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

Lore Hostens, Koenraad van Meerbeek, Dymphna Wiegmans, Keith Larson, Jonathan Roger Michel Henri Lenoir, Jan Clavel, Ronja Wedegärtner, Amber Pirée, Ivan Nijs, Jonas Lembrechts

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

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

Short title: Dark diversity can be metric-dependent

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 mountains. 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 mountains around Abisko, northern Sweden.
Methods: We calculated the dark diversity in 107 plots spread out across four mountain trails using four different methods: two co-occurrence-based (Beals' index and hypergeometric method) and two niche-based (climatic niche model and climatic niche model followed by species-specific threshold) methods. We then applied multiple generalized linear mixed-effects models and general linear models to determine which habitat characteristics and species traits contributed the most to 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 17% higher in areas at low elevations and 31% higher in areas with a low species richness.
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 the 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 concept introduced by Pärtel et al. (2011). To be part of the dark diversity, the absent species must have a reasonable probability of dispersing to and establishing viable populations in the area (i.e., by belonging to the regional species pool) and its ecological requirements (depending on the methodology used that may incorporate 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 (Keddy, 1992; 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 can be expected, as explained below. One of the main benefits of the dark diversity concept is that it enables us to compare biodiversity across various habitats or ecosystems despite significant differences in local diversity by deriving a relativized biodiversity index from the dark diversity, known as community completeness (Pärtel et al., 2011; Pärtel et al., 2013).

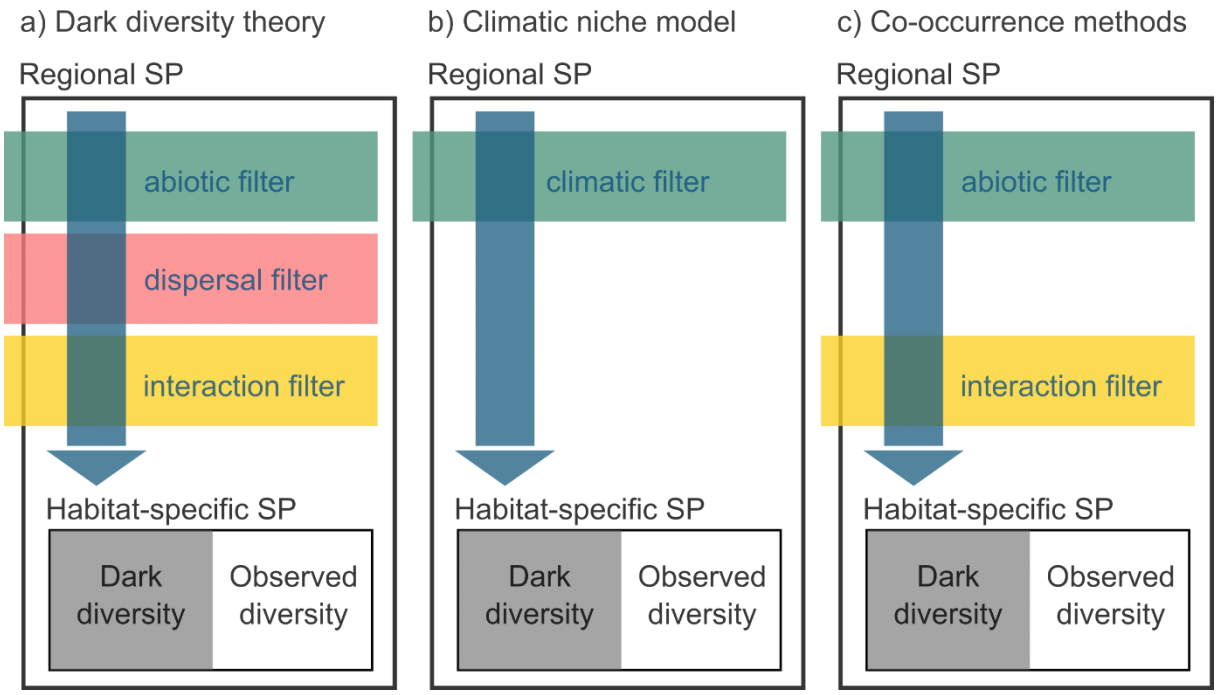

Figure 1: Schematic overview of three approaches used to estimate the habitat-specific species pool (SP). a) the theoretical concept of dark diversity, where the dark diversity is the non-observed set of species in a certain location, after filtering the regional species pool based on abiotic, dispersal 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 commonly used co-occurrence-based methods, which integrate both abiotic and interaction filters. Figure adjusted from Stephenson (2016). The combination of dark- and observed diversity encompasses the habitat-specific species pool. Note that for both the methods in b) and c), several other methodological decisions can still be made that might affect the outcome.

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 the focal 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 (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 a proxy for species' ecological requirements and are often used to characterize environmental conditions. 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). Additionally, predicting a species' habitat suitability based only on occurrence observations and

environmental data may prove to be difficult since processes like competition can play a crucial role, especially at the local scale (Cadotte & Tucker, 2017).

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 relies on 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 major difference between the Beals' probability index and hypergeometric method is that the hypergeometric method compares the actual number of co-occurrences between two species to the association of random pairs of species (i.e. under the assumption that there is no association). The difference between the observed and random association provides the index value, whereas for the Beals' index, the index value is only based on the observed patterns of co-occurrence (Carmona & Pärtel, 2020; Trindade et al., 2023). The advantage of these two co-occurrence-based approaches is that one only requires species composition data in the community without the need for 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 methods share a common purpose: they help recognize species that belong to the habitat-specific species pool. The species not recorded in the observed diversity, but belonging to the habitat-specific species pool of the focal site are part of the dark diversity (Figure 1; Pärtel et al., 2011). Considering the absence of a standard method for calculating the habitat-specific species pool and, by extension, the dark diversity, we used both niche- and co-occurrence-based approaches. Our aim was to estimate the dark diversity around Abisko, Sweden. We wanted to explore whether these different methods would yield varying estimates of dark diversity due to their inherent filters (Figure 1). We then further explored the drivers behind the spatial patterns of this dark diversity and assessed the impact of the different methods on these drivers. 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 subarctic environments (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 mountains 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).

The composition of dark diversity can be influenced by not only plot characteristics but also species traits. Certain traits might make some species more likely to be absent from plots, thereby contributing to the dark diversity (Moeslund et al., 2017). Therefore, we have selected six species traits related to resource-use efficiency and dispersal as these can play a key factor in plant recruitment and persistence. 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 soils in the alpine habitats of the study area are nutrient-poor (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 (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-idaea* 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).

