The drivers of dark diversity in the Scandinavian tundra are metric-dependent

Hostens Lore[1], Van Meerbeek Koenraad[2,3], Wiegmans Dymphna[1], Larson Keith[4], Lenoir Jonathan[5], Clavel Jan[1], Wedegärtner Ronja[6], Pirée Amber[1], Nijs Ivan[1], Lembrechts J. Jonas[1]

Affiliations

[1]Research Group Plants and Ecosystems (PLECO), University of Antwerp, Belgium
[2]Department Earth of Environmental Science, KU Leuven, Leuven, Belgium
[3]KU Leuven Plant Institute, KU Leuven, Leuven, Belgium
[4]Climate Impacts Research Centre, Department of Ecology and Environmental Sciences, Umeå University, Sweden
[5]UMR CNRS 7058, Ecologie et Dynamique des Systèmes Anthropisés (EDYSAN), Université de Picardie Jules Verne, Amiens, France
[6]Department of Biology, Norwegian University of Science and Technology, Trondheim, Norway

Correspondence
Lore Hostens, Research Group Plants and Ecosystems (PLECO), University of Antwerp, Belgium
Email: lore.hostens@kuleuven.be
Orcid: https://orcid.org/0000-0001-8245-1152

Jonas J. Lembrechts, Research Center Plants and Ecosystems (PLECO), University of Antwerp, Belgium
Email: jonas.lembrechts@uantwerpen.be
Orcid: https://orcid.org/0000-0002-1933-0750

Funding information

This project was funded by FWO projects G018919N, 12P1819N and W001919N, as well as by ANR-20-EBI5-0004, BiodivERsA, BiodivClim call 2019–2020.

Abstract

Aim: Dark diversity refers to the set of species that are not observed in an area but could potentially occur based on suitable local environmental conditions. In this paper, we applied both niche-based and co-occurrence-based methods to estimate the dark diversity of vascular plant species in the subarctic tundra. We then aimed to unravel the drivers explaining (1) why some locations were missing relatively more suitable species than others, and (2) why certain plant species were more often absent from suitable locations than others.
Location: The Scandinavian tundra around Abisko, northern Sweden.
Methods: We calculated the dark diversity in 107 plots spread out across four mountain trails using four different methods. Two niche-based (Beals' index and hypergeometric method) and two co-occurrences-based (climatic niche model and climatic niche model followed by species-specific threshold) methods. This was then followed by multiple generalized linear mixed models and general linear models to determine which habitat characteristics and species traits contributed most to the dark diversity.
Results: The study showed a notable divergence in the predicted drivers of dark diversity depending on the method used. Nevertheless, we can conclude that plot-level dark diversity was generally 18% higher in areas at low elevations and 30% and 10% higher in areas with a low species richness or low levels of habitat disturbance, respectively.
Conclusion: Our findings call for caution when interpreting statistical findings of dark diversity estimates. Even so, all analyses point towards an important role for natural processes such as

competitive dominance as main driver of the spatial patterns found in dark diversity in the northern Scandes.

Key-words: plant ecology, Beals' index, co-occurrence-model, niche-model, method comparison, plant diversity, regional species pool, plant traits, habitat characteristics

Introduction

Terrestrial ecosystems are increasingly affected by land-use and climate change, leading to large-scale biodiversity loss and community turnover (Theurillat & Guisan, 2001; Mooney et al., 2009; Newbold et al., 2015). Biodiversity plays an important role in ecosystem health and its loss alters ecosystem function (Hooper et al., 2012; Tilman et al., 2014). While most research has focused on the set of species that occur in an area, much less attention has gone to those species that are missing but could potentially inhabit the area (Pärtel et al., 2011). Nevertheless, to get a better understanding of community patterns and their underlying processes, such species absences hold viable additional information (Pärtel, 2014). Knowing which species from the regional species pool are absent within a given locality and identifying why, can help fine-tune conservation planning (Lewis et al., 2017). For example, if many of the absent – yet expected based on climate conditions – species are dispersal limited or cannot access the focal area due to strong dispersal barriers (i.e., habitat fragmentation), then some form of facilitated dispersal through assisted migration or actions to restore habitat connectivity is needed to restore biodiversity. However, if the nutrient conditions in the soil of the focal area are unsuitable for many of the missing species, then only providing assisted migration towards climatically suitable locations or restoring suitable climatic corridors would not be sufficient as restoration measures.

Species belonging to the missing part of the environmentally filtered regional species pool are defined as the so-called "dark diversity" (see Figure 1a), a term introduced by Pärtel et al. (2011). To be part of the dark diversity, the absent species must have a reasonable probability of dispersing to the area (i.e., by belonging to the regional species pool) and its ecological requirements (depending on the methodology either only its climatic or all environmental requirements) must match the local conditions (Pärtel, 2014). As a result, species that are present in the regional surroundings of the focal locality can be locally missing because they have a lower competitive ability, are dispersal limited, are ill-adapted to abiotic conditions, or due to stochastic processes (Riibak et al., 2015). Understanding how extrinsic abiotic conditions and intrinsic species characteristics related to competition and dispersal abilities influence a species' absence can consequently give a better view of the community assembly (Belinchón et al., 2020).

The dark diversity concept does not encompass the total regional species pool across different habitats but focuses on the environmentally filtered, or habitat-specific, regional species pool (Lewis et al., 2017). Combining this habitat-specific regional species pool with the local observed species composition can result in an estimate of the dark diversity (Figure 1). However, there are several methods that use different biotic and abiotic filters to estimate the habitat-specific species pool (Figure 1). Depending on the method, different outcomes will be obtained, as explained below. One of the main benefits of the dark diversity concept is that it allows us to compare the status of biodiversity in different habitats or ecosystems even if the local diversity differs by multiple orders of magnitude (Pärtel et al., 2011).

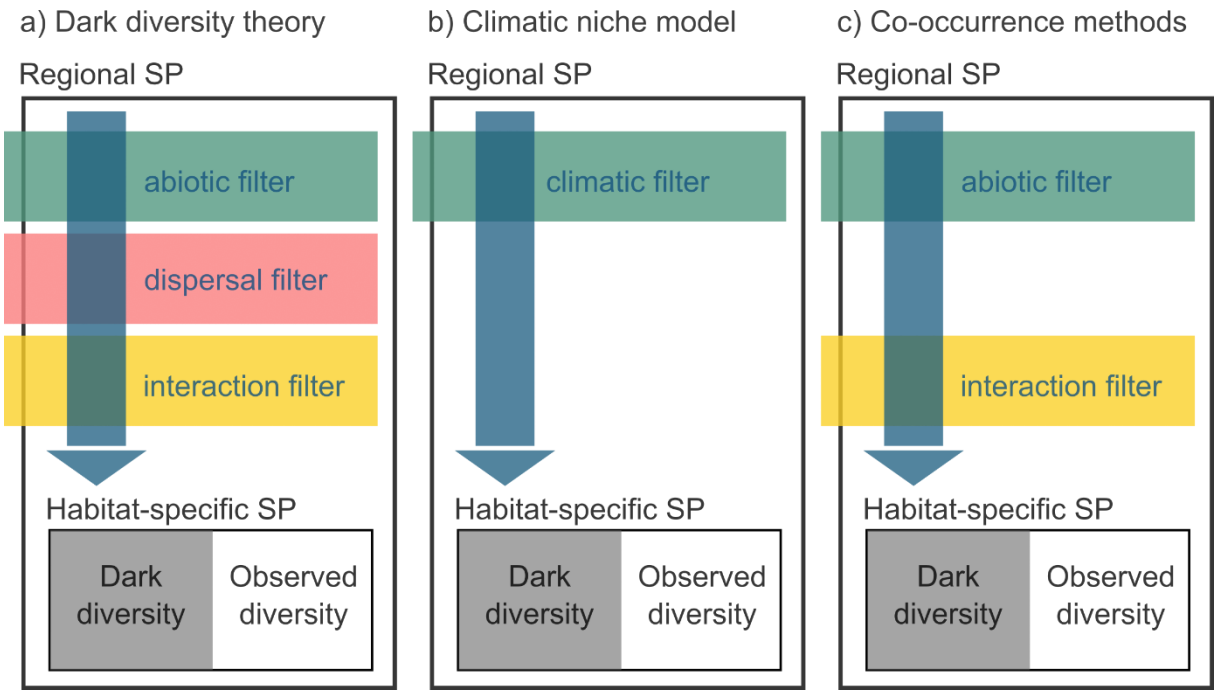

Figure 1: Schematic overview of three approaches to estimate the habitat specific species pool (SP). A) the theoretical concept of dark diversity, where the dark diversity is the non-observed set of species in a certain location, after filtering the regional species pool based on abiotic, dispersal and biotic interaction limitations. In b), dark diversity is calculated using climatic filtering of the regional species pool (e.g. using climatic niche models to estimate which species could occur at a certain location), while c) represents the commonly used co-occurrence based method, which integrates both abiotic and interaction filters. Figure adjusted from Stephenson (2016). The combination of dark- and observed diversity encompasses the habitat-specific species pool.

From a conservation perspective, the fact that species can be part of a community does not always mean that they should be, since not all species are desirable (Lewis et al., 2017). By only looking at suitable environmental conditions, non-native species adapted to those conditions will, for example, also be included. Very often, however, established non-native species have become part of the habitat-specific regional species pool and can, as a result, also be part of the dark diversity in the focal site. Dark diversity can therefore not only be used to investigate unwanted absences and ways to mitigate them, but also to prevent unwanted presences when non-native species are considered part of the habitat-specific regional species pool (Lewis et al., 2017). Monitoring both the dark and observed diversity can then give an early warning signal when non-native species are part of the dark diversity of a location, i.e., even before they are entering the observed diversity. Additionally, the approach can also inform us when species are leaving the dark and entering the observed diversity in response to, for instance, restoration actions (Lewis et al., 2017).

Estimating dark diversity is not straightforward but can be done in multiple ways (Lewis et al., 2016, Figure 1). The difficulty lies in estimating the habitat-specific species pool, which is, as explained above, the set of species in a region that can persist in the environmental conditions of the target site (Pärtel et al., 2011). It encompasses both the observed and dark diversity of a habitat. One could perform extensive sampling of habitat types in a region to estimate the habitat-specific species pool of each habitat type but this can be costly and time-consuming (de Bello et al., 2016). Therefore, computational approaches are often implemented. Most commonly, two types of methods are used to estimate the habitat-specific species pool, either (1) based on the abiotic niche of the species (e.g., using ecological indicator values or species distribution models) or (2) based on metrics of species' co-occurrence with other species in the region (e.g., the Beals' probability index or the hypergeometric method) (Lenoir et al., 2010; de Bello et al., 2016; Carmona and Pärtel, 2020).

Ecological indicator values are often used as an estimate for a species' ecological requirements. The approach allows to identify species from the regional species pool along environmental gradients based on their ecological preferences (Ellenberg et al., 1991). A downside of this method is the difficulty of defining the realized niche of species since such indicator values are rough estimates of the niche optimum along a few specific ecological gradients, often based on expert knowledge (Lewis et al., 2016). Potentially more accurate approaches based on abiotic conditions make use of habitat suitability models to estimate species' environmental niches (Guisan & Thuiller, 2005). These models can be used to determine the environmental conditions suitable for a species (Parolo et al., 2008). In this method, the accuracy of the models highly depends on the resolution as well as on the selected set of environmental data (de Bello et al., 2016).

In both the above-mentioned methods, the aim is to estimate the suitability of a location based only on the environmental niche of the species, regardless of the other species co-occurring in said location. By contrast, one could also estimate the potential of finding a species at a certain location based on the presence of its associated species. The Beals' probability index can be used to calculate species co-occurrence patterns (Beals, 1984). It uses the idea that the presence of a species that is frequently found together with another species could indicate shared suitable abiotic conditions (Ewald, 2002). If the associated species of a target species are observed, but the target species itself is not, it is part of the dark diversity. The hypergeometric method works similarly by verifying if certain species associations occur more often than predicted by chance and by estimating the dark diversity of a given species at a location from the likelihood of its co-occurrence with species present at that location (Carmona & Pärtel, 2020). The advantage of these approaches is that one only requires species observations and no environmental conditions. However, the prediction of the probability of a given species to belong to the dark diversity is dependent on the distribution of other species, which is especially challenging for species that are not strongly confined to particular communities or for environments where traditional communities and thus species associations are truncated (e.g., due to habitat disturbances).

