Master's thesis

NTNU
Norwegian University of Science and Technology
NTNU University Museum
Department of Natural History

Nils Einar Brandtzæg

Whole genome data reveals possible speciation in the common peatmoss *Sphagnum compactum*

Master's thesis in Biology
Supervisor: Kristian Hassel
Co-supervisor: Olena Meleshko & Magni Olsen Kyrkjeeide
June 2023

Photo: Kjell Ivar Flatberg | NTNU Vitenskapsmuseet

NTNU
Norwegian University of
Science and Technology

Nils Einar Brandtzæg

Whole genome data reveals possible speciation in the common peatmoss *Sphagnum compactum*

Master's thesis in Biology
Supervisor: Kristian Hassel
Co-supervisor: Olena Meleshko & Magni Olsen Kyrkjeeide
June 2023

Norwegian University of Science and Technology
NTNU University Museum
Department of Natural History

NTNU
Norwegian University of
Science and Technology

Abstract

A long standing view in bryophyte biogeography is that bryophyte species have wide distribution ranges and low speciation rates. Here I study the population structure of the common and widespread peatmoss *Sphagnum compactum* Lam. & DC. in Europe. I use low coverage whole genome sequencing data from 41 individual samples, sampled from five populations across Europe. The genetic clustering of the data was assessed using model-based evolutionary clustering (ADMIXTURE) and principal component analysis. The genetic differentiation among the clusters was inferred based on F_{ST}, and phylogenetic relationship was assessed using a maximum likelihood approach. Split time between the clusters was estimated using a coalescent based approach. The results show that the sampled populations contain three genetically distinct clusters. The genetic differentiation, low gene flow, and split times between the clusters could indicate that they represent separate species. My results add to the increasing amount of evidence for genetic structuring in widely distributed bryophytes species.

Acknowledgements

First, I want to thank my supervisors; Kristian Hassel, Lena Meleshko, and Magni Olsen Kyrkjeeide, for invaluable academic guidance, encouragement, and patience throughout this project. And especially Lena for all the lab work and help with bioinformatics, I don't think this would have been possible without you, thank you!

I also want to thank the rest of the Bryology Group, especially Kjell Ivar Flatberg and Tommy Prestø, for taking part in discussions about my results. Discussions with and advice from members of the NTNU Museum Holomuseomics group is also greatly appreciated.

Thank you to my fellow students Ingvill, Charlotte, and Tiril. Sharing a reading room with you throughout this process has been wonderful, thanks for great discussions, and fun coffee breaks.

I am grateful to Ole, my twin brother and my father for practical help and encouragement, informatics is not always easy for a biologist. I am grateful to my mom for emotional support, and for reminding me to eat good food and drink nice wine. Thank you to Astrid and Trygve for always supporting me. Finally, I want to thank all of my friends for the good times that kept me sane during this project. Especially Thea for putting a smile on my face every day.

Contents

Abstract ..... **i**
Acknowledgements ..... **ii**
Introduction **2**
Materials and Methods **4**
Sampling 4
DNA extraction, Library Preparation and Sequencing 5
Sequencing Data Processing 5
SNP Calling and Filtering 5
Principal Component Analysis 6
Admixture 6
FST analysis 6
Phylogenetic analysis 7
Demographic history 7
Results **8**
Sequencing Summary, Mapping, SNP Calling and Filtering 8
PCA 8
Admixture 9
FST analysis 10
Phylogenetic analysis 11
Demographic history 13
Discussion **14**
Genetic distinctiveness of the clusters 14
Demographic history of the clusters 15
Implications for bryophyte biogeography 15
References **17**
Appendix A - Supporting tables **23**
Appendix B - Supporting figures **29**

Introduction

Speciation is an evolutionary process that results in the reproductive isolation of populations. Speciation can occur in allopatric or sympatric populations; in the former there is a geographic barrier to gene flow, while in the latter there are no spatial barriers to gene flow (Sætre and Ravinet, 2019, pp. 194–195). However, these two modes of speciation are at each end of a continuum of gene flow among diverging populations, and most speciation events fall somewhere in between (Butlin et al., 2008). Since speciation is the formation of reproductive isolation between populations, studying population structure can give insight into past and ongoing speciation events.

Population structure is the distribution of alleles in time and space. This distribution is the result of evolutionary and ecological processes. Dispersal is one important process that affects population structure (Alsos et al., 2007; Kyrkjeeide et al., 2016a; Muñoz et al., 2004). Dispersal (gene flow) is an important factor in speciation, either constraining speciation by homogenising the gene pool, or promoting speciation by providing the raw material for adaptive divergence (Morjan and Rieseberg, 2004; Slatkin, 1987). Lack of dispersal, on the other hand, results in isolated populations that evolve separately (Slatkin, 1987).

Population structure of extant northern species is affected by climate fluctuations during the Quaternary that resulted in expansion and retraction of the Eurasian ice sheets due to changes in latitudinal and altitudinal ranges (Hewitt, 2004; Stewart et al., 2010; Taberlet et al., 1998). During the last glacial maximum (LGM, ~20 Kya), the Eurasian ice sheet covered most of Fennoscandia and extended to continental Europe and the British Isles (Svendsen et al., 2004). The extant northern species could have survived the LGM in two types of refugia: refugia outside the ice sheet and cryptic northern refugia (in-situ survival) (Stewart et al., 2010). Cryptic refugia were ice-free areas along the coast or mountains protruding from the ice (nunataks) (Westergaard et al., 2019, 2011). Post-LGM recolonization of glaciated areas from refugia outside the ice sheet is expected to result in populations with reduced genetic diversity due to repeated founder events during recolonization, so-called bottleneck effects (Hewitt, 1996). High genetic diversity of populations living in previously glaciated areas can indicate that the area is a contact zone i.e. the species has survived in different refugia and recolonized the same area after the LGM (Provan and Bennett, 2008).

Bryophytes (mosses, liverworts, and hornworts) mainly disperse by small spores, and are capable of frequent long-distance dispersal (Muñoz et al., 2004). Bryophytes often have exceptionally broad geographical distribution ranges (Shaw, 2001), that could be explained by their dispersal capabilities. Based on that, one could expect a homogeneous distribution of genetic variants across their distribution range (Klein et al., 2006; Szövényi et al., 2012), but there is evidence for geographic genetic structure in bryophytes populations (Duffy et al., 2022; Kyrkjeeide et al., 2016a). Recently, cosmopolitan bryophyte species from various

families have been circumscribed as multiple species based on molecular and thorough morphological analysis (Hassel et al., 2018; Vigalondo et al., 2019).

Sphagnum L., the peatmosses, is a species rich (300-500 species) moss genus which dominates northern peatlands (Shaw et al., 2003). In peatlands, peatmosses create conditions which hamper decomposition, resulting in a build-up of peat (organic material) and net carbon fixation (Rydin and Jeglum, 2013). Thus, peatmoss dominated peatlands serve an important role for the global climate, storing about 30% of all soil organic carbon (Weston et al., 2015). Peatmosses exhibit great variability in life history, ecology, and morphological plasticity (Stenøien et al., 1997, 2014). The life cycle of peatmosses is dominated by a haploid gametophyte that produces gametes (in archegonia and antheridia), and fertilisation results in a diploid sporophyte that is dependent on the gametophyte. The sporophyte produces spores through meiosis. The genus includes both monoicous (archegonia and antheridia on the same gametophyte) and dioicous species (archegonia and antheridia on different gametophytes), and the former are more frequently found with sporophytes (Cronberg, 1992). Selfing is common in monoicous *Sphagnum* species, but outcrossing do also occur (Johnson and Shaw, 2015).

The extant species richness in *Sphagnum* originates from a rapid and relatively recent (7-20 Mya) burst of diversification likely associated with the Miocene cooling (Shaw et al., 2010). Speciation is an ongoing process, a more "recent" example of speciation in *Sphagnum* is the split between *S. divinum* Flatberg & Hassel vs. *S. medium* Limpr. which is estimated to be ~28,900 generations ago (Yousefi et al., 2017).

Sphagnum compactum Lam. & DC. is a monoicous peatmoss species with a nearly cosmopolitan distribution which frequently produces spores (Cronberg, 1992). Johnson and Shaw (2016) studied mating patterns in *Sphagnum*: sampling two *S. compactum* populations, they found evidence for selfing and out-crossing in both populations. Suzuki (1965) describes *S. compactum* as an easily identifiable species, although with some morphological differences between samples from Japan and Germany. Daniels (1985) studied isozyme variation in *S. compactum* samples from England and Finland, and found overall a high similarity, but genetic distance was greater between the English and Finnish populations than among the Finnish populations. Furthermore, in a phylogenomic analysis of 12 common *Sphagnum* species sampled in three metapopulations from central Norway, Austria, and Germany, Meleshko et al. (2021) found that *S. compactum* was the only species with evidence for strong geographical structure, with samples from each metapopulation forming distinctive, well-supported clades in the analysis.

In this study, I will explore the genetic structure of *S. compactum* in Europe by expanding the dataset by Meleshko et al. (2021) and using low-coverage whole-genome sequencing data analyses. I hypothesise that (1) the observed European clades are highly differentiated due to speciation; and (2) the gene flow between them is low.

I aim to (1) explore the genetic variation of the three European *S. compactum* metapopulations from Meleshko et al. (2021), (2) estimate genetic differentiation and admixture between the populations, and (3) use a coalescence model to estimate time of divergence between the populations.

Materials and Methods

Sampling

A total of 41 individual samples of *Sphagnum compactum* were included in the study: 24 from Norway, 12 from Austria, 5 from Germany, and 10 from Slovakia. Raw sequencing data for 13 samples (8 from Norway, 3 from Austria, and 2 from Germany) was generated by Meleshko et al. (2021). The 28 additional samples were selected based on a pilot study using microsatellite markers following the method by Kyrkjeeide (2016b), the goal of the pilot study was to identify samples that genetically cluster with the existing samples to increase the reliability of the results. The geographical distribution of the samples is shown in fig. 1 (see Appendix A Table A-1 for information about the specimens).

Figure 1. Sampling localities for *Sphagnum compactum* individuals (each population indicated with a different shape) included in the study.

DNA extraction, Library Preparation and Sequencing

Dried capitula tissue was lysed using a Qiagen Tissues Lyser II (Qiagen) for 120s at 30Hz, and total DNA was extracted following the protocol (with minor modifications) for dried plant samples for the E.Z.N.A.® HP Plant DNA Mini Kit (Omega Bio-tek). DNA concentration of all extracts was measured using a Qubit 2.0 Fluorometer (Thermo Fisher Scientific). The extracted DNA was fragmented to a target length of 400 bp using a Biorupter® Pico (Diagenode).

Genomic libraries were built following the BEST (Blunt-End-Singel-Tube) protocol presented in Carøe et al., (2018). The resulting libraries were quality-checked using Qubit (Thermo Fisher Scientific) and TapeStation (Agilent) and pooled equimolarly into two pools. DNA sequencing was performed on Illumina NovaSeq platform in 150-bp paired-end mode at Novogene (UK).

Sequencing Data Processing

The raw sequencing data were processed using the Paleomix pipeline v.1.3.7 (Schubert et al., 2014). Adapter contamination was trimmed using AdapterRemoval v2.3.3 (Schubert et al., 2016), trimmed reads shorter than 25 bases were discarded. The trimmed reads were mapped to a reference genome assembly of *Sphagnum angustifolium* (C.E.O.Jensen ex Russow) C.E.O.Jensen (Healey et al., 2023) using the mem algorithm of BWA v0.7.17 (Li and Durbin, 2009). Reads with a mapping quality below 30 and unmapped reads were discarded during mapping. SAMtools v1.16.1 (Danecek et al., 2021) were used to index and sort the resulting BAM files. Picard tools v2.27 ("Picard toolkit," 2019) were used to mark PCR duplicates and validate the BAM files. The reads with inserts length below 50 were discarded.