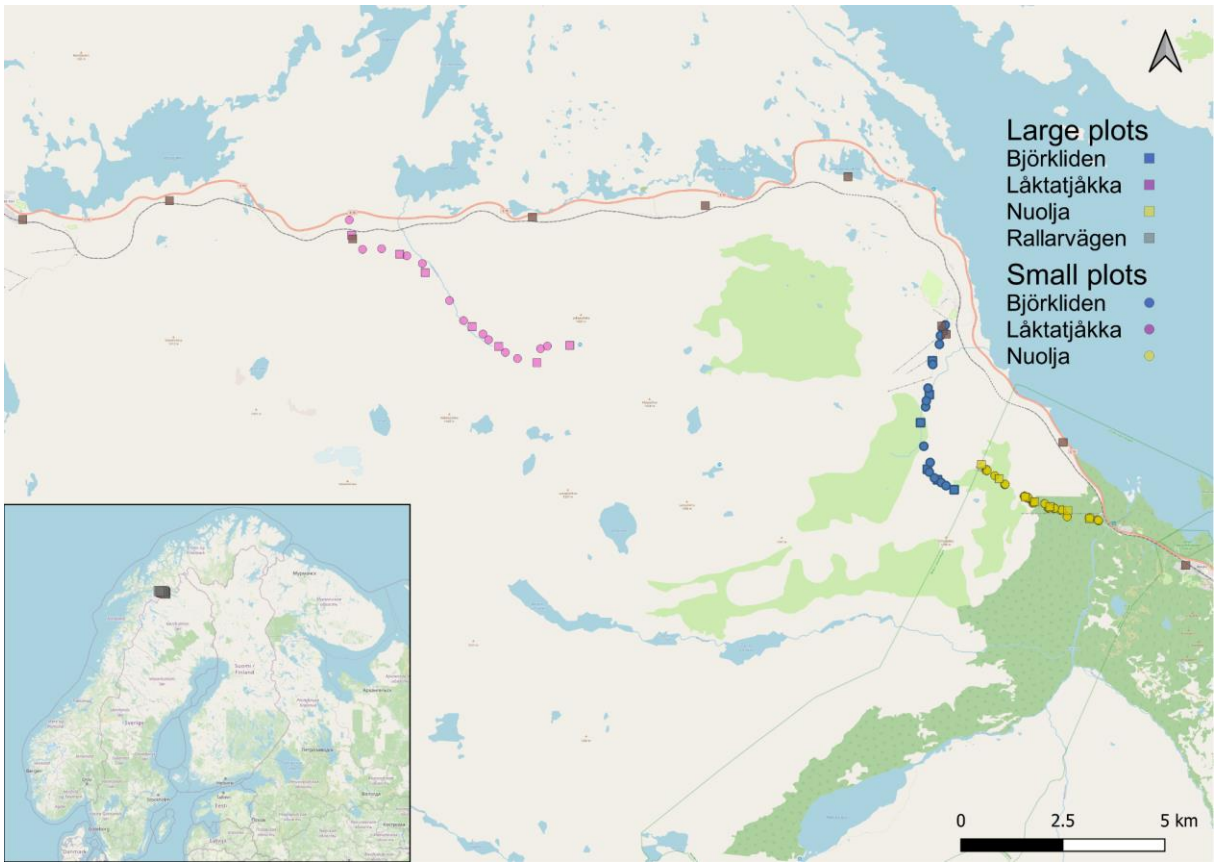

Figure 2: Map of the study area around Abisko, Sweden (grey dot on the inset), with 107 surveyed plots along the four hiking trails (colors) and the different survey methods (symbols).

Data from new and ongoing vegetation surveys were combined, with two different methodologies: 73 1 m × 1 m plots from a long-term vegetation composition monitoring project in the area (hereafter called 'small plots'), as well as 34 large (10 m × 10 m) plots established in the framework of the global DarkDivNet network (Pärtel et al., 2019). Of the 107 plots, 40 were situated 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 m × 10 m) 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 GARMIN GPSMAP® 66i 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 m × 1 m 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 km at the equator) for annual mean air temperature, annual precipitation, mean maximum air temperature of the warmest month, and mean minimum air temperature of the coldest month. Gridded data were downloaded 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). The SoilTemp maps were derived from CHELSA monthly air temperature maps and the offset between gridded air temperature and in-situ soil temperature measurements stored in 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 arcseconds.

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 layer covering Europe (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 gridded data were handled in R version 4.2.1 (R Core Team, 2021) using the raster (Hijmans et al., 2012), sp (Pebesma et al., 2005), and rgdal (Keitt et al., 2010) packages to overlay the spatial coordinates of all 107 plots and extract climatic information at the plot-level.

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 (QGIS Development Team, 2021). 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. While a continuous variable for distance to the disturbance would have allowed for more nuance, adding a separate parameter for distance to the trail, to the road and to the railroad was not possible, as all plots were at a fixed distance of 10 m from a trail, and the distance to road and railroad were too strongly correlated.

2.4.3 Amount of bare ground

Disturbances can generate patches of bare ground that can open empty niches for new species to colonize and establish themselves (Lembrechts et al., 2014). The amount of bare ground (%), here used 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 specific leaf area (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).

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).

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). All species 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.

306 $$NI = \frac{\sqrt{year\ (2020) - year\ first\ occurrence\ species}}{\sqrt{year\ (2020) - year\ first\ survey\ (1850)}}$$

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

311 Note that the region was poor in non-native species, and those present were mostly introduced
312 already over a century ago (Wiegmans et al., 2022). This is reflected in the high values of our nativeness
313 index (mean = 0.98, 5% lowest = 0.88). Consequently, one should not expect strong effects of
314 nativeness on dark diversity patterns in the northern Scandes.

315 **2.4.6 Mycorrhizal associations**

316 The association of plant species with the main types of mycorrhizal fungi (AM = arbuscular mycorrhiza,
317 EcM = ectomycorrhiza, ErM = ericoid mycorrhiza and NM = no mycorrhiza) was retrieved from the
318 FungalRoot database (Soudzilovskaia et al., 2020).

319 More details on the online data collection can be found in Appendix S1.

320 **2.5 Data-analysis**

321 **2.5.1 Dark diversity modeling**

322 For further analysis, only species with 10 or more occurrences, were included (n=49), as sufficient
323 observations were needed to calibrate climatic niche models and build co-occurrence matrices. We
324 then used the same dataset in four different approaches to estimate dark diversity. All statistical
325 analyses were conducted in R version 4.2.1 (R Core Team, 2021).

326 Climatic niche modeling

327 The presence and absence of all species in every plot was used to make climatic niche models. For
328 every species, a generalized linear model (GLM; Bates et al., 2015) was calibrated, with a binomial
329 distribution containing all climatic variables and their quadratic terms as explanatory variables (i.e.,
330 annual mean air temperature, annual precipitation, maximum air temperature of the warmest month,
331 minimum air temperature of the coldest month, annual mean soil temperature, minimum soil
332 temperature of the coldest month, and maximum soil temperature of the warmest month) and
333 presence/absence (1/0) of a species per plot as the response variable. Multicollinearity was checked
334 using the Variance Inflation Factor (VIF) from the *car* package (Fox & Weisberg, 2011) and variables
335 that increased the VIF to 5 or more were removed. The final models contained: annual precipitation,
336 minimum soil temperature of the coldest month, maximum soil temperature of the warmest month,
337 and their quadratic terms. No further model selection was done as we were not interested in a model
338 identifying the drivers of the species' climatic niche, but rather wanted to approximate their climatic
339 niche as consistently as possible.