All these approaches have in common that they help identify species that are part of the habitat-specific species pool but that are not recorded in the observed diversity at a certain location, making them part of the dark diversity (Figure 1; Pärtel et al., 2011). In this paper, we applied both niche- and co-occurrence-based methods to estimate dark diversity in a case study on the subarctic tundra around Abisko, northern Sweden. We then further explored the drivers behind the spatial patterns of this dark diversity. The concept of dark diversity is still in its infancy and therefore only a handful of studies have explored why species are part of the dark diversity, none of which were to our knowledge conducted in a tundra climate (Belinchón et al., 2020; Moeslund et al., 2017; Riibak et al., 2015). In this study, we wanted to unravel the drivers behind (1) why some locations are missing relatively more suitable species than others, and (2) why certain vascular plants of the Scandinavian tundra are more often absent from suitable locations than others.

In light of the first research question, we expected locations with a higher relative dark diversity, hereafter referred to as plot-level dark diversity (i.e., a higher percentage of missing species from the habitat-specific species pool) to: (1) appear at lower elevations, as more intense competition will exclude a higher proportion of species (Jones & Gilbert, 2016); (2) be at the extreme ends of disturbance gradients, based on the intermediate disturbance hypothesis (Lembrechts et al., 2014; Rashid et al., 2021); (3) be at the extreme end of low pH and/or moisture gradients, since such conditions can be tolerated by a few species only (Gough et al., 2000; Vonlanthen et al., 2006); or (4) have low observed species richness, as these locations will be dominated by highly competitive species preventing specialist species from co-occurring (Pellissier et al., 2010). Of course, these factors would

act in addition to the stochasticity that always explains part of the variation in species occurrences at small spatial scales (Mohd et al., 2016).

Secondly, we predict that plant species with a higher dark diversity probability, hereafter referred to as species-level dark diversity (i.e., absent in a higher percentage of plots where they were predicted to occur) to: (1) have a higher specific leaf area (SLA), since the Scandinavian tundra is a nutrient-poor environment (Westoby, 1998); (2) have a lower maximum vegetative plant height, as smaller plants would be more easily outcompeted in plots were they could theoretically occur; (3) have a higher seed mass or short-distance dispersal, since these are (loosely) correlated to a limited dispersal ability and lower seed abundance, which decreases the number of successful dispersal events (Howe & Smallwood, 1982; Ozinga et al., 2005); (4) be more recently introduced in the region, as non-native species have a more limited distribution and show possible time-lags in niche filling (Alexander et al., 2016; Crooks, 2005); or finally, (5) be associated with arbuscular mycorrhizal (AM) or ectomycorrhizal (EcM) fungi, as the native vegetation in the region is dominated by ericoid mycorrhizal (ErM) species (Clavel, 2022 unpubl.; Finlay, 2008; Tedersoo, 2017).

2 Materials and methods

2.1 Study area

The field data collection was performed in July and August 2021 in the Abisko area, northern Sweden (68°21'N, 18°49'E). The region has a subarctic montane climate with an average annual temperature of -0.6°C (1913-2020, although average annual temperatures have not dropped below 0°C since 2011) and average annual precipitation of 310 mm (Abisko Scientific Research Station, 400 m above sea level (a.s.l.); https://polar.se/). The soil is comprised of till, colluvium, and glacio-fluvial deposits (Callaghan et al., 2013). At high elevations, the area is covered in snow for about 27 weeks of the year (Callaghan et al., 2013). At low elevations, the vegetation is dominated by open birch forests (*Betula pubescens* Ehrh.), with additional presence of rowan (*Sorbus aucuparia* L.) and several willow species (*Salix* sp.). The understory vegetation often consists of heath species (e.g., dwarf birch (*Betula nana* L.), European blueberry (*Vaccinium myrtillus* L.) and black crowberry (*Empetrum nigrum* L.)), or meadow species (e.g., Alpine bistort (*Bistorta vivipara* L.), globeflower (*Trollius europaeus* L.) and Alpine saw-wort (*Saussurea alpina* DC.)) (Sonesson & Lundberg, 1974). Above the treeline (520 m a.s.l), the vegetation is dominated by alpine/arctic heathland species (e.g., blue heath (*Phyllodoce caerulea* L.), bog blueberry (*Vaccinium uliginosum* L.) and lingonberry (*Vaccinium vitis-ideae* L.)) (Kullman, 2015).

2.2 Field data collection

2.2.1 Study sites

A total of 107 plots were surveyed in the vicinity of four mountain trails: Björkliden, Låktatjåkka, Nuolja, and Rallarvägen (Figure 2).

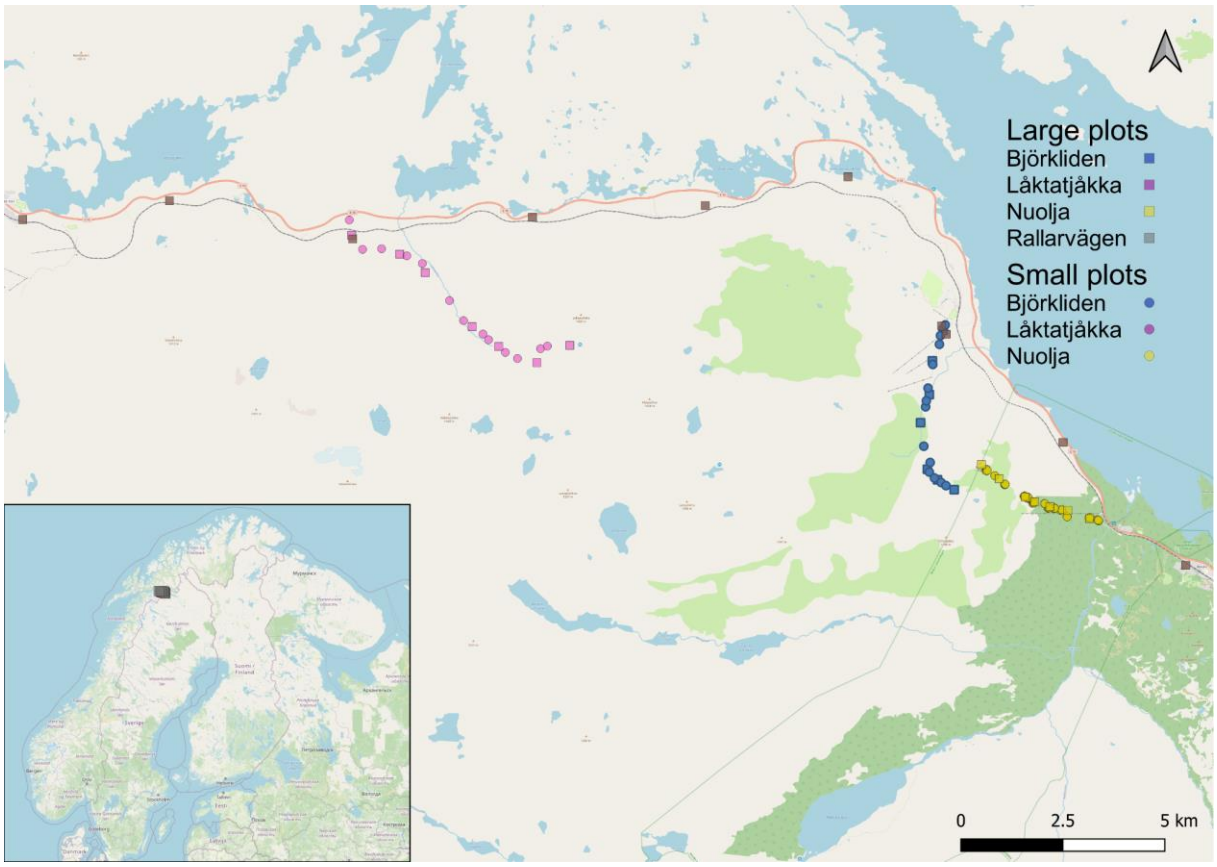

Figure 2: Map of the study area around Abisko, Sweden (grey dot on the inset), with 107 surveyed plots along the four hiking trails (colors) and the different survey methods (symbols).

Data from new and ongoing vegetation surveys were combined, with two different methodologies: 73 1 × 1 m² plots from a long-term vegetation composition monitoring project in the area (hereafter called 'small plots'), as well as 34 large (10 × 10 m²) plots established in the framework of the global DarkDivNet network (Pärtel et al., 2019). Of the 107 plots, two times 20 were located along trails close to Björkliden and around Låktatjåkka (Wedegärtner et al., 2022), 57 in the Abisko National Park on Mount Nuolja (MacDougall et al., 2021), and 10 along the Rallarvägen.

2.2.2 Large plots

The vegetation monitoring method used in the large plots was based on the DarkDivNet protocol (Pärtel et al., 2019). The plots (10 × 10 m²) were placed at a 10 m perpendicular distance from the trail. In each plot, all vascular plants were recorded. Species were identified using the Fjällflora (Mossberg & Stenberg 2008). Observations that could not be identified to the species level (e.g., *Alchemilla* sp.) were removed from the species list and thus also from the regional species pool. Furthermore, following the DarkDivNet protocol, the maximum vegetative height (cm) was measured with a ruler for the tallest individual of each species in all plots.

In every plot, we visually estimated the cover (%) of total vegetation, bare ground, rock, litter, herbaceous vegetation, bryophytes, lichen, shrubs, and trees (> 200 cm). At the center of every plot, the exact location was recorded with a hand-held GPS receiver. Soil samples were collected using the protocol explained below (see 2.3).

2.2.3 Small plots

The small plots were surveyed using the pin-point or point intercept method, which is often used to assess plant cover (Jonasson, 1988). A 1 × 1 m² plot was placed at 10 m from the trail. In one plot, 100 pins were vertically dropped in 10-cm increments from left to right and top to bottom. With every pin-drop, we recorded the vascular plant species touching the pin, multiple recordings for the same species occurred when more than one individual of that species touched the pin. When the pin touched only the ground, the observation was categorized as either litter, bryophytes, bare soil, or lichen, a single hit was noted. Soil samples were collected using the same protocol as explained below (see 2.3).

2.3 Soil sample analysis

Soil samples were collected in 50 out of the 107 plots (both large and small plots). During sampling, the litter covering the soil was removed and a minimum of 300 g of soil was taken from the top 10 cm of the ground. Soil samples could not be collected along the Nuolja trail (57/107 plots) as this trail is in the Abisko National Park and no sampling permission was obtained in the year of the survey. However, 50 of these plots were long-term permanent plots for which soil pH measurements were available from previous soil sampling campaigns conducted in 2018 (using the same sampling and analysis procedure). The seven remaining plots were in very close (<10 m) proximity to small plots for which pH was measured in 2018, and we therefore used the mean pH of those plots. Ultimately, pH could be obtained for all but one plot, assuming that when largely undisturbed – as was the case in the system – pH-values would only change slightly over time.

All soil samples were stored in a fridge at 4°C until they were analyzed between September and December 2021 at the University of Antwerp, Belgium. To measure soil pH, 25 mL of a KCl solution was added to 10 g (9.9-10.1 g) of soil. The samples were put in a shaker for an hour and afterward rested for another 60 min. Then, soil pH was measured with a 914 pH/Conductometer by Metrohm© in the liquid layer at the top of the sample after shortly manually shaking the tubes.

2.4 Online data collection

2.4.1 Gridded data products

To create the climatic niche models, we collected gridded climate data with a resolution of 30 arcseconds (c. 1 × 1 km² at the equator) for the annual mean air temperature, annual precipitation, the mean maximum air temperature of the warmest month, and the mean minimum air temperature of the coldest month from CHELSA version 1.2, representing the long-term (1979-2013) climatic conditions (Karger et al., 2017).

Soil temperature estimates (i.e., annual mean soil temperature, mean soil minimum temperature of the coldest month and mean soil maximum temperature of the warmest month) were obtained from the SoilTemp global maps of soil temperature (Lembrechts et al., 2021). These maps calculated the soil bioclimatic variables using CHELSA monthly air temperature maps and the offset between gridded air temperature and in-situ soil temperature measurements from the SoilTemp database (Lembrechts et al. 2020). The gridded data, representative of the upper soil layer (top 5 cm), had the same resolution as the CHELSA data, namely 30".