SNP Calling and Filtering

Using GATK v4.1.9.0 (McKenna et al., 2010), SNPs and indels were called for each sample with *HaplotypeCaller* using the "-ploidy 1" argument. Using *CombineGVCFs*, the resulting per-sample files were combined into one multi-sample file, which was genotyped using *GenotypeGVCFs*. Based on the GATK's best practices pipeline (Van der Auwera and O'Connor, 2020), the SNPs meeting any of the following criteria were excluded from the resulting variant call format (VCF) file: QualByDepth < 2.0, FisherStrand > 60.0, RMSMappingQuality < 40.0, MappingQualityRankSumTest < -12.5, ReadPosRankSumTest < -8.0. Using VCFtools (Danecek et al., 2011), the SNPs were further filtered out based on the per-individual depth (<5 and >40), sample size (<¾ of individuals), and minor allele frequency (<0.05). The resulting dataset hereafter is called the 'nuclear SNP dataset'.

Principal Component Analysis

To visualise genetic differences between the sampled populations, a principal component analysis (PCA) was carried out using ANGSD v0.941-6-g67b6b3b (Korneliussen et al., 2014) by calculating a covariance matrix based on randomly sampled bases from each individual at each position. First, per-Base Alignment Quality (BAQ) was calculated (with the "-baq 1" option in ANGSD) to adjust quality scores around indels (Li, 2011), and the mapping quality (MAPQ) score was adjusted to 50 for excessive mismatches. Then, bad reads (with flag above 255), reads with low MAPQ score (<30), low base quality (<20), or that did not map uniquely were discarded. The same read filtering as described here was also applied in other analyses in ANGSD (hereafter 'ANGSD read filtering'). Then, sites were filtered to a per-individual depth between 5 and 40. Major and minor alleles were inferred by counting bases across individuals and keeping the two most frequent bases, and then inferring the least frequent of the kept bases as minor allele (Li et al., 2010). Then, sites were filtered based on minor allele frequency (<0.05) and sample size (<¾ of individuals). The resulting covariance matrix was loaded into R 4.3.0 (R Core Team, 2023), where the function *eigen* was used to compute eigenvalues and eigenvectors. To determine the significance of the eigenvalues, a Tracy-Wisdom test was performed (Patterson et al., 2006) using the package "AssocTest" (Wang et al., 2020).

Admixture

To infer population structure and admixture, the program ADMIXTURE v1.3 (Alexander et al., 2009) was used. The model implemented in ADMIXTURE does not explicitly take linkage disequilibrium into consideration, hence the nuclear SNP dataset was pruned using the "`--thin`" option of VCFtools (Danecek et al., 2011) so that the distance between each SNP is ≥2000 bp. PLINK v1.90b7 (Chang et al., 2015) was used to convert the variant file into a binary PLINK file which was used as input for ADMIXTURE. To select the best number of genetic clusters (*K)*, 10-fold cross-validation was used testing K=1 to K=6. For each K, 10 independent runs were performed, and mean cross-validation error across all the runs for each K was calculated to identify the best K. The ancestry coefficient matrix from ADMIXTURE for the best K was loaded into R 4.3.0 (R Core Team, 2023) and visualised.

F_{ST} analysis

Genetic differentiation was measured by the fixation index (F_{ST}). Pairwise F_{ST} was calculated for each population pair, and for each pair of the observed genetic clusters (identified with the PCA and ADMIXTURE analysis) using *RealSFS* in ANGSD v0.941-6-g67b6b3b (Korneliussen et al., 2014) based on two-dimensional site frequency spectrum (2D SFS). First, ANGSD read filtering was performed. Then, the sites were filtered based on

per-individual depth (2-40) and sample size (<¾ of individuals). The site allele frequency (SAF) likelihood was estimated assuming Hardy-Weinberg equilibrium and based on individual genotype likelihoods estimated with the GATK likelihood model (McKenna et al., 2010). To indicate that the data was haploid, the "-isHap 1" option was used. Then, a folded maximum likelihood estimate of the 2D SFS was obtained for each population/genetic cluster pair, from which pairwise F_{ST} was estimated using an extended version of the method-of-moments estimator (Reynolds et al., 1983) as implemented in ANGSD (Fumagalli et al., 2013).

Phylogenetic analysis

Phylogenetic relationships between the sampled populations were inferred using the maximum likelihood (ML) software RAxML v8.2.12 (Stamatakis, 2014). Using BCFtools v1.5 (Danecek et al., 2021), the nuclear SNP dataset was merged with a VCF file containing SNPs of *Sphagnum lindbergii* Schimp. that was obtained in the same manner as the nuclear dataset described above. The resulting VCF file was converted into a multisample FASTA file using a custom Python script and SeqKit (Shen et al., 2016). RAxML was used with 200 rapid bootstrap inferences followed by 20 ML searches under the generalised time reversible model for nucleotide substitution and the gamma model of rate heterogeneity (GTRGAMMA).

RAxML was used with the same settings and models on a chloroplast dataset. The dataset was obtained by mapping raw sequencing data from all samples and the outgroup to a reference genome assembly of *S. compactum*'s chloroplast genome (GenBank accession number KU725453.1) using the same approach as described above in "Sequencing Data Processing". Then, ANGSD v0.941-6-g67b6b3b (Korneliussen et al., 2014) was used to convert the mapped reads to per-sample FASTA files. First, ANGSD read filtering was performed. Then, sites were filtered based on per-individual depth (100-3000). Finally, FASTA files were produced by using the base with the highest effective depth (-doFasta 3). The individual FASTA files were concatenated into a multisample FASTA file using SeqKit (Shen et al., 2016).

The resulting best-scoring ML trees were visualised using FigTree v1.4.4.

Demographic history

To infer the changes in demographic history of the studied populations, I used the pairwise sequentially Markovian coalescent method implemented in the software package PSMC (Li and Durbin, 2011). The method infers the effective population size (N_e) through time based on the patterns of heterozygosity across regions of a single diploid genome (Li and Durbin, 2011). In haploids, PSMC can be used to estimate divergence time between populations

(Cahill et al., 2016). To achieve this, pseudo-diploid sequences were created for all pairs of populations using the sample with the highest sequencing coverage per population (>6x, see Appendix A, Table A-2). First, the mapped reads were filtered for base (<20) and mapping quality (<30) and converted into genotype likelihoods using the 'mpileup' tool by SAMtools v1.14 (Danecek et al., 2021) . Then, the variants were called using 'call' by BCFtools v1.5 (Danecek et al., 2021) with the "-c" and "-ploidy 1" options. The resulting VCF file was converted to a FASTQ-file using PSMC's utility "vcfutils.pl" with "vcf2fq" option keeping the sites with per-individual depth of 10-50. For each pair, the resulting files were merged using Seqtk v1.3 (Li, 2023) and converted to a PSMC input file (psmcfa) using PSMC's utility "fq2psmcfa" (Li and Durbin, 2011). Then, PSMC was run on each pair with the following settings: the upper limit for TMRCA set to 10 (-t option), initial θ/ρ value set to 5 (-r option), and standard atomic time intervals (4+25*2+4+6). The results were visualised using PSMC's utility "psmc_plot.pl", assuming a generation time of 5 years and mutation rate of $1.0*10^{-8}$ per site per generation ($2.0*10^{-9}$ per site per year) (Linde et al., 2021).

Results

Sequencing Summary, Mapping, SNP Calling and Filtering

An average of 112±56M (SD) reads per sample was retained after quality filtering of the raw sequencing reads, 22±12% (SD) of the reads mapped to the *S. angustifolium* reference genome, and 22±8% (SD) of the reads that mapped to the *S. angustifolium* reference genome were PCR duplicates. Coverage varied from 1.18 to 21.22 (mean coverage was 4.64±3.59 SD). After SNP calling and filtering, the nuclear SNP dataset contained 673K SNPs.

PCA

After filtering the raw dataset, the PCA analysis was based on 265,417 sites. The first four principal components were statistically significant (P<0.001, see Appendix B fig. B-1). The first axis (PC1) explains 30.34% of the variation, while the second axis (PC2) explains 16.92% of the variation (fig. 2 A). PC1 separates 10 of the Austrian samples (the 'blue cluster') from the remaining samples (including two Austrian samples), and PC2 separates the German samples from the other samples. The third axis (PC3) explains 4.48% of the variation, and reveals additional variation among the 10 Austrian samples within the 'blue cluster' (fig. 2 B). The Norwegian, Slovakian, and two Austrian (LI89 and UI91) samples cluster together and show less variation compared to the other clusters.

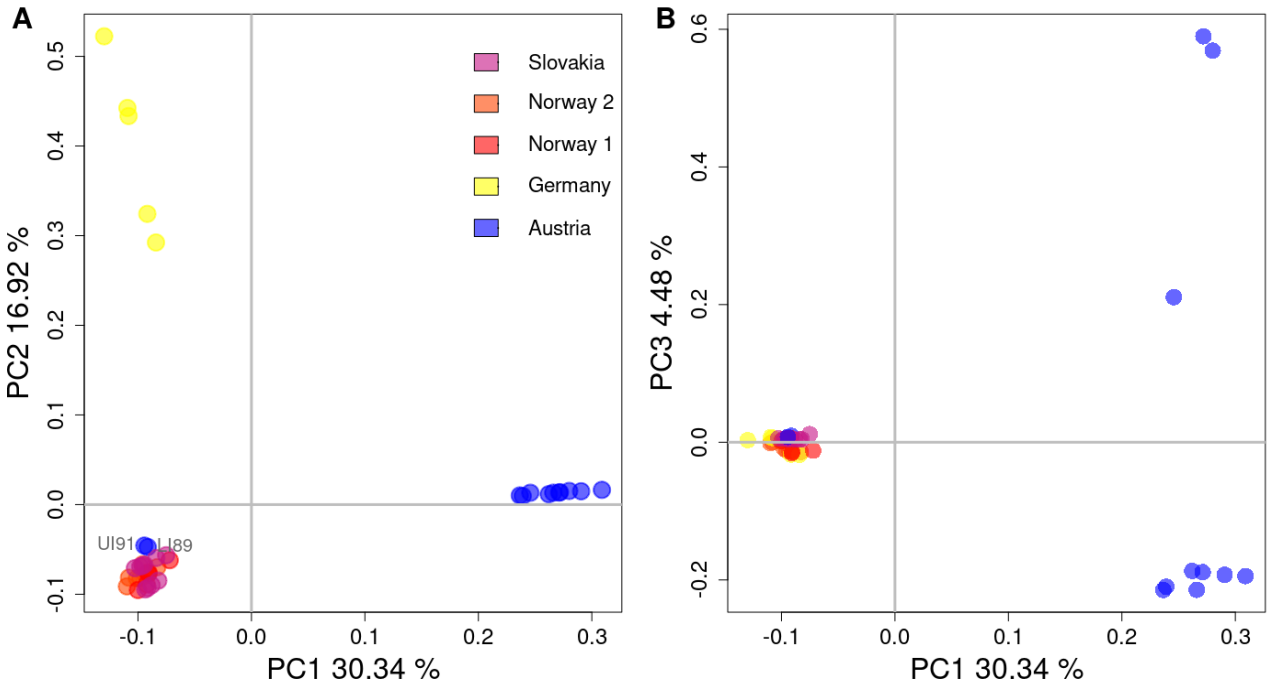

Figure 2. Principal component analysis of 41 *Sphagnum compactum* individuals, samples are coloured according to sampling location. **A.** All individuals in the space of the first two principal components and **B.** in the space of the first and the third principal components.

Admixture

The pruned nuclear SNP dataset used in the admixture analysis contained 69,182 SNPs. The runs with *K*=3 had the lowest mean cross-validation error across all runs, and the run with the lowest cross-validation error which equaled 0.339 (Appendix B fig. B-2). All German samples are assigned to one cluster (‘yellow cluster’), while Norwegian and Slovakian samples are assigned to a second cluster (‘red cluster’) (fig. 3). Two Austrian samples (LI89 and UI91) are admixed, with most of the ancestry being assigned to the ‘red cluster’ and a small part (~7%) to the ‘yellow cluster’. The remaining Austrian samples are assigned to their own cluster (‘blue cluster’).

