448.

Thompson, F. J., Moles, A. T., Auld, T. D., & Kingsford, R. T. (2011). Seed dispersal distance is more strongly correlated with plant height than with seed mass. *Journal of Ecology*, 99, 1299–1307.

Titeux, N., Maes, D., Daele, T. V., Onkelinx, T., Heikkinen, R. K., Romo, H., …, Luoto, M. (2017). The need for large-scale distribution data to estimate regional changes in species richness under future climate change. *Diversity and Distributions*, 23, 1393–1407.

Tralau, H. (1969–1981). Index Holmiensis Vol. 1–5. Stockholm, Sweden: Swedish Museum of Natural History.

Troia, M. J. & McManamay, R. A. (2016). Filling in the GAPS: evaluating completeness and coverage of open-access biodiversity databases in the United States. *Ecology and Evolution*, 6, 4654–4669.

USDA, NRCS. The PLANTS Database. National Plant Data Team, Greensboro, NC 27401-4901 USA. [cited November 20, 2020]. Available at: http://plants.usda.gov.

Van Couwenberghe, R., Collet, C., Pierrat, J. C., Verheyen, K., & Gégout, J. C. (2013). Can species distribution models be used to describe plant abundance patterns? *Ecography*, 36, 665–674.

van der Maarel, E. (1979). Transformation of cover-abundance values in phytosociology and its effects on community similarity. *Vegetatio*, 39, 97–114.

Van der Veken, S., Bellemare, J., Verheyen, K., & Hermy, M. (2007). Life-history traits are correlated with geographical distribution patterns of western European forest herb species. *Journal of Biogeography*, 34, 1723–1735.

VanDerWal, J., Shoo, L. P., Johnson, C. N., & Williams, S. E. (2009). Abundance and the environmental niche: environmental suitability estimated from niche models predicts the upper limit of local abundance. *The American Naturalist*, 174, 282–291.

Viechtbauer, W. (2010). Conducting meta-analyses in R with the metafor package. *Journal of Statistical Software*, 36, 1-48. [cited November 20, 2020]. Available at: http://www.jstatsoft.org/v36/i03/

Violle, C., Navas, M.L., Vile, D., Kazakou, E., Fortunel, C., Hummel, I., & Garnier, E. (2007). Let the concept of trait be functional! *Oikos*, 116(5), 882–892.

Wasof, S., Lenoir, J., Aarrestad, P. A., Alsos, I. G., Armbruster, W. S., Austrheim, G., ..., Decocq, G. (2015). Disjunct populations of European vascular plant species keep the same climatic niches. *Global Ecology and Biogeography*, 24, 1401–1412.

Weber, M. M., Stevens, R. D., Diniz-Filho, J. A. F., & Grelle, C. E. V. (2017). Is there a correlation between abundance and environmental suitability derived from ecological niche modelling? A meta-analysis. *Ecography*, 40, 817–828.

Welk, A., Welk, E., & Bruelheide, H. (2014). Biotic interactions overrule plant responses to climate, depending on the species' biogeography. *PLOS ONE*, 9, e111023.

Wellstein, C., Chelli, S., Campetella, G., Bartha, S., Galie, M., Spada, F., & Canullo, R. (2013). Intraspecific phenotypic variability of plant functional traits in contrasting mountain grasslands habitats. *Biodiversity and Conservation*, 22(10), 2353–2374.

Wiegand, T. & Moloney, K. A. (2013). Handbook of Spatial Point-Pattern Analysis in Ecology. Boca Raton, FL, US: CRC Press.

Wiens, J. A. (1989). Spatial scaling in ecology. *Functional Ecology*, 3, 385–397.

Willis, K. J. & Whittaker, R. J. (2002). Species Diversity-Scale Matters. *Science*, 295, 1245–1248.

Winemiller, K. O., Fitzgerald, D. B., Bower, L. M., & Pianka, E. R. (2015). Functional traits, convergent evolution, and periodic tables of niches. *Ecology Letters*, 18(8), 737–751.