340 To predict the probability of a species' occurrence in a specific plot, the GLM was calibrated on all
341 remaining plots (Lembrechts et al., 2019) and the probability was estimated for that specific plot
342 excluded from the model calibration. This leave-one-out procedure was then repeated for all plots and
343 all species, each time predicting the probability of occurrence of a species in a plot based on a model
344 calibrated on its occurrence pattern in all other plots. We then calculated the relative dark diversity
345 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 et al. (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 et al., 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 et al., 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 and a conservative threshold of 0.9 was applied as done by Trindade et al. (2023). All values below 0.9 were

given a 0 since we did not expect the species to be present here. To calculate the relative plot-level dark diversity, per plot the mean was taken from the remaining values (i.e. all values larger than 0.9). 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-effects models (GLMMs) with a beta distribution and logit-link function using the *glmmTMB* package (Brooks et al., 2017). Predictions from each of the four dark diversity indices (the two approaches based on niche models, the Beals' index, and the hypergeometric approach) were used as the 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 interaction terms 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). Community completeness was also calculated for each plot and every method as ln(observed richness/dark diversity) (Pärtel et al., 2013). A linear mixed model was created using the *lmer* function from the *lme4* package (Bates et al., 2015) with plot as a random factor to compare whether the community completeness differed significantly depending on the method. The distribution of the residuals was checked using the *DHARMa* package (Hartig, 2022) and assumptions were not deemed violated. As one needs to assess community completeness using species numbers, the community completeness based on the climatic niche models had to be calculated using a species-specific threshold as well, thus resulting in the same values as in the original dark diversity assessment using climatic niche models with a threshold. We thus maintained only one of these in the comparison.

2.5.3 Drivers of species-level dark diversity probability

To investigate why certain species had a higher dark diversity probability, we created 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 interaction terms 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 87% of the variance in plot-level dark diversity. In two cases (climatic niche models and hypergeometric method), elevation was responsible for the largest share, while in the two other cases (species-specific threshold and Beals' index) species richness was the most dominant factor (Figure 3). On average across all models, elevation explained 13% of the variance, species richness 7%, and plot size, type of disturbance, amount of bare ground, pH and TWI explained an additional 2%, 1%, 2%, 4% and 2%, respectively. 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.

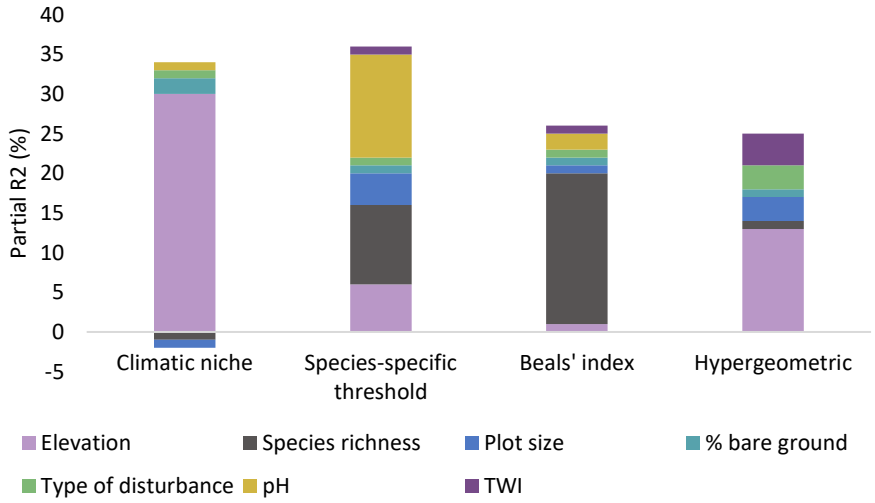

Figure 3: Variance partitioning (expressed in % and calculated using the marginal R^2) of the different explanatory variables in the GLMMs of the plot-level analyses on the predictions of each of the four different dark diversity methods. TWI = topographic wetness index.

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. Beals = Beals' index; CN = climatic niche models; Hyper = hypergeometric method; SS = species-specific threshold; Elev = elevation; SR = species richness; TOD = type of disturbance; TWI = topographic wetness index.*

Model	Intercept (Road)	Elev	SR	Plot size (10m x 10m)	TOD Hiking trail	TOD Railroad	% bare ground	pH	TWI	AIC
Beals	1.04*	10^{-4}	-0.082***	-0.121	-0.112	-0.345	0.001	0.032	0.001	-195
CN	-0.327	-0.001***	-0.029***	0.079	0.524	-0.251	-0.001*	-0.022	-0.014	-370
Hyper	0.101	-0.001**	0.01	-0.304	0.416	-0.177	-0.006	-0.001	-0.071	-137
SS	3.00***	-0.001**	-0.105***	-0.408*	0.281	-0.165	-0.001	-0.262***	0.038	-140

