NTNU

Norwegian University of
Science and Technology

Patterns and Drivers of Phylogenetic Diversity and Endemism in the Norwegian Vascular Flora

Ida Marielle Mienna

MSc in Biology
Submission date: May 2018
Supervisor: Michael D Martin, IBI
Co-supervisor: James Speed, IBI
Mika Bendiksby, IBI

Norwegian University of Science and Technology
Department of Biology

ACKNOWLEDGEMENTS

This Master's thesis in Biology is written at the Department of Natural History, Norwegian University of Science and Technology (NTNU) University museum. The project was funded by the NTNU University museum and Nord-Troms Regional Council DA. I want to thank my three supervisors; Mike (Michael), James, and Mika. Thank you for helping me with the project, sending me relevant literature, and guidance when I needed it. The organised group meetings in systematics by Mike and Mika every second week helped me in progressing forward. Thanks to Mike for teaching me how to extract and sequence DNA, and for guiding me with the alignments and phylogeny. I want to especially thank him for being so optimistic about the project, which has really motivated me! Thanks to James for always being available when I needed answers about statistics, R and GIS, and also for discussing the results in an ecological way. I am also very thankful that he introduced me for this project, for letting me attend one symposium and two conferences, and for always believing in me. Thanks to Mika for introducing me to the world of plant systematics, for encouraging me to present the project at ForBio in 2018, and for helping me with the alignments and inferring the phylogeny. Her thorough proof-reading and commenting on all written material have taught me valuable lessons that I will carry with me in the future. I would like to thank Mika and Charlotte Bjorå for letting me borrow herbarium specimens from the herbaria in Trondheim (TRH) and Oslo (O), respectively. Thanks to Malene and Mika for assisting me in sampling the herbarium specimens from TRH and O, respectively, and for teaching me how to extract plant tissue. Thanks to Sonia Nosratinia (UC Berkeley) for technical advice on amplifying the markers, Erik, Narcis, Markus and Andreas for guiding me in the lab when I needed it, and Marie Brunier for assisting me with the lab work. Thanks to Brent Mishler and Andrew Thornhill for giving me feedback on my preliminary results, and especially Andrew for modifying and helping me with the Biodiverse pipeline. I thank Asbjørn Moen for discussions on my preliminary results and for giving me guidance on the Norwegian vascular flora and its drivers. Thanks to Siri, for all the hours at the gym, and for always being there when I needed it. Thanks to Ingegjerd for being my museum study partner and all the wine we have drunk. Thanks to Veronica and Live for keeping my motivation for biology where it was supposed to be, and to Victoria for all the conversations we have had at the museum. Thanks to Sivert for always making me dinner when needed, and to Torstein for keeping me sane. In the end, I would like to thank the lunch group at the museum, and all others that have given me feedback on the project and helped me in any way. This Master's thesis would not have been the same without you!

ABSTRACT

Biodiversity is commonly measured as the local abundance of discrete species. This approach fails to incorporate evolutionary relationships among species and gives an incomplete picture of the diversity in a given region. Recently developed methods for *spatial phylogenetics* combine species occurrence datasets with molecular phylogenetic data to recover information about the spatial distribution of phylogenetic diversity and endemism across a geographic region. In this study, we applied these methods on vascular plants across Norway, a region that was covered by the Fennoscandian ice sheet until about 10 kyr BP. Our primary aim was to test whether patterns of phylogenetic diversity (PD) and phylogenetic endemism (PE) for 1238 native vascular plant species are non-randomly distributed across Norway and to seek causal mechanisms for the revealed patterns. A multi-locus concatenated alignment was produced using a combination of GenBank data and newly produced sequence data from herbarium specimens, and a corresponding phylogenetic hypothesis was inferred. We combined this data with occurrence data from GBIF, and then calculated PD, PE, relative phylogenetic diversity (RPD), and relative phylogenetic endemism (RPE). To assess areas of statistical significance we performed a spatial randomisation of these indices. To determine if PD outlier regions are associated with various environmental explanatory variables, we performed pairwise linear correlation analyses. We also performed range-weighted turnover analyses and compared these maps with previous studies on vegetation zones in Norway. Our results revealed significantly high PD and RPD (phylogenetic overdispersion) along the coast of southern Norway and significantly low PD and RPD (phylogenetic clustering) in the mountainous areas. The best fitted statistical models showed that the PD pattern could best be explained by temperature, time since last glaciation cover, and topographic heterogeneity. For PE and RPE, we found a concentration of short restricted branches (neo-endemics) in northern Norway, and a mix of short and long restricted branches (both neo-endemics and paleo-endemics) on the coast of southern Norway, with temperature inferred as a likely driver for the revealed patterns. The range-weighted turnover analyses showed an evolutionary signal in the way the flora is distributed. This study emphasizes the importance of incorporating evolutionary relationships between species to see patterns of diversity and endemism that would normally not be seen by examining species diversity alone. Moreover, it shows that *spatial phylogenetics* can be meaningfully applied to a region of short evolutionary history as opposed to previous studies restricted to regions with long evolutionary history (e.g. Australia and California).

SAMMENDRAG

Biodiversitet er til vanlig målt som det lokale antallet arter. Denne fremgangsmåten inkorporerer ikke evolusjonært slektskap mellom arter, og gir derfor et ufullstendig bilde av biodiversiteten i et gitt område. Nylig utviklede metoder innenfor *spatial phylogenetics* kombinerer forekomstdata av arter med molekylær fylogenetisk data for å undersøke fordelingen av fylogenetisk diversitet og endemisme på tvers av en geografisk region. I denne studien har vi anvendt disse metodene på karplanter i Norge, en region som var dekket av is så sent som for 10 tusen år siden. Vårt primære mål har vært å teste om plantefylogenetisk diversitet ikke er tilfeldig fordelt i Norge, samt identifisere mulige årsaks relaterte mekanismer for det observerte mønsteret. En hypotese om evolusjonært slektskap (dvs. et fylogenetisk tre) mellom artene ble produsert ved å bruke flere gener både fra GenBank og fra egenproduserte sekvenser fra herbarium belegg. Vi kombinerte det fylogenetiske treet med forekomstdata av fra GBIF, og beregnet så fylogenetisk diversitet, fylogenetisk endemisme, relativ fylogenetisk diversitet og relativ fylogenetisk endemisme. Vi utførte en romlig randomisering for å evaluere områder med statistisk signifikans. For å undersøke om områdene er assosiert med forskjellige miljøbaserte forklaringsvariabler utførte vi parvis lineære korrelasjonsanalyser. Vi utførte også *range-weighted turnover* analyser og sammenlignet disse kartene med tidligere studier av vegetasjonssoner i Norge. Våre resultater viste signifikant høy fylogenetisk diversitet langs kysten av Sør-Norge, og signifikant lav fylogenetisk diversitet i fjellområder. De beste statistiske modellene viste at dette mønsteret var best forklart av temperatur, tid siden området var dekket av is, og topografisk heterogenitet. For fylogenetisk endemisme fant vi en konsentrasjon av korte begrensete grener (neo-endemismer) i Nord-Norge, og en blanding av korte og lange begrensete grener (både neo-endemismer og paleo-endemismer) på kysten av Sør-Norge. Temperatur hadde den beste forklaringsevnen for det observerte mønsteret. *Range-weighted turnover* analysen viste et evolusjonært signal i måten floraen i Norge er utbredt. Dette studiet viser viktigheten av å inkorporere evolusjonært slektskap mellom arter for å identifisere diversitet- og endemismemønstre som vanligvis ikke er detekterbart ved kun å undersøke artsdiversiteten i et område. Vår studie viser også at fagfeltet *spatial phylogenetics* er relevant også i områder med kortere evolusjonær historie sammenlignet med tidligere studier som kun har blitt utført i områder med lang evolusjonær historie (f.eks. Australia og California).

TABLE OF CONTENTS

ABBREVIATIONS

AIC	Akaike Information Criteria
APG IV	Angiosperm Phylogeny Group IV
Cal kyr	Thousand years
CANAPE	Categorical Analysis of Neo- And Paleo-Endemism
GBIF	Global Biodiversity Information Facility
MSA	Multiple sequence alignment
O	Oslo herbarium
PD	Phylogenetic diversity
PE	Phylogenetic endemism
PE_alt	Phylogenetic endemism alternative tree
PE_orig	Phylogenetic endemism original tree
PhyloRWTurnover	Range-weighted phylogenetic turnover
RPD	Relative phylogenetic diversity
RPE	Relative phylogenetic endemism
RWTurnover	Range-weighted turnover
TRH	Trondheim herbarium
WE	Weighted endemism

1 INTRODUCTION

Biodiversity is commonly measured as the local abundance of discrete species, i.e. species richness (SR), and it is this metric that is most widely used in biodiversity conservation. This is because each species is genetically unique and each has a particular function in the ecosystem (Purvis and Hector, 2000). Today, we are facing the sixth mass extinction of species on Earth (Barnosky et al., 2011). Certain species may have strong effects on ecosystem processes (Chapin et al., 2000, Schwartz and Simberloff, 2001, Hooper et al., 2005). Loss of biodiversity may therefore have dramatic consequences and biodiversity conservation issues are important to assess.

Using SR to measure biodiversity may, however, mislead conservation assessments, as this measure fails to incorporate information about relationships between the species. This information is important for multiple reasons. Firstly, considering species alone does not necessarily provide adequate information concerning the processes behind the observed biodiversity. Knowledge of evolutionary relationships may illuminate the origin of diversity patterns as well as the drivers that gave rise to it (Hawkins et al., 2014, Lagomarsino et al., 2016). Secondly, ecological and functional similarities between species are generally shaped by their common ancestry. Previous studies have found a positive relationship between ecosystem functioning and low relatedness between the species present, even when functional traits and other diversity metrics were non-significant (Cadotte et al., 2008, Cadotte et al., 2009, Tan et al., 2012) underlining the importance of taking species relatedness into consideration. Thirdly, because closely related species often are more functionally and ecologically similar than distantly related species (i.e. phylogenetic autocorrelation; (Gittleman and Kot, 1990), species should not be assessed as independent units in biodiversity assessments (Felsenstein, 1985). Fourthly, integrating evolutionary history in biodiversity conservation planning of species with extinction risks is thought to be important (Forest et al., 2015).

The restriction of biodiversity to certain areas (i.e. endemism) is also often measured by looking at species alone. This approach gives an incomplete picture of endemism in an area as endemism can occur for clades at all phylogenetic levels. Endemism is often used to identify areas for conservation priority (Myers et al., 2000). Assessing meaningful measures of endemism therefore becomes important. The species' evolutionary history is relevant for various measures of endemicity, such as neo-endemism (i.e. recently evolved) or paleo-

endemism (i.e. anciently evolved). Endemism category is informative about the biogeographic history of the area (Casal-Lopez et al., 2018) and can aid in identifying potential areas of particular biological interest (Nieto-Blázquez et al., 2017). It is also important to look at the differences in diversity between regions (i.e. beta diversity). Beta diversity is commonly measured as species composition change between sites (i.e. turnover), which describes species that the different areas have in common (Whittaker, 1960). Patterns of change in the beta diversity can be used to group regions of similar plant diversity into phytogeographic regions and communities (Cavender-Bares et al., 2009). Ecologists have long tried to understand what shapes a community, and positive (i.e. facilitation) and negative (i.e. competition) interactions between species have been suggested to play a role (Bertness and Callaway, 1994). However, other drivers may have an impact on the composition that cannot be identified by studying the community at species level alone. Examining a region's biodiversity in a phylogenetic context may reveal unidentified drivers.

The approach of *spatial phylogenetics* combines species occurrence data with molecular phylogenetics to recover information about the spatial distribution of phylogenetic diversity (PD) and phylogenetic endemism (PE) across an arbitrary geographic region. Faith (1992) described PD as the sum of the lengths of the phylogenetic branches separating all species occurring in a region. As opposed to SR, which estimates the level of biodiversity in an area by summarizing the number of species present, PD considers the relatedness between the species present. Hence, areas that have more terminal branches (i.e. more species) than other areas may have less PD than a species-poor area if the species in the species rich area are overall more closely related (short branches in the phylogeny). An area with low PD has less genetic diversity than an area with high PD (i.e. more distantly related species). Just as with PD, PE uses the sum of branch lengths in an area, as opposed to endemism, which uses species. By doing so, an area's level of endemism will be measured by how restricted each branch is (Rosauer et al., 2009). Other measures of PD and PE enables us to look at the distribution of branch lengths in the study area. An area with high amounts of long branches can be interpreted as having species with no or few close relatives present (Mishler et al., 2014). If the species or branch also is very restricted, it can possibly be paleo-endemic (i.e. a relict species). Areas with high amounts of short branches can however be interpreted as having many close relatives present in the same area (Mishler et al., 2014), but it can also be neo-endemic (i.e. a recently evolved species). Looking at these patterns is interesting as they can say something about what forces have formed the community as discussed above. Turnover can also be measured phylogenetically,

where the phylogenetic relatedness between the taxa in the region is measured using the branch lengths from a phylogeny (Graham and Fine, 2008). Incorporating phylogenetic information in the regionalization can help to identify regions that are evolutionary distinct from each other (Daru et al., 2017a) and to discover areas of interest that would not be identified by using species only (Daru et al., 2017b).

Approaches for studying PD and PE patterns on a large-scale are possible today for various reasons. Available data from GenBank (Benson et al., 2013) with uploaded sequences of broad taxonomic coverage make it possible to make multi sequence alignments and phylogenies for different taxa. The value of natural history collections is also apparent here because of exchange agreements between the herbaria where one can borrow specific specimens for research use. This makes acquiring species needed for a study easier, as the herbarium collection may have rare and endangered species that one normally would not be able to sample from nature (Bieker and Martin, 2018). For the occurrence data, the major digitation efforts of herbaria make it possible to use the coordinates of both preserved specimens and observed occurrences of taxa (James et al., 2018, Sweeney et al., 2018).

In this study, we take advantage of these newly available data to study vascular plants in Norway. This Northern European country has digitized herbarium collections and a total of over five million vascular plant occurrences available on the online Global Biodiversity Information Facility (GBIF). Earlier studies have investigated spatial patterns of PD and PE of vascular plants in Australia (Mishler et al., 2014, Nagalingum et al., 2015, Thornhill et al., 2016), Chile (Scherson et al., 2017), and California (Thornhill et al., 2017). What these areas have in common is that they have relatively long evolutionary histories with limited disturbance over the evolutionary timescale. Because of this, the organisms in the area have had time to adapt to the area and form new species. These previously studied areas also contain several hotspots for vascular plant diversity (Myers et al., 2000).

Norway, however, does not fall within these categories. This region was covered by the Fennoscandian Ice Sheet 22-9.7 thousand years ago (kyr BP) (Stroeven et al., 2016). The plants and animals now inhabiting the area had to disperse from glacial refugia to where they are today. Studies have found *tabula rasa* refugia (Nordal, 1987) in southern Europe (Hewitt, 1999, Nordal et al., 2005), but some studies also argument for *in situ* refugia, or so called *nunataks*, rising above the ice sheet (Westergaard et al., 2011, Parducci et al., 2012). Earlier studies have tried to delimit the immigration routes plants and animals may have used when dispersing to

the Fennoscandian region. Some mammals, like *Ursus arctos* and *Microtus agrestis* (Taberlet et al., 1998), have a hybrid zone in the middle of Sweden, suggesting two dispersal routes: one from southern and western Europe and the other from Finland and eastern Europe. This has also been seen for some vascular plants, e.g. *Cerastium alpinum* (Berglund and Westerbergh, 2001). Some plant species, e.g. *Picea abies* (Lagercrantz and Ryman, 1990) and *Fagus sylvatica* (Demesure et al., 1996), have a one-way colonization route.

Norway is not a hotspot for vascular plants. Approximately 2880 known vascular plant species occur in Norway, but only about 46% of these are native to the area (i.e. not introduced by humans) (Lid and Lid, 2005). However, Norway contains some species that are highly restricted, like *Primula scandinavica*, which is endemic to Norway and Sweden (Guggisberg et al., 2006), and *Sorbus lancifolia*, which can be found in only one location in Norway (Westergaard, 2012). Wide variation in topography and climate can be found in Norway, as can all vegetation zones that are recognized in Northern Europe (Bakkestuen et al., 2008), making Norway an interesting region to study.

Moen et al. (1999) grouped the flora of Norway into floristic elements based on where the different species are distributed (e.g. western species from western Europe and eastern species from north-eastern Europe and Russia.) and their potential immigration routes to Norway. The distributions of the species can be linked to climate. For example, western species often demand milder winter temperatures, and eastern species avoid mild winters (Moen et al., 1999). Because the different species have different climatic requirements for survival, it is possible to group the vegetation based on these requirements. Norway's different vegetation zones have already been described in the past (Wille, 1915, Dahl, 1998, Moen et al., 1999, Bakkestuen et al., 2008), and Moen et al. (1999) classified Norway into six vegetation zones mainly based on climate: alpine, northern boreal, middle boreal, southern boreal, boreonemoral and nemoral.

In this study, our primary aim is to estimate and analyse the PD and PE patterns of the native vascular plants across Norway. We also seek to find causal mechanisms for the revealed patterns. Our final aim is to estimate different compositions of plant communities in Norway, and differences in the relatedness between these communities. We expect SR and PD to be highly correlated (Faith, 1992), and to observe higher values of these two metrics in areas with high sampling intensity. Earlier studies have found significantly low PD in areas with extreme values of temperature and little precipitation (Thornhill et al., 2016, Thornhill et al., 2017).

Grytnes (2000) found that there were more species per family in northern Fennoscandia and in higher altitudes in southern Fennoscandia. Because of this, we hypothesise that these same areas will have low PD as this suggests that the species present are more closely related to each other. Earlier studies have found significantly high PD in areas with more precipitation (Thornhill et al., 2017, Scherson et al., 2017), and we thus expect the same here.

For the causal mechanisms, we expect temperature and precipitation to have a positive effect on both SR and PD. This is because climate has been found to be a strong predictor for plant diversity, where warmer temperatures and higher values of annual precipitation are correlated with plant SR (Kreft and Jetz, 2007). For significantly high and low PD, we expect lower temperatures and precipitation to be a predictor for low PD because of more stressful and harsher climates leading to a phylogenetic clustering. Because these environments are more stressful and harsh for plants, we expect there to be an environmental or habitat filtering for traits that make these plants tolerate this type of environment (Cavender-Bares et al., 2009). For significantly high PD, we expect the opposite. We expect regional spatial heterogeneity in terms of habitat (number of vegetation types) and topography (the variation in elevation) will have positive effects on SR and PD, as these variables have been found to be strong predictors of plant richness (Bruun et al., 2006, Kreft and Jetz, 2007). Because earlier studies have found plant diversity to increase with increasing soil pH (Pärtel, 2002), we also expect a positive relationship here. Since the number of vascular plants in Norway is still increasing (Felde et al., 2017), and because earlier studies have found that recently glaciated areas inhabit less species because of succession (Erschbamer et al., 2008), we expect diversity to be positively correlated with the time since the area was covered by the Fennoscandian ice sheet.

Just as with SR and PD, we expect weighted endemism and PE to be correlated with each other and show a bias towards areas with high sampling intensity. We expect to find centres of endemism close to potential immigration routes from the south and the northeast, as these areas would have had the longest time for species to arrive, and because recently arrived species may still have restricted distributions. As paleo-endemism is correlated with high water availability (Jordan et al., 2016), we expect to find this endemism category in areas with higher amounts of precipitation.

For the different compositions of Norwegian plant communities, we expect to find a close resemblance with Moen's vegetation zones from 1999. However, in concordance to previous studies on immigration routes and hybridization zones (Taberlet et al., 1998, Berglund

and Westerbergh, 2001) and the grouping of plants into floristic elements (Moen et al., 1999), we expect to see a boundary in the middle of Norway where north-eastern species meet south-western species.