Woodward, F. I. (1986). Climate and plant distribution. Cambridge, UK: Cambridge University Press.

Wright, I. J., Dong, N., Maire, V., Prentice, I. C., Westoby, M., Díaz, S., …, Leishman, M. R. (2017). Global climatic drivers of leaf size. *Science*, 357(6353), 917–921.

Wright, I. J., Reich, P. B., Westoby, M., Ackerly, D. D., Baruch, Z., Bongers, F., …, Flexas, J. (2004). The worldwide leaf economics spectrum. *Nature*, 428(6985), 821–827.

Yang, W., Ma, K., & Kreft, H. (2013). Geographical sampling bias in a large distributional database and its effects on species richness–environment models. *Journal of Biogeography*, 40, 1415–1426.

Zizka, A., Antonelli, A., & Silvestro, D. (2020). Sampbias, a method for quantifying geographic sampling biases in species distribution data. *Ecography*, 43, 1–7.

Appendices

Author Contributions

Chapter II

EW and MS developed the DMC concept, with considerable input by GS and HB. MS wrote the first draft of the manuscript, with considerable input by EW, HB, PK and UJ. MS and GS harmonized data retrieved from EVA and CDH. GS wrote R code for DMC calculation. PK wrote R code for the null model application for DMC calculations. MS carried out statistical analyses and produced the graphs. All other authors contributed data. All authors contributed to writing the manuscript. My overall contribution was 70%.

Chapter III

MS and EW conceived the study with considerable input by HB and PK. GS and MS harmonized data retrieved from EVA and CDH. MS carried out statistical analyses with support by GS. MS produced the graphs. MS and PK wrote the paper, MS led the writing. All other authors contributed data. All authors contributed to the writing of the manuscript. My overall contribution was 75%.

Chapter IV

MS, EW and HB conceived the study. GS and MS harmonized data retrieved from EVA and CDH. MS harmonized data retrieved from TRY with data retrieved from EVA and CDH. EW, GS and MS developed the measures for niche properties and abundance skewness. MS and HB carried out statistical analyses. MS produced the graphs. MS and HB wrote the paper, MS led the writing. All other authors contributed data. All authors discussed the results and commented on the manuscript. My overall contribution was 85%.

Acknowledgement/ Danksagung

First of all, I would like to thank my main supervisors Helge Bruelheide and Erik Welk. You gave me the opportunity to develop my own project and were most supportive and motivating at any stage of this project. Thank you, Helge, for always taking the time to discuss my work, for teaching me statistics, for all your supportive feedback, for supporting me in building my scientific network, and for pushing me when needed, - I can not imagine a better Doktorvater. Thank you, Erik, for sharing your knowledge, for the fun (and headaches) while developing paper ideas, for teaching me to think critically, and for really always taking the time and having an open ear to discuss all my worries and doubts - I don´t take all that for granted.

A huge thanks goes to Gunnar Seidler, for your infinite patience in teaching me coding and GIS, for your support in data merging and keeping the computers busy with our calculations (and running it again and again, …). I will never forget the hours and days of discussion with you and Erik - I learned so much from you guys and I really enjoyed the years working with you up there in the attic (ok, except during heat waves, to be honest).

A warm thanks to Petr Keil, you were such a great co-supervisor – thanks for all your support and motivating words during this whole project (thanks for calling!), for teaching me that negative results can be a cool thing, for making paper writing fun, and especially for sharing your excitement and passion for science with me.

A big thanks to Ute Jandt, without your support in the data handling process this project would not have been possible. Thanks for all the discussions, critical and helpful edits to manuscripts and talks, and for introducing me to the EVA community. With this, a big shoutout to all my great co-authors. To be honest, I was kind of afraid in the beginning to take the lead in a project with such a bunch of experienced researchers. But you all were so kind and a true inspiration, always supportive, and sharing your thoughts and discussing science with me really have increased my horizon.