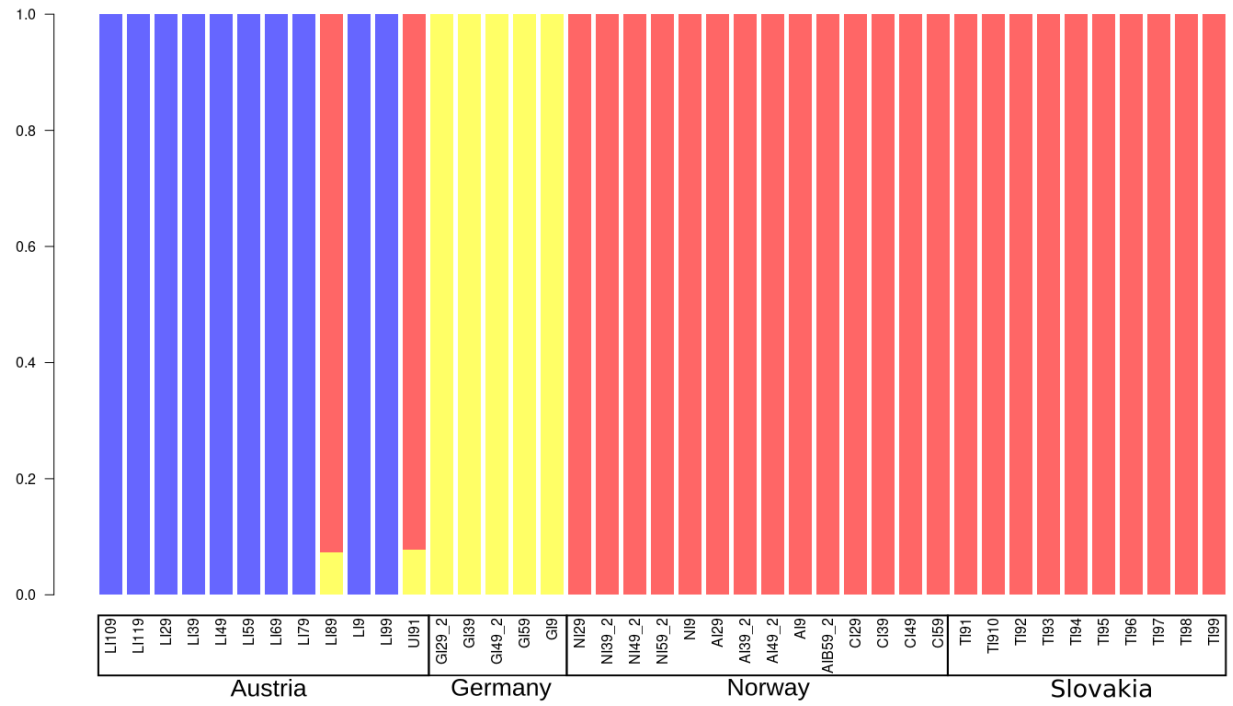

Figure 3. Barplot showing the results of ADMIXTURE analysis with K=3 for 41 *Sphagnum compactum* samples.

F_{ST} analysis

The F_{ST} values among the genetic clusters vary from 0.80 to 0.90 (Table 1), indicating that the clusters are well-differentiated. The pairwise F_{ST} estimates based on sample location vary from 0.22 to 0.86 (Table 2). The lower F_{ST} values (0.22-0.32) between the Norwegian and Slovakian samples is in accordance with the PCA and ADMIXTURE results.

Table 1. Genome-wide pairwise weighted F_{ST} among observed genetic clusters

	Blue	Yellow	Red
Blue			
Yellow	**0.90**		
Red	**0.87**	**0.80**	

Table 2. Global pairwise weighted F_{ST} among five European *S. compactum* populations

	Austria	Germany	Norway 1	Norway 2	Slovakia
Austria					
Germany	**0.73**				
Norway 1	**0.69**	**0.86**			
Norway 2	**0.71**	**0.82**	**0.22**		
Slovakia	**0.72**	**0.85**	**0.32**	**0.29**	

Phylogenetic analysis

The phylogenetic relationships between the sampled specimens reconstructed using RAxML based on nuclear (664,013 alignment patterns) and chloroplast (4,380 alignment patterns) markers are incongruent. The ML tree based on the nuclear markers show high support (bootstrap values >80) for two clades, one clade containing all German and Austrian samples in addition to five Slovakian samples, the second clade containing the remaining samples (fig. 4). Thus, only the blue genetic clusters identified in the PCA and ADMIXTURE analysis represent a monophyletic clade. The branch leading to the blue cluster is very long compared to other branches in the tree (fig. 5).

In the ML tree based on chloroplast markers, the three genetic clusters form well supported clades, except for one Slovakian sample (TI93) which is placed in the German clade (fig. 6). The red cluster is inferred as the sister clade to the blue and yellow cluster.

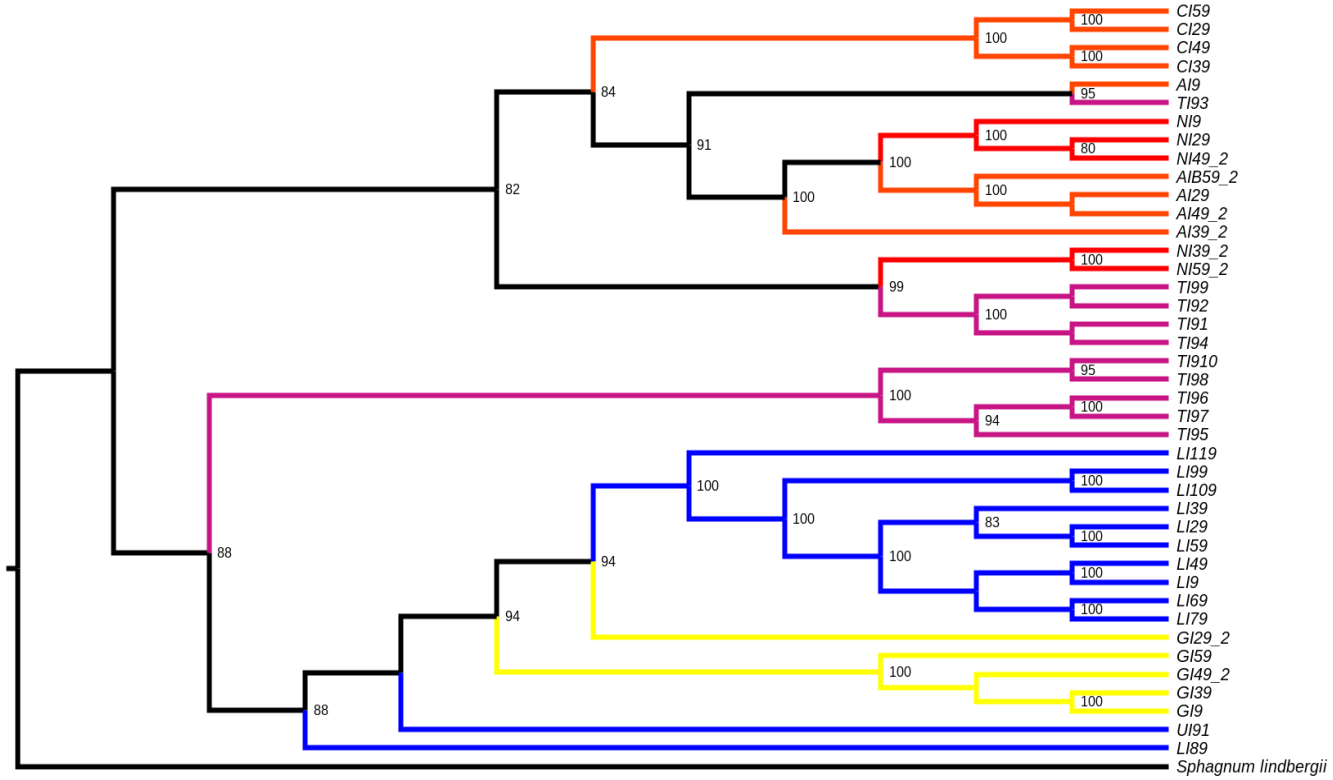

Figure 4. Phylogenetic tree for *Sphagnum compactum* sampled from five European populations based on nuclear SNPs (673K SNPs). Bootstrap support values are printed at the nodes. Branches are transformed. Branches are coloured according to sampling location blue (Austria), yellow (Germany), red/orange (Norway), and magenta (Slovakia).

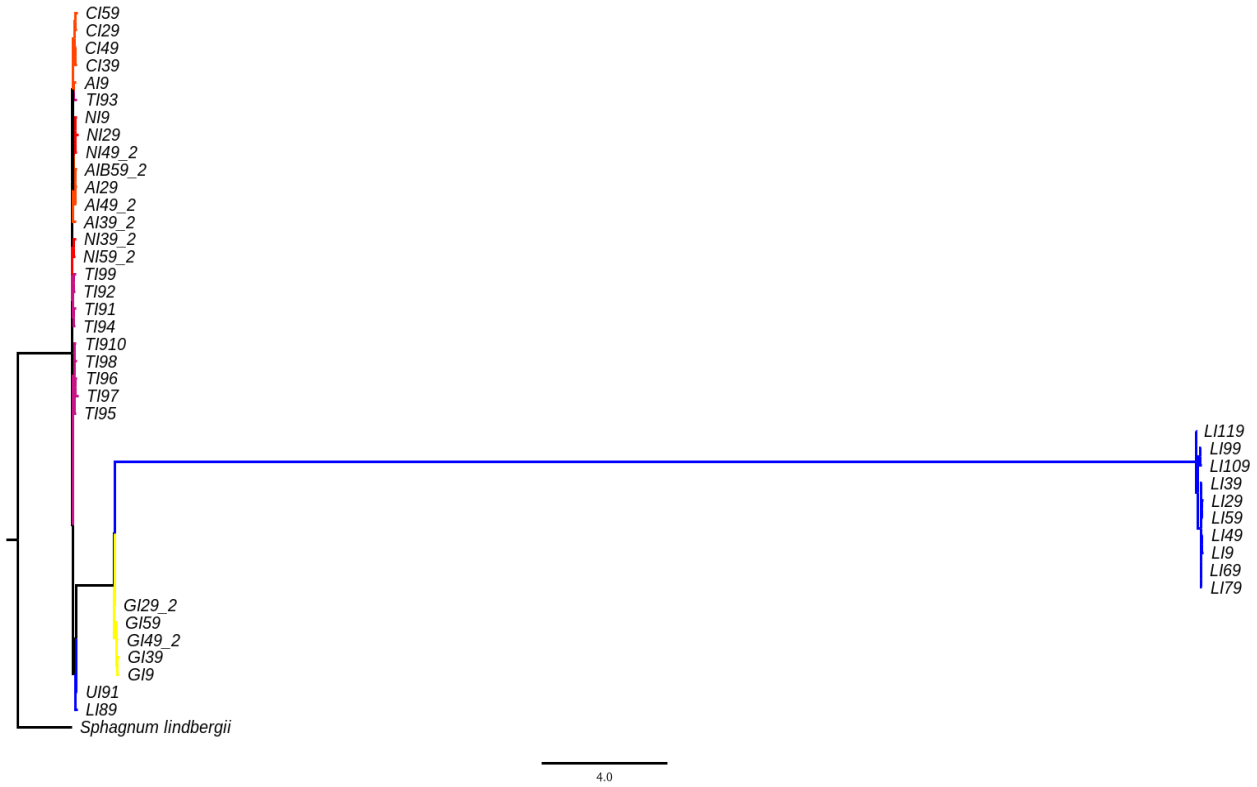

Figure 5. Phylogenetic tree for *Sphagnum compactum* sampled from five European populations based on nuclear SNPs (673K SNPs). Branches are untransformed with branch lengths given in average number of nucleotide substitutions per site. Branches are coloured according to sampling location blue (Austria), yellow (Germany), red/orange (Norway), and magenta (Slovakia).