456

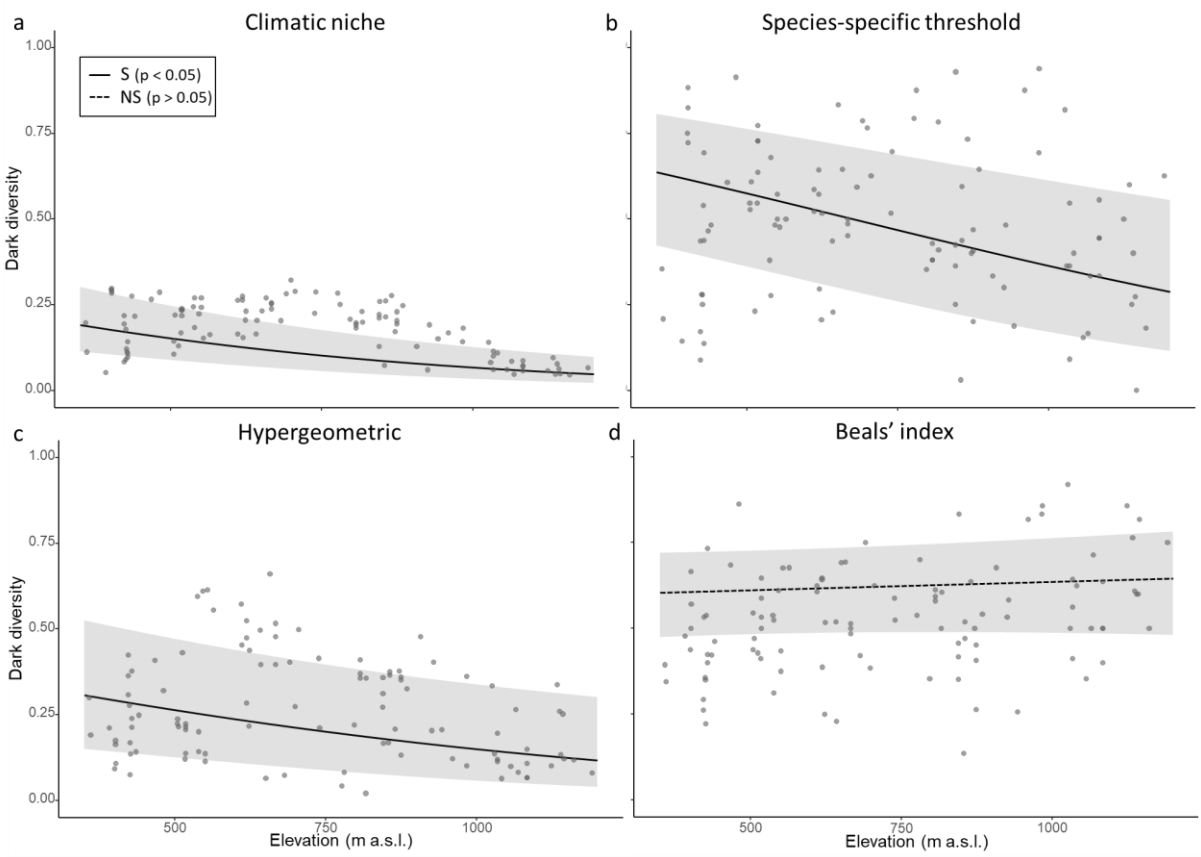

Figure 4: Marginal effects plots of the plot-level dark diversity as a function of elevation (m a.s.l.). The grey area indicates the 95% confidence interval and the grey dots are the raw data points. Dark diversity was estimated using a) the climatic niche models, b) the climatic niche models followed by the species-specific threshold, c) the hypergeometric method and d) the Beals' index. S = significant; NS = non-significant.

The plot-level dark diversity decreased significantly with increasing species richness in three cases (Table 1; Figure 5a, 5b, 5d), yet increased with increasing species richness when using the hypergeometric method, albeit not significantly (Table 1; Figure 5c).

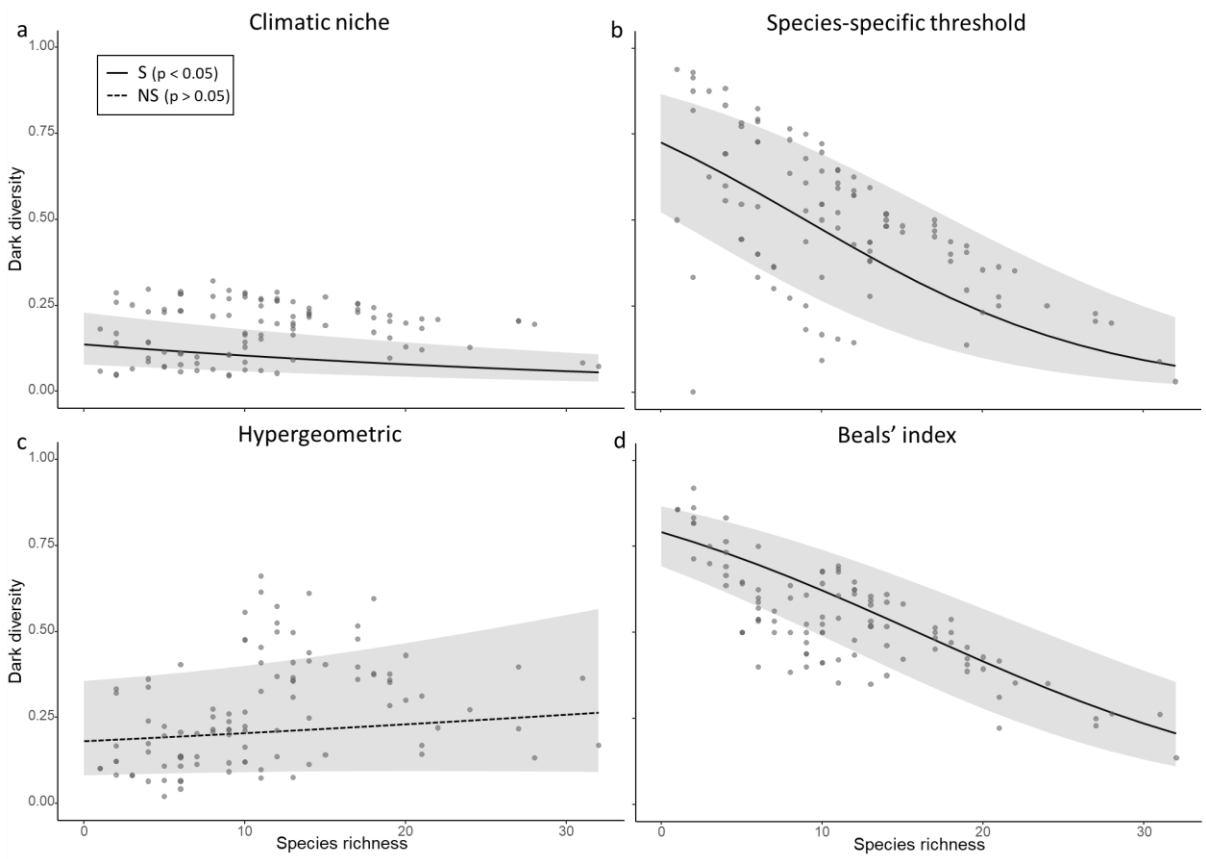

Figure 5: Marginal effects plots of the plot-level dark diversity as a function of species richness. The grey area indicates the 95% confidence interval, and the grey dots are the raw data points. a) the climatic niche models, b) the climatic niche models followed by the species-specific threshold, c) the hypergeometric method and d) the Beals' index. S = significant; NS = non-significant.

Furthermore, our results indicate that only the climatic niche model had a significant relationship between dark diversity and bare ground. Moreover, only the climatic niche model followed by the species-specific threshold had significant relationships with plot size and pH (Appendix S2, Figure S1a). In the remaining two models, none of the other variables were found to be significant predictors of dark diversity.

Lastly, the community completeness based on the Beals' index was significantly lower than the community completeness based on the other two methods (Appendix S2, Figure S2).

3.2 Species-level dark diversity

Depending on the method, we could explain between 8% and 45% of the variance in species-level dark diversity (Figure 6). In all cases, mycorrhizal association was responsible for the largest share (Figure 6). On average across all models, mycorrhizal association explained 16% of the variance, seed mass 9%, SLA 3% and the NI and the maximum vegetative plant height an additional 2% and 3%, respectively. 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.