2 MATERIALS AND METHODS

2.1 Species selection

A list of vascular plant species occurring in Norway was obtained from Artsdatabanken (Supplementary Table B1). Only species native to Norway were used. We obtained the information concerning native status from Lid and Lid (2005), which defines native species as those that have come to an area without the help of humans. Species that were assessed as native in Lid and Lid (2005), but marked as not evaluated (NE) or data deficient (DD) in the original list from Artsdatabanken (Supplementary Table B1), were included in the list. Species were selected so that each terminal node in the phylogeny would represent a single species (Supplementary Table B2). All molecular and species occurrence data were based on the same species name (or synonym). For the occurrences, the data had to have geographic coordinates within the Norwegian mainland border (i.e. did not include overseas areas like Svalbard and Jan Mayen). However, for the molecular data the specimen did not have to be collected within Norway.

Species from the original list that had no occurrences on GBIF, or where GBIF considered the occurrences as doubtful, were removed. Further, only species with available data on GenBank (Benson et al., 2013) or BOLDSYSTEMS (Ratnasingham and Hebert, 2007), or were available for loan from the herbaria in Trondheim (TRH) and Oslo (O) (see Materials and methods – molecular data), were selected.

Not all species from the genera *Hieracium*, *Pilosella* and *Taraxacum* were included in this study, as species delimitation for these groups is unclear (Lid and Lid, 2005). A few species from these three genera were selected so that the genera would be represented in the tree. These were chosen by subjectively selecting representative species from each genus section in Lid and Lid (2005). Only species with sequences on GenBank/BOLDSYSTEMS or herbarium specimens at TRH, and with occurrences on GBIF were included.

2.2 Species occurrence data

We downloaded occurrence data from the Global Biodiversity Information Facility (GBIF) for all included species using the package *rgbif* (Chamberlain et al., 2017) in R Studio Version 3.4.1 (R Core Team, 2017) between November 21st – 27th 2017 (GBIF.org, 2017). We used the

function *name_backbone* to extract the species keys needed to search for the occurrence data of each species (Supplementary table B2). For taxa that Lid and Lid (2005) or Artsdatabanken (Supplementary table B1) defined as a species, but where GBIF defined the taxa as a synonym, hybrid, or subspecies, we used a usage key rather than a species key to search for and download occurrences. We downloaded both occurrence records stated as human observations as well as those with accompanying preserved specimens; the latter normally preferred (Speed et al., 2018). This, in order to get the best possible coverage of the species' distributional ranges. Species that lacked a species key or occurrences were examined to find possible errors, and if no errors were found, the species were excluded from the study. Species that had only one occurrence (e.g. *Hieracium bifidum*) and/or very old occurrences (older than the year 1900; e.g. *Coleanthus subtilis*) were excluded from the study.

An outline of the Norwegian border with a 2-km buffer acquired from the packages *raster* (Hijmans, 2017) and *rgeos* (Bivand and Rundel, 2017) was used to remove occurrences outside of Norway. Occurrences with a spatial uncertainty larger than 2000 meters were removed from the data set. A total of 3 597 865 occurrences were used in the analyses and converted into a species-level presence-absence grid using 1094 20x20-km grid cells. To ensure that spatial analyses of diversity was limited to well sampled regions, the data was tested for completeness (Chao, 1987). Here, the expected number of species in a cell was calculated by using the function *estimateR* from the package *vegan* (Oksanen et al., 2017) in R, and completeness was calculated by dividing the observed species occurrence number in a cell by the expected species number. A total of 44 cells which with a completeness of less than 50% were removed from the dataset. One species (*Chamaedaphne calyculata*) was removed in the process as it only occurred in a cell with low completeness. The data set was projected to an equal area UTM grid (WGS 84 / UTM zone 33N) and converted into comma-separated value (CSV) format for further analyses.

2.3 Molecular data

For the Norwegian vascular plants, the nuclear marker *ITS* (internal transcribed spacer) and the two chloroplast genes *matK* (maturase K gene) and *rbcL* (RuBisCO large subunit gene) were found to be most common in GenBank (Benson et al., 2013) and BOLDSYSTEMS (Ratnasingham and Hebert, 2007). These were therefore chosen to be used to reconstruct the phylogeny. Out of the 1239 species in the species occurrence dataset, 1139 were found to already have available data on GenBank or BOLD for at least two of the markers. For the

remaining 99 species, we sought to generate the missing sequence data by obtaining representative specimens via institutional loan from the herbarium collections in Trondheim (TRH) and Oslo (O). Three species (*Trisetum subalpestre, Gymnocarpium continentale,* and *Scilla verna*) with no available molecular data and with no recent (collected after 1979) specimens available at TRH or O were excluded from the study.

Approximately 50 mg tissue was sampled from herbarium specimens into 2-mL microfuge tubes using flame-sterilized scissors and forceps. After placing a sterile steel bead in each tube, the tissue was disrupted using a Qiagen TissueLyser II for 1 min at 25 Hz, and this step was repeated after rotating the plates. Another bead was placed in tubes that still contained un-disrupted tissue, and these tubes were again agitated for 1 min at 25 Hz. We used the Qiagen DNeasy Plant Mini Kit to extract genomic DNA from tissue sampled from each herbarium specimen. The manufacturer's recommended protocol was performed except that 50-µL 10 or 20 mg/mL proteinase K was added after the lysis and RNA removal step, and the lysis solution was incubated for an additional 13-18 hours at 45ºC. Extract DNA concentration was measured by using the Qubit™ dsDNA HS assay kit and Qubit 2.0 Fluorometer. The polymerase chain reaction (PCR) was performed using universal primers for the three loci *ITS*, *matK* and *rbcL* (Table 1) using 50-µL volumes with the following reaction components: 1.0 µL template DNA (~0.01–157 ng/µL), 0.25 µL 1.25 units AmpliTaq Gold™ DNA Polymerase, 1.0 µL 0.4 mg/mL bovine serum albumin (BSA), 1.0 µL 0.2 µM each primer, 3.0 µL 1.5 mM $MgCl_2$, 0.5 µL 2.5 mM dNTPs, and 5.0 µL GeneAmp™ 1 x PCR buffer II.

Table 1 Oligonucleotide primers and associated PCR thermocycling protocol used to amplify and Sanger sequence the markers *ITS*, *matK* and *rbcL*.

Marker	Primer	Source	PCR conditions
ITS	ITS5a	(Stanford et al., 2000)	1 min of initial denaturation at 96°C; 45 cycles: 96°C 10 s, 48°C 30 s, 72°C 20 s + 4 s/cycle; final extension at 72°C 7 min.
	ITS4	(White et al., 1990)	
matK	matK-xf	(Ford et al., 2009)	5 min of initial denaturation at 95°C; 45 cycles: 95°C 30 s, 49°C 1 min, 72°C 1 min 40 s; final extension at 72°C 7 min.
	matK-MALP	(Dunning and Savolainen, 2010)	
rbcL	aF	(Pryer et al., 2001)	5 min of initial denaturation at 80°C; 45 cycles: 95°C 1 min, 60°C 1 min, 65°C 2 min; final extension 72°C 5 min.
	1379R	(Pryer et al., 2001)	

Performing gel electrophoresis on a 1.5% TAE gel stained with Invitrogen™ SYBR™ Safe, PCR products were assessed with a 100-bp ladder for a single, appropriately sized band of DNA. Some samples failed to amplify in the initial PCR step. For these the PCR product was used as the template in another round of PCR. Successful PCR products were Sanger sequenced by the commercial service Eurofins, and the pre-sequencing enzymatic purification of the samples was performed using illustra™ ExoProStar™ and following the manufacturer's protocol.

2.4 Sequence alignment and phylogeny

In the software Geneious 10.2.3 (Kearse et al., 2012), forward and reverse chromatograms were manually inspected, edited and trimmed to obtain high-quality sequences for the final alignment. Low-quality sequences were excluded from the alignments. The generated sequences were checked for potential contamination or misidentification by using the function BLAST (Basic Local Alignment Search Tool) in Geneious, and misidentified sequences were removed. If possible, sequences with few base pairs were concatenated with sequences from a different specimen. We used MAFFT version 7 (Katoh and Standley, 2013) to perform an automated alignment of the generated sequences with the downloaded sequences from GenBank and BOLD. The alignments were inspected and manually adjusted if needed. Single-

marker trees were produced in RAxML to identify potential issues with the sequences, and species that were not concordant with the work of the Angiosperm Phylogeny Group IV (APG IV) (Byng et al., 2016) were further inspected. The three loci alignments were concatenated into a multiple sequence alignment (MSA) with a total length of 7601 nucleotides. A partitioned maximum-likelihood analysis of the MSA was performed using RAxML (Stamatakis, 2006) under the GTRGAMMA nucleotide substitution model and three data partitions. The topology of the tree was verified by comparing it to the work of the APG IV: each species was checked to see if they were in the correct order, family and genus. Alignment corrections were performed if the species were wrongly placed according to APG IV. If a genus was found to be paraphyletic or polyphyletic, and corrections of the alignment did not help, then the genus was not inspected any further. The branches within a node were also examined if a node had low support (< 70%). If corrections of the alignment did not remove the low support, but the placement of the taxa was correct according to the APG IV system, the branch was not inspected any further. FigTree version 1.4.3 (Rambaut, 2012) was used to view the produced tree and to convert the tree into Nexus format for further analyses. The monophyletic pteridophyte clade was chosen as the outgroup of the phylogeny (Bateman et al., 2006). There were in total 111 families and 40 orders that the species were divided into (Supplementary Table B2).

2.5 Spatial phylogenetic analyses

The spatial data (CSV format) and the phylogeny (Nexus format) were imported into Biodiverse version 1.99_088 (Laffan et al., 2010) for combined analyses. The following calculations were generated in R using a Biodiverse pipeline (https://github.com/NunzioKnerr/biodiverse_pipeline): SR, PD, weighted endemism (WE), PE, relative phylogenetic diversity (RPD), and relative phylogenetic endemism (RPE). Calculations of SR was done by summing the number of species occurring in each cell. We calculated PD by summing the lengths of the branches occurring in each cell (Faith, 1992). For a cell, WE was calculated by weighting each species occurring in a cell by multiplying it by the proportion of its range which occurs in that cell (Crisp et al., 2001). All weighted species were then summed together to get the WE score of a cell. By doing so, the WE is not a metric of absolute endemism where a species is only occurring in one region (e.g. only in Norway) and nowhere else. Rather, it is a metric of relative endemism where it measures how rare or abundant a species is in a study region by giving each species a weight between 1 (i.e. occurring in only one cell in the study region) and 0 (i.e. occurring in all cells in the study region). PE was

calculated in the same way as WE only that it was the branches that were weighted (Rosauer et al., 2009). RPD and RPE were studied by identifying areas where there was a significant amount of long or short branch tips (Mishler et al., 2014). We found the RPD of an area by calculating the PD of the branches occurring there from the original tree divided by a comparison tree in which all the branches are of equal length but the topology of the tree is still the same (Mishler et al., 2014). The same method was done for RPE, but here PE was calculated instead.

Spatial randomisations of PD, PE, RPD and RPE were performed to assess the statistical significance of the different phylogenetic indices (Mishler et al., 2014). This was done to find areas with higher and lower values of the indices than expected by chance. In the process, a null model was created to compare with the observed result by randomly reassigning the species occurring within a grid cell without replacement. By doing this, the total number of grid cells and the number of species within each grid cell remains constant (Laffan and Crisp, 2003). The randomisations ran 999 times, and the different phylogenetic indices were calculated for each run. A two-tailed test was performed to identify cells with an observed value significantly higher or lower than the null model. Grid cells were classified as significantly high or significantly low if its observed value was higher than 97.5% or lower than 2.5% of the randomisations, respectively.

Different types of endemism were categorized by doing a Categorical Analysis of Neo- And Paleo-Endemism (CANAPE), which is based on the randomisation tests of RPE and PE (Mishler et al., 2014). For a grid cell to be endemic, it had to be significantly high (one-tailed test with $\alpha=0.05$) for one of these: PE of the original tree (PE_orig), PE of the comparison tree (PE_alt; see description above for RPD), or both. If PE_orig was significantly high in the grid cell and PE_alt not, then the cell was categorized as having paleo-endemism. If PE_alt was significantly high in the grid cell and PE_orig not, then the cell was categorized as having neo-endemism. If both PE_alt and PE_orig were significantly high, but RPE was not significant, then the cell was categorized as being a centre of mixed endemism. This means that the grid cell has a mix of short and long restricted branches, and neither paleo- nor neo-endemism was found to be dominating the cell. Centre of mixed endemism were further divided into two categories, where cells having a significantly high value of both PE_orig and PE_alt ($\alpha=0.01$) were categorized as having super-endemism.

Turnover among the presence-absence grid cells was investigated by clustering the cells using the Unweighted Pair Group Method with Arithmetic Mean (UPGMA) in Biodiverse. The

result is a map and a tree representing the similarities between the cells where similar cells are closer to each other in the tree and form clusters (clades). We did both a range-weighted metric turnover (RWTurnover) analysis and a range-weighted phylogenetic turnover (PhyloRWTurnover) analysis. The former analysis weights each species present in a grid cell with the fraction of its geographic range, and the latter weights the branch lengths instead of species. These two methods result in less obscure transition zones between geographic clusters as they give less weight to widespread species and branches (Laffan et al., 2016). The colourisation of the clusters was done manually, where each cluster was chosen subjectively. If the cells in a cluster were spatially clustered together (relatively close with each other on the map), the cluster they created was given a colour. A cluster was still coloured if the cells did not make a clear spatial pattern (cells were scattered across the map), but had more than 50 branches (cells) in the cluster.

2.6 Assembly of explanatory variables for diversity and endemism patterns

We used seven explanatory variables to analyse the diversity and endemism patterns (Table 2; Supplementary Figure A1). The three bioclimatic variables annual precipitation, mean temperature of warmest quarter and precipitation seasonality were chosen for this study as they have previously been found to explain 89% of the bioclimatic variation in Norway (Speed and Austrheim, 2017). Respectively, the three variables correlated with a precipitation axis, temperature axis, and a third axis with a lower correlation with precipitation seasonality. All variables were projected to WGS 84 / UTM zone 33n and then rasterized and resampled to 20-km grid cells. The bilinear method was used for the resampling for all variables except for habitat heterogeneity where the nearest neighbour method was used. Possible correlation between the variables was investigated (Supplementary Figure A2), and the maximum absolute correlations were observed between mean temperature of warmest quarter and soil pH, and between mean temperature of warmest quarter and topographic heterogeneity (0.48).

Table 2 Variables used for explaining the diversity and endemism patterns.

Explanatory variables	**Description**
Annual precipitation (Bio 12)	Sum of all monthly precipitation values in a year (O'Donnel and Ignizio, 2012). Original layer was derived from WorldClim through the package *dismo* (Hijmans et al., 2015) for the tiles 6 and 16, and had a 30 arc second resolution (~1x1 km).
Habitat heterogeneity	Sum of area types within a 20-km grid cell from AR50 (Source: NIBIO). Seven different area types are included: built-up area, agriculture, forest, bare land, peatland, glacier, and fresh water. Original layer was a polygon.
Mean temperature of warmest quarter (Bio 10)	The average temperature (ºC) for the three warmest months of the year (O'Donnel and Ignizio, 2012). See description of original layer under Annual precipitation.
Precipitation seasonality (Bio 15)	Measure of variation in the monthly precipitation values in a year expressed in percentage (O'Donnel and Ignizio, 2012). High values indicate greater variation in precipitation through the year. See description of original layer under Annual precipitation.
Soil pH	Soil pH index (H_2O solution) at 0.05m depth (Hengl et al., 2017). Original layer had a resolution of 250x250 m.
Time since last glaciation cover	The time in thousand years (kyr) since the area was covered by the Fennoscandian Ice Sheet (Stroeven et al., 2016). Original layer represented as isoclines, and was rasterized by using the function "Topo to Raster" in ArcGIS 10.4 (http://www.esri.com/software/arcgis).
Topographic heterogeneity	The difference in altitude between the highest and lowest point in a 20-grid km cell. Original layer was derived from the package *raster* (Hijmans, 2017) in R using the function *getData*, and had a resolution of 1x1 km.

2.7 Statistical analyses of diversity patterns

The four diversity indices SR, PD, significantly high PD (referred to as high PD henceforth), and significantly low PD (referred to as low PD henceforth) were statistically modelled to find which explanatory variables that could best predict the four diversity patterns. High and low PD were derived from the randomising of PD, and divided into two binomial datasets before the modelling. For high PD, cells were set to 1 when they had a value higher than 0.975 and cells with values below this were set to 0. The same was done for low PD only that cells with values smaller than 0.025 were set to 1. Possible correlation between the indices was investigated (Supplementary Figure A3).

We fitted all models as generalized linear models (GLMs). For SR and PD, gaussian distribution was chosen as the distribution model, and high and low PD as binomial. To account for the spatial autocorrelation between the data we used the Moran Eigenvector spatial filtering function from the package *spdep* (Bivand and Piras, 2015). The function removes the spatial autocorrelation in the residuals by selecting eigenvectors (Griffith, 2000) as explanatory variables in the model until the residuals fall below a specified tolerance level (Wang et al., 2013). Because of the high number of possible models, we used the function *dredge* from the package *MuMIn* (Barton, 2018) in R to calculate the Akaike Information Criteria (AIC) value of all possible models. Model-averaging was used to infer the relative importance of the explanatory variables for the three diversity patterns by giving the models with lowest AIC value more weight than models with higher AIC values. Coefficient estimates for the different explanatory variables were averaged and weighted by model importance across the models where the variable was present. This was done to find the relative effect of the variable on the diversity patterns.

3 RESULTS

3.1 Spatial and molecular data

The final spatial dataset contained a total of 1050 grid cells (Supplementary Figure A4). For the statistical analyses, 35 grid cells were removed due to missing values in the explanatory variables topographic heterogeneity, soil pH and the three bioclimatic variables.

The final molecular dataset was a three-locus, 7601-bp MSA containing sequences for 1239 species. The MSA included 1221, 1019, and 1048 sequences of *ITS*, *matK* and *rbcL*, respectively. In total, the MSA contained 31.2 % missing data. One species (*Chamaedaphne calyculata*) was trimmed from the topology in Biodiverse as it was not within the final spatial data set (see Methods – Species occurrence data). Therefore, the final phylogeny contained 1238 branch tips, each representing a single species (Figure 1). All families and orders were monophyletic in the phylogeny, and had in general very high support (Supplementary Figure B1). A few genera fell out as paraphyletic in the phylogeny (e.g. *Crepis, Campanula, Silene, Arenaria*, and *Galium*), and some branches were very long (e.g. *Cuscuta europaea*, *Viscum album, Cryptogramma crispa*, *Utricularia* sp., and Pteridophyta in general).