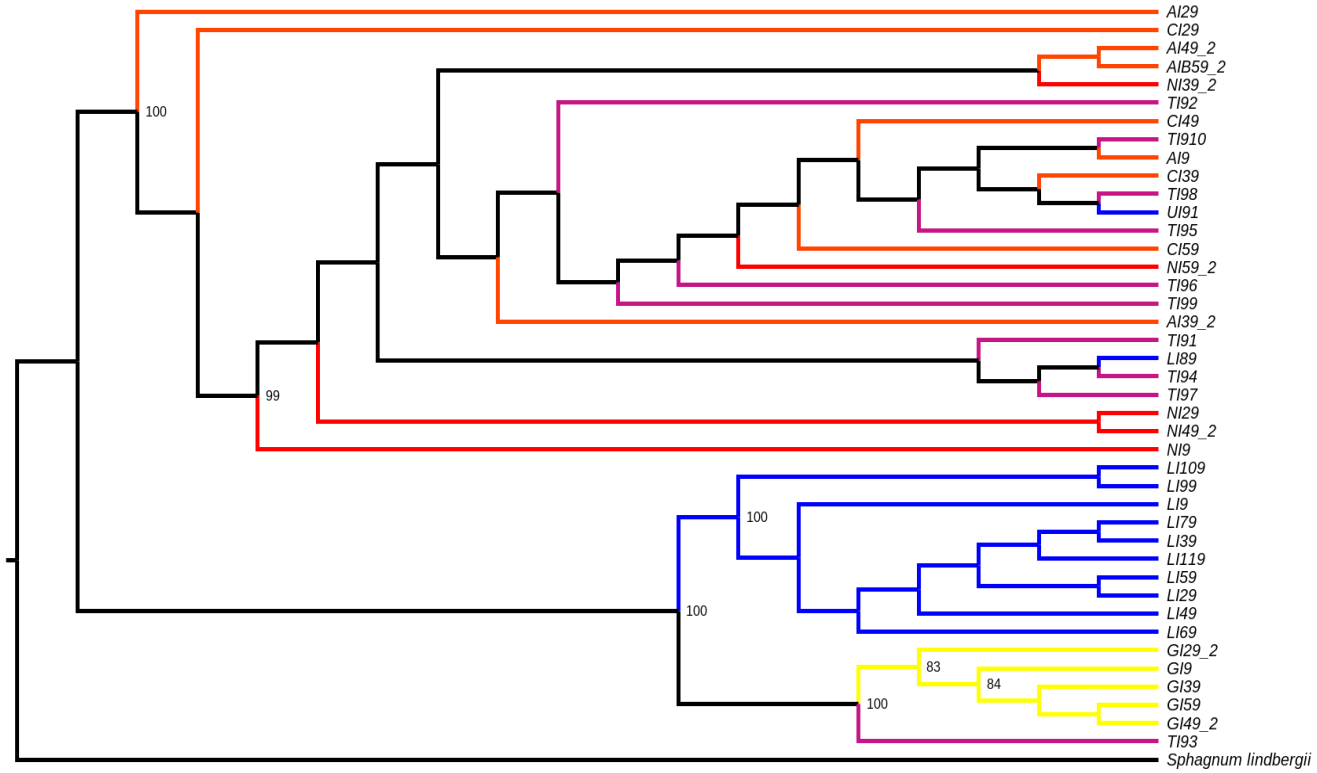

Figure 6. Phylogenetic tree for *Sphagnum compactum* sampled from five European populations based on chloroplast markers. Bootstrap support values are printed at the nodes. Branches are transformed. Branches are coloured according to sampling location blue (Austria), yellow (Germany), red/orange (Norway), and magenta (Slovakia).

Demographic history

The estimate of N_e will tend towards infinite after the split of the two populations when the PSMC model is applied to a pseudo-diploid genome sequence made from haploid sequences from different populations (Cahill et al., 2016). PSMC plots of four pseudo-diploids made by combining an Austrian sample with a sample from each of the other four sampled populations indicate that the Austrian population started to diverge from the other populations more than 1 million years ago (Mya) based on the assumed mutation rate and generation time (fig. 7). The split between the German population and the Norwegian populations started 110-120 thousand years ago (Kya). The N_e of the German-Slovakian pseudo-diploid starts to increase rapidly around 110 Kya, indicating reduction in gene flow, however, it stops increasing at 100 Kya and then starts to reduce. The pseudo-diploids made by combining Norwegian and Slovakian samples show a sudden increase in N_e close to 10 Kya. The Austrian sample used in the PSMC analysis belongs to the 'blue' cluster, so the estimated splits can be interpreted as the split between clusters. The 'blue' cluster diverged first, followed by the divergence between the 'red' and 'yellow' cluster.

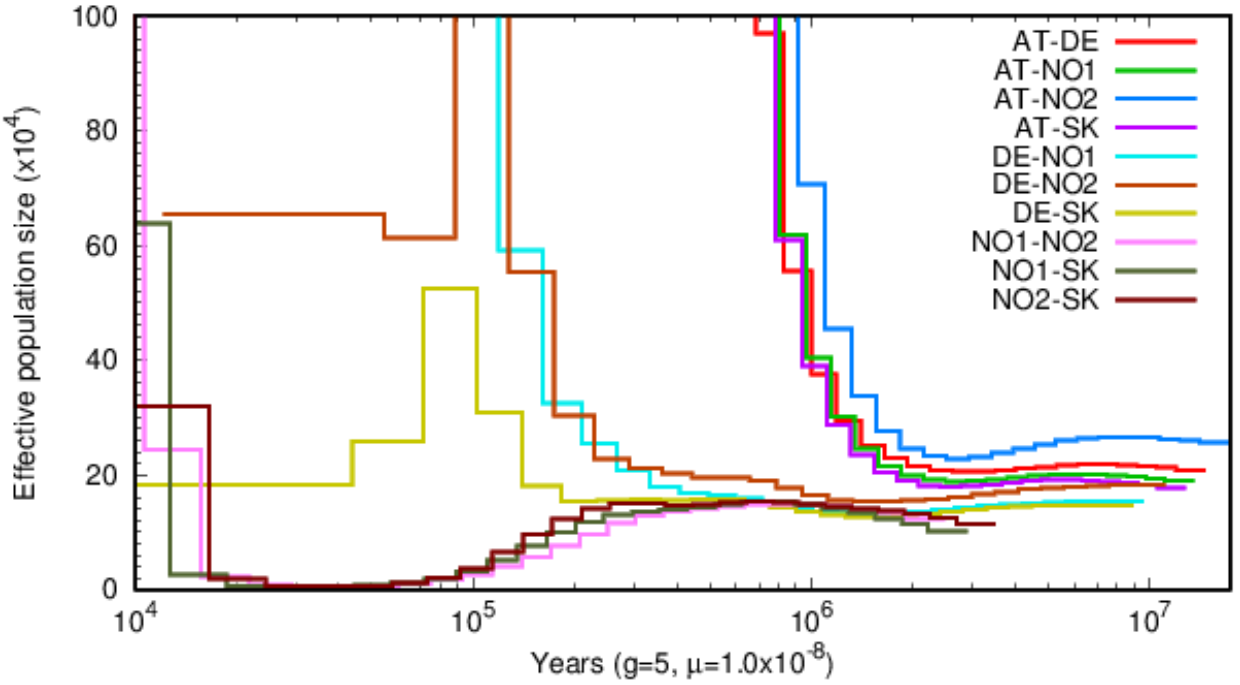

Figure 7. Effective population size (N_e) estimated with PSMC for pseudo diploids made by combining haploid sequences from five different populations. Each line represents a population pair. When N_e goes to infinity the population pair have diverged. The plots are scaled based on a generation

time of 5 years and a mutation rate of $1.0*10^{-8}$ per site per generation.
Key to population abbreviation: AT - Austria, DE - Germany, NO1 - Norway 1, NO2 - Norway 2, SK - Slovakia.

Discussion

Lately, genomic data has been used to discover spatial genetic structure in bryophyte species (Kyrkjeeide et al., 2016b; Vigalondo et al., 2019). This has resulted in species with wide geographical ranges being circumscribed as several separate species (Hassel et al., 2018; Shaw et al., 2023; Vigalondo et al., 2020). In this study, I confirmed that the three clades observed by Meleshko et al. (2021) represent three distinct genetic clusters in *Sphagnum compactum* sampled from five populations in Europe based on analysis of low coverage whole genome sequencing data. The genetic differentiation, low gene flow, and deep split times between the clusters could indicate that they represent separate species.

Genetic distinctiveness of the clusters

The number of genetic clusters in the data set was explored using PCA and ADMIXTURE analysis. Both analyses identified three distinct genetic clusters. The ADMIXTURE analysis indicates no recent gene flow between the clusters, except for two of the Austrian samples that show a small amount of admixture. ADMIXTURE is suited to detect recent gene flow because it does not fit a historical model (Alter et al., 2017). Thus, the inferred admixture could be explained by recent gene flow. However, the overall lack of admixture indicates that there are barriers to gene flow between the clusters and that the clusters could represent separate species.

The pairwise F_{ST} values among the genetic clusters are high (0.80-0.90), and highest when comparing the 'blue cluster' to the two others (0.87-0.90). This strongly suggests that the clusters are highly differentiated from one another. The F_{ST} values are higher than those reported for closely related species, such as *S. divinum* Flatberg & Hassel vs. *S. medium* Limpr. (Yousefi et al., 2017), *S. fuscum* (Schimp.) H.Klinggr. vs. *S. beothuk* R.E. Andrus (Kyrkjeeide et al., 2015), and *S. balticum* (Russow) C.E.O.Jensen vs. *S. tenellum* (Brid.) Brid. (Stenøien et al., 2011). However, one should keep in mind that estimates of F_{ST} vary depending on method for estimating and data type (Holsinger and Weir, 2009), and thus comparing results from different studies should be done with care. Genetic differentiation between *S. divinum* and *S. medium* which diverged from each other ~28,900 generations ago was estimated from genome-wide data (Yousefi et al., 2017), and is therefore comparable to my results. Also, Meleshko et al. (2021) estimated F_{ST} among *Sphagnum* species using the same methods as here, and the reported F_{ST} values among species pairs varied between 0.76 to 0.98. In turn, differentiation among morphologically defined morphs of *S. warnstorfii* was

estimated as 0.21, which is lower than inferred here (Yousefi et al., 2019). The combination of these results strongly suggest that the genetic cluster identified here represent genetically separate species.

Demographic history of the clusters

The PSMC analysis of the pseudo-diploids estimated that the 'blue cluster' diverged first, followed by the divergence between the 'red' and 'yellow' cluster. But, two of the plots ('DE-NO2' and 'DE-SK') plateaued after an increase in N_e, which can be interpreted as there being some gene flow between the clusters after initial separation (Sato et al., 2020).

The results of a PSMC analysis need to be scaled based on estimates of generation time and mutation rate. Erroneous estimates will affect how the PSMC-plot scales along the axes, but not the principal shape of the curve (Nadachowska-Brzyska et al., 2016). Thus, the timing of events will change when changing estimates of mutation rate, but the order will stay intact. The assumed short generation time (5 years) used in this study is an educated guess based on the fact that *S. compactum* is a pioneer species of disturbed peat and frequently found with spores. The mutation rate ($2.0*10^{-9}$ per site per year) used was estimated by Linde et al. (2021) based on several moss lineages with differing life histories. Mutation rate is linked to life span (Linde et al., 2021; Thomas et al., 2010). Thus, it is uncertain how precise the estimated mutation rate is for *S. compactum* and the estimated split times could be biased. However, the order of population splitting events is not affected since the principal shape of the curve is not affected by generation time or mutation rate. Combined with the other results, especially the long branch leading to the 'blue' cluster in the nuclear phylogeny, the splitting order of the genetic clusters makes sense.