Elevation was extracted from the European Digital Elevation Model (DEM) with a resolution of 25 m, obtained from Copernicus Land Monitoring Service version 1.1 (European Union, 2021).

Lastly, the Topographic Wetness Index, a topographical proxy for soil moisture, was obtained from a TWI raster across Scandinavia, data acquired from Haesen et al. (2021). The TWI raster, which had a spatial resolution of 25 m, was generated using the method developed by Kopecký et al. (2021).

All data were extracted to the according coordinates of every plot in R version 4.2.1 (R Core Team, 2021) using the raster (Hijmans et al., 2012), sp (Pebesma et al., 2005), and GDAL (Mitchel et al., 2014) packages.

2.4.2 Type of disturbance

For every plot, we assigned a type of disturbance based on its proximity to hiking trails, roads, and railroad. By visual assessment in QGIS, one of the three disturbance types (hiking trail, road or railroad) was assigned to every plot. All plots were close to hiking trails, yet whenever the railroad or a road was within 150 m of the plot, its impact was considered dominant, and the hiking trail classification thus overruled.

2.4.3 Amount of bare ground

The amount of bare ground (%) as a proxy of disturbance was estimated or calculated for every plot. For the large plots, this was estimated from the percentage cover of litter and bare ground. This was calculated for the small plots by summing up all the pins that touched bare ground and litter, dividing this by the total number of pins in a plot.

2.4.4 Plant functional traits

Average maximum vegetative plant height per species was calculated from the measurements done in the large plots.

The SLA for every species was retrieved from data collected in the framework of the Mountain Invasion Research Network (MIREN) in the region in 2017 (published as part of the Tundra Trait Team database (TTT); Bjorkman et al., 2018). The SLA was calculated as leaf area (cm^2)/dry weight (g). Whenever there were two or fewer observations (30% of species, n=15) or no data at all (13%, n=6), the data was completed with values from the global Tundra Trait Team database (Bjorkman et al., 2018). Ultimately, for 28% (n=14) of the species, we extracted data from the TTT database and for 2% (n=1) of the species there was no data available. Within the TTT database, only single measurements on an individual - and thus not e.g., site-specific means - were used.

Average seed mass per species was obtained from the global TTT database or – if not available there - the LEDA Traitbase (Bjorkman et al., 2018; Kleyer et al., 2008). Again, from the TTT database, only 'single measurements on an individual' were used, whereas from LEDA only 'actual measurements on individuals' (no estimations) were used. The TTT database contained seed mass for 28% (n=14) of the species. An additional 47% (n=23) of the species could be supplemented with seed mass from the LEDA database.

The dispersal type per species was also retrieved from the LEDA Traitbase and used to categorize species according to their potential for long-distance dispersal (LDD) and short-distance dispersal (SDD) (Kleyer et al., 2008). For 86% (n=42) of the species this data was available and all were considered long-distance dispersers, hence this variable was not included in further analyses.

2.4.5 Nativeness Index

We used a continuous rather than a binary measure of the status of a species within a region, to get a more accurate view of the history of the species. Our Nativeness Index (NI) used historical surveys from the Global Biodiversity Information Facility (GBIF) database. It considered the first year a species was observed (year first occurrence species) and the first year in which more than 50 species were observed in the region (year first survey). If the NI was close to 1, the species was already observed at the time of the first survey. As the value approached 0, the species was observed increasingly recently for the first time and was thus more likely to be non-native.

$$NI = \frac{\sqrt{year\ (2020) - year\ first\ occurrence\ species}}{\sqrt{year\ (2020) - year\ first\ survey\ (1850)}}$$

Square roots were used in the formula to give more weight to recent differences (e.g., a first observation in 2010 vs 2020 is considered a more substantial difference than one in 1900 vs 1910). The first occurrence and the year of the first survey were obtained using the *rgbif* package (Chamberlain et al., 2021).

2.4.6 Mycorrhizal associations

The association of plant species with the main types of mycorrhizal fungi (AM = arbuscular mycorrhiza, EcM = ectomycorrhiza, ErM = ericoid mycorrhiza and NM = no mycorrhiza) was retrieved from the FungalRoot database (Soudzilovskaia et al., 2020). This data was available for 92% (n=42) of the species. For the remaining 8% (n=7) for which species-level information was absent, the mycorrhizal type was derived at the family level (i.e., the most common mycorrhizal type within the family).

2.5 Data-analysis

All statistical analyses were conducted in R version 4.2.1 (R Core Team, 2021).

2.5.1 Dark diversity modeling

For further analysis, only the most common species, i.e., species with 10 or more observations, were included (n=49), as sufficient observations were needed to calibrate climatic niche models and build co-occurrence matrices. We then used the same dataset in four different approaches to estimate dark diversity.

Climatic niche modeling

The presence and absence of all species in every plot was used to make climatic niche models. For every species, a Generalized Linear Model was made, with a binomial distribution containing all climatic variables and their quadratic terms as explanatory variables (i.e., annual mean air temperature, annual precipitation, maximum air temperature of the warmest month, minimum air temperature of the coldest month, annual mean soil temperature, minimum soil temperature of the coldest month, and maximum soil temperature of the warmest month) and presence/absence (1/0) of a species per plot as the response variable. Multicollinearity was checked using the Variance Inflation Factor (VIF) from the *car* package (Fox & Weisberg, 2019) and variables that increased the VIF to 5 or more were removed. The final models contained: annual precipitation, minimum soil temperature of the coldest month, maximum soil temperature of the warmest month, and their quadratic terms. No further model selection was done as we were not interested in a model identifying the drivers of the species' climatic niche, but rather wanted to approximate their climatic niche as consistently as possible.

To predict the probability of a species' occurrence in a specific plot, the GLM was fit on the remaining plots (Lembrechts et al., 2019) and the probability was estimated for that specific plot. This leave-one-out procedure was then repeated for all plots and all species. We then calculated the relative dark diversity per plot by averaging the predicted presence of each absent species in a plot and the dark diversity probability per species by averaging the predicted presence of a species across all plots where it was absent.

The second method to estimate the dark diversity used the same climatic niche model as above. Yet, instead of continuous probability estimates, we converted niche model predictions into presence/absence estimates. For this, we calculated species-specific thresholds for presence using the function *ecospat.max.tss* from the *ecospat* package (Broennimann et al., 2022) which chooses the threshold that maximizes values for the True Skill Statistic (TSS), which assesses the accuracy of species distribution models (Allouche et al., 2006). Based on this, we created a binary dataset where the values below the threshold got a 0 (predicted to be absent) and the values above got a 1 (predicted to be present). Afterward, we removed the values where the species was observed to be present based on the vegetation surveys. To calculate the species-level dark diversity probability, we used the formula proposed by Moeslund et al. (2017), using the number of plot-level observations and predictions:

$$\frac{\#\ times\ in\ dark\ diversity}{\#\ times\ in\ species\ pool}$$

To calculate the relative plot-level dark diversity:

$$\frac{\#\ species\ in\ dark\ diversity}{\#\ species\ in\ species\ pool}$$

The habitat-specific species pool consisted of both the observed and dark species. Note that at the species level, we are estimating the probability that a species belongs to the dark diversity (dark diversity probability), while at the plot-level, we are estimating the percentage of species from the species pool that is absent (dark diversity *per se*).

Beals' method

Two co-occurrence-based methods to estimate the dark diversity were used, with the first being the Beals' index (Beals, 1984), as applied by Lewis, Szava-Kovats & Pärtel (2016). We first built a species co-occurrence matrix, then calculated the Beals' index, using the *beals* function from the *vegan* package, for each species in every plot, excluding the focal species as suggested by Oksanen et al. (2022). The thresholds used to decide whether a species was part of the regional species pool were species-specific and defined as the 5th percentile of the Beals' index value for the species (Gijbels, Adriaens & Honnay, 2012). Before calculating each threshold, the lowest value of the Beals' index was determined among the plots containing occurrences of the species in question, and all plots with values below this lowest value were discarded (Moeslund et al., 2017). For each plot, the dark diversity then consisted of all species from the habitat-specific species pool, except those present (Pärtel, Szava-Kovats & Zobel, 2011). To calculate the plot- and species-level dark diversity probability the same formulae as for the species-specific threshold were used.

Hypergeometric method

The second method used to estimate the dark diversity was the hypergeometric method, as proposed by Carmona & Pärtel (2020). This method avoids the binary form in which dark diversity is often defined. The co-occurrence matrix used for the Beals' method was also employed in this case. To get estimates of the dark diversity, we used the function *DarkDiv* from the *DarkDiv* package, with the

argument 'method' containing 'Hypergeometric' (Carmona & Pärtel, 2020). We applied this method to all species in all plots for which we obtained a probability that the species could be present in that plot. Afterward, all values for plots where the species were observed to be present were removed. To calculate the relative plot-level dark diversity, per plot the mean was taken from the remaining values. The same was done for the species-level dark diversity, yet here the mean was taken per species.

2.5.2 Drivers of relative plot-level dark diversity

To investigate why certain plots had a higher relative dark diversity, we created generalized linear mixed models (GLMMs) with a beta distribution and logit-link function using the *glmmTMB* package (Brooks et al., 2017), with predictions from each of the four dark diversity indices (the two approaches based on niche models, the Beals' index, and the hypergeometric approach) as a response variable.

These plot-level models contained elevation, soil pH, type of disturbance, amount of bare ground, TWI, observed species richness and plot size as explanatory variables. The plots were situated along various trails. To account for this hierarchical sampling design, the model included a random intercept for plot number nested within trail identity. Multicollinearity and distribution of residuals were checked using the Variance Inflation Factor (VIF) and the *DHARMa* package (Hartig, 2022) and deemed not violated. Due to the low sample size, we limited ourselves to linear patterns and did not include two-way interactions since these more complex models could not converge. For the same reason, quadratic effects were not tested, even though theoretically they could be expected for pH and soil moisture. However, within our study system both the pH and moisture gradient only reached extreme values on one side of the gradient (e.g., highly acidic yet no highly basic soils).

No further model selection was performed (Hartig, 2018). The variance explained by the full model was obtained using the *performance* function from the *performance* package (Lüdecke et al., 2021). To determine the proportion of explained variance of every variable, we followed a variation partitioning approach. First, the variance of the full model was calculated. Afterward, for every explanatory variable, a model was made consisting of all variables except the focal variable. By extracting the marginal R^2 of the individual models from the R^2 of the full model, the variance of the focal variable was obtained (Legendre & Legendre, 1998).

2.5.3 Drivers of species-level dark diversity probability

To investigate why certain species had a higher dark diversity probability, we created general linear models (GLMs) with a beta distribution and logit-link function using the *betareg* package (Cribari-Neto & Zeilis, 2010) with predictions from each of the used dark diversity indices (based on the niche models, the Beals' index, and the hypergeometric approach) as a response variable.

First, full models were made separately for each dark diversity index that contained the nativeness index, maximum vegetative plant height, specific leaf area, dispersal type, seed mass, and mycorrhizal association as explanatory variables and species-level dark diversity as the response variable. Assumptions of multicollinearity and distribution of residuals were tested and not violated. Here as well, two-way interactions could not be tested and no further model selection was performed (Hartig, 2018). Afterward, pairwise comparisons were conducted on the categorical parameters using the *emmeans* package (Lenth, 2022).

3 Results

3.1 Plot-level dark diversity

Depending on the method, we could explain between 39% and 80% of the variance in plot-level dark diversity. In one case (climatic niche models), elevation was responsible for the largest share, while in the three other cases (species-specific threshold, Beals' index and hypergeometric method) species richness was the most dominant factor (Figure 3). On average across all models, elevation explained 11%, species richness 10%, and plot size, type of disturbance, amount of bare ground, pH and TWI an additional 2%, 2%, 3%, 4% and 1%. Note that due to the nature of the variance partitioning calculations, variances do not necessarily add up to the total variance of the full model.