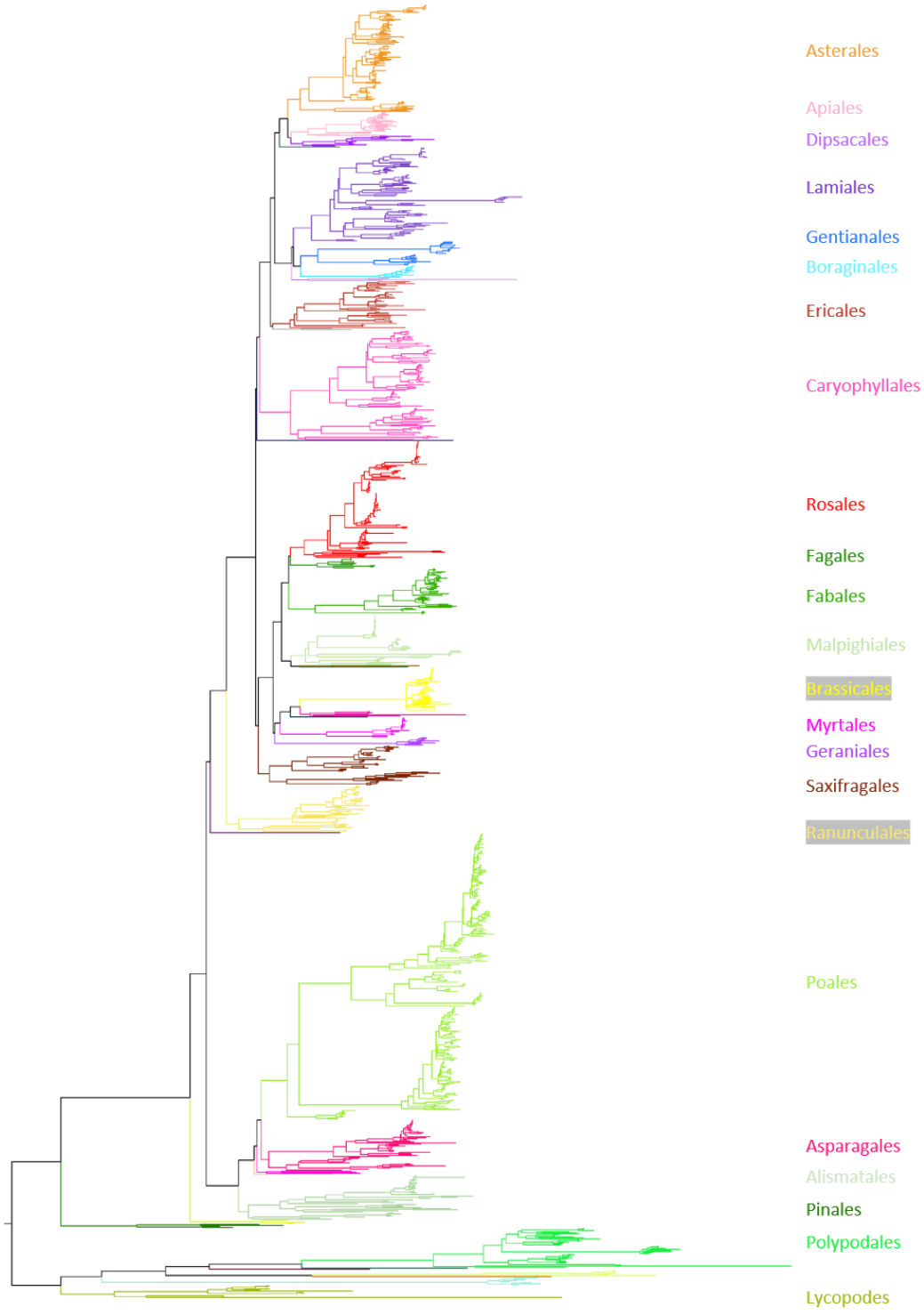

Figure 1 Topology of the inferred phylogeny. The phylogeny was created using a partitioned maximum-likelihood approach with a GTRGAMMA nucleotide substitution model on a three-locus (*ITS*, *matK*, *rbcL*) and 7601-bp MSA. Branches are coloured by the orders they belong to according to the APG IV system (Byng et al., 2016). Names to the right of the phylogeny represent the most species-rich orders with their corresponding colours. The tree is visualised using FigTree v1.4.3 (Rambaut, 2012). See Supplementary Figure B1 for full phylogeny.

3.2 Diversity patterns and drivers

The observed distribution patterns of SR and PD were very similar (Figure 2), and showed elevated diversity along the coastal areas of southern Norway. The highest number of species was found in the innermost parts of the Oslofjord where ~50-65% of all the species in Norway occurred. The highest values of PD were also found in the Oslofjord, and here ~71-84% of all the branches in the phylogeny were represented. In general, PD had higher values across Norway. The mean percentage of the total number of species and branches per cell were 24% and 46%, respectively. For both indices, less diversity was found in mountainous areas and northern Norway. For SR, 76 out of the 1050 cells had less than 10% of the total amount of species, and these were mainly located along the Norwegian border and in northern Norway. As expected, both SR and PD had high values in areas with higher number of occurrences (Supplementary Figure A5).

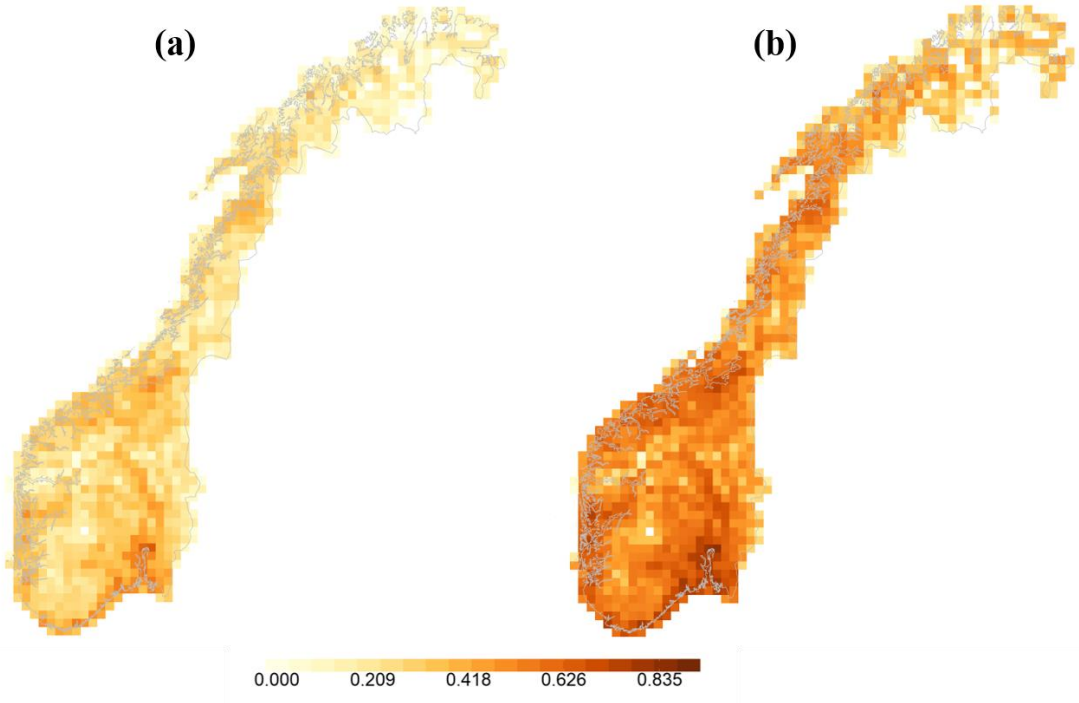

Figure 2 Observed diversity patterns of the 1238 vascular plant species in Norway. Both indices are visualised as the proportion of the total amount of species or branches. (a) SR as sum of species found in a cell. Around 50-65% of the total amount of species were found in the Oslofjord. (b) PD as the sum of branch lengths found in a cell. The lowest amount of species and branches was in general found around the Norwegian border and in northern Norway. Both maps are outlined with the Norwegian border. White cells scattered across Norway were not included in the analyses.

The spatial randomisations of PD and RPD (Figure 3) showed highly different patterns compared with the observed PD (Figure 2). For both randomisations (Figure 3), there was a distinct spatial distinction between significantly high (> 0.975) and low (< 0.025) PD of vascular plants in Norway. Regions of significantly low PD were found in the mountainous areas. Significantly high PD was found on the coastal regions in the south. RPD had a pattern that was in general similar with PD. However, RPD had more cells which were significantly high (154 cells) and less cells which were significantly low (146 cells) when compared with PD (80 and 286 cells, respectively).

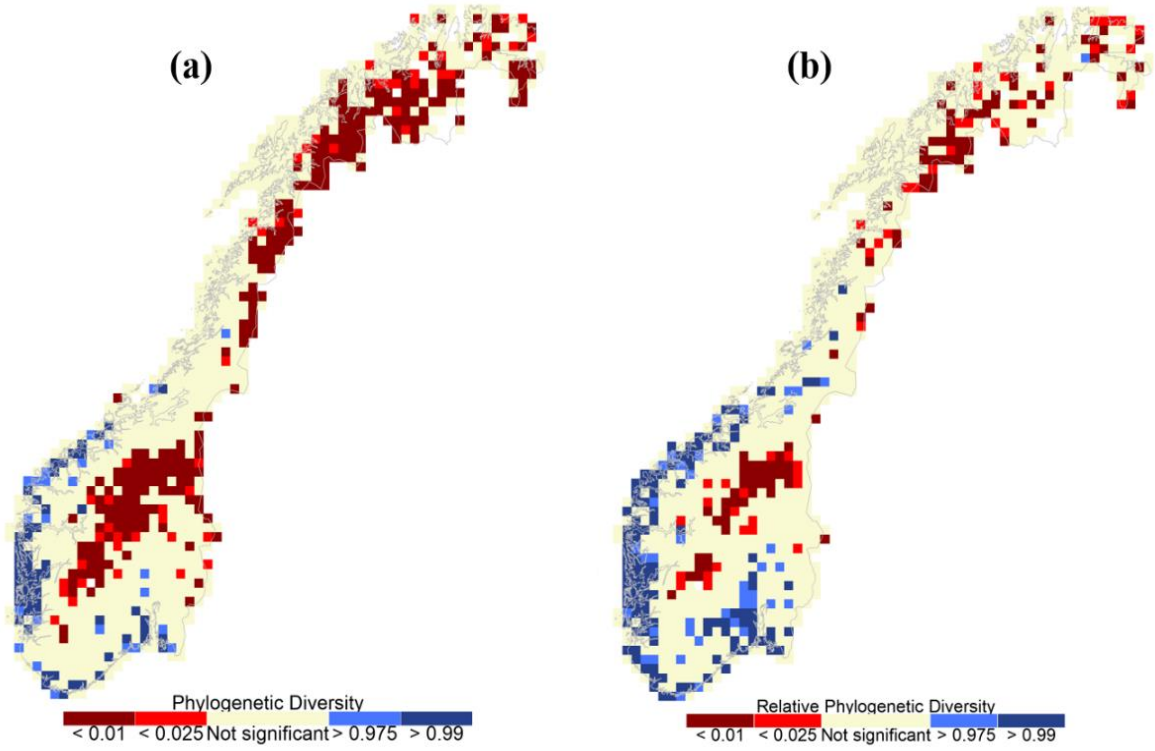

Figure 3 Results of spatial randomisations of PD and RPD. (a) Randomised PD. Blue cells show areas with significantly high PD (> 0.975) and red cells show significantly low PD (< 0.025). (b) Randomised RPD. Blue cells show significantly high concentrations of long branches and red cells show significantly high concentrations of short branches. Both maps are outlined with the Norwegian border. White cells scattered across Norway were not included in the analyses.

The relative importance of each of the seven potential drivers showed how many times a variable was included in each possible model (Figure 4). For all four diversity patterns (i.e., SR, PD, high PD, and low PD), mean temperature of warmest quarter and topographic heterogeneity had very high relative importance (100%; Figure 4). The other variables varied in their importance for each diversity pattern. Annual precipitation was important for all patterns (100%) except for PD (26%). The same was seen for time since last glaciation cover, where the variable was less important for PD (94%). For high PD, habitat heterogeneity and precipitation seasonality was less important (93% and 91%, respectively) compared to the other diversity patterns (~99-100%). For all diversity patterns, soil pH was the least important variable (~27-62%).

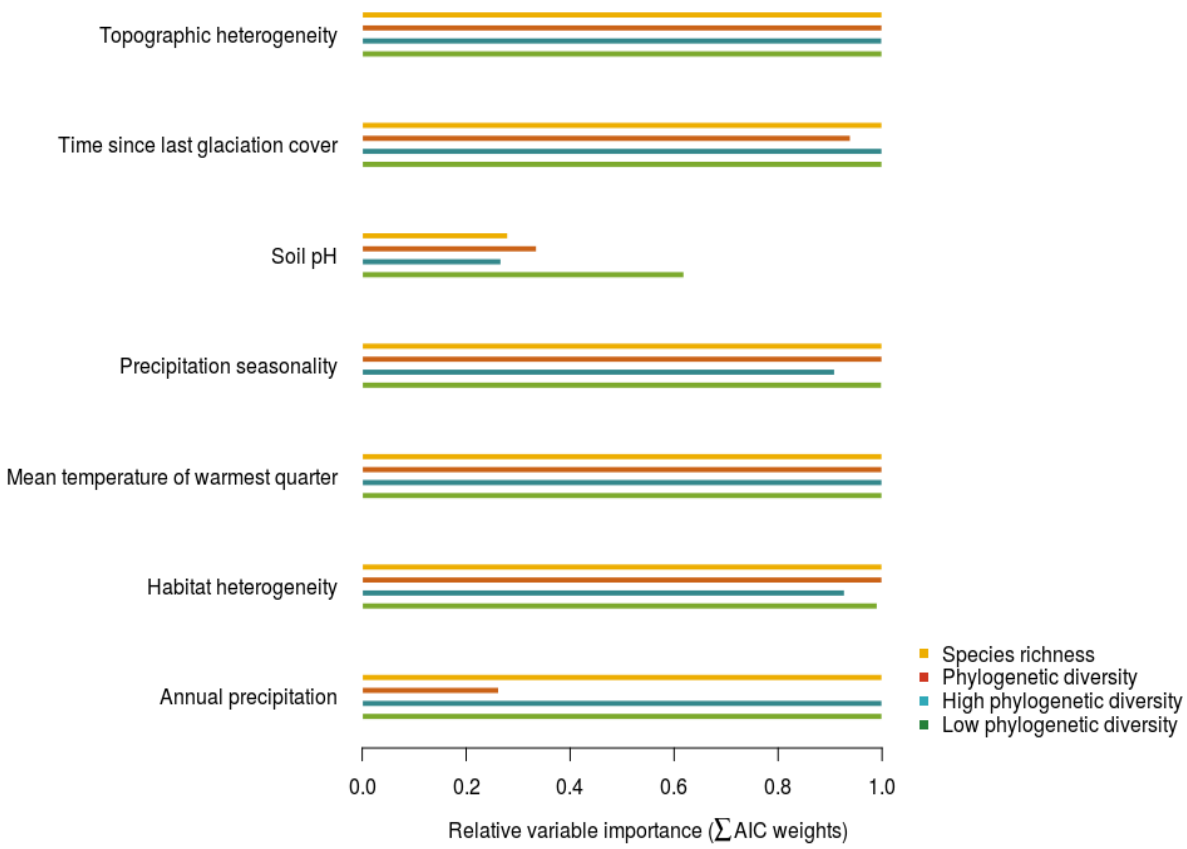

Figure 4 Importance of the seven explanatory variables measured as the AIC weight for the four diversity patterns: SR (species richness; yellow), PD (phylogenetic diversity; red), high PD (significantly high phylogenetic diversity; green), and low PD (significantly high phylogenetic diversity; blue). There were in total 128 possible models. Variables that were in models with low AIC were given more weight than models with high AIC.

The model-averaged coefficient estimates for the explanatory variables and their effects on the four diversity indices (Figure 5). The four models were divided into two plots because of their difference in parameter estimates due to different model families. For SR, PD, and high PD, topographic heterogeneity, time since last glaciation cover, and mean temperature of warmest quarter had a positive effect. Precipitation seasonality had a negative effect on SR and PD, but a positive effect on high PD. Annual precipitation had a negative effect on SR, no effect on PD, and a positive effect on high PD. For low PD, habitat heterogeneity had a positive effect. All other variables except for soil pH had a negative effect on low PD. Soil pH had no significant effect on any of the patterns.

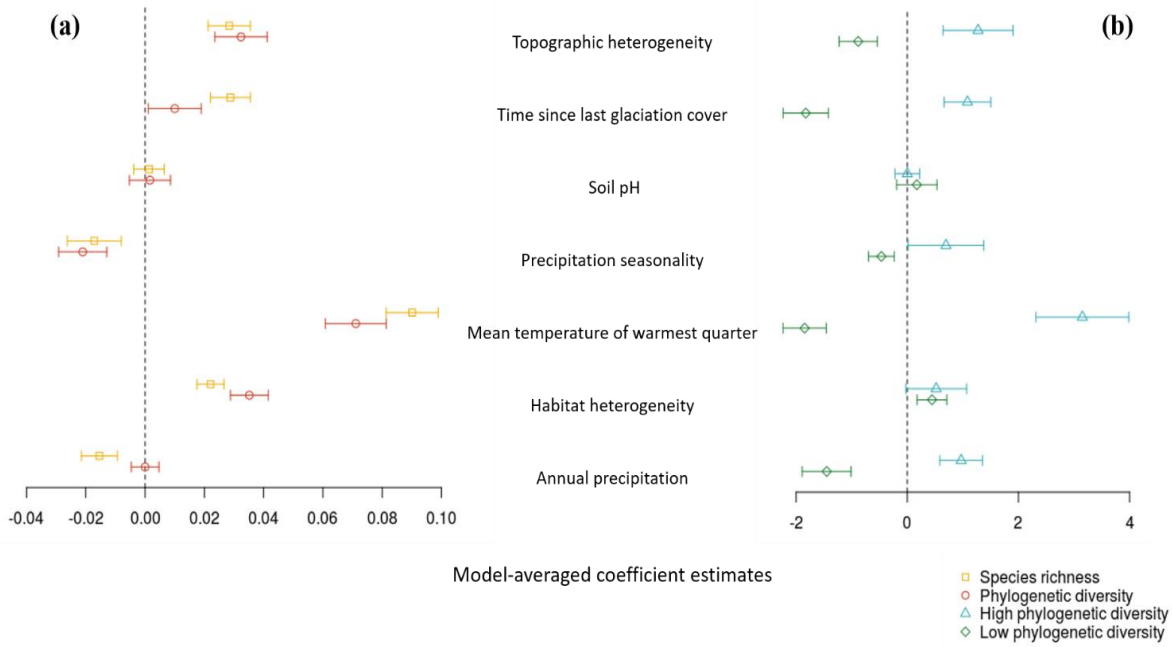

Figure 5 Model-averaged coefficient estimates with confidence interval bars for the four diversity indices SR (species richness; yellow square), PD (phylogenetic diversity; red circle), high PD (high phylogenetic diversity; blue triangle), and low PD (low phylogenetic diversity; green diamond). All explanatory variables are standardised so that each parameter estimate is comparable with the others, and their effects are interpreted relatively to the other effects. The effects are divided into two plots: coefficients in (a) are from a Gaussian model, and the coefficients in (b) are from a binomial model. Points that are overlapping with the stippled line had a non-significant effect.

3.3 Endemism patterns and potential drivers

Observed patterns of WE and PE were in general very similar, and both indices show low values of observed endemism across Norway (Figure 6). More cells with high values of endemism were seen for WE (Figure 6a) than PE (Figure 6b). Some cells in mid-southern Norway were also showing an WE pattern (Figure 6a) that was not apparent in PE (Figure 6b). Just as with SR and PD, WE and PE show expected patterns with higher values biased towards well-sampled areas in southern Norway (Supplementary Figure A4).