Population structure and demographic history of extant northern species is affected by LGM (~20 Kya) (Hewitt, 2004; Stewart et al., 2010; Taberlet et al., 1998). However, even with some uncertainty related to the estimated split times in *S. compactum*, the observed differentiation among the clusters are much greater than differentiation linked to LGM in bryophytes (Kyrkjeeide et al., 2014). Thus, it seems likely that the splits between the clusters happened before LGM. There are three main routes of recolonization in bryophytes post LGM: western, southern, and eastern route (Kyrkjeeide et al., 2014). The sampling in this study is not broad enough to conclude where the clusters could have survived, but the presence of the 'red' cluster in Norway and Slovakia could indicate an eastern refugia and recolonization route.

Implications for bryophyte biogeography

The sampling in this study is limited to five populations across Europe, thus, geographic range is not wide enough to say anything conclusive about the distribution ranges of genetic

clusters. However, the results show that there is considerable genetic structure within *S. compactum*, adding to the increasing amount of evidence that widely distributed peatmosses are in fact genetically structured (Kyrkjeeide et al., 2016b, 2016a). A long standing view in bryophyte biogeography is that bryophyte species have wide distribution ranges and low speciation rates (Patiño and Vanderpoorten, 2018). Vigalando et al. (2019) challenged this view in their study of the *Lewinskya affinis* complex, where they discovered seven species with narrow distribution ranges, and hypothesise that the low levels of endemism documented in bryophytes could be due to limitations in recognising species within plants with reduced morphology. Thus, my results add to the growing number of bryophyte species with wide distribution ranges that are being split up based on molecular evidence (Escolástico-Ortiz et al., 2023; Fuselier et al., 2009; Hedenäs, 2008; Kyrkjeeide et al., 2016b; Vigalondo et al., 2019). Although this study doesn't address the taxonomic status of *S. compactum*, the results indicate that *S. compactum* could potentially include several species. Future studies addressing the taxonomic status of *S. compactum* should employ an integrative approach, combining morphological and molecular data, to reveal if the clusters are morphologically different.

References

Alexander, D.H., Novembre, J., Lange, K., 2009. Fast model-based estimation of ancestry in unrelated individuals. Genome Res. 19, 1655–1664. https://doi.org/10.1101/gr.094052.109

Alsos, I.G., Eidesen, P.B., Ehrich, D., Skrede, I., Westergaard, K., Jacobsen, G.H., Landvik, J.Y., Taberlet, P., Brochmann, C., 2007. Frequent Long-Distance Plant Colonization in the Changing Arctic. Science 316, 1606–1609. https://doi.org/10.1126/science.1139178

Alter, S.E., Munshi-South, J., Stiassny, M.L.J., 2017. Genomewide SNP data reveal cryptic phylogeographic structure and microallopatric divergence in a rapids-adapted clade of cichlids from the Congo River. Mol. Ecol. 26, 1401–1419. https://doi.org/10.1111/mec.13973

Butlin, R.K., Galindo, J., Grahame, J.W., 2008. Sympatric, parapatric or allopatric: the most important way to classify speciation? Philos. Trans. R. Soc. B Biol. Sci. 363, 2997–3007. https://doi.org/10.1098/rstb.2008.0076

Cahill, J.A., Soares, A.E.R., Green, R.E., Shapiro, B., 2016. Inferring species divergence times using pairwise sequential Markovian coalescent modelling and low-coverage genomic data. Philos. Trans. R. Soc. B Biol. Sci. 371, 20150138. https://doi.org/10.1098/rstb.2015.0138

Carøe, C., Gopalakrishnan, S., Vinner, L., Mak, S.S.T., Sinding, M.H.S., Samaniego, J.A., Wales, N., Sicheritz-Pontén, T., Gilbert, M.T.P., 2018. Single-tube library preparation for degraded DNA. Methods Ecol. Evol. 9, 410–419. https://doi.org/10.1111/2041-210X.12871

Chang, C.C., Chow, C.C., Tellier, L.C., Vattikuti, S., Purcell, S.M., Lee, J.J., 2015. Second-generation PLINK: rising to the challenge of larger and richer datasets. GigaScience 4, s13742-015-0047–8. https://doi.org/10.1186/s13742-015-0047-8

Cronberg, N., 1992. Reproductive Biology of Sphagnum. Lindbergia 17, 69–82.

Danecek, P., Auton, A., Abecasis, G., Albers, C.A., Banks, E., DePristo, M.A., Handsaker, R.E., Lunter, G., Marth, G.T., Sherry, S.T., McVean, G., Durbin, R., Genomes Project Analysis Group, 2011. The variant call format and VCFtools. Bioinformatics 27, 2156–2158. https://doi.org/10.1093/bioinformatics/btr330

Danecek, P., Bonfield, J.K., Liddle, J., Marshall, J., Ohan, V., Pollard, M.O., Whitwham, A., Keane, T., McCarthy, S.A., Davies, R.M., Li, H., 2021. Twelve years of SAMtools and BCFtools. GigaScience 10, giab008. https://doi.org/10.1093/gigascience/giab008

Daniels, R.E., 1985. Isozyme Variation in Finnish and British Populations of *Sphagnum compactum*. Ann. Bot. Fenn. 22, 275–279.

Duffy, A.M., Ricca, M., Robinson, S., Aguero, B., Johnson, M.G., Stenøien, H.K., Flatberg, K.I., Hassel, K., Shaw, A.J., 2022. Heterogeneous genetic structure in eastern North American peat mosses (*Sphagnum*). Biol. J. Linn. Soc. 135, 692–707. https://doi.org/10.1093/biolinnean/blab175

Escolástico-Ortiz, D.A., Hedenäs, L., Quandt, D., Harpke, D., Larraín, J., Stech, M., Villarreal A, J.C., 2023. Cryptic speciation shapes the biogeographic history of a northern distributed moss. Bot. J. Linn. Soc. 201, 114–134. https://doi.org/10.1093/botlinnean/boac027

Fumagalli, M., Vieira, F.G., Korneliussen, T.S., Linderoth, T., Huerta-Sánchez, E., Albrechtsen, A., Nielsen, R., 2013. Quantifying Population Genetic Differentiation from Next-Generation Sequencing Data. Genetics 195, 979–992. https://doi.org/10.1534/genetics.113.154740

Fuselier, L., Davison, P.G., Clements, M., Shaw, B., Devos, N., Heinrichs, J., Hentschel, J.,

Sabovljevic, M., Szövényi, P., Schuette, S., Hofbauer, W., Shaw, A.J., 2009. Phylogeographic analyses reveal distinct lineages of the liverworts *Metzgeria furcata* (L.) Dumort. and *Metzgeria conjugata* Lindb. (Metzgeriaceae) in Europe and North America. Biol. J. Linn. Soc. 98, 745–756. https://doi.org/10.1111/j.1095-8312.2009.01319.x

Hassel, K., Kyrkjeeide, M.O., Yousefi, N., Prestø, T., Stenøien, H.K., Shaw, J.A., Flatberg, K.I., 2018. *Sphagnum divinum* (sp. nov.) and *S. medium* Limpr. and their relationship to *S. magellanicum* Brid. J. Bryol. 40, 197–222. https://doi.org/10.1080/03736687.2018.1474424

Healey, A.L., Piatkowski, B., Lovell, J.T., Sreedasyam, A., Carey, S.B., Mamidi, S., Shu, S., Plott, C., Jenkins, J., Lawrence, T., Aguero, B., Carrell, A.A., Nieto-Lugilde, M., Talag, J., Duffy, A., Jawdy, S., Carter, K.R., Boston, L.-B., Jones, T., Jaramillo-Chico, J., Harkess, A., Barry, K., Keymanesh, K., Bauer, D., Grimwood, J., Gunter, L., Schmutz, J., Weston, D.J., Shaw, A.J., 2023. Newly identified sex chromosomes in the *Sphagnum* (peat moss) genome alter carbon sequestration and ecosystem dynamics. Nat. Plants 1–17. https://doi.org/10.1038/s41477-022-01333-5

Hedenäs, L., 2008. Molecular variation and speciation in *Antitrichia curtipendula* s.l. (Leucodontaceae, Bryophyta). Bot. J. Linn. Soc. 156, 341–354. https://doi.org/10.1111/j.1095-8339.2007.00775.x

Hewitt, G.M., 2004. Genetic consequences of climatic oscillations in the Quaternary. Philos. Trans. R. Soc. B Biol. Sci. 359, 183–195.

Hewitt, G.M., 1996. Some genetic consequences of ice ages, and their role in divergence and speciation. Biol. J. Linn. Soc. 58, 247–276. https://doi.org/10.1006/bijl.1996.0035

Holsinger, K.E., Weir, B.S., 2009. Genetics in geographically structured populations: defining, estimating and interpreting FST. Nat. Rev. Genet. 10, 639–650. https://doi.org/10.1038/nrg2611

Johnson, M.G., Shaw, A.J., 2015. Genetic diversity, sexual condition, and microhabitat preference determine mating patterns in *Sphagnum* (Sphagnaceae) peat-mosses. Biol. J. Linn. Soc. 115, 96–113. https://doi.org/10.1111/bij.12497

Klein, E.K., Lavigne, C., Gouyon, P.-H., 2006. Mixing of propagules from discrete sources at long distance: comparing a dispersal tail to an exponential. BMC Ecol. 6, 3. https://doi.org/10.1186/1472-6785-6-3

Korneliussen, T.S., Albrechtsen, A., Nielsen, R., 2014. ANGSD: Analysis of Next Generation Sequencing Data. BMC Bioinformatics 15, 356. https://doi.org/10.1186/s12859-014-0356-4

Kyrkjeeide, M.O., Hassel, K., Flatberg, K.I., Shaw, A.J., Brochmann, C., Stenøien, H.K., 2016a. Long-distance dispersal and barriers shape genetic structure of peatmosses (*Sphagnum*) across the Northern Hemisphere. J. Biogeogr. 43, 1215–1226. https://doi.org/10.1111/jbi.12716

Kyrkjeeide, M.O., Hassel, K., Flatberg, K.I., Shaw, A.J., Yousefi, N., Stenøien, H.K., 2016b. Spatial Genetic Structure of the Abundant and Widespread Peatmoss *Sphagnum magellanicum* Brid. PLOS ONE 11, e0148447. https://doi.org/10.1371/journal.pone.0148447

Kyrkjeeide, M.O., Hassel, K., Stenøien, H.K., Prestø, T., Boström, E., Shaw, A.J., Flatberg, K.I., 2015. The dark morph of *Sphagnum fuscum* (Schimp.) H.Klinggr. in Europe is conspecific with the North American *S. beothuk*. J. Bryol. 37, 251–266. https://doi.org/10.1179/1743282015Y.0000000020

Kyrkjeeide, M.O., Stenøien, H.K., Flatberg, K.I., Hassel, K., 2014. Glacial refugia and post-glacial colonization patterns in European bryophytes. Lindbergia 37, 47–59, 13.

Li, H., 2023. seqtk.