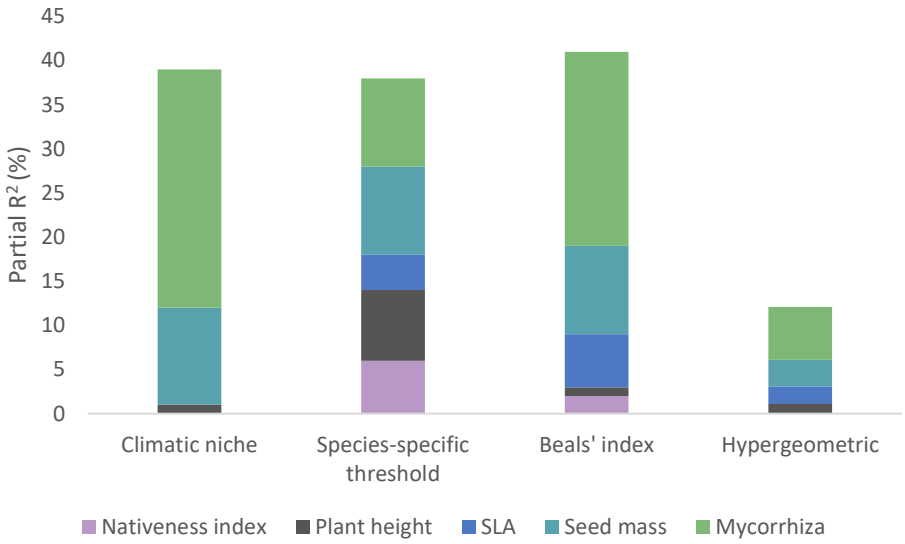

Figure 6: *Variance partitioning (expressed in % and calculated using the marginal R^2) of the different explanatory variables in the GLMs of the species-level analyses on the predictions of each of the four different dark diversity methods. SLA = specific leaf area.*

Mycorrhizal status was the only significant parameter in the climate niche model approach, with ericoid mycorrhizae differing significantly from AM, EcM and NM (Figure 7; Table 2; Appendix S3). 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 Beal's index and climatic niche model followed by the species-specific threshold (Table 2; Figure 7b, d). For the Beals' index the contrast test also revealed that ericoid mycorrhizae differed significantly from AM, EcM and NM (Figure 7d, Appendix S3). For the species-specific threshold, the contrast test only showed a borderline significant difference between ErM and NM (Figure 7b; Appendix S3). Lastly, and even more contrasting, species with a symbiotic ectomycorrhizal association had a significantly lower dark diversity than all other associations when using the hypergeometric method (Figure 7c). The contrast test also revealed that EcM differed significantly from AM and NM (Appendix S3).

*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. Beals = Beals' index; CN = climatic niche models; Hyper = hypergeometric method; SS = species-specific threshold; 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
Beals	1.04*	0.170	-1.17***	0.193	-10^{-4}	-0.001	-3.28	-0.089	-19

CN	-0.043	0.108	1.17***	-0.101	-10^{-4}	-10^{-4}	-1.37	-0.068	-72
Hyper	-3.66	-1.40*	-0.597	0.314	0.001	-0.001	3.42	0.121	-38
SS	-5.96	0.453	-0.916*	0.208	-0.001	-0.001	6.95	0.082	-11

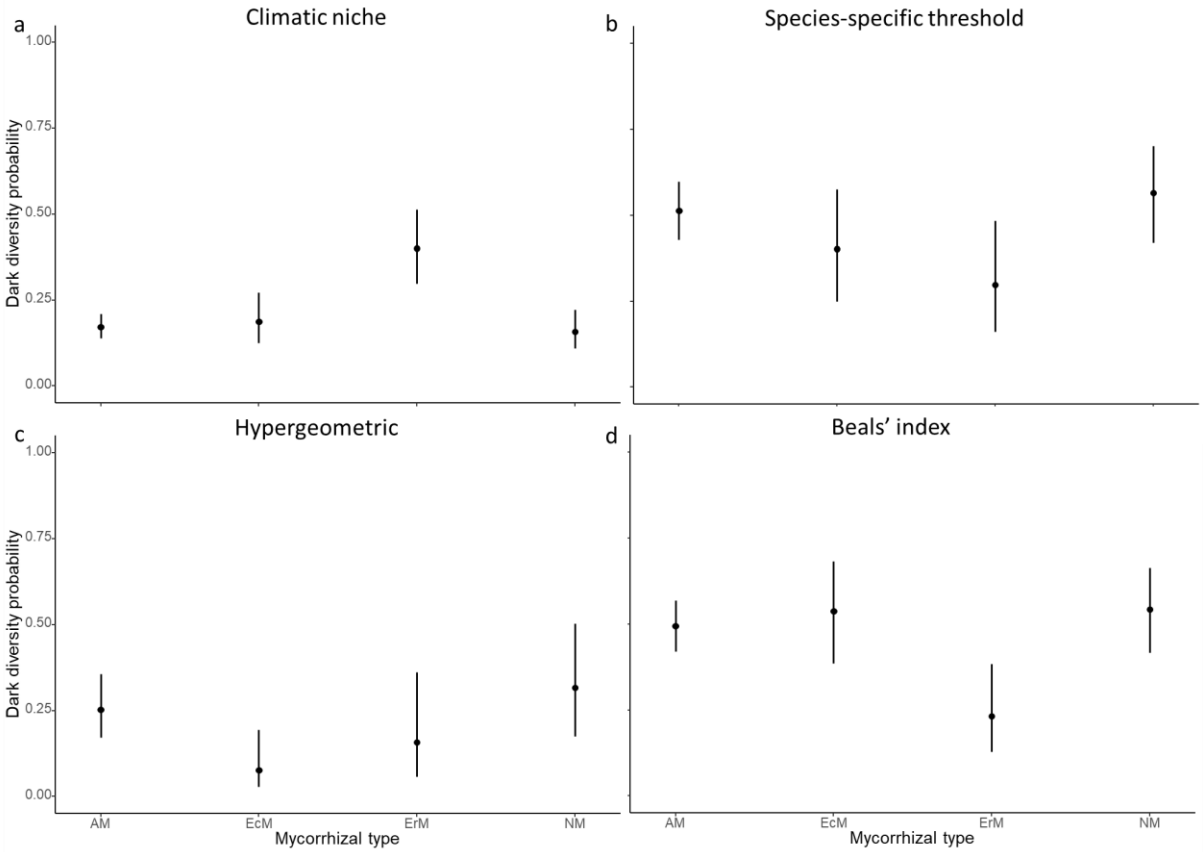

Figure 7: Prediction of the species-level dark diversity in relation to the mycorrhizal type based on the beta regression model. The black dots show the average dark diversity per individual factor whereas the error bars show the standard deviation. AM = arbuscular mycorrhiza; EcM = ectomycorrhiza; ErM = 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.

None of the other variables had a significant influence on the species-level dark diversity.

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

Species richness was identified as a key driver of plot-level dark diversity in three out of the four methods. Its effect was negative for all but the hypergeometric method for which it was not significant, thus largely following our hypothesis. In this system, plots with a low number of species are likely to be dominated by highly competitive species, which can prevent the establishment of several species that could in theory occur there (Pellissier et al., 2010). Indeed, plots with a low species richness in the study system were often dominated by *Empetrum nigrum*. It is an efficient competitor for nutrients, can grow on soils with low pH, and has allelopathic effects against seed germination and the growth of surrounding species (Tybirk et al., 2000), and can possibly direct several species from the regional species pool locally to the dark diversity. Our results seem to support the study by Fløjgaard et al. (2020) who found that competitive species have an adverse effect on species richness, leading to an increase in dark diversity. Nevertheless, it is possible that approaches based on species co-occurrences, such as the hypergeometric method and the Beals' index, already account for this effect of competition.

Finally, the amount of bare ground, soil pH and plot size also appeared to have a significant effect on the plot-level dark diversity, but this was only the case for the niche-based methods. No other variables were significant for the other two methods which already indicates that these models should be handled with caution.