My PhD studies were supported by a graduate stipend of the federal state of Saxony-Anhalt. I am thankful for the possibility to be part of iDivs graduate school yDiv, I much enjoyed the great courses, our retreats with the best PhD fellows, and my time at the iDiv Council - special thanks to Nicole Sachmerda-Schulz and Johanna Müller for all your effort and support.

Thanks to all the colleagues from the Geobotany working group Halle for creating a lively, supportive, open-minded and welcoming working atmosphere – it was a pleasure to work with you. A special thanks to Sylvia Haider, you became an inspirational and supportive mentor to me. A warm thanks to the bunch of wonderful PhD colleagues that became good friends, I much enjoyed our scientific and non-scientific discussions ;). Special thanks to Carolin Plos, Yolanda Caceres and Julia Dieskau, you ladies were always an important support for me from the beginning - let´s get this PhD done!

Ich danke von Herzen meiner Familie, im Besonderen meinen Eltern, für eure uneingeschränkte Unterstützung und Motivation während all der Jahre seit Beginn meines Studiums. Danke auch an meine engsten Freunde, die für die nicht zu unterschätzende Zerstreuung neben der Arbeit gesorgt haben. Und abschließend: Danke Philipp, für so Vieles, aber vor Allem dafür, dass ich immer auf dich zählen kann. Was für eine Reise, oder?!

Curriculum Vitae

Personal Information

Name: Maria Sporbert
Email: maria.sporbert@gmail.com

Scientific Employment

2019 – present	**Scientific Coordinator (Post doc position)** of the project "PhenObs - Botanical Gardens as a Global Phenological Observation Network", German Centre for Integrative Biodiversity Research (iDiv) Halle-Jena-Leipzig, Germany & Plant Biodiversity, Institute of Ecology and Evolution with Herbarium Haussknecht and Botanical Garden, Friedrich Schiller University Jena, Germany PI: Prof. Dr. Christine Römermann

Education

2015 –present	**PhD student**, Institute of Biology / Geobotany & Botanical Garden, Martin-Luther University, Halle-Wittenberg Research topic: "Relationship between local abundance, range-wide niche characteristics and functional traits in vascular plants" Supervisor: Prof. Dr. Helge Bruelheide Co-Supervisors: Dr. Erik Welk, Dr. Ute Jandt, Dr. Petr Keil Funding: Graduiertenförderung Sachsen-Anhalt (scholarship), and additional support through institutional funds of Martin-Luther University Halle-Wittenberg
2011 – 2015	**Master of Science Biology**, MLU Halle Title master thesis: "Range and niche dynamics in plant invasion processes" Supervisors: Prof. Dr. Helge Bruelheide, Dr. Erik Welk
2008 - 2011	**Bachelor of Science Biology**, MLU Halle Title bachelor thesis: „Auswirkung von Trockenstress auf Wachstum und Etablierung südwest- und nordost-europäisch verbreiteter Gräser und Kräuter" Supervisors: Prof. Dr. Isabell Hensen, Dr. Susanne Lachmuth

Teaching expertise

2016 – 2019 annual, summer semester	**Course: plant identification** (Modul: Systematic Botany and Biodiversity) **Practical course Müritz Nationalpark** (Modul: Evolution and Biodiversity)

Scientific Memberships and Activities

Since 2018	International Association of Vegetation Science (IAVS)
Since 2015	yDiv Graduate School of iDiv, Halle-Jena-Leipzig Elected PhD representative in the iDiv Council 2017-2019 PhD representative in the iDiv Conference Organization Team 2017

Publications and Conference Contributions

Publications

Sporbert, M., Welk, E., Seidler, G., Jandt, U., Aćić, S., Biurrun, I., Campos, J.A., Čarni, A., Cerabolini, B.E.L., Chytrý, M., Ćušterevska, R., Dengler, J., De Sanctis, M., Dziuba, T., Fagúndez, J., Field, R., Golub, V., He, T., Jansen, F., Lenoir, J., Marcenò, C., Martín-Forés, I., Moeslund, J.E., Moretti, M., Niinemets, Ü., Penuelas, J., Pérez-Haase, A., Vandvik, V., Vassilev, K., Vynokurov, D. & Bruelheide, H. (2021). Different sets of traits explain abundance and distribution patterns of European plants at different spatial scales. *Journal of Vegetation Science*. Accepted. https://doi.org/10.1111/jvs.13016