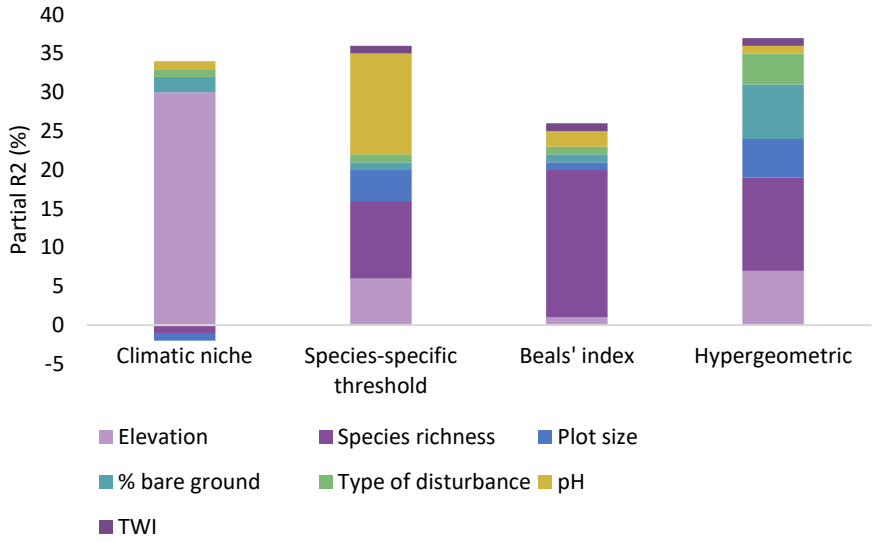

Figure 3: Variance partitioning (expressed in % and calculated using the marginal R^2) of the different explanatory variables in the GLMMs of the plot-level analyses on the predictions of each of the four different dark diversity methods. TWI = Topographic Wetness Index.

In three out of the four methods used, the plot-level dark diversity decreased significantly across the elevation gradient (Table 1; Figure 4). Only in the model based on the Beals' index did elevation not have a significant influence (Table 1; Figure 4D)

*Table 1: Models explaining the plot-level dark diversity using the different dark diversity estimation methods: coefficients (p-values: * p≤0.05; ** p≤0.01; *** p≤0.001). The factor used for the intercept was allocated alphabetically and all other factors were compared to this baseline. CN = Climatic Niche models; SS = Species-specific threshold; Hyper = Hypergeometric method; Beals = Beals' index; Elev = elevation; SR = Species Richness; TOD = Type Of Disturbance; TWI = Topographic Wetness Index.*

Model	**Intercept (Road)**	**Elev**	**SR**	**Plot size (10x10 m^2)**	**TOD Hiking trail**	**TOD Railroad**	**% bare ground**	**pH**	**TWI**	**AIC**
CN	-0.327	-0.001***	-0.029***	0.079	0.524	-0.251	-0.001*	-0.022	-0.014	-370
SS	3.00***	-0.001**	-0.105***	-0.408*	0.281	-0.165	-0.001	-0.262***	0.038	-140
Hyper	0.102	-0.001**	0.001***	-0.183	0.425	0.001	-0.012***	0.065	-0.011	-230
Beals	1.04*	10^{-4}	-0.082***	-0.121	-0.112	-0.345	0.001	0.032	0.001	-195

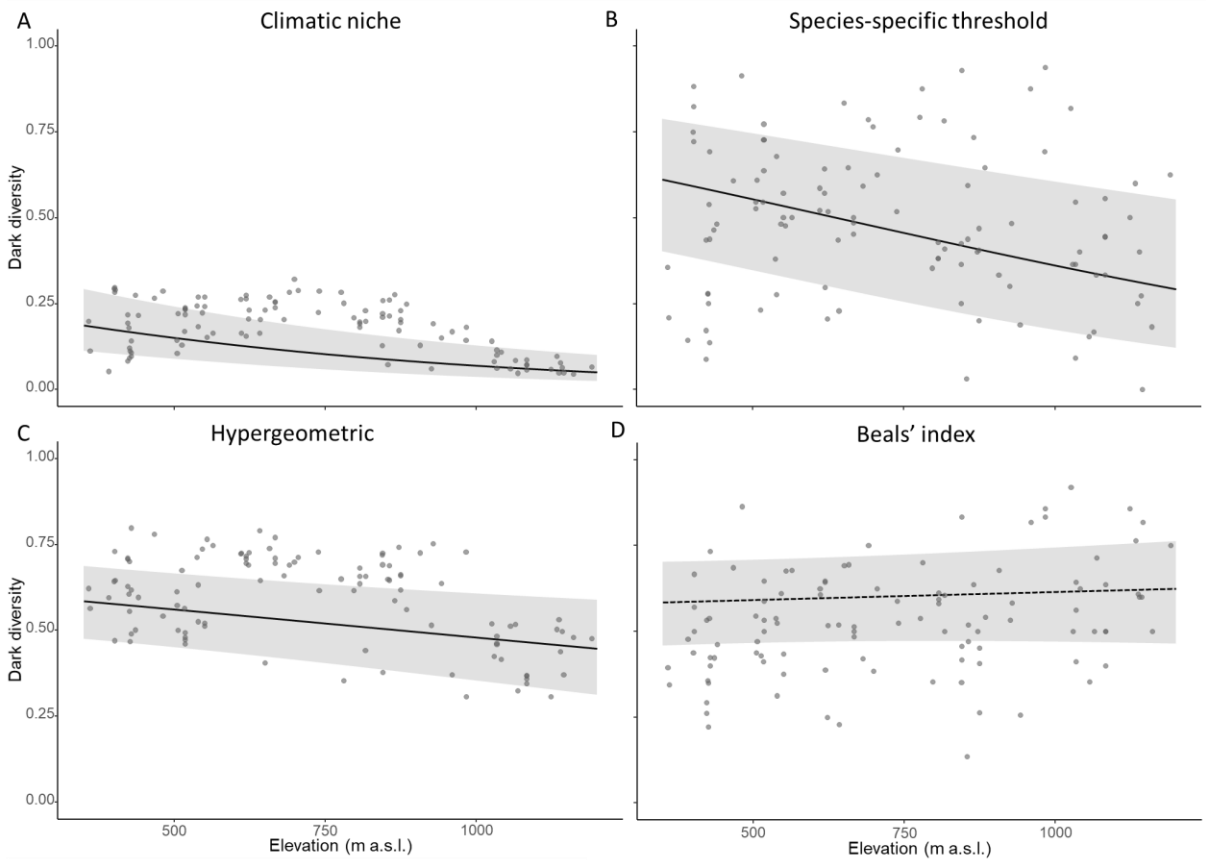

Figure 4: Marginal effects plots of the plot-level dark diversity as a function of elevation (m a.s.l.). The grey area indicates the 95% confidence interval, the grey dots are the raw data points, and the dotted line indicates non-significance. Dark diversity estimated using A) the climatic niche models, B) the climatic niche models followed by the species-specific threshold, C) the hypergeometric method and D) the Beals' index.

The plot-level dark diversity decreased significantly with increasing species richness in three cases (Table 1; Figure 5A, 5B, 5D), yet increased significantly with increasing species richness when using the hypergeometric method (Table 1; Figure 5C).

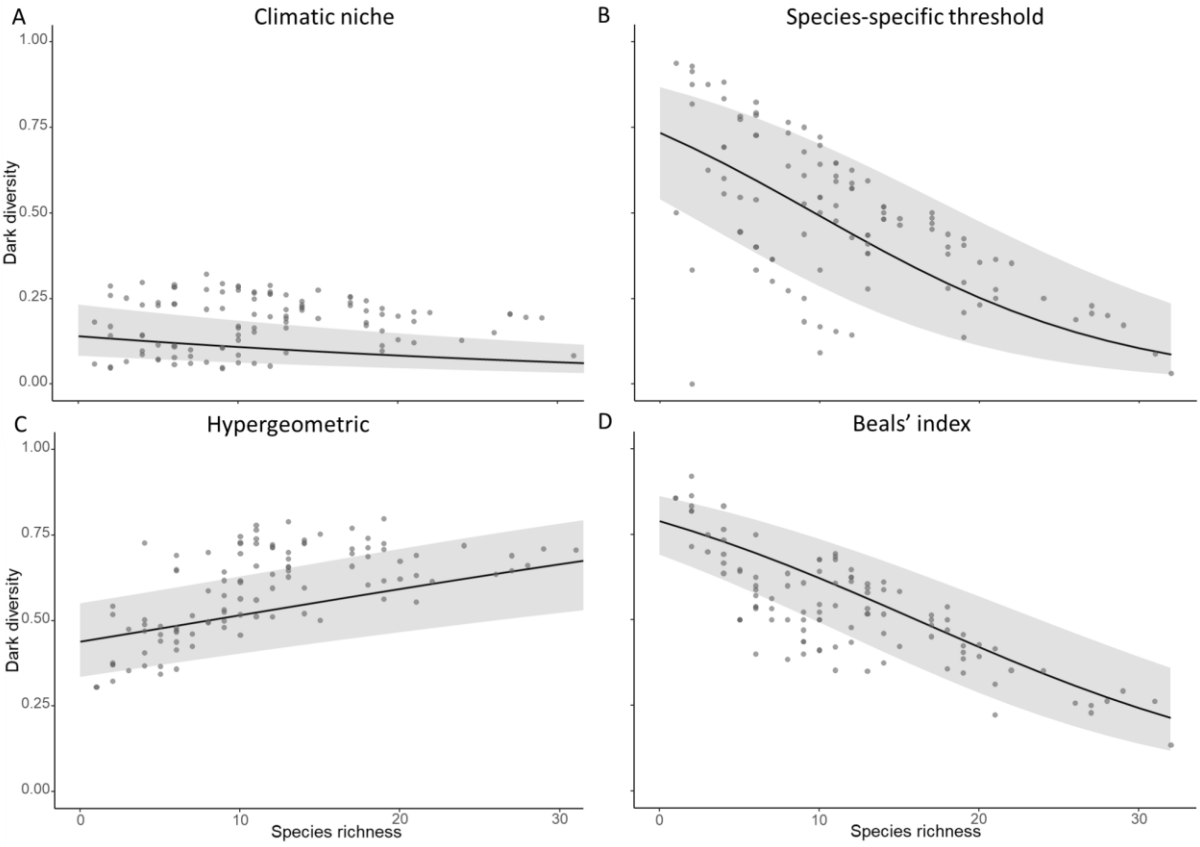

Figure 5: Marginal effects plots of the plot-level dark diversity as a function of species richness. The grey area indicates the 95% confidence interval, and the grey dots are the raw data points. A) the climatic niche models, B) the climatic niche models followed by the species-specific threshold, C) the hypergeometric method and D) the Beals' index.

Moreover, the dark diversity decreased significantly with increasing amounts of bare ground when using the hypergeometric method (Table 1; Figure 6B) and the climatic niche models (Table 1; Figure 6A). In these models there was also some support that plots near a hiking trail showed a slightly higher dark diversity probability than plots closer to a road or railroad, although not significantly so (climatic niche: z = 1.65, *p = 0.099*; hypergeometric: z = 1.80, *p = 0.071*; Table 1; Figure 6C, D). The larger plots (10 x 10 m^2) also had a significantly lower dark diversity compared to the smaller plots (1 x 1 m^2) when using the climatic niche models followed by the species-specific threshold (Table 1; Figure 6E) and the hypergeometric method (Figure 6F), although in the latter not significantly (Table 1). Lastly, when using the climatic niche model followed by the species-specific threshold, the dark diversity decreased significantly with increasing pH (Table 1; Figure 6G), while the opposite was true for the hypergeometric method, albeit not significantly (z = 1.81; *p = 0.057*) (Table 1; Figure 6H).