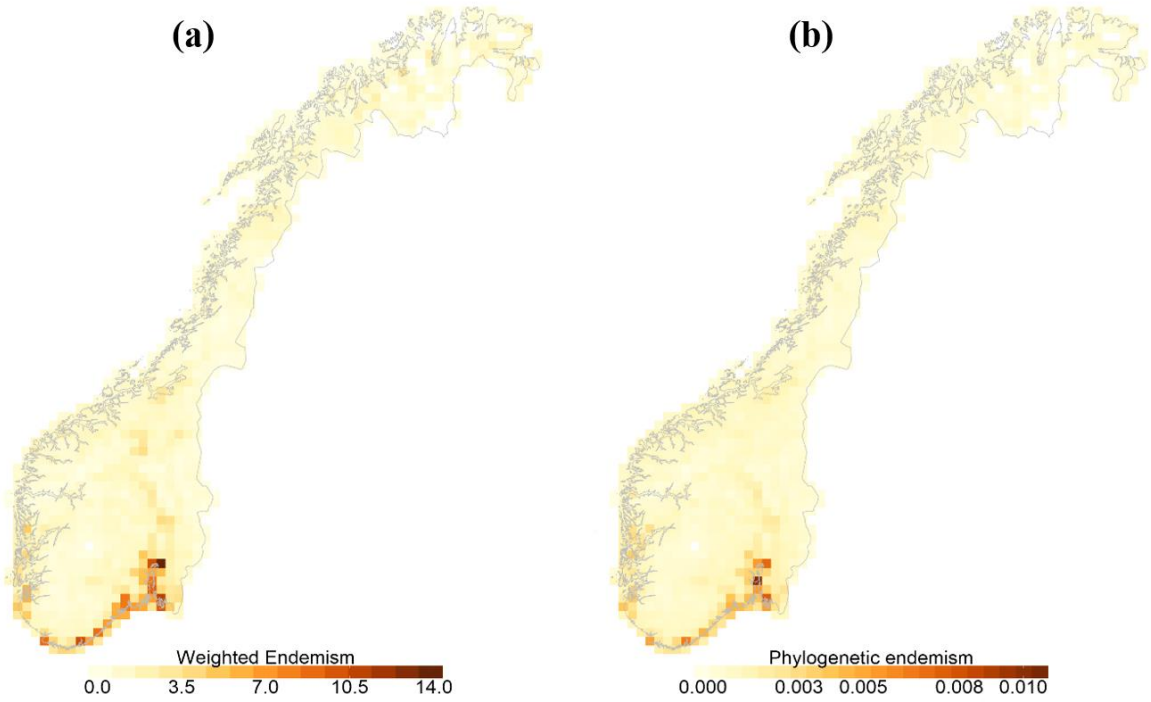

Figure 6 Observed patterns of endemism in Norway. (a) Weighted endemism (WE) as number of species present in a cell multiplied by each species' restriction proportion. (b) Phylogenetic endemism (PE) as sum of branch lengths present in a cell multiplied by each branch's restriction proportion. Both maps are outlined with the Norwegian border. White cells scattered across Norway were not included in the analyses.

The results from the CANAPE show patterns of the different endemism categories scattered across Norway (Figure 7). The plot was made by categorizing spatial randomised PE and RPE (Supplementary Figure A6). A concentration of neo-endemism (i.e. short restricted branches) was found in the northernmost parts of Norway and more sparsely further south in the mountainous areas (109 cells). Paleo-endemism had few cells (16 blue cells) which were found mainly on the south-west coast with a few other cells scattered across the region. Centres of endemism (118 purple cells) were found mainly on the coast of southern Norway, but also scattered across middle and northern Norway.

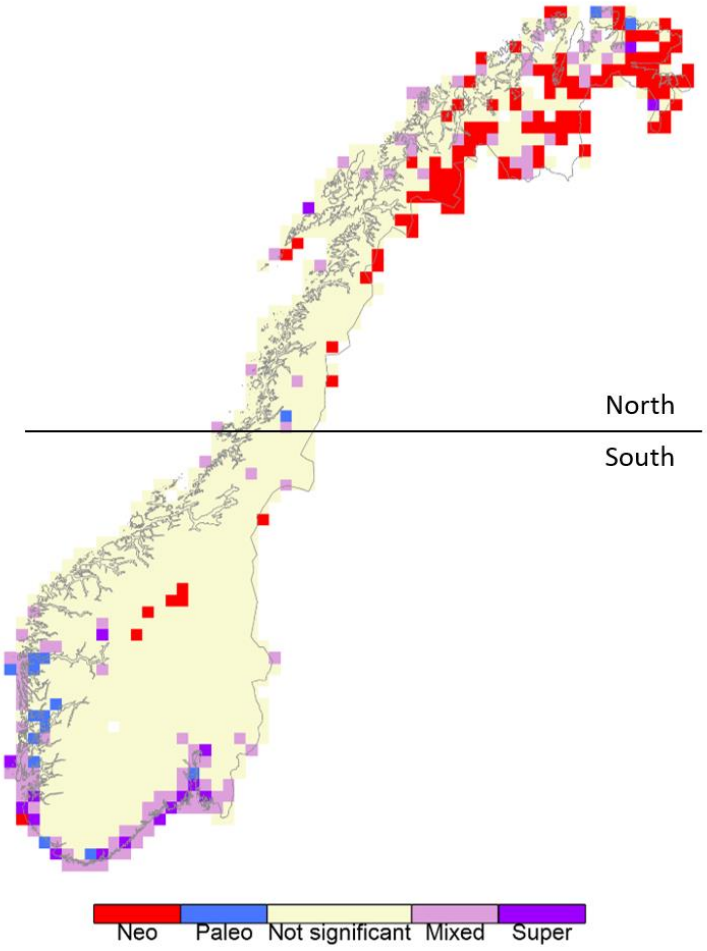

Figure 7 Patterns of PE and RPE from the CANAPE. Red cells show concentrations of significantly short restricted branches (neo-endemism) and blue cells show concentrations of significantly long restricted branches (paleo-endemism). The two shades of purple show areas with concentrations of both short and long branches. Beige cells are not significant. The black line represents the middle of the northernmost and southernmost extremes of Norway. White cells scattered across Norway were not included in the analyses.

Potential drivers of PE (Figure 8) show how the different endemism categories (Figure 7) were distributed on the different variable scales. For annual precipitation (Figure 8a), neo-endemism was highly clustered below ~600 mm (75% of the neo-endemic cells) and paleo-endemism was mainly located in areas with annual precipitation above ~1500 mm (73.3% of the paleo-endemic cells). Mixed endemism was mostly scattered across the whole scale. Habitat heterogeneity (Figure 8b) had the most amount of endemism in cells with seven habitat types, where also paleo-endemism was mainly found (86.6% of the paleo-endemic cells). Mixed endemism was found in cells with almost all number of vegetation types, but was most common in cells with seven vegetation types. Neo-endemic cells were found in almost all number of vegetation types except in cells with only two types. For mean temperature of warmest quarter (Figure 8c), there was a clear clustering of neo-endemic cells below 11ºC (100 % of neo-endemic cells). We also saw a distinct disjunction between mixed endemism in the north and in the south (Figure 7), where northern mixed endemism was found below 11ºC (97% northern mixed endemism cells) and southern mix endemism closed was mainly found above 12.5ºC (84.5% southern mixed endemism cells). The same pattern was seen for paleo-endemism (84.5% of paleo-endemic cells above 11ºC). Precipitation seasonality (Figure 8d) did not show a clear pattern for neo-endemism as its cells were scattered across the scale. However, both cells with paleo-endemism and mixed endemism were mainly clustered between 26% and 34% in precipitation variability. Soil pH (Figure 8e) did not show a distinct pattern for any of the endemism categories. However, almost all paleo-endemic cells (73.3% of paleo-endemic cells) were in grid cells with a soil pH below 5. Time since last glaciation cover (Figure 8f) did also not show a clear pattern for any of the endemism categories. Paleo-endemic cells were only found in cells that were older than 11 thousand years. For topographic heterogeneity (Figure 8g), almost all mixed endemism cells were clustered below 100 m in altitude variation. Paleo-endemic cells were only found in cells with a variation between ~ 60 m - 285 m. Cells with neo-endemism were found between ~ 10 - 322 m in altitude variation.

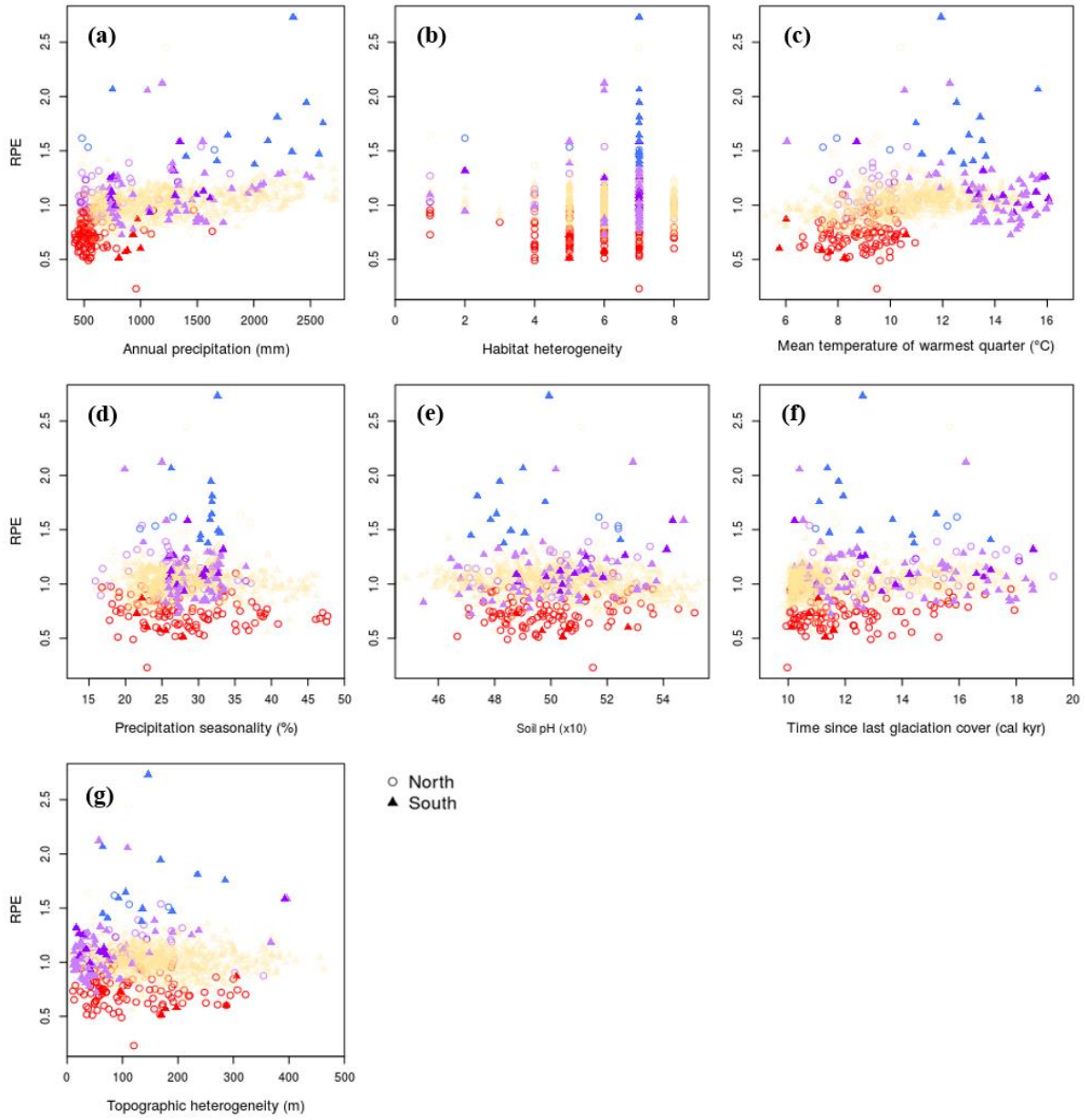

Figure 8 Potential drivers of CANAPE of Norwegian vascular plants. The variables were plotted against RPE, and the cells are coloured according to their endemism category, i.e. neo-endemism (red), paleo-endemism (blue), mixed endemism (light purple), and super endemism (dark purple). Non-significant cells are coloured in beige. The cells were also divided into two groups based on their location in Norway: north (open circle) and south (solid triangle) of the black line (Figure 7) representing the middle of the northernmost and southernmost extremes of Norway. (a) Annual precipitation. (b) Habitat heterogeneity. (c) Mean temperature of warmest quarter. (d) Precipitation seasonality. (e) Soil pH. (f) Time since last glaciation cover. (g) Topographic heterogeneity.

3.4 Turnover patterns

The RWTurnover and PhyloRWTurnover analyses recovered multiple geographically distinct clusters representing the native vascular plants in Norway. The RWTurnover analysis showed at least six major clusters (Figure 9). A total of 57% of the grid cells were either in cluster 1 or cluster 2, and these two clusters were more similar to each other than with any other cluster. Both clusters were further divided into four sub clusters (a-d). For cluster 1, the cells of the sub clusters were grouped by their closeness to the coast, where 1a and b were found inland, and 1c and d were found on the western coast. For cluster 2, the cells were divided similarly as cluster 1. Cluster 2a and b were found inland, while cluster 2c and d were found on the coast of mid and northern Norway, and in the fjords in the southwest, respectively. Cluster 3 was the cluster that was most similar to 1 and 2, and the cells here were found in the mountainous inland of southern Norway and in the mountains of northern Norway. This cluster had three sub clusters, which showed a latitudinal pattern with 3b in the south, and 3a and 3c in the north, respectively. There was a clear boundary between cluster 2a and 3a in the middle of Norway. Cluster 4 contained cells on the coast of middle and northern Norway, and had also two sub clusters with a latitudinal pattern. Cluster 5 was located along the coast and inland of southeast Norway (5a and b), and a few cells are also located on the southwest coast (5c). Cluster 6 was mainly located in the most northern part of the region, but one cell was found inland in the southern part.

The PhyloRWTurnover analysis had also six major clusters (Figure 10). A total of 68% of the grid cells were either in cluster 1 or 2, and these two were also more similar with each other than with any other cluster. Just as with the RWTurnover analysis (Figure 9), cluster 1 in the PhyloRWTurnover analysis had several minor clusters within it (a-d). Cluster 1 in the PhyloRWTurnover analysis were relatively similar to cluster 1 in the RWTurnover analysis in the southern part of the region (Figure 9). However, in the PhyloRWTurnover analysis (Figure 10), the cells in cluster 2c in the RWTurnover analysis were more similar with the other cells in cluster 1 than with cluster 2. Cluster 2 in the PhyloRWTurnover analysis (Figure 10) was also different from the patterns seen in the RWTurnover analysis (Figure 9) in that the cells that were in two other distinct major clusters (3, 6) were clustered with cells in cluster 2. Cluster 3 was located in the southeast inland and expanded inwards of Gudbrandsdalen. Cluster 4 was mainly located in the north of the region and did not show a clear spatial clustering as the previous clusters. The same can be said for cluster 6. Cluster 5 was located at the coast of southern Norway, and could be divided into two sub clusters where 5a was located to the west and 5b to the east.

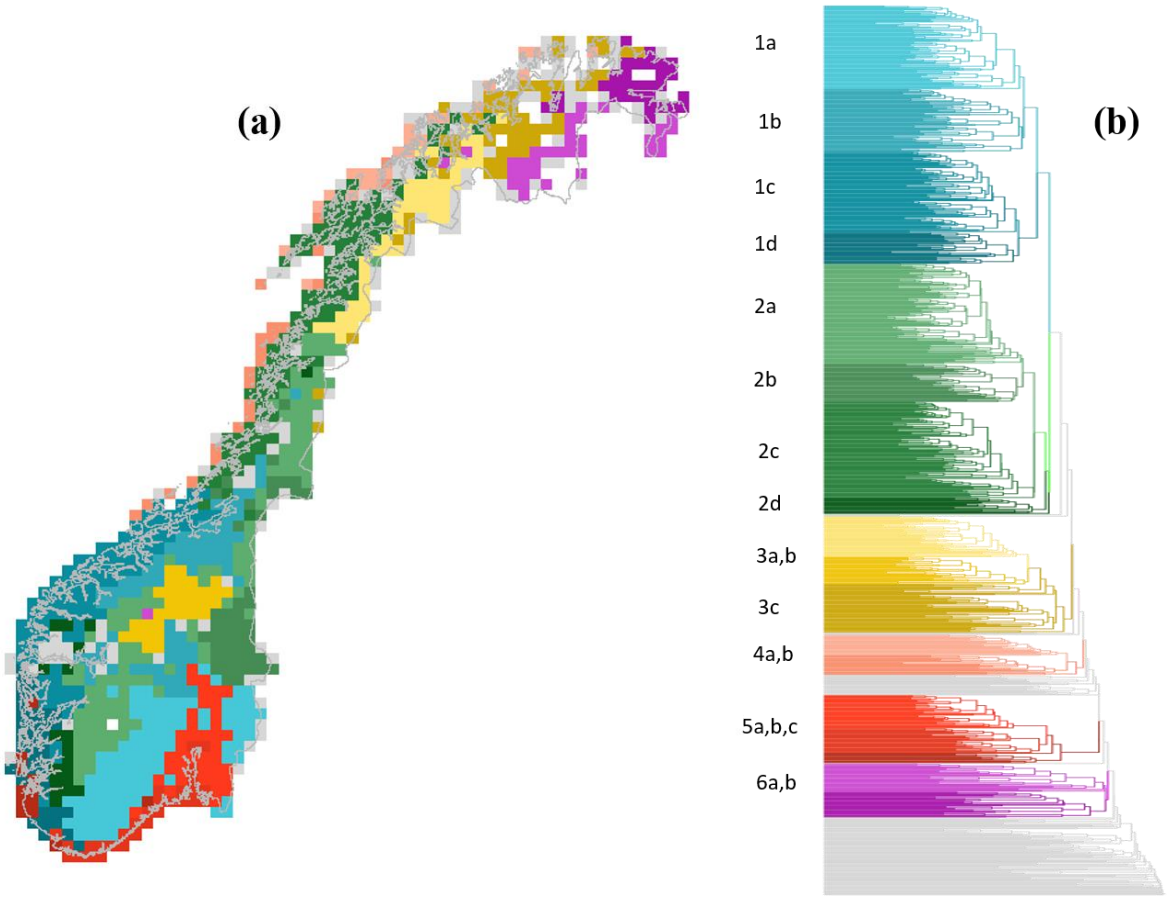

Figure 9 Range-weighted turnover (RWTurnover) analysis of Norwegian vascular plants. The analysis compares all grid cells, and cluster the cells together by looking at the species they share, and at the same time gives less weight to wide-ranged species. (a) Map of Norway that indicates the geographic position of the raster cell clusters in the turnover topology (b). (b) The turnover topology that shows how similar the grid cells are; cells that are closely clustered and have the same colours are more similar. White cells scattered across Norway were not included in the analyses.

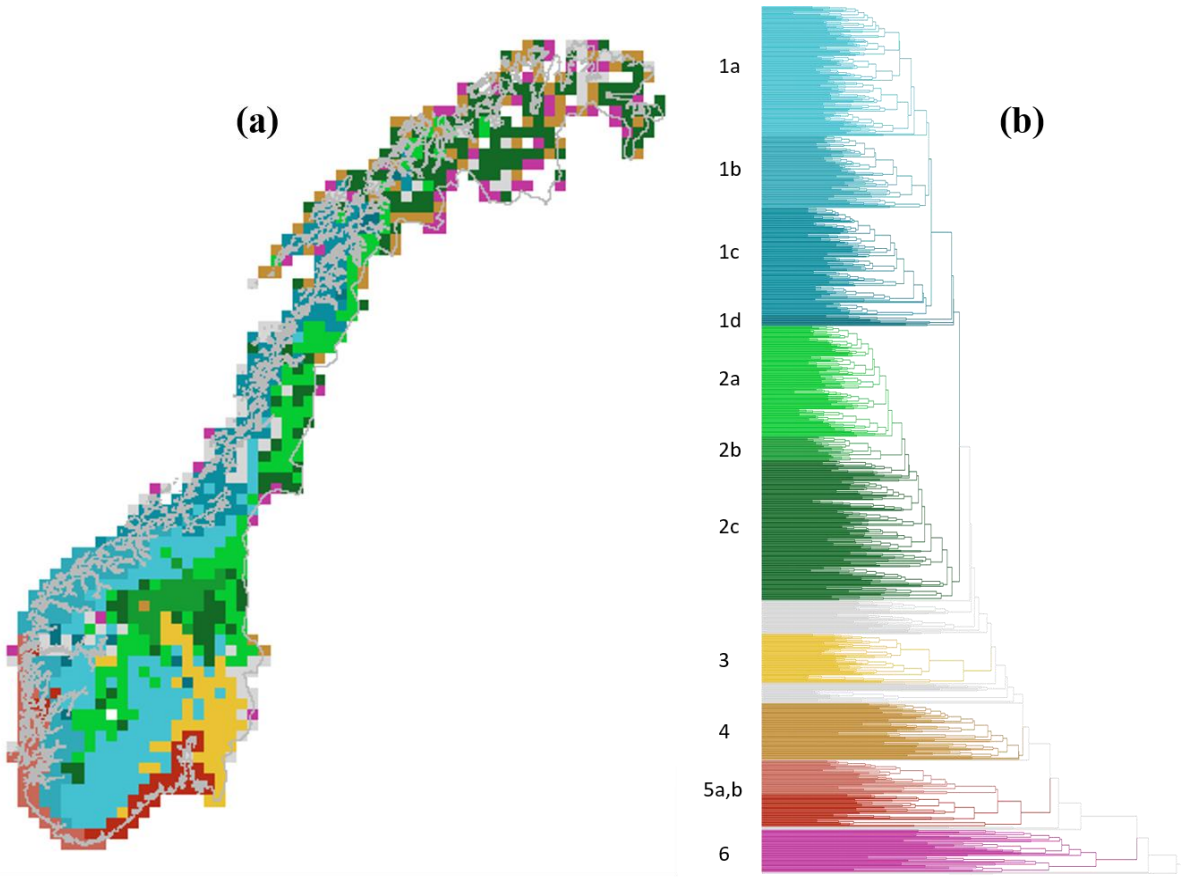

Figure 10 Range-weighted phylogenetic turnover (PhyloRWTurnover) analysis of Norwegian vascular plants. The analysis compares all grid cells, and cluster the cells together by looking at the phylogenetic branches they share, and at the same time gives less weight to wide-ranged species. (a) Map of Norway that indicates the geographic position of the raster cell clusters in the turnover topology (b). (b) The turnover topology that shows how similar the grid cells are; cells that are closely clustered and have the same colours are more similar. White cells scattered across Norway were not included in the analyses.