Nordt, B., Hensen, I., Bucher, S. F., Freiberg, M., Primack, R., Stevens, A.-D., Bonn, A., Wirth, C., Jakubka, D., Plos, C., **Sporbert, M.**, & Römermann, C. (2021). The PhenObs initiative — A standardised protocol for monitoring phenological responses to climate change using herbaceous plant species in botanical gardens. *Functional Ecology,* 35, 821-834. https://doi.org/10.1111/1365-2435.13747

Proß, T., Bruelheide, H., Potvin, C., **Sporbert, M.**, Trogisch, S., & Haider, S. (2021). Drivers of within-tree leaf variation in tropical forests of different tree richness. *Basic and Applied Ecology,* 50, 203-216. https://doi.org/10.1016/j.baae.2020.11.001

Sporbert, M.*,* Keil, P., Bruelheide, H., Seidler, G., Jandt, U., Acic, S., Biurrun, I., Campos, J.A., Čarni, A., Casella, L., Chytrý, M., Custerevska, R., Dengler, J., Golub, V., Jansen, F., Kuzemko, A., Lenoir, J., Marceno, C., Moeslund, J.E., Pérez-Haase, A., Rūsiņa, S., Šilc, U., Tsiripidris, I., Vandvik, V., Vasilev, K., Virtanen, R., & Welk, E. (2020). Testing macroecological abundance patterns: the relationship between local abundance and range size, range position and climatic suitability among European vascular plants. *Journal of Biogeography,* 47, 2210–2222. https://doi.org/10.1111/jbi.13926

Hock, M., Plos, C., **Sporbert, M.***,* & Erfmeier, A. (2020). Combined effects of UV-B and drought on native and exotic populations of *Verbascum thapsus* L. *Plants*, 9, 269.
https://doi.org/10.3390/plants9020269

Sporbert, M.*,* Bruelheide, H., Seidler, G., Keil P., Jandt, U., Austrheim, G, Biurrun, I., Campos, J.A., Čarni, A., Chytrý, M., Csiky, J., De Bie, E., Dengler, J., Golub, V., Grytnes, J.-A., Indreica, A., Jansen, F., Jiroušek, M., Lenoir, J., Luoto, M., Marcenò, C., Moeslund, J.E., Pérez-Haase, A., Rūsiņa, S., Vandvik, V., Vassilev, K., & Welk, E. (2019). Assessing sampling coverage of species distribution in biodiversity databases. *Journal of Vegetation Science*, 30, 620-632. https://doi.org/10.1111/jvs.12763

Bruelheide, H., Dengler, J., Purschke, O., Lenoir, J., Jiménez-Alfaro, B., Hennekens, S.M., Botta-Dukát, Z., Chytrý, M., Field, R., Jansen, F., Kattge, J., Pillar, V.D., Schrodt, F., Mahecha, M.D., Peet, R.K., Sandel, B., van Bodegom, P., ..., **Sporbert, M.**, ... & Jandt, U. (2018). Global trait–environment relationships of plant communities. *Nature Ecology & Evolution*, 2, 1906-1917.
https://doi.org/10.1038/s41559-018-0699-8

Conference Contributions

Sporbert, M., Keil, P., Jandt, U., Bruelheide, H. & Welk, E. (2019): Linking local abundance to broad-scale distribution of European plants. 62nd Annual Symposium of the International Association of Vegetation Science, Bremen, Germany, 14.-19.07.2019. (Talk)