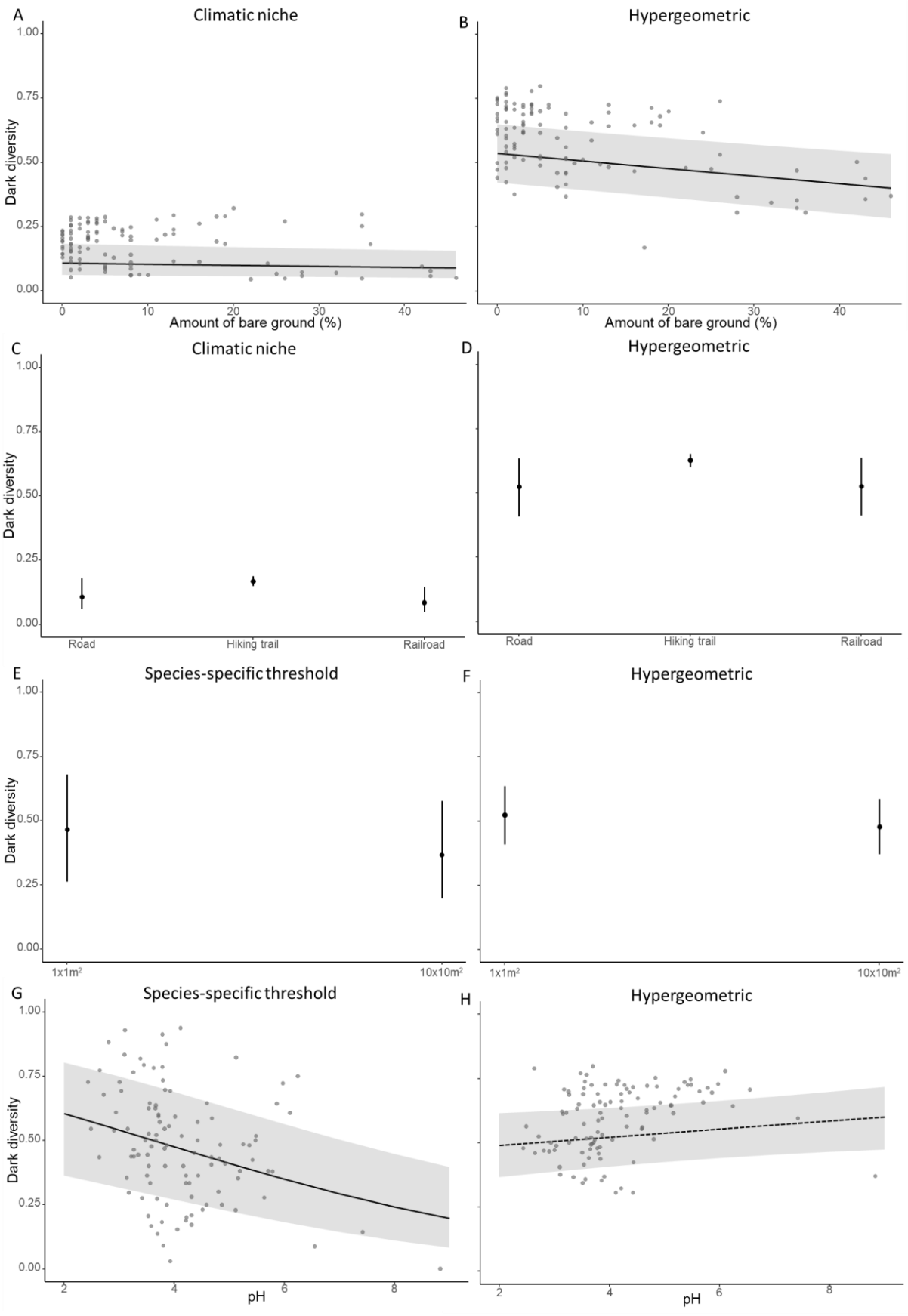

Figure 6: Marginal effects plots of the plot-level dark diversity as a function of the A/B) amount of bare ground, C/D) type of disturbance, E/F) plot size and G/H) pH. The black dots show the average dark

diversity per individual factor whereas the error bars show the standard deviation. The grey area indicates the 95% confidence interval, the grey dots are the raw data points and the dotted line indicates non-significance. Dark diversity was estimated using A/C) the climatic niche models, E/G) the climatic niche models followed by the species-specific threshold and B/D/F/H) the hypergeometric method.

3.2 Species-level dark diversity

Depending on the method, we could explain between 15% and 45% of the variance in species-level dark diversity (Figure 7). In all cases, mycorrhizal association was responsible for the largest share (Figure 7). On average across all models, mycorrhizal association explained 18%, seed mass 8%, specific leaf area 3% and the nativeness index and the maximum vegetative plant height an additional 2%. Note that due to the nature of the variance partitioning calculations, variances do not necessarily add up to the total variance of the full model.

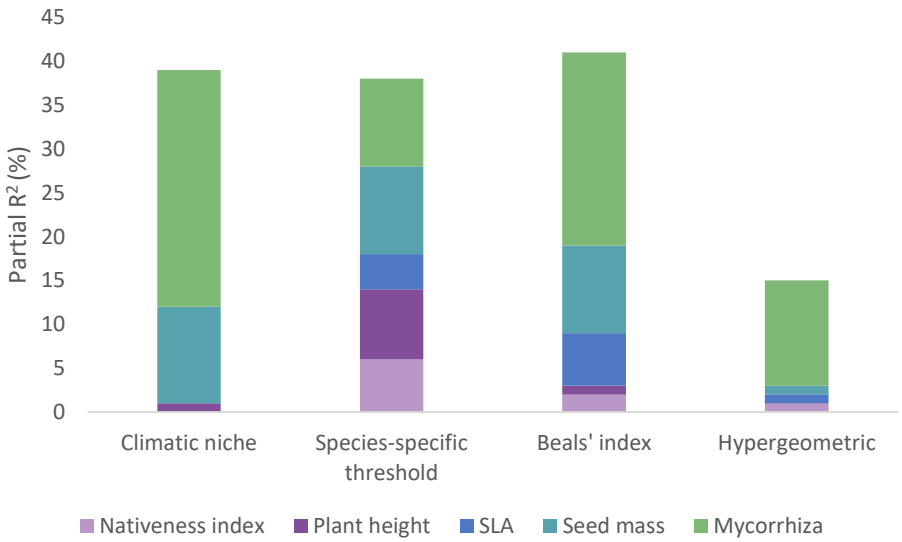

Figure 7: *Variance partitioning (expressed in % and calculated using the marginal R^2) of the different explanatory variables in the GLMMs of the species-level analyses on the predictions of each of the four different dark diversity methods. SLA = Specific Leaf Area.*

Mycorrhizal status was the only significant parameter in the climate niche model approach, with ericoid mycorrhizae differing significantly from AM, EcM and NM (Figure 8; Table 2; Appendix S1). Species with a symbiotic ericoid mycorrhizal association had a significantly higher dark diversity than all other associations when using the climatic niche models (Table 2; Figure 8A). However, the opposite was true when using the three other methods (Table 2; Figure 8B, C, D). For the Beals' index the contrast test also revealed ericoid mycorrhizae differing significantly from AM, EcM and NM (Figure 8D, Appendix S1). For the other two methods, the contrast test only showed a borderline significant difference between ErM and NM for the species-specific threshold and ErM AM for the hypergeometric method (Figure 8B, C; Appendix S1).

*Table 2: Models explaining the plot-level dark diversity using the different dark diversity estimate methods: coefficients (p-values: * p≤0.05; ** p≤0.01; *** p≤0.001). The factor used for the intercept was allocated alphabetically and all other factors were compared to this baseline. CN = Climatic Niche model; SS = Species-specific threshold; Hyper = Hypergeometric method; Beals = Beals' index; AM = Arbuscular Mycorrhiza; EcM = Ectomycorrhiza; ErM = Ericoid Mycorrhiza; NM = No Mycorrhiza; MVH = Maximum Vegetative plant Height; SLA = Specific Leaf Area; NI = Nativeness Index; SM = Seed Mass.*

Model	**Intercept (AM)**	**EcM**	**ErM**	**NM**	**MVH**	**SLA**	**NI**	**SM**	**AIC**
CN	-0.043	0.108	1.17***	-0.101	-10^{-4}	-10^{-4}	-1.37	-0.068	-72
SS	-5.96	0.453	-0.916*	0.208	-10^{-3}	-10^{-3}	6.95	0.082	-11
Hyper	0.102	-0.167	-0.965*	-0.067	10^{-4}	0.001	1.20	-0.040	-12
Beals	1.04*	0.170	-1.17***	0.193	-10^{-4}	-0.001	-3.28	-0.089	-19

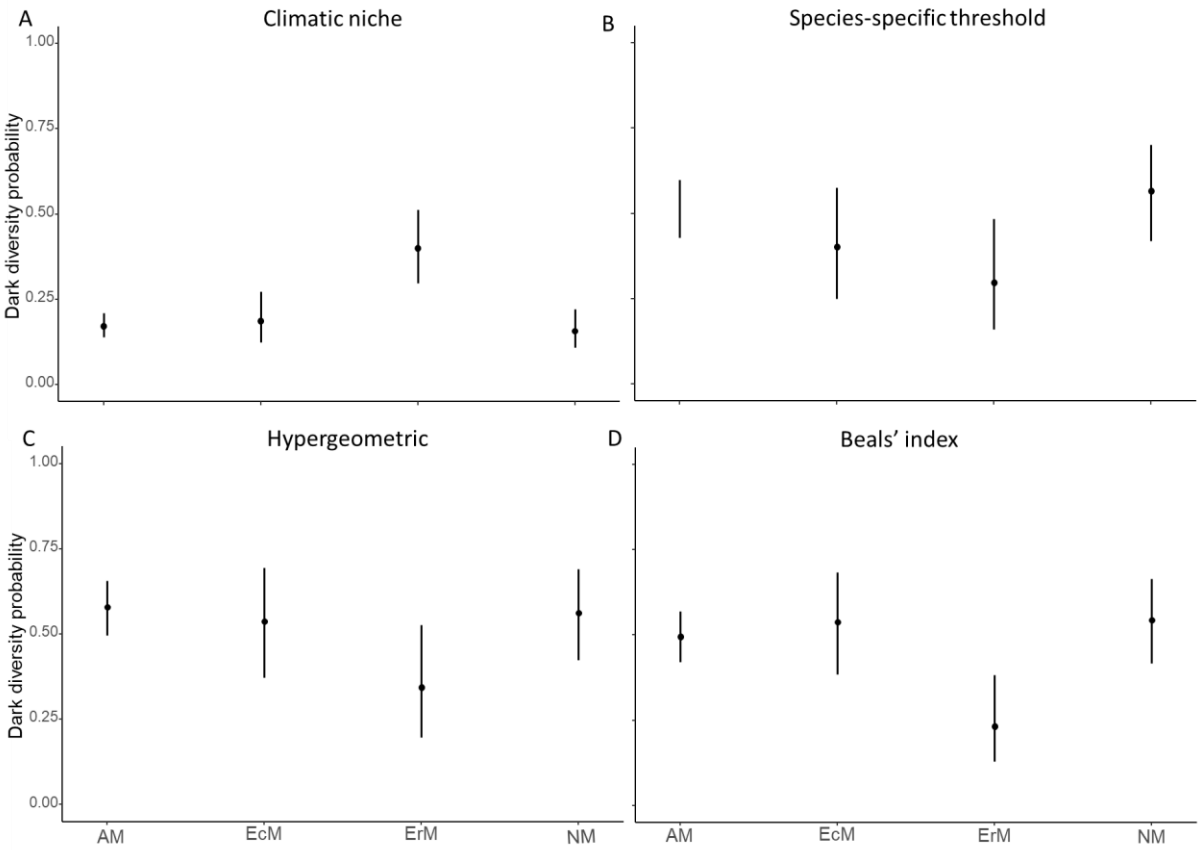

Figure 8: Prediction of the species-level dark diversity in relation to the type of mycorrhizal association based on the beta regression model. The black dots show the average dark diversity per individual factor whereas the error bars show the standard deviation. AM = Arbuscular Mycorrhiza; EcM = Ectomycorrhiza; ErM = Ericoid Mycorrhiza; NM = No Mycorrhiza. Dark diversity estimated using A) the climatic niche models, B) Species-specific threshold, C) the hypergeometric method and D) the Beals' index.

Even though none of the other parameters had a significant influence on the species-level dark diversity, there is some support that the species-level dark diversity increased with increasing nativeness index (Figure 9B; z = 1.84, *p = 0.067*) and decreased with increasing maximum vegetative plant height (Figure 9A; z = -1.75, *p = 0.080*) when using the climatic niche model followed by the species-specific threshold, albeit both not significantly so. Lastly, there is also an indication (z = -1.86, *p = 0.062*) that the dark diversity decreased with increasing SLA when using the Beals' index (Figure 9C).