Li, H., 2011. Improving SNP discovery by base alignment quality. Bioinformatics 27, 1157–1158. https://doi.org/10.1093/bioinformatics/btr076
Li, H., Durbin, R., 2011. Inference of human population history from individual whole-genome sequences. Nature 475, 493–496. https://doi.org/10.1038/nature10231
Li, H., Durbin, R., 2009. Fast and accurate short read alignment with Burrows–Wheeler transform. Bioinformatics 25, 1754–1760. https://doi.org/10.1093/bioinformatics/btp324
Li, Y., Vinckenbosch, N., Tian, G., Huerta-Sanchez, E., Jiang, T., Jiang, H., Albrechtsen, A., Andersen, G., Cao, H., Korneliussen, T., Grarup, N., Guo, Y., Hellman, I., Jin, X., Li, Q., Liu, J., Liu, X., Sparsø, T., Tang, M., Wu, H., Wu, R., Yu, C., Zheng, H., Astrup, A., Bolund, L., Holmkvist, J., Jørgensen, T., Kristiansen, K., Schmitz, O., Schwartz, T.W., Zhang, X., Li, R., Yang, H., Wang, Jian, Hansen, T., Pedersen, O., Nielsen, R., Wang, Jun, 2010. Resequencing of 200 human exomes identifies an excess of low-frequency non-synonymous coding variants. Nat. Genet. 42, 969–972. https://doi.org/10.1038/ng.680
Linde, A.-M., Eklund, D.M., Cronberg, N., Bowman, J.L., Lagercrantz, U., 2021. Rates and patterns of molecular evolution in bryophyte genomes, with focus on complex thalloid liverworts, *Marchantiopsida*. Mol. Phylogenet. Evol. 165, 107295. https://doi.org/10.1016/j.ympev.2021.107295
M. C. Urban, G. Bocedi, A. P. Hendry, J. -B. Mihoub, G. Pe'er, A. Singer, J. R. Bridle, L. G. Crozier, L. De Meester, W. Godsoe, A. Gonzalez, J. J. Hellmann, R. D. Holt, A. Huth, K. Johst, C. B. Krug, P. W. Leadley, S. C. F. Palmer, J. H. Pantel, A. Schmitz, P. A. Zollner, J. M. J. Travis, 2016. Improving the forecast for biodiversity under climate change. Science 353, aad8466. https://doi.org/doi:10.1126/science.aad8466
McKenna, A., Hanna, M., Banks, E., Sivachenko, A., Cibulskis, K., Kernytsky, A., Garimella, K., Altshuler, D., Gabriel, S., Daly, M., DePristo, M.A., 2010. The Genome Analysis Toolkit: A MapReduce framework for analyzing next-generation DNA sequencing data. Genome Res. 20, 1297–1303. https://doi.org/10.1101/gr.107524.110
Meleshko, O., Martin, M.D., Korneliussen, T.S., Schröck, C., Lamkowski, P., Schmutz, J., Healey, A., Piatkowski, B.T., Shaw, A.J., Weston, D.J., Flatberg, K.I., Szövényi, P., Hassel, K., Stenøien, H.K., 2021. Extensive Genome-Wide Phylogenetic Discordance Is Due to Incomplete Lineage Sorting and Not Ongoing Introgression in a Rapidly Radiated Bryophyte Genus. Mol. Biol. Evol. 38, 2750–2766. https://doi.org/10.1093/molbev/msab063
Morjan, C.L., Rieseberg, L.H., 2004. How species evolve collectively: implications of gene flow and selection for the spread of advantageous alleles. Mol. Ecol. 13, 1341–1356. https://doi.org/10.1111/j.1365-294X.2004.02164.x
Muñoz, J., Felicísimo, Á.M., Cabezas, F., Burgaz, A.R., Martínez, I., 2004. Wind as a Long-Distance Dispersal Vehicle in the Southern Hemisphere. Science 304, 1144–1147. https://doi.org/10.1126/science.1095210
Nadachowska-Brzyska, K., Burri, R., Smeds, L., Ellegren, H., 2016. PSMC analysis of effective population sizes in molecular ecology and its application to black-and-white *Ficedula* flycatchers. Mol. Ecol. 25, 1058–1072. https://doi.org/10.1111/mec.13540
Patiño, J., Vanderpoorten, A., 2018. Bryophyte Biogeography. Crit. Rev. Plant Sci. 37, 175–209. https://doi.org/10.1080/07352689.2018.1482444
Patterson, N., Price, A.L., Reich, D., 2006. Population Structure and Eigenanalysis. PLOS Genet. 2, e190. https://doi.org/10.1371/journal.pgen.0020190
Picard toolkit, 2019. . Broad Inst. GitHub Repos.
Provan, J., Bennett, K.D., 2008. Phylogeographic insights into cryptic glacial refugia. Trends

Ecol. Evol. 23, 564–571. https://doi.org/10.1016/j.tree.2008.06.010
R Core Team, 2023. R: A Language and Environment for Statistical Computing.
Reynolds, J., Weir, B.S., Cockerham, C.C., 1983. Estimation of the Coancestry Coefficient: Basis for a Short-Term Genetic Distance. Genetics 105, 767–779. https://doi.org/10.1093/genetics/105.3.767
Rydin, H., Jeglum, J.K., 2013. Productivity and peat accumulation, in: Rydin, H., Jeglum, J.K. (Eds.), The Biology of Peatlands. Oxford University Press, pp. 254–273. https://doi.org/10.1093/acprof:osobl/9780199602995.003.0012
Sætre, G.-P., Ravinet, M., 2019. Evolutionary genetics: concepts, analysis, and practice. University Press, Oxford.
Sato, Y., Ogden, R., Kishida, T., Nakajima, N., Maeda, T., Inoue-Murayama, M., 2020. Population history of the golden eagle inferred from whole-genome sequencing of three of its subspecies. Biol. J. Linn. Soc. 130, 826–838. https://doi.org/10.1093/biolinnean/blaa068
Schubert, M., Ermini, L., Sarkissian, C.D., Jónsson, H., Ginolhac, A., Schaefer, R., Martin, M.D., Fernández, R., Kircher, M., McCue, M., Willerslev, E., Orlando, L., 2014. Characterization of ancient and modern genomes by SNP detection and phylogenomic and metagenomic analysis using PALEOMIX. Nat. Protoc. 9, 1056–1082. https://doi.org/10.1038/nprot.2014.063
Schubert, M., Lindgreen, S., Orlando, L., 2016. AdapterRemoval v2: rapid adapter trimming, identification, and read merging. BMC Res. Notes 9, 88. https://doi.org/10.1186/s13104-016-1900-2
Shaw, A.J., Cox, C.J., Boles, S.B., 2003. Global patterns in peatmoss biodiversity. Mol. Ecol. 12, 2553–2570. https://doi.org/10.1046/j.1365-294X.2003.01929.x
Shaw, A.J., Devos, N., Cox, C.J., Boles, S.B., Shaw, B., Buchanan, A.M., Cave, L., Seppelt, R., 2010. Peatmoss (*Sphagnum*) diversification associated with Miocene Northern Hemisphere climatic cooling? Mol. Phylogenet. Evol. 55, 1139–1145. https://doi.org/10.1016/j.ympev.2010.01.020
Shaw, A.J., Nieto-Lugilde, M., Aguero, B., Duffy, A., Piatkowski, B.T., Jaramillo-Chico, J., Robinson, S., Hassel, K., Flatberg, K.I., Weston, D.J., Schuette, S., Hicks, K.A., 2023. *Sphagnum diabolicum* sp. nov. and *S. magniae* sp. nov.; morphological variation and taxonomy of the "*S. magellanicum* complex." The Bryologist 126, 69–89. https://doi.org/10.1639/0007-2745-126.1.069
Shaw, J., 2001. Biogeographic patterns and cryptic speciation in bryophytes. J. Biogeogr. 28, 253–261. https://doi.org/10.1046/j.1365-2699.2001.00530.x
Shen, W., Le, S., Li, Y., Hu, F., 2016. SeqKit: A Cross-Platform and Ultrafast Toolkit for FASTA/Q File Manipulation. PLOS ONE 11, e0163962. https://doi.org/10.1371/journal.pone.0163962
Slatkin, M., 1987. Gene Flow and the Geographic Structure of Natural Populations. Science 236, 787–792.
Stamatakis, A., 2014. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. Bioinformatics 30, 1312–1313. https://doi.org/10.1093/bioinformatics/btu033
Stenøien, H., Bakken, S., Flatberg, K.I., 1997. Phenotypic variation in the *Sphagnum recurvum* complex: a cultivation experiment. J. Bryol. 19, 731–750. https://doi.org/10.1179/jbr.1997.19.4.731
Stenøien, H.K., Hassel, K., Segreto, R., Gabriel, R., Karlin, E.F., Shaw, A.J., Flatberg, K.I., 2014. High morphological diversity in remote island populations of the peat moss *Sphagnum palustre*: glacial refugium, adaptive radiation or just plasticity? The Bryologist 117, 95–109. https://doi.org/10.1639/0007-2745-117.2.095

Stenøien, H.K., Shaw, A.J., Stengrundet, K., Flatberg, K.I., 2011. The narrow endemic Norwegian peat moss *Sphagnum troendelagicum* originated before the last glacial maximum. Heredity 106, 370–382. https://doi.org/10.1038/hdy.2010.96

Stewart, J.R., Lister, A.M., Barnes, I., Dalén, L., 2010. Refugia revisited: individualistic responses of species in space and time. Proc. R. Soc. B Biol. Sci. 277, 661–671. https://doi.org/doi:10.1098/rspb.2009.1272

Suzuki, H., 1965. Observations on *Sphagnum compactum* DC. in Japan. Hikobia 4, 303–317.

Svendsen, J.I., Alexanderson, H., Astakhov, V.I., Demidov, I., Dowdeswell, J.A., Funder, S., Gataullin, V., Henriksen, M., Hjort, C., Houmark-Nielsen, M., Hubberten, H.W., Ingólfsson, Ó., Jakobsson, M., Kjær, K.H., Larsen, E., Lokrantz, H., Lunkka, J.P., Lyså, A., Mangerud, J., Matiouchkov, A., Murray, A., Möller, P., Niessen, F., Nikolskaya, O., Polyak, L., Saarnisto, M., Siegert, C., Siegert, M.J., Spielhagen, R.F., Stein, R., 2004. Late Quaternary ice sheet history of northern Eurasia. Quat. Sci. Rev. 23, 1229–1271. https://doi.org/10.1016/j.quascirev.2003.12.008

Szövényi, P., Sundberg, S., Shaw, A.J., 2012. Long-distance dispersal and genetic structure of natural populations: an assessment of the inverse isolation hypothesis in peat mosses. Mol. Ecol. 21, 5461–5472. https://doi.org/10.1111/mec.12055

Taberlet, P., Fumagalli, L., Wust-Saucy, A.-G., Cosson, J.-F., 1998. Comparative phylogeography and postglacial colonization routes in Europe. Mol. Ecol. 7, 453–464. https://doi.org/10.1046/j.1365-294x.1998.00289.x

Thomas, J.A., Welch, J.J., Lanfear, R., Bromham, L., 2010. A Generation Time Effect on the Rate of Molecular Evolution in Invertebrates. Mol. Biol. Evol. 27, 1173–1180. https://doi.org/10.1093/molbev/msq009

Van der Auwera, G.A., O'Connor, B.D., 2020. Genomics in the Cloud: Using Docker, GATK, and WDL in Terra, 1st Edition. ed. O'Reilly Media.