4.2 Species-level dark diversity

Mycorrhizal association was the only variable with significant influence, across all methods, 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 using the climatic niche models, the opposite was true for the climatic niche model followed by a species-specific threshold and the Beals' method. Noteworthy, when using the hypergeometric method species with a symbiotic ectomycorrhizal association had a significantly lower dark diversity than all other associations. These contrasting results highlight the differences between the different methods used to estimate dark diversity. In the Scandinavian mountains, the species with an ErM association

(e.g. *E. nigrum* and *V. vitis-idaea*) 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 modeling approach. We aimed to correct this issue by using species-specific thresholds, yet here again mycorrhizal type was withheld as significant.

These ErM-associated species not only dominated the studied 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 Beals' index. 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 based on 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, which are often used interchangeably in the scientific literature. However, our results suggest 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 from the neighboring 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-level dark diversity, when using climatic niches with or without species-specific thresholds, as well as when using the hypergeometric method versus the Beals' index. Additionally, the community completeness also differed significantly, depending on the method used. As such, our results highlight the need for caution and transparency when calculating and interpreting dark diversity estimates, as the conclusions depend heavily on the methodological decisions one makes, and methods should thus be tailored to the specific research questions.

Of course, several alternative methods could still be used to estimate dark diversity, and many adjustments to the methods used above could be proposed. For example, one could use global datasets such as GBIF to model the climatic niche, rather than data from the study region only. Using global datasets for such broader-scale niche models could result in a more accurate estimate of the climatic niche since the entire climatic niche could be modelled, rather than a truncated version as results from regional data (Bazzichetto et al., 2023). However, most of these global datasets lack absence data and presences are obtained using a wide variety of methodologies and spatial resolutions (Tessarolo et al., 2014), while abiotic data is at the global scale often only available at coarser resolution (Lembrechts et al., 2019). This could also make the predictions less accurate. Additionally, there is the possibility of mismatches, especially for rare species, since global datasets can be spatially biased (Meyer et al., 2016). Therefore, predicting local climatic niches based on global data can make it more difficult to figure out whether the absences are due to a bias in the global dataset or the drivers under investigation. Furthermore, it is worth mentioning that alternative thresholds could be used for the species-specific method, such as Cohen's Kappa or Area Under the Curve (AUC). The chosen TSS threshold in this study may be affected by the low prevalence of species (Leroy et al., 2018). However, since we only used the relatively common species, the issue of low prevalence should not pose a notable concern (Allouche et al., 2006; Wunderlich et al., 2019). These alternative threshold methods were not examined in this particular study as this may further complicate methodological decisions for dark diversity estimation. Hence, we suggest that more research is needed to investigate the impact of alternative thresholds when using species-specific methods.

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

4.5 Conclusions

The concept of dark diversity is still in its infancy, yet its contribution to understanding community completeness and its use in nature conservation has already been shown to be significant (Lewis et al., 2017; Riibak et al., 2015). In this context, it is crucial to determine whether a species' absence is a result of species-specific traits or plot characteristics, be it abiotic factors or biotic interactions, which is something traditional biodiversity studies that only focus on species presences cannot provide. We here compared different methodological approaches to estimate dark diversity and showed significant divergence in predicted drivers of dark diversity based on the method used, calling for caution when interpreting statistical findings on dark diversity estimates. Given the high level of variation in outcome between methods, it is currently not possible to recommend one or the other. More comparative studies in different environments are thus necessary to elaborate further on the search for a robust methodology to estimate dark diversity. Nevertheless, we can generally conclude that areas at low elevations, and, to a certain extent, with a low species richness showed a higher plot-level dark diversity, largely due to natural processes such as competitive dominance. How valid these findings are for patterns in dark diversity in other (mountain) areas across the globe remains to be seen, yet the significant effect of methodological decisions on conclusions should remind us that any other regional study on dark diversity should be cautious in its conclusions. Nonetheless, one could assume

651 that dark diversity will indeed decrease with increasing elevation since only more specialized species
652 can survive at higher elevations, and competition is lower.

653 **Acknowledgments**

654 We thank the master students Renée Lejeune and Jasmine Spreeuwers for their assistance in gathering
655 data during the summer of 2021. Additionally, we extend our appreciation to Stef Haesen for supplying
656 us with the raster layer for the topographic wetness index (TWI).

657 **Author contribution**

658 J.J.L. and L.H. conceived the research idea; L.H., D.W. and J.C. collected data; L.H. performed statistical
659 analyses with guidance from J.J.L.; L.H. and J.J.L. wrote the paper with contributions from K.V.M.; all
660 authors discussed the results and commented on the manuscript.

661 **Data availability statement**

662 All data and codes that support the findings of this study are available on Zenodo
663 https://zenodo.org/record/8059877.

664 **References**

665 Alexander, J. M., Lembrechts, J. J., Cavieres, L. A., Daehler, C., Haider, S., Kueffer, C. et al., (2016)
666 Plant invasions into mountains and alpine ecosystems: current status and future challenges.
667 *Alpine Botany*, 126, 89–103. https://doi.org/10.1007/s00035-016-0172-8
668 Allouche, O., Tsoar, A., & Kadmon, R. (2006) Assessing the accuracy of species distribution models:
669 prevalence, kappa and the true skill statistic (TSS). *Journal of applied ecology*, 43, 1223-1232.
670 https://doi.org/10.1111/j.1365-2664.2006.01214.x
671 Bates, D., Mächler, M., Bolker, B., Walker, S. (2015) Fitting linear mixed-effects models using
672 lme4. *Journal of Statistical Software*, 67, 1–48. doi:10.18637/jss.v067.i01
673 Bazzichetto, M., Lenoir, J., Da Re, D., Tordoni, E., Rocchini, D., Malavasi, M. et al. (2023) Sampling
674 strategy matters to accurately estimate response curves' parameters in species distribution
675 models*. Global Ecology and Biogeography*. https://doi.org/10.1111/geb.13725
676 Beals, E. W. (1984) Bray-curtis ordination: An effective strategy for analysis of multivariate ecological
677 data. *Advances in Ecological Research*, 14, 1–55. https://doi.org/10.1111/oik.07308
678 Belinchón, R., Hemrová, L., & Münzbergová, Z. (2020) Functional traits determine why species belong
679 to the dark diversity in a dry grassland fragmented landscape. *Oikos*, 129, 1468–1480.
680 https://doi.org/10.1111/oik.07308
681 Bjorkman, A. D., Myers-Smith, I. H., Elmendorf, S. C., Normand, S., Thomas, H. J. D., Alatalo, J. M. et al.
682 (2018) Tundra Trait Team: A database of plant traits spanning the tundra biome. *Global*
683 *Ecology and Biogeography*, 27, 1402–1411. https://doi.org/10.1111/geb.12821
684 Broennimann, O., Di Cola, V., Guisan, A. (2022). *Ecospat: spatial ecology miscellaneous methods.*
685 *Version 3.4.* Available at https://CRAN.R-project.org/package=ecospat [Accessed 16 March
686 2022]
687 Brooks, M. E., Kristensen, K., van Benthem, K. J., Magnusson, A., Berg, C. W., Nielsen, A. et al.
688 (2017) "glmmTMB Balances Speed and Flexibility Among Packages for Zero-inflated
689 Generalized Linear Mixed Modeling." *The R Journal*, 9, 378–400.
690 Cadotte, M. W., Tucker, C. M. (2017) Should environmental filtering be abandoned? *Trends in ecology*
691 *and evolution*, 32, 429-437. https://doi.org/10.1016/j.tree.2017.03.004