4 DISCUSSION

The approaches of spatial phylogenetics show the importance of including evolutionary relationships between taxa. By combining a molecular phylogeny of vascular plants with their occurrences in Norway, this study is the first to map the PD and PE of Norway's native vascular plants. Our results advance our understanding of the diversity and endemism of northern European vascular plants, and which environmental and historical factors that drive these patterns.

4.1 Phylogenetic analyses

The topology of the phylogeny did in general coincide with the APG IV system. Paraphyletic genera were however common. As the three-loci dataset was not complete for all species, and because the three genetic markers are differently conserved, this could have affected the placement of species within the correct genus. Another possible explanation is that one or more of the sequences used for the species were misidentified (i.e. actually came from a different species than the one in question). During the creation of the MSA, at least 60 sequences were found to be contaminated, and there is a possibility that undiscovered contamination is still present. Taxonomic misidentification in public sequence repositories is a well-known problem that is not easily remedied (Vilgalys, 2003). One solution for this could have been to use a BLAST search on all the sequences to determine if the sequence was of correct genus. This is not a solution, however, if the contamination is from a close relative and not a distant one.

The phylogeny also consisted of many long branches. An explanation for this is that the long-branched taxa do not have any close relatives which the taxa can be compared with, e.g. *Ceratophyllum demersum*, *Parnassia palustris* and *Oxalis acetosella*, which are the only species in their orders. Species or taxa could also have experienced rapid evolution and therefore real variants in the sequence alignments. This has been shown for parasitic species like *Cuscuta europea* and *Viscum album* (Zuber and Widmer, 2000, Barkman et al., 2007) or the carnivorous plants in the *Utricularia* genus (Albert et al., 1992). The long branches seen for Pteridophyta were also found for vascular plants in Chile (Scherson et al., 2017) and California (Thornhill et al., 2017). This is probably because Pteridophyta is distantly related to the rest of the phylogeny. The overall effect these long branches have on the results is uncertain. A possible way of test this pattern further is to remove long-branched taxa and do the analyses

without them to see if the results change. This was done by Scherson et al. (2017) for the vascular plants in Chile, and they found that doing the analyses on seed plants alone made the paleo-endemic cells they found doing the analyses on all vascular plants, disappear.

4.2 Spatial patterns and drivers of diversity

The observed PD had a pattern that was almost identical with SR (Figure 2), which was expected. In general, PD had higher values than SR across Norway. This is because PD and SR are not linearly correlated (Supplementary Figure A3). Areas of higher diversity coincide with high numbers of occurrences (Supplementary Figure A4). However, some areas did have high diversity (both SR and PD) although the occurrence numbers were relatively low. This was especially apparent in the western and northern part of Norway where there in general were less than 5000 occurrences per 20x20km grid cell. This is most likely because the number of species does not increase linearly with sampling intensity, but increases rapidly with low sampling intensity and then flatten out where few or no more species are found with increasing sampling intensity (Gotelli and Colwell, 2001). The same can be said for the well-sampled area in southeast Norway (Supplementary Figure A4), which does not show any clear pattern in either of the diversity maps. Even though the diversity patterns do not have an overall correlation with number of occurrences, there is still a sampling bias that can lead to incorrect interpretations (Brose et al., 2003, Beck et al., 2014). A recent study on observation and specimen based occurrence data in Norway found indeed that both occurrence types were biased towards areas with higher population density than expected in Norway (Speed et al., 2018). They also found species occurrences to be overrepresented in the warmer regions of Norway, which thus can lead to wrong interpretations of a species' ecology and distribution.

The species that were examined in this study were mainly extracted from a list derived from Artsdatabanken (Supplementary table B1) concerning vascular plant species in Norway. However, as not all species in Norway have been evaluated in this list and because Artsdatabanken considers a species as native as long as it appeared in Norway before the year 1800, information about species were also extracted from Lid and Lid (2005). This is because we wanted to see if the species' distribution in Norway were affected by their evolutionary history and not by human activity. However, some species present in Norway are not native all over the region, e.g. *Picea abies* which has been planted outside its natural distributions (Lid and Lid, 2005). Another species, *Sorbus intermedia*, is probably native in the south-eastern part of Norway (Lid and Lid 2005), but is on the national black list of species (Gederaas et al., 2012). The distributions of each species were not corrected for as done in other studies (Thornhill et

al., 2017). As both CANAPE and the two turnover analyses weights the species/branches by their ranges, a potential phylogenetic signal and floristic pattern could possibly have been removed by not doing the correction.

Earlier studies concerning spatial phylogenetics used only occurrence data from preserved specimens (Mishler et al., 2014, Thornhill et al., 2017, Scherson et al., 2017), and this study was the first to include human observations in the spatial data. This was done for two reasons: it was seen that rare species like *Nigritella nigra* had occurrences other places than where the preserved specimens had been sampled. By not including these data, some species could potentially have a very much smaller distribution range than expected by including all occurrences. Occurrences that are categorized as human observations do not distinguish between those of professional botanists and amateurs. The Norwegian Botanical Society contributed 32% of all vascular plant occurrence records in Norway, and data collected by these botanists would be lost if only preserved specimens were used. Still, a total of 406 990 records were removed because of their high uncertainty in meters, and this removal of occurrences was especially clear for areas that had less sampling intensity in general. Another potential error in this study is the resampling of the occurrence grid cells to get the correct map projection. This may have changed the initial number of occurrences per grid cell, and therefore also affected the results. This is apparent on the borders of northern Norway (Supplementary Figure A3) where some cells are located outside of Norway. This could have been avoided by deciding on a map projection from the beginning of so that resampling of occurrence cells would be unnecessary.

Patterns of the randomised PD and RPD (Figure 3) were very different from the observed PD (Figure 2). Significantly high values of PD (Figure 3a) were located mainly on the south coast of Norway. These areas may indicate phylogenetic overdispersion (Webb et al., 2002), where the species occurring in the area are distant from each other in the phylogeny. This can be explained ecologically, where branches close to each other in the phylogeny have niche similarities that lead to competitive exclusion (Darwin, 1859, Violle et al., 2011). Phylogenetic overdispersion is also found in abiotically low-stress environments (Butterfield et al., 2013). The cells with significantly low PD show a possible phylogenetic clustering of taxa (Webb et al., 2002). This can indicate that the species occurring there are close relatives that are adapted for specific conditions (i.e. presence of niche conservatism). The species also occurring there can be closely related as there is less competition in more physically stressful habitats (Callaway et al., 2002, Bruun et al., 2006).

RPD resembled randomised PD in that significantly high values were located on the coast of southern Norway and significantly low values in the mountainous areas. However, the mountainous areas had in general fewer cells with significantly low RPD than low PD, indicating that the branches there are not shorter than expected by chance. An interpretation of this is that the observed significantly low PD is a result of species being close to each other on the tree, and not a result of short branches. Significantly low RPD may indicate recent radiations in the region, but can also be a result of niche conservatism just as with PD. Significantly high RPD however had more cells than significantly high PD. Thus, these areas have an over-representation of long branches, which may indicate either a presence of relictual species or that the species there have no close relatives. Cells that have both significantly high PD and RPD may indicate that it is the long branches that lead to the cells being significantly high in PD. As previously discussed, long branches from e.g. Pteridophyta can lead to biased patterns, and the analyses could be redone without these branches to see if the patterns are affected.

When it came to the drivers of diversity patterns, all were in general important in explaining the diversity patterns (Figure 4). However, soil pH had a low relative importance compared with the other drivers (Figure 4) and no significant effect on any of the patterns (Figure 5). This is not surprising as soil drives diversity on a more local level than on a regional one (Moen et al., 1999), and was therefore insignificant because of the low grid cell resolution.

Mean temperature of warmest quarter had the largest effect on all diversity patterns. This is not surprising, as temperature in general has been found to drive plant diversity (Rohde, 1992, Allen et al., 2002). A limitation for vascular plants in Norway is not necessarily that the temperature becomes too high, which may be an issue in areas with low water availability, but rather the opposite. As low PD is found in areas with lower temperatures (Figure 5), this can be interpreted as that the clades present in those areas are adapted towards cold tolerance (Donoghue, 2008). For SR and PD, it is however important to remember the potential sampling bias towards warmer regions (Speed et al., 2018), but since the effect of temperature on these two diversity patterns was found to be relatively high, it is still possible to say that this is factual.

Both precipitation variables (annual precipitation and precipitation seasonality) had surprising results when it came to their model-averaged effects (Figure 5). Precipitation seasonality had a negative effect on SR, PD, and low PD, but a positive effect on high PD. However, the magnitude of the effect was small compared to the other variables. Almost the same pattern was seen for annual precipitation. This variable had negative effects on both SR

and low PD, a positive effect on high PD, and no significant effect on PD. Annual precipitation had a relatively bigger effect on high and low PD compared with precipitation seasonality, and a lower effect on SR. An interpretation of these two variables and their effects is that there are more species in areas that have less precipitation and yearly variation in precipitation, but these species are also closely related to each other, explaining why annual precipitation is insignificant for PD. Also, as previously stated about Pteridophyta having very long branches in the phylogeny (Figure 1), that possibly this group is contributing to annual precipitation being significant for high PD. This was seen for ferns in Australia (Nagalingum et al., 2015). Previous studies have found precipitation to be a strong predictor of plant SR (Kreft and Jetz, 2007), which was not seen here. A possible explanation for this is that precipitation possibly has a unimodal effect as seen in other studies on plants in Norway (Speed and Austrheim, 2017) and not a linear one as modelled here.

The spatial heterogeneity variables (habitat and topography) were both highly important for all four diversity patterns. Habitat heterogeneity had a positive effect on SR and PD, which was expected, because different habitats provide ecological space for more species with different specialisations and therefore an increase in species and PD. However, this variable had also a significant effect on low PD and an insignificant effect on high PD, suggesting that there also is a phylogenetic clustering with increasing habitat numbers. Topographic heterogeneity and mean temperature in warmest quarter were the most important variables in general (Figure 4). The effects of topographic heterogeneity on the different diversity patterns were as expected: positive effects on SR, PD and high PD, and negative effects on low PD (Figure 5). Thus, more variation in altitude gives room for more species and with it more PD, and the species occurring in these areas could possibly also be less related to each other.

Time since last glaciation cover had a positive effect on increased diversity (SR, PD, and high PD) and a negative effect on low PD. An interpretation of this is that there is more diversity in older areas and lower diversity in areas with a shorter time since it was glaciated. A possible explanation for this pattern is succession, since older areas have had longer time to recruit new species than young areas (Erschbamer et al., 2008).

4.3 Spatial patterns and potential drivers of endemism

The observed patterns of endemism in Norway (Figure 6) were highly correlated (Supplementary Figure A2), which was expected. Both endemism patterns had also a high correlation (>0.6) with SR and PD, which was also as expected (Crisp et al., 2001). Some cells

of WE had higher values than PE (e.g. around Dovre area in the mountains of southern Norway and the southern coast). This may indicate that these areas have very restricted species, but that the total PD (i.e. the branch sums) is small, suggesting range-restricted but closely-related species. Some cells for PE had higher values than WE. An interpretation of this is that the species in those cells are range-restricted and phylogenetically distant. The general pattern for both endemism maps is that the highest levels of endemism are found along the coast of southern Norway and in the Oslofjord, and that most of Norway has close to zero endemism.

For the CANAPE there was a very different pattern. Paleo-endemic cells (i.e. long, restricted branches) were located mainly on the coast of southern Norway. Previous studies have found that paleo-endemics seem to be found in well-watered environments with less extreme temperatures (Jordan et al., 2016). This was also seen here, where almost all paleo-endemic cells were located in areas with an annual precipitation between ~1400 – 2600 mm and with an annual variation in precipitation around 33 % (Figure 8a and d). These areas also had a mean temperature of warmest quarter between ~11ºC-14ºC (Figure 8c). Paleo-endemic cells were also found to be in areas with heterogenous habitats, which may suggest differences in niche preferences.

It is important to keep in mind that although these areas are classified as areas with paleo-endemism, it does not necessarily mean that the species present in that area have originated there (Grandcolas et al., 2014). This category rather means that the species have no close related species (i.e. long branch) and that it also is restricted, meaning that it could possibly have been the only species within its genus or family that has dispersed to the area. However, previous discussions on potential *nunataks* in Norway have stated that there could possibly have been areas in the southwestern part of Norway that were ice-free during the last glacial maximum (Brook et al., 1996). Also, earlier studies on spatial phylogenetics have stated that restricted, long branches may reflect species that have been isolated in an area for a long time and where close relatives have gone extinct (Mishler et al., 2014). However, there were not found any paleo-endemic cells in other proposed areas with *nunataks*, i.e. Andøya (Vorren et al., 1988, Alm, 1993, Parducci et al., 2012) and other nearby mountain tops on the coast of northern Norway (Vorren et al., 1988). As for now, climatic drivers seem more probable in explaining patterns of paleo-endemism in Norway.

Neo-endemic cells (i.e. short, restricted branches) were mainly located in the northernmost areas of Norway. Recently evolved polyploid species, which have been found to

increase with latitude and altitude (Grytnes, 2003, Mable, 2004), are a possible cause of this pattern. For neo-endemic cells, annual precipitation and mean temperature of warmest quarter seem to be clear potential drivers for this endemism category (Figure 8a and c). As previously discussed, low temperatures and precipitation often create stressful environments for plants. However, few studies have been done on how stress can lead to speciation in plants (Bijlsma and Loeschcke, 2005, Lexer and Fay, 2005), but this possibility is still underlined as being an important factor and could possibly explain the neo-endemic pattern.

To enable distinguishing between eastern or western species, the CANAPE cells were categorized into north and south based on if they were located north or south of the middle of Norway (Figure 7). Centres of endemism (i.e. areas with a mix of long and short restricted branches) were mainly found two places: the coast of southern Norway, and scattered in the northernmost part of Norway. As the grid cells were categorized into either being in the north (open circle) or the south (closed triangle) (Figure 7), we see a clear difference in the mean temperature of warmest quarter (Figure 8c). Almost all northern mixed endemism cells were found below 11ºC, and the southern cells were found above 12.5ºC, suggesting differences in immigration routes and floristic elements (Moen et al., 1999). The northern mixed endemism cells may reflect the eastern species, which are limited by mild winters (Moen et al., 1999). Since the temperature variable used in this study was highly correlated with other temperature variables, among them winter temperature (Speed and Austrheim, 2017), we can argue that there would be a positive correlation between cold winter temperatures and northern mixed endemism cells. For the southern mixed endemism, most likely western and southern species are represented in these grid cells as these species are limited by both cold summers and winters (Moen et al., 1999). Another interesting pattern is the clustering of mixed endemism cells (both northern and southern) below 100 m in altitude variation (Figure 8g). In comparison, both neo-endemism and paleo-endemism have their grid cells mostly scattered between ~11 and 322 m. Possible explanations is that variation in topography is a limiting factor for these species, or because most cells are found in areas with this variation in topography (Supplementary Figure A2). Statistical testing of this pattern is necessary for us to say anything about this for sure.

4.4 Species and phylogenetic turnover patterns

The RWTurnover analysis (Figure 9) and the PhyloRWTurnover analysis (Figure 10) showed distinct clusters of the vascular plants in Norway based on species and on branch lengths, respectively. In the PhyloRWTurnover analysis (Figure 10), almost all the clusters were

geographically clustered and structured together, which suggests that there is an evolutionary signal in the distribution of the vascular flora of Norway. If the species were randomly distributed, then the analysis would have shown a more random distribution of clusters. This is especially interesting for cluster 3, which does not appear in the RWTurnover analysis (Figure 9). Comparing the RWTurnover and PhyloRWTurnover analyses (Figure 9-10), we see that both have two very distinct clusters: cluster 1 and cluster 2. Comparing these two clusters with Moen et al.'s (1999) vegetation zones (Supplementary Figure A7), we see that they are very similar to the boreal (northern, middle and southern combined) and alpine zone, respectively. The low resolution of the grid cells may explain why the three different boreal zones (Supplementary Figure A7) did not appear distinctly in the maps. In the RWTurnover analysis (Figure 9) we see the geographic cluster 1c in mid-Norway is within cluster 2 in the PhyloRWTurnover analysis (Figure 10). This suggest that the species present in that area are more similar to the alpine zone defined in Moen et al.'s (1999), but they are more related to the species in the boreal zone. In the RWTurnover analysis (Figure 9), the alpine zone was divided into two distinct clusters: 3 and 6. Cluster 3 is most likely a result of centric alpine plants with a disjunct distribution (Moen et al., 1999). For the sub clusters, 3a and 3b were more similar to each other than any were with the northernmost cluster 3c (Figure 9b). This may suggest that cluster 3a and 3b have more species in common, and a potential immigration pattern could be discussed from this. Interestingly, the results from Berglund and Westerbergh (2001) on *Cerastium alpinum* in Norway coincide with the boundary between cluster 2a and 3a in the RWTurnover analysis (Figure 9). A hybrid zone has also been observed for mammals in mid-Norway and mid-Sweden (Taberlet et al., 1998), underpinning the importance of this area as a biologically interesting zone between species coming from the south-west and east.

Another cluster that is observed in both analyses (Figure 9-10) is cluster 5, which especially resembles the boreonemoral and nemoral zones of Moen et al. (1999) (Supplementary Figure A7). This cluster also coincides with the levels of observed PD (Figure 2b). High levels of PD on the coast of southern Norway match both the areas with high numbers of species (Figure 2a), areas of mixed endemism (Figure 7), and cluster 5 (Figure 10). Therefore, because it has neither the relatively highest number of long branches (Figure 3b) nor restricted long branches (Supplementary Figure A6b), this area can be interpreted as having the highest diversity of vascular taxa in Norway. This pattern can be explained by looking at both the drivers of diversity patterns (Figure 5) and the potential driver of endemism (Figure 8). Temperature of the warmest quarter is the most important driver for all the diversity patterns,

as it had the highest effect for higher values of diversity and the most negative effect on lower values of PD (Figure 5). This is not surprising, as temperature is found be to very important in driving vascular plant richness as previously discussed.

In general, the PhyloRWTurnover analysis resembles previous maps of Norway's vegetation (Supplementary Figure A7) more than the RWTurnover analysis does. The seen regionalisation of plants in the PhyloRWTurnover analysis (Figure 10) may be explained by phenotypic sorting where communities differ from each other environmentally (Webb et al., 2002). This may cause closely related species to co-occur because of traits important for the different environments stays conserved (Cavender-Bares et al., 2006). This is however not seen in the RWTurnover analysis (Figure 9) as it is based on species only, which does not take into account ecological similarities between species.