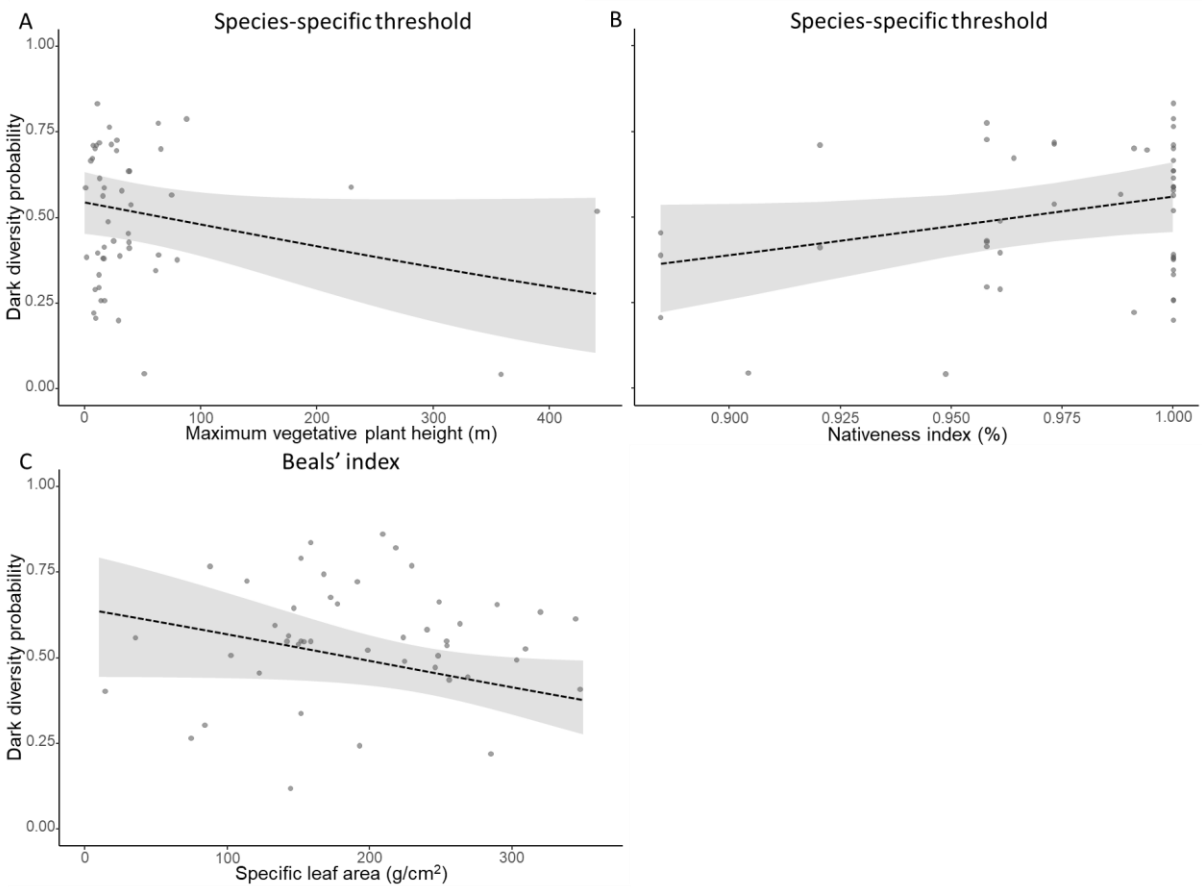

Figure 9: Marginal effects plots of the species-level dark diversity as a function of the A) Maximum vegetative plant height, B) Nativeness index and C) Specific leaf area. The dotted line indicates non-significance whereas the grey area indicates the 95% confidence interval and the grey dots are the raw data points. Dark diversity was estimated using A/B) the climatic niche models followed by the species-specific threshold, C) the Beals' index.

4. Discussion

4.1. Plot-level dark diversity

We found relatively consistent patterns in the drivers of dark diversity at the plot-level, but much less consistency was observed at the species level. Plot-level dark diversity was most consistently related to elevation, with plots at higher elevations having a lower plot-level dark diversity - and thus fewer expected species missing - than plots at lower elevations. This was true for both niche-based methods as well as for the hypergeometric method, yet not for the Beals' index, in which elevation was not significant. Such a decline with elevation is in line with ecological theory. Indeed, under harsh environmental conditions, competitive interactions are often replaced by mutualistic ones, or competition is at least lowered in intensity, thereby reducing the exclusion of less competitive species with a lower dark diversity as a result (Callaway et al., 2002; Klanderud, 2010; Lembrechts et al., 2018). Additionally, the presence of more ruderal and competitive species in the lowlands compared to the stress-tolerant species higher up in the mountains along roadsides also suggests that reduced competition can be one of the main drivers behind the lower dark diversity at higher elevations (Lembrechts et al., 2014). Furthermore, climatic conditions are usually milder in the lowlands, making them suitable for a broader set of species (Körner, 2021). Consequently, since more species can be

549 present in these plots, it is also more likely that at least some of them are excluded, resulting in a
550 higher number of species belonging to the dark diversity. As the co-occurrence-based metrics
551 accounted for some of these factors (e.g., lower expectancy of species in plots dominated by species
552 that traditionally outcompete them), it should come as no surprise that elevation was not significant
553 in the model for the Beals' index, and that the decline in dark diversity with elevation was the least
554 steep for the hypergeometric method.

555 Species richness was identified as a key driver of plot-level dark diversity with all four methods. Its
556 effect was negative for all but the hypergeometric method, thus largely following our hypothesis. In
557 this system, plots with a low number of species are likely to be dominated by highly competitive
558 species, which will prevent the establishment of several species that could in theory occur there
559 (Pellissier et al., 2010). Indeed, plots with a low species richness in the study system were often
560 dominated by black crowberry. It is an efficient competitor for nutrients, can grow on soils with low
561 pH, and has allelopathic effects against seed germination and the growth of surrounding species
562 (Tybirk et al., 2000), thus effectively referring several species from the regional species pool locally to
563 the dark diversity. Nevertheless, it is possible that approaches based on species co-occurrences, such
564 as the hypergeometric method and the Beals' index, already account for this effect of competition.
565 The latter potentially explains the observed positive relationship between species richness and dark
566 diversity when using the hypergeometric method. Given the consistent negative relationship of dark
567 diversity with elevation (see above), it is not unlikely that the effects of species richness and elevation
568 would interact, with higher dark diversity especially in lowland plots with higher species richness, yet
569 we lacked the sample size to explore such interactions statistically.

570 Finally, we found some support for the type of disturbance and percentage of bare ground as drivers
571 of the plot-level dark diversity, while the results for soil pH were largely inconclusive. In two models,
572 dark diversity tended to be lower in plots closer to the road - though none significantly so. Similarly, in
573 two methods the dark diversity was found to decline significantly with increasing bare ground cover.
574 Both point in the direction of a decreased dark diversity with increasing disturbance, a finding in line
575 with other studies in the region. Indeed, bare ground cover could serve as a proxy for disturbance, with
576 plots with virtually no bare ground usually being in late-successional stages (Lembrechts et al., 2014).
577 In the Scandinavian tundra and other arctic, subarctic or alpine areas, late-successional stages –
578 especially at low elevations - are often dominated by black crowberry, lingonberry or related species
579 (Tybirk et al., 2000). As a result, these undisturbed plots are dominated by a few late-successional
580 competitors, which can explain their higher dark diversity. More frequent disturbances, on the other
581 hand, can result in more bare ground, which creates opportunities (empty niches) for species to
582 establish and thus result in a lower dark diversity in these plots (Prieur-Richard & Lavorel, 2000).
583 Lembrechts et al. (2014), for example, found in the same region a higher species diversity along
584 disturbed roadsides than in natural communities owing largely to competitive release, a conclusion
585 that found some support here in our observed lower dark diversity close to a road and the railroad, at
586 least with some methods. Initially, we expected that disturbance by roads or railroads would result in
587 a higher plot-level dark diversity because many stress-tolerant mountain species would not be able to
588 cope with these disturbances (Auerbach et al., 1997). However, this would only be the case if the
589 disturbance would be sufficiently intense to reduce species richness, which rarely – if ever – happened
590 in our study plots, given our sampling focus on undisturbed vegetation only. Therefore, the same
591 explanation as for bare ground stands, i.e., plots that were more disturbed created opportunities for
592 species to establish, resulting in a lower plot-level dark diversity (Lembrechts et al., 2014).

593

4.2 Species-level dark diversity

Mycorrhizal association was the only variable with significant influence on the species-level dark diversity across all methods. However, while species with a symbiotic ericoid mycorrhizal association had a significantly higher dark diversity than all other associations when based on the climatic niche models, the opposite was true when using the other three methods. These contrasting results highlight the differences between these dark diversity estimation methods. In this system, the species with an ErM association (e.g. black crowberry and lingonberry) were virtually not climate-limited (occurring in 64 and 63 out of the 107 plots, respectively) and could in theory, based on their climatic niche, be present in all plots. Therefore, their dark diversity probability ended up being very high in any plot where they were absent, simply because of the underlying modelling approach. This issue was corrected by using species-specific thresholds, in which case mycorrhizal type was not withheld as significant.

These ErM-associated species not only dominated the landscape, but they were also often found in strong association with each other, resulting in clear predictions of their presence once one of them was present, when using the co-occurrence-based method. As their spatial connection in the field was so consistent, their estimated dark diversity using these methods ended up relatively low. Additionally, as ErM-fungi are the most dominant and widespread fungi in tundra regions (Tendersoo, 2017), in theory, there ought to be enough coverage of ErM-fungi so that the establishment of species associated with them should not be hampered. Consequently, there should be less reason for the species to be absent in areas where they could potentially occur than for AM-associated species (Tendersoo, 2017). All of this suggests that the observed higher dark diversity estimates for ErM-associated species by the climatic niche-approach are most likely a methodological artefact. These methodological issues could also explain why such little consistency was observed for the other studied drivers of species-level dark diversity, calling for caution when interpreting findings from any such dark diversity estimate separately.

4.3 Comparison of methods and uncertainties

In this paper, we estimated dark diversity using both niche-based and co-occurrence-based methods with both approaches often used interchangeably in literature. However, our results show that both approaches have significantly different assumptions and, as a result, get relatively incomparable results. Indeed, the niche-based approaches estimate the dark diversity as the set of species that could occur at a certain location based on their climatic niche or other environmental filters. The latter drivers are then often used as explanatory variables for the observed dark diversity, as done in the underlying study. For example, reduced competitive interactions in sites with larger percentages of bare ground would result in lower dark diversity, as is hinted at by our results.

Co-occurrence-based methods, on the other hand, estimate dark diversity simply by the neighbouring species with which a target species is usually associated. These approaches incorporate biotic interactions inherently in the dark diversity estimate. However, they do exclude species from the dark diversity for which the climatic conditions fall within their climatic limits, yet whose co-occurring species are also missing at a site. The latter could be especially problematic in diverse communities with high beta diversity, or areas with truncated, reduced, or novel communities as a result of anthropogenic land use or climatic changes (Christensen et al., 2021).

Perhaps more worryingly, within each type of dark diversity estimation method, results were not necessarily in agreement with each other. We found largely different findings, especially for species-

637 level dark diversity, when using climatic niches with or without species-specific thresholds, as well as
638 when using the hypergeometric method versus the Beals' index. As such, our results highlight the need
639 for caution when calculating and interpreting dark diversity estimates, as the conclusions depend
640 heavily on the methodological decisions made, and methods should thus be tailored to the specific
641 research questions.

642 Of course, several alternative methods could still be used to estimate dark diversity, and many
643 adjustments to the methods used above could be proposed. For example, one could use global
644 datasets such as GBIF to model the climatic niche, rather than data from the study region only. Using
645 global datasets for such broader-scale niche models could result in a more accurate estimate of the
646 climatic niche since the entire climatic niche could be modelled, rather than a truncated version as
647 results from regional data. However, most of these global datasets lack absence data and presences
648 are obtained using a wide variety of methodologies and spatial resolutions (Tessarolo et al., 2014),
649 while abiotic data is at the global scale often only available at coarser resolution (Lembrechts et al.,
650 2019). This could also make the predictions less accurate. Additionally, there is the possibility of
651 mismatches, especially for rare species, since global datasets can be strongly spatially biased (Meyer
652 et al., 2016). Therefore, predicting local climatic niches based on global data can make it more difficult
653 to figure out whether the absences are due to a bias in the global dataset or the drivers under
654 investigation.