692 Callaghan, T. V., Jonasson, C., Thierfelder, T., Yang, Z., Hedenås, H., Johansson, M. et al. (2013)
693 Ecosystem change and stability over multiple decades in the Swedish subarctic: Complex
694 processes and multiple drivers. *Philosophical Transactions of the Royal Society B: Biological*
695 *Sciences*, 368, 1–17. https://doi.org/10.1098/rstb.2012.0488
696 Callaway, R. M., Brooker, R. W., Choler, P., Kikvidze, Z., Lortie, C. J., Michalet, R. et al. (2002) Positive
697 interactions among alpine plants increase with stress. *Nature*, 417, 844–848.
698 https://doi.org/10.1038/nature00812
699 Carmona, C. P., & Pärtel, M. (2020) Estimating probabilistic site-specific species pools and dark
700 diversity from co-occurrence data. *Global Ecology and Biogeography*, 30, 316–326.
701 https://doi.org/10.1111/geb.13203
702 Chamberlain, S., Barve, V., Mcglinn, D., Oldoni, D., Desmet, P., Geffert, L., & Ram, K. (2021) *rgbif:*
703 *Interface to the Global Biodiversity Information Facility API*. *Version 3.6.0*. Available at
704 https://CRAN.R-project.org/package=rgbif [Accessed 15 May 2021]
705 Christensen, E., Christensen, B., & Christensen, S. (2021) Problems in using Beals' index to detect
706 species trends in incomplete floristic monitoring data (Reply to Bruelheide et al. (2020)).
707 *Diversity and Distributions*, 27, 1324–1327. https://doi.org/10.1111/ddi.13276
708 Cribari-Neto, F., Zeileis, A. (2010) Beta Regression in R. *Journal of Statistical Software,* 34, 1–24.
709 Crooks, J. A. (2005) Lag times and exotic species: The ecology and management of biological invasions
710 in slow-motion. *Ecoscience*, 12, 316–329. https://doi.org/10.2980/i1195-6860-12-3-316.1
711 de Bello, F., Fibich, P., Zelený, D., Kopecký, M., Mudrák, O., Chytrý, M. Et al. (2016) Measuring size and
712 composition of species pools: a comparison of dark diversity estimates. *Ecology and Evolution*,
713 6, 4088–4101. https://doi.org/10.1002/ece3.2169
714 Ellenberg, H., Weber, H.E., Düll, R., Wirth, V., Werner, W. & Paulißen, D. (1991) *Zeigerwerte von*
715 *Pflanzen in Mitteleuropa.* Scripta Geobotanica.
716 European Union, Copernicus Land Monitoring Service 2021, European Environment Agency (EEA).
717 Ewald, J. (2002) A probabilistic approach to estimating species pools from large compositional
718 matrices. *Journal of Vegetation Science*, 13, 191–198. https://doi.org/10.1111/j.1654-
719 1103.2002.tb02039.x
720 Finlay, R. D. (2008) Ecological aspects of mycorrhizal symbiosis: With special emphasis on the
721 functional diversity of interactions involving the extraradical mycelium. *Journal of*
722 *Experimental Botany*, 59, 1115–1126. https://doi.org/10.1093/jxb/ern059
723 Fløjgaard, C., Valdez, J. W., Dalby, L., Moeslund, J. E., Clausen, K. K., Ejrnæs, R. et al. (2020) Dark
724 diversity reveals importance of biotic resources and competition for plant diversity across
725 habitats. *Ecology and Evolution*, 10, 6078-6088. https://doi.org/10.1002/ece3.6351
726 Fox, J., & Weisberg, S. (2011). An R companion to applied regression (2nd edition). SAGE Publications
727 Inc.
728 Gijbels, P., Adriaens, D. & Honnay, O. (2012) An orchid colonization credit in restored calcareous
729 grasslands. *Ecoscience,* 19, 21–28. https://doi.org/10.2980/19-1-3460
730 Gough, L., Shaver, G. R., Carroll, J., Royer, D. L., & Laundre, J. A. (2000) Vascular plant species richness
731 in Alaskan arctic tundra: The importance of soil pH. *Journal of Ecology*, 88, 54–66.
732 https://doi.org/10.1046/j.1365-2745.2000.00426.x
733 Guisan, A., & Thuiller, W. (2005) Predicting species distribution: Offering more than simple habitat
734 models. *Ecology Letters*, 8, 993–1009. https://doi.org/10.1111/j.1461-0248.2005.00792.x
735 Haesen, S., Lembrechts, J. J., De Frenne, P., Lenoir, J., Aalto, J., Ashcroft, M. et al. (2021) ForestTemp
736 Sub-canopy microclimate temperatures of European forests. *Global Change Biology*, 27, 6307–
737 6319. https://doi.org/10.1111/gcb.15892
738 Hartig, F. (2018) *Yes, statistical errors are slowing down scientific progress! Theoretical ecology.*