4.5 Future implications

Today, the climate is changing rapidly with increasing temperatures especially in Europe and the Arctic (Smith et al., 2015). As mean temperature of warmest quarter had such a positive effect on SR, PD and high PD and a negative effect on low PD (Figure 5), we can predict that an increase in this variable due to global warming may lead to a shift in the vegetation zones. We predict that western species (i.e. species that are mainly found in areas with mild winters) will tend to shift their ranges towards higher altitudes and latitudes with increasing temperatures. Because areas with significantly low PD are mainly found in the alpine vegetation zone (Supplementary Figure A7), we expect that this zone will either move northwards or disappear as alpine biota is also highly sensitive to increased temperatures (Bertrand et al., 2011). Indeed, a recent study found that there has been an increase in number of species in the Norwegian mountains (Steinbauer et al., 2018). Precipitation is also predicted to increase in the future (Engler et al., 2011). Because annual precipitation had a positive effect on SR and high PD as well as a negative effect on low PD, we expect increasing precipitation to drive further increases in the number and diversity of plant species in Norway. Although areas of significantly low PD do not score as high in diversity, it does not mean that conserving these areas are of low importance. In these areas, the loss of evolutionary history may be less important if the areas are high in species diversity (Huang and Roy, 2015). As low PD is mainly clustered within the alpine vegetation zone, a potential change to its environment may have dramatic causes for the biodiversity as areas of low PD may have less robust ecosystems (Tan et al., 2012).

It is also important to remember that species may not have reached all currently suitable habitats and that shifts in distributions may still be happening (Felde et al., 2017). Using the information from this study may therefore be of help in predicting future distributions. Knowing the difference between the vegetation zones based on the species (Figure 9) and their branches (Figure 10) will help in foreseeing how the zones may potentially change and shift.

4.6 Conclusions

In this study, we have seen that the vascular flora of Norway is non-randomly distributed across the region, and that the flora makes up different patterns on both a discrete species level and on a phylogenetic one. We found patterns that coincided with previous studies. Drivers of diversity patterns were found to be both of ecological (spatial heterogeneity and climatic) and historical (time since last glaciation cover) importance (Grytnes et al., 1999). There was a high relatedness between species in higher altitudes and latitudes (Grytnes et al., 2000). The PhyloRWTurnover analysis coincided with previous works on vegetation zones (Moen et al., 1999; Bakkestuen et al., 2008), and the results show that there is an evolutionary signal to the way that the flora is distributed in Norway.

We have also found patterns that have not previously been identified in Norway. Endemic branches in Norway seemed to be mainly driven by climatic variables (Figure 8). Neo-endemism (short restricted branches) was only found in areas with lower temperatures (below 11ºC in mean temperature of warmest quarter) and mainly found in areas with low levels of precipitation (below ~ 600 mm of annual precipitation). Mixed endemism (combination of short and long restricted branches) could be divided into two groups (north and south), where the northern mixed endemism cells were mainly found in lower temperatures (below 11ºC; Figure 7 and Figure 8c) and the southern mixed endemism cells were mainly found in areas with warmer temperatures (above 12.5ºC; Figure 7 and Figure 8c). This pattern could possibly be a result of different immigration routes of species from different floristic elements (Moen et al., 1999). We also saw a biotic transition zone in mid-northern Norway (Figure 9a), which could possibly be a biologically interesting area where species from the south-west and east meet. These results can be of help in future conservation assessments of vascular plants, where especially areas rich in endemic species could be looked.

5 REFERENCES

Albert, V. A., Williams, S. E. & Chase, M. W. 1992. Carnivorous plants: phylogeny and structural evolution. *Science,* 257**,** 1491-1495.

Allen, A. P., Brown, J. H. & Gillooly, J. F. 2002. Global biodiversity, biochemical kinetics, and the energetic-equivalence rule. *Science,* 297**,** 1545-1548.

Alm, T. 1993. Øvre Æråsvatn-palynostratigraphy of a 22,000 to 10,000 BP lacustrine record on Andøya, northern Norway. *Boreas,* 22**,** 171-188.

Bakkestuen, V., Erikstad, L. & Halvorsen, R. 2008. Step-less models for regional environmental variation in Norway. *Journal of biogeography,* 35**,** 1906-1922.

Barkman, T. J., McNeal, J. R., Lim, S.-H., Coat, G., Croom, H. B. & Young, N. D. 2007. Mitochondrial DNA suggests at least 11 origins of parasitism in angiosperms and reveals genomic chimerism in parasitic plants. *BMC Evolutionary Biology,* 7**,** 248.

Barnosky, A. D., Matzke, N., Tomiya, S., Wogan, G. O. U., Swartz, B., Quental, T. B., Marshall, C., McGuire, J. L., Lindsey, E. L., Maguire, K. C., Mersey, B. & Ferrer, E. A. 2011. Has the Earth's sixth mass extinction already arrived? *Nature,* 471**,** 51-57.

Barton, K. 2018. MuMIn: multi-model inference. R package version 1.40.4. https://CRAN.R-project.org/package=MuMIn.

Bateman, R. M., Hilton, J. & Rudall, P. J. 2006. Morphological and molecular phylogenetic context of the angiosperms: contrasting the ‘top-down’and ‘bottom-up’approaches used to infer the likely characteristics of the first flowers. *Journal of Experimental Botany,* 57**,** 3471-3503.

Beck, J., Böller, M., Erhardt, A. & Schwanghart, W. 2014. Spatial bias in the GBIF database and its effect on modeling species' geographic distributions. *Ecological Informatics,* 19**,** 10-15.

Benson, D. A., Cavanaugh, M., Clark, K., Karsch-Mizrachi, I., Lipman, D. J., Ostell, J. & Sayers, E. W. 2013. GenBank. *Nucleic acids research,* 41**,** D36-D42.

Berglund, A. B. N. & Westerbergh, A. 2001. Two postglacial immigration lineages of the polyploid Cerastium alpinum (Caryophyllaceae). *Hereditas,* 134**,** 171-183.

Bertness, M. D. & Callaway, R. 1994. POSITIVE INTERACTIONS IN COMMUNITIES. *Trends in Ecology & Evolution,* 9**,** 191-193.

Bertrand, R., Lenoir, J., Piedallu, C., Riofrío-Dillon, G., De Ruffray, P., Vidal, C., Pierrat, J.-C. & Gégout, J.-C. 2011. Changes in plant community composition lag behind climate warming in lowland forests. *Nature,* 479**,** 517.

Bieker, V. C. & Martin, M. D. 2018. Implications and future prospects for evolutionary analyses of DNA in historical herbarium collections. *Botany Letters***,** 1-10.

Bijlsma, R. & Loeschcke, V. 2005. Environmental stress, adaptation and evolution: an overview. *Journal of evolutionary biology,* 18**,** 744-749.

Bivand, R. & Piras, G. Comparing implementations of estimation methods for spatial econometrics. 2015. American Statistical Association.

Bivand, R. & Rundel, C. 2017. *rgeos: Interface to Geometry Engine - Open Source ('GEOS'). R package version 0.3-26.* https://CRAN.R-project.org/package=rgeos.

Brook, E. J., Nesje, A., Lehman, S. J., Raisbeck, G. M. & Yiou, F. o. 1996. Cosmogenic nuclide exposure ages along a vertical transect in western Norway: Implications for the height of the Fennoscandian ice sheet. *Geology,* 24**,** 207-210.

Brose, U., Martinez, N. D. & Williams, R. J. 2003. Estimating species richness: sensitivity to sample coverage and insensitivity to spatial patterns. *Ecology,* 84**,** 2364-2377.

Bruun, H. H., Moen, J., Virtanen, R., Grytnes, J.-A., Oksanen, L. & Angerbjörn, A. 2006. Effects of altitude and topography on species richness of vascular plants, bryophytes and lichens in alpine communities. *Journal of Vegetation Science,* 17**,** 37-46.

Butterfield, B. J., Cavieres, L. A., Callaway, R. M., Cook, B. J., Kikvidze, Z., Lortie, C. J., Michalet, R., Pugnaire, F. I., Schöb, C. & Xiao, S. 2013. Alpine cushion plants inhibit the loss of phylogenetic diversity in severe environments. *Ecology Letters,* 16**,** 478-486.

Byng, J. W., Chase, M. W., Christenhusz, M. J., Fay, M. F., Judd, W. S., Mabberley, D. J., Sennikov, A. N., Soltis, D. E., Soltis, P. S. & Stevens, P. F. 2016. An update of the Angiosperm Phylogeny Group classification for the orders and families of flowering plants: APG IV. *Botanical Journal of the Linnean Society,* 181**,** 1-20.

Cadotte, M. W., Cardinale, B. J. & Oakley, T. H. 2008. Evolutionary history and the effect of biodiversity on plant productivity. *Proceedings of the National Academy of Sciences,* 105**,** 17012-17017.

Cadotte, M. W., Cavender-Bares, J., Tilman, D. & Oakley, T. H. 2009. Using phylogenetic, functional and trait diversity to understand patterns of plant community productivity. *PloS one,* 4**,** e5695.

Callaway, R. M., Brooker, R., Choler, P., Kikvidze, Z., Lortie, C. J., Michalet, R., Paolini, L., Pugnaire, F. I., Newingham, B. & Aschehoug, E. T. 2002. Positive interactions among alpine plants increase with stress. *Nature,* 417**,** 844.

Casal-Lopez, M., Perea, S., Sousa-Santos, C., Robalo, J. I., Torralva, M., Oliva-Paterna, F. J. & Doadrio, I. 2018. Paleobiogeography of an Iberian endemic species, Luciobarbus sclateri (Gunther, 1868) (Actinopterygii, Cyprinidae), inferred from mitochondrial and nuclear markers. *Journal of Zoological Systematics and Evolutionary Research,* 56**,** 127.

Cavender-Bares, J., Keen, A. & Miles, B. 2006. Phylogenetic structure of Floridian plant communities depends on taxonomic and spatial scale. *Ecology,* 87.

Cavender-Bares, J., Kozak, K. H., Fine, P. V. A. & Kembel, S. W. 2009. The merging of community ecology and phylogenetic biology. *Ecology Letters,* 12**,** 693-715.

Chamberlain, S., Ram, K., Barve, V. & Mcglinn, D. 2017. rgbif: Interface to the Global Biodiversity Information Facility API. *R package version 0.9.9.*

Chao, A. 1987. Estimating the population size for capture-recapture data with unequal catchability. *Biometrics***,** 783-791.

Chapin, F. S., Zavaleta, E. S., Eviner, V. T., Naylor, R. L., Vitousek, P. M., Reynolds, H. L., Hooper, D. U., Lavorel, S., Sala, O. E., Hobbie, S. E., Mack, M. C. & Diaz, S. 2000. Consequences of changing biodiversity. *Nature,* 405**,** 234-242.

Crisp, M. D., Laffan, S., Linder, H. P. & Monro, A. 2001. Endemism in the Australian flora. *Journal of Biogeography,* 28**,** 183-198.

Dahl, E. 1998. *The phytogeography of northern Europe: British Isles, Fennoscandia, and adjacent areas,* Cambridge university press Cambridge.

Daru, B. H., Elliott, T. L., Park, D. S. & Davies, T. J. 2017a. Understanding the Processes Underpinning Patterns of Phylogenetic Regionalization. *Trends in Ecology & Evolution,* 32**,** 845-860.

Daru, B. H., Holt, B. G., Lessard, J.-P., Yessoufou, K. & Davies, T. J. 2017b. Phylogenetic regionalization of marine plants reveals close evolutionary affinities among disjunct temperate assemblages. *Biological Conservation,* 213**,** 351-356.

Darwin, C. 1859. *On the origin of species by means of natural selection, or, the preservation of favoured races in the struggle for life,* London, J. Murray.

Demesure, B., Comps, B. & Petit, R. J. 1996. Chloroplast DNA phylogeography of the common beech (Fagus sylvatica L.) in Europe. *Evolution,* 50**,** 2515-2520.

Donoghue, M. J. 2008. A phylogenetic perspective on the distribution of plant diversity. *Proceedings of the National Academy of Sciences,* 105**,** 11549-11555.

Dunning, L. T. & Savolainen, V. 2010. Broad-scale amplification of matK for DNA barcoding plants, a technical note. *Botanical Journal of the Linnean Society,* 164**,** 1-9.

Engler, R., Randin, C. F., Thuiller, W., Dullinger, S., Zimmermann, N. E., Araújo, M. B., Pearman, P. B., Le Lay, G., Piedallu, C. & Albert, C. H. 2011. 21st century climate change threatens mountain flora unequally across Europe. *Global Change Biology,* 17**,** 2330-2341.

Erschbamer, B., Ruth, N. S. & Winkler, E. 2008. Colonization processes on a central Alpine glacier foreland. *Journal of Vegetation Science,* 19**,** 855-862.

Faith, D. P. 1992. Conservation evaluation and phylogenetic diversity. *Biological Conservation,* 61**,** 1-10.

Felde, V. A., Grytnes, J. A., Bjune, A. E., Peglar, S. M. & Birks, H. J. B. 2017. Are diversity trends in western Scandinavia influenced by post-glacial dispersal limitation? *Journal of Vegetation Science*.

Felsenstein, J. 1985. Phylogenies and the comparative method. *The American Naturalist,* 125**,** 1-15.

Ford, C. S., Ayres, K. L., Toomey, N., Haider, N., Van Alphen Stahl, J., Kelly, L. J., Wikström, N., Hollingsworth, P. M., Duff, R. J. & Hoot, S. B. 2009. Selection of candidate coding DNA barcoding regions for use on land plants. *Botanical Journal of the Linnean Society,* 159**,** 1-11.

Forest, F., Crandall, K. A., Chase, M. W. & Faith, D. P. 2015. Phylogeny, extinction and conservation: embracing uncertainties in a time of urgency. *Philosophical Transactions of the Royal Society B-Biological Sciences,* 370**,** 8.

Gederaas, L., Moen, T., Skjelseth, S. & Larsen, L. 2012. Fremmede arter i Norge–med norsk svarteliste 2012. *Artsdatabanken, Trondheim,* 20**,** 2012.

Gittleman, J. L. & Kot, M. 1990. Adaptation: statistics and a null model for estimating phylogenetic effects. *Systematic Zoology,* 39**,** 227-241.

Gotelli, N. J. & Colwell, R. K. 2001. Quantifying biodiversity: procedures and pitfalls in the measurement and comparison of species richness. *Ecology letters,* 4**,** 379-391.

Graham, C. H. & Fine, P. V. 2008. Phylogenetic beta diversity: linking ecological and evolutionary processes across space in time. *Ecology letters,* 11**,** 1265-1277.

Grandcolas, P., Nattier, R. & Trewick, S. 2014. Relict species: a relict concept? *Trends in ecology & evolution,* 29**,** 655-663.

Griffith, D. A. 2000. Eigenfunction properties and approximations of selected incidence matrices employed in spatial analyses. *Linear Algebra and its Applications,* 321**,** 95-112.

Grytnes, J. A. 2003. Species-richness patterns of vascular plants along seven altitudinal transects in Norway. *Ecography,* 26**,** 291-300.

Grytnes, J. A., Birks, H., Heegaard, E. & Peglar, S. M. 2000. Geographical trends in the species-to-family ratio of vascular plants in Fennoscandia. *Norsk Geografisk Tidsskrift,* 54**,** 60-64.

Grytnes, J. A., Birks, H. & Peglar, S. M. 1999. Plant species richness in Fennoscandia: evaluating the relative importance of climate and history. *Nordic Journal of Botany,* 19**,** 489-503.

Guggisberg, A., Mansion, G., Kelso, S. & Conti, E. 2006. Evolution of biogeographic patterns, ploidy levels, and breeding systems in a diploid–polyploid species complex of Primula. *New Phytologist,* 171**,** 617-632.

Hawkins, B. A., Rueda, M., Rangel, T. F., Field, R. & Diniz-Filho, J. A. F. 2014. Community phylogenetics at the biogeographical scale: cold tolerance, niche conservatism and the structure of North American forests. *Journal of Biogeography,* 41**,** 23-38.

Hengl, T., de Jesus, J. M., Heuvelink, G. B., Gonzalez, M. R., Kilibarda, M., Blagotić, A., Shangguan, W., Wright, M. N., Geng, X. & Bauer-Marschallinger, B. 2017. SoilGrids250m: Global gridded soil information based on machine learning. *PLoS one,* 12**,** e0169748.

Hewitt, G. M. 1999. Post-glacial re-colonization of European biota. *Biological journal of the Linnean Society,* 68**,** 87-112.

Hijmans, R. J. 2017. *raster: Geographic Data Analysis and Modeling. R package version 2.6-7.* [Online]. Available: https://CRAN.R-project.org/package=raster [Accessed].

Hijmans, R. J., Phillips, S., Leathwick, J. & Elith, J. 2015. dismo: Species distribution modeling. R package version 1.0-12. *The R Foundation for Statistical Computing, Vienna* http://cran. *r-project. org*.

Hooper, D. U., Chapin, F. S., Ewel, J. J., Hector, A., Inchausti, P., Lavorel, S., Lawton, J. H., Lodge, D. M., Loreau, M., Naeem, S., Schmid, B., Setala, H., Symstad, A. J., Vandermeer, J. & Wardle, D. A. 2005. Effects of biodiversity on ecosystem functioning: A consensus of current knowledge. *Ecological Monographs,* 75**,** 3-35.

Huang, D. & Roy, K. 2015. The future of evolutionary diversity in reef corals. *Phil. Trans. R. Soc. B,* 370**,** 20140010.

James, S. A., Soltis, P. S., Belbin, L., Chapman, A. D., Nelson, G., Paul, D. L. & Collins, M. 2018. Herbarium data: Global biodiversity and societal botanical needs for novel research. *Applications in Plant Sciences,* 6**,** 8.

Jordan, G. J., Harrison, P. A., Worth, J. R., Williamson, G. J. & Kirkpatrick, J. B. 2016. Palaeoendemic plants provide evidence for persistence of open, well-watered vegetation since the Cretaceous. *Global Ecology and Biogeography,* 25**,** 127-140.

Katoh, K. & Standley, D. M. 2013. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Molecular biology and evolution,* 30**,** 772-780.

Kearse, M., Moir, R., Wilson, A., Stones-Havas, S., Cheung, M., Sturrock, S., Buxton, S., Cooper, A., Markowitz, S. & Duran, C. 2012. Geneious Basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data. *Bioinformatics,* 28**,** 1647-1649.

Kreft, H. & Jetz, W. 2007. Global patterns and determinants of vascular plant diversity. *Proceedings of the National Academy of Sciences of the United States of America,* 104**,** 5925-5930.

Laffan, S., Thornhill, A., Miller, J., Knerr, N., Gonzales-Orozco, C. & Mishler, B. Understanding spatial patterns of biodiversity: How sensitive is phylogenetic endemism to the randomisation model? International Conference on GIScience Short Paper Proceedings, 2016.

Laffan, S. W. & Crisp, M. D. 2003. Assessing endemism at multiple spatial scales, with an example from the Australian vascular flora. *Journal of Biogeography,* 30**,** 511-520.

Laffan, S. W., Lubarsky, E. & Rosauer, D. F. 2010. Biodiverse, a tool for the spatial analysis of biological and related diversity. *Ecography,* 33**,** 643-647.

Lagercrantz, U. & Ryman, N. 1990. Genetic structure of Norway spruce (Picea abies): concordance of morphological and allozymic variation. *Evolution,* 44**,** 38-53.

Lagomarsino, L. P., Condamine, F. L., Antonelli, A., Mulch, A. & Davis, C. C. 2016. The abiotic and biotic drivers of rapid diversification in Andean bellflowers (Campanulaceae). *New Phytologist,* 210**,** 1430-1442.

Lexer, C. & Fay, M. F. 2005. Adaptation to environmental stress: a rare or frequent driver of speciation? *Journal of Evolutionary Biology,* 18**,** 893-900.

Lid, J. & Lid, D. T. 2005. *Norsk flora. 7. utg. ved R. Elven,* Oslo, Det Norske Samlaget.