655 The most promising avenue could perhaps come from an approach that combines both climatic niches
656 with co-occurrences, such as joint Species Distribution Models (jSDMs; Pollock et al., 2014). This recent
657 class of distribution models draws information from species co-occurrences and explains spatial
658 variation in species distributions by extending standard species distribution models with species–
659 species associations. Such an approach could potentially allow distinguishing through one model
660 between absences driven by environmental unsuitability, biotic interactions, or other drivers.
661 Nevertheless, Carmona & Pärtel (2020) did find that jSDMs could not outperform the hypergeometric
662 method, yet they do substantially increase computational time.

663 **4.5 Conclusions**

664 The concept of dark diversity is still in its infancy, yet its contribution to understanding community
665 completeness and nature conservation has already been shown to be significant (Lewis et al., 2017;
666 Riibak et al., 2015). In this context, it is crucial to determine whether a species' absence is a result of
667 species-specific traits or plot characteristics, be it abiotic factors or biotic interactions, which is
668 something traditional biodiversity studies that only focus on species presences cannot provide. We
669 here compared different methodological approaches to estimate dark diversity and showed significant
670 divergence in predicted drivers of dark diversity based on the method used, calling for caution when
671 interpreting statistical findings on dark diversity estimates. Nevertheless, we can generally conclude
672 that areas at low elevations, and, to a certain extent, with a low species richness or low levels of
673 disturbance showed a higher plot-level dark diversity, largely due to natural processes such as
674 competitive dominance.

675 **Acknowledgments**

676 We thank the master students Renée Lejeune and Jasmine Spreeuwers for their assistance in gathering
677 data during the summer of 2021. Additionally, we extend our appreciation to Stef Haesen for supplying
678 us with the Topographical Wetness Index raster.

Data availability statement

Data will be made available on Zenodo upon acceptance of the paper.

References

Alexander, J. M., Lembrechts, J. J., Cavieres, L. A., Daehler, C., Haider, S., Kueffer, C. et al., (2016) Plant invasions into mountains and alpine ecosystems: current status and future challenges. *Alpine Botany*, 126, 89–103. https://doi.org/10.1007/s00035-016-0172-8

Auerbach, N. A., Walker, M. D., & Walker, D. A. (1997) Effects of roadside disturbance on substrate and vegetation properties in arctic tundra. *Ecological Applications*, 7, 218–235. https://doi.org/10.1890/1051-0761(1997)007[0218:EORDOS]2.0.CO;2

Beals, E. W. (1984) Bray-curtis ordination: An effective strategy for analysis of multivariate ecological data. *Advances in Ecological Research*, 14, 1–55. https://doi.org/10.1111/oik.07308

Belinchón, R., Hemrová, L., & Münzbergová, Z. (2020) Functional traits determine why species belong to the dark diversity in a dry grassland fragmented landscape. *Oikos*, 129, 1468–1480. https://doi.org/10.1111/oik.07308

Bjorkman, A. D., Myers-Smith, I. H., Elmendorf, S. C., Normand, S., Thomas, H. J. D., Alatalo, J. M. et al (2018) Tundra Trait Team: A database of plant traits spanning the tundra biome. *Global Ecology and Biogeography*, 27, 1402–1411. https://doi.org/10.1111/geb.12821

Broennimann, O., Di Cola, V., Guisan, A. (2022). *Ecospat: spatial ecology miscellaneous methods. Version 3.4.* Available at https://CRAN.R-project.org/package=ecospat [Accessed 16 March 2022]

Brooks, M. E., Kristensen, K., van Benthem, K. J., Magnusson, A., Berg, C. W., Nielsen, A. et al. (2017) "glmmTMB Balances Speed and Flexibility Among Packages for Zero-inflated Generalized Linear Mixed Modeling." *The R Journal*, 9, 378–400.

Callaghan, T. V., Jonasson, C., Thierfelder, T., Yang, Z., Hedenås, H., Johansson, M. et al. (2013) Ecosystem change and stability over multiple decades in the Swedish subarctic: Complex processes and multiple drivers. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 368, 1–17. https://doi.org/10.1098/rstb.2012.0488

Callaway, R. M., Brooker, R. W., Choler, P., Kikvidze, Z., Lortie, C. J., Michalet, R. et al. (2002) Positive interactions among alpine plants increase with stress. *Nature*, 417, 844–848. https://doi.org/10.1038/nature00812

Carmona, C. P., & Pärtel, M. (2020) Estimating probabilistic site-specific species pools and dark diversity from co-occurrence data. *Global Ecology and Biogeography*, 30, 316–326. https://doi.org/10.1111/geb.13203

Chamberlain, S., Barve, V., Mcglinn, D., Oldoni, D., Desmet, P., Geffert, L., & Ram, K. (2021) *rgbif: Interface to the Global Biodiversity Information Facility API*. *Version 3.6.0*. Available at https://CRAN.R-project.org/package=rgbif [Accessed 15 May 2021]

Christensen, E., Christensen, B., & Christensen, S. (2021) Problems in using Beals' index to detect species trends in incomplete floristic monitoring data (Reply to Bruelheide et al. (2020)). *Diversity and Distributions*, 27, 1324–1327. https://doi.org/10.1111/ddi.13276

Clavel, J. (2022) Roadside disturbance promotes arbuscular mycorrhizal communities in mountain regions worldwide [unpublished].

Clavel, J., Lembrechts, J., Alexander, J., Haider, S., Lenoir, J., Milbau, A. et al. (2021) The role of arbuscular mycorrhizal fungi in nonnative plant invasion along mountain roads. *New Phytologist*, 230, 1156–1168. https://doi.org/10.1111/nph.16954

Cribari-Neto, F., Zeileis, A. (2010) Beta Regression in R. *Journal of Statistical Software,* 34, 1–24.

724 Crooks, J. A. (2005) Lag times and exotic species: The ecology and management of biological invasions
725 in slow-motion. *Ecoscience*, 12, 316–329. https://doi.org/10.2980/i1195-6860-12-3-316.1
726 de Bello, F., Fibich, P., Zelený, D., Kopecký, M., Mudrák, O., Chytrý, M. et al. (2016) Measuring size and
727 composition of species pools: a comparison of dark diversity estimates. *Ecology and Evolution*,
728 6, 4088–4101. https://doi.org/10.1002/ece3.2169
729 Ellenberg, H., Weber, H.E., Düll, R., Wirth, V., Werner, W. & Paulißen, D. (1991) *Zeigerwerte von*
730 *Pflanzen in Mitteleuropa.* Scripta Geobotanica.
731 European Union, Copernicus Land Monitoring Service 2021, European Environment Agency (EEA).
732 Ewald, J. (2002) A probabilistic approach to estimating species pools from large compositional
733 matrices. *Journal of Vegetation Science*, 13, 191–198. https://doi.org/10.1111/j.1654-
734 1103.2002.tb02039.x
735 Finlay, R. D. (2008) Ecological aspects of mycorrhizal symbiosis: With special emphasis on the
736 functional diversity of interactions involving the extraradical mycelium. *Journal of*
737 *Experimental Botany*, 59, 1115–1126. https://doi.org/10.1093/jxb/ern059
738 Fox, J., & Weisberg, S. (2011). An R companion to applied regression (2nd edition). SAGE Publications
739 Inc.
740 Gijbels, P., Adriaens, D. & Honnay, O. (2012) An orchid colonization credit in restored calcareous
741 grasslands. *Ecoscience,* 19, 21–28. https://doi.org/10.2980/19-1-3460
742 Gough, L., Shaver, G. R., Carroll, J., Royer, D. L., & Laundre, J. A. (2000) Vascular plant species richness
743 in Alaskan arctic tundra: The importance of soil pH. *Journal of Ecology*, 88, 54–66.
744 https://doi.org/10.1046/j.1365-2745.2000.00426.x
745 Grieve, I. C. (2000) Effects of human disturbance and cryoturbation on soil iron and organic matter
746 distributions and on carbon storage at high elevations in the Cairngorm Mountains, Scotland.
747 *Geoderma*, 95, 1–14. https://doi.org/10.1016/S0016-7061(99)00060-9
748 Guisan, A., & Thuiller, W. (2005) Predicting species distribution: Offering more than simple habitat
749 models. *Ecology Letters*, 8, 993–1009. https://doi.org/10.1111/j.1461-0248.2005.00792.x
750 Haesen, S., Lembrechts, J. J., De Frenne, P., Lenoir, J., Aalto, J., Ashcroft, M. et al. (2021) ForestTemp
751 Sub-canopy microclimate temperatures of European forests. *Global Change Biology*, 27, 6307–
752 6319. https://doi.org/10.1111/gcb.15892
753 Hartig, F. (2018) *Yes, statistical errors are slowing down scientific progress! Theoretical ecology.*
754 Available at https://theoreticalecology.wordpress.com/2018/05/03/yes-statistical-errors-are-
755 slowing-down-scientific-progress/ [Accessed 2 November 2022]
756 Hartig, F. (2022) *DHARMa: Residual diagnostics for hierarchical (multi-level / mixed) regression*
757 *Models. R package version 0.4.5.* Available at https://CRAN.R-project.org/package=DHARMa
758 [Accessed 14 November 2021]
759 Hijmans, R. J., van Etten, J. (2012) *Raster: Geographic data analysis and modeling. R package version*
760 *3.6-3.* Available at http:// CRAN.R-project.org/package=raster [Accessed 5 October 2021]
761 Hooper, D. U., Adair, E. C., Cardinale, B. J., Byrnes, J. E. K., Hungate, B. A., Matulich, K. L. et al. (2012)
762 A global synthesis reveals biodiversity loss as a major driver of ecosystem change. *Nature*, 486,
763 105–108. https://doi.org/10.1038/nature11118
764 Howe, F., & Smallwood, J. (1982) Ecology of seed dispersal. *Annual Review of Ecology and Systematics,*
765 *13*, 201–228. https://doi.org/10.1146/annurev.es.13.110182.001221
766 Hu, F. S., Higuera, P. E., Duffy, P., Chipman, M. L., Rocha, A. V., Young, A. M. et al. (2015) Arctic tundra
767 fires: Natural variability and responses to climate change. *Frontiers in Ecology and the*
768 *Environment*, 13, 369–377. https://doi.org/10.1890/150063
769 Jonasson, S. (1988) Evaluation of the point intercept method for the estimation of plant biomass.
770 *Oikos*, *52*, 101–106.
771 Jones, N. T., & Gilbert, B. (2016) Biotic forcing: the push–pull of plant ranges. *Plant Ecology*, 217, 1331–