739 Available at https://theoreticalecology.wordpress.com/2018/05/03/yes-statistical-errors-are-
740 slowing-down-scientific-progress/ [Accessed 2 November 2022]
741 Hartig, F. (2022) *DHARMa: Residual diagnostics for hierarchical (multi-level / mixed) regression*
742 *Models. R package version 0.4.5.* Available at https://CRAN.R-project.org/package=DHARMa
743 [Accessed 14 November 2021]
744 Hijmans, R. J., van Etten, J. (2012) *Raster: Geographic data analysis and modeling. R package version*
745 *3.6-3.* Available at http:// CRAN.R-project.org/package=raster [Accessed 5 October 2021]
746 Hooper, D. U., Adair, E. C., Cardinale, B. J., Byrnes, J. E. K., Hungate, B. A., Matulich, K. L. et al. (2012)
747 A global synthesis reveals biodiversity loss as a major driver of ecosystem change. *Nature*, 486,
748 105–108. https://doi.org/10.1038/nature11118
749 Howe, F., & Smallwood, J. (1982) Ecology of seed dispersal. *Annual Review of Ecology and Systematics,*
750 *13*, 201–228. https://doi.org/10.1146/annurev.es.13.110182.001221
751 Jonasson, S. (1988) Evaluation of the point intercept method for the estimation of plant biomass.
752 *Oikos*, *52*, 101–106.
753 Jones, N. T., & Gilbert, B. (2016) Biotic forcing: the push–pull of plant ranges. *Plant Ecology*, 217, 1331–
754 1344. https://doi.org/10.1007/s11258-016-0603-z
755 Karger, D. N., Conrad, O., Böhner, J., Kawohl, T., Kreft, H., Soria-Auza, R. W. et al. (2017) Climatologies
756 at high resolution for the Earth land surface areas. *Scientific Data,* 4, 170122.
757 https://doi.org/10.1038/sdata.2017.122
758 Keddy, P. A. (1992) Assembly and response rules: two goals for predictive community ecology. *Journal*
759 *of vegetation science*, 3, 157-164. https://doi.org/10.2307/3235676
760 Keitt, T., Bivand, R., Pebesma, E. & Rowlingson, B. (2010) *rgdal: Bindings for the Geospatial Data*
761 *Abstraction Library. Version 1.6-5.* Available at: https://CRAN.R-project.org/package=rgdal
762 [Accessed 28 March 2023]
763 Klanderud, K. (2010) Species recruitment in alpine plant communities: The role of species interactions
764 and productivity. *Journal of Ecology*, 98, 1128–1133. https://doi.org/10.1111/j.1365-
765 2745.2010.01703.x
766 Kleyer, M., Bekker, R. M., Knevel, I. C., Bakker, J. P., Thompson, K., Sonnenschein, M. et al. (2008) The
767 LEDA traitbase: a database of life- history traits of the northwest European flora. *Journal of*
768 *Ecology,* 96, 1266–1274. https://doi.org/10.1111/j.1365-2745.2008.01430.x
769 Kopecký, M., Macek, M., & Wild, J. (2021) Science of the Total Environment Topographic Wetness
770 Index calculation guidelines based on measured soil moisture and plant species composition.
771 *Science of the Total Environment*, 757, 143785.
772 https://doi.org/10.1016/j.scitotenv.2020.143785
773 Körner, C. (2021) *Alpine plant life: Functional plant ecology of high mountain ecosystems*, 3rd edition.
774 Springer Nature Switzerland AG 2021. https://doi.org/10.1007/978-3-030-59538-8
775 Kullman, L. (2015) Recent and past trees and climates at the Arctic/Alpine margin in Swedish Lapland:
776 An Abisko case study Review. *Journal of Biodiversity Management & Forestry*, 4, 1–12.
777 doi:10.4172/2327-4417.1000150
778 Legendre P, Legendre J (1998) *Numerical Ecology*, 3rd edition. Elsevier, Amsterdam.
779 Lembrechts, J. J., Lenoir, J., Nuñez, M. A., Pauchard, A., Geron, C., Bussé, G. et al. (2018) Microclimate
780 variability in alpine ecosystems as stepping stones for non-native plant establishment above
781 their current elevational limit. *Ecography*, 41, 900–909. https://doi.org/10.1111/ecog.03263
782 Lembrechts, J. J., Lenoir, J., Roth, N., Hattab, T., Milbau, A., Haider, S. et al (2019) Comparing
783 temperature data sources for use in species distribution models: From in-situ logging to
784 remote sensing. *Global Ecology and Biogeography*, 28, 1578–1596.
785 https://doi.org/10.1111/geb.12974
786 Lembrechts, J. J., Milbau, A., & Nijs, I. (2014) Alien roadside species more easily invade alpine than

lowland plant communities in a subarctic mountain ecosystem. *PLoS ONE*, 9, 1–10. https://doi.org/10.1371/journal.pone.0089664

Lembrechts, J. J., Aalto, J., Ashcroft, M. B., De Frenne, P., Kopecký, M., Lenoir, J. et al. (2020) SoilTemp: A global database of near-surface temperature. *Global Change Biology,* 26, 6616-6629. https://doi.org/10.1111/gcb.15123

Lembrechts, J. J., van den Hoogen, J., Aalto, J., Ashcroft, M., De Frenne, P., Kemppinen, J. et al (2021) Global maps of soil temperature. *Global Change Biology,* 28, 3110-3144. https://doi.org/10.1111/gcb.16060

Lenoir, J., Gégout, J. C., Guisan, A., Vittoz, P., Wohlgemuth, T., Zimmermann, N. E. et al. (2010) Cross-scale analysis of the region effect on vascular plant species diversity in southern and northern European mountain ranges. *PLoS One,* 5, e15734. https://doi.org/10.1371/journal.pone.0015734

Lenth, R. (2022) *Emmeans: Estimated marginal means, aka least-squares means. Version 1.8.2.* Available at: https://CRAN.R-project.org/package=emmeans [Accessed 14 November 2021]

Leroy, B., Delsol, R., Hugueny, B., Meynard, C. N., Barhoumi, C., Barbet-Massin, M., & Bellard, C. (2018) Without quality presence–absence data, discrimination metrics such as TSS can be misleading measures of model performance. *Journal of Biogeography*, 45, 1994-2002. https://doi.org/10.1111/jbi.13402

Lewis, R. J., de Bello, F., Bennett, J. A., Fibich, P., Finerty, G. E., Götzenberger, L. et al. (2017) Applying the dark diversity concept to nature conservation. *Conservation Biology*, 31, 40–47. https://doi.org/10.1111/cobi.12723

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

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]

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

Pärtel, M., Szava-Kovats, R., & Zobel, M. (2013) Community completeness: Linking local and dark diversity within the species pool concept. *Folia Geobotanica,* 48, 307–317. https://doi.org/10.1007/s12224-013-9169-x

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

QGIS Development Team. (2021) QGIS geographic information system. Open Source Geospatial Foundation Project. Available from: http://www.qgis.org

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]

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
Trindade, D. P., Carmona, C. P., Reitalu, T., & Pärtel, M. (2023) Observed and dark diversity dynamics over millennial time scales: fast life-history traits linked to expansion lags of plants in northern Europe. *Proceedings of the Royal Society B*, 290, 20221904. https://doi.org/10.1098/rspb.2022.1904
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 species richness: The importance of daily maximum temperature and pH. *Plant Ecology*, 184, 13–25. https://doi.org/10.1007/s11258-005-9048-5
Wedegärtner, R. E., Lembrechts, J. J., van der Wal, R., Barros, A., Chauvin, A., Janssens et al. (2022) Hiking trails shift plant species' realized climatic niches and locally increase species richness. *Diversity and Distributions*, 28, 1416-1429 https://doi.org/10.1111/ddi.13552
Westoby, M. (1998) A leaf-height-seed (LHS) plant ecology strategy scheme. *Plant and Soil*, 199, 213–227. https://doi.org/10.1023/A:1004327224729
Wiegmans, D., Larson, K., Clavel, J., Spreeuwers, J., Pirée, A., Nijs, I., & Lembrechts, J. (2023) Historic disturbance events may overrule climatic factors as drivers of ruderal species distributions in the Scandinavian mountains. *Authorea Preprints*. 10.22541/au.167515746.69858832/v1
Wunderlich, R. F., Lin, Y. P., Anthony, J., & Petway, J. R. (2019) Two alternative evaluation metrics to replace the true skill statistic in the assessment of species distribution models. *Nature Conservation*, 35, 97-116.

Appendices

Appendix S1. Details online data collection
Appendix S2. Additional results
Appendix S3. Contrast tests mycorrhizal associations