Vigalondo, B., Draper, I., Mazimpaka, V., Calleja, J.A., Lara, F., Garilleti, R., 2020. The *Lewinskya affinis* complex (Orthotrichaceae) revisited: species description and differentiation. The Bryologist 123, 455–482. https://doi.org/10.1639/0007-2745-123.3.454

Vigalondo, B., Garilleti, R., Vanderpoorten, A., Patiño, J., Draper, I., Calleja, J.A., Mazimpaka, V., Lara, F., 2019. Do mosses really exhibit so large distribution ranges? Insights from the integrative taxonomic study of the Lewinskya affinis complex (Orthotrichaceae, Bryopsida). Mol. Phylogenet. Evol. 140, 106598. https://doi.org/10.1016/j.ympev.2019.106598

Wang, L., Zhang, W., Li, Q., 2020. AssocTests: An R Package for Genetic Association Studies. J. Stat. Softw. 94, 1–26. https://doi.org/10.18637/jss.v094.i05

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. Mol. Ecol. 20, 376–393. https://doi.org/10.1111/j.1365-294X.2010.04928.x

Westergaard, K.B., Zemp, N., Bruederle, L.P., Stenøien, H.K., Widmer, A., Fior, S., 2019. Population genomic evidence for plant glacial survival in Scandinavia. Mol. Ecol. 28, 818–832. https://doi.org/10.1111/mec.14994

Weston, D.J., Timm, C.M., Walker, A.P., Gu, L., Muchero, W., Schmutz, J., Shaw, A.J., Tuskan, G.A., Warren, J.M., Wullschleger, S.D., 2015. *Sphagnum* physiology in the context of changing climate: emergent influences of genomics, modelling and host–microbiome interactions on understanding ecosystem function. Plant Cell Environ. 38, 1737–1751. https://doi.org/10.1111/pce.12458

Yousefi, N., Hassel, K., Flatberg, K.I., Kemppainen, P., Trucchi, E., Shaw, A.J., Kyrkjeeide, M.O., Szövényi, P., Stenøien, H.K., 2017. Divergent evolution and niche

differentiation within the common peatmoss *Sphagnum magellanicum*. Am. J. Bot. 104, 1060–1072. https://doi.org/10.3732/ajb.1700163

Yousefi, N., Mikulášková, E., Stenøien, H.K., Flatberg, K.I., Košuthová, A., Hájek, M., Hassel, K., 2019. Genetic and morphological variation in the circumpolar distribution range of *Sphagnum warnstorfii*: indications of vicariant divergence in a common peatmoss. Bot. J. Linn. Soc. 189, 408–423. https://doi.org/10.1093/botlinnean/boy086

Appendix A - Supporting tables

Table A-1. Information summary for *Sphagnum compactum* individuals included in this study. Sample NJ210 is a *Sphagnum lindbergii* sample.

sampleID	Herbarium accession ID	Date sampling	Population	Country	Locality	Ecology	Altitude	Latitude	Longitude
LI109	LI819854	*2009*	*Austria*	Austria	Murauer Berge NE Tamsweg, Überlinggebiet, Überling-Sonnseite, Moor S der Überlinghutte, Ostteil	mountain pine bog, trichophoretum	1733	47.16967	13.91169
LI119	comp_2022_1	*2022*	*Austria*	Austria	Bundschuhtal, Mehrlhütte	transition mire	1720	46.98062	13.77496
LI29	TRH-B111709	01/06/17	Austria	Austria	Distr. Lungau, Sauerfelder Berg, Salzriegelmoor	Grassed fen/meadow with Nardus stricta, Andromeda polifolia and Eriophorum vaginatum	1867	47.10812	13.88392
LI39	TRH-B111710	01/06/17	Austria	Austria	Distr. Lungau, Sauerfelder Berg, Salzriegelmoor	Grassed fen/meadow with Nardus stricta, Andromeda polifolia and Eriophorum vaginatum	1867	47.10812	13.88392
LI49	TRH-B111711	*2017*	*Austria*	Austria	Distr. Lungau, Sauerfelder Berg, Salzriegelmoor		1867		
LI59	TRH-B111712	01/06/17	Austria	Austria	Distr. Lungau, Sauerfelder Berg, Salzriegelmoor	Grassed fen/meadow with Nardus stricta, Andromeda polifolia and	1867	47.10812	13.88392

						Eriophorum vaginatum			
LI69	TRH-B111713	*2017*	*Austria*	Austria	Distr. Lungau, Sauerfelder Berg, Salzriegelmoor		1867		
LI79	TRH-B111714	*2017*	*Austria*	Austria	Distr. Lungau, Sauerfelder Berg, Salzriegelmoor		1867		
LI89	LI819850	*2009*	*Austria*	Austria	Murauer Berge NE Tamsweg, Überlinggebiet, Überling-Sonnseite, Moor W des Zechenergrabens, Zentralteil	transition mire, trichophoretum	1671	47.16622	13.89681
LI9	TRH-B111708	*2017*	*Austria*	Austria	Distr. Lungau, Sauerfelder Berg, Salzriegelmoor		1867		
LI99	LI826120	*2009*	*Austria*	Austria	Murauer Berge NE Tamsweg, Überlinggebiet, Überling-Schattseite, Großes Schattseitmoor	mountain pine bog, dangling edge	1759	47.17289	13.90187
UI91	94027	01/06/17	*Austria*	Austria	Distr. Lungau, Sauerfelder Berg	Ombrotrophic, low lawn	1761	47.11799	13.8955
GI29_2	TRH-B111846	*2017*	*Germany*	Germany	Lower Saxony, Bispingen, Benninghofer Heide				

GI39	TRH-B111847	03/11/17	Germany	Germany	Lower Saxony, Bispingen, Benninghofer Heide			53.10247	9.92533
GI49_2	TRH-B111848	*2017*	*Germany*	Germany	Lower Saxony, Bispingen, Benninghofer Heide				
GI59	TRH-B111849	*2017*	*Germany*	Germany	Lower Saxony, Bispingen, Benninghofer Heide				
GI9	TRH-B111845	03/11/17	Germany	Germany	Lower Saxony, Bispingen, Benninghofer Heide			53.10247	9.92533
AI29	TRH-B111911	09/08/17	Norway_2	Norway	Draksten, mire N of Svarttjønnåsen	Pure fen lawn	354	63.31293	10.66901
AI39_2	TRH-B111912	*2017*	*Norway_2*	Norway	Draksten, mire N of Svarttjønnåsen		354		
AI49_2	TRH-B111913	*2017*	*Norway_2*	Norway	Draksten, mire N of Svarttjønnåsen		354		
AI9	TRH-B111910	09/08/17	Norway_2	Norway	Draksten, mire N of Svarttjønnåsen	Pure fen lawn	354	63.31293	10.66901
AIB59_2	TRH-B111914	09/08/17	Norway_2	Norway	Draksten, mire N of Svarttjønnåsen	Pure fen lawn	354	63.31293	10.66901
CI29	TRH-B111905	09/08/17	Norway_2	Norway	Mire S of Greistad, NW of Digresmyra	Ombrotrophic carpet	177	63.36555	10.52418
CI39	TRH-B111906	*2017*	*Norway_2*	Norway	Mire S of Greistad, NW of Digresmyra		177		

CI49	TRH-B111907	*2017*	*Norway_2*	Norway	Mire S of Greistad, NW of Digresmyra		177		
CI59	TRH-B111908	09/08/17	Norway_2	Norway	Mire S of Greistad, NW of Digresmyra	Ombrotrophic carpet	177	63.36555	10.52418
NI29	TRH-B111917	*25/10/17*	*Norway_1*	Norway	E of Langvatnet by Daltjønnbakken (Daltjønna)	Ombrotrophic disturbed hollows, transition to carpets	99	64.26518	11.45628
NI39_2	TRH-B111918	*2017*	*Norway_1*	Norway	E of Langvatnet by Daltjønnbakken (Daltjønna)		99		
NI49_2	TRH-B111919	25/10/17	Norway_1	Norway	E of Langvatnet by Daltjønnbakken (Daltjønna)	Ombrotrophic disturbed hollows, transition to carpets	99	64.26518	11.45628
NI59_2	TRH-B111920	*2017*	*Norway_1*	Norway	E of Langvatnet by Daltjønnbakken (Daltjønna)		99		
NI9	TRH-B111916	25/10/17	Norway_1	Norway	E of Langvatnet by Daltjønnbakken (Daltjønna)	Ombrotrophic disturbed hollows, transition to carpets	99	64.26518	11.45628
TI91	TRH-B113698	*2016*	*Slovakia*	Slovakia		Ombrotrophic bog by lake, depression	1618	49.2195	20.22992
TI910	TRH-B113699	*2016*	*Slovakia*	Slovakia		Ombrotrophic bog by lake, depression	1618	49.2195	20.22992
TI92	TRH-B113698	*2016*	*Slovakia*	Slovakia		Ombrotrophic bog by lake, depression	1618	49.2195	20.22992
TI93	TRH-B113698	*2016*	*Slovakia*	Slovakia		Ombrotrophic bog by lake, depression	1618	49.2195	20.22992
TI94	TRH-B113698	*2016*	*Slovakia*	Slovakia		Ombrotrophic bog by lake, depression	1618	49.2195	20.22992
TI95	TRH-B113698	*2016*	*Slovakia*	Slovakia		Ombrotrophic bog by lake, depression	1618	49.2195	20.22992

TI96	TRH-B113699	*2016*	*Slovakia*	Slovakia		Ombrotrophic bog by lake, depression	1618	49.2195	20.22992
TI97	TRH-B113699	*2016*	*Slovakia*	Slovakia		Ombrotrophic bog by lake, depression	1618	49.2195	20.22992
TI98	TRH-B113699	*2016*	*Slovakia*	Slovakia		Ombrotrophic bog by lake, depression	1618	49.2195	20.22992
TI99	TRH-B113699	*2016*	*Slovakia*	Slovakia		Ombrotrophic bog by lake, depression	1618	49.2195	20.22992
NJ210	TRH-B111989	*10/03/17*		Norway	E of Langvatnet by Daltjønnbakken (Daltjønna)	Intermediate fen carpet		64.26684	11.45651

Table A-2. Sequencing summary for all samples in the study. Samples in bold are used in PSMC analysis.

sampleID	Total number of retained reads	Total number of hits	Fraction of reads that mapped	Fraction of hits that were PCR duplicates	Mean coverage
LI109	81,743,574	26,230,018	0.32	0.16	3.21
LI119	119,531,562	19,729,666	0.17	0.19	3.99
LI29	106,045,325	10,969,341	0.10	0.17	2.96
LI39	103,366,533	30,268,436	0.29	0.07	7.18
LI49	118,281,366	20,403,581	0.17	0.17	4.05
LI59	164,245,506	12,815,549	0.08	0.09	3.23
LI69	**180,903,852**	**64,279,994**	**0.36**	**0.22**	**12.36**
LI79	223,263,462	49,254,816	0.22	0.21	9.53
LI89	66,460,534	17,461,530	0.26	0.45	2.23
LI9	95,693,604	15,642,994	0.16	0.20	3.33
LI99	85,378,725	38,970,105	0.46	0.17	5.85
UI91	110,629,017	15,351,809	0.14	0.21	2.92

GI29_2	**250,432,904**	**115,295,292**	**0.46**	**0.22**	**21.22**
GI39	19,555,078	7,115,749	0.36	0.13	1.78
GI49_2	94,309,371	23,787,627	0.25	0.27	4.30
GI59	62,691,003	19,736,467	0.31	0.15	4.01
GI9	51,039,380	10,975,805	0.22	0.18	2.88
AI29	105,785,366	19,420,117	0.18	0.16	4.14
AI39_2	182,910,698	48,577,246	0.27	0.22	8.90
AI49_2	**146,973,540**	**53,156,879**	**0.36**	**0.25**	**10.26**
AI9	36,405,206	19,361,276	0.53	0.10	5.72
AIB59_2	37,885,102	15,566,147	0.41	0.11	4.02
CI29	34,604,748	8,445,505	0.24	0.18	2.34
CI39	78,028,100	13,860,839	0.18	0.25	2.77
CI49	116,287,790	35,404,805	0.30	0.25	6.93
CI59	124,186,209	12,458,509	0.10	0.39	2.50
NI29	90,976,813	6,523,826	0.07	0.25	1.43
NI39_2	89,655,174	16,611,359	0.19	0.18	3.21
NI49_2	42,592,514	12,956,487	0.30	0.20	3.35
NI59_2	**158,212,300**	**34,863,250**	**0.22**	**0.26**	**6.63**
NI9	269,534,230	20,397,230	0.08	0.42	3.49
TI91	101,719,476	14,004,013	0.14	0.30	2.31
TI910	86,114,818	19,436,097	0.23	0.21	3.79
TI92	64,015,568	15,901,122	0.25	0.20	2.99
TI93	99,607,336	12,583,607	0.13	0.30	2.27
TI94	152,523,049	16,272,550	0.11	0.19	2.99
TI95	**209,614,333**	**33,268,723**	**0.16**	**0.27**	**6.01**

TI96	97,717,397	12,677,359	0.13	0.18	2.69
TI97	87,774,705	6,707,728	0.08	0.28	1.18
TI98	108,427,810	17,135,933	0.16	0.19	3.46
TI99	126,955,037	9,985,935	0.08	0.25	1.88
NJ210	55,896,288	28,139,159	0.50	0.12	8.15