772 1344. https://doi.org/10.1007/s11258-016-0603-z
773 Kaarlejärvi, E., & Olofsson, J. (2014) Concurrent biotic interactions influence plant performance at
774 their altitudinal distribution margins. *Oikos*, 123, 943–952. https://doi.org/10.1111/oik.01261
775 Karger, D. N., Conrad, O., Böhner, J., Kawohl, T., Kreft, H., Soria-Auza, R. W. et al. (2017) Climatologies
776 at high resolution for the Earth land surface areas. *Scientific Data,* 4, 170122.
777 https://doi.org/10.1038/sdata.2017.122
778 Klanderud, K. (2010) Species recruitment in alpine plant communities: The role of species interactions
779 and productivity. *Journal of Ecology*, 98, 1128–1133. https://doi.org/10.1111/j.1365-
780 2745.2010.01703.x
781 Kleyer, M., Bekker, R. M., Knevel, I. C., Bakker, J. P., Thompson, K., Sonnenschein, M. et al. (2008) The
782 LEDA traitbase: a database of life- history traits of the northwest European flora. *Journal of*
783 *Ecology,* 96, 1266–1274. https://doi.org/10.1111/j.1365-2745.2008.01430.x
784 Kopecký, M., Macek, M., & Wild, J. (2021) Science of the Total Environment Topographic Wetness
785 Index calculation guidelines based on measured soil moisture and plant species composition.
786 *Science of the Total Environment*, 757, 143785.
787 https://doi.org/10.1016/j.scitotenv.2020.143785
788 Körner, C. (2021) *Alpine plant life: Functional plant ecology of high mountain ecosystems*, 3rd edition.
789 Springer Nature Switzerland AG 2021. https://doi.org/10.1007/978-3-030-59538-8
790 Kullman, L. (2015) Recent and past trees and climates at the Arctic/Alpine margin in Swedish Lapland:
791 An Abisko case study Review. *Journal of Biodiversity Management & Forestry*, 4, 1–12.
792 doi:10.4172/2327-4417.1000150
793 Legendre P, Legendre J (1998) *Numerical Ecology*, 3rd edition. Elsevier, Amsterdam.
794 Lembrechts, J. J., Lenoir, J., Nuñez, M. A., Pauchard, A., Geron, C., Bussé, G. et al. (2018) Microclimate
795 variability in alpine ecosystems as stepping stones for non-native plant establishment above
796 their current elevational limit. *Ecography*, 41, 900–909. https://doi.org/10.1111/ecog.03263
797 Lembrechts, J. J., Lenoir, J., Roth, N., Hattab, T., Milbau, A., Haider, S. et al (2019) Comparing
798 temperature data sources for use in species distribution models: From in-situ logging to
799 remote sensing. *Global Ecology and Biogeography*, 28, 1578–1596.
800 https://doi.org/10.1111/geb.12974
801 Lembrechts, J. J., Milbau, A., & Nijs, I. (2014) Alien roadside species more easily invade alpine than
802 lowland plant communities in a subarctic mountain ecosystem. *PLoS ONE*, 9, 1–10.
803 https://doi.org/10.1371/journal.pone.0089664
804 Lembrechts, J. J., Aalto, J., Ashcroft, M. B., De Frenne, P., Kopecký, M., Lenoir, J. et al. (2020) SoilTemp:
805 A global database of near-surface temperature. *Global Change Biology,* 26, 6616-6629.
806 https://doi.org/10.1111/gcb.15123
807 Lembrechts, J. J., van den Hoogen, J., Aalto, J., Ashcroft, M., De Frenne, P., Kemppinen, J. et al (2021)
808 Global maps of soil temperature. *Global Change Biology,* 28, 3110-3144.
809 https://doi.org/10.1111/gcb.16060
810 Lenoir, J., Gégout, J. C., Guisan, A., Vittoz, P., Wohlgemuth, T., Zimmermann, N. E. et al. (2010) Cross-
811 scale analysis of the region effect on vascular plant species diversity in southern and northern
812 European mountain ranges. *PLoS One,* 5, e15734.
813 https://doi.org/10.1371/journal.pone.0015734
814 Lenth, R. (2022) *Emmeans: Estimated marginal means, aka least-squares means. Version 1.8.2.*
815 Available at: https://CRAN.R-project.org/package=emmeans [Accessed 14 November 2021]
816 Lewis, R. J., de Bello, F., Bennett, J. A., Fibich, P., Finerty, G. E., Götzenberger, L. et al. (2017) Applying
817 the dark diversity concept to nature conservation. *Conservation Biology*, 31, 40–47.
818 https://doi.org/10.1111/cobi.12723
819 Lewis, R. J., Szava-Kovats, R., & Pärtel, M. (2016) Estimating dark diversity and species pools: An

empirical assessment of two methods. *Methods in Ecology and Evolution*, 7, 104–113. https://doi.org/10.1111/2041-210X.12443

Lindén, E., Gough, L., & Olofsson, J. (2021) Large and small herbivores have strong effects on tundra vegetation in Scandinavia and Alaska. *Ecology and Evolution*, 11, 12141–12152. https://doi.org/10.1002/ece3.7977

Lüdecke, D., Ben-Shachar, M., Patil, I., Waggoner, P. & Makowski, D. (2021) "performance: An R package for assessment, comparison and testing of statistical models." *Journal of Open Source Software*, 6, 3139. https://doi.org/10.21105/joss.03139

MacDougall, A. S., Caplat, P., Olofsson, J., Siewert, M. B., Bonner, C., Esch et al. (2021) Comparison of the distribution and phenology of Arctic Mountain plants between the early 20th and 21st centuries. *Global Change Biology,* 27, 5070-5083. https://doi.org/10.1111/gcb.15767

Meyer, C., Weigelt, P., & Kreft, H. (2016) Multidimensional biases, gaps and uncertainties in global plant occurrence information. *Ecology Letters*, 19, 992–1006. https://doi.org/10.1111/ele.12624

Mitchell, T. (2014) Developers G. Geospatial Power Tools: GDAL Raster & Vector Commands. *Locate Press.*

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. https://doi.org/10.1111/1365-2664.12867

Mohd, M. H., Murray, R., Plank, M. J., & Godsoe, W. (2016). Effects of dispersal and stochasticity on the presence–absence of multiple species. *Ecological Modelling,* 342, 49-59. https://doi.org/10.1016/j.ecolmodel.2016.09.026

Mooney, H., Larigauderie, A., Cesario, M., Elmquist, T., Hoegh-Guldberg, O., Lavorel, S. et al. (2009) Biodiversity, climate change, and ecosystem services. *Current Opinion in Environmental Sustainability*, 1, 46–54. https://doi.org/10.1016/j.cosust.2009.07.006

Mossberg, B. & Stenberg, L. (2008) *Fjällflora: Sverige, Finland, Norge, Svalbard*. Wahlström & Widstrand.

Newbold, T., Hudson, L. N., Hill, S. L. L., Contu, S., Lysenko, I., Senior, R. A. et al. (2015) Global effects of land use on local terrestrial biodiversity. *Nature*, 520, 45–50. https://doi.org/10.1038/nature14324

Oksanen, J., Blanchet, F.G., Kindt, R. et al. (2022) *Vegan: Community ecology package. Version 2.6-4.* Available at: https://CRAN.R-project.org/package=vegan [Accessed 2 November 2022]

Olofsson, J. (2001) Influence of herbivory and abiotic factors on the distribution of tall forbs along a productivity gradient: A transplantation experiment. *Oikos*, 94, 351–357. https://doi.org/10.1034/j.1600-0706.2001.940216.x

Ozinga, W. A., Hennekens, S. M., Schaminée, J. H. J., Bekker, R. M., Prinzing, A., Bonn, S. et al. (2005) Assessing the relative importance of dispersal in plant communities using an ecoinformatics approach. *Folia Geobotanica*, 40, 53–67. https://doi.org/10.1007/BF02803044

Parolo, G., Rossi, G., & Ferrarini, A. (2008). Toward improved species niche modelling: *Arnica montana* in the Alps as a case study. *Journal of Applied Ecology,* 45, 1410–1418. https://doi.org/10.1111/j.1365-2664.2008.01516.x

Pärtel, M. (2014) Community ecology of absent species: Hidden and dark diversity. *Journal of Vegetation Science*, 25, 1154–1159. https://doi.org/10.1111/jvs.12169

Pärtel, M., Carmona, C. P., Zobel, M., Moora, M., Riibak, K., & Tamme, R. (2019) DarkDivNet – A global research collaboration to explore the dark diversity of plant communities. *Journal of Vegetation Science*, 30, 1039–1043. https://doi.org/10.1111/jvs.12798

Pärtel, M., Szava-Kovats, R., & Zobel, M. (2011) Dark diversity: Shedding light on absent species. *Trends*

in Ecology and Evolution, 26, 124–128. https://doi.org/10.1016/j.tree.2010.12.004

Pebesma, E. J., Bivand, R. S. (2005) Classes and methods for spatial data in R. *R news*, 5, 9–13.

Pellissier, L., Bråthen, K. A., Pottier, J., Randin, C. F., Vittoz, P., Dubuis, A. et al. (2010) Species distribution models reveal apparent competitive and facilitative effects of a dominant species on the distribution of tundra plants. *Ecography*, 33, 1004–1014. https://doi.org/10.1111/j.1600-0587.2010.06386.x

Pollock, L. J., Tingley, R., Morris, W. K., Golding, N., O'Hara, R. B., Parris, K. M. et al. (2014) Understanding co-occurrence by modelling species simultaneously with a Joint Species Distribution Model (JSDM). *Methods in Ecology and Evolution,* 5, 397-406. https://doi.org/10.1111/2041-210X.12180

Prieur-Richard, A. H., & Lavorel, S. (2000) Invasions: The perspective of diverse plant communities. (*Austral Ecology*, 25, 1–7. https://doi.org/10.1046/j.1442-9993.2000.01033.x

QGIS.org, 2021. QGIS Geographic Information System. QGIS Association. Available at: http://www.qgis.org

Quinn, Q. P., Keough, M. J. (2002) *Experimental design and data analysis for biologists*. Cambridge University Press, Cambridge. https://doi.org/10.1017/CBO9780511806384

R Core Team (2021) R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. Available at: https://www.R-project.org/

Rashid, I., Haq, S. M., Lembrechts, J. J., Khuroo, A. A., Pauchard, A., & Dukes, J. S. (2021) Railways redistribute plant species in mountain landscapes. *Journal of Applied Ecology*, 58, 1967–1980. https://doi.org/10.1111/1365-2664.13961

Riibak, K., Reitalu, T., Tamme, R., Helm, A., Gerhold, P., Znamenskiy, S. et al. (2015) Dark diversity in dry calcareous grasslands is determined by dispersal ability and stress-tolerance. *Ecography*, 38, 713–721. https://doi.org/10.1111/ecog.01312

Sonesson, M., & Lundberg, B. (1974) Late Quaternary forest development of the Tornetrask area, North Sweden. *Oikos*, 25, 121–133. https://doi.org/10.2307/3543947

Soudzilovskaia, N. A., Vaessen, S., Barcelo, M., He, J., Rahimlou, S., Abarenkov, K. et al. (2020) FungalRoot: global online database of plant mycorrhizal associations. *New Phytologist*, 227, 955–966. https://doi.org/10.1111/nph.16569

Stephenson, I. (2016) *What is Dark Diversity?* Methods blog. Available at https://methodsblog.com/2016/05/22/dark-diversity/ [Accessed 4 April 2022]

Sugihara, N. G., Van Wagtendonk, J. W., & Fites-Kaufman, J. (2006) Fire as an ecological process. In: Van Wagtendonk, J. W., Sugihara, N. G., Stephens, S. L., Thode, A. E., Shaffer, K. E., & Fites-Kaufman, J. A. (Eds), *Fire in California's ecosystems*, 1st edition. University of California Press, pp. 58-74.

Tendersoo, L. (Ed) (2017) *Biogeography of Mycorrhizal Symbiosis*, 1st edition. Springer International Publishing 2017. https://doi.org/10.1007/978-3-319-56363-3

Tessarolo, G., Rangel, T. F., Araújo, M. B., & Hortal, J. (2014) Uncertainty associated with survey design in Species Distribution Models. *Diversity and Distributions*, 20, 1258–1269. https://doi.org/10.1111/ddi.12236

Tilman, D., Isbell, F., & Cowles, J. M. (2014) Biodiversity and ecosystem functioning. *Annual Review of Ecology, Evolution, and Systematics*, 45, 471–493. https://doi.org/10.1146/annurev-ecolsys-120213-091917

Tybirk, K., Nilsson, M. C., Michelsen, A., Kristensen, H. L., Sheytsova, A., Strandberg, M. T. et al. (2000) Nordic *Empetrum* dominated ecosystems: Function and susceptibility to environmental changes. *Ambio*, 29, 90–97. https://doi.org/10.1579/0044-7447-29.2.90

Vonlanthen, C. M., Kammer, P. M., Eugster, W., Bühler, A., & Veit, H. (2006) Alpine vascular plant

915 species richness: The importance of daily maximum temperature and pH. *Plant Ecology*, 184,
916 13–25. https://doi.org/10.1007/s11258-005-9048-5
917 Wedegärtner, R. E., Lembrechts, J. J., van der Wal, R., Barros, A., Chauvin, A., Janssens et al. (2022)
918 Hiking trails shift plant species' realized climatic niches and locally increase species richness.
919 *Diversity and Distributions*, 28, 1416-1429 https://doi.org/10.1111/ddi.13552
920 Westoby, M. (1998) A leaf-height-seed (LHS) plant ecology strategy scheme. *Plant and Soil*, 199, 213–
921 227. https://doi.org/10.1023/A:1004327224729