Mable, B. 2004. 'Why polyploidy is rarer in animals than in plants': myths and mechanisms. *Biological Journal of the Linnean Society,* 82**,** 453-466.

Mishler, B. D., Knerr, N., Gonzalez-Orozco, C. E., Thornhill, A. H., Laffan, S. W. & Miller, J. T. 2014. Phylogenetic measures of biodiversity and neo- and paleo-endemism in Australian Acacia. *Nature Communications,* 5**,** 10.

Moen, A., Lillethun, A. & Odland, A. 1999. *Vegetation,* Hønefoss, Norwegian Mapping Authority.

Myers, N., Mittermeier, R. A., Mittermeier, C. G., da Fonseca, G. A. B. & Kent, J. 2000. Biodiversity hotspots for conservation priorities. *Nature,* 403**,** 853-858.

Nagalingum, N. S., Knerr, N., Laffan, S. W., González-Orozco, C. E., Thornhill, A. H., Miller, J. T. & Mishler, B. D. 2015. Continental scale patterns and predictors of fern richness and phylogenetic diversity. *Frontiers in Genetics,* 6.

Nieto-Blázquez, M. E., Antonelli, A. & Roncal, J. 2017. Historical Biogeography of endemic seed plant genera in the Caribbean: Did GAARlandia play a role? *Ecology and Evolution,* 7**,** 10158-10174.

Nordal, I. 1987. Tabula rasa after all? Botanical evidence for ice-free refugia in Scandinavia reviewed. *Journal of Biogeography***,** 377-388.

Nordal, I., Jonsell, B. & Marcussen, T. 2005. Viola rupestris: molecular analyses to elucidate postglacial migration in Western Europe. *Journal of biogeography,* 32**,** 1453-1459.

O'Donnel, M. S. & Ignizio, D. A. 2012. Bioclimatic predictors for supporting ecological applications in the conterminous United States. US Geological Survey.

Oksanen, J., Blanchet F.G., Friendly M., Kindt R, Legendre P, McGlinn D, Minchin P.R., O'Hara R. B. , Simpson G.L. , Solymos P., Stevens M.H.H, Szoecs E. & H., W. 2017. vegan: Community Ecology Package. R package version 2.4-5. https://CRAN.R-project.org/package=vegan.

Parducci, L., Jorgensen, T., Tollefsrud, M. M., Elverland, E., Alm, T., Fontana, S. L., Bennett, K. D., Haile, J., Matetovici, I., Suyama, Y., Edwards, M. E., Andersen, K., Rasmussen, M., Boessenkool, S., Coissac, E., Brochmann, C., Taberlet, P., Houmark-Nielsen, M., Larsen, N. K., Orlando, L., Gilbert, M. T. P., Kjaer, K. H., Alsos, I. G. & Willerslev, E. 2012. Glacial Survival of Boreal Trees in Northern Scandinavia. *Science,* 335**,** 1083-1086.

Pryer, K. M., Smith, A. R., Hunt, J. S. & Dubuisson, J.-Y. 2001. rbcL data reveal two monophyletic groups of filmy ferns (Filicopsida: Hymenophyllaceae). *American Journal of Botany,* 88**,** 1118-1130.

Purvis, A. & Hector, A. 2000. Getting the measure of biodiversity. *Nature,* 405**,** 212-219.

Pärtel, M. 2002. Local plant diversity patterns and evolutionary history at the regional scale. *Ecology,* 83**,** 2361-2366.

R Core Team 2017. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/.

Rambaut, A. 2012. FigTree v1. 4. *Molecular evolution, phylogenetics and epidemiology. Edinburgh, UK: University of Edinburgh, Institute of Evolutionary Biology*.

Ratnasingham, S. & Hebert, P. D. 2007. BOLD: The Barcode of Life Data System (http://www. barcodinglife. org). *Molecular Ecology Resources,* 7**,** 355-364.

Rohde, K. 1992. Latitudinal gradients in species diversity: the search for the primary cause. *Oikos***,** 514-527.

Rosauer, D., Laffan, S. W., Crisp, M. D., Donnellan, S. C. & Cook, L. G. 2009. Phylogenetic endemism: a new approach for identifying geographical concentrations of evolutionary history. *Molecular Ecology,* 18**,** 4061-4072.

Scherson, R. A., Thornhill, A. H., Urbina-Casanova, R., Freyman, W. A., Pliscoff, P. A. & Mishler, B. D. 2017. Spatial phylogenetics of the vascular flora of Chile. *Molecular Phylogenetics and Evolution,* 112**,** 88-95.

Schwartz, M. W. & Simberloff, D. 2001. Taxon size predicts rates of rarity in vascular plants. *Ecology Letters,* 4**,** 464-469.

Smith, S. J., Edmonds, J., Hartin, C. A., Mundra, A. & Calvin, K. 2015. Near-term acceleration in the rate of temperature change. *Nature Climate Change,* 5**,** 333.

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

Speed, J. D. M. & Austrheim, G. 2017. The importance of herbivore density and management as determinants of the distribution of rare plant species. *Biological Conservation,* 205**,** 77-84.

Stamatakis, A. 2006. RAxML-VI-HPC: maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. *Bioinformatics,* 22**,** 2688-2690.

Stanford, A. M., Harden, R. & Parks, C. R. 2000. Phylogeny and biogeography of Juglans (Juglandaceae) based on matK and ITS sequence data. *American Journal of Botany,* 87**,** 872-882.

Steinbauer, M. J., Grytnes, J. A., Jurasinski, G., Kulonen, A., Lenoir, J., Pauli, H., Rixen, C., Winkler, M., Bardy-Durchhalter, M., Barni, E., Bjorkman, A. D., Breiner, F. T., Burg, S., Czortek, P., Dawes, M. A., Delimat, A., Dullinger, S., Erschbamer, B., Felde, V. A., Fernandez-Arberas, O., Fossheim, K. F., Gomez-Garcia, D., Georges, D., Grindrud, E. T., Haider, S., Haugum, S. V., Henriksen, H., Herreros, M. J., Jaroszewicz, B., Jaroszynska, F., Kanka, R., Kapfer, J., Klanderud, K., Kuhn, I., Lamprecht, A., Matteodo, M., di Cella, U. M., Normand, S., Odland, A., Olsen, S. L., Palacio, S., Petey, M., Piscova, V., Sedlakova, B., Steinbauer, K., Stockli, V., Svenning, J. C., Teppa, G., Theurillat, J. P., Vittoz, P., Woodin, S. J., Zimmermann, N. E. & Wipf, S. 2018. Accelerated increase in plant species richness on mountain summits is linked to warming. *Nature,* 556**,** 231.

Stroeven, A. P., Hättestrand, C., Kleman, J., Heyman, J., Fabel, D., Fredin, O., Goodfellow, B. W., Harbor, J. M., Jansen, J. D., Olsen, L., Caffee, M. W., Fink, D., Lundqvist, J., Rosqvist, G. C., Strömberg, B. & Jansson, K. N. 2016. Deglaciation of Fennoscandia. *Quaternary Science Reviews,* 147**,** 91-121.

Sweeney, P. W., Starly, B., Morris, P. J., Xu, Y. M., Jones, A., Radhakrishnan, S., Grassa, C. J. & Davis, C. C. 2018. Large-scale digitization of herbarium specimens: Development and usage of an automated, high-throughput conveyor system. *Taxon,* 67**,** 165-178.

Taberlet, P., Fumagalli, L., Wust-Saucy, A. G. & Cosson, J. F. 1998. Comparative phylogeography and postglacial colonization routes in Europe. *Molecular Ecology,* 7**,** 453-464.

Tan, J., Pu, Z., Ryberg, W. A. & Jiang, L. 2012. Species phylogenetic relatedness, priority effects, and ecosystem functioning. *Ecology,* 93**,** 1164-1172.

Thornhill, A. H., Baldwin, B. G., Freyman, W. A., Nosratinia, S., Kling, M. M., Morueta-Holme, N., Madsen, T. P., Ackerly, D. D. & Mishler, B. D. 2017. Spatial phylogenetics of the native California flora. *BMC biology,* 15**,** 96.

Thornhill, A. H., Mishler, B. D., Knerr, N. J., González-Orozco, C. E., Costion, C. M., Crayn, D. M., Laffan, S. W. & Miller, J. T. 2016. Continental-scale spatial phylogenetics of Australian angiosperms provides insights into ecology, evolution and conservation. *Journal of Biogeography,* 43**,** 2085-2098.

Vilgalys, R. 2003. Taxonomic misidentification in public DNA databases. *New Phytologist,* 160**,** 4-5.

Violle, C., Nemergut, D. R., Pu, Z. & Jiang, L. 2011. Phylogenetic limiting similarity and competitive exclusion. *Ecology letters,* 14**,** 782-787.

Vorren, T. O., Vorren, K. D., Alm, T., Gulliksen, S. & Løvlie, R. 1988. The last deglaciation (20,000 to 11,000 BP) on Andoya, Northern Norway. *Boreas,* 17**,** 41-77.

Wang, Y., Kockelman, K. M. & Wang, X. C. 2013. Understanding spatial filtering for analysis of land use-transport data. *Journal of Transport Geography,* 31**,** 123-131.

Webb, C. O., Ackerly, D. D., McPeek, M. A. & Donoghue, M. J. 2002. Phylogenies and community ecology. *Annual review of ecology and systematics,* 33**,** 475-505.

Westergaard, K. 2012. Deler av faggrunnlaget for smalasal Sorbus lancifolia og nordlandasal Sorbus neglecta.[The scientific basis for a management action plan for Sorbus lancifolia and Sorbus neglecta]. *NINA Report,* 835**,** 1-27.

Westergaard, K. B., Alsos, I. G., Popp, M., Engelskjøn, T., Flatberg, K. I. & Brochmann, C. 2011. Glacial survival may matter after all: nunatak signatures in the rare European populations of two west-arctic species. *Molecular Ecology,* 20**,** 376-393.

White, T. J., Bruns, T., Lee, S. & Taylor, J. 1990. Amplification and direct sequencing of fungal ribosomal RNA genes for phylogenetics. *PCR protocols: a guide to methods and applications,* 18**,** 315-322.

Whittaker, R. H. 1960. Vegetation of the Siskiyou mountains, Oregon and California. *Ecological monographs,* 30**,** 279-338.

Wille, N. 1915. The flora of Norway and its immigration. *Annals of the Missouri Botanical Garden,* 2**,** 59-108.

Zuber, D. & Widmer, A. 2000. Genetic evidence for host specificity in the hemi-parasitic Viscum album L.(Viscaceae). *Molecular Ecology,* 9**,** 1069-1073.

APPENDIX A

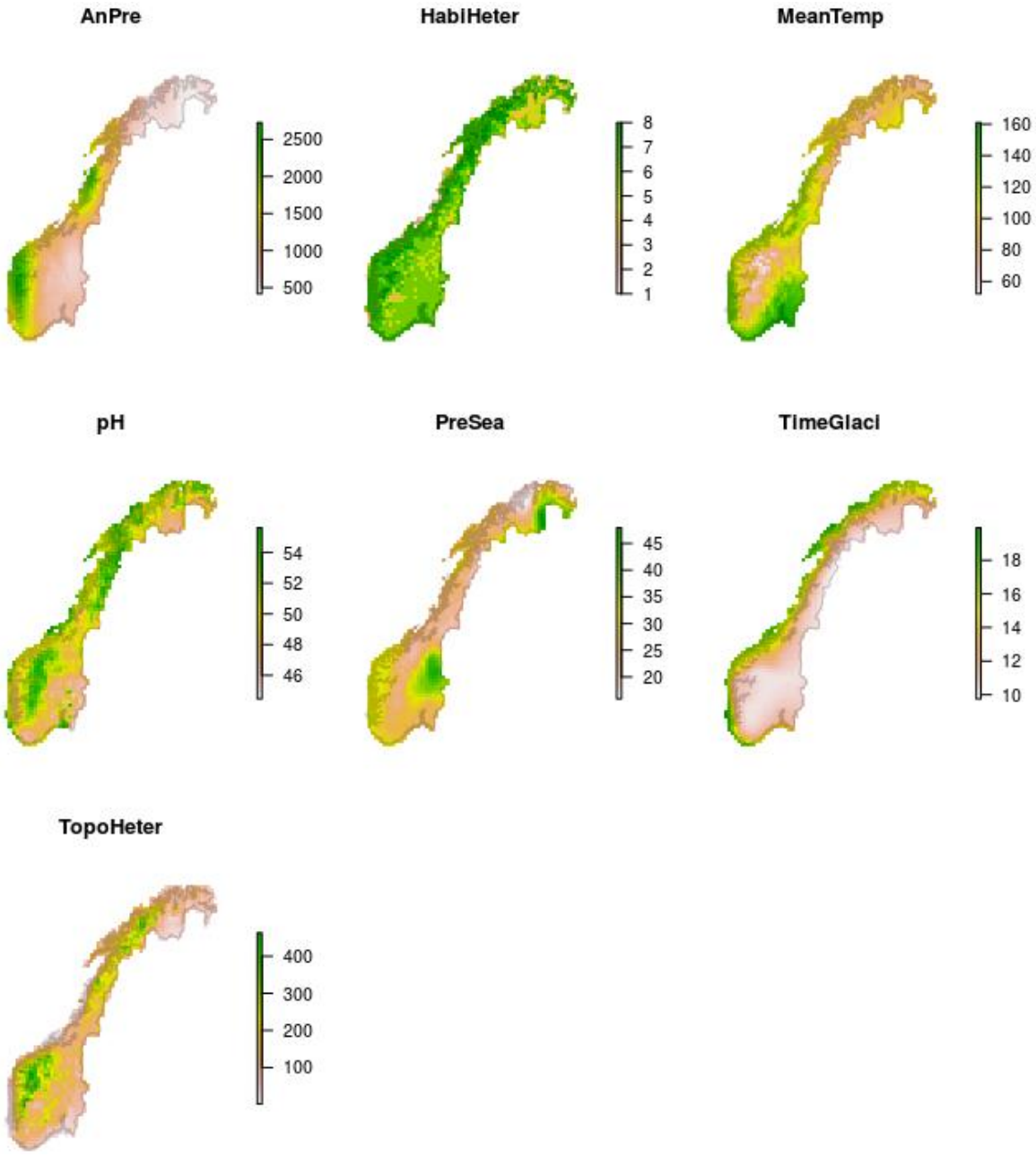

Figure A1 Explanatory variables and their distribution in Norway for analysing the diversity and endemism patterns: topographic heterogeneity (TopoHeter), habitat heterogeneity (HabiHeter), time since last glaciation cover (TimeGlaci), soil pH (pH), annual precipitation (AnPre), mean temperature of warmest quarter (MeanTemp), and precipitation seasonality (PreSea). All variables are in 20x20-km grid cells and projected to WGS 84 / UTM zone 33n. The plots represent the map of each variable before the occurrence plot was overlaid for the analyses (Supplementary Figure A4).

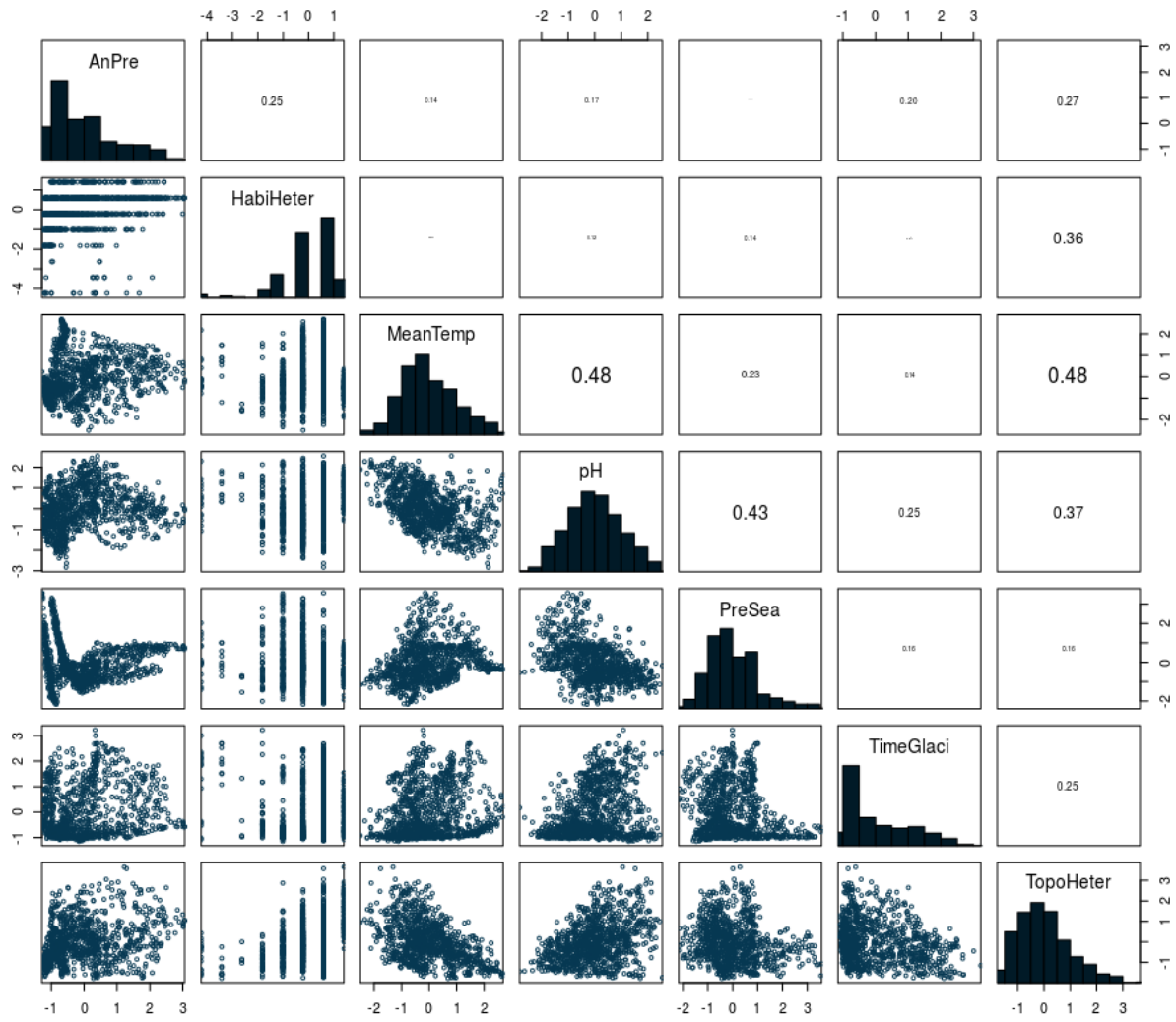

Figure A2 Pairwise correlation plots between the explanatory variables: annual precipitation (AnPre), habitat heterogeneity (HabiHeter), mean temperature of warmest quarter (MeanTemp), soil pH (pH), precipitation seasonality (PreSea), time since last glaciation cover (TimeGlaci), and topographic heterogeneity (TopoHeter). The diagonal plots show the frequency distribution of the different explanatory variables. The correlation plots (lower panels) show the different variables plotted against each other. The numbers in the upper panels are the correlation coefficients between the indices with 1 being fully correlated and 0 having no correlation at all.