Appendix B - Supporting figures

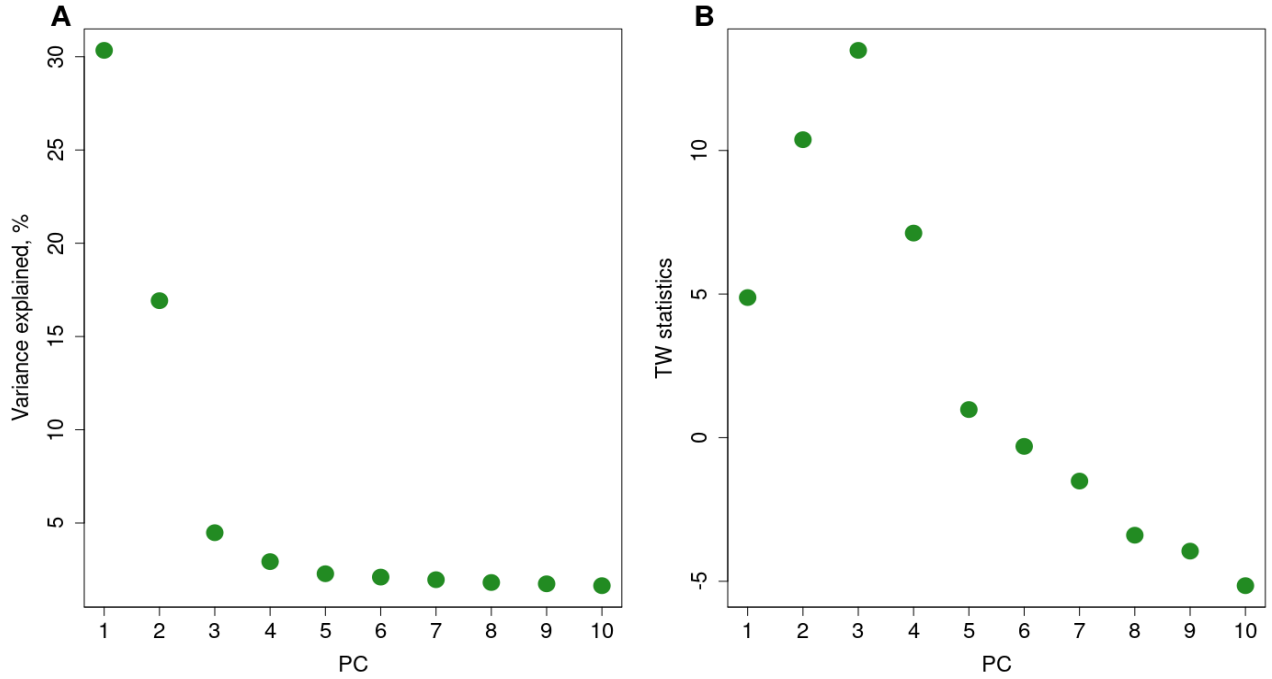

Figure B-1. Statistics for PCA. (A) Variance explained by each principal component and (B) Tracy-Wisdom statistics for each principal component.

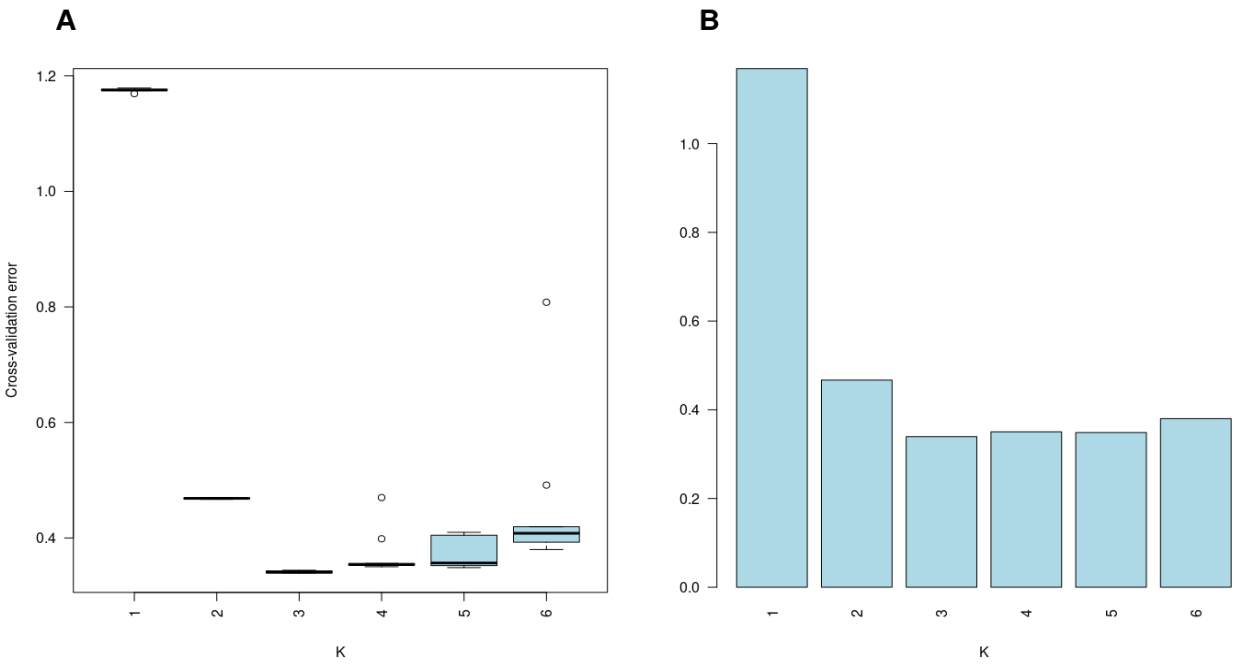

Figure B-2. ADMIXTURE results summary. (A) Boxplot of 10-fold cross-validation error (y axis) in all replicates for the corresponding number of *K* (x axis). (B) Replicates with lowest 10-fold cross-validation error.

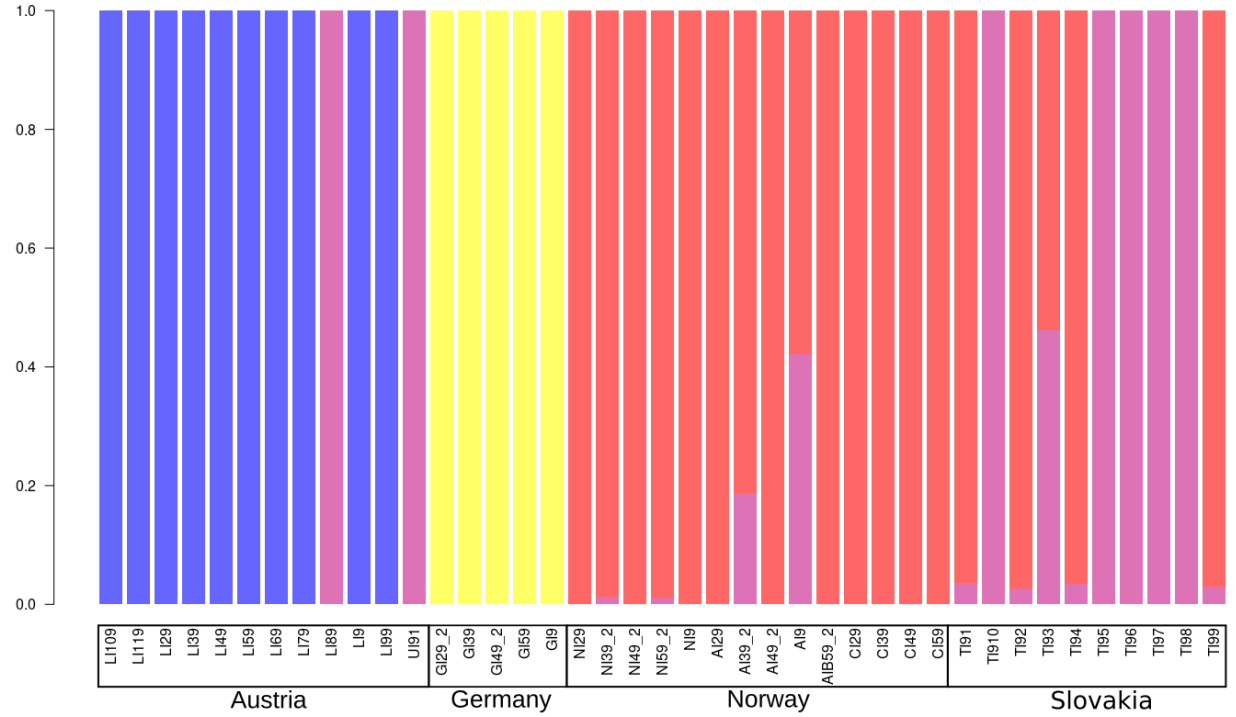

Figure B-3. Barplot showing the results of ADMIXTURE analysis with *K*=4 for 41 *Sphagnum compactum* samples.

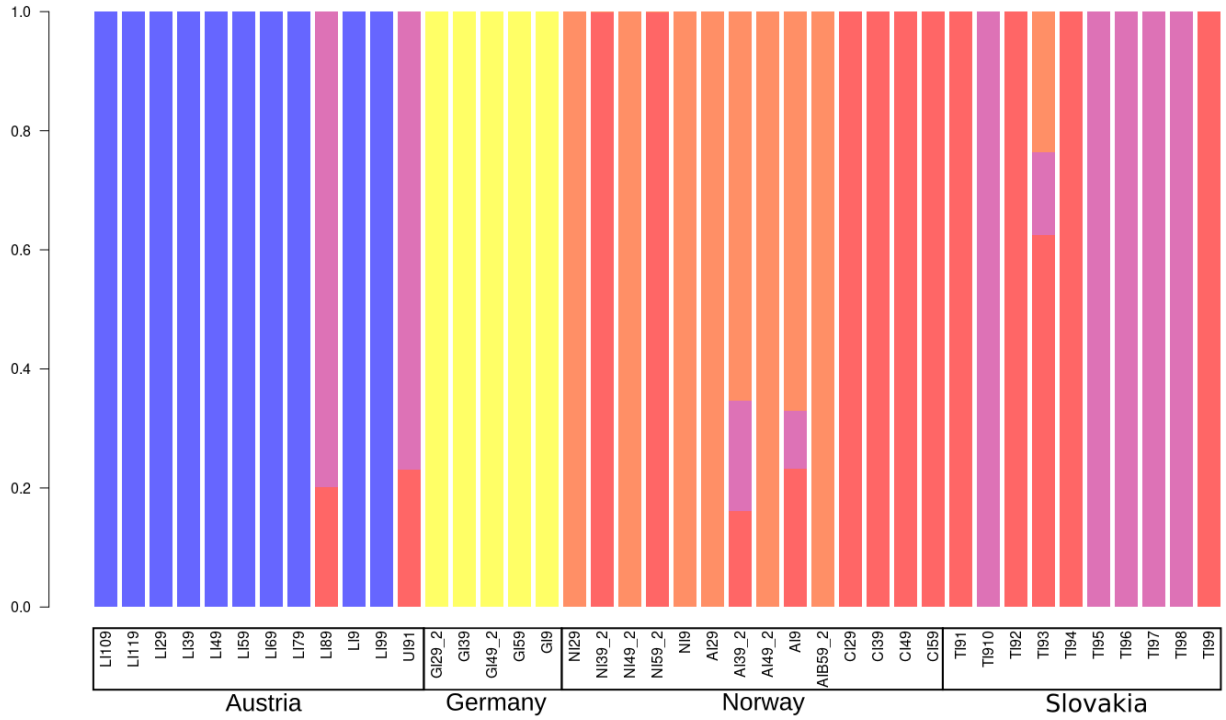

Figure B-4. Barplot showing the results of ADMIXTURE analysis with *K*=5 for 41 *Sphagnum compactum* samples.

NTNU

Norwegian University of
Science and Technology