Proß, T., Bruelheide, H., Potvin, C., *Sporbert, M.*, Trogisch, S. & Haider, S. (2019): Leaf traits vary within tree crowns and depend on local neighborhood species richness in a tropical rain forest. Meeting of the Sino-German Cooperation Group "Linkages between plant diversity, microbial diversity and ecosystem functioning in subtropical forest", Haikou University, Hainan, China, 24.03.2019. (Talk)

Welk, E., Seidler, G., Jäger, E.J., *Sporbert, M.* & Bruelheide, H. (2018): CDH - The Chorological Database Halle -content, applications, prospects. iDiv Annual Conference, Leipzig, Germany, 11.-12.12.2018. (Poster)

Proß, T., Bruelheide, H. Potvin, C., *Sporbert, M.*, Trogisch, S. & Haider, S. (2018): Neighborhood species richness and canopy height influence leaf trait variation in a tropical forest. iDiv Annual Conference, Leipzig, Germany, 11.-12.12.2018. (Poster)

Sporbert M., Welk E., Jandt U. & Bruelheide, H. (2018): Inter-relationships between species biotic, climatic and geographic niche properties. Annual Meeting of the Specialist Group for Macroecology of the Ecological Society of Germany, Austria and Switzerland, Birmensdorf, Switzerland, 10.-13.04.2018. (Poster)

Sporbert, M., Welk, E., Jandt, U. & Bruelheide, H. (2018): Inter-relationships between species biotic, climatic and geographic niche properties. 17th workshop of the German working group on vegetation databases, Jena, Germany, 14.-16.03.2018. (Poster)

Sporbert, M., Welk, E., Jandt, U. & Bruelheide, H. (2017): The relationship between macroclimatic niche volume, regional niche width and local species abundance. Joint Annual Meeting of the BES, GfÖ, NecoV and EEF, Ghent, Belgium, 11.-14.12.2017. (Poster)

Sporbert, M., Welk, E., Jandt, U. & Bruelheide, H. (2017): The relationship between macroclimatic niche volume, regional niche width and local species abundance. iDiv Annual Conference, Leipzig, Germany, 19.-20.09.2017. (Poster)

Sporbert, M., Welk, E., Jandt, U. & Bruelheide, H. (2016): Matching large plot databases with distribution range data to evaluate spatial and ecological representativeness. iDiv Annual Conference, Leipzig, Germany, 07.- 08.11.2016. (Poster)

Welk, E., *Sporbert, M.* & Bruelheide, H. (2015): Invasive range dynamics of weeds - The role of plant functional traits and climatic niche attributes. 45th Annual Meeting of the Ecological Society of Germany, Austria and Switzerland, Göttingen, Germany, 31.08-04.09.2015. (Talk)

Hock, M., Plos, C., *Sporbert, M.* & Erfmeier, A. (2014): UV-B radiation and drought – Combined stress effects on native and exotic origins of *Verbascum thapsus* L. 44th Annual Meeting of the Ecological Society of Germany, Austria and Switzerland, Hildesheim, Germany, 08.-12.09.2014. (Poster)

Eigenständigkeitserklärung

Hiermit erkläre ich, dass ich die vorliegende Doktorarbeit mit dem Titel „Relationships of local abundance of vascular plants with range-wide niche characteristics, and the role of functional traits“ eigenständig und ohne fremde Hilfe verfasst sowie keine anderen als die im Text angegebenen Quellen und Hilfsmittel verwendet habe. Textstellen, welche aus verwendeten Werken wörtlich oder inhaltlich übernommen wurden, wurden von mir als solche kenntlich gemacht. Ich erkläre weiterhin, dass ich mich bisher noch nie um einen Doktorgrad beworben habe. Die vorliegende Doktorarbeit wurde bis zu diesem Zeitpunkt weder bei der Naturwissenschaftlichen Fakultät I – Biowissenschaften der Martin-Luther-Universität Halle-Wittenberg noch einer anderen wissenschaftlichen Einrichtung zum Zweck der Promotion vorgelegt.

Maria Sporbert, Halle (Saale), 26.11.2020