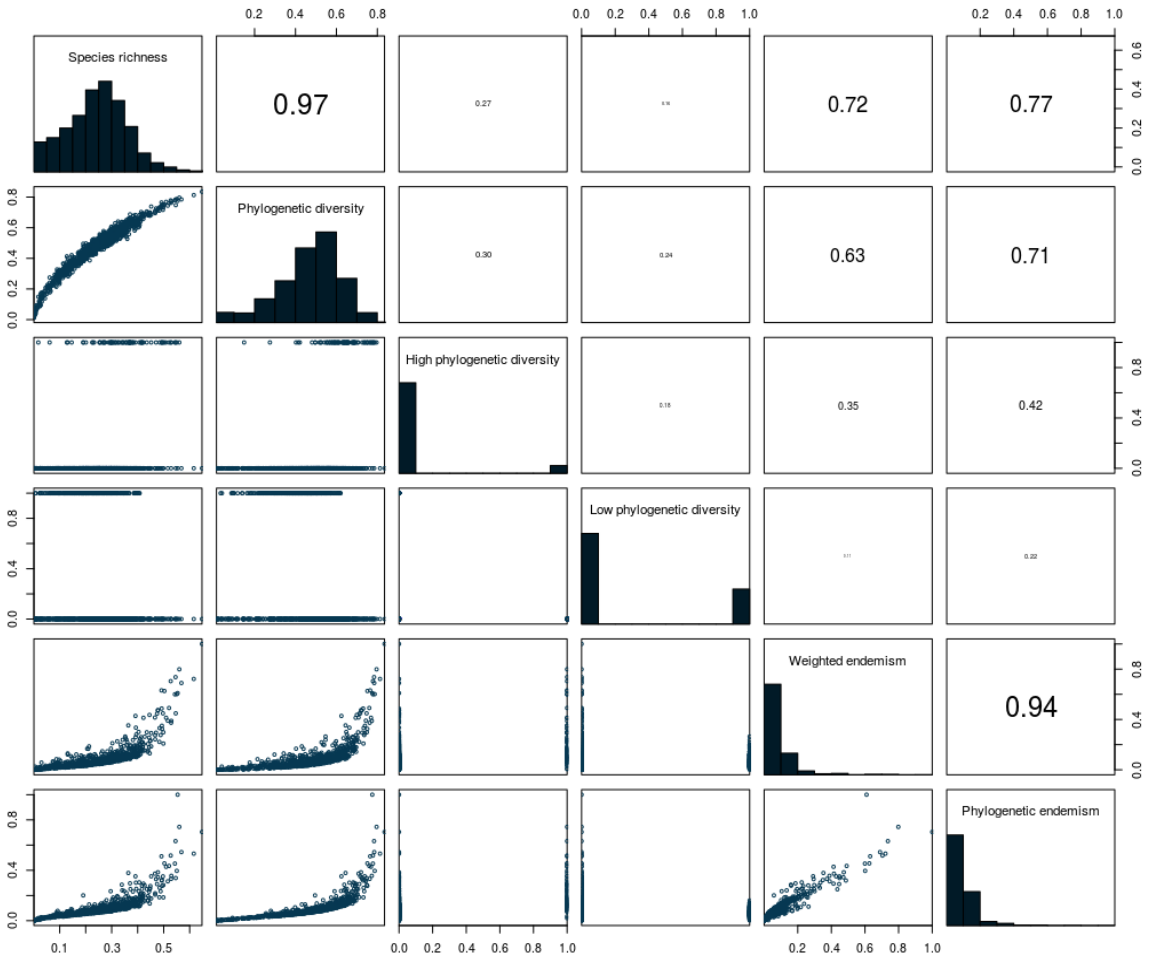

Figure A3 Distribution plots of species richness, phylogenetic diversity, high phylogenetic diversity, low phylogenetic diversity, weighted endemism, and phylogenetic endemism. The diagonal plots show the frequency distribution of the different diversity patterns. The correlation plots (lower panels) show the different indices plotted against each other. The numbers in the upper panels are the correlation coefficients between the indices with 1 being fully correlated and 0 having no correlation at all.

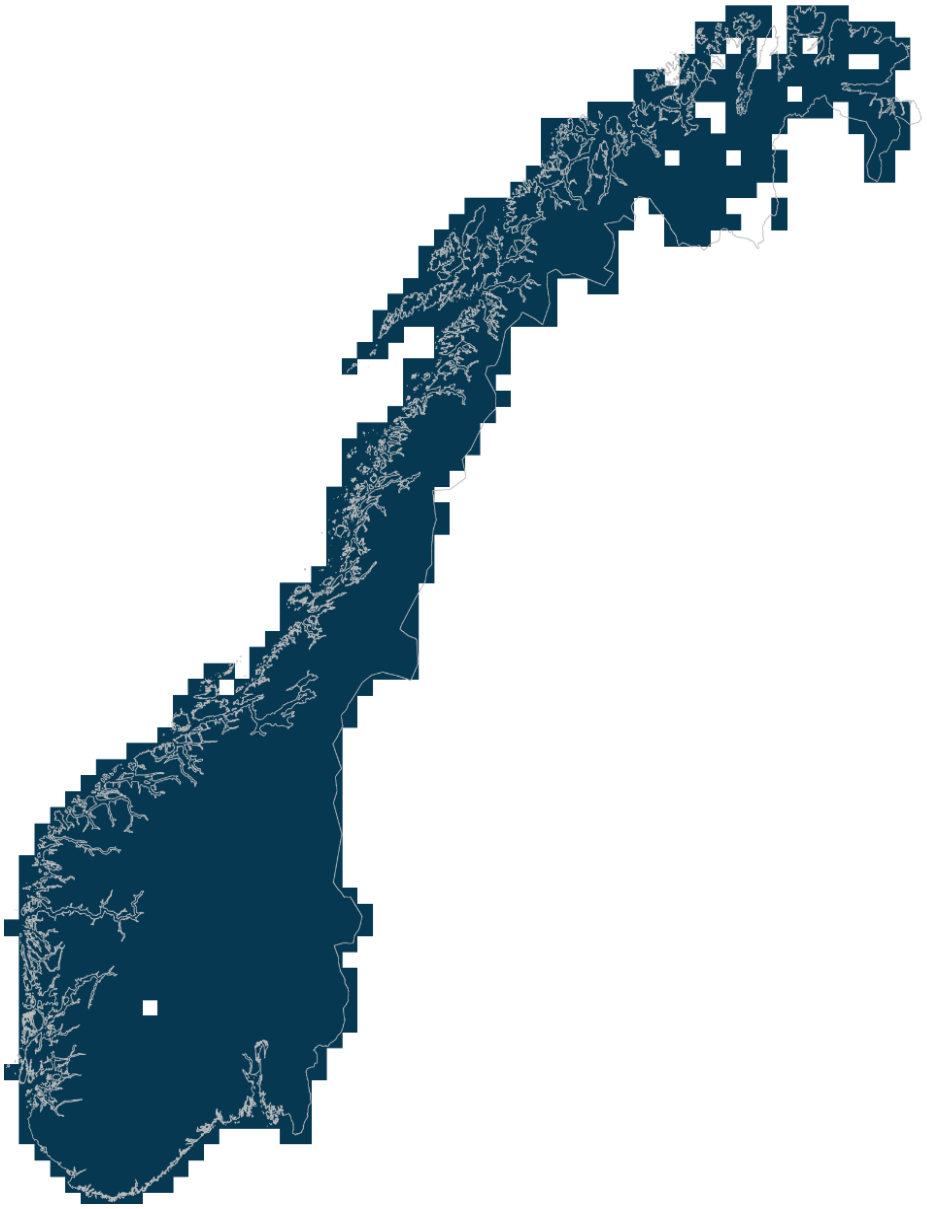

Figure A4 Map of the grid cells that were used in the different analyses. The map is projected in WGS 84 / UTM zone 33N and with 20x20-km grid cells. Coloured cells show which cells that were included in the analyses. White cells within the Norwegian border had a completeness less than 50%, and where thus not included in the analyses (See Methods – Species occurrence data).

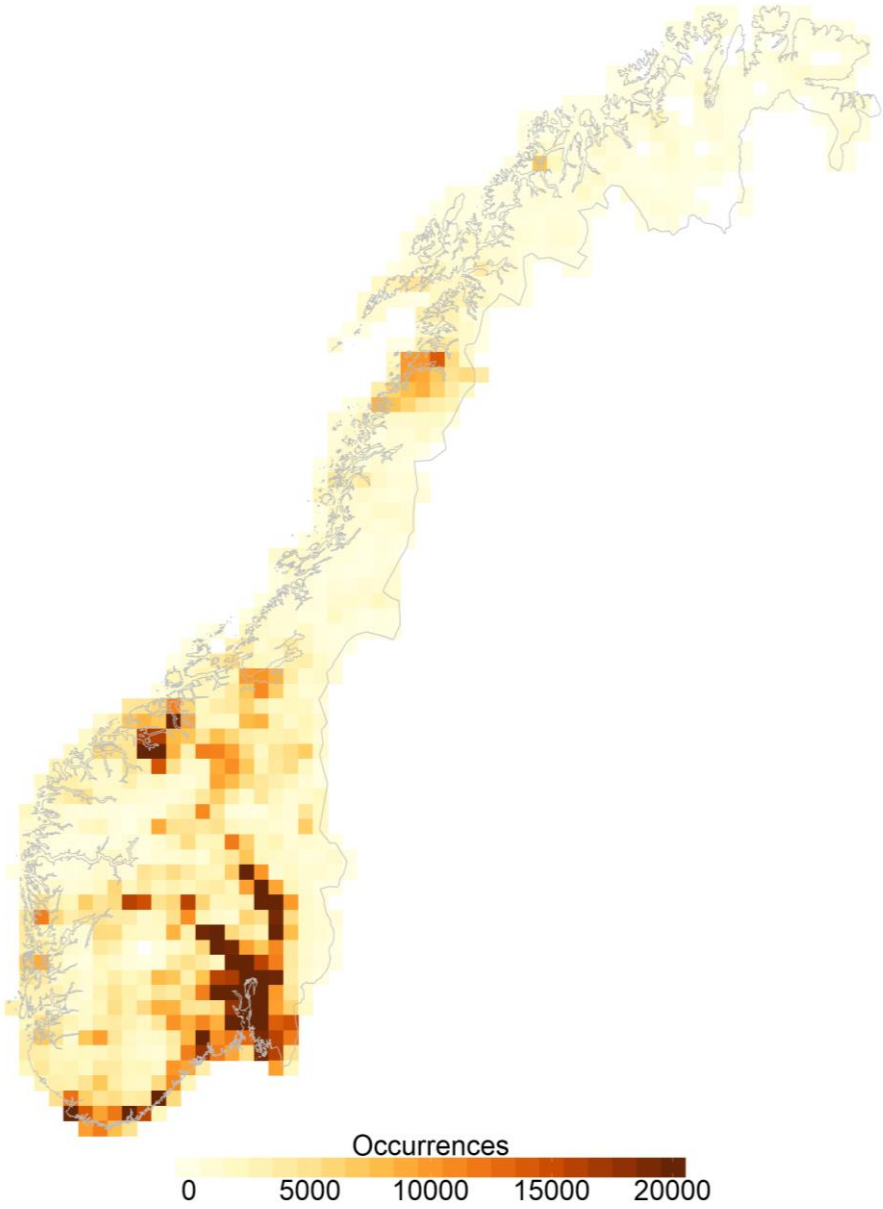

Figure A5 Number of occurrences of vascular plant species per 20x20km grid cell. The darkest cells in the legend (20000 occurrences) range from 20000 to the maximum number of occurrences in a grid cell, which is 65113.

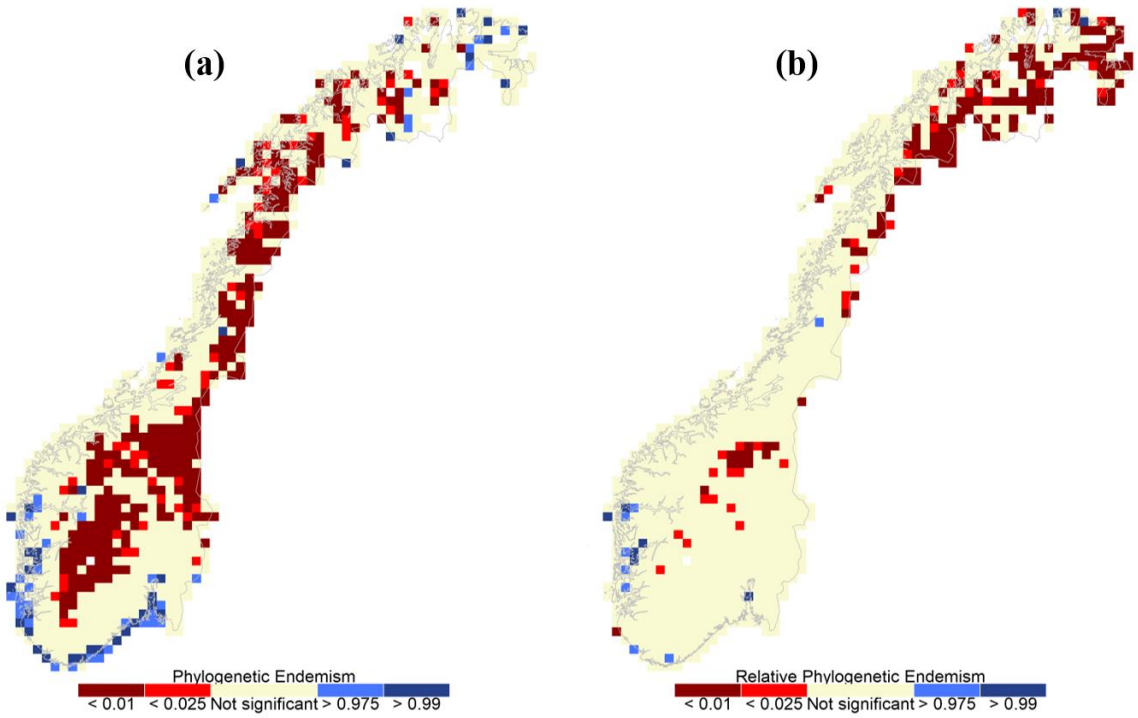

Figure A6 Results of the randomised phylogenetic endemism pattern in Figure 3. (a) Randomised phylogenetic endemism. Blue cells show areas with significantly high phylogenetic endemism (> 0.975) and red cells show significantly low phylogenetic endemism (< 0.025). (b) Randomised relative phylogenetic endemism. Blue cells show significantly high over-representation of long, restricted branches, and red cells show significantly low over-representation of short, restricted branches. Beige cells have no significant values. White cells scattered across Norway were not included in the analyses.

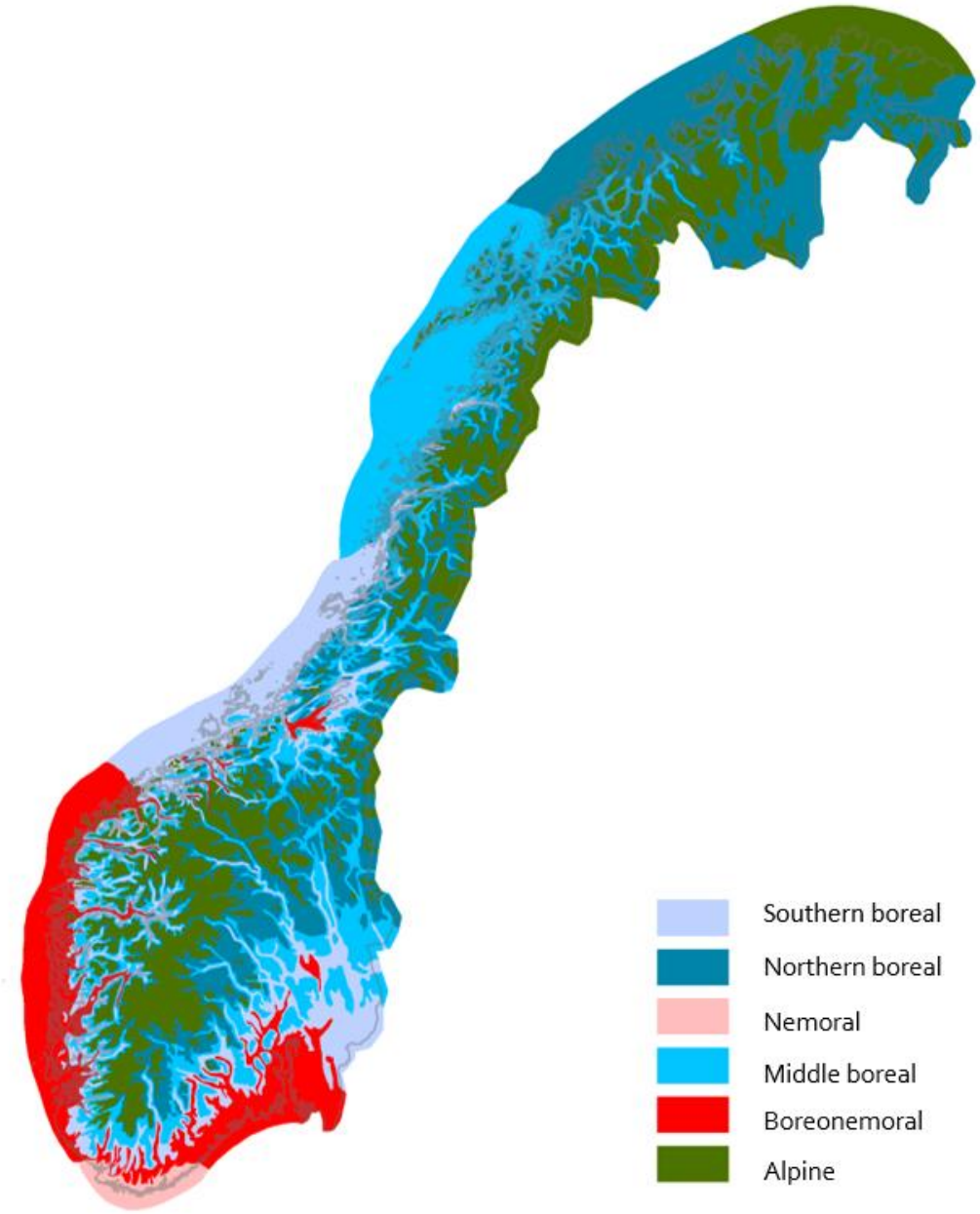

Figure A7 The vegetation zones in Norway modified from Moen et al. (1999). The zones are divided into six categories: southern boreal, northern boreal, nemoral, middle boreal, boreonemoral, and alpine. The layer is a polygon, and is outlined with the Norwegian border (grey).

Table A1 The top-ranked models for each of the response variables: species richness, phylogenetic diversity, significantly high phylogenetic diversity, and significantly low phylogenetic diversity. For the model to be top-ranked, the ΔAIC < 3. k represents number of parameters in the model. Weight is the weight each model is given and is based on the probability of the model being the single best model. Log-likelihood is the likelihood of the model fitting the data. The explanatory variables are annual precipitation (AnPre), habitat heterogeneity (HabiHeter), mean temperature of warmest quarter (MeanTemp), precipitation sesasonality (PreSea), soil pH (pH), time since last glaciation cover (TimeGlaci), and topographic heterogeneity (TopoHeter). The different Moran's eigenvectors that were also included in each model are not included in the table.

	k	**ΔAIC**	**Weight**	**Log-likelihood**
Species richness				
~ AnPre + HabiHeter + MeanTemp + PreSea + TimeGlaci + TopoHeter	6	0.00	0.72	1438.27
~ AnPre + HabiHeter + MeanTemp + PreSea + pH + TimeGlaci + TopoHeter	7	1.89	0.28	1438.45
Phylogenetic diversity				
~ HabiHeter + MeanTemp + PreSea + TimeGlaci + TopoHeter	5	0.00	0.48	1074.90
~ HabiHeter + MeanTemp + PreSea + pH + TimeGlaci + TopoHeter	6	1.60	0.22	1075.18
~ AnPre + HabiHeter + MeanTemp + PreSea + TimeGlaci + TopoHeter	6	2.14	0.17	1074.90
High phylogenetic diversity				
~ AnPre + HabiHeter + MeanTemp + PreSea + TimeGlaci + TopoHeter	6	0.00	0.60	-144.22
~ AnPre + HabiHeter + MeanTemp + PreSea + pH + TimeGlaci + TopoHeter	7	2.04	0.22	-144.21

	k	ΔAIC	Weight	Log-likelihood
Low phylogenetic diversity				
~ AnPre + HabiHeter + MeanTemp + PreSea + pH + TimeGlaci + TopoHeter	7	0.00	0.61	-306.88
~ AnPre + HabiHeter + MeanTemp + PreSea + TimeGlaci + TopoHeter	6	0.97	0.38	